

Marcos Cirne¹, Jaime Bacardit¹, Rahele Kafieh², Siyue Song¹, Dexter Canoy¹, David Steel¹, Anya Hurlbert¹

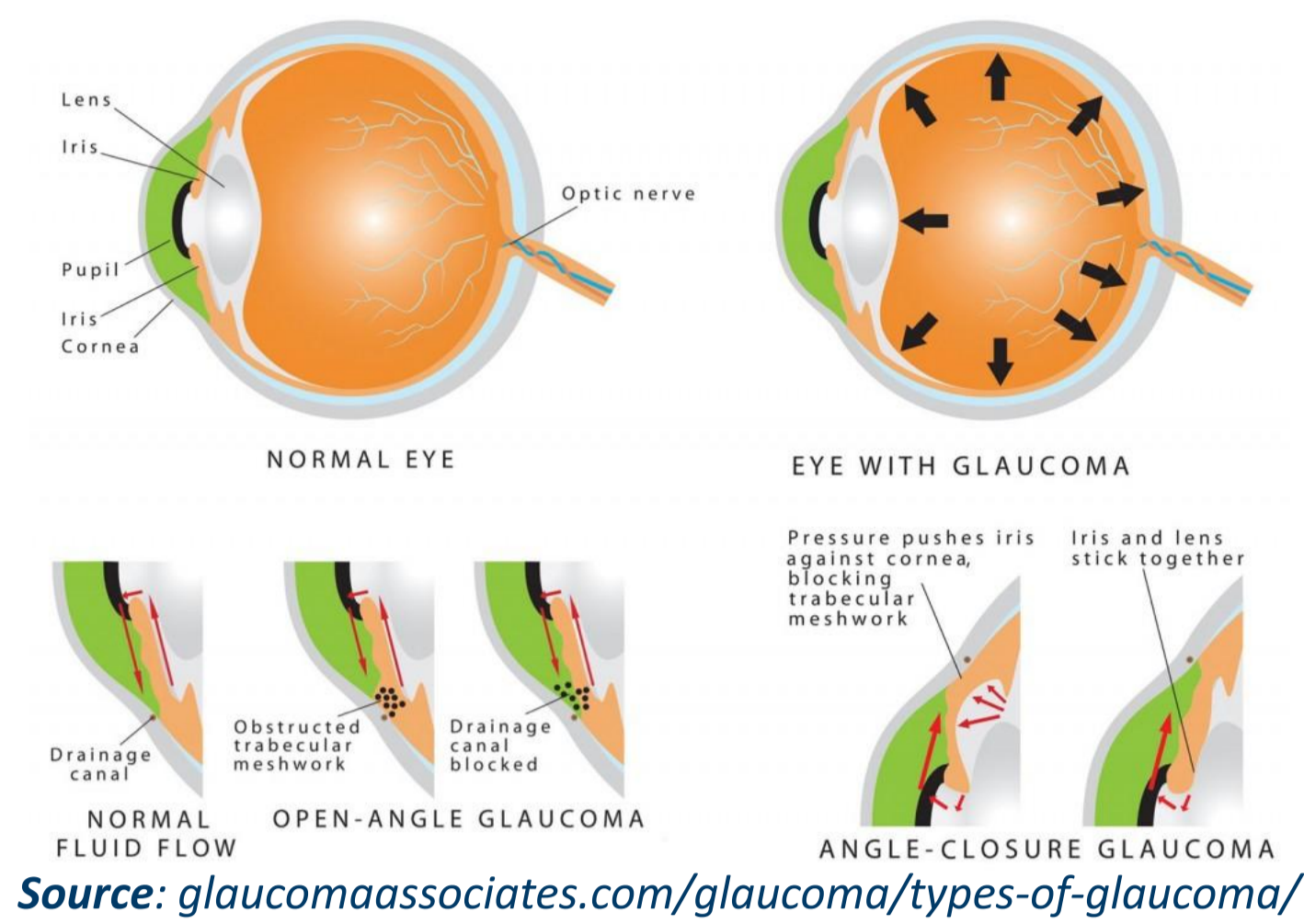
¹ Newcastle University, Newcastle upon Tyne, United Kingdom, NE1 7RU

² Durham University, Durham, United Kingdom, DH1 3LE

1. Introduction

Glaucoma is the leading cause of irreversible blindness worldwide.

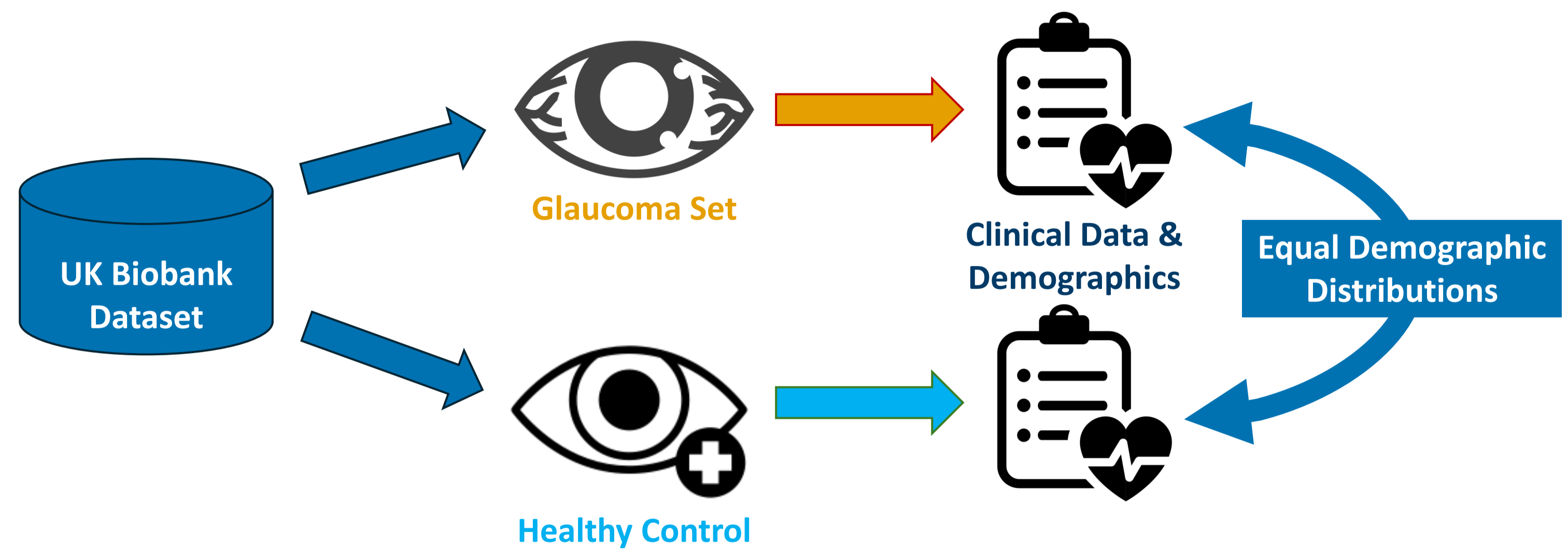
In **EYESAVE**, we aim to develop an AI-based tool for **risk stratification** of glaucoma patients. **AI-based systems** require **well-curated datasets** for optimal performance on training and evaluation.



We are constructing a **longitudinal dataset** of retinal imaging and clinical data of patients from local hospital records.

The current preliminary analysis of **UK Biobank** data illustrates statistical analyses used to identify key differences between **glaucoma** and **healthy** patients, thereby mitigating bias.

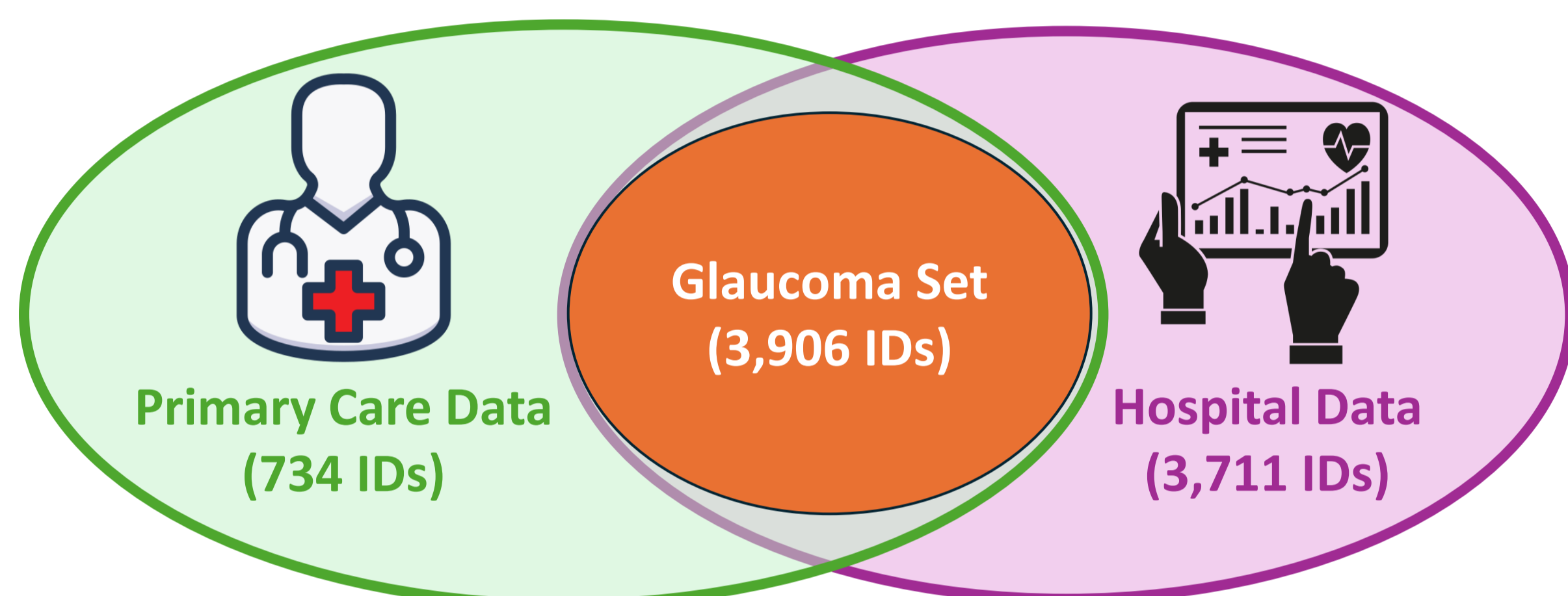
2. Materials and Methods



Subsets are matched on **Demographic Features**: Age, Gender, Ethnicity and Townsend Deprivation Index (TDI); then analysed for differences in **Clinical Features**: intraocular pressure (IOP), blood pressure, body mass index (BMI), corneal thickness and comorbidities (e.g. diabetes, hypertension, smoking status, ischaemic heart diseases).

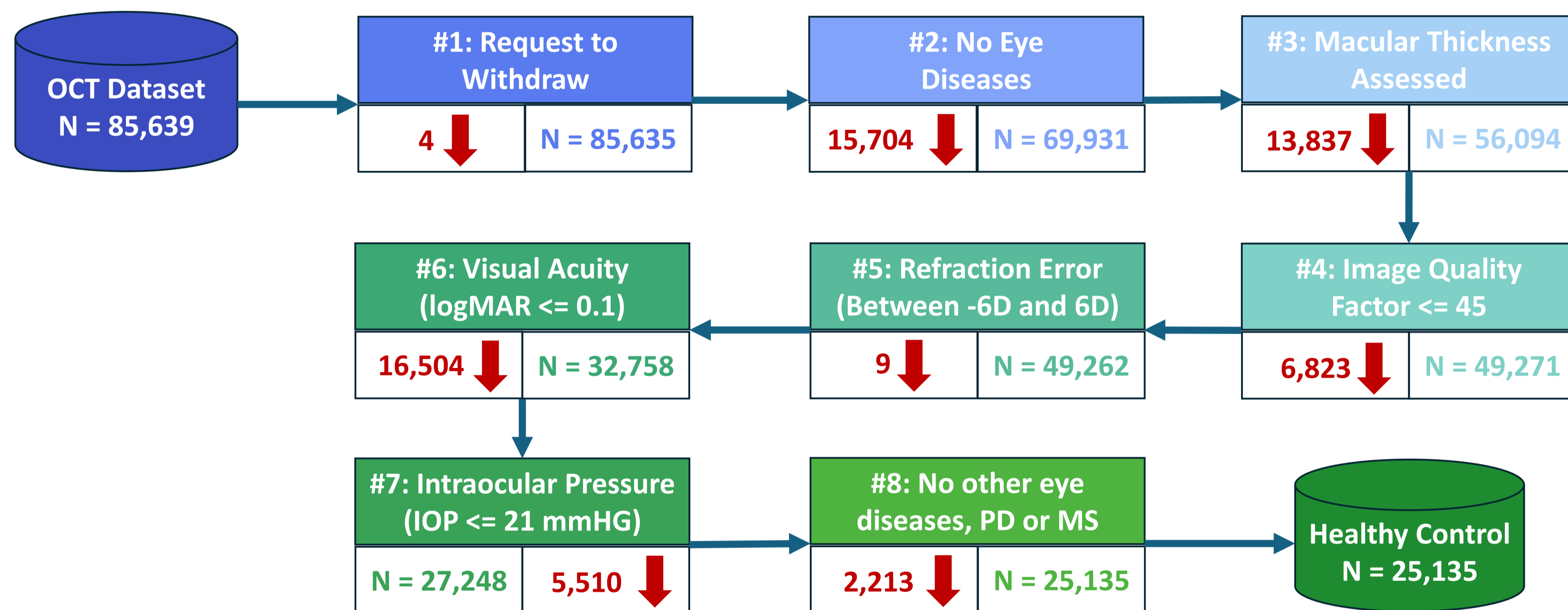
3. Glaucoma Filters

Applied only on patients aged between 40 and 69 years, with **OCT scans** available for both eyes (N = 85,639)



4. Healthy Control Filters

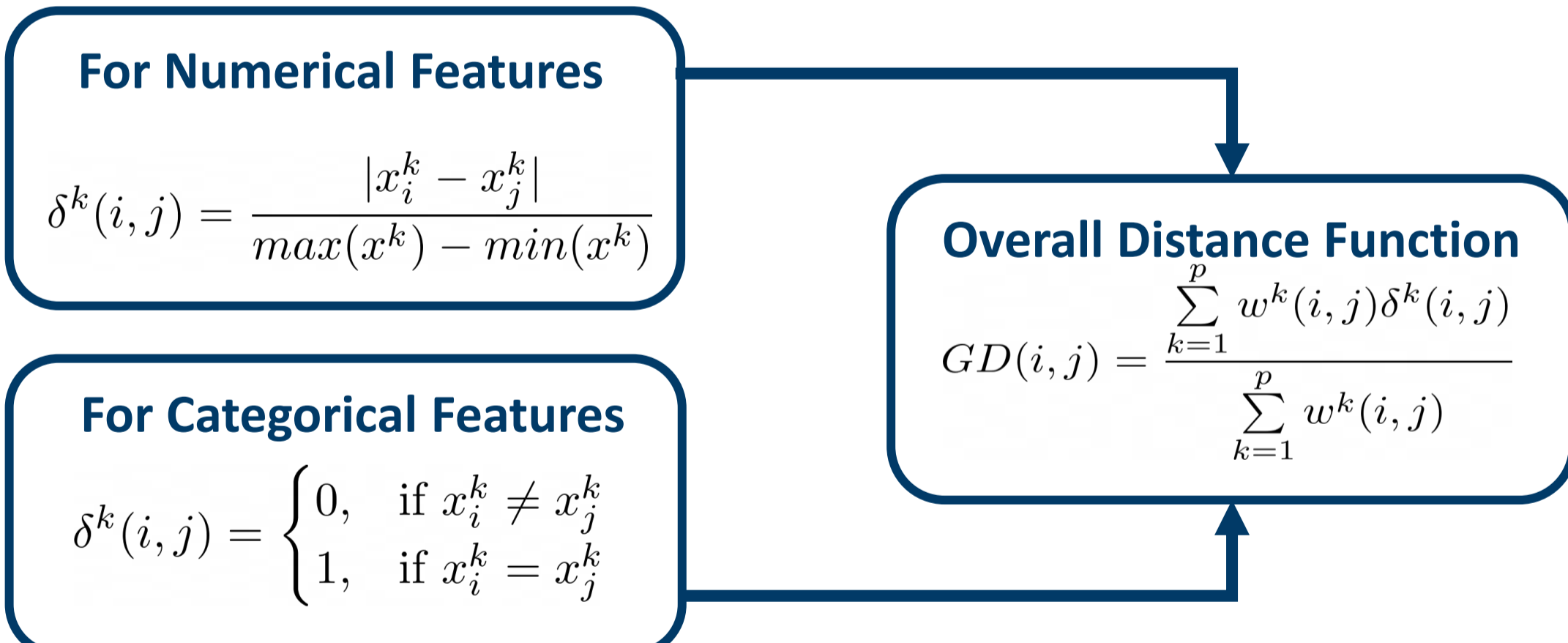
Filter sequence based on the work of Patel et al. [1]



5. Subset Matching

Finds the “best subset” of the **healthy control** group with the same size as the **glaucoma set**.

Minimizes the **total distance** (weighted average) of the distributions of all demographic features between all possible pairs (i,j) of patients based on **Gower’s Distance (GD)** [2].

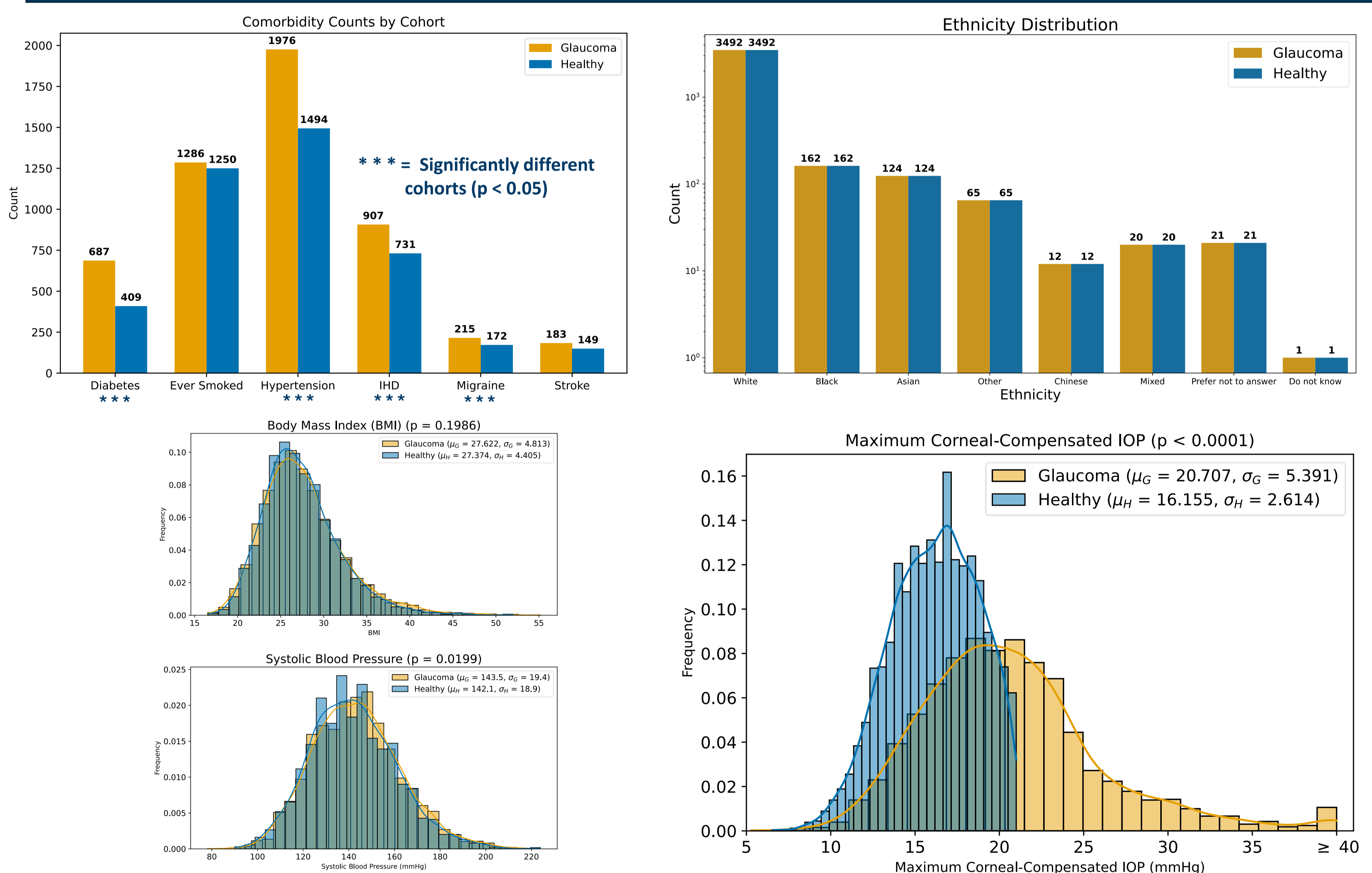


6. Population Description

After subset matching, the **healthy control** subset matched the **glaucoma set** in distributions of age, sex, ethnicity and socioeconomic status (by TDI).

For age, both subsets have a mean value of 60.8 years, with a standard deviation of 6.4 (**p = 0.1139**).

7. Statistical Analysis



No significant differences between the subsets are found for smoking status, body mass index (BMI) and systolic blood pressure.

8. Conclusions

The **glaucoma set** has significantly greater proportions of diabetes, hypertension, ischaemic heart disease, and migraine, compared with the **healthy control subset**. The statistical analysis identifies these co-morbidities, together with IOP, as important factors in glaucoma diagnoses. A further in-depth analysis of smoking status will also be conducted.

Non-white British patients are still highly underrepresented, despite demographic feature matching, thus requiring **data augmentation methods** to handle class imbalance.

References

- [1] Patel, Praveen J., et al. *Spectral-domain optical coherence tomography imaging in 67 321 adults: associations with macular thickness in the UK Biobank study.* *Ophthalmology* 123.4 (2016): 829-840.
- [2] Tuerhong, G. and Kim, S. B. *Gower distance-based multivariate control charts for a mixture of continuous and categorical variable.* Volume 41, Issue 4, Part 2, (2014): 1701-1707

Acknowledgement

This work was supported by the Vivensa Foundation, grant number ARHVF2402\7, for project EYESAVE at Newcastle University, with partners Durham University, and the Newcastle upon Tyne Hospitals and Sunderland and South Tyneside NHS Trusts.

Contact

marcos.cirne@ncl.ac.uk



[linkedin.com/in/marcoscirne/](https://www.linkedin.com/in/marcoscirne/)