# Commentary: Big data in Ophthalmology: A big game changer on horizon

The current issue of the Indian Journal of Ophthalmology analyzes the relationship between age-related macular degeneration (AMD) and diabetic retinopathy (DR) using big data.[1] AMD is one of the leading causes of blindness in the elderly, while DR is the leading cause of severe vision loss in adults of working age.[2] To prove a causal relationship between two common conditions, a large sample size data is required. Data, which in the current age is available in abundance, is now being claimed to be the new oil fueling the fourth industrial revolution.[3]

Medical research has been driven by a common phrase "Minimum Sample Size." The aim of collecting the minimal sample size is to acquire a sufficiently large amount of data to detect a minimal meaningful detectable difference (effect size), and its calculation has been hypothesized by multiple authors depending upon the methodology employed.[4] For detecting a small effect size, which translates to rare incidences and subtle associations among various pathologies, a large number of participants are required. Such large data has previously been unimaginable to obtain and process.

Big data is an evolving concept, first introduced in the 1990s. It has been defined by the Oxford dictionary as "extremely large data sets that may be analyzed computationally to reveal patterns, trends, and associations, especially relating to human behavior and interactions." Important attributes of big data include a large volume of data, velocity (speed of data flow in real-time), variety (diversity), veracity (quality or accuracy), and value (usefulness of the data).[3]

Big data can be obtained from electronic health records (EHR) or electronic medical records (EMR). The current article[1] employs one such EHR. Big data is like crude oil, which is of value, but is of no utility unless refined. A simple example can be found in this article. Records of a total of 43,153 diabetics were analyzed to filter out 2048 participants who were age and diabetes mellitus duration matched for finding the association. With the induction of big data, the future of medical research is now ready to move beyond the minimum sample size.

Artificial intelligence (AI) can be employed to automatically analyze big data to find out ocular conditions, side effects, and novel biomarkers, which would otherwise have been disregarded as rare incidences or anomalies. A novel association between glaucoma progression and pulmonary function is one of the most exciting and unexpected outcomes from a machine learning algorithm employed in the United Kingdom.[5] Automation of diagnosis by analyzing the images such as fundus photographs, optical coherence tomography (OCT) images, or OCT-angiogram outputs is also being attempted by various groups by using AI, machine learning, and deep learning.[3] The Food and Drug Administration, USA has approved IDx-DR (Digital Diagnostics, Coralville, IA, USA) and EyeArt (Eyenuk, Woodland Hills, CA, USA), which are based on AI to "detect more than mild diabetic retinopathy (mtmDR) in adults diagnosed with diabetes who have not been previously diagnosed with diabetic retinopathy." The analyzed big data, in Ophthalmology, has applications also beyond medical research. It can improve patient care, planning of policy, and hospital administration as well.

The limitations of big data include difficult analysis, the need to formulate suitable algorithms, and a high risk of false-positive findings. Challenges regarding the security of data include data privacy, the potential for cyber theft, misuse, and the need for legislation. Barriers to using big data include the cost of infrastructure, dependence on network connectivity, and skilled manpower to analyze the data. EHR and EMR are well known to contain incomplete records, misclassifications, nonstandard nomenclatures, and different classifications/grading of the same clinical finding/condition.[6] Maintaining the quality of big data is one of the most important challenges.

Also, the employment of a large amount of data can be ironic. While rare incidences can be picked up, minor deviations can also be calculated as statistically significant while analyzing thousands and lakhs of data points. To avoid such a situation, the biological plausibility and its clinical implication of the outcomes of any trial using big data should be humanly inferred. The accompanying article, for example, showed patients with AMD to have lower odds of developing DR. The ischemia in viable retina required for the pathogenesis of DR can be counteracted by the atrophy and subsequent reduced oxygen demand in patients with AMD, giving a biological plausibility for the results.

We look forward to the next decade for trials employing big data and AI to provide us with newer insights and better diagnostic algorithms.

***Nilesh Kumar, Ashish Sharma[1], Koushik Tripathy[2]***

Department of Retina, Ayodhya Eye Hospital, Tulsibari, Ayodhya, Uttar Pradesh, [1]Department of Retina, Lotus Eye Hospital and Institute, Avinashi Road, Coimbatore, Tamil Nadu, [2]Department of Retina, ASG Eye Hospital, 149 BT Road, Kolkata, West Bengal, India

Correspondence to: Dr. Koushik Tripathy, Department of Retina, ASG Eye Hospital, 149 BT Road, Kolkata - 700 058, West Bengal, India. E-mail: koushiktripathy@gmail.com

## References

1. Sahoo NK, Mehta MC, Rani PK, Khanna RC, Raman R, Bhattacharya J, *et al*. Impact of age-related macular degeneration on diabetic retinopathy: An electronic health record based big data analysis from a tertiary eye centre in South India. Indian J Ophthalmol 2021;69:3184-8.
2. Shukla UV, Tripathy K. Diabetic Retinopathy. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2021. Available from: http://www.ncbi.nlm.nih.gov/books/NBK560805/. [Last accessed on 2021 Sep 30].
3. Cheng C-Y, Soh ZD, Majithia S, Thakur S, Rim TH, Tham YC, *et al*. Big data in ophthalmology. Asia Pac J Ophthalmol (Phila) 2020;9:291–8.
4. Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. J Clin Diagn Res 2016;10:YE01–6.
5. Mehta P, Petersen CA, Wen JC, Banitt MR, Chen PP, Bojikian KD, *et al*. Automated detection of glaucoma with interpretable machine learning using clinical data and multimodal retinal images. Am J Ophthalmol 2021;231:154–69.
6. Bowman S. Impact of electronic health record systems on information integrity: Quality and safety implications. Perspect Health Inf Manag 2013;10:1c.



**Access this article online**

**Quick Response Code:**

**Website:** www.ijo.in

**DOI:** 10.4103/ijo.IJO_2566_21

**Cite this article as:** Kumar N, Sharma A, Tripathy K. Commentary: Big data in Ophthalmology: A big game changer on horizon. Indian J Ophthalmol 2021;69:3188-9.