JAMA Ophthalmology | Original Investigation

Evaluation of an Algorithm for Identifying Ocular Conditions in Electronic Health Record Data

Joshua D. Stein, MD, MS; Moshiur Rahman, PhD; Chris Andrews, PhD; Joshua R. Ehrlich, MD, MPH; Shivani Kamat, MD; Manjool Shah, MD; Erin A. Boese, MD; Maria A. Woodward, MD, MSc; Jeff Cowall, BS; Edward H. Trager, MS; Prabha Narayanaswamy, MS; David A. Hanauer, MD, MS

+ Invited Commentary

IMPORTANCE For research involving big data, researchers must accurately identify patients with ocular diseases or phenotypes of interest. Reliance on administrative billing codes alone for this purpose is limiting.

OBJECTIVE To develop a method to accurately identify the presence or absence of ocular conditions of interest using electronic health record (EHR) data.

DESIGN, SETTING, AND PARTICIPANTS This study is a retrospective analysis of the EHR data of patients (n = 122 339) in the Sight Outcomes Research Collaborative Ophthalmology Data Repository who received eye care at participating academic medical centers between August 1, 2012, and August 31, 2017. An algorithm that searches structured and unstructured (free-text) EHR data for conditions of interest was developed and then tested to determine how well it could detect the presence or absence of exfoliation syndrome (XFS). The algorithm was trained to search for evidence of XFS among a sample of patients with and without XFS (n = 200) by reviewing *International Classification of Diseases, Ninth Revision* or *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision* (*ICD-9* or *ICD-10*) billing codes, the patient's problem list, and text within the ocular examination section and unstructured (free-text) data in the EHR. The likelihood that each patient had XFS was estimated using logistic least absolute shrinkage and selection operator (LASSO) regression. The EHR data of all patients were run through the algorithm to generate an XFS probability score for each patient. The algorithm was validated with review of EHRs by glaucoma specialists.

MAIN OUTCOMES AND MEASURES Positive predictive value (PPV) and negative predictive value (NPV) of the algorithm were computed as the proportion of patients correctly classified with XFS or without XFS.

RESULTS This study included 122 339 patients, with a mean (SD) age of 52.4 (25.1) years. Of these patients, 69 002 (56.4%) were female and 99 579 (81.4%) were white. The algorithm assigned a less than 10% probability of XFS for 121 085 patients (99.0%) as well as an XFS probability score of more than 75% for 543 patients (0.4%), more than 90% for 353 patients (0.3%), and more than 99% for 83 patients (0.07%). Validated by glaucoma specialists, the algorithm had a PPV of 95.0% (95% CI, 89.5%-97.7%) and an NPV of 100% (95% CI, 91.2%-100%). When there was *ICD-9* or *ICD-10* billing code documentation of XFS, in 86% or 96% of the records, respectively, evidence of XFS was also recorded elsewhere in the EHR. Conversely, when there was clinical examination or free-text evidence of XFS, it was documented with *ICD-9* codes only approximately 40% of the time and even less often with *ICD-10* codes.

CONCLUSIONS AND RELEVANCE The algorithm developed, tested, and validated in this study appears to be better at identifying the presence or absence of XFS in EHR data than the conventional approach of assessing only billing codes; such an algorithm may enhance the ability of investigators to use EHR data to study patients with ocular diseases.

JAMA Ophthalmol. doi:10.1001/jamaophthalmol.2018.7051
Published online February 21, 2019.

Author Affiliations: Author affiliations are listed at the end of this article.

Corresponding Author: Joshua D. Stein, MD, MS, W. K. Kellogg Eye Center, Department of Ophthalmology and Visual Sciences, University of Michigan Medical School, 1000 Wall St, Ann Arbor, MI 48105 (jdstein@med.umich.edu).

Researchers have used claims data for many years[1] and, more recently, registry data to study the epidemiologic characteristics of ocular diseases,[1,2] eye care service utilization,[3-5] disparities in care,[6,7] and outcomes of ophthalmic interventions.[8-11] Although these studies involving big data have led to many insights that have informed patient care and health policymaking, they typically share a key limitation: the sole reliance on *International Classification of Diseases, Ninth Revision,* or *International Statistical Classification of Diseases and Related Health Problems, Tenth Revision,* (*ICD-9* or *ICD-10*) billing codes to identify the presence or absence of ocular conditions of interest.[12]

Although investigators have found relatively good concordance between billing and medical record documentation for most ocular conditions studied,[13,14] reliance on billing codes to identify diseases has limitations, including miscoding by clinicians, insufficient granularity of some codes, and an expectation that the health care practitioner would properly document all conditions present during a given encounter.[12] Fortunately, for most analyses, the sample sizes are so large that, even if some patients are misclassified, researchers can still answer pertinent research questions. However, if researchers can devise alternative methods to more accurately identify the presence or absence of ocular diseases, it could enhance the quality of these big data studies.

Along with the 2009 Health Information Technology for Economic and Clinical Health Act has come widespread adoption of electronic health records (EHRs) to capture medical encounters.[15] Nearly three-quarters of all encounters with eye care practitioners in the United States are documented using EHRs.[16] In the EHR system for each patient visit, clinicians routinely document historical information, ocular examination findings, diagnostic test results, and an assessment and plan as well as the accompanying billing codes to submit for reimbursement. Researchers can now use these richly detailed information sources, not just the billing codes, to develop algorithms to detect the presence or absence of a disease. Outside of ophthalmology, researchers have begun using these approaches to study selected conditions[17]; to our knowledge, no validated algorithms have been built to identify the presence of ocular diseases in EHRs.

We developed a method to accurately identify the presence or absence of an ocular phenotype of interest using EHR data. Using exfoliation syndrome (XFS), a common and debilitating cause of glaucoma, as a test case, we assessed how well our algorithm identifies the presence or absence of XFS compared with the performance of criterion standard evaluation of EHRs by several glaucoma subspecialists.

Methods

Data Source

The data source we used was the Sight Outcomes Research Collaborative (SOURCE) Ophthalmology Data Repository, which captures the EHR data of all patients receiving any eye care at academic medical centers participating in this research collaborative. All contributing sites used the EPIC EHR system (EPIC Systems Corporation). SOURCE captures patient demographics, diagnoses identified by *ICD-9* and *ICD-10* billing codes, and structured and unstructured (free-text) data from all clinical encounters (eg, clinic visits, operative reports). The sample includes persons in SOURCE who received any eye care between August 1, 2012, and August 31, 2017. Persons younger than 18 years of age were excluded. The University of Michigan Institutional Review Board approved this study and waived the need for informed consent because it deemed it impractical to obtain consent from all of the patients in the repository.

Key Points

Question What method other than assessing administrative billing codes can researchers, using big data, apply to accurately identify patients with ocular diseases of interest?

Findings In this study of the electronic health records of 122 339 eye care recipients, a newly developed and validated algorithm that searches structured and unstructured data in electronic health records successfully detected most patients with and without exfoliation syndrome.

Meaning Algorithms may enhance the ability of researchers to make use of big data to study patients with ocular diseases.

Evidence of Exfoliation Syndrome in EHRs

Clinicians have various ways of documenting the presence of XFS in EHRs. We searched 4 areas of the EHR to identify evidence of XFS. The first area was billing data, comprising the billing codes for XFS or XFS glaucoma (*ICD-9* codes 365.52 and 366.11, and *ICD-10* codes H40.14XX).[18,19] The second area was the problem list, a running list of medical conditions identified for each patient during the person's time in the practice. Examples included pseudoexfoliation glaucoma; glaucoma, pseudoexfoliation; and pseudoexfoliative glaucoma. The third area was clinical examination findings. Through an automated process, we searched all clinical notes for findings in the iris and lens sections of the ocular examination that indicated the presence of XFS, including iris transillumination defects, exfoliative material on the lens capsule, iridodonesis, phacodonesis, and lens dislocation or subluxation.

The fourth area was the unstructured data elsewhere in the medical record. Using natural language processing (NLP), we searched the rest of the EHR for documentation of XFS, including these terms and abbreviations in the search algorithm: *XFS*, *pex*, *pxe*, *pxg*, *xfg*, *pseudoexfoliation*, *pseudoexfoliative*, *pseudo-x*, *pseudo-ex*, *px material*, *pxe material*, *pseudoex*, *exfoliation*, and *exfoliation syndrome*. This list reflects a review of the literature and the EHRs of multiple glaucoma specialists, who contribute to SOURCE, to identify other ways clinicians describe XFS and its clinical findings. The search algorithm reviewed the text immediately before and after 1 of these words or abbreviations. If evidence of negation terms (ie, *no*, *none*, or *without*) existed, the algorithm did not count this as evidence of XFS. The algorithm also excluded mentions of *exfoliation* unrelated to the eye, such as skin exfoliation in burn patients and descriptions of someone other than the patient

Table 1. Variables in the Electronic Health Record Evaluated to Assess for the Presence of Exfoliation Syndrome

Variable	Health Record Entry Documenting No.	Health Record Entry Documenting %
Demographic	Age	NA
	Sex	NA
	Race/ethnicity	NA
Clinical encounter	No. of clinical encounters	NA
Billing code documentation	No. of visits with *ICD-9* documentation of XFS	% of Visits with *ICD-9* documentation of XFS
	No. of visits with *ICD-10* documentation of XFS	% of Visits with *ICD-10* documentation of XFS
Clinical examination note	No. of visits with XFS mentioned in iris examination	% of Visits with XFS mentioned in iris examination
	No. of visits with XFS mentioned in lens examination	% of Visits with XFS mentioned in lens examination
	No. of visits with iris TIDs mentioned in iris examination	% of Visits with iris TIDs mentioned in iris examination
	Documentation of lens dislocation/subluxation	NA
	Documentation of phacodonesis	NA
Problem list	No. of visits with problem list documentation of XFS	% of Visits with problem list documentation of XFS
Unstructured data	No. of mentions of XFS detected using NLP across all clinical encounters	% of All clinical encounters with ≥1 mention of XFS detected using NLP

Abbreviations: *ICD*, *International Classification of Diseases*; NA, not applicable; NLP, natural language processing; XFS, exfoliation syndrome; SOURCE, Sight Outcomes Research Collaborative; TID, transillumination defect.

(ie, *family history of exfoliation glaucoma*). The algorithm used regular expressions and generalized Levenshtein edit distance to identify close misspellings of the key terms of interest.

Algorithm for Identifying Exfoliation Syndrome Using EHRs

We created an algorithm to estimate the likelihood that a person had XFS according to EHR documentation. The algorithm considered 20 potential variables (**Table 1**). Female sex, white race/ethnicity, and older age are associated with XFS,[20,21] so we included those demographic variables. Because XFS is primarily captured during visits to eye care professionals, we included number of visits to eye care practitioners as an algorithm covariate. The *ICD-9* and *ICD-10* billing codes were summarized for use in the algorithm by the logarithm of the number of clinical encounters (log count) that included an *ICD-9* or *ICD-10* code for XFS and the percentage of all eye encounters that included such codes. The log count of lens and iris examinations that recorded exfoliation material presence and the percentage of all lens and iris examinations that recorded exfoliation material were included in the algorithm. The presence of iris transillumination defects was summarized similarly. Binary (ever or never) variables were created to indicate the presence of phacodonesis and dislocation or subluxation of the native lens or intraocular lens in the structured and unstructured data at any encounter. The log count and the percentage of total visits with XFS listed in the problem list area of the EHR were 2 additional covariates. Finally, using the NLP algorithm described earlier, we captured the log count and percentage of clinical encounters with evidence of the presence of XFS in the unstructured free text of all clinical notes.

Statistical Analysis

Data processing and analysis were performed in SAS, version 9.4 (SAS Institute Inc) and R, version 4.3.2 (R Foundation for Statistical Computing). The NLP used the EMERSE software (Project EMERSE).[22]

Training Set

A random sample of 200 patients was selected as the training set for building the algorithm. This set contained a balance of patients with and without XFS. Four of us (glaucoma subspecialists S.K., M.S., J.R.E., and E.A.B.) each evaluated the entire EHR of 50 patients and rated each patient as having definite, possible, or nonexistent XFS. On the basis of these ratings, the variable definite XFS (yes or no) was used as the criterion standard for building the algorithm. Persons classified as possible cases were treated as noncases in the algorithm. The glaucoma subspecialists who reviewed these records were masked to avoid biasing their assessments.

Prediction Modeling

Logistic LASSO (least absolute shrinkage and selection operator) regression determined which of the 20 algorithm variables together best indicated the likely presence or absence of XFS compared with the criterion standard ratings of the 200 patients by our glaucoma subspecialists.[23] Logistic LASSO regression constrains the magnitude of the regression coefficients and favors models with fewer variables, thereby inducing variable selection in the fitting process. The optimality criterion was the 10-fold cross-validated mean prediction error. One advantage of logistic LASSO regression over other variable selection methods (forward, backward, or best subset) is its flexibility to include a variable, yet it penalizes its coefficient when a strict inclusion or exclusion decision is too restrictive. Elastic net and several discriminant models gave similar results to those presented here from the logistic LASSO regression. After we identified the algorithm variables that best detected which patients likely had XFS, we applied the algorithm to all 122 339 patients in SOURCE to estimate each

Table 2. Demographic Characteristics of All Patients and Patients With High and Low Probabilities of Exfoliation Syndrome in the SOURCE Repository

	No. (%)		
		Patients With Estimated XFS Probability	
Variable	All Patients	<20% Score	>90% Score
Total	122 339	121 273	353
Age, mean (SD), y	52.4 (25.1)	52.2 (25.1)	78.8 (10.1)
Sex			
Male	53 337 (43.6)	52 959 (43.7)	115 (32.6)
Female	69 002 (56.4)	68 314 (56.3)	238 (67.4)
Race/ethnicity			
White	99 579 (81.4)	98 599 (81.3)	330 (93.5)
Black	11 279 (9.2)	11 242 (9.3)	7 (2.0)
Latino	3113 (2.5)	3098 (2.5)	6 (1.7)
Asian	6417 (5.3)	6393 (5.3)	7 (2.0)
Other	1951 (1.6)	1941 (1.6)	3 (0.8)

Abbreviations: XFS, exfoliation syndrome; SOURCE, Sight Outcomes Research Collaborative.

Table 3. Conditional Probability of Evidence of Exfoliation Syndrome in Each Area of the Electronic Health Record (EHR)

EHR Area	No. of Patients With ≥1 Record of XFS	*ICD-9* Evidence, %	*ICD-10* Evidence, %[a]	Problem List Evidence, %	Iris and Lens Examination Data Evidence, %	NLP Evidence, %
ICD-9	491	NA	28	100	87	86
ICD-10	226	63	NA	100	92	96
Problem list	610	81	37	NA	87	87
Iris and lens examination data	1216	38	18	47	NA	70
NLP	1205	40	20	50	75	NA

Abbreviations: *ICD*, *International Classification of Diseases*; NA, not applicable; NLP, natural language processing; XFS, exfoliation syndrome.

[a] Because *ICD-10* codes were not used prior to 2015, persons who received all of their care during 2012-2015 would not have received an *ICD-10* code.

person's probability (score range, 0%-100%, with the highest percentage indicating greatest likelihood) of having XFS.

Validation Set

Next, 120 patients whose XFS probability, according to the algorithm, surpassed 90% were labeled as *cases* and selected to estimate the positive predictive value (PPV) of the algorithm. Similarly, 40 patients with an XFS probability score below 20% were labeled as *noncases* and selected to estimate the negative predictive value (NPV) of the algorithm. Our glaucoma subspecialists each evaluated the entire EHR entries for 30 cases and 10 noncases, and they scored each patient's evidence of XFS as definite, possible, or nonexistent. The PPV was computed as the proportion of cases classified as definite XFS; the NPV was the proportion of noncases classified as nonexistent XFS.

Results

SOURCE contains data on 122 339 patients, with a mean (SD) age of 52.4 (25.1) years. Of these patients, 69 002 (56.4%) were female and 99 579 (81.4%) were white (**Table 2**).

The number of patients who had 1 or more records of XFS according to *ICD-9* codes was 491 (0.4%), *ICD-10* codes was 226 (0.2%), and the problem list of the EHR was 610 (0.5%). In total, 1216 persons (0.9%) had 1 or more mentions of XFS in the iris or lens section of the encounters; 1205 persons (0.9%) had 1 or more such mentions in free text elsewhere in the records. When there was *ICD-9* or *ICD-10* billing code documentation of XFS, in 86% or 96% of the records, respectively, some evidence of the presence of exfoliative material was found in the clinic examination records or detected using NLP searching of the free text of the EHR data. Conversely, with clinical examination evidence or free-text evidence of XFS detected using NLP, XFS was documented with *ICD-9* codes only approximately 40% of the time and even less often (approximately 20%) with *ICD-10* codes (**Table 3**). In an ancillary analysis, we determined that when XFS was coded as the primary *ICD-9* or *ICD-10* diagnosis at any clinical encounter, evidence of XFS was found in the clinical notes or free text using NLP 87% to 97% of the time. When evidence of XFS was detected in the clinical notes or free text using NLP, only 12% to 26% of the time was XFS recorded as the primary *ICD-9* or *ICD-10* diagnosis code.

Among the 20 variables we considered, the 5 that best indicated definite XFS as identified by the glaucoma subspecialists were the proportions of encounters with evidence of XFS in the iris and lens sections of the clinical notes, visits with iris transillumination defects documented, visits with problem list documentation of XFS, and visits with 1 or more mentions of XFS in the free text of the clinical notes.

We ran all 122 339 patients' data through the algorithm to generate an XFS probability score for each patient. The

Table 4. Comparison of More Than 90% Algorithm-Assigned Probability of Exfoliation Syndrome With Clinician-Indicated Absence of Exfoliation Syndrome

Patient With Discordant Scores	Basis for Algorithm-Assigned Probability Score	Independent Clinician Assessment[a]
Case 1	Evidence of XFS in OH and lens examination; however, in the A/P section the clinician specified glaucoma type as "pigment dispersion."	Patient likely had XFS
Case 2	Evidence of XFS in the lens examination and NLP but not in the problem list or billing codes.	Patient likely had XFS
Case 3	Patient had 21 total visits. XFS was documented in several of the earlier visits prior to the patient undergoing cataract surgery. After cataract surgery, the patient was under the care of a different clinician who instead classified the patient as open-angle glaucoma.	Patient likely had XFS
Case 4	The clinician billed the patient as XFS, but the demographic profile of the patient and clinical exam findings were all consistent with pigment dispersion instead.	Patient likely did not have XFS
Case 5	A glaucoma specialist noted increased pigmentation and speculated that he patient may have XFS but stated in the progress note she was not completely sure of this.	It is possible the patient had XFS
Case 6	There was evidence of XFS in the clinical examination and NLP but not in the problem list or billing codes. The ophthalmology records were mixed in with dozens of non-ophthalmology encounters, which may have made it difficult for the grader to locate evidence of XFS.	Patient likely had XFS

Abbreviations: A/P, assessment and plan; NLP, natural language processing; OH, ocular history; XFS, exfoliation syndrome.

[a] Assessment performed by Joshua D. Stein, MD, MS.

algorithm assigned a less than 10% probability of XFS for 121 085 patients (99.0%). The algorithm also assigned an XFS probability score of more than 75% for 543 patients (0.4%), more than 90% for 353 patients (0.3%), and more than 99% for 83 patients (0.07%).

Next, our glaucoma subspecialists each reviewed the EHRs for another 40 patients. Some records were of patients whose algorithm-assigned XFS probability score was more than 90%; others were of patients whose algorithm-assigned XFS probability score was less than 20%. We compared how well the algorithm fared with the performance of criterion standard assessments of the glaucoma specialists. The PPV of the algorithm was 95.0% (95% CI, 89.5%-97.7%), and the NPV was 100% (95% CI, 91.2%-100%).

One of us (J.D.S.) reviewed the records of the 6 patients for whom the algorithm assigned a probability score of more than 90% but whose classification by the glaucoma specialists were characterized as nonexistent XFS. In 4 cases, the evidence indicated that the patient likely had XFS; in 1 case, the clinician indicated possible XFS; and in only 1 case, the evidence to support a diagnosis was insufficient (**Table 4**).

Discussion

We developed, tested, and validated an algorithm to identify the presence or absence of XFS in EHR data. Our approach searched for evidence of XFS not only using billing codes but also considering evidence of XFS in the clinical examination notes and in unstructured data from clinic visits, operative reports, and elsewhere in the medical record. When we compared the algorithm's performance with that of a criterion standard evaluation by 4 glaucoma subspecialists, the algorithm properly detected the presence or absence of XFS in nearly every EHR reviewed. In a few cases, the algorithm may have outperformed the clinician rater.

Nearly all previously published analyses of health care claims or EHR data have relied on *ICD-9* or *ICD-10* billing codes to identify the presence of a given disease. As reflected in Table 3, in this study sample and for this particular condition, when the eye care practitioner billed for XFS, we found evidence in the examination records to substantiate the diagnosis 86% to 96% of the time. Therefore, researchers can feel confident that most patients with billing codes indicating XFS do have this phenotype. However, the reverse appears not to be true: In the absence of *ICD-9* or *ICD-10* documentation of XFS, the patient may still have the phenotype. In the case of XFS, more than 50% of persons with evidence of the disease in the examination records or the unstructured data never received a billing code for XFS. The reasons may include the clinician's use of more general billing codes (ie, open-angle glaucoma), the clinician's prioritization of coding for other conditions that are more relevant for that particular encounter (eg, a retina specialist evaluating diabetic retinopathy may take note of concomitant XFS without billing for it), or the EHR system's limit on the number of diagnosis codes that can be captured in a clinical encounter.

In addition, once clinicians document a sufficient number of diagnoses to justify billing for a certain level of service, they may deem it unnecessary to code others. Furthermore, some patients simultaneously receive care from multiple eye care practitioners, and some practitioners may be more adept at diagnosing or coding this condition than others. Fortunately, the very large sample sizes found in data sources, such as the Standard Analytic Files (Medicare claims data) and IRIS (Intelligent Research in Sight) Registry, may be able to handle the misclassification of some patients as not having a given disease when indeed they have it. Furthermore, such a misclassification would likely bias the findings to the null, such that if researchers were observing an association between a given condition and another condition or outcome and if the misclassified patients were excluded, the magnitude of the association would actually be larger than what would occur if using billing codes alone to identify these conditions. Nevertheless, if researchers can devise better ways to identify conditions of interest by using an algorithm like the one presented here, the accuracy of big data analyses will be enhanced and smaller sample sizes will be sufficient to demonstrate the associations among conditions of interest.

This study aimed in part to develop an algorithm to effectively identify the presence or absence of XFS, a sight-threatening condition that can lead to difficult-to-manage glaucoma,[24] but the more global objective of this work was to devise and test a methodologic approach that researchers can use to identify the presence or absence of many different ocular and nonocular conditions. Such an approach may be particularly helpful in identifying patients with ocular conditions that lack specific billing codes or conditions for which the codes may be ambiguous or nonspecific or that require a more accurate classification of disease than simply using billing codes. Proper identification and characterization of phenotypes using an algorithm such as this one is a prerequisite for future genotype-phenotype association studies.

Researchers frequently design their studies such that every patient in the cohort is assigned a binary value (yes or no) for having the disease or outcome of interest. For example, if a study is exploring endophthalmitis, every patient is classified with endophthalmitis or not. Rather than simply characterizing each patient as having the condition of interest or not, our algorithm instead assigns each person a probability score from 0% to 100% that indicates whether they have the condition. An advantage of this approach is that researchers can decide, on a case by case basis, how stringent to make the probability cutoff score that classifies the presence or absence of the perusal condition depending on the nature of the study. For studies in which it is critically important that all patients are properly classified, the researcher may assign a very high probability cutoff score for participants to be classified as cases. For other studies in which case identification is of lesser importance, the researcher can set a lower cutoff value for classifying cases so that a larger number of patients are identified with the outcome, even if some patients may have been misclassified.

Limitations

This study has several limitations. First, all of the patients in SOURCE were receiving eye care from clinicians at academic medical centers that used the EPIC EHR system. Additional validation is required to determine how well the algorithm identifies XFS among patients receiving care in other settings and with other EHR systems. For example, the level of detail of documentation may vary according to the practice setting, the clinical volume of the clinician, and the use of trainees or scribes to perform some of the documentation activities.

Second, although the algorithm enhances the identification of patients who were documented as having XFS, it would not detect this condition if a patient received a misdiagnosis or the clinician did not document the characteristic findings of XFS. Third, the training set used to develop the algorithm was based on review of 200 EHRs. With a greater number of EHRs to train the algorithm, it may be possible to achieve even greater sensitivity and specificity for detecting this condition. When applying this algorithm to other conditions, the appropriate size for the training set depends on the similarity or differences in the way clinicians describe the condition of interest in their records. When there are many ways to describe a particular condition, a larger training set is needed.

Conclusions

The algorithm developed and tested to identify the presence or absence of XFS in EHRs seems to perform quite well. We believe that similar algorithms can be developed to identify other ocular conditions and phenotypes. These methods may enhance the quality of analyses using big data and the types of questions that researchers can answer using these resources.

ARTICLE INFORMATION

Accepted for Publication: December 5, 2018.

Published Online: February 21, 2019. doi:10.1001/jamaophthalmol.2018.7051

Author Affiliations: W. K. Kellogg Eye Center, Department of Ophthalmology and Visual Sciences, University of Michigan Medical School, Ann Arbor (Stein, Rahman, Andrews, Ehrlich, Kamat, Shah, Boese, Woodward, Trager, Narayanaswamy); Department of Health Management and Policy, University of Michigan School of Public Health, Ann Arbor (Stein, Ehrlich); Center for Eye Policy and Innovation, University of Michigan, Ann Arbor (Stein, Rahman, Andrews, Ehrlich, Woodward, Narayanaswamy); Data Office for Clinical and Translational Research, University of Michigan Medical School, Ann Arbor (Cowall); Department of Pediatrics, University of Michigan Medical School, Ann Arbor (Hanauer).

Author Contributions: Dr Stein had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.
Study concept and design: Stein.
Acquisition, analysis, or interpretation of data: All authors.
Drafting of the manuscript: Stein, Kamat, Woodward, Cowall.
Critical revision of the manuscript for important intellectual content: Rahman, Andrews, Ehrlich, Shah, Boese, Woodward, Trager, Narayanaswamy, Hanauer.
Statistical analysis: Rahman, Andrews, Narayanaswamy.
Obtained funding: Ehrlich.
Administrative, technical, or material support: Stein, Cowall, Hanauer.
Supervision: Andrews, Woodward.

Conflict of Interest Disclosures: Dr Stein reported grants from NIH, grants from Lighthouse Guild, and grants from Research to Prevent Blindness during the conduct of the study. Dr Ehrlich reported grants from National Institutes of Health during the conduct of the study; grants from Blue Cross Blue Shield of Michigan Foundation and grants from American Society of Cataract and Refractive Surgery Foundation outside the submitted work. Dr Shah reported personal fees from Allergan and personal fees from Glaukos outside the submitted work. Dr Woodward reported grants from National Eye Institute during the conduct of the study. No other disclosures were reported.

Funding/Support: This study was supported by grants R01 EY026641 (Dr Stein), K23EY027848 (Dr Ehrlich), and K23EY023596 (Dr Woodward) from the National Institutes of Health; a grant from the Lighthouse Guild; and an unrestricted grant to the University of Michigan Department of Ophthalmology and Visual Sciences from Research to Prevent Blindness.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Meeting Presentation: The results of this study were presented at the 2018 American Ophthalmological Society Annual Meeting; May 18, 2018; Dana Point, California.

REFERENCES

1. Musch DC, Niziol LM, Stein JD, Kamyar RM, Sugar A. Prevalence of corneal dystrophies in the United States: estimates from claims data. *Invest Ophthalmol Vis Sci*. 2011;52(9):6959-6963. doi:10.1167/iovs.11-7771

2. Wang SY, Andrews CA, Herman WH, Gardner TW, Stein JD. Incidence and risk factors for developing diabetic retinopathy among youths with type 1 or type 2 diabetes throughout the United States. *Ophthalmology*. 2017;124(4):424-430. doi:10.1016/j.ophtha.2016.10.031

3. Wu AM, Wu CM, Tseng VL, et al. Characteristics associated with receiving cataract surgery in the US Medicare and Veterans Health Administration populations. *JAMA Ophthalmol*. 2018;136(7):738-745. doi:10.1001/jamaophthalmol.2018.1361

4. Coleman AL, Yu F, Evans SJ. Use of gonioscopy in Medicare beneficiaries before glaucoma surgery. *J Glaucoma*. 2006;15(6):486-493. doi:10.1097/01.ijg.0000212287.62798.8f

5. Coleman AL, Yu F, Rowe S. Visual field testing in glaucoma Medicare beneficiaries before surgery. *Ophthalmology*. 2005;112(3):401-406. doi:10.1016/j.ophtha.2004.09.034

6. Stein JD, Andrews C, Musch DC, Green C, Lee PP. Sight-threatening ocular diseases remain underdiagnosed among children of less affluent families. *Health Aff (Millwood)*. 2016;35(8):1359-1366. doi:10.1377/hlthaff.2015.1007

7. Elam AR, Andrews C, Musch DC, Lee PP, Stein JD. Large disparities in receipt of glaucoma care between enrollees in Medicaid and those with commercial health insurance. *Ophthalmology*. 2017;124(10):1442-1448. doi:10.1016/j.ophtha.2017.05.003

8. Stein JD, Grossman DS, Mundy KM, Sugar A, Sloan FA. Severe adverse events after cataract surgery among Medicare beneficiaries. *Ophthalmology*. 2011;118(9):1716-1723. doi:10.1016/j.ophtha.2011.02.024

9. Atchison EA, Wood KM, Mattox CG, Barry CN, Lum F, MacCumber MW. The real-world effect of intravitreous anti-vascular endothelial growth factor drugs on intraocular pressure: an analysis using the IRIS registry. *Ophthalmology*. 2018;125(5):676-682. doi:10.1016/j.ophtha.2017.11.027

10. VanderBeek BL, Bonaffini SG, Ma L. Association of compounded bevacizumab with postinjection endophthalmitis. *JAMA Ophthalmol*. 2015;133(10):1159-1164. doi:10.1001/jamaophthalmol.2015.2556

11. Gower EW, Keay LJ, Stare DE, et al. Characteristics of endophthalmitis after cataract surgery in the United States Medicare population. *Ophthalmology*. 2015;122(8):1625-1632. doi:10.1016/j.ophtha.2015.04.036

12. Stein JD, Lum F, Lee PP, Rich WL III, Coleman AL. Use of health care claims data to study patients with ophthalmologic conditions. *Ophthalmology*. 2014;121(5):1134-1141. doi:10.1016/j.ophtha.2013.11.038

13. Muir KW, Gupta C, Gill P, Stein JD. Accuracy of*International Classification of Diseases, Ninth Revision, Clinical Modification* billing codes for common ophthalmic conditions. *JAMA Ophthalmol*. 2013;131(1):119-120. doi:10.1001/jamaophthalmol.2013.577

14. Lau M, Prenner JL, Brucker AJ, VanderBeek BL. Accuracy of billing codes used in the therapeutic care of diabetic retinopathy. *JAMA Ophthalmol*. 2017;135(7):791-794. doi:10.1001/jamaophthalmol.2017.1595

15. Adler-Milstein J, Jha AK. HITECH Act drove large gains in hospital electronic health record adoption. *Health Aff (Millwood)*. 2017;36(8):1416-1422. doi:10.1377/hlthaff.2016.1651

16. Lim MC, Boland MV, McCannel CA, et al. Adoption of electronic health records and perceptions of financial and clinical outcomes among ophthalmologists in the United States. *JAMA Ophthalmol*. 2018;136(2):164-170. doi:10.1001/jamaophthalmol.2017.5978

17. Liao KP, Cai T, Savova GK, et al. Development of phenotype algorithms using electronic medical records and incorporating natural language processing. *BMJ*. 2015;350:h1885. doi:10.1136/bmj.h1885

18. American Medical Association. *Physician ICD-9-CM 2006: International Classification of Diseases, 9th Revision, Clinical Modification*. Vol 1. Chicago, IL: AMA Press; 2006.

19. American Medical Association. *ICD-10-CM 2017: The Complete Official Code Book*. Chicago, IL: AMA Press; 2016.

20. Anastasopoulos E, Topouzis F, Wilson MR, et al. Characteristics of pseudoexfoliation in the Thessaloniki Eye Study. *J Glaucoma*. 2011;20(3):160-166. doi:10.1097/IJG.0b013e3181d9d8bd

21. Ariga M, Nivean M, Utkarsha P. Pseudoexfoliation syndrome. *J Curr Glaucoma Pract*. 2013;7(3):118-120. doi:10.5005/jp-journals-10008-1148

22. Hanauer DA. EMERSE: the electronic medical record search engine. *AMIA Annu Symp Proc*. 2006;941.

23. Kim SM, Kim Y, Jeong K, Jeong H, Kim J. Logistic LASSO regression for the diagnosis of breast cancer using clinical demographic data and the BI-RADS lexicon for ultrasonography. *Ultrasonography*. 2018;37(1):36-42. doi:10.14366/usg.16045

24. Naumann GO, Schlötzer-Schrehardt U, Küchle M. Pseudoexfoliation syndrome for the comprehensive ophthalmologist: intraocular and systemic manifestations. *Ophthalmology*. 1998;105(6):951-968. doi:10.1016/S0161-6420(98)96020-1