Study on the Importance Effect of fasting Gluocose for the Detection of Retinopathy

Yalu Tong yalut/yalut@andrew.cmu.edu
Heinz College
Carnegie Mellon University
Pittsburgh, PA, United States

Youngjoo Son yson/yson@andrew.cmu.edu
School of Architecture
Carnegie Mellon University
Pittsburgh, PA, United States

Sidharth Srikumar ssrikuma/ssrikuma@andrew.cmu.edu
College of Engineering
Carnegie Mellon University
Pittsburgh, PA, United States

Abstract

Diabetic Retinopathy is an eye disease that affects the blood vessels of the retina: a light sensitive tissue at the back of the eye which triggers nerve impulses that are passed to the brain via the optic nerve to form images (National Eye Institute, 2015).

Estimates from the National Eye Institute indicate that between 40 to 45 percent of Americans diagnosed with diabetes have some stage of retinopathy. However, only half are aware of it since early stage retinopathy usually shows no symptoms. Given its effect on the retina, diabetic retinopathy is the leading cause of vision loss among diabetic patients as well as the leading cause of blindness and vision impairment among working age adults in America. This highlights the importance of early diagnosis and prompt treatment.

This study examines data from the NHANES (National Health and Nutrition Examination Survey) surveys in 2005-06 and 2007-08, and designs a lesion study by comparing the effect of fasting glucose in different situations and demonstrates the importance of fasting glucose measurements as a predictive variable in detecting diabetic retinopathy.

1. Introduction

Diabetic Retinopathy is a disease that affects the blood vessels of the retina that lines the back of the eye. It is the most common cause of vision loss among people with diabetes and the leading cause of the impairment and loss of vision among working age adults in America (National Eye Institute, 2015).

High blood sugar levels associated with diabetes can cause damage to the blood vessels in the retina, which in turn, is associated with retinopathy (National Eye Institute, 2015).

Diabetic retinopathy cause the blood vessels to leak fluid or hemorrhage, which can cause an impairment of the vision. In its most advanced stages, diabetic retinopathy can lead to the proliferation of new abnormal blood vessels on the surface of the retina, which can lead to scarring and cell loss in the retina. Diabetic retinopathy progresses in four stages (National Eye Institute, 2015):

1. Mild non-proliferative retinopathy: This stage involves small areas of swelling in the retinal blood vessels. These swellings are known as microaneurysms, and may leak fluid into the retina.

2. Moderate non-proliferative retinopathy: As the disease progresses, blood vessels that nourish the retina swell and in extreme cases, lose their ability to transport blood.

3. Severe non-proliferative retinopathy: More blood vessels are blocked, further depleting the blood supply to the retina. This, in turn, signals the retina to grow new blood vessels.

4. Proliferative diabetic retinopathy: At this advanced stage, there is a proliferation of new blood vessels, which grow on the surface of the retina and into the vitreous gel (the fluid that fills the eye). However, these new blood vessels are fragile and therefore, more likely to leak and bleed. This can cause the contraction of scar tissues, which can cause retinal detachment. Retinal detachment can lead to permanent vision loss.

Despite the harsh symptoms described above, early stages of retinopathy often show no symptoms. In fact, the disease often goes unnoticed until it begins to impair vision. Bleeding from the abnormal blood vessels contracted on the surface of the retina can cause the appearance of "floating" spots. Often, these spots clear on their own. However, in the absence of prompt treatment, the bleeding often recurs, and can cause retinal detachment which leads to permanent vision loss.

Current estimates indicate that between 40 and 45 percent of Americans diagnosed with diabetes have some stage of diabetic retinopathy. However, only about half of them are aware of it. Women who develop diabetes during pregnancy often experience rapid onset diabetic retinopathy (National Eye Institute, 2015).

The above described symptoms and effects associated with the progression of diabetic retinopathy make apparent the importance of early detection and prompt treatment of diabetic retinopathy. On the other hand, the lack of symptoms in the early stages of the disease make apparent the difficulty associated with detecting retinopathy.

To detect retinopathy, existing studies investigated fasting glucose as a predictor of retinopathy (Wong et al., 2008; Zoppini et al., 2009), and sometimes they correlated hemoglobin A1C as well (Cheng et al., 2009). They applied correlation analysis, statistical test, or built prediction models. Multiple predictors were being considered to give a precise prediction on retinopathy but the best combination of the predictors is not yet discovered.

This study looks at data made available via NHANES from nationwide surveys conducted in 2005-06 and 2007-08 and uses results from ophthalmology examinations to get insight into the prevelance of diabetic retinopathy among the survey respondents. Based on the unique sequence number assigned to each survey respondent, further data about the fasting glucose levels, demographic (gender, age), weight and sugar levels of these survey respondents. This data is then used to designs a lesion study by comparing the effect of fasting glucose in different situations and demonstrates the importance of fasting glucose measurements as a predictive variable in detecting diabetic retinopathy.

2. Background

Before getting into the experimental setup, the collected data and the methodology used to develop the models, some underlying concepts are explained.

1. The dependent variable: It is the variable that is being studied in the experiment. In this case, the dependent variable is the presence of diabetic retinopathy in the patient being studied. In the dataset, there is data about whether the patient has retinopathy in the left or right eye. These data are combined to create our dependent variable.

2. Independent variable: It is the variable that is being varied or controlled in the experiment in order to study its effect on the dependent variable. As stated, the study is aimed to demonstrate the importance of the fasting glucose levels in predicting the presence or absence of retinopathy. Thus, the fasting glucose levels in the patients being studied is the independent variable.

3. Co-variates: These are variables present in the model which can be expected to affect the value of the dependent variable. In this study, factors such as the results of ophthalmology examinations, the weight and body measurements of the patients being studied, energy consumption over the last 30 days etc.

4. Logistic Regression: For the purpose of our predictive model, a logistic regression model is used. Logistic regression is a classification algorithm which gives an output between 0 and 1. For the purposes of this study, a binary classification logistic regression model is used to classify whether or not the patient under examination has diabetic retinopathy. An output of 1 indicated presence of diabetic retinopathy while an output of 0 zero indicates absence.

5. Training data: It is the data using which the logistic regression model is developed. For the purpose of this study, 70 percent of the total available data from the NHANES surveys is used by the machine learning algorithm to learn the predictive model.

6. Test data: Once the predictive model is learned on the training data, the accuracy of the learned model is assessed using its performance on the test data set.

7. Retinopathy: Diabetic Retinopathy is a disease that affects the blood vessels of the retina that lines the back of the eye. High blood sugar levels associated with diabetes can cause damage to the blood vessels in the retina, which in turn, is associated with retinopathy (National Eye Institute, 2015).

The study design that we used is inspired by lesion study which is a research method in which areas of the brain are removed or disabled in order to determine their specific functions(n.d, Lesioning Studies.). In our project, we keep or remove the fasting glucose variables in two different situations and compare the results. Details about study design can be seen in section 3.1.

3. Experimental Setup

3.1 Cohort Selection

The dataset that we choose are from National Health and Nutrition Exanimation Survey (NHANES) dataset, which contains examination, demographics, laboratory, questionnaire and so on. Our sample are from people who took the Ophthalmology examination in the survey during 2007-2008 and 2005-2006 of retinopathy with related information about their

demographic, basic body measure, fast glucose testing data. Details about feature choosing will be introduced in 5.3. Below (Table 1) is the samples descriptive statistics.

OPDUHEM	OPDUHE	OPDUSE	RIAGENDR	DMDHRAGE
Min. :0.00000	Min. :0.00000	Min. :0.00000	Min. :1.000	Min. :40.00
1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:0.00000	1st Qu.:1.000	1st Qu.:49.00
Median :0.00000	Median :0.00000	Median :0.00000	Median :2.000	Median :60.00
Mean :0.03828	Mean :0.05338	Mean :0.05726	Mean :1.503	Mean :59.86
3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:0.00000	3rd Qu.:2.000	3rd Qu.:70.00
Max. :1.00000	Max. :2.00000	Max. :2.00000	Max. :2.000	Max. :85.00
NA's :215				

RIDRETH1	DMDFMSIZ	DMDHHSIZ	VIQ200	VIDLOVA	VIDROVA
Min. :1.000	Min. :1.000	Min. :1.000	Min. :1.000	Min. : 25.00	Min. : 25.00
1st Qu.:3.000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.:2.000	1st Qu.: 25.00	1st Qu.: 25.00
Median :3.000	Median :2.000	Median :2.000	Median :2.000	Median : 25.00	Median : 25.00
Mean :2.909	Mean :2.578	Mean :2.677	Mean :1.887	Mean : 45.09	Mean : 45.05
3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:3.000	3rd Qu.:2.000	3rd Qu.: 30.00	3rd Qu.: 30.00
Max. :5.000	Max. :7.000	Max. :7.000	Max. :9.000	Max. :666.00	Max. :666.00
			NA's :2	NA's :3506	NA's :3535

VIQ150	BMXBMI	BMXWT	LBXGLU	RETI
Min. :1.000	Min. : 13.36	Min. : 35.90	Min. : 45.0	Min. :0.0000
1st Qu.:1.000	1st Qu.: 24.91	1st Qu.: 68.10	1st Qu.: 96.0	1st Qu.:0.0000
Median :1.000	Median : 28.35	Median : 79.80	Median :103.0	Median :0.0000
Mean :1.138	Mean : 29.25	Mean : 82.23	Mean :113.8	Mean :0.1289
3rd Qu.:1.000	3rd Qu.: 32.41	3rd Qu.: 92.90	3rd Qu.:115.0	3rd Qu.:0.0000
Max. :9.000	Max. :130.21	Max. :371.00	Max. :470.0	Max. :1.0000
NA's :1350	NA's :36	NA's :31	NA's :2725	

Table 1: Samples Descriptive Statistics

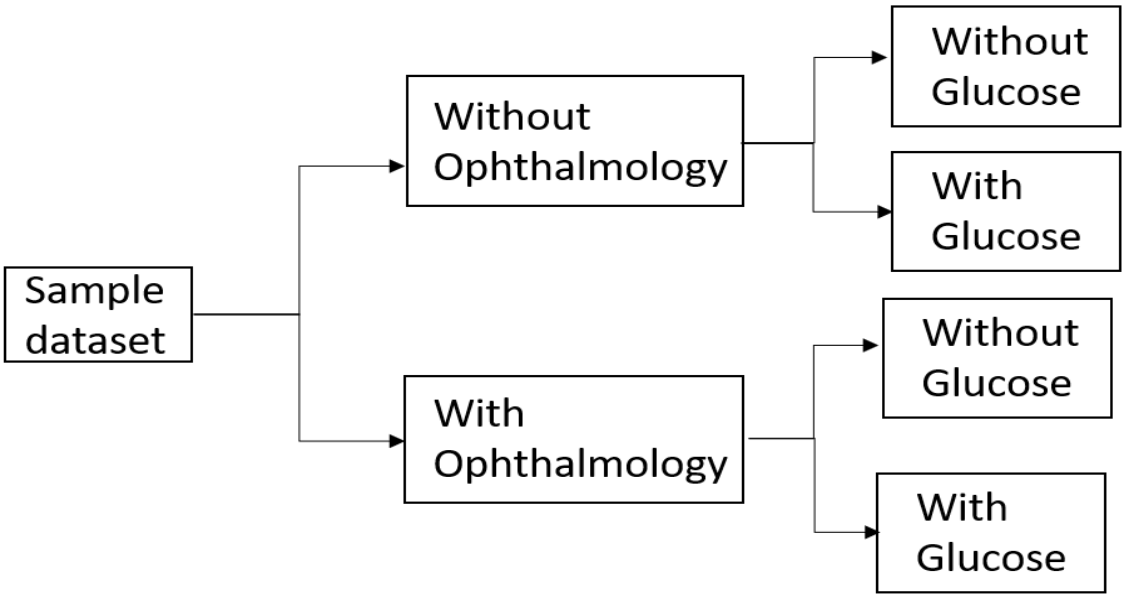
Study Design: The objective of the project is to test if retinopathy can be detected by fasting glucose for people who are over 40. The reason we narrow our study for age of 40 is because that 95 percent people who took this examination is at least 40 years old.

The study design is basically testing the prediction performance based on dataset with and without fasting glucose variables. However, we also make assumptions that in different situations which involve whether the individuals took Ophthalmology examination, the effect of glucose variables helping detecting retinopathy may fluctuate. Thus, we split in to two big parts and compared the effect of glucose under each circumstance.

Without Ophthalmology related variables: This analysis is intended for the situation that we assume that when we have normal knowledge about this person with basically demographic and basic body measure data. We want to see whether by adding fasting glucose variable, it will help to predict the present of retinopathy and by what extent.

With Ophthalmology related variables: This analysis is intended for the situation that we assume that this person already took a further examination of ophthalmology, and we have some more information about the results. We want to see whether by adding fasting glucose variable, it will help to predict the present of retinopathy and by what extent.

To help understanding the former description, please see the graph of structure of study design.



Graph 1 Structure of Study Design

3.2 Data Extraction

We are using National Health and Nutrition Examination Survey (NHANES) for the analysis. An R package 'NHANES' enabled us to download all the data we need. There are examination, demographics, dietary, laboratory, questionnaire, and limited access datasets.

Raw data is extracted from 10 different tables in 5 categories in 2 survey year (2005-2006,2007-2008) including demographic, Ophthalmology, vision, body measure, and glucose.

The assumption that we made:

a) The records for person who may get survey in both different survey year counted as two independent individuals.

b) A sample's either eye got retinopathy count as having retinopathy.

Based on that, after we join all the tables together by their unique number SEQN, we created a new column called RETI to show if a person has retinopathy by check if a persons right or left eye presented retinopathy and dropped the original two columns for left and right eye retinopathy test. We also dropped SEQN since it will not make any sense going further. We also dropped the samples that are smaller than 40 years old to fit with our objective.

Missing Data:

1) For cases missing retinopathy (Y) variables, we directly dropped these samples.

2) In original raw data, we also included some dietary variables, but later we found that in 2005-2006 year, it doesnt have the data that we need, so we decided just drops these columns.

3) For other missing data, since it is MCAR, we decided to use multiple imputation. We first divided the sample data randomly into train data (70 percent) and test data (30 percent), and we later imputed them separately. The training data and test data that we used in final are pooled from imputed data which contain all 5 imputed data that we set.

3.3 Feature Choices

To generate a dependent variable (Y), two variables indicating presence or absence of retinopathy for right and left eyes were chosen in ophthalmology dataset. By combining the binary results of the variables, the dependent variable is coded '1' for either right or left eye has retinopathy, and '0' otherwise.

The variable of our interest, fasting glucose (U), was used as it was.

In different circumstance, there are 3 covariates(V) related with Ophthalmology will only be used in the situation we want to test the effect of fasting glucose when we have more information on the people's Ophthalmology. And they are OPDUHEM(Retinal Blot hemorrhage), OPDUHE(Retinal hard exudate), OPDUSE(Retinal soft exudate)

Some covariates (V) were also considered in both of the analysis in this study. Gender, age, race, number of people in the family and in household, Body Mass Index (BMI)(kg/m*2), and weight (kg) were included in the model. Also, four variables of basic vision relevant data were selected, which indicated whether a participant had surgery for cataract before, whether a participant was wearing glasses, and visual acuity of left eyes and right eyes.

As 95 percent of our dependent variable was for the population over 40 years old, this study looked at population over 40 years old (W).

3.4 Comparison Methods

There is a study used NHANES dataset for discriminating prevalence of retinopathy utlizing fasting glucose, A1C, and retinal images (Cheng et al., 2009). Each of fasting glucose and A1C effectively detected prevalence of retinopathy and A1C turned out to be better than fasting glucose. While A1C has diagnostic cutoff, fasting glucose does not have clear diagnostic cutoff (Sabanayagam et al., 2009). This implies that fasting glucose needs to be further studied to see if it has significant impact on retinopathy detection. Thus, this study discovers the value of fasting glucose in retinopathy detection, especially when ophthalmology data is or is not available.

3.5 Evaluation Criteria

There are three evaluation criteria that we used to judge the effect of fasting glucose variables to help detect retinopathy.

a. P-value: this value shows the significant relationship between fasting glucose and retinopathy, but this just give us a sense of its importance.

b. AUROC curve: this used to show in general, how our model is performed at distinguishing retinopathy and non-retinopathy. We got results mainly from comparing the AUC value under different circumstance to see if adding fasting glucose will bring improvement to the model.

c. Confusion matrix: since we know that our dataset is imbalance in number of retinopathy patients and non-retinopathy, accuracy may not make much sense to show the result. Thus, we used confusion matrix to dig a little deeper to show how model get improved after adding fasting glucose by calculating the specificity and NPV from confusion matrix.

4. Model: Logistic Regression

The model that we used is logistic regression. We use it to do classification for our model and get comparison results based on this model.

5. Results

From two analysis, we find that when there is no Ophthalmology information, the fasting glucose can have a significant positive effect on detection of retinopathy. However, for a patient who has information about their Ophthalmology examination, the fasting glucose does not have obvious improvement on the results.

5.1 Results on analysis that without Ophthalmology related variables

a) The P-value for fasting glucose is smaller than 0.001 so that we can see it will significantly influence the Retinopathy.

b) AUROC: From the table 2 we can see that in first situation where we do not have much information about a person. The AUC is 0.594 without fasting glucose variable but after adding glucose variable, the AUC increases to 0.646 which improves 5.2 percent.

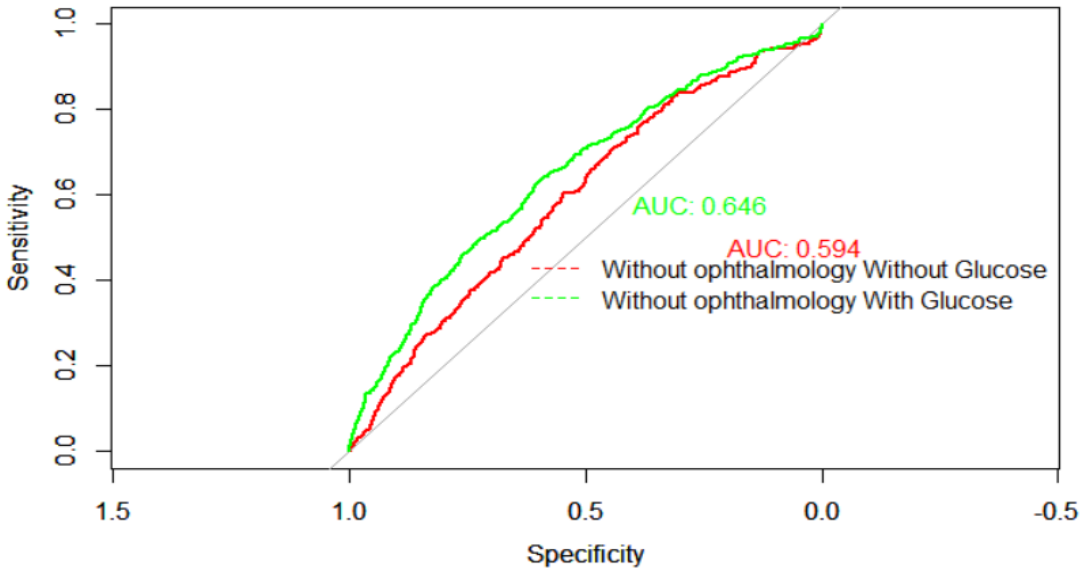


Table 2 ROC curve for situation 1 without ophthalmology variables

c) Confusion Matrix

The following table shows the confusion matrix (Upper is without glucose)

Specificity: The ability to detect retinopathy improving 40 times than without glucose.

NPV: The rate that correctly predict retinopathy is 4.3 times higher than without glucose.

glm.pred1	0	1
0	6768	949
1	12	1

glm.pred2	0	1
0	6723	911
1	57	39

Table 3 Confusion Matrix for situation 1 without ophthalmology variables

5.2 Results on analysis that with Ophthalmology related variables

a) The P-value for fasting glucose is smaller than 0.001 so that we can see it will significantly influence the Retinopathy.

b) AUROC: From the table 4 we can see that in first situation where we do not have much information about a person. The AUC is 0.770 without fasting glucose variable but after adding glucose variable, the AUC increases to 0.786 which improves 1.6 percent. The improvement is slightly.

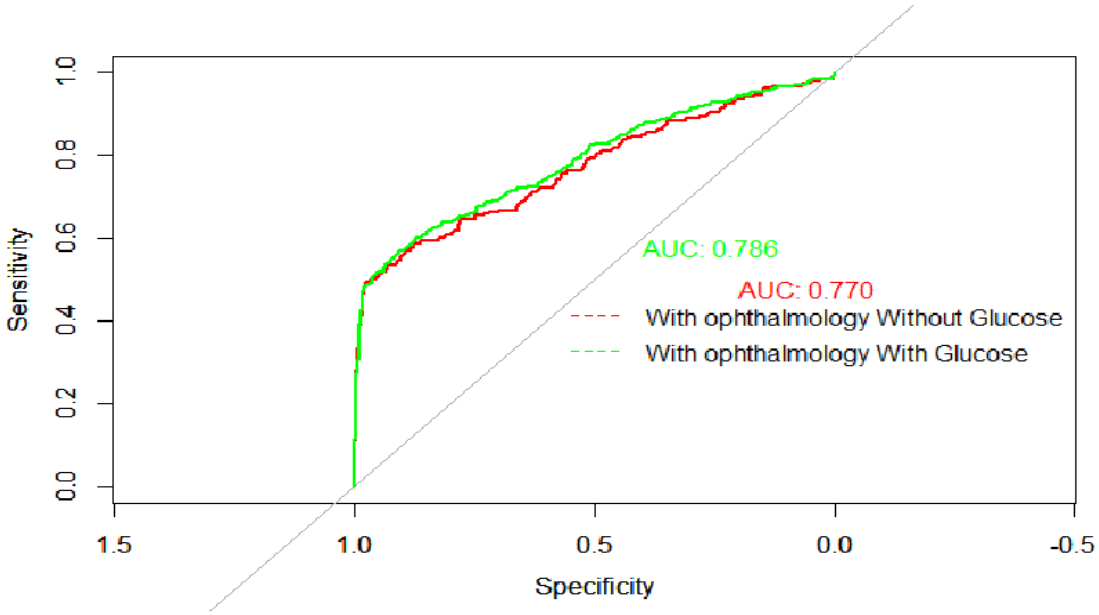


Table 4 ROC curve for situation 2 with ophthalmology variables

c) Confusion Matrix

The following table shows the confusion matrix (Upper is without glucose)

Specificity: The ability to detect retinopathy improving 1 percent than without glucose.

NPV: The rate that correctly predict retinopathy basically the same as without glucose.

glm.pred3	0	1
0	6664	527
1	116	423

glm.pred4	0	1
0	6662	520
1	118	430

Table 5 Confusion Matrix for situation 2 with ophthalmology variables

6. Discussion and Related Work

Significant influence of fasting plasma glucose on detecting retinopathy has been adressed in existing studies (Patel et al., 2017; Takao et al., 2018). This study further tested whether fasting glucose helps detect retinopathy with and without ophthalmology data. The results from this study suggest hospital to measure fasting glucose first rather than collect multiple ophthalmology data to detect retinophthy because fasting glucose raises correct retinopathy examination. The assumption of this suggestion is that cost, time, and labor are lower for measuring fasting glucose than collecting multiple ophthalmology data. After the exact cost is calculated, hospital may want to use fasting glucose as a cost-effective way of detecting retinopathy.

7. Conclusion

This study investigated whether fasting glucose has a positive impact on detecting retinopathy. The results presented in this study showed that knowing fasting glucose enhances model prediction on detecting retinopathy, where ophthalmology data is not available. This implies that fasting glucose can be a substitute of ophthalmology data, which may enable hospital to detect retinopathy in a cost-effective way.

References

National Eye Institue. Internet: https://nei.nih.gov/health/diabetic/retinopathy , December 12, 2018.

Lesioning Studies. (n.d.). Retrieved from "https://www.alleydog.com/glossary/definition.php?term=Lesion Studies"

Cheng, Y. J., Gregg, E. W., Geiss, L. S., Imperatore, G., Williams, D. E., Zhang, X., ... Saaddine, J. B. (2009). Association of A1C and fasting plasma glucose levels with diabetic retinopathy prevalence in the US population: implications for diabetes diagnostic thresholds. Diabetes care, 32(11), 2027-2032.

Patel, Y. R., Kirkman, M. S., Considine, R. V., Hannon, T. S., Mather, K. J. (2017). Retinopathy predicts progression of fasting plasma glucose: An Early Diabetes Intervention Program (EDIP) analysis. Journal of diabetes and its complications, 31(3), 605-610.

Sabanayagam, C., Liew, G., Tai, E. S., Shankar, A., Lim, S. C., Subramaniam, T., Wong, T. Y. (2009). Relationship between glycated haemoglobin and microvascular com-

plications: is there a natural cut-off point for the diagnosis of diabetes?. Diabetologia, 52(7), 1279.

Takao, T., Inoue, K., Suka, M., Yanagisawa, H., Iwamoto, Y. (2018). Optimal cutoff values of fasting plasma glucose (FPG) variability for detecting retinopathy and the threshold of FPG levels for predicting the risk of retinopathy in type 2 diabetes: A longitudinal study over 27 years. Diabetes research and clinical practice, 140, 228-235.

Wong, T. Y., Liew, G., Tapp, R. J., Schmidt, M. I., Wang, J. J., Mitchell, P., ... Shaw, J. (2008). Relation between fasting glucose and retinopathy for diagnosis of diabetes: three population-based cross-sectional studies. The Lancet, 371(9614), 736-743.

Zoppini, G., Verlato, G., Targher, G., Casati, S., Gusson, E., Biasi, V., ... Muggeo, M. (2009). Is fasting glucose variability a risk factor for retinopathy in people with type 2 diabetes?. Nutrition, Metabolism and Cardiovascular Diseases, 19(5), 334-339.