Searching in the Dark: Phenotyping Diabetic Retinopathy in a De-Identified Electronic Medical Record Sample of African Americans

Nicole A. Restrepo, PhD[1], Eric Farber-Eger, B.S.[2], Dana C. Crawford, PhD[1,3]

Case Western Reserve University, Department of Epidemiology and Biostatistics, Cleveland, Ohio; Vanderbilt University Medical Center, Vanderbilt Institute for Clinical and Translational Research, Nashville, Tennessee; Case Western Reserve University, Institute for Computational Biology, Cleveland, Ohio

Abstract

A hurdle to EMR-based studies is the characterization and extraction of complex phenotypes not readily defined by single diagnostic/procedural codes. Here we developed an algorithm utilizing data mining techniques to identify a diabetic retinopathy (DR) cohort of type-2 diabetic African Americans from the Vanderbilt University de-identified EMR system. The algorithm incorporates a combination of diagnostic codes, current procedural terminology billing codes, medications, and text matching to identify DR when gold-standard digital photography results were unavailable. DR cases were identified with a positive predictive value of 75.3% and an accuracy of 84.8%. Controls were classified with a negative predictive value of 1.0% as could be assessed. Limited studies of DR have been performed in African Americans who are at an elevated risk of DR. Identification of EMR-based African American cohorts may help stimulate new biomedical studies that could elucidate differences in risk for the development of DR and other complex diseases.

Introduction

The Precision Medicine Initiative (PMI)(1) promises resources and much needed epidemiologic and clinical data designed to provide a better understanding of the factors underlying the known inter-individual differences in susceptibility, onset, prognosis, and treatment of disease(2). In parsing out the finer details of molecular and genetic variability, precision medicine (PM) offers the opportunity to "tailor better treatment options" for patients. This concept is best highlighted by near-term PMI objectives in precision oncology with marked success in improving overall survival rates with tailored treatment regimens for patients with triple negative breast cancer(3). Long-term objectives aim to target other diseases such as diabetes, which afflicts an estimated 9.3% of the total American population. Although much progress has been made in studying the environmental and genetic risk factors for diabetes and diabetic complications(4,5), the role that ancestry-specific genetics plays is still being discovered.

One major complication of diabetes that disproportionally develops in African Americans compared with other groups is diabetic retinopathy(6,7), a disease of the retina that can lead to severe vision loss and blindness. The American Diabetes Association currently sets diagnostic standards for diabetes as a fasting blood glucose level of >126 mg/dl, glycated hemoglobin of > 6.5%, and oral glucose tolerance test level of > 200 mg/dl(8). Clinical testing thresholds for glycated hemoglobin levels were set by the International Expert Committee as a result of trials that found that European Americans with HbA1c at 6.5% had a prevalence of diabetic retinopathy of 6.3%(9). African Americans have a prevalence of diabetic retinopathy of 13.1% at the same HbA1c level, suggesting that current diagnostic criteria should be reevaluated for individuals of differing racial/ethnic backgrounds(10,11). After accounting for traditional risk factors(12), African Americans from the Veterans Affairs Diabetes Trial had a higher frequency of severe DR and a greater risk (OR = 2.30) of macular edema(13) compared to their European American counterparts.

Utilizing genomic data to more effectively classify patients into clinically distinct subpopulations is one of the major concepts and hope of PM. Unfortunately, in DR very few genetic association studies have been conducted in diverse populations(14–17) and to date, none have been performed in African-descent populations. Electronic medical records (EMRs) coupled with DNA databanks offer the opportunity to study the genetic and molecular profiles of clinical subpopulations within the clinic setting. EMRs typically contain several years of medical information that can be interrogated for a range of phenotypes. Also, large sample sizes can be acquired from the development of phenotype algorithms standardized for use across multiple institutions(18,19).

The promise of PM is not the development of new pharmaceuticals, but a greater understanding of inter-individual variability at the molecular and genomic level and how this knowledge can lead to selection of the best prevention and treatment options for the individual. Part and parcel of this process will be the understanding of how racial/ethnic differences affect disease risk and progression. In an effort to begin this process, we have developed phenotype algorithms to identify African American type 2 diabetics (T2D) with and without diabetic retinopathy. The algorithms were deployed in a de-identified EMR and tested for reliability and accuracy. In going forward, differentiating between population-specific risk factors may allow for more targeted studies to elucidate differences in the underlying etiology of common, complex diseases.

Materials and Methods

Study population

The Vanderbilt University Medical Center (VUMC) Synthetic Derivative (SD) is a de-identified version of Vanderbilt's institutionally developed EMR system StarChart. The SD contains inpatient and outpatient medical records collected at VUMC and affiliated clinics. Patient records consist of both structured (e.g., billing codes, procedure codes, laboratory values) and unstructured (e.g., clinical free text) data. To date, the Vanderbilt EMR contains over 2.2 million records with each record containing on average 6.5 years of medical history and an average of eight prescriptions. The SD is linked with VUMC's DNA repository (BioVU)(20). These DNA samples are extracted from discarded blood samples collected from outpatient clinical laboratories for use by researchers. We as part of the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed all DNA samples and data from non-European Americans within BioVU as of 2011 for genetic association studies(21). These data are collectively referred to here as "EAGLE BioVU."

A challenge with the use of the Vanderbilt EMR for ocular research is that the specialty eye clinic (Vanderbilt Eye Institute (VEI)) lacks structured fields for upload of ocular specific test results, such as intraocular pressure and cup-to-disk ratios. Additionally, due to the resources required to de-identify images from fundus exams, these images have not yet been made available in the SD. Given these limitations, we sought to develop a phenotype algorithm using searchable and parsable elements available in the SD for the identification of ocular cohorts that could be utilized in biomedical studies down the line. With this goal in mind, we have 1) developed and implemented a data-mining algorithm to identify African American type-2 diabetics with DR and those who are free of DR; and 2) manually verified the case/control status of individuals to evaluate the algorithm's performance.

Ethics statement

BioVU study subjects are not consented. DNA is collected from discarded blood samples remaining after routine clinical testing and is linked to de-identified medical records. According to the Vanderbilt Institutional Review Board (IRB) and the Federal Office of Human Research Protections provisions, the Vanderbilt protocol is considered nonhuman subjects research because no identifying information is available to the investigators. (The Code of Federal Regulations, 45 CFR 46.102 (f)). The IRB at Vanderbilt University approved this research.

Development of Phenotype Algorithm

The DR algorithms relied on the use of billing codes (International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes), procedure codes (Current Procedural Terminology (CPT) codes), medications, and free text searches to discriminate between cases and controls from among patients with T2D (Figure 1). To identify type-2 diabetics among African Americans in BioVU, we applied a T2D algorithm developed as part of the Electronic Medical Records and Genomics (eMERGE) Network(19). The eMERGE Network utilized the clinical diagnostic criteria set forth by the American Diabetes Association and categorized individuals based on data extracted from the EMR. In brief, the eMERGE Network excluded individuals with ICD-9-CM codes for type 1 diabetes (T1D). For individuals with ICD-9-CM codes for T2D, cases were required to have 1) a prescription for insulin or 2) a prescription for a T2D medication. Then in conjunction with either insulin/T2D medications the individual must have *either* 1) more than two clinic visits with a recorded T2D diagnosis or 2) a prescription of T2D medication prior to the insulin prescription. Cases were also identified among individuals without a prescription for insulin but with a prescription for a T2D medication and an abnormal glucose or glycated hemoglobin test result. As previously described(19), the eMERGE Network's T2D algorithm, deployed and validated in part within the VUMC SD, achieved 98% and 100%

positive predictive values for case and control identification of T2D.

We applied the eMERGE Network's T2D algorithm to 11,521 African Americans in BioVU. A total of 630 cases of T2D were identified in BioVU. These type 2 diabetics were then included in the following study for the identification of DR cases and controls.

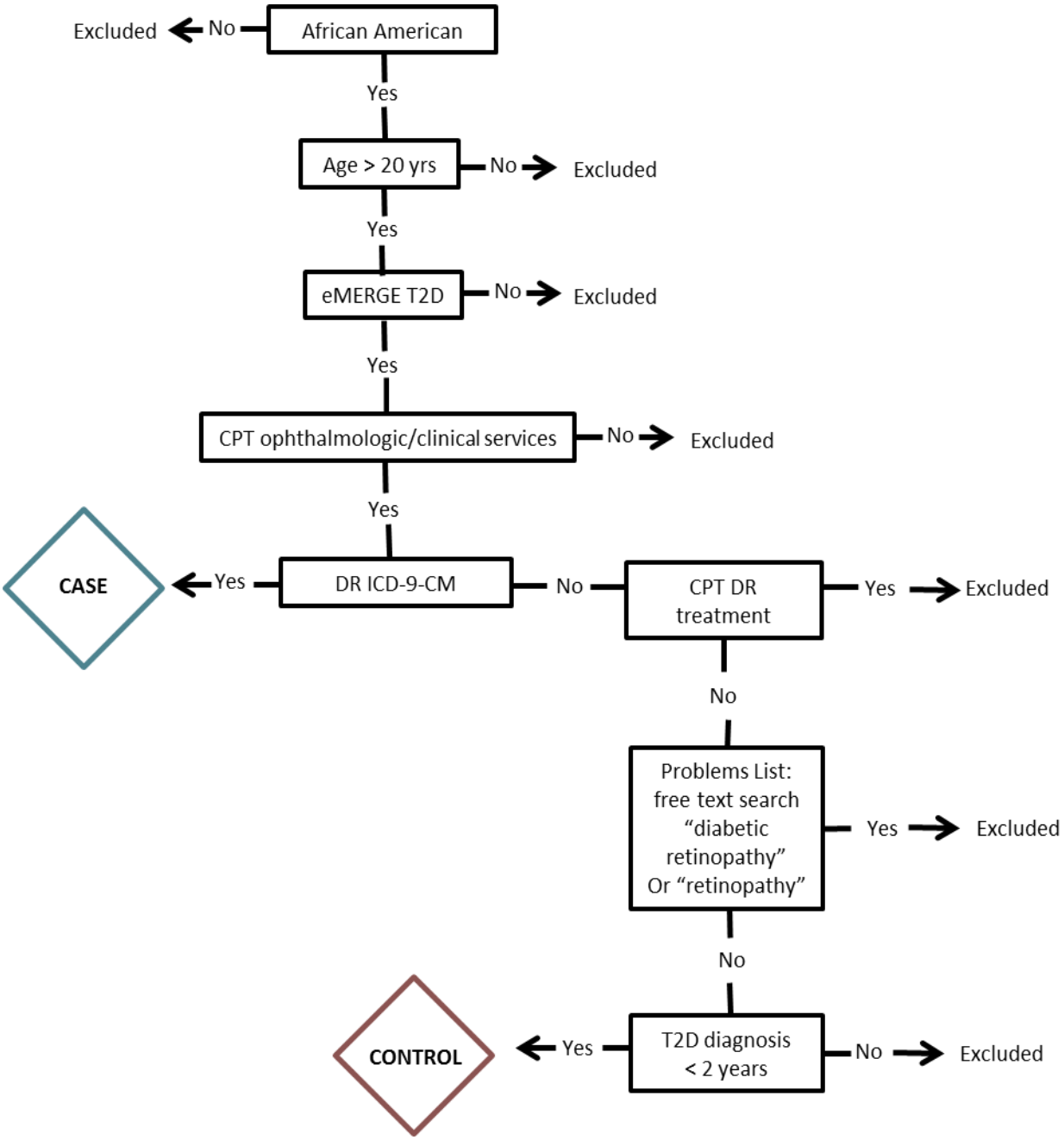

Figure 1: Decision tree for diabetic retinopathy case and control algorithms.

Initial screening criteria for study population

Individuals included for this study were African American adults with T2D over the age of 20 years as of March 20, 2013 with a minimum of either one mention of a CPT code for ophthalmology or a CPT code for general clinic procedures (Table 1) in the medical record. As part of the eMERGE Network algorithm for T2D, individuals were excluded if medical records contained an ICD-9-CM for T2D (250.xx) before the age of twenty years as this individual is likely a type 1 diabetic.

Table 1: Ophthalmology and clinical CPT codes	
92002, 92004	Ophthalmological services: medical examination and evaluation with initiation of diagnostic and treatment program
92012, 92014	Ophthalmological services: medical examination and evaluation, with initiation or continuation of diagnostic and treatment program
99201, 99202, 99203, 99204, 99205	Office/outpatient visit for evaluation of new patient, problem focused history, exam, counseling, and/or coordination with other physicians or health care professionals
99212, 99213, 99214, 99215	Office/outpatient visit for evaluation and management of established patient: problem focused history, exam, medical decision making, counseling, and/or coordination with other physicians or health care professionals
99307, 99308, 99309, 99310	Nursing facility care for evaluation and management of patient, problem focused interval history, exam, medical decision making, counseling, and/or coordination with other physicians or health care professionals
99324, 99325, 99326, 99327, 99328	Domiciliary or rest home visit for evaluation and management of new patient, problem focused history, exam, medical decision making, counseling, and/or coordination with other physicians or health care professionals
99334, 99335, 99336, 99337	Domiciliary or rest home visit for evaluation and management of established patient, problem focused history, exam, medical decision making, counseling, and/or coordination with other physicians or health care professionals

Diabetic retinopathy cases

DR cases were individuals identified as having T2D as defined by the eMERGE Network and at least one mention of a DR ICD-9-CM code (Table 2), excluding ICD-9-CM 362.01, in conjunction with at least one mention of a clinic or ophthalmology CPT code (Table 3). ICD-9-CM 362.01 was excluded from the case definition given that background diabetic retinopathy, the presence of microaneurisms and hemorrhages, may resolve on its own and does not impede vision.

Table 2: International Classification of Disease, Ninth Edition, Clinical Modification (ICD-9-CM) codes for diabetic retinopathy

362.0	Diabetic retinopathy
362.01	Background diabetic retinopathy
362.02	Proliferative diabetic retinopathy
362.03	Non-proliferative diabetic retinopathy
362.04	Mild non-proliferative diabetic retinopathy NOS
362.05	Moderate non-proliferative diabetic retinopathy
362.06	Severe non-proliferative diabetic retinopathy

Diabetic retinopathy controls

Controls are defined as cases of T2D whose medical records included a CPT code for clinic/ophthalmology visit (Table 1) but excluded any of the following three: 1) any ICD-9-CM for DR (Table 2), 2) any CPT code for treatment procedures commonly used in the treatment of DR (Table 3), and 3) any text mention of "diabetic retinopathy" or "retinopathy" in the problems lists. Lastly, controls with T2D duration of less than two years, with date of diagnosis determined from the first mention of T2D ICD-9-CM, were removed as they could potentially develop into future incident cases. It should be noted that in the case of ophthalmology visit CPT codes, it cannot be presumed that patients had a dilated fundus examination or that physicians looked for or made note of the presence of any retinopathy.

After the initial manual review of controls, it was noted that some were being excluded. Part of the T2D routine medical plan at Vanderbilt includes an annual eye exam to screen for potential development of DR. As such, compliant T2D cases will have DR ICD-9-CM codes in his or her EMR marking these annual eye exams regardless of DR diagnosis. After additional review, we adjusted the inclusion/exclusion criteria for controls to allow for the inclusion of ICD-9-CM codes for background diabetic retinopathy (362.01) and mild nonproliferative DR (362.04). These two codes were selected based on initial case chart reviews that were determined to be false positives. DR controls with ICD-9-CM codes (362.01 and 362.04) were flagged for additional screening. Inclusion of ICD-9-CM codes 362.01 and 362.04 resulted in an additional 15 individuals meeting control criteria.

Table 3: Current procedural terminology codes for common treatment options for diabetic retinopathy

Procedures	**CPT Codes**
Impltj Intravitreal Drug Dlvr Sys Rmvl Vts Implantation of intravitreal drug delivery system (eg, ganciclovir implant), includes concomitant removal of vitreous	67027
Vtrc Mchnl Pars Plna Vtrc Mchnl Pars Plna Focal Endolaser Pc Vtrc Mchnl Pars Plna Endolaser Panrta Pc Vitrectomy Pars Plana Remove Preretinal Membrane Vitrectomy, mechanical, pars plana approach;……. (eg, macular pucker) photocoagulation, drainage of subretinal fluid, scleral buckling, and/or removal of lens	67039 67040 67041 67113
Destruction of localized lesion of retina (e.g., macular edema, tumors), 1 or more sessions; cryotherapy, diathermy	67208

Extraction and calculation of individual demographic elements

Demographic data and laboratory measurements were extracted and calculated as follows: for cases, age at DR diagnosis was determined by the date in the records for the first mention of a DR ICD-9-CM (Table 2). For controls, age at last clinic visit (LCV) was taken as the date of the last CPT mentioned in the records. Median values were calculated for the following laboratory measurements within a two-year window of an individual's DR diagnosis or LCV for controls: blood pressure (systolic and diastolic), lipids (total cholesterol, high-density cholesterol, low-density cholesterol, and triglycerides), and body mass index (height and weight).

Manual review

The SD records of *all* BioVU African American DR cases and a random sample of 100 controls identified by the algorithm were manually reviewed to verify DR case/control status. Of the 158 individuals initially identified as cases, 119 were determined to be definite cases based on records retrieved from the SD. The records accessed for verification of case status were primarily composed of but not limited to surgical reports, optometry, and ophthalmology clinic notes. Other data taken into account included medication lists, general and specialty clinic reports, clinical communications, and problems lists. DR cases were classified as either a definite or potential case. An individual was classified as a definite case if they met one of two criteria: 1) a written diagnosis by a Vanderbilt ophthalmologist/optometrist as pertained to a patient's diabetic retinopathy status (Example, Figure 2) or 2) the patient's EMR contained *both* of the following: at least three independent mentions of a DR ICD-9-CM code and a surgical procedure for treatment of DR complications as identified by a surgical report. Surgical procedures for treatment of DR complications may include membrane peel, pars plana vitrectomy, scleral buckle, laser/photo coagulation, silicone oil, and intraocular/expanding gas. We classified individuals whose records were positive for text mention of DR case status but lacked surgical notes or ophthalmology clinic notes as "potential cases." The criteria for potential cases include *at least one* of the following: at least three mentions of DR ICD-9-CM with text mention of "diabetic retinopathy", text mention(s) for "surgery for diabetic retinopathy" in a clinic note or problems history, or thorough ophthalmology/optometry notes with diagnosis of background DR. Of the records reviewed, we identified 119 definite cases and 26 potential cases. Thirteen individuals were determined to be false positives. False positives tended to be T2D cases with thorough ophthalmology/optometry notes that explicitly state an individual was clear of signs of DR at the time of last visit (n = 7), records that lacked sufficient data to determine a diagnosis (n = 2), T1D (n = 1), or individuals diagnosed with other clinical forms of retinopathy such as "hypertensive" or "herpes retinitis" (n = 3).

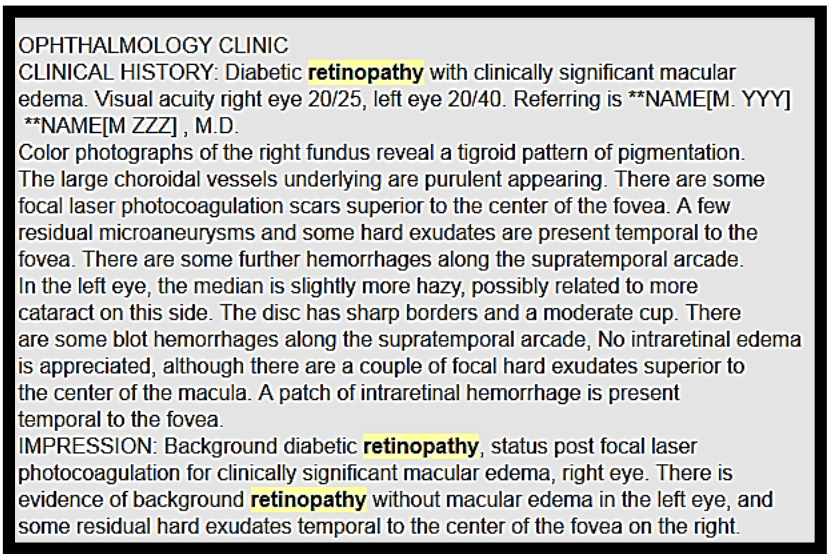
OPHTHALMOLOGY CLINIC
CLINICAL HISTORY: Diabetic **retinopathy** with clinically significant macular edema. Visual acuity right eye 20/25, left eye 20/40. Referring is **NAME[M. YYY] **NAME[M ZZZ] , M.D.
Color photographs of the right fundus reveal a tigroid pattern of pigmentation. The large choroidal vessels underlying are purulent appearing. There are some focal laser photocoagulation scars superior to the center of the fovea. A few residual microaneurysms and some hard exudates are present temporal to the fovea. There are some further hemorrhages along the supratemporal arcade. In the left eye, the median is slightly more hazy, possibly related to more cataract on this side. The disc has sharp borders and a moderate cup. There are some blot hemorrhages along the supratemporal arcade, No intraretinal edema is appreciated, although there are a couple of focal hard exudates superior to the center of the macula. A patch of intraretinal hemorrhage is present temporal to the fovea.
IMPRESSION: Background diabetic **retinopathy**, status post focal laser photocoagulation for clinically significant macular edema, right eye. There is evidence of background **retinopathy** without macular edema in the left eye, and some residual hard exudates temporal to the center of the fovea on the right.

Figure 2 : Screen shot of a de-identified clinic note from the Ophthalmology Clinic at Vanderbilt University Medical Center, as seen in the Synthetic Derivative. Notes pertain to a patient's retinal eye exam and diagnosis.

At total of 434 individuals met DR control criteria. Of these, 100 controls were randomly selected for manual chart review. Each record was searched for text mention of "diabetic retinopathy" in all clinic notes and problems lists. Of these 100, all were clear of either a text mention of DR or else contained an ophthalmology/optometry report with negative findings for DR.

Calculation of PPV, NPV, and Accuracy

Performance of case and control algorithms were calculated as follows: positive predictive value (PPV) was calculated as the ratio of true positives (TP) over TP + false positive (FP) [PPV = TP/(TP + FP)]. A TP was an individual who was identified by the case algorithm as a case and was then confirmed by manual review to be a true case. A FP in turn was an individual identified as a case that was determined not to be a case during manual review. Negative predictive value (NPV) is the ratio of true negative (TN) (i.e., a control who was confirmed as a control) over TN and FN [NPV = TN/(TN + FN)]. Lastly, accuracy was calculated as the ratio of the sum of TP and TN over the sum of all positives and all negatives [Accuracy = (TP + TN)/(Positives (TP + FP) + Negatives (TN + FN))].

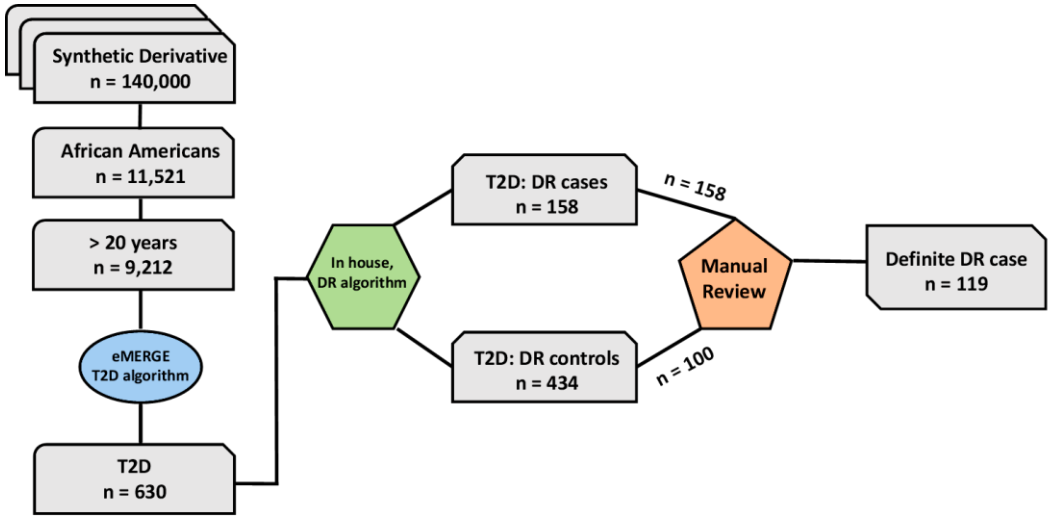

Figure 3: Flow chart of the global study design.

Results

Algorithm performance

The PPV for definite DR cases was 75.3% with an accuracy of 84.8% (Table 5). With the inclusion of potential cases, PPV increased to 91.7% and accuracy of 94.9% (Table 5). We manually reviewed the de-identified medical records in the SD of 100 randomly selected controls identified by the algorithm to calculate NPV and accuracy. Performance of the DR control algorithm was found to have a NPV of 1.0% (Table 5).

Table 5: Evaluation of diabetic retinopathy phenotype algorithm in African Americans from EAGLE BioVU

	Sample Size	Manually reviewed	PPV	NPV	Accuracy
Cases	158	158	-	-	-
-Definite		119	75.3%	-	84.8%
-Potential		26	91.7%	-	94.9%
Controls	434	100	-	1.0	-

Study population characteristics

The study population consisted of a total of 630 cases of T2D among African Americans, of which 119 were identified as definite DR cases and 434 as DR controls (Figure 3). As expected, the median age of cases was older than controls (65.0 versus 63.5 years; Table 6). Controls were predominately female and tended to have better control of their diabetes compared with cases (HbA1c = 6.7% vs. 10.6%). On average, both cases and controls were obese (median body mass index > 30.0 kg/m^2).

Table 6: Demographics of EAGLE BioVU diabetic retinopathy cases and controls in African Americans

	Diabetic retinopathy	
	Cases (> 20 yrs)	Controls (>60 yrs)
N	119	434
Median Age (yrs)	65.0	63.5
% female	39%	61%
Median BMI (kg/m^2)	31.1	32.4
Systolic (mmHg)	139.1	132.7
Diastolic (mmHg)	76.3	77.9
Total Cholesterol (mg/dL)	194.5	165.0
Glucose (mg/dL)	222.0	117.0
LDL-C (mg/dL)	111	91.0
HbA1c	10.6	6.7

Discussion

In this study, under definite case status, the DR case algorithm had a PPV of 75.3% and an accuracy of 84.8 % with a NPV of 1.0% for the control algorithm. In a similar study performed in the eMERGE network, a diabetic retinopathy algorithm was developed and deployed at both the Marshfield Clinic and Vanderbilt University (https://phekb.org/phenotype/diabetic-retinopathy). While our case algorithm did not perform as well as the eMERGE algorithm in the Marshfield Clinic (case: PPV = 80%) it outperformed the eMERGE algorithm deployed at Vanderbilt University (case: PPV = 67.5%)(18). The difference in performance may be due to dissimilarities between the algorithms. The eMERGE DR case algorithm may have performed better in the Marshfield clinic as a result of screening out patients with a "negative mention of DR" as part of their initial, whereas, we performed this function manually after cases were automatically screened by our algorithm. While we cannot directly test for performance enhancing elements across the two algorithms, our case algorithm versus the eMERGE algorithm likely performed better at Vanderbilt due in part that we did not include patients without an ICD-9 for DR. In comparison, the eMERGE algorithm allowed for patients with a text mention of DR in the Problems List that were absent an ICD-9 for DR. While more work is needed to refine these phenotype algorithms, they provide an excellent starting point for ascertaining study populations for use in biomedical studies.

This study was impacted by challenges such as missing data, which can introduce misclassification bias. The control algorithm developed here was designed around the concept that an individual is free of DR. However, without a complete eye exam, there is the potential that a control is an undiagnosed case. Misclassification and potential ascertainment bias are also possible in case identification where cases are only those individuals who have been evaluated by a specialist and are therefore potentially extreme or overly symptomatic cases. A well-known barrier for individuals seeking medical attention is low socioeconomic status which disproportionally affects African Americans(22). This may in part explain the limited number of African American DR cases in BioVU (n=119). The rate of DR among adults >40 years in BioVU T2D individuals (21.6%) is notably lower than would be expected in the general U.S. African American population of diabetics over 40 years (36.7%)(23). Epidemiologic surveys conduct

eye examinations on all participants regardless of health status at the time of exam, a protocol designed to accurately estimate the prevalence or incidence of DR. In comparison, clinical cases of DR such as those identified in BioVU may represent cases being diagnosed once vision loss becomes severe, thereby reflecting a seemingly decreased prevalence compared with epidemiologic surveys.

Despite the limitations in data access, such as lack of access to fundus photographs, we were able to characterize diabetic retinopathy from the VUMC's SD without the gold standard data available for research. We have a diverse population connected to a depth of medical data, even if not all of it is easily searchable. Despite the small sample sizes, the present study has made available more case counts for African Americans with DR available for study in the larger, collaborative scientific community. The ability to extract complex ocular phenotypes(24) from EMRs will provide researchers with previously unaccessed datasets to further advances in ocular genetic research and vision-loss prevention.

In conclusion, utilization of EMRs will allow for furthering the Precision Medicine Initiative by enabling studies of population genetics of highly prevalent diseases, such as diabetic retinopathy. In understanding how disease susceptibility varies across racial/ethnic populations we can better understand how such differences may affect disease onset and progression in individuals.

Acknowledgments

This work was supported in part by NIH grant U01 HG004798 and its ARRA supplements. The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant funded by the National Center for Research Resources, Grant UL1 RR024975-01, which is now at the National Center for Advancing Translational Sciences, Grant 2 UL1 TR000445-06.

References

1. Precision Medicine Initiative - National Institutes of Health (NIH) [Internet]. [cited 2015 Jul 19]. Available from: http://www.nih.gov/precisionmedicine/

2. Collins FS, Varmus H. A New Initiative on Precision Medicine. N Engl J Med. 2015 Feb 26;372(9):793–5.

3. Minckwitz G von, Martin M. Neoadjuvant treatments for triple-negative breast cancer (TNBC). Ann Oncol. 2012 Aug 1;23(suppl 6):vi35–9.

4. Mohlke KL, Boehnke M. Recent advances in understanding the genetic architecture of type 2 diabetes. Hum Mol Genet. 2015 Jul 9;

5. Kahn SE, Cooper ME, Del Prato S. Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. Lancet Lond Engl. 2014 Mar 22;383(9922):1068–83.

6. Zhang X, Cotch MF, Ryskulova A, Primo SA, Nair P, Chou C-F, et al. Vision health disparities in the United States by race/ethnicity, education, and economic status: findings from two nationally representative surveys. Am J Ophthalmol. 2012 Dec;154(6 Suppl):S53–62.e1.

7. Klein R, Klein BEK. The Prevalence of Age-Related Eye Diseases and Visual Impairment in Aging: Current Estimates. Invest Ophthalmol Vis Sci. 2013 Dec 1;54(14):ORSF5–13.

8. Alex ADA 1701 NBS, ria, 1-800-Diabetes V 22311. Diagnosing Diabetes and Learning About Prediabetes [Internet]. American Diabetes Association. [cited 2015 Jul 20]. Available from: http://www.diabetes.org/diabetes-basics/diagnosis/

9. Committee* TIE. International Expert Committee Report on the Role of the A1C Assay in the Diagnosis of Diabetes. Diabetes Care. 2009 Jul 1;32(7):1327–34.

10. Cheng YJ, Gregg EW, Geiss LS, Imperatore G, Williams DE, Zhang X, et al. Association of A1C and fasting plasma glucose levels with diabetic retinopathy prevalence in the U.S. population: Implications for diabetes diagnostic thresholds. Diabetes Care. 2009 Nov;32(11):2027–32.

11. Tsugawa Y, Mukamal KJ, Davis RB, Taylor WC, Wee CC. Should the hemoglobin A(1c) diagnostic cutoff differ between blacks and whites?: a cross-sectional study. Ann Intern Med. 2012 Aug 7;157(3):153–9.

12. Emanuele N, Sacks J, Klein R, Reda D, Anderson R, Duckworth W, et al. Ethnicity, race, and baseline retinopathy correlates in the veterans affairs diabetes trial. Diabetes Care. 2005 Aug;28(8):1954–8.

13. Emanuele N, Moritz T, Klein R, Davis MD, Glander K, Khanna A, et al. Ethnicity, race, and clinically significant macular edema in the Veterans Affairs Diabetes Trial (VADT). Diabetes Res Clin Pract. 2009 Nov;86(2):104–10.

14. Sheu WH-H, Kuo JZ, Lee I-T, Hung Y-J, Lee W-J, Tsai H-Y, et al. Genome-wide association study in a Chinese population with diabetic retinopathy. Hum Mol Genet. 2013 Aug 1;22(15):3165–73.

15. Fu Y-P, Hallman DM, Gonzalez VH, Klein BEK, Klein R, Hayes MG, et al. Identification of Diabetic Retinopathy Genes through a Genome-Wide Association Study among Mexican-Americans from Starr County, Texas. J Ophthalmol. 2010;2010.

16. Huang Y-C, Lin J-M, Lin H-J, Chen C-C, Chen S-Y, Tsai C-H, et al. Genome-wide association study of diabetic retinopathy in a Taiwanese population. Ophthalmology. 2011 Apr;118(4):642–8.

17. Awata T, Yamashita H, Kurihara S, Morita-Ohkubo T, Miyashita Y, Katayama S, et al. A genome-wide association study for diabetic retinopathy in a Japanese population: potential association with a long intergenic non-coding RNA. PloS One. 2014;9(11):e111715.

18. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. J Am Med Inform Assoc JAMIA. 2013 Jun;20(e1):e147–54.

19. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. J Am Med Inform Assoc JAMIA. 2012 Apr;19(2):212–8.

20. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balser JR, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. Clin Pharmacol Ther. 2008 Sep;84(3):362–9.

21. Crawford DC, Goodloe R, Farber-Eger E, Boston J, Pendergrass SA, Haines JL, et al. Leveraging Epidemiologic and Clinical Collections for Genomic Studies of Complex Traits. Hum Hered. 2015;79(3-4):137–46.

22. Anderson NB, Bulatao RA, Cohen B, National Research Council (US) Panel on Race E. Race/Ethnicity, Socioeconomic Status, and Health [Internet]. 2004 [cited 2013 Nov 11]. Available from: http://www.ncbi.nlm.nih.gov/books/NBK25526/

23. Wong TY, Klein R, Islam FMA, Cotch MF, Folsom AR, Klein BEK, et al. Diabetic Retinopathy in a Multi-ethnic Cohort in the United States. Am J Ophthalmol. 2006 Mar;141(3):446–55.

24. Restrepo NA, Farber-Eger E, Goodloe R, Haines JL, Crawford DC. Extracting Primary Open-Angle Glaucoma from Electronic Medical Records for Genetic Association Studies. PloS One. 2015;10(6):e0127817.