Review Article - Narrative Review

Taiwan J Ophthalmol 2024;14:352-359

Access this article online

Quick Response Code:

Website: http://journals.lww.com/TJOP

DOI: 10.4103/tjo.TJO-D-24-00055

Big data and electronic health records for glaucoma research

Isaac A. Bernstein[1], Karen S. Fernandez[1], Joshua D. Stein[2], Suzann Pershing[1], Sophia Y. Wang[1]*

Abstract:

The digitization of health records through electronic health records (EHRs) has transformed the landscape of ophthalmic research, particularly in the study of glaucoma. EHRs offer a wealth of structured and unstructured data, allowing for comprehensive analyses of patient characteristics, treatment histories, and outcomes. This review comprehensively discusses different EHR data sources, their strengths, limitations, and applicability towards glaucoma research. Institutional EHR repositories provide detailed multimodal clinical data, enabling in-depth investigations into conditions such as glaucoma and facilitating the development of artificial intelligence applications. Multicenter initiatives such as the Sight Outcomes Research Collaborative and the Intelligent Research In Sight registry offer larger, more diverse datasets, enhancing the generalizability of findings and supporting large-scale studies on glaucoma epidemiology, treatment outcomes, and practice patterns. The All of Us Research Program, with a special emphasis on diversity and inclusivity, presents a unique opportunity for glaucoma research by including underrepresented populations and offering comprehensive health data even beyond the EHR. Challenges persist, such as data access restrictions and standardization issues, but may be addressed through continued collaborative efforts between researchers, institutions, and regulatory bodies. Standardized data formats and improved data linkage methods, especially for ophthalmic imaging and testing, would further enhance the utility of EHR datasets for ophthalmic research, ultimately advancing our understanding and treatment of glaucoma and other ocular diseases on a global scale.

Keywords:

Electronic health records, glaucoma, information storage and retrieval, registries

Introduction

Electronic health records (EHRs) have been increasingly adopted in both hospital and office settings over the past decade.[1,2] In the United States, results from a 2018 survey showed that 72.1% of ophthalmologists had reported implementing EHRs in their practice, compared to 47% in 2011.[3] The increasing adoption of EHRs worldwide has enabled research studies in ophthalmology of unprecedented breadth and depth. EHRs contain a wide array of structured and unstructured data, including patient demographics, medication histories, laboratory results, imaging, and free-text clinical progress notes. From this data, detailed information on patient characteristics, treatment histories, and intervention outcomes can be determined. The digitization of ophthalmic health records has not only facilitated the collection of highly detailed data but has also elevated the scale of analysis to heights previously unattainable with paper records.

EHR data has several distinct advantages over other sources of large-scale observational data, such as insurance claims or national survey data. Insurance claims data is generated for the purposes of billing for medical service, and thus does not directly include details of history, clinical exam, laboratory and other test results, imaging,

[1]Department of Ophthalmology, Byers Eye Institute, Stanford University, California, [2]Division of Ophthalmology and Visual Sciences, University of Michigan Kellogg Eye Center, Ann Arbor, MI, USA

***Address for correspondence:**
Dr. Sophia Y. Wang, 2370 Watson Ct, Palo Alto, CA 94030, USA.
E-mail: sywang@stanford.edu

Submission: 15-05-2024
Accepted: 05-06-2024
Published: 13-09-2024

How to cite this article: Bernstein IA, Fernandez KS, Stein JD, Pershing S, Wang SY. Big data and electronic health records for glaucoma research. Taiwan J Ophthalmol 2024;14:352-9.

or free-text clinical notes. National survey data, such as National Health and Nutrition Examination Survey, is often very broad, with limited information on ophthalmic care.[4] Furthermore, the data rely upon participant self-report, which can be unreliable. In contrast, EHR data is generally entered by clinical providers and staff and is derived directly from the clinical setting. It may include physical exam findings such as visual acuity (VA) or intraocular pressure (IOP), which are especially relevant to the study of ophthalmic disease. Clinical free-text notes, test results, and/or ophthalmic imaging may also be available, providing additional depth.

Given the many advantages of EHR data, researchers have increasingly turned to its use to investigate many ophthalmic diseases, including glaucoma, a leading cause of irreversible blindness worldwide.[5] The clinical diagnosis and management of glaucoma involves integration of multiple data modalities, which is particularly suited for analyses using EHR data. This paper aims to review the role of big data and EHRs in glaucoma research. We will review major sources and different types of ophthalmology-relevant big data derived from EHR [Table 1] and how they have been used to advance our understanding of glaucoma. We will explore the challenges and opportunities associated with leveraging EHR data for glaucoma research, highlighting key achievements and future directions. By tapping into the power of big data analytics, researchers and clinicians can enhance glaucoma diagnosis, treatment, and patient care, ultimately reducing the global burden of this sight-threatening disease.

Individual Institutional Electronic Health Records Repositories

With the adoption of EHR systems into medical practice, several academic institutions have collected and organized clinical EHR data into repositories designed for research. Because the data comes from a single institution, there may be greater control over data quality and consistency, as the institution can implement standardized data quality assurance procedures. Other benefits of institutional datasets include availability of specific clinical variables, limited only by their capture in the EHR. For example, the Stanford translational research integrated database environment contains data on clinical encounters, diagnoses, procedures, radiology, pathology, and laboratory test reports, inpatient pharmacy orders, biospecimens, and other clinical documents for over 3.4 million patients cared for at Stanford University Medical Center since 1995.[6,7] Gui *et al*. recently used Stanford EHR data to compare IOP measurements from Tono-Pen against Goldmann applanation in 4550 patients. Because the Stanford dataset included data on both types of IOP measurements, central corneal thickness, and details on each patient's diagnosis, researchers could ultimately demonstrate a mean difference of 0.15 ± 5.5 mmHg, also finding Tono-Pen overestimating IOP <16.5 mmHg and underestimated IOP >16.5 mmHg compared to Goldmann.[8]

Another benefit of institutional datasets is the relative ease of integrating standard structured EHR with ophthalmic imaging and testing data into ophthalmology-specific

Table 1: Sources and characteristics of electronic health records data repositories for ophthalmic research

Dataset	Data types	Size	Population	Access
Individual institutional repository	All data types may potentially be included, such as structured EHRs with ophthalmic examination data, free-text clinical progress notes, ophthalmic imaging, and testing data	Varies by individual institution. For example, Stanford Medicine Research Data Repository: >3.4 million patients[7] Duke ophthalmic registry: >400,000 patients[9] Bascom palmer glaucoma repository >73,000 patients[12]	Patients seen at that center	Generally limited to researchers at the specific institution
SOURCE	Structured EHRs data including ophthalmic examination data, some social determinants of health information	>640,000 subjects[20]	Patients seen at affiliated academic centers across the US	Must be affiliated with a SOURCE member institution
IRIS	Structured EHRs data including ophthalmic examination data	>72 million patients[27]	Mostly private ophthalmology practices from across the US	Available through academic institutions in the IRIS Registry Analytic Center Consortium and other specific research mechanisms
All of US	Structured EHRs data, patient self-reported survey data, genomic data, wearable sensor data. No availability of ophthalmic examination data	>790,000 subjects[50]	Eligible US adults	Must be affiliated with a registered institution available here: https://www.researchallofus.org/institutional-agreements/

SOURCE=Sight Outcomes Research Collaborative, IRIS=Intelligent Research in Sight, EHRs=Electronic health records

data repositories. Ophthalmic imaging and testing data is of vital importance to ophthalmology outcomes and often resides in separate systems from the standard structured EHR data. As the data reside within a single institution, repositories can more easily combine silos of data comprising different modalities. For instance, the Duke Ophthalmic Registry contains EHR data, optical coherence tomography (OCT) images, and visual field data from over 400,000 patients.[9] Jammal *et al*. leveraged this multimodal data to study the relationship between blood pressure and structural damage over time in glaucoma. Pairing EHR data containing blood and IOPs with spectral-domain OCT imaging data enabled the authors to find that mean arterial pressure and diastolic arterial pressure were associated with significantly faster rates of retinal nerve fiber layer change over time.[10] In another Duke Ophthalmic Registry glaucoma study, over 19,000 OCT and standard automated perimetry tests were analyzed to quantify structural and functional changes to the retinal nerve fiber layer.[11] Similarly, using multimodal data in the Bascom Palmer Glaucoma Repository, Gallo Afflitto and Swaminathan evaluated racial-ethnic differences in key structural glaucoma parameters. Analyzing over 20,000 OCT peripapillary retinal nerve fiber layer (pRNFL) and over 14,000 macular ganglion cell-inner plexiform layer (mGCIPL) scans from 2002 eyes, they demonstrated faster rates of pRNFL thinning and mGCIPL loss among non-Hispanic Black glaucoma suspects.[12]

Institutional EHR datasets can also include free-text clinical progress notes, especially helpful for modern natural language processing (NLP) applications. Progress notes often contain information on patient exam findings, procedures, and medical, family, and social history which are much more detailed and accurate than can be captured in structured codes. In one previous study investigating glaucoma surgery outcomes, variables for type of glaucoma surgery (model of tube shunt implant used, type of minimally invasive glaucoma surgery [MIGS] performed) and laterality were extracted from free-text operative notes using an NLP algorithm.[13,14] These informatics approaches then enabled the detailed comparison of real-world outcomes of trabeculectomies, Ex-Press shunts, Ahmed tube shunts, and Baerveldt tube shunts.[14] Free text data in institutional EHR datasets has also been used in artificial intelligence (AI) applications for ophthalmology, for example to train domain-specific word embeddings and develop multimodal models or use large language models to predict which glaucoma patients will progress to require surgery.[15-17]

One major limitation of institutional datasets is that they are limited to the patient population of a given tertiary academic medical center. Thus, findings may not generalize well to other patient populations. The local population may be limited in racial, ethnic, or socioeconomic diversity, and/or limited in disease phenotypes and presentations. Cohort sizes may also be relatively small, especially for rare diseases. Local practice patterns may also not reflect broader practice patterns, especially in glaucoma, where surgical practices often reflect the preference of individual surgeons and the characteristics of glaucoma presentation in the local population. Furthermore, use of institutional data is typically limited to investigators affiliated with the institution. The data structure of the institutional dataset may be idiosyncratic or specific to that institution, thus limiting the ability to integrate other institutions' data in the future. To facilitate more diverse multi-institutional studies, future efforts should encourage data sharing across institutions and could include standardizing data structures and federated learning.[18,19] Nonetheless, to those with access, institutional datasets offer greater control over structure and variables, relative ease of access within the institution, integration of EHR with imaging data, and availability of free text data.

The Sight Outcomes Research Collaborative Repository

The Sight Outcomes Research Collaborative (SOURCE) (sourcecollaborative.org) is a consortium that comprises academic ophthalmology departments that share de-identified EHR and ocular diagnostic test data.[20] There are at least 18 health systems actively contributing their data, with more institutions in the process of joining. SOURCE includes data on patient demographics; VA, IOP, and other eye exam findings; systemic and ocular diagnosis (International Classification of Disease [ICD-9] and ICD-10) and procedure (CPT-4) codes; electronic orders and charges; systemic and ocular medications prescribed; as well as encounter records for close to 6 million eye care recipients. A variety of measures of social determinants of health are also available, including neighborhood-level measures such as Distressed Communities Index and rural-urban commuting area codes, as well as individual-level measures including personal income, household net worth, education, and others that have been connected to the clinical data in SOURCE using privacy-preserving linkages.[21,22] To contribute data to SOURCE, member institutions use a standard script to extract and format data on all eye care recipients seen at their institution from the activation of their EHR system to the present, and continuing into the future. Protected health information is removed prior to transmission to SOURCE Data Center. Faculty and trainees at active sites are then welcome to submit proposals to access the pooled data to conduct research

or quality improvement studies, leveraging the power and scale of multicenter data.

Researchers have leveraged data from SOURCE to study an array of different ocular diseases, including glaucoma. Use cases have included epidemiological studies, outcomes research, AI predictive models, and describing disparities and inequities in eye care. To date, there have been several glaucoma-related studies using SOURCE repository data. For example, researchers used SOURCE data to create algorithms to help prioritize timing of glaucoma patient visits during the COVID-19 pandemic based on glaucoma severity.[23] In the context of tele-ophthalmology, SOURCE data was used to study low-contact glaucoma monitoring during the COVID-19 pandemic. Sun *et al*. examined the reliability of drive-through IOP measurements using a handheld tonometer during the COVID-19 pandemic, finding that these measurements tend to overestimate IOP compared to in-clinic measurements, suggesting the need for caution and confirmation with in-clinic measurements for management decisions, despite the potential of drive-through tonometry as a low-contact monitoring method.[24] SOURCE data has also been used to develop AI applications for glaucoma. Using demographic data, medication prescriptions, IOP, VA, and other variables from more than 36,000 patients with glaucoma in SOURCE, logistic regression, tree-based, and deep learning models were trained to predict progression to requiring incisional glaucoma surgery. Because SOURCE data is derived from numerous academic centers, investigators in this study were able to reserve data from one site as an external test set for model external validation, making this data particularly useful in AI research.[21]

The multicenter nature of SOURCE overcomes many limitations of single-center data repositories. Participating centers are located throughout the country and bring much more geographic as well as racial and ethnic diversity to the patient population than would be possible in single-center studies. The number of patients, including patients with uncommon or rare ocular diseases is much greater than could be attained at any one site. Data is harmonized across sites, ensuring that the resulting repository is relatively clean and easy to work with, for a resource of its size. It is also easier for researchers to access SOURCE data relative to some other Big Data resources. There are several limitations to SOURCE. Presently, SOURCE is only able to work with institutions that use Epic as their EHR vendor; this limits private practices, Veterans Affairs Medical Centers, and many international sites from being able to participate. Although SOURCE captures ocular as well as nonocular care received by patients seen within the health systems, if a patient goes outside of any of these health systems for care, that care would not be captured. While all the participating sites are sharing their EHR data, it is optional for sites to share ancillary data such as data from ocular diagnostic tests. While some of the active sites have shared this data, not all have done so. Hopefully as SOURCE continues to grow, more and more sites will share this data with the consortium.

The Intelligent Research in Sight Registry

Developed by the American Academy of Ophthalmology, the Intelligent Research in Sight (IRIS®) Registry is considered the world's largest medical specialty data registry.[25,26] The IRIS Registry enables ophthalmologists throughout the United States to submit data from their practice's EHR database, thereby assisting them in fulfilling the quality reporting requirements of the merit-based incentive payment system mandated by the Medicare access and CHIP reauthorization act. IRIS currently includes data on over 72 million patients from almost 14,000 ophthalmology practices, with most data coming from ophthalmology practices unaffiliated with academic medical centers.[27] As of October 1, 2023, the IRIS Registry contains aggregated, de-identified data on over 434.5 million visits and 72.7 million unique patients.[27]

To reduce the burden of data submission for participating practices, the IRIS Registry can integrate with 41 different EHR systems that can report data on a nightly or weekly basis. IRIS Registry data includes patient demographics, examination findings such as IOP and cup-disk ratio, code-based diagnoses and procedures, laterality, and select history variables (e.g. smoking, in some iterations of the dataset), as well as limited data on medication use. One key strength of the IRIS Registry is its large size, representing over 70 million US patients seen in ophthalmology practices. For instance, Lee *et al*. used IRIS Registry data to examine the relationship between IOP and smoking. Examining over 12 million patients, they found a statistically significant higher IOP in current and past smokers after adjusting for demographics and ophthalmic morbidity.[28]

The large size of the registry not only facilitates the study of rare conditions or outcomes, which would be challenging in smaller datasets, but also enhances national-scale research of relatively common conditions like glaucoma. For instance, Sun *et al*.[29] investigated a cohort of 3123 subjects on anti-CTLA-4 or anti-PD-1 therapy to determine incidence of ophthalmic adverse events, such as uveitis. Similarly, Shah *et al*. evaluated IRIS Registry data to determine the risk factors for blindness among patients with primary angle closure glaucoma, with findings including higher odds of blindness among subjects with Medicaid or Medicare insurance, after

adjusting for ocular co-morbidities.[30] Given its large population, the IRIS Registry is exceptionally well-suited for studying rare outcomes such as blindness or other rare diseases, offering insights that smaller datasets could not provide.

Previous IRIS studies have examined glaucoma practice patterns at scale. The tube versus trabeculectomy IRIS Registry Study evaluated 419 glaucomatous eyes and found no significant differences in 1-year IOP reduction or failure risk between the two treatments.[31,32] Another study examined 263,480 glaucomatous eyes to identify the factors linked to the effectiveness of laser trabeculoplasty. Higher baseline IOP was associated with response, whereas angle recession and uveitis were associated with nonresponse.[33] In a retrospective cohort study, Hall *et al*. examined over 44,000 patients having undergone tube shunt implantation to uncover factors associated with risk of device revision/removal surgery, such as chronic angle closure glaucoma or dry eye disease.[34] Usage and outcomes of MIGS methods have also been assessed.[35,36] One IRIS study demonstrates an increase in the proportion of MIGS procedures performed over 5 years whereas another analyzes how MIGS practice patterns vary by glaucoma type.[37,38] For instance, iStent was the most common procedure in open angle glaucoma, whereas more invasive glaucoma drainage devices were more common in secondary glaucoma.[38] Since IRIS includes data from nonacademic settings, Skuta *et al*. could assess how glaucoma practice patterns differ between academic and non-academic settings, highlighting differences in workup and management. For instance, academic practices performed significantly more tube shunt procedures, whereas nonacademic settings did more endoscopic cyclophotocoagulation.[39]

However, the use of the IRIS Registry comes with several key limitations. Despite its size, the IRIS Registry lacks the capacity to generate nationally representative population prevalence estimates since data are limited to patients managed at ophthalmology clinics.[40] Due to privacy limitations and contractual agreements with participating practices, only rigorously de-identified IRIS Registry data can be used for research. Furthermore, access to aggregate de-identified data for research is available through specific mechanisms, including American Academy of Ophthalmology research initiatives, limited organizational grants such as through the American Glaucoma Society and Research to Prevent Blindness, commercially sponsored projects through the registry vendor, or projects performed through academic institutions in the IRIS Registry Analytic Center Consortium. Diagnosis and procedure information is based on ICD and Current Procedural Terminology/Healthcare Common Procedure Coding System codes, with inherent potential for missing or inaccurate data. Since data is derived almost exclusively from ophthalmology practices, information on non-ophthalmic healthcare and systemic disease is incomplete and unreliable. Furthermore, IRIS data does not include free-text notes, laboratory and other test results, or imaging data.[41] Despite these considerations, the IRIS Registry is ideal for larger, nationwide studies specifically focused on ophthalmic disease, and in particular rare ophthalmic diseases.

The All of Us Research Program

The All of Us Research Program is a national, longitudinal cohort study administered by the National Institutes of Health. All of Us began enrolment in May 2018 of participants aged 18 years and above from a network of more than 340 recruitment sites. This research program aims to advance precision medicine and improve health outcomes. In addition, it places a strong emphasis on developing a diverse cohort that comprises people from the backgrounds historically underrepresented in biomedical research including those with limited access to care and from racial/ethnic minority groups. Prospective participants enroll digitally through the program's website or mobile application, providing consent for a comprehensive data collection process encompassing EHRs, biospecimens, and self-reported surveys. As of July 2023, the initiative has enrolled approximately half a million participants.

The breadth of EHR data captured in this repository facilitated Baxter *et al*. to develop a machine learning model that performed better than single center models to predict the need for surgery among glaucoma patients.[42] Notably, the program's intentionality to include marginalized patients holds particular significance in glaucoma research where minorities, specifically African Americans, share a disproportionate disease burden. Another study leveraging All of Us data reported an association between low blood pressure and increased risk of developing open-angle glaucoma. The platform was key for this study because it provided blood pressure measurements that are absent in claims datasets but available in EHR databases.[43] All Of Us data undergoes standardization using the observational medical outcomes partnership common data model (OMOP-CDM), ensuring interoperability, consistency, streamline processes, accuracy, and lays the groundwork for collaborative and effective data management.[44]

Despite its extensive size and depth of data, the All of Us database presently lacks dedicated ophthalmic data such as visual acuities, IOPs, or any ophthalmic imaging. This is partly due to gaps in coverage in the

OMOP-CDM for ophthalmic concepts, which is an area of active development.[18] Furthermore, while All Of Us contains substantial self-reported survey data on social determinants of health, survey completion rates for most surveys beyond basic demographic information are low.[45] To address this, the program has introduced computer-assisted telephone interviews to enhance future completion rates. The repository also houses de-identified genomic data, genotyping arrays, and Fitbit device data, although only on a subset of the whole cohort.[46]

Access to data from All of Us is restricted to researchers affiliated with specific institutions which have completed a Data Use and Research Agreement with All of Us.[47] Researchers submit proposals through the program's research center to ensure compliance with ethical standards and data security protocols. Upon approval, researchers gain access to de-identified data through a cloud-based platform accessible exclusively through the program.[48] While this approach safeguards data integrity and privacy, this can hinder broader collaboration, pose a challenge for research requiring detailed individual-level information, and create barriers for researchers unfamiliar with the program's interface.

Discussion and Conclusion

By capturing and organizing data generated in routine clinical care on a large scale, EHR datasets have revolutionized the study of ocular disease, particularly glaucoma. The suitability of using institutional, multicenter, and national EHR repositories for research depends on data accessibility and desired data types, such as eye exam findings, ophthalmic imaging results, or free text notes. Although institutional datasets are necessarily limited in size and scope compared to multi-center datasets, they remain best suited for analyses which require integration of structured EHR data with ophthalmic imaging and/or clinical free-text data. As such, multimodal data from institutional datasets prove especially useful for clinical AI applications. For the study of diseases or outcomes in larger populations, multicenter or national registries such as the SOURCE repository, IRIS Registry, or All of Us are best suited. Where detailed eye examination information, such as VA or IOP is required, SOURCE and IRIS Registry are superior. Given its substantial size, the IRIS Registry is ideal for conducting high-powered studies on common diseases, as well as investigating rare conditions. The All of Us dataset, conversely, is the most diverse due to emphasis and intentional inclusion of underrepresented populations. Another unique benefit of All of Us is inclusion of a wider variety of clinical data, including medications, measurements, laboratory test results, and unique survey-gathered information on social determinants of health, though it currently lacks key ophthalmic variables, like IOP.

Several challenges remain in the development and use of EHR datasets. Access to EHR datasets remain limited, with institutional affiliation required for most. Healthcare data is subject to various regulations and laws, such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States; hence, compliance with these regulations may be enforced by limiting access to datasets and ensuring that researchers adhere to strict guidelines when handling patient data. Access restrictions may also be imposed to uphold ethical standards in research, to prevent misuse of data, and to protect vulnerable populations from potential harm or exploitation. Limiting access to datasets may also be a strategic decision to manage resources effectively. With large costs of developing and maintaining large datasets, governing bodies prioritize access for affiliated researchers or grant awardees who have demonstrated a commitment to the institution or research objectives. Several strategies may facilitate access to empower qualified yet unaffiliated scientists in contributing to ophthalmic research. Misuse of data may be mitigated with a rigorous registration process including verification of identity, credentials, and affiliations. Training programs may also be useful to this end. For instance, All of Us requires completion of a training program prior to accessing their data, and requires investigators to conduct all analyses within their dedicated coding environments.[49]

Larger datasets such as the IRIS Registry, SOURCE repository, and All of Us would also benefit from multimodal data linkage, incorporating data from OCT imaging, visual fields, or fundus photographs into analyses or applications. This remains a challenge for the future, subject to data sharing agreements, development and adherence to imaging data standards, and compatibility across platforms.

All EHR datasets would benefit from a more standardized format to facilitate larger, more diverse, and more well-powered analyses or applications. To this end, standardized formats like the Observational Health Data Sciences and Informatics OMOP CDM must be revamped to include a full coverage of ophthalmic concepts.[18]

By providing large sizes and varieties of data, EHR datasets empower researchers to investigate ophthalmic disease. The investigation of glaucoma especially benefits from this. Glaucoma Researchers, institutions, and regulatory bodies must work together to improve data accessibility, standardization, and integration, thereby advancing our understanding and treatment of ocular diseases on a global scale.

Data availability statement

The datasets generated during and/or analyzed during the current study are publicly available.

Financial support and sponsorship

This work was funded in part by: National Eye Institute K23EY03263501 (SYW); Career Development Award from Research to Prevent Blindness (SYW); unrestricted departmental grant from Research to Prevent Blindness (SYW, IAB, KSF, SP); departmental grant National Eye Institute P30-EY026877 (SYW, IAB, KSF, SP).

Conflicts of interest

The authors declare that there are no conflicts of interests of this paper.

References

1. Jiang JX, Qi K, Bai G, Schulman K. Pre-pandemic assessment: A decade of progress in electronic health record adoption among U.S. hospitals. Health Aff Sch 2023;1:qxad056.
2. National Trends in Hospital and Physician Adoption of Electronic Health Records. Available from: https://www.healthit.gov/data/quickstats/national-trends-hospital-and-physician-adoption-electronic-health-records. [Last accessed on 2024 Apr 22].
3. Lim MC, Boland MV, McCannel CA, Saini A, Chiang MF, Epley KD, *et al.* Adoption of electronic health records and perceptions of financial and clinical outcomes among ophthalmologists in the United States. JAMA Ophthalmol 2018;136:164-70.
4. NHANES Questionnaires, Datasets, and Related Documentation. Available from: https://www.n.cdc.gov/nchs/nhanes/continuousnhanes/default.aspx. [Last accessed on 2024 Apr 22].
5. Jayaram H, Kolko M, Friedman DS, Gazzard G. Glaucoma: Now and beyond. Lancet 2023;402:1788-801.
6. Lowe HJ, Ferris TA, Hernandez PM, Weber SC. STRIDE – An integrated standards-based translational research informatics platform. AMIA Annu Symp Proc 2009;2009:391-5.
7. Callahan A, Ashley E, Datta S, Desai P, Ferris TA, Fries JA, *et al.* The Stanford medicine data science ecosystem for clinical and translational research. JAMIA Open 2023;6:ooad054.
8. Gui H, Zhang Y, Chang RT, Wang SY. Real-world agreement of same-visit Tono-Pen versus Goldmann applanation intraocular pressure measurements using electronic health records. Heliyon 2023;9:e18703.
9. Using Big Data to Drive Advancements in Ophthalmology. Duke Department of Ophthalmology; 2022. Available from: https://www.dukeeyecenter.duke.edu/news/using-big-data-drive-advancements-ophthalmology. [Last accessed on 2024 Apr 22].
10. Jammal AA, Berchuck SI, Mariottoni EB, Tanna AP, Costa VP, Medeiros FA. Blood pressure and glaucomatous progression in a large clinical population. Ophthalmology 2022;129:161-70.
11. Jammal AA, Thompson AC, Mariottoni EB, Urata CN, Estrela T, Berchuck SI, *et al.* Rates of glaucomatous structural and functional change from a large clinical population: The duke glaucoma registry study. Am J Ophthalmol 2021;222:238-47.
12. Gallo Afflitto G, Swaminathan SS. Racial-ethnic disparities in concurrent rates of peripapillary and macular OCT parameters among a large glaucomatous clinical population. Eye (Lond) 2024; [doi: 10.1038/s41433-024-03103-3].
13. Wang SY, Pershing S, Tran E, Hernandez-Boussard T. Automated extraction of ophthalmic surgery outcomes from the electronic health record. Int J Med Inform 2020;133:104007.
14. Sun MT, Singh K, Wang SY. Real-world outcomes of glaucoma filtration surgery using electronic health records: An informatics study. J Glaucoma 2022;31:847-53.
15. Wang SY, Tseng B, Hernandez-Boussard T. Deep learning approaches for predicting glaucoma progression using electronic health records and natural language processing. Ophthalmol Sci 2022;2:100127.
16. Jalamangala Shivananjaiah SK, Kumari S, Majid I, Wang SY. Predicting near-term glaucoma progression: An artificial intelligence approach using clinical free-text notes and data from electronic health records. Front Med (Lausanne) 2023;10:1157016.
17. Hu W, Wang SY. Predicting glaucoma progression requiring surgery using clinical free-text notes and transfer learning with transformers. Transl Vis Sci Technol 2022;11:37.
18. Cai CX, Halfpenny W, Boland MV, Lehmann HP, Hribar M, Goetz KE, *et al.* Advancing toward a common data model in ophthalmology: Gap analysis of general eye examination concepts to standard Observational Medical Outcomes Partnership (OMOP) concepts. Ophthalmol Sci 2023;3:100391.
19. Hanif A, Lu C, Chang K, Singh P, Coyner AS, Brown JM, *et al.* Federated learning for multicenter collaboration in ophthalmology: Implications for clinical diagnosis and disease epidemiology. Ophthalmol Retina 2022;6:650-6.
20. Sight Outcomes Research Collaborative. Available from: https://www.sourcecollaborative.org/. [Last accessed on 2024 Apr 22].
21. Wang SY, Ravindranath R, Stein JD, SOURCE Consortium. Prediction models for glaucoma in a multicenter electronic health records consortium: The sight outcomes research collaborative. Ophthalmol Sci 2024;4:100445.
22. Brandon. Distressed Communities Index. Economic Innovation Group. Available from: https://www.eig.org/distressed-communities/. [Last accessed on 2024 Apr 22].
23. Bommakanti NK, Zhou Y, Ehrlich JR, Elam AR, John D, Kamat SS, *et al.* Application of the sight outcomes research collaborative ophthalmology data repository for triaging patients with glaucoma and clinic appointments during pandemics such as COVID-19. JAMA Ophthalmol 2020;138:974-80.
24. Sun LW, Zhou Y, Stein JD, Weizer JS. Reliability of intraocular pressure measurements in a low-contact drive-through setting. J Glaucoma 2024;33:24-7.
25. Parke Ii DW, Lum F, Rich WL. The IRIS® Registry: Purpose and perspectives. Ophthalmologe 2017;114:1-6.
26. Lee CS, Blazes M, Lorch A, Pershing S, Hyman L, Ho AC, *et al.* American Academy of Ophthalmology Intelligent Research in Sight (IRIS®) Registry and the IRIS registry analytic center consortium. Ophthalmol Sci 2022;2:100112.
27. IRIS Registry Data Analysis: Overview. American Academy of Ophthalmology; 2017. Available from: https://www.aao.org/iris-registry/data-analysis/requirements. [Last accessed on 2024 May 11].
28. Lee CS, Owen JP, Yanagihara RT, Lorch A, Pershing S, Hyman L, *et al.* Smoking is associated with higher intraocular pressure regardless of glaucoma: A retrospective study of 12.5 million patients using the Intelligent Research in Sight (IRIS®) registry. Ophthalmol Glaucoma 2020;3:253-61.
29. Sun MM, Kelly SP, Mylavarapu Bs AL, Holland GN, Coleman AL, Yu F, *et al.* Ophthalmic Immune-Related Adverse Events after Anti-CTLA-4 or PD-1 Therapy Recorded in the American Academy of Ophthalmology Intelligent Research in Sight Registry. Ophthalmology. 2021 Jun;128(6):910-919. doi: 10.1016/j.ophtha.2020.11.001. Epub 2020 Nov 6. PMID: 33166553.
30. Shah SN, Zhou S, Sanvicente C, Burkemper B, Apolo G, Li C, *et al.* Prevalence and risk factors of blindness among primary angle closure glaucoma patients in the United States: An IRIS registry analysis. Am J Ophthalmol 2024;259:131-40.
31. Sun CQ, McSoley MJ, Lum F, Chang TC, Gedde SJ, Vanner EA. Tube versus trabeculectomy IRIS registry study: 1-year secondary

treatment outcomes. Am J Ophthalmol 2023;256:97-107.
32. Vanner EA, Sun CQ, McSoley MJ, Persad PJ, Feuer WJ, Lum F, *et al.* Tube versus trabeculectomy IRIS(Ⅱ) registry 1-year composite outcome analysis with comparisons to the randomized controlled trial. Am J Ophthalmol 2021;227:87-99.
33. Chang TC, Parrish RK, Fujino D, Kelly SP, Vanner EA. Factors associated with favorable laser trabeculoplasty response: IRIS registry analysis. Am J Ophthalmol 2021;223:149-58.
34. Hall NE, Chang EK, Samuel S, Gupta S, Klug E, Elze T, *et al.* Risk factors for glaucoma drainage device revision or removal using the IRIS registry. Am J Ophthalmol 2022;240:302-20.
35. Yang SA, Ciociola EC, Mitchell W, Hall N, Lorch AC, Miller JW, *et al.* Effectiveness of microinvasive glaucoma surgery in the United States: Intelligent research in sight registry analysis 2013-2019. Ophthalmology 2023;130:242-55.
36. Olivier MM, Smith OU, Croteau-Chonka CC, VanderBeek BL, Maguire MG, Lum F, *et al.* Demographic and clinical characteristics associated with minimally invasive glaucoma surgery use: An Intelligent Research in Sight (IRIS®) registry retrospective cohort analysis. Ophthalmology 2021;128:1292-9.
37. Yang SA, Mitchell W, Hall N, Elze T, Lorch AC, Miller JW, *et al.* Trends and usage patterns of minimally invasive glaucoma surgery in the United States: IRIS® registry analysis 2013-2018. Ophthalmol Glaucoma 2021;4:558-68.
38. Yang SA, Mitchell WG, Hall N, Elze T, Miller JW, Lorch AC, *et al.* Usage patterns of Minimally Invasive Glaucoma Surgery (MIGS) differ by glaucoma type: IRIS registry analysis 2013-2018. Ophthalmic Epidemiol 2022;29:443-51.
39. Skuta GL, Ding K, Lum F, Coleman AL. An IRIS registry-based assessment of primary open-angle glaucoma practice patterns in academic versus nonacademic settings. Am J Ophthalmol 2022;242:228-42.
40. Brant A, Kolomeyer N, Goldberg JL, Haller J, Lee CS, Lee AY, *et al.* United States population disparities in ophthalmic care: Blindness and visual impairment in the IRIS® registry (Intelligent Research in Sight). Ophthalmology 2023;130:1121-37.
41. Pershing S, Lum F. The American Academy of Ophthalmology IRIS registry (Intelligent Research in Sight): Current and future state of big data analytics. Curr Opin Ophthalmol 2022;33:394-8.
42. Baxter SL, Saseendrakumar BR, Paul P, Kim J, Bonomi L, Kuo TT, *et al.* Predictive analytics for glaucoma using data from the all of us research program. Am J Ophthalmol 2021;227:74-86.
43. Lee EB, Hu W, Singh K, Wang SY. The association among blood pressure, blood pressure medications, and glaucoma in a nationwide electronic health records database. Ophthalmology 2022;129:276-84.
44. Data Standardization – OHDSI. Available from: https://www.ohdsi.org/data-standardization/. [Last accessed on 2024 Apr 22].
45. Sulieman L, Cronin RM, Carroll RJ, Natarajan K, Marginean K, Mapes B, *et al.* Comparing medical history data derived from electronic health records and survey answers in the all of us research program. J Am Med Inform Assoc 2022;29:1131-41.
46. All of Us Research Program Genomics Investigators. Genomic data in the All of Us Research Program. Nature 2024;627:340-6. doi: 10.1038/s41586-023-06957-x. Epub 2024 Feb 19.
47. Registered Institutions – All of Us Research Hub. Available from: https://www.researchallofus.org/institutional-agreements/. [Last accessed on 2024 Apr 22].
48. Ramirez AH, Sulieman L, Schlueter DJ, Halvorson A, Qian J, Ratsimbazafy F, *et al.* The all of us research program: Data quality, utility, and diversity. Patterns (N Y) 2022;3:100570.
49. Researcher Workbench – All of Us Research Hub. Available from: https://www.researchallofus.org/data-tools/workbench/. [Last accessed on 2024 May 11].
50. Data Snapshots – All of Us Research Hub. Available from: https://www.researchallofus.org/data-tools/data-snapshots/. [Last accessed on 2024 May 11].