YALE JOURNAL OF BIOLOGY AND MEDICINE 96 (2023), pp.421-426.

PERSPECTIVES

YALE JOURNAL OF BIOLOGY AND MEDICINE
YJBM
CELEBRATING 95 YEARS 1928–2023

Ophthalmology at the Forefront of Big Data Integration in Medicine: Insights from the IRIS Registry Database

Austen N. Knapp[a], Theodore Leng[a], and Ehsan Rahimy[a,b,*]

[a]*Department of Ophthalmology, Byers Eye Institute, Stanford University School of Medicine, Palo Alto, CA, USA;* [b]*Department of Ophthalmology, Palo Alto Medical Foundation, Palo Alto, CA, USA*

Ophthalmology stands at the vanguard of incorporating big data into medicine, as exemplified by the integration of The Intelligent Research in Sight (IRIS) Registry. This synergy cultivates patient-centered care, demonstrates real world efficacy and safety data for new therapies, and facilitates comprehensive population health insights. By evaluating the creation and utilization of the world's largest specialty clinical data registry, we underscore the transformative capacity of data-driven medical paradigms, current shortcomings, and future directions. We aim to provide a scaffold for other specialties to adopt big data integration into medicine.

INTRODUCTION – THE BEGINNINGS OF BIG DATA

The use of real-world data (RWD) in clinical research is a growing trend that has the potential to revolutionize the way we understand and treat disease. Big data, defined to encompass both large quantities and diverse types of data, was originally described in the 1990s [1]. However, the impact on the medical field was not realized until much more recently due to the passive collection of large amounts of patient data catalyzed by the advances in electronic health records (EHRs), medical imaging, and data analytics technology.

Ophthalmology has been at the forefront of big data utilization in healthcare research, partially because many ophthalmic conditions can be objectively measured and tracked over time, and the use of noninvasive, fast imaging modalities provide detailed information on disease progression and treatment response [2]. Big data is an aggregation of RWD - information that is routinely collected during the delivery of healthcare. Systematic analysis of RWD yields real-world evidence (RWE). This information has been used to evaluate clinical outcomes, practice patterns among physicians, epidemiology of different diseases, and adverse events of different therapies. As the collection and quality of big data continues to improve, there continues to be an increased impact. In 2021, the FDA released a draft guidance establishing the use of big data (claims and electronic health records) for regulatory decision-making [3]. Herein, we give an overview of the use of big data in ophthalmology to inspire further innovation and utilization, not just in our field, but across all specialties in medicine.

*To whom all correspondence should be addressed: Ehsan Rahimy, MD , Department of Ophthalmology, Palo Alto Medical Foundation, Palo Alto, CA; Email: erahimy@gmail.com.

Abbreviations: RWD, real-world data; EHR, electronic health record; RWE, real-world evidence; RCTs, randomized controlled trials; IRIS, The Intelligent Research in Sight Registry; VEGF, anti-vascular endothelial growth factor; DME, diabetic macular edema; nAMD, neovascular age-related macular degeneration; MIGS, microinvasive glaucoma surgery; AMD, age-related macular degeneration; GHI, global horizontal irradiance; DNI, direct normal irradiance; AI, artificial intelligence.

Keywords: big data, iris registry, ophthalmology, clinical trials, real-world data

FROM CLINICAL TRIALS TO BIG DATA

Prior to the widespread adoption of EHRs, the mainstay of robust clinical research in ophthalmology was through traditional, randomized controlled trials (RCTs). While RCTs remain a critical tool for evaluating the safety and efficacy of new treatments, they have limitations in ability to reflect real-world patient populations and treatment patterns due to limited sample sizes and relative homogeneity of patients due to strict inclusion and exclusion criteria. Often, patients in RCTs are not as diverse as the populations to which their answers are applied. For example, in an observational study of patients with treatment naïve neovascular age related macular degeneration, 44.7% did not meet eligibility criteria for a clinical trial whose conclusions on efficacy and safety of medical therapy in this population of patients were widely adapted [4]. In comparison, the use of RWD can enhance clinical trial outcomes by evaluating clinical questions on an expansive sample size in a heterogenous group that is likely more generalizable to other populations. In addition, RWD reflects the integration of both quantitative (measurable, scientific outcomes) data and non-quantifiable nuances (physician and patient preference, motivations, etc.) that influences patient outcomes. More so, randomized clinical trials are using registries to reduce the time and costs of recruitment of patients and collection of data [5]. Certainly, clinical trials remain the most rigorous form of medical research, but there is an expanding role for big data to complement RCTs and augment our understanding of trial results in the real-world setting.

THE START OF BIG DATA IN OPHTHALMOLOGY– THE IRIS REGISTRY

The concept of clinical registries can be traced back to over a century ago, when a disease-specific registry was developed to track outcomes associated with leprosy [6]. In the mid-20th century, the World Health Organization (WHO) formally defined clinical registries [7], and several were subsequently developed to monitor and track the incidence and prevalence of disease, such as cancer [8]. These early registries were often maintained by government agencies or research institutions and primarily served as surveillance tools.

Over time, the scope and purpose of clinical registries expanded to include monitoring of treatment patterns, outcomes, and quality of care. In the 1990s, the use of electronic health records (EHRs) became more widespread, and this enabled the development of more comprehensive and centralized clinical registries. In this space, ophthalmology has been a leader in the medical field. The Intelligent Research in Sight (IRIS) Registry, developed by the American Academy of Ophthalmology (AAO), was launched in 2014 [9,10]. Data from dozens of different EHRs are deidentified and aggregated into a centralized database from physicians across the United States. Specifically, the database includes data on patient demographics, medical and ocular history, clinical examination findings (including laterality, visual acuity, intraocular pressure, etc.), diagnoses, procedures, and medications. Currently, the database is not open-access but requires approved access that is limited and funding for research utilization. As of January 1, 2023, the IRIS Registry is the world's largest specialty clinical data registry with data from 78.6 million patients, over 438.6 million visits, contributed by 15,799 clinicians [11].

IRIS REGISTRY IN USE – VARIOUS ADVANTAGES OF BIG DATA

This use of RWD has been instrumental in advancing our understanding of real-world treatment outcomes in ophthalmology [12]. By providing a comprehensive database of patient outcomes and treatment patterns, the IRIS Registry has generated valuable insights into the real-world effectiveness of various ophthalmic treatments. Moreover, the use of RWD can also help identify areas where treatments may be underutilized, overutilized, or ineffective, and develop targeted interventions to address these issues. Below, we provide real examples of how big data has been used in ophthalmology. We present different methodologies and uses of RWD to hopefully inspire new applications across medical specialties.

Effectiveness of Current Treatments

Real-world outcomes of anti-vascular endothelial growth factor (VEGF) therapy for diabetic macular edema (DME) and neovascular age-related macular degeneration (nAMD) were evaluated. In one study, patients with DME were found to have better visual acuity outcomes if they received at least 6 monthly injections of anti-VEGF therapy compared to those who had fewer [13]. Similarly, another study found that patients who received at least 3 monthly injections of anti-VEGF had significantly better outcomes than those who had fewer in patients with nAMD [14]. It also found that longer treatment duration and adherence to injection schedules were associated with better outcomes. These findings have important implications for clinical practice, as they suggest more frequent anti-VEGF injections may be necessary to achieve optimal outcomes in patients with DME and highlight the need for regular and consistent follow up in patients with nAMD.

In a different realm of ophthalmology, microinvasive glaucoma surgery (MIGS) are newer surgical procedures/devices for the treatment glaucoma. Utilizing a sample

size of 79,363 eyes, it was found that MIGS were more effective with significantly lower reoperation rates when performed concurrently with phacoemulsification than when performed alone [15]. This demonstrates relatively real time results of emerging technologies on a large scale and informs clinician decision making when adapting new technologies and techniques.

Comparing Therapies

Big data can be used to compare the effectiveness of different medical and surgical treatments in real-world settings, complementing findings from randomized controlled trials. One study published in the journal *Ophthalmology* used RWD to compare the effectiveness of different anti-VEGF therapies – aflibercept, ranibizumab, and bevacizumab – for nAMD. The study found that monotherapy with each drug had similar effectiveness in real-world settings at one year despite differences reported in other studies [16].

Similarly, a study in 2021 compared one year success of two different surgical interventions for glaucoma from the IRIS Registry [17], trabeculectomy or tube, and compared the results to those from the Tube Versus Trabeculectomy RCT. While the two interventions had no difference in failure risk, when the results were compared between studies, there was significantly greater risk of tube failures in eyes from the IRIS Registry than the RCT. These discrepancies may in part be due to different practice patterns and greater severity of disease in the real world compared to the clinical trial and highlight the value big data can add in conjunction with RCTs.

Impact of Comorbidities and Environment on Patient Outcomes

An interesting study evaluated the association of environmental factors on exudative and nonexudative age-related macular degeneration (AMD) [18]. Provider zip codes were used to collect environmental data including elevation, latitude, global horizontal irradiance (GHI) and direct normal irradiance (DNI), temperature and precipitation variables, and pollution variables. The association of each variable was quantified in three models (AMD versus non-AMD, exudative versus nonexudative AMD, active exudative AMD versus inactive exudative, and nonexudative AMD). Researchers found solar variables and latitude were significantly associated with active exudative AMD while pollution variables were associated with any AMD.

In another study of 87,774 adults with Graves disease, it was found that current smokers more than former smokers had significantly increased risk of surgical intervention for thyroid eye disease compared to never smokers [19]. These studies demonstrate how big data can potentially impact public health and environment policies.

Identifying Health Disparities

Real-world data can be used to identify patient subgroups who may have different treatment outcomes based on demographic, socioeconomic, and/or other factors. Evaluation of 18,841 children aged 3 to 7 and 9,762 children aged 8 to 12 with amblyopia revealed there were differences in success rates of treatment [20]. Amblyopia treatment outcomes were significantly worse in children with Medicaid insurance and Black children at both age groups. In a study of adults with DME and treatment with anti-VEGF therapy, White race, non-Hispanic/Latino ethnicity, and private insurance were associated with more anti-VEGF injections and better visual acuity over the study period [21]. Although registry studies cannot provide data on the causality of disparities, it is an important step in identifying challenges certain groups face in the real world that can inform our advocacy and policy efforts.

Monitoring Safety Outcomes

RWD can be used to monitor the safety of treatments in real-world settings, allowing for the detection of rare or unexpected adverse events that may not have been detected in clinical trials. Shortly after a new anti-VEGF medication, brolucizumab, was approved by the US Food and Drug Administration (FDA) for the treatment of nAMD in 2019 [22], there were a plethora of reports of adverse events related to its use, specifically intraocular inflammation and retinal vasculitis [23,24]. A study using the IRIS registry evaluated the incidence of this adverse event in over 10,000 patients receiving the injection compared to the less than 2,000 in the clinical trials and found it was approximately 2.4% [25]. This is one example of how big data can be useful for evaluating the safety and adverse events after widespread use of a recently launched therapy.

Emulation of Clinical Trials in the Real World

A proof-of-concept study evaluated outcomes of anti-VEGF therapy in nAMD patients in the real world using the methodology of pivotal clinical trials, the Comparison of AMD Treatment Trials (CATT) [26]. The study matched patients (age, gender, baseline visual acuity) from the trial to treatment naïve patients in the IRIS registry to evaluate real-world outcomes and treatment patterns [27]. Patients in the real world, as compared to the clinical trial, received significantly fewer injections and worse visual acuity outcomes at 1 year after initiation of treatment. This study established a new methodology for using big data and patient level clinical trial data. Another analysis utilized the IRIS registry to replicate a widely

cited clinical trial for treatment of nAMD and was able to produce a study population size capable of emulating the primary endpoint [28]. This highlights another potential utilization of big data – to create synthetic control arms, ie, digital twins, to decrease time and cost of clinical trials.

Health Economic Analyses

Real-world data can be used to perform health economic analyses, which can inform reimbursement decisions, identify areas for cost savings, and help allocate healthcare resources more effectively. A study published in the journal *Ophthalmology* in 2020 used real-world data from the IRIS registry and Medicare claims to explore Medicare Part B and patient savings associated with increases in bevacizumab use compared to ranibizumab or aflibercept in four different models [29]. In this scenario, increasing bevacizumab reimbursement to $125.78 (equalizing the same dollar margin as the more expensive aflibercept) would potentially eliminate financial disincentive to its use and would save Medicare Part B $468 million along with $119 million in patient savings.

Inform Clinical Trial Design

There is a discrepancy in racial and ethnic demographic representation in ophthalmologic clinical trials in the US compared to the population distribution of the diseases studied [30]. RWD can be used to increase diversity in clinical trial populations by identifying underrepresented groups and designing trials that specifically target these groups. This is a long-standing problem across all RCTs in the US. Recently, the FDA released a draft guidance regarding the use of Race and Ethnicity Diversity Plan in clinical trials whereby they encourage leverage of RWD to help address these problems [31]. Real-world data can be used to identify gaps in current knowledge or areas where more research is needed, which can inform the design of clinical trials. RWD can also model effects of different eligibility criteria and trial endpoints to help enable more diversity and inclusivity in trial design.

These studies demonstrate the significant impact that the IRIS Registry has had on the field of ophthalmology and highlight the potential for clinical registries and real-world data to inform and improve clinical practice in all of medicine. This also demonstrates the different methodologies possible with analyzing big data.

CHALLENGES

While big data has big potential to transform health care value and complement more traditional methods of research, there are several challenges with its use. One of the main concerns is data quality. The accuracy and completeness of EHRs can vary widely between different providers and systems which then impacts the validity of the findings, hence the saying "garbage in, garbage out." In most database and registries research, patients are identified by International Statistical Classification of Disease and Related Health Problems (ICD) and Current Procedural Terminology (CPT) codes. Certainly, there are miscoding and billing errors with a magnitude unknown. The current systems also do not allow for free text extraction nor the ability to automate imaging analysis, which would further increase its value.

Another challenge is the need for standardized data collection and analysis methods, as different healthcare providers may use different EHR systems and data formats, making it difficult to compare and aggregate data across different practices. This can constrain contained variables of a database or registry. While one of the biggest strengths of big data is the sheer number of data points, this can also be a weakness. Statistical significance becomes less meaningful when we are discussing N of millions. Thus, it is imperative that clinical significance be stressed when designing queries interpreting analyses. Another perceived strength is the reflection of real-world data and utilization. However, the flip side must be recognized – there is no randomization in patient populations being compared in any of these analyses.

Finally, even though there is strength in numbers, there must be a clear understanding of the data being used, as the composition of any given database may still not be generalizable to the general population. The IRIS Registry has struggled to incentivize academic institutions to contribute data to the network. As of 2021, only 24% of Association of University Professors of Ophthalmology (AUPO) institutions were contributing to the database [32].

THE FUTURE OF BIG DATA

While the use of big data in medical research has dramatically increased over the past decade, we are still in the beginnings of realizing the potential of this new research method. Innovations in novel statistical approaches, standardization of datasets, and data sharing are all necessary to advance the utilization and impact. Going beyond phenotype, linkage between genotype (whole-exome sequencing, genome-wide genotyping, whole-genome sequencing) and RWD is a key step in advancing personalized medicine and tailoring treatment, prognosis, and outcomes to the individual. Another step forward is the successful integration of artificial intelligence (AI) and big data. Again, ophthalmology has been at the forefront of the application of machine learning and deep learning approaches thanks in part to our utilization of noninvasive imaging technology. Deep learning algo-

rithms to detect nAMD, DME, and drusen from retinal optical coherence tomography images [33] and an algorithm to predict visual field deficits from a single baseline visual field [34] are exciting examples of the possibilities.

CONCLUSION

The IRIS Registry represents an important step forward in the use of big data in medical research. The use of RWD approaches in ophthalmology has already led to significant improvements in patient outcomes and has the potential to drive further innovation in the field. As other medical specialties continue to adopt real-world data approaches and integrate AI, they can look to ophthalmology as a model for successful implementation and use of these technologies. As with any new approach to research, there are challenges and limitations, but the potential benefits are significant and warrant continued investment and development.

REFERENCES

1. Lohr S. The origins of "big data": an etymological detective story. The New York Times. Available at: https://bits.blogs.nytimes.com/2013/02/01/the-origins-of-big-data-an-etymological-detective-story/. 01.02.13. Accessed 04.05.23.
2. Lee CS, Brandt JD, Lee AY. Big Data and Artificial Intelligence in Ophthalmology: Where Are We Now? Ophthalmol Sci. 2021 Jun;1(2):100036.
3. US Food and Drug Administration. Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products: Draft Guidance for Industry. December 2022. Accessed April 5, 2023. Available at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory
4. Kim JH, Kim JW, Kim CG. Eyes that Do Not Meet the Eligibility Criteria of Clinical Trials on Age-Related Macular Degeneration: Proportion of the Real-World Patient Population and Reasons for Exclusion. J Ophthalmol. 2021 Apr;2021:6635467.
5. Lauer MS, D'Agostino RB Sr. The randomized registry trial—the next disruptive technology in clinical research? N Engl J Med. 2013 Oct;369(17):1579–81.
6. Irgens LM. The origin of registry-based medical research and care. Acta Neurol Scand Suppl. 2012;126 195:4–6.
7. WHO Expert Committee on Health Statistics & World Health Organization. Epidemiological methods in the study of chronic diseases: eleventh report of the WHO Expert Committee on Health Statistics [meeting held in Geneva from 15 to 21 Nov 1966]World Health Organization; 1967.
8. Bleicher RJ, Ruth K, Sigurdson ER, Beck JR, Ross E, Wong YN, et al. Time to surgery and breast cancer survival in the United States. JAMA Oncol. 2016 Mar;2(3):330–9.
9. Chiang MF, Sommer A, Rich WL, et al. The 2016 American Academy of Ophthalmology IRIS Registry (Intelligent Research in Sight) database: characteristics and methods. Ophthalmology. 2018;125(8):1143e1148.
10. Parke II DW, Lum F, Rich WL. The IRIS Registry: Purpose and perspectives. Ophthalmologe. 2017;114(Suppl 1):1e6.
11. American Academy of Ophthalmology. Iris Registry Data Analytics. The American of Ophthalmology web site. https://www.aao.org/iris-registry/data-analysis/requirements. Accessed: 04.05.23.
12. Parke DW 2nd, Rich WL 3rd, Sommer A, Lum F. The American Academy of Ophthalmology's IRIS® Registry (Intelligent Research in Sight Clinical Data): A Look Back and a Look to the Future. Ophthalmology. 2017 Nov;124(11):1572-1574.
13. Chandra S, Aronow ME, Peden MC, et al. Anti-VEGF Use and Diabetic Macular Edema: A Time-Course Analysis of the American Academy of Ophthalmology Intelligent Research in Sight Registry. Ophthalmology. 2016;123(11):2300–6.
14. Holekamp NM, Liu Y, Yeh WS, et al. Clinical utilization of anti-VEGF agents and disease monitoring in neovascular age-related macular degeneration. JAMA Ophthalmol. 2018;136(9):941–5.
15. Yang SA, Ciociola EC, Mitchell W, Hall N, Lorch AC, Miller JW, et al.; IRIS® Registry Analytic Center Consortium. Effectiveness of microinvasive glaucoma surgery in the United States: Intelligent Research in Sight Registry analysis 2013-2019. Ophthalmology. 2023 Mar;130(3):242–55.
16. Rao P, Lum F, Wood K, Salman C, Burugapalli B, Hall R, et al. Real-World Vision in Age-Related Macular Degeneration Patients Treated with Single Anti-VEGF Drug Type for 1 Year in the IRIS Registry. Ophthalmology. 2018 Apr;125(4):522–8.
17. Vanner EA, Sun CQ, McSoley MJ, Persad PJ, Feuer WJ, Lum F, et al. Tube Versus Trabeculectomy IRIS® Registry 1-Year Composite Outcome Analysis with Comparisons to the Randomized Controlled Trial. Am J Ophthalmol. 2021 Jul;227:87–99.
18. Hunt MS, Chee YE, Saraf SS, Chew EY, Lee CS, Lee AY, et al. Association of environmental factors with age-related macular degeneration using the Intelligent Research in Sight Registry. Ophthalmol Sci. 2022 Jul;2(4):100195.
19. Oke I, Reshef ER, Elze T, et al. Smoking is associated with a higher risk of surgical intervention for thyroid eye disease in the IRIS® Registry. Am J Ophthalmol. 2023;S0002-9394(23)00029-6.
20. Repka MX, Li C, Lum F. Multivariable Analyses of Amblyopia Treatment Outcomes from a Clinical Data Registry. Ophthalmology. 2023 Feb;130(2):164–6.
21. Greenlee TE, Malhotra NA, Iyer AI, Conti TF, Chen AX, Singh RP. Association of socioeconomic health care disparities with use of anti-vascular endothelial growth factor and visual acuity outcomes in patients with diabetic macular edema. Ophthalmic Surg Lasers Imaging Retina. 2022 Jul;53(7):380–91.
22. Novartis Pharmaceuticals Corporation. Beovu package insert. Published; 2020.
23. Baumal CR, Spaide RF, Vajzovic L, Freund KB, Walter SD, John V, et al. Retinal vasculitis and intraocular

inflammation after intravitreal injection of brolucizumab. Ophthalmology. 2020 Oct;127(10):1345–59.
24. Jain A, Chea S, Matsumiya W, et al.. Severe vision loss secondary to retinal arteriolar occlusions after multiple intravitreal brolucizumab administrations. Am J Ophthalmol Case Rep. 2020;18:100687.24. https://doi.org/10.1016/j.ajoc.2020.100687.
25. Khanani AM, Zarbin MA, Barakat MR, Albini TA, Kaiser PK, B G, Agashivala N, Yu JS, Wykoff CC, MacCumber MW. Safety Outcomes of Brolucizumab in Neovascular Age-Related Macular Degeneration: Results From the IRIS Registry and Komodo Healthcare Map. JAMA Ophthalmol. 2022 Jan 1;140(1):20-28.
26. Martin DF, Maguire MG, Fine SL, Ying GS, Jaffe GJ, Grunwald JE, et al.; Comparison of Age-related Macular Degeneration Treatments Trials (CATT) Research Group. Ranibizumab and bevacizumab for treatment of neovascular age-related macular degeneration: two-year results. Ophthalmology. 2012 Jul;119(7):1388–98.
27. Borkar DS, LaPris A, Fevrier H, Torres AZ, Mbagwu M, Lum F, Leng T. Clinical Trial Emulation of the Comparison of AMD Treatment Trials (CATT) Using Data from IRIS Registry. Presented at: Duke Advanced Vitreous Surgery Course. Durham, NC. March 25, 2023.
28. Gallivan MD, Garcia KM, Torres AZ, Lum F, Li C, Mbagwu M, et al. Emulating VIEW1 and VIEW2 clinical trial outcome data using the American Academy of Ophthalmology IRIS Registry. Ophthalmic Surg Lasers Imaging Retina. 2023 Jan;54(1):6–14.
29. Glasser DB, Parikh R, Lum F, Williams GA. Intravitreal Anti-Vascular Endothelial Growth Factor Cost Savings Achievable with Increased Bevacizumab Reimbursement and Use. Ophthalmology. 2020 Dec;127(12):1688–92.
30. Kaakour AH, Hua HU, Rachitskaya A. Representation of Race and Ethnicity in Randomized Clinical Trials of Diabetic Macular Edema and Retinal Vein Occlusion Compared to 2010 US Census Data. JAMA Ophthalmol. 2022 Nov;140(11):1096–102.
31. US Food and Drug Administration. Diversity plans to improve enrollment of participants from underrepresented racial and ethnic populations in clinical trials; draft guidance for industry. April 2022. Accessed April 05, 2023. https://www-fda-gov.laneproxy.stanford.edu/regulatory-information/search-fda-guidance-documents/diversity-plans-improve-enrollment-participants-underrepresented-racial-and-ethnic-populations
32. Weng CY, Kim JE. When the N Is 65 Million-The Rise of Database Research. JAMA Ophthalmol. 2021 Jun;139(8):855–6.
33. Faes L, Wagner SK, Fu DJ, et al. Automated deep learning design for medical image classification by health-care professionals with no coding experience: a feasibility study. Lancet Digit Health. 2019;1(5):e232ee242. https://doi.org/10.1016/S2589-7500(19)30108-6.
34. Wen JC, Lee CS, Keane PA, Xiao S, Rokem AS, Chen PP, et al. Forecasting future Humphrey Visual Fields using deep learning. PLoS One. 2019 Apr;14(4):e0214875.