Predictive Technologies in Healthcare: Public Perspectives and Health System Governance in the Context of Structural Inequity

by

Paige Nong

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Health Services Organization and Policy)
in the University of Michigan
2023

Doctoral Committee:

        Associate Professor Jodyn Platt
        Professor Julia Adler-Milstein
        Professor Denise Anthony
        Assistant Professor Melissa Creary
        Professor Sharon Kardia

Paige Nong

ptassie@umich.edu

ORCID iD: 0000-0002-2849-9005

Dedication

This dissertation is dedicated to my family – Neo, Ratu, and Kabo. It is also dedicated to Mamane Mpho, a uniquely remarkable woman, who was always confident that I would be "the first doctor in the family". The way she spoke about it helped me believe it.

inspiration, thought partner, and comrade. Dori Cross has been a truly tireless mentor since I was a newly graduated project manager. Her generosity is unique. Corina Simonelli has encouraged me and helped me translate my work to make policy impact. Tiffany Cornwall, Ariella Hoffman Peterson, and N'dea Moore-Petinak have been wonderful teammates and collaborators.

Most importantly, I want to acknowledge and thank my family. Neo has been my constant. He has been the persistent voice quietly telling me I could do it all along. He has been my partner in the truest sense of the word, and I will forever be deeply grateful to him.

To my daughter who genuinely cared about this project- for asking insightful questions and being constantly hungry to learn. For being genuinely considerate and patient beyond what could ever be expected of someone her age. For her truly tireless wit and sass.

To my son who could not have possibly cared less about this project- for never being deterred by the presence of my laptop but confidently pushing it aside so I could perform all the voices for each Doc McStuffins character in his book. For elbowing his way onto my lap in a fuzzy tiger costume to see who could give the most kisses when I should have been writing.

 Thank you to my parents, Peter and Colleen, who made this possible in innumerable ways, especially by tirelessly loving us and being there for us as a family. Thank you to Ma, Ntate Nong, and my entire South African family for their love and patience with this process. Thank you to my sister, Bryn, for being my best friend and constant supporter. Finally, to Mamane Mpho who made me believe I could do this.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

The health data ecosystem is increasingly focused on the design and implementation predictions in the form of AI-enabled clinical decision support, risk calculation, and resource allocation. This system of prediction in healthcare is developing rapidly in the context of limited regulation and structural inequity. The stakes for patients and health systems are high as predictive models are deployed more widely, affecting multiple aspects of care from appointment wait times to treatment for sepsis. Risks of racism, bias, and other inequities in the data used to build these models are increasingly recognized. However, public perspectives and values related to predictive modeling in healthcare have not yet been studied at the national level. It is also unclear how health systems are currently governing prediction, especially in the context of structural inequity. In this dissertation, I analyze an original national survey of the public to understand their perspectives on prediction in healthcare. I also analyze qualitative in-depth interviews with health system leadership to examine their governance strategies for predictive models. This approach treats both health system leadership and members of the public as key stakeholders engaged in and affected by the sociotechnical system of prediction.

In the first study, I analyze public comfort with data use for prediction using survey responses from a national sample of US adults. I identify that the public differentiates between the use of various data types for prediction and observe higher comfort among 1) white respondents and 2) those who have not experienced discrimination while seeking healthcare.

In the second study of a national sample of US adults, I identify misalignment between public perspectives and current regulatory frameworks. Analyzing original survey measures of

comfort with six specific predictive models in healthcare, I find that the public is less comfortable with administrative applications of prediction (e.g., predicting missed appointments) than with clinical applications (e.g., predicting stroke).

The third study presents findings from qualitative interviews with leadership from academic medical centers across the country about how they manage and design governance processes. This project focuses on understanding how predictive models are currently governed, how regulation shapes that governance, and whether equity is a consideration in health system governance processes. I identify variation among academic medical centers in their governance structures and the degree to which they consider equity when evaluating predictive models. I also find that current regulation is ambiguous for these decision-makers and could be strengthened to provide important guidance for health system policy.

As patients are increasingly exposed to predictive technologies and healthcare systems are expected to govern them, there is a critical need for empirical evidence on both stakeholders' needs, perspectives, and expectations. Policymakers, model developers, and health system leadership have roles to play in leveraging this evidence to design more responsive and equitable predictive systems in healthcare.

Chapter 1 Introduction

In 2021, 78% of surveyed healthcare systems reported the use of predictive analytics and artificial intelligence or machine learning (AI/ML) (Apathy, Holmgren, and Adler-Milstein 2021; CHIME 2021). Interest and investment in these tools are expanding every year. They are deployed in the healthcare system for a growing variety of purposes and stand to impact an increasing proportion of patients (Gupta, Frosch, and Kaplan 2021; Ngiam and Khor 2019; Wang and Preininger 2019). These models inform management, administration, and clinical care across the country (Tan et al. 2020).

In this context, a small but growing literature has identified the ways prediction and related methodologies can reflect and entrench structural racism. There is an urgent need to better understand how this occurs in order to prevent its continuation. This dissertation treats predictive technologies in healthcare as social tools that co-constitute social positions, power, and inequities (Bowker and Star 1999; Burrell and Fourcade 2021; Fourcade and Healy 2013) because they reflect and contribute to people's lived experiences and social identities. Rather than focusing on methodological details or data quality issues that currently dominate discussions of fairness in health information technology, this project focuses on prediction as a method of resource allocation that may reflect and entrench larger structural inequities. It also considers how the health system may better govern and implement predictive tools to mitigate the effects of structural racism. This project places at its core the ethical appreciation of the consequences of prediction for patients and explicitly considers how racism and inequity may be reflected in predictive systems (Bowker and Star 1999).

Predictive models

Predictive models are often promoted as evidence-based, data-driven decision support for healthcare providers and health systems (Chen and Asch 2017; Desai 2020; Ghassemi, Oakden-Rayner, and Beam 2021; Shilo, Rossman, and Segal 2020). These tools rely on increasingly large amounts of data, where greater precision is a consistent goal. While the tools themselves are applied to more data sources to manage patients with increasing precision, very little data on patient perspectives has been analyzed (Richardson et al. 2021). Attempts to compute, predict, and manage patients ("understand" them in one sense) are driving significant investment and effort in healthcare, but deeper understanding of the system of prediction and its social implications is limited. Patient perspectives and the role of health systems in managing predictive models have not been extensively analyzed. This is especially true in relation to inequity and structural racism.

The fields of sociology and science and technology studies (STS) offer critical analyses of predictive technologies and algorithmic risk classification. Much of this literature considers how racism may be reflected or further entrenched by predictive tools, building on concepts of structural racism and classification (Campolo and Crawford 2020; Hanna et al. 2020; McMillan Cottom 2020; Sewell, Jefferson, and Lee 2016). Some work engages the concept of prediction or algorithmic classification while other literatures focus on specific types of prediction and their implications for racial inequity, like credit scoring or predictive policing (Burrell and Fourcade 2021; Fourcade and Healy 2013; Richardson, Schultz, and Crawford 2019). However, we lack a similarly critical and in-depth literature on prediction in healthcare. Our understanding of the relationships between structural racism, inequity, and the system of prediction in healthcare is limited. Discussions of how racism and SES inequities are reflected in health-related predictions

are typically concentrated in editorials or commentaries that focus on methodological specifics of AI/ML (McCradden et al. 2020; Rajkomar et al. 2018). Some empirical work has identified racist outcomes from specific predictive models (Benjamin 2019; Obermeyer et al. 2019; Vyas, Eisenstein, and Jones 2020), but the connection between healthcare and sociological or STS insights remains limited, with patient perspectives severely under-analyzed (Richardson et al. 2021; Zhang et al. 2021).

Terminology, distinctions, and scope

Predictive models are described in a multitude of ways in the literature and healthcare practice. These include, but are not limited to; risk scores, predictions, eligibility criteria, algorithms, and AI/ML (Desai 2020; Obermeyer and Emanuel 2016; Wynants et al. 2019). These terms are often used interchangeably, despite meaningful distinctions between them, including the adaptability of an algorithm, its complexity, and the degree to which these characteristics of an algorithm are opaque (Burrell 2016; Ngiam and Khor 2019). Fundamentally, predictive models draw on historical data to identify patterns and use those patterns to make a prediction (or provide a risk score) related to some outcome like a patients' health, behaviors, needs, or risks (Waljee, Higgins, and Singal 2014).

This project generally eschews the AI and ML labels, although those methodologies are certainly used to construct predictive models and risk scores. This project is concerned with; 1) predictive classification systems and how they relate to public perceptions and structural racism, and 2) how health systems govern these technologies in the context of structural inequity. These are the core foci of the project, regardless of the computational methodology used to generate the predictions discussed. Thus, rather than focusing on the potential differences or connections

between AI and ML, I use predictive model, risk score, and classification most often throughout this dissertation.

Although this project is agnostic with regard to predictive methodologies, the use of AI/ML methods may be relevant to the ways predictions or risk scores interact with structural racism. There are myriad issues with these methods and their relationships to inequity in predictive analytics. These include model opacity, or the degree to which a given model is a "black box", even to its developers or implementers. For example, opaque unsupervised machine learning may produce output that is semantically uninterpretable for human beings and embed inequity into predictions (Burrell 2016; Gupta et al. 2021; Wang, Kaushal, and Khullar 2020). Specific population validation approaches necessary for AI/ML models are also important for the quality of predictions according to patients' social positions (Barda et al. 2021). However, these issues are out of scope for this specific project and will be addressed in future work.

Conceptual framework

Social conditions, like socioeconomic status and racism, are fundamental causes of health that operate through multiple replaceable mechanisms over time (Link and Phelan 1995; Phelan and Link 2015). The fundamental cause of racism, as described by Link and Phelan, operates through replaceable mechanisms by providing differential access to "flexible resources" like power, freedom, knowledge, financial resources, and social capital that impact more proximal causes of health. For example, racism structures peoples' living environments. Though the specific laws, policies, and drivers of such segregation have changed over time, they continue to result in disproportionate exposure to harmful physical environments for people who belong to marginalized racial and ethnic groups (Riley 2018; Williams and Collins 2001). Link and Phelan

point out that rather than focusing on the proximate cause of health (environmental exposures), deeper understanding of the cause of inequity focuses on the more distal cause of racism (Phelan and Link 2015). This is often framed as identifying what puts people "at risk of risk" (Link and Phelan 1995).

Not only does racism external to the healthcare system structure people's chances for living healthy lives as described above, but the healthcare system itself reflects these inequalities and perpetuates them. This has been observed in racially inequitable access to high quality care, racial discrimination in treatment and prescribing practices, inequity in health insurance policies, and interpersonal racism perpetrated against patients when they seek care (LaVeist, Rolley, and Diala 2003; Lutfey and Freese 2005; van Ryn et al. 2011; Spencer and Grace 2016; White, Haas, and Williams 2012). These healthcare-related mechanisms are replaceable mechanisms in the Link and Phelan model that are proximal to racial differences in health outcomes. They operate in the context of differences in flexible resources caused by systemic racism.

The conceptual framework for this project combines insights from Link and Phelan's fundamental cause framework with Bowker and Star's conceptualization of classification as a mechanism of inequality. Drawing on the fundamental cause approach explained above, racism and socioeconomic inequality are fundamental causes of inequalities in health. They operate through multiple replaceable mechanisms to differentially distribute the risk of disease. Healthcare itself can function as one of these mechanisms through which race and health are related (Phelan and Link 2015). Within healthcare, systems of classification operate as specific methods of resource allocation (Bowker and Star 1999) as depicted in Figure 1.1. Here, in the context of racism and socioeconomic inequality, classification and prediction are operating as a mechanism of racism in healthcare through their impact on healthcare resources (Bowker and

Star 1999; Link and Phelan 1995). Put another way, prediction sorts people into categories. Because these categories are connected to resources like care management services or medications, they can function as a mechanism of the fundamental causes of racism and socioeconomic inequality.

Figure 1.1 Conceptual framework for prediction and inequality
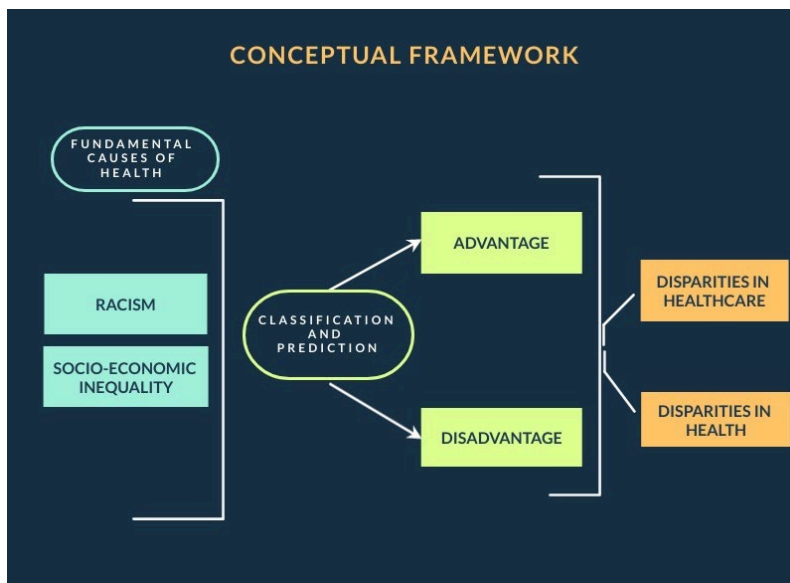
Building on this conceptual approach, this project is concerned with how racism may be codified in customs, practices, and policies around prediction in healthcare. Following the public health critical race praxis, this work engages with racism as ordinary and structural rather than an individual attitude or aberration (Ford and Airhihenbuwa 2010). In this framing institutional customs, practices, and policies shape the differential access to resources observed between racial and ethnic groups (Alang et al. 2021; Jones 2000). As Bowker and Star explain, classification systems operating in this context are often used to distribute resources (Bowker and Star 1999) and have serious consequences for inequity.

Because structural racism is observed in differential access to resources like healthcare (Alang et al. 2021; Elias and Feagin 2016; Hardeman et al. 2018; Jones 2000), methods of resource allocation are particularly important fields of interest in identifying and dismantling structural racism. To understand how structural racism may be operating in predictive systems, it is important to analyze the customs, practices, and policies around prediction in healthcare to better understand; 1) whose values are reflected in the system's functions and policies and 2) the implications of structural inequity in health information technology. This dissertation considers how predictive technologies reflect or contradict patient values and needs across racial differences. It also analyzes how health system decision-making informs predictive systems.

Prediction in healthcare

Predictive models use large amounts of patient data to make predictions about a variety of patient diagnoses, outcomes, and behaviors. They are part of an effort in the US healthcare system to increase the precision with which patients are identified, managed, and treated. Although the evidence of predictive effectiveness is highly variable (Chen and Asch 2017; Christodoulou et al. 2019), the logic of increasing data collection for precision healthcare continues to drive investment and predictive model implementation across the health system (Ginsburg and Phillips 2018).

Not only are predictive tools promoted as highly innovative and cost-effective, but health systems are growing their investments in these types of tools and applying them to both clinical and administrative operations (Ding et al. 2018; S. Murray, Wachter, and Cucina 2020; Siwicki 2021). Clinical prediction models are used to anticipate the onset of sepsis, COVID-19 deterioration, cardiac events, and kidney disease progression, for example (Nemati et al. 2018;

Niederer, Lumens, and Trayanova 2019; Singh et al. 2020; Wong et al. 2021). These models may be available to providers through the electronic health record (EHR) as clinical decision support (CDS) for patient care. Increasingly, predictive tools draw on genetic information as part of precision medicine efforts, wherein multiple data types are analyzed to predict patient outcomes with increasing precision (Ginsburg and Phillips 2018; Kaplanis et al. 2020).

In addition to clinical applications, predictions can be used to allocate health system resources. Figure 1.2 includes examples of predictive models that fall into two categories: administrative and clinical. In the case of COVID-19 deterioration referenced in the figure, models were constructed to predict the risk of deterioration in order to potentially triage patients to ICUs or remote care locations (Singh et al. 2016). Models are also used to predict health service utilization, readmissions, and patient appointment no-shows, for example (Futoma, Morris, and Lucas 2015; S. Murray et al. 2020). Administrative or operational applications are often intended to lower costs and target resource utilization. Prediction of missed appointments, for example, is motivated by a desire to decrease lost revenue associated with unused appointment slots (Ding et al. 2018; S. Murray et al. 2020). In this case, if a patient's risk of missing an appointment is calculated as high, the health system may double-book that patient's appointment.

Figure 1.2 Types of predictive model applications in healthcare
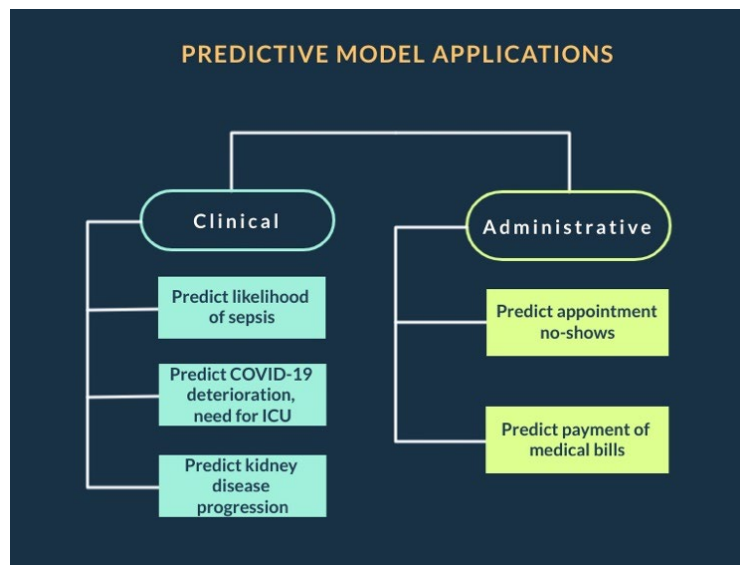
Classification and inequity in healthcare

In healthcare, predictive models predict health status, inform resource management, and determine access to care (Escobar et al. 2016; S. Murray et al. 2020; Obermeyer et al. 2019). In the context of persistent health inequities driven by racism and SES inequality (Link and Phelan 1995; Phelan and Link 2015) critical analysis of tools that serve these functions is necessary. Acknowledging that classification does not simply reproduce inequalities, but co-constructs and generates them (Fourcade and Healy 2013), predictive models may not only be reflecting health inequity but actively creating it as well.

A wide variety of negative implications or risks related to these models have been identified in the literature (Barda et al. 2021; Holmberg and Vickers 2013; M. D. McCradden et al. 2020; Rajkomar et al. 2018; Veinot, Mitchell, and Ancker 2018). Although only some of these risks have been empirically analyzed in specific models (Obermeyer et al. 2019, 2021), the literature provides a long list of potential ways predictive tools can be biased or perpetuate inequality (Table 1.1). The model itself can perpetuate inequality through the way it functions

(e.g., the model does not generalize to all populations). The data used to build a model can also be a source if inequality, for example, when data is systematically missing for certain groups of people. Finally, the structural inequalities that affect health and healthcare can also lead to inequality when a predictive model is deployed through inequitable access to care or disparate treatment of patients.

Table 1.1 Examples of inequality in predictive models

Sources of Inequality	
Model	Labeling: a variable is chosen as a proxy for something that it does not accurately measure for all patients or groups (Obermeyer et al. 2019)
	Lack of generalizability: a model is built without verification that its accuracy holds for different patient populations (Seyyed-Kalantari et al. 2021)
	Purpose: models applied for the purpose of efficiencies or restricting access to treatments can negatively impact marginalized patients (S. Murray et al. 2020)
Data	Missingness: data is systematically missing or missingness is not random across patient populations (Barda et al. 2021)
	Entry or coding decisions: bias in how patient data is captured/coded that reflect social inequity or discrimination (Beach et al. 2021)
	Sample size: patients are underrepresented in the training data sets used to build models (Gianfrancesco et al. 2018)
Structural inequality	Privilege bias (Rajkomar et al. 2018): Tools may be unavailable to the providers of disadvantaged patient populations
	Undertreatment: white patients receive higher cost, more aggressive treatment which can further bias models against interventions for racial and ethnic minority patients (Bonham 2001)

Predictive models can thus perpetuate bias, construct new forms of inequity, and obfuscate decision-making processes in healthcare. In the following section I describe each of these three ways predictive models can contribute to inequity.

Perpetuate existing biases

The empirical literature on predictive biases and risks is relatively nascent, but early examples indicate that existing inequities can be reified, and new ones created, by predictive models in healthcare. In one of the most well-known empirical examples, analysis of an insurance algorithm used to allocate care management resources demonstrated a clear racial bias against Black patients (Obermeyer et al. 2019). The algorithm used health expenditures as a proxy measure of clinical need to predict need for support services. Because Black patients receive less aggressive, less expensive medical treatments even when their clinical need is equal to that of white patients, the algorithm reflected this longstanding racial bias in the healthcare system (Benjamin 2019). At any given risk score, Black patients had significantly higher clinical need than white patients with the same score. In this way, the algorithm was predicting health expenditures based on historical data structured by systemic racism and perpetuating that bias in access to care management.

Construct new forms or mechanisms of inequity

Other tools generate new mechanisms of inequity. Arkansas' Medicaid program, for example, deployed an algorithm to predict the need for in-home care (Lecher 2018). While the process had previously relied on clinically trained staff to calculate appropriate care hours, a predictive model was deployed to incorporate different data types and allocate resources. Certain information that previously informed in-home care allocation was missing from this model, and previously excluded or unavailable data types were included. This facilitated new types of bias not previously observed in the previous system of distributing Arkansas Medicaid's healthcare resources. For example, cerebral palsy diagnoses and other diagnoses or conditions were not

included in the models, resulting in the exclusion of whole patient populations from access to the care they needed (Lecher 2018). This introduced explicit discrimination against people with disabilities and certain chronic conditions that did not exist prior to the deployment of the model.

Facilitate opaque decision-making

Not only can predictive models drive inequity through classification, but they are also opaque to the people whose lives and care are affected by them. One vendor built a predictive model for a large health system in Minnesota to predict patients' propensity to pay their medical bills, for example. This model was deployed to stratify patients based on the probability that they would pay their bills, then determine whether and how the system would contact patients before referring them to a collections agency (HealthCatalyst 2019). Similarly, the leading EHR vendor in the country provides a model that is purported to predict the likelihood that a patient will miss a scheduled appointment. These predictions can then be used to double-book appointment slots (S. Murray et al. 2020). Patients are not aware that these models are being applied to determine whether they receive notice of outstanding medical bills or whether their appointment is double booked. Not only does this opacity prevent understanding on the patient's behalf, but it also functionally prevents any contestation by the patient. This is especially true for the most marginalized because they are not well positioned to contest the tools that are being used to determine their experiences (Burrell and Fourcade 2021). Bureaucratic opacity functions as a way of "making decisions without seeming to decide" (Porter 2020), obscuring who is responsible. This makes contestation nearly impossible except for those with the most advantage (Pasquale 2015).

Identifying ways to maximize benefits and minimize harms of predictive models

In order to mitigate unequal consequences of predictive models or classifications, discussions in healthcare literatures often focus on data bias, labeling decisions, and model fairness (Barda et al. 2021; Obermeyer and Topol 2021; Rajkomar et al. 2018; Vyas et al. 2020). This kind of work is increasingly common and provides insight on some of the potential issues with predictive modeling in healthcare. However, the healthcare system continues to build and implement these tools, making deeper questions and extensive analysis of predictive systems critically important. Recognizing that the consequences of predictive technologies are not pre-determined but depend on the uses to which they are applied (S. Murray et al. 2020; Pierson et al. 2021), critical analysis of the decision-making process about predictive model use in healthcare is necessary.

As Bowker and Star describe, "one may get ever more precise knowledge without having resolved deeper questions, and indeed, by burying those questions" (Bowker and Star 1999). Resolution of those deeper questions about the effects of prediction on people's lives and an inequitable healthcare system are pressing. Although predictive precision may increase, a deeper understanding of how prediction as a system reflects structural racism, and the role of health systems as decision-makers, remains underexplored.

Power and patient perspectives

The stakes of predictive approaches to healthcare are highest for patients. Whether classification shapes a patient's diagnosis or leads to delayed care through overbooking their appointment, the patient's experience of the healthcare system can be directly affected by predictive technologies. As the sociological literature describes, the life circumstances and

histories of the subjects of classification are circumscribed and defined by their position within the classification system (Burrell and Fourcade 2021). Classification systems "torque" people's lives (Bowker and Star 1999), meaning that the act of sorting people into categories has significant material effects on their lives. Receiving a diagnosis, for example, can legitimize a patient's symptoms and open possibilities of treatment. Lack of diagnosis can be delegitimizing and confusing (Timmermans and Buchbinder 2010). On the other hand, when some patients are classified as having a certain illness, their rights and autonomy can be threatened (Conrad and Schneider 2008). When patients are classified as high risk for a negative health outcome, they can receive additional resources or support. Alternatively, if a model predicts that they are likely to miss an appointment, they may receive lower quality care. Predictions of 30-day readmissions result in patients either receiving additional transition support or not (Leppin et al. 2014). Risk scores for overdose grant or deny access to certain treatments for patients who use substances (Apriss 2022). Because classification systems are used to grant or deny resources to individuals and groups of patients, patients are the stakeholders with the most to lose from the negative consequences of predictive modeling.

Research and policy have long emphasized the importance of patient engagement and participation in healthcare (Berwick, Nolan, and Whittington 2008). This participatory role has evolved as patients collect, manage, and report their own data to healthcare providers in different ways. Patients are increasingly expected to participate, partner, and contribute data for their care. They use medical devices like glucometers and blood pressure cuffs to collect their own data and report it to their provider (Adler-Milstein and Nong 2019). They use apps to monitor their nutrition and exercise (Gordon et al. 2020). They use smart scales and a variety of other tools to actively manage their own health and participate in their care (Walker et al. 2019). This kind of

patient role, whereby patients use digital tools to increasingly take on responsibility for their health has been termed the "digital health citizen" (Ziebland, Hyde, and Powell 2021). Not only are patients expected to actively engage, but they have a crucial role in collecting and reporting their data to their provider or health system as digital health citizens. As such, patient values, perspectives, and responses to predictive technologies are necessary to inform policy and use of these tools.

Summary of approach

In this dissertation, I analyze public perceptions of predictive models in healthcare as well as health system approaches to governing and implementing these models. I do this using national survey data and qualitative interviews with health system informatics and analytics leaders. Specifically, I ask:

1. Does public comfort co-vary between different types of data used for prediction? What are the underlying, co-varying dimensions (multivariate factors) of public comfort with use of various data types for prediction in healthcare?

2. What individual-level variables (e.g., racial/ethnic identity, experiences of discrimination, beliefs) predict the multivariate factors of comfort with data use for predictive modeling in healthcare?

3. Does public comfort with predictive models reflect the regulatory emphasis on clinical software applications? Does public comfort differ between clinical and administrative applications?

4. What are the individual-level predictors of comfort with specific predictive models?

5. How do academic medical centers govern predictive models? What are their key priorities and current governance processes?

6. How does federal regulation affect the governance practices of academic medical centers?

In **Chapter 2**, I examine variation in the public's comfort with the use of different data types to build predictive models. Analyzing original national survey data, I present findings on public attitudes about specific data types being used to drive prediction in healthcare. This chapter focuses on racial/ethnic differences in comfort with the current customs and practices around prediction in healthcare. It also examines the relationship between comfort with these practices and experiences of discrimination in the healthcare system. This is the first national survey of US adults to analyze perceptions of data use for prediction in healthcare.

Chapter 3 focuses on public perceptions of specific predictive models in healthcare. Reflecting important distinctions in the current policy landscape, these models include both clinical and administrative applications. By comparing comfort with these two types of models, this study identifies misalignment between public comfort and current policy.

In **Chapter 4**, I present the results of in-depth interviews with health information technology decision-makers from across the country to understand how they conceptualize and make decisions about predictive technologies. This analysis provides understanding of how academic medical centers currently govern predictive models. It includes key insights on how equity and regulatory frameworks are incorporated in academic medical centers' governance processes.

In the final chapter, I summarize the findings of the dissertation and reflect on the implications for healthcare delivery and inequity in healthcare. I consider responsive policy approaches and include important topics for future research that will expand on the findings presented here.

CONCLUSION

Predictive models in healthcare carry potential risk and benefit. The goal of equitably balancing these risks and benefits is complex and requires empirical evidence to drive appropriate policy and healthcare delivery. Building on insights about how predictive technologies can produce racially inequitable outcomes (Benjamin 2019; Burrell and Fourcade 2021; Campolo and Crawford 2020; G. F. Murray, Rodriguez, and Lewis 2020), this dissertation seeks to identify how current practices (e.g. data use, governance) in predictive systems may also perpetuate structural racism. It contributes important empirical evidence of both public perspectives and health system governance using the "technology-in-practice" concept described by Orlikowski (Orlikowski 2000). The findings presented here can inform best practices for healthcare systems as they evaluate and govern predictive technologies that are increasingly made available to them. This work also provides important insights about public perspectives that respond to new and emerging policy efforts to engage with public and patient concerns.

References

Adler-Milstein, Julia, and Paige Nong. 2019. "Early Experiences with Patient Generated Health Data: Health System and Patient Perspectives." *Journal of the American Medical Informatics Association* 26(10):952–59. doi: 10.1093/jamia/ocz045.

Alang, Sirry, Rachel Hardeman, J'Mag Karbeah, Odichinma Akosionu, Cydney McGuire, Hamdi Abdi, and Donna McAlpine. 2021. "White Supremacy and the Core Functions of Public Health." *American Journal of Public Health* 111(5):815–19. doi: 10.2105/AJPH.2020.306137.

Apathy, Nate C., A. Jay Holmgren, and Julia Adler-Milstein. 2021. "A Decade Post-HITECH: Critical Access Hospitals Have Electronic Health Records but Struggle to Keep up with Other Advanced Functions." *Journal of the American Medical Informatics Association* ocab102. doi: 10.1093/jamia/ocab102.

Apriss. 2022. *NarxCare: Analytics, Tools and Technology to Help Care Teams Address Substance Use Disorder and Improve Patient Outcomes*. State of Michigan Department of Licensing and Regulatory Affairs.

Bai, P., Y. Zhou, Y. Liu, G. Li, Z. Li, T. Wang, and X. Guo. 2020. "Risk Factors of Cerebral Infarction and Myocardial Infarction after Carotid Endarterectomy Analyzed by Machine Learning." *Computational and Mathematical Methods in Medicine* 2020. doi: 10.1155/2020/6217392.

Barda, Noam, Gal Yona, Guy N. Rothblum, Philip Greenland, Morton Leibowitz, Ran Balicer, Eitan Bachmat, and Noa Dagan. 2021. "Addressing Bias in Prediction Models by Improving Subpopulation Calibration." *Journal of the American Medical Informatics Association*. doi: 10.1093/jamia/ocaa283.

Beach, Mary Catherine, Somnath Saha, Jenny Park, Janiece Taylor, Paul Drew, Eve Plank, Lisa A. Cooper, and Brant Chee. 2021. "Testimonial Injustice: Linguistic Bias in the Medical Records of Black Patients and Women." *Journal of General Internal Medicine*. doi: 10.1007/s11606-021-06682-z.

Benjamin, Ruha. 2019. "Assessing Risk, Automating Racism." *Science* 366(6464):421–22. doi: 10.1126/science.aaz3873.

Berwick, Donald M., Thomas W. Nolan, and John Whittington. 2008. "The Triple Aim: Care, Health, And Cost." *Health Affairs* 27(3):759–69. doi: 10.1377/hlthaff.27.3.759.

Bonham, Vence L. 2001. "Race, Ethnicity, and Pain Treatment: Striving to Understand the Causes and Solutions to the Disparities in Pain Treatment." *The Journal of Law, Medicine & Ethics* 28(s4):52–68. doi: 10.1111/j.1748-720X.2001.tb00039.x.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things out: Classification and Its Consequences*. Cambridge, Mass: MIT Press.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15(5):662–79. doi: 10.1080/1369118X.2012.678878.

Bracic, Ana, Shawneequa Callier, and W. Nicholson Price. 2022. "Exclusion Cycles: Reinforcing Disparities in Medicine." *Science* 377(6611).

Brayne, Sarah. 2021. *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York, NY: Oxford University Press.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3(1):2053951715622512. doi: 10.1177/2053951715622512.

Burrell, Jenna, and Marion Fourcade. 2021. "The Society of Algorithms." *Annual Review of Sociology* 47(1):annurev-soc-090820-020800. doi: 10.1146/annurev-soc-090820-020800.

Campolo, Alexander, and Kate Crawford. 2020. "Enchanted Determinism: Power without Responsibility in Artificial Intelligence." *Engaging Science, Technology, and Society* 6(0):1–19. doi: 10.17351/ests2020.277.

Campos-Castillo, Celeste, Benjamin Woodson, Elizabeth Theiss-Morse, Tina Sacks, Michelle Fleig-Palmer, and Monica Peek. 2016. "Examining the Relationship Between Interpersonal and Institutional Trust in Political and Health Care Contexts." Pp. 99–115 in.

Casper, Monica J., and Daniel R. Morrison. 2010. "Medical Sociology and Technology: Critical Engagements." *Journal of Health and Social Behavior* 51(1_suppl):S120–32. doi: 10.1177/0022146510383493.

Chen, Jonathan H., and Steven M. Asch. 2017. "Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations." *The New England Journal of Medicine* 376(26):2507–9. doi: 10.1056/NEJMp1702071.

CHIME. 2021. *Healthcare's Most Wired: National Trends 2021*. Survey.

Christodoulou, Evangelia, Jie Ma, Gary S. Collins, Ewout W. Steyerberg, Jan Y. Verbakel, and Ben Van Calster. 2019. "A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models." *Journal of Clinical Epidemiology* 110:12–22. doi: 10.1016/j.jclinepi.2019.02.004.

Conrad, Peter, and Joseph Schneider. 2008. "Deviance and Medicalization." in *Sociology: Exploring the Architecture of Everyday Life*. Vol. 1980. Pine Forge Press.

Conrad, Peter, and Joseph W. Schneider. 1992. *Deviance and Medicalization: From Badness to Sickness*. Temple University Press.

Desai, Angel. 2020. "Artificial Intelligence: Promise, Pitfalls, and Perspective." *JAMA Network Open* 323(24).

Ding, Xiruo, Ziad F. Gellad, Chad Mather, Pamela Barth, Eric G. Poon, Mark Newman, and Benjamin A. Goldstein. 2018. "Designing Risk Prediction Models for Ambulatory No-Shows across Different Specialties and Clinics." *Journal of the American Medical Informatics Association* 25(8):924–30. doi: 10.1093/jamia/ocy002.

Elias, Sean, and Joe R. Feagin. 2016. *Racial Theories in Social Science: A Systemic Racism Critique*. New York, NY: Routledge.

Escobar, Gabriel J., Benjamin J. Turk, Arona Ragins, Jason Ha, Brian Hoberman, Steven M. LeVine, Manuel A. Ballesca, Vincent Liu, and Patricia Kipnis. 2016. "Piloting Electronic Medical Record–Based Early Detection of Inpatient Deterioration in Community Hospitals." *Journal of Hospital Medicine* 11(S1):S18–24. doi: https://doi.org/10.1002/jhm.2652.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. St. Martin's Publishing Group.

Feagin, Joe, and Zinobia Bennefield. 2014. "Systemic Racism and U.S. Health Care." *Social Science & Medicine* 103:7–14. doi: 10.1016/j.socscimed.2013.09.006.

Ferguson, Andrew Guthrie. 2016. "Policing Predictive Policing." *Washington University Law Review* 94:1109.

Ford, Chandra L., and Collins O. Airhihenbuwa. 2010. "The Public Health Critical Race Methodology: Praxis for Antiracism Research." *Social Science & Medicine* 71(8):1390–98. doi: 10.1016/j.socscimed.2010.07.030.

Fourcade, Marion, and Kieran Healy. 2013. "Classification Situations: Life-Chances in the Neoliberal Era." *Accounting, Organizations and Society* 38(8):559–72. doi: 10.1016/j.aos.2013.11.002.

Fourcade, Marion, and Kieran Healy. 2016. "Seeing like a Market." *Socio-Economic Review* mww033. doi: 10.1093/ser/mww033.

Futoma, Joseph, Jonathan Morris, and Joseph Lucas. 2015. "A Comparison of Models for Predicting Early Hospital Readmissions." *Journal of Biomedical Informatics* 56:229–38. doi: 10.1016/j.jbi.2015.05.016.

Ghassemi, Marzyeh, Luke Oakden-Rayner, and Andrew L. Beam. 2021. "The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care." *The Lancet Digital Health* 3(11):e745–50. doi: 10.1016/S2589-7500(21)00208-9.

Gianfrancesco, Milena A., Suzanne Tamang, Jinoos Yazdany, and Gabriela Schmajuk. 2018. "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data." *JAMA Internal Medicine* 178(11):1544–47. doi: 10.1001/jamainternmed.2018.3763.

Gilson, Lucy. 2003. "Trust and the Development of Health Care as a Social Institution." *Social Science & Medicine* 56(7):1453–68. doi: 10.1016/S0277-9536(02)00142-9.

Ginsburg, Geoffrey S., and Kathryn A. Phillips. 2018. "Precision Medicine: From Science to Value." *Health Affairs* 37(5):694–701. doi: 10.1377/hlthaff.2017.1624.

Gordon, William J., Adam Landman, Haipeng Zhang, and David W. Bates. 2020. "Beyond Validation: Getting Health Apps into Clinical Practice." *Npj Digital Medicine* 3(1):1–6. doi: 10.1038/s41746-019-0212-z.

Gupta, Kush, Dominick Frosch, and Robert Kaplan. 2021. "Opening The Black Box Of Digital Health Care: Making Sense Of 'Evidence.'" *Health Affairs Blog*.

Hall, Mark A., Elizabeth Dugan, Beiyao Zheng, and Aneil K. Mishra. 2001. "Trust in Physicians and Medical Institutions: What Is It, Can It Be Measured, and Does It Matter?" *The Milbank Quarterly* 79(4):613–39. doi: 10.1111/1468-0009.00223.

Hanna, Alex, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. "Towards a Critical Race Methodology in Algorithmic Fairness." Pp. 501–12 in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, *FAT* '20*. New York, NY, USA: Association for Computing Machinery.

Hardeman, Rachel R., Katy A. Murphy, J'Mag Karbeah, and Katy Backes Kozhimannil. 2018. "Naming Institutionalized Racism in the Public Health Literature: A Systematic Literature Review." *Public Health Reports* 133(3):240–49. doi: 10.1177/0033354918760574.

HealthCatalyst. 2019. "Effectively Predicting Propensity to Pay with Machine Learning." *Health Catalyst*. Retrieved March 25, 2021 (https://www.healthcatalyst.com/success_stories/predicting-propensity-to-pay-allina-health).

Heaven, Will Douglas. 2020. "Predictive Policing Algorithms Are Racist. They Need to Be Dismantled." *MIT Technology Review*, July 17.

Hoffmann, Anna Lauren. 2019. "Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse." *Information, Communication & Society* 22(7):900–915. doi: 10.1080/1369118X.2019.1573912.

Holmberg, Lars, and Andrew Vickers. 2013. "Evaluation of Prediction Models for Decision-Making: Beyond Calibration and Discrimination." *PLoS Medicine* 10(7):e1001491. doi: 10.1371/journal.pmed.1001491.

Jones, Camara Phyllis. 2000. "Levels of Racism: A Theoretic Framework and a Gardener's Tale." *American Journal of Public Health* 90(8):1212–15.

Kaplanis, Joanna, Kaitlin E. Samocha, Laurens Wiel, Zhancheng Zhang, Kevin J. Arvai, Ruth Y. Eberhardt, Giuseppe Gallone, Stefan H. Lelieveld, Hilary C. Martin, Jeremy F. McRae, Patrick J. Short, Rebecca I. Torene, Elke de Boer, Petr Danecek, Eugene J. Gardner, Ni Huang, Jenny Lord, Iñigo Martincorena, Rolph Pfundt, Margot R. F. Reijnders, Alison Yeung, Helger G. Yntema, Lisenka E. L. M. Vissers, Jane Juusola, Caroline F. Wright, Han G. Brunner, Helen V. Firth, David R. FitzPatrick, Jeffrey C. Barrett, Matthew E. Hurles, Christian Gilissen, and Kyle Retterer. 2020. "Evidence for 28 Genetic Disorders Discovered by Combining Healthcare and Research Data." *Nature* 586(7831):757–62. doi: 10.1038/s41586-020-2832-5.

LaVeist, Thomas A., Lydia A. Isaac, and Karen Patricia Williams. 2009. "Mistrust of Health Care Organizations Is Associated with Underutilization of Health Services." *Health Services Research* 44(6):2093–2105. doi: 10.1111/j.1475-6773.2009.01017.x.

LaVeist, Thomas A., Nicole C. Rolley, and Chamberlain Diala. 2003. "Prevalence and Patterns of Discrimination among U.S. Health Care Consumers." *International Journal of Health Services* 33(2):331–44. doi: 10.2190/TCAC-P90F-ATM5-B5U0.

Lecher, Colin. 2018. "What Happens When an Algorithm Cuts Your Healthcare." *The Verge*, March 21.

Leppin, Aaron L., Michael R. Gionfriddo, Maya Kessler, Juan Pablo Brito, Frances S. Mair, Katie Gallacher, Zhen Wang, Patricia J. Erwin, Tanya Sylvester, Kasey Boehmer, Henry H. Ting, M. Hassan Murad, Nathan D. Shippee, and Victor M. Montori. 2014. "Preventing 30-Day Hospital Readmissions: A Systematic Review and Meta-Analysis of Randomized Trials." *JAMA Internal Medicine* 174(7):1095–1107. doi: 10.1001/jamainternmed.2014.1608.

Levy, Andrea Gurmankin, Aaron M. Scherer, Brian J. Zikmund-Fisher, Knoll Larkin, Geoffrey D. Barnes, and Angela Fagerlin. 2018. "Prevalence of and Factors Associated With Patient Nondisclosure of Medically Relevant Information to Clinicians." *JAMA Network Open* 1(7):e185293. doi: 10.1001/jamanetworkopen.2018.5293.

Link, Bruce G., and Jo Phelan. 1995. "Social Conditions As Fundamental Causes of Disease." *Journal of Health and Social Behavior* 80–94. doi: 10.2307/2626958.

Lutfey, Karen, and Jeremy Freese. 2005. "Toward Some Fundamentals of Fundamental Causality: Socioeconomic Status and Health in the Routine Clinic Visit for Diabetes." *American Journal of Sociology* 110(5):1326–72. doi: 10.1086/428914.

Matthiesen, Stina, Søren Zöga Diederichsen, Mikkel Klitzing Hartmann Hansen, Christina Villumsen, Mats Christian Højbjerg Lassen, Peter Karl Jacobsen, Niels Risum, Bo Gregers Winkel, Berit T. Philbert, Jesper Hastrup Svendsen, and Tariq Osman Andersen. 2021. "Clinician Preimplementation Perspectives of a Decision-Support Tool for the Prediction of Cardiac Arrhythmia Based on Machine Learning: Near-Live Feasibility and Qualitative Study." *JMIR Human Factors* 8(4):e26964. doi: 10.2196/26964.

McCradden, Melissa D., Shalmali Joshi, James A. Anderson, Mjaye Mazwi, Anna Goldenberg, and Randi Zlotnik Shaul. 2020. "Patient Safety and Quality Improvement: Ethical Principles for a Regulatory Approach to Bias in Healthcare Machine Learning." *Journal of the American Medical Informatics Association* 27(12):2024–27. doi: 10.1093/jamia/ocaa085.

McCradden, Melissa, Shalmali Joshi, Mjaye Mazwi, and James A. Anderson. 2020. "Ethical Limitations of Algorithmic Fairness Solutions in Health Care Machine Learning." *The Lancet Digital Health* 2(5):e221–23. doi: 10.1016/S2589-7500(20)30065-0.

McMillan Cottom, Tressie. 2020. "Where Platform Capitalism and Racial Capitalism Meet: The Sociology of Race and Racism in the Digital Society." *Sociology of Race and Ethnicity* 6(4):441–49. doi: 10.1177/2332649220949473.

Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45:22.

Murray, Genevra F., Hector P. Rodriguez, and Valerie A. Lewis. 2020. "Upstream With A Small Paddle: How ACOs Are Working Against The Current To Meet Patients' Social Needs." *Health Affairs* 39(2):199–206. doi: 10.1377/hlthaff.2019.01266.

Murray, Sara, Robert Wachter, and Russell Cucina. 2020a. "Discrimination by Artificial Intelligence in a Commercial Electronic Health Record- a Case Study." *Health Affairs Blog*. doi: 10.1377/hblog20200128.626576.

Murray, Sara, Robert Wachter, and Russell Cucina. 2020b. "Discrimination By Artificial Intelligence In A Commercial Electronic Health Record—A Case Study | Health Affairs."

Nemati, Shamim, Andre Holder, Fereshteh Razmi, Matthew D. Stanley, Gari D. Clifford, and Timothy G. Buchman. 2018. "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU." *Critical Care Medicine* 46(4):547–53. doi: 10.1097/CCM.0000000000002936.

Ngiam, Kee Yuan, and Ing Wei Khor. 2019. "Big Data and Machine Learning Algorithms for Health-Care Delivery." *The Lancet Oncology* 20(5):e262–73. doi: 10.1016/S1470-2045(19)30149-4.

Niederer, Steven A., Joost Lumens, and Natalia A. Trayanova. 2019. "Computational Models in Cardiology." *Nature Reviews Cardiology* 16(2):100–111. doi: 10.1038/s41569-018-0104-y.

Nong, Paige, Alicia Williamson, Denise Anthony, Jodyn Platt, and Sharon Kardia. 2022. "Discrimination, Trust, and Withholding Information from Providers: Implications for Missing Data and Inequity." *SSM - Population Health* 18:101092. doi: 10.1016/j.ssmph.2022.101092.

Obermeyer, Ziad, and Ezekiel J. Emanuel. 2016. "Predicting the Future - Big Data, Machine Learning, and Clinical Medicine." *The New England Journal of Medicine* 375(13):1216–19. doi: 10.1056/NEJMp1606181.

Obermeyer, Ziad, Rebecca Nissan, Michael Stern, Emily Joy Bembeneck, and Sendhil Mullainathan. 2021. *Algorithmic Bias Playbook*. Chicago Booth Center for Applied Artifical Intelligence.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366(6464):447–53. doi: 10.1126/science.aax2342.

Obermeyer, Ziad, and Eric J. Topol. 2021. "Artificial Intelligence, Bias, and Patients'
        Perspectives." *The Lancet* 397(10289):2038. doi: 10.1016/S0140-6736(21)01152-1.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and
        Threatens Democracy*. Crown.

Orlikowski, Wanda. 2000. "Using Technology and Constituting Structures: A Practice Lens for
        Studying Technology in Organizations." *Organization Science* 11(4). doi:
        10.1287/orsc.11.4.404.14600.

Pasquale, Frank. 2015. *The Black Box Society*. Harvard University Press.

Phelan, Jo C., and Bruce G. Link. 2015. "Is Racism a Fundamental Cause of Inequalities in
        Health?" *Annual Review of Sociology* 41:311–30.

Pierson, Emma, David M. Cutler, Jure Leskovec, Sendhil Mullainathan, and Ziad Obermeyer.
        2021. "An Algorithmic Approach to Reducing Unexplained Pain Disparities in
        Underserved Populations." *Nature Medicine* 27(1):136–40. doi: 10.1038/s41591-020-
        01192-7.

Platt, Jodyn, and Sharon Kardia. 2015. "Public Trust in Health Information Sharing: Implications
        for Biobanking and Electronic Health Record Systems." *Journal of Personalized
        Medicine* 5(1):3–21. doi: 10.3390/jpm5010003.

Porter, Theodore M. 2020. *Trust in Numbers: The Pursuit of Objectivity in Science and Public
        Life*. Princeton University Press.

Price, W. Nicholson, and I. Glenn Cohen. 2019. "Privacy in the Age of Medical Big Data."
        Nature Medicine 25(1):37–43. doi: 10.1038/s41591-018-0272-7.

Rajkomar, Alvin, Michaela Hardt, Michael D. Howell, Greg Corrado, and Marshall H. Chin.
        2018. "Ensuring Fairness in Machine Learning to Advance Health Equity." *Annals of
        Internal Medicine* 169(12):866–72. doi: 10.7326/M18-1990.

Richardson, Jordan P., Cambray Smith, Susan Curtis, Sara Watson, Xuan Zhu, Barbara Barry,
        and Richard R. Sharp. 2021. "Patient Apprehensions about the Use of Artificial
        Intelligence in Healthcare." *Npj Digital Medicine* 4(1):1–6. doi: 10.1038/s41746-021-
        00509-1.

Richardson, Rashida, Jason Schultz, and Kate Crawford. 2019. *Dirty Data, Bad Predictions:
        How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and
        Justice. SSRN Scholarly Paper*. ID 3333423. Rochester, NY: Social Science Research
        Network.

Riley, Alicia R. 2018. "Neighborhood Disadvantage, Residential Segregation, and Beyond—
        Lessons for Studying Structural Racism and Health." *Journal of Racial and Ethnic
        Health Disparities* 5(2):357–65. doi: 10.1007/s40615-017-0378-5.

van Ryn, Michelle, Diana J. Burgess, John F. Dovidio, Sean M. Phelan, Somnath Saha, Jennifer
        Malat, Joan M. Griffin, Steven S. Fu, and Sylvia Perry. 2011. "The Impact of Racism on
        Clinician Cognition, Behavior, and Clinical Decision Making." *Du Bois Review : Social
        Science Research on Race* 8(1):199–218. doi: 10.1017/S1742058X11000191.

Schoorman, F. David, Roger C. Mayer, and James H. Davis. 2007. "An Integrative Model of
        Organizational Trust: Past, Present, and Future." *The Academy of Management Review*
        32(2):344–54. doi: 10.2307/20159304.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet
        Vertesi. 2019. "Fairness and Abstraction in Sociotechnical Systems." Pp. 59–68 in
        Proceedings of the Conference on Fairness, Accountability, and Transparency. Atlanta
        GA USA: ACM.

Sewell, Alyasah, Kevin Jefferson, and Hedwig Lee. 2016. "Living under Surveillance: Gender, Psychological Distress, and Stop-Question-and-Frisk Policing in New York City." *Social Science & Medicine* 159. doi: 10.1016/j.socscimed.2016.04.024.

Seyyed-Kalantari, Laleh, Haoran Zhang, Matthew B. A. McDermott, Irene Y. Chen, and Marzyeh Ghassemi. 2021. "Underdiagnosis Bias of Artificial Intelligence Algorithms Applied to Chest Radiographs in Under-Served Patient Populations." *Nature Medicine*. doi: 10.1038/s41591-021-01595-0.

Shilo, Smadar, Hagai Rossman, and Eran Segal. 2020. "Axes of a Revolution: Challenges and Promises of Big Data in Healthcare." *Nature Medicine* 26(1):29–38. doi: 10.1038/s41591-019-0727-5.

Singh, Karandeep, Rebecca A. Betensky, Adam Wright, Gary C. Curhan, David W. Bates, and Sushrut S. Waikar. 2016. "A Concept–Wide Association Study of Clinical Notes to Discover New Predictors of Kidney Failure." *Clinical Journal of the American Society of Nephrology* 11(12):2150–58. doi: 10.2215/CJN.02420316.

Singh, Karandeep, Thomas S. Valley, Shengpu Tang, Benjamin Y. Li, Fahad Kamran, Michael W. Sjoding, Jenna Wiens, Erkin Otles, John P. Donnelly, Melissa Y. Wei, Jonathon P. McBride, Jie Cao, Carleen Penoza, John Z. Ayanian, and Brahmajee K. Nallamothu. 2020. "Validating a Widely Implemented Deterioration Index Model Among Hospitalized COVID-19 Patients." *MedRxiv* 2020.04.24.20079012. doi: 10.1101/2020.04.24.20079012.

Siwicki, Bill. 2021. "How CIOs Are Prioritizing AI Investments for the next 5 Years." *Healthcare IT News*, October 18.

Spencer, Karen Lutfey, and Matthew Grace. 2016. "Social Foundations of Health Care Inequality and Treatment Bias." *Annual Review of Sociology* 42(1):101–20. doi: 10.1146/annurev-soc-081715-074226.

Tan, Marissa, Elham Hatef, Delaram Taghipour, Kinjel Vyas, Hadi Kharrazi, Laura Gottlieb, and Jonathan Weiner. 2020. "Including Social and Behavioral Determinants in Predictive Models: Trends, Challenges, and Opportunities." *JMIR Medical Informatics* 8(9):e18084. doi: 10.2196/18084.

Timmermans, Stefan, and Marc Berg. 2003. "The Practice of Medical Technology." *Sociology of Health & Illness* 25(3):97–114. doi: 10.1111/1467-9566.00342.

Timmermans, Stefan, and Mara Buchbinder. 2010. "Patients-in-Waiting: Living between Sickness and Health in the Genomics Era." *Journal of Health and Social Behavior* 51(4):408–23. doi: 10.1177/0022146510386794.

Veinot, Tiffany C., Hannah Mitchell, and Jessica S. Ancker. 2018. "Good Intentions Are Not Enough: How Informatics Interventions Can Worsen Inequality." *Journal of the American Medical Informatics Association* 25(8):1080–88. doi: 10.1093/jamia/ocy052.

Vyas, Darshali A., Leo G. Eisenstein, and David S. Jones. 2020. "Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms." *New England Journal of Medicine* 383(9):874–82. doi: 10.1056/NEJMms2004740.

Waljee, Akbar K., Peter D. R. Higgins, and Amit G. Singal. 2014. "A Primer on Predictive Models." *Clinical and Translational Gastroenterology* 5(1):e44. doi: 10.1038/ctg.2013.19.

Walker, Rachael C., Allison Tong, Kirsten Howard, and Suetonia C. Palmer. 2019. "Patient Expectations and Experiences of Remote Monitoring for Chronic Diseases: Systematic

Review and Thematic Synthesis of Qualitative Studies." *International Journal of Medical Informatics* 124:78–85. doi: 10.1016/j.ijmedinf.2019.01.013.

Wang, Fei, Rainu Kaushal, and Dhruv Khullar. 2020. "Should Health Care Demand Interpretable Artificial Intelligence or Accept 'Black Box' Medicine?" *Annals of Internal Medicine* 172(1):59–60. doi: 10.7326/M19-2548.

Wang, Fei, and Anita Preininger. 2019. "AI in Health: State of the Art, Challenges, and Future Directions." *Yearbook of Medical Informatics* 28(1):16–26. doi: 10.1055/s-0039-1677908.

White, Kellee, Jennifer S. Haas, and David R. Williams. 2012. "Elucidating the Role of Place in Health Care Disparities: The Example of Racial/Ethnic Residential Segregation." *Health Services Research* 47(3 Pt 2):1278–99. doi: 10.1111/j.1475-6773.2012.01410.x.

Williams, David R., and Chiquita Collins. 2001. "Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health." *Public Health Reports* 116(5):404–16. doi: 10.1093/phr/116.5.404.

Wong, Andrew, Erkin Otles, John P. Donnelly, Andrew Krumm, Jeffrey McCullough, Olivia DeTroyer-Cooley, Justin Pestrue, Marie Phillips, Judy Konye, Carleen Penoza, Muhammad Ghous, and Karandeep Singh. 2021. "External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients." *JAMA Internal Medicine*. doi: 10.1001/jamainternmed.2021.2626.

Wynants, Laure, Maarten van Smeden, David J. McLernon, Dirk Timmerman, Ewout W. Steyerberg, Ben Van Calster, and on behalf of the Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. 2019. "Three Myths about Risk Thresholds for Prediction Models." *BMC Medicine* 17(1):192. doi: 10.1186/s12916-019-1425-3.

Zhang, Simone, Rebecca A. Johnson, John Novembre, Edward Freeland, and Dalton Conley. 2021. "Public Attitudes toward Genetic Risk Scoring in Medicine and Beyond." *Social Science & Medicine* 274:113796. doi: 10.1016/j.socscimed.2021.113796.

Ziebland, Sue, Emma Hyde, and John Powell. 2021. "Power, Paradox and Pessimism: On the Unintended Consequences of Digital Health Technologies in Primary Care." *Social Science & Medicine* 289:114419. doi: 10.1016/j.socscimed.2021.114419.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. First edition. New York: PublicAffairs.

Chapter 2 Public Perspectives on Data Use and Indicators of Structural Racism

BACKGROUND

Predictive models and risk scores are often promoted as evidence-based, data-driven decision support for healthcare providers and health systems (Chen and Asch 2017; Desai 2020; Ghassemi et al. 2021; Shilo et al. 2020). Such data-intensive approaches to healthcare, while not new, are deployed in the healthcare system for a growing variety of purposes (Gupta et al. 2021; Ngiam and Khor 2019; Wang and Preininger 2019). These models inform management, administration, and clinical care across the country (Tan et al. 2020) and are used in a wide range of medical fields. In 2021, a large majority of acute care organizations were using predictive analytics and risk-based stratification to manage their patient populations (Apathy et al. 2021; CHIME 2021).

Predictive models in healthcare rely on increasingly large amounts of data, with a consistent goal of greater precision and more accurate predictive capability. Data largely comes from electronic health record (EHR) systems within and across health systems but is also sourced from places outside clinical delivery sites such as industry (e.g., pharmaceutical sources, labs) and patient-facing mobile applications (e.g., nutrition tracking apps). The types of data drawn from these varied sources differ between predictive models. Some models use data on clinical indicators and diagnoses (Bai et al. 2020; Matthiesen et al. 2021; Wong et al. 2021). Others use patient ethnicity, religious affiliation, and payment of medical bills (Ding et al. 2018; Murray, Wachter, and Cucina 2020). Still others use administrative claims data to make predictions about patient need (Obermeyer et al. 2019).

Even as predictive tools themselves are applied to more data sources to manage patients with increasing precision, very little data on patient perspectives has been analyzed (Richardson et al. 2021). Attempts to compute, predict, and manage patients through their data are receiving significant investment and effort in healthcare, but deeper understanding of the system of prediction and its social implications is limited. Patient perspectives, values, and priorities in relation to this type of advanced health IT have not been extensively analyzed.

Prediction operates as a method of classification, whereby patients are sorted and stratified. Concerns about the implications for patients' lives, access to healthcare, and quality of care have grown alongside the excitement about the potential benefits of prediction, spurred by multiple analyses of discrimination and bias in predictive classification systems (Murray et al. 2020; Obermeyer et al. 2019). This is discussed in detail below. First, prediction as classification is described based on the sociological literature. Subsequently, the implications for structural racism are considered. This is followed by the conceptualization of patients as key stakeholders in the system of prediction in healthcare and description of the objective for this work that meets the current need for 1) evidence-based understanding of patients' perspectives on data use, and 2) analysis of racial inequities.

Prediction as classification

Predictive models are integrated into healthcare to create systems of classification that sort and analyze people in a wide variety of contexts. Using various computational techniques, these models draw on high volumes of data from individuals and groups to classify people according to their predicted behaviors and other outcomes (Burrell and Fourcade 2021; Hoffmann 2019; Molina and Garip 2019; Selbst et al. 2019). Demographic information, online

behavior, social media engagement, credit histories, location data, biometric data, and a host of other data types are used to predict peoples' behavior or outcomes across a range of domains (Zuboff 2019).

Critical literatures on prediction in policing, financial institutions, insurance, and education are considerable (Brayne 2021; Burrell and Fourcade 2021; Ferguson 2016; Fourcade and Healy 2016; McMillan Cottom 2020; O'Neil 2016). Here, algorithmic predictions are understood as systems of classification, or ways of sorting people, that reflect the needs and goals of various institutions with serious consequences for people's lives (Fourcade and Healy 2013). Sociological literature specifically analyzes algorithms and implications for social position and opacity in modeling (Burrell 2016). Running through these analyses is an emphasis on why algorithms or predictive models are important: because they have serious implications for social position, mobility, and inequality. Predictive financial modeling can determine whether people can purchase a home. Predictive policing can expose Black people in particular to police surveillance, wrongful arrest, and fatal violence (Heaven 2020; Richardson, Schultz, and Crawford 2019). Algorithmic classification can expose Black families and other racial and ethnic minorities to disproportionate surveillance by child protective services and removal of their children from their homes (Eubanks 2018). These examples represent key ways classification can relate to inequity, including perpetuating existing biases or inequities in society, constructing new forms or mechanisms of inequity, and facilitating opaque decision-making that marginalizes the subjects of classification (boyd and Crawford 2012; Burrell and Fourcade 2021; Campolo and Crawford 2020).

Sociological work on systemic inequity and medical technologies is particularly informative related to the social implications of prediction and risk scoring (Casper and Morrison

2010; Link and Phelan 1995; Timmermans and Berg 2003). Whether classification shapes a patient's diagnosis or leads to delayed care through overbooking their appointment, the patient's experience of the health system can be directly affected by predictive technologies. As the sociological literature describes, the life circumstances and histories of the subjects of classification are circumscribed and defined by their position within the classification system (Burrell and Fourcade 2021). Classification systems "torque" people's lives (Bowker and Star 1999) based on the categories into which they are sorted. Receiving some diagnoses, for example, can legitimize a patient's symptoms, provide clarity to them, and open possibilities of treatment. Lack of diagnosis can be delegitimizing and confusing (Timmermans and Buchbinder 2010). Alternatively, diagnosis and classification as a certain "type" of patient can expose patients to social control, surveillance, and institutionalization. This was done through medicalizing diverse gender expressions and sexualities, for example (Conrad and Schneider 1992).

Depending on the specific prediction, patients' health outcomes can also be impacted by this kind of categorization. In the example analyzed by Obermeyer and colleagues, patients were either granted or denied access to additional care management support by an insurance algorithm (Obermeyer et al. 2019). Predictions of 30-day readmissions result in patients either receiving additional transition support or not (Leppin et al. 2014). Risk scores for overdose grant or deny access to certain treatments for patients who use substances (Apriss 2022). While classification for access to resources has a long history in the healthcare system, the literature highlights how the nature of classification changes when it is automated: increased opacity and decreased possibility of contestation (Burrell and Fourcade 2021). These classificatory tools allow the healthcare system to "make decisions without seeming to decide" (Porter 2020), obscuring who

is responsible for resource allocation. This makes contestation nearly impossible except for those with the most advantage (Pasquale 2015). Patients' avenues for contestation and knowledge of classification are thus foreclosed. Because classification systems grant or deny resources to individuals and groups of patients, patients have the most at stake in the world of predictive modeling in healthcare.

Structural racism and the system of prediction

The risks patients face in relation to predictive technologies are not equally distributed. Increasingly, the literature addresses how inequities might function in health IT. This includes algorithms, predictive models, and devices like the pulse oximeter. Such inquiries are fundamental to identifying the specific mechanisms within the tools themselves that drive inequity. However, less attention has been paid to how the system of health IT and advanced analytics is developing.

Identifying structural racism in healthcare institutions requires attention to the ways the structures are built. As Feagin and Bennefield describe, healthcare in the US is generally responsive to the needs, priorities, and values of white patients compared to racially minoritized patients (Feagin and Bennefield 2014). Should patients who are already racially marginalized in the healthcare system feel greater discomfort with the use of certain data for predictive models than white patients, this may indicate that the healthcare system's practices and policies are perpetuating racism. Conversely, if patients who already enjoy privileged status in the healthcare system are more comfortable with the use of their data, this may signal that the system is inequitably reflecting their interests, priorities, and values.

Questions about whose values this system reflects and how these dynamics can perpetuate structural racism are typically left unaddressed. This may reflect the paternalism observed in both medical and technology fields that converge in health IT where patients are assumed to be ignorant of the technology or incapable of engaging with complex medical technologies. The analysis presented here explicitly eschews both of those approaches and treats patients as fundamentally important to the sociotechnical system of healthcare.

The public as stakeholders

A focus on public consultation and patient engagement has been a longstanding priority in healthcare (Berwick et al. 2008). Information technologies have evolved to increasingly rely on patients to curate their data and repot it to healthcare providers in various ways. The public is increasingly expected to partner with providers in the use of data-drive information technologies as they seek and receive care (Gordon et al. 2020; Walker et al. 2019). As described above, patients using digital tools to take on growing responsibility for their health has been termed the role of the "digital health citizen" (Ziebland et al. 2021). This evolving role requires that patients trust the health system to protect that data and use it appropriately (Platt and Kardia 2015; Price and Cohen 2019). Understanding public trust in systems to use data appropriately is an important piece of analyzing patient values and perspectives related to predictive technologies in order to design the system as patient-centered.

However, little is known about how the public perceives and understands predictive tools. Their perspectives are not routinely evaluated or considered in the design or implementation of advanced health IT. Patient feedback, when it is gathered, is concentrated on patient-facing tools

like portals. While this is important, it does not address the much larger data infrastructure that operates in healthcare, affecting patient experiences or outcomes.

It is especially important to analyze public perceptions and comfort at this point in the expansion of predictive systems in healthcare. Without empirical analysis of patients as stakeholders, the health system runs the risk of designing predictive systems that perpetuate and create inequities (Bracic, Callier, and Price 2022). This is how preventable unintended consequences occur. They are unintended insofar as they are unexamined prior to design and implementation. A lack of specific and preventive consideration of harms results in a system that widens inequities. Thus, the analysis presented here specifically centers patient perspectives in order to identify potential signals of inequity before the system of prediction in healthcare is established and characterized by racism, sexism, and other types of structural inequities. It eschews the white racial frame and examines potential indicators of structural racism in healthcare predictions.

As part of this approach to the public as stakeholders, the analysis also accounts for specific dimensions of trust in the healthcare system's use of patient data. Trust in health systems or organizations has been identified as an important aspect of patient perspectives, with impacts on a variety of indicators of patient engagement or utilization (Gilson 2003; Hall et al. 2001; LaVeist, Isaac, and Williams 2009). Certain dimensions of trust play particularly significant roles in relation to information disclosure and information technology (Campos-Castillo et al. 2016; Schoorman, Mayer, and Davis 2007). Specifically, competence and integrity as dimensions of trust are included in this analysis to identify how public assessments of the systems' abilities and honesty are related to comfort with data use for prediction. Prior work has identified that low trust is related to withholding information from providers, which

demonstrates the stakes of trust in a data-driven healthcare system (Levy et al. 2018; Nong et al. 2022).

How the public perceives data use

Currently, there is not a large body of evidence analyzing how patients or the public understand and respond to data collection for prediction or related informatics tools like AI/ML (Richardson et al. 2021). However, evidence on patient perspectives around the collection of race/ethnicity data and analytics in other domains provide some insight on potential concerns and predictors of comfort.

Prior analysis has identified sensitivity about the collection of some specific data types among patients and the public. For example, qualitative interviews and small surveys indicate that patients have complicated feelings about race/ethnicity data being collected by healthcare providers and systems (Baker et al. 2005; Varcoe et al. 2009). In some circumstances, race/ethnicity data is viewed as information required for ameliorating disparities. However, there patients are also concerned that this kind of data collection could lead to harm for themselves or for marginalized racial groups (Varcoe et al. 2009). Although these studies focus on data collection rather than informatics applications, they indicate that patients can see both potential benefits and harms related to the use of race/ethnicity data. While comfort with data use for research generally has been extensively analyzed, specific data types outside of race/ethnicity are not frequently analyzed empirically.

Public attitudes about data use for complex computations in other domains also provide some insight into how the public understand data use for prediction in healthcare. For example, one study analyzed survey data from a national sample of US adults on the acceptability of data

use for polygenic risk score. In this case, acceptance of data use for varied by race/ethnicity (Zhang et al. 2021). The literature contains multiple calls for empirical analysis of patient perspectives on AI/ML and predictive modeling for healthcare, especially in the context of concerns about trustworthiness and the risk of exacerbation of racial inequities (Nundy, Montgomery, and Wachter 2019; Obermeyer and Topol 2021).

OBJECTIVE

This chapter examines variation in public comfort with the use of different data types for prediction in healthcare using a national survey of US adults. The analysis focuses on potential racial/ethnic variation in comfort with the current customs and practices of prediction in healthcare as one method of understanding structural racism. It examines predictors of comfort with data use, with an emphasis on understanding the role of experiences of interpersonal discrimination in the healthcare system. The research questions are as follows:

1. Does public comfort co-vary between different types of data used for prediction? What are the underlying, co-varying dimensions (multivariate factors) of public comfort with use of various data types for prediction in healthcare?

2. What individual-level variables (e.g., racial/ethnic identity, experiences of discrimination, and attitudes or beliefs) predict the multivariate factors of comfort with the use of data for predictive modeling in healthcare?

METHODS

This analysis uses cross-sectional data from a nationally representative survey of US adults who can speak English. The survey sample is the National Opinion Research Center's

(NORC) AmeriSpeak Panel. A total of 1,541 participants completed the survey with oversamples of African American respondents, Hispanic respondents, and respondents earning less than 200% of the federal poverty level. Observations with missing data for demographics, comfort with data use, and the other covariates listed below in Table 2.2 were excluded. This resulted in an analytic sample of 1,436. NORC calculated poststratification survey weights based on demographics and Census division from the Current Population Survey. Weights for non-response are also included. The survey was tested for comprehension through cognitive interviews (n=17) during which participants "thought aloud" as they took the survey verbally in order to identify survey items that were unclear or confusing. The survey was edited based on these interviews, then pre-tested using MTurk (n=550) and pilot tested with a sample of AmeriSpeak panel participants (n=150).

The survey instrument included a 90-second explanatory video describing how health information is used and shared in the US healthcare system. Experts on health data sharing reviewed the video and provided input on the content, concepts, and wording. Definitions of key terms (health system, healthcare provider, etc.) were provided and available to respondents as hover-over text each time the term was used in the survey.

Predictive models were defined and described in a short paragraph (Flesch-Kincaid score 8.7) immediately preceding the survey questions on prediction in healthcare. The paragraph included popular examples of predictive technology applications outside of and within the healthcare system. This description and the key term definitions are available in the Appendix.

Measures

The outcome measure of this analysis is comfort with use of specific data types for the creation of predictive models in healthcare. Survey respondents indicated their comfort level on a four-point scale (1=Not comfortable to 4=Very comfortable) with each of the 15 data types listed on the survey. These data types were presented to respondents in randomized order to prevent the introduction of bias based on the order in which the survey items were viewed. All 15 data types presented to survey respondents are presented in Table 2.1.

Table 2.1 Survey measure: types of data used for prediction

	How comfortable are you with predictive models using each type of information?
1	Age
2	Sex
3	Race/ethnicity
4	Income
5	Weight
6	Marital status
7	Cancer diagnosis or family history of cancer
8	Blood pressure
9	Health behavior (smoking status, alcohol use, etc.)
10	Mental health diagnoses
11	Employment status
12	Genetic information
13	History of incarceration
14	Zip code
15	Health insurer

Independent variables include self-reported age in years, sex, and race/ethnicity (Multiracial, Hispanic, non-Hispanic Asian, non-Hispanic Black, non-Hispanic white, and other). Respondents reported their annual household income and education level (no high school

diploma, high school or equivalent, some college, BA or more). Health-related independent variables include health insurance status (insured/uninsured), a binary measure of healthcare utilization in the past 12 months, self-reported health status (poor to very good), and experiences of discrimination in the healthcare system (yes/no).

Additional independent variables of interest include composite measures of aspects of trust in how the health system manages and uses patient data. These include trust in competence and integrity related to data use. Competence was measured by two survey items using a four-point true scale (1=Not true, 4=Very true). Respondents indicated how true it was that their health systems "*Have a good track record of using health information responsibly*" and "*Can be trusted to keep health information secure".* Integrity, defined as honesty about information use, indicates respondents' perceptions that the health system is forthcoming about how their health information is used ("*Tell me how my health information is used"* and "*Would never mislead me about how my health information is used"*).

Ethical considerations

This study was approved by the institutional review board. Participants were paid for their time according to NORC's standards for participant incentives based on the duration of the survey.

ANALYSIS

To answer research question one, distributions of comfort with all 15 data types were examined along with correlations between them in a correlation matrix. Principal components factor analysis was conducted as a method of data reduction for the 15 data types. This produced

multivariate factors of public comfort with data use for prediction. The resulting factors were confirmed using Cronbach's alpha with the full sample, and by race/ethnicity. The Cronbach's alpha testing confirmed that the dimensions identified in the factor analysis were valid for the full sample and for each racial/ethnic group identified in the data.

The three dimensions identified in the factor analysis were named (personal characteristic data, health-related data, and sensitive data types). These data types were then analyzed as dependent variables in bivariable and multivariable regressions to answer research question two. This allowed for the examination of relationships between each data category and the independent variables of interest, as well as the demographic and health-related covariates. Bivariable regressions were conducted separately of each data type on race/ethnicity, experiences of discrimination, and the demographic and health-related covariates listed in the measures section above. These include; insurance status, self-reported health status, income, education, utilization in the last 12 months. These models also included measures of system competence and integrity with data use. Three multivariable linear regressions were then conducted with each data category as the dependent variable.

In the results section below, descriptive statistics for the full analytic sample are presented. These are followed by the results related to research question one, which are described in full detail. Following this section, the results for research question two are described.

RESULTS

As reported in Table 2.2, half of the sample was female, and representation of racial and ethnic groups generally reflected the US population. A large majority had health insurance (93.5%) and had seen a healthcare provider within the previous year (84.5%). The sample was

roughly evenly divided between individuals reporting less than $50,000 in annual household income and those earning more than $50,000 per year.

Table 2.2 Descriptive statistics (n=1,436)

Measure	n	%
Sex		
Female	718	50
Male	718	50
Age		
18-29	87	6.1
30-44	411	28.6
45-59	364	25.4
60+	574	40
Race/ethnicity		
White	895	62.3
Hispanic	255	17.8
Black	191	13.3
Multiracial and other	57	4
Asian	38	2.7
Education		
Less than high school	37	2.6
High school	242	16.9
Some college	666	46.4
BA or more	491	34.2
Annual household income		

<$50,000	727	50.6
At least $50,000	709	49.4
Health insurance coverage		
No	94	6.6
Yes	1,342	93.5
Self-reported health		
Poor to fair	336	23.4
Good	612	42.6
Very good to excellent	488	34
Last healthcare visit		
Longer than 1 year	222	15.5
Within past year	1,214	84.5
System competence with data use (mean, SD)	2.3	0.82
System integrity with data use (mean, SD)	2.4	0.9

Descriptive statistics indicate that comfort with data use for prediction is low (Figure 2.1). Only two of the data types analyzed here, blood pressure and health behaviors, were acceptable to a majority of the sample. Comfort with the use of income for prediction in healthcare was lowest (18.62%), followed closely by history of incarceration (19.86%).

Figure 2.1 Comfort with 15 data types for prediction in healthcare

Proportion of respondents reporting comfort with data use for prediction (n=1,436)

Data type	Value
Income	18.38
History of incarceration	19.57
Employment status	22.77
Zip code	26.53
Health insurer	27.09
Marital status	30.15
Race/ethnicity	34.33
Mental health diagnoses	39.14
Genetic information	40.39
Sex	42.06
Weight	44.15
Age	45.96
Health behaviors	47.42
Cancer diagnosis	52.72
Blood pressure	52.79

In Figure 1, comfort is defined as "very comfortable" or "somewhat comfortable" with the use of each type of data for prediction.

Research Question 1

What are the underlying, co-varying dimensions (multivariate factors) of public comfort with use of various data types for prediction in healthcare?

Exploratory factor analysis was conducted for the full sample to identify multivariate factors underlying public comfort with the use of all 15 data types. An oblique promax rotation was used. The Kaiser-Meyer-Olkin (KMO) value was 0.95, indicating that sampling adequacy was very strong. Variables were considered to load on a factor when the loading was 0.4 or above. Using this criteria, three factors were identified and confirmed with a scree test indicating that three factors should be retained. The variables and their factor loadings are listed in Table 2.3.

Table 2.3 Factor analysis of comfort with the use of all data types for prediction

Variable	Factor 1	Factor2	Factor3
Eigen value	8.43	1.34	0.69
Age	**0.79**	0.18	0.005
Sex	**0.80**	0.09	0.01
Race/ethnicity	**0.66**	0.02	0.14
Weight	**0.73**	0.24	0.01
Cancer diagnosis or family history	0.06	**0.89**	0.06
Blood pressure	0.12	**0.88**	0.03
Health behaviors	0.11	**0.67**	0.05
Mental health diagnoses	0.03	**0.69**	0.17
Genetic information	0.17	**0.44**	0.21
Income	0.17	0.06	**0.65**
Marital status	0.38	0.04	**0.49**
Employment status	0.02	0.08	**0.80**
History of incarceration	0.05	0.04	**0.82**
Zip code	0.09	0.02	**0.71**
Health insurer	0.01	0.06	**0.76**

Eigen values indicate that factors 1-3 should be retained (Figure 2.1), which is confirmed by the Scree plot (Figure 2.2). The eigen value of factor 1 was 8.43. Factor 2 was 1.34 and Factor 3 was 0.69. Cumulative variance explained by the three factors is approximately 100%.

Figure 2.2 Scree plot

Scree plot of eigenvalues after factor
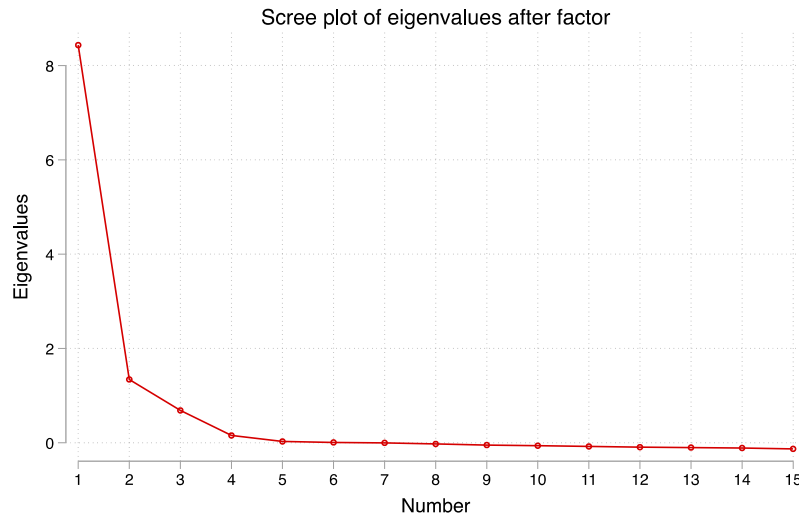
Figure 2.3 shows the distributions of comfort with each data type according to the factor on which the variables loaded. The variables that loaded on factor 1 were largely demographic and personal characteristic data, including age, weight, sex, and race/ethnicity. Variables that met the criteria for loading on factor 2 included data related to health status and diagnoses (e.g., blood pressure, cancer diagnosis, health behaviors). The third factor included variables describing social experience or position (e.g., income, marital status, history of incarceration). For further analysis, these factors are labeled as follows: 1) personal characteristic data, 2) health-related data, and 3) sensitive data. Three variables loaded on factor 1 (personal characteristic data). Five variables loaded on factor 2 (health-related data), and seven variables loaded on factor 3 (sensitive data).

Figure 2.3 Distributions of comfort with each data type by factor

Factor 1: Personal Characteristic Data

Factor 2: Health-related Data

Factor 3: Sensitive Data
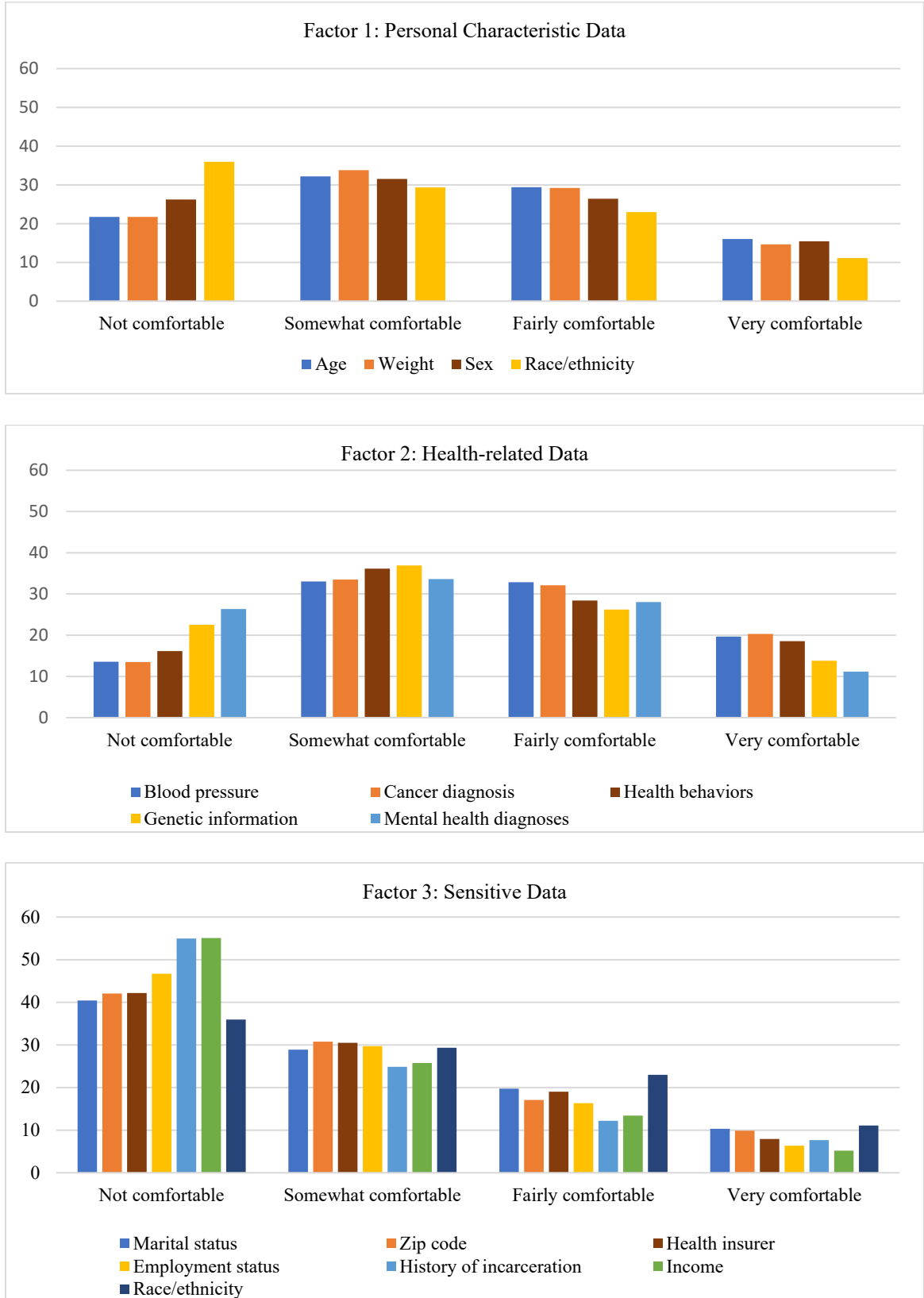
Factor analysis was repeated separately by race/ethnicity to confirm that the factor loadings were valid for all racial and ethnic groups represented in the data. Factor loadings were similar between the full sample and the analyses by race/ethnicity. There were two exceptions. Marital status loaded with demographics and personal characteristic data for Black and Hispanic respondents. Genetic information also loaded on this factor for Hispanic respondents. Otherwise, the factors were consistent with the same variables loading together across racial and ethnic groups represented in the data. For the smallest demographic groups (Asian, Multiracial and Other) there was slightly more difference compared to the full sample. The loadings were not quite as strong, which may be a function of the relatively small sample sizes.

However, concerns about factor validity across racial and ethnic groups are resolved by the measures of internal consistency. These Cronbach's alpha measures for each factor, created according to factor loadings for the full sample, were high and well above the acceptability threshold of 0.65 in the full sample and by racial/ethnic group (Table 2.4). This confirms that the three multivariate factors identified in the full sample are valid and do not disproportionately reflect the opinions or experiences of the white respondents who are predominant in the sample.

Additionally, as demonstrated in the Factor 1 panel of Figure 2.3, comfort with the use of race/ethnicity data is distributed slightly differently than the other variables that loaded on that factor. To ensure that the factor loading was appropriate, additional analysis was conducted. Alternative measures of internal consistency were calculated and compared to the measures generated by the factor analysis. Although race/ethnicity did not meet the criteria (at least 0.4) for loading on factor 3 (loading factor 0.14) in the factor analysis, Cronbach's alpha was re-calculated for factor 3 with race/ethnicity included. The Cronbach's alpha for factor 3 did not increase when it was calculated to include race/ethnicity compared to the initial calculation.

However, the Cronbach's alpha decreased for factor 1 when race/ethnicity was excluded. This finding, combined with the very low factor loading for race/ethnicity with factor 3, confirmed that race/ethnicity data loaded on factor 1. However, its slightly different distribution from the other variables in factor 1 indicated that further analysis of this particular variable was warranted. This analysis is presented below in the results section.

With the full sample, Cronbach's alpha for each identified factor was high (at least 0.88) and well above the acceptability threshold of 0.65. The scale reliability coefficients by race/ethnicity were similar to the overall Cronbach's alpha and all remained above 0.85 (Table 2.4). This confirmed the validity of the multivariate factors identified in the exploratory factor analysis.

Table 2.4 Measures of internal consistency for each multivariate factor

Factor	Items	Cronbach's α for full sample	α for Black respondents	α for Hispanic respondents	α for white respondents	α for multiracial respondents	α for Asian respondents
			n=191	n=255	n=895	n=57	38
1	Age Sex Race Weight	0.93	0.91	0.94	0.94	0.94	0.87
2	Cancer diagnosis Blood pressure Health behaviors Mental health Genetic information	0.92	0.90	0.92	0.93	0.93	0.91
3	Income Marital status Employment status History of incarceration Zip code Insurer	0.90	0.88	0.90	0.90	0.91	0.93

This analysis identifies that there are three multivariate factors of public comfort with use of various data types for prediction in healthcare. The factors are robust and confirmed through multiple statistical tests. This analysis confirms that the multivariate factors are valid across all racial and ethnic groups represented in this dataset.

Research Question 2

What individual-level variables (e.g., racial/ethnic identity, experiences of discrimination, and attitudes or beliefs) predict the multivariate factors of comfort with the use of data for predictive modeling in healthcare?

Descriptive statistics for the retained factors are included in Table 2.5. Composite measures of comfort were calculated as the mean of comfort scores for all the variables that loaded on each factor. The resulting measures were then analyzed as dependent variables. Each of these data categories, as composite measures, was analyzed using multivariable OLS regression. Specifically, each data category was regressed on the same combination of independent variables of interest (Table 2.6). Demographic and health-related data were included as covariates in these models.

Table 2.5 Descriptive statistics of retained factors

Factor	Factor Label	Mean	SD
1	Personal characteristic data	2.36	0.94
2	Health-related data	2.45	0.84
3	Sensitive data	1.88	0.77

Table 2.6 Multivariable OLS regression results for each composite measure of comfort

	Personal characteristic data		Health-related data		Sensitive data	
	b	p-value	b	p-value	b	p-value
Female (ref. male)	-0.21	<0.001***	-0.13	0.002**	-0.16	<0.001***
Age						
18-29	ref		ref		ref	
30-44	0.003	0.97	-0.16	0.07	-0.13	0.14
45-59	-0.13	0.21	-0.27	0.004**	-0.29	<0.001***
60+	-0.14	0.17	-0.22	0.02*	-0.32	<0.001***
Race/ethnicity						
White	ref		ref		ref	
Black	-0.19	0.004**	-0.16	0.01*	0.002	0.97
Hispanic	-0.02	0.71	0.003	0.96	0.06	0.26
Asian	0.04	0.8	0.03	0.83	0.15	0.2
Other	-0.16	0.18	-0.08	0.48	0.08	0.39
Education						
Less than high school	ref		ref		ref	
High school	-0.24	0.1	-0.13	0.34	-0.03	0.8
Some college	-0.18	0.2	-0.07	0.6	-0.02	0.84
BA or more	-0.05	0.75	0.04	0.73	0.01	0.93
Annual household income						
<$50,000	ref		ref		ref	
At least $50,000	0.1	0.03*	0.1	0.02*	-0.03	0.45
Health insurance coverage						
No	ref		ref		ref	
Yes	0.2	0.03*	0.12	0.16	0.07	0.34
Self-reported health						
Poor to fair	ref		ref		ref	
Good	-0.03	0.57	-0.05	0.32	-0.01	0.77
Very good to excellent	-0.02	0.74	-0.08	0.17	0.03	0.55
Last healthcare visit						
Longer than one year ago	ref		ref		ref	
Within past year	-0.04	0.53	-0.05	0.44	0.001	0.99
Experienced discrimination						
No	ref		ref		ref	
Yes	-0.13	0.03*	-0.02	0.69	-0.13	0.008**
System integrity with data	0.08	0.01*	0.09	<0.001***	0.11	<0.001***
System competence with data	0.35	<0.001***	0.36	<0.001***	0.27	<0.001***

Across all three types of data use for prediction in healthcare, female respondents were less comfortable compared to male respondents. Age was inconsistently associated with the dependent variables, whereby adults 45 and older were less comfortable than 18- to 29-year-olds with the use of health-related and sensitive data types.

Race and ethnicity were not generally statistically significant predictors in these models. However, there were statistically significant differences between Black and white respondents in comfort with personal characteristic data and health-related data. White respondents were significantly more comfortable with the use of these data types than Black respondents, holding all other covariates constant. It is possible that this relationship is not observed for sensitive data types because there is relatively low variation to be explained. Comfort with this type of data use was generally very low.

Trust in system integrity and competence with data were consistently positively predictive of comfort with use of all three data categories. Experiences of discrimination in healthcare were statistically significant negative predictors of comfort in the multivariable regressions for personal characteristic data and sensitive data. This may indicate a concern among people who have experienced discrimination in healthcare that they are vulnerable to digital marginalization or harm based on their social identities.

Use of race and ethnicity data

To understand the particularities of public perceptions of the use of race and ethnicity data, additional bivariable and multivariable regressions were conducted. Comfort with race and ethnicity data use for prediction was regressed on the full list of independent variables. Generally, comfort with race/ethnicity data in prediction is low (34.3%). This analysis identifies

that those who have experienced discrimination in healthcare are significantly less comfortable

with the use of race/ethnicity data for prediction (b=0.69, p=0.02). Female respondents were

significantly less comfortable with this kind of data use (b=0.64, p=<0.001). Compared to white

respondents, Black respondents are less likely to feel comfortable (b=0.63, p=0.01). Trust in

system competence with information use was a significant predictor of comfort with

race/ethnicity in prediction (b=1.99, p<0.001), while trust in system integrity was not.

Table 2.7 Abbreviated results of multivariable logistic regression of comfort with the use of race/ethnicity data for prediction

	Odds Ratio	p-value
Experienced discrimination		
No	ref	ref
Yes	0.69	0.02*
Sex		
Male	ref	
Female	0.64	<0.001***
Race/ethnicity		
White	ref	
Black	0.63	0.01*
Hispanic	1.28	0.12
Multiple	0.81	0.5
Asian	0.71	0.37
Income		
<$50,000	ref	ref
At least $50,000	1.27	0.06
System integrity with data	1.14	0.09
System competence with data	1.99	<0.001***

These results indicate the importance of perceived system competence in predicting

comfort with data use, which is expected based on prior literature on competence broadly

(Armstrong et al. 2008; Ozawa and Sripad 2013). They also demonstrate the salience of

experiences of discrimination for patients. Black respondents' significantly lower comfort with

the use of race/ethnicity data than white patients is important. It is a potential indicator that the system of prediction is being built to replicate existing structural racism, whereby white people's preferences and priorities are reflected by the system and Black people's preferences and priorities are not (Feagin and Bennefield 2014).

LIMITATIONS

There are limitations to this study that should inform interpretation. First, identity is limited to binary self-identified sex in this sample, which is an incomplete measure. The relationships between gender identities and comfort with data use for prediction are not presented here. Second, Native American and Alaskan Native respondents are not identified in this sample. This is a limitation in the data's representativeness and future work should specifically ensure that Native American and Alaskan Native respondents are 1) specifically identified and 2) adequately represented in analysis of comfort with data use.

Additionally, the cross-sectional nature of the data limits the inferences that can be made about the results. For example, it is not possible to identify longitudinal relationships between experiences of discrimination and perceptions of data use for prediction. Additionally, we do not observe measures of structural racism, which is likely to inform various aspects of how the public understands and responds to the concept of data use for prediction. Although experiences of discrimination are relevant, they capture only interpersonal discrimination rather than structural inequities that inform how patients experience the collection and use of their data. Future work should explore additional structural inequities and measures of structural racism in relation to comfort with data use. Qualitative work will also provide a more in-depth

understanding of how and why the public feels the way they do about prediction in health broadly, and the use of individual data categories specifically.

DISCUSSION

This analysis identifies how the public perceives data use for prediction in healthcare, indicating that current practice around prediction in healthcare is misaligned with public comfort. The public is comparatively more comfortable with the use of more overtly clinical data (blood pressure, cancer diagnoses, health behaviors) than demographic or social data types. This is reflective of prior analysis of public perceptions of polygenic risk scores where acceptability was higher for the use of data that seemed most directly relevant to a given score's purpose (Zhang et al. 2021). However, the data types collected and used by health systems to make predictions about patients are not restricted to data that seems directly relevant to clinical care. Prediction in practice involves the use of widely varied data types like religious identity and whether patients have outstanding medical bills (Ding et al. 2018; Murray et al. 2020). This may indicate that the system of prediction is developing in a way that could undermine trust by violating patient comfort and expectations.

Comfort with use of personal characteristic data (e.g., age, sex) for prediction was lower than health-related data with 34.1-45.6% of respondents reporting comfort this data category. Experiences of discrimination in healthcare was a significant negative predictor of comfort with the use of personal characteristic data. This is aligned with prior work indicating that those who experience discrimination are more likely to withhold information from their providers (Nong et al. 2022). Interpersonal harms perpetrated in the healthcare system become reflected in the data captured in EHRs and other sources because they inform how people relate to the healthcare

53

system's data ecosystem. The findings presented here indicate a similar dynamic, where people who have experienced discrimination are less comfortable with the use of data on their personal characteristics and demographic data for prediction.

Public comfort was lowest with the use of sensitive data types (e.g., income, history of incarceration) for prediction in healthcare. There was little variation in comfort with the use of type of data, with 87.6% of respondents indicating discomfort. While experiences of discrimination, system competence, and system integrity were predictive of comfort with sensitive data use, it is likely that the minimal variation in comfort with this data type precluded the identification of additional statistically significant predictors. Qualitative work exploring each data type comprising the composite measure, as well as why the public feels so negatively toward the use of this kind of data, will provide important insight on the dynamics underlying the relative lack of variation observed here.

This study has multiple implications for understanding how structural racism operates in advanced health IT. Specifically, white respondents are comparatively more comfortable with current practices of data use for prediction, indicating that the system is reflecting the white racial frame that historically shaped and contemporarily characterizes the healthcare system.

These findings identify some racial and ethnic variation in comfort with data use for prediction. Specifically, comfort with the use of data on race/ethnicity in prediction is significantly higher for white participants than for Black participants. This indicates that predictive models that use race/ethnicity data, or proxies for the same, violate the preferences of Black people in comparison to white people. Among experts on the perpetuation of racism in predictive or algorithmic tools, there is a diversity of opinions about the use of race/ethnicity data. Similarly, the literature has identified complex patient preferences and concerns related to

this kind of data collection (Baker et al. 2005; Varcoe et al. 2009). When race data is used to explicitly penalize Black patients, it is understood that this is an egregious act of racism. Some racial equity efforts, on the other hand, may be able to make progress in mitigating racist inequities by correcting or compensating for structural racism through an algorithmic reparations approach (Davis, Williams, and Yang 2021). The nuances in how the public perceives the implications of health-related predictions using race data are likely to be similarly complex (Varcoe et al. 2009). It is possible, for example, that some are concerned about the use of race as a potential penalty, as has occurred in a variety of clinical algorithms from VBAC scores to kidney transplant eligibility determinations (Vyas, Eisenstein, and Jones 2020). On the other hand, some people may see the inclusion of race as either a potential advantage for themselves or as a necessary method of quantifying and ameliorating inequity (Baker et al. 2005). While some prior qualitative work has been conducted on patient concerns about AI in healthcare (Richardson et al. 2021), specific qualitative analysis of the use of race data for prediction will expand on the insights gleaned from the results we present here. Necessary clarification on perceptions of what is driving the results identified here will come from future qualitative work.

CONCLUSION

This analysis identifies that comfort with the use of data for prediction in healthcare is generally low among US adults. The use of only two types of data, blood pressure and health behaviors, was acceptable for a majority of participants. Because predictive models in healthcare draw on increasingly diverse data types, and because known models have explicitly included some of the data types identified as unacceptable to a large majority of participants in a national sample, we identify misalignment between public comfort and current practice.

This work reveals three multivariate factors of public comfort with use of data for prediction. These factors and their internal consistency are confirmed using multiple methods and demonstrate conceptual coherence: 1) personal characteristic data, 2) health-related data, and 3) sensitive data. These factors are internally consistent in the full sample and for all racial/ethnic groups represented in the data. Multivariable analysis identifies that respondent sex, perception of system competence in data use, and perceived system integrity in data use are consistently positively associated with comfort with all three categories of data for prediction. Experiences of discrimination in healthcare are negatively predictive of comfort with personal characteristic data and sensitive data use.

Comfort with the use of personal characteristic and health-related data is higher for white respondents than for Black respondents. This indicates that the use of these data types may reflect existing structural racism whereby white people's perspectives and priorities are valued over those of Black people. As predictive technologies expand in the US healthcare system, these findings demonstrate that patient perspectives and anti-racist approaches to system governance are necessary.

References

Adler-Milstein, Julia, and Paige Nong. 2019. "Early Experiences with Patient Generated Health Data: Health System and Patient Perspectives." *Journal of the American Medical Informatics Association* 26(10):952–59. doi: 10.1093/jamia/ocz045.

Apathy, Nate C., A. Jay Holmgren, and Julia Adler-Milstein. 2021. "A Decade Post-HITECH: Critical Access Hospitals Have Electronic Health Records but Struggle to Keep up with Other Advanced Functions." *Journal of the American Medical Informatics Association* ocab102. doi: 10.1093/jamia/ocab102.

Apriss. 2022. *NarxCare: Analytics, Tools and Technology to Help Care Teams Address Substance Use Disorder and Improve Patient Outcomes*. State of Michigan Department of Licensing and Regulatory Affairs.

Armstrong, Katrina, Suzanne McMurphy, Lorraine T. Dean, Ellyn Micco, Mary Putt, Chanita Hughes Halbert, J. Sanford Schwartz, Pamela Sankar, Reed E. Pyeritz, Barbara Bernhardt, and Judy A. Shea. 2008. "Differences in the Patterns of Health Care System Distrust between Blacks and Whites." *Journal of General Internal Medicine* 23(6):827–33. doi: 10.1007/s11606-008-0561-9.

Bai, P., Y. Zhou, Y. Liu, G. Li, Z. Li, T. Wang, and X. Guo. 2020. "Risk Factors of Cerebral Infarction and Myocardial Infarction after Carotid Endarterectomy Analyzed by Machine Learning." *Computational and Mathematical Methods in Medicine* 2020. doi: 10.1155/2020/6217392.

Baker, David W., Kenzie A. Cameron, Joseph Feinglass, Patricia Georgas, Shawn Foster, Deborah Pierce, Jason A. Thompson, and Romana Hasnain-Wynia. 2005. "Patients' Attitudes toward Health Care Providers Collecting Information about Their Race and Ethnicity." *Journal of General Internal Medicine* 20(10):895–900. doi: 10.1111/j.1525-1497.2005.0195.x.

Berwick, Donald M., Thomas W. Nolan, and John Whittington. 2008. "The Triple Aim: Care, Health, And Cost." *Health Affairs* 27(3):759–69. doi: 10.1377/hlthaff.27.3.759.

Bowker, Geoffrey C., and Susan Leigh Star. 1999. *Sorting Things out: Classification and Its Consequences*. Cambridge, Mass: MIT Press.

boyd, danah, and Kate Crawford. 2012. "Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon." *Information, Communication & Society* 15(5):662–79. doi: 10.1080/1369118X.2012.678878.

Bracic, Ana, Shawneequa Callier, and W. Nicholson Price. 2022. "Exclusion Cycles: Reinforcing Disparities in Medicine." *Science* 377(6611).

Brayne, Sarah. 2021. *Predict and Surveil: Data, Discretion, and the Future of Policing*. New York, NY: Oxford University Press.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3(1):2053951715622512. doi: 10.1177/2053951715622512.

Burrell, Jenna, and Marion Fourcade. 2021. "The Society of Algorithms." *Annual Review of Sociology* 47(1):annurev-soc-090820-020800. doi: 10.1146/annurev-soc-090820-020800.

Campolo, Alexander, and Kate Crawford. 2020. "Enchanted Determinism: Power without Responsibility in Artificial Intelligence." *Engaging Science, Technology, and Society* 6(0):1–19. doi: 10.17351/ests2020.277.

Campos-Castillo, Celeste, Benjamin Woodson, Elizabeth Theiss-Morse, Tina Sacks, Michelle
        Fleig-Palmer, and Monica Peek. 2016. "Examining the Relationship Between
        Interpersonal and Institutional Trust in Political and Health Care Contexts." Pp. 99–115
        in.

Casper, Monica J., and Daniel R. Morrison. 2010. "Medical Sociology and Technology: Critical
        Engagements." *Journal of Health and Social Behavior* 51(1_suppl):S120–32. doi:
        10.1177/0022146510383493.

Chen, Jonathan H., and Steven M. Asch. 2017. "Machine Learning and Prediction in Medicine
        — Beyond the Peak of Inflated Expectations." *The New England Journal of Medicine*
        376(26):2507–9. doi: 10.1056/NEJMp1702071.

CHIME. 2021. *Healthcare's Most Wired: National Trends 2021*. *Survey*.

Conrad, Peter, and Joseph W. Schneider. 1992. *Deviance and Medicalization: From Badness to
        Sickness*. Temple University Press.

Davis, Jenny L., Apryl Williams, and Michael W. Yang. 2021. "Algorithmic Reparation." *Big
        Data & Society* 8(2):20539517211044810. doi: 10.1177/20539517211044808.

Desai, Angel. 2020. "Artificial Intelligence: Promise, Pitfalls, and Perspective." *JAMA Network
        Open* 323(24).

Ding, Xiruo, Ziad F. Gellad, Chad Mather, Pamela Barth, Eric G. Poon, Mark Newman, and
        Benjamin A. Goldstein. 2018. "Designing Risk Prediction Models for Ambulatory No-
        Shows across Different Specialties and Clinics." *Journal of the American Medical
        Informatics Association* 25(8):924–30. doi: 10.1093/jamia/ocy002.

Eubanks, Virginia. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and
        Punish the Poor*. St. Martin's Publishing Group.

Feagin, Joe, and Zinobia Bennefield. 2014. "Systemic Racism and U.S. Health Care." *Social
        Science & Medicine* 103:7–14. doi: 10.1016/j.socscimed.2013.09.006.

Ferguson, Andrew Guthrie. 2016. "Policing Predictive Policing." *Washington University Law
        Review* 94:1109.

Fourcade, Marion, and Kieran Healy. 2013. "Classification Situations: Life-Chances in the
        Neoliberal Era." *Accounting, Organizations and Society* 38(8):559–72. doi:
        10.1016/j.aos.2013.11.002.

Fourcade, Marion, and Kieran Healy. 2016. "Seeing like a Market." *Socio-Economic Review*
        mww033. doi: 10.1093/ser/mww033.

Ghassemi, Marzyeh, Luke Oakden-Rayner, and Andrew L. Beam. 2021. "The False Hope of
        Current Approaches to Explainable Artificial Intelligence in Health Care." *The Lancet
        Digital Health* 3(11):e745–50. doi: 10.1016/S2589-7500(21)00208-9.

Gilson, Lucy. 2003. "Trust and the Development of Health Care as a Social Institution." *Social
        Science & Medicine* 56(7):1453–68. doi: 10.1016/S0277-9536(02)00142-9.

Gupta, Kush, Dominick Frosch, and Robert Kaplan. 2021. "Opening The Black Box Of Digital
        Health Care: Making Sense Of 'Evidence.'" *Health Affairs Blog*.

Hall, Mark A., Elizabeth Dugan, Beiyao Zheng, and Aneil K. Mishra. 2001. "Trust in Physicians
        and Medical Institutions: What Is It, Can It Be Measured, and Does It Matter?" *The
        Milbank Quarterly* 79(4):613–39. doi: 10.1111/1468-0009.00223.

Heaven, Will Douglas. 2020. "Predictive Policing Algorithms Are Racist. They Need to Be
        Dismantled." *MIT Technology Review*, July 17.

Hoffmann, Anna Lauren. 2019. "Where Fairness Fails: Data, Algorithms, and the Limits of Antidiscrimination Discourse." *Information, Communication & Society* 22(7):900–915. doi: 10.1080/1369118X.2019.1573912.

LaVeist, Thomas A., Lydia A. Isaac, and Karen Patricia Williams. 2009. "Mistrust of Health Care Organizations Is Associated with Underutilization of Health Services." *Health Services Research* 44(6):2093–2105. doi: 10.1111/j.1475-6773.2009.01017.x.

Leppin, Aaron L., Michael R. Gionfriddo, Maya Kessler, Juan Pablo Brito, Frances S. Mair, Katie Gallacher, Zhen Wang, Patricia J. Erwin, Tanya Sylvester, Kasey Boehmer, Henry H. Ting, M. Hassan Murad, Nathan D. Shippee, and Victor M. Montori. 2014. "Preventing 30-Day Hospital Readmissions: A Systematic Review and Meta-Analysis of Randomized Trials." *JAMA Internal Medicine* 174(7):1095–1107. doi: 10.1001/jamainternmed.2014.1608.

Levy, Andrea Gurmankin, Aaron M. Scherer, Brian J. Zikmund-Fisher, Knoll Larkin, Geoffrey D. Barnes, and Angela Fagerlin. 2018. "Prevalence of and Factors Associated With Patient Nondisclosure of Medically Relevant Information to Clinicians." *JAMA Network Open* 1(7):e185293. doi: 10.1001/jamanetworkopen.2018.5293.

Link, Bruce G., and Jo Phelan. 1995. "Social Conditions As Fundamental Causes of Disease." *Journal of Health and Social Behavior* 80–94. doi: 10.2307/2626958.

Matthiesen, Stina, Søren Zöga Diederichsen, Mikkel Klitzing Hartmann Hansen, Christina Villumsen, Mats Christian Højbjerg Lassen, Peter Karl Jacobsen, Niels Risum, Bo Gregers Winkel, Berit T. Philbert, Jesper Hastrup Svendsen, and Tariq Osman Andersen. 2021. "Clinician Preimplementation Perspectives of a Decision-Support Tool for the Prediction of Cardiac Arrhythmia Based on Machine Learning: Near-Live Feasibility and Qualitative Study." *JMIR Human Factors* 8(4):e26964. doi: 10.2196/26964.

McMillan Cottom, Tressie. 2020. "Where Platform Capitalism and Racial Capitalism Meet: The Sociology of Race and Racism in the Digital Society." *Sociology of Race and Ethnicity* 6(4):441–49. doi: 10.1177/2332649220949473.

Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45:22.

Murray, Sara, Robert Wachter, and Russell Cucina. 2020. "Discrimination By Artificial Intelligence In A Commercial Electronic Health Record—A Case Study | Health Affairs."

Ngiam, Kee Yuan, and Ing Wei Khor. 2019. "Big Data and Machine Learning Algorithms for Health-Care Delivery." *The Lancet Oncology* 20(5):e262–73. doi: 10.1016/S1470-2045(19)30149-4.

Nong, Paige, Alicia Williamson, Denise Anthony, Jodyn Platt, and Sharon Kardia. 2022. "Discrimination, Trust, and Withholding Information from Providers: Implications for Missing Data and Inequity." *SSM - Population Health* 18:101092. doi: 10.1016/j.ssmph.2022.101092.

Nundy, Shantanu, Tara Montgomery, and Robert M. Wachter. 2019. "Promoting Trust Between Patients and Physicians in the Era of Artificial Intelligence." *JAMA* 322(6):497–98. doi: 10.1001/jama.2018.20563.

Obermeyer, Ziad, and Ezekiel J. Emanuel. 2016. "Predicting the Future - Big Data, Machine Learning, and Clinical Medicine." *The New England Journal of Medicine* 375(13):1216–19. doi: 10.1056/NEJMp1606181.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366(6464):447–53. doi: 10.1126/science.aax2342.

Obermeyer, Ziad, and Eric J. Topol. 2021. "Artificial Intelligence, Bias, and Patients' Perspectives." *The Lancet* 397(10289):2038. doi: 10.1016/S0140-6736(21)01152-1.

O'Neil, Cathy. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

Ozawa, Sachiko, and Pooja Sripad. 2013. "How Do You Measure Trust in the Health System? A Systematic Review of the Literature." *Social Science & Medicine* 91:10–14. doi: 10.1016/j.socscimed.2013.05.005.

Platt, Jodyn, and Sharon Kardia. 2015. "Public Trust in Health Information Sharing: Implications for Biobanking and Electronic Health Record Systems." *Journal of Personalized Medicine* 5(1):3–21. doi: 10.3390/jpm5010003.

Richardson, Jordan P., Cambray Smith, Susan Curtis, Sara Watson, Xuan Zhu, Barbara Barry, and Richard R. Sharp. 2021. "Patient Apprehensions about the Use of Artificial Intelligence in Healthcare." *Npj Digital Medicine* 4(1):1–6. doi: 10.1038/s41746-021-00509-1.

Richardson, Rashida, Jason Schultz, and Kate Crawford. 2019. *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice. SSRN Scholarly Paper*. ID 3333423. Rochester, NY: Social Science Research Network.

Schoorman, F. David, Roger C. Mayer, and James H. Davis. 2007. "An Integrative Model of Organizational Trust: Past, Present, and Future." *The Academy of Management Review* 32(2):344–54. doi: 10.2307/20159304.

Selbst, Andrew D., Danah Boyd, Sorelle A. Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. "Fairness and Abstraction in Sociotechnical Systems." Pp. 59–68 in *Proceedings of the Conference on Fairness, Accountability, and Transparency*. Atlanta GA USA: ACM.

Shilo, Smadar, Hagai Rossman, and Eran Segal. 2020. "Axes of a Revolution: Challenges and Promises of Big Data in Healthcare." *Nature Medicine* 26(1):29–38. doi: 10.1038/s41591-019-0727-5.

Tan, Marissa, Elham Hatef, Delaram Taghipour, Kinjel Vyas, Hadi Kharrazi, Laura Gottlieb, and Jonathan Weiner. 2020. "Including Social and Behavioral Determinants in Predictive Models: Trends, Challenges, and Opportunities." *JMIR Medical Informatics* 8(9):e18084. doi: 10.2196/18084.

Timmermans, Stefan, and Marc Berg. 2003. "The Practice of Medical Technology." *Sociology of Health & Illness* 25(3):97–114. doi: 10.1111/1467-9566.00342.

Timmermans, Stefan, and Mara Buchbinder. 2010. "Patients-in-Waiting: Living between Sickness and Health in the Genomics Era." *Journal of Health and Social Behavior* 51(4):408–23. doi: 10.1177/0022146510386794.

Varcoe, Colleen, Annette J. Browne, Sabrina Wong, and Victoria L. Smye. 2009. "Harms and Benefits: Collecting Ethnicity Data in a Clinical Context." *Social Science & Medicine* 68(9):1659–66. doi: 10.1016/j.socscimed.2009.02.034.

Vyas, Darshali A., Leo G. Eisenstein, and David S. Jones. 2020. "Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms." *New England Journal of Medicine* 383(9):874–82. doi: 10.1056/NEJMms2004740.

Waljee, Akbar K., Peter D. R. Higgins, and Amit G. Singal. 2014. "A Primer on Predictive Models." *Clinical and Translational Gastroenterology* 5(1):e44. doi: 10.1038/ctg.2013.19.

Wang, Fei, and Anita Preininger. 2019. "AI in Health: State of the Art, Challenges, and Future Directions." *Yearbook of Medical Informatics* 28(1):16–26. doi: 10.1055/s-0039-1677908.

Wong, Andrew, Erkin Otles, John P. Donnelly, Andrew Krumm, Jeffrey McCullough, Olivia DeTroyer-Cooley, Justin Pestrue, Marie Phillips, Judy Konye, Carleen Penoza, Muhammad Ghous, and Karandeep Singh. 2021. "External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients." *JAMA Internal Medicine*. doi: 10.1001/jamainternmed.2021.2626.

Wynants, Laure, Maarten van Smeden, David J. McLernon, Dirk Timmerman, Ewout W. Steyerberg, Ben Van Calster, and on behalf of the Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. 2019. "Three Myths about Risk Thresholds for Prediction Models." *BMC Medicine* 17(1):192. doi: 10.1186/s12916-019-1425-3.

Zhang, Simone, Rebecca A. Johnson, John Novembre, Edward Freeland, and Dalton Conley. 2021. "Public Attitudes toward Genetic Risk Scoring in Medicine and Beyond." *Social Science & Medicine* 274:113796. doi: 10.1016/j.socscimed.2021.113796.

Ziebland, Sue, Emma Hyde, and John Powell. 2021. "Power, Paradox and Pessimism: On the Unintended Consequences of Digital Health Technologies in Primary Care." *Social Science & Medicine* 289:114419. doi: 10.1016/j.socscimed.2021.114419.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. First edition. New York: PublicAffairs.

Chapter 3 Limitations of the Regulatory Distinction between Clinical and Administrative Predictions in Healthcare

BACKGROUND

Software, including predictive models and artificial intelligence and machine learning (AI/ML) tools, is widely used across the healthcare system with applications in clinical settings ranging from primary care to inpatient care, as well as administrative offices. Federal regulation of this software in healthcare is built on the foundation of existing approaches to medical device regulation (FDA 2019a; FDA 2022). Software receives varying degrees of oversight from the Food and Drug Administration (FDA) depending on its alignment with the definition of a medical device. If a software application addresses certain conditions in specific contexts, it can be considered a medical device and receive oversight. This device framework does not translate particularly well to contemporary developments in software tools like predictive models that use AI/ML methods, through which software can change over time (Wu et al. 2021). The concept of software as a static object that can be overseen like a device and managed by clinicians is increasingly unresponsive to the realities of software use in healthcare. Furthermore, the medical device framework constrains the scope of regulatory oversight to explicitly clinical software applications that can be defined as medical devices, leaving administrative or managerial software in healthcare unregulated.

The current FDA regulatory approach to Software as a Medical Device (SaMD) focuses on risk to health and safety and is concentrated solely on clinical applications and diagnosis (FDA 2019b). However, there is a growing body of literature indicating that administrative

applications also present risks to patients (Murray, Wachter, and Cucina 2020; Obermeyer et al. 2019). Whether by building barriers to care or restricting patient access to beneficial programs and resources, software applied to administrative or managerial operations can harm patients. These types of harm are unaddressed by current policy. In the context of a generally lax regulatory approach at the federal level (Amarasingham et al. 2014), it is possible that the exclusion of administrative software applications has additional negative implications for patient safety and quality care.

Below, the policy distinction between clinical and administrative applications is examined using the case of predictive models in healthcare. These models are discussed within the frameworks of FDA's SaMD classifications, FDA's final guidance on clinical decision support, and the 21st Century Cures Act. This analysis highlights policy gaps related to predictive models and related methods in healthcare, including the lack of empirical information about public attitudes that would inform how future policy might become responsive to the public's concerns. This discussion is followed by an overview of implications of the current policy environment for patients and health inequities. This includes considerations of the evidence needed on public perspectives and values.

Predictive model applications

Predictive technologies are often applied in the healthcare system for both clinical decision-making and analysis of administrative data (Apathy, Holmgren, and Adler-Milstein 2021; CHIME 2021). These models use large amounts of historical patient data to make predictions or produce risk scores about a variety of diagnoses, outcomes, and behaviors (Waljee, Higgins, and Singal 2014). Predictive models are used to anticipate the onset of sepsis,

COVID-19 deterioration, cardiac events, and kidney disease progression, for example (Nemati et al. 2018; Niederer, Lumens, and Trayanova 2019; Singh et al. 2020; Wong et al. 2021). Empirical analysis of predictive models is typically focused on these types of clinical applications (Beaulieu-Jones et al. 2021; Brenner et al. 2016; Middleton, Sittig, and Wright 2016; Sendak et al. 2019).

However, predictive technologies are also applied to administrative or managerial functions. They are used to predict health service utilization, staffing needs, and missed appointments (Ding et al. 2018; Futoma, Morris, and Lucas 2015). These administrative models may be designed and implemented to lower costs and target resource utilization (Murray et al. 2020). Health systems use models that predict missed appointments, for example, to decrease lost revenue associated with unused appointment slots. In this case, if a patient's risk of missing an appointment is predicted to be high, the health system may double-book their appointment. There are multiple potential negative consequences related to this kind of administrative model, including the barriers it creates for patients who struggle to access healthcare. Specifically, this model application can disadvantage marginalized patients and exacerbate socioeconomic and racial inequities in care. Despite these issues and the other potential negative impacts of administrative prediction on health inequities, these models remain unregulated (FDA 2019b; FDA 2019a; FDA 2022).

Policy distinctions between clinical and administrative models

The federal regulatory framework for software is built on policy that draws a clear distinction between clinical and administrative applications with the definition of a medical device. The 21st Century Cures Act defines this distinction and FDA's regulation and guidance

reflect its salience. These policies and regulations indicate that when software fulfill clinical functions like a medical device, they should be regulated. Software that is not defined as a medical device is excluded from the regulation and oversight defined in these policies and frameworks.

Healthcare technology has long been regulated to the extent that it exists as part of physical devices, like glucometers or pulse oximeters. The FDA has specific, well-established requirements for these devices that include registration, premarket notification, quality regulations, labeling requirements, and reporting procedures (FDA 2020a). As the use of non-device technologies and software expanded, FDA used the International Medical Device Regulators Forum (IMDRF) approach to software regulation. This approach, called Software as Medical Device (SaMD) (FDA 2020b), uses the regulation of medical devices as a template for the complex task of regulating multiple types of medical software.

Software, including predictive models, is categorized by the SaMD framework according to 1) the significance of information provided by the software and 2) the severity of the health condition being addressed (Figure 3.1). Models that fall on the clinical and critical end of the spectrum (e.g., levels III and IV) are most likely to be regulated. A model used to treat or diagnose a patient with a critical condition would receive the most oversight from FDA. On the other end of the spectrum, models that inform clinical management, like nutrition planning, are of least concern to regulators. This category falls under enforcement discretion, which means the FDA does not enforce compliance (FDA 2020b; IMDRF Software as a Medical Device Working Group 2014).

Figure 3.1 FDA Software as Medical Device classification

State of healthcare situation or condition	Significance of information provided by SaMD to healthcare decision		
	Treat or diagnose	Drive clinical management	Inform clinical management
Critical	IV	III	II
Serious	III	II	I
Non-serious	II	I	I

Source: IMDRF Software as a Medical Device Working Group. 2014. *"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations*. Higher numbers indicate a greater degree of oversight, with level IV being the greatest amount of FDA oversight.

The Cures Act forms the basis of this distinction between clinical and administrative functions under the paradigm of device regulation. Here, the definition of a medical device excludes tools focused on the practice of healthcare, like digital tools used for resource management or quality improvement purposes. The Cures Act explicitly excludes models used for healthcare management from the definition of medical devices unless FDA sees a likelihood of serious health consequences related to their use (Price II, Sachs, and Eisenberg 2021). Given FDA's assessment of risk described above, administrative models receive no oversight.

FDA recently released its final guidance on clinical decision support software (FDA 2022). While some language in the final guidance is more specific than the draft guidance, there has been no change to the lack of oversight of administrative or managerial software. There are relatively few applications that would receive FDA oversight under the final guidance and none of these would include administrative models.

In the absence of comprehensive federal regulation, alternatives have been proposed. Industry "self-governance" is one of these approaches. In this proposed approach, the industry of healthcare software would develop best practices and self-monitor for patient safety, quality, and accuracy (Roski et al. 2021). Multiple efforts are underway to establish fairness guidelines and

measures of bias to encourage the industry to adopt standards (Bedoya et al. 2022; Smith 2020). However, there is no consensus on whether or how administrative applications would be included. Despite routine calls for the design of patient-centered systems, there is also a lack of robust evidence on patient perspectives, trust in these tools, or concerns about their effects (Agency for Healthcare Research and Quality 2021).

Public values and regulation

Equity and patient-centeredness have received some attention as important aspects of health IT and policy. The American Medical Informatics Association has made patient-centeredness a key theme of their annual symposia and the Patient Centered Outcomes Research Institute (PCORI) continues to promote evidence-based approaches to IT-driven care. The Agency for Healthcare Research and Quality (AHRQ) has also increased its attention to public perspectives on data use in the last six years through funding and public reports (Agency for Healthcare Research and Quality 2021). Some of this work is motivated by a realization that previous data collection and use in healthcare has violated patient trust and negatively impacted patients (Brandon, Isaac, and LaVeist 2005; Scharff et al. 2010).

Patients as the intended beneficiaries of healthcare are, by definition, stakeholders in predictive modeling and informatics more broadly. Efforts to 1) incorporate patient perspectives, 2) build a patient-centered healthcare system, and 3) engender public trust in data-driven care have been proliferating for years. In order to pursue evidence-based patient-centeredness, empirical data analysis is fundamentally necessary. The Office of Science and Technology Policy (OSTP) specifically highlights the urgent need of public engagement and policy that

responds to public concerns, needs, and priorities (The Office of Science and Technology Policy 2022).

Previous examples of data use in healthcare demonstrate that unresponsive policy and violations of trust have negative consequences for patient trust and engagement (Wachter and Cassel 2020). For example, the revelation that Ascension Health was sharing patient data with Google led to public outcry and mistrust. Similarly, the realization that patient data was being commercialized by Sloan Kettering researchers through data sharing with Paige.AI led to legal review and system-wide revisions of conflict of interest policies after significant public attention (Dyer 2019). Although these commercial partnerships and uses of patient data were not illegal, they caused concern about patient trust and privacy. When patient data is used without patient knowledge, it can undermine trust and engagement with the healthcare system.

Patients' experiences in the healthcare system may also lead to withholding information from providers (Nong et al. 2022). When patient data is systematically missing due to systematic violations of trust or discrimination, the data available on those patients is of poorer quality and can produce information technologies that do not perform as well for them. This dynamic of poor data and resulting lower performance of the tools built with that data has been termed "exclusion cycles" (Bracic, Callier, and Price 2022). In these exclusion cycles patient trust is violated and they may withhold their data. The data quality in their medical record is then comparatively poorer, resulting in lower quality predictive outputs. Because predictive models rely on high volumes of patient data, these exclusion cycles can have increasingly negative impacts on patients who are already excluded, discriminated against, and marginalized.

Empirical evidence of public perspectives on the use of prediction in healthcare is necessary to inform a patient-centered predictive system in healthcare. This work responds to federal calls for evidence on how the public and patients understand and respond to the use of these kinds of digital tools (US Department of Health and Human Services 2021).

OBJECTIVE

This chapter empirically analyzes how the public perceives predictive modeling in healthcare. It tests whether the policy distinction drawn between clinical and administrative prediction is reflective of the US adult population's comfort and discomfort with these models. It also analyzes predictors of comfort with these models in order to empirically identify systematic variation.

1. Does public comfort with predictive models reflect the regulatory emphasis on clinical software applications? Does public comfort differ between clinical and administrative applications?

2. What are the individual-level predictors of comfort with specific predictive models?

METHODS

This analysis uses data from a cross-sectional national survey of US adults who can speak English. The survey was fielded with the National Opinion Research Center's (NORC) AmeriSpeak Panel from mid-November to December 2021. A total of 1,541 participants completed the survey. This included oversamples of African American respondents, Hispanic respondents, and respondents earning less than 200% of the federal poverty level. NORC calculated poststratification survey weights based on demographics and Census division from the

Current Population Survey. Weights for non-response were also calculated. The survey was pre-tested using MTurk (n=550) and pilot tested with a sample of AmeriSpeak panel participants (n=150). After excluding observations with missing data for the variables of interest, the final analytic sample was 1,488.

Simple, accessible definitions of key terms like health system and healthcare provider were provided to participants. The full definitions were also available to participants as hover-over text each time the term was mentioned in the text of the survey (see Appendix B for these definitions). Participants also viewed a short explanatory video describing how health information is used and shared in the healthcare system. As described elsewhere, the video has been reviewed by experts in the field, tested, and used in multiple previous surveys (Amara et al. 2022; Spector-Bagdady et al. 2022; Trinidad, Platt, and Kardia 2020).

Predictive models were defined and described in a short paragraph (Flesch-Kincaid score 8.7) immediately preceding survey questions about predictive models. It included popular examples of predictive technologies external to and within the healthcare system.

Measures

The outcome measure for this analysis is comfort with specific applications of predictive models, spanning the SaMD categories (Table 3.1). Respondents indicated their comfort level on a four-point scale (1=Not comfortable to 4=Very comfortable) with each of the 6 predictive model applications listed on the survey. Model applications were displayed in random order.

Table 3.1 Measures of comfort with predictive model applications

	How comfortable are you with each type of predictive model?
1	Models that predict which patients might develop colon cancer
2	Models that predict which patients might miss their appointments
3	Models used to diagnose stroke in an emergency
4	Models that predict which patients might not pay their medical bills
5	Models used to determine if a patient is eligible for a kidney transplant
6	Models that predict which patients will develop sepsis (a serious blood infection)

Independent variables include self-reported age in years, a binary measure of sex (male, female), and race/ethnicity (Multiracial, Hispanic, non-Hispanic Asian, Black, white, and other). Respondents reported their annual household income and personal education level (no high school diploma, high school or equivalent, some college, BA or more). Health-related independent variables include health insurance status (insured/uninsured), healthcare utilization in the past 12 months, self-reported health status (poor to very good), previous cancer diagnosis, and experiences of discrimination in the healthcare system (yes/no).

The independent variables of interest include patient experiences of discrimination in the healthcare system. Used in multiple previous surveys, the survey measure is adapted from the Williams Major and Everyday Discrimination Measures (Krieger et al. 2005; Nong et al. 2020). To account for perceptions of the health system, a measure of perceived clarity of privacy policies was also included as an independent variable of interest. For this measure, participants used a four-point scale (1=Not true, 4=Very true) to indicate their agreement with the statement "*The privacy policies of my healthcare system are clear to me*".

Analysis

　　To answer research question one, correlations between comfort with all six predictive model applications were calculated. Measures of internal consistency (Cronbach's alpha) were also calculated to identify the degree to which participant responses indicated that clinical and administrative categories of predictive models were valid. Paired t-tests were used to analyze whether there were differences in mean comfort between the predictive model categories. Comfort with the two categories of predictive models (clinical and administrative) were then used to create composite measures for additional analysis.

　　To answer research question two, bivariable and multivariable logistic regressions were conducted. Bivariable logistic regressions of comfort with each model category were conducted with the independent variables of interest and covariates described above and listed below in Table 3.2.

　　Six multivariable logistic regression models were run to identify individual-level predictors of comfort with each predictive model. Multivariable logistic regression models were also run on the composite measures of comfort with clinical and administrative models. Independent variables of interest were experiences of discrimination and clarity of system privacy policies, described above. Covariates included self-reported demographics (sex, race/ethnicity, age), education, income, healthcare utilization, self-reported health status, insurance, and previous cancer diagnosis.

　　In the results section below, descriptive statistics for the sample are presented along with the distributions of comfort with each predictive model application. This is followed by the full results for research question one, including the correlation matrix of comfort with predictive model applications, Cronbach's alpha for internal consistency of model types, and t-tests of

differences between clinical and administrative model applications. Results related to research question two include the regression analyses of individual-level predictors of comfort with each type of predictive model and the composite measure of comfort with clinical predictions.

Ethical considerations

This study was determined to be exempt by the University of Michigan Institutional Review Board. Participants were compensated for their time according to standard NORC remuneration policies.

RESULTS

The sample was 49.7% female. Racial and ethnic representation was generally reflective of the US population. Most respondents had either some college education (46.4%) or a bachelor's degree or more education (33.8%). Half of respondents reported at least $50,000 in annual household income (50.1%) and most reported having health insurance (93.2%).

Table 3.2 Sample demographics (n=1,488)

Measure	n	%
Sex		
Female	740	49.7
Male	748	50.3
Age		
18-29	86	5.8
30-44	422	28.4
45-59	380	25.5
60+	600	40.3
Race/ethnicity		
White	915	61.7
Hispanic	269	18.1
Black	205	13.8
Multiracial and other	60	4.0

Asian	39	2.6
Education		
Less than high school	39	2.6
High school	256	17.2
Some college	690	46.4
BA or more	503	33.8
Annual household income		
<$50,000	743	49.9
At least $50,000	745	50.1
Health insurance coverage		
No	101	6.8
Yes	1,387	93.2
Self-reported health		
Poor to fair	353	23.7
Good	633	42.5
Very good to excellent	502	33.7
Last healthcare visit		
Longer than 1 year	235	15.8
Within past year	1,253	84.2
Experienced discrimination		
No	1,166	78.4
Yes	322	21.6
Transparent privacy policies		
No	873	58.7
Yes	615	41.3
Cancer diagnosis (self)		
No	1,299	87.3
Yes	189	12.7
Family history of cancer		
No	744	50.0
Yes	744	50.0

Research Question 1

Does public comfort with predictive model applications reflect the regulatory emphasis

on clinical software applications?

As depicted in Figure 3.2, discomfort with prediction of bill payment and missed

appointments was very high (78.4% and 63.4%, respectively). This contrasts participants'

reported comfort with clinical predictive models. Comfort was highest with models used to diagnose stroke in an emergency (55.8%). This indicates least public concern about the predictive model with the largest role in time-sensitive diagnosis. This type of model is most likely to be regulated by FDA because of its significance in clinical decision-making and the emergent nature of the application (FDA 2022). Comfort with other clinical predictions was around 50%, with 50.6% of participants reporting comfort with predicting sepsis and 52.3% predicting colon cancer. Comfort with prediction for kidney transplant eligibility was 46.5%.

Figure 3.2 Comfort with predictive model applications

Distributions of Comfort with Predictive Model Applications
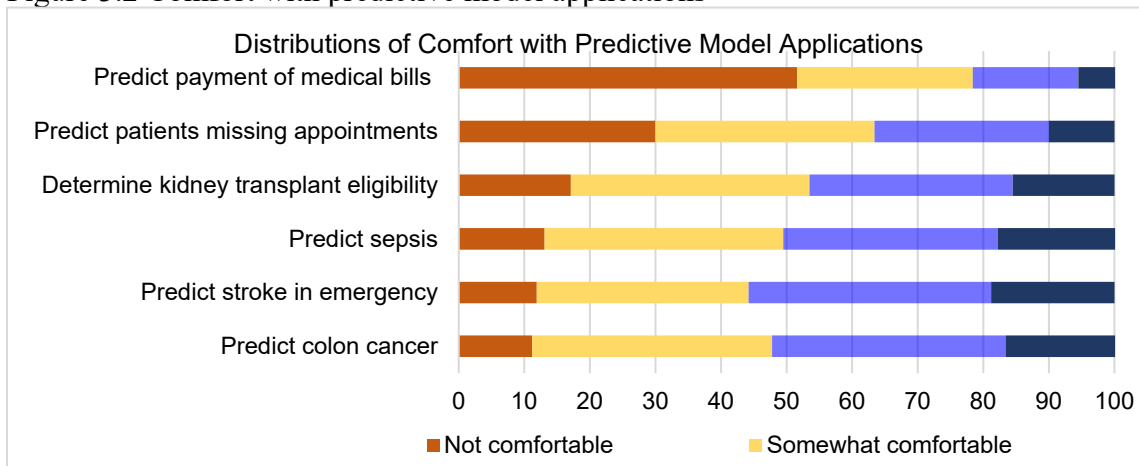
Correlations between comfort with the predictive model types were calculated. The correlation matrix is presented in Table 3.3. Correlations are high between the clinical model types (r=0.59-0.73). Administrative model types are also correlated (r=0.55). Cronbach's alpha was calculated to determine whether internal consistency of these measures was high enough to construct composite measures. The Cronbach's alpha was 0.89 for clinical model types (colon cancer, kidney transplant, sepsis, and stroke) and 0.71 for the administrative model types

(payment and missed appointments). Both scores are above the acceptability threshold of 0.65

and indicate that composite measures are internally consistent.

Table 3.3 Correlation matrix of comfort with each predictive model

	Colon cancer	Kidney transplant	Sepsis	Stroke	Payment	Missed appointment
Colon cancer	1					
Kidney transplant	0.586	1				
Sepsis	0.697	0.711	1			
Stroke	0.734	0.610	0.677	1		
Payment	0.199	0.266	0.239	0.128	1	
Missed appointment	0.360	0.308	0.327	0.270	0.547	1

Mean comfort with clinical model types was 2.55. For administrative models, mean

comfort was 1.96. A paired t-test indicates that these means are significantly different from each

other (p<0.001), demonstrating that public comfort is different for clinical models compared to

administrative ones. However, public discomfort is highest for the model type that will receive

no oversight according to current regulation.

Research Question 2

What are individual-level predictors of comfort with different predictive model

applications?

To identify individual-level predictors of comfort, bivariable and multivariable logistic

regressions were conducted with 1) binary measures of comfort with each individual model type,

and 2) binary composite measures of comfort with administrative and clinical models.

Dichotomized measures of comfort were created where responses of not comfortable and

somewhat comfortable with a model were equal to zero. Responses of fairly and very

comfortable were coded as 1. Bivariable and multivariable logistic regressions were run with these binary measures.

Composite measures of comfort with predictive models were also created. For each model category (clinical and administrative), the binary measures of comfort with each model were used. These composite measures were top coded so that participants who indicated comfort with all predictive models that comprise the category were coded as 1. All other participants were coded as zero. In this way, participants who indicated comfort with all four clinical models (colon cancer, kidney transplant, sepsis, stroke) are considered comfortable with the composite measure. The same approach was used for the administrative models (payments and missed appointments).

Comfort with clinical and administrative models was regressed on the independent variables of interest and all covariates (Table 3.4). These analyses identified that both independent variables of interest (experiences of discrimination and transparent privacy policies) were statistically significant predictors of comfort with both clinical and administrative models. Sex, age, and health insurance coverage were statistically significantly associated with comfort with clinical models but not with administrative models.

Table 3.4 Bivariable logistic regressions of comfort with predictive models by type

	Clinical models		Administrative models	
	Odds Ratio	p-value	Odds Ratio	p-value
Experienced discrimination				
No	ref.		ref.	
Yes	0.75	0.04*	0.44	<0.001***
Transparent privacy policy				
No	ref.		ref.	
Yes	2.27	<0.001***	1.7	<0.001***
Sex				
Male	ref.		ref.	
Female	0.79	0.03*	0.78	0.08
Age				
18-29	ref.		ref.	
30-44	0.75	0.24	0.92	0.78
45-59	0.55	0.02*	0.6	0.1
60+	0.78	0.3	0.88	0.66
Race/ethnicity				
White	ref.		ref.	
Black	0.82	0.26	0.69	0.11
Hispanic	1.03	0.86	1.17	0.39
Asian	1.02	0.96	1.96	0.07
Other	0.62	0.13	1.12	0.75
Education				
Less than high school	ref.		ref.	
High school	1.33	0.48	1.3	0.58
Some college	1.48	0.32	0.95	0.92
BA or more	1.91	0.1	1.26	0.61
Annual household income				
<$50,000	ref.		ref.	
At least $50,000	1.19	0.12	0.97	0.74
Health insurance coverage				
No	ref.		ref.	
Yes	1.75	0.02*	1.08	0.78
Self-reported health				
Poor to fair	ref.		ref.	
Good	0.1	0.97	0.93	0.7
Very good to excellent	1.01	0.94	1.19	0.35
Last healthcare visit				
Longer than one year ago	ref.		ref.	
Within past year	1.06	0.78	1.29	0.21
Cancer diagnosis				
No	ref.		ref.	
Yes	1.32	0.08	1.44	0.06
Family history of cancer				
No	ref.		ref.	
Yes	1.11	0.35	0.99	0.95

Multivariable logistic regressions were run for each individual predictive model application (Table 3.5). The perceived clarity of health system privacy policies was a significant positive predictor of comfort with every individual predictive model application. Experience of discrimination was a significant negative predictor of comfort with the administrative models and kidney transplant eligibility. Insurance status was significantly and positively associated with comfort with sepsis, stroke, and colon cancer predictions.

Table 3.5 Multivariable logistic regression of comfort with each predictive model application

	OR, p-value										
	Stroke		Colon cancer		Sepsis		Kidney transplant		Missed appointment		Bill payment	
Experienced discrimination												
No	ref.		ref.		ref.		ref.		ref.		ref.	
Yes	0.87	0.31	0.98	0.87	0.88	0.38	0.71	0.01*	0.64	0.002**	0.57	0.002**
Transparent privacy policy												
No	ref.		ref.		ref.		ref.		ref.		ref.	
Yes	1.87	<0.001***	2.4	<0.001***	2.21	<0.001***	2.0	<0.001***	2.0	<0.001***	1.76	<0.001***
Sex												
Male	ref.		ref.		ref.		ref.		ref.		ref.	
Female	0.82	0.08	0.81	0.06	0.78	0.03*	0.9	0.31	0.81	0.07	0.81	0.12
Age												
18-29	ref.		ref.		ref.		ref.		ref.		ref.	
30-44	0.89	0.62	0.64	0.08	0.56	0.02*	0.67	0.1	1.33	0.29	0.98	0.93
45-59	0.6	0.04*	0.4	<0.001***	0.35	<0.001***	0.51	0.008**	1.01	0.97	0.64	0.14
60+	0.93	0.77	0.56	0.03*	0.48	0.005**	0.58	0.03*	1.28	0.36	0.73	0.29
Race/ethnicity												
White	ref.		ref.		ref.		ref.		ref.		ref.	
Black	0.92	0.6	1.07	0.68	1.05	0.78	0.99	0.93	0.85	0.36	1.02	0.93
Hispanic	1.07	0.64	1.07	0.65	0.95	0.75	1.02	0.89	1.08	0.62	1.21	0.29
Asian	1.45	0.29	1.23	0.55	1.2	0.6	1.18	0.63	1.62	0.17	3.04	0.002**
Other	0.91	0.74	0.87	0.63	0.55	0.04*	0.83	0.52	0.85	0.6	1.2	0.58
Education												

Less than high school	ref.		ref.		ref.		ref.		ref.		ref.	
High school	0.61	0.18	1.08	0.83	0.85	0.65	0.89	0.75	0.93	0.84	1.07	0.87
Some college	0.66	0.24	1.14	0.71	0.94	0.87	0.92	0.81	0.81	0.55	0.85	0.67
BA or more	0.9	0.77	1.51	0.25	1.25	0.53	1.19	0.61	1.0	0.99	0.86	0.71
Annual household income												
<$50,000	ref.		ref.		ref.		ref.		ref.		ref.	
At least $50,000	1.14	0.26	1.25	0.06	1.3	0.03*	1.11	0.34	1.12	0.31	0.97	0.8
Health insurance												
No	ref.		ref.		ref.		ref.		ref.		ref.	
Yes	1.7	0.02*	2.07	0.002**	1.67	0.03*	1.33	0.21	0.84	0.44	1.06	0.82
Self-reported health												
Poor to fair	ref.		ref.		ref.		ref.		ref.		ref.	
Good	0.86	0.27	0.92	0.56	0.98	0.86	1.02	0.88	1.14	0.38	0.87	0.42
Very good to excellent	0.81	0.16	0.84	0.25	0.95	0.73	0.92	0.57	1.14	0.41	1.01	0.97
Last healthcare visit												
Longer than one year	ref.		ref.		ref.		ref.		ref.		ref.	
Within past year	1.12	0.48	2.4	<0.001***	1.05	0.77	1.15	0.37	1.29	0.05	1.23	0.29
Cancer diagnosis												
No	ref.		ref.		ref.		ref.		ref.		ref.	
Yes	1.41	0.05	1.35	0.08	1.27	0.16	1.17	0.34	1.25	0.2	1.56	0.02*
Family history of cancer												
No	ref.		ref.		ref.		ref.		ref.		ref.	
Yes	1.19	0.12	1.25	0.04*	1.31	0.01*	1.11	0.33	1.02	0.86	0.92	0.51

Multivariable logistic regressions were also run with the composite measures of comfort as the dependent variables (Table 3.6). In these multivariable regressions, experiences of discrimination were a negative predictor of comfort with administrative models (OR 0.48, p=0.001). Transparent privacy policies remained statistically significant positive predictors of comfort with both clinical (OR 2.35, p<0.001) and administrative models (OR 1.68, p<0.001). As was observed in the bivariable logistic regressions, sex (female (ref. male), OR 0.74, p=0.01) and age (45-59 (ref. 138-29), OR 0.51, p=0.01) were significantly negatively predictive of comfort with clinical models.

Table 3.6 Multivariable logistic regression of predictive model by type

	OR, p-value			
	Clinical models		Administrative models	
Experienced discrimination				
No	ref.		ref.	
Yes	0.85	0.27	0.48	0.001**
Transparent privacy policy				
No	ref.		ref.	
Yes	2.35	<0.001***	1.68	<0.001***
Sex				
Male	ref.		ref.	
Female	0.74	0.01*	0.77	0.07
Age				
18-29	ref.		ref.	
30-44	0.75	0.25	0.92	0.8
45-59	0.51	0.01*	0.58	0.09
60+	0.67	0.12	0.73	0.32
Race/ethnicity				
White	ref.		ref.	
Black	0.88	0.47	0.76	0.25
Hispanic	1.16	0.36	1.27	0.22
Asian	1.01	0.98	2.05	0.07
Other	0.68	0.23	1.27	0.51
Education				
Less than high school	ref.		ref.	
High school	1.55	0.29	1.48	0.41
Some college	1.78	0.15	1.15	0.76
BA or more	2.26	0.05	1.43	0.45

Annual household income				
<$50,000	ref.		ref.	
At least $50,000	1.15	0.26	0.93	0.65
Health insurance coverage				
No	ref.		ref.	
Yes	1.54	0.1	0.96	0.88
Self-reported health				
Poor to fair	ref.		ref.	
Good	0.9	0.48	0.87	0.46
Very good to excellent	0.84	0.29	1.01	0.94
Last healthcare visit				
Longer than one year ago	ref.		ref.	
Within past year	0.95	0.76	1.38	0.14
Cancer diagnosis				
No	ref.		ref.	
Yes	1.25	0.2	1.42	0.09
Family history of cancer				
No	ref.		ref.	
Yes	1.09	0.45	1.01	0.96

LIMITATIONS

There are limitations to this study. First, it does not include a gender variable but is restricted to self-reported sex. Additional work on a multitude of sex and gender identities in relation to comfort with predictive modeling will be an important next step. Second, Native American and Alaskan Native respondents are not identified in this sample because the multiple-choice options for racial and ethnic identities did not include these options. This is a limitation in the data's representativeness and future work should specifically ensure that Native American and Alaskan Native respondents are 1) specifically identified and 2) adequately represented in analysis of comfort with predictive modeling.

The data analyzed here is cross-sectional. This limits the inferences that can be made about predictors of comfort with prediction and how this might change over time. For example, it is not possible to identify longitudinal relationships between experiences of discrimination and perceptions of data use for prediction. Additionally, measures of exposure to structural racism

are not available in this dataset. Because of the importance of structural racism in relation to public trust, it is likely to inform various aspects of how the public feels about predictive modeling. Although experiences of discrimination are relevant, they capture only interpersonal discrimination rather than structural inequities that inform how patients experience the collection healthcare system. Future work will explore additional structural inequities and measures of structural racism in relation to comfort with prediction. Qualitative work will also provide a more in-depth understanding of how and why the public feels the way they do about prediction in health broadly, and the use of individual data categories specifically.

DISCUSSION

These results identify that public discomfort with administrative models is significantly higher than discomfort with clinical models. This suggests that current policy is misaligned with public perceptions and concerns. It is possible that the public is more comfortable with clinical prediction because they perceive or assume that clinicians' expertise protects them from potential harms (Richardson et al. 2021). However, in contrast to the expertise and professional boundaries around clinical knowledge and predictive models, patients may feel more qualified to express discomfort or concern related to the use of administrative models. For example, a patient may recognize they do not have the expertise to identify equity or bias issues in relation to an oncology predictive model. Their experience or expertise may not be in oncology. They do, however, have experience with their hospital's billing and scheduling processes. Their experience trying to schedule an appointment with a specialist or attempting to contact a representative about a billing error may make potential risks or discomfort with these models more salient to them (Anderson, Camacho, and Balkrishnan 2007). Patients have also indicated

that they expect clinicians to understand and manage AI (Richardson et al. 2021). For clinical predictive models, this assumption may be engendering comfort whereas administrative predictions are removed from clinician engagement and the assumed protection that offers.

This analysis identifies that participants were least comfortable with the prediction of bill payment. Patients tend to be uncomfortable with commercial activities in healthcare (Trinidad et al. 2020). Aggressive debt collection practices and disclosure of commercialization of health data have elicited strong negative reactions among patients and the public (O'Toole, Arbelaez, and Lawrence 2004; Thomas 2018). The expectation that patients behave as consumers can contradict aspects of cultural expectations around the role of medicine (Khullar 2019). Thus, the observation of discomfort with prediction of payment in this analysis is expected.

Predictors of comfort differ for clinical and administrative prediction models with one exception. Perception that one's health system has transparent privacy policies is positively predictive of comfort with both model categories. Experiences of discrimination are predictive of comfort with administrative models only. Prior literature indicates that cancer patients, or those with a family history of cancer, may be more willing to share their health data and exhibit more enthusiasm for tools like precision medicine (Grande et al. 2015). However, this analysis does not identify that cancer patients or those with a family history of cancer are more comfortable with the use of prediction. With the exception of sex and age for comfort with clinical predictions, self-reported demographic characteristics are not significantly associated with comfort.

CONCLUSION

This analysis provides empirical evidence of misalignment between public conceptualization of predictive model applications and the current regulatory framework, with implications for federal policy as well as health system practice. Because the applications patients perceive as high-risk fall outside of FDA's oversight framework, policies that respond to these concerns at the health system and federal policy levels will be important as predictive tools proliferate.

Given federal prioritization of engendering trust and confidence in artificial intelligence and similar methodologies among the public, the results of this study are particularly important. The narrow regulatory purview of FDA has multiple implications for patient access to care, equity, and quality that are not overtly clinical but stand to impact patients across the healthcare system. These issues of access, equity, and quality are inextricably linked to the clinical care patients receive, although they may not be directly used to diagnose a patient. Regulatory frameworks that seriously engage with administrative or managerial predictive modeling are needed to make the system of prediction more patient-centered and equitable.

References

Agency for Healthcare Research and Quality. 2021. *CDS Innovation Collaborative*.

Amara, Philip S., Jodyn E. Platt, Minakshi Raj, and Paige Nong. 2022. "Learning about COVID-19: Sources of Information, Public Trust, and Contact Tracing during the Pandemic." *BMC Public Health* 22(1):1348. doi: 10.1186/s12889-022-13731-7.

Amarasingham, Ruben, Rachel Patzer, Marco Huesch, Nam Nguyen, and Bin Xie. 2014. "Implementing Electronic Health Care Predictive Analytics: Considerations And Challenges." *Health Affairs (Project Hope)* 33:1148–54. doi: 10.1377/hlthaff.2014.0352.

Anderson, Roger T., Fabian T. Camacho, and Rajesh Balkrishnan. 2007. "Willing to Wait?: The Influence of Patient Wait Time on Satisfaction with Primary Care." *BMC Health Services Research* 7(1):31. doi: 10.1186/1472-6963-7-31.

Apathy, Nate C., A. Jay Holmgren, and Julia Adler-Milstein. 2021. "A Decade Post-HITECH: Critical Access Hospitals Have Electronic Health Records but Struggle to Keep up with Other Advanced Functions." *Journal of the American Medical Informatics Association* ocab102. doi: 10.1093/jamia/ocab102.

Beaulieu-Jones, Brett K., William Yuan, Gabriel A. Brat, Andrew L. Beam, Griffin Weber, Marshall Ruffin, and Isaac S. Kohane. 2021. "Machine Learning for Patient Risk Stratification: Standing on, or Looking over, the Shoulders of Clinicians?" *Npj Digital Medicine* 4(1):1–6. doi: 10.1038/s41746-021-00426-3.

Bedoya, Armando D., Nicoleta J. Economou-Zavlanos, Benjamin A. Goldstein, Allison Young, J. Eric Jelovsek, Cara O'Brien, Amanda B. Parrish, Scott Elengold, Kay Lytle, Suresh Balu, Erich Huang, Eric G. Poon, and Michael J. Pencina. 2022. "A Framework for the Oversight and Local Deployment of Safe and High-Quality Prediction Models." *Journal of the American Medical Informatics Association* 29(9):1631–36. doi: 10.1093/jamia/ocac078.

Bracic, Ana, Shawneequa Callier, and W. Nicholson Price. 2022. "Exclusion Cycles: Reinforcing Disparities in Medicine." *Science* 377(6611).

Brandon, Dwayne T., Lydia A. Isaac, and Thomas A. LaVeist. 2005. "The Legacy of Tuskegee and Trust in Medical Care: Is Tuskegee Responsible for Race Differences in Mistrust of Medical Care?" *JOURNAL OF THE NATIONAL MEDICAL ASSOCIATION* 97(7):6.

Brenner, Samantha K., Rainu Kaushal, Zachary Grinspan, Christine Joyce, Inho Kim, Rhonda J. Allard, Diana Delgado, and Erika L. Abramson. 2016. "Effects of Health Information Technology on Patient Outcomes: A Systematic Review." *Journal of the American Medical Informatics Association* 23(5):1016–36. doi: 10.1093/jamia/ocv138.

CHIME. 2021. *Healthcare's Most Wired: National Trends 2021*. Survey.

Ding, Xiruo, Ziad F. Gellad, Chad Mather, Pamela Barth, Eric G. Poon, Mark Newman, and Benjamin A. Goldstein. 2018. "Designing Risk Prediction Models for Ambulatory No-Shows across Different Specialties and Clinics." *Journal of the American Medical Informatics Association* 25(8):924–30. doi: 10.1093/jamia/ocy002.

Dyer, Owen. 2019. "Memorial Sloan Kettering Cancer Center Tightens Conflict of Interest Rules after Scandals." *BMJ : British Medical Journal (Online)* 365:l1762. doi: 10.1136/bmj.l1762.

FDA. 2019a. *Clinical Decision Support Software - Draft Guidance for Industry and Food and Drug Administration Staff*. US Department of Health and Human Services.

FDA. 2019b. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD)*. US Department of Health and Human Services.

FDA. 2020a. *Overview of Device Regulation*.

FDA. 2020b. *Software as a Medical Device (SaMD)*.

FDA. 2022. *Clinical Decision Support Software - Guidance for Industry and Food and Drug Administration Staff*. US Department of Health and Human Services.

Futoma, Joseph, Jonathan Morris, and Joseph Lucas. 2015. "A Comparison of Models for Predicting Early Hospital Readmissions." *Journal of Biomedical Informatics* 56:229–38. doi: 10.1016/j.jbi.2015.05.016.

Grande, David, David A. Asch, Fei Wan, Angela R. Bradbury, Reshma Jagsi, and Nandita Mitra. 2015. "Are Patients With Cancer Less Willing to Share Their Health Information? Privacy, Sensitivity, and Social Purpose." *Journal of Oncology Practice* 11(5):378–83. doi: 10.1200/JOP.2015.004820.

IMDRF Software as a Medical Device Working Group. 2014. *"Software as a Medical Device": Possible Framework for Risk Categorization and Corresponding Considerations*.

Khullar, Dhruv. 2019. "Building Trust in Health Care—Why, Where, and How." *JAMA* 322(6):507. doi: 10.1001/jama.2019.4892.

Krieger, Nancy, Kevin Smith, Deepa Naishadham, Cathy Hartman, and Elizabeth M. Barbeau. 2005. "Experiences of Discrimination: Validity and Reliability of a Self-Report Measure for Population Health Research on Racism and Health." *Social Science & Medicine* 61(7):1576–96. doi: 10.1016/j.socscimed.2005.03.006.

Middleton, B., D. F. Sittig, and A. Wright. 2016. "Clinical Decision Support: A 25 Year Retrospective and a 25 Year Vision." *Yearbook of Medical Informatics* (Suppl 1):S103–16. doi: 10.15265/IYS-2016-s034.

Murray, Sara, Robert Wachter, and Russell Cucina. 2020. "Discrimination by Artificial Intelligence in a Commercial Electronic Health Record- a Case Study." *Health Affairs Blog*. doi: 10.1377/hblog20200128.626576.

Nemati, Shamim, Andre Holder, Fereshteh Razmi, Matthew D. Stanley, Gari D. Clifford, and Timothy G. Buchman. 2018. "An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU." *Critical Care Medicine* 46(4):547–53. doi: 10.1097/CCM.0000000000002936.

Niederer, Steven A., Joost Lumens, and Natalia A. Trayanova. 2019. "Computational Models in Cardiology." *Nature Reviews Cardiology* 16(2):100–111. doi: 10.1038/s41569-018-0104-y.

Nong, Paige, Minakshi Raj, Melissa Creary, Sharon L. R. Kardia, and Jodyn E. Platt. 2020. "Patient-Reported Experiences of Discrimination in the US Health Care System." *JAMA Network Open* 3(12):e2029650. doi: 10.1001/jamanetworkopen.2020.29650.

Nong, Paige, Alicia Williamson, Denise Anthony, Jodyn Platt, and Sharon Kardia. 2022. "Discrimination, Trust, and Withholding Information from Providers: Implications for Missing Data and Inequity." *SSM - Population Health* 18:101092. doi: 10.1016/j.ssmph.2022.101092.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366(6464):447–53. doi: 10.1126/science.aax2342.

O'Toole, Thomas P., Jose J. Arbelaez, and Robert S. Lawrence. 2004. "Medical Debt and Aggressive Debt Restitution Practices." *Journal of General Internal Medicine* 19(7):772–78. doi: https://doi.org/10.1111/j.1525-1497.2004.30099.x.

Price II, William Nicholson, Rachel Sachs, and Rebecca S. Eisenberg. 2021. "New Innovation Models in Medical AI." *SSRN Electronic Journal*. doi: 10.2139/ssrn.3783879.

Richardson, Jordan P., Cambray Smith, Susan Curtis, Sara Watson, Xuan Zhu, Barbara Barry, and Richard R. Sharp. 2021. "Patient Apprehensions about the Use of Artificial Intelligence in Healthcare." *Npj Digital Medicine* 4(1):1–6. doi: 10.1038/s41746-021-00509-1.

Roski, Joachim, Ezekiel J. Maier, Kevin Vigilante, Elizabeth A. Kane, and Michael E. Matheny. 2021. "Enhancing Trust in AI through Industry Self-Governance." *Journal of the American Medical Informatics Association* 28(7):1582–90. doi: 10.1093/jamia/ocab065.

Scharff, Darcell P., Katherine J. Mathews, Pamela Jackson, Jonathan Hoffsuemmer, Emeobong Martin, and Dorothy Edwards. 2010. "More than Tuskegee: Understanding Mistrust about Research Participation." *Journal of Health Care for the Poor and Underserved* 21(3):879–97. doi: 10.1353/hpu.0.0323.

Sendak, Mark, Madeleine Elish, Michael Gao, Joseph Futoma, William Ratliff, Marshall Nichols, Armando Bedoya, Suresh Balu, and Cara O'Brien. 2019. "'The Human Body Is a Black Box': Supporting Clinical Decision-Making with Deep Learning." *ArXiv:1911.08089 [Cs]*.

Singh, Karandeep, Thomas S. Valley, Shengpu Tang, Benjamin Y. Li, Fahad Kamran, Michael W. Sjoding, Jenna Wiens, Erkin Otles, John P. Donnelly, Melissa Y. Wei, Jonathon P. McBride, Jie Cao, Carleen Penoza, John Z. Ayanian, and Brahmajee K. Nallamothu. 2020. "Validating a Widely Implemented Deterioration Index Model Among Hospitalized COVID-19 Patients." *MedRxiv* 2020.04.24.20079012. doi: 10.1101/2020.04.24.20079012.

Smith, Jeffery. 2020. "Setting the Agenda: An Informatics-Led Policy Framework for Adaptive CDS." *Journal of the American Medical Informatics Association* 27(12):1831–33. doi: 10.1093/jamia/ocaa239.

Spector-Bagdady, Kayte, Grace Trinidad, Sharon Kardia, Chris D. Krenz, Paige Nong, Minakshi Raj, and Jodyn E. Platt. 2022. "Reported Interest in Notification Regarding Use of Health Information and Biospecimens." *JAMA* 328(5):474–76. doi: 10.1001/jama.2022.9740.

The Office of Science and Technology Policy. 2022. *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. The White House.

Thomas, Charles Ornstein, Katie. 2018. "Sloan Kettering's Cozy Deal With Start-Up Ignites a New Uproar." *ProPublica*. Retrieved September 6, 2022 (https://www.propublica.org/article/sloan-kettering-cozy-deal-with-start-up-paige-ai-ignites-new-uproar?token=D33sffVktWYYZbliQA-IClmxqRRoMh71).

Trinidad, M. Grace, Jodyn Platt, and Sharon L. R. Kardia. 2020. "The Public's Comfort with Sharing Health Data with Third-Party Commercial Companies." *Humanities and Social Sciences Communications* 7(1):1–10. doi: 10.1057/s41599-020-00641-5.

US Department of Health and Human Services. 2021. *Trustworthy AI (TAI) Playbook*.

Wachter, Robert M., and Christine K. Cassel. 2020. "Sharing Health Care Data With Digital Giants: Overcoming Obstacles and Reaping Benefits While Protecting Patients." *JAMA* 323(6):507–8. doi: 10.1001/jama.2019.21215.

Waljee, Akbar K., Peter D. R. Higgins, and Amit G. Singal. 2014. "A Primer on Predictive Models." *Clinical and Translational Gastroenterology* 5(1):e44. doi: 10.1038/ctg.2013.19.

Wong, Andrew, Erkin Otles, John P. Donnelly, Andrew Krumm, Jeffrey McCullough, Olivia DeTroyer-Cooley, Justin Pestrue, Marie Phillips, Judy Konye, Carleen Penoza, Muhammad Ghous, and Karandeep Singh. 2021. "External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients." *JAMA Internal Medicine*. doi: 10.1001/jamainternmed.2021.2626.

Wu, Eric, Kevin Wu, Roxana Daneshjou, David Ouyang, Daniel E. Ho, and James Zou. 2021. "How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals." *Nature Medicine* 1–3. doi: 10.1038/s41591-021-01312-x.

Chapter 4 "It's a Little Bit Scary": How Academic Medical Centers Govern Prediction in the Context of Uncertainty and Inequity

BACKGROUND

Advanced health informatics tools like artificial intelligence (AI), machine learning (ML), and prediction are becoming increasingly central to healthcare delivery in the US (Adler-Milstein, Nong, and Friedman 2019; Apathy, Holmgren, and Adler-Milstein 2021; CHIME 2021). Research on predictive models and implementation of AI/ML methodologies has grown significantly in the last 15 years (Panch et al. 2022). PubMed results for predictive models have increased every year for the past two decades (Figure 4.1). Despite this proliferation, evidence on the value of these tools is limited with some analysis showing that clinical prediction models did not provide any benefit or advantage compared to traditional statistical modeling (Christodoulou et al. 2019).

Health systems and providers are left with little guidance on how to govern and evaluate predictive models (Walsh et al. 2021), which is a pressing concern for multiple reasons. First, the lack of empirical evidence on the value of prediction means these systems are operating with limited information as they make important decisions about patient care. Second, there are major risks of entrenching racism and other inequities through prediction and similar tools (Benjamin 2019; Dhiman et al. 2022; Goodman, Morgan, and Hoffmann 2023). It is currently unclear how governance can best address this issue at the health system level, particularly in the absence of strong federal oversight focused on discrimination or equity. Third, regulation is attempting to

adapt to the ways software is developing but does not yet provide clear and robust oversight guidelines for health systems.

Figure 4.1 PubMed search results for predictive models
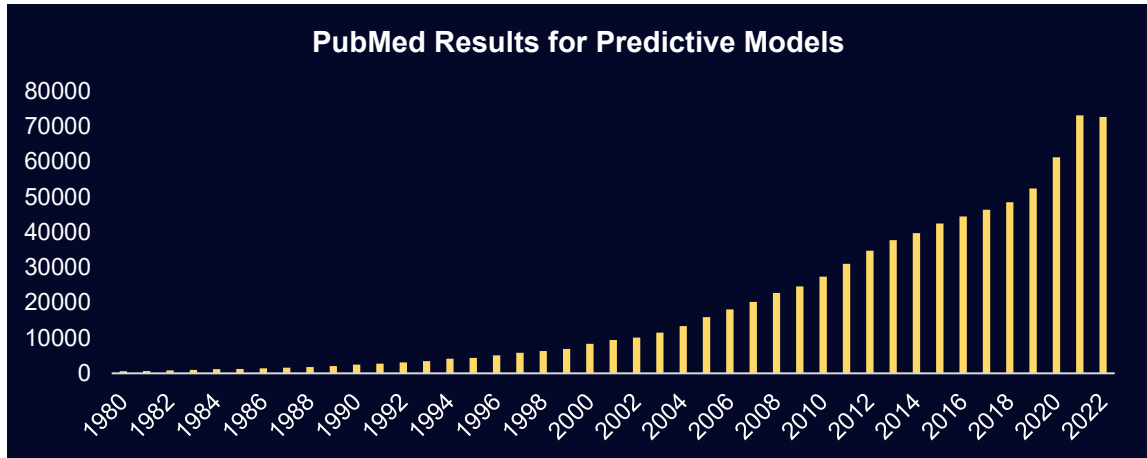
Source: Author creation. PubMed search conducted 12/12/2022

As regulatory agencies and policymakers grapple with the pace of technological development, health systems are making decisions about which models to deploy and how to use them. This project of deciding what tools to implement and how to do so is referred to as governance, which includes evaluation processes and leadership structures tasked with designing policies for review and implementation (Wright et al. 2011). For prediction, this means health systems are deciding which predictive tools to implement, and especially among academic medical centers, how to develop them. Health systems are thus designing governance practices for prediction while responding to changing regulation, shifting technologies, threats to provider and patient trust in these tools, and increasing provider burden (Asan, Bayrak, and Choudhury 2020; Christodoulou et al. 2019; Roski et al. 2021; Skeff et al. 2022). Although hospitals across the country have experience governing electronic health records (EHRs) and clinical decision

support (CDS) (Fennelly et al. 2020; Middleton, Sittig, and Wright 2016; Reddy et al. 2019; Wright et al. 2011), these existing governance approaches have key limitations when it comes to prediction and related methods. There is growing recognition of the need for more detailed, specific evidence on best governance practices and multiple layers of oversight that could provide appropriate guardrails at the organizational and federal policy levels (Reddy et al. 2019; Roski et al. 2021). These issues are explored below through discussions of the literature on oversight, healthcare software regulation, and organizational software governance.

Oversight

Technical infrastructure and quality safeguards are critical aspects of managing algorithmic tools (Chmielinski et al. 2022; Gebru et al. 2021; Price II 2022). Governance includes policies around designing and implementing IT tools or systems (Wright et al. 2011) Oversight, quality assurance, safety and effectiveness reviews, technical evaluations, and the processes for each of these aspects of management fall under governance. For prediction, these processes might include population validation, transparent quality metrics, evaluating data representativeness, and verifying independence of training datasets, which are broadly understood as important aspects of quality tool development and oversight (FDA 2022b; Gianfrancesco et al. 2018; Wong et al. 2021). Additional concerns include provider uptake, potential implications for staff workflow, and operational or financial implications for health systems. The evaluation and governance of these tools is thus multilayered and complex (Wu et al. 2021).

The Food and Drug Administration (FDA), while not tasked with establishing industry standards, does regulate predictive software applications to some extent (FDA 2019). However,

with recent changes to FDA's regulatory framework it is unclear how regulatory considerations may shape health systems' governance approaches (FDA 2022a). The specifics of federal regulation continue to evolve, but they generally lag behind the development of prediction. In the context of limited federal oversight of predictive models, governance remains concentrated at the organizational level (Roski et al. 2021). There are currently no federal policies requiring or ensuring good governance. Understanding organizational governance is thus critically important for developing best practice frameworks, ensuring patient safety, and preventing the exacerbation of inequities.

Various additional frameworks have been proposed for governance of prediction and similar tools, with the recent release of guiding principles from the Office of Science and Technology Policy and Department of Health and Human Services, for example (The Office of Science and Technology Policy 2022; US Department of Health and Human Services 2021). The National Institute of Standards and Technology has also released a framework for managing risk related to AI applications (Tabassi 2023). NIST defines governance as the processes of risk management through policies, procedures, infrastructure, and capacity building. The report includes key considerations for governance, like the evaluation, measurement, and management of risks related to AI systems. While equity is one important component of oversight, it is not as central to the document as it is for other policies and rules.

While the focus on equity in policy at the national level is relatively limited, some states have pursued action to encourage governance that centers equity. The California State Attorney General recently initiated an inquiry into racial bias in healthcare algorithms (State of California Department of Justice 2022). Although this particular action is relatively limited, it is an

important consideration for healthcare systems as they build their governance processes (Bedoya et al. 2022).

A recently proposed rule on Section 1557 of the Affordable Care Act also includes specific discussion of the risks of discrimination in clinical models (Department of Health and Human Services, Centers for Medicare and Medicaid Services 2022). These proposed rules and frameworks are largely focused on broad, high-level principles rather than specific workflows and processes that can concretely inform health system practices (Reddy et al. 2019; US Department of Health and Human Services 2021). These frameworks importantly emphasize principles like fairness, accountability, transparency, and trustworthiness but do not guide health systems with evidence on best practices or provide specifics of the governance capacity, infrastructure, and processes required to effectively pursue these principles.

Governing predictive models in health systems

Governance of predictive models strains current organizational structures, raising questions of capacity and quality. While existing policy frameworks for medical devices and EHRs or clinical decision support (CDS) can provide a starting point for predictive governance, certain characteristics of prediction present unique challenges (Price II 2022). Adaptive tools can change over time such that the decisions made based on original models may not be relevant after implementation. Similarly, model drift can occur after a model is implemented in a healthcare system. Models often require ongoing maintenance and evaluation to prevent this kind of change (Panch et al. 2022). While the importance of continuous monitoring for these issues is widely recognized, sustained evaluation like this is not required by any policy or regulation.

A survey of 25 member sites of one non-profit health system identified considerable variation in governance capacity and approaches, indicating a lack of evidence-based guidance on best practices even within a single system (Rojas et al. 2022). Other qualitative work has described a similar diversity of broad governance approaches, even among similarly well-resourced institutions (Price II 2022). Thus, there is a pressing need for understanding current governance and best practices for managing the design and implementation of prediction. Widely available information about best practices is also crucial for more equitably distributing the potential risks and benefits related to prediction in healthcare.

The literature includes some high-level characterizations of how some health systems have begun to manage AI/ML or prediction (Bedoya et al. 2022; Price II 2022). However, little is known about how health systems govern these tools. Especially because the application of these technologies is changing, it is particularly important that procedural infrastructure and best practices for governance be established.

Organizational strategies, like designing rules and allocating resources for predictive model use, are likely to reflect historic variation among health systems in response to regulation and evolving technologies (Anthony, Appari, and Johnson 2014). As described by Orlikowski, this analysis is focused on the practices and enacted rules of technology use ("technology-in-practice") as opposed to the specific characteristics of particular models (Orlikowski 2000). Mapping the rules and resources of predictive model governance in academic medical centers in this way contributes to understanding of healthcare system strategies. It also facilitates consideration of the ways these organizations, under competing pressures, respond to regulation (Anthony et al. 2014). As the regulatory landscape continues to engage with prediction, these considerations will be particularly important.

OBJECTIVE

Healthcare delivery, policy, and research would benefit from more information about 1) how health systems govern predictive models, 2) how this variation reflects best practices or potential oversights, and 3) how current frameworks or regulation inform the current state of governance. This information is necessary in order to build specific, high-quality governance frameworks that protect patients and begin to balance the potential risks and benefits of prediction in healthcare.

The objective of this chapter is to qualitatively analyze and understand predictive model governance among US academic medical centers. The research questions for this analysis are as follows:

1. How do academic medical centers govern predictive models? What are their key priorities and current governance processes?
2. How does federal regulation affect the governance practices of academic medical centers?

METHODS

Reporting for this analysis follows COREQ reporting guidelines (Tong, Sainsbury, and Craig 2007) and guidance for qualitative work in informatics (Ancker et al. 2021). This study was determined to be exempt by the University of Michigan Institutional Review Board.

The sample was constructed from a database of all Association of American Medical College Hospital and Health System Members in 2022. Veteran's Affairs systems (n=40), pediatric and children's hospitals (n=22), and other specialty hospitals (n=5) were excluded. The resulting sample of 246 hospitals was stratified by Census region (West, Midwest, South,

Northeast). Random samples of 10 hospitals were drawn from each region using Stata (v16). A committee of health informatics experts reviewed the sample and suggested additional academic medical centers to include in the sample based on their knowledge of the use of advanced informatics use across the country.

Design

An original semi-structured interview guide was developed and used for all interviews. The interview guide focused on predictive modeling capacity, governance, regulation, and model evaluation. The interview guide was piloted with three experts in multiple professional roles at different academic medical centers. These preliminary interviews informed scoping of the interview content. They were also helpful for ensuring the language used in the interview guide was widely accessible for potential interviewees.

Because governance of prediction is in its early stages, a flexible recruitment approach was used with the goal of identifying the most appropriate contact at each hospital who could provide insight on predictive model governance. Initially, one health informatics executive or leader was identified for outreach at each academic medical center. Initial recruitment contacts were most often directed to the hospital CMIO or directors of analytics and/or informatics. The individual identified at each academic medical center in the sample received at least two email contacts, with a request for referral to a colleague if necessary. In some cases, the CMIO identified a colleague who was more closely engaged in the work of governing prediction. Outreach included a description of the interviewer (e.g., name, role) and the purpose of the research project in accordance with COREQ guidelines. Participants were then scheduled for a 30 to 45-minute interview via Zoom.

The job titles of interviewees included data analytics officers, informatics officers, AI leads, and data governance directors. A total of 17 individuals participated in interviews conducted via Zoom. These individuals represented 13 academic medical centers across the country. Participating institutions are categorized by geographic region in Table 4.1. Interviews were conducted from October 2022 to January 2023 and lasted approximately 30-60 minutes.

Table 4.1 Participating academic medical centers by region

Census region	Institutions (n=13)
Northeast	4
Midwest	3
South	3
West	3

Interviews were recorded and professionally transcribed through a transcription service (Rev). I reviewed and edited every transcript to ensure accurate transcription of jargon and to correct any errors. De-identified transcripts were coded using MaxQDA software. In this chapter, I refer to participants using "P" followed by their participant ID number.

ANALYSIS

Analysis of interview data was inductive. Memos were written throughout the process of interviewing and analysis (Creswell 2007). These memos included summaries of interviews, connections between them, and key concepts or ideas that emerged from each interview (Deterding and Waters 2021). Each interview transcript was reviewed in full. Additional memos were produced for every transcript that included notes and initial inductive codes. These codes were then categorized, collapsed, or combined as needed. An initial codebook with all codes and definitions was created. These codes were applied to a subset of 4 transcripts. Necessary adjustments were made to the codebook, and codes with definitions were finalized. All

transcripts were coded using the final codebook (included in Appendix C). Themes were identified using the analytic memos, reviews of all coded segments of transcripts, and data visualizations in MaxQDA (Deterding and Waters 2021).

RESULTS

In the following sections, I describe the themes identified in this analysis (e.g., committee structure, effect on clinicians, hype). First, the topic area is indicated by bold italics and described. Under each bolded topic area, italicized themes are listed and described. Illustrative quotes and explanations of each theme are included.

Governance and CDS

Many participants described their approach to governance in relation to existing CDS governance. Because this kind of work has a longer history in healthcare, it often provided a template or comparison for participants as they thought about governing prediction. The key themes under governance and CDS were committee structure and service lines. Committee structure refers to the ways a governance committee's worked is defined and scoped in relation to CDS. Service lines are also described as clinical domains, where governance was centered on how various service lines would be affected by a certain model.

Committee structure

Participants described two primary approaches to governing prediction at academic medical centers across the country. The first was based on existing understanding of rules or rules-based clinical decision support (CDS). The second was a departure from rules-based CDS

governance that engaged with the methodological specifics of predictive tools, especially those drawing on AI or ML.

Some health system approaches did not differentiate between rules-based CDS tools and predictive models in governance. In this approach, any predictive tool was governed the same way as rules-based CDS. This often meant that predictive models would undergo the same evaluation or decision-making process as rules-based CDS tools. Other systems drew a clear distinction between these tools. While the governance structure and processes might have been designed based on CDS governance, the priorities for these health systems expanded beyond traditional CDS. They often included additional methodological questions and concerns in their governance and evaluation processes. Thus, a predictive model would undergo specific types of validation or evaluation that differed from a rules-based CDS tool.

The structure of governance committees depended on how health system leadership conceptualized predictive models. If prediction was conceptualized as a form of CDS, it was largely governed by existing CDS committees that have historically managed algorithmic alerts in the EHR. For these health systems, there was no distinction drawn between a rules-based algorithm and a ML-driven prediction. This meant the governance process was not specifically designed to manage the potential risks, benefits, or quality issues related to prediction.

> **P12** "At this point, we wouldn't make any differentiation between those two [rules-based and predictive tools]. I know that some places have machine learning or AI committees, and they separate it out, but we have not done that... It's an orders and clinical decision support governance [group], and so anything clinical would go through that group - any clinical decision support, whether it's machine learning/AI-based or more traditional rules-based CDS. "

> **P2** "I don't think we have anything that's specific, that because it's a clinical predictive model we treat it differently than any other sort of clinical decision support system from our end."

If, however, prediction was understood as something separate from traditional rules-based decision support, there were typically steps taken to govern the methodological specifics of prediction. For example, academic medical centers that identified specific potential harms related to prediction typically designed steps of a governance process that attempted to identify and mitigate these harms. Issues like population validation and performance by subpopulation were particularly important for systems governing prediction beyond traditional CDS approaches. These concerns were central to evaluation and decision-making, whereas a similar approach was largely absent if a system treated prediction as a form of CDS governed by existing CDS committees. Two participants clearly describe this distinction within their own systems. Traditional CDS governance was not designed to engage with the specifics of prediction, so a separate committee was created to oversee predictive tools.

> **P10** "And our clinical decision support subcommittee recognized 'this is not really our wheelhouse. Best practice alerts based on logic that was driven by clinical guidelines? That's what we do. But all these kinds of tools where you have to know if the actual tool is working as intended, if the score's any good, that's just not in our wheelhouse'."

> **P7** "They don't need to think about 'how predictive is this tool' like what are the performance characteristics of the model? What is the bias in the model? They're not the right people to think about that. They're the right people to think about 'how are our users going to uptake the output of the model? What does the clinical decision support look like? Is it going to fire on every single patient 10 times a day?'."

> **P1** "[The predictive model committee] was pretty much stood up specifically for these tools because we recognized that there was a need that was not met by our more traditional governance process. And the fact that… the skillsets that are typically present in the traditional governance groups were not equipped to handle these requests and actually weren't empowered to ask the difficult questions related to predictive models."

Service lines

Most health systems structured governance around service lines or clinical domains, in addition to a centralized governance committee. This was true whether or not the system had a

separate committee specifically for predictive model governance. Typically, service line or clinical leadership was represented on a central committee or was regularly included in governance based on a given model's relevance to their work.

One key decision point in governance processes was whether a model would impact clinicians across service lines or clinical domains. If a model only impacted clinicians and staff in one clinical domain, governance would often occur through a central committee after the approval of the clinical leadership. However, if a model crossed service lines or clinical domains, an additional layer of governance occurred. For example, if a cardiologist requested a model that would affect primary care providers, the approval of the primary care leadership was required. Similarly, nursing leadership approval was required if the implementation of a model would affect nurse workflow in some systems, service line and clinical leadership was represented on the central CDS or predictive model committee. In others, they were consulted on a more ad hoc basis when a requested model would affect their staff.

Model characteristics

Participants reported multiple key characteristics of predictive models themselves that shaped governance. They described specific priorities and concerns related to model characteristics that could either trigger additional governance or alter the governance process. The source or provenance of the model was one of these defining characteristics. The other was the type of model (e.g., clinical or administrative) being considered for implementation.

Model provenance

Health system governance structures usually differed between models produced by the EHR vendor, developers within their system ("homegrown"), and third parties. The particular concern about model provenance was most often related to population validation. Participants explained that because vendors typically created models using national samples of patients, those models might not function well for their patient populations. Homegrown models, on the other hand, were built using data from the health system itself. Thus, population validation was sometimes less of a concern for homegrown models. Multiple participants described how they ran vendor models "silently" (without affecting patients or visibly presenting to clinicians) to verify that they were functioning properly before considering implementation. In some cases, evaluation was not as rigorous or strictly required for homegrown models.

> **P9.1** "There are the ones that come out of the box from [the EHR vendor], and then there are those which you could either make yourself or buy from somebody else and then implement into the back end of [the EHR]. But a big component of our evaluation process is model performance in the local environment. So, a lot of what we do is run the models in the dark, so to speak, just collecting data about them. Only now have we accumulated enough data where our governance committee can actually start taking a look at it and trying to understand what the value of the model is."

For other systems, the goal was a consistent governance approach for all models regardless of origin. Here, both homegrown and vendor models were evaluated and governed the same way. This standard governance and evaluation was not reported by a majority of participants, but for those health systems with a consistent approach, it was described in these ways:

> **P6** "If something's built locally in terms of a model, I would say it's important not to skip that step in terms of reviewing the evidence with the local group and that they agree that this model is one that's worth following, that it performs to their satisfaction"

P7 "If someone builds a model from scratch, it's different than someone signing with a company on the side and them starting to do work. But our goal is to actually really unify our approach to governance such that whether things come from the research community, the health system, vendors, they go through the same steps of evaluation. But what may be different is prioritization."

Some health systems ran vendor models to test and evaluate them, while collaborating with developers of homegrown models in a peer-review style working relationship.

P10 "We are doing a lot of analyses and helping with evaluating if models are good, **particularly if it's a vendor model, then we try to be that second check of is it any good. If it's a model developed by a research group at [institution], by and large we're not involved in trying to evaluate it.** We're having the group who's in involved in building it come and talk to us about how it's doing here and what are some weak points and stuff."

Model types: administrative or clinical

Most health systems described a different approach to administrative or operational models compared to clinical models. While two health systems had committees that oversaw every predictive model deployed across the health system regardless of model type, this was the exception. Although some individual CMIOs or similarly positioned individuals may have had knowledge of administrative models, they were not generally involved in decision-making about whether to deploy them. A business operations committee or individual decision-maker was usually assumed to oversee those kinds of models.

Generally, clinical models received more oversight than non-clinical models. The governance structures and processes described above, with the exception of two health systems, were largely only relevant to clinical applications. The processes of population validation, running models "in the background", and cross-service line governance were not usually applied to administrative models. One participant described the understanding of risk that drove the lower standards for governing administrative models:

P9.1 "So, some of these business operations models where there's basically no clinical risk, you're not going to mistreat somebody because of it. **Those we might take the vendor's word for it and just try it for a while to see if it does save us money** the way they say it [will]. Whereas for a clinical thing, we really have a higher standard."

This understanding generally reflects current regulatory frameworks that focus only on clinical risk and do not engage with other types of risks to patients. This is not to say that most health systems completely ignored administrative applications. Some CMIOs or similarly positioned individuals would occasionally consult with operations managers on administrative tools, but there was not a required vetting or approval process like there was for clinical applications.

The two health systems that employed the same governance approach for both clinical and administrative models described a broader conceptualization of risk. They included things like inequitable outcomes and barriers to care as risks. These two systems generally had the most detailed and centralized governance approaches of all the health systems represented in this study. For them, the centralized governance of both clinical and administrative models meant there was a unified approach to all tools being deployed that could affect patients and their care.

P1 "We include within [our purview] the administrative models as well. The reasoning was that administrative decisions in healthcare still end up touching patients."

One example of non-clinical risk was described by this participant. They referred to a predictive model built to identify patients at risk of missing appointments in order to double-book their appointment slots. They described how this kind of model directly impacted the quality of care even though it was not strictly clinical in terms of diagnosis or treatment.

P8 "I guess it's a different kind of risk. But if you predict that all these people [will] not show up and you double book those spots, then you have all these patients who are not going to get optimal care because there's too many of them there."

Some participants that did not work in health systems with this unified governance of clinical and administrative models reported that the ideal governance structure would be full review and evaluation of all models that impacted patients or their care. However, their governance structures or resources were not equipped to manage such coordinated or centralized management.

Governance process

This topic area includes the specific practices that comprised governance for the health systems included in this study. The themes under this topic include 1) the procedures followed by various health systems, 2) how governance procedures were initiated, and 3) how model requests were made.

Procedures

Governance processes varied and fell into three main categories. The first is a well-defined process of review and evaluation across the institution. The second is an emerging process, typically evolving from existing CDS and EHR governance processes. The third is a governance process reliant on interpersonal relationships or individual decisions (see Figure 4.2).

Figure 4.2 Governance procedures
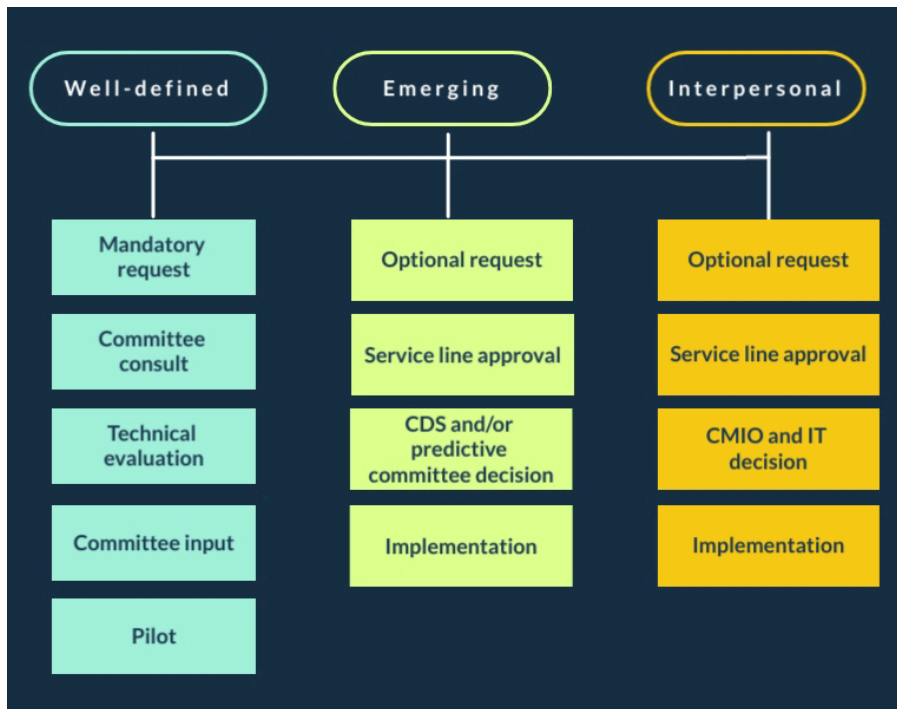
In well-defined governance processes, every model received the same review and

evaluation. A request to the governance committee was required. The committee would review

the model in question and consult with the requester. They would evaluate the model and decide

if it would proceed for further piloting and review. If the model proceeded, it would then be

piloted. Implementation decisions were made based on the pilot results. This participant

describes some of the details of a well-defined governance process.

> **P10** "If it's a model that needs to sit in the [EHR] it has to go through us. And the reason
> it has to go through us is who's going to actually integrate it, who's going to wire it up,
> who's going to connect it? It's going to be one of a handful of people. So, our job is to
> help decide is this something that one of those analysts should be spending their time on?
> That's part of why we exist. That's the governance piece. And the other is, let's say we
> have five requests, what should they work on first? What's a priority? What's less of a
> priority? That's why we exist."

Here, the predictive model governance committee decided what to implement and how to prioritize the many requests they received. Their well-defined governance procedures were focused on resource allocation and strategic priorities.

In the emerging and interpersonal categories, health system governance depended more on individual decisionmakers and circumstances than the well-defined approaches described above. This participant explains how a model approval process could happen in an interpersonal governance system.

> **P12** "It depends on who the cardiologist is. It depends on who they know. It depends on all sorts of stuff, right. That cardiologist might just go put a ticket in the ticketing system and say, "I'd like this." They might know me, and they might shoot me an email or a text message. They might know some analyst who works in the IT department and send them a message. Or they might take it to somebody else in their department, like their business administrator or their chair or their division director, and say, "Hey, please go do this." So then, depending on which of those entry points it starts at, there's a variety of different ways that it can go."

Initiating governance

Especially systems without the well-defined governance structure that specified exact governance processes, there was variation in how a model came to be reviewed by a committee. There were some models or tools that could be implemented without committee oversight. These were most often vendor-produced models that did not 1) impact multiple service lines and clinical domains or 2) require extensive IT staff time or financial resources.

If a model could be implemented without extensive IT resources or effort, it was less likely to be reviewed by a governance committee. This was especially true for vendor-produced models. In these cases, the models were available through the EHR vendor. They already existed and could be "turned on" without extensive customization or IT staff time. If there was a clinician sponsor and buy-in from their clinical domain, additional committee oversight was not

routinely conducted in health systems without a well-defined governance structure in **Figure 2**.

If a model implementation required substantial IT staff time, it was more likely to reach a

governance committee.

Another common pathway to committee review or oversight was if a given tool impacted

clinicians across service lines or clinical domains.

> **P10** "For anything that touches multiple stakeholders though, it goes across departments, or it affects a nurse and a physician or a radiology tech or it affects different types of stakeholders - that often goes to us. And anything where the health system has to spend money where they want us to review the thing, that goes through us."

When a model was specific to the operations and clinical care of one specific area of the

health system, whether defined clinically or by service lines, the leadership of that domain

typically governed the tool. Barring financial implications or IT staff effort, these models could

be turned on in systems without well-defined governance structures. In systems with well-

defined governance, all these models would be reviewed regardless of financial or service line

implications.

> **P1** "So in general, there are service line specific governance groups that make decisions on which request gets approved and what priority they get approved at. The predictive model governance process was set up because, in part, that particular service line-based governance process was not well set up to ask additional questions beyond what each of those service line-specific stakeholder groups asked in terms of lifecycle management of models, validity of the models, effectiveness of the models, issues around equity and fairness, and the regulatory nature of predictive models."

Requests or rubrics

Multiple health systems, especially those with the most robust and explicit governance

workflows, described specific evaluation rubrics or request forms they used as a first step to

evaluate a model. These forms or checklists were often succinct and focused on actionable

outcomes. They were often very brief and asked about 1) what model was being requested, 2)

what problem it would solve, and 3) what intervention the model would lead to or support. This rubric, when required, was used to trigger the governance process. The appropriate committee would review the request form and work with the requester to determine if a given model was appropriate, identify a good model, or consider alternative approaches like altering existing alerts or using rules--based algorithms.

Governance priorities and considerations

Participants described multiple priorities that shaped the governance process in their health system. These included 1) effect on clinicians and 2) actionability or impact on patients. This participant described the common range of priorities reported in this study.

> **P4** "Cost-benefit, which includes obviously a financial component, but also an outcomes component related to clinical outcomes, related to patient satisfaction, related to employee efficiency and satisfaction. Those are probably the major [priorities]. Is this going to save our doctors time, make their lives easier. Is this going to improve the experience for our patients? Is this going to improve a clinical outcome? Sometimes there are regulatory or quality imperatives, I think, that we're trying to address as well."

In this example, the effect of a model on clinicians was important but also integrated with actionability, the effect on patients, and financial implications of a given model.

Effect on clinicians

Effect on clinicians was one of the most commonly described priorities or guiding concerns in governance processes. As described above, this often grew from the CDS governance paradigm that focused on minimizing alerts and interruptions to clinician workflow. Here, prioritization of requests or new model implementation centered around the impact on the

clinician in terms of interruptions or "clicks" required to deal with them. The priority of

minimizing interruptive alerts was top of mind for governance leaders.

> **P2** "Governance is primarily designed to prevent extraneous things from showing up [in the EHR] that people don't want."

This framing reflected the typical approach to prediction as CDS, whereby the primary

goal is minimizing inappropriate, irrelevant, or interruptive alerts that frustrate clinicians.

Additionally, clinician sponsors were often required to initiate the request process for a new

model. Clinician buy-in was central to the process.

> **P5.1** "We're very careful about things that might impact on a patient's care and make sure that we have clinical stakeholders who may or may not be directly involved in the project… you need to be working very closely with your clinical champions and experts. So, for sepsis, we've used a mix of infectious disease and intensive care specialists to buy in or not buy in to the proposed approach."

Actionability and impact on patients

Fewer health systems explicitly described the impact of models on patients as a core

principle of governance. When participants described patient impact as central to decision-

making, it was typically related to the number of patients that would be impacted. It was often

described as "how many patients will this touch?". Priority was usually granted to models with

the greatest potential impact on patients. Safety was a key theme related to patient-focused

priorities.

> **P2** "If it's a safety or compliance issue, it's prioritized. Otherwise, it's kind of a mix of financial and clinical [priorities]. There may be formal governance processes. There may not be. It kind of depends."

> **P3** "There are people that have even said in our organization 'what's happening right now is not safe'. And not that we wouldn't have created the oversight committee, but it definitely has pushed us to get it in place with legal there, with quality and patient safety there."

Regulation and best practices

As part of their governance approaches, participants considered organizations and regulators outside of their own health system to different degrees. As described above, industry self-governance and FDA regulation have both been suggested as pathways for oversight and engendering trust in tools like prediction. Below, participants' perceptions of regulation and best governance practices are described.

Considering regulation

FDA software regulation was not typically a core concern among participants. Most participants were aware of how the FDA categorized and regulated software, but they did not see FDA regulation as very impactful to their own work or governance. Regulation was generally seen as an issue that might become more relevant in the future. However, in its current form, FDA's guidelines were not prioritized.

> **P9.1** "The fact that they in the future might have to submit that to the FDA for some clearance, I don't know whether it will have any impact on the commercial side of this industry or not. But that part of the FDA's arm, the device arm, they're classifying these things as medical devices and the barrier to get a medical device approved is minuscule compared to a drug. I don't think on the commercial side it will have a big impact."

> **P7** "Well, nothing's been regulated to date really… Our regulatory framework is very challenging. But I mean, I think since it hasn't existed, we have functioned under the belief that we can be marketed crap at any time basically. And that's what happens. [The vendor] markets a model and the r-squared is 0.09 and it says that in the model brief. But unless you have people who actually understand biostatistics read that small fine print, they're literally marketing that. But people don't know that that means the model doesn't do anything for you."

Some participants believed that FDA regulation was designed for EHR vendors and third-party model developers rather than health systems that might design or implement a tool that

could be classified as a medical device. Thus, for the health system itself, regulation was

interesting but only indirectly relevant.

> **P7** "We're going to need to be more accountable for the vendor-provided models than homegrown ones. It's going to be the commercial models that are regulated."

> **P2** "We don't really… Obviously, we think about it a little bit, but we don't think it's likely that for institutionally developed and implemented things, even if it's potentially under FDA regulation, they're going to be coming after our health system, saying 'you implemented something that we consider a medical device, and we're going to fine you'… I think there would be too much blowback for the FDA to do that."

The role of vendors

Most participants described considerable excitement among vendors and some members

of health system leadership related to prediction. Although many participants did not share this

optimism, it was important context for their decision-making. External excitement among

vendors and leadership about prediction at times interfered with participants' ability to evaluate

and govern models. It also obscured the potential value, or lack thereof, of some models. Some

participants also described issues with reliance on vendors as an alternative or supplementary

stakeholder that could shoulder some of the oversight burden.

Hype

For many participants, designing a governance process in the context of considerable

hype around advanced analytics was challenging. In combination with the rubrics explained

above, they described a concerted effort to cut through the hype and identify actionable

healthcare challenges that could benefit from tools like predictive models.

> **P4** "We really focus a lot on trying to separate the myth or the fantasy of how [this] is going to solve every problem from just realistic, what can we actually do?"

P9 "Yeah, I think the potential for real damage to the long-term ability to leverage EHR data for the benefit of patients is real. **I think the hype could really prematurely sour the whole community's understanding of this benefit.** Taking it out of the hands of people who want to sell it and putting it in the hands of people who want to find the places where it's really clearly of value and then responsibly expand from there in a careful way is a key thing to be done."

Multiple participants described skepticism of vendor and third-party solutions because of the financial incentives that drove their product pitches. While participants expressed this feeling personally, some also described it as affecting system's governance process. Put another way, the governance of prediction was characterized by a need to measure or validate the claims of a vendor about their products.

This was especially the case because of issues with population validation or performance of these models in the health system.

P5.1 "So, basically, at this point, I'm very skeptical of vendor input, because it's all going to be self-inflating and self-aggrandizing. What I find most useful is talking to [other academic medical centers], places that have enough medical informatics, intelligence, etc. where there are strong computers in medicine teams, and swapping notes."

P7 "So I just perceive the whole field and landscape to be a mess… This needs to be regulated because all these health systems who don't have someone in their clinical systems who knows what an R squared is, are like, "cool, let's take it. It must be a little bit helpful". And I've actually even heard that from academic medical centers where you'd think they wouldn't think that. So… **we function with the understanding [here] at least, that these models are meaningless until proven otherwise.** We have to get under the hood of these, test them ourselves."

Many participants mentioned the recently published analysis of performance issues with the sepsis prediction designed by a large EHR vendor. This was top of mind as an example of why vendor promotion and hype gave them pause. The gap between the vendor's reported performance and the observed performance published in the academic literature was like an alarm for many participants when they thought about implementing models in their own systems.

P9.1 "Then the other thing frankly is some of the work [about sepsis models] that you cannot take these vendor's word that this thing does what it's supposed to do. That I would say, more so than anything, helped the other leaders in our organization… pay attention. They just as well drink the [vendor] Kool-Aid and be like, '[the vendor] says we should do this.' But then when you show them work [on sepsis models], it's like, 'Oh, okay, now I'm glad that there's a smart person somewhere thinking about whether this is the right thing to do or not'. Then we start to sanction it."

P2 "So yeah, I think we were proven right that you should check. I think [the vendor] has gotten more careful about it after that sort of debacle. But it really makes sense, right? Like, why would you implement something you don't understand? Just because your vendor says it's a good idea? I don't know, that just seems reckless."

Reliance and vendor expertise

Although many participants represented academic medical centers with large IT and data science teams or resources, they described some issues with reliance on vendor expertise. Part of this reliance was technical (e.g., taking vendor metrics at face value). However, they also reported concern about relying on vendors or third parties to take care of governance. Participants described this as farming out the responsibility for oversight and management of predictive tools.

P11 "I suspect that the more nascent health IT places are probably just... pulling things from the vendor platform. Or they make some big deal with some third-party company to say, 'Now you're responsible for all of these things'."

P5 "I'm surprised vendors have actually gotten into this space with their own set of models without a required infrastructure to take that on. Obviously, they're paying for that already with the press…[they] really paid the price for trying to do this without the appropriate infrastructure. Having said that, I noticed that **a lot of organizations tend to third-party their complex solutions, whether it's pathways or predictive models. I think that's very risky** as well [as] not having a designed, robust, appropriate team to be able to handle all the elements that come with algorithms that are relatively complex… I think organizations suffer now and will continue to really hurt until they take it very seriously."

For some health systems, the vendor encouraged this kind of reliance. They would emphasize their expertise and product quality to system leadership, making it difficult for

participants to advocate for the resources they needed to conduct appropriate model oversight and management.

> **P9.1** "So, I would say that's one of the big barriers to this is getting a team together that's funded to do the work, because like I said, on one hand, [the vendor] is saying, 'You don't need to worry about it.' So, it can be hard to convince the organizational leadership that the investment's worth it."

> **P5.1** "[The EHR vendor] is producing AI but it's not an AI company. I had our guys working with their AI team and the [vendor] team was much less experienced than our guys. So, to expect them to produce AI that's going to be turned on and deployed at the top 20 academic medical centers in the country, when you realize who's behind it, it's a little bit off-putting, scary."

Equity

A minority of participants mentioned concerns about equity or bias in predictive models. Some participants described that their health systems had recently become engaged in conversations about algorithmic bias. Four out of thirteen systems had algorithmic bias and equity as core to their priorities for predictive model governance. When they discussed equity, participants described it in two ways. First, and more common, was an understanding of equity in the models themselves through aspects of the prediction or the data used to train the model. Second, and less common, was a discussion of equity in relation to the application of a model and the consequences for patients. Finally, some participants described how structural differences between academic medical centers and smaller hospitals could have negative effects on health inequities.

Inequity and models

For some participants, the idea of equity was limited to algorithmic bias or statistical fairness. The concept of bias or fairness was only discussed as an aspect of the models

themselves (e.g., equal performance across populations, representativeness of training datasets).

One participant described thinking about equity and how the models function.

> **P1** "We ask about the performance of models across racial groups. We ask them if their features include racial groups or other demographics that could be predictors for socioeconomic statuses. We ask them 'why did you do that, or did you just do it because you have done it in the modeling process', and to think through the implications and to think about whether those features might inadvertently exacerbate inequities. We have experts that provide guidance for how to make modifications to the models or the model approaches if there's a concern about inequities."

For others, equity was a larger concept than just the predictors in a model or the model's performance across populations. It was related to how the model was used or to what purpose it was applied. One participant described this in the context of a model predicting patients missing appointments.

> **P10** "We look at the number side, but then we also look at the implementation side. So, it wasn't the race and religion [variables] that actually tipped off the equity, it was 'who is the patient it's identifying?'. It's identifying people who aren't going to show up. Who are the sorts of people who don't show up? The people who need a wheelchair, the people who don't have transportation, who have other issues. And so, it was actually just regardless of the predictors, who are we trying to identify? And if we use that to double book, if the people show up, it's going to be the worst people to get double booked who will show up and get double booked. And so that was the bigger concern than anything around what was going on in the model.

> **P5** "The Data Science Team had taken into consideration bias and health equity. So, they put their models through this audit process, and also work very closely with our Health Equity Office and Diversity Group. Is the model performance bad for a certain group, because that represents indeed a health difference? Or is it because of the systems' effect, or the bias of systems?... There is an audit in the model inputs, or other proxies for race. If you provided zip code, or something like that, that ends up being a proxy for race. So, we do an audit of the inputs that are being fed to the model, as well."

This health system had a clear workflow for the data scientists vetting models that included an extensive list of questions about effects by race/ethnicity and specifically required an analysis of whether a potential difference in performance was driven by health disparities or

118

disparate health*care*. Importantly, this type of analysis was done for both clinical and operational

models.

System-level resource inequalities

Most participants discussed the challenge of resource limitations in governing and

managing predictive models. Even among these leading academic medical centers, there were

consistent concerns about allocating limited resources. The IT staff time required to implement

new models or configure and customize vendor models was a limiting factor for most health

systems. When participants considered the implications of these constraints for less well-

resourced health systems or hospitals, this concern became even more pressing. Put another way,

these participants who generally had an IT team and data scientists evaluating, implementing,

and managing models were concerned about their counterparts at smaller hospitals who did not

have similar resources. Part of this concern was that smaller hospitals were left relying on

vendors to provide and evaluate models.

> **P2** "Smaller hospitals, community hospitals, etc. they typically just take the vendor package, right? They do not want to spend any effort on this, and it is just what they have. I think it is what it is until we can get inoperable decision support that's easier to spread. There really is a…**It's kind of like, kind of almost a luxury for the places that have more resources.**"

> **P1** "What [the vendor] does not necessarily require is for the clients to undergo a clinical effectiveness evaluation because it does take a little bit more resources to look at whether a model has its intended impact clinically and in terms of processes. I would encourage other clients to at least think about, in a minimal way, if you've done a silent evaluation and the model's continuously promising, when you put it in the hands of the clinicians …can you do a simple evaluation of a sample of usages to see whether it led to the intended outcome? Can you do some surveys to see what people thought of it and to ask periodically whether this model should continue to be used?"

> **P9** "We need as a community to be focused exactly on what the information needs of these people are. Because it's easy, I think, to attract them with the hype and then for

them to just flounder for a couple of years, not really knowing whether they're making a misstep or having the right meetings with the right people and everything else."

P9.1 "Well, I mean I would say that the hardest part of all this is figuring out how to fund it. So, without the toolkit, I mentioned we have probably half of a statistician's time and my time and [a faculty member's] time. It's expensive… I went to the [vendor] meeting and heard an interesting presentation about how a community hospital is using some of the other analytical tools inside [the EHR]. The thing is they're using them straight out of the box. They're not doing anything because they just can't. So, anyway, it's a little bit scary."

When this participant said, "they're not doing anything", they meant that the community hospital was not doing any customization or adjustments based on the tools' performance in the hospital itself. For this CMIO, using a vendor model this way was concerning because the models may not perform as well in a particular patient population when compared to their performance with a national sample of patients.

LIMITATIONS

There are important limitations to consider when interpreting these findings. First, the participants included in this study represent academic medical centers. The specific work done by these health systems, and the resources available to them, differ from other types of health systems. The findings here, specifically related to the development of homegrown models, may not be relevant to or representative of non-academic medical centers. Additionally, the positions of the participants interviewed for this project is relevant to interpretation. The perspectives of CMIOs and chief data analytics directors differ from the perspectives of health administrators, legal experts, and clinicians or end users. Qualitative work that includes the perspectives of these various stakeholders will be necessary to understand predictive model governance more holistically.

DISCUSSION

This analysis identifies key characteristics of predictive model governance in academic medical centers across the US (see Table 4.2 for a summary of results). In addition to revealing the governance structures and processes currently operating at academic medical centers, it analyzes the role of regulation and equity considerations in these processes. This work identifies multiple aspects of governance that should inform industry-wide best practices and provide insight into current governance approaches for policymakers and regulators as they consider changes to current frameworks.

Table 4.2 Summary of results

Topic	Theme	Description
Governance and CDS	Committee structure	An existing CDS committee governs prediction Or A separate committee exists for prediction, AI/ML
	Service lines	Governance is focused around clinical areas and/or service lines, much like traditional CDS governance
Model characteristics	Provenance	The designer or provider of a predictive model shapes the governance process (e.g., vendor-produced models receive different oversight than homegrown models)
	Model type	Different model types receive different types of governance according to a clinical/administrative application distinction Or All models, regardless of model type, undergo the same governance process
Governance process	Processes/procedures	1. Well-defined 2. Emerging 3. Interpersonal or individual
	Initiating governance	In systems where models are governed differently, there are some models that do not receive significant oversight. In these cases, a trigger or initiator for governance is required (e.g., if there are financial implications of implementing a model)
	Model requests	Governance processes depend on request forms or rubrics that are submitted to a

		governance committee. Decision-making is shaped or guided by the answers to the prompts on the request form Or Rubrics/request forms are available but not required for the governance committee's review
Priorities and considerations	Effect on clinicians	Governance prioritizes impacts on clinician workflow, akin to traditional CDS governance
	Impact/actionability	Decision-making focuses on the potential effects of a model on patients and the types of action that can be taken based on the use of a model
Regulation and best practices	Consideration of regulation	Regulation is seen as relevant, but not central to the work of the health system Or Regulation shapes and is centered within health system governance structures/processes
	Best practices	Decision-makers attempt to identify best practices through collaborations and discussions with their own colleagues/networks
The role of vendors	Hype	Hype around predictive models, while not limited to vendors, is often promoted by EHR vendors that are pitching various predictive tools to health system leadership
	Expertise and reliance	Some health systems rely on vendor expertise in evaluation or governing models Or Health system leadership is concerned about over-reliance on EHR vendor expertise and treat vendor pitches with skepticism
Equity	Conceptualizing equity	Considerations of equity, when made, are focused on algorithmic bias and discrimination *within the tool itself* Or Equity is understood to incorporate not only the model itself, but also the decisions made and actions taken based on the output of a model
	System-level resource inequalities	Among academic medical centers, there is variation in resources and capacity for model evaluation and validation And There is considerable concern about the structural differences between well-resourced health systems and others without the staff or capacity to extensively evaluate predictive models prior to implementation

Health systems are making decisions about predictive model implementation in the context of limited regulatory guidance and fragmented knowledge about best governance practices (Roski et al. 2021). As key stakeholders, they exercise considerable power over decisions about models that currently affect patient care. As identified in this analysis, decision-makers are concerned with the quality of the models available to them. They either have explicit governance procedures or recognize the important of this kind of clarity as an aspiration. However, they report skepticism or mistrust around predictive models designed and sold by vendors. This may be a result of the hype around prediction and related methods like AI/ML.

Many of the academic medical centers represented in this analysis saw regulation as focused on model vendors. This perception is important for how prediction and software more broadly are governed. It results in health systems designing governance processes based on their experience and understanding of best practices from peer institutions rather than federal agencies or professional organizations. More specific industry standards, guidelines, and best practices are necessary.

Participants described a central focus on how predictive model implementation affects clinician workflow. Considering the large literature on clinician burnout, this is an important consideration. However, a focus on minimizing EHR alerts represents a limited understanding of the impact of prediction on clinicians and clinical care more broadly. It is possible that the paradigm of alert fatigue is obscuring deeper potential issues that should be prioritized. For example, as described above, even traditional CDS governance at leading academic medical centers can miss crucial questions about specific model characteristics that may indicate quality. It can also lead to implementation of models that may not perform well in a specific patient population or perpetuate biases in multiple ways. Based on the data analyzed here, best practices

may include a combination of traditional CDS governance prioritizing impact on clinicians with specific methodological expertise related to prediction and related methods.

Although a growing body of literature highlights the risk of racism and other inequities inherent in prediction, most participants did not mention bias or inequity as a central priority in their governance processes. This indicates a need for improved equity literacy across the health system. Specific guidelines and best practices for incorporating equity audits and evaluations are also necessary. Other fields and collaborative efforts have produced specific steps health systems can take in relation to countering structural racism (Wyatt et al. 2023). Similarly specific guidance and best practices would benefit governance of predictive model governance in healthcare.

CONCLUSION

This analysis identifies how academic medical centers across the US govern prediction. Specifically, it highlights three main types of governance processes: 1) well-defined, 2) emerging, and 3) interpersonal or individual. In well-defined governance processes, health systems have explicit procedures for review and evaluation of predictive models. In emerging governance processes, systems are adjusting or applying previous CDS governance approaches to prediction and related methods. In health systems with interpersonal or individual-driven governance approaches, there is less explicit or consistent model governance. In systems with this approach, an individual is most often making decisions about model implementation. Among academic medical centers, there is considerable variation in approaches to predictive model governance.

The range of governance structures and processes described here indicate a need for additional guidance, whether regulatory or otherwise, for health systems as prediction continues to proliferate. Rather than concentrating responsibility for governance within organizations, multiple levels of governance that include the industry and regulators would benefit patient care and safety. Guidance and oversight from both regulators and professional organizations could support consistent protections for patients. These efforts, combined with additional evidence on the risks and benefits of prediction, would benefit decision-makers at health systems across the country.

Based on the data analyzed here, there is also considerable work to be done to build health equity literacy among leading decision-makers at academic medical centers across the country. There is limited structural engagement with the relationships between prediction, racism, and other inequities. Health systems will need to prioritize capacity building on this topic to prevent further exacerbation of digital racism and protect patients.

References

Adler-Milstein, Julia, Paige Nong, and Charles P. Friedman. 2019. "Preparing Healthcare Delivery Organizations for Managing Computable Knowledge." *Learning Health Systems* 3(2):e10070. doi: 10.1002/lrh2.10070.

Ancker, Jessica S., Natalie C. Benda, Madhu Reddy, Kim M. Unertl, and Tiffany Veinot. 2021. "Guidance for Publishing Qualitative Research in Informatics." *Journal of the American Medical Informatics Association* (ocab195). doi: 10.1093/jamia/ocab195.

Anthony, Denise L., Ajit Appari, and M. Eric Johnson. 2014. "Institutionalizing HIPAA Compliance: Organizations and Competing Logics in U.S. Health Care." *Journal of Health and Social Behavior* 55(1):108–24. doi: 10.1177/0022146513520431.

Apathy, Nate C., A. Jay Holmgren, and Julia Adler-Milstein. 2021. "A Decade Post-HITECH: Critical Access Hospitals Have Electronic Health Records but Struggle to Keep up with Other Advanced Functions." *Journal of the American Medical Informatics Association* ocab102. doi: 10.1093/jamia/ocab102.

Asan, Onur, Alparslan Emrah Bayrak, and Avishek Choudhury. 2020. "Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians." *Journal of Medical Internet Research* 22(6):e15154. doi: 10.2196/15154.

"Attorney General Bonta Launches Inquiry into Racial and Ethnic Bias in Healthcare Algorithms." *State of California - Department of Justice - Office of the Attorney General*. Retrieved October 13, 2022 (https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-inquiry-racial-and-ethnic-bias-healthcare).

Bedoya, Armando D., Nicoleta J. Economou-Zavlanos, Benjamin A. Goldstein, Allison Young, J. Eric Jelovsek, Cara O'Brien, Amanda B. Parrish, Scott Elengold, Kay Lytle, Suresh Balu, Erich Huang, Eric G. Poon, and Michael J. Pencina. 2022. "A Framework for the Oversight and Local Deployment of Safe and High-Quality Prediction Models." *Journal of the American Medical Informatics Association* 29(9):1631–36. doi: 10.1093/jamia/ocac078.

Benjamin, Ruha. 2019. "Assessing Risk, Automating Racism." *Science* 366(6464):421–22. doi: 10.1126/science.aaz3873.

CHIME. 2021. *Healthcare's Most Wired: National Trends 2021*. *Survey*.

Chmielinski, Kasia S., Sarah Newman, Matt Taylor, Josh Joseph, Kemi Thomas, Jessica Yurkofsky, and Yue Chelsea Qiu. 2022. "The Dataset Nutrition Label (2nd Gen): Leveraging Context to Mitigate Harms in Artificial Intelligence."

Christodoulou, Evangelia, Jie Ma, Gary S. Collins, Ewout W. Steyerberg, Jan Y. Verbakel, and Ben Van Calster. 2019. "A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models." *Journal of Clinical Epidemiology* 110:12–22. doi: 10.1016/j.jclinepi.2019.02.004.

Creswell, John W. 2007. *Qualitative Inquiry & Research Design: Choosing among Five Approaches*. 2nd ed. Thousand Oaks: Sage Publications.

Department of Health and Human Services, Centers for Medicare and Medicaid Services. 2022. *Nondiscrimination in Health Programs and Activities*. *Notice of proposed rulemaking*. 42 CFR Parts 438, 440, 457, and 460.

Deterding, Nicole M., and Mary C. Waters. 2021. "Flexible Coding of In-Depth Interviews: A Twenty-First-Century Approach." *Sociological Methods & Research* 50(2):708–39. doi: 10.1177/0049124118799377.

Dhiman, Paula, Jie Ma, Constanza L. Andaur Navarro, Benjamin Speich, Garrett Bullock, Johanna A. A. Damen, Lotty Hooft, Shona Kirtley, Richard D. Riley, Ben Van Calster, Karel G. M. Moons, and Gary S. Collins. 2022. "Risk of Bias of Prognostic Models Developed Using Machine Learning: A Systematic Review in Oncology." *Diagnostic and Prognostic Research* 6(1):13. doi: 10.1186/s41512-022-00126-w.

FDA. 2019. *Clinical Decision Support Software - Draft Guidance for Industry and Food and Drug Administration Staff*. US Department of Health and Human Services.

FDA. 2022a. *Clinical Decision Support Software - Guidance for Industry and Food and Drug Administration Staff*. US Department of Health and Human Services.

FDA. 2022b. *Good Machine Learning Practice for Medical Device Development: Guiding Principles*.

Fennelly, Orna, Caitriona Cunningham, Loretto Grogan, Heather Cronin, Conor O'Shea, Miriam Roche, Fiona Lawlor, and Neil O'Hare. 2020. "Successfully Implementing a National Electronic Health Record: A Rapid Umbrella Review." *International Journal of Medical Informatics* 144:104281. doi: 10.1016/j.ijmedinf.2020.104281.

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford. 2021. "Datasheets for Datasets." *Communications of the ACM* 64(12):86–92. doi: 10.1145/3458723.

Gianfrancesco, Milena A., Suzanne Tamang, Jinoos Yazdany, and Gabriela Schmajuk. 2018. "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data." *JAMA Internal Medicine* 178(11):1544–47. doi: 10.1001/jamainternmed.2018.3763.

Goodman, Katherine E., Daniel J. Morgan, and Diane E. Hoffmann. 2023. "Clinical Algorithms, Antidiscrimination Laws, and Medical Device Regulation." *JAMA*. doi: 10.1001/jama.2022.23870.

Middleton, B., D. F. Sittig, and A. Wright. 2016. "Clinical Decision Support: A 25 Year Retrospective and a 25 Year Vision." *Yearbook of Medical Informatics* (Suppl 1):S103–16. doi: 10.15265/IYS-2016-s034.

Orlikowski, Wanda. 2000. "Using Technology and Constituting Structures: A Practice Lens for Studying Technology in Organizations." *Organization Science* 11(4). doi: 10.1287/orsc.11.4.404.14600.

Panch, Trishan, Erin Duralde, Heather Mattie, Gopal Kotecha, Leo Anthony Celi, Melanie Wright, and Felix Greaves. 2022. "A Distributed Approach to the Regulation of Clinical AI." *PLOS Digital Health* 1(5):e0000040. doi: 10.1371/journal.pdig.0000040.

Price II, William Nicholson. 2022. "Distributed Governance of Medical AI." *SSRN Electronic Journal*. doi: 10.2139/ssrn.4051834.

Reddy, Sandeep, Sonia Allan, Simon Coghlan, and Paul Cooper. 2019. "A Governance Model for the Application of AI in Health Care." *Journal of the American Medical Informatics Association : JAMIA* 27(3):491–97. doi: 10.1093/jamia/ocz192.

Rojas, Juan C., Gordon Rohweder, Janet Guptill, Vineet M. Arora, and Craig A. Umscheid. 2022. "Predictive Analytics Programs at Large Healthcare Systems in the USA: A National Survey." *Journal of General Internal Medicine*. doi: 10.1007/s11606-022-07517-1.

Roski, Joachim, Ezekiel J. Maier, Kevin Vigilante, Elizabeth A. Kane, and Michael E. Matheny. 2021. "Enhancing Trust in AI through Industry Self-Governance." *Journal of the American Medical Informatics Association* 28(7):1582–90. doi: 10.1093/jamia/ocab065.

Skeff, Kelley M., Cati G. Brown-Johnson, Steven M. Asch, Dani L. Zionts, Marcy Winget, and Yaniv Kerem. 2022. "Professional Behavior and Value Erosion: A Qualitative Study of

Physicians and the Electronic Health Record." *Journal of Healthcare Management* 67(5):339–52. doi: 10.1097/JHM-D-21-00070.

Tabassi, Elham. 2023. *AI Risk Management Framework: AI RMF (1.0)*. error: NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology. doi: 10.6028/NIST.AI.100-1.

The Office of Science and Technology Policy. 2022. *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. The White House.

Tong, A., P. Sainsbury, and J. Craig. 2007. "Consolidated Criteria for Reporting Qualitative Research (COREQ): A 32-Item Checklist for Interviews and Focus Groups." *International Journal for Quality in Health Care* 19(6):349–57. doi: 10.1093/intqhc/mzm042.

US Department of Health and Human Services. 2021. *Trustworthy AI (TAI) Playbook*.

Walsh, Colin G., Mollie M. McKillop, Patricia Lee, Joyce W. Harris, Christopher Simpson, and Laurie Lovett Novak. 2021. "Risky Business: A Scoping Review for Communicating Results of Predictive Models between Providers and Patients." *JAMIA Open* 4(4):ooab092. doi: 10.1093/jamiaopen/ooab092.

Wong, Andrew, Erkin Otles, John P. Donnelly, Andrew Krumm, Jeffrey McCullough, Olivia DeTroyer-Cooley, Justin Pestrue, Marie Phillips, Judy Konye, Carleen Penoza, Muhammad Ghous, and Karandeep Singh. 2021. "External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients." *JAMA Internal Medicine*. doi: 10.1001/jamainternmed.2021.2626.

Wright, Adam, Dean F. Sittig, Joan S. Ash, David W. Bates, Joshua Feblowitz, Greg Fraser, Saverio M. Maviglia, Carmit McMullen, W. Paul Nichol, Justine E. Pang, Jack Starmer, and Blackford Middleton. 2011. "Governance for Clinical Decision Support: Case Studies and Recommended Practices from Leading Institutions." *Journal of the American Medical Informatics Association* 18(2):187–94. doi: 10.1136/jamia.2009.002030.

Wu, Eric, Kevin Wu, Roxana Daneshjou, David Ouyang, Daniel E. Ho, and James Zou. 2021. "How Medical AI Devices Are Evaluated: Limitations and Recommendations from an Analysis of FDA Approvals." *Nature Medicine* 1–3. doi: 10.1038/s41591-021-01312-x.

Wyatt, Ron, Leslie Tucker, Kedar Mate, Fred Cerise, Alicia Fernandez, Sachin Jain, Ignatius Bau, and Daniel Wolfson. 2023. "A Matter of Trust: Commitment to Act for Health Equity." *Healthcare* 11(1):100675. doi: 10.1016/j.hjdsi.2023.100675.

Chapter 5 Conclusion

In this dissertation, I examined public perspectives on predictive models and analyzed how health systems design and implement prediction. In the studies on public perspectives, I analyzed original survey measures fielded with a national sample of US adults. In Chapter 2, I analyzed this survey data and demonstrated how the public perceives data use for prediction. In Chapter 3, I examined public perspectives on specific predictive models in healthcare. Chapter 4 focused on the results of qualitative interviews with health system governance leadership. In this chapter I identified how academic medical centers currently govern prediction, analyzed the role of regulation, and described how equity is included in current governance processes.

The growing literature on algorithmic fairness and bias in predictive models indicates an important development in the field. It represents an acknowledgement of the ways information technologies perpetuate racism and other types of inequity. However, it is insufficient. As Denton and Hanna explain, "In contrast to the significant efforts that have focused on statistical properties of training datasets, comparatively little attention has been paid to the various modes of their constitution; that is…what and whose values influence the choices of data to collect" (Hanna et al. 2020). Similar dynamics shape the literature on values and prediction more broadly. Questions about whose values are shaping prediction and the governance approaches that determine how prediction operates in healthcare have remained unanswered while the tools themselves have been implemented. This project asks specifically about those values and choices. It analyzes public perspectives and values around the use of data and prediction in healthcare. It identifies policy gaps and the need for more robust guidance for governance.

Equity

In Chapter 2, I found that people who have experienced discrimination in the healthcare system are less comfortable with the use of personal characteristic and sensitive data types for prediction than people who have not experienced discrimination. I also identified that white respondents are more comfortable with personal characteristic and health-related data being used for prediction than Black respondents. These findings signal potential inequity in the system of prediction in healthcare. Chapter 3 expanded on this analysis, identifying that the public is less comfortable with administrative predictive model applications than clinical ones. Taken together, the findings from these two chapters indicate misalignment between current practice and patient values, in some cases across identity and experiences of discrimination.

Some of the ways predictive models in healthcare can perpetuate bias and inequity have been analyzed and explored in the literature (Obermeyer et al. 2019; Rajkomar et al. 2018; Vyas, Eisenstein, and Jones 2020). However, these analyses do not typically engage with the reality of prediction for patients' lived experiences. For example, removing the power of decision-making from individuals or administrators and using predictive models to determine appointment times can affect patients in multiple ways. Not only can their appointment be double-booked, but they also have no way of knowing why their appointment time was cut in half. The reason for their shorter appointment time is unclear. This kind of bureaucratic opacity functions as a way of "making decisions without seeming to decide" (Porter 2020) by obscuring who is responsible for the shorter appointment time or for other clinical and administrative predictions that shape patient experiences of the healthcare system. This makes contestation nearly impossible except for those with the most advantage (Burrell and Fourcade 2021; Pasquale 2015). The implications for health inequities are concerning and require further investigation.

Implications for policy and practice

As described in Chapter 4, academic medical centers are struggling to design effective and well-defined governance approaches to predictive models. There are multiple implications of these findings for healthcare in practice. First, health systems would benefit from clearly defined and described industry-wide best practices. Some efforts to design maturity models are currently underway that may support this work (Knosp et al. 2018), but organizational governance is only one necessary layer of prediction management in healthcare (Roski et al. 2021). While many participants described using their professional networks to learn from their peers, this is not a sufficient or reliable method of identifying and implementing best governance practices. Rather, industry-wide, publicly available guidelines on strategies for managing the challenges of designing, implementing, and monitoring predictive models are necessary.

Second, health systems would benefit from specific guidance on how to evaluate predictive models. Critical evaluation of vendors' and developers' claims about the value of a given model is an important aspect of evaluation (Christodoulou et al. 2019; Murray, Wachter, and Cucina 2020). However, many health systems do not have the capacity to conduct this critical evaluation. Current preliminary work on model cards and explainability could support these efforts and help health systems make informed decisions about the use of prediction (Mitchell et al. 2019; Nong, Raj, and Platt 2022). To complement these efforts, widely available guidelines for model evaluation would benefit the field. Clarifying policy around software oversight is also important. While FDA's recent final guidance on CDS is a positive step, the findings presented in Chapter 4 indicate that there is a need for greater clarity on the implications for health systems.

Finally, healthcare practice should draw on the growing empirical evidence on patient concerns. The role of trust in healthcare has been an important topic in the literature for over fifty years (Anderson and Griffith 2022; Bajaj and Stanford 2021; Benkert et al. 2019; Taylor, Nong, and Platt 2023). Over this period, myriad analyses have identified the significant role of trust in patients sharing data, engaging with providers, and participating in research (Armstrong et al. 2006; Campos-Castillo and Anthony 2019; Catellier and Yang 2012; Nong, Williamson, et al. 2022). If health systems are interested in engendering or preserving trust, engagement with patient perspectives is critical. Drawing on the findings presented in Chapters 2 and 3 of this dissertation, healthcare practice can incorporate patient concerns in their governance and system policies. They can prioritize signaling competence and integrity to their patients as they use data to build predictive models. They can also consult patients and the public on how to communicate around and manage prediction in a responsive and trustworthy way. While some frameworks from regulators and federal agencies promote transparency, incentives to implement practice and policy that reflect these values could promote progress.

Future research

This dissertation provides novel evidence and insights into 1) public and health system administrator perspectives on the role predictive models play in healthcare, 2) misalignment between public values and current policy, and 3) how the system of prediction in healthcare reflects structural racism. For example, my research suggests that administrative applications are particularly problematic for patients, and yet there is no systematic oversight for this use of prediction. While the focus of this dissertation was predictive models, there are implications for

empirically evaluating equity in the context of learning health systems, clinical decision support, and precision health more broadly.

In Chapters 2 and 3 of this dissertation, I analyze public comfort with 1) data use for prediction and 2) specific applications of predictive models. Future work should expand on these findings to analyze public perspectives on potential risks of prediction, anticipated consequences, policy expectations, desire for notification, and expectations of providers. Empirical evidence on how patients think about tradeoffs between predictive accuracy and potential privacy issues or uses of sensitive data, for example, could guide responsive policy and practice that takes into account the potential benefits of prediction in healthcare.

The research in this dissertation elucidates the need for: 1) research questions that more accurately address the knowledge gaps impacting system operations and populations, including naming and monitoring /measuring how racism operates in information systems; 2) comprehensive standards around predictive modeling and governance; 3) production of knowledge that is more reliable, responsive, and robust to the needs of minoritized communities; and 4) practice recommendations that comprehensively address implications for patients and systems that engage with inequities in all their forms.

This dissertation also highlights the need for future studies that will contribute to a comprehensive view of the predictive model ecosystem, and the stakeholder interests and tensions therein. Two critical next steps include expanding research to include vendor perspectives on implementation of predictive models as well as community-based participatory research that centers the voices of racially minoritized people in the development and application of predictive models in healthcare. These are described below.

The role of vendors

One key stakeholder that is especially important in understanding customs and practices around prediction in healthcare are the EHR vendors who often design and provide predictive models to health systems. Given the power of these stakeholders, and the reach that some of their models have (Chen and Asch 2017; Ding et al. 2018; Murray et al. 2020; Wong et al. 2021), it is critically important to analyze how they manage prediction. The degree to which these organizations consider or engage with the inequity implications of their models is especially important in considering how policy may better protect the public from the negative consequences of prediction. Future work should also explore how vendors may be able to inform health systems about the appropriate (and inappropriate) applications of models that may contribute to minimizing racial health inequities. Building on the findings presented in Chapter 4, this kind of research should incorporate the needs and perspectives of non-academic medical centers as well. Such an approach will identify potential misalignments between how vendors promote predictive models and the needs of smaller or non-academic medical centers in governing or implementing prediction.

AI/ML methodologies

As mentioned in the introductory chapter, the specific methodologies driving predictive models and risk scores in healthcare are fundamental to our understanding of customs and practices around prediction. Because of the importance of methodological decisions in the construction of predictions, future work should engage with model developers, health systems, and patients to better understand their interpretations of AI/ML-driven predictions specifically. While there is some existing work on patient perspectives on AI, the samples are

disproportionately white and employed or previously employed in healthcare (Richardson et al. 2021). Expanding on the insights derived from the proposed project, future work should involve community based participatory research (CBPR) with patients about their perspectives on AI/ML predictions. It will also engage with model developers and health systems on more specific methodological choices and their implications for racial inequities.

CONCLUSION

There is considerable excitement about big data analytics in healthcare. EHR vendors, technology startup companies, and health systems themselves are increasingly investing in tools like predictive models (Ding et al. 2018; Siwicki 2021). Some of these stakeholders are indicating awareness of health equity and expressing concern about negative impacts of these tools on patients. However, this work is limited in multiple ways. Among the many limitations of the fairness approach (Davis, Williams, and Yang 2021) is a narrow focus on the ways predictive models operate and the data they use. Larger questions about the values these systems prioritize remain largely unasked and unanswered in the healthcare literature, with some important exceptions in popular media (Ross 2023).

In order to understand predictive tools in context and identify the many ways they can affect the public, it is critical that public values and concerns are better understood. Various frameworks and policy documents emphasize the importance of patient engagement and transparency (Tabassi 2023; US Department of Health and Human Services 2021), but these goals are not achievable without empirical evidence on public perspectives and values.

References

Anderson, Andrew, and Derek M. Griffith. 2022. "Measuring the Trustworthiness of Health Care Organizations and Systems." *The Milbank Quarterly* 100(2):345–64. doi: 10.1111/1468-0009.12564.

Armstrong, Katrina, Abigail Rose, Nikki Peters, Judith A. Long, Suzanne McMurphy, and Judy A. Shea. 2006. "Distrust of the Health Care System and Self-Reported Health in the United States." *Journal of General Internal Medicine* 21(4):292–97. doi: 10.1111/j.1525-1497.2006.00396.x.

Bajaj, Simar Singh, and Fatima Cody Stanford. 2021. "Beyond Tuskegee — Vaccine Distrust and Everyday Racism." *New England Journal of Medicine* 384(5):e12. doi: 10.1056/NEJMpv2035827.

Benkert, Ramona, Adolfo Cuevas, Hayley S. Thompson, Emily Dove-Meadows, and Donulae Knuckles. 2019. "Ubiquitous Yet Unclear: A Systematic Review of Medical Mistrust." *Behavioral Medicine* 45(2):86–101. doi: 10.1080/08964289.2019.1588220.

Burrell, Jenna, and Marion Fourcade. 2021. "The Society of Algorithms." *Annual Review of Sociology* 47(1):annurev-soc-090820-020800. doi: 10.1146/annurev-soc-090820-020800.

Campos-Castillo, Celeste, and Denise Anthony. 2019. "Situated Trust in a Physician: Patient Health Characteristics and Trust in Physician Confidentiality." *The Sociological Quarterly* 60(4):559–82. doi: 10.1080/00380253.2018.1547174.

Catellier, Jennifer R. Allen, and Z. Janet Yang. 2012. "Trust and Affect: How Do They Impact Risk Information Seeking in a Health Context?" *Journal of Risk Research* 15(8):897–911. doi: 10.1080/13669877.2012.686048.

Chen, Jonathan H., and Steven M. Asch. 2017. "Machine Learning and Prediction in Medicine — Beyond the Peak of Inflated Expectations." *The New England Journal of Medicine* 376(26):2507–9. doi: 10.1056/NEJMp1702071.

Christodoulou, Evangelia, Jie Ma, Gary S. Collins, Ewout W. Steyerberg, Jan Y. Verbakel, and Ben Van Calster. 2019. "A Systematic Review Shows No Performance Benefit of Machine Learning over Logistic Regression for Clinical Prediction Models." *Journal of Clinical Epidemiology* 110:12–22. doi: 10.1016/j.jclinepi.2019.02.004.

Davis, Jenny L., Apryl Williams, and Michael W. Yang. 2021. "Algorithmic Reparation." *Big Data & Society* 8(2):20539517211044810. doi: 10.1177/20539517211044808.

Ding, Xiruo, Ziad F. Gellad, Chad Mather, Pamela Barth, Eric G. Poon, Mark Newman, and Benjamin A. Goldstein. 2018. "Designing Risk Prediction Models for Ambulatory No-Shows across Different Specialties and Clinics." *Journal of the American Medical Informatics Association* 25(8):924–30. doi: 10.1093/jamia/ocy002.

Hanna, Alex, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. "Towards a Critical Race Methodology in Algorithmic Fairness." Pp. 501–12 in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, *FAT* '20*. New York, NY, USA: Association for Computing Machinery.

Knosp, Boyd M., William K. Barnett, Nicholas R. Anderson, and Peter J. Embi. 2018. "Research IT Maturity Models for Academic Health Centers: Early Development and Initial Evaluation." *Journal of Clinical and Translational Science* 2(5):289–94. doi: 10.1017/cts.2018.339.

Mitchell, Margaret, Simone Wu, Andrew Zaldivar, Parker Barnes, Lucy Vasserman, Ben Hutchinson, Elena Spitzer, Inioluwa Deborah Raji, and Timnit Gebru. 2019. "Model
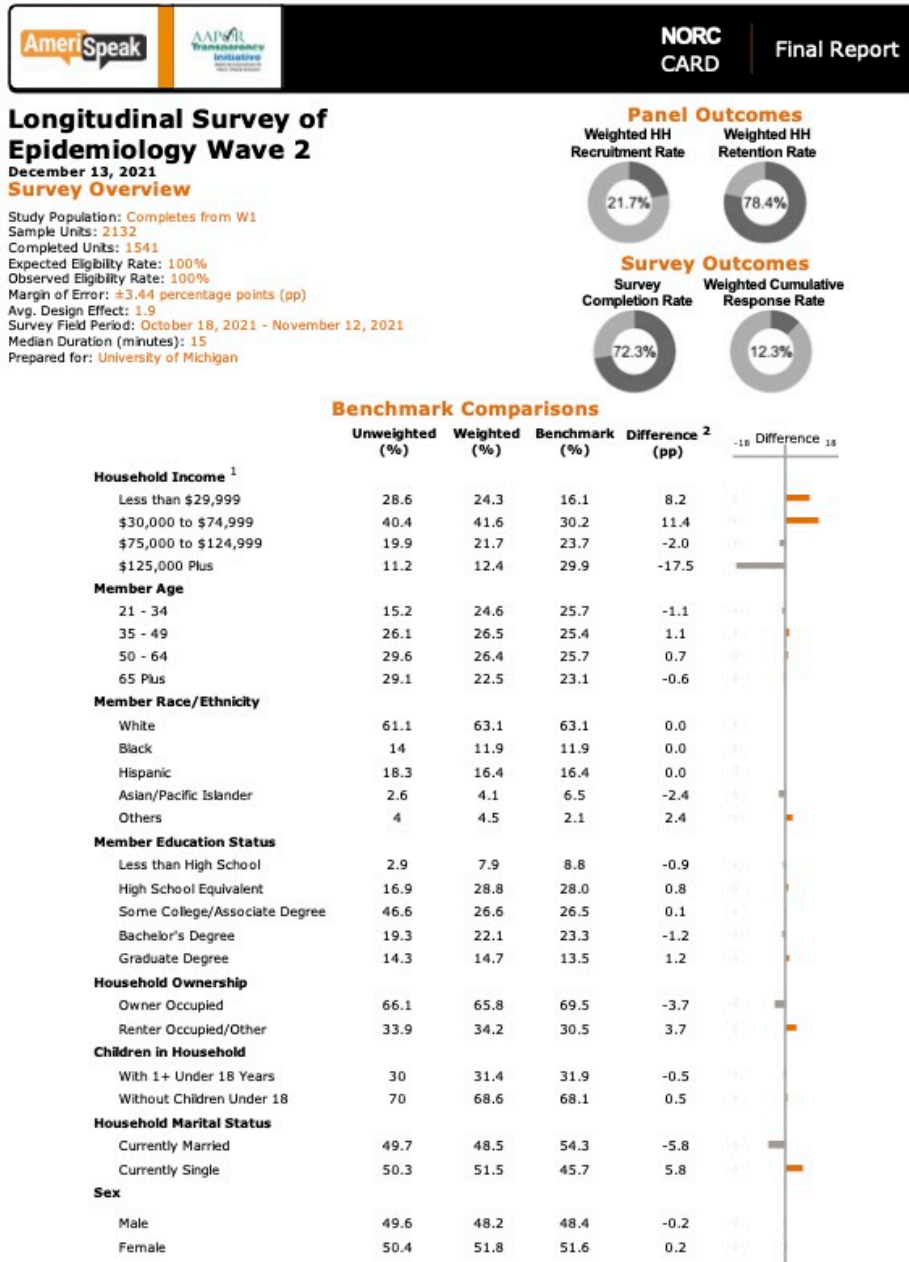
Cards for Model Reporting." Pp. 220–29 in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, *FAT* '19*. New York, NY, USA: Association for Computing Machinery.

Murray, Sara, Robert Wachter, and Russell Cucina. 2020. "Discrimination by Artificial Intelligence in a Commercial Electronic Health Record- a Case Study." *Health Affairs Blog*. doi: 10.1377/hblog20200128.626576.

Nong, Paige, Minakshi Raj, and Jodyn Platt. 2022. "Integrating Predictive Models into Care: Facilitating Informed Decision-Making and Communicating Equity Issues." *The American Journal of Managed Care* 28(1):18–24. doi: 10.37765/ajmc.2022.88812.

Nong, Paige, Alicia Williamson, Denise Anthony, Jodyn Platt, and Sharon Kardia. 2022. "Discrimination, Trust, and Withholding Information from Providers: Implications for Missing Data and Inequity." *SSM - Population Health* 18:101092. doi: 10.1016/j.ssmph.2022.101092.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366(6464):447–53. doi: 10.1126/science.aax2342.

Pasquale, Frank. 2015. *The Black Box Society*. Harvard University Press.

Porter, Theodore M. 2020. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.

Rajkomar, Alvin, Michaela Hardt, Michael D. Howell, Greg Corrado, and Marshall H. Chin. 2018. "Ensuring Fairness in Machine Learning to Advance Health Equity." *Annals of Internal Medicine* 169(12):866–72. doi: 10.7326/M18-1990.

Richardson, Jordan P., Cambray Smith, Susan Curtis, Sara Watson, Xuan Zhu, Barbara Barry, and Richard R. Sharp. 2021. "Patient Apprehensions about the Use of Artificial Intelligence in Healthcare." *Npj Digital Medicine* 4(1):1–6. doi: 10.1038/s41746-021-00509-1.

Roski, Joachim, Ezekiel J. Maier, Kevin Vigilante, Elizabeth A. Kane, and Michael E. Matheny. 2021. "Enhancing Trust in AI through Industry Self-Governance." *Journal of the American Medical Informatics Association* 28(7):1582–90. doi: 10.1093/jamia/ocab065.

Ross, Casey. 2023. "Denied by AI: How Medicare Advantage Plans Use Algorithms to Cut off Care for Seniors in Need." *STAT*. Retrieved March 13, 2023 (https://www.statnews.com/2023/03/13/medicare-advantage-plans-denial-artificial-intelligence/).

Siwicki, Bill. 2021. "How CIOs Are Prioritizing AI Investments for the next 5 Years." *Healthcare IT News*, October 18.

Tabassi, Elham. 2023. *AI Risk Management Framework: AI RMF (1.0)*. NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology. doi: 10.6028/NIST.AI.100-1.

Taylor, Lauren A., Paige Nong, and Jodyn Platt. 2023. "Fifty Years of Trust Research in Health Care: A Synthetic Review." *The Milbank Quarterly* 1–53. doi: 10.1111/1468-0009.12598.

US Department of Health and Human Services. 2021. *Trustworthy AI (TAI) Playbook*.

Vyas, Darshali A., Leo G. Eisenstein, and David S. Jones. 2020. "Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms." *New England Journal of Medicine* 383(9):874–82. doi: 10.1056/NEJMms2004740.

Wong, Andrew, Erkin Otles, John P. Donnelly, Andrew Krumm, Jeffrey McCullough, Olivia DeTroyer-Cooley, Justin Pestrue, Marie Phillips, Judy Konye, Carleen Penoza, Muhammad Ghous, and Karandeep Singh. 2021. "External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients." *JAMA Internal Medicine*. doi: 10.1001/jamainternmed.2021.2626.

Appendices

Appendix A Survey Sample and Measures

Appendix Table A.1 NORC survey sample description

AmeriSpeak | AAPOR Transparency Initiative | **NORC CARD** | **Final Report**

Longitudinal Survey of Epidemiology Wave 2
December 13, 2021
Survey Overview

Study Population: Completes from W1
Sample Units: 2132
Completed Units: 1541
Expected Eligibility Rate: 100%
Observed Eligibility Rate: 100%
Margin of Error: ±3.44 percentage points (pp)
Avg. Design Effect: 1.9
Survey Field Period: October 18, 2021 - November 12, 2021
Median Duration (minutes): 15
Prepared for: University of Michigan

Panel Outcomes
Weighted HH Recruitment Rate: 21.7%
Weighted HH Retention Rate: 78.4%

Survey Outcomes
Survey Completion Rate: 72.3%
Weighted Cumulative Response Rate: 12.3%

Benchmark Comparisons

	Unweighted (%)	Weighted (%)	Benchmark (%)	Difference [2] (pp)
Household Income [1]				
Less than $29,999	28.6	24.3	16.1	8.2
$30,000 to $74,999	40.4	41.6	30.2	11.4
$75,000 to $124,999	19.9	21.7	23.7	-2.0
$125,000 Plus	11.2	12.4	29.9	-17.5
Member Age				
21 - 34	15.2	24.6	25.7	-1.1
35 - 49	26.1	26.5	25.4	1.1
50 - 64	29.6	26.4	25.7	0.7
65 Plus	29.1	22.5	23.1	-0.6
Member Race/Ethnicity				
White	61.1	63.1	63.1	0.0
Black	14	11.9	11.9	0.0
Hispanic	18.3	16.4	16.4	0.0
Asian/Pacific Islander	2.6	4.1	6.5	-2.4
Others	4	4.5	2.1	2.4
Member Education Status				
Less than High School	2.9	7.9	8.8	-0.9
High School Equivalent	16.9	28.8	28.0	0.8
Some College/Associate Degree	46.6	26.6	26.5	0.1
Bachelor's Degree	19.3	22.1	23.3	-1.2
Graduate Degree	14.3	14.7	13.5	1.2
Household Ownership				
Owner Occupied	66.1	65.8	69.5	-3.7
Renter Occupied/Other	33.9	34.2	30.5	3.7
Children in Household				
With 1+ Under 18 Years	30	31.4	31.9	-0.5
Without Children Under 18	70	68.6	68.1	0.5
Household Marital Status				
Currently Married	49.7	48.5	54.3	-5.8
Currently Single	50.3	51.5	45.7	5.8
Sex				
Male	49.6	48.2	48.4	-0.2
Female	50.4	51.8	51.6	0.2

[1] Race/Ethnicity, Household Ownership, Income, Education, Number of Children, Marital Status, Gender, and Geographic Region benchmarks are from the March 2021 Census Bureau Current Population Survey. [2] The difference between the Weighted and Benchmark columns. [3] 21+ General Population

Appendix Table A.2 Survey description of predictive models

Predictive models use information about you, and other people like you, to predict what you might want or do in the future. For example, Netflix predicts what you might watch based on the movies you have already watched. Amazon looks at what you have bought, and what other people buy, to guess what you might buy next.

Predictive models are also used in healthcare. They can help a doctor anticipate if you are at high risk of an allergic reaction. They can help identify the most effective treatments for you based on information about you and other patients like you. Predictive models can also help health systems decide how many patients to schedule based on whether people have missed their appointments in the past.

Appendix B Survey Definitions

Appendix Table B.1 Survey definitions of key terms

My healthcare system
"Your healthcare system" refers to the healthcare professionals and institutions that you personally interact with when getting health care.

The healthcare system
"The healthcare system" refers generally to the healthcare system in this country.

Healthcare providers
Health care providers include people such as doctors and nurses who provide medical treatment.

Electronic health record
A digital version of your paper chart or medical record. An electronic health record contains your medical and treatment history including diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results.

Health information
Health information includes information about you and your medical treatment history including diagnoses, medications, treatment plans, immunization dates, allergies, radiology images, and laboratory and test results.

De-identified [health information or biospecimens]
De-identified means that "identifying information" about you is removed from your health information. Identifying information includes things like your name, address, date of birth, etc.

Appendix C Codebook

Appendix Table C.1 Codebook

Role	
Chief information officer	CIO, CMIO, etc.
Data analytics chief/officer	Designing models, implementation
Other	
Regulation	
Don't engage	FDA regulation doesn't factor into governance
Consider	FDA regulation isn't dismissed, but isn't central to decisions
Prioritize	FDA regulation shapes the institution's approach
Priorities	
Workflow/minimize alerts	Prioritizes impact on clinical workflow/desire to streamline and keep alerts at a manageable level
Model characteristics (stats, validation)	Prioritizes quality of the model, as measured by things like AUC/stats/validation/quantitative analysis
Actionability	Prioritizes an action or change based on a model/tool – "what are you going to do based on the output?"
Patient safety/harm	Prioritizes minimizing safety risks
Business case/$	Prioritizes financial implications. This can include HIT staff time, data analyst hours, etc.
# patients impacted	Prioritizes scale – "how many patients are actually going to be affected. This is going to include "no one is seeing this" or "no one is using it anymore"
Governance	
Source	
Third party	Non-EHR vendor and non-homegrown models or tools (e.g. startup)
Homegrown	Designed within the healthcare system
Vendor	Designed by the EHR vendor
Oversight requirement	Is the oversight process required prior to implementation?
None/interpersonal	There is no requirement for vetting/governance process for a model to be implemented or turned on OR this process depends on personal relationships ("If they know the CMIO they can get it moved up the list")
Optional	Vetting/governance is encouraged but there isn't a hard requirement for involvement of an oversight committee (gray area between interpersonal and required)
Required	Vetting/governance is required before a model can be turned on
Rubric	Does the institution have a specific rubric or list of questions for model requesters? ("rubric" or "request form")
	This code will include what's on that list: does it require silent validation? Does it require effectiveness evaluation? Etc.

Structure	How is governance structured?
Service line/clinical area	Clinical area or service line is engaged in the approval and review (e.g., nothing can impact primary care if the primary care committee/leadership isn't involved)
Effect on clinicians/ sponsor	A clinician or faculty member has to sponsor or request a model for it to be reviewed, or the governance process is triggered by a model affecting clinicians.
Separate committee	Relationship between traditional/logic-based model management and prediction/AI. e.g., Work of a previous committee wasn't sufficient for AI/prediction, so a new structure was stood up, or
CDS committee governs	Prediction is covered by CDS committee – no separate committee or structure
Centralized	There is a central governance committee for prediction/AI/ML/etc.
Previous failure	Prior attempts at governance dissolved/did not work
Type	
Clinical	Clinical models get oversight
Admin/operational	Admin/operational models get oversight
Equity	
Consider	Equity concerns are mentioned and inform the governance process but are not core to the decision-making
Prioritize	Equity is a priority to governance/included in a specific step of the evaluation
Best practices	Looking for best practices from colleagues or other institutions "consult with my colleagues" or "I'm part of this consortium"
Vendors	How are vendors relevant to the governance process? How do they factor into the decision making?
Expertise	Considerations of vendor expertise – do they know what makes a good model? Are they considered automatically good?
Skepticism	Skepticism of vendor models – an indicator that the system does not accept what the vendor offers
Pitches	Vendors pitching models to CMIOS/clinicians etc., exerting pressure or promoting the use of models
Support	Vendor assistance or support in managing/governing
Hype	Any mention of excitement or hype around prediction/AI/ML