International Journal of Applied Mathematics

Volume 35 No. 5 2022, 735-744
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)
doi: http://dx.doi.org/10.12732/ijam.v35i5.7

THE PREDICTIVE ROLE OF GENETIC MARKERS IN THE DIFFERENT TYPES DIABETES – RELATED COMPLICATIONS

Sylvia V. Pashkunova

Military Medical Academy
3 St. Georgi Sofiiski str.
Sofia - 1606, BULGARIA

Abstract: We apply a Hardy-Weinberg mathematical equilibrium model for a prospective study of the comparison of genetic markers Calpain, Leptin, MnSod, eNos, eNosGly, LPL and AC in 100 healthy persons (control group) and 200 diabetics patients classified into five groups. The association of the genetic markers and the different type of diabetes-related complications are studied.

The associations between genetic markers and micro- and macrovascular complications were investigated for the first time for the population of diabetics in Bulgaria. The software package used for statistical modeling of real data is STATISTICA 13.

AMS Subject Classification: 62P10
Key Words: diabetes-related complications; genetic markers; mathematical equilibrium model; statistical modeling of real data

1. Introduction

Diabetes mellitus comprises a heterogeneous group of diseases which have chronic hyperglycaemia in common as well as the resulting microvascular, macrovas-

Received: September 10, 2022

cular and neurological complications of this condition. Familial studies have provided strong evidence for the existence of genetic determinants in the different types of diabetes. In particular, monozygotic twin studies have indicated a higher rate of concordance in non-insulin-dependent (NIDDM) than in insulin-dependent diabetes mellitus (IDDM). Precise analysis of phenotypes in the remaining families or systematic screening of the genome could allow the genes of each subtype to be identified. Maturity-onset diabetes of the young (MODY) is a genetically and clinically heterogeneous subtype of non-insulin-dependent diabetes mellitus (NIDDM) characterised by early onset, autosomal dominant inheritance and a primary defect in insulin secretion. A majority of Chinese MODY patients for example are due to defects in unknown genes and appear to be characterized by insulin resistance.

A survey of a four diabetics related complications in patients of the Department of Endocrinology of Military Medical Academy, Sofia is presented. Genetic markers Calpain, Leptin, MnSod, eNos, eNosGly, LPL and AC are considered to be prognostic factors of these complications. The patients are classified into five groups subject to the type of the diabetes. A Hardy-Weinberg equilibrium mathematical model is performed in order to predict how gene frequencies will be inherited from generation to generation given a specific set of assumptions. The software package used for statistical modeling of real data was STATISTICA 13.

2. Measurements

During the study period 94 healthy subjects (control group) and 206 patients (male and female) have been prospectively studied. The patients were divided into 5 groups depending on the mode of inheritance, the amount of insulin secretion, treatment and course of illness:

- Insulin-dependent type 1 diabetes (juvenile diabetes) – 43 patients;
- Insulin- not dependent type 2 diabetes mellitus (DM Age) – 60 patients;
- Insulin-dependent type 2 diabetes mellitus – 42 patients;
- Diabetes type MODY – 29 patients;
- Diabetes mellitus type LADA – 32 patients.

The data obtained for each patient included 12 clinical parameters as: Age (in years), Sex, Glucosed Hemoglobin (HbAlc%), C-peptide, Creatinin,

C-reactive Protein (CRP), Cholesterol, TRG, HDL, LDL, Duration (in years). The complications which are observed are retinopathy, neuropathy, nephropathy and macro vascular.

The primary data analysis is as follows: minimum age is 19 and maximum 81 years, 35% of patients were male and 65% women, 23% was observed retinopathy, minimum of duration is 1 year and maximum is 40 years, 75% had neuropathy, in 21% nephropathy is observed and 18% was observed with macro vascular complications. The sum of percentages is more than 100 because there are patients with 2, 3 or 4 complications.

In the study we obtain data for genotypes of 7 genetic markers – Calpain, Leptin, MnSod, eNos, eNosGly, LPL and AC in the patients and in the control group.

An individual's genotype is the combination of alleles found in that individual at a given genetic locus [9]. Allele is a variant of the DNA level. In biological systems, each parent transmits one allele with germ cells. So each person has two alleles from their parents.

If there are two alleles in a population at locus A (denoted by $\boldsymbol{A}$ and $\boldsymbol{a}$), then the possible genotypes in that population are $\boldsymbol{AA}$, $\boldsymbol{Aa}$, and $\boldsymbol{aa}$. Individuals with genotypes $\boldsymbol{AA}$ and aa are homozygous (i.e., they have two copies of the same allele). Individuals with genotype $\boldsymbol{Aa}$ are heterozygous (i.e., they have two different alleles at the A locus). If the heterozygous is phenotypically identical to one of the homozygotes, the allele found in that homozygote is said to be dominant, and the allele found in the other homozygote is recessive.

Allele demonstrates the power of gene mutation. If the allele is recessive, it will be associated with a phenotype - for example have or not a neurological complication only when is in homozygous state – $\boldsymbol{aa}$ or $\boldsymbol{AA}$. So in these phenotypic characteristics, which are given the absence or presence of symptoms – nephrology, ophthalmology, etc., they may be associated with dominant or recessive mark.

In the nomenclature there is several signs for denoting the different allels: "+" and "-" , $\boldsymbol{T}$ and $\boldsymbol{L}$, $\boldsymbol{T}$ and $\boldsymbol{G}$, $\boldsymbol{G}$ and $\boldsymbol{A}$, etc. The comparison of the genotypes distribution of considered gene markers in our control group and in a healthy subjects from other countries shows that the distribution is similar to other European populations.

3. Statistical methods and models

3.1. Hardy-Weinberg equilibrium model

Mathematical models have great potentialities as regards their utility in different disciplines of medicine and health [16, 17]. Mathematical concept, called the Hardy-Weinberg equilibrium model or principle (HWP), is a crucial concept in population genetics, [10, 11]. It predicts how gene frequencies will be inherited from generation to generation given a specific set of assumptions.

The definition of HWP is the following [10, 11]: The stable frequency distribution of genotypes, $\boldsymbol{AA}$, $\boldsymbol{Aa}$, and $\boldsymbol{aa}$, in the proportions p^2, $2pq$, and q^2 respectively (where p and q are the frequencies of the alleles, $\boldsymbol{A}$ and $\boldsymbol{a}$, $p+q=1$) that is a consequence of random mating in the absence of mutation, migration, natural selection, or random drift.

This principle is important because it gives biologists a standard for measuring changes in allele frequency in a population. When a population meets all of the Hardy-Weinberg conditions, it is said to be in Hardy-Weinberg equilibrium (HWE). Human populations do not meet all of the conditions of HWE exactly, and their allele frequencies will change from one generation to the next and the population will evolve. How far a population deviates from HWE can be measured using the χ^2 "goodness of fit" statistical test.

To test null hypothesis H_0: The frequencies of the genotypes in control group and in the group of patients with some type of diabetes are in Hardy-Weinberg equilibrium we calculate expected frequencies of the genotypes $E_i\,(i=1,2,3)$ using observed frequencies $Q_i\,(i=1,2,3)$.

Table 1: The genotypes and their observed and expected frequencies

Genotype	Expected frequencies	Observed frequencies
Common homozygote ($\boldsymbol{AA}$)	E_1	Q_1
Heterozygote ($\boldsymbol{Aa}$)	E_1	Q_1
Rare homozygote ($\boldsymbol{aa}$)	E_1	Q_3

First we calculate the expected values of p and q:

$$p=\frac{Q_1}{n}+\frac{1}{2}\frac{Q_2}{n},\quad q=1-p, \tag{1}$$

where $n=\sum_{i=1}^{3}Q_i$. Then, the expected genotype frequencies are calculated:

$E_1 = p^2$, $E_2 = 2pq$, $E_3 = q^2$. Now using the statistic chi^2,

$$\chi^2 = \sum_{i=1}^{3} \frac{(Q_i - E)^2}{E_i} \tag{2}$$

with 2 degrees of freedoms, we test the null hypothesis at the 5%-level.

3.2. Fisher's exact test for association of alleles frequencies

To test the null hypothesis H_0: The frequencies of allels of given marker in control group and in the group of patients with some type of diabetes are equally likely we consider 2×2 contingency table:

Table 2: 2×2 contingency table for Fisher's exact test

	Allel $\boldsymbol{A}$	Allel $\boldsymbol{a}$	Total
Frequencies in the control group	r	b	$r+b$
Frequencies in group LADA	c	d	$c+d$

As pointed out in [12, 13], this leads under a null hypothesis of independence to a hypergeometric distribution of the numbers in the cells of Table 2. To determine whether or not there is an association, the P-value is calculated by the formula

$$P = \frac{(r+d)!(c+d)!(r+c)!(b+d)!}{r!\,b!\,c!\,d!\,n!}, \tag{3}$$

where $r + cb + dr + b + c + d = n$.

In order to calculate the significance of the observed data, i.e. the total probability of observing data as extreme or more extreme if the null hypothesis is true, statistical package STATISTICA 13 [8] computes the P-value by summing the probabilities for all tables with probabilities less than or equal to that of the observed table.

4. Application to the real data. Results and conclusions

To test the null hypothesis, that frequencies of the genotypes in control group and in all other groups of diabetics are in Hardy-Weinberg equilibrium, we use the package WinStat 4.3, [14]. In Table 3 the result of the test for the marker eNOS in the control group is given. The conclusion is that the differences between expected and observed frequencies is not statistically significant at the

level of significance $\alpha = 0.05$, so the genotypes of marker eNOS in control sample meets the Hardy-Weinberg equilibrium.

Table 3: Hardy-Weinberg equilibrium of genotypes for marker eNOS in the control group

Genotype eNOS Control group	Expected	Observed
Common homozygote ($\boldsymbol{TT}$)	60	61
Heterozygote ($\boldsymbol{TL}$)	30.99	29
Rare homozygote ($\boldsymbol{LL}$)	4	5
p - allele $\boldsymbol{T}$ frequency	0.79	
P-level of (3)	0.39	

The same tests for HWE about all other 7 markers are made for the control group and for the other five groups of diabetics. For all samples we derive to be in Hardy-Weinberg equilibrium. Fisher's exact test for association between the control group and the each of the other five groups of diabetics (for all 7 markers) have been carried out. In Table 4 the result of Fisher's exact test for Insulin-dependent type 1 diabetes is showed.

Table 4: Fisher's exact test for marker eNOS in type 1 DM

	Column 1 Alel $\boldsymbol{T}$	Column 2 Alel $\boldsymbol{L}$	Row - Totals
Frequencies Control Group	151	39	190
Frequencies Type 1	65	21	86
Column totals	216	60	276
Fisher exact P, one-tailed		p= 0.282	
Fisher exact P, two-tailed		p= 0.529	

The two-tailed P value equals 0.529 and the conclusion is that the association between rows (groups) and columns (outcomes) is not statistically significant at 5%-level for marker eNOS in Type 1 diabetes.

The result of Fisher's exact test is that the association is statistically significant for the gene marker MnSod in patients of dibetes type LADA:

Our result for MnSod in LADA DM is similar as in [2]. The results of all other Fisher's exact tests have shown that the association is not statistically significant at 5%-level as it is in [4, 6, 7]. Our results for eNOS are different from (see [1] and [5]), where the association is discovered to be significant. Similar results are published in [1, 2, 4, 18]. The main part of study is to investigate the association of the genetic markers with the different types diabetes-related complications. For this purpose we make two types of code clustering:

Table 5: Fisher's exact test for marker MnSod in LADA DM

	Allele (+)	Allele (-)	Total two-tailed	P-value,
LADA	34 (60.7%)	22 (39.3%)	56	$P = 0.042$
Control group	84 (44.7%)	104 (55.3%)	188	

- genotypes ***AA*** and ***Aa*** (code1) versus ***aa*** (code2);
- genotypes ***aa*** and ***Aa*** (code1) versus ***AA*** (code2),

and than the respective dichotomous response models are built.

It turns out that if we find differences in these two groups, we will know if allele ***A*** or allele ***a*** is dominant.

The association between pathogenesis and gene genotypes of all types of diabetes are analysed using logistic regression.

Table 6: Models of inheritance ***AA*** and ***Aa*** versus ***aa*** for the group of type 2 DM for diabetes-related complications

Diabetics type 2: Genotypes ***AA*** and ***Aa*** (code1) versus ***aa*** (code2):						
Complic-ation	Gene marker	$\hat{\beta}_0$	$\hat{\beta}_1$	p-level (Wald's χ^2)	Probab. P(code1)	Probab. P(code2)
NEPHRO	Calpain	-5.128	2.014	0.041	0.04	0.25
NEPHRO	LPL	-2.793	2.1	0.04	0.33	0.8
NEURO	MnSod	-1.38	1.098	0.044	0.2	0.43
MACRO	Calpain	-4.678	1.992	0.019	0.06	0.33
MACRO	eNOS	-4.644	2.322	0.035	0.09	0.49
RETINO	eNosGly	-4.388	1.471	0.011	0.05	0.19
RETINO	AC	-4.685	2.14	0.041	0.07	0.4

The results, when we assumed a recessive model of inheritance (i.e., ***AA*** and ***Aa*** versus ***aa***) for the group of diabetes type 2, are given in Table 7.

Lack of an increased risk of complication posed by the ***A*** allele is not dominantly expressed and that the increased risk is confined to ***aa*** homozygotes for:

- Calpain of neuropathy and macro vascular complications;
- LPL of nephropathy;
- MnSod of neuropathy;

Table 7: Models of inheritance ***aa*** and ***Aa*** versus ***AA*** for the group of type 2 DM for diabetes-related complications

Diabetics type 2: Genotypes ***AA*** and ***Aa*** (code1) versus ***aa*** (code2):						
Complic-ation	Gene marker	$\hat{\beta}_0$	$\hat{\beta}_1$	p-level (Wald's χ^2)	Probab. P(code1)	Probab. P(code2)
NEPHRO	AC	0.209	-1.868	0.014	0.16	0.03
RETINO	eNOS	-4.738	1.742	0.047	0.05	0.22

- eNosGly of retinopathy;
- AC of retinopaty and nephropathy.

By contrast, the increased risk of retinopathy (Table 7) is confined to eNOS ***AA*** homozygotes for code cluster "***aa*** and ***Aa*** versus ***AA***". Our results imply that homozygosity for the eNos may be involved in predisposition to retinopathy and macro vascular complications.

The results of association between pathogenesis and gene genotypes in other groups of DM patients are published in [19] and are similar to those in [1]-[7].

To our knowledge, this was the first study in Bulgaria to implicate polymorphisms of Calpain, Leptin, MnSod, eNOS, eNosGly, LPL and AC as a genetic risk factor for retinopathy, neuropathy, nephropathy and macro vascular complications. In conclusion, we believe that our findings could contribute to any future analyses validating these relationship.

References

[1] E. Noiri, H. Satoh, J. Taguchi, S.V. Brodsky, A. Nakao, Y. Ogawa, S. Nishijima, T. Yokomizo, K. Tokunaga, T. Fujita, Association of eNOS Glu298Asp Polymorphism with End-Stage Renal Disease, *Hypertension*, **40** (2002), 535–540.

[2] M. Tarek, E. Masry, A.M. Mona, Z. Abou, M. Mohamed, E. Tawil, A. Rawhia, A. Khalifa, Manganese Superoxide Dismutase Alanine to Valine Polymorphism and risk of neuropathy and nephropathy in Egyptian Type 1 Diabetic patients, *Review of Diabetic Studies*, **2** (2005), 70–74.

[3] F. Thameem, S. Puppala, N.H. Arar, M.P. Stern, J. Blangero, R. Duggirala, H.E. Abboud, Endothelian nitric oxide synthase gene polymorphisms

and their association with type 2 diabetes-related traits in Mexican Americans, *Diabetes and Vascular Disease Research*, **5** (2008), 109–113.

[4] A. Zanchi, D.K. Moczulski, L.S. Hanna, M. Wantman, J.H. Warram, A.S. Krolewski, Risk of advanced diabetic nephropathy in type 1 diabetes is associated with endothelial nitric oxide synthase (eNOS) gene polymorphism, *Kidney International*, **57** (2000), 405–413.

[5] K. Hibi, T. Ishigami, K. Tamura, S. Mizushima, K. Kimura, M. Ishii, S. Umemura, T. Fujita, H. Ochiai, M. Kosuge, Y. Watanabe, Infarction Endothelial Nitric Oxide Synthase (eNOS) Gene Polymorphism and Acute Myocardial, *Hypertension*, **32** (1998), 521–526.

[6] L. Haddad, J.C. Evans, N. Gharani, C. Robertson, K. Rush, S. Wiltshire, T. Frayling, T. Wilkin, A. Demaine, A. Millward, A. Hattersley, G. Conway, N.Cox, G. Bell, S. Franks, M. Mccarthy, Variation within the Type 2 Diabetes Susceptibility Gene Calpain10 and Polycystic Ovary Syndrome, *The Journal of Clinical Endocrinology & Metabolism*, **87**, No 6 (2002), 2606–2610.

[7] J. Vaškù, A. Vaškù, Z. Dostálová, P. Bienert, Association of Leptin genetic polymorphism with gestational diabetes mellitus, *Genes & Nutrition*, **1**, No 2 (2006), 117–124.

[8] StatSoft, Inc., *STATISTICA Manual (Data Analysis Software System)*, Version 13.0 (2019).

[9] F. Griffiths, J.F. Anthony, *An Introduction to Genetic Analysis*. 6th Ed., W. H. Freeman, New York (1998).

[10] G.H. Hardy, Mendelian proportions in a mixed population, Source: *Science, New Series*, **28**, No 706 (1908), 49–50.

[11] W. Weinberg, Über den Nachweis der Vererbung beim Menschen, *Jahreshefte des Vereins für vaterländische Naturkunde in Württemberg*, **64** (1908), 368–382.

[12] R.A. Fisher, A new test for 2×2 tables, *Nature*, **156** (1945), 783–784.

[13] R.M. Cyrus, P. Senchaudhuri, Conditional versus Unconditional Exact Tests for Comparing Two Binomials (2003), Retrieved 20 November 2009 from http://www.cytel.com/Papers/twobinomials.pdf.

[14] H.M. Stauss, *Doing Statistics with WinStat*, Univ. of Iowa (2008), 1–75.

[15] J.D. Jobson, *Applied Multivariate Data Analysis* Vol. II, Springer Verlag (1991), 278–303.

[16] Y. Uzunova, K. Prodanova, Comparison of binary regression models for the outcome of pediatric liver transplantation, *International Journal of Applied Mathematics*, **35**, No 4 (2022), 601–609; doi:10.12732/ijam.v35i4.8.

[17] L. Lazarova, N. Stoikovikj, A. Stojanova, M. Mileva, M. Ljubenovska, Mathematical model for prediction of Covid-19 dynamics, *International Journal of Applied Mathematics*, **35**, No 1 (2022), 119–133; doi:10.12732/ijamv35i1.9.

[18] I. Larson, M.M. Hoffmann, J.M. Ordovas, E.J. Schaefer, W. Marz, J. Kreuzer, The lipoprotein lipase HindIII polymorphism: association with total cholesterol and LDL-choleterol, but not with HDL and triglycerides in 342 females, *Clin. Chem.*, **45** (1999), 963–968.

[19] S. Pashkunova, *Associative Investigations of Genetic Markers Determing Type of Dibetes and Chronic Diabetes-Related Complications*, PhD Thesis, Medical Academy of Bulgarian Army, Sofia, 2012, pp. 1–35 [in Bulgarian].