

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

**Minghong Yuan**

University of Texas at Austin  
2110 Speedway  
Austin, Texas  
United States  
78712

[minghong.yuan@mcombs.utexas.edu](mailto:minghong.yuan@mcombs.utexas.edu)

**Indranil Bardhan**

University of Texas at Austin  
2110 Speedway  
Austin, Texas  
United States  
78712

[Indranil.Bardhan@mcombs.utexas.edu](mailto:Indranil.Bardhan@mcombs.utexas.edu)

**Wen Wen**

University of Texas at Austin  
2110 Speedway  
Austin, Texas  
United States  
78712

[Wen.Wen@mcombs.utexas.edu](mailto:Wen.Wen@mcombs.utexas.edu)

**Online Appendix**

## Appendix A: Model Development

In this Appendix, we provide detailed explanation of how the quality of disease signals (i.e., signal noise  $\sigma_{jr}$ ) may impact expected disease risk  $E(Z_i^*|S_{ij})$ , and thereby, disparate impact. First, to study the implications of signal noise  $\sigma_{jr}$  for expected disease risk  $E(Z_i^*|S_{ij})$ , we take the first-order derivative of  $E(Z_i^*|S_{ij})$  with respect to  $\sigma_{jr}$ :

$$\frac{\partial E(Z_i^*|S_{ij})}{\partial \sigma_{jr}} = \frac{2\sigma_{jr}\sigma_Z^2(\mu - S_{ij})}{(\sigma_Z^2 + \sigma_{jr}^2)^2} \quad (A1)$$

When  $S_{ij} > \mu$ , the derivative  $\frac{\partial E(Z_i^*|S_{ij})}{\partial \sigma_{jr}}$  becomes negative. This suggests that higher  $\sigma_{jr}$  leads to lower  $E(Z_i^*|S_{ij})$ . Since we argue that the noise of disease signals is larger for minority patients than that of majority patients (i.e.,  $\sigma_{jMinority}^2 > \sigma_{jMajority}^2$ ), the estimated severity  $E(Z_i^*|S_{ij})$  for minority patients would be consistently lower than majority patients when  $S_{ij} > \mu$ .

Furthermore, we show when  $S_{ij} = \mu$ , this is where the two lines in Figure A.1 intersect. In particular,  $E(Z_i^*|S_{ij}) = (1 - \beta_{jr})\mu + \beta_{jr}S_{ij}$ , where  $\beta_{jr} = \frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_{jr}^2}$ . Hence, the horizontal value  $S_{ij}^*$  at the intersection of the two lines for minority and majority patients, respectively, is shown as:

$$\begin{aligned} (1 - \beta_{jMinority})\mu + \beta_{jMinority}S_{ij}^* &= (1 - \beta_{jMajority})\mu + \beta_{jMajority}S_{ij}^* \\ \Leftrightarrow \left(\frac{\sigma_{jMinority}^2}{\sigma_Z^2 + \sigma_{jMinority}^2}\right)\mu + \left(\frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_{jMinority}^2}\right)S_{ij}^* &= \left(\frac{\sigma_{jMajority}^2}{\sigma_Z^2 + \sigma_{jMajority}^2}\right)\mu + \left(\frac{\sigma_Z^2}{\sigma_Z^2 + \sigma_{jMajority}^2}\right)S_{ij}^* \end{aligned} \quad (A2)$$

After algebraic transformation, we obtain:

$$(\sigma_{jMinority}^2 - \sigma_{jMajority}^2)\sigma_Z^2\mu = (\sigma_{jMinority}^2 - \sigma_{jMajority}^2)\sigma_Z^2S_{ij}^* \quad (A3)$$

Since both  $\sigma_{jMinority}^2 - \sigma_{jMajority}^2$  and  $\sigma_Z^2$  are non-zero, it follows that  $S_{ij}^* = \mu$ .

Next, we demonstrate how the quality of disease signals (i.e., signal noise  $\sigma_{jr}$ ) may impact disparate impact. As discussed in Section 2, suppose physician  $j$  uses Bayes' rule to update their priors about a patient's disease risk after observing patient  $i$ 's noisy signals  $S_{ij} = Z_i^* + \varepsilon_{ij}$ . Assume that the underlying disease risk  $Z_i^*$  follows a normal distribution  $N(\mu, \sigma_Z^2)$ .  $\varepsilon_{ij}$  also follows a normal distribution  $N(0, \sigma_{jr}^2)$ . The likelihood of being diagnosed with the disease conditional on the underlying disease risk  $Z_i^*$  and race  $R_i$  is expressed as:

$$Pr(D_{ij} = 1|Z_i^* = z, R_i = r) = P_j(S_{ij}, R_i)|Z_i^* = z, R_i = r \quad (A4)$$

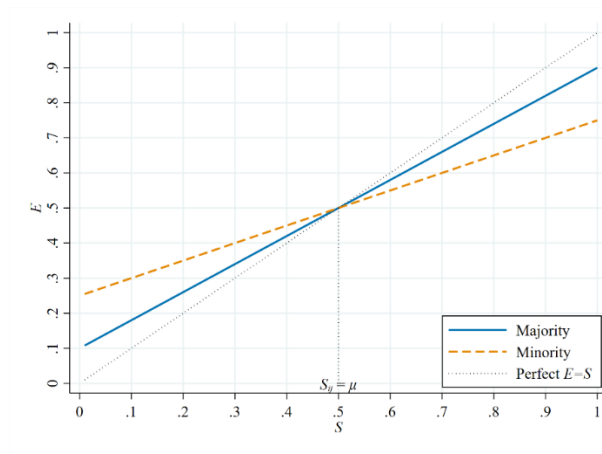
$$\frac{\partial Pr(D_{ij} = 1|Z_i^* = z, R_i = r)}{\partial \sigma_{jr}} = \frac{\partial f}{\partial E(Z_i^*|S_{ij})} \frac{\partial E(Z_i^*|S_{ij})}{\partial \sigma_{jr}} |Z_i^* = z, R_i = r \quad (A5)$$

Because  $f(\cdot)$  is a strictly increasing function of  $E(Z_i^*|S_{ij})$ , then  $\frac{\partial f}{\partial E(Z_i^*|S_{ij})} > 0$ . We have shown

that when  $S_{ij} > \mu$ , the derivative  $\frac{\partial E(Z_i^*|S_{ij})}{\partial \sigma_{jr}} < 0$ , and thus  $Pr(D_{ij} = 1|Z_i^* = z, R_i = r)$  is an decreasing function of  $\sigma_{jr}$  when  $S_{ij} > \mu$ .

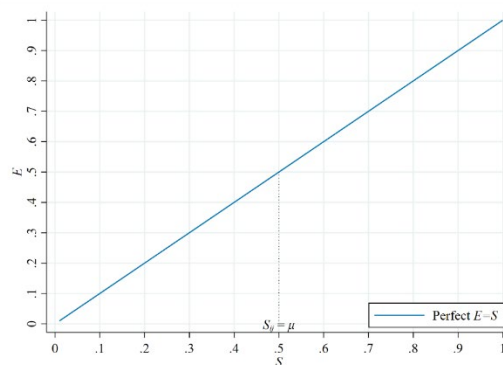
Given that  $\sigma_{jMinority}^2 > \sigma_{jMajority}^2$ , the conditional probability of being diagnosed with a certain disease would generally be lower for minority patients than majority patients (i.e.,  $Pr(D_{ij} = 1|Z_i^* = z, R_i = Minority) - Pr(D_{ij} = 1|Z_i^* = z, R_i = Majority) < 0$ ). In summary, we show that a difference in signal (noise)  $\sigma_{jr}$  between majority and minority patients can affect the difference in expected disease risk  $E(Z_i^*|S_{ij})$  between the two races and lead to racial disparate impact in diagnoses.

**Figure A.1: Disparate impact caused by a difference in the quality of disease signals**



*Notes:* The x-axis indicates the observed disease signals  $S_{ij}$  and the y-axis indicates the expected disease risk  $E(Z_i^*|S_{ij})$ . When  $S_{ij} > \mu$ , the estimated disease risk  $E(Z_i^*|S_{ij})$  for minority patients will be consistently lower than that for majority patients.

**Figure A.2: Scenario with no disparate impact**



*Notes:* The x-axis indicates the observed disease signals  $S_{ij}$  and the y-axis indicates the expected disease risk  $E(Z_i^*|S_{ij})$ . When  $\sigma_{jMinority}^2 = \sigma_{jMajority}^2 = 0$  and  $\beta_{jr} = 1$  across majority and minority patients, the expected disease risk  $E(Z_i^*|S_{ij})$  would become the same for the two groups of patients with the same signals  $S_{ij}$  and would be equal to  $S_{ij}$ .

## Appendix B: Supporting Empirical Analyses

**Table B.1: Quasi-random Physician Assignment: Relationship between Physician Skill and Patient Characteristics**

	(1)	(2)	(3)	(4)
Dependent variable: Physician skill				
Hispanic	0.001 (0.012)			
Comorbidities		-0.003 (0.003)		
Male			0.007 (0.010)	
Age				-0.007 (0.021)
Constant	22.225*** (0.006)	22.244*** (0.022)	22.222*** (0.004)	22.252*** (0.086)
Observations	285,448	285,448	285,448	285,448
R-squared	0.781	0.781	0.781	0.781
Physician FE	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES
Time FE	YES	YES	YES	YES

*Notes:* Robust standard errors clustered at physician level are reported in parentheses. The dependent variable, physician skill, is measured based on the average 30-day readmission rate of minority patients treated by a given physician in the prior year. We take the log transformation of Age to reduce its skewness. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.2: Quasi-random Physician Assignment: Relationship between Physician Experience and Patient Characteristics**

	(1)	(2)	(3)	(4)
Dependent variable: Physician Experience				
Hispanic	0.465 (0.545)			
Comorbidities		-0.295 (0.257)		
Male			-0.965 (0.739)	
Age				-1.591 (0.971)
Constant	319.975*** (0.289)	322.367*** (1.869)	320.660*** (0.335)	326.682*** (3.943)
Observations	285,448	285,448	285,448	285,448
R-squared	0.599	0.599	0.599	0.599
Physician FE	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES
Time FE	YES	YES	YES	YES

*Notes:* Robust standard errors clustered at physician level are reported in parentheses. The dependent variable, physician experience, is measured by the cumulative count of ER encounters (in thousands) with cardiovascular conditions diagnosed by a given physician in the prior year. We take the log transformation of Age to reduce its skewness. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.3: Quasi-random Physician Assignment: Physician Experience in Treating Hispanic Patients and Patient Characteristics**

	(1)	(2)	(3)	(4)
<u>Dependent variable: Physician experience treating Hispanic patients</u>				
Hispanic	-0.018 (0.027)			
Comorbidities		-0.004 (0.005)		
Male			-0.006 (0.020)	
Age				-0.027 (0.045)
Constant	41.607*** (0.014)	41.623*** (0.040)	41.600*** (0.009)	41.706*** (0.184)
Observations	285,448	285,448	285,448	285,448
R-squared	0.864	0.864	0.864	0.864
Physician FE	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES
Time FE	YES	YES	YES	YES

*Notes:* Robust standard errors clustered at physician level are reported in parentheses. The dependent variable, physician experience treating Hispanic patients, is measured by the proportion of Hispanic patients in ER encounters with cardiovascular conditions diagnosed by a given physician in the prior year. We take the log transformation of Age to reduce its skewness. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.4: Effect of Health Information Sharing on Disparate Impact Using Quadratic Extrapolation**

Dependent Variable	Absolute Disparate Impact (1)	Absolute Disparate Impact (2)	Severe Disparate Impact (3)	Severe Disparate Impact (4)
InfoShare	-0.016** (0.007)	-0.014** (0.007)	-0.064*** (0.024)	-0.059** (0.024)
Cumulative ER encounters (in 1000s)		-0.007*** (0.002)		-0.014** (0.006)
Share of Hispanic Patients		0.030 (0.073)		0.020 (0.205)
Average Comorbidities		-0.001 (0.004)		-0.013 (0.011)
Observations	2,469	2,469	2,469	2,469
Number of Physician_hospitals	622	622	622	622
Physician FE	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

*Notes:* Robust standard errors clustered at the physician level are reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.5: Effect of Health Information Sharing on Disparate Impact Using Local Linear Extrapolation**

Dependent Variable	Absolute Disparate Impact (1)	Absolute Disparate Impact (2)	Severe Disparate Impact (3)	Severe Disparate Impact (4)
InfoShare	-0.017*** (0.007)	-0.016** (0.007)	-0.072*** (0.024)	-0.066*** (0.024)
Cumulative ER encounters (in 1000s)		-0.007*** (0.002)		-0.013** (0.006)
Share of Hispanic Patients		0.033 (0.073)		-0.012 (0.206)
Average Comorbidities		-0.001 (0.004)		-0.015 (0.011)
Observations	2,469	2,469	2,469	2,469
Number of Physician_hospitals	622	622	622	622
Physician FE	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

*Notes:* Robust standard errors clustered at the physician level are reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.6: IV Exogeneity Tests**

VARIABLES	(1) Number of ER Encounters	(2) Number of Hispanic ER Encounters	(3) Readmission Rate
Remote Partner InfoShare	-16.440 (23.170)	-7.228 (13.547)	-0.013 (0.010)
Constant	144.629*** (16.979)	74.981*** (9.927)	0.238*** (0.008)
Observations	2,469	2,469	1,965
R-squared	0.678	0.729	0.696
Physician FE	YES	YES	YES
Hospital FE	YES	YES	YES
Year FE	YES	YES	YES

*Notes:* Robust standard errors clustered at the physician level are reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table B.7: Effect of Health Information Sharing on Severe Disparate Impact using IV regressions**

Dependent Variable	First Stage		Dependent Variable	Second Stage	
	(1) InfoShare	(2) InfoShare		(3) Severe Disparate Impact	(4) Severe Disparate Impact
Remote Partner InfoShare	0.958*** (0.033)	0.951*** (0.033)	InfoShare	-0.133** (0.057)	-0.106* (0.060)
Cumulative ER encounters (in 1000s)		-0.002 (0.007)	Cumulative ER encounters (in 1000s)		-0.012** (0.005)
Share of Hispanics		-0.078 (0.192)	Share of Hispanics		-0.031 (0.208)
Average Comorbidities		0.009 (0.009)	Average Comorbidities		-0.013 (0.011)
Observations	2,469	2,469	Observations	2,469	2,469
Number of Physician_hospitals	622	622	Number of Physician_hospitals	622	622
Physician FE	YES	YES	Physician FE	YES	YES
Hospital FE	YES	YES	Hospital FE	YES	YES
Year FE	YES	YES	Year FE	YES	YES

Notes: Robust standard errors clustered at the physician level are reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

**Table B.8: Falsification Tests based on Patient Subsample Analysis**

Dependent Variable	Absolute Disparate Impact		Severe Disparate Impact	
	(1)	(2)	(3)	(4)
InfoShare	-0.030 (0.031)	-0.025 (0.030)	0.058 (0.093)	0.051 (0.093)
Cumulative ER encounters (in 1000s)		-0.030*** (0.003)		0.004 (0.008)
Share of Hispanics		0.237 (0.384)		0.783 (1.322)
Average Comorbidities		-0.009 (0.010)		0.049* (0.028)
Observations	861	861	861	861
Number of Physician_hospitals	244	244	244	244
Physician FE	YES	YES	YES	YES
Hospital FE	YES	YES	YES	YES
Year FE	YES	YES	YES	YES

Notes: Robust standard errors clustered at the physician level are reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table B.9: Estimation Results of Disparate Impact on Alternative Outcomes**

	(1)	(2)	(3)	(4)
	Revisit with heart disease in 6 months (Baseline)	Revisit with chest pain in 6 months	Emergency revisit with heart disease in 6 months	Emergency Revisit with chest pain in 6 months
<i>Panel A: Risk of Heart disease</i>				
White Patients	0.163*** (0.003)	0.151*** (0.003)	0.138*** (0.003)	0.136*** (0.003)
Hispanic Patients	0.116*** (0.002)	0.186*** (0.003)	0.098*** (0.002)	0.164*** (0.003)
<i>Panel B: Disparate Impact</i>				
Mean Across Cases	-0.033*** (0.002)	-0.041*** (0.002)	-0.034*** (0.002)	-0.041*** (0.002)
Physicians	401	401	401	401

*Notes:* This table summarizes estimates of mean risk and disparate impact from linear extrapolations of the variation in Figure 2. Panel A reports the 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, are in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1.

**Table B.10: Effect of Health Information Sharing on Disparate Impact based on Hispanic Population**

Dependent Variable	Absolute Disparate Impact	Absolute Disparate Impact	Severe Disparate Impact	Severe Disparate Impact
	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)
InfoShare	-0.016** (0.007)	-0.053** (0.023)	-0.074*** (0.026)	-0.123 (0.078)
InfoShare × Low Hispanic Region	-0.007 (0.009)	-0.001 (0.014)	0.014 (0.032)	0.021 (0.047)
Cumulative ER encounters (in 1000s)	-0.007*** (0.002)	-0.006*** (0.001)	-0.013** (0.005)	-0.012** (0.005)
Share of Hispanics	0.034 (0.072)	0.038 (0.073)	-0.038 (0.209)	-0.032 (0.208)
Average Comorbidities	-0.001 (0.004)	0.001 (0.004)	-0.015 (0.011)	-0.013 (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

*Notes:* Low Hispanic Region is defined as a region where the proportion of Hispanics is below the 25th percentile value (19.65%). Robust standard errors clustered at the physician level are reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1

### Appendix C: Effect of Health Information Sharing on Disparity in Missed Diagnosis

As discussed in the main paper, missed diagnosis refers to cases where a patient does not receive a diagnosis for a particular disease during their initial visit but is later diagnosed with the disease in subsequent visits (Kwok et al. 2021). Although this concept is closely related to diagnostic disparate impact, it represents a slightly different aspect of the diagnoses process. Diagnostic disparate impact emphasizes the broader implications of unequal diagnostic rates which can disproportionately affect minority groups in high-risk settings. On the other hand, missed diagnosis focuses on the disparity in the accuracy of diagnoses. Given the important distinction, we consider the question on how health information sharing influences the disparity in missed diagnosis between majority and minority patients in this supplemental analysis.

We focus on ER visits for cardiovascular conditions and restrict our sample to cases where no heart disease diagnosis was made in a focal visit ( $D_i = 0$ ). Missed diagnosis is defined as a revisit (that occurred) to the ER or EMS with a heart disease diagnosis within one week of the focal visit (Chang et al. 2019; Moy et al. 2015). We exclude cases with inpatient admissions within one week of the focal visit, as inpatient settings involve different diagnostic protocols and resources. Our model is specified in equation (C1).

$$\begin{aligned} \text{Missed Diag}_{ijt} = & \beta_0 + \beta_1 \text{InfoShare}_{jht} + \beta_2 \text{Hispanic}_i + \beta_3 \text{InfoShare}_{jht} * \text{Hispanic}_i \\ & + \beta_4 X_i + \text{Year}_t + \text{Physician}_j + \text{Hospital}_h + \varepsilon_{ijt} \end{aligned} \quad (\text{C1})$$

$\text{Missed Diag}_{ijt}$  is the dependent variable indicating a missed diagnosis for patient  $i$ , treated by physician  $j$  in year  $t$ .  $\text{InfoShare}_{jht}$  captures the level of health information used by physician  $j$  that is shared by outside providers in hospital  $h$  in year  $t$ .  $\text{Hispanic}_i$  is an indicator for Hispanic patients, and  $X_i$  includes patient-level covariates such as gender, age, and number of comorbidities.  $\text{Physician}_j$ ,  $\text{Hospital}_h$ , and  $\text{Year}_t$  are physician, hospital, and year fixed effects, respectively.

The results, as presented in Table C.1, provide several interesting insights. First, as shown in column (1), Hispanic patients are significantly more likely to experience missed diagnoses compared to White patients. However, as observed by the coefficient estimate for the interaction term  $\text{Hispanic} \times \text{InfoShare}$ , greater health information sharing effectively reduces this difference. Next, we split our

sample based on physician skill, as shown in columns (2) and (3). They seem to suggest that the impact of health information sharing is more pronounced among low-skilled physicians.

Overall, our results highlight how health information sharing can potentially serve as a critical tool to mitigate disparities in missed diagnosis. It is consistent with our key argument that health information sharing can effectively narrow the gap in the quality of risk signals between majority and minority patients. Due to the theoretical distinction between disparate impact, which is the primary focus of our study, and missed diagnosis, as well as challenges in measuring missed diagnosis, we do not include this analyses in our paper. Nevertheless, we hope our preliminary evidence can encourage further studies to provide more systematic results on this important issue.

**Table C.1: Effect of Health Information Sharing on Missed Diagnosis**

Dependent Variable	Missed Diagnosis		
	(1) All Physician	(2) Low Skill Physician	(3) High Skill Physician
Hispanic	0.002** (0.001)	0.005** (0.002)	0.002** (0.001)
InfoShare	0.003 (0.002)	0.003 (0.003)	0.003 (0.002)
Hispanic × InfoShare	-0.002* (0.001)	-0.005** (0.002)	-0.001 (0.002)
Male	0.006*** (0.001)	0.006*** (0.001)	0.005*** (0.001)
Age	0.001*** (0.001)	0.001*** (0.001)	0.001*** (0.001)
Comorbidities	0.001*** (0.001)	0.001*** (0.001)	0.001*** (0.001)
Observations	273,153	114,445	115,095
R-squared	0.002	0.002	0.002
Physician FE	YES	YES	YES
Hospital FE	YES	YES	YES
Year FE	YES	YES	YES
Marginal Effect of Hispanic Under high <i>InfoShare</i>	-0.00005 (0.0007)	-0.0004 (0.0008)	0.001 (0.001)

Note: Robust standard errors clustered at physician level are reported in parentheses. \*\*\* p<0.01, \*\* p<0.05, \* p<0.1