An individual-level socioeconomic measure for assessing algorithmic bias in health care settings: A case for HOUSES index

Running Head: A socioeconomic approach for assessing algorithmic bias in asthma care

Young J. Juhn^{1,2*}, M.D., M.P.H., Euijung Ryu^{3*}, Ph.D., Chung-Il Wi^{1,2}, M.D., Katherine S. King³, M.S., Santiago Romero Brufau, M.D., Ph.D.,⁴ Chunhua Weng, Ph.D.,⁵ Sunghwan Sohn, Ph.D.,⁶ Richard Sharp, Ph.D.,⁷ John D. Halamka, M.D., M.S.^{4,8}

¹Precision Population Science Lab, Mayo Clinic, Rochester, Minnesota;

²Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, Minnesota;

³Department of Quantitative Health Sciences, Mayo Clinic, Rochester, Minnesota

⁴Department of Internal Medicine, Mayo Clinic Rochester, Minnesota

⁵Department of Biomedical Informatics, Columbia University

⁶Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, Minnesota;

⁷Biomedical Ethics Program, Mayo Clinic, Rochester, Minnesota

⁸Mayo Clinic Platform

*These authors contributed equally to this work.

* Corresponding author

Email: juhn.young@mayo.edu (YJ) and ryu.euijung@mayo.edu (ER)

Sources of Support/Funding and their role: This work was supported by National Institute of Health (NIH)-funded R01 grant (R01 HL126667), R21 grant (R21AG65639-01A1) and R21 grant (R21AI142702).

Financial Disclosure: Young J. Juhn is Principal Investigator (PI) of the Respiratory Syncytial Virus incidence study supported by GlaxoSmithKline but they have no relationship with the presented work.

Conflicts of interest: The authors declare no conflict of interest pertaining to the presented work.

Data Sharing Statement: The datasets generated and/or analyzed during the current study are not publicly available as they include protected health information. Access to data could be discussed per the institutional policy after approval of the IRB at Mayo Clinic.

Abstract

While artificial intelligence (AI) algorithms hold great potential for improving health and reducing health disparities, biased AI algorithms have a potential to negatively impact the health of under-resourced communities or racial/ethnic minority populations. Our study highlights the major role of socioeconomic status (SES) in AI algorithm bias and (in)completeness of electronic health records (EHRs) data, which is commonly used for algorithm development. Understanding the extent to which SES impacts algorithmic bias and its pathways through which SES operates its impact on algorithmic bias such as differential (in)completeness of EHRs will be important for assessing and mitigating algorithmic bias. Despite its importance, the role of SES in the AI fairness science literature is currently under-recognized and under-studied, largely because objective and scalable individual-level SES measures are frequently unavailable in commonly used data sources such as EHRs. We addressed this challenge by applying a validated individual-level socioeconomic measure that we call the HOUSES index. This tool allows AI researchers to assess algorithmic bias due to SES. Although our study used a cohort with a relatively small sample size, these study results highlight a novel conceptual strategy for quantifying AI bias by SES.

Keywords:

Artificial intelligence, algorithmic bias, ethics, social determinants of health, fairness, health equity, pediatric asthma

Abbreviations:

AE: Asthma exacerbation

AI: Artificial Intelligence

EHRs: Electronic health records

FN: False negatives

FP: False positives

GBM: Gradient Boosting Machine

HOUSES: HOUsing-based SocioEconomic Status measure;

ML: Machine Learning

NB: Naïve Bayes

NAEPP: National Asthma Education and Prevention Program;

PAC: Predetermined Asthma Criteria

SDH: Social Determinants of Health

SES: Socioeconomic status

TN: True negatives

TP: True positives

Introduction

Augmented computing power, storage capability and predictive analytics have accelerated the adoption and deployment of artificial (augmented or machine) intelligence (AI) tools in health care system of the United States (US).^{1,2} As of 2017, 96% of U.S. hospitals adopted certified EHRs and 98% of U.S. hospitals demonstrated meaningful use of at least one certified health information technology (HIT).¹ A recent survey showed that 90% of health care executives in the US reported they had AI tools and automation strategy in 2020 compared to 53% in 2019.² As of 2020, more than 80 imaging-related AI algorithms have been cleared by the US Food and Drug Administration (FDA).^{3,4} A research group at Mayo Clinic showed that an electrocardiogram (ECG)-based, AI-powered clinical decision support (CDS) tool demonstrated early diagnosis of low ejection fraction, a condition that is underdiagnosed but treatable, via a randomized clinical trial (RCT).⁵ The intervention increased the diagnosis of low ejection fraction in the overall cohort (1.6% in the control arm versus 2.1% in the intervention arm, odds ratio (OR) 1.32 (1.01–1.61), $P = 0.007$).

While AI tools hold a great potential for improving health and even reducing health disparities, biased AI algorithm can negatively impact health in under-resourced or racial/ethnic minority populations as reported in the literature.⁶⁻⁹ For example, a recent study of a widely used commercial AI algorithm identified 18% of Black patients needing additional care for chronic disease management (compared to White patients). When a mitigation strategy was applied to reduce algorithmic bias, the percentage increased to 47% for Black patients⁶, demonstrating the potential harm of AI tools when the bias is not mitigated before implementation. Others reported algorithmic biases by race and socioeconomic status (SES) in post-partum depression,¹⁰ ICU mortality¹¹ and 30-day psychiatric readmission.¹¹ Therefore, while differential predictability of

models by patient characteristics such as SES is a hardly new phenomenon in biomedical research,¹² it has a major implication on health equity because of the potential to exacerbate inequities on a monumental scale when deployed on a large scale in clinical care.

Given the significant impact of SES on health risk and healthcare access, quantifying the degree of algorithmic bias by SES has important ethical implications for AI research as well as health care delivery research. SES is a key element of social determinants of health (SDH) in health care delivery and research¹³⁻²² and considered as a major factor accounting for differential health outcomes through their associated biological (eg, epigenetics, gene expression or telomere length), behavioral and environmental factors.²²⁻²⁵ Through influences on health care utilization, SES may shape the completeness of EHRs as a data source for developing AI algorithms and result in algorithmic bias by SES. Thus, SES is an important variable for understanding the nature of algorithmic bias stemming from differential health risk, healthcare access and completeness of EHRs and for addressing algorithmic bias in health care.

Unfortunately, individual-level SES measures that are objective, precise and scalable are frequently unavailable in commonly used data sources for clinical care and research²⁶ posing a major barrier to health care delivery and research as acknowledged by National Academy of Medicine and National Quality Forum.^{8,27,28} Also, the limited availability of suitable individual-level SES in EHRs is a major challenge in addressing algorithmic bias in health care settings. Consequently, current AI fairness science focuses on readily available limited demographic factors such as age, sex and race/ethnicity,²⁹ leaving the role of SES in AI bias poorly understood.

To address this major roadblock in the equitable implementation of health care AI, we propose to use the HOUSES (individual HOUSing-based SES) index to assess key features of

SES in AI research such as accuracy, precision, objectivity and scalability. In this work, we 1) assessed differential data availability and validity of EHRs among study subjects according to SES as measured by HOUSES and 2) applied HOUSES index to quantify algorithmic bias by SES based on commonly used metrics for assessing AI bias.

Methods

Study Population and Setting: The study was conducted in primary care practices (i.e., including teaching pediatric faculty, residents, and nurse practitioners) in Olmsted County in southeastern Minnesota (MN). The study population and setting were described in our recent report.³⁰ Briefly, Olmsted County, is a virtually self-contained health care environment with only two health care systems provide clinical care to nearly all residents. About 98% of residents authorize use of their medical records for research.³¹ According to U.S. census data in 2010, the age, sex, and ethnic characteristics of Olmsted County residents were similar to those of the state of Minnesota and the Upper Midwest.³² However, Olmsted County has become more diverse as indicated by the racial/ethnic characteristics of children enrolled in public schools (in 2019, 35.2 % reported to be a racial/ethnic minority groups). Mayo Clinic Primary Care Pediatric Practices offers primary care service at four locations within Olmsted County and this study was conducted in Baldwin Primary Care Practice site, the largest of the four practice sites. Asthma is the most prevalent chronic illness with the third highest health care expenditures in children and adolescents in Olmsted County, MN.³³ The asthma prevalence in the primary care practice (14%) is slightly lower than that of the county (17.6%).³⁴

Study Design and Subjects: The study design and subjects were described in our recent report.³⁰ Briefly, the study was designed as a cross-sectional study. We used the data from the training

and testing cohorts that were used to develop ML algorithms applied to a previous study that developed a novel AI-assisted clinical decision support system named A-GPS (Asthma-Guidance and Prediction System), which is based on a single-center pragmatic randomized clinical trial (RCT). A-GPS-based intervention in the original study provided clinicians with the summary of most relevant information for asthma management and the risk prediction for asthma exacerbation using ML algorithms.³⁰ It significantly reduced clinicians' burden for EHRs review resulting in more efficient asthma management by providing more actionable guidance. The focus of the original study was to assess the effectiveness of A-GPS on asthma outcomes (e.g., asthma exacerbation, asthma control, asthma-related health care utilization, asthma care quality and health care costs). The original study used data from subjects who had persistent asthma or met Predetermined Asthma Criteria (PAC). In this present report, we limited the primary analysis assessing algorithmic bias to those with persistent asthma in order to focus on more homogeneous patient groups with persistent asthma. Details of the original study have been reported.³⁰ For an additional analysis, we used the subjects in the original study who met PAC definition but were not yet diagnosed with asthma at the time of enrollment. This study (IRB number:15-004435) was approved by the Mayo Clinic Institutional Review Board (IRB).

ML algorithms for predicting AE: For A-GPS, we trained and tested two ML algorithms: Naïve Bayes (NB) model and gradient boosting machine (GBM) model for predicting 1-year AE risk among children with asthma. We extracted 29 candidate variables based on the literature including sociodemographics, risk factors, and asthma outcomes from EHR over a prior three - year period. The original study included 590 subjects (300 in the training and 290 in the test set) who had persistent asthma or met Predetermined Asthma Criteria from Mayo Clinic pediatric practice panel, respectively. Receiver operating characteristic (ROC)- Areas Under Curve for NB

and GBM model were 0.78 and 0.74 on the testing cohort, respectively. These algorithms were used to quantify algorithmic bias by SES in this report.

Algorithmic fairness metrics: We used the widely reported common metrics as defined in Table 1. As it is impossible to satisfy all metrics (‘impossibility theorem’) ^{35,36} and the literature suggests to use false positive rate (FPR) and false negative rate (FNR),^{35,36} we used error rate, defined as the sum of FPR and FNR, as the primary metric for assessing algorithmic bias in this presented work. For each metric, we calculated the ratio comparing least privileged group (e.g., HOUSES

Table 1. Metrics for assessing algorithmic fairness used in this study

Metric*	Definition**
Accuracy equality	$(TP+TN)/(TP+FP+TN+FN)$
Equal opportunity (sensitivity)	$TP/(TP+FN)$
Predictive equality (false positive rate)	$FP/(FP+TN)$
Predictive parity (positive predictive value)	$TP/(TP+FP)$
Error rate (FPR+FNR)	$FP/(FP+TN) + FN/(TP+FN)$
* For each metric, ratio was by scores from unprivileged group, divided by that from privileged group. For FPR and error rate, with the ratio greater than 1 meaning that the algorithm is more favorable to privileged group, while more favorable to unprivileged group for the other three metrics.	
** TP: true positives; FP: false positives; TN: true negatives; FN: false negatives; FPR: false positive rate; FNR: false negative rate	

Q1, see below) with the privileged group (HOUSES Q2-Q4). For FPR and error rate, ratio greater than 1 means that the algorithm is more favorable to privileged group, while more favorable to less privileged groups for the other three metrics (accuracy equality, equal opportunity and predictive parity). As a rule of thumb, a ratio that is less than 0.8 or greater than 1.25 (1/0.8) is considered as meaningful difference, which is implemented in an open source program such as AIF360.³⁷

HOUSES: HOUSES is an individual-level SES measure based on 4 real property data variables of an individual housing unit after principal component factor analysis: *housing value, square footage, number of bedrooms, and number of bathrooms*. An individual’s address from the EHR

is directly linked to the publicly available assessor's data (which is a basis for property tax and thus is available throughout US counties and cities).²⁶ We formulated a standardized HOUSES index score by summing these variables after z-score transformation. The greater the HOUSES index, the higher the SES. Since its development, HOUSES has been extensively applied as a validated SES measure that has shown association with numerous health-related outcomes, including acute/chronic conditions, healthcare access issues, healthcare utilization, and other health-related behaviors such as smoking and vaccine status as summarized in Table 2.

Table 2. The reported health outcomes predicted by HOUSES

A. Chronic conditions and mortality
Adults: 1) Risk of rheumatoid arthritis (RA) and post-RA mortality, ³⁸ 2) risks of coronary heart disease, asthma, diabetes, hypertension, mood disorder, ³⁹ Post-Myocardial Infarction mortality, ⁴⁰ 3) multiple chronic conditions, ⁴¹ 4) post-Glioma mortality, ⁴² 5) renal transplantation outcome ⁴³ , 6) ICU mortality, ⁴⁴ 7) persistent depressive symptoms and remission of depressive symptoms, ⁴⁵
Children: 8) Risks of asthma, epilepsy, mood disorders, ⁴⁶ 9) overweight, ^{26,47} 8) asthma outcome ⁴⁸ , 10) multiple complex chronic conditions, ⁴⁹ 11) serum 25 (OH)D concentration ⁵⁰
B. Acute conditions
Adults: 12) Risk of accidental falls ⁵¹ , 13) osteoporotic fracture incidence ⁵²
Children: 14) Risk of invasive pneumococcal disease, ⁵³ 15) risk of bronchiolitis, pneumonia, urinary tract infection, adverse childhood experiences ⁴⁶
C. Other health outcomes
Adults: 16) All-cause hospitalization, ⁴¹ 17) smoking status, ⁵⁴ 18) detecting survey biases, ⁵⁵ 19) end of life care access (advance directives, social work consultation) ⁵⁶
Children: 20) Low birth weight, household smoking status, ^{26,47} 21) pertussis vaccine up-to-date status, ⁵⁷ 22) self-rated health ⁵⁸ , 23) HPV vaccine (initiation and completion) up-to-date status ⁵⁹

Other pertinent variables: While our focus is to quantify algorithmic bias by SES, we also considered other readily available demographic characteristics (age, sex, and race/ethnicity), and pediatric chronic conditions defined by Feudtner et al (an accepted measure of pediatric chronic conditions in literature).⁴⁹ These variables are extracted from patient's EHR. For chronic conditions, ICD-9 diagnostic and procedure codes were used. For simplicity, age (<12 vs ≥12 years) and chronic conditions (yes vs no) are dichotomized. To demonstrate the impact of SES on completeness of EHR, we compared availability of 7 key variables that are clinically relevant to childhood asthma management (health maintenance visit, asthma compliance, asthma severity, asthma type, NAEPP recommendation, smoking status, and missing school; See Table 2). These

variables were extracted from EHR in the 3 years prior to the study index date. Additionally, we assessed data validity defined as having ICD-9 codes for asthma among those who met PAC definition but were not yet diagnosed with asthma at the time of the study.³⁰ Specifically, we previously reported a significant number of children had undiagnosed asthma by comparing asthma prevalence by ICD code-based asthma ascertainment with that by NLP-based ascertainment using PAC (sensitivity 31% (ICD-9) vs. 81% (NLP logic) and 85% (NLP ML)).⁶⁰⁻

62

Table 3. Subject characteristics used in the study		
	Training cohort (N=133)	Testing cohort (N=113)
Age (in years), n (%)		
< 12	94 (71%)	80 (71%)
≥ 12	39 (29%)	33 (29%)
Sex, n (%)		
Male	76 (57%)	67 (59%)
Female	57 (43%)	46 (41%)
Race/ethnicity, n (%)		
Non-Hispanic Whites	76 (57%)	67 (59%)
Others	57 (43%)	46 (41%)
HOUSES, n (%)		
Q1 (the lowest SES)	22 (18%)	15 (14%)
Q2-Q4	102 (82%)	92 (86%)
Missing	9	6
Chronic condition, n (%)		
Yes	30 (23%)	19 (17%)
No	103 (77%)	94 (83%)
Asthma exacerbation, n (%)		
Yes	34 (26%)	40 (35%)
No	99 (74%)	73 (65%)

Data analysis: In this presented work, we quantified algorithmic bias for two ML algorithms (NB and GBM models for predicting 1-year AE risk among pediatric asthmatics) by demographic factors (age, sex, race/ethnicity), SES (HOUSES), and chronic condition. This was done using the testing cohort, because model performance metrics in the training cohort are generally overestimated. To see the impact of SES on completeness of EHR, we also calculated

proportions of subjects with missing or unknown information for 7 key variables for asthma management measuring data availability. Based on our earlier work, we assessed one variable measuring data validity (i.e., diagnosed vs. undiagnosed asthma by ICD codes for those who met PAC^{61,62}) by SES. This calculation was done in both the training and testing cohorts.

Table 4. Proportion of subjects with asthma exacerbation (AE) by subject characteristics		
	Percentage of subjects with AE	
	Training cohort	Testing cohort
Age (in years), n (%)		
< 12	28 (29.8%)	30 (37.5%)
≥ 12	6 (15.4%)	10 (30.3%)
Sex, n (%)		
Male	25 (32.9%)	23 (34.3%)
Female	9 (15.8%)	17 (37.0%)
Race/ethnicity, n (%)		
Non-Hispanic Whites	19 (25.0%)	25 (37.3%)
Others	15 (26.3%)	15 (32.6%)
HOUSES, n (%)		
Q1 (the lowest SES)	6 (27.3%)	8 (53.3%)
Q2-Q4	23 (22.5%)	29 (31.5%)
Chronic condition, n (%)		
Yes	10 (33.3%)	7 (36.8%)
No	24 (23.3%)	33 (35.1%)

Results

Subject characteristics: Subjects included in the training cohort were younger (71% with younger than 12 years old), more male (57%), more non-Hispanic White (57%)) as shown in Table 3. Roughly 20% of the subjects were in the low-SES (HOUSES Q1) group and 20 % had at least one chronic condition. Subject characteristics were similar between training and testing cohorts. Roughly 30% of subjects had asthma exacerbation within 1-year follow-up period (26% in the training cohort and 35% in the testing cohort: Table 3). Table 4 showed that proportion of AE differed by subject characteristics. In general, the proportion was higher in subjects who were younger, male, lower SES, and those with chronic conditions.

Algorithmic bias: Using the testing cohort, Table 5 summarizes the results of algorithmic bias for both NB and GBM models for predicting 1-year AE risk. Overall, algorithmic performance depended on patient characteristics such as age, sex, and chronic diseases as expected. Also, the degree of differential algorithmic performance by these factors depends on ML algorithm used (NB vs GBM model) in a way that no specific ML algorithm was more or less susceptible to algorithmic bias.

Table 5. Assessment of algorithmic bias for two machine learning algorithms (Naïve Bayes [NB] and gradient boosting machine [GBM] models) predicting 1-year asthma exacerbation risk in childhood asthma using 5 commonly used bias metrics.

	Accuracy equality		Equal opportunity (sensitivity)		Predictive parity (PPV)		Predictive equality (FPR)		Error rate (FPR+FNR)	
Groups	NB model	GBM model	NB model	GBM model	NB model	GBM model	NB model	GBM model	NB model	GBM model
SES (HOUES)										
Q1 (lowest SES)	0.47	0.47	0.38	0.50	0.50	0.50	0.43	0.57	1.05	1.07
Q2-Q4	0.62	0.50	0.59	0.76	0.43	0.36	0.37	0.62	0.78	0.86
Ratio (Q1/Q2-4) (1=no diff)	0.76	0.93	0.64	0.66	1.18	1.39	1.18	0.92	1.35	1.25
Age										
<12	0.53	0.45	0.57	0.70	0.41	0.38	0.50	0.70	0.93	1.0
≥12	0.76	0.64	0.40	0.80	0.67	0.44	0.09	0.44	0.69	0.64
Ratio (<12/≥12) (1=no diff)	0.69	0.71	1.42	0.88	0.61	0.84	5.75	1.61	1.36	1.57
Sex										
Male	0.49	0.45	0.48	0.78	0.33	0.36	0.50	0.73	1.02	0.94
Female	0.74	0.59	0.59	0.65	0.67	0.46	0.17	0.45	0.58	0.80
Ratio (male/female) (1=no diff)	0.67	0.76	0.81	1.21	0.5	0.79	2.9	1.62	1.75	1.18
Race/Ethnicity										
Others	0.54	0.39	0.47	0.60	0.35	0.29	0.42	0.55	0.95	0.95
Non-Hispanic White	0.63	0.58	0.56	0.80	0.50	0.47	0.33	0.71	0.77	0.91
Ratio (others/white) (1=no diff)	0.87	0.67	0.83	0.75	0.7	0.62	1.26	1.30	1.23	1.04
Chronic condition										
At least one	0.53	0.47	0.20	0.80	0.20	0.33	0.33	0.67	1.13	0.87
None	0.61	0.50	0.59	0.69	0.46	0.39	0.38	0.66	0.79	0.98
Ratio (≥1/none) (1=no diff)	0.87	0.94	0.34	1.16	0.43	0.86	0.88	1.11	1.44	0.89

SES as measured by HOUSES index *greatly* impacted algorithmic performance.

Specifically, children in lower SES groups had higher error rates than those in the higher SES group in both ML models (ratio = 1.35 for NB model and 1.25 for GBM model) which exceed those for race/ethnicity (1.23 and 1.04, respectively). This differential algorithmic bias by SES was driven more by FNR (=1-sensitivity; ratio=1.51 by NB and 2.01 by GBM model) than FPR

(1.18 by NB and 0.92 by GBM model). This was also true for equal opportunity (i.e., sensitivity) metric. Children in the higher SES group had significantly higher sensitivity of both algorithms, compared to those in the lower SES group in a way exceeding the impact of other demographic factors.

Table 6: Summary of data availability for variables relevant to asthma management and data validity by SES for each cohort (training and testing cohort)				
	Training		Testing	
Data unavailability, n (%)	Q1 (n=22)	Q2-Q4 (n=102)	Q1 (n=15)	Q2-Q4 (92)
Missing health maintenance visit	3 (14%)	12 (12%)	2 (13%)	11 (12%)
Missing asthma compliance	14 (64%)	43 (42%)	9 (60%)	44 (48%)
Missing asthma severity	9 (41%)	24 (24%)	8 (53%)	23 (25%)
Missing asthma type	22 (100%)	95 (93%)	14 (93%)	79 (86%)
NAEPP recommendation missing	13 (59%)	43 (42%)	8 (53%)	37 (40%)
Missing smoking status	16 (73%)	39 (38%)	8 (53%)	34 (37%)
Missing data on missing school	15 (68%)	41 (40%)	8 (53%)	34 (37%)
	Training		Testing	
Data validity*	Q1 (n=34)	Q2-Q4 (n=112)	Q1 (n=37)	Q2-Q4 (n=121)
Undiagnosed (ICD) asthma	4 (12%)	11 (9.8%)	3 (8.1%)	8 (6.6%)
* Data validity was calculated for subjects who met PAC criteria but did not have physician diagnosis of asthma				

Availability and accuracy of data relevant to asthma management: We compared data availability for the key variables associated with the risk of AE in the training and testing cohorts. As shown in Table 6, compared to children in the higher SES group, those from lower SES background had higher unavailability of the key variables for asthma (eg, compliance data, severity and smoking exposure) associated with the risk of AE. Additionally, children with lower SES had higher prevalence of undiagnosed asthma, compared to those with higher SES, although they met the criteria for asthma.

Discussion

Our study results suggest that SES as measured by HOUSES may impact algorithmic bias in a way resulting in greater algorithmic bias in patients with lower SES, compared to those with

higher SES. Additionally, children with asthma from lower SES groups had a greater degree of unavailable and inaccurate EHRs data for asthma management, compared to those with higher SES. One noteworthy finding is disparities in undiagnosed or delayed diagnosed asthma by SES as the lack of timely diagnosis of asthma will deter access to preventive and therapeutic interventions^{63,64} and may impact long-term respiratory outcomes. We postulate SES might impact algorithmic bias through differential completeness of EHRs data as SES impacts health risk, healthcare access and completeness of EHRs.

As discussed earlier, SES is a key variable for understanding the nature of algorithmic bias stemming from differential health risk, health care access, and completeness of available EHRs and for assessing and mitigating algorithmic bias in health care. However, objective, scalable and well-validated individual-level SES measures are frequently unavailable in commonly used data sources for clinical care and research²⁶ posing a major barrier to health care delivery and research as acknowledged by National Academy of Medicine and National Quality Forum.^{8,27,28} In this respect, HOUSES measuring individual-level SES, can be a useful tool for health care research including AI research as it overcomes such unavailability of individual-level SES measures in commonly used data sources such as EHRs.

Our previous work demonstrated that SES defined by HOUSES index predicted a broad range of health outcomes and care quality as summarized in Table 2. Relevant to this present report, we showed that HOUSES was associated with inconsistent self-reporting.⁵⁵ We found that lower HOUSES (SES) was associated with higher rates of inconsistency (inaccuracy) in self-reporting a diagnosed disease for the given (documented) diseases between the baseline and 4-year follow-up survey, and the association remained significant after pertinent characteristics such as age and perceived general health (adjusted OR=1.46; 95% CI 1.17 to 1.84 for the lowest

compared with the highest HOUSES decile). Given that self-reported information is captured in EHR and often used clinically (e.g., a history of pediatric asthma), higher proportion of inconsistent self-reporting among patients with low SES may produce less reliable ML algorithms (if used). In addition, our unpublished data showed that availability of patient's online portal system (a proxy for healthcare access) was significantly lower among families with lower SES (68% in Q1 [lowest SES]), compared to 74% in Q2, 88% in Q3, and 92% in Q4 (highest SES) ($p=.02$). As online portal is an important tool for chronic disease management such as childhood asthma (eg, communicating with care providers, patient reported outcomes [PROs], medication update, etc, being captured in EHRs), it significantly affected availability of a key PROs on asthma (i.e., Asthma Control Test results; 99% for those with portal vs. 77% for those without portal) at the end of a clinical trial. Indeed, our study results in Table 5 showed the potential impact of SES as measured by HOUSES on algorithmic bias. For example, error rates were higher for children with lower SES for both algorithms predicting the risk of AE, compared to those with higher SES and exceed the impact of other demographic factors (age, sex and race/ethnicity). This was also true for sensitivity. A recent study also showed SES defined by health insurance (public vs. commercial health insurance) influenced ML algorithms predicting ICU mortality¹¹ and 30-day psychiatric readmission (people with lower SES had poorer prediction performance of their ML algorithms, compared to those with higher SES).¹¹ Overall, our study results and the literature suggest that SES impacts differential (in)completeness and validity of PROs which may influence differential algorithmic performance by SES if algorithms are trained by using skewed training cohort by individual-level SES.

It is also important to recognize differential performance of SES measures in predicting health outcomes because researchers routinely use aggregate-level SES measures^{18,65-67} or other

SES measures in research. Our recent study showed that HOUSES predicted that kidney transplant recipients with lowest HOUSES (Q1) had a significantly higher risk of graft failure than those with highest HOUSES (Q2-4) (adjusted hazard ratio: 2.12; 95% CI, 1.08-4.16).⁴³ Importantly, other SES measures such as individual educational levels and census-block group level education and income failed to predict outcomes on graft failure. Therefore, in assessing and mitigating algorithmic bias by SES, it is important to use an SES measure with accuracy and precision in capturing individual-level SES measure and in this respect, HOUSES fulfills these requirements and can be a complementary SES measure beyond the existing conventional SES measures. HOUSES has several conceptual and methodological merits for clinical and translational research: First, HOUSES is able to capture health effects of SES (defined as ‘one’s ability to access desired resources’)⁶⁸ which predicts 39 health care access, care quality, and health outcomes as summarized in Table 2. Second, it assesses an objective individual-level SES measure, in contrast to self-reported (e.g., income) or aggregate-level (e.g., zip-code based Census data) measures. Third, it can retrospectively measure SES at any given point in time whenever address information at the index date of events is available (not relying on recalls). Fourth, as spatial coordinates are intrinsic to HOUSES, it enables geospatial analysis to identify geographic hotspots of interest (e.g., COVID-19 cases) to be used as a feature in predictive models.⁶⁹⁻⁷¹ Finally, unlike other SES measures (e.g., educational level which is relatively static), it can capture longitudinal changes as real property data are regularly updated, and relocation of residence often reflects changes in a subject’s SES. This feature allows us to use HOUSES as a financial outcome across life stages. Taken together, these features highlight how the HOUSES

measure can help to address issues of algorithmic fairness, ultimately helping to achieve greater levels of health equity across populations.

Our study has a few strengths. First, our study is based on a real world setting where patients have a wide range of EHR completeness, instead of studies based on highly selected subjects. Second, we used an objective individual-level SES measure instead of self-reported or aggregate-level SES measures (e.g., Census level data). Therefore, it does not suffer from biases such as recall bias or inaccuracy due to aggregation. Third, we assessed data availability and validity for features relevant to AE risk, which is not commonly done in AI research despite its importance. Our study also has limitations. First, the analysis was based on a small sample size. However, this pilot study provides a conceptual framework for using SES when assessing AI bias. Second, our study subjects may not represent the general pediatric population. However, it represents patient population (source population) as this study was based on those who receive care at Mayo Clinic without involving any recruitment steps.

In conclusion, our study findings highlight the important role of SES in assessing algorithmic bias. Understanding the extent to which SES impacts contributes to algorithmic bias, and examining the potential impact of SES on emerging applications of AI in healthcare will be crucially important for recognizing and mitigating algorithmic bias, ultimately supporting efforts to promote health equity and fairness. We believe the HOUSES index can play an important role in those efforts.

Acknowledgement

We would like to acknowledge the HOUSES program of the Mayo Clinic and Precision Population Science Lab staff, as well as thank Ms. Kelly Okeson for her administrative assistance.

References

1. (ONC) TootNCfHIT. Health IT Dashboard. <https://dashboard.healthit.gov/apps/health-information-technology-data-summaries.php?state=National&cat9=all+data>. Published 2021. Accessed January 17, 2021.
2. Partners SG. *The State of Healthcare Automation: urgent need, growing awareness and tremendous potential*. Sage Growth Partners;2021.
3. Radiology DSIACo. FDA Cleared AI Algorithms. <https://www.acrdsi.org/DSI-Services/FDA-cleared-ai-algorithms>. Accessed.
4. Benjamins S, Dhunoo P, Meskó B. The state of artificial intelligence-based FDA-approved medical devices and algorithms: an online database. *npj Digital Medicine*. 2020;3(1):118.
5. Yao X, Rushlow DR, Inselman JW, et al. Artificial intelligence-enabled electrocardiograms for identification of patients with low ejection fraction: a pragmatic, randomized clinical trial. *Nature Medicine*. 2021;27(5):815-819.
6. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447.
7. Institute of Medicine. *Unequal Treatment: Confronting racial and ethnic disparities in health care*. Washington DC: The National Academy of Science;2003.
8. National Academy of Medicine. *Accounting for Social Risk Factors in Medicare Payment: Health and Medicine Division*. 2017.
9. Dzau VJ, McClellan MB, McGinnis J, et al. Vital directions for health and health care: Priorities from a national academy of medicine initiative. *JAMA*. 2017.
10. Park Y, Hu J, Singh M, et al. Comparison of Methods to Reduce Bias From Clinical Prediction Models of Postpartum Depression. *JAMA Network Open*. 2021;4(4):e213909-e213909.
11. Irene Y. Chen PS, PhD, and Marzyeh Ghassemi, PhD. Can AI Help Reduce Disparities in General Medical and Mental Health Care? *AMA Journal of Ethics*. 2019;21(2):E167-179.
12. Hoskins KF, Danciu OC, Ko NY, Calip GS. Association of Race/Ethnicity and the 21-Gene Recurrence Score With Breast Cancer-Specific Mortality Among US Women. *JAMA Oncology*. 2021.
13. Bach PB, Pham HH, Schrag D, Tate RC, Hargraves JL. Primary Care Physicians Who Treat Blacks and Whites. *New England Journal of Medicine*. 2004;351(6):575-584.
14. Warnecke RB, Oh A, Breen N, et al. Approaching Health Disparities From a Population Perspective: The National Institutes of Health Centers for Population Health and Health Disparities. *Am J Public Health*. 2008;98(9):1608-1615.
15. Adler NE, Newman K. Socioeconomic disparities in health: pathways and policies. *Health affairs*. 2002;21(2):60-76.

16. Bernheim SM, Ross JS, Krumholz HM, Bradley EH. Influence of Patients' Socioeconomic Status on Clinical Management Decisions: A Qualitative Study. *The Annals of Family Medicine*. 2008;6(1):53-59.
17. Franks P, Fiscella K. Effect of patient socioeconomic status on physician profiles for prevention, disease management, and diagnostic testing costs. *Med Care*. 2002;40(8):717-724.
18. Sills MR, Hall M, Colvin JD, et al. Association of social determinants with children's hospitals' preventable readmissions performance. *JAMA Pediatrics*. 2016;170(4):350-358.
19. Roberts ET, Zaslavsky AM, Barnett ML, Landon BE, Ding L, McWilliams J. Assessment of the effect of adjustment for patient characteristics on hospital readmission rates: Implications for pay for performance. *JAMA Internal Medicine*. 2018;178(11):1498-1507.
20. Baker DW, Chassin MR. Holding providers accountable for health care outcomes. *Annals of internal medicine*. 2017;167(6):418-423.
21. Jha AK, Zaslavsky AM. Quality reporting that addresses disparities in health care. *JAMA*. 2014;312(3):225-226.
22. Snyder-Mackler N, Burger JR, Gaydos L, et al. Social determinants of health and survival in humans and other animals. *Science*. 2020;368(6493):eaax9553.
23. Belsky DW, Snyder-Mackler N. Invited Commentary: Integrating Genomics and Social Epidemiology—Analysis of Late-Life Low Socioeconomic Status and the Conserved Transcriptional Response to Adversity. *American Journal of Epidemiology*. 2017;186(5):510-513.
24. Martens DS, Janssen BG, Bijlens EM, et al. Association of Parental Socioeconomic Status and Newborn Telomere Length. *JAMA Network Open*. 2020;3(5):e204057.
25. Warnecke RB, Oh A, Breen N, et al. Approaching health disparities from a population perspective: the National Institutes of Health Centers for Population Health and Health Disparities. *Am J Public Health*. 2008;98(9):1608-1615.
26. Juhn YJ, Beebe TJ, Finnie DM, et al. Development and initial testing of a new socioeconomic status measure based on housing data. *Journal of urban health : bulletin of the New York Academy of Medicine*. 2011;88(5):933-944.
27. National Quality Forum Technical Report. *Risk Adjustment for Socioeconomic Status or Other Sociodemographic Factors*. 2014. The report is funded by the DHHS under contract HHSM-500-2012-000091 task order 7.
28. National Quality Forum. *Evaluation of NQF's Trial Period for Risk Adjustment for Social Risk Factors* June 8, 2017 2017.
29. Administration UFaD. EXECUTIVE SUMMARY FOR THE PATIENT ENGAGEMENT ADVISORY COMMITTEE MEETING:Artificial Intelligence (AI) and Machine Learning (ML) in Medical Devices. In: Services DoHaH, ed: FDA; 2020.
30. Seol HY, Shrestha P, Sohn S, et al. Artificial Intelligence-Assisted Clinical Decision Support for Childhood Asthma Management: A Randomized Clinical Trial. *PloS one*. 2021;doi.org/10.1371/journal.pone.0255261.
31. St Sauver JL, Grossardt BR, Yawn BP, et al. Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. *Int J Epidemiol*. 2012;41(6):1614-1624.
32. St. Sauver JL, Grossardt BR, Leibson CL, Yawn BP, Melton Iii LJ, Rocca WA. Generalizability of Epidemiological Findings and Public Health Decisions: An Illustration From the Rochester Epidemiology Project. *Mayo Clinic Proceedings*. 2012;87(2):151-160.
33. Zhong W, Finnie DM, Shah ND, et al. Effect of Multiple Chronic Diseases on Health Care Expenditures in Childhood. *J Prim Care Community Health*. 2015;6(1):2-9.
34. Yawn BP, Wollan P, Kurland M, Scanlon P. A longitudinal study of the prevalence of asthma in a community population of school-age children. *Journal of Pediatrics*. 2002;140(5):576-581.

35. Narayanan A. Translation tutorial: 21 fairness definitions and their politics. Paper presented at: Proc. Conf. Fairness Accountability Transp., New York, USA2018.
36. Chouldechova A. Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big data*. 2017;5(2):153-163.
37. Bellamy RKE, Dey K, Hind M, et al. AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*. 2019;63(4/5):4:1-4:15.
38. Ghawi H, Crowson CS, Rand-Weaver J, Krusemark E, Gabriel SE, Juhn YJ. A novel measure of socioeconomic status using individual housing data to assess the association of SES with rheumatoid arthritis and its mortality: a population-based case-control study. *BMJ Open*. 2015;5(4):e006469.
39. Wi CI, St Sauver JL, Jacobson DJ, et al. Ethnicity, Socioeconomic Status, and Health Disparities in a Mixed Rural-Urban US Community-Olmsted County, Minnesota. *Mayo Clin Proc*. 2016;91(5):612-622.
40. Bang DW, Manemann SM, Gerber Y, et al. A novel socioeconomic measure using individual housing data in cardiovascular outcome research. *Int J Environ Res Public Health*. 2014;11(11):11597-11615.
41. Takahashi PY, Ryu E, Hathcock MA, et al. A novel housing-based socioeconomic measure predicts hospitalisation and multiple chronic conditions in a community population. *J Epidemiol Community Health*. 2016;70(3):286-291.
42. Ryan CS, Juhn YJ, Kaur H, et al. Long-term incidence of glioma in Olmsted County, Minnesota, and disparities in postglioma survival rate: a population-based study. *Neurooncol Pract*. 2020;7(3):288-298.
43. Stevens MA, Beebe TJ, Wi C-I, Taler SJ, St. Sauver JL, Juhn YJ. HOUSES index as an innovative socioeconomic measure predicts graft failure among kidney transplant recipients. *Transplantation*. 2020;Online First.
44. Barwise A, Wi CI, Frank R, et al. An Innovative Individual-Level Socioeconomic Measure Predicts Critical Care Outcomes in Older Adults: A Population-Based Study. *Journal of intensive care medicine*. 2020;885066620931020.
45. Angstman KB, Wi CI, Williams MD, Bohn BA, Garrison GM. Impact of socioeconomic status on depression clinical outcomes at six months in a Midwestern, United States community. *J Affect Disord*. 2021;292:751-756.
46. Bjur KA, Wi CI, Ryu E, et al. Socioeconomic Status, Race/Ethnicity, and Health Disparities in Children and Adolescents in a Mixed Rural-Urban Community-Olmsted County, Minnesota. *Mayo Clin Proc*. 2019;94(1):44-53.
47. Ryu E, Wi CI, Crow SS, et al. Assessing health disparities in children using a modified housing-related socioeconomic status measure: a cross-sectional study. *BMJ Open*. 2016;6(7):e011564.
48. Harris MN, Lundien MC, Finnie DM, et al. Application of a novel socioeconomic measure using individual housing data in asthma research: an exploratory study. *NPJ primary care respiratory medicine*. 2014;24:14018.
49. Bjur KA, Wi CI, Ryu E, Crow SS, King KS, Juhn YJ. Epidemiology of Children With Multiple Complex Chronic Conditions in a Mixed Urban-Rural US Community. *Hospital pediatrics*. 2019;9(4):281-290.
50. Thacher TD, Dudenkov DV, Mara KC, Maxson JA, Wi CI, Juhn YJ. The relationship of 25-hydroxyvitamin D concentrations and individual-level socioeconomic status. *The Journal of steroid biochemistry and molecular biology*. 2020;197:105545.
51. Ryu E, Juhn YJ, Wheeler PH, et al. Individual housing-based socioeconomic status predicts risk of accidental falls among adults. *Ann Epidemiol*. 2017;27(7):415-420.e412.

52. Aul AJ, Dudenkov DV, Mara KC, et al. The relationship of 25-hydroxyvitamin D values and risk of fracture: a population-based retrospective cohort study. *Osteoporos Int.* 2020;31(9):1787-1799.
53. Johnson MD, Urm SH, Jung JA, et al. Housing data-based socioeconomic index and risk of invasive pneumococcal disease: an exploratory study. *Epidemiology and infection.* 2013;141(4):880-887.
54. Wi CI, Gauger J, Bachman M, et al. Role of individual-housing-based socioeconomic status measure in relation to smoking status among late adolescents with asthma. *Ann Epidemiol.* 2016;26(7):455-460.
55. Ryu E, Olson JE, Juhn YJ, et al. Association between an individual housing-based socioeconomic index and inconsistent self-reporting of health conditions: a prospective cohort study in the Mayo Clinic Biobank. *BMJ Open.* 2018;8(5):e020054.
56. Barwise A, Juhn YJ, Wi CI, et al. An Individual Housing-Based Socioeconomic Status Measure Predicts Advance Care Planning and Nursing Home Utilization. *Am J Hosp Palliat Care.* 2018.
57. Hammer R, Capili C, Wi C-I, Ryu E, Rand-Weaver J, Juhn YJ. A new socioeconomic status measure for vaccine research in children using individual housing data: a population-based case-control study. *BMC Public Health.* 2016;16(1):1-9.
58. Butterfield MC, Williams AR, Beebe T, et al. A two-county comparison of the HOUSES index on predicting self-rated health. *Journal of epidemiology and community health.* 2011;65(3):254-259.
59. MacLaughlin KL, Jacobson RM, Sauver JLS, et al. An innovative housing-related measure for individual socioeconomic status and human papillomavirus vaccination coverage: A population-based cross-sectional study. *Vaccine.* 2020;38(39):6112-6119.
60. Wu ST, Sohn S, Ravikumar KE, et al. Automated chart review for asthma cohort identification using natural language processing: an exploratory study. *Annals of allergy, asthma & immunology : official publication of the American College of Allergy, Asthma, & Immunology.* 2013;111(5):364-369.
61. Wi CI, Sohn S, Rolfes MC, et al. Application of a Natural Language Processing Algorithm to Asthma Ascertainment. An Automated Chart Review. *American journal of respiratory and critical care medicine.* 2017;196(4):430-437.
62. Wi CI, Sohn S, Ali M, et al. Natural Language Processing for Asthma Ascertainment in Different Practice Settings. *The journal of allergy and clinical immunology In practice.* 2018;6(1):126-131.
63. Bisgaard H, Szeffler S. Prevalence of asthma-like symptoms in young children. *Pediatr Pulmonol.* 2007;42(8):723-728.
64. Bloom CI, Franklin C, Bush A, Saglani S, Quint JK. Burden of preschool wheeze and progression to asthma in the UK: Population-based cohort 2007 to 2017. *Journal of Allergy and Clinical Immunology.* 2021;147(5):1949-1958.
65. Ash AS, Mick EO, Ellis RP, Kiefe CI, Allison JJ, Clark MA. Social determinants of health in managed care payment formulas. *JAMA Internal Medicine.* 2017;177(10):1424-1430.
66. Knighton AJ, Savitz L, Belnap T, Stephenson B, VanDerslice J. Introduction of an Area Deprivation Index Measuring Patient Socioeconomic Status in an Integrated Health System: Implications for Population Health. *eGEMs.* 2016;4(3):1238.
67. Chien AT, Wroblewski K, Damberg C, et al. Do physician organizations located in lower socioeconomic status areas score lower on pay-for-performance measures? *Journal of general internal medicine.* 2012;27(5):548-554.
68. Oakes JM RP. The measurement of SES in health research: current practice and steps toward a new approach. *Social Science & Medicine.* 2003;56:769-784.
69. Juhn YJ, Wheeler P, Wi CI, et al. Role of Geographic Risk Factors in COVID-19 Epidemiology: Longitudinal Geospatial Analysis. *Mayo Clin Proc Innov Qual Outcomes.* 2021.

70. Wi C-I, Wheeler PH, Kaur H, Ryu E, Kim D, Juhn Y. Spatio-temporal comparison of pertussis outbreaks in Olmsted County, Minnesota, 2004–2005 and 2012: a population-based study. *BMJ Open*. 2019;9(5):e025521.
71. Patel AA, Wheeler PH, Wi C-I, et al. Mobile home residence as a risk factor for adverse events among children in a mixed rural–urban community: A case for geospatial analysis. *Journal of Clinical and Translational Science*. 2020;4(5):443-450.