AMERICAN ACADEMY™
OF OPHTHALMOLOGY

Editorial

Big Data, Big Challenges

Michael V. Boland, MD, PhD - *Baltimore, Maryland*

The relatively rapid and recent adoption of electronic health records (EHRs) in ophthalmology[1,2] has been associated with the promise that the accumulation of large volumes of clinical data would facilitate quality improvement and help answer a variety of research questions. Given that EHRs are relatively new in most practices and that clinical data are inherently more complex than other fields that have been altered by the digital revolution, these proposed benefits have yet to be realized.[3] The results reported by Shen et al[4] in this issue of *Ophthalmology* (see p. 92) represent an early glimpse of just how ophthalmology may ultimately benefit from "big data."

The Kaiser Permanente health system was a relatively early adopter of EHRs (1995) and therefore is in a position to demonstrate how a large database of clinical information can be used to assess risk factors for disease without having to undertake costly population-based studies. By querying the data from one regional Kaiser system of 3.5 million patients, the authors were able to analyze the information from >400 000 patients to evaluate the relationship between refractive error and glaucoma. A study with this many subjects is revolutionary in ophthalmology based on size alone—some 100 times larger than prior population-based studies in the field.

A study with this many subjects is revolutionary in ophthalmology based on size alone—some 100 times larger than prior population-based studies in the field.

This study's size is indeed its main strength. By being able to assess relationships among demographics, clinical findings, diagnoses, and procedures across so many subjects, one can hope to find subtle but important results that would have remained statistically insignificant in smaller studies. Another important advantage of studies like this is that we have data for an entire population, so issues of random selection and bias are reduced or eliminated. These advantages assume that the population in the database is the one of interest, however.

Although the Kaiser study used some clinical information, it is still limited by the issues inherent in the analysis of claims data.[5] First among these is misclassification bias. By relying on diagnostic codes (*International Classification of Diseases, 9th edition* [ICD-9]) to determine the presence of absence of glaucoma, one has to forego any systematic clinical definition and assume that all of the physicians involved are coding their patients consistently. Another issue inherent in this setting is confounding, which the authors do discuss. Specifically, patients seeking eye care for refractive error are relatively overrepresented and therefore more likely to be diagnosed with glaucoma. There is also some selection bias in that the patients being evaluated have health insurance and sought eye care. These factors may result in conclusions that are not generalizable to the overall population, which does include the uninsured.

Diagnostic billing codes may also introduce bias in favor of confirming prior studies or beliefs about disease. Based on such beliefs, physicians may be more likely to diagnose pseudoexfoliation in white patients and angle closure in Asian patients, for example. Even if such associations are real, they may seem to be more significant than they really are owing to the unconscious bias of the physicians making the diagnoses and entering codes in the computer. This issue can be made less relevant by relying more on analysis of clinical examination findings and test results to determine the presence or absence of disease.

These kinds of cross-sectional and observational studies are a way to evaluate hypotheses regarding the association between risk factors and disease. However, retrospective and observational studies have not had their results borne out in randomized trials. The Nurses Health Study (NHS) provides examples of this phenomenon. Analyses of the NHS suggested a protective effect of estrogen and progesterone on cardiovascular disease and a protective effect of antioxidants on certain cancers. When evaluated in randomized trials, however, neither finding was confirmed and hormone replacement therapy was, in fact, found to be harmful. A systematic review of NHS results found agreement only 20% of the time between those results and subsequent randomized trials.[6] Findings like this call into question the role of retrospective or observational studies that might be conducted with "big data." In any case, we will all need to continue to be vigilant regarding our interpretation of studies even if they are based on data from large populations.

As large institutions like Kaiser accumulate and organize their clinical data and as the American Academy of Ophthalmology does the same with the Intelligent Research In Sight registry, we will need to ask appropriate questions using those data and then follow up with randomized trials whenever possible to evaluate the most important questions and hypotheses we generate.

To do the kind of analyses described using a single EHR is one thing, but the real promise of shared clinical data requires that we be able to combine data from multiple EHRs from multiple practices. Although it may seem like it should be easy to combine clinical data from multiple systems, we are faced with 2 levels of interoperability that

http://dx.doi.org/10.1016/j.ophtha.2015.08.041
ISSN 0161-6420/15

have proven elusive over many years and many attempts to integrate clinical data across regions of the country. Coupled with the lack of a viable financial model for regional health information exchange, such organizations have languished.[7]

The first requirement for sharing clinical data is that the syntax of the 2 systems must be the same—both must record visual acuity in the same way, for example. If one system in a network records visual acuity in a single field ("20/40") and another records it in 2 fields (numerator "20" and denominator "40"), there must be some translation that takes place to combine those data in a common database. This is also perhaps the simplest example of a problem that only gets more complex as the data become less structured. Imagine the various ways one can record a slit-lamp examination, for example.

Useful exchange and integration of data also requires that data collected in one system mean the same thing as the corresponding data collected in another. If the value "20/40" is recorded in both systems, but in one system it represents visual acuity with correction and in the other system without correction, then the semantics (meaning) of the two are not the same. This is the concept of semantic interoperability and it will be important for ophthalmology to clearly define and record the data we want to share, combine, and analyze. We have examples from other fields of medicine,[8,9] but we will need to help ourselves by defining the details required for exchanging data with one another to truly transform the way we learn about disease and provide care to our patients.

References

1. Chiang MF, Boland MV, Margolis JW, et al. Adoption and perceptions of electronic health record systems by ophthalmologists: an American Academy of Ophthalmology survey. Ophthalmology 2008;115:1591–1597.e5.
2. Boland MV, Chiang MF, Lim MC, et al. Adoption of electronic health records and preparations for demonstrating meaningful use: an American Academy of Ophthalmology survey. Ophthalmology 2013;120:1702–10.
3. Rahurkar S, Vest JR, Menachemi N. Despite the spread of health information exchange, there is little evidence of its impact on cost, use, and quality of care. Health Aff (Millwood) 2015;34:477–83.
4. Shen L, Melles RB, Metlapally R, et al. The association of refractive error with glaucoma in a multiethnic population. Ophthalmology 2015;123:92–101.
5. Coleman AL, Morgenstern H. Use of insurance claims databases to evaluate the outcomes of ophthalmic surgery. Surv Ophthalmol 1997;42:271–8.
6. Tai V, Grey A, Bolland MJ. Results of observational studies: analysis of findings from the Nurses' Health Study. PLoS One 2014;9:e110403.
7. Adler-Milstein J, Bates DW, Jha AK. A survey of health information exchange organizations in the United States: implications for meaningful use. Ann Intern Med 2011;154:666–71.
8. Richesson RL, Chute CG. Health information technology data standards get down to business: maturation within domains and the emergence of interoperability. J Am Med Inform Assoc 2015;22:492–4.
9. Manion FJ, Harris MR, Buyuktur AG, et al. Leveraging EHR data for outcomes and comparative effectiveness research in oncology. Curr Oncol Rep 2012;14:494–501.

Footnotes and Financial Disclosures

Financial Disclosure(s):
The authors have no proprietary or commercial interest in any materials discussed in this article.

Correspondence:
Michael V. Boland, MD, PhD, Wilmer Eye Institute and Health Sciences Informatics, The Johns Hopkins University, 600 North Wolfe Street, Wilmer 131, Baltimore, MD 21287. E-mail: boland@jhu.edu.