Exploring Diabetic Retinopathy and Macular Edema Progression through Data Analytics

Maria Mayorga PhD, Julie Ivy PhD, Erik Rosenstrom B.S., Priscille Koutouan B.S., Zack Horton

1. Introduction

DR is a growing issue in America, and the world, so it is important that we understand it, and learn to predict its progression:

- From 2000 to 2010, the number of DR cases grew by 89%, from 4.06 million to 7.69 million (National Eye Institute, 2020)
- Diabetic retinopathy is a leading cause of blindness in American adults, and the most common diabetes disease (National Eye Institute, 2020)

Partnering with Retinal Care Inc., we set out to explore, understand, and predict the progression of diabetic retinopathy (DR) and macular edema (ME) based on patient data they had collected.

There are four main areas of focus for our research:

1. Apply statistical and analytical methods to understand trends, anomalies, potential issues, and human error
2. Predict the time until a patient progresses to the next DR or ME
3. Build a Natural History Model to simulate patient progression
4. Create a Machine Learning Classification model that will predict if a patient will develop DR or ME within the next year

Diabetic Retinopathy Data and Statistics | National Eye Institute. (2020, November 19). Www.nei.nih.gov; National Eye Institute. https://www.nei.nih.gov/learn-about-eye-health/outreach-campaigns-and-resources/eye-health-data-and-statistics/diabetic-retinopathy-data-and-statistics

2. Modeling Patient Progression

Unlike previous research that has been conducted, we have sporadic and messy patient data which gives us a look at how diabetic retinopathy progresses in patients outside of a controlled, experimental setting.

We began step 1 of our main areas of focus, *understand the data*:

- Identified 5 different stages of DR severity
- Identified 2 different stages ME severity
- Created a continuous-time Markov chain (CTMC) to model patient progression, during the process, we also found the average years until transition for DR and ME

Descriptions of our 5 DR States

Number	Description
-1	No DR
1	Unspecified OR Mild DR
3	Moderate DR
4	Severe DR
5	Proliferative DR

Average Number of Years to Transition (DR)

	No DR	Mild DR	Moderate DR	Severe DR	Proliferative DR
No DR		7.025	5.767	5.793	6.239
Mild DR			1.589	1.333	1.997
Moderate DR				1.458	1.266
Severe DR					1.268

Description of our 2 ME States

Number	Description
1	No ME
2	ME Present

Average Number of Years to Transition (ME)

	No ME	ME Present
No ME		2.143

We found that some patients had multiple entries on the same day, sometimes for different eyes. This would skew our estimates to be lower-than-actual, and we will need to rerun our analysis once we better understand how this occurs, and how to work around it.

Patient	Problem	Status	Date	Severity
1	Type 2 Diab with Mild DR	Inactive	5/15/04	2
1	Type 1 Diab with Severe DR	Chronic	7/12/05	4
1	Type 1 Diab with Prolif. DR	Chronic	7/12/05	5

3. Methods of Analysis

Before beginning step 2, *prediction*, we wanted to understand the basic statistics of the different severity states, and compare them:

Units (days)	No DR	Mild DR	Moderate DR	Severe DR
Mean	2503	1038	619	617
Median	2277	702	483	540
Range	7865	7055	2119	1924
Std Dev	1791	1156	558	500

When creating our CTMC, we assumed a few things:

- Time between DR or ME states was exponentially distributed
- Patient will carry-forward their most severe state so far
- Patients could not get better, and could only become more severe

After making our general CTMC, we used statistical tests (Chi-Sq, A-D, and K-S) to make sure that our assumption was supported

Due to how different each state's progression seemed, we created 5 different linear regressions to identify important factors:

1. 1 General Model for all patients, regardless of current severity
2. 4 Models for DR severity states from -1 to 4 (No DR, Mild DR, Moderate DR, and Severe DR)

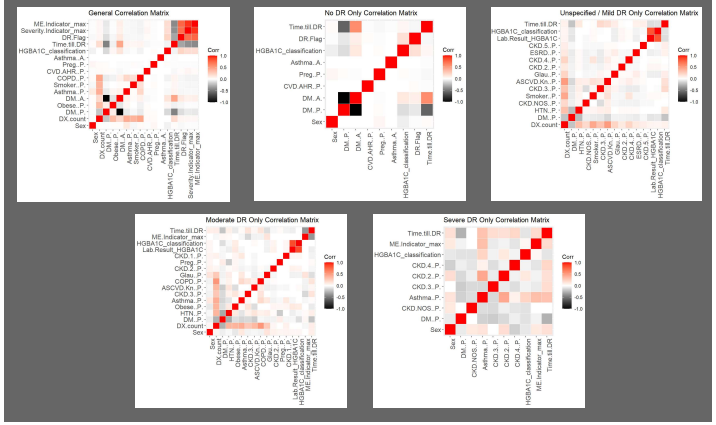
As we better understand what drives progression, it will improve our prediction, and simulation through the Natural History Model. Allowing us to look at how certain types of patients are affected depending on this age, gender, race, and other conditions

4. Insights from Linear Models

Model Type	Top 4 Key Factors of Progression Time
General Model	DM A, DM P, Asthma, and DR Status
No DR	DM A, DM P, Asthma, and CVD AHR P
Mild DR	CKD, HTN P, DM P, and ASCVD
Moderate DR	ME Status, CKD P, COPD, and Glaucoma
Severe DR	ME Status, CKD P, DM P, and Glaucoma

CKD – Chronic Kidney Disease
DM A / DM P – Diabetes Mellitus (A for Active, P for Pre)
CVD – Cardiovascular Disease
HTN – Hypertension
COPD – Chronic Obtrusive Pulmonary Disease
ASCVD - Atherosclerotic CVD

5. Correlation Matrices from Linear Models



6. Future Enhancements

- Re-run analysis on data for CTMC and average years to transition once the issue of multiple entries is fully understood
- To better predict the time until progression, explore other types of models outside linear regression
- Create specialized CTMC matrices for the Natural History Model
- Obtain additional data to improve accuracy of prediction, simulation, and classification
- Automate preliminary analytics for new data releases