



Social Determinants of Health Data Availability for Patients with Eye Conditions

Terrence C. Lee, BS,^{1,2} Bharanidharan Radha Saseendrakumar, MS,^{1,2} Mahasweta Nayak,² Alison X. Chan, BS,^{1,2} John J. McDermott IV, BA,^{1,2} Bita Shahrvini, BS,^{1,2} Gordon Y. Ye,^{1,2} Amy M. Sitapati, MD,^{2,3} Camille Nebeker, EdD, MS,⁴ Sally L. Baxter, MD, MSc^{1,2}

Purpose: To quantify and characterize social determinants of health (SDoH) data coverage using single-center electronic health records (EHRs) and the National Institutes of Health All of Us research program.

Design: Retrospective cohort study from June 2014 through June 2021.

Participants: Adults 18 years of age or older with a diagnosis of diabetic retinopathy, glaucoma, cataracts, or age-related macular degeneration.

Methods: For All of Us, research participants completed online survey forms as part of a nationwide prospective cohort study. In local EHRs, patients were selected based on diagnosis codes.

Main Outcome Measures: Social determinants of health data coverage, characterized by the proportion of each disease cohort with available data regarding demographics and socioeconomic factors.

Results: In All of Us, we identified 23 806 unique adult patients, of whom 2246 had a diagnosis of diabetic retinopathy, 13 448 had a diagnosis of glaucoma, 6634 had a diagnosis of cataracts, and 1478 had a diagnosis of age-related macular degeneration. Survey completion rates were high (99.5%–100%) across all cohorts for demographic information, overall health, income, education, and lifestyle. However, health care access (12.7%–29.4%), housing (0.7%–1.1%), social isolation (0.2%–0.3%), and food security (0–0.1%) showed significantly lower response rates. In local EHRs, we identified 80 548 adult patients, of whom 6616 had a diagnosis of diabetic retinopathy, 26 793 had a diagnosis of glaucoma, 40 427 had a diagnosis of cataracts, and 6712 had a diagnosis of age-related macular degeneration. High data coverage was found across all cohorts for variables related to tobacco use (82.84%–89.07%), alcohol use (77.45%–83.66%), and intravenous drug use (84.76%–93.14%). However, low data coverage (< 50% completion) was found for all other variables, including education, finances, social isolation, stress, physical activity, food insecurity, and transportation. We used chi-square testing to assess whether the data coverage varied across different disease cohorts and found that all fields varied significantly ($P < 0.001$).

Conclusions: The limited and highly variable data coverage in both local EHRs and All of Us highlights the need for researchers and providers to develop SDoH data collection strategies and to assemble complete datasets. *Ophthalmology Science* 2022;2:100151 © 2022 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



Supplemental material is available at www.ophthalmologyscience.org

At the beginning of the 21st century, the Institute of Medicine (IOM) identified gaps in health care quality in the United States and called for systemic changes to eliminate socioeconomic and racial or ethnic disparities.^{1,2} In response, the Centers for Medicare and Medicaid Services (CMS) 2016 Quality Strategy emphasized the need for care providers to identify and address social determinants of health (SDoH), broadly defined as conditions in which people live, work, and grow, including social, political, economic, and environmental factors.³ Specifically, the IOM recommended collection of 11 SDoH domains: race or ethnic group, education, financial resource strain, stress, depression, physical activity, tobacco use, alcohol use, social isolation,

intimate partner violence, and residential address.⁴ As of January 7, 2021, CMS also issued new guidance to shift health care toward value-based models to address these health disparities.⁵ With recent major developments in health information technology, substantial research has been carried out regarding how health data could be used to bridge health care disparities. Gold et al⁶ highlighted the benefits of standardized SDoH data collection and presentation using electronic health record (EHR) tools in improving patient and population health outcomes in various care settings. However, other studies such as that by Zhang et al⁵ also reported challenges such as training requirements and disproportionate data access that may worsen disparities in clinical informatics.

In ophthalmology, interest in SDoH is growing. Recent studies have shown that several ophthalmic conditions are associated with various socioeconomic or racial and ethnic factors. Congdon⁷ prepared prevalence estimates stratified by race and found that the leading cause of blindness differed by race. Similarly, studies in the United Kingdom and Australia reported a higher prevalence of visual impairment in Blacks and South Asians with diabetes and a higher rate of diabetic retinopathy in Indigenous Australian, Pacific Islander, and Indian populations.^{8,9} Healthcare disparities may also contribute to worsened postoperative care in patients who receive cataract surgery.¹⁰ The relationship between SDoH and ophthalmic conditions also is bidirectional; Constantino et al⁸ and Brezin et al¹¹ reported that low visual acuity could in fact be a risk factor for socioeconomic factors such as adverse social outcomes, worsened mental health, and poverty.¹²

Despite the growing importance of SDoH in ophthalmology, no studies have reported on the documentation of SDoH data in EHR for patients with eye conditions. One concern is that despite available SDoH data fields in local EHRs, compliance is low in documenting that data successfully. Cottrell et al¹³ found that most SDoH screening of EHRs at community health centers included responses for only 1 of the 11 domains recommended by the IOM. Based on an extensive literature search of databases (e.g., PubMed, Google Scholar, Web of Science), this is a novel investigation that specifically examines data coverage for SDoH for patients with ophthalmic conditions. In this study, we used both local EHR data and nationwide data from the National Institutes of Health All of Us research program. Established in 2015, All of Us aims to build a nationally representative database with an emphasis on enrolling diverse and underrepresented research participants.^{14,15} All of Us has enrolled more than 440 000 adults as of December 2021, with a projected goal of 1 million participants.^{16,17} All of Us collects data such as health questionnaire responses, individual EHR, and physical measurements and allows research teams to have access via their Researcher Workbench. By leveraging both local and nationwide data repositories, this study reports on the availability of SDoH data in patients with eye conditions to inform new strategies for care providers and health systems to improve and standardize SDoH data collection.

Methods

Eligibility Criteria

This study identified patients with the 4 leading causes of blindness and vision impairment as defined by the United States Centers for Disease Control and Prevention: diabetic retinopathy (DR), glaucoma, cataracts, or age-related macular degeneration (AMD).¹⁸ Inclusion criteria were all adults (18 years of age or older) with discrete International Classification of Disease or Systemized Nomenclature of Medicine—Clinical Terms diagnosis codes for DR, glaucoma, cataracts, or AMD during the study period from June 2014 through June 2021 to reflect the IOM's statement in 2014 on capturing social and behavioral domains in EHR. Both survey data and EHR data were extracted from the nationwide National Institutes of Health All of Us data repository, and EHR

data were extracted from the University of California, San Diego (UCSD) Health System. The objective was to understand SDoH data coverage in a large nationwide research-oriented database as well as in a local or institutional EHR system (Epic) used for routine clinical practice. This study adhered to the tenets of the Declaration of Helsinki and was approved by the UCSD Institutional Review Board.

Cohort Building and Data Extraction in All of Us

The recruitment methods and scientific rationale for All of Us have been described previously.¹⁹ The data extraction was performed on EHR domains and survey results that were available via the All of Us Researcher Workbench, a cloud-based platform that enables researchers to cluster participants into cohorts, select certain health information within each cohort, and perform direct analysis and query using R (R Foundation for Statistical Computing) and Python 3.0 (Python Software Foundation) programming languages within Jupyter Notebooks. The EHR data derived from captured data including billing codes and encounter records were used to cluster participants into disease cohorts based on Systemized Nomenclature of Medicine—Clinical Terms diagnosis codes (the standardized vocabulary in All of Us sourced from corresponding International Classification of Diseases codes), whereas SDoH data were derived from survey responses (Supplemental Table A). Examples of the surveys can be found through the publicly available Data Browser.²⁰ Some surveys (e.g., basic demographics, overall health, income, education, and lifestyle) were mandatory, whereas others (e.g., social isolation, housing, health care access, and food security) were optional. Both survey data and EHR data are mapped to the Observational Health and Medicines Outcomes Partnership common data model version 5.2. Individuals participating in All of Us provided written informed consent, and study procedures were approved by the All of Us Institutional Review Board.

To evaluate SDoH data coverage in All of Us, we identified demographic and social variables that were included in the IOM's recommended domains in addition to other variables with high clinical value.²¹ These variables include basic demographics, social isolation, overall health (e.g., health literacy, quality of life, activities of daily living, and mental health), income, housing, health care access, food security, education, and lifestyle (e.g., tobacco use, alcohol use, or other drug use) that were derived from survey data. Because of All of Us data sharing policies, which prohibit displaying counts of fewer than 20, proportions and percentages were not specified when the numerator was less than 20.

Cohort Building and Data Extraction in Local Electronic Health Records

In the UCSD EHR database, an EHR-based data exploration tool (Epic SlicerDicer; Epic) was used to identify patients with DR, glaucoma, cataracts, or AMD based on a search query using patients' medical information found in the medical history, chief complaint, problem list, and encounter diagnosis fields (Supplemental Table B). Discrete patient medical record numbers then were paired with local EHR data elements to derive SDoH variables. Similar to All of Us data extraction, variables were selected to include the IOM's recommended domains and other variables with high clinical value. These variables include demographics, tobacco use, cigarette use, alcohol use, other drug use, education, financial strain, living situation, daily stress, socialization, physical activity, food insecurity, and transportation. The cohorts and extracted variables then were compiled in an external dataset and exported to an R notebook for subsequent analyses.

Synthesis and Analysis of Results

For both All of Us and UCSD EHR data, descriptive statistics of each disease cohort were generated for age, gender, and race. Data coverage was defined as the proportion of patients within the cohort with any data available for that specific variable. The SDoH data coverage also was compared across disease cohorts as well as across data sources. Categorical and continuous variables were compared using Pearson's chi-square test and Student's *t* test, respectively. For the latter, assumptions for parametric hypothesis testing were confirmed. Statistical analysis was performed using R software version 4.0.3 with 2-sided *P* values of less than 0.05 considered statistically significant.

Results

General Cohort Characteristics in All of Us

Of the 298 827 adults whose data were included in All of Us at the time of data extraction in November 2020, we identified 2246 with DR, 13 448 with glaucoma, 6634 with cataracts, and 1478 with AMD (Table 1). The mean ages of patients with DR, glaucoma, cataracts, and AMD were 62.2 years, 67.0 years, 69.9 years, and 73.6 years, respectively. Most patients (57.97%–63.86%) across all cohorts were women. Across disease cohorts, White participants (31.27%–74.25%), Black participants (10.0%–30.32%), and patients who did not indicate a race (10.21%–33.17%) were more represented than Asian participants (2.03%–3.12%). Hispanic or Latino/a patient representation ranged from 10.82% in the AMD cohort to 34.84% in the DR cohort.

Social Determinants of Health Data Coverage in All of Us

In the All of Us database, survey completion rates were high (99.5%–100%) across all cohorts for demographic information, overall health, income, education, and lifestyle (Table 2). However, health care access (12.7%–29.4%),

housing (< 1.35%), social isolation (< 1.35%), and food security (< 1.35%) showed significantly lower response rates. The variability between highly completed fields across disease cohorts was minimal and not statistically significant. Of the variables with lower response rates, availability of data regarding health care access varied significantly (*P* < 0.001) across different disease cohorts, with the lowest in patients with DR at 12.7% compared with patients with AMD at 29.4%.

General Cohort Characteristics in Local Electronic Health Records

In the UCSD EHR database, we identified 6616 adults with DR, 26 793 adults with glaucoma, 40 427 adults with cataracts, and 6712 adults with AMD (Table 3). The mean age of patients with DR, glaucoma, cataracts, and AMD were 64.8 years, 69.8 years, 71.4 years, and 80.5 years, respectively. Most patients (54.69%–59.51%) in the DR and AMD cohort were men, whereas most patients (55.84%–56.03%) in the glaucoma and cataracts cohort were women. Across disease cohorts, White patients (42.25%–71.34%), other or mixed-race patients (11.57%–32.63%), and Asian patients (8.97%–12.57%) were more represented, whereas Black patients (1.40%–6.70%) were the least represented.

Social Determinants of Health Data Coverage in Local Electronic Health Records

For most patients in the UCSD EHR database, data were available regarding whether they used tobacco (82.84%–89.07%), cigarettes (84.76%–93.14%), alcohol (77.45%–83.66%), or intravenous drugs (84.76%–93.14%; Table 4; Fig 1). However, low data coverage (< 50%) was present for variables related to specific characteristics of tobacco use (packs per day, number of years used) and alcohol use (ounces per week, frequency, and self-characterization as a binge user). Moreover, data coverage

Table 1. General Cohort Characteristics in the All of Us Research Program

	All Adults (n = 298 827)	Diabetic Retinopathy (n = 2246)	Glaucoma (n = 13 448)	Cataract (n = 6634)	Age-Related Macular Degeneration (n = 1478)
Age (yrs)	52.86 ± 16.73	62.18 ± 12.23	66.97 ± 12.18	69.85 ± 9.89	73.59 ± 10.06
Sex					
Male	114 543 (38.33)	906 (40.34)	2391 (36.04)	4643 (34.53)	529 (35.79)
Female	184 284 (61.67)	1302 (57.97)	4132 (62.29)	8587 (63.85)	925 (62.58)
Nonbinary or preferred not to respond		38 (1.69)	111 (1.67)	217 (1.61)	23 (1.56)
Race					
Black	66 302 (22.19)	671 (30.32)	1776 (27.12)	2717 (20.49)	146 (10.0)
White	163 347 (54.66)	692 (31.27)	3293 (50.29)	7888 (59.5)	1084 (74.25)
Asian	10 377 (3.47)	45 (2.03)	204 (3.12)	335 (2.53)	36 (2.47)
Other	7622 (2.55)	71 (3.21)	208 (3.18)	401 (3.02)	45 (3.08)
None indicated	51 179 (17.13)	734 (33.17)	1067 (16.3)	1916 (14.45)	149 (10.21)
Ethnicity					
Not Hispanic or Latino	240 968 (80.64)	1414 (63.9)	5303 (80.99)	11 030 (83.2)	1291 (88.42)
Hispanic or Latino	57 859 (19.36)	771 (34.84)	1164 (17.78)	2074 (15.64)	158 (10.82)

Data are presented as mean ± standard deviation or no. (%).

Table 2. Data Coverage of Variables Related to Social Determinants of Health in the All of Us Database

Survey Name	Diabetic Retinopathy (N = 2246)	Glaucoma (N = 13448)	Cataract (N = 6634)	Age-Related Macular Degeneration (N = 1478)
Basic Demographics	2246 (100)	13 448 (100)	6634 (100)	1478 (100)
Social Isolation	< 20 (< 0.89)	27 (0.2)	< 20 (< 0.30)	< 20 (< 1.35)
Overall Health*	2239 (99.69)	13 411 (99.72)	6619 (99.77)	1474 (99.73)
Income	2246 (100)	13 448 (100)	6632 (99.97)	1478 (100)
Housing	< 20 (< 0.89)	144 (1.07)	73 (1.1)	< 20 (< 1.35)
Health Care Access	286 (12.73)	3396 (25.25)	1565 (23.59)	434 (29.36)
Food Security	< 20 (< 0.89)	< 20 (< 0.15)	< 20 (< 0.30)	< 20 (< 1.35)
Education	2246 (100)	13 448 (100)	6634 (100)	1478 (100)
Lifestyle [†]	2235 (99.51)	13 403 (99.67)	6612 (99.67)	1474 (99.73)

Data are presented as no. (%).

*Includes health literacy, quality of life, activities of daily living, and mental health.

[†]Includes use of tobacco, alcohol, and recreational drugs.

was also low for the following SDoH variables: education, finances, social isolation, stress, physical activity, food insecurity, and transportation (all less than 7%). Data coverage for all variables differed significantly across different disease cohorts ($P < 0.001$).

Discussion

Summary of Key Findings in All of Us

In reference to the most recent United States Census data, the demographic distribution of All of Us reflected national demographics, with increased representation of Black and Hispanic participants,²² demonstrating an encouraging trend toward the program's aims of building a diverse and nationally representative database. The local EHRs included larger disease cohorts and increased

representation of Asian patients, which was more reflective of local San Diego County demographics.²² In both databases, an increased proportion of Hispanic patients was found in the DR cohort and a higher mean age and increased proportion of White patients was found in the AMD cohort, which follows the demographic trends observed in prior studies.^{23,24}

All of Us showed high coverage for survey responses that were required for participants to register for the program. Data coverage for these variables exceeded 99%, indicating that mandatory survey items were an effective measure for All of Us data collection. However, other variables such as housing, social isolation, and food security were populated for less than 2% of patients across all disease cohorts, highlighting the need for SDoH variables to be grouped into mandatory surveys to ensure complete data collection. Also, significant variability was found between disease cohorts for data concerning health care

Table 3. General Cohort Characteristics in the University of California, San Diego, Electronic Health Record Clinical Database

	Diabetic Retinopathy (n = 6616)	Glaucoma (n = 26 793)	Cataract (n = 40 427)	Age-Related Macular Degeneration (n = 6712)
Age (yrs)	64.8 ± 13.89	69.8 ± 16.95	71.4 ± 12.13	80.5 ± 12.57
Sex				
Male	3618 (54.69)	11 831 (44.16)	17 774 (43.96)	3995 (59.51)
Female	2998 (45.31)	14 962 (55.84)	22 653 (56.03)	2717 (40.47)
Race				
Black	443 (6.7)	1639 (6.12)	1606 (3.97)	94 (1.40)
White	2795 (42.25)	14 978 (55.90)	25 011 (61.86)	4789 (71.34)
Asian	760 (11.49)	3368 (12.57)	5014 (12.40)	602 (8.97)
Native Hawaiian or other Pacific Islander	49 (0.74)	81 (0.30)	159 (0.39)	<20 (<0.30)
American Indian or Alaska Native	40 (0.60)	95 (0.35)	133 (0.33)	<20 (<0.30)
Other race or mixed race	2159 (32.63)	4766 (17.79)	6151 (15.21)	777 (11.57)
None indicated	370 (5.59)	1867 (6.97)	2355 (5.83)	428 (6.37)
Ethnicity				
Not Hispanic or Latino	3727 (56.33)	19 830 (74.01)	31 653 (78.29)	5493 (81.83)
Hispanic or Latino	2459 (37.17)	4174 (15.58)	5194 (12.85)	494 (7.36)
Other or unknown	430 (6.50)	2790 (10.41)	3580 (8.86)	725 (10.81)

Data are presented as mean ± standard deviation or no. (%).

Table 4. Data Coverage of Variables Related to Social Determinants of Health in the University of California, San Diego, Electronic Health Record Clinical Database

Variable	Diabetic Retinopathy (n = 6616)	Glaucoma (n = 26 793)	Cataract (n = 40 427)	Age-Related Macular Degeneration (n = 6712)
Tobacco use				
Tobacco user	5889 (89.01)	22 802 (85.10)	36 010 (89.07)	5561 (82.84)
Packs/day	1378 (20.83)	4933 (18.41)	9072 (22.44)	1459 (21.73)
Used years	1459 (22.05)	5086 (18.98)	9459 (23.40)	1566 (23.33)
Cigarettes Y/N	6162 (93.14)	23 428 (87.44)	36 741 (90.88)	5690 (84.76)
Alcohol use				
Alcohol user	5352 (80.89)	21 027 (78.48)	33 821 (83.66)	5199 (77.45)
Oz/wk	1949 (29.46)	8640 (32.25)	15 580 (38.54)	2464 (36.70)
Frequency	1014 (15.33)	4906 (18.31)	8690 (21.49)	1273 (18.96)
Alcoholic binge	137 (2.07)	767 (2.86)	1406 (3.48)	218 (3.25)
Drug use				
IV drug user	6162 (93.14)	23 428 (87.44)	36 741 (90.88)	5690 (84.76)
Education				
No. of yrs	138 (2.09)	533 (1.99)	1000 (2.47)	155 (2.31)
Level	124 (1.87)	367 (1.37)	729 (1.80)	107 (1.59)
Finances				
Financial resource strain	165 (2.49)	427 (1.59)	843 (2.09)	159 (2.37)
Social isolation				
Living with spouse	119 (1.80)	360 (1.34)	654 (1.62)	126 (1.88)
Phone communication	111 (1.68)	345 (1.29)	656 (1.62)	116 (1.73)
Socialization frequency	109 (1.65)	344 (1.28)	659 (1.63)	112 (1.67)
Stress level				
Daily stress	95 (1.44)	269 (1.00)	480 (1.19)	99 (1.47)
Physical activity				
Days/wk	219 (3.31)	1161 (4.33)	2429 (6.01)	454 (6.76)
Min/session	206 (3.11)	1097 (4.09)	2313 (5.72)	431 (6.42)
Food insecurity				
Scarcity or worry	157 (2.37)	413 (1.54)	824 (2.04)	156 (2.32)
Transportation				
Medical or other transportation need	120 (1.81)	343 (1.28)	703 (1.74)	122 (1.82)

IV = intravenous; N = no; Y = yes.
Data are presented as no. (%).

access, which may be explained by varying demographic trends found among disease cohorts. Extensive research in racial or ethnic disparities have found that Black and Hispanic patients have lower odds of ambulatory visits and establishing care with specialists.²⁵ These findings are consistent with the decreased proportion of data available for patients with DR, which showed the highest Black and Hispanic representation among disease cohorts. Prior studies also found that patients with DR underuse health care, possibly because of structural barriers inclusive of issues with insurance, transportation, or poor referral patterns.^{26,27} In contrast, more data from the AMD cohort were available regarding health care access, which is consistent with prior studies that have found increased health care use in aging and White populations.²⁸ Variability in health care access also may be the result of the pathophysiologic features and social consequences of each disease: a metasynthesis of patient experiences with AMD identified themes of functional limitations and frequent interactions with health services.²⁹ Similarly, patients with early-presenting or severe eye conditions may report more SDoH data if the disease causes profound disability or requires regular follow-up.

Summary of Key Findings in Local Electronic Health Records

The local EHRs showed high coverage for certain variables (e.g., tobacco, alcohol, and other drug use) likely driven by federal and state quality metrics such as the Merit-Based Incentive Payment System,³⁰ which indicates the success of these institutional quality measures in motivating data collection. In addition, these lifestyle habits are of high interest for providers to risk-stratify patients. However, the relative scarcity of other SDoH variables may stem from known barriers to data collection faced by both patients and providers.³¹ For example, education in local EHRs showed significantly fewer available data ($P < 0.001$) than in All of Us, demonstrating the effectiveness of requiring certain fields that reflect clinical interest. Patients may not readily offer a comprehensive social or behavioral history if not prompted during clinic visits, and providers may lack awareness and ask only for social or behavioral history if it is deemed clinically relevant.

Also significant variability was found across all SDoH variables between disease cohorts in the local EHRs, which is consistent with a study by Wang et al³² that also revealed

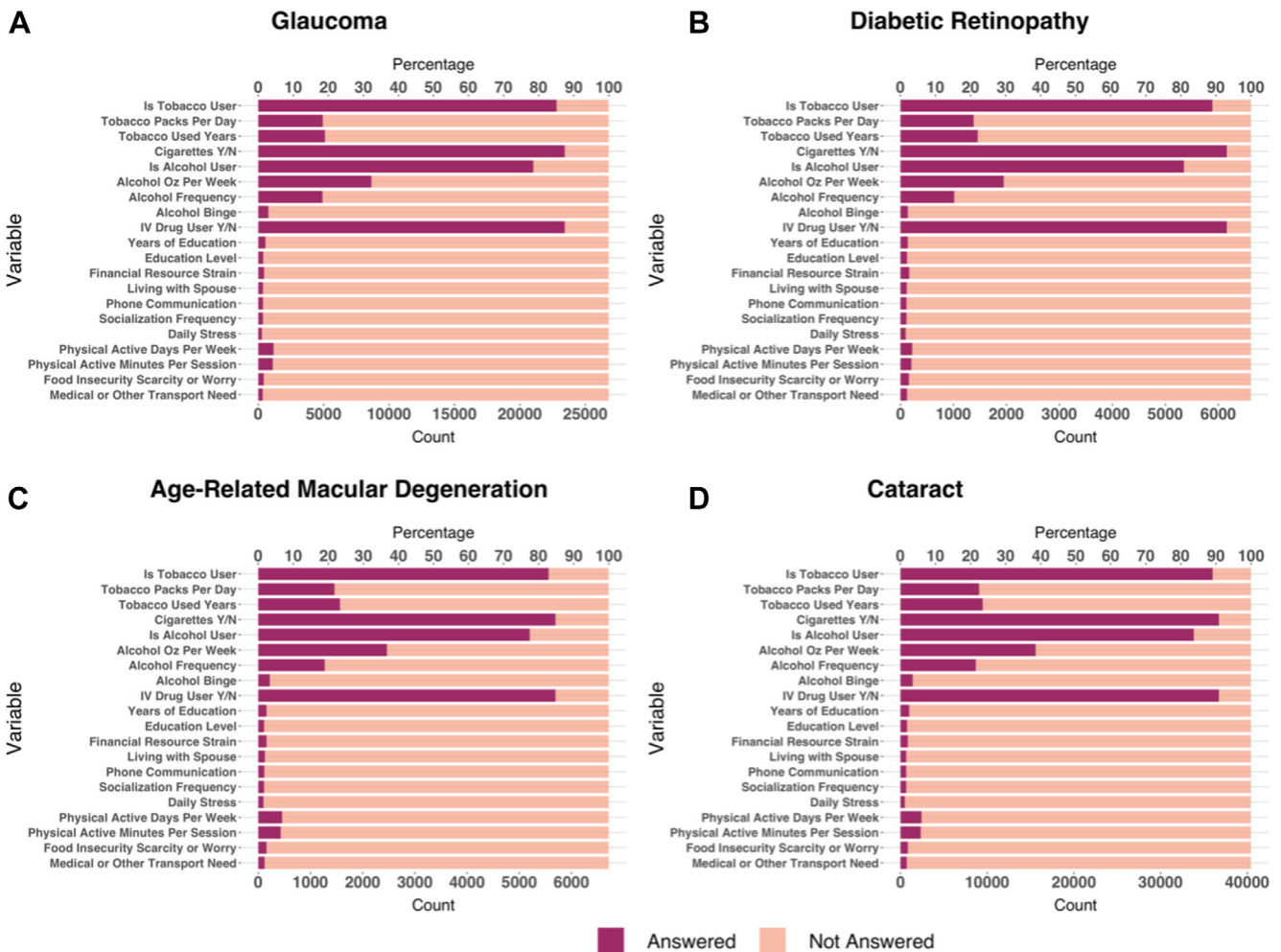


Figure 1. Bar graphs showing data regarding patient use of tobacco, cigarettes, alcohol, or intravenous drugs and variables related to specific characteristics of tobacco use and alcohol use patients with a diagnosis of (A) glaucoma, (B) diabetic retinopathy, (C) age-related macular degeneration, and (D) cataract. N = no; Y = yes.

substantial variation across SDoH data types. Because data in local EHRs are heavily reliant on provider input, a need exists for increased EHR-user education and targeted workflow integration; several studies have reported that adding dedicated SDoH fields and subsets of standardized codes to EHRs are insufficient solutions for increasing documentation of SDoH data.³³ Although past works, including that of Gold et al,⁶ developed SDoH-related workflows to address this need, more research is needed to test these workflows empirically.

Improvements in Data Coverage

Interest is growing in the ophthalmology community in understanding SDoH, as evidenced by the American Academy of Ophthalmology including a chapter on SDoH in its forthcoming Basic and Clinical Science Course.³⁴ Improvements in data coverage must be aimed at increasing provider awareness and facilitating ease of entry into EHR systems. In the All of Us dataset, SDoH

data primarily are collected via self-report survey responses, and one barrier faced by participants to complete data collection is the length and time burden of these surveys. In clinical settings, patients may have privacy concerns or have inadequate time to complete intake forms. Online patient portals that allow data collection before the visit, in addition to streamlined intake forms, are interventions that may improve workflow and overall quality of care.³⁵ In local EHRs, providers face significant time constraints during clinic visits and may overlook SDoH variables if they are not clearly connected to the presenting medical issue. Specialist providers also may defer SDoH documentation to primary care providers; however, primary care clinics often face the most time constraints; prior time-motion studies have highlighted the disproportionate time burden of EHR use specifically in primary care settings.³⁶ In modern-day collaborative practices, providers thus would benefit from institutional guidelines that outline SDoH-related workflows and delineate when (e.g., initial visit vs. follow-up visits), where

(e.g., SDoH-specific data fields vs. free text), and by whom (e.g., primary care vs. specialty clinics) SDoH data should be collected. Data collection also should be streamlined to include clinically relevant survey questions while minimizing irrelevant or repetitive questions.

Electronic health records now are ubiquitous in health care systems, and as such, all stakeholders will benefit from making the EHR database as comprehensive and complete as possible. The time burden of EHR documentation has been a well-studied contributor to physician burnout and decreased face-to-face communication with patients.³⁷ However, EHR tools and predictive models can be used to optimize documentation and practice efficiency.³⁸ Prior studies have shown that although predictive modeling of EHRs has advanced rapidly, research describing actual implementation of these predictive models in clinical settings is lacking.³⁹ Finally, administrators may improve data collection by modifying the existing payment system (e.g., offering new incentives or developing new CMS evaluation and management coding guidelines) to reimburse providers for SDoH documentation. In 2015, CMS developed SDoH-related Z-codes to document SDoH data; however, recent studies have reported underuse of these codes and highlighted a need for alternative solutions.³³

Limitations and Opportunities for Future Investigations

One limitation is that SDoH variables in All of Us were divided between mandatory and optional surveys, thus leading to a lack of data completeness. We mediated this limitation by using chi-square testing to analyze differences between disease cohorts within each survey response. Other limitations include enrollment bias in the All of Us patient cohort and the lack of inclusion of unstructured data in the EHRs.

The data obtained from local EHRs similarly are limited to the fields available in the vendor interface. One limitation

is that providers document SDoH variables in unstructured or free-narrative text format, rather than documenting within the provided fields, highlighting the need for health care institutions to standardize SDoH documentation in EHRs and for researchers to use natural language processing when assembling complete databases. Future investigations should explore the relative impact of SDoH variables to stratify which variables are the most pertinent to the patient's clinical outcome, and therefore more likely to lead to actionable interventions (e.g., food insecurity may be less clinically relevant in glaucoma, but more relevant in DR). In addition, future studies using All of Us may elucidate differences in data coverage among racial or ethnic cohorts as inclusive data collection continues to evolve.

Conclusions

Social determinants of health data play a significant role in understanding the risk factors and management for common eye conditions, yet data coverage is highly variable in both the national All of Us data repository and local EHRs. The variability and paucity of complete SDoH data in both All of Us and local EHRs highlights the need to approach data collection from multiple stakeholder perspectives. For researchers, assembly of complete SDoH datasets is necessary to develop predictive models and to risk-stratify patients based on socioeconomic predictors. For providers, collection of SDoH data presents a significant workflow consideration; however, it is necessary to input data into designated EHR fields so that this information can aid in future encounters. For patients, reporting SDoH variables may impact directly their ability to access care and to adhere to clinical recommendations, and it allows their providers to treat them holistically. Social determinants of health data are pivotal in addressing the underpinning factors that contribute to health inequity, and strategies to improve data collection should account for the role of EHRs in modern-day clinical practice.

Footnotes and Disclosures

Originally received: January 19, 2022.

Final revision: March 23, 2022.

Accepted: March 30, 2022.

Available online: April 5, 2022. Manuscript no. XOPS-D-22-00009

¹ Viterbi Family Department of Ophthalmology and Shiley Eye Institute, University of California, San Diego, La Jolla, California.

² UCSD Health Department of Biomedical Informatics, University of California, San Diego, La Jolla, California.

³ Department of Medicine, University of California, San Diego, La Jolla, California.

⁴ Herbert Wertheim School of Public Health and Human Longevity Science, University of California, San Diego, La Jolla, California.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s): C.N.: Lecturer – American Psychological Association, Dartmouth University, Academy Health; Chair – American Association for the Advancement of Science, Committee on Scientific Freedom and Responsibility; Member – IEEE Organization Governance of Artificial Intelligences Standards Committee. S.L.B.: Consultant – Voxelcloud.io; Lecturer – National Eye Institute, iVista Medical Education

Supported by the National Institutes of Health, Bethesda, Maryland (grant nos.: P30EY022589, 1DP50D029610 [S.L.B.], and 5T15LM01127109 [M.N.]); Research to Prevent Blindness, Inc., New York, New York (unrestricted departmental grant); The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The All of Us Research Program is supported (or funded) by the National Institutes of Health, Office of the Director, Regional Medical Centers, Bethesda, Maryland (grant nos.: 1 OT2 OD026549, 1 OT2 OD026554, 1 OT2 OD026557, 1 OT2 OD026556, 1

OT2 OD026550, 1 OT2 OD 026552, 1 OT2 OD026553, 1 OT2 OD026548, 1 OT2 OD026551, 1 OT2 OD026555).

Presented at: Association for Research in Vision and Ophthalmology Annual Meeting, Virtual, May 1-7, 2021.

HUMAN SUBJECTS: Human subjects were included in this study. This study was approved by the UCSD Institutional Review Board and adheres to the tenets of the Declaration of Helsinki. Individuals participating in All of Us provided written informed consent, and study procedures were approved by the All of Us Institutional Review Board.

No animal subjects were included in this study.

Author Contributions:

Conception and design: Lee, Saseendrakumar, Nayak, Chan, McDermott, Shahrvini, Ye, Sitapati, Nebeker, Baxter

Analysis and interpretation: Lee, Saseendrakumar, Nayak, Chan, McDermott, Shahrvini, Ye, Sitapati, Nebeker, Baxter

Data collection: Lee, Saseendrakumar, Nayak, Chan, McDermott, Shahrvini, Ye, Sitapati, Nebeker, Baxter

Obtained funding: Baxter; Study was performed as part of the authors' regular employment duties. No additional funding was provided.

Overall responsibility: Lee, Saseendrakumar, Nayak, Chan, McDermott, Shahrvini, Ye, Sitapati, Nebeker, Baxter

Abbreviations and Acronyms:

AMD = age-related macular degeneration; **CMS** = Centers for Medicare and Medicaid Services; **DR** = diabetic retinopathy; **EHR** = electronic health record; **IOM** = Institute of Medicine; **SDoH** = social determinants of health; **UCSD** = University of California, San Diego.

Keywords:

Biomedical informatics, Clinical informatics, Data availability, Electronic health record, Social determinants of health.

Correspondence:

Sally L. Baxter, MD, MSc, Viterbi Family Department of Ophthalmology and Shiley Eye Institute, University of California, San Diego, 9415 Campus Point Drive, MC 0946, La Jolla, CA 92093-0946. E-mail: s1baxter@health.ucsd.edu.

References

- Committee on the Recommended Social and Behavioral Domains and Measures for Electronic Health Records; Board on Population Health and Public Health Practice; Institute of Medicine. Implementation issues. In: Capturing Social and Behavioral Domains and Measures in Electronic Health Records: Phase 2. Available at: https://www.ncbi.nlm.nih.gov/books/NBK269333/#sec_000257. January 8, 2015. Accessed 28.11.21.
- Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies. (Vol. 1: Series Overview and Methodology). Rockville, MD: Agency for Healthcare Research and Quality. In: Shojania KG, McDonald KM, Wachter RM, Owens DK, eds; 2004:27–37. Available at: <https://www.ncbi.nlm.nih.gov/books/NBK43908/>. Accessed 27.11.21.
- Centers for Medicare and Medicaid Services. Quality strategy. Available at: <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityInitiativesGenInfo/Legacy-Quality-Strategy>; 2021. Accessed 28.11.21.
- Adler NE, Stead WW. Patients in context—EHR capture of social and behavioral determinants of health. *N Engl J Med*. 2015;372(8):698–701.
- Zhang X, Pérez-Stable EJ, Bourne PE, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. *Ethn Dis*. 2020;27(2):95–106.
- Gold R, Cottrell E, Bunce A, et al. Developing electronic health record (EHR) strategies related to health center patients' social determinants of health. *J Am Board Fam Med*. 2017;30(4):428–447.
- Congdon N. Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol*. 2004;122(4):477–485.
- Constantino MI, Molyneux L, Wu T, et al. Data collection on retinopathy as a public health tool: the Hubble telescope equivalent of looking back in time. *J Diabetes Complications*. 2017;31(4):721–725.
- Sivaprasad S, Gupta B, Gulliford MC, et al. Ethnic variation in the prevalence of visual impairment in people attending diabetic retinopathy screening in the United Kingdom (DRIVE UK). *PLoS One*. 2012;7(6):e39608.
- Moustafa G, Borkar D, et al. Healthcare disparities contribute to missed follow-up visits after cataract surgery in the USA: results from the perioperative care for intraocular lens study. *BMJ Open*. 2021;11(3). <https://doi.org/10.1136/bmjopen-2020-038565>.
- Brézin AP, Lafuma A, Fagnani F, et al. Prevalence and burden of self-reported blindness, low vision, and visual impairment in the french community: a nationwide survey. *Arch Ophthalmol*. 2005;123(8):1117–1124.
- Cumberland PM, Rahi JS; UK Biobank Eye and Vision Consortium. Visual function, social position, and health and life chances: the UK Biobank Study. *JAMA Ophthalmol*. 2016;134(9):959–966.
- Cottrell EK, Dambrun K, Cowburn S, et al. Variation in electronic health record documentation of social determinants of health across a national network of community health centers. *Am J Prev Med*. 2019;57(6 Suppl 1):S65–S73.
- Cronin RM, Jerome RN, Mapes B, et al. Development of the initial surveys for the All of Us Research Program. *Epidemiology*. 2019;30(4):597.
- Mapes BM, Foster CS, Kusnoor SV, et al. Diversity and inclusion for the All of Us research program: a scoping review. *PLoS One*. 2020;15(7).
- Ramirez AH, Gebo KA, Johns Hopkins M, et al. Progress with the All of Us Research Program: opening access for researchers. *JAMA*. 2021;325(24):2441–2442.
- All of Us Research Hub. Data snapshots. Available at: <https://www.researchallofus.org/data-tools/data-snapshots/>; 2021. Accessed 17.12.21.
- United States Centers for Disease Control and Prevention. Common eye disorders and diseases. Available at: <https://www.cdc.gov/visionhealth/basics/ced/index.html>; 2020. Accessed 28.11.21.
- The “All of Us” Research Program. *N Engl J Med*. 2019;381(7):668–676.
- All of Us Public Data Browser. View survey questions and answers. Available at: <https://databrowser.researchallofus.org/survey/family-health-history>; 2021. Accessed 28.11.21.
- Giuse NB, Koonce TY, Kusnoor SV, et al. Institute of Medicine measures of social and behavioral determinants of health: a feasibility study. *Am J Prev Med*. 2017;52(2):199.
- United States Census Bureau. QuickFacts: United States. Available at: <https://www.census.gov/quickfacts/fact/table/US/PST045219>; 2021. Accessed 28.11.21.

23. Jonas JB, Cheung CMG, Panda-Jonas S. Updates on the epidemiology of age-related macular degeneration. *Asia Pac J Ophthalmol*. 2017;6(6):493–497.
24. Wong TY, Klein R, Islam FMA, et al. Diabetic retinopathy in a multi-ethnic cohort in the United States. *Am J Ophthalmol*. 2006;141(3):446–455.
25. Manuel JI. Racial/ethnic and gender disparities in health care use and access. *Health Serv Res*. 2018;53(3):1407–1429.
26. Willis JR, Doan QV, Gleeson M, et al. Self-reported healthcare utilization by adults with diabetic retinopathy in the United States. *Ophthalmic Epidemiol*. 2018;25(5–6):365–372.
27. Syed ST, Gerber BS, Sharp LK. Traveling towards disease: transportation barriers to health care access. *J Community Health*. 2013;38(5):976.
28. Atella V, Piano Mortari A, Kopinska J, et al. Trends in age-related disease burden and healthcare utilization. *Aging Cell*. 2019;18(1):e12861.
29. Thier A, Holmberg C. The patients' view: age-related macular degeneration and its effects—a meta-synthesis. *Disabil Rehabil*. 2022;44(5):661–671.
30. Centers for Medicare and Medicaid Services. Merit-Based Incentive Payment System Explore Measures—QPP. Available at: <https://qpp.cms.gov/mips/explore-measures>; 2021. Accessed 28.11.21.
31. Groß I, Bunce A, Davis J, et al. Initiating and implementing social determinants of health data collection in community health centers. *Popul Health Manag*. 2021;24(1):52–58.
32. Wang M, Pantell MS, Gottlieb LM, Adler-Milstein J. Documentation and review of social determinants of health data in the EHR: measures and associated insights. *J Am Med Inform Assoc*. 2021;28(12):2608–2616.
33. Truong HP, Luke AA, Hammond G, et al. Utilization of social determinants of health ICD-10 z-codes among hospitalized patients in the United States, 2016–2017. *Med Care*. 2020;58(12):1037–1043.
34. American Academy of Ophthalmology. Social determinants of health. In: *Basic and Clinical Science Course*. San Francisco: American Academy of Ophthalmology; 2022:311–316.
35. Bucher S, Maury A, Rosso J, et al. Time and feasibility of prevention in primary care. *Fam Pract*. 2017;34(1):49–56.
36. Arndt BG, Beasley JW, Watkinson MD, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med*. 2017;15(5):419–426.
37. Baxter SL, Gali HE, Mehta MC, et al. Multicenter analysis of electronic health record use among ophthalmologists. *Ophthalmology*. 2021;128(1):165–166.
38. Baxter SL, Gali HE, Chiang MF, et al. Promoting quality face-to-face communication during ophthalmology encounters in the electronic health record era. *Appl Clin Inform*. 2020;11(1):130.
39. Lee TC, Shah NU, Haack A, Baxter SL. Clinical implementation of predictive models embedded within electronic health record systems: a systematic review. *Informatics*. 2020;7(3):25.