



Management Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Can Information Sharing Reduce Diagnostic Disparate Impact? Evidence from a Health Information Exchange

Minghong Yuan, Indranil Bardhan, Wen Wen

To cite this article:

Minghong Yuan, Indranil Bardhan, Wen Wen (2026) Can Information Sharing Reduce Diagnostic Disparate Impact? Evidence from a Health Information Exchange. *Management Science*

Published online in Articles in Advance 08 May 2026

. <https://doi.org/10.1287/mnsc.2024.05617>

This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as “*Management Science*. Copyright © 2026 The Author(s). <https://doi.org/10.1287/mnsc.2024.05617>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>.”

Copyright © 2026 The Author(s)

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Can Information Sharing Reduce Diagnostic Disparate Impact? Evidence from a Health Information Exchange

 Minghong Yuan,^a Indranil Bardhan,^a Wen Wen^{a,*}
^aUniversity of Texas at Austin, Austin, Texas 78712

*Corresponding author

Contact: minghong.yuan@mcombs.utexas.edu,  <https://orcid.org/0009-0001-5785-4571> (MY); indranil.bardhan@mcombs.utexas.edu,  <https://orcid.org/0000-0002-4265-5780> (IB); wen.wen@mcombs.utexas.edu,  <https://orcid.org/0000-0002-8086-6386> (WW)

Received: April 8, 2024

Revised: June 13, 2025; October 16, 2025

Accepted: December 30, 2025

Published Online in Articles in Advance: May 8, 2026

<https://doi.org/10.1287/mnsc.2024.05617>
Copyright: © 2026 The Author(s)

Abstract. The literature has documented qualitative evidence of disparities in physician diagnoses, which can lead to disparities in healthcare delivery, patients' perception of care, and health outcomes. In this research, we seek to understand whether (a) disparities in physician diagnoses can be attributed to disparate impact based on patient race, and (b) information technology-enabled health information sharing among healthcare providers can mitigate such disparate impact. Our empirical context focuses on racial disparities in diagnoses of heart disease between Hispanic patients and non-Hispanic, White patients. Utilizing patient-level, emergency room (ER) encounter data from 2015 to 2022, we find statistical evidence of diagnostic disparate impact where the likelihood of Hispanic patients being diagnosed with heart disease is around three percentage points lower than White patients after accounting for their underlying race-specific risk of heart disease. However, we find that health information sharing can reduce the level of diagnostic disparate impact against Hispanic patients by 18% and the likelihood of severe disparate impact by seven percentage points. We evaluate the robustness of our results using a range of specifications, such as instrumental variable estimation, falsification tests, and alternative measures of disparate impact. We also highlight the underlying mechanism that explains the role of health information sharing in mitigating diagnostic disparate impact. Specifically, we show that health information sharing between healthcare providers can reduce diagnostic uncertainty, especially for Hispanic patients, and low-skilled physicians benefit more from health information sharing compared with highly skilled physicians.

History: Accepted by D. J. Wu, information systems.


Open Access Statement: This work is licensed under a Creative Commons Attribution 4.0 International License. You are free to copy, distribute, transmit and adapt this work, but you must attribute this work as "Management Science. Copyright © 2026 The Author(s). <https://doi.org/10.1287/mnsc.2024.05617>, used under a Creative Commons Attribution License: <https://creativecommons.org/licenses/by/4.0/>."

Funding: The authors gratefully acknowledge the McCombs Dean's Excellence Research Grant for this research.

Supplemental Material: The online appendix and data files are available at <https://doi.org/10.1287/mnsc.2024.05617>.

Keywords: health information sharing • diagnostic disparity • disparate impact • healthcare providers • information exchange

1. Introduction

Health disparities pose a significant challenge to the U.S. healthcare system because they represent significant differences in health outcomes across different population groups (Arcaya et al. 2015). Such disparities may be caused, in part, by differences in diagnostic outcomes for different demographic groups, which are commonly referred to as diagnostic disparities. Observed diagnostic disparities may indicate heterogeneity in disease risk among patients, but they may also reveal *disparate impact in physician diagnoses*, the primary focus of our study. Specifically, disparate impact in diagnoses refers to systematic differences among population groups in

their diagnosis rates of a disease after controlling for the underlying, group-specific risk of the disease.

Whereas a large body of research has investigated the impact of information technology (IT) and IT-enabled health information sharing on the quality and cost of patient care (Vest et al. 2015, Yaraghi 2015, Adjerdid et al. 2018, Janakiraman et al. 2023), a noticeable gap exists in studying the role of health information sharing through the lens of disparities. A small body of recent literature has suggested that health IT may mitigate disparities in access to care that are driven by geographic factors and disparities in treatment with respect to certain medical procedures (Ganju et al. 2020, Hwang et al. 2022).

However, there has been limited attention on the role of IT-enabled health information sharing in mitigating diagnostic disparities, particularly disparate impact in diagnoses.

Motivated by these observations, we seek to answer two research questions. First, *is there evidence of disparate impact in physician diagnoses based on patient race?* To the best of our knowledge, little empirical research has systematically quantified disparate impact in diagnostic decisions. Establishing evidence of diagnostic disparate impact is critical to understand how seemingly neutral diagnostic practices may disproportionately disadvantage certain demographic groups, even in the absence of explicit intent to discriminate. Furthermore, it provides an essential foundation to explore strategies and policies that can mitigate these disparities and improve equity in healthcare delivery.

Besides identifying the presence of disparate impact in physician diagnoses, a more important question is: *how can it be effectively mitigated?* Answering this question can contribute valuable insights to ongoing efforts to improve healthcare delivery and promote healthcare equity and fairness. Motivated by recent government-led efforts (such as Healthy People 2030) that aim to increase the proportion of hospitals that exchange electronic health information with outside providers, our second research question can be posited as follows: *how does diagnostic disparate impact change when healthcare providers have access to patient health information shared by other hospitals?* In other words, our aim is not only to detect whether there is evidence of disparate impact in physician diagnoses but, more importantly, to propose potential solutions to alleviate such health disparities.

Building on the prior literature on disparate impact, our key argument is that diagnostic disparate impact based on patient race may exist because of noisier risk signals received from a minority patient group during the diagnostic process (Aigner and Cain 1977, Hull 2021, Arnold et al. 2022, Bohren et al. 2022).¹ When patients are at high risk for a disease, lower-quality risk signals from minority patients could lead to systematically lower diagnosis rates compared with majority patients, even after controlling for their underlying disease risk. However, because IT-enabled health information sharing enables physicians to access more comprehensive health information for minority patients, it improves the quality of risk signals and reduces diagnostic uncertainty, particularly for the minority group. As a result, enhanced access to integrated patient data by physicians can effectively reduce the gap in risk signal quality between the majority and minority groups, leading to a reduction in disparate impact.

To test these predictions, we focus on racial disparities in physician diagnoses of heart disease between non-Hispanic, White patients and Hispanic patients as our empirical context. Our choice of empirical setting is

motivated by the “Hispanic paradox” reported in the literature (Swenson et al. 2002, Willey et al. 2012, Medina-Inojosa et al. 2014). Specifically, even though Hispanics are reported to have poorer overall health status and higher prevalence of risk factors for heart disease such as diabetes and obesity, they also exhibit a lower heart disease rate compared with the majority (White) population in the United States. Although a range of factors could explain such a disparity, we are interested in examining whether the observed disparity may be partly attributed to potential disparate impact in diagnoses and, if so, whether IT-enabled health information sharing can address this important concern.

Our primary data set is longitudinal patient encounter data obtained from a regional health information exchange (HIE) in central Texas. We focus on emergency room (ER) encounters because they allow for quasirandom assignment of physicians to patients, which is critical for identification of disparate impact. Our analysis sample comprises 336,413 ER encounters involving 91,580 White and Hispanic patients with cardiovascular conditions and 401 physicians from 2015 to 2022. In this setting, the overall heart disease diagnosis rate for Hispanic patients is 5.8 percentage points lower than the diagnosis rate for White patients. To investigate whether this observed disparity can be partly attributed to potential disparate impact, we adopt a novel approach to control for the underlying heart disease risk of the two patient populations (Arnold et al. 2022). Our results provide robust evidence of the existence of diagnostic disparate impact in this setting, where the likelihood of Hispanic patients being diagnosed with heart disease is around three percentage points lower than White patients, even after accounting for the differences in their underlying race-specific heart disease risk.

Next, we estimate the impact of health information sharing in alleviating diagnostic disparate impact at the physician-hospital-year level.² Specifically, we measure the extent of interorganizational health information sharing experienced by physicians at a given hospital. Our results indicate that health information sharing reduces the level of diagnostic disparate impact by 18% and the likelihood of severe disparate impact by seven percentage points. Our results are consistent across a range of robustness checks such as instrumental variable estimation and falsification tests.

Furthermore, we provide direct empirical evidence on the difference in diagnostic uncertainty between White and Hispanic patients. We show that higher levels of health information sharing can reduce diagnostic uncertainty, and such an effect is especially stronger for Hispanic patients. Our analysis also suggests that health information sharing is particularly beneficial to physicians with lower skill, who may encounter a bigger gap in the quality of risk signals between majority and minority patients. These findings are consistent

with our proposed mechanism that health information sharing can mitigate diagnostic disparate impact by closing the gap in the quality of risk signals between majority and minority patients.³

Our research represents one of the first studies to provide empirical evidence of racial disparate impact in physician diagnoses. Although the extant literature has acknowledged healthcare disparities in diagnosis (e.g., Kim et al. 2018, Shi et al. 2021), access to care (e.g., Moriya and Chakravarty 2023), and treatment (Cooper and Roter 2003, Ashraf et al. 2023), it does not distinguish between disparate treatment and disparate impact, which exhibit distinct legal elements and require unique empirical identification. To the best of our knowledge, ours is among the first to measure *disparate impact* in diagnostic decisions. We believe this is an important health policy question because disparate impact based on race reflects broader systemic inequalities, which not only includes a “direct” effect based on race but also an “indirect” effect based on nonrace characteristics such as socioeconomic barriers that are highly correlated with race.

Empirically identifying the existence of disparate impact is challenging because it requires controlling for legitimate justification (such as the underlying disease risk) to address omitted variable bias (OVB) and exclusion of nonrace factors as controls to address included variable bias (IVB) (Pager and Shepherd 2008, Rubineau and Kang 2012, Arnold et al. 2022).⁴ Following a novel approach proposed by Arnold et al. (2022), we take a first step toward measuring diagnostic disparate impact in a healthcare context by leveraging the quasirandom assignment of physicians in the ER setting, which effectively addresses both IVB and OVB concerns.

More importantly, our research contributes to the literature on the impact of IT-enabled information sharing on patient outcomes (Ayabakan et al. 2017, Janakiraman et al. 2023). Our study is most relevant to the recent body of research that has studied the influence of health IT on disparities related to access to care and treatment decisions among various population groups (Ganju et al. 2020, Hwang et al. 2022). We contribute to this literature by demonstrating that IT-enabled health information can narrow the gap in risk signal quality between majority and minority patients and reduce diagnostic uncertainty, particularly for minority patients, thereby mitigating diagnostic disparate impact based on race. Furthermore, we study the effect of health information sharing in mitigating disparate impact across physicians of different skill levels. In doing so, our research not only highlights the potential of health information sharing to alleviate diagnostic disparate impact but also provides policymakers with targeted strategies to improve health equity.

Our research also contributes to the ongoing debate on the effectiveness of national initiatives aimed at

promoting health information sharing. Since the passage of the Health Information Technology for Clinical Health (HITECH) Act in 2009, concerted national efforts have been directed at standardizing the process of sharing patient health data across healthcare providers. Notably, we observe that Fast Healthcare Interoperability Resources (FHIR) standards were developed to expedite, simplify, and improve methods for sharing clinical and administrative data and are now being adopted by hospitals.⁵ Our study extends recent research by demonstrating the role of health information sharing in reducing racial disparate impact in diagnoses. It also offers valuable guidance for policymakers in designing an equitable healthcare system that addresses the needs of individuals from diverse racial and ethnic backgrounds.

2. Theoretical Motivation

2.1. Disparate Impact in Physician Diagnoses

The legal foundation for the concept of disparate impact can be traced back to the landmark case of *Griggs v. Duke Power Company* (U.S. Supreme Court 1971). In this case, Duke Power Company faced restrictions in deploying prerequisites, such as high school diplomas or intelligence tests, because these requirements were correlated with race and could result in an indirect discriminatory impact on minority workers. In essence, disparate impact refers to the overall discriminatory effects of a policy or a practice. Therefore, it can encompass both direct effects based on certain group labels such as race and indirect effects through other factors related to the group label, such as socioeconomic status. A policy or decision-making process that appears race neutral can still result in disparate impact. For example, the literature reveals the presence of disparate impact in algorithmic decision making across various domains, including mortgage lending, personalized medicine dosage, and insurance rate setting (Fu et al. 2022, Kallus et al. 2022, Zhang and Xu 2023). In healthcare, even when patient race is not directly used to predict disease risk, the diagnostic decision process can still lead to disparate impact because of indirect effects.

To motivate our theoretical framework, we draw on the prior literature to develop a model that illustrates why disparate impact based on patient race may exist in physician diagnoses and how patient health information sharing can mitigate disparate impact (Balsa and McGuire 2001, McGuire et al. 2008, Arnold et al. 2022). Physician j 's diagnosis decision for patient i can be denoted as $D_{ij} \in \{0, 1\}$. Then, patient i is diagnosed with a certain disease by physician j (i.e., $D_{ij} = 1$) with the likelihood of $P_j(S_{ij}, R_i)$. Here, R_i indicates patient race, which could belong to either a majority or a minority group. Physician j observes patient i 's noisy signals $S_{ij} = Z_i^* + \varepsilon_{ij}$. Suppose that the underlying disease risk Z_i^* follows a normal distribution $N(\mu, \sigma_z^2)$ and the error

terms ε_{ij} are also normally distributed, with mean zero and variance σ_{jr}^2 . A lower σ_{jr}^2 indicates a less noisy risk signal observed by physician j regarding race r (McGuire et al. 2008, Arnold et al. 2022).

Following the literature, we assume that physicians use Bayes' rule to update their priors about patients' disease risks (Balsa and McGuire 2001). Specifically, given the signal S_{ij} , the expected disease risk $E(Z_i^*|S_{ij}) = (1 - \beta_{jr})\mu + \beta_{jr}S_{ij}$. It is a weighted average of the population mean risk μ and observed individual signal S_{ij} . The weight on the observed individual signal is β_{jr} , where $\beta_{jr} = \sigma_Z^2 / (\sigma_Z^2 + \sigma_{jr}^2)$. The noisier the signal (σ_{jr}^2), the smaller the weight (β_{jr}) attributed to the observed individual signal (S_{ij}). Furthermore, we consider the probability of patient i being diagnosed by physician j (denoted as $P_j(S_{ij}, R_i)$) as a strictly increasing function of the expected risk $E(Z_i^*|S_{ij})$. This is because the higher the expected risk of disease, the higher the probability of being diagnosed with the disease. Hence, the probability of patient i being diagnosed with the disease by physician j can be expressed as shown in Equation (1).

$$P_j(S_{ij}, R_i) = f(E(Z_i^*|S_{ij})) = f\left(\left(1 - \beta_{jr}\right)\mu + \beta_{jr}S_{ij}\right),$$

$$\text{where } \beta_{jr} = \sigma_Z^2 / (\sigma_Z^2 + \sigma_{jr}^2). \quad (1)$$

Majority and minority patients may differ significantly in the quality of risk signals regarding their disease. That is, we would expect that the signal noise is larger for minority patients compared with majority patients (i.e., $\sigma_{jMinority}^2 > \sigma_{jMajority}^2$), for reasons as follows. A critical factor is the underrepresentation of minority populations in clinical research and development of diagnostic tools. Minority patients also tend to have lower access to healthcare because of socioeconomic barriers, and hence, the amount and accuracy of data on minority patients are limited. Accordingly, many clinical risk factors and diagnostic models are developed using clinical studies or trials that predominantly include data from majority populations (Bibbins-Domingo and Helman 2022, pp. 23–33). Hence, underrepresentation of minority populations can lead to less clarity in interpreting their risk signals.

Second, social determinants of health, such as income, insurance, education, and living conditions, disproportionately affect minority populations and contribute further to the noise associated with risk signals (Cole and Nguyen 2020, White-Williams et al. 2020, Hill-Briggs et al. 2021). Hispanic populations experience greater social disadvantages compared with White populations, which influences their healthcare utilization and outcomes (Javed et al. 2022). These social factors can impact clinical presentations, introducing additional variability that complicates diagnosis and treatment.

Minority patients are also reported to experience poorer communication with physicians because of language barriers or cultural differences (Cooper et al. 2002, Saha et al. 2011). These factors can impair physicians' understanding of their health conditions, contributing to noisier risk signals.

Third, minority patients often have more fragmented medical histories across hospitals than the majority group (López et al. 2011, Kaltenborn et al. 2021). This fragmentation of patient data can undermine the effectiveness of care coordination, complicating the process of consolidating health information from different providers (Pinheiro et al. 2021, Patiño-Benavidez et al. 2024). This can hinder physicians' ability to make informed decisions, increasing the uncertainty in their risk assessments and diagnosis of minority patients.

In high-risk settings such as the ER, it is reasonable to assume that $S_{ij} > \mu$. In such settings, if $\sigma_{jMinority}^2 > \sigma_{jMajority}^2$, the expected disease risk $E(Z_i^*|S_{ij})$ for minority patients will be consistently lower than that of majority patients. Accordingly, the likelihood $P_j(S_{ij}, R_i)$ of being diagnosed with the disease for minority patients ($R_i = \text{minority}$) is systematically lower compared with majority patients ($R_i = \text{majority}$), resulting in disparate impact. When disparate impact is caused by systematic differences in the quality of signals between two groups, it is often referred to as statistical discrimination (Arrow 1971, Phelps 1972, Aigner and Cain 1977). Prior research has documented qualitative evidence of statistical discrimination in patient care, such as diabetes and depression treatment (Lutfey and Ketcham 2005, McGuire et al. 2008), but there has been limited empirical evidence that systematically quantifies this effect in patient diagnoses.

Figure A.1 in Online Appendix A further illustrates how disparate impact based on patient race may exist in diagnoses. Because $\sigma_{jMinority}^2 > \sigma_{jMajority}^2$, the rise in expected disease risk $E(Z_i^*|S_{ij})$ is less pronounced for minority patients than majority patients per unit increase in the observed signals (i.e., $\Delta E(Z_i^*|S_{ij}) / \Delta S_{ij} = \beta_{jr} = \sigma_Z^2 / (\sigma_Z^2 + \sigma_{jr}^2)$ and $\beta_{jMinority} < \beta_{jMajority}$). Further, when the disease signals of patient i detected by physician j is greater than the population mean (i.e., $S_{ij} > \mu$) for the same observed disease signal S_{ij} , the low signal quality of minority patients may lead them to exhibit lower expected risk $E(Z_i^*|S_{ij})$. That is, the orange dashed line (representing minority patients) will always be lower than the blue solid line (representing majority patients) when $S_{ij} > \mu$. Because the probability of patient i being diagnosed with the disease by physician j (i.e., $P_j(S_{ij}, R_i)$) is a strictly increasing function of $E(Z_i^*|S_{ij})$, we expect minority patients to have a lower likelihood of being diagnosed than majority patients when $S_{ij} > \mu$.

2.2. Health Information Sharing and Disparate Impact

Interorganizational information exchange between healthcare providers enables sharing of patient health information across hospitals, physicians, and laboratories (Adjerid et al. 2018). By granting physicians timely access to comprehensive patient information, health information sharing facilitates thorough evaluation of a patient's risk of disease, including access to their prior medical history, diagnoses, medications, and test results, from various healthcare providers (Ayabakan et al. 2017). Access to patient health information empowers physicians to make well-informed decisions. For example, health information sharing ensures that physicians can obtain patient medication history and allergies, allowing them to make treatment decisions that do not adversely interact with other ongoing treatments or medications. Furthermore, health information sharing contributes to enhanced coordination of care plans across disparate healthcare providers, reducing duplicative tests and treatments and avoiding medical errors. Prior research has demonstrated that health information sharing is associated with several benefits, such as reduction in redundant medical procedures, lower length of stay, and a reduction in readmission rates (Lammers et al. 2014, Ayabakan et al. 2017, Janakiraman et al. 2023).⁶

As discussed previously, minority patients often have noisier risk signals compared with majority patients because of factors such as fragmented care histories, socioeconomic constraints, and limited access to diagnostic resources. This gap in signal quality is reflected in the differences in diagnostic uncertainty between minority and majority groups (Lossos et al. 1989, Balsa et al. 2003, Alam et al. 2017, Atolagbe et al. 2024). Health information sharing can significantly improve the quality of disease signals and reduce diagnostic uncertainty especially for minority patients because of the following reasons. First, information sharing addresses gaps in signal quality caused by fragmented care histories, which are more common among minority patients, because they tend to receive care across multiple providers (López et al. 2011, Kaltenborn et al. 2021). By aggregating care records from different providers, including diagnoses, treatments, and test results, health information sharing mitigates the informational gaps that lead to uncertainty in diagnosis. Second, by integrating historical medical data and test results, health information sharing reduces the need for redundant or costly diagnostic procedures. This, in turn, enables physicians to make precise diagnostic decisions without imposing additional financial burden on patients, particularly for minority patients who are often constrained by insurance coverage. Third, health information sharing can provide critical care-related information that patients may find difficult to convey themselves. This includes

details such as past visit records, test results, and medication history, which may otherwise be hindered by challenges related to communication and education.

In sum, information sharing of patient health information across healthcare providers can help physicians improve the quality of risk signals about individual patients and reduce uncertainty in diagnosis, particularly for minority patients. Based on Equation (1), we argue that health information sharing can enhance signal quality (σ_{jr}^2), particularly among minority patients, so that the gap between $\sigma_{jMinority}^2$ and $\sigma_{jMajority}^2$ becomes narrower. Consequently, with the same observed risk signal S_{ij} , the difference in $E(Z_i^* | S_{ij})$ between majority and minority patients should become smaller, which, in turn, reduces disparate impact. Consider an extreme case where health information sharing completely removes the noise for all patients (i.e., $\sigma_{jMinority}^2 = \sigma_{jMajority}^2 = 0$ and $\beta_{jr} = 1$ across two races). As shown in Figure A.2 in Online Appendix A, the expected disease risk $E(Z_i^* | S_{ij})$ would be the same for majority and minority patients for the same signals S_{ij} . Hence, when physicians have greater access to patient health information, it can help narrow the gap in signal quality and diagnostic uncertainty between majority and minority patient groups. As a result, patient health information sharing may alleviate disparate impact.

3. Research Setting

We study racial disparities in physician diagnoses of heart disease between Hispanic and White patients in the context of ER visits. We choose ER visits as our research setting because of several reasons. First, as discussed in our theoretical motivation, we are interested in high-risk settings where the likelihood of being diagnosed with a certain disease for minority patients is systematically lower compared with majority patients. The ER setting serves as an ideal empirical context for this purpose, as the health status of ER patients is typically more severe and acute compared with those observed in routine outpatient settings. Second, the ER setting allows for quasirandom assignment of physicians to patients, which is critical for identification of disparate impact. According to the U.S. Emergency Medical Treatment and Active Labor Act (EMTALA), if an emergency medical condition such as heart attack is reported, the individual is typically treated at the closest hospital equipped to provide the necessary treatment to stabilize the patient. Further, EMTALA requires that emergency departments provide a medical screening examination and stabilizing treatment to all individuals, regardless of their ability to pay or their insurance status. Because ER visits related to heart conditions are often accompanied by acute symptoms such as chest pain, patients are unlikely to plan for an ER visit and self-select specific physicians a priori.

On the physician side, the process of assigning physicians to the ER is based on hospital-wide scheduling as well as individual physician availability, which further contributes to the random assignment of physicians to patients (Lu and Rui 2018). Additional analyses, as reported in Online Appendix B, suggests no evidence that patients with certain demographic or clinical characteristics are assigned to physicians with specific levels of skill or experience. These analyses support the assumption of quasirandom assignment of patients in the ER, which is consistent with the literature.

In this study, we examine three types of conditions that fall under the category of heart disease, namely, acute myocardial infarction (heart attack), coronary heart disease, and angina. Our focus on heart disease stems from the concern that it represents a substantial health burden and a leading cause of death in the United States and worldwide. According to the Centers for Disease Control and Prevention (CDC), it accounted for more than 650,000 deaths in 2019 (CDC 2021).⁷ Further, heart disease ranks among the most commonly reported conditions for ER visits (Eichelberger et al. 2020). Hence, the insights from studying this disease can add to a growing body of literature that seeks to understand racial disparities associated with heart disease (Jha et al. 2003, Brown et al. 2022, Javed et al. 2022).

Our interest in investigating disparate impact in the diagnoses of heart disease between White and Hispanic patients is driven by the “Hispanic paradox” in the literature (Markides and Coreil 1986). First, as shown in the top half of Figure 1, data from the National Center for Health Statistics indicate that Hispanics exhibit lower overall disease rates for the three most common types of

heart disease compared with the White population. However, the data in the bottom half of Figure 1 indicate that Hispanics experience poorer overall health status and higher prevalence of risk factors for heart disease, such as diabetes and obesity, compared with Whites. A variety of factors, such as diet, lifestyle, cultural factors, and potential measurement error, have been proposed to explain this paradox (Swenson et al. 2002, Willey et al. 2012, Medina-Inojosa et al. 2014). Although it is beyond the scope of our research to identify the underlying causes of the observed (lower) disease rate in the Hispanic population, we seek to understand whether such observed disparities may be partially explained by disparate impact during the diagnostic process.

4. Empirical Approach

4.1. Measuring Disparate Impact

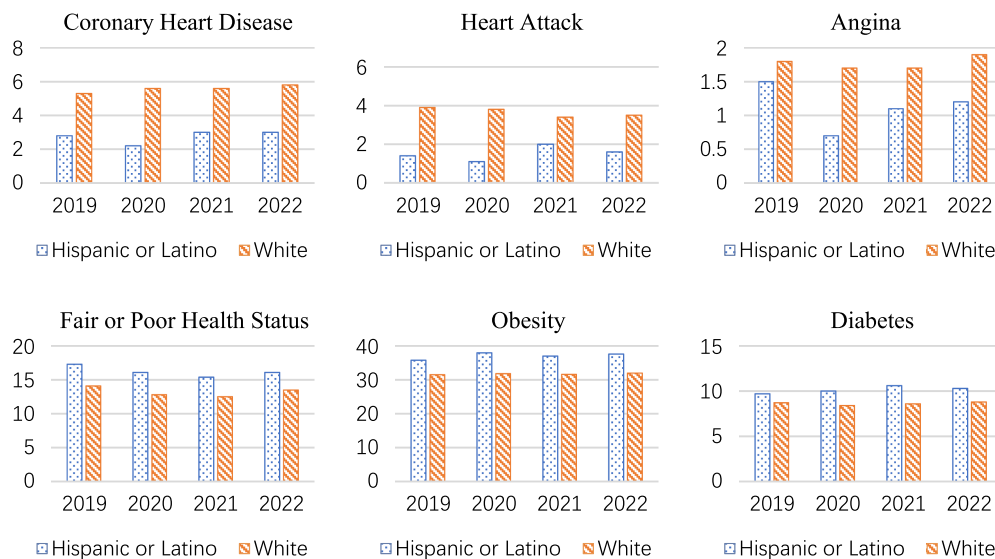
In this section, we describe our adaptation of the novel approach developed by Arnold et al. (2022) to measure racial disparate impact in physician diagnosis. Our empirical framework is described in Equations (2) to (4) as follows.

$$\Delta_{j0} = E[D_{ij} | R_i = \text{Hispanic}, H_i^* = 0] - E[D_{ij} | R_i = \text{White}, H_i^* = 0] \quad (2)$$

$$\Delta_{j1} = E[D_{ij} | R_i = \text{Hispanic}, H_i^* = 1] - E[D_{ij} | R_i = \text{White}, H_i^* = 1] \quad (3)$$

$$\begin{aligned} \Delta_j &= E[E[D_{ij} | R_i = \text{Hispanic}, H_i^*] - E[D_{ij} | R_i = \text{White}, H_i^*]] \\ &= \Delta_{j0}(1 - \mu) + \Delta_{j1}\mu. \end{aligned} \quad (4)$$

Figure 1. (Color online) Incidence (%) of Various Diseases for Adults 18 Years and Older, United States, 2019–2022



Source. National Center for Health Statistics. National Health Interview Survey. Generated on September 30, 2023, from https://www.cdc.gov/NHISDataQueryTool/SHS_adult/index.html.

$D_{ij} \in \{0, 1\}$ is the binary diagnosis decision made by physician j for patient i who has an ER visit because of certain cardiovascular conditions that share similar symptoms with heart disease. D_{ij} is equal to one if patient i is diagnosed by physician j with heart disease, and zero otherwise. R_i represents the race of patient i , and H_i^* represents patient i 's risk level of heart disease. We focus on the racial disparity in diagnosis rate between Hispanic and White patients. Hence, patients of other races, such as African American, Asian, American Indian, Native Hawaiian, and Pacific Islanders, are excluded. For simplicity, we assume that $H_i^* \in \{0, 1\}$, so H_i^* is equal to one if patient i has a high risk of heart disease when left undiagnosed, and zero if patient i has a low risk of heart disease when left undiagnosed.⁸ The mean risk of heart disease $-\mu$ can then be written as $-\mu = E[H_i^*] = \sum_r E[H_i^* | R_i = r] * Prob(R_i = r)$.

Δ_{j0} is the disparity in diagnosis rates between Hispanic and White patients, given their hidden risk of heart disease being low. Δ_{j1} is the disparity in the diagnosis rate between Hispanic and White patients, given their hidden risk of heart disease being high. In the first line of Equation (4), the inner difference represents the difference in physician j 's diagnosis rate between Hispanic and White patients with similar risk of heart disease. The outer expectation is the average across the entire distribution based on patient heart disease risk. Hence, Δ_j measures the *average diagnostic disparate impact* exhibited by physician j . A negative (and positive) Δ_j indicates that physician j tends to diagnose Hispanic patients with heart disease at a lower (and higher) rate than White patients, given the same underlying heart disease risk. If $\Delta_j = 0$, then physician j exhibits no difference in diagnosis rate of heart disease between the two racial groups with the same risk.

In our data, we observe patient demographic information such as race ($R_i \in \{Hispanic, White\}$) for patient i , the assignment of patient i to physician j ($P_{ij} \in \{0, 1\}$), and the diagnosis made by physician j regarding patient i in the focal ER visit (D_{ij}). The diagnosis outcome of patient i (D_i) can be written in more general form as $D_i = \sum_j D_{ij} \times P_{ij}$. We can also observe diagnosis decisions made by physician j for patient i across subsequent hospital visits and their corresponding diagnosis outcomes (denoted as $D2_{ij} \in \{0, 1\}$ and $D2_i = \sum_j D2_{ij} \times P_{ij}$).

Following Arnold et al. (2022), a key assumption in our empirical approach is as follows. If patient i is not diagnosed with heart disease during a focal visit ($D_i = 0$) but receives a positive diagnosis during *any* subsequent hospital visit ($D2_i = 1$) within a given time window after the focal visit, the risk of having heart disease would be high; that is, $H_i^* = 1$. If patient i is not diagnosed with heart disease during the focal and *all* subsequent hospital visits within this time window (i.e., $D_i = 0$ and $D2_i = 0$), the risk of heart disease would be low; that is, $H_i^* = 0$.⁹

However, an important empirical challenge occurs when patient i is diagnosed with heart disease during the focal visit ($D_i = 1$), but we do not know patient i 's risk of heart disease (H_i^*) because they have already been treated. In other words, we cannot tell from future visits ($D2_i$) what would have happened if patient i had not been diagnosed during the focal visit. Because H_i^* is not observed for patients with positive diagnosis of heart disease in the focal visit, we cannot simply add it as a control variable in the regression. Hence, without controlling H_i^* , the observed disparity for physician j (α_j) is expressed as

$$\begin{aligned} \alpha_j &= E[D_{ij} | R_i = Hispanic, P_{ij} = 1] \\ &\quad - E[D_{ij} | R_i = White, P_{ij} = 1] \quad (5) \\ &= E[D_{ij} | R_i = Hispanic] - E[D_{ij} | R_i = White] \quad (6) \end{aligned}$$

Equation (6) is based on the random assignment of a physician to the patient. Compared with Equation (4), the measured disparity α_j can be affected by the underlying risk of heart disease. For instance, it is possible that Hispanic patients are diagnosed at a lower rate because their underlying risk of heart disease may be lower than White patients. Failing to account for this risk (H_i^*) can lead to omitted variable bias, so α_j may deviate from Δ_j .

A common approach is to include patient-specific factors as controls for the underlying risk of heart disease. However, as suggested by several studies, measuring disparate impact using this approach can introduce IVB because disparate impact measures overall discriminatory effects (Ayres 2010, Jung et al. 2024). For example, adding a control such as insurance status, which is strongly correlated with race, may lead to underestimation of disparate impact in diagnoses.

To address these challenges, we adopt the identification strategy proposed by Arnold et al. (2022) and Baron et al. (2024). The key idea is to estimate disparate impact (Δ_j) through a quasiexperimental approach. Specifically, expanding the inner component in the first line of Equation (4), we obtain the following expressions in Equations (7) and (8). The second line in both Equations (7) and (8) follows the assumption of random assignment of physicians to patients.

$$\begin{aligned} E[D_{ij} | H_i^* = 0, R_i = r] &= 1 - \frac{E[(1 - D_{ij})(1 - H_i^*) | R_i = r]}{E[1 - H_i^* | R_i = r]} \\ &= 1 - \frac{E[(1 - D_i)(1 - H_i) | R_i = r, P_{ij} = 1]}{1 - \mu_r} \quad (7) \end{aligned}$$

$$\begin{aligned} E[D_{ij} | H_i^* = 1, R_i = r] &= 1 - \frac{E[(1 - D_{ij})H_i^* | R_i = r]}{E[H_i^* | R_i = r]} \\ &= 1 - \frac{E[(1 - D_i)H_i | R_i = r, P_{ij} = 1]}{\mu_r} \quad (8) \end{aligned}$$

It is worth noting that because D_{ij} is directly observed, the numerator in the first line of Equation (7) and (8) can still be estimated when $D_{ij} = 1$, even if we do not know the underlying risk of heart disease (H_i^*). When $D_{ij} = 0$, as noted earlier, we use the patient’s future diagnosis of heart disease to approximate H_i^* , that is, $H_i^* = H_i$, where H_i indicates the observed diagnosis outcome of heart disease in future visits for the focal patient. The formal expression of disparate impact can be written by plugging Equations (7) and (8) into Equation (4).

$$\begin{aligned} \Delta_j = & E[(D_i - 1)(1 - H_i) | R_i = \text{Hispanic}, P_{ij} = 1] \frac{1 - \bar{\mu}}{1 - \mu_{\text{Hispanic}}} \\ & + E[(D_i - 1)H_i | R_i = \text{Hispanic}, P_{ij} = 1] \frac{\bar{\mu}}{\mu_{\text{Hispanic}}} \\ & - E[(D_i - 1)(1 - H_i) | R_i = \text{White}, P_{ij} = 1] \frac{1 - \bar{\mu}}{1 - \mu_{\text{White}}} \\ & - E[(D_i - 1)H_i | R_i = \text{White}, P_{ij} = 1] \frac{\bar{\mu}}{\mu_{\text{White}}} \\ = & E[\Omega_i(D_i - 1) | R_i = \text{Hispanic}, P_{ij} = 1] \\ & - E[\Omega_i(D_i - 1) | R_i = \text{White}, P_{ij} = 1], \end{aligned} \quad (9)$$

where $\Omega_i = (1 - H_i) \frac{1 - \bar{\mu}}{1 - \mu_{R_i}} + H_i \frac{\bar{\mu}}{\mu_{R_i}} > 0$.

Because $\bar{\mu} = \mu_{\text{Hispanic}} \text{Prob}_{\text{Hispanic}} + \mu_{\text{White}} \text{Prob}_{\text{White}}$, Ω_i is essentially a function of observed H_i and the unobserved mean heart disease risk for the two racial groups combined. In order to determine this factor Ω_i , the key is to determine the two parameters for race-specific heart disease risk: $\mu_{\text{Hispanic}} = E[H_i^* | R_i = \text{Hispanic}]$ and $\mu_{\text{White}} = E[H_i^* | R_i = \text{White}]$.

The average race-specific risk of heart disease (i.e., $\mu_{\text{Hispanic}}, \mu_{\text{White}}$) can be estimated from an extrapolation method proposed by Arnold et al. (2022). Suppose there is a physician who always diagnoses patients as not having heart disease, and, as a result, none of the patients diagnosed by this physician receive relevant treatment during the focal visit. Based on their future hospital visits, we can estimate their risk of heart disease (i.e., H_i^*). Assuming that this physician is randomly assigned to patients, the sample average of heart disease risk in Equation (10) provides a consistent estimate of population-level average risk of heart disease, that is, μ_{Hispanic} and μ_{White} .

$$\hat{\mu}_{\text{Hispanic}} = \text{Avg}_{R_i=\text{Hispanic}}(H_i); \hat{\mu}_{\text{White}} = \text{Avg}_{R_i=\text{White}}(H_i) \quad (10)$$

In the real world, however, it may not be possible to observe such a physician. In most cases, for each physician j , some of their patients are likely to receive a positive diagnosis ($D_{ij} = 1$), whereas others do not ($D_{ij} = 0$). Hence, we estimate the average risk of heart disease by extrapolating the variation in diagnosis rate of heart disease across physicians who are randomly assigned to patients.¹⁰ Specifically, we fit various regression models

(linear, quadratic, and local linear) of the estimated heart disease rate when left undiagnosed $E[H_i^* | D_{ij} = 0, R_i = r]$ on estimated diagnosis rates $E[D_{ij} | R_i = r]$ across physician j within each race r . The estimated mean risk of heart disease for the two racial groups (i.e., $\hat{\mu}_{\text{Hispanic}}, \hat{\mu}_{\text{White}}$) can be derived from the vertical intercept at diagnosis rate of zero (i.e., $D_{ij} = 0 \forall i$). Ω_i can then be estimated accordingly.

4.2. Effect of Patient Health Information Sharing on Disparate Impact

As suggested in Section 2, because of the noisier risk signals of minority patients in high-risk settings, they encounter systematically lower diagnosis rates than majority patients. Accordingly, the measure of disparate impact Δ_j in Equation (9) would not only show a high magnitude but also be negative. Hence, in our analyses, we are particularly interested in whether health information sharing influences the following outcome variables: the magnitude of Δ_j and an indicator that turns on when Δ_j is significantly negative. The first outcome is a direct measure of whether health information sharing narrows the gap in risk signal quality and thereby reduces the difference in diagnosis rates between majority and minority patients after controlling for the underlying disease risk. We examine the latter outcome because if minority patients experience a systematically lower diagnosis rate than majority patients (i.e., a very negative Δ_j), the progression of untreated disease could be exacerbated by their socioeconomic conditions, resulting in worse health outcomes. Hence, in our discussion, we denote this latter outcome as severe disparate impact against minority patients.

Accordingly, we deploy the following Model Specification (11) with the unit of analysis at the physician (j) – hospital (h) – year (t) level.

$$DI_{jht} = \beta_0 + \beta_1 \text{InfoShare}_{jht} + \beta_2 \mathbf{X}_{jt} + \text{Year}_{t+} \text{Physician}_j + \text{Hospital}_h + \varepsilon_{jht}. \quad (11)$$

In this model, the dependent variable DI_{jht} is measured in two ways, as discussed above. First, DI_{jht} is measured using the magnitude of Δ_{jht} , the level of diagnostic disparate impact of a physician j in hospital h in year t . Whereas Section 4.1 discusses our approach to estimate the overall disparate impact of physician j (Δ_j) based on all patients diagnosed by physician j , the diagnostic disparate impact of physician j can also be measured at the hospital h level for a particular year t (denoted as Δ_{jht}). That is, Δ_{jht} is measured based on Equation (9) using the set of patients diagnosed by physician j in hospital h in year t . Second, DI_{jht} is measured using the indicator of severe disparate impact against minority patients, which is equal to one if Δ_{jht} falls below its sample mean value minus one standard deviation, and equal to zero otherwise.

$InfoShare_{jht}$ is a dummy variable that indicates whether the level of shared patient health information used by physician j in hospital h in year t is high versus low. The vector X_{jt} includes a set of control variables. It includes a few variables that represent physician j 's experience in diagnosing patients with cardiovascular conditions in the ER setting, based on the cumulative count of ER encounters with cardiovascular conditions diagnosed by physician j up to year t and the proportion of Hispanic patients in ER encounters with cardiovascular conditions diagnosed by physician j up to year t . To control for the severity of patients treated by physician j in year t , we use the average number of comorbidities of ER encounters with cardiovascular conditions diagnosed by physician j in year t . Comorbidities are measured by the number of diagnosis codes assigned during each encounter.

We include year fixed effects ($year_t$) to control for general time trends. Physician fixed effects ($Physician_j$) and hospital fixed effects ($Hospital_h$) are also included to account for unobserved time-invariant differences at the physician and hospital levels. For analyses using the magnitude of Δ_{jht} as the outcome variable, we employ a linear model with ordinary least squares (OLS) regression. For analyses using the indicator of severe disparate impact as the outcome variable, we employ a linear probability regression model for better interpretability.

5. Data and Measures

We utilize proprietary data from a regional HIE in central Texas to examine diagnostic disparities between Hispanic and White populations for heart disease. Although Hispanics comprise approximately 33% of the population in this region, it is worth noting that the minority status of a racial group is defined by their social and economic disempowerment rather than only based on numerical representation (Blume 2013). A substantial body of research has documented that Hispanics face significant socioeconomic barriers even in states such as California and Texas, where they represent a relatively larger share of the population (Morales et al. 2002, Chando et al. 2013, Escobedo et al. 2023). Whereas a higher proportion of Hispanic patients may improve physician familiarity with this population, their underrepresentation in clinical research, the fragmentation of medical records, and broader social disadvantages can complicate diagnosis and contribute to noisier risk signals for Hispanic patients.

The HIE provides comprehensive encounter-level records of patient visits from 2015 to 2022 and covers a large majority of hospitals in central Texas. For each patient, the data set includes demographics such as race, gender, age, and residential zip code. For each patient encounter, the data set contains the encounter type, so we can distinguish ER visits from other types of

visits (such as scheduled outpatient visits), physician and hospital identifiers, and related diagnosis codes (ICD-9/ICD-10 codes and descriptions). Hence, the data set allows us to track each physician throughout this period to identify their diagnosis decisions for a particular patient, including the patient's encounter history and demographic information.

We focus on ER encounters associated with Hispanic and White patients. Further, to examine the diagnosis rate of heart disease, we focus on the sample of ER encounters for which patients experienced a range of cardiovascular conditions based on the set of diagnosis codes shown in Table 1. Last, to estimate disparate impact as shown in Equation (9), we need to ensure that a physician has diagnosed enough patients across the two racial groups. Therefore, we focus on physicians who had more than 100 ER patient encounters with cardiovascular conditions during our sample period. The final sample used to estimate the overall disparate impact by physician j (i.e., Δ_j) comprises 336,413 ER encounters, including 91,580 patients and 401 physicians.

The descriptive statistics of our sample are presented in Table 2. We observe that 52.5% of all patient encounters in the ER for cardiovascular conditions were attributed to Hispanic patients. Only 51% of Hispanic patients are insured, in contrast to 69% of insured White patients. These statistics are consistent with the generally lower socioeconomic status of Hispanic patients compared with their White counterparts. Hispanic patients also had more hospital visits than Whites during the six-month time window before a focal visit, suggesting poorer health status. However, consistent with the previously discussed Hispanic paradox, our data indicate that the diagnosis rate of heart disease for Hispanic patients is 5.8 percentage points (= 0.163 – 0.105) lower than that for

Table 1. Diagnosis Codes for Patient Sample Construction

Diagnosis code (ICD-10)	Description
I00-I02	Acute rheumatic fever
I05-I09	Chronic rheumatic heart diseases
I10-I16	Hypertensive diseases
I20-I25	Ischemic heart diseases
I26-I28	Pulmonary heart disease and diseases of pulmonary circulation
I30-I5A	Other forms of heart disease
I60-I69	Cerebrovascular diseases
I70-I79	Diseases of arteries, arterioles, and capillaries
I80-I89	Diseases of veins, lymphatic vessels, and lymph nodes, not elsewhere classified
I95-I99	Other and unspecified disorders of the circulatory system

Notes. This table provides a list of diagnoses that can be attributed to cardiovascular conditions in an ER setting. Although we only list the ICD-10 codes for brevity, the corresponding ICD-9 codes are also included in our analyses to identify patients with cardiovascular conditions.

Table 2. Descriptive Statistics

Variable	All patients (1)	Hispanic patients (2)	White patients (3)
<i>Hispanic</i>	0.525	—	—
<i>Male</i>	0.452	0.435	0.472
<i>Age</i>	59.59	54.39	65.33
<i>Insured</i>	0.597	0.514	0.690
<i>Hospital visits in the last six months</i>	6.275	6.982	5.495
<i>Diagnosis outcome</i>	0.133	0.105	0.163
<i>Risk of heart disease, when not diagnosed</i>	0.113	0.100	0.128
<i>Total Encounters</i>	336,413	176,557	159,856
<i>Total Patients</i>	91,580	41,126	50,454
<i>Encounter with negative diagnosis</i>	291,808	157,932	133,876

Notes. Our sample consists of ER encounters for patients with cardiovascular conditions from 2015–2022. *Diagnosis outcome* is defined as whether a patient is diagnosed with heart disease in their focal ER visit, that is, D_i , discussed in Section 4. *Risk of heart disease, when not diagnosed* is measured based on instances where a patient is not diagnosed with heart disease on a focal visit but diagnosed with heart disease in any subsequent hospital visits within six months of the focal visit, that is, $D2_i$, discussed in Section 4.

White patients. Our primary interest is to examine whether such an observed disparity can be partially explained based on diagnostic disparate impact.

5.1. ER Setting: A Quasirandom Assignment

As noted above, one of the reasons we focus on the ER is that it provides an ideal quasirandom assignment setting, which is critical for our empirical identification. To test this assumption, we investigate whether patients with specific demographic or clinical characteristics are systematically more likely to be treated by physicians with certain characteristics. Using a patient encounter as the unit of analysis, we study how patient characteristics, including race, comorbidities, gender, and age, are correlated with physician characteristics such as physician skill, overall experience, and experience in treating Hispanic patients.

The results of our analysis are reported in Tables B.1, B.2, and B.3 of Online Appendix B. Table B.1 examines

the relationship between physician skill and patient characteristics. Overall, we find no systematic pattern that suggests patients with certain demographic or clinical characteristics are assigned to physicians with specific skills. Similarly, Table B.2 shows no significant correlations between a physician’s overall experience and patient demographics or clinical characteristics. Table B.3 also indicates no statistically significant relationship between physician experience in treating Hispanic patients and patient race or other characteristics. Collectively, these results support the assumption of quasirandom assignment of physicians to patients.¹¹

5.2. Health Information Sharing

Whereas the HIE data allow us to estimate Δ_j in Equation (9) and Δ_{jht} and X_{jt} in Equation (11), we integrate the HIE data with data from the AHA IT supplement at the hospital-year level to measure $InfoShare_{jht}$ in Equation (11).¹² The AHA data set includes survey-based measurements of health information sharing at the hospital level from 2015 to 2022. Specifically, two variables, *PHIOUT* and *SOCINT*, are relevant for our study. Table 3 provides detailed descriptions of these variables. *PHIOUT* indicates the level of patient health information from outside sources used by physicians in the hospital. *SOCINT* indicates the level of patient health information from outside sources integrated into the electronic health record system of a hospital. Hence, the combination of these two variables forms our baseline measure of $InfoShare_{jht}$. It is equal to one when the values of both *PHIOUT* and *SOCINT*, which measure the use and availability of patient health information to the physician, respectively, are equal to one or two, and zero otherwise.

6. Results

We first present our empirical estimation of disparate impact against Hispanic patients, followed by the analyses on the role of health information sharing. Next, we provide additional empirical evidence that supports our underlying mechanism that explains how health information sharing between providers may reduce disparate diagnostic impact.

Table 3. Measures of Health Information Sharing

Variable	Description	Measure
<i>PHIOUT</i>	Do providers at your hospital use patient health information received electronically (not eFax) from outside providers or sources when treating a patient?	1 = Often; 2 = Sometimes; 3 = Rarely; 4 = Never; 5 = Do not know
<i>SOCINT</i>	Does your EHR integrate the information contained in summary of care records received electronically (not eFax) without the need for manual entry?	1 = Yes, routinely; 2 = Yes, but not routinely; 3 = No; 4 = Do not know; 5 = NA

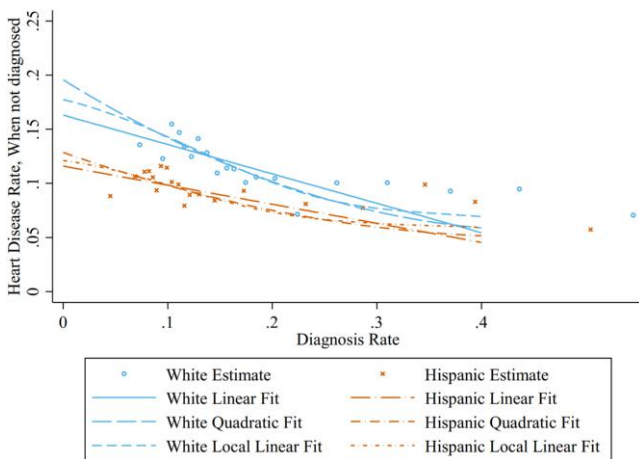
Notes. In the AHA data set, two other variables, *CIAOUT* and *EQPHIOS*, are also related to the level of availability of patient health information. However, we do not use *CIAOUT* because it asks the availability of information specifically on patients who were previously seen only by outside providers. Similarly, we do not use *EQPHIOS* because the question itself changed in 2019.

6.1. Disparate Impact

Figure 2 shows the extrapolation-based estimates of the mean risk of heart disease for Hispanic versus White patients in our sample. The horizontal axis plots the estimated physician- and race-specific diagnosis rates of heart disease for a focal encounter. The vertical axis plots the regression-adjusted, physician- and race-specific heart disease rates based on subsequent visits (within six months) for patients not diagnosed with heart disease for a focal encounter (i.e., $D2_i$ in Section 4.1). Following the linear extrapolation approach in Arnold et al. (2022), we plot two solid lines based on OLS regressions of physician-specific heart disease estimates among patients not diagnosed with heart disease (in their focal encounter) on physician-specific diagnosis rate estimates, weighted inversely by the variance of the former variable. We adopted a binned scatterplot approach as utilized in Baron et al. (2024). Specifically, we binned the estimates from 401 physicians into 20 equal-sized groups based on the estimated race-specific diagnosis rates and calculated the average estimated heart disease rates within each bin. Overall, Figure 2 reveals that a lower diagnosis rate in the focal encounter (on the horizontal axis) corresponds to a higher heart disease rate diagnosed in subsequent visits (on the vertical axis).

The vertical intercepts of the two solid lines in Figure 2 represent the estimates of the race-specific heart disease risk for Hispanic and White patients, that is, $\mu_{Hispanic}$ and

Figure 2. (Color online) Physician-Specific Diagnosis Rates and Conditional Heart Disease Rates



Notes. This figure shows a bin scatterplot (with 20 equal-sized bins) of race-specific diagnosis rates for 401 physicians against heart disease rates among patients not diagnosed with heart disease in their focal ER visits. All estimates are adjusted for hospital-time fixed effects. The figure also plots race-specific linear, quadratic, and local linear curves of best fit, obtained from physician-level regressions that are inverse weighted by the variance of the estimated heart disease rate among patients not diagnosed with heart disease in their focal ER visits. The local linear regressions use a Gaussian kernel with a race-specific rule-of-thumb bandwidth.

μ_{White} in Equation (9). These estimates are also reported in column (1) of panel A in Table 4. They suggest that the average (observed) heart disease risk of Hispanic patients in ER encounters with cardiovascular conditions is 4.7 percentage points (= 0.163 – 0.116) lower than that of White patients.

Next, we check the robustness of our results using quadratic extrapolation and local linear extrapolation methods. The quadratic and local linear curves are also plotted in Figure 2 and show consistent results. Columns (2) and (3) of panel A in Table 4 show the estimates of mean risk of heart disease by race based on the two alternate extrapolation methods. The estimated risk of both races is somewhat higher than the linear extrapolation method. We observe a sizable difference between the two races with respect to their mean risk, suggesting a difference of 6.7 (and 5.6) percentage points based on quadratic (and local linear) extrapolation methods, respectively.

Panel B in Table 4 reports the estimates of system-wide disparate impact Δ_j in Equation (9), using estimates of the race-specific heart disease risk for Hispanic and White patients from Figure 2. To calculate $\bar{\mu}$ in Equation (9), we use the sample share of Hispanic versus White patients to determine $Prob_{Hispanic}$ and $Prob_{White}$. Overall, as shown in column (1) of panel B, after controlling for the race-specific heart disease risk based on linear extrapolation, there still exists a significant difference in the diagnosis rates of heart disease between the two racial groups. Specifically, given the same underlying risk, Hispanic patients are 3.3 percentage points less likely to be diagnosed with heart disease compared with White patients. This represents around 57% of their observed disparity in heart disease diagnosis rate (which is 5.8 percentage points), as shown in Table 2. The results are consistent if we estimate race-specific heart disease risk based on the quadratic and local linear extrapolation methods, as reported in columns (2) and (3) of Table 4, respectively. Overall, our findings suggest both statistically and economically significant disparate impact in the diagnosis of heart disease among Hispanic and White patients, even after adjusting for their underlying risk of heart disease.

6.2. Role of Health Information Sharing

Next, we seek to understand whether IT-enabled health information sharing can mitigate diagnostic disparate impact. To estimate the diagnostic disparate impact Δ_{jht} in Equation (11), each physician j needs to diagnose enough patients across the two racial groups at the physician-hospital-year level. Hence, we focus on physicians who had more than 15 patient encounters with cardiovascular conditions in hospital h in year t .¹³ This results in a sample comprised of 622 physician-hospital-year combinations, with a total of 2,469 observations to estimate the impact of *InfoShare* on disparate impact in Equation (11).¹⁴

Table 4. Patient Risk of Heart Disease and Estimation of Disparate Impact in Diagnoses

Panel A: Mean risk of heart disease by race			
	Linear extrapolation (1)	Quadratic extrapolation (2)	Local linear extrapolation (3)
White patients	0.163*** (0.003)	0.196*** (0.007)	0.177*** (0.007)
Hispanic patients	0.116*** (0.002)	0.129*** (0.006)	0.121*** (0.006)
Panel B: System-wide disparate impact			
Mean across cases	−0.033*** (0.002)	−0.022*** (0.003)	−0.028*** (0.003)
Physicians	401	401	401

Notes. This table summarizes estimates of mean risk and disparate impact from different extrapolations of the variation in Figure 2. Panel A reports estimates of race-specific average heart disease risk, and panel B reports estimates of system-wide (case-weighted) disparate impact. Robust two-way standard errors, clustered at the patient and physician level, appear in parentheses.

*** $p < 0.01$.

Table 5 presents the estimation results for disparate impact. In our analyses, we use the linear extrapolation method as the baseline approach to estimate race-specific heart disease risk and disparate impact at the physician-hospital-year level. First, we report the estimation results without time-varying controls in column (1), followed by the estimates after their inclusion in column (2). We first report the results using the magnitude of Δ_{jit} (denoted as *Absolute Disparate Impact*) as the dependent variable and then the results using the indicator of severe disparate impact (denoted as *Severe Disparate Impact*) as the dependent variable. The estimate of *InfoShare* suggests that in hospitals where shared patient health information was available and utilized by physicians, there is a 0.017 (i.e., 18%) reduction in the level of diagnostic disparate impact. Furthermore, the impact of health information sharing reduces the likelihood of severe disparate impact against Hispanic patients by seven percentage points (i.e., 64%).

Next, we test the robustness of our results using both quadratic extrapolation and local linear extrapolation methods to estimate racial disparate impact in diagnosis. The results are reported in Tables B.4 and B.5 in Online Appendix B. In Table B.4, where disparate impact is estimated through quadratic extrapolation, utilization of shared patient health information leads to a 15% reduction in the level of disparate impact; it also reduces the likelihood of severe disparate impact against Hispanic patients by 5.9 percentage points. As shown in Table B.5, we obtain similar results if disparate impact is estimated through local linear extrapolation.

In sum, these results suggest that information sharing plays a crucial role in alleviating racial disparate impact in diagnosis between Hispanic patients and White patients. This could potentially lead to an overall improvement in the diagnosis and treatment of minority groups. Our results also highlight the importance of initiatives to promote data standards that enable

Table 5. OLS Estimation of Health Information Sharing on Disparate Impact

Dependent variable	Absolute disparate impact (1)	Absolute disparate impact (2)	Severe disparate impact (3)	Severe disparate impact (4)
<i>InfoShare</i>	−0.019*** (0.007)	−0.017*** (0.007)	−0.076*** (0.025)	−0.070*** (0.025)
<i>Cumulative ER encounters (in 1000s)</i>		−0.007*** (0.002)		−0.013*** (0.005)
<i>Share of Hispanic Patients</i>		0.033 (0.072)		−0.036 (0.209)
<i>Average Comorbidities</i>		−0.001 (0.004)		−0.014 (0.011)
Observations	2,469	2,469	2,469	2,469
Physician FE	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Note. Robust standard errors clustered at the physician level are reported in parentheses.

*** $p < 0.01$.

sharing of patient health information across healthcare providers.

6.3. Instrumental Variable Estimation

An important concern about the identification strategy in estimating Equation (11) is that omitted variables may be correlated with both the likelihood of disparate impact by a physician in a hospital and the level of health information sharing enabled by the hospital in a given year. For example, better-performing hospitals may be more likely to collaborate with external healthcare providers to promote health information sharing. By the same token, they may also be more likely to have physicians who are less likely to exhibit disparate impact against Hispanic patients.

To address this concern, we use the level of health information sharing in remote partner hospitals as the instrumental variable (IV) for $InfoShare_{jht}$, based on the prior literature (Atasoy et al. 2021, Bardhan et al. 2023). Remote partner hospitals are defined as those within the same health system but located in different hospital referral regions (HRRs) than the focal hospital. The IV is based on the average level of health information sharing among remote partner hospitals. Institutional theory posits that organizations are influenced by the structures, norms, and practices within their institutional environment, which leads them to exhibit similar behavior (Angst et al. 2017). Furthermore, Miller and Tucker (2009) suggest a network effect where the adoption of health IT at one hospital is influenced by adoption decisions of other hospitals within the same health system. Building on both institutional theory and network effects, greater use of health information sharing at remote partner hospitals increases the likelihood that physicians at the focal hospital adopt and use patient information shared by other providers. Hence, we argue that our choice of IV satisfies the relevance condition.

Next, this IV also satisfies the exclusion restriction for the following reasons. Patients typically receive a majority of healthcare services from hospitals within their own HRR (Lee et al. 2011, Miller and Tucker 2014, Atasoy et al. 2018). A study on patient care-seeking behavior shows that most patients seek care within a 10-mile radius of their primary residence and rarely travel more than 20 miles for healthcare services (Yen 2013). Therefore, decisions made by remote partner hospitals, such as HIE adoption, as well as the quality of information from patients outside the HRR, should not be correlated with patient diagnosis or outcomes at the focal hospital.

To assess the validity of our IV with respect to the exclusion restriction, we conduct empirical tests examining the relationship between our IV and the focal physician's clinical characteristics. Specifically, we estimate whether the IV is correlated with *patient flow* and *physician skill* in treating minority patients. Patient flow is

proxied using two variables—the number of ER encounters with cardiovascular conditions and number of ER encounters with cardiovascular conditions related to Hispanic patients—at the physician-hospital-year level. Physician skill in treating minority patients is measured as the average 30-day readmission rate of minority patients treated by physician j in year $t-1$.¹⁵ The exogeneity tests, as presented in Table B.6 of Online Appendix B, suggest no association between the IV and either patient flow or physician skill. Our results indicate that information sharing at remote partner hospitals does not directly affect the diagnostic decisions of physicians at the focal hospital. Hence, we conclude that the IV satisfies both relevance and exogeneity assumptions.

Following Miller and Tucker (2009) and Angst et al. (2017), we construct our IV for $InfoShare_{jht}$. Specifically, for physician j in hospital h in year t , we first identify hospital h 's HRR code and hospital system code. Then, we identified remote partner hospitals of h within the same hospital system but having a different HRR code. Next, based on the AHA data, we obtained the status of health information sharing availability of these remote partner hospitals using a similar measure as before (i.e., it is equal to one if $SOCINT = 1$ or 2 , and zero otherwise). We calculate the average level of health information sharing across remote partner hospitals, denoted as $Remote\ Partner\ InfoShare_{jht}$, and use this variable as the IV for $InfoShare_{jht}$.

The estimation results are reported in Table 6. We use this IV in a two-stage least squares (2SLS) regression with fixed effects to estimate the impact of patient health information sharing on the level of physician diagnostic disparate impact. The first stage results based on the IV estimation are reported in columns (1) and (2). The coefficient of the IV, $Remote\ Partner\ InfoShare$, is significantly and positively correlated with $InfoShare$ for both models, that is, without and with control variables. The F -statistic is 814.16 and above the critical value of 10 (Staiger and Stock 1997). This suggests that our IV satisfies the relevance condition. The second-stage results are reported in columns (3) and (4) for both models. The estimated coefficient of -0.054 for $InfoShare$ is significantly negative, suggesting that higher utilization of shared patient health information can significantly reduce physician diagnostic disparate impact. Overall, these results support our main finding on the role of health information sharing in mitigating diagnostic disparate impact.

Next, we conduct robustness analyses using 2SLS regressions with the same IV to estimate the impact of health information sharing on the likelihood of severe disparate impact where Hispanic patients experienced a significantly lower likelihood of being diagnosed with heart diseases than White patients. The results, as reported in Table B.7 of Online Appendix B (coefficient

Table 6. IV Estimation of Health Information Sharing on Disparate Impact

Dependent variable	First stage		Dependent variable	Second stage	
	(1)	(2)		(3)	(4)
	<i>InfoShare</i>	<i>InfoShare</i>		<i>Absolute Disparate Impact</i>	<i>Absolute Disparate Impact</i>
<i>Remote Partner InfoShare</i>	0.958*** (0.033)	0.951*** (0.033)	<i>InfoShare</i>	−0.059*** (0.018)	−0.054*** (0.019)
<i>Cumulative ER encounters (in 1000s)</i>		−0.002 (0.007)	<i>Cumulative ER encounters (in 1000s)</i>		−0.006*** (0.001)
<i>Share of Hispanic Patients</i>		−0.078 (0.192)	<i>Share of Hispanics</i>		0.038 (0.073)
<i>Average Comorbidities</i>		0.009 (0.009)	<i>Average Comorbidities</i>		0.001 (0.004)
Observations	2,469	2,469	Observations	2,469	2,469
Physician FE	Yes	Yes	Physician FE	Yes	Yes
Hospital FE	Yes	Yes	Hospital FE	Yes	Yes
Year FE	Yes	Yes	Year FE	Yes	Yes

Note. Robust standard errors clustered at the physician level are reported in parentheses.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

= −0.106; $p < 0.1$), are consistent with our baseline estimates.

6.4. Falsification Test

Although we have demonstrated that the use of shared patient health information mitigates diagnostic disparate impact, one concern is that its use is measured at the hospital level, whereas the likelihood of disparate impact (i.e., our dependent variable) is measured at the physician-hospital level. In other words, the extent to which a physician uses shared patient health information for diagnosis decisions may not be precisely captured by hospital-level reporting. Although we acknowledge this as a data limitation, our focus on the ER setting could partially mitigate this concern. That is, different from clinical settings where physicians may have greater discretion on whether to use prior patient health information shared by external providers, hospitals have established protocols that guide the use of shared patient health information in an ER setting, which physicians are required to follow. Hence, reporting physician use of shared patient health information (at the hospital level) likely reflects their actual behavior.

Nevertheless, we implement the following falsification exercise to further address this concern. Because our HIE data include a large majority of hospitals in central Texas, the observed health information about a patient mostly reflects what is shared among hospitals. A subset of patients in our sample had little recent health information at the time that they visited the ER for cardiovascular conditions. This subset provides a good opportunity to conduct a falsification test. That is, even if a hospital reports a high level of patient health information use by physicians, because these patients have no (or very limited) recent information that physicians can retrieve to make diagnostic decisions, we do

not expect to observe a reduction in disparate impact between Hispanic and White patients for these patients. On the other hand, if reporting of information use by hospitals does not reflect actual physician behavior and the observed effect is driven by unobservable variables, we would expect a similar effect for *InfoShare*.

We use a two-year window to determine the availability of recent health information for a given patient. That is, if a patient did not have any record in the HIE data for two years, we include them in this subsample analysis. Because our HIE data spans the period from 2015 to 2022, our subsample analysis focuses on the period from 2017 to 2022 so that we can determine patient health information-sharing status from 2017 onward, resulting in 27,515 ER encounters based on 22,717 White and Hispanic patients. We follow the same approach to measure disparate impact by physician j in hospital h in year t (i.e., Δ_{jht}) using this subsample. *InfoShare* is measured in the same way as in our baseline analysis, and we employ the same model specified in Equation (11).

The results based on this subsample of patients with no health information are reported in Table B.8 of Online Appendix B. Overall, we do not observe a significant effect of health information sharing on the likelihood on diagnostic disparate impact or severe diagnostic disparate impact against Hispanic patients. This falsification test provides further evidence that our hospital-level reporting of the use of shared patient health information does reflect *actual use* of such information by physicians, and the observed effect in our baseline analysis is not driven by omitted variables.

6.5. Mechanism

As discussed in Section 2, the key mechanism that we propose is that disparate impact in diagnosis may arise

from a systematic difference in the quality of risk signals between two racial groups (i.e., diagnostic uncertainty). Patient health information sharing mitigates diagnostic disparate impact because it reduces such a gap in signal quality between majority and minority patients. To further substantiate this underlying mechanism, we analyze the effect of health information sharing on disparities in signal quality, as reflected in diagnostic uncertainty.

According to the CMS ICD-10-CM guidelines, symptom codes are used as a substitute for diagnosis codes when a definitive diagnosis has not yet been established, and they are often used as a proxy for uncertainty in diagnoses (Atolagbe et al. 2024).¹⁶ For example, symptom codes are utilized if a patient presents with chest pain but the physician is uncertain about determination of a specific diagnosis. Further, it is inappropriate to select a specific code that is unsupported by medical documentation or perform unnecessary diagnostic testing solely to assign a more specific code. Hence, the use of *symptom codes* indicates insufficient or ambiguous diagnostic signals, suggesting a lack of clear information needed to establish a diagnosis. In other words, a higher probability of receiving symptom codes indicates greater uncertainty in diagnosis and, consequently, lower signal quality.

Building on the approach by Lossos et al. (1989), we utilize the likelihood of receiving a chest pain symptom code to study how health information sharing influences diagnostic uncertainty.¹⁷ Specifically, we deploy the model in Equation (12) to study whether (a) Hispanic patients face more diagnostic uncertainty, and (b) health information sharing reduces diagnostic uncertainty, particularly for Hispanic patients.

$$\begin{aligned}
 \text{Diagnostic Uncertainty}_{ijt} = & \beta_0 + \beta_1 \text{InfoShare}_{jht} \\
 & + \beta_2 \text{Hispanic}_i + \beta_3 \text{InfoShare}_{jht} \\
 & \times \text{Hispanic}_i + \beta_4 \mathbf{X}_i + \text{Year}_t \\
 & + \text{Physician}_j + \text{Hospital}_h + \varepsilon_{ijt}.
 \end{aligned} \tag{12}$$

$\text{Diagnostic Uncertainty}_{ijt}$ is a dependent variable that indicates whether a chest pain symptom code is used for patient i treated by physician j in year t . Hispanic_i is an indicator for Hispanic patients, and other variables are the same as defined previously.

The results, as reported in Table 7, include the OLS estimates in column (1) and IV estimates in column (2). The coefficient estimate of *Hispanic* indicator reveals that Hispanic patients are more likely to receive uncertain diagnoses compared with White patients. The coefficient estimate of *InfoShare* is also statistically significant and suggests health information sharing may effectively reduce diagnostic uncertainty.¹⁸ More importantly, the interaction term for $\text{Hispanic} \times \text{InfoShare}$ (coefficient =

Table 7. Estimation of the Effect of Health Information Sharing on Diagnostic Uncertainty

Dependent variable	Diagnostic uncertainty	
	(1) OLS	(2) IV
<i>Hispanic</i>	0.021*** (0.002)	0.024*** (0.003)
<i>InfoShare</i>	-0.009*** (0.003)	-0.024*** (0.009)
<i>Hispanic</i> × <i>InfoShare</i>	-0.009*** (0.003)	-0.013*** (0.004)
<i>Male</i>	0.003* (0.001)	0.003** (0.001)
<i>Age</i>	-0.001*** (0.00005)	-0.001*** (0.00005)
<i>Comorbidities</i>	0.005*** (0.0003)	0.005*** (0.0003)
Observations	336,413	328,698
Physician FE	Yes	Yes
Hospital FE	Yes	Yes
Year FE	Yes	Yes

Notes. The number of observations dropped in column (2) because of missing values in the instrumental variable. Robust standard errors clustered at the physician level are reported in parentheses.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

– 0.013, $p < 0.01$) indicates that greater health information sharing significantly reduces diagnostic uncertainty among Hispanic patients. These findings provide empirical support for our proposed mechanism, demonstrating that health information sharing helps mitigate the difference in signal quality between minority and majority patient groups.

6.6. Health Information Sharing and Physician Skill

Extant research on health disparities provides compelling evidence that variations in physician skill are associated with disparities in diagnoses and drug prescribing (Chan et al. 2022). Specifically, greater skill in minority patient care enables physicians to better interpret noisy risk signals, leading to more accurate diagnosis. For instance, physicians with better skill in communicating and educating minority patients can effectively overcome language barriers and improve their ability to recognize health conditions early (Balsa et al. 2005, McGuire et al. 2008). These enhanced abilities can help physicians assess patient risk from noisy signals. Therefore, physicians with better skills in caring for minority patients could face a narrower gap between $\sigma_{jMinority}^2$ and $\sigma_{jMajority}^2$. On the other hand, the gap in risk signals between majority and minority patients (i.e., $\sigma_{jMinority}^2$ and $\sigma_{jMajority}^2$) may be higher for physicians with lower skill in treating minority patients, as they are less adept at interpreting risk signals.

Given our proposed mechanism that health information sharing is likely to reduce disparate impact by

narrowing the gap in risk signals between majority and minority patients, we expect that the impact of health information sharing on reducing disparate impact to be most salient among physicians with lower skill for minority patient care. In contrast, because highly skilled physicians already face a narrower gap in risk signals between majority and minority patient groups, the shared patient health information may not have as strong an effect on their diagnostic decisions as it does for lower-skilled physicians.

Because our data set does not include information on physician characteristics such as years of experience or board certifications, we follow the prior literature, which uses patient readmission rate as an alternative metric to evaluate physician skill (Fischer et al. 2014, Kripalani et al. 2014). Specifically, we utilize the average readmission rate of minority patients treated by physician j in year $t-1$ to measure physician j 's skill in treating minority patients in year t . For each inpatient and ER encounter related to minority patients, we identify whether the treated patient had revisits (inpatient or ER) with any care provider in the region within 30 days after discharge. Hence, the 30-day readmission rate of physician j in year $t-1$ is the number of minority patients that experienced readmissions within one month divided by their total number of minority patients in year $t-1$.

The model specification to test this prediction is shown in Equation (13). $ReadmissionRate_{jt}$ is a continuous variable with greater values representing lower

physician skill. All other variables are the same as specified in Equation (11).

$$DI_{jht} = \beta_0 + \beta_1 InfoShare_{jht} + \beta_2 ReadmissionRate_{jt} + \beta_3 InfoShare_{jht} \times ReadmissionRate_{jt} + X_{jt} + Year_t + Physician_j + Hospital_h + \varepsilon_{jht}. \tag{13}$$

As noted above, an important endogeneity concern is that hospitals with better performance may be more likely to share patient health information with outside providers; such hospitals also have physicians with higher skill. In this case, the estimates β_1 and β_3 may be biased. Therefore, in Table 8, we report the 2SLS results using the same IV as in Section 6.3, though we also test the robustness of the results using OLS regressions. Similar to our prior approach, we use the indicator of severe disparate impact as an alternate dependent variable. We also test the robustness of our results with and without control variables.

Table 8 indicates that the coefficient estimate of $InfoShare \times ReadmissionRate$ is significantly negative, indicating that the effect of health information sharing on reducing disparate rate is higher among low-skilled physicians (i.e., with higher readmission rates). The bottom panel of Table 8 reports the marginal effects of $InfoShare$ at different levels of physician skill. Specifically, as shown in columns (2) and (4), the use of shared patient health information by low-skilled physicians (measured

Table 8. Heterogeneous Effect of Health Information Sharing Based on Physician Skill

Dependent variable	Absolute disparate impact (1)	Absolute disparate impact (2)	Severe disparate impact (3)	Severe disparate impact (4)
<i>InfoShare</i>	0.004 (0.039)	0.015 (0.039)	0.050 (0.108)	0.074 (0.111)
<i>ReadmissionRate</i>	0.115 (0.124)	0.127 (0.129)	0.248 (0.310)	0.237 (0.312)
<i>InfoShare × ReadmissionRate</i>	-0.294* (0.154)	-0.319** (0.156)	-0.902** (0.430)	-0.927** (0.438)
<i>Cumulative ER encounters (in 1000s)</i>		-0.007*** (0.002)		-0.015** (0.006)
<i>Share of Hispanics</i>		0.031 (0.107)		0.068 (0.271)
<i>Average comorbidities</i>		-0.002 (0.005)		-0.018 (0.014)
Observations	1,965	1,965	1,965	1,965
Physician FE	Yes	Yes	Yes	Yes
Hospital FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
<i>Marginal Effect of InfoShare at High Skill (i.e., 25th percentile of ReadmissionRate)</i>	-0.054*** (0.017)	-0.047*** (0.017)	-0.126** (0.052)	-0.107** (0.053)
<i>Marginal Effect of InfoShare at Low Skill (i.e., 75th percentile of ReadmissionRate)</i>	-0.073*** (0.018)	-0.068*** (0.018)	-0.184*** (0.054)	-0.166*** (0.054)
<i>Difference Between High Skill and Low Skill (p-value)</i>	0.056	0.041	0.036	0.034

Note. Robust standard errors clustered at the physician level are reported in parentheses.

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$.

using the 75th percentile of *Readmission Rate*) is associated with a 0.068 reduction (i.e., 73%) in the level of disparate impact and a reduction in the likelihood of severe disparate impact against Hispanic patients by 16.6 percentage points. However, the marginal effect of shared patient health information on high-skilled physicians (measured at the 25th percentile of *Readmission Rate*) is both statistically and economically smaller.

In sum, we show that the effect of health information sharing on mitigating diagnostic disparate impact is more pronounced for physicians with lower skill in treating minority patients. This evidence is consistent with our argument that if health information sharing narrows the gap in the quality of risk signals between majority and minority patients, its effect should be stronger for physicians who are less adept at obtaining high-quality risk signals from minority patients.

7. Robustness Checks

7.1. Alternative Measure of Disparate Impact

One concern is that if disparate impact in patient diagnosis affects their subsequent diagnoses, it may impact our estimation of heart disease risk. Specifically, because Hispanic patients are less likely to be diagnosed with heart disease, their underlying risk of heart disease may be systematically underestimated. This would result in underestimation of disparate impact. It is worth noting that such underestimation of disparate impact would work against us, making it less likely to identify an effect of health information sharing on reducing disparate impact. Similar discussions of underestimation of disparate impact are available in Arnold et al. (2022, p. 3022) and Baron et al. (2024, p. 21).

Nevertheless, to further address this concern directly, we deployed alternative proxies for heart disease risk to evaluate the consistency of our findings. First, we utilized a symptom-based indicator, that is, whether the patient had any revisits with chest pain symptoms within six months of the focal visit. Because this alternative measure is based on symptoms instead of diagnosis, it may be less influenced by prior diagnosis. The results are reported in column (2) in Table B.9 of Online Appendix B, which are consistent with our baseline results shown in column (1). Second, we also utilized proxies based on more urgent outcomes. Specifically, we examined whether patients who were not diagnosed with heart disease on the focal visit and had an emergency revisit with heart disease diagnosis or chest pain within six months. The results, as presented in columns (3) and (4) of Table B.9, are consistent with the baseline results and support the prevalence of significant diagnostic disparate impact.

7.2. Generalizability

Because our sample data come from a regional HIE in central Texas with relatively more Hispanic patients

than other regions in the United States, one concern is whether the reduction in disparate impact attributed to health information sharing can be generalized to other regions with different proportions of Hispanic population. As discussed earlier, several factors may contribute to differences in the quality of risk signals between majority and minority patients. These factors include development of diagnostic tools based on data that underrepresent minority patients in clinical research, social determinants of health, and fragmentation of their health records. The effect of these factors on reducing the clarity of risk signals of minority patients could persist, regardless of the Hispanic population in a specific geographic region. If health information sharing can reduce the difference in risk signals between majority and minority patients, then we expect health information sharing to reduce disparate impact in other regions with different proportions of Hispanic populations.

To further address this concern, we conduct a supplementary analysis examining the impact of health information sharing in different neighborhoods across central Texas with different proportions of Hispanic populations (because the Hispanic population is unevenly distributed within this region). Using the U.S. Census Bureau 2020 data (Jensen et al. 2021), we calculated the proportion of Hispanic residents in each hospital's zip code, denoted as *Regional Share of Hispanics*, and used the 25th percentile (19.65%) as a threshold to create an indicator for regions with relatively lower Hispanic population representation, denoted as *Low Hispanic Region*.¹⁹ Our model specification is shown in Equation (14).

$$DI_{jht} = \beta_0 + \beta_1 \text{InfoShare}_{jht} + \beta_3 \text{InfoShare}_{jht} \\
 \times \text{Low Hispanic Region}_h + \mathbf{X}_{jt} + \text{Year}_t \\
 + \text{Physician}_j + \text{Hospital}_h + \varepsilon_{jht}. \quad (14)$$

DI_{jht} and InfoShare_{jht} are defined in the same way as in Equation (11). $\text{Low Hispanic Region}_h$ is an indicator variable that is equal to one for hospitals located in regions with a low proportion of Hispanics (i.e., lower than the 25th percentile threshold).

The results are reported in Online Appendix Table B.10, where we first show the OLS results and then the 2SLS regression results based on the same IV used earlier. Overall, they indicate there is no statistically significant difference in the effect of health information sharing between regions with low Hispanic population and other regions. The overall effect of health information sharing remains significantly negative and is consistent with our observation on the effect of health information sharing on mitigating diagnostic disparate impact, regardless of the Hispanic population in a specific region.

7.3. Effect on Missed Diagnosis

In this study, we seek to understand the existence of diagnostic disparate impact, defined as the difference in

likelihood of receiving a certain disease diagnosis among different racial groups with similar underlying risk. A related concept is *missed diagnosis*, which often refers to the accuracy of the diagnosis decision based on the ground truth of disease status. Although both concepts are related to disparities in the diagnostic process, there is an important distinction. Diagnostic disparate impact emphasizes the broader implications of unequal diagnosis that disproportionately affects minority groups in high-risk settings. On the other hand, missed diagnosis is a consequence of the disparity in the accuracy of diagnosis. A large body of literature has highlighted the challenges in defining and measuring missed diagnoses (e.g., Kahraman-Koytak et al. 2019, Kwok et al. 2021). Because of these reasons, we do not focus on the impact of patient health information on differences in missed diagnosis rates between majority and minority patients, as it is outside the scope of this study.

Nevertheless, we conduct preliminary analysis to examine whether there is evidence of a disparity in missed diagnosis for heart disease between White and Hispanic patients in the ER setting, as well as the circumstances under which such a disparity can be mitigated. The details of our analyses are explained in greater detail in Online Appendix C. Overall, the results presented in Table C.1 suggest that Hispanic patients are significantly more likely to experience missed diagnoses compared with White patients. We also observe that greater health information sharing can effectively reduce this difference, as observed by the negative coefficient of $Hispanic \times InfoShare$. Further, this effect is more pronounced among low-skilled physicians. This evidence is consistent with our key argument that health information sharing can narrow the gap in the quality of risk signals between Hispanic and White patients, which, in turn, mitigates the disparity in missed diagnosis between the two racial groups.

8. Discussion

In this research, we study the role of patient health information sharing in addressing health disparities in physician diagnoses, specifically focusing on diagnoses of heart disease patients. Our findings make several important contributions to the literature and offer a range of policy implications. First, in contrast to the prior literature, which primarily focuses on disparities in access to care because of racial or ethnic differences, our study contributes to the literature on health disparities by providing empirically validated evidence of racial disparate impact in physician diagnoses. Based on our analyses of ER patients with cardiovascular conditions, we observe that Hispanic patients are three percentage points less likely to be diagnosed with heart disease than White patients, after accounting for their underlying race-specific risk of heart disease. This disparate impact

represents around 57% of the observed disparity in heart disease diagnosis rate (of 5.8%) in our data, which is both statistically and economically significant. Prevalence of diagnostic disparate impact could have important implications for patients' subsequent treatment decisions and health outcomes of minority populations. For example, a lower likelihood of diagnosing minority patients at risk may leave them untreated, leading to adverse health outcomes and even higher mortality (McCarthy et al. 2021). Our study demonstrates the significance of disparate impact as a driver of health disparities and highlights the importance of IT-enabled solutions to mitigate such disparate impact.

Second, prior research has observed that health information sharing positively influences healthcare quality (Ayer et al. 2019, Janakiraman et al. 2023). Whereas most of these studies focus on patient-level outcomes, there is less understanding of *physician-level* outcomes, particularly involving diagnostic disparate impact. Our findings show that health information sharing holds significant promise in addressing disparities related to the care decisions of minority groups and advancement of health equity. Our research echoes recent calls for more resources aimed at enhancing the utilization of shared health information by medical professionals and clinicians.

Third, our results on the effect of health information sharing on diagnostic uncertainty and its effect on reducing disparate impact among low-skilled physicians are consistent with the proposed mechanism; that is, health information sharing can mitigate diagnostic disparate impact by narrowing the gap in the quality of risk signals between majority and minority patients, thereby reducing statistical discrimination. Consistent with Rubineau and Kang (2012), we do not assert signal quality difference as the only mechanism driving racial disparate impact in physician diagnoses. Rather, we emphasize that health information sharing may effectively reduce signal quality differences by providing physicians with useful information that they may otherwise be unable to obtain because of challenges in treating underserved patient populations.

Our results hold important implications for various stakeholders. For racial minorities, our research suggests that seeking care at hospitals with higher levels of health information sharing may reduce any discriminatory effects in diagnoses and potentially result in better health outcomes. To harness the full benefit of health IT systems, hospitals should go beyond simply adoption and integration of health information systems such as electronic health records (EHRs). Our findings underscore the importance of training physicians and promoting the use of shared information in their clinical routines. For example, hospitals should facilitate access and ease of use of patient health information to physicians. From a physician perspective, it is critical to

recognize the potential for disparate impact in patient diagnoses, which can adversely affect subsequent health outcomes, and how effective use of shared patient health information can mitigate disparate impact.

For policymakers, our study highlights the importance of addressing disparities in diagnosis, particularly disparate impact against certain demographic groups. It highlights an urgent need to accelerate the adoption and utilization of health information sharing, through health information exchanges and adherence to common data standards using FHIR-enabled technologies, to effectively address healthcare disparities in the diagnoses and treatment decisions of minority patients (Vorsek et al. 2022). By doing so, we contribute to building a more equitable and inclusive healthcare system that benefits individuals across diverse racial and ethnic backgrounds.

9. Conclusion

Disparate impact in patient diagnoses poses a significant obstacle on the path to achieving health equity. Our research provides empirical evidence of disparate impact in physician diagnoses of Hispanic patients with cardiovascular disease, based on a longitudinal study of patient-physician encounters across a large metropolitan region. Specifically, after accounting for race-specific heart disease risk, we observe that Hispanic patients are systematically less likely to be diagnosed with heart disease than White patients. We also observe that health information sharing across healthcare providers can effectively reduce the level of disparate impact and likelihood of severe disparate impact against Hispanic patients. We enhance the robustness of our results by using a range of estimation models, such as instrumental variable estimation, falsification test, and alternative measures of disparate impact and disease risk.

We show that Hispanic patients are more likely to receive uncertain diagnoses compared with White patients, and greater health information sharing can reduce the likelihood of uncertain diagnoses, especially for Hispanic patients. Further, our analysis suggests that health information sharing is particularly beneficial to physicians with lower skill, who may face a bigger gap in the quality of risk signals between majority and minority patients. These findings are consistent with our proposed mechanism that information sharing can mitigate diagnostic disparate impact by closing the gap in quality of risk signals between majority and minority patients.

Our research represents one of the first studies to emphasize the potential impact of enhanced patient health information sharing on mitigating disparate impact in physician diagnoses based on patient race. Hence, our work has actionable policy implications,

underscoring the significance of promoting standardized health information sharing across healthcare providers as a pivotal strategy for alleviating health disparities and promoting health equity.

Nevertheless, given our data availability and research design, our study still has several limitations. First, we focus on the ER setting because it represents a high-risk environment where the likelihood of being diagnosed with a certain disease for minority patients could be systematically lower compared with majority patients and because it also allows for quasirandom assignment of physicians to patients, which is critical for identification of disparate impact. Future research could examine whether and how diagnostic disparate impact based on patient race arises in low-risk settings such as preventive care and routine outpatient visits. Whereas we expect health information sharing to similarly improve the quality of disease signals, particularly for minority patients, future research can provide more systematic evidence in low-risk settings.

Second, it is important to evaluate the existence of disparate impact among other minority races and other types of diseases. Future research can extend our study to evaluate diagnostic impact across other geographic regions with diverse patient populations. Third, our data do not provide physician demographic information. Future research may include race and ethnicity of physicians to study how disparate impact varies among physicians of different race/ethnic groups. Understanding such variation is crucial for developing policy solutions regarding the importance of physician-patient racial concordance and its impact on patient diagnoses and health outcomes.

Fourth, an important assumption of our study is that hospital-level reporting of physician utilization of shared patient health information reflects its actual use by the focal physician at a given hospital. We believe that this is a valid assumption given our focus on the ER setting, where physicians are required to follow standard protocols mandated by the hospital. Our falsification analyses based on a subsample of patients without recent health information also support our main results. Nevertheless, an interesting avenue for future research would be to obtain granular data on physician use of patient health information, based on system access logs, to precisely measure the intensity and frequency of information use by each physician for patient diagnoses.

Another concern is related to the identification challenges inherent in this study. It is neither practical nor ethical to randomly provide physicians with different levels of patient health information shared by external providers. Hence, we implement a panel data approach to control for a range of time-invariant, unobservable variables at the physician-hospital level, as well as general time trends. Our empirical results, along with falsification and robustness tests, suggest that the estimated

effects reflect the impact of health information sharing on reducing diagnostic disparate impact. Still, future research could identify exogenous shocks, such as regulatory or policy changes, that may influence physician utilization of shared patient health information to identify its impact on patients' diagnostic outcomes.

Finally, our study does not investigate whether disparate impact in physician diagnoses may have downstream implications on patient health outcomes and utilization of healthcare resources. Examining potential solutions that not only mitigate diagnostic disparities but also improve patient outcomes could offer promising avenues for future research. Despite these limitations, we believe that our research represents an important step toward evaluating diagnostic disparate impact among healthcare providers and proposing potential solutions to address disparities in healthcare. Our research can serve as a catalyst to explore new research questions related to the role of information systems and artificial intelligence in improving health equity.

Acknowledgments

The authors gratefully acknowledge the guidance and constructive feedback from the editors and review team. I. Bardhan is grateful for the financial support of the Charles and Elizabeth Prothro Regents Chair in Healthcare Management at the McCombs School of Business. Earlier versions of this research were presented at the Annual Conference on Health IT and Analytics in 2023 and 2024, INFORMS Annual Meetings in 2023, the International Conference on Information Systems in 2024, the Symposium on Statistical Challenges in E-commerce Research in 2024, and the Workshop on Information Systems and Economics in 2023. The authors thank seminar participants at UT Austin, UT Dallas, University of Houston, University of Florida, University of Miami, University of Arizona, University of California San Diego, University of Minnesota, University of Wisconsin–Madison, and New York University for their helpful feedback.

Endnotes

¹ Minority patients are defined as those belonging to a racial minority group, which include Hispanics, African Americans, Asian Americans, American Indians, and Native Hawaiians. Agency for Healthcare Research and Quality, retrieved from <https://www.ahrq.gov/topics/racial-ethnic-minorities.html>.

² In high-stakes ER settings, if Hispanic patients with the same underlying heart disease risk are less likely to be diagnosed than White patients, it could result in more severe outcomes because of potential delay in treatment. For simplicity, we use the phrase “disparate impact against Hispanic patients” to refer to the scenario where they have a lower diagnosis rate of heart disease compared to White patients, after controlling for underlying disease risk.

³ Our definition of a minority group follows conventional norm in the healthcare literature, which defines minority status not solely by numerical representation in society but by social and economic disempowerment (Blume 2013).

⁴ That is, if certain characteristics related to the group label can mediate disparate impact, conditioning on such characteristics may

introduce included variable bias. For example, certain races may be correlated with low socioeconomic status which, in turn, would lead to observed diagnostic disparities among different racial groups.

⁵ For details, see the announcement (NOT-OD-19-122) by the Office of the Director, National Institutes of Health, retrieved from <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-19-122.html>.

⁶ Although sharing of patient health information can be facilitated by an HIE, it can also be achieved through other IT-enabled channels. Examples of other types of health information sharing between providers include (a) direct interfaces between electronic health record (EHR) systems, (b) query access to other organizations' EHR systems using login credentials, (c) EHR vendor-vendor integration that enables patient record identification within the network, and (d) access to national or state-level health information networks.

⁷ See Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System, Mortality 2018–2021 on CDC WONDER Online Database. Data are from the Multiple Cause of Death Files, 2018–2021, as compiled from data provided by the 57 vital statistics jurisdictions through the Vital Statistics Cooperative Program. Accessed at <http://wonder.cdc.gov/ucd-icd10-expanded.html> on May 15, 2023.

⁸ This approach to estimate disparate impact can be extended to settings with continuous H_i^* (Arnold et al. 2022).

⁹ Because heart disease may not always manifest within a short period after occurrence of symptoms, we use six months as our baseline time window (after the focal visit) to measure the potential risk of heart disease (Zaya et al. 2012, Grootven et al. 2021, Li et al. 2022). We also run robustness checks using shorter (i.e., three-month) and longer (i.e., one-year) time windows to estimate patient i 's heart disease risk. The results remain consistent and are available upon request.

¹⁰ For details of the extrapolation method, see Arnold et al. (2022, pp. 3012–3013).

¹¹ We thank the anonymous reviewer for suggesting this test of the quasirandom assignment assumption.

¹² Although the regional HIE has comprehensive encounter-level records of patient visits, it does not include timestamps of HIE access by physicians. Hence, we are unable to measure *InfoShare* directly using the HIE data.

¹³ As a robustness check, we also use the median value of encounter count at the physician-hospital-year level, which is equal to 17 patient encounters. The results are qualitatively similar and available upon request.

¹⁴ In the first part of our study, where we estimate diagnostic disparate impact, the unit of analysis is at the patient encounter level, which includes 336,413 encounters for 91,580 patients. In the latter part on the role of health information sharing, because our unit of analysis shifts to the physician-year level, this panel data structure has resulted in fewer observations, although it is still based on the same underlying sample.

¹⁵ A detailed discussion of the physician skill measure is available in Section 6.6.

¹⁶ Detailed guidelines available at <https://www.cms.gov/files/document/fy-2022-icd-10-cm-coding-guidelines-updated-02012022.pdf>.

¹⁷ We focus on chest pain because it is the most common symptom for heart disease and is highly relevant to the diagnostic process for heart disease (DeVon et al. 2020, Jurgens et al. 2022).

¹⁸ An alternate explanation is that access to prior diagnosis codes through health information sharing may change physician coding behavior, making them less likely to use symptom codes. However, the ICD-10-CM guidelines require physicians to make a

comprehensive assessment based on patient symptoms, medical history, examination, test results, and clinical judgment. Even without prior diagnosis codes, physicians can assign a definitive diagnosis code if sufficient evidence exists. Thus, health information sharing influences diagnosis by providing rich clinical information, rather than by directly changing physicians' tendency to use symptom codes.

¹⁹ This threshold is close to the U.S. national average, that is, 18.7%, in 2020 (U.S. Census Bureau 2020). See the report from Census 2020 (Jensen et al. 2021).

References

- Adjerid I, Adler-Milstein J, Angst C (2018) Reducing Medicare spending through electronic health information exchange: The role of incentives and exchange maturity. *Inform. Systems Res.* 29(2):341–361.
- Aigner DJ, Cain GG (1977) Statistical theories of discrimination in labor markets. *ILR Rev.* 30(2):175–187.
- Alam R, Cheraghi-Sohi S, Panagioti M, Esmail A, Campbell S, Panagopoulou E (2017) Managing diagnostic uncertainty in primary care: A systematic critical review. *BMC Family Practice* 18(1):79.
- Angst CM, Wowak KD, Handley SM, Kelley K (2017) Antecedents of information systems sourcing strategies in U.S. hospitals: A longitudinal study. *MIS Quart.* 41(4):1129–1152.
- Arcaya MC, Arcaya AL, Subramanian SV (2015) Inequalities in health: Definitions, concepts, and theories. *Global Health Action* 8(1):27106.
- Arnold D, Dobbie W, Hull P (2022) Measuring racial discrimination in bail decisions. *Amer. Econom. Rev.* 112(9):2992–3038.
- Arrow K (1971) The theory of discrimination. Working Paper No. 403, Department of Economics, Industrial Relations Section, Princeton University, Princeton, NJ.
- Ashraf M, Zlochiver V, Sajed SM, Sajed S, Bajwa T, Allaqaband SQ, Jan MF (2023) Racial disparities in diagnostic evaluation and revascularization in patients with acute myocardial infarction—A 15-year longitudinal study. *Current Problems Cardiology* 48(8):101733.
- Atasoy H, Chen P, Ganju K (2018) The spillover effects of health IT investments on regional healthcare costs. *Management Sci.* 64(6):2515–2534.
- Atasoy H, Demirezen EM, Chen P (2021) Impacts of patient characteristics and care fragmentation on the value of HIEs. *Production Oper. Management* 30(2):563–583.
- Atolagbe OO, Romano PS, Southern DA, Wongtanasarasin W, Ghali WA (2024) Coding rules for uncertain and “ruled out” diagnoses in ICD-10 and ICD-11. *BMC Medical Inform. Decision Making* 21(6):386.
- Ayabakan S, Bardhan I, Zheng Z (E), Kirksey K (2017) The impact of health information sharing on duplicate testing. *MIS Quart.* 41(4):1083–1104.
- Ayer T, Ayvaci MUS, Karaca Z, Vlachy J (2019) The impact of health information exchanges on emergency department length of stay. *Production Oper. Management* 28(3):740–758.
- Ayres I (2010) Testing for discrimination and the problem of “included variable bias.” Working paper, Yale Law School, New Haven, CT.
- Balsa AI, McGuire TG (2001) Statistical discrimination in health care. *J. Health Econom.* 20(6):881–907.
- Balsa AI, McGuire TG, Meredith LS (2005) Testing for statistical discrimination in health care. *Health Services Res.* 40(1):227–252.
- Balsa AI, Seiler N, McGuire TG, Bloche MG (2003) Clinical uncertainty and healthcare disparities. *Amer. J. Law Medicine* 29(2–3): 203–219.
- Bardhan IR, Bao C, Ayabakan S (2023) Value implications of sourcing electronic health records: The role of physician practice integration. *Inform. Systems Res.* 34(3):1169–1190.
- Baron EJ, Doyle JJ, Emanuel N, Hull P, Ryan J (2024) Discrimination in multi-phase systems: evidence from child protection. *Quart. J. Econom.* 139(3):1611–1664.
- Bibbins-Domingo K, Helman A, eds. (2022) Why diverse representation in clinical research matters and the current state of representation within the clinical research ecosystem. *Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups* (National Academies Press, Washington, DC), 23–33.
- Blume AW (2013) Minority groups and addictions. Miller PM, ed. *Principles of Addiction, Comprehensive Addictive Behaviors and Disorders*, vol. 1 (Academic Press, San Diego), 149–158.
- Bohren JA, Hull P, Imas A (2022) Systemic discrimination: Theory and measurement. NBER Working Paper No. 29820, National Bureau of Economic Research, Cambridge, MA.
- Brown R-M, Tamazi S, Weinberg CR, Dwivedi A, Mieres JH (2022) Racial disparities in cardiovascular risk and cardiovascular care in women. *Current Cardiology Rep.* 24(9):1197–1208.
- Chan DC, Gentzkow M, Yu C (2022) Selection with variation in diagnostic skill: Evidence from radiologists. *Quart. J. Econom.* 137(2):729–783.
- Chando S, Tiro JA, Harris TR, Kobrin S, Breen N (2013) Effects of socioeconomic status and health care access on low levels of human papillomavirus vaccination among Spanish-speaking Hispanics in California. *Amer. J. Public Health* 103(2):270–272.
- Cole MB, Nguyen KH (2020) Unmet social needs among low-income adults in the United States: Associations with health care access and quality. *Health Services Res.* 55(S2):873–882.
- Cooper LA, Roter DL (2003) Patient-provider communication: The effect of race and ethnicity on process and outcomes of health-care. Smedley BD, Stith AY, Nelson AR, eds. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care* (National Academies Press, Washington, DC), 552–593.
- Cooper LA, Hill MN, Powe NR (2002) Designing and evaluating interventions to eliminate racial and ethnic disparities in health care. *J. General Internal Medicine* 17(6):477–486.
- DeVon HA, Mirzaei S, Zègre-Hemsey J (2020) Typical and atypical symptoms of acute coronary syndrome: Time to retire the terms? *J. Amer. Heart Assoc. Cardiovascular Cerebrovascular Disease* 9(7):e015539.
- Eichelberger C, Patel A, Ding Z, Pericone CD, Lin JH, Baugh CW (2020) Emergency department visits and subsequent hospital admission trends for patients with chest pain and a history of coronary artery disease. *Cardiology Therapy* 9(1):153–165.
- Escobedo LE, Cervantes L, Havranek E (2023) Barriers in healthcare for LatinX patients with limited English proficiency—A narrative review. *J. General Internal Medicine* 38(5):1264–1271.
- Fischer C, Lingsma HF, Marang-van de Mheen PJ, Kringos DS, Klazinga NS, Steyerberg EW (2014) Is the readmission rate a valid quality indicator? A review of the evidence. *PLoS One* 9(11):e112282.
- Fu R, Aseri M, Singh PV, Srinivasan K (2022) “Un”fair machine learning algorithms. *Management Sci.* 68(6):4173–4195.
- Ganju KK, Atasoy H, McCullough J, Greenwood B (2020) The role of decision support systems in attenuating racial biases in healthcare delivery. *Management Sci.* 66(11):5171–5181.
- Grootven BV, Jepma P, Rijkema C, Verweij L, Leeflang M, Daams J, Deschodt M, Milisen K, Flamaing J, Buurman B (2021) Prediction models for hospital readmissions in patients with heart disease: A systematic review and meta-analysis. *BMJ Open* 11(8):e047576.
- Hill-Briggs F, Adler NE, Berkowitz SA, Chin MH, Gary-Webb TL, Navas-Acien A, Thornton PL, Haire-Joshu D (2021) Social determinants of health and diabetes: A scientific review. *Diabetes Care* 44(1):258–279.
- Hull P (2021) What marginal outcome tests can tell us about racially biased decision-making. Preprint, submitted March 2, <https://doi.org/10.2139/ssrn.3795454>.

- Hwang EH, Guo X, Tan Y, Dang Y (2022) Delivering healthcare through teleconsultations: Implications for offline healthcare disparity. *Inform. Systems Res.* 33(2):515–539.
- Janakiraman R, Park E, M. Demirezen E, Kumar S (2023) The effects of health information exchange access on healthcare quality and efficiency: An empirical investigation. *Management Sci.* 69(2):791–811.
- Javed Z, Haisum Maqsood M, Yahya T, Amin Z, Acquah I, Valero-Elizondo J, Andrieni J, et al. (2022) Race, racism, and cardiovascular health: Applying a social determinants of health framework to racial/ethnic disparities in cardiovascular disease. *Circulation Cardiovascular Quality Outcomes* 15(1):e007917.
- Jensen E, Jones N, Rabe M, Pratt B, Medina L, Orozco K, Spell L (2021) The chance that two people chosen at random are of different race or ethnicity groups has increased since 2010. *U.S. Census Bureau* (August 12), <https://www.census.gov/library/stories/2021/08/2020-united-states-population-more-racially-ethnically-diverse-than-2010.html>.
- Jha AK, Varosy PD, Kanaya AM, Hunninghake DB, Hlatky MA, Waters DD, Furberg CD, Shlipak MG (2003) Differences in medical care and disease outcomes among Black and White women with heart disease. *Circulation* 108(9):1089–1094.
- Jung J, Corbett-Davies S, Gaebler JD, Shroff R, Goel S (2024) Mitigating included- and omitted-variable bias in estimates of disparate impact. Preprint, submitted September 15, <https://arxiv.org/abs/1809.05651>.
- Jurgens CY, Lee CS, Aycok DM, Masterson Creber R, Denfeld QE, DeVon HA, Evers LR, et al. (2022) State of the science: The relevance of symptoms in cardiovascular disease and research: A scientific statement from the American Heart Association. *Circulation* 146(12):e173–e184.
- Kahraman-Koytak P, Bruce BB, Peragallo JH, Newman NJ, Biousse V (2019) Diagnostic errors in initial misdiagnosis of optic nerve sheath meningiomas. *JAMA Neurology* 76(3):326–332.
- Kallus N, Mao X, Zhou A (2022) Assessing algorithmic fairness with unobserved protected class using data combination. *Management Sci.* 68(3):1959–1981.
- Kaltenborn Z, Paul K, Kirsch JD, Aylward M, Rogers EA, Rhodes MT, Usher MG (2021) Super fragmented: A nationally representative cross-sectional study exploring the fragmentation of inpatient care among super-utilizers. *BMC Health Services Res.* 21(1):338.
- Kim EJ, Kim T, Conigliaro J, Liebschutz JM, Paasche-Orlow MK, Hanchate AD (2018) Racial and ethnic disparities in diagnosis of chronic medical conditions in the USA. *J. General Internal Medicine* 33(7):1116–1123.
- Kripalani S, Theobald CN, Anctil B, Vasilevskis EE (2014) Reducing hospital readmission rates: Current strategies and future directions. *Annual Rev. Medicine* 65(1):471–485.
- Kwok CS, Bennett S, Azam Z, Welsh V, Potluri R, Loke YK, Mallen CD (2021) Misdiagnosis of acute myocardial infarction: A systematic review of the literature. *Critical Pathways Cardiology* 20(3):155–162.
- Lammers EJ, Adler-Milstein J, Kocher KE (2014) Does health information exchange reduce redundant imaging? Evidence From emergency departments. *Medical Care* 52(3):227–234.
- Lee BY, McGlone SM, Song Y, Avery TR, Eubank S, Chang C-C, Bailey RR, et al. (2011) Social network analysis of patient sharing among hospitals in Orange County, California. *Amer. J. Public Health* 101(4):707–713.
- Li D, Lin Y, Dong W, Hu Y, Li K (2022) A nomogram for predicting the readmission within 6 months after treatment in patients with acute coronary syndrome. *BMC Cardiovascular Disorders* 22(1):448.
- López L, Green AR, Tan-McGrory A, King RS, Betancourt JR (2011) Bridging the digital divide in health care: The role of health information technology in addressing racial and ethnic disparities. *Joint Commission J. Quality Patient Safety* 37(10):437–445.
- Lossos I, Israeli A, Zajicek G, Berry EM (1989) Diagnosis deferred—The clinical spectrum of diagnostic uncertainty. *J. Clinical Epidemiology* 42(7):649–657.
- Lu SF, Rui H (2018) Can we trust online physician ratings? Evidence from cardiac surgeons in Florida. *Management Sci.* 64(6):2557–2573.
- Lutfey KE, Ketcham JD (2005) Patient and provider assessments of adherence and the sources of disparities: Evidence from diabetes care. *Health Services Res.* 40(6p1):1803–1817.
- Markides KS, Coreil J (1986) The health of Hispanics in the southwestern United States: An epidemiologic paradox. *Public Health Rep.* 101(3):253–265.
- McCarthy CP, Olshan DS, Rehman S, Jones-O'Connor M, Murphy S, Cohen JA, Singh A, Vaduganathan M, Januzzi JL, Wasfy JH (2021) Cardiologist evaluation of patients with type 2 myocardial infarction. *Circulation Cardiovascular Quality Outcomes* 14(1):e007440.
- McGuire TG, Ayanian JZ, Ford DE, Henke REM, Rost KM, Zaslavsky AM (2008) Testing for statistical discrimination by race/ethnicity in panel data for depression treatment in primary care. *Health Services Res.* 43(2):531–551.
- Medina-Inojosa J, Jean N, Cortes-Bergoderi M, Lopez-Jimenez F (2014) The Hispanic paradox in cardiovascular disease and total mortality. *Progress Cardiovascular Diseases* 57(3):286–292.
- Miller AR, Tucker C (2009) Privacy protection and technology diffusion: The case of electronic medical records. *Management Sci.* 55(7):1077–1093.
- Miller AR, Tucker C (2014) Health information exchange, system size and information silos. *J. Health Econom.* 33(1):28–42.
- Morales LS, Lara M, Kington RS, Valdez RO, Escarce JJ (2002) Socio-economic, cultural, and behavioral factors affecting Hispanic health outcomes. *J. Health Care Poor Underserved* 13(4):477–503.
- Moriya AS, Chakravarty S (2023) Racial and ethnic disparities in preventable hospitalizations and ED visits five years after ACA Medicaid expansions. *Health Affairs* 42(1):26–34.
- Pager D, Shepherd H (2008) The sociology of discrimination: Racial discrimination in employment, housing, credit, and consumer markets. *Annual Rev. Sociol.* 34(1):181–209.
- Patiño-Benavidez AF, Buitrago G, Roza-Agudelo N, Saldaña-Espinel LE, Gamboa-Garay ÓA, Eslava-Schmalbach J, Bonilla-González C, et al. (2024) Association of Healthcare Fragmentation and the survival of patients with colorectal cancer in Colombia. *Value Health Reg. Issues* 41:63–71.
- Phelps ES (1972) The statistical theory of racism and sexism. *Amer. Econom. Rev.* 62(4):659–661.
- Pinheiro LC, Reshetnyak E, Safford MM, Kern LM (2021) Racial disparities in preventable adverse events attributed to poor care coordination reported in a national study of older U.S. adults. *Medical Care* 59(10):901–906.
- Rubineau B, Kang Y (2012) Bias in White: A longitudinal natural experiment measuring changes in discrimination. *Management Sci.* 58(4):660–677.
- Saha S, Arbelaez JJ, Cooper LA (2011) Patient–physician relationships and racial disparities in the quality of health care. *Amer. J. Public Health* 93(10):1713–1719.
- Shi Y, Hunter Guevara LR, Dykhoff HJ, Sangaralingham LR, Phelan S, Zaccariello MJ, Warner DO (2021) Racial disparities in diagnosis of attention-deficit/hyperactivity disorder in a US national birth cohort. *JAMA Network Open* 4(3):e210321.
- Staiger D, Stock JH (1997) Instrumental variables regression with weak instruments. *Econometrica* 65(3):557.
- Swenson CJ, Trepka MJ, Rewers MJ, Scarbro S, Hiatt WR, Hamman RF (2002) Cardiovascular disease mortality in Hispanics and non-Hispanic Whites. *Amer. J. Epidemiology* 156(10):919–928.
- U.S. Supreme Court (1971) *Griggs v. Duke Power Co.*, 401 U.S. 424 (1971). Retrieved November 11, <https://supreme.justia.com/cases/federal/us/401/424/>.
- Vest JR, Kern LM, Silver MD, Kaushal R & for the HITEC investigators (2015) The potential for community-based health

- information exchange systems to reduce hospital readmissions. *J. Amer. Medical Inform. Assoc.* 22(2):435–442.
- Vorisek CN, Lehne M, Klopfenstein SAI, Mayer PJ, Bartschke A, Haese T, Thun S (2022) Fast healthcare interoperability resources (FHIR) for interoperability in health research: Systematic review. *JMIR Med. Inform.* 10(7):e35724.
- White-Williams C, Rossi LP, Bittner VA, Driscoll A, Durant RW, Granger BB, et al. (2020) Addressing social determinants of health in the care of patients with heart failure: A Scientific statement from the American Heart Association. *Circulation* 141(22):e841–e863.
- Willey JZ, Rodriguez CJ, Moon YP, Paik MC, Di Tullio MR, Homma S, Sacco RL, Elkind MSV (2012) Coronary death and myocardial infarction among Hispanics in the Northern Manhattan Study: Exploring the Hispanic paradox. *Ann. Epidemiology* 22(5):303–309.
- Yaraghi N (2015) An empirical analysis of the financial benefits of health information exchange in emergency departments. *J. Amer. Medical Inform. Assoc.* 22(6):1169–1172.
- Yen W (2013) How long and how far do adults travel and will adults travel for primary care? Research Brief No. 70, Washington State Office of Financial Management, Olympia, WA, <https://www.ofm.wa.gov/sites/default/files/public/legacy/researchbriefs/2013/brief070.pdf>.
- Zaya M, Phan A, Schwarz ER (2012) Predictors of re-hospitalization in patients with chronic heart failure. *World J. Cardiology* 4(2):23–30.
- Zhang N, Xu H (2023) Fairness of ratemaking for catastrophe insurance: Lessons from machine learning. *Inform. Systems Res.* 35(2): 469–488.