Review Article

Taiwan J Ophthalmol 2023;13:123-132

Access this article online

Quick Response Code:

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

DOI:
10.4103/tjo.TJO-D-23-00012

Application of big data in ophthalmology

Zhi Da Soh[1,2], Ching-Yu Cheng[1,2,3,4*]

Abstract:

The advents of information technologies have led to the creation of ever-larger datasets. Also known as *big data*, these large datasets are characterized by its volume, variety, velocity, veracity, and value. More importantly, big data has the potential to expand traditional research capabilities, inform clinical practice based on real-world data, and improve the health system and service delivery. This review first identified the different sources of big data in ophthalmology, including electronic medical records, data registries, research consortia, administrative databases, and biobanks. Then, we provided an in-depth look at how big data analytics have been applied in ophthalmology for disease surveillance, and evaluation on disease associations, detection, management, and prognostication. Finally, we discussed the challenges involved in big data analytics, such as data suitability and quality, data security, and analytical methodologies.

Keywords:

Artificial intelligence, big data, machine learning, ophthalmology, precision medicine

[1]Singapore National Eye Centre, Singapore Eye Research Institute, [2]Department of Ophthalmology, Yong Loo Lin School of Medicine, National University of Singapore, [3]Centre for Innovation and Precision Eye Health, Yong Loo Lin School of Medicine, National University of Singapore, [4]Ophthalmology and Visual Sciences Academic Clinical Program, Duke-NUS Medical School, Singapore

***Address for correspondence:**
Dr. Ching-Yu Cheng, The Academia, Discovery Tower Level 6, 20 College Road, 169856 Singapore.
E-mail: chengyu.cheng@duke-nus.edu.sg

Submission: 20-01-2023
Accepted: 02-04-2023
Published: 23-05-2023

Introduction

The advents in information technologies (IT), such as the internet-of-things (IoT) and artificial intelligence (AI), have fundamentally changed the way we live and work. The IoT allows data from interconnected digital devices (e.g., smartwatches and electronic medical records [EMR]) to be amalgamated seamlessly in real time, leading to the creation of ever larger datasets or *big data*.[1]

Big data refer to the rapid aggregation of a large amount of diverse and constantly changing data points that are too complex or "big" to be handled by traditional methods.[1] In general, big data is characterized by its volume (how big), variety (how diverse), velocity (how fast), veracity (how accurate), and value (how useful).[2] Although "big" is emphasized in big data, the sheer amount of data *per se* do not provide significant advantages. Instead, it is the ability to draw in-depth knowledge from big data that is germane.

In this aspect, the advents of AI, especially its sub-domains of machine learning (ML) and deep learning (DL), have been instrumental in translating the messy *"sea of information"* in big data into meaningful and actionable insights.[3,4] For example, although conventional statistical methods remain imperative in analyzing structured quantitative data, they are unable to analyze unstructured data like text recordings in medical records and scan images. In contrast, ML techniques such as natural language processing and convolutional neural network were developed specifically to analyze free text and images respectively.[5,6] This allows the full spectrum of data collected to be analyzed simultaneously to draw insights that were not possible previously. Consequently, the application of AI techniques to big data are seen as the backbone of mankind's fourth industrial revolution.[7]

How to cite this article: Soh ZD, Cheng CY. Application of big data in ophthalmology. Taiwan J Ophthalmol 2023;13:123-32.

Benefits of Big Data Analytics

The ability to leverage big data may bring about significant benefits to biomedical research, clinical practice, and health system strengthening.[8,9] First, big data analytics may expand the boundaries of traditional research methodologies and enhance our ability to generate new knowledge. Big data are better powered statistically to generate scientific insights into hypotheses that may otherwise by unanswered due to the prohibitive cost of primary data collection or answered inadequately due to limited funding and sample size.[10,11] This includes detection of novel biomarkers, analysis of subtle or inconclusive risk factors, longer term observation on pharmacodynamics, and greater understanding of disease pathogenesis through the inclusion of socio-economical, environmental, and molecular data.[9,12] In addition, the diversity of big data further aids in improving the generalizability of findings.[11,12]

Second, big data analytics may inform and improve clinical practice through the development of sophisticated algorithms.[1] These algorithms may come in the form of a screening tool for timely detection, or as a decision support tool that provides diagnostics and/or therapeutic suggestions based on real-time analysis of aggregated inputs from fellow physicians and/or other resources.[9] As a result, these algorithms aid in harmonizing the standard-of-care among practitioners.[10]

Third, big data analytics may be utilized to identify gaps and evaluate the quality and efficiency of public health policies and healthcare delivery.[12] For example, big data analytics is projected to create a value of more than US$300 billion annually for the United States (US) healthcare system, the majority of which would come in the form of reduced healthcare expenditure.[13] Furthermore, big data obtained from digital devices, such as smart watches, are now delivering health information directly to individuals. This not only empowers individuals to play a more active role in managing their health but may also alter the way in which health-care services are sought and delivered.[10]

Sources of Big Data

Ophthalmology is well-placed to benefit from the insights curated from big data analytics due to the sheer amount of data generated in clinical and research settings. This includes clinical and surgical notes, pharmacology records, reimbursement claims, test measurements (e.g., refractive error, intra-ocular pressure), two-dimension images (e.g., fundus photographs), and three-dimension scans (e.g., optical coherence tomography [OCT]).

Electronic medical records and data registry

Medical records and auxiliary test results are increasingly digitalized into electronic format (i.e., EMR) in healthcare settings across the world.[14] The advents of IoT further enable different EMRs and databases to be linked automatically in real-time.[15] This creates a "one-stop" portal and data registry that allows physicians to track the pattern and effectiveness of care, administrators to identify gaps and efficiency in service delivery, and researchers in analyzing disease trends in real-world settings.

Intelligent Research in Sight (IRIS) is a cloud-based ophthalmic data registry developed in 2014 by the American Academy of Ophthalmology.[16] The aim of IRIS was to improve the provision of eye care services, promote population health through adequate eye coverage, and generate evidence-based scientific knowledge.[17] Clinical data from participating clinics are aggregated automatically in real-time and comprise fifteen control measures and 22 outcome measures from over sixty million patients.[17]

The Sight Outcomes Research Collaborative (SOURCE) ophthalmic data registry was initiated by various academic ophthalmology institutions across the US to share de-identified EMR and diagnostic test data for research and quality improvement projects.[18]

The Save Sight Registries (SSR) is made up of different specific registries, such as the Fight Corneal Blindness!, Fight Glaucoma Blindness!, Fight Tumor Blindness!, and Fight Retinal Blindness (FRB!).[19] The FRB! is the flagship database of SSR and was developed in 2009 to track data on the outcomes of retinal diseases (e.g., age-related macular degeneration and macular edema) from Asia, Europe, and the Middle East.[20] The FRB! further incorporates data from observational studies to establish treatment regimens that are feasible and effective for routine clinical practice. This is unlike treatment regimens used in pivotal clinical trials where patients and practitioners are unlikely to comply even if they wanted to.[21]

Administrative database

In healthcare, administrative and insurance databases provide a vital source of information for epidemiology, pharmacoepidemiologic and health economic evaluations.[12] For example, insurance databases have been used to identify surgery trend and safety profiles of ophthalmic drugs.

In Europe, the EPISAFE collaboration program utilized the French national health insurance database, the système national d'information interrégimes de l'assurance maladie, to evaluate the epidemiology and safety of interventions used in ophthalmology.[22] Other administrative databases used for similar evaluation in the West include the US Medicare,[23] the UK clinical practice research datalink,[24] and the régie d'assurance maladie du québec in Canada.[25]

In Asia, data from the national health insurance program in South Korea and Taiwan are frequently utilized for research purpose.[26,27] In South Korea, a database was created to include 2% (~1 million) of data from the Korean National Health Insurance Service, along with other cohort studies to provide de-identified data on claims, health screening, and mortality.[27]

Research consortium

Research consortium or network is collaborative initiatives that bring researchers across different domains and/or countries together in a shared platform to build and share research capabilities. In epidemiology, consortia research provides an aggregated view of the burden of diseases and its impact in a particular geographical region. In addition, the increased statistical power from combined databases is often used to evaluate research questions that are answered inadequately by individual groups.

Internationally, the Vision Loss Expert Group (VLEG), which comprises of 78 leading ophthalmologists, optometrists, and epidemiologists from across the world, was formed by the Global Burden of Disease in 2007.[28] The aim of VLEG was to conduct retrospective and prospective, consistent, and comparative systematic reviews on the burden of disease, injuries, and risk factors due to vision impairment. Other international consortium includes the Meta-analysis for Eye Disease study group,[29] the International Rare Disease Research Consortium,[30] and the International Eye Disease Consortium.[31]

In Europe, the European Eye Epidemiology (E3) consortium consists of 29 study groups from twelve European countries.[32] This includes population-based studies, such as the Rotterdam study from the Netherlands, and the Guttenberg Health study from Germany. The E3 consortium was set up to promote research collaboration and sharing of data in Europe, and to focus on standardizing methods for future research.

In Asia, the Asian Eye Epidemiology Consortium (AEEC) is a collaborative network of forty population-based study groups from nine Asian countries.[33] This includes the Beijing Eye Study from China and the Singapore Epidemiology or Eye Diseases study from Singapore.[34] The overall aim of AEEC was to utilize big data analytics to generate deeper insights into the trends and associated risk factors of major age-related eye diseases among Asians.

Biobank

Handling of biospecimens has evolved from storage in a few freezers and manual handling, to large repositories with computerized databases and robotic processing of samples. These advancements led to the emergence of biobanks, which may include biological samples from epidemiology studies, clinical trials, and diagnostic studies.

The UK biobank is a large-scale multi-site cohort study that was established to investigate the effects of genetic, lifestyle, and environmental risk factors on a wide range of diseases.[35,36] The UK biobank eye and vision data include phenotypic data, biomarker variables, dense genotyping, and lifestyle variables, as well as a large collection of fundus photographs and OCT scan images. The open-access nature of the UK biobank allows comparative research to be conducted, and the rich diversity of data allows for the evaluation of novel disease etiology and biomarkers.

Importantly, data from the UK biobank has been made available for application with a fee, unlike the other sources described above. The application includes selecting the data fields required, and indicating the personnel with access to the data.[37] Thereafter, a material transfer agreement will be initiated along with a fee based on the number and type of data fields applied. In addition, there are also various open-source databases made available for ophthalmology research.[38] This includes fundus photographs from the Asia Pacific Tele-Ophthalmology Society (n = 5590 images) and Eye Picture Archive Communication System (n = 88,702), and OCT scans from Duke OCT (n = 38,400) and Kermany (n = 109,312).[38]

Hybrid databases

Big data from different sources may not be confined solely to one of the above categories. For example, the Vision and Eye Health Surveillance System in the US incorporates data from the IRIS data registry, along with other national surveys, population-based studies, and administrative databases.[39] This virtual surveillance system was initiated to estimate the prevalence of vision loss and eye diseases at the local and national levels, identify disparities in access to eye care, monitor trends of eye disease, and promote eye health education.

Application of Big Data

Big data analytics is gaining increased prominence in healthcare and has been applied in eye care for disease surveillance, and evaluation on disease associations, detection, management, and prognostication [Figure 1].

Service utilization and improvement

Big data analytics have been applied to analyze the utilization of eye care services, and the profile of patients seeking eye care services. In India, the L. V. Prasad Eye Institute developed and utilized an EMR system to evaluate the distribution of patient workload, demographic characteristics of patients, and the type and frequency of ocular diseases seen in its operation network.[40,41]

This system has also been used to evaluate the profile and magnitude of diabetic retinopathy (DR) in its patient pool,[42] and in assessing the biogeographical distribution of senile cataracts and its association with environmental factors, such as terrain altitude and ultraviolet exposure.[43] These analyses show that although there was gender equality in seeking eye care services,[40] increased provision of eye care services in high altitude terrains may be needed to improve the detection of senile cataracts.[43] Furthermore, anterior segment diseases and refractive errors comprised two-thirds of consultations, suggesting the need for more resource allocation in these domains.[40] As a result, this application informed administrators and physicians on the gaps in service provision, barriers to access eye care services, and resource planning.

In addition, data registries have been used to inform and improve operational processes. For example, SOURCE data were used to develop an algorithm that could search for ocular diseases based on both structured and unstructured data instead of relying solely on billing codes.[44] This algorithm was subsequently tested on its effectiveness in searching for pseudo-exfoliation syndrome and achieved a positive predictive value of 95% and negative predictive value of 100%. Crucially, 60% of cases identified would have been missed if the algorithm had relied solely on billing codes. The SOURCE data were further used to develop an algorithm to triage patient appointments during the COVID-19 pandemic.[45] This algorithm weighted the risk of disease progression due to delayed care (i.e., postponement of appointment) to the morbidity risk of COVID-19. Consequently, this algorithm was not only applied to identify cases that could be safely postponed during the

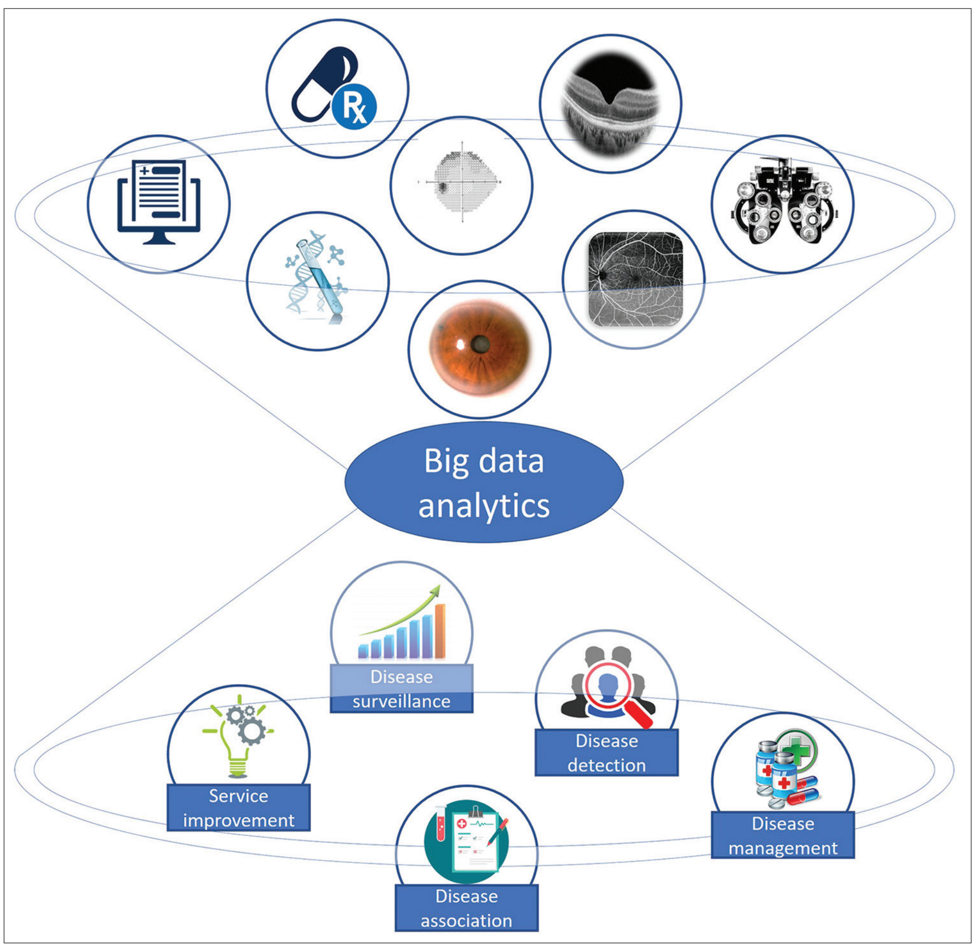

Figure 1: Input and output of big data analytics

pandemic but also in prioritizing appointments during the reopening phase.

Likewise, a patient appointment system from a university hospital EMR was used in simulation analyses, and the best model derived was found to reduce the total waiting time of patients by 21% upon implementation.[46] Furthermore, EMR data from a glaucoma clinic have been used to establish reference values for monitoring glaucoma in a virtual clinic.[47]

Disease surveillance

Several big data analytics have been applied to evaluate the burden and impact of eye diseases on various populations. For example, EMR data from 28 optometry centers were combined with an administrative database of a large European lens manufacturer to estimate the distribution of refractive error (Rx).[48] When matched by age and gender, estimates obtained from this approach were comparable to those reported by the E3 consortium, suggesting the viability of using these big data sources as an alternative to population surveys.

The E3 consortium further reported that Rx affected slightly over half of all European adults with the greatest burden being myopia.[49] In addition, the prevalence of myopia was found to be higher in later birth cohorts and was associated with higher education levels.[50] Likewise, the prevalence of high myopia and high progressive myopia was reported to impose a relatively high burden on adults in the US based on data triangulated from the IRIS registry, US population census, and NHANES database.[51]

The VLEG utilized further aggregated data and estimated that 65% of blind and 76% of moderate and severe visual impairment (VI) cases were either preventable or treatable.[52] This big data analysis further identified the higher risk of blindness among women, and the increasing risk of blindness due to Rx and age-related diseases (e.g., age-related macular degeneration, glaucoma). The VLEG has further reported on the effective coverage rate for Rx globally,[53] as well as the prevalence and cause of blindness and VI in various geographical regions.[54-57]

Disease association

Big data analytics have also been applied to evaluate the socioeconomical, systemic, and genetic risk factors of eye diseases. For example, data from the UK biobank found that moderate VI was associated with older age, and observed more in females and ethnic minorities.[58] In addition, all causes of VI were associated with poorer social outcome measures, as well as impaired general and mental health.[59] These findings highlight the importance of considering non-clinical variables in the clinical course of eye diseases and the comorbidities of VI.

The genetic data available in UK biobank has also been used, either in silos or in combination with other genetic databases, to assess the risk of age-related eye diseases.[60-63] These databases include the Australian and New Zealand Registry of Advanced Glaucoma, the National Eye Institute Glaucoma Human Genetics Collaboration consortium, and the International Glaucoma Genetic consortium.[64] For example, in glaucoma, the genome-wide analysis identified 101 significant single nucleotide polymorphisms associated with intraocular pressure, and the top decile of allele score was associated with a 5.6-fold increase in odds of glaucoma.[65] Using genome-wide polygenic risk score (PRS), the prevalence of primary angle closure glaucoma (PACG) was observed to increase with each decile of higher PRS, and the use of psychotropic medication was further associated with a higher risk of PACG at each decile of PRS.[61] In a separate work, the top PRS decile reached an absolute risk for glaucoma 10 years earlier than the bottom decile, and had a 15-fold increase in the risk of developing advanced glaucoma.[60] Separately, the AEEC has aggregated data to evaluate findings that were inconclusive from individual studies. For example, the consortium's meta-analysis suggested the association between chronic kidney disease and primary open-angle glaucoma (POAG) may be present only among East Asians.[66] In addition, a separate meta-analysis from AEEC confirmed the inverse association between body mass index and DR reported in three previous studies that were not included in their meta-analysis.[67] When analyzed individually, this novel but rather controversial finding may be downplayed due to the lack of statistical power. However, the finding from AEEC increased credibility and the importance of further evaluation.

In addition, AEEC reported different normative distributions of retinal nerve fiber layer (RNFL) among Asians and suggested the need for population-specific normative databases.[33] The E3 consortium further reported on the association between systemic vascular and neurovascular diseases and reduced peripapillary RNFL thickness,[68] while data from the UK biobank suggested an association between thinner RNFL and poorer cognitive function.[69]

Disease detection

Big data is particularly useful in the development of deep-learning algorithms for disease detection. In ophthalmology, DL algorithms are often developed to detect diseases, such as DR and glaucoma, from ocular images.[70,71] For example, SELENA + is a DL algorithm that was approved in Singapore for screening DR. This algorithm was developed using close to half a million fundus photographs from the Singapore National DR Screening Programme and ten cohort studies.[72]

In addition, EMR data from eight ophthalmic centers and data from two cohort studies were utilized to develop an ML algorithm to predict the development of high myopia among school children in China.[73] This ML algorithm achieved an area-under-the-curve of >0.80 in predicting the onset of high myopia (defined as spherical equivalent [SE] ≤-6D) in 3, 5 and 8 years, as well as the onset of high myopia at 18 years old in both internal testing and external validation. Furthermore, 95% of predicted Rx by the algorithm were within 0.50–0.80D of the true SE measured at year 8.

Disease management

Big data obtained from real-world registries are often used to analyze the pattern of care among fellow physicians in real-world practice. For example, data from IRIS showed that 73% of myopic choroidal neovascularization cases were treated within 1 year of diagnosis, of which 99.3% were treated with anti-vascular endothelial growth factor (VEGF) injections.[51] The FRB! Data further showed that the prescription rate for Ranibizumab and Aflibercept was similar among physicians in Australia although the former was prescribed more often in older patients while the latter was in eyes with larger lesions.[74]

In addition to management pattern, longitudinal analysis of prescription trends has also been reported. For example, FRB! data showed that the use of macular lasers and intra-vitreal triamcinolone in treating diabetic macular edema (DME) declined progressively after 2009.[75] This decline coincided with the shift in preference towards anti-VEGF injection. By 2015, 99% of DME were treated with anti-VEGF, with the choice of anti-VEGF changing from Bevacizumab (2009–2011) to Ranibizumab (2012–2015) and Aflibercept from 2016 onwards.

Furthermore, pattern-of-care in academic and non-academic settings has also been compared using big data from data registry. For example, data from IRIS showed more black patients and more severe cases of POAG cases were seen in academic settings.[76] Gonioscopy, pachymetry, and VF testing were performed more often in academic settings, as well as shunt procedures as compared to microinvasive glaucoma surgery and endoscopic cyclophotocoagulation were preferred in nonacademic settings. Such analyses not only summarize the changing of pattern-of-care over time but also highlight disparities in care provision. For example, gonioscopy remained under-performed in nonacademic settings despite the continued emphasis on the AAO’s preferred practice patterns.

Treatment outcome

Big data analytics have been applied extensively to evaluate the effectiveness and safety of interventions in real-world practice, especially in cataract surgery and anti-VEGF treatment. For example, the SOURCE data showed that mean signed prediction errors in the modern intra-ocular lens (IOL) formulas were significantly affected by gender, with more hyperopic prediction in males and vice versa for females.[77] This suggested the need for gender consideration in optimizing lens constantly to reduce prediction errors in formulas such as SRK/T and Hoffer Q. Separately, data from IRIS showed that 1.3% of monofocal toric IOL implants required repositioning in 1st year postsurgery, with younger adults at higher risk.[78] The risk of repositioning was also higher in TECNIS (3.1%) as compared to Acrysof toric IOL (0.6%). Furthermore, analysis on IRIS and the US Medicare database showed that the rate of endophthalmitis was between 0.08% and 0.14%, respectively.[79] The risk of endophthalmitis 4 weeks’ postsurgery was similar between people who underwent sequential cataract surgery in both eyes and those with cataract surgery delayed by ≥1 day in the second eye.[80]

In anti-VEGF treatment, data from FRB! showed that eyes treated with Bevacizumab initially before switching to either Ranibizumab and Aflibercept over a 1-year period did not improve visual outcomes despite further reduction in macular thickness.[81] Similarly, data from IRIS showed that all 3 types of anti-VEGF improved visual acuity similarly in neovascular age-related macular degeneration over 1 year of mono-therapy,[82] while FRB! data further showed similar vision outcomes over 3 years.[83] In addition, similar vision outcome was also observed at month 12 and 24 in fixed bimonthly and treat-and-extend regimen.[84] Nonetheless, higher rates of non-infectious endophthalmitis were observed with Bevacizumab as compared to Ranibizumab and Aflibercept.[85]

Interpretation and Consideration

Although significant benefits can be derived from big data analytics, careful consideration on the infrastructures and processes needed to adopt big data and understanding the potential limitations and biases in results interpretation is imperative.

Data quality and suitability

First, the aggregation of large and diverse data is inherently messy, especially if appropriate systems and handling protocols are not in place.[86,87] This raises questions with regard to the quality and suitability of big data for analytical purposes. For example, EMR was not developed for research purposes, and recording in a standardized format is not mandatory, leading to incomplete documentation and difficulties in combing disparate data.[12,88] In addition, the majority of EMR data are likely to be unstructured (e.g., free text), which may be missed in systems that rely solely or heavily on diagnostic codes, such as the international classification of diseases codes, during data extraction.[88,89] In addition, misclassification, error in coding, and inadequate representation of diseases may happen during documentation.[89]

Data security

Second, data security is a major concern in big data analytics, and proper data governance and ethics are imperative to build trust in using this tool.[87] For example, protocols and audit trail to ensure only de-identified data are used for evaluation are needed to preserve data privacy.[12] Furthemore, these large databases are attractive targets for cyber theft.[90] Thus, a secured and scalable data security network, along with protocols to handle cyber threats, must be in place beforehand.

Data analysis

Third, the results obtained from bigger data analysis should be interpreted with caution. For instance, the sheer amount of data in big data analytics inherently results in smaller *P* values, thereby indicating statistical significance.[91] However, considerations of the clinical significance or implications remain vital during interpretation.[91] In addition, it is important to look out for undesirable practices, such as performing multiple testing to obtain p-significant outcomes, in big data analytics.[92] Also known as p-hacking or p-fishing, these practices increase the risk of false-positive results that are not only not reproducible but also misleading.[93]

Furthermore, common research considerations, such as confounding, selection and measurement bias, or reverse causation, are not eliminated by simply using big data.[93] The adage “garbage in, garbage out” remains relevant in big data analytics. Thus, careful consideration of study methodology remains imperative in mitigating these shortcomings. For example, EMR data depend on the catchment area of the institution and may not appropriately or adequately represents the general population for epidemiology evaluation.[12] Similarly, EMR from a specialized institution contains much higher risk of detection bias as compared to a general practice, and as such, may be more appropriate in analyzing risk factors rather than the burden of diseases. Nonetheless, study design and statistical methods to mitigate these errors are available. This includes the use of propensity score adjustment, and the use of sensitivity and stratified analyses when appropriate.[94,95]

Conclusion

Ophthalmology is well-placed to benefit from big data. Multiple sources of big data already exist and are increasingly utilized to expand research capabilities, inform clinical practice, and improve service provision. Nonetheless, big data must be harnessed systematically in safe and secured infrastructures, and proper consideration for data suitability and quality, as well as analytical methodologies, are imperative.

Financial support and sponsorship

Nil.

Conflicts of interest

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

References

1. Panesar A. Machine Learning and AI for Healthcare. New York, United States: Springer; 2019.
2. Baro E, Degoul S, Beuscart R, Chazard E. Toward a literature-driven definition of big data in healthcare. Biomed Res Int 2015;2015:639021.
3. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-44.
4. Bote-Curiel L, Munoz-Romero S, Gerrero-Curieses A, Rojo-Álvarez JL. Deep learning and big data in healthcare: A double review for critical beginners. Appl Sci 2019;9:2331.
5. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: An introduction. J Am Med Inform Assoc 2011;18:544-51.
6. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, *et al*. Recent advances in convolutional neural networks. Pattern Recognit 2018;77:354-77.
7. Forum WE. The Fourth Industrial Revolution: What it Means, How to Respond; 2016. Available from: https://wwwweforum.org/agenda/2016/01/the-fourth-industrial-revolutionwhat-it-means-and-how-to-respond/. [Last accessed on 2023 Mar 04].
8. Shilo S, Rossman H, Segal E. Axes of a revolution: Challenges and promises of big data in healthcare. Nat Med 2020;26:29-38.
9. Mehta N, Pandit A. Concurrence of big data analytics and healthcare: A systematic review. Int J Med Inform 2018;114:57-65.
10. Murdoch TB, Detsky AS. The inevitable application of big data to health care. JAMA 2013;309:1351-2.
11. Boland MV. Big data, big challenges. Ophthalmology 2016;123:7-8.
12. Clark A, Ng JQ, Morlet N, Semmens JB. Big data and ophthalmic research. Surv Ophthalmol 2016;61:443-65.
13. Manyika J, Chui M, Brown B, Bughin J, Dobbs R, Roxburgh C, *et al.* Big Data: The Next Frontier for Innovation, Competition, and Productivity. New York, United States: McKinsey Global Institute; 2011.
14. Sheikh A, Bates DW, Wright A, Cresswell K. Key Advances in Clinical Informatics: Transforming Health Care through Health Information Technology. Massachusetts, United States: Academic Press; 2017.
15. Dash S, Shakyawar SK, Sharma M, Kaushik S. Big data in healthcare: Management, analysis and future prospects. J Big Data 2019;6:1-25.
16. Chiang MF, Sommer A, Rich WL, Lum F, Parke DW 2[nd]. The 2016 American Academy of Ophthalmology IRIS(®)

registry (intelligent research in sight) database: Characteristics and methods. Ophthalmology 2018;125:1143-8.
17. 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;124:1572-4.
18. SOURCE; 2022. Available from: https://www.sourcecollaborative.org/. [Last accessed on 2023 Mar 03].
19. SSR; 2022. Available from: https://savesightregistries.org/. [Last accessed on 2023 Mar 03].
20. SSR; 2022. Available from: https://savesightregistries.org/publication_category/retina/. [Last accessed on 2023 Mar 03].
21. Nguyen V, Barthelmes D, Gillies MC. Neovascular age-related macular degeneration: A review of findings from the real-world Fight Retinal Blindness! Registry. Clin Exp Ophthalmol 2021;49:652-63.
22. Daien V, Korobelnik JF, Delcourt C, Cougnard-Gregoire A, Delyfer MN, Bron AM, *et al.* French medical-administrative database for epidemiology and safety in ophthalmology (EPISAFE): The EPISAFE collaboration program in cataract surgery. Ophthalmic Res 2017;58:67-73.
23. Rector TS, Wickstrom SL, Shah M, Thomas Greeenlee N, Rheault P, Rogowski J, *et al.* Specificity and sensitivity of claims-based algorithms for identifying members of Medicare+Choice health plans that have chronic medical conditions. Health Serv Res 2004;39:1839-57.
24. Jameson K, D'Oca K, Leigh P, Murray-Thomas T. Adherence to NICE guidance on glucagon-like peptide-1 receptor agonists among patients with type 2 diabetes mellitus: An evaluation using the clinical practice research datalink. Curr Med Res Opin 2016;32:49-60.
25. Dixon WG, Abrahamowicz M, Beauchamp ME, Ray DW, Bernatsky S, Suissa S, *et al.* Immediate and delayed impact of oral glucocorticoid therapy on risk of serious infection in older patients with rheumatoid arthritis: A nested case-control analysis. Ann Rheum Dis 2012;71:1128-33.
26. Cheng TM. Taiwan's new national health insurance program: Genesis and experience so far. Health Aff (Millwood) 2003;22:61-76.
27. Song SO, Jung CH, Song YD, Park CY, Kwon HS, Cha BS, *et al.* Background and data configuration process of a nationwide population-based study using the Korean national health insurance system. Diabetes Metab J 2014;38:395-403.
28. Bourne R, Price H, Taylor H, Leasher J, Keeffe J, Glanville J, *et al.* New systematic review methodology for visual impairment and blindness for the 2010 global burden of disease study. Ophthalmic Epidemiol 2013;20:33-9.
29. Yau JW, Rogers SL, Kawasaki R, Lamoureux EL, Kowalski JW, Bek T, *et al.* Global prevalence and major risk factors of diabetic retinopathy. Diabetes Care 2012;35:556-64.
30. IRDiRC; 2022. Available from: https://irdirc.org/who-we-are-2/. [Last accessed on 2023 Mar 05].
31. Rogers S, McIntosh RL, Cheung N, Lim L, Wang JJ, Mitchell P, *et al.* The prevalence of retinal vein occlusion: Pooled data from population studies from the United States, Europe, Asia, and Australia. Ophthalmology 2010;117:313-9.e1.
32. Delcourt C, Korobelnik JF, Buitendijk GH, Foster PJ, Hammond CJ, Piermarocchi S, *et al.* Ophthalmic epidemiology in Europe: The "European Eye Epidemiology" (E3) consortium. Eur J Epidemiol 2016;31:197-210.
33. Majithia S, Tham YC, Chong CC, Yu M, Cheung CY, Bikbov MM, *et al.* Retinal nerve fiber layer thickness and rim area profiles in Asians: Pooled analysis from the Asian eye epidemiology consortium. Ophthalmology 2022;129:552-61.
34. Majithia S, Tham YC, Chee ML, Nusinovici S, Teo CL, Chee ML, *et al.* Cohort profile: The Singapore Epidemiology of Eye Diseases study (SEED). Int J Epidemiol 2021;50:41-52.
35. Chua SY, Thomas D, Allen N, Lotery A, Desai P, Patel P, *et al.* Cohort profile: Design and methods in the eye and vision consortium of UK Biobank. BMJ Open 2019;9:e025077.
36. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, *et al.* The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562:203-9.
37. Available from: https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access. [Last accessed on 2023 Mar 13].
38. Khan SM, Liu X, Nath S, Korot E, Faes L, Wagner SK, *et al.* A global review of publicly available datasets for ophthalmological imaging: Barriers to access, usability, and generalisability. Lancet Digit Health 2021;3:e51-66.
39. Rein DB, Wittenborn JS, Phillips EA, Saaddine JB, Vision and Eye Health Surveillance System Study Group. Establishing a vision and eye health surveillance system for the nation: A status update on the vision and eye health surveillance system. Ophthalmology 2018;125:471-3.
40. Das AV, Kammari P, Vadapalli R, Basu S. Big data and the eyeSmart electronic medical record system – An 8-year experience from a three-tier eye care network in India. Indian J Ophthalmol 2020;68:427-32.
41. Alalawi A, Sztandera L, Lalakia P, Das AV, Gumpili SP, Derman R. Leveraging big data for pattern recognition of socio-demographic and climatic factors in correlation with eye disorders in Telangana State, India. Indian J Ophthalmol 2021;69:1894-900.
42. Das AV, Prashanthi GS, Das T, Narayanan R, Rani PK. Clinical profile and magnitude of diabetic retinopathy: An electronic medical record-driven big data analytics from an eye care network in India. Indian J Ophthalmol 2021;69:3110-7.
43. Garrigan H, Ifantides C, Prashanthi GS, Das AV. Biogeographical and altitudinal distribution of cataract: A nine-year experience using electronic medical record-driven big data analytics in India. Ophthalmic Epidemiol 2021;28:392-9.
44. Stein JD, Rahman M, Andrews C, Ehrlich JR, Kamat S, Shah M, *et al.* Evaluation of an algorithm for identifying ocular conditions in electronic health record data. JAMA Ophthalmol 2019;137:491-7.
45. 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.
46. Kern C, König A, Fu DJ, Schworm B, Wolf A, Priglinger S, *et al.* Big data simulations for capacity improvement in a general ophthalmology clinic. Graefes Arch Clin Exp Ophthalmol 2021;259:1289-96.
47. Jones L, Bryan SR, Miranda MA, Crabb DP, Kotecha A. Example of monitoring measurements in a virtual eye clinic using 'big data'. Br J Ophthalmol 2018;102:911-5.
48. Moore M, Loughman J, Butler JS, Ohlendorf A, Wahl S, Flitcroft DI. Application of big-data for epidemiological studies of refractive error. PLoS One 2021;16:e0250468.
49. Williams KM, Verhoeven VJ, Cumberland P, Bertelsen G, Wolfram C, Buitendijk GH, *et al.* Prevalence of refractive error in Europe: The European Eye Epidemiology (E (3)) consortium. Eur J Epidemiol 2015;30:305-15.
50. Williams KM, Bertelsen G, Cumberland P, Wolfram C, Verhoeven VJ, Anastasopoulos E, *et al.* Increasing prevalence of myopia in Europe and the impact of education. Ophthalmology 2015;122:1489-97.
51. Willis JR, Vitale S, Morse L, Parke DW 2nd, Rich WL, Lum F, *et al.* The prevalence of myopic choroidal neovascularization in the United States: Analysis of the IRIS(®) data registry and NHANES. Ophthalmology 2016;123:1771-82.
52. Bourne RR, Stevens GA, White RA, Smith JL, Flaxman SR, Price H, *et al.* Causes of vision loss worldwide, 1990-2010: A systematic analysis. Lancet Glob Health 2013;1:e339-49.
53. Bourne RR, Cicinelli MV, Sedighi T, Tapply IH, McCormick I, Jonas JB, *et al.* Effective refractive error coverage in adults aged

50 years and older: Estimates from population-based surveys in 61 countries. Lancet Glob Health 2022;10:e1754-63.
54. Bourne RR, Jonas JB, Flaxman SR, Keeffe J, Leasher J, Naidoo K, *et al.* Prevalence and causes of vision loss in high-income countries and in Eastern and Central Europe: 1990-2010. Br J Ophthalmol 2014;98:629-38.
55. Jonas JB, George R, Asokan R, Flaxman SR, Keeffe J, Leasher J, *et al.* Prevalence and causes of vision loss in Central and South Asia: 1990-2010. Br J Ophthalmol 2014;98:592-8.
56. Khairallah M, Kahloun R, Flaxman SR, Jonas JB, Keeffe J, Leasher J, *et al.* Prevalence and causes of vision loss in North Africa and the Middle East: 1990-2010. Br J Ophthalmol 2014;98:605-11.
57. Leasher JL, Lansingh V, Flaxman SR, Jonas JB, Keeffe J, Naidoo K, *et al.* Prevalence and causes of vision loss in Latin America and the Caribbean: 1990-2010. Br J Ophthalmol 2014;98:619-28.
58. Cumberland PM, Rahi JS. Visual health inequalities: Findings from UK Biobank. Lancet 2014;384:S27.
59. 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;134:959-66.
60. Craig JE, Han X, Qassim A, Hassall M, Cooke Bailey JN, Kinzy TG, *et al.* Multitrait analysis of glaucoma identifies new risk loci and enables polygenic prediction of disease susceptibility and progression. Nat Genet 2020;52:160-6.
61. Sekimitsu S, Wang J, Elze T, Segrè AV, Wiggs JL, Zebardast N. Interaction of background genetic risk, psychotropic medications, and primary angle closure glaucoma in the UK Biobank. PLoS One 2022;17:e0270530.
62. Kaye RA, Patasova K, Patel PJ, Hysi P, Lotery AJ, UK Biobank Eye and Vision Consortium. Macular thickness varies with age-related macular degeneration genetic risk variants in the UK Biobank cohort. Sci Rep 2021;11:23255.
63. Gao XR, Huang H, Kim H. Genome-wide association analyses identify 139 loci associated with macular thickness in the UK Biobank cohort. Hum Mol Genet 2019;28:1162-72.
64. Choquet H, Wiggs JL, Khawaja AP. Clinical implications of recent advances in primary open-angle glaucoma genetics. Eye (Lond) 2020;34:29-39.
65. MacGregor S, Ong JS, An J, Han X, Zhou T, Siggs OM, *et al.* Genome-wide association study of intraocular pressure uncovers new pathways to glaucoma. Nat Genet 2018;50:1067-71.
66. Tham YC, Tao Y, Zhang L, Rim TH, Thakur S, Lim ZW, *et al.* Is kidney function associated with primary open-angle glaucoma? Findings from the Asian Eye Epidemiology consortium. Br J Ophthalmol 2020;104:1298-303.
67. Sabanayagam C, Sultana R, Banu R, Rim T, Tham YC, Mohan S, *et al.* Association between body mass index and diabetic retinopathy in Asians: The Asian Eye Epidemiology Consortium (AEEC) study. Br J Ophthalmol 2022;106:980-6.
68. Mauschitz MM, Bonnemaijer PW, Diers K, Rauscher FG, Elze T, Engel C, *et al.* Systemic and ocular determinants of peripapillary retinal nerve fiber layer thickness measurements in the European Eye Epidemiology (E3) population. Ophthalmology 2018;125:1526-36.
69. Ko F, Muthy ZA, Gallacher J, Sudlow C, Rees G, Yang Q, *et al.* Association of retinal nerve fiber layer thinning with current and future cognitive decline: A study using optical coherence tomography. JAMA Neurol 2018;75:1198-205.
70. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, *et al.* Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402-10.
71. Ran AR, Tham CC, Chan PP, Cheng CY, Tham YC, Rim TH, *et al.* Deep learning in glaucoma with optical coherence tomography: A review. Eye (Lond) 2021;35:188-201.
72. Ting DS, Cheung CY, Lim G, Tan GS, Quang ND, Gan A, *et al.* Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. JAMA 2017;318:2211-23.
73. Lin H, Long E, Ding X, Diao H, Chen Z, Liu R, *et al.* Prediction of myopia development among Chinese school-aged children using refraction data from electronic medical records: A retrospective, multicentre machine learning study. PLoS Med 2018;15:e1002674.
74. Barthelmes D, Nguyen V, Walton R, Gillies MC, Daien V, Fight Retinal Blindness Study Group. A pharmacoepidemiologic study of ranibizumab and aflibercept use 2013-2016. The Fight Retinal Blindness! Project. Graefes Arch Clin Exp Ophthalmol 2018;256:1839-46.
75. Biechl AC, Bhandari S, Nguyen V, Arnold JJ, Young S, Fraser-Bell S, *et al.* Changes in real-world treatment patterns for diabetic macular oedema from 2009 to 2019 and 5-year outcomes: Data from the Fight Retinal Blindness! Registry. Clin Exp Ophthalmol 2020;48:802-12.
76. 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.
77. Zhang Y, Li T, Reddy A, Nallasamy N. Gender differences in refraction prediction error of five formulas for cataract surgery. BMC Ophthalmol 2021;21:183.
78. Kramer BA, Berdahl J, Gu X, Merchea M. Real-world incidence of monofocal toric intraocular lens repositioning: Analysis of the American Academy of Ophthalmology IRIS registry. J Cataract Refract Surg 2022;48:298-303.
79. Coleman AL. How big data informs us about cataract surgery: The LXXII Edward Jackson memorial lecture. Am J Ophthalmol 2015;160:1091-103.e3.
80. Lacy M, Kung TH, Owen JP, Yanagihara RT, Blazes M, Pershing S, *et al.* Endophthalmitis rate in immediately sequential versus delayed sequential bilateral cataract surgery within the intelligent research in sight (IRIS®) registry data. Ophthalmology 2022;129:129-38.
81. Bhandari S, Squirrell D, Nguyen V, Wang N, Wells JM, Tan T, *et al.* Bevacizumab for diabetic macular oedema: One-year treatment outcomes from the Fight Retinal Blindness! Registry. Eye (Lond) 2022;36:594-602.
82. 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;125:522-8.
83. Bhandari S, Nguyen V, Arnold J, Young S, Banerjee G, Gillies M, *et al.* Treatment outcomes of ranibizumab versus aflibercept for neovascular age-related macular degeneration: Data from the Fight Retinal Blindness! Registry. Ophthalmology 2020;127:369-76.
84. Figueras-Roca M, Parrado-Carrillo A, Nguyen V, Casaroli-Marano RP, Moll-Udina A, Gillies MC, *et al.* Treat-and-extend versus fixed bimonthly treatment regimens for treatment-naive neovascular age-related macular degeneration: Real world data from the Fight Retinal Blindness registry. Graefes Arch Clin Exp Ophthalmol 2021;259:1463-70.
85. Daien V, Nguyen V, Essex RW, Morlet N, Barthelmes D, Gillies MC, *et al.* Incidence and outcomes of infectious and noninfectious endophthalmitis after intravitreal injections for age-related macular degeneration. Ophthalmology 2018;125:66-74.
86. Rough K, Thompson JT. When does size matter? Promises, pitfalls, and appropriate interpretation of "big" medical records data. Ophthalmology 2018;125:1136-8.
87. Househ MS, Borycki EM. Big Data, Big Challenges: A Healthcare Perspective. Berlin, Germany: Springer; 2019.
88. Bowman S. Impact of electronic health record systems on information integrity: Quality and safety implications. Perspect Health Inf Manag 2013;10:1c.
89. Coleman AL, Morgenstern H. Use of insurance claims databases

to evaluate the outcomes of ophthalmic surgery. Surv Ophthalmol 1997;42:271-8.
90. Liu V, Musen MA, Chou T. Data breaches of protected health information in the United States. JAMA 2015;313:1471-3.
91. Madondo SM. The American Statistical Association (ASA) statement of 2016 on statistical significance and *P* value: A critical thought. Sci J Appl Math Stat 2017;5:41-8.
92. Nuzzo R. Scientific method: Statistical errors. Nature 2014;506:150-2.
93. Ehrenstein V, Nielsen H, Pedersen AB, Johnsen SP, Pedersen L. Clinical epidemiology in the era of big data: New opportunities, familiar challenges. Clin Epidemiol 2017;9:245-50.
94. Lonjon G, Porcher R, Ergina P, Fouet M, Boutron I. Potential pitfalls of reporting and bias in observational studies with propensity score analysis assessing a surgical procedure: A methodological systematic review. Ann Surg 2017;265:901-9.
95. Bower JK, Patel S, Rudy JE, Felix AS. Addressing bias in electronic health record-based surveillance of cardiovascular disease risk: Finding the signal through the noise. Curr Epidemiol Rep 2017;4:346-52.