Access to Care and its Implications for EHR Reliability and Clinical Risk Prediction Model Performance

Anonymous Authors ABC@SAMPLE.COM
University, Country

Abstract

Disparities in access to healthcare have been well-documented in the United States, but their effects on electronic health record (EHR) data reliability and resulting clinical models is poorly understood. Using an All of Us dataset of 134,513 participants, we investigate the effects of access to care on the medical machine learning pipeline, including medical condition rates, data quality, outcome label accuracy, and prediction performance. Our findings reveal that patients with lower access to care have worse EHR reliability as measured by patient self-reported conditions for 78% of examined medical conditions. In two medical settings, we demonstrate that clinical risk predictive performance can be worse for patients with lower access to care with AUC gaps of 3-7%. We develop solutions to mitigate these disparities and find that including patient self-reported conditions improves performance for patients with lower access to care with up to 4.5% higher AUC. More broadly, our results illustrate that access to care may have significant effects on medical data and model development.

Keywords: Access to care, clinical prediction, algorithmic bias

Data and Code Availability Data and code will be made available within the community workspace on the All of Us platform, which can be accessed by registered users of All of Us.

Institutional Review Board (IRB) This study was reviewed and determined to be Not Human Subjects Research by the institional IRB.

1. Introduction

Access to care is a critical component of high quality healthcare. Defined as "the timely use of health services to achieve the best possible health outcome" (Millman, 1993), access to care can be limited by various factors, including affordability, insurance status, availability of doctors and appointments, ability to take time-off from work, and lack of transportation (Call et al., 2014). Barriers to accessing healthcare services have been extensively documented across medical settings. Additionally, disparities in access to care among people of color and other marginalized groups contribute to ongoing health disparities across numerous health conditions (Caraballo et al., 2022). However, our ability to understand the effects of these health disparities remains limited due to insufficient documentation of access in EHRs.

As clinical risk prediction models become more relied upon in healthcare settings to inform clinical decision-making (Davenport and Kalakota, 2019), it is crucial to not only evaluate how these models work on patients with different levels of access, but also investigate how access to healthcare affects EHR collection and reliability. The idea that low access to care will reduce clinical model performance has been widely suggested, but previous studies have lacked a direct measure of health care access to test this hypothesis with race or insurance status being used as potential proxies (Chen et al., 2021; Seyyed-Kalantari et al., 2021; Obermeyer et al., 2019). Given that many people in the United States lack the ability to seek, reach, and afford care, this issue demands careful attention and rigorous study.

Towards this goal, we investigate three research questions (Figure S1a) in our study. First, we quantify how access to care affects the reliability of EHR data through patient engagement with the healthcare system and collected clinical features by examining 37 medical conditions. Second, we evaluate the extent to which state-of-the-art clinical machine learning methods may create biased predictions for patients with low access to care on two clinical tasks. Finally, we develop and compare data-centered and algorithmic mitigation strategies to improve predictions, particularly for patients with low access to care.

2. Data & Methods

We use data from the All of Us Research Program (Murray, 2019). Developed by the National Institute of Health to create a diverse dataset, the All of Us dataset contains over a million people across 50 states and is one of the largest biomedical data resources in the United States. Data include survey and (contingent on participant agreement) linked EHR data from participating provider organizations in the All of Us program. Surveys are administered to participants upon their enrollment in the All of Us program. EHR data are current through July 1, 2022. Our analysis sample included all participants with linked EHR data that answered the "Health Access & Utilization" survey.

We used responses from the All of Us "Health Access & Utilization" survey to define healthcare access for participants. The survey includes questions about affordability and reasons for delaying care within the last 12 months (Section A.1). We defined a participant as having "high access" to healthcare services if they indicated that they did not have issues with affording or delaying care. We defined a participant as having "low access" if they indicated that they had issues affording care *and* had delayed care.

2.1. Computing EHR Reliability

Data quality in the EHR is a well-explored topic (Sauer et al., 2022), with many studies documenting the effect of EHR quality on research results (Bower et al., 2017; Beaulieu-Jones et al., 2021). Prior research has found high rates of discordance between survey data and EHR data on data elements including conditions and demographics (Wagaw et al., 2018; O'Brien et al., 2022).

To test rates of discordance by access to care, we compared self-reported conditions in the All of Us "Personal and Family Health History" survey to health conditions documented in the EHR. We quantified the percent of participants with a self-reported condition that had no record of the same condition in the EHR (or "missing EHR diagnosis rate") for 37 self-reported conditions.

2.2. Implications for Clinical Prediction Models

We study the impact of access on two clinical tasks: predicting 2-year diabetes incidence and predicting 30-day hospital readmission.

	Overall (N=134,513)	**High Access** (N=62,454)	**Low Access** (N=32,177)
Access (%)			
Affordability	42.9	–	100.0
Delayed Care	34.6	–	100.0
Age (Average)	54.8	59.4	46.6
Gender (%)			
Male	33.5	39.4	24.0
Female	63.7	58.2	72.0
Other	2.8	2.4	4.0
Race (%)			
White	70.3	75.4	62.5
Black	10.4	8.8	13.2
Other	19.3	15.7	24.3
Ethnicity (%)			
Hispanic	12.1	8.4	17.1

Table 1: Study sample characteristics. See Table S1 for additional access to care group characteristics

For the 2-year diabetes task, the prediction sample included participants with no record of type 2 diabetes as of the date they completed the All of Us survey data, following prior work on diabetes prediction (Razavian et al., 2015). We predicted whether someone developed type 2 diabetes in the following two years, with the following EHR-based features in a 2-year lookback window: demographics (age, gender, race/ethnicity), condition codes, number of medications, and clinical measures (Section A.2). See Section A.3 for readmission task setup.

Logistic regression and random forests algorithms were trained with 5-fold cross validation. Logistic regression models were developed with a regularization parameter search from 1e−4 to 1e4 in ten logarithmically spaced intervals and across L1 and L2 regularization. Random forest models were developed with a regularization parameter search over 5, 50, 500, and 1000 decision trees. We selected the cutoff point that maximized Youden's J statistic. Model performance (AUC and balanced accuracy) was assessed overall and by access levels. For uncertainty quantification, we compute the numerical 95% confidence intervals as derived from the Central Limit Theorem.

Finally, we tested how the inclusion of access information and (predicted) self-reported data changed the predictive performance by levels of access. Predicted self-reported conditions were constructed using a multi-output model with a similar 5-fold cross-validation setup and parameter search as the baseline predictive models.

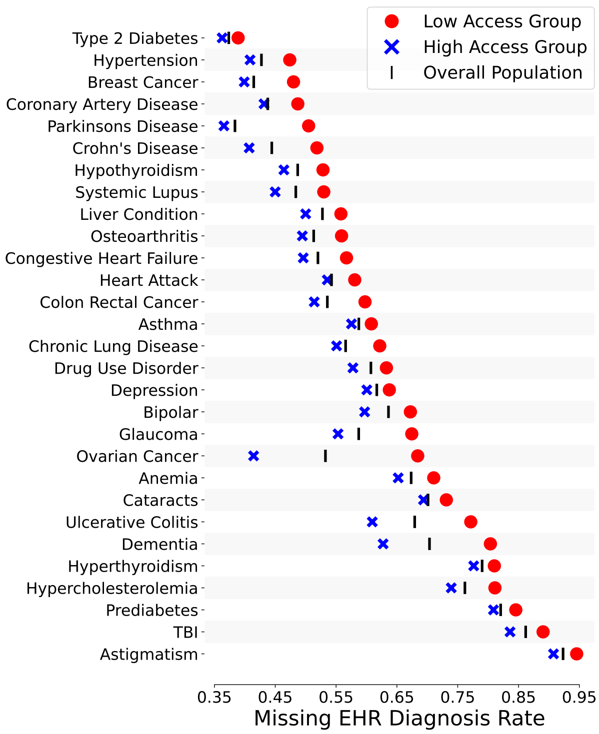

Figure 1: Missing EHR Diagnosis Rate in Low and High Access by Condition

3. Results

3.1. Patients with low access to care have different condition distributions

Our sample included 134,513 All of Us participants for which we had self-reported data on healthcare access and linked EHR data (Figure S1b). More than half of the participants indicated that they had trouble accessing care, either reporting affordability problems (42.9%) and/or that they had delayed care (34.6%) in the previous year (Figure S2).

In Table 1, we compare participants with high versus low access to care. Participants with low access to care were more likely to be younger, female, Black, and Hispanic. After adjusting for age (using 10-year age weights based on the age distribution of the entire sample), participants with low access to care were more likely to report anxiety (26.6% versus 13.9%), depression (36.3% versus 20.2%), asthma (21.8% versus 14.1%), hypertension (32.5% versus 27.2%), high cholesterol (31.4% versus 28.7%), prediabetes (10.5% versus 7.5%), and diabetes (12.5% versus 7.7%). All differences were significant at the $p<0.05$ level, based on a two-sided t-test.

3.2. Patients with lower access to care have lower EHR reliability

When comparing self-reported conditions to EHR-reported conditions, we found high-rates of missingness across the sample: among participants reporting any condition, 87.3% had at least one self-reported condition that was never recorded in the EHR. Rates of missingness varied significantly depending on the type of condition. For example, conditions like type 2 diabetes and Parkinson's disease were more likely to be present in the EHR compared to conditions like acid reflux, astigmatism, and stroke. This difference in missing diagnosis rate could be explained by the fact that the latter conditions might only be addressed in specific care settings.

Participants with low access to care had statistically significant higher rates of missingness for 29 out of 37 conditions compared to the high access group (Figure 1), computed using a chi-squared test with a Benjamani-Hochberg false discovery correction of $\alpha = 0.05$. We observed the largest differences for ovarian cancer (68% versus 41%), and dementia (80% versus 63%). These conditions were fairly rare, reported by less than 1% of participants. Among more common conditions such as depression and hypertension (reported by more than 25% of participants), we still observed significantly higher rates of missingness for low versus high access groups (depression: 64% vs 60%, hypertension: 47% vs 41%). When EHR data is used for predictive algorithms, these differences in medical condition reliability indicate that patients with lower access to care may have worse quality feature data or outcome labels (e.g., predicting type 2 diabetes).

3.3. Algorithmic predictive performance is lower for patients without access to care

In the diabetes prediction task with 52,059 participants, the logistic regression model outperformed the random forests with an AUC of 0.733 (95% CI: 0.729-0.737). In Figure 2, we plot performance results from the logistic regression prediction, stratified by four access to care levels. The high access group had the best AUC with a value of 0.753 (95% CI: 0.747-0.759). The AUC for the high access group was significantly higher than the affordability (0.727, 95% CI: 0.721-0.733) and delayed care (0.725, 95% CI: 0.718-0.733)

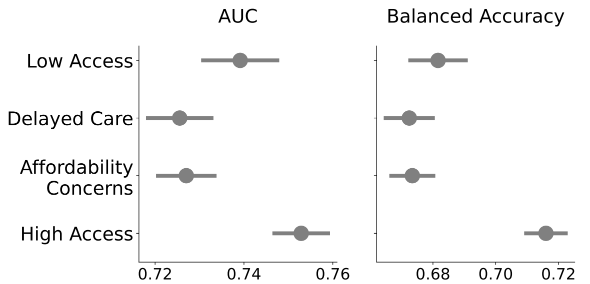

Figure 2: Diabetes diagnosis predictive performance across access to care groups.

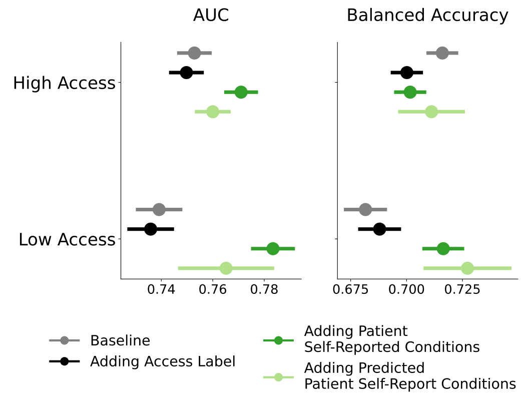

Figure 3: Mitigation strategies for improving predictive performance for diabetes diagnosis across access to care groups.

groups. Balanced accuracy was also lower for these groups due to a large reduction in sensitivity (compared to only modest increases in specificity). For example, we were only able to correctly classify 57% and 54% of new diabetes cases for the affordability and delayed care groups (respectively) compared to 66% for the high access group. Lower sensitivity for the low access group may lead to less comparative clinical interventions.

Similar to the diabetes prediction task, we observe a slightly higher performance for patients with high access to care compared to patients with delayed care and affordability concerns in the readmission prediction task (Figure S3). Results are more stark in balanced accuracy compared to AUC.

3.4. Self-report medical history can improve performance for patients with low access to care

In Figure 3, we evaluate two strategies for improving performance for patients with low access to care: (1) including access groups as features in the clinical prediction model and (2) including (observed or predicted) self-reported condition data.

Incorporating the access groups (e.g, “low access”) could improve prediction if there was differential risk of type 2 diabetes incidence for those with low access to care which was not captured by the features in the model (Zink et al., 2024). Surprisingly, we did not find any differences in the predictive performance when adding in this label: the performance results were almost exactly the same as the baseline model.

Next, we tested whether including self-reported conditions as features improved the predictive performance for those with low access to care. When we included all self-reported conditions in the diabetes prediction model (except for the set of self-reported diabetic conditions), we observed a significant increase in AUC for both access groups, but particularly those with lower access to care: the AUC for those with low access went from 0.74 (95% CI: 0.731-0.748) in the baseline model to 0.78 (95% CI: 0.776-0.791) in the model that included self-reported health conditions. This improvement was reflected in an increased sensitivity for those with lower access to care. When patient self-reported conditions may not be available, we found that even the inclusion of *predicted* self-reported conditions led to a statistically-significant improvement in both AUC and balanced accuracy.

4. Discussion

In our study, we found lower EHR reliability among All of Us participants that reported low access to care. Furthermore, we report disparities in access to care in the predictive performance of two clinical tasks. Our analysis suggests that this performance gap is partly driven by missing and noisy data: when we incorporated self-reported condition data into the type 2 diabetes model, we significantly improved the predictive performance for those with low access to care. Potential next steps may include extending our framework to more granular barriers to care and investigating the implications for clinical interventions. This work underscores the importance of using multimodal and patient-centered data sources to quantify, understand, and address health disparities based on access to care.

References

Brett K Beaulieu-Jones, William Yuan, Gabriel A Brat, Andrew L Beam, Griffin Weber, Marshall Ruffin, and Isaac S Kohane. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? *NPJ Digit. Med.*, 4(1):62, March 2021.

Julie K Bower, Sejal Patel, Joyce E Rudy, and Ashley S Felix. Addressing bias in electronic health record-based surveillance of cardiovascular disease risk: Finding the signal through the noise. *Curr. Epidemiol. Rep.*, 4(4):346–352, December 2017.

Kathleen T Call, Donna D McAlpine, Carolyn M Garcia, Nathan Shippee, Timothy Beebe, Titilope Cole Adeniyi, and Tetyana Shippee. Barriers to care in an ethnically diverse publicly insured population: is health care reform enough? *Med. Care*, 52(8):720–727, August 2014.

César Caraballo, Chima D Ndumele, Brita Roy, Yuan Lu, Carley Riley, Jeph Herrin, and Harlan M Krumholz. Trends in Racial and Ethnic Disparities in Barriers to Timely Medical Care Among Adults in the US, 1999 to 2018. *JAMA Health Forum*, 3 (10):e223856, October 2022.

Irene Y Chen, Emma Pierson, Sherri Rose, Shalmali Joshi, Kadija Ferryman, and Marzyeh Ghassemi. Ethical machine learning in healthcare. *Annu Rev Biomed Data Sci*, 4:123–144, July 2021.

Thomas Davenport and Ravi Kalakota. The potential for artificial intelligence in healthcare. *Future Healthc. J.*, 6(2):94–98, June 2019.

M Millman. Access to health care in america. 1993.

Jeff Murray. The "all of us" research program. *N. Engl. J. Med.*, 381(19):1884, November 2019.

Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464):447–453, October 2019.

Emily C O'Brien, Hillary Mulder, W Schuyler Jones, Bradley G Hammill, Amber Sharlow, Adrian F Hernandez, and Lesley H Curtis. Concordance between patient-reported health data and electronic health data in the ADAPTABLE trial. *JAMA Cardiol.*, 7(12):1235–1243, December 2022.

Narges Razavian, Saul Blecker, Ann Marie Schmidt, Aaron Smith-McLallen, Somesh Nigam, and David Sontag. Population-Level Prediction of Type 2 Diabetes From Claims Data and Analysis of Risk Factors. *Big Data*, 3(4):277–287, December 2015.

Christopher M Sauer, Li-Ching Chen, Stephanie L Hyland, Armand Girbes, Paul Elbers, and Leo A Celi. Leveraging electronic health records for data science: common pitfalls and how to avoid them. *Lancet Digit. Health*, 4(12):e893–e898, December 2022.

Laleh Seyyed-Kalantari, Haoran Zhang, Matthew B A McDermott, Irene Y Chen, and Marzyeh Ghassemi. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nat. Med.*, 27 (12):2176–2182, December 2021.

Fikirte Wagaw, Catherine A Okoro, Sunkyung Kim, Jessica Park, and Fred Rachman. Linking data from health surveys and electronic health records: A demonstration project in two chicago health center clinics. *Prev. Chronic Dis.*, 15:E09, January 2018.

Anna Zink, Ziad Obermeyer, and Emma Pierson. Race adjustments in clinical algorithms can help correct for racial disparities in data quality. *Proc. Natl. Acad. Sci. U. S. A.*, 121(34):e2402267121, August 2024.

Appendix A. Supplementary Material

A.1. Reasons for low access to care

In Figure S2, we report the rates of response for each question in the affordability and delayed care categories. Cited reasons for delaying care included having to pay out of pocket, nervousness, and inability to take time off from work. Cited affordability issues included having to take a lower cost prescription to save money, having to pay out of pocket, and nervousness about affording care.

A.2. Defining features for the diabetes prediction model

The All of Us EHR data are collected in a set of data tables standardized using the OMOP format.[1] Condition data are stored in the "condition_occurrence" table. To create condition features, we created a binary indicator for each condition with a prevalence of at least 1% in the prediction sample, resulting in 354 features. Number of medications was calculated using the "drug_exposure" table by counting the number of unique drug concept ids per participant over the two-year lookback window. Clinical measures were calculated using the "measurement" table. Value ranges for each clinical measure can be found in Table S2.

A.3. 30-day hospital readmission prediction task

We created a prediction sample by identifying all participants with an inpatient hospitalization on or after the date on which the participant completed the All of Us surveys. Our outcome is an indicator for whether the patient was readmitted to the hospital within 30 days (1 = Yes, 0 = No). We used the following features: age, gender, race/ethnicity, length of stay, and clinical information for the visit: condition codes (indicator for the presence of any condition code occurring in at least 1% of the prediction sample), the number of unique procedures recorded during the hospitalization, the number of unique prescription drugs administered during the admission, and an indicator for clinical measures found in Table S2.

1. https://www.ohdsi.org/data-standardization/

A.4. EHR reliability, continued

On average, those with low access were missing 2.1 diagnoses compared to 1.7 for those with high access ($p < 0.001$). When we ran a regression predicting the presence of an EHR condition given the presence of the self-reported condition and access variable, controlling for conditions, we found that access to care was negatively correlated with the presence of an EHR condition by 1% ($p < 0.001$).

Our main calculation included all EHR data prior to the date of self report, but we additionally assessed how rates of missingness changed under various lookback windows, as is common in prediction models. If we only looked in the EHR in the prior year, this amount doubled to 2%.

A.5. Experiment results, continued

For the diabetes prediction task, the random forest model results show similar results: patients with high access to care have stronger predictive performance with an AUC of 0.723 (95% CI: 0.717-0.729) compared to than patients with lower access to care, e.g., patients with affordability with an AUC of 0.704 (95% CI: 0.698-0.711). Similarly, adding patient self-reported conditions improved performance, increasing the AUC of 0.749 (95% CI: 0.740-0.757) for baseline model to 0.807 (95% CI: 0.790-0.823) for the model which includes patient self-reported conditions.

In the rehospitalization task, we observe similar logistic regression results (Figure S3) as the diabetes task. Predictions demonstrate a higher algorithmic performance for patients with high access to care than patients with low access to care, but the differences were not statistically significant potentially due to the smaller sample size.

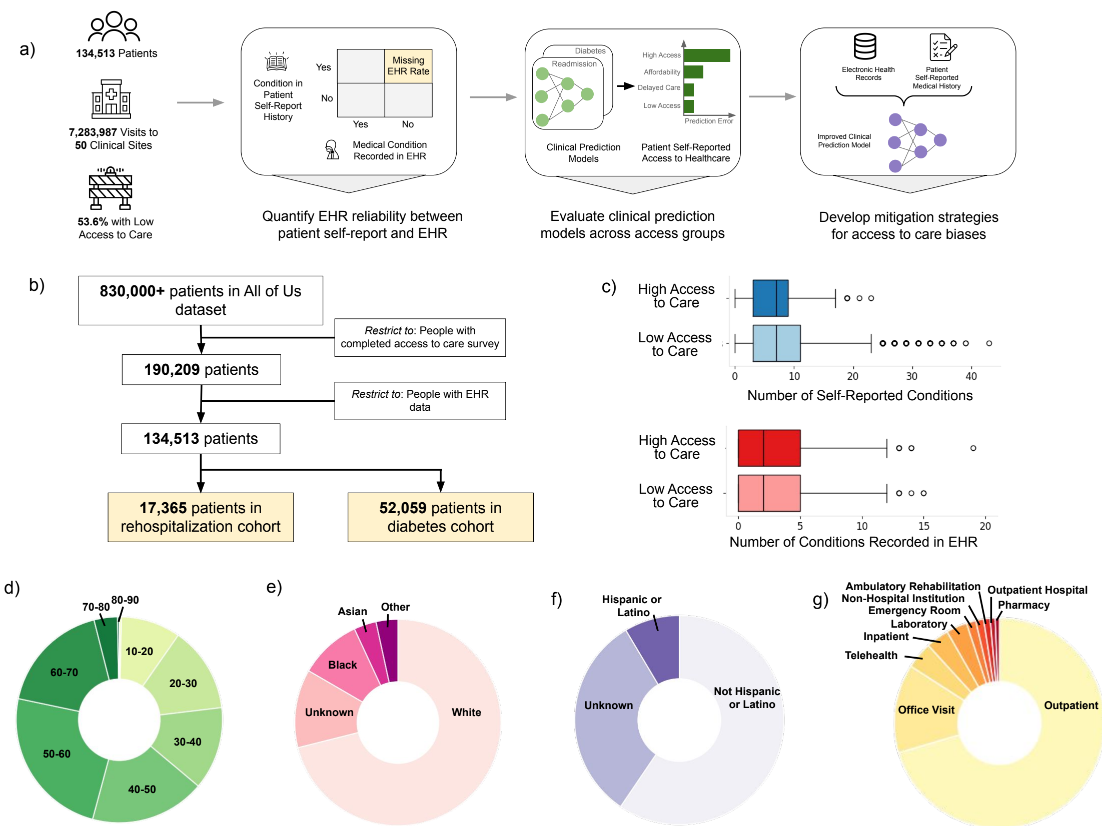

Figure S1: Study overview

	Overall (N=134,513)	High Access (N=62,454)	Low Access (N=32,177)	Afford-ability (N=57,654)	Delayed Care (N=46,582)
Access (%)					
Affordability	42.9	–	100.0	100.0	69.1
Delayed Care	34.6	–	100.0	55.8	100.0
Age (Average)	54.8	59.4	46.6	51.2	47.3
Gender (%)					
Male	33.5	39.4	24.0	28.1	25.7
Female	63.7	58.2	72.0	68.6	70.6
Other	2.8	2.4	4.0	3.3	3.7
Race (%)					
White	70.3	75.4	62.5	65.1	64.6
Black	10.4	8.8	13.2	12.1	12.3
Other	19.3	15.7	24.3	22.8	23.0
Ethnicity (%)					
Hispanic	12.1	8.4	17.1	15.8	15.8

Table S1: Expanded study sample characteristics

Condition	Lab Measurement (Range)
Obesity	BMI (>30)
Fever	Temperature (>100.4 Fahrenheit)
Fever	Temperature (>38 Celsius)
Hypocalcemia	Calcium (< 8.5)
Hypercalcemia	Calcium (>10.5)
Creatinine	Creatinine (>1.3 and < 3.4)
High Blood Pressure	Systolic BP (> 130)
High Blood Pressure	Diastolic BP (>80)
Diabetes	Glucose (>126)
Tachycardia	Heart Rate (> 100)
Anemia	Hemoglobin (<12)
High Hemoglobin	Hemoglobin (>15)
Hypoxemia	Oxygen Value (<.90)
Hyperkalemia	Potassium (>6)
Hypokalemia	Potassium (<2.5)
Tachypneic	Respiratory Rate (>20)
Bradypnea	Respiratory Rate (<12)
Hypernatremia	Sodium (>145)
Hyponatremia	Sodium (<135)
Low Urea	Urea (<8)
High Urea	Urea (>24)

Table S2: Lab-based EHR Conditions

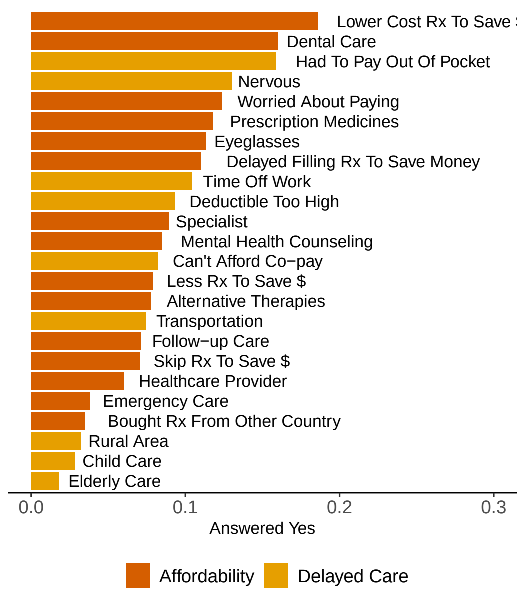

Figure S2: Response Rate to Access to Care Questions

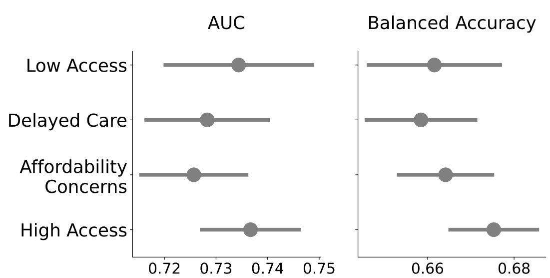

Figure S3: Comparison of rehospitalization performance across access to care patient groups.