A Datasheet for the INSIGHT University Hospitals Birmingham Retinal Vein Occlusion Data Set

Bilton, Edward J.; Guggenheim, Emily J.; Baranyi, Balazs; Radovanovic, Charlotte; Williams, Rowena L.; Bradlow, William; Denniston, Alastair K.; Mollan, Susan P.

[Link to publication on Research at Birmingham portal](#)

A Datasheet for the INSIGHT University Hospitals Birmingham Retinal Vein Occlusion Data Set

Edward J. Bilton, MBChB (MBBS),[1,][*] Emily J. Guggenheim, PhD,[1,][*] Balazs Baranyi,[1] Charlotte Radovanovic,[1] Rowena L. Williams,[1] William Bradlow, FRCP,[2,3,4] Alastair K. Denniston, PhD,[1,5,6] Susan P. Mollan, FRCOphth[1,6]

Purpose: Retinal vein occlusion (RVO) is the second leading cause of visual loss due to retinal disease. Retinal vein occlusion increases the risk of cardiovascular mortality and the risk of stroke. This article describes the data contained within the INSIGHT eye health data set for RVO and cardiovascular disease.

Design: Data set descriptor for routinely collected eye and systemic disease data.

Participants: All people who had suffered an RVO aged ≥ 18 years old, attending the Ophthalmology Clinic at Queen Elizabeth Hospital, University Hospitals Birmingham (UHB) National Health Service (NHS) Trust were included.

Methods: The INSIGHT Health Data Research Hub for Eye Health is an NHS-led ophthalmic bioresource. It provides researchers with safe access to anonymized routinely collected data from contributing NHS hospitals to advance research for patient benefit. This report describes the INSIGHT UHB RVO and major adverse cardiovascular events data set, a data set of ophthalmology and systemic data derived from the United Kingdom's largest acute care trust.

Main Outcome Measures: This data set consists of routinely collected data from the hospital's electronic patient records. The data set primarily includes structured data (relating to their hospital eye care and any cardiovascular data held for the individual) and OCT ocular images. Further details regarding the available data points are available in the supplementary information.

Results: At the time point of this analysis (September 30, 2022) the data set was composed of clinical data from 1521 patients, from Medisoft records inception. The data set includes 2196 occurrences of RVO affecting 2026 eyes, longitudinal eye follow-up clinical parameters, over 6217 eye-related procedures, and 982 encountered complications. The data set contains information on 2534 major adverse cardiovascular event occurrences, their subtype, number experienced per patient, and chronological relation to RVO event. Longitudinal follow-up data including laboratory results, regular medications, and all-cause mortality are also available within the data set.

Conclusions: This data set descriptor article summarizes the data set contents, the process of its curation, and potential uses. The data set is available through the structured application process that ensures research studies are for patient benefit. Further information regarding the data repository and contact details can be found at https://www.insight.hdrhub.org/.

Financial Disclosure(s): Proprietary or commercial disclosure may be found after the references. *Ophthalmology Science 2023;3:100388 © 2023 by the American Academy of Ophthalmology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).*

Supplemental material available at www.ophthalmologyscience.org.

Retinal vein occlusion (RVO) is the second leading cause of vision loss due to retinal disease, after diabetic retinopathy.[1] Retinal vein occlusion is caused by a blockage of the veins in the retina and can be subdivided into branch (BRVO), central (CRVO), or hemiretinal (HRVO), depending on the location of the occlusion. These can be further clinically categorized depending upon the presence of complications such as macular edema, macular ischemia, and neovascularization.

Globally, BRVO has a higher incidence rate than CRVO of 0.5% to 1.2%; however, there is limited prevalence and incidence data in the United Kingdom (UK). Secondary macular edema will develop in approximately 5% to 15% of BRVO eyes within 12 months and lead to vision loss in 50% of cases.[2] Conversely, CRVO has a lower incidence rate of approximately 0.1% to 0.4%; however, patients typically have worse visual acuity at diagnosis and often have evidence of cystoid macular edema.[1,3] Of CRVO patients,

75% will go on to develop a complication of cystoid macular edema within 2 months, and all will experience vision loss.[3]

Importantly, aside from the ocular sequelae, RVO is associated with adverse cardiovascular health due to the strong relationship to atherosclerosis, which can underpin RVO's pathophysiology. Retinal vein occlusion doubles the risk of cardiovascular mortality in patients < 70 years old[4] and increases the risk of stroke (odds ratio, 1.73; 95% confidence interval, 1.4−2.12, $P < 0.001$).[5] There are currently conflicting views regarding RVO and the development of other major adverse cardiovascular events (MACEs), such as myocardial infarction.[6,7]

The Queen Elizabeth Hospital is a large UK hospital and part of the University Hospitals Birmingham (UHB) National Health Service (NHS) Foundation Trust. The trust provides general and specialist care to patients living across South Birmingham. Among its various services, it delivers ophthalmic specialist care provision to patients who experience RVOs. During their health care admissions, patients routinely have multiple parameters recorded relevant to eye disease and cardiovascular health. The data set presented here represents a cohort of patients who have been diagnosed with an RVO while under the care of the Queen Elizabeth Hospital or who have previously suffered an RVO, as mentioned in their patient history, and includes subsequent procedures, eye examinations, medications, and test results after diagnosis.

INSIGHT is one of a number of Health Data Research Hubs established by UK Research & Innovation through Health Data Research UK. INSIGHT enables access to anonymized routinely collected patient data from UHB and Moorfields Eye Hospital. There is a focus on eye health and oculomics,[8] where the eye and eye data are used to discover novel biomarkers for systemic diseases such as dementia and ischemic heart disease.[9] Data are presented in a research-ready format where NHS patient-level data are curated, pseudo-anonymized, and anonymized through an irreversible deidentification process before access. Distribution of the data in this way, in tandem with INSIGHT's governance processes, allows safe access to research data in keeping with the Office of National Statistics' "Five Safes" framework and has been approved by the National Research Ethics Service.[10] With rigorous pipelining, INSIGHT offers consistent and continual uploading of up-to-date clinical data to provide an "evergreen" source of research informatics.

This article describes the INSIGHT UHB RVO and MACE data set, a longitudinal record of routinely collected ophthalmic and cardiovascular data relevant to RVO and MACE episodes. Compilation of the data set adhered to the tenets of the Declaration of Helsinki and has approval from the relevant institutional review board and ethical approval from the West of Scotland Research Ethics Service (REC reference: 20/WS/0087). Approval has been given to retrospectively collate patient data without explicit written consent. The data set is presented in line with datasheets for data set guidance, which have been developed to enhance the transparency and accountability of data sets, while attempting to minimize societal biases data sets

might be susceptible to when used for machine learning purposes.[11] Headings such as motivation, composition, collection process, and recommended uses, with bespoke headings relevant to RVO and MACE descriptive purposes, will be discussed in this article.

Datasheet

Motivation for Data Set Creation

An estimated 16.4 million adults are affected by RVO worldwide, with a prevalence of 5.20 per 1000 across Europe, the United States, Asia, and Australia.[12] The pathogenesis of RVO involves obstruction of retinal veins secondary to external mechanical compression or internal thrombotic luminal occlusion, leading to venous congestion and consequent retinal ischemia. In response to ischemia, VEGF production increases, causing retinal neovascularization and subsequent risk of edema, which can cause permanent vision loss if untreated.[13,14]

Retinal vein occlusion has been associated with increased risk of Alzheimer's disease and vascular and all-cause dementia.[15] There is much debate whether RVO may predict the development of cardiovascular or cerebrovascular events or whether their associations are due to shared risk factors, of which there are many. Recent systematic reviews have attempted to find associations between RVO, MACEs, and stroke.[6,7] Despite finding positive associations between the diseases, these reviews have highlighted the lack of studies providing large RVO populations, adequate long-term follow-up, and analyses assessing cardiovascular risk factors in different RVO subtypes.

Cardiovascular disease is the leading cause of death worldwide, conveying a significant service provision and financial burden to health care services, costing the NHS £7.4 billion per year.[16] Consistent promotion of a healthy lifestyle is one of the most important preventative measures to avoid future atherosclerotic cardiovascular disease and is often used in conjunction with lipid and blood pressure-lowering therapeutic interventions.[17] By investigating the link between RVO and MACE, we may be able to provide observable, objective evidence denoting the cardiovascular health of patients, helping to inform decision-making processes regarding preventative and therapeutic options from both patient and professional perspectives.

The utility of artificial intelligence in the diagnosis of diabetic retinopathy, glaucoma, and age-related macular degeneration has been demonstrated by multiple studies over the past decade. Although it is a common ophthalmic condition, there have been relatively few studies investigating the use of artificial intelligence in diagnosing RVO.[18] Key barriers limiting the implementation of artificial intelligence surround the availability of diverse patient cohorts with sufficient size and longitudinal data for models to train on. In this regard, the RVO data set holds great promise, given the wide range of ethnic and socioeconomic patient groups that UHB represents in its locality, and access to a constant, self-sustaining stream of new data sources to add to and enhance its repertoire.

Data Set Composition

This data set comprises a longitudinal record of routinely collected ophthalmic and cardiovascular data relevant to RVO and MACE episodes at UHB. University Hospitals Birmingham provides care to a diverse socioethnic patient population with a greater-than-average number of minority ethnic groups compared with other NHS Trusts.

The data set contains all patients that have been diagnosed with any form of RVO, either as an active diagnosis or historical disease in the electronic health record (EHR) system, Medisoft (Heidelberg Engineering GmbH). The analysis presented here is time-locked and covers a reporting time period ranging from historic records as early as 1978 until a data set lock date of September 30, 2022. The majority of patient data were recorded after the introduction of Medisoft in 2010; records before this date were uploaded to Medisoft retrospectively. The breakdown per year along with patient demographics is detailed in Figure 1 and Table 1.

The data set comprises information regarding 1521 people. Of the total number of patients, after classification into historical or diagnoses, 1237 patients (1361 eyes) had an RVO diagnosis, 58 patients had both a history recorded and a diagnosis, and 226 patients only had a history of RVO. Within the data set, there are a total of 2534 MACE occurrences, 1290 post-RVO diagnosis and 1244 pre-RVO diagnosis, experienced by a total of 723 patients. The number of MACEs per year is summarized in Figure 2. A total of 291 patients experienced a MACE only after an RVO event, 242 patients had events both pre- and post-RVO, and 190 patients experienced MACE only before an RVO incident. The majority of patients suffered 1 or 2 MACEs; however, small numbers of patients experienced > 10 events (Table 2).

A breakdown of RVO type and the number of patients and eyes affected is detailed in Table S3 (available at www.ophthalmologyscience.org). Complications such as edema and neovascularization can affect the macula or other areas of the eye in RVO. Figure 3 and Table S4 (available at www.ophthalmologyscience.org) display the number of patients who developed complications in conjunction with their RVO classification. The categorization of complications is detailed in Table S5 (available at www.ophthalmologyscience.org). A subset of RVO patients undergo various procedures at UHB. These range from eye procedures, cardiac-related procedures, and vascular procedures to any other not classified in these groups (Figure 4; Figure S5 and Table S6, available at www.ophthalmologyscience.org). Of this cohort, almost all patients (1513 of 1521 patients) underwent ≥ 1 procedure in one of the categories. Of these, 1047 patients, who may have had a history or diagnosis, or both, underwent > 1 eye procedure.

A portion of the data set is eye-linked, where eyecode laterality was used to link the first instance of an RVO per eye with subsequent follow-up eye examinations, complications, and ophthalmic procedures, if available. In the case of 290 patients, historical information was available that predates a diagnosis with no eyecode or earliest diagnosis date information; these were therefore excluded from the presented numbers pertaining to eye-linked procedures and eye-linked eye examinations. A further 6 patients had a concurrent giant cell arteritis (GCA) diagnosis and were excluded from the eye-linked data set due to their differences in underlying pathophysiology when compared with RVO. As a result, 1225 patients (1349 eyes) were included in the subsequent eye-linked data set. Although beyond the scope of this datasheet, supplemental analysis could potentially be carried out to identify additional eyecodes and dates in such cases, if required.

Of the RVO patients with eyecodes and diagnosis dates available, 746 patients had 6225 eye procedures after their diagnosis, of which 4276 were anti-VEGF injections, 241 steroid injections, and 887 laser procedures (Table 7). Procedures are categorized as front of the eye (anterior segment procedures), back of the eye (posterior segment procedures), or unable to be determined (Table S8, available at www.ophthalmologyscience.org). The majority of procedures were posterior segment procedures, and people with CRVO and BRVO had roughly equivalent numbers of procedures overall.

The data set includes longitudinal follow-up eye examinations for the RVO cohort for eyes after a diagnosis of RVO, including their lens status, intraocular pressure, refraction, optic disc information, visual acuity measurements, visual field tests, and a large number of OCT exams, which are recorded in Medisoft (Table 9). The average and median time between follow-up tests in the first year is reported. The data set also contains diabetic retinopathy grading data; in total, grading data was present for 293 patients.

A multitude of laboratory and test results are available for the patients, within 3 months and 1 year of diagnosis, including alcohol and smoking assessments, body mass index, lipid profiles, blood pressure, and glycated hemoglobin (see Table S10, available at www.ophthalmologyscience.org, for full list; others may be available upon request).

Medications taken by the entire cohort after RVO diagnosis were collated and were categorized into "Lipid lowering," "Hypertension," or "Diabetes" drugs (Figure 6 and Table S11, available at www.ophthalmologyscience.org). Categorizations can be found in Table S12 (available at www.ophthalmology science.org).

Mortality data, including time and cause of death, was populated via death certificate information from EHRs. Cause of death was categorized to cardiac-related, stroke-related, other vascular-related, or other, and is an impression of the overall cause of death considering sections 1 (a, b, and c) and 2 of certificates. At the time of extraction (January 10, 2022), 527 people had died, with the average time from first RVO record to death of 1604 days (4.4 years). Of the patients who died, 21 suffered a cardiac-related death, and 10 suffered a vascular-related death.

Key data included in this data set are available in a tabular format (Appendix A) and include:

- Total number of patients with recorded diagnosis of RVO and a medical history mention of RVO covering a period from 1978 to 2022

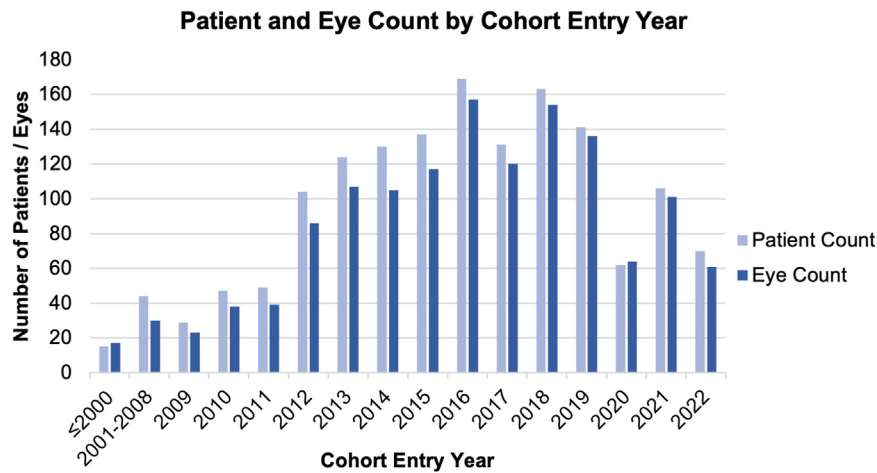**Patient and Eye Count by Cohort Entry Year**

Figure 1. Patient and eye count by cohort entry year*. Number of patients and eyes presenting with retinal vein occlusion (RVO) grouped by year. Entry year calculated by earliest diagnosis date, or date of first episode for patients with a history of RVO. Patients may have new diagnoses in their second eye in subsequent years. This does not include patients with a history of RVO, or those with no eyecode available. *Medisoft was first introduced in 2010, and paper noting was phased out by 2014.

- Demographic information including age, sex, ethnicity, indices of multiple deprivation deciles, and age at diagnosis as well as information relating to alcohol consumption and smoking status
- Diagnosis RVO type and affected eyes, where available
- Ophthalmic complications including macular edema and neovascularization
- Occurrence of MACEs before and after a recorded RVO diagnosis
- Procedures carried out after an RVO diagnosis
- Longitudinal follow-up visits for a range of eye examinations including: visual acuity measurements, lens status, intraocular pressure, OCT, optic disc size, refraction, and visual fields
- Hospital tests for a range of related markers, including blood pressure, body mass index, serum cholesterol, low-density lipoproteins, high-density lipoproteins, triglycerides, estimated glomerular filtration rate, hemoglobin, mean cell volume, glycated hemoglobin, and N-terminal pro B-type natriuretic peptide within 3 months and 1 year of a diagnosis
- Information on related medications, such as hypertension drugs, lipid lowering drugs, and diabetes drugs
- Diabetic status and type
- Mortality and cause of death

Due to INSIGHT's potential, the contents of this data set do not resemble an exhaustive list of the available data points for these patients. As all patients attended UHB, health data relevant to eye and systemic health from other visits would be available to researchers through application, allowing new research potential into RVO or MACE association with other imaging, laboratory, and disease entities.

Collection Process

Identification of Patient Cohort from Medisoft. Patients were identified using the Medisoft Ophthalmology database at UHB. Patients were identified from 2 database locations, diagnoses, and clinical observations, and were subdivided into 2 categories: (1) Diagnosis — those having an active diagnosis of any form of RVO recorded in Medisoft either via a clinical diagnosis, a final diagnosis, or a patient complaint; or (2) History — those having RVO mentioned on their past ophthalmic history or patient history.

Briefly, these patients were identified by searching 2 tables, the diagnosis table and the clinical observations table, for descriptions that were like "RVO" or like "vein occlusion." Certain exclusions were applied, including exclusion of "SH" (Social history), "FH" (Family History), and "Family History of Blindness." Specific exclusions were applied to prevent misclassification of patients who did not have RVO but might have it mentioned, i.e., "No evidence of RVO" (full exclusions listed in Supplementary Methods, available at www.ophthalmologyscience.org). Of the patients identified, 3 had no NHS number and were excluded from the data set. Together, this identified a total of 1521 patients that had a positive mention of RVO in their EHR. Of those with NHS numbers, 1367 patients had ≥ 1 diagnosis of RVO, and 503 had ≥ 1 mention of history of RVO (with some having both a diagnosis and history). For the 1521 patients, there were a total of 2196 occurrences of RVO affecting 2026 eyes available within this data set. These figures do not represent distinct cases of RVO because episodes describing multiple RVO types, or those where eye laterality was unknown, were counted as separate episodes. As such, a single episode of RVO may have been represented multiple times, making it impossible to determine the number of individual RVO cases. A subset of the patient cohort was manually validated to ensure data accuracy.

Reassigning Clinical Diagnosis as Historical. In some cases, clinical diagnoses were recorded in clinical observation tables that referenced historical diagnoses or disease. These cases were identified by clinical diagnoses that also included the terms "old," "previous," or "longstanding," or patient complaints that included "old vein," "resolved," "previous," "old right," "old left," and "longstanding." These cases were subsequently relabeled into the History category.

Reassigning History as Diagnosis. In situations where a diagnosis and a history were reported concurrently, it was assumed that this was the first instance of a patient being diagnosed with RVO. This permitted reclassification as a patient diagnosis, assuming the patient had no earlier mention of RVO.

Earliest Date of RVO. The historical disease and diagnoses were combined to populate a row per patient and an eyecode that contained the following: presence of historical disease and first date in history, earliest diagnosis date for that eye, and overall earliest date that RVO as a disease was mentioned regardless of source. A subset of patient records was manually checked to ensure accurate dates were included.

Eyecode Inferral. Eyecode information is readily available for those with a clinical diagnosis or a final diagnosis and not for those with historical information or patient complaint (PC). In cases where a PC and a diagnosis were reported concurrently, the eyecode and diagnosis date were inferred for the PC. Those without any available eyecode or without a diagnosis date (e.g., historical disease) were excluded from the eye-linked procedure and follow-up eye examination data sets.

In total, after classification of disease using the aforementioned methods, 290 patients were excluded, leaving 1231 patients (1355 eyes) with an RVO diagnosis. Of the 290 excluded patients, 284 had historical disease in total; 226 had a history alone, and 58 of these also had a diagnosis but would be excluded from eyecode analysis due to a lack of historical eyecode, and 6 patients had a PC of RVO and no eyecode. A further 6 patients had a concurrent GCA diagnosis and so were removed from the eye-linked data set, leaving 1225 patients (1349 eyes) available to be included in the eye-linked data set. Despite being eye-linked, reported figures for eye-linked investigations and procedures were subject to patients receiving these as part of their routine care.

Preprocessing/Cleaning/Labeling

Diagnosis Type. Diagnosis was readily available for those highlighted in the diagnosis table or from those with a clinical diagnosis that had been input into the clinical observations table. Some were mentioned only in the clinical notes, particularly in the case of historical disease. In these cases, to label patients as having specific diagnoses the clinical notes were searched for terms such as "BRVO," "CRVO," "HRVO," "branch," "hemi," and "central vein," and those that could not be identified as BRVO, CRVO, or HRVO were assigned unspecified RVO.

Table 1. Patient Demographics

Category	Subcategory	Patient Count
Diagnosis origin*	Diagnosis	1237
	History	226
	Diagnosis and history	58
	Total number of patients	1521
Age at entry to cohort (yrs)	≤ 30	10
	31–40	15
	41–50	58
	51–60	165
	61–70	295
	71–80	486
	81–90	427
	≥ 91	65
Sex	Female	783
	Male	738
Ethnicity	White – British	995
	Other	171
	Unknown	102
	Asian – Pakistani	59
	White – Irish	47
	Asian – Indian	40
	Black – Caribbean	39
	Asian - Bangladeshi	10
IMD decile (1st decile most deprived, 10th decile least)	1	377
	2	221
	3	190
	4	192
	5	165
	6	90
	7	128
	8	51
	9	48
	10	48
	Unknown	11
Diabetes	Type 1	24
	Type 2	484
	Not diabetic	367
	Unknown	645

Data set information for the number of patients, whether they have a diagnosis of RVO or a history or both, age at entry to cohort, sex, ethnicity, diabetic status, and IMD. Age at entry calculated from earliest diagnosis date, or date of first episode with a history of RVO.
IMD = indices of multiple deprivation; RVO = retinal vein occlusion.
*Medisoft was first introduced in 2010, and paper noting was phased out by 2014.

Identification of Retinal Artery Occlusion. A similar method as above was used to identify patients with retinal artery occlusion (RAO). Diagnoses were identified again from diagnosis and clinical observations tables by searching for the following terms: "artery occlusion" and "RAO." Additional exclusions were applied, including an exclusion of "SH" (Social History), "FH" (Family History), and "Family History of Blindness." There were also specific exclusions applied to prevent the misclassification of patients with the terms "intraocular" and "rayone" (rayone has RAO800c in the observation notes) and additional exclusions as listed in the Supplementary Methods.
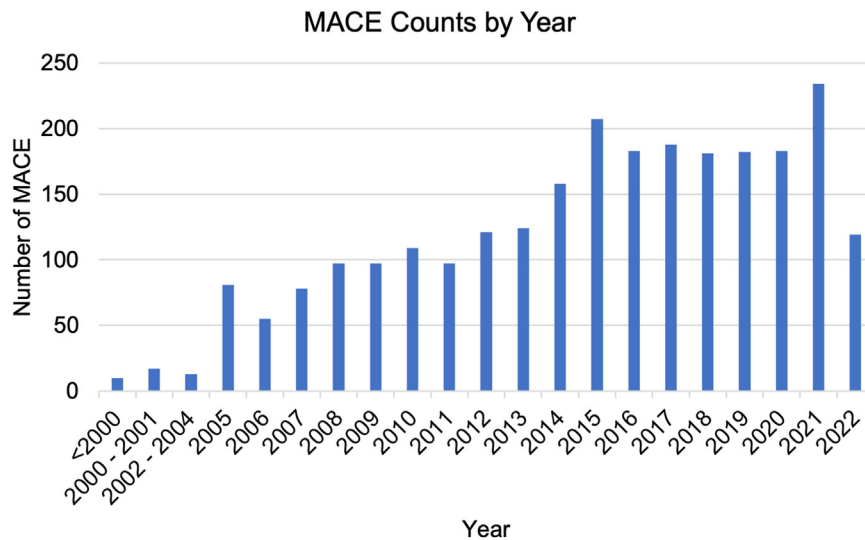
Figure 2. Major adverse cardiovascular event (MACE) counts by year*. The number of MACEs encountered by patients per year as of September 30, 2022. *Medisoft was first introduced in 2010, and paper noting was phased out by 2014.

Identification of GCA. Patients with GCA were identified from the UHB data warehouse using the International Classification of Diseases, 10th Revision codes M315 and M316 or SNOMED 239938009 and 239939001. These were categorized as pre- or post-RVO and were excluded

Table 2. Number of MACEs

Number of MACEs	Number of Patients Who Experience MACEs Before RVO	Number of Patients Who Experience MACEs after RVO	
1	144	224	
2	103	126	
3	63	74	
4	39	48	
5	34	20	
6	13	16	
7	15	11	
8	<	<	
9	<	<	
10	<	<	
11	<	<	
12	<	0	
13	0	0	
14	0	0	
Total patients	432	533	723*
Total MACEs	1244	1290	2534

The number of patients who experienced specific numbers of MACEs pre and post their first RVO diagnosis. If a MACE occurred on the same day as a diagnosis, this was classed as "post."
Patient groups with < 10 participants are denoted as "<".
MACE = major adverse cardiovascular events; RVO = retinal vein occlusion.
*There were instances when the same patient had both MACE before and after their RVO.
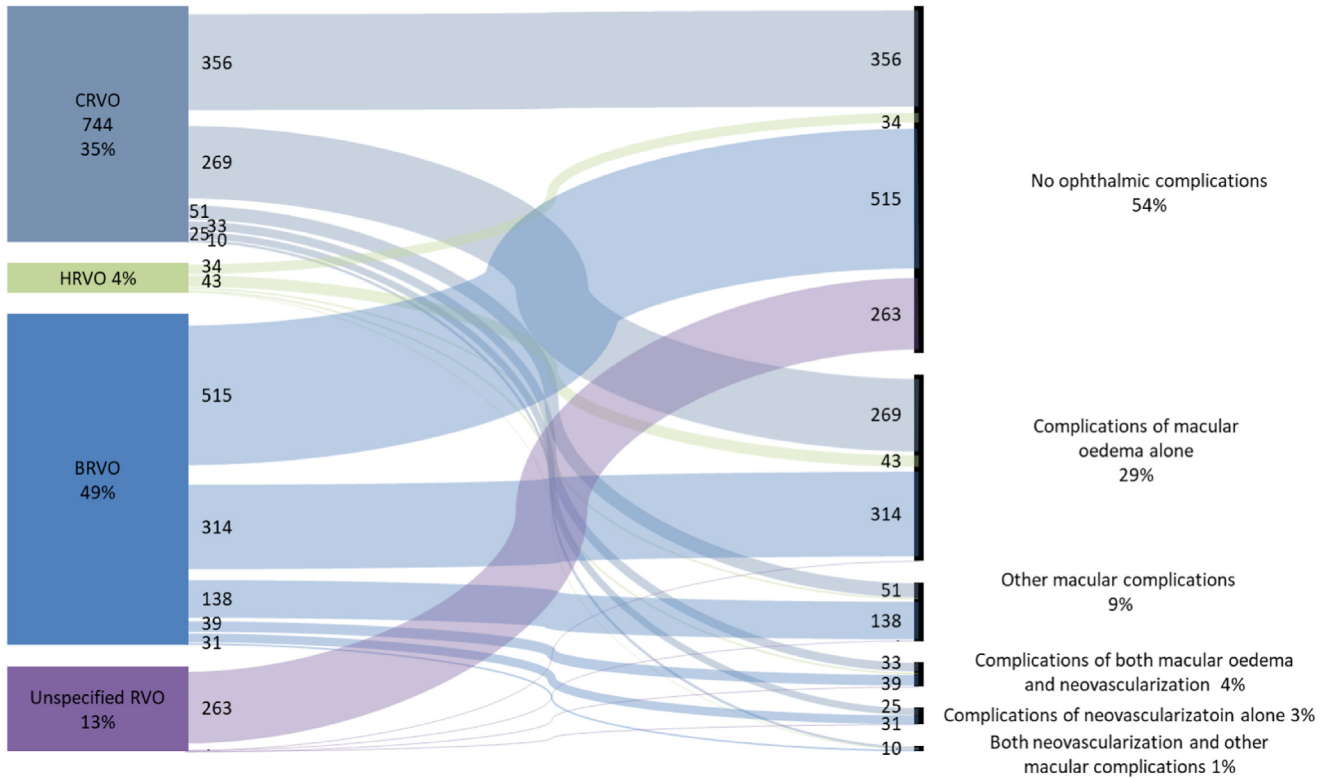
from later analyses due to their different underlying pathophysiology as compared with RVO and RAO.

Identification of Complications. Complications were identified from the diagnosis table and the clinical diagnoses in the observations table. Only diagnoses that occurred on or after the initial RVO diagnosis date were included.

For macular complications, the patient cohort diagnoses were searched for terms like "acular" before the lockdate of January 10, 2022. Neovascular complications were identified by applying search terms like "neovas" and "remodeling" to the patient cohort diagnoses before the lockdate of January 10, 2022. Exclusions were made for diagnoses highlighted by the search terms "no macular," "without macular," "no iris neovascularization," and "no evidence of."

MACE Codes. The data warehouse contained information relating to all patient diagnoses entered in the hospital EHRs. These diagnoses were searched for the International Classification of Diseases, Tenth Revision MACE codes (see Appendix B), and those identified were classed as pre-RVO diagnosis if it occurred before the first identified RVO date on the patient record, and post-RVO if it occurred on the same day as or after the first RVO date. Only the first instance of each International Classification of Diseases, Tenth Revision code was included in the presented numbers, as diagnoses may be entered at each visit.

General Procedure Analysis. Procedure analysis was performed generally on the data set to identify general, cardiac, vascular, and eye procedures, regardless of eyecode. Categorization of procedure codes is listed in Table S6.

Eyecode Procedure Analysis. Eye-linked procedures were included only for those diagnoses with eyecodes. Eye procedures, procedure operating procedure codes, supplement codes, and dates were included. Eye procedures were categorized as laser or injection, and injections were

Figure 3. Retinal vein occlusion (RVO) diagnosis and complications. Sankey diagram showing the breakdown of eyes, RVO diagnosis and complications after RVO diagnosis (all patients − unknown eyecodes are included as "N"). Note: some eyes may have multiple diagnoses (i.e., branch retinal vein occlusion [BRVO] and central retinal vein occlusion [CRVO]). HRVO = hemiretinal vein occlusion.
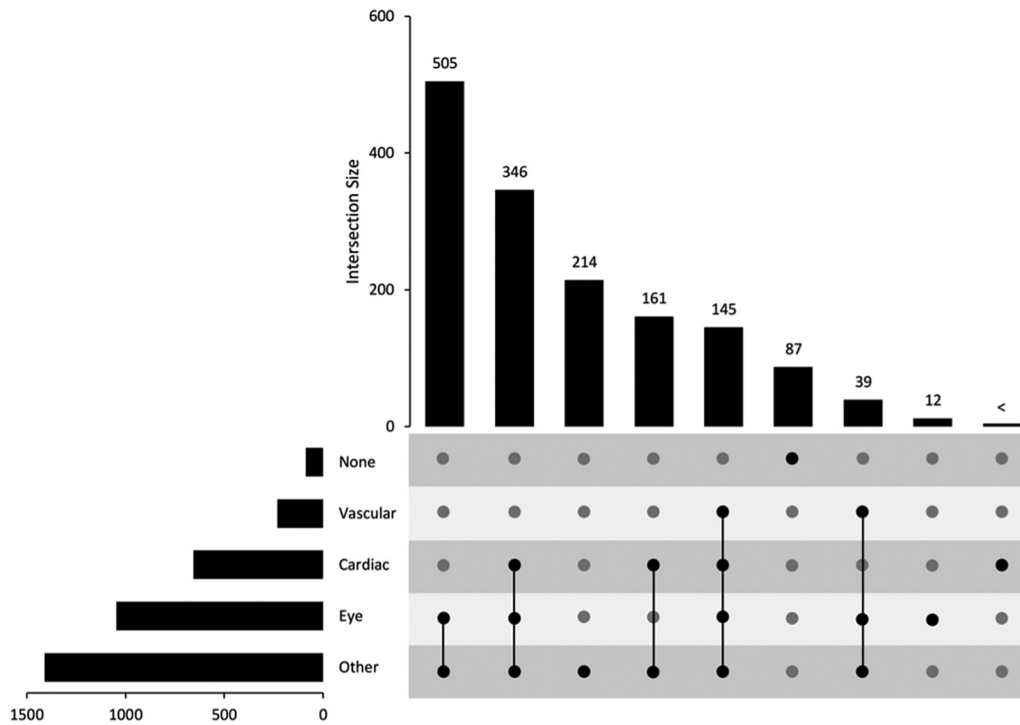

Figure 4. Retinal vein occlusion procedures. Upset diagram showing the breakdown of procedures that patients undergo. Patients may undergo > 1 procedure type, as indicated by the intersections between procedure types (none, vascular, cardiac, eye, and other). Patients may also have multiple procedures of each type. Patient groups with < 10 participants are denoted as "<".

Table 7. RVO Eye-Linked Procedures

Procedure Type (Eye-Linked)	Number of Patients	Number of Procedures	Laser	Total Injections	anti-VEGF Injections	Steroid Injections
Posterior segment	614	5290	656	4526	4276	241
Anterior segment	353	927	231	0	0	0
Uncertain	<	<	0	0	0	0

Procedure Type (Eye-Linked)	CRVO	BRVO	HRVO	Other
Posterior segment	2830	2881	606	11
Anterior segment	424	544	64	13
Uncertain	<	<	0	<

Eye procedures for those with a diagnosis of RVO and an eyecode available (eye-linked procedures).
Patient groups with < 10 participants are denoted as "<".
BRVO = branch retinal vein occlusion; CRVO = central retinal vein occlusion; HRVO = hemiretinal vein occlusion; RVO = retinal vein occlusion.

subdivided further to anti-VEGF and steroid injections. Procedures were then further categorized into front of the eye (anterior segment procedures), back of the eye (posterior segment procedures), or unable to be determined as detailed in Table S8, with numbers of each procedure type presented in Table 7.

Eyecode Eye Examinations. Eye examination data for the most common eye tests, such as visual acuity measurements, intraocular pressure, and visual field tests, were gathered via Medisoft or UHB data warehouse. OCT information is stored outside of Medisoft but available for inclusion in the data description and for data requests.

Further analysis permitted calculation of descriptive statistics such as the mean and median time to first test after diagnosis and follow-up interval within one year of diagnosis. Time to follow-up was calculated by ordering the tests by datetime (grouped by observation type) and calculating the days to the next follow-up in each case. This was then averaged to produce an average follow-up time. Follow-ups in the first year were designated (first year = 1), which enabled separation of analysis into first year and other years. Time to first test was ascertained by taking the earliest test date on or after the diagnosis. The mean and median

times to and between tests were calculated; median time was included because it is less skewed by outlier results.

Diabetic grading data as well as diabetic type are readily available for diabetic patients in Medisoft. Diabetic type was enriched from diagnosis records from UHB data warehouse.

Hospital Assessments. University Hospitals Birmingham data warehouse holds patient data relating to a wide range of tests and investigations. A small number of relevant tests were included in the data set. The number of patients and total number of tests in 3 months before or after and 1 year before or after diagnosis were summed (Table S10).

Medications. Patient medications are recorded in Medisoft and in the data warehouse. Medications were categorized into the following relevant groups: "Lipid lowering," "Hypertension," and "Diabetes" medications. Categorization was performed as detailed in Table S12.

Recommended Uses

The INSIGHT UHB RVO and MACE data set has been curated to provide tabular data to support research into RVO and MACEs. Currently this data set is being used to: describe the RVO cohort in UHB; improve understanding of

Table 9. Ophthalmic Examination Findings

Observation	Patients	Total Tests	Average Number of Tests Per Patient	Median Time to First Test	Median Time to Follow-up Test	Average Time to Follow-up Test	Average Total Tests Per Patient
Intraocular pressure	1200	12 940	3	0	117	187	9
Visual acuity test	1214	14 649	4	0	100	171	10
Visual field test	231	766	1	299	336	415	3
OCT images	919	6561	2	0	153	203	6
OCT	266	672	1	285	184	273	2
Optic disc size	556	1748	1	0	255	410	2
Lens status	1153	6609	2	0	203	338	5
Refraction	229	366	1	568	125	317	1

Number of patients, examinations, and eye-linked tests after a retinal vein occlusion diagnosis, including the median time to first test, and median/average time between follow-up tests (days) within the first year. Average total number of tests per patient in the data set is also reported. Multiple tests can occur on the same day; however, these were treated as a single test for presentation purposes.
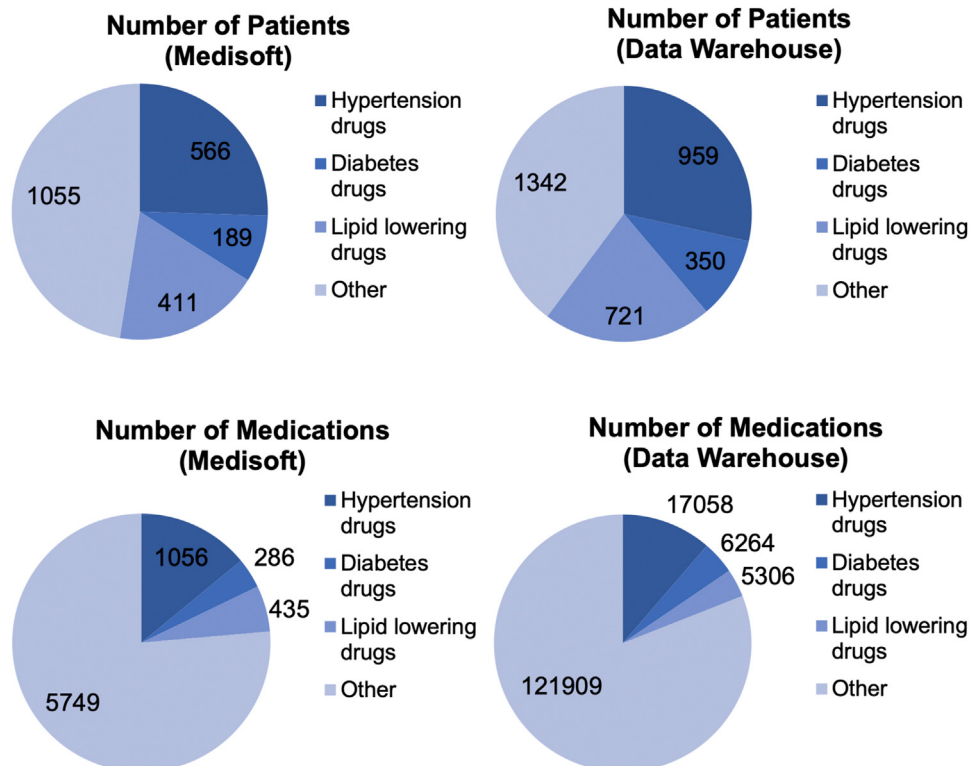
Figure 6. Retinal vein occlusion (RVO) patient medications. Breakdown of medications taken by RVO patients, split into categories of cardiac drugs, diabetes drugs, lipid lowering drugs, and any other drugs.

the course of RVO pathology in relation to subtype and visual outcome; aid discovery of novel associations between RVO and MACEs; and identify MACE critical risk time periods after RVO diagnosis. Further use of this data set holds great potential for future researchers. This data set could be recommended to: map RVO or MACEs to other systemic disease entities using routinely collected health data to establish novel laboratory or imaging biomarkers; explore therapeutic interventions and their impact on MACEs and all-cause mortality; provide nationwide epidemiological RVO data in conjunction with data sets from other locations across the UK; interrogate eye images for the training of machine learning models; and develop and validate artificial intelligence models thanks to the cohort's quality, scale, and diverse population.

Distribution

The INSIGHT UHB RVO and MACE data set will be available through submission of a data use application form to the Health Data Research UK Innovation gateway (https://www.healthdatagateway.org/) or the INSIGHT Hub (www.insight.hdrhub.org). Applicants are asked to detail their institution, researcher information, intended use of the data set, and where data will be accessed. Applications should demonstrate how the proposed project may benefit patients and the public, with the provision of a plain language summary of the project. Data use applications undergo a rigorous 3-stage appraisal process by the INSIGHT team and an independent Data Trust Advisory Board before

access to the data set is granted.[19,20] Throughout the appraisal, applications must demonstrate compliance with UK's general data protection regulation and Data Protection Act, national law, and "Five Safes" framework.[10,21,22]

Strengths and Limitations

This data set benefits from its diverse population, wealth of clinical information, and extensive follow-up. It is the largest RVO data set to date and, due to the continual pipelining of new patient information into the data set, it is anticipated to grow. Missing datafields are a likely consequence of the data set's method of data acquisition. Even with this limitation, this data set is likely to have good generalizability due to its inclusivity and high-quality data sourced from everyday healthcare encounters and clinical episodes.

Medisoft was used as this data set's primary source for data extraction. Medisoft was first introduced in 2010, with paper noting phased out by 2014. As such, periods before 2014 are likely to underrepresent the total number of RVO cases receiving care at UHB.

Data extraction covered periods encompassing the coronavirus disease 2019 pandemic. The pandemic caused a significant disruption across the globe, impacting on service provision to all specialties. During this period, there may have been fewer hospital attendees than expected, and those who attended may represent a cohort with more severe RVO disease and profound vision loss on presentation.

Summary

The INSIGHT UHB RVO and MACE data set is a longitudinal, anonymized data resource generated from routinely collected data of RVO NHS patients presenting to UHB. This datasheet was designed to outline the data set in a transparent and standardized manner to enhance its utility to researchers employing machine learning.[11] The datasheet contains an excerpt of the data set, covering a time-locked period from as early as 1978 to September 30, 2022. The data set contains clinical information of 1521 patients with 2196 occurrences of RVO and 2534 of MACEs. Further information related to demographics, visual acuity, laboratory test results, ophthalmic complications, and all-cause mortality are available within the data set. Access to the data set is granted by the Data Controller at UHB via the INSIGHT Health Data Research Hub (www.insight.hdrhub.org)

Footnotes and Disclosures

[1] INSIGHT, Health Data Research Hub for Eye Health, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom.

[2] Institute of Cardiovascular Sciences, College of Medical and Dental Sciences, University of Birmingham, United Kingdom.

[3] University Hospitals Birmingham, NHS Foundation Trust, United Kingdom.

[4] MRC Health Data Research UK (HDR UK) Midlands, United Kingdom.

[5] Academic Unit of Ophthalmology, Institute of Inflammation & Ageing, College of Medical and Dental Sciences, University of Birmingham, United Kingdom.

[6] Ophthalmology Department, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom.

*E.J.B. and E.J.G. contributed equally to this work.

Disclosure(s):

All authors have completed and submitted the ICMJE disclosures form.

The author(s) have made the following disclosure(s): A.K.D.: Grant — Health Data Research UK.

S.P.M.: Research support — UK Space Agency; Royalties — Springer publishing: Neuro-Ophthalmology, Global Trends in Diagnosis, Treatment and Management; Consultant — Neurodiem, Invex Therapeutics; Speaker fees — Allergan, Chiesi, Chugai-Roche Ltd, Heidelberg Engineering, Novartis, Santen, Santhera; Travel expenses — AbbVie, Santhera; Advisory boards — Invex Therapeutics, Janssen, Roche; Leadership — North American Neuro-ophthalmology Society, Journal of Neuro-Ophthalmology, European Neuro-ophthalmology Society, European Association of Neurological Surgeons, United Kingdom Neuro-ophthalmology Society, British Isle Neuro-Ophthalmology Club.

HUMAN SUBJECTS: Human subjects were included in this study. This study was conducted with the tenets of the Declaration of Helsinki and has approval from the relevant institutional review board and ethical approval from the West of Scotland Research Ethics Service (REC reference: 20/WS/0087). Approval has been given to retrospectively collate patient data without explicit written consent.

No animal subjects were used in this study.

Author Contributions:

Conception and design: Bradlow, Denniston, Mollan

Data collection: Bilton, Guggenheim, Baranyi, Radovanovic, Williams, Mollan

Analysis and interpretation: Bilton, Guggenheim, Baranyi, Mollan

Obtained funding: N/A.

Overall responsibility: Bilton, Guggenheim, Bradlow, Denniston, Mollan, Baranyi.

Abbreviations and Acronyms:

BRVO = branch retinal vein occlusion; **CRVO** = central retinal vein occlusion; **EHR** = electronic hospital record; **GCA** = giant cell arteritis; **HRVO** = hemiretinal vein occlusion; **MACE** = major adverse cardiovascular event; **NHS** = National Health Service; **PC** = patient complaint; **RAO** = retinal artery occlusion; **RVO** = retinal vein occlusion; **UHB** = University Hospitals Birmingham; **UK** = United Kingdom.

Keywords:

Biomedical data, Data set, Major Cardiovascular events, Myocardial infarction, Retinal vein occlusion.

Correspondence:
Edward J. Bilton, MBChB (MBBS), INSIGHT, Health Data Research Hub for Eye Health, University Hospitals Birmingham NHS Foundation Trust, Birmingham, West Midlands B15 2GW, United Kingdom. E-mail: edward.bilton2@nhs.net.

References

1. Laouri M, Chen E, Looman M, Gallagher M. The burden of disease of retinal vein occlusion: review of the literature. *Eye (Lond)*. 2011;25:981−988. https://doi.org/10.1038/eye.2011.92.
2. Gale R, Pikoula M, Lee AY, et al. Real world evidence on 5661 patients treated for macular oedema secondary to branch retinal vein occlusion with intravitreal anti-vascular endothelial growth factor, intravitreal dexamethasone or macular laser. *Br J Ophthalmol*. 2021;105:549−554. https://doi.org/10.1136/bjophthalmol-2020-315836.
3. Gale R, Gill C, Pikoula M, et al. Multicentre study of 4626 patients assesses the effectiveness, safety and burden of two categories of treatments for central retinal vein occlusion: intravitreal anti-vascular endothelial growth factor injections and intravitreal Ozurdex injections. *Br J Ophthalmol*.

2021;105:1571−1576. https://doi.org/10.1136/bjophthalmol-2020-317306.

4. Cugati S, Wang JJ, Knudtson MD, et al. Retinal vein occlusion and vascular mortality: pooled data analysis of 2 population-based cohorts. *Ophthalmology*. 2007;114:520−524. https://doi.org/10.1016/j.ophtha.2006.06.061.

5. Bakhoum CY, Madala S, Long CK, et al. Retinal vein occlusion is associated with stroke independent of underlying cardiovascular disease. *Eye (Lond)*. 2023;37:764−767. https://doi.org/10.1038/s41433-022-02038-x.

6. Woo SCY, Lip GYH, Lip PL. Associations of retinal artery occlusion and retinal vein occlusion to mortality, stroke, and myocardial infarction: a systematic review. *Eye (Lond)*. 2016;30:1031−1038. https://doi.org/10.1038/eye.2016.111.

7. Zhong C, You S, Zhong X, et al. Retinal vein occlusion and risk of cerebrovascular disease and myocardial infarction: a meta-analysis of cohort studies. *Atherosclerosis*. 2016;247:170−176. https://doi.org/10.1016/j.atherosclerosis.2016.02.024.

8. Mollan SP, Fu DJ, Chuo CY, et al. Predicting the immediate impact of national lockdown on neovascular age-related macular degeneration and associated visual morbidity: an INSIGHT Health Data Research Hub for Eye Health report. *Br J Ophthalmol*. 2023;107:267−274. https://doi.org/10.1136/bjophthalmol-2021-319383.

9. Wagner SK, Fu DJ, Faes L, et al. Insights into systemic disease through retinal imaging-based oculomics. *Transl Vis Sci Technol*. 2020;9:6. https://doi.org/10.1167/tvst.9.2.6.

10. UK Data Service. What is the Five Safes framework?. https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/. Accessed November 30, 2022.

11. Gebru T, Morgenstern J, Vecchione B, et al. Datasheets for datasets. *Commun ACM*. 2021;64:86−92. https://doi.org/10.1145/3458723.

12. Rogers S, McIntosh RL, Cheung N, et al. The prevalence of retinal vein occlusion: pooled data from population studies from the United States, Europe, Asia, and Australia. *Ophthalmology*. 2010;117:313−319.e1. https://doi.org/10.1016/j.ophtha.2009.07.017.

13. The Central Vein Occlusion Study Group. A randomized clinical trial of early panretinal photocoagulation for ischermic central vein occlusion: the Central Vein Occlusion Study Group N Report. *Ophthalmology*. 1995;102:1434−1444. https://doi.org/10.1016/S0161-6420(95)30848-2.

14. Ehlers JP, Fekrat S. Retinal vein occlusion: beyond the acute event. *Surv Ophthalmol*. 2011;56:281−299. https://doi.org/10.1016/j.survophthal.2010.11.006.

15. Nam GE, Han K, Park SH, et al. Retinal vein occlusion and the risk of dementia: a nationwide cohort study. *Am J Ophthalmol*. 2021;221:181−189. https://doi.org/10.1016/j.ajo.2020.07.050.

16. Public Health England. Health matters: preventing cardiovascular disease. https://www.gov.uk/government/publications/health-matters-preventing-cardiovascular-disease/health-matters-preventing-cardiovascular-disease. Accessed October 16, 2022.

17. Arnett DK, Blumenthal RS, Albert MA, et al. 2019 ACC/AHA guideline on the primary prevention of cardiovascular disease: executive summary: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *Circulation*. 2019;140:e563−e595. https://doi.org/10.1161/cir.0000000000000677.

18. Chen Q, Yu WH, Lin S, et al. Artificial intelligence can assist with diagnosing retinal vein occlusion. *Int J Ophthalmol*. 2021;14:1895−1902. https://doi.org/10.18240/ijo.2021.12.13.

19. INSIGHT. The Data Trust Advisory Board is a key aspect of INSIGHT's governance. https://www.insight.hdrhub.org/data-trust-advisory-board. Accessed November 30, 2022.

20. Denniston AK, Kale AU, Lee WH, Mollan SP, Keane PA. Building trust in real-world data: lessons from INSIGHT, the UK's health data research hub for eye health and oculomics. *Curr Opin Ophthalmol*. 2022;33:399−406. https://doi.org/10.1097/icu.0000000000000887.

21. Information Commissioner's Office. UK GDPR guidance and resources. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/. Accessed November 30, 2022.

22. Stokes P. Office for National Statistics. The "Five Safes" - Data Privacy at ONS. https://blog.ons.gov.uk/2017/01/27/the-five-safes-data-privacy-at-ons/. Accessed November 30, 2022.