



HHS Public Access

Author manuscript

J Neuroophthalmol. Author manuscript; available in PMC 2020 December 01.

Published in final edited form as:

J Neuroophthalmol. 2019 December ; 39(4): 480–486. doi:10.1097/WNO.0000000000000751.

Big Data Research in Neuro-ophthalmology: Promises and Pitfalls

Heather E. Moss, MD PhD,

Stanford University, Departments of Ophthalmology and Neurology & Neurological Sciences, Palo Alto, CA, USA, hemoss@stanford.edu

Charlotte E. Joslin, OD, PhD,

University of Illinois, Departments of Ophthalmology and Visual Sciences, College of Medicine, and Epidemiology and Public Health, School of Public Health, Chicago, IL, USA

Daniel S. Rubin, MD,

University of Chicago, Department of Anesthesia and Critical Care, Chicago, IL, USA

Steven Roth, MD

University of Illinois, College of Medicine, Departments of Anesthesiology, and Ophthalmology and Visual Sciences, Chicago, IL, USA

Abstract

Background: Big data clinical research involves application of large data sets to the study of disease. It is of interest to neuro-ophthalmologists but also a challenge because of the relative rarity of many of the diseases treated limit prospective investigation.

Corresponding Author: Heather Moss, MD PhD, Spencer Center for Vision Research, 2370 Watson Court, MC 5349, Palo Alto, CA 94303, hemoss@stanford.edu, 650-498-9370.

Statement of Authorship

Category 1:

a. Conception and design

HM, CJ, DR, SR

b. Acquisition of data

HM, CJ, DR, SR

c. Analysis and interpretation of data

HM, CJ, DR, SR

Category 2:

a. Drafting the manuscript

HM, CJ, DR, SR

b. Revising it for intellectual content

HM, CJ, DR, SR

Category 3:

a. Final approval of the completed manuscript

HM, CJ, DR, SR

Conflict of interest disclosure: none

Evidence acquisition: Evidence for this review was gathered from the authors' experiences performing analysis of large data sets and review of the literature.

Results: Big data sets are heterogenous including prospective surveys, medical administrative and claims data and registries compiled from medical records. High quality studies must pay careful attention to data set selection including bias, data management including missing data, variable definition and statistical modeling in order to generate appropriate conclusions. There are many studies of neuro-ophthalmic diseases that utilize big data approaches.

Conclusions: Big data clinical research studies complement other research methodologies to advance understanding of human disease. A rigorous and careful approach to data set selection, data management, data analysis and data interpretation characterizes high quality studies.

Keywords

big data; databases; machine learning; outcomes; neuro-ophthalmology

Big data refers to large volume data sets that are analyzed for use in a variety of fields including government, media, retail and healthcare, and the methods used to perform the analyses. There is no size threshold that defines big data, rather it is characterized by being challenging to manage and process, which is a subjective and time-dependent definition (1). For the purposes of this review, we consider big data to be defined by large sample sizes that was collected in a general manner (i.e. not targeted to study a specific disease). Big data in medical research is not new, with epidemiology having its roots in analysis of large collections of data, and medical cohort studies having been in process for decades (e.g. Framingham Heart Study, founded in 1948) (2). Interest in big data-based research has accelerated recently with the rapid growth of data collection in minable formats, improved storage capabilities and statistical methods, and faster computing speed, facilitating the use of bigger, broader data sets to investigate novel questions of disease risk and outcome (3).

The last decade has witnessed an enormous explosion in availability of health care-related databases and pursuit of "secondary analyses" from this information (4). There is considerable enthusiasm in the clinical research community for "real world" data collected outside the confines of rigidly structured clinical trials to study risk factors, treatment effects, incidence, prevalence and outcome of diseases, as well as treatment strategies. However, the quality and variability of the data, often collected for another purpose, and the complexity of the analytic techniques necessitate increasing expertise by both researchers and stakeholders to interpret results appropriately (5). As a result, the "promise" that big data might change medical practice is still met with skepticism from some practitioners.

Germane to practicing neuro-ophthalmologists is big data-based clinical research, which will be our focus. We highlight salient features of different types of big data sets, common pitfalls in analysis and interpretation, new developments in analytics including machine learning, and provide examples of recent, relevant studies.

Data Sources

Big data sets are heterogeneous in origin with varied sampling strategies and contents. A common result is inherent selection bias, due to non-random selection of individuals. The sample may not represent the population, influencing generalizability of the studies (6). Examples of different kinds of big data sources are given below. For further details, the reader is referred to the Vision and Eye Health Surveillance System(7) (VEHSS) maintained by the United States Centers for Disease Control (CDC), which compiles and maintains excellent summaries of many relevant data sources in vision research.

National Health and Nutrition Examination Survey (NHANES) – a survey data set:

Surveys are collected prospectively as population-based samples. The NHANES assesses nutrition and health of the US by administering a survey to approximately 5000 individuals every 2 years. The data has included vision questions, eye examinations, visual field testing, and retinal imaging. Although data is available for the 2015-16 administration, eye data was most recently collected in 2008. Data collection procedures are rigorous (8), with ample literature and publicly available characterization of the data set (9),

Though data validity and reliability are typically excellent, use of survey data to study neuro-ophthalmic disease is limited by sample sizes. A population sample of 5000 individuals is unlikely to contain very many cases of neuro-ophthalmic disease. An alternative is a survey with a sample enriched in disease by sampling from a population seeking medical care. For example, the National Ambulatory Medical Care Survey and the National Hospital Ambulatory Medical Care Survey have been applied to quantify volume and types of medical visits for diplopia (10). Survey data sets have demonstrated utility in study of ophthalmic markers for common neurological diseases (e.g., retinal vessel measurements for risk stratification of cerebrovascular disease in the Atherosclerosis Risk in Communities Study study)(11, 12) and study of the relationship between vision impairment and common neurological conditions (e.g., visual and cognitive impairment in the Salisbury Eye Study) (13).

There is no direct opportunity for physicians to contribute to NHANES, which is administered by the US government. The data is publicly available for “purposes of health statistical reporting and analysis.” Some information and linkage to other government data sets is available via a fee through application to the National Center for Health Statistics Research Data Center.(14) Access to other survey data sets such as cohort studies is at the discretion of study management.

National Inpatient Sample (NIS) – an administrative data set: These are compiled by sampling medical administrative data. The NIS is a 20% stratified sample of non-federal US hospital discharges starting in 1988 and currently available through 2016. It is compiled and maintained by the US Agency for Healthcare Research and Quality (15). It contains 5-8 million hospital stays and weighting can be applied to obtain population estimates. The data are cross sectional and not identifiable by individual, with data points including demographics, medical diagnoses, procedures, costs and hospital variables extracted from hospital administrative databases maintained by states. The diagnoses and procedures are

coded using International Classification of Diseases (ICD) taxonomy. Both under- and over-coding are possible leading to information (misclassification) bias.

Utility of administrative data for study of neuro-ophthalmic disease depends on sample size of the disease of interest. Administrative samples are usually population samples with the advantage of generalizability and they tend to be larger than survey data sets, likely because data collection is less intensive. Due to the larger database size and enrichment with neuro-ophthalmic diseases due to derivation from health care records, this format has been applied effectively to the study of neuro-ophthalmic conditions including perioperative ischemic optic neuropathy and retinal artery occlusion with spine and cardiac surgery (16–20). A major limitation of NIS for neuro-ophthalmic conditions is the inability to capture those events that do not result in hospital admission or medical coding.

There is no opportunity for physicians to contribute to the NIS, which is sampled from state managed data bases. Purchase of data costs between \$160-\$500 per year of data. Databases such as NIS that contain a risk of identifying patients require a data use agreement and may require training in proper use and protection of data.

Clinformatics datamart (Optum Inc) – a commercial insurance claims data set:

These are compiled by amalgamating insurance claims submitted by health care providers for medical care. Clinformatics contains longitudinal information for over 50 million covered individuals by a large US national health insurer (21). Thus, it is not a population sample. Data points for diagnoses and procedures rely on coding taxonomies (ICD and current procedural technology (CPT)), a source of information bias. In addition to diagnoses and procedure information, Clinformatics includes demographics, inpatient and outpatient provider, facility and pharmacy claims. There is laboratory data available for a subset of individuals. Linkage to zip codes, socioeconomic status or death records is available.

Despite potential selection bias due to sampling and information errors from over and under coding, the longitudinal data and very large sample size facilitate study of long-term outcomes and identification of risk factors in neuro-ophthalmic diseases. Clinformatics has been applied to study risk of thyroid associated orbitopathy in Graves disease (22), risk factors for branch retinal vein occlusion(23) and risk factors nonarteritic anterior ischemic optic neuropathy (NAION) (21). Other claims databases that have been applied to neuro-ophthalmic diseases include the Medicare 5% claims sample to study association between diabetes mellitus and NAION (24), the LifeLink database (IMS Health Inc.) to study association between medications and secondary pseudotumor cerebri (25) or NAION(26) and MarketScan (Truven Health Analytics Inc.) to study the association between uveitis and optic neuritis (27). These data bases offer excellent opportunity for further study of neuro-ophthalmic disorders.

There is no direct opportunity for physicians to contribute to Clinformatics or other claims data bases. Purchase of commercial claims datasets (e.g. Clinformatics, MarketScan, LifeLink) carries substantial cost and typically is negotiated at an institutional level, often with prices exceeding \$15,000. Medicare data fees are structured according to the number of subjects, amount of data and time period, estimated at over \$10,000 for basic claims

information for up to 1 million beneficiaries for a year (28). Data use agreements usually are required.

IRIS (American Academy of Ophthalmology) & Axon (American Academy of Neurology) electronic medical records data sets:

Medical records data sets are compiled by amalgamating medical records from providers. Ambitious efforts have recently resulted in the American Academy of Ophthalmology Intelligent Research in Sight (IRIS) registry (29, 30) and the American Academy of Neurology Axon Registry (31) (32). In 2016, IRIS collected data from over 36 million clinical visits for over 17 million unique patients seen by over 10 000 providers in ophthalmology practices (42% of US practicing ophthalmologists). Prevalence of conditions relevant to neuro-ophthalmologists was: 2.04% optic nerve disorders (excluding glaucoma), and 1.96% strabismus (33). Axon has captured over 4 million visits for 1.3 million unique patients and has over 1000 participating neurology providers, 20 of whom are neuro-ophthalmology related (Personal communication, Katie Hentges, Program Manager, Registry, American Academy of Neurology, June 29, 2018). Both registries are actively enrolling new providers. There is selection bias since the registries are not population samples. Rather, the data is based on which practices participate, with academic practices currently under represented in both registries due to challenges in setting up data extraction from electronic health record (EHR) used in academic medical health systems.

It is important to point out that the IRIS and Axon registries were established primarily for quality improvement, benchmarking, and to comply with insurance-based incentive provider payment systems, rather than for research. Both registries collect the entirety of the medical record for visits with the participating provider. Extracted data points including performance measures (both Axon and IRIS) and some other data fields (IRIS) are based on mapping between the provider's EHR and the registries. While the potentially available information is large, analyzable information is limited by what has been mapped (34). There is future opportunity through additional mapping to specific EHR fields and application of natural language processing to create comprehensive data sets that capture clinical neurologic and ophthalmic care. A limitation is that both registries are limited to the visit record of the providers and do not include external records, provider notes in other specialties, operating room data and images except as referenced in the enrolled provider's record.

Neither the Axon nor IRIS registries have published research related to neuro-ophthalmic disease. However, there is rich publication in regional and national registries including studies of third nerve palsy and idiopathic intracranial hypertension (IIH) using the Rochester Epidemiology Project (35, 36), quantifying the incidence of ocular symptoms following a diagnosis of giant cell arteritis using the Swedish Hospital Registry (37), and study of cranial nerve palsies in diabetic patients using a Saudi diabetes registry (38).

Ophthalmology practices can participate in IRIS and Neurology practices can participate in Axon through application with the sponsoring society. Setup time is required to establish practice-specific mapping of variables and data transmission. There is no fee for participation, but membership in the sponsoring organization and US practice are required. Participants are given access to their performance metrics. Access to IRIS for research

currently is limited to a competitive grant process administered jointly by the AAO and Research to Prevent Blindness Inc. A separate AAO fund is being established to support young investigators to perform IRIS based research, but application details are not yet available. Subspecialty societies are invited to apply to the AAO sponsor IRIS based research and the American Glaucoma Society is sponsoring an award this year.(39) Axon is not currently accessible for research use, although this is planned for the future and expected to have a similar process to IRIS.

Opportunities for neuro-ophthalmology:

With regards to data sources there is much research that can be pursued in neuro-ophthalmic diseases as well as vision outcomes in neurological diseases using existing data.

Participation in professional organization sponsored registries will enrich the data sets as well as provide individual benefits with regards to performance measurements. There is an opportunity for neuro-ophthalmology professional organizations to sponsor awards that would enable their members to access restricted data sets such as the IRIS registry and to sponsor neuro-ophthalmology focused data sets.

Analytical techniques

Consultation with a biostatistician or other individual well versed in big data analysis is extremely important throughout the research process to ensure appropriate formulation of the research question, data management practices, modeling and interpretation.

Research questions:

In the context of observational studies, big data sets offer the promise of large sample sizes and are particularly attractive for studying rare diseases such as those seen in neuro-ophthalmology. These large samples can be used in cross-sectional, retrospective cohort, case-control, and other study designs to define incidence and prevalence and identify potential risk factors. This has application to risk stratification and prediction which typically requires a large sample and population breadth to generate robust, clinically useful results (40). Another promise of big data, particularly those derived from clinical care records is “real world” experience in contrast to the idealized treatment and testing structure of gold standard randomized interventional trials. The data can be leveraged to investigate questions that may be practically difficult, financially prohibitive or ethically challenging to address prospectively (e.g., drug-induced diseases). Studies can be exploratory and hypothesis generating but also can test hypotheses directly.

Power analysis is essential to ensure adequate sample size for the research questions asked and analytical techniques used. While big data sets are attractive for their size, a rare condition may still have a relatively small sample size in a big data set and this can limit both the analyses and the conclusions (16, 19).

Data Management

Data management decisions including how raw data is used to define variables are an important foundation for subsequent analysis. Too broad a definition biases towards the null,

while too narrow a definition can limit sample size and study power. When done appropriately, these can help to address problems with completeness and consistency in the raw data, but otherwise can skew the results. For example, requiring an observation period of 5 years vs. 1 year before a diagnosis to define it as incident, reduced over-estimation of glaucoma incidence from 135% to <30% (41). With regards to idiopathic intracranial hypertension, ICD codes from emergency department visits have a 55% positive predictive value (42). One strategy is to require certain tests in addition to a diagnosis code. However, this is not fool proof with less than 70% of patients with an ICD code for IIH and CPT codes for neuro-imaging and lumbar puncture meeting criteria for IIH on medical record review (43). Due to the nuances of diagnosing neuro-ophthalmic conditions, neuro-ophthalmic experts have raised serious concerns regarding the accuracy of diagnostic definitions for optic neuritis and ischemic optic neuropathy used in recent studies led by non-neuro-ophthalmologists (44, 45). One study found a false positive rate of 60% for optic neuritis diagnoses by non-neuro-ophthalmologists (46).

Selection of control subjects (i.e. those without the exposure or outcome of interest) also requires careful thought to ensure accurate classification of dependent and independent variables (47). For example, in a study of IIH using medical claims data, selecting controls from a population with a prior eye exam or without a diagnosis of headache may decrease misclassification bias. Similarly, if glaucoma is an independent variable of interest, then it is important that cases and controls have similar eye examination histories. Sparse data bias can occur when there are insufficient cases with some combinations of predictive variables and this can bias away from the null, predicting large effect sizes (48). One strategy to address this is matching controls to cases with strong risk factors, those for which the exposure – outcome relationship has a large effect size.

Missing data, unavoidable and common with epidemiologic research, clinical trials, and big data in particular, can lead to biased estimates and reduced precision that significantly affect conclusions. Therefore, it is one of the most critical elements that must be acknowledged, described and addressed in any study. There are three general types of missing data: missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) (49). Only MCAR, in which missingness does not depend on other variables but rather on only random events, yields unbiased estimates. MAR is missingness dependent upon another observed variable, while MNAR missingness, sometimes called “non-ignorable non-response” or “informative missingness” is dependent upon another unobserved variable. As an example, income that is missing based on an observed variable (e.g., sex) would be MAR, but income missing based on whether the income is high or low, which is unobserved for subjects with missing income variables, would be MNAR.

General analytic strategies to address missing data include complete case analysis, multiple imputation (49), likelihood-based methods(50) and inverse probability weighting. Complete case analysis is based on the subset without missing data in either the outcome or covariates and will produce biased samples unless 1) data missingness is MCAR; or, 2) the overall rate of missing data is small (e.g., < 5% of total sample; impact of bias likely to be small). Complete case analysis results in a loss of efficiency (e.g., larger standard errors). Recommendations for missing data generally lean towards complete case analysis when the

overall missing rate of the total sample is small (e.g., < 5%), regardless of the type of missing data as the impact is low.

Multiple imputation is a common approach to MAR. Multiple imputed values for each missing observation are generated using carefully specified joint imputation models that reflect the uncertainty of the missing value. A statistical model is fit to each complete dataset, and the results of multiple separate analyses (for each imputation) are combined to account for the uncertainty in the imputation (49). More naïve imputations (e.g., substitution of means, last observation carried forward), can be worse than complete case analysis. Multiple imputation can be valuable when missing data are restricted to covariates and not outcome data.

Likelihood-based approaches model subjects with complete and partial data together and exclude only observations with both covariates and outcomes missing but require sophisticated analytic techniques with assumptions based on the nuisances of covariate distribution. Inverse probability weighting corrects for bias of estimates obtained with complete case analysis, in which each individual is given a sampling weight and the probability for selection proportional to this weight (50). Inverse probability weighting may be most applicable if there is a large amount of missing outcome data, which cannot be strongly modeled with covariates through multiple imputation. Finally, when the same variables are used in analysis, multiple imputation and likelihood-based methods yield similar results.

Modeling:

With regards to evaluating relationships between data one approach is to use multivariable models with outcome as a function of exposure and covariates. These include logistic regression, Cox regression, mixed models and generalized estimating equations. Included variables are typically based on prior information from the literature or the investigators' framework. Often univariate comparisons of outcome to each predictive variable and stratified comparisons of outcome to exposure by predictive variable level are used to inform the initial models. Other strategies include incrementally adding terms to models or incrementally removing them to arrive at the final model. Techniques such as propensity scores and mediation analysis can be used to address issues of confounding and questions of causality respectively. The sample size determines the number parameters that can be accommodated to identify associations between dependent and independent variables as well as interactions between independent variables, while accounting for a wide range of potentially confounding variables. For example, a general rule of thumb for logistic regression is analyzing a minimum of 10 events per analysis variable (51).

Another approach is data mining or machine learning techniques without prior identification of relevant variables. The techniques are typically supervised (as opposed to unsupervised) in that the outcomes of interest are defined by the investigator and the goal of analysis is to identify patterns in the data associated with the outcome (52). From a research perspective, these are hypothesis-generating rather than hypothesis-testing. They have particular application to the development of predictive models. Machine learning also is increasingly

popular for its ability to analyze image data, with clear relevance to neuro-ophthalmology, where fundus and OCT imaging have been evaluated by this methodology (53–55).

Given a large data set and commonly available statistical software, it becomes relatively straightforward for a novice to run models and generate numerical statistical results. However, planning and selecting the appropriate analyses requires expertise. A practical note is that analysis of large quantities of data, (e.g., from Medicare or NIS), may require high speed computing resources due to insufficient memory on personal computers. Usually, this will result in increased costs and necessity for programming expertise.

Opportunities for neuro-ophthalmology:

Beyond utilizing big data set research as a tool to investigate and advance understanding of neuro-ophthalmic disease, the neuro-ophthalmology community can make important contributions in development and validation of algorithms for accurate classification of neuro-ophthalmic disease from claims and administrative data.

Interpretation—Research study results are only applicable to the extent that the data they are based on is appropriate with acknowledged limitations, the analysis is appropriate and the conclusions reasonable. With large sample sizes it becomes more likely to have statistically significant associations with small effect sizes that make them clinically irrelevant. Due to the observational nature of most big data sets the analyses detect association and do not imply causation. Another risk is spurious correlations, which are statistically identified associations that are either coincidental or related to a common cause. As with any research study, a big data study does not stand alone but must be interpreted in the context of the broad literature for the disease of interest.

Attempts to improve study design and reporting include STROBE and RECORD reporting standards (56–58). Concerns have been raised about approaches to report and evaluate data collected in longitudinal big data studies impacting drug therapy (59).

Conclusion

The increasing ease of data collection, storage, and analysis has increased enthusiasm for big data analysis in medical research, and clinical care as well as many applications outside of medicine. The large sample sizes and real-world observations are promising as a basis for research questions that cannot practically be answered using clinical trials and for hypothesis generation. However, big data remains simply a collection of data sources. Careful data selection, management analysis and interpretation are critical to generate meaningful conclusions. There are numerous and increasing opportunities for further research using these databases to study neuro-ophthalmic diseases, most of which are low in incidence and prevalence.

Acknowledgments

Funding: National Institutes of Health (Bethesda, MD) grants R21 EY027447 (to Dr. Roth), K23 EY 024345 (to Dr. Moss), P30 EY 026877 to the Department of Ophthalmology at Stanford University, P30 EY001792 to the Department of Ophthalmology at the University of Illinois at Chicago, unrestricted grants from Research to Prevent

Blindness, Inc. (New York, NY) to the Stanford Department of Ophthalmology and to the University of Illinois at Chicago Department of Ophthalmology and Visual Sciences, and by the Michael Reese Foundation (Chicago, IL) Pioneers Award to Dr. Roth.

References

1. Press G 12 Big Data Definitions: What's Yours? Forbes [Internet]. 2014 [cited 2018 October 3]. Available from: <https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#19f939ca13ae>.
2. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. *Lancet*. 2014;383:999–1008. [PubMed: 24084292]
3. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med*. 2016;375:1216–1219. [PubMed: 27682033]
4. Whiteley WN, Emberson J, Lees KR, Blackwell L, Albers G, Bluhmki E, Brott T, Cohen G, Davis S, Donnan G, Grotta J, Howard G, Kaste M, Koga M, von Kummer R, Lansberg MG, Lindley RI, Lyden P, Olivot JM, Parsons M, Toni D, Toyoda K, Wahlgren N, Wardlaw J, Del Zoppo GJ, Sandercock P, Hacke W, Baigent C. Risk of intracerebral haemorrhage with alteplase after acute ischaemic stroke: a secondary analysis of an individual patient data meta-analysis. *Lancet Neurol*. 2016;15:925–933. [PubMed: 27289487]
5. Schneeweiss S Learning from big health care data. *N Engl J Med*. 2014;370:2161–2163. [PubMed: 24897079]
6. Sessler DI, Imrey PB. Clinical research methodology 2: observational clinical research. *Anesth Analg*. 2015;121:1043–1051. [PubMed: 26378704]
7. Dupont WD, Plummer WD. Power and sample size calculations: a review and computer program. *Control Clin Trials*. 1990;11:116–128. [PubMed: 2161310]
8. National Health and Nutrition Examination Survey (NHANES): Ophthalmology Procedures Manual [CDC NHANES document]. 9 2005 Available at: <https://wwwn.cdc.gov/nchs/data/nhanes/2005-2006/manuals/OP.pdf>. Accessed June 6, 2018.
9. Patton N, Aslam T, MacGillivray T, Pattie A, Deary IJ, Dhillon B. Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: a rationale based on homology between cerebral and retinal microvasculatures. *J Anat*. 2005;206:319–348. [PubMed: 15817102]
10. De Lott LB, Kerber KA, Lee PP, Brown DL, Burke JF. Diplopia-related ambulatory and emergency department visits in the united states, 2003-2012. *JAMA Ophthalmol*. 2017;135:1339–1344. [PubMed: 29075739]
11. Hubbard LD, Brothers RJ, King WN, Clegg LX, Klein R, Cooper LS, Sharrett AR, Davis MD, Cai J. Methods for evaluation of retinal microvascular abnormalities associated with hypertension/sclerosis in the Atherosclerosis Risk in Communities Study. *Ophthalmology*. 1999;106:2269–2280. [PubMed: 10599656]
12. Seidelmann SB, Claggett B, Bravo PE, Gupta A, Farhad H, Klein BE, Klein R, Di Carli M, Solomon SD. Retinal vessel calibers in predicting long-term cardiovascular outcomes: the Atherosclerosis Risk in Communities Study. *Circulation*. 2016;124:1328–1338.
13. Zheng DD, Swenor BK, Christ SL, West SK, Lam BL, Lee DJ. Longitudinal associations between visual impairment and cognitive functioning: The Salisbury Eye Evaluation Study. *JAMA Ophthalmol*. 2018.
14. Sinclair AJ, Burdon MA, Nightingale PG, Ball AK, Good P, Matthews TD, Jacks A, Lawden M, Clarke CE, Stewart PM. Low energy diet and intracranial pressure in women with idiopathic intracranial hypertension: prospective cohort study. *BMJ*. 2010;341.
15. Sugerman HJ, Felton WL III, Sismanis A, Kellum JM, DeMaria EJ, Sugerman EL. Gastric surgery for pseudotumor cerebri associated with severe obesity. *Ann Surg*. 1999;229:634–640. [PubMed: 10235521]
16. Rubin DS, Matsumoto MM, Moss HE, Joslin CE, Tung A, Roth S. Ischemic optic neuropathy in cardiac surgery: incidence and risk factors in the United States from the National Inpatient Sample 1998 to 2013. *Anesthesiology*. 2017; 126:810–821. [PubMed: 28244936]

17. Calway T, Rubin DS, Moss HE, Joslin CE, Beckmann K, Roth S. Perioperative Retinal Artery occlusion: Risk Factors in Cardiac Surgery from the United States National Inpatient Sample 1998-2013. *Ophthalmology*. 2017;124:189–196. [PubMed: 27914836]
18. Calway T, Rubin DS, Moss HE, Joslin CE, Mehta AI, Roth S. Perioperative retinal artery occlusion: incidence and risk factors in spinal fusion surgery from the US National Inpatient Sample 1998-2013. *J Neuroophthalmol*. 2018;38:36–41. [PubMed: 28665867]
19. Rubin DS, Parakati I, Lee LA, Moss HE, Joslin CE, Roth S. Perioperative visual loss in spine fusion surgery: ischemic optic neuropathy in the United States from 1998 to 2012 in the Nationwide Inpatient Sample. *Anesthesiology*. 2016;125:457–464. [PubMed: 27362870]
20. Lee YC, Wang JH, Huang TL, Tsai RK. Increased risk of stroke in patients with nonarteritic anterior ischemic optic neuropathy: A nationwide retrospective cohort study. *Am J Ophthalmol*. 2016;170:183–189. [PubMed: 27521605]
21. Cestari DM, Gaier ED, Bouzika P, Blachley TS, De Lott LB, Rizzo JF, Wiggs JL, Kang JH, Pasquale LR, Stein JD. Demographic, systemic, and ocular factors associated with nonarteritic anterior ischemic optic neuropathy. *Ophthalmology*. 2016;123:2446–2455. [PubMed: 27659545]
22. Stein JD, Childers D, Gupta S, Talwar N, Nan B, Lee BJ, Smith TJ, Douglas R. Risk factors for developing thyroid-associated ophthalmopathy among individuals with Graves disease. *JAMA Ophthalmol*. 2015;133:290–296. [PubMed: 25502604]
23. Newman-Casey PA, Stem M, Talwar N, Musch DC, Besirli CG, Stein JD. Risk factors associated with developing branch retinal vein occlusion among enrollees in a United States managed care plan. *Ophthalmology*. 2014;121:1939–1948. [PubMed: 24953793]
24. Lee MS, Grossman D, Arnold AC, Sloan FA. Incidence of nonarteritic anterior ischemic optic neuropathy: increased risk among diabetic patients. *Ophthalmology*. 2011;118:959–963. [PubMed: 21439645]
25. Sodhi M, Sheldon CA, Carleton B, Etminan M. Oral fluoroquinolones and risk of secondary pseudotumor cerebri syndrome: nested case-control study. *Neurology*. 2017;89:792–795. [PubMed: 28754842]
26. Nathoo NA, Etminan M, Mikelberg FS. Association between phosphodiesterase-5 inhibitors and nonarteritic anterior ischemic optic neuropathy. *J Neuroophthalmol*. 2015;35:12–15. [PubMed: 25295683]
27. Guo D, Liu J, Gao R, Tari S, Islam S. Prevalence and incidence of optic neuritis in patients with different types of uveitis. *Ophthalmic Epidemiol*. 2018;25:39–44. [PubMed: 28895765]
28. Corbett JJ, Savino PJ, Thompson HS, Kansu T, Schatz NJ, Orr LS, Hopson D. Visual loss in pseudotumor cerebri. Follow-up of 57 patients from five to 41 years and a profile of 14 patients with permanent severe visual loss. *Arch Neurol*. 1982;39:461–474. [PubMed: 7103794]
29. Parke DW 2nd, Rich WL 3rd, Sommer A, Lum F. The American Academy of Ophthalmology's IRIS((R)) Registry (Intelligent Research in Sight Clinical Data): A look back and a look to the future. *Ophthalmology*. 2017;124:1572–1574. [PubMed: 29055362]
30. Chiang MF, Sommer A, Rich WL, Lum F, Parke DW II. The 2016 American Academy of Ophthalmology IRIS® Registry (Intelligent Research in Sight) Database. *Ophthalmology*. 2018;DOI: 10.1016/j.ophtha.2017.12.001.
31. Sigsbee B, Goldenberg JN, Bever CT Jr., Schierman B, Jones LK Jr., Introducing the Axon Registry: An opportunity to improve quality of neurologic care. *Neurology*. 2016;87:2254–2258. [PubMed: 27694258]
32. Busis NA, Franklin GM. The AAN's Axon Registry. Mastering how we are measured. 2016;87:2180–2181.
33. Tso MO, Hayreh SS. Optic disc edema in raised intracranial pressure: IV. Axoplasmic transport in experimental papilledema. *Arch Ophthalmol*. 1977;95:1458. [PubMed: 70201]
34. Kesler A, Vakhapova V, Korczyn AD, Drory VE. Visual evoked potentials in idiopathic intracranial hypertension. *Clin Neurol Neurosurg*. 2009;111:433–436. [PubMed: 19345474]
35. Fang C, Leavitt JA, Hodge DO, Holmes JM, Mohny BG, Chen JJ. Incidence and etiologies of acquired third nerve palsy using a population-based method. *JAMA Ophthalmol*. 2017;135:23–28. [PubMed: 27893002]

36. Kilgore KP, Lee MS, Leavitt JA, Mokri B, Hodge DO, Frank RD, Chen JJ. Re-evaluating the incidence of idiopathic intracranial hypertension in an era of increasing obesity. *Ophthalmology*. 2017;124:697–700. [PubMed: 28187976]
37. Ji J, Dimitrijevic I, Sundquist J, Sundquist K, Zoller B. Risk of ocular manifestations in patients with giant cell arteritis: a nationwide study in Sweden. *Scand J Rheumatol*. 2017;46:484–489. [PubMed: 28276953]
38. Al Kahtani ES, Khandekar R, Al-Rubeaan K, Youssef AM, Ibrahim HM, Al-Sharqawi AH. Assessment of the prevalence and risk factors of ophthalmoplegia among diabetic patients in a large national diabetes registry cohort. *BMC Ophthalmol*. 2016;16:118. [PubMed: 27449153]
39. Sorensen PS, Trojaborg W, Gjerris F, Krogsaa B. Visual evoked potentials in pseudotumor cerebri. *Arch Neurol*. 1985;42:150–153. [PubMed: 3977643]
40. Lee YH, Bang H, Kim DJ. How to Establish Clinical Prediction Models. *Endocrinol Metab (Seoul)*. 2016;31:38–44. [PubMed: 26996421]
41. Stein JD, Blachley TS, Musch DC. Identification of persons with incident ocular diseases using health care claims databases. *Am J Ophthalmol*. 2013;156:1169–1175.e1163. [PubMed: 23972306]
42. Koerner JC, Friedman DI. Inpatient and emergency service utilization in patients with idiopathic intracranial hypertension. *J Neuroophthalmol*. 2014;34:229–232. [PubMed: 25136774]
43. Mudopivic I, Shirazi Z, Moss H, editors. Predictive value of international classification of disease code for idiopathic intracranial hypertension (IIH) in a university health system. North American Neuro-ophthalmology Society Annual Meeting; 2018; Kailua, HI.
44. Eggenberger E Initiation of anti-TNF therapy and the risk of optic neuritis: from the Safety Assessment of Biologic ThERapy (SABER) Study. *Am J Ophthalmol*. 2013;156:407–408. [PubMed: 23870360]
45. Hayreh SS. Increased risk of stroke in patients with nonarteritic anterior ischemic optic neuropathy: A nationwide retrospective cohort study. *Am J Ophthalmol*. 2017;175:213–214. [PubMed: 28093100]
46. Stunkel L, Kung NH, Wilson B, McClelland CM, Van Stavern GP. Incidence and causes of overdiagnosis of optic neuritis. *JAMA Ophthalmol*. 2018;136:76–81. [PubMed: 29222573]
47. Wacholder S, McLaughlin JK, Silverman DT, Mandel JS. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol*. 1992;135:1019–1028. [PubMed: 1595688]
48. Greenland S, Mansournia MA, Altman DG. Sparse data bias: a problem hiding in plain sight. *BMJ*. 2016;352:i1981. [PubMed: 27121591]
49. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987 258 p. 258.
50. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. *Stat Methods Med Res*. 2013;22:278–295. [PubMed: 21220355]
51. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49:1373–1379. [PubMed: 8970487]
52. Lee CH, Yoon HJ. Medical big data: promise and challenges. *Kidney Res Clin Pract*. 2017;36:3–11. [PubMed: 28392994]
53. Rohm M, Tresp V, Muller M, Kern C, Manakov I, Weiss M, Sim DA, Priglinger S, Keane PA, Kortuem K. Predicting visual acuity by using machine learning in patients treated for neovascular age-related macular degeneration. *Ophthalmology*. 2018; 125:1028–1026. [PubMed: 29454659]
54. Mazzaferri J, Larrivee B, Cakir B, Sapieha P, Costantino S. A machine learning approach for automated assessment of retinal vasculature in the oxygen induced retinopathy model. *Sci Rep*. 2018;8:3916. [PubMed: 29500375]
55. Ting DSW, Cheung CY, Lim G, Tan GSW, Quang ND, Gan A, Hamzah H, Garcia-Franco R, San Yeo IY, Lee SY, Wong EYM, Sabanayagam C, Baskaran M, Ibrahim F, Tan NC, Finkelstein EA, Lamoureux EL, Wong IY, Bressler NM, Sivaprasad S, Varma R, Jonas JB, He MG, Cheng CY, Cheung GCM, Aung T, Hsu W, Lee ML, Wong TY. 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;318:2211–2223. [PubMed: 29234807]

56. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Epidemiology*. 2007;18:800–804. [PubMed: 18049194]
57. Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, Sorensen HT, von Elm E, Langan SM. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12:e1001885. [PubMed: 26440803]
58. Fitchett EJA, Seale AC, Vergnano S, Sharland M, Heath PT, Saha SK, Agarwal R, Ayede AI, Bhutta ZA, Black R, Bojang K, Campbell H, Cousens S, Darmstadt GL, Madhi SA, Meulen AS, Modi N, Patterson J, Qazi S, Schrag SJ, Stoll BJ, Wall SN, Wammanda RD, Lawn JE. Strengthening the Reporting of Observational Studies in Epidemiology for Newborn Infection (STROBE-NI): an extension of the STROBE statement for neonatal infection research. *Lancet Infect Dis*. 2016;16:e202–e213. [PubMed: 27633910]
59. Wang SV, Schneeweiss S, Berger ML, Brown J, de Vries F, Douglas I, Gagne JJ, Gini R, Klungel O, Mullins CD, Nguyen MD, Rassen JA, Smeeth L, Sturkenboom M. Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiol Drug Saf*. 2017;26:1018–1032. [PubMed: 28913963]