

# The importance of the chosen statistical methods in medical research: study over modelling in estimation superficial varicose vein risk factors in young male population

*Tıbbi arařtırmalarda seçilen istatistiksel yöntemlerin önemi: genç erkek popülasyonda yüzeysel variköz ven risk faktörlerinin kestirimi üzerine modelleme çalışması*

 Feray Aydın<sup>1</sup>,  Dilek Dülger<sup>2</sup>,  Özgür Albuz<sup>3</sup>

<sup>1</sup>Ankara 29 Mayıs State Hospital, Department of General Surgery, Ankara, Turkey

<sup>2</sup>Karabük University, School of Medicine, Department of Microbiology, Karabük, Turkey

<sup>3</sup>Keçiören Training and Research Hospital, Department of General Surgery, Ankara, Turkey

Cite this article as / Bu makaleye atf için: Aydın F, Dülger D, Albuz Ö. The importance of the chosen statistical methods in medical research: Study over modelling in estimation superficial varicose vein risk factors in young male population. J Health Sci Med 2020; 3(3): 231-239.

## ABSTRACT

**Objective:** To clarify the under which conditions univariate analysis is insufficient and the importance of multiple logistic regression in medical research is the primary objective. For this purpose, lower extremity superficial venous risk factors were selected as a model in a young male population retrospectively planned.

**Material and Method:** All patients who presented to the General Surgery Clinic of 29 Mayıs State Hospital with symptoms of pruritus, edema, leg pain and swelling between May 30, 2016 and January 1, 2019 were retrospectively analyzed for this study.

**Results:** The results of univariate analysis showed that the development of varicose veins was significantly higher in terms of variables than in the control group. Family history [34 (61.8%) vs. 21 (38.2%)], p<0.008, smoking [37 (59.7%) and 25 (40.3%), p<0.04], long-term standing, respectively [90 (53.6%) vs. 78 (46.4%), p<0.01], constipation [64 (55.2%) vs. 52 (44.8%), p<0.001] and the urban lifestyle [49 (56.7%) - 22 (43.3%), p<0.04] were found to be effective on the development of varicose veins. However, when we do binary logistic regression for important univariate results; We found different results in terms of risk factors of varicose veins. The results showed that family history, smoking, long standing and constipation contributed to the development of varicose veins [(61.8% vs. 38.2%, AOR: 2.62 (1.35, 5.07), p (0.04)], [(59.7% vs. 40.3%, AOR: 2.08 (1.13, 3.8) p<0.02)], [(53.6% vs. 46.4%, AOR: 1.83 (1.06, 3.16) p<0.03)], [(55.2% and 44.8%) AOR: 1.74 (1.03, 2.92) p<0.04]. On the other hand, urban lifestyle was found to be insignificant [(56.7% and 43.3%, AOR: 1.73 (0.96, 3.15) p<0.07)\*].

**Conclusion:** In cases where there is more than one independent variable affecting the dependent variable; In determining how and in what way the dependent variable is affected by independent variables; multiple logistic regression tests are more successful than univariate analysis. False or incomplete statistical studies selected in medical research may lead to very important false clinical interpretations.

**Keywords:** Statistical and mathematical methods, medical research, modelling in estimation

## ÖZ

**Amaç:** Hangi koşullar altında tek değişkenli analizin yetersiz olduğunu ve tıbbi arařtırmalarda multipl lojistik regresyonun önemini açıklığa kavuşturmak birincil amaçtır. Bu amaçla retrospektif olarak planlanan genç erkek popülasyonda alt ekstremitte yüzeysel venöz risk faktörleri model olarak seçildi.

**Gereç ve Yöntem:** Bu çalışma için 30 Mayıs 2016-1 Ocak 2019 tarihleri arasında 29 Mayıs Devlet Hastanesi Genel Cerrahi Kliniği'ne kaşıntı, ödem, bacak ağrısı ve şişlik gibi semptomları olan tüm hastalar retrospektif olarak incelendi.

**Bulgular:** Tek değişkenli analiz sonuçları, variköz venlerin gelişiminin, değişkenler açısından kontrol grubundan anlamlı derecede yüksek olduğunu göstermiştir. Sırasıyla, aile öyküsü [34 (%61,8) vs. 21 (%38,2)], p<0,008, sigara içme [37 (%59,7) ve 25 (%40,3)], p<0,04], uzun süreli [90 (%53,6) 78'e karşı (%46,4), p<0,01], kabızlık [64 (%55,2) - 52 (%44,8)], p<0,001] ve kentsel yaşam tarzı [49 (%56,7) - 22 (%43,3)], p<0,04], variköz venlerinin gelişimi üzerine etkili bulunmuştur. Oysa, önemli tek değişkenli sonuçlar için ikili lojistik regresyonunu yaptığımızda; Variköz venlerin risk faktörleri açısından farklı sonuçlar saptadık; Sonuçlar aile öyküsünün, sigara içmenin, uzun süre ayakta durmanın ve kabızlığın variköz venlerin gelişiminde katkıda bulunduğunu ortaya koydu [(%61,8'e karşı %38,2, AOR: 2,62 (1,35 ve 5,07), p<0,04)], [(%59,7'ye karşı %40,3, AOR: 2,08 (1,13 ve 3,8) p<0,02)], [(%53,6'a karşı %46,4, AOR: 1,83 (1,06 ve 3,16) p<0,03)], [(%55,2 ve %44,8) AOR: 1,74 (1,03 ve 2,92) p<0,04]]. Öte yandan, kentsel yaşam tarzı ise önemsiz bulundu [(%56,7 ve %43,3, AOR: 1,73 (0,96 ve 3,15) p<0,07)\*].

**Sonuç:** Bağımlı değişkeni etkileyen birden fazla bağımsız değişken olduğu durumlarda; bağımlı değişkenin bağımsız değişkenlerden nasıl ve ne şekilde etkilendiğinin tespitinde; çoklu lojistik regresyon testleri, tek değişkenli analizden daha başarılıdır. Tıbbi arařtırmalarda seçilen yanlış veya eksik istatistiksel çalışmalar çok önemli yanlış klinik yorumlara neden olabilir.

**Anahtar Kelimeler:** İstatistiksel ve matematiksel yöntemler, tıbbi araştırma, kestirimde modelleme

**Corresponding Author / Sorumlu Yazar:** Feray Aydın, 29 Mayıs Devlet Hastanesi/Aydınlar, Dikmen Cd. No:312, Çankaya, Ankara, Türkiye

**E-mail / E-posta:** ferayaydin2008@hotmail.com

**Received / Geliş:** 27.02.2020 **Accepted / Kabul:** 31.03.2020



## INTRODUCTION

Statistics can be defined as a science that includes techniques and methods for collecting, classifying, analyzing and interpreting data for a specific purpose (1). German statistician Egel (2) reported 180 different statistical definitions in 1896. Fisher had been underlined that of statistics are a branch of applied mathematics and can be accepted as mathematics applied over observational data (3). The scientific value of a research in the field of health; In addition to being current, useful and important, it is also related to the value of biostatistics. In addition, the statistical errors caused by waste of resources and information pollution are a big problem in terms of ethics.

In order to avoid these errors, research must be correctly applied with the statistical science at all stages including preparation, planning, data collection, evaluation and interpretation(4). When there are various assumptions, logistic regression analysis is used as an alternative to discriminant analysis and cross table applications. The most basic approach as the reason for use can be made in linear regression analysis; When the dependent variable is a discrete variable with binary such as 0 and 1, the normality assumption is disrupted and linear regression analysis cannot be applied. The most distinctive feature which differentiates logistic regression from linear regression is the binary or multiple result variable in logistic regression (5). The results of the statistical method to be selected in the retrospective or prospective research of medical disorders are very important.

This importance has such a situation that, depending on these results, clinical manifestations may be reorganized when necessary. Therefore, if there are different arguments, which may affect a dependent variable, in particular; Multivariate analysis method should be applied without neglecting the effects of all independent variables.

For the varicose veins which are medical subjects of this study; Clinical condition and epidemiological studies have showed that around one quarter of the adult population have visible varicose veins (6). Chronic venous insufficiency (CVI) is a condition associated with multiple complications, such as venous stasis ulcers, which account for a loss of work days, additionally individual discomfort and disability (7). In recently conducted a study; Cannon et al. (8) found that varicose veins in the lower extremities were an important risk factor in the development of leg cellulitis and infections, in 2018. The clinical expression of venous insufficiency was first described with the CEAP (Clinical-Etiological-Anatomical - Pathophysiological) for classification, (9)

established in 1994, which has since been universally accepted. The prevalence of varicose veins reported in studies ranges from 2%-56% in men and from 1%-60% in women (10). As it is understood from the above figures, different results are mentioned in wide ranges for varicose veins of lower extremities. We cannot say that the mathematical and statistical methods chosen for these results have no effect. Because of all these reasons we aimed to clarify that the effect of univariate and multivariate analysis methods over the medical issues. In this study is to determine between the multiple logistic regression and univariate analysis methods success in detecting varicose vein risk factors in the lower extremity of young male population. In this reason as first; Smoking, familial history, constipation, long standing, urban life style effects of variables on the risk of developing varicose veins in the lower extremity of the young male population were investigated with univariate analysis methods. Then, the values which were found to be significant in the univariate analysis methods were examined by binary multiple logistic regression. The aim of this study the results of both statistical methods, definition the importance of statistical methods in the interpretation of medical studies examining and discuss how statistical results can lead to interpretive differences in medical studies.

## MATERIAL AND METHOD

### Model Formulation

There were totally 270 patients PVI, with PVI like complaints, in our model and in this model the cases of whom diagnosed PVI(+) and PVI(-) was compared with the two different statistical method.

This study was approved by the university /local human research ethics committee and all procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The study was carried out with the permission of T.C. Ministry of Health, Etlik Zübeyde Hanım Gynecology Training and Research Hospital (Permission granted /date-TUEK: 04/07/2019, Number of meeting: 11, Decision no: 90057706-799). After the approving we screened 270 retrospective cases from hospital registration of whom applied with symptoms similar the varicose veins. We identified of whom 125 (46.3%) were diagnosed with symptomatic superficial venous insufficiency, in the total population.

Due to the compared the chosen statistical methods in investigation of risk factors for development superficial

varicose veins in the young male population. Only patients diagnosed with PVI, defined as CEAP clinical classes I-III, were included in further analyses. In CEAP classification; C1 corresponds to telangiectasia or reticular veins smaller than 4 mm, while C2 corresponds to veins larger than 4 mm. C3 describes the edematous structure of the lower extremity due to varicose veins. C4 can include venous eczema, ulcers and lipodermatosclerosis, C5 has described a healed skin ulcer whereas C6 has an active skin ulcer. In this reason, the cases with C4, C5, C6 were excluded from the study. Patients with inconsistent diagnosis data, no documentation of primary venous insufficiency (PVI), or missing outcomes were excluded. All patients were young male population. For this study, all patients who attended with symptoms such as itching, edema, leg pain and swelling to the 29 Mayıs State Hospital general surgery clinic department between 30 January 2016 and 1 January 2019 screened data was that reviewed retrospectively. As a result of examination of records; The patients who underwent physical examination and doppler USG on suspicion of varicose veins were retrospectively examined. While the first (patient) group were consisted of the cases that had been detected PVI, the second (control) group were consisted of the cases that hadn't been detected PVI by physical examination or doppler ultrasound. There were not included any cases with thrombosis anamnesis, in this population. The patient group and control group were compared for risk factors such as, smoking, constipation, low fiber diet, long time standing, family history, using of cigarette and living in a urban life style. According to the anamnesis reports in hospital registration; Family history of varicose veins was restricted to first degree relatives, and was evaluated according to the information obtained from the patients. (For patients with primary venous insufficiency in first-degree relatives for one person or the number of up, family history was accepted positive. Were taken in consideration, the six hour or the higher duration, for long -time standing variable. For per day at least half box cigarette was accepted smoking risk factor.

The potential independent risk factors which determined in parallel with the risk factors, previously mentioned in the literature were included to the study, as risk factors that could contribute to the primary venous insufficiency, such as family history, smoking, chronic constipation, urban area and prolonged standing. The factors that significantly contributed to the development of PVI were chosen as the candidate variables for effective over the dependent variable (in terms of PVI development) . Patients with inconsistent diagnosis data, no documentation of varicose veins, or missing outcomes were excluded. In patient demographics, mean and

standart deviation was used in continuous variables, and percentage and frequency was used for dichotomous variabes. To reveal that the importance of mathematics and statistics in medical studies; The study of statistics was planned in two stages. In first stage; Chi-square test was used for binary variables. In second stage; Multipl binary logistic regression was used for significant results in univariate analysis. The clinical significance of the results obtained in both stages was evaluated separately as medically. In the comparison between the PVI positive and PVI negative groups, differences were considered significant at  $p < 0.05$ . In this study null hypotesis is that there are no statistically significant difference cases between who applied to the general surgery clinic for the like varicose veins symtomp complaints, in terms of the risk factors between varicose veins positive patients and negative cases. On the other side the N1 hypotesis is that there are differences statistically significant between cases who applied to the hospital for the like symptos of varicose veins enviromental, personal and behaviors as like smoking, genetically status etc. , for varicose veins positive and negative cases. In line with the extant studies reported in the pertinent literature, the group characterized by PVI and the control group were assessed in terms of risk factors that could contribute to the primary venous insufficiency, such as family history, smoking, urban life style, prolonged standing and constipation. Our primary outcome for this study is the estimation of the main risk factors in terms of the development of PVI.

## RESULTS

Total population had been consisting from 270 young male patients. Their mean age was  $20.5 \pm 1.1$ . In this group, varicose veins were documented in 124 cases (45.9%), and it thus served as a study group, while the remaining 146 (54.1%) cases were asymptomatic and formed the control group. In 45 patients that took part in the study, grade III and grade IV varicose veins were identified by Doppler ultrasound. According to the CEAP classification, 61 (22.4%) cases were CEAP I, 49 (18%) cases were CEAP II, and 14 (5.1%) cases were rated as CEAP III (**Table 1**).

**Table 1.** The descriptive analysis results in terms of the CEAP classification, in PVI group

The distrubition of CEAP classification	PVI(+)
Age\year (Mean±Std. Dv.)	21.6±1.77
CEAP I	61 (22.4%)
CEAP II	49 (18%)
CEAPIII	14 (5.1%)
Total	124
*PVI (Primary venous insufficiency)	

In term’s of univariate analysis results; The between comparison of the control group and varicose vein group results was showed in that Table II; The patients consequently suffering from varicose veins had a significantly higher rate. The results for independent variables such as family history (34 vs. 21, p<0.008), smoking (37 vs. 25, p<0.04), prolonged standing (90 vs. 78, p<0.01), constipation (64 vs. 52, p<0.001) and urban life style (49 vs. 22, p<0.02) were found respectively. (**Table 2**).

**Table 2.** The risk factors with calculated univariate  $\chi^2$  test, for primary venous insufficiency

	PVI(+)	PVI(-)	P<0.05
Age\year (Mean±SD)	21.76±1.73	21.47±1.81	0.187
Family history	34 (61.8%)	21 (38.2%)	0.008
Smoking	37 (59.7%)	25 (40.3%)	0.04
Prolonged standing	90 (53.6%)	78 (46.4%)	0.01
Urban life style	38 (56.7%)	29 (43.3%)	0.04
Constipation	64 (55.2%)	52 (44.8%)	0.001

\*PVI (Primary venous insufficiency)

Based on according the **Table 2** test results, our interpretation; The familial history, long-term survival, urban lifestyle, constipation, and smoking factors had been found to be statistically significant risk factors for varicose vein development. For these results, when we done a deeper analysis with stepwise logistic regression, we encountered a different situation. If we look at our the multipl binary logistic regression for significant univariate results (family history, smoking, prolonged standing, provincial area and constipation); We detected a different results, in terms of varicose veins risk factors; While the results revealed that family history, urban life style, smoking, prolonged standing and constipation were contributing factors in the development of varicose veins, with [(61.8% vs. 38.2%, AOR: 2.62 (1.35 vs. 5.07), p<0.04)], [(59.7% vs. 40.3%, AOR: 2.08 (1.13 vs. 3.8) p<0.02)], [(53.6% vs. 46.4%, AOR: 1.83 (1.06 vs. 3.16) p<0.03)], [(55.2% vs.44.8% AOR: 1.74 (1.03 vs. 2.92) p<0.04)], respectively. Whereas the urban life style was found insignificant[(56.7% vs. 43.3%, AOR: 1.73 (0.96 vs. 3.15) p<0.07)\*] (**Table 3**).

## DISCUSSION

Superficial venous insufficiency has been recognized and treated since the time of the ancient Egyptians (11) The incidence of venous reflux, including those without clinically significant disease in the general population, is closely related to the presence of venous disease. Empirical evidence indicates that reflux in the deep veins is more prevalent in men than in women. Biostatistical methods; It is used both on a community and personal level. At the social level, monitoring and evaluation of health conditions, determining the effectiveness of the programs implemented for any protection purposes, as well as it is used determining the effect of prognostic course or different types of treatment at the personal level. It is necessary for the planning, realization, analysis and interpretation of all kinds of researches such as clinical and field researches in the fields of medicine and health. The researches that are not carried out without the use of biostatistics techniques are not scientific researches. Therefore, the selection of the correct statistical methods is especially important in medical research (12). Logistic regression (LR) is a multivariable method that was devised for dichotomous outcomes (1–4). It is particularly appropriate for models involving disease state (diseased/ healthy) and decision making (yes/no), and therefore is widely used in studies in the health sciences (13-16).

If we look at the main principles and developments of the tests we have applied in this study; Although regression and correlation analyzes are used to determine risk factors, logistic regression method has been widely used recently. Logistic or logistic models are models used to study the variation of a single binary response variable depending on one or more discrete or continuous type explanatory variables (17). In this method mainly interested in the probability of realization (P) . Here, the probability of p is in the range [0,1] and cannot be defined directly with a linear model (18). The aim of the use of the logistic regression analysis; The same with the other model creation techniques are the same. In other words, logistic regression; Using the minimum number of variables possible, it can described as biologically significant and with good harmony a meaningful model

**Table 3.** The comparison of univariate and multivariate logistic regression results

	Univariate analysis results			Multivariate analysis results		
	PVI(+)	PVI(-)	P<0.05	*Adjusted Odds Ratio [Exp(B)] (95% CI) / Adjusted Mean Difference (95% CI)	Wald	P<0.05
Family history	34 (61.8%)	21 (38.2%)	0.008	2.62 (1.35, 5.07)	8,127	0.04
Smoking	37 (59.7%)	25 (40.3%)	0.04	2.08 (1.13, 3.8)	5,615	0.02
Prolonged standing	90 (53.6%)	78 (46.4%)	0.01	1.83 (1.06, 3.16)	4,791	0.03
Urban life style	38 (56.7%)	29 (43.3%)	0.04	1.73 (0.96, 3.15)	3,299	0.07**
Constipation	64 (55.2%)	52 (44.8%)	0.001	1.74 (1.03, 2.92)	4,306	0.04

\* Logistic regression was performed with adjusting for differences in family history, provincial area, prolonging standing, constipation and smoking. In order to determine the variables to be included in the multivariate logistic regression model, univariate analysis test results for each of the candidate variables are given in Table III. The effect of urban lifestyle was found different in terms of univariate and multipl logistic regression results.

that clarify the relationship between the result variable and the independent variable with correctly (19) . The logistic regression analysis, in case of various hypothesis corruption in which does not have normal distribution and lack of common covariance; It is an alternative to discriminant analysis and cross table (19). Regression Equation: It is an equation that mathematically explains the relationship between two or more variables. This equation shows the degree of how some variables can be estimated from other known variables, or the degree of the relationship between variables. Regression line is the most appropriate line that passes through the points on a scattering diagram created to summarize the relationship between the variables, and so as to be as close to them as possible and to best summarize them. If the line has a downward slope, the relationship is negative. If the slope of the line is upwards, the relationship is positive. On the other hand, in a regression equation (Y=a+bX); Regression coefficient is known as a (b) inclination of the line . After the regression equation (Y<sub>i</sub>=a+bX) is obtained, it is possible to draw on the scatter diagram of the line by knowing the coefficients a and b. In the linear regression analysis, the dependent variable and the independent variable(s) are numerical (continuous or discrete numerical) as specified. The method to be applied according to the category number of the dependent variable differs. If the categorical differences of each of the dependent variables result in two different conditions, binominal ie binary logistic regression should be selected (5,20-22). Due to the we too had two different categorical independent variables, we performed binary logistic regression.

**On the other hand, if we take a look at the necessities of the logistics model;**  
**Logistics model;**

It is represented by the equation shown below

$$\lambda(\chi_i) = E(y_i | \chi_i) = \log[P(\chi_i)/(1 - P(\chi_i))] = \sum \beta_k \chi_{ik} \quad (1)$$

(i=1, 2, ....., n; k=1, 2, ....., p; xi0=1)

Here, the probability of P (xi) is

$$P(\chi_i) = \frac{e^{\sum \beta_k \chi_{ik}}}{1 + e^{\sum \beta_k \chi_{ik}}} = \frac{e^{\lambda(\chi_i)}}{1 + e^{\lambda(\chi_i)}} \quad (2)$$

in that way. (5) This equation is called “logistics function “.

**Estimation of parameters**

For parameter estimation in logistic regression, mostly likelihood method is widely used. First of all, for the implementation of this method ; The maximum likelihood function should be created. This function is that as a function of unknown parameters gives the possibility of the observed data. The contribution to likelihood function of the (xi, yi) pair writeable as the below;

$$P(y_i | \chi_i) = P(\chi_i)^{y_i} [1 - P(\chi_i)]^{1-y_i} \quad (3)$$

Due to the assuming that the observations are independent of each other; The likelihood function is obtained by multiplying of the terms in equation, for n observation (4,5,23). According to this description, the likelihood function is written as below;

$$L(y|x, \beta) = \prod_{i=1}^n P(\chi)^{y_i} [1 - P(\chi)]^{1-y_i} \quad (4,23,24)$$

The test of the importance of the logistic regression coefficients

The using with the likelihood function; The process of the comparison between with the expected values and observed values is made with the following statement

$$D = -2 \log \left[ \frac{\text{The probability of the reduced model}}{\text{The probability of the all model}} \right] \quad (5)$$

The expression in parentheses indicates likelihood ratio (25);

If it will be written that denominated in the log probability function; The expression in parentheses indicates likelihood ratio (25). The results are showed the below that with formulation.

$$D = -2 \sum_{i=1}^n \left[ y_i \log \left( \frac{\hat{p}_i}{y_i} \right) + (1 - y_i) \log \left( \frac{1 - \hat{p}_i}{1 - y_i} \right) \right] \quad (4,5,23,24)$$

In here  $\hat{p}_i = \hat{p}(x_i)$  (5). D statistic value play a role, while determination of the in harmony goodness.

In order to determine the importance of an independent variable,

where this independent variable exists and does not exist

D values are compared in cases.

Due to the whether the independent variable exists or not change the emerging change in D.

Expressed in, G=D (Reduced model) - D (Whole model) (5,26).

**Interpretation of coefficients**

For the interpretation of coefficients in logistic regression ‘odds ratio’s’ are utilized. Odds Ratio (Ω), The dividing of the calculated for x=1 odds value to calculated for calculated x=0 the odds value. According to this odds ratio;

$$\text{Writable as this; } \Omega(1,0) = \frac{P(1)/[1 - P(1)]}{P(0)/[1 - P(0)]} \quad (5,27)$$

Accordingly, logistic regression the argument is binary and it is in the case of coding as 0, 1 odds ratio (5,24).

It is expressed as;  $\Omega = e \quad (28,29)$

In determining the goodness of harmony Hosmer Lemeshow C^ test statistic is used and ‘2m’ can be calculated as (m: risk group). C^ statistic is t-2 shows that of degrees of freedom ki kare distribution (16) .

**Determination of goodness of fit**

In determining the goodness of harmony Hosmer Lemeshow

C test statistic is used and ‘2m’ can be calculated as (m: risk group).

C statistic is t-2 shows that of degrees of freedom ki kare distribution.

$$\hat{C} = \sum_{m=1}^t \left[ \frac{(g_{1m} - b_{1m})^2}{b_{1m}} + \frac{(g_{0m} - b_{0m})^2}{b_{0m}} \right] \quad (16)$$

Logit is the natural logarithm of the Odds ratio and odds ratio is asymmetric. Taking its natural logarithm; It is converted to symmetrical. Logistic coefficients (logistic) are the equivalent of the ‘coefficient b’ in the linear regression analysis. The packet programs give the standard error of the coefficient b, the Wald statistics for significance, the odds ratio and the confidence interval of the odds ratio. In multivariable methods, the calculation of the coefficients from the original data is more complex than can be conveniently performed by hand. The calculation of the coefficients for logistic models involves equations that cannot be solved

explicitly but can be solved by an iterative procedure, easily expressed in computational form, and now widely available in statistical software packages. The quality of the regression analysis depends heavily on researchers understanding the assumptions inherent in the method and following principles developed to ensure their sound application (30). In the varicose veins, Trendelenburg’s proposal implicating has received wide acceptance in the late 1800s (31). But on the other hand varicose veins are common and many affected peoples remain without any symptom, just only a minority present for treatment. Consequently, population-based surveys of these venous disorders are rare, and usually restricted to a few subsets. To our knowledge, this study is the first to report the examined varicose vein disease risk factors in the young adult population with the comparison different statistical methods. Clinical manifestations are vein-related skin trophic changes, ranging from pigmented dermatitis to lipodermatosclerosis, white atrophy, and leg ulcers, pitting ankle edema and symptoms attributable to venous dysfunction, such as aching, pain, congestion, skin irritation, and muscle cramps (32). Community-based studies show that on average 15% of men and 25% of women have visible varicose structures (33). In one of the recent studies conducted in 2018, it was reported that venous insufficiency was found in approximately 50.91% of cases with PVI and chronic deep venous thrombosis. Unfortunately it is medically obvious that this is an important risk factor for embolism and mortality (34). Early detection is essential in the treatment of varicose veins and it can prevent further complications. Thus, correct management of varicose veins need to accurate identification of the source of superficial venous incompetence. Treatment, especially tailored to abolish of venous reflux, should relieve any symptoms attributable to superficial venous incompetence, as well as impede complications and ameliorate cosmesis. Pakdemirli A et al. (35) had stated that mesenchymal stem cells (MSCs) are a type of adult stem cell consisting of a heterogeneous subset of stromal stem cells that can be isolated from adult tissues. Additionally thet had stated that folic acid is another important contributor to tissue regeneration and repair, which affects the synthesis of some building block molecules used for wound healing. On the other side Athanerey A et al. (36) they had stated that mesenchymal stem cells (MSC) are capable and proved of accelerating wound healing in vivo and their study with human chronic wounds is currently awaited especially as venous leg ulcers wounds. In this point we too think that especially thanks to early diagnosis maybe without any invasive venous leg ulcers process this can be the more slowly slower with mesenchymal stem cells (MSC) by

powered by folic acid. Without any invasive venous leg ulcers, this pathologic process can be slowing and maybe stopping. However, in order to reach these conclusions, we would like to emphasize that the statistical methods chosen especially in medical studies should be accurate and complete.

The optimal outcome is low morbidity, low recurrence rates and, if possible, a short recovery time (37). Numerous risk factors related to genetic, lifestyle, and physiological characteristics have been suggested as having a contributing role in the development of varicose veins. Those attracting considerable interest include prolonged standing or sitting, obesity, constipation, pregnancy, family history of varicose veins, and a personal history of thrombosis (38). Cornu-Thenard et al. (29) conducted a study in which 67 patients with varicose veins and their parents were compared with 67 non-symptomatic (control) patients and their parents (39). Our results too indicated that family history was a contributing factor in the occurrence of varicose veins, with 2.62 ratio ( $p < 0.04$ ). Additionally, family history was found to be the most influential factor in the development of varicose veins. Furthermore correlation between cigarette smoking and varicose veins was found among men in the Musil D et al. (40) found that there had been declared a risk factor for venous clearance in cigarettes in 2016. Smoking contributes to a myriad of health problems, including high blood pressure, known to cause varicose veins. Significantly, in our study, smoking was confirmed as the risk factor, with 2.1 odds ratio ( $p < 0.02$ ). Additionally; The standing had been posited to be associated with the development of varicose veins too. In another population-based study showed that long-term sitting or standing shows an important role in lifestyle not only in varicose veins but also in the pathogenesis of venous symptoms with similar risk factors (41). In this study, prolonged standing was found another influential factor in the development of varicose veins, with the odds ratio of 1.83 ( $p < 0.03$ ). Another factor examined in pertinent studies was diet. Cleave (42) was the first to postulate that a diet deficient in fiber-rich foods contributes to the development of varicose veins. He hypothesized that the load of stagnant cecum causes compression of the iliac veins. Burkitt (37) expanded on this work, claiming that constipation and the resultant straining during defecation cause an increase in intra-abdominal pressure, which repeated over time, leads to vein dilation. In our study, constipation was also a significant factor in increasing the risk of primary venous insufficiency, with the odds ratio of 1.74 ( $p < 0.04$ ). Siddharth Mishra and colleagues (38) have found a higher rate of urban lifestyle in the development of varicose veins (62% vs 38%). In

our study, although we found the urban lifestyle to be effective in the development of varicose veins in the univariate analysis, we found that the urban lifestyle was not a risk factor in the development of varicose veins due to multivariate logistic regression analysis [1.73 odds ratio ( $p = 0.07^*$ )]. CVI may start early in life and progress over time. For example, 10% of schoolchildren aged between 10 and 12 years that took part in the Bochum study (43) had slight varicose veins. In this study focusing on young male population in their twenties, we have shown a strong relationship between early development of varicose veins and the examined risk factors like longstanding, smoking, family story and constipation. Although urban life style was found as a risk factor in results of univariate analysis for development early varicose veins etiopathogenesis; In the next step statistical analysis with multipl binary logistic regression, it was found not significant as statistically. Stocking K et al. (44) have underlined the importance of the medical statistics with the power and precision in trials and meta-analyses with their a review article in 2019. Furthermore Baveja CP and Aggarwal P. had been stated that no study in medical science could be completed without being dependent on statistical principles, additionally they that misinterpretation of the statistical tests cause of the mis results and misinterpretation (45). Zhou J et al. (46) in their study in construction of a logistic regression model for patients; They had stated that according to the results of univariate analysis, the multivariate analysis was performed to determine the independent prognostic factors and construct a LRM. For univariate statistic results Bircan H. (5) too had stated, even if any results was found important in the univariate model, the variables which are not significant in the multivariate regression model should be excluding from the model and thus the final model should be obtained with this way. Additionally in 2019, Sunderland KM et al. (47) conducted had claim that insufficiency of the univariate analysis in the utility of multivariate outlier detection techniques for data quality evaluation in large studies. In first part of the study with univariate analysis; We have identified five critical risk factors in etiopathogenesis of primary chronic venous disease. On the other hand; We have identified four critical risk factor in etiopathogenesis of primary chronic venous disease. Urban factor was not detected as a risk factor in multiple regression. In this study, we see the importance of statistics and mathematics in medical studies. We found five different risk factors in the univariate analysis of our statistical application. However, in multi-logistic regression results, where independent risk factors were evaluated together in advanced statistical analysis, we found that urban lifestyle, which we found not to have

a reliability level of  $p < 0.05$ , was excluded from the risk factors. As a result of this study; We believe that the analysis of multivariate risk factors should be kept in mind that the analyzes that contain only univariate analysis methods are not sufficient and that statistics that ignore the effect of each variable on the other may cause very bad clinical interpretations. If we determine our medical practice and control and policies according to based on first statistical part of this study, we see that an extra wasted time, effort and money will be. Our data also show that these four risk factors precede further complications and can thus assist in increasing the patient's awareness of the disease progress. On the other hand; Wrongly chosen statistical methods are creates resource of the wrong results and unfortunately this too source for wrong medical comments and treatments or preventive precautions.

## CONCLUSION

The choice of the statistical method has critical important. Due to the with the wrong choice occurrence of the wrong results can be cause of the occurrence of the wrong medical bias. According to the results of multiple logistic regression results; The urban lifestyle is not effective in the development of lower extremity varicose veins whereas we believe that varicose veins, especially constipation, smoking, and long-term standing risk factors can be decreased by preventive medical measures and prevent further serious complications, such as cellulitis, venous stasis ulcers or individual discomfort and disabilities.

## ETHICAL DECLARATIONS

**Ethics Committee Approval:** The study was carried out with the permission of T.C. Ministry of Health, Etlik Zübeyde Hanım Gynecology Training and Research Hospital (Permission granted /date-TUEK: 04/07/2019, Number of meeting: 11, Decision no: 90057706-799).

**Informed Consent:** Because the study was designed retrospectively, no written informed consent form was obtained from patients.

**Referee Evaluation Process:** Externally peer-reviewed.

**Conflict of Interest Statement:** The authors have no conflicts of interest to declare.

**Financial Disclosure:** The authors declared that this study has received no financial support.

**Author Contributions:** All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

## REFERENCES

- Köklü N, Büyüköztürk Ş, Çokluk –Bökeoğlu Ö. Sosyal bilimler için istatistik. 2. Baskı. Pegem A Yayıncılık, 2007.
- Arıcan E. Moleküler biyolojide kullanılan biyoistatistiksel yöntemler. file:///C:/Users/user/ Downloads /biyoistatistik\_1\_1.pdf.2017. Erişim Tarihi: 08.05.2017.
- Belle GV, Fisher LD, Heagerty PJ, Lumley T. Biostatistics: A methodology for the health sciences. 2st ed. John Wiley-Sons; 2010; 552-61.
- Taşdelen B, Kanık EA. Sağlık araştırmalarında biyoistatistiksel yöntemlerin doğru kullanımı ve sunumu. Mersin Univ. Sağlık Bilimleri Derg 2009; 1: 1-13.
- Bircan H. Lojistik regresyon analizi: Tıp verileri üzerine bir uygulama. Kocaeli Üniversitesi Sosyal Bilimler Enstitüsü Derg 2001; 2: 185-208.
- Callam MJ. Epidemiology of varicose veins. Br J Surg 1994; 81: 167-73.
- Rabe E, Partsch H, Jünger M, et al. Guidelines for clinical studies with compression devices in patients with venous disorders of the lower limb. Eur J Vasc Endovasc Surg 2008; 35: 494-500.
- Cannon J, Rajakaruna G, Dyer J, Carapetis J, Manning L. Severe lower limb cellulitis: defining the epidemiology and risk factors for primary episodes in a population-based case-control study. Clin Microbiol Infect 2018; 24: 1089-94.
- Porter JM, Moneta GL. Reporting standards in venous disease: an update. International Consensus Committee on Chronic Venous Disease. J Vasc Surg.1995; 21: 635-45.
- Robertson L, Evans C, Fowkes FG. Epidemiology of chronic venous disease. Phlebology 2008; 23: 103-11.
- Carradice D. Superficial venous insufficiency from the infernal to the endothermal. Ann R Coll Surg Engl 2014; 96: 5-10.
- Öncü Öner T, Can Ş. Sağlıkta biyoistatistiksel uygulamalar. İzmir Kâtip Çelebi Üniversitesi Sağlık Bilimleri Fak Derg 2018; 3: 39-45.
- Vollmer RT. Multivariate statistical analysis for pathologists. Part I, The logistic model. Am J Clin Pathol 1996; 105: 115-26.
- Lemeshow S, Hosmer DW. Logistic regression. In: Armitage P, Colton T, Eds. Encyclopedia of Biostatistics. New York: J. Wiley, 1998. p. 2316-27.
- Glantz SA, Slinker BK. Primer of applied regression and analysis of variance. New York: McGraw-Hill, Inc., 1990. ISBN:0070234078.
- Hosmer WD, Lemeshow S, Klar J. Goodness of fit testing for multiple logistic regression analysis when the estimated probabilities are small. Biometrical J 1988; 30: 911-24.
- Robert G, Rao JNK, Kumar s. Logistic regression analysis of sample survey data. Biometrika 1987; 74: 1-12,
- Şahin M, Efe E. Lojistik regresyon yöntemi ile doğum ağılığına etki eden faktörlerin belirlenmesi. Black Sea J Health Sci 2018; 1: 22-7.
- Chatfield C, Collins A. Introduction to Multivariate Analysis. Chapman & Hall, London, 1992.
- Gibbons RD, Hedeker D. Random effects probit and logistic regression models for three-level data. Biometrika 1997; 53: 1527-37.
- Scott A, Wild CJ. Fitting logistic regression models in stratified case-control studies. Biometrics 1991; 47: 497-510.
- Feinstein AR. Multivariable analysis: an introduction. New Haven, CT: Yale University Press, 1996.
- Dietz K, Gail M, Krickeberg K, Tsiatis A, Samet J. Statistics for Biology and Health. Logistic Regression A Self-Learning Text. 2 st ed. With Contributions by Erica Rihl Pryor 2002; 102-24.

24. Imai K, Enamorado T. POL572 Quantitative Analysis II Spring 2016; 63-5.
25. Van Doornmalen JPCM, Kopinga K. Temperature dependence of F-, D- and z-values used in steam sterilization processes. *J Applied Microbiol* 2009; 107: 1054-60.
26. Cox DR, Snell ES. *Analysis of Binary Data*. 2nd ed. Chapman and Hall, London. 1989.
27. Jerome C. Joint dependence of risk of coronary heart disease on serum cholesterol and systolic blood pressure: a discriminant function analysis. *Fed Proc* 1962; 21: 58-61.
28. Korkmaz M, Güney S, Yiğiter ŞY. The importance of logistic regression implementations in the Turkish livestock sector and logistic regression implementations/fields. *J Agric Fac HRU* 2012; 16: 25-36.
29. Frasin BA. Coefficient bounds for certain classes of bi-univalent functions. *Hacettepe J Mathematics Statistics* 2014; 43: 383-9.
30. Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. *J Clin Epidemiol* 2001; 54: 979-85.
31. Trendelenburg F. "Über die Unterbindung der Vena saphena magna bei Unterschenkelvaricen". [Brun's] *Beiträge zur Klinischen Chirurgie* 1891; 7: 195-210.
32. Scott TE, Lamorte WW, Gorin DR, et al. Risk factors for chronic venous insufficiency: A dual case-control study. *J Vasc Surg* 1995; 22: 622-8.
33. Bradbury A, Evans C, Allan P, Lee A, Ruckley CV, Fowkes FGR. What are the symptoms of varicose veins? Edinburgh vein study cross sectional population survey. *BMJ* 1999; 318: 353-6.
34. Tolu İ, Durmaz MS. Frequency and significance of perforating venous insufficiency in patients with chronic venous insufficiency of lower extremity. *Eurasian J Med* 2018; 50: 99-104.
35. Pakdemirli A, Toksöz F, Karadağ A, Mısırlıoğlu HK, Başpınar Y, Ellidokuz H, Açıkgöz O. Role of mesenchymal stem cell-derived soluble factors and folic acid in wound healing. *Turk J Med Sci* 2019; 49: 914-21.
36. Athanery A, Patra PK, Kumar A. J. Mesenchymal stem cell in venous leg ulcer: An intoxicating therapy. *Tissue Viability* 2017; 26: 216-23.
37. Burkitt DP. Varicose veins, deep vein thrombosis and hemorrhoids: Epidemiology and suggested aetiology. *Br Med J* 1972; 2: 556-61.
38. Mishra S, Ali I, Singh G. A study of epidemiological factors and clinical profile of primary varicose veins. *Med J Dr. DY Patil University* 2016; 9: 617-21.
39. Cornu-Thenard A, Boivin P, Baud JM, De Vincenzi I, Carpentier PH. Importance of the familial factor in varicose disease. Clinical study of 134 families. *J Dermatol Surg Oncol* 1994; 20: 318-26.
40. Musil D, Kaletova M, Herman J. Vasa. Risk factors for superficial vein thrombosis in patients with primary chronic venous disease. *Vasa* 2016; 45: 63-6.
41. Yun MJ, Kim YK, Kang DM, et al. A study on prevalence and risk factors for varicose veins in nurses at a university. *Hospital Saf Health Work* 2018; 9: 79-83.
42. Cleave TL. The neglect of natural principles in current medical practice. *J R Nav Med Serv* 1956; 42: 55-83.
43. Schultz-Ehrenburg U, Weindorf N, Von Uslar D, et al. Prospektive epidemiologische Studie & uml; ber die Entstehungsweise der Krampfader bei Kindern und Jugendlichen (Bochumer Studie I und II). *Phlebol Proktol* 1989; 18: 3-11.
44. Stocking K, Wilkinson J, Lensen S, Brison DR, Roberts SA, Vail A. Are interventions in reproductive medicine assessed for plausible and clinically relevant effects? A systematic review of power and precision in trials and meta-analyses. *Hum Reprod* 2019; 34: 659-65.
45. Baveja C P, Aggarwal P. Statistical analysis of microbiological diagnostic tests. *Indian J Med Microbiol* 2017; 35: 184-93.
46. Zhou J, Li Q, Gong G, Gong H, Hou Z. Analysis of prognostic factors and construction of a logistic regression model for patients with drug-induced liver failure. *Zhong Nan Da Xue Xue Bao Yi Xue Ban=J Central South Univ Med Sci* 2018; 43: 1337-44.
47. Sunderland KM, Beaton D, Fraser J, et al. The utility of multivariate outlier detection techniques for data quality evaluation in large studies: an application within the ONDRI Project. *Manuel Montero-Odasso. BMC Medical Research Methodology* 2019; 19: 1-16.