

Research Article

Aneta Cymbaluk-Płoska*, Anita Chudecka-Głaz, Agnieszka Sompolska-Rzechuła, Kamila Rasinska, Paulina Dubiel, Janusz Menkiszak

Risk model in women with ovarian cancer without mutations

<https://doi.org/10.1515/med-2018-0084>

received October 2, 2017; accepted August 9, 2018

Abstract: Ovarian cancer is characterised by the greatest mortality among all tumors of the reproductive tract.

This study included 246 patients which consisted of 136 women with ovarian cancer without genetic mutation and 110 women with benign ovarian cysts.

We created two mathematical logic models containing positive and negative risk factors of ovarian cancer such as: age at last menstruation cycle, patient age, OC, HRT, smoking, education status, and alcohol consumption. The calculated cut-off point for the first model was 0.5117. Classification determined on the basis of that cut-off point yielded 87.19% of correctly classified cases, of which 91.38% are “case” and 81.61% - „noncase”. For the second model the designated cut-off point was set at 0.5149 and the percentage of correctly classified patients was 88.12%, with 92.24% correctly rated as cancer patients and 82.56% of the cases rightly recognised as having no ovarian cancer.

Logit is a simple mathematical model that can be a useful tool for identification of patients with increased risk of ovarian cancer.

Keywords: Logit; Mathematical model; Ovarian cancer; Risk factor.

*Corresponding author: Aneta Cymbaluk-Płoska, Department of Gynecological Surgery and Gynecological Oncology of Adults and Adolescents, Pomeranian Medical University, Szczecin, Poland, postal address: al. Powstańców Wielkopolskich 72, 70-111 Szczecin, Poland, anetac@data.pl

Anita Chudecka-Głaz, Kamila Rasinska, Paulina Dubiel, Janusz Menkiszak, Department of Gynecological Surgery and Gynecological Oncology of Adults and Adolescents, Pomeranian Medical University, Szczecin, Poland

Agnieszka Sompolska-Rzechuła, Department of Statistics, West Pomeranian University of Technology, Szczecin, Poland

1 Introduction

Ovarian cancer remains a serious problem of modern oncological gynecology. It is the fifth malignancy in the structure of disease in women in Europe and the USA [1].

In the majority of cases, ovarian cancer is diagnosed at the late stage of clinical development (III-IV according to FIGO), which significantly reduces the chances of a complete cure.

The five-year survival rate for patients with stage I disease is 70-80%, while it drops to 15% in patients with stage IV. The cause of late diagnosis is the asymptomatic course of the disease in its early stages and non-specificity of symptoms, which are often confused with other diseases.

Ovarian cancer is characterized by the highest mortality among all gynecological cancers, which is why numerous scientific researches are being carried out to seek new causative agents as well as new biomarkers.

Apart from the recognized and well researched gene mutations responsible for increased risk of occurrence of ovarian cancer, the impact of environmental factors remains unclear to date [2-4]. We grouped all environmental factors known for increasing the risk of ovarian cancer in a way that enables stratification of patients into groups at high or low risk of ovarian cancer development. It will allow extending closer medical care to those high-risk individuals.

2 Material and methods

All subsequent patients reporting to the Department of Surgical Gynecology Adult and Adolescent Gynecologic Oncology of Pomeranian Medical University in Szczecin in years 2013-2015 with changes in the adnexa in the ultrasound examination were qualified for the study. After receiving the histopathological result, 246 patients were enrolled in the study.

Patients with endometrial cysts, dermoid cysts and patients with ovarian border tumors were excluded from the study.

All patients with confirmed BRCA1 and BRCA2 mutations were excluded from the study. Our interests focused on several aspects, primarily on patient age and age at the first menstruation.

The age distribution in the studied population is characterised by weak left-sided asymmetry, while the distribution of the age at first and last menstruation is moderately asymmetrical. The direction of distribution of the age at menarche is positive; the distribution of the age at last menstruation has a negative direction.

In our study we analysed the characteristics that affect the odds ratio for ovarian cancer occurrence. For that purpose, we produced contingency tables where ovarian cancer was the tested phenomenon and specific properties constituted the risk factors.

Contingency table.

The odds ratio (OR) with 95% confidence interval was also calculated:

$$OR_{AxB} = \frac{S(A)}{S(B)} = \frac{P(A)}{1 - P(A)} : \frac{P(B)}{1 - P(B)}$$

Weak or moderate asymmetry of the distribution of selected characteristics was taken into account: patient age, age at menarche, and age at last menstruation, and only mean values were considered.

The logistic regression model allows for examining the influence of multiple independent variables X_1, \dots, X_k on a dichotomous dependent Y variable. The dependent variable is coded in the following way: 1 - case (success), 0 - noncase (failure). Logistic function of values in the (0;1) interval and an S-shaped graph function are used describe logit regression; its analytic expression is such:

$$f(z) = \frac{e^z}{1 + e^z} = \frac{1}{1 + e^{-z}}, \quad z \in R$$

The logistic regression model for dichotomous Y variable determines the conditional probability of this variable assuming the “case” value and is expressed as follows:

$$P(Y = 1 / X_1, \dots, X_k) = \frac{e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}}{1 + e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}}$$

where: $\alpha_0, \alpha_1, \dots, \alpha_k$ are the parameters of the model, X_1, \dots, X_k are independent variables that may be either qualitative or quantitative.

The logistic regression model coefficients may be searched for using the maximal likelihood method and the generalized least squares estimator.

Due to the non-linear character of the model with respect to independent variables and parameters, the logarithm is used to transform the logistic model into a linear one. For that purpose, the concept of the odds ratio is introduced, which is the ratio of the probability of occurrence of a given event to the probability of it not occurring; it is expressed in the following way:

$$\frac{P(Y = 1 / X_1, \dots, X_k)}{1 - P(Y = 1 / X_1, \dots, X_k)} = \frac{e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}}{1 + e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}} : \frac{1}{1 + e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}} = e^{\alpha_0 + \alpha_1 X_1 + \dots + \alpha_k X_k}$$

Thus, the odds ratio expresses how many times the probability of a given event occurring increases or decreases when the independent variable changes by a single unit (at fixed values of the remaining independent variables).

The natural logarithm of the odds ratio is linear with respect to independent variables and the parameters of the model, which makes the estimation easier. It is known as the logit function or the logit form of the logistic model, therefore [5,6]:

$$\text{logit}P = \ln \frac{P(Y = 1 / X_1, \dots, X_k)}{1 - P(Y = 1 / X_1, \dots, X_k)} = \alpha_0 + \sum_{i=1}^k \alpha_i X_i$$

Based on the logit model we assessed the accuracy of the estimated model, calculating the numbers of correctly and incorrectly classified cases (table 2).

In order to assess the goodness of fit of the regression model to the empirical model we may use the R2 coefficient, which assumes the values from the (0,1) interval, and is defined as follows (7):

$$R^2_{zlicz} = \frac{n_{11} + n_{22}}{n}$$

Table 1: Risk factors for patients with ovarian cancer.

Occurs		Risk factor		Total
		Does not occur	Total	
Examined phenomenon	Occur	a	b	a+b
	Does not occur	c	d	c+d
	Total	a+c	b+d	a+b+c+d

The closer this coefficient is to one, the better the logistic model fits empirical data, where R^2_{zlicz} signifies the percentage of correctly classified cases. This model works well in the prediction of the studied phenomenon when $R^2_{zlicz} > 50\%$. This means that model-based classification is better than random classification.

The quality of constructed logistic regression model may be also assessed with other measures, e.g.:

- Pseudo $R^2_{Pseudo} = 1 - \frac{LnL_{FM}}{LnL_0}$, where: L_{FM} - maximum likelihood function of the model containing all variables, L_0 - maximum likelihood function for the model containing the intercept only. It assumes values from the $(0, 1)$ interval. Values close to 1 denote a good fit, while values close to 0 denote lack thereof. Since the R2 coefficient does not assume the value of 1 and is sensitive to a number of variables in the model, its corrected value is determined: R2Nagelkerke and R2Cox-Snell

$$R^2_{Nagelkerke} = \frac{1 - e^{-(2/n)(LnL_{FM} - LnL_0)}}{1 - e^{(2/n)LnL_0}}$$

$$R^2_{Cox-Snell} = 1 - e^{-2(LnL_{FM} - LnL_0)/n}$$

- The Hosmer-Lemeshow test - a test, which for different subsets of data, compares the observed frequencies of occurrence of subjects with the studied characteristic (Og) in a given subset of objects and expected frequencies of occurrence (Eg) of the examined values. If Og and Eg are sufficiently close, it may be assumed that a well-fitting model was built.
- AUC (Area Under the Curie) – area under the ROC curve, which is produced by connecting points in a Cartesian coordinate system: (1-specificity, sensitivity).

The sensitivity of the test is the ratio of the true positive to the true positive and false negative results. The sensitivity of 100% in a medical test means that all people who are ill or generally with specific sought disorders will be recognized. The term is interpreted as the ability of the

test to correctly diagnose the disease where it occurs. The specificity of the test is the ratio of true negative results to the sum of true negative and false positives. The 100% specificity means that all healthy people in the diagnostic test will be marked as healthy.

Area under the ROC curve depicts the evaluative quality of the model. A good model is such that it minimizes the number of errors, that is n_{12} and n_{21} values. Thus, a model characterized by high evaluative quality gives sensitivity and specificity values that are as large as possible, which for such a model the area under the ROC curve should be close to 1. When the ROC curve overlaps with the diagonal $x = y$, a decision on assigning a case to a given class made based on a model is as good as random assortment. Evaluative value of a model is high when the area under ROC curve is significantly greater than the area under the line $x=y$; thus, greater than 0.5. If the chances for occurrence of a given phenomenon increase or decrease together with an increase in the value of the variable, then we are looking for a so-called optimal cut-off point of predicted probability for the ROC curve, for which the dependent variable best divides the population into groups: one where the phenomenon occurs and one where it does not occur.

Ethical approval: The research related to human use has been complied with all the relevant national regulations, institutional policies and in accordance the tenets of the Helsinki Declaration, and has been approved by the authors' institutional review board or equivalent committee.

Informed consent: Informed consent has been obtained from all individuals included in this study.

3 Results

The mean patient age was 55.4 years. The mean age at menarche in the entire study population was 13.6 years; the mean age at the last menstruation was 47.2 years. In the group of patients with benign ovarian cysts the mean age was 41.5 years, in the group with ovarian cancer – 60.7 years (Table 3). Table 4 presents percentage distribution of various characteristics that might influence the risk of ovarian cancer in the two groups.

We calculated the risk of developing malignant ovarian tumor in the presence and the absence of a risk factor. We also calculated the odds ratio (OR) with 95% confidence interval using the test for OR. The test was performed to verify the hypothesis that a chance of occur-

Table 2: The matrix of case classifications.

Expected values	Observed values		Total
	yi = 1	yi = 0	
ŷi = 1	n11	n12	n1·
ŷi = 0	n21	n22	n2·
Total	n·1	n·2	n

Table 3: Age distribution of first and the last menstruation in the whole study group.

	N valid	Mean	Median	Minimum	Maximum	Lower quartile	Upper quartile	SD	Coefficient of variation
(FM) first menstruation	222	13.698	13.000	7.000	20.000	13.000	15.000	1.771	12.925
(LM) last menstruation	227	47.229	47.000	18.000	82.000	40.000	50.000	9.016	20.385
age	234	55.432	54.000	18.000	86.000	41.000	65.000	15.765	30.068

rence of a given phenomenon is the same in the group exposed to a specific risk factor as it is in the unexposed group.

Examined features and their categories:

1. Cancer (1 - yes, 0 - no),
2. Age at menarche (< 13 years, or > 13 years),
3. Age at last menstruation (< 48 years, or > 48 years),
4. Patient age (< 56 years, or > 56 years),
5. Menopause (yes, no),
6. Nulliparous (yes, no),
7. Uniparous (yes, no),
8. Multiparous (yes, no),
9. OC (yes, no),
10. HRT (yes, no),
11. Higher education (yes, no),
12. Smoking (yes, no),
13. Alcohol consumption (yes, no).
14. Family history of cancer (yes/no)

Age was the first parameter subject to the analysis in the following age groups: below and above 56 years of age.

The chance of developing ovarian cancer is almost 6.5-fold higher in women older than 56 years compared to women below 56 years of age (Table 5).

The chance of developing cancer is almost 7 times higher in women whose last menstrual cycle occurred above the age of 48 years compared to women whose last menstrual cycle occurred before the 48th year of life. Chance of developing cancer is 7-fold higher in menopausal women compared to those in premenopausal age (Table 6).

The chance of developing cancer is 2-fold greater in nulliparous women compared to those who have given

Table 4: Age analysis in individual groups

Age in years	Odds ratio OR	95% confidence interval	P
< 56 >	6.42	2.3 – 14.1	0.000001

Table 5: Percentage distribution of the various risk factors in ovarian cancer and control group.

Study group	Yes	No
Ovarian Cancer	53%	47%
Menopause	57%	43%
Nulliparous	24%	76%
Multiparous	57%	43%
Used OC	23%	77%
Used HRT	26%	74%
Higher education	30%	70%
Smoking	40%	60%
Alcohol	16%	84%

Table 6: The risk of ovarian cancer including the first and last menstruation.

First menstruation	Odds ratio OR	95% confidence interval	P
< 13 >	1.756	0.7 – 4.3	0.079
Last menstruation			
<48 >	6.82	3.1 – 16.2	0.00001
Menopause	7.09	4.4 – 17.3	0.00001

birth. The likelihood of developing cancer is 2.2 times lower among women who gave multiple births (Table 7).

The chance of developing ovarian cancer is over 9 times lower in women who have used OC compared to those who have not. The likelihood of developing ovarian cancer is 7 times higher for women who have used HRT compared to those who have not (Table 8).

The chance of cancer development is 3.7-fold higher in women with history of smoking compared to non-smoking women. The likelihood of cancer development is 2.3

Table 7: The risk of developing ovarian cancer, taking into account the parity of patients.

Parity	OR	95% confidence interval	P
Nulliparous	1.27	0.1 – 3.4	0.000087
Uniparous	1.4	0.6 – 8.2	0.29
Multiparous	2.17	0.9 – 7.9	0.0038

Table 8: The risk of ovarian cancer in patients taking OC and RT.

Use	OR	95% confidence interval	P
OC	9.47	4.7 – 19.5	0.00032
HRT	6.93	2.8 – 16.2	0.011

times lower in women who consumed alcohol compared to those who did not (Table 9).

Women with higher education have 10 times lower chance of developing cancer compared to women with other levels of education. The chance of developing ovarian cancer was 15 times greater in women who have a family history of more than one cancer.

Due to the presence of factors that increase or reduce the risk of ovarian cancer in isolation, we decided to build a model in which several factors were substituted simultaneously in order to identify patients with the greatest risk of developing ovarian cancer.

To assess the likelihood of ovarian cancer development we adopted the following set of potential variables:

- Y – ovarian cancer (yes, no),
- X1 – age at menarche (years),
- X2 – age at last menstruation (years),
- X3 – patient age (years),
- X4 – menopause (yes, no),
- X5 – parity (nulliparous, uniparous, multiparous),
- X6 – HRT use (yes, no),
- X7 – OC use (yes, no),
- X9 – smoking (yes, no),
- X10 – alcohol consumption (yes, no).

In order to find the best combination of features significantly influencing development of ovarian cancer we formally selected features using backwards stepwise regression analysis and obtained the following set of variables:

- X2 – age at last menstruation (years),
- X3 – patient age (years),
- X6 – HRT use (yes, no),
- X8 – higher education (yes, no),

Table 9: Chance of getting ovarian cancer depending on the usage of drugs.

Substance use	OR	95% confidence interval	p
Smoking	3.77	1.9 – 10.8	0.00004
Alcohol consumption	2.33	1.2 – 6.9	0.028

Table 10: Chance of getting ovarian cancer depending on other factors.

Other risk factors	OR	95% confidence interval	P
Higher education	9.7	5.4 – 18.2	0.000002
Family history of cancer	14.8	8.2 – 21.6	0.000001

Table 11: Evaluation of parameters of the logit model.

Variable	Variable	Parameter estimate	p value	Quotient
	Constant	-9.4880	0.0000	
X ₂	Age at last menstruation cycle (years)	0.1374	0.0019	1.1472
X ₃	Patient age	0.0621	0.0033	1.0641
X ₆	HRT	4.0709	0.0001	58.6122
X ₈	Higher education	-2.6223	0.0000	0.0726
X ₉	Smoking	1.7400	0.0004	5.6975

The resulting variables are weakly correlated with each other and, at the same time, strongly correlated with other characteristics excluded from the compilation of diagnostic features. Evaluation of the parameters of the logit model is presented in table 10.

In our model, the following characteristics have statistically significant positive effect on the dependent variable: age at last menstruation cycle, patient age, HRT, smoking, whereas higher education has a negative, statistically significant effect on the dependent variable.

Interpreting the odds ratios for the i-th variable (assuming that the other variables included in the model remain unchanged) we obtain the following information:

- later occurrence of last menstrual period (by one year) increases the risk of developing ovarian cancer by 15%,
- each additional year of life increases the chance of developing ovarian cancer by 6.5%,

- use of HRT increases the chance of developing the disease by 18 times compared to subjects who have not used HRT,
- higher education decreases the likelihood of ovarian cancer by 92.8% compared to women without higher education,
- smoking increases the risk of ovarian cancer 5.7-fold compared to non-smokers.

We evaluated the correctness of the estimated model by counting the accuracy of the classification of women on the basis of selected cut-off point 0.5117 (Table 11).

In the logit model the parity variable turned out to be irrelevant. Analysis of the odds ratio for ovarian cancer development, in a one-dimensional approach, variables denoted as “nulliparous” and “multiparous” demonstrated statistical significance. In another logit model we created three dichotomous dummy variables: nulliparous, uniparous and multiparous, in order to investigate whether they significantly affect the risk of developing ovarian cancer.

The following set of variables was used:

- Y – ovarian cancer (1 - yes, 0 - no),
- X1 – age at menarche (years),
- X2 – age at last menstruation (years),
- X3 – patient age (< 56 years, > 56 years),
- X4 – menopause (yes, no),
- X5 – nulliparous (yes, no),
- X6 – uniparous (yes, no),
- X7 – multiparous (yes, no),
- X8 – OC use (yes, no),
- X9 – HRT use (yes, no),
- X10 – higher education (yes, no),
- X11 – smoking (yes, no),
- X12 – alcohol consumption (yes, no).

In the logit model regression analysis was used to determine variables exerting significant effect on the dependent variable. We received a set of variables:

- X2 – age of last menstruation (years),
- X3 – patient age (< 56 years, > 56 years),
- X8 – HRT use (yes, no),
- X10 – higher education (yes, no),
- X11 – smoking (yes, no),
- X12 – alcohol consumption (yes, no).

No variable associated with fertility was found among the independent variables. An additional X12 variable (alcohol consumption) was included in the set. Evaluation of the parameters of the logit model is presented in table 12.

In this model the following factors have a positive, statistically significant effect on the dependent variable: age at the last menstrual period, age of the patient, HRT, smoking and alcohol consumption. In contrast, the higher education variable has a negative, statistically significant effect on the dependent variable.

Interpreting the odds ratios for the *i*-th variable (assuming that the other variables included in the model remain unchanged) we obtained the following information:

- later occurrence of the last menstrual period (by one year) increases the risk of developing ovarian cancer by 15%,
- each additional year of life increases the chance of developing ovarian cancer by 6.8%,
- use of HRT increases the chance of developing the disease 13 times compared to women who have not used HRT,
- having higher education decreases the likelihood of ovarian cancer by 33.4% compared to women without higher education,
- smoking increases the risk of ovarian cancer 5.2-fold compared to non-smokers,
- alcohol consumption reduces the risk of developing ovarian cancer 4.4-fold compared to non-drinking subjects.

We assessed the correctness of the estimated model by calculating the accuracy of classification of women on the basis of the identified cut-off point 0.5149 (table 13).

Accuracy of classification of both logit models were assessed using: R^2_{zlicz} , $R^2_{Nagelkerke}$, R^2_{Pseudo} , $R^2_{Cox-Snell}$, Hosmer-Lemeshow test and the ROC curve and its components. The results are shown in table 14.

Based on the results presented in table 14 it may be concluded that both logistic regression models are characterized by a very good fit to the empirical data. The R^2_{zlicz} coefficients in both models indicate very high accuracy of classification of patients based on logistic regression. The results of Hosmer – Lemeshow test indicate no significant differences between empirical and theoretical

Table 12: Accuracy of classification according to the logit model.

Classification on the basis of the logit model	Observed classification		General accuracy of classification
	yi=1	yi=0	
$\hat{y}_i=1$	106	16	87.19%
$\hat{y}_i=0$	10	71	
Sensitivity	91.38%	84.4%	
Specificity	90.6%	81.61%	

numbers obtained from the estimated logistic regression models.

Based on the analysis of the graphs shown in Figure 1 and calculations presented in Table 5, it can be concluded that the area under the ROC curve for both models is significantly greater than 0.5 (the significance level greater than 0.000001 for both models). Therefore, patients may be classified on the basis of constructed models. The calculated cut-off point for the first model was 0.5117. The classification determined on the basis of that cut-off point yielded 87.19% of correctly classified cases, of which 91.38% are “case” and 81.61% - “noncase”. For the second model the designated cut-off point was set at 0.5149 and the percentage of correctly classified patients was 88.12%, with 92.24% correctly rated as cancer patients and 82.56% of the cases rightly recognized as having no ovarian cancer – Figure 1.

4 Discussion

Genetic and environmental risk factors for ovarian cancer are well established. Genetic factors have been described relatively wellso our studies focused on environmental factors.

The current data indicate that ovarian cancer develops mainly in postmenopausal women, reaching peak of onset around 60 years of age [8]. Rosner and collaborators demonstrated with a mathematical model that increased risk among 55-year-old patients is 62% greater compared to 45-year-olds [9]. Our analyses show that the chance of developing ovarian cancer after the 56th year of life increases 6.5-fold per year compared to younger patients. We incorporated age into the logit as it is one of the factors that significantly increases the probability of illness.

In a study conducted by Yang et al., an increased risk of developing ovarian cancer was found in patients taking HRT. Carcinoma mucinosum ovarii is an exception. This type of histological ovarian cancer is less common in patients on HRT supplementation, RR - 0.37 [10]. In our

study group carcinoma papillare serosum was the histopathological type of cancer in 90% of patients. Therefore, we are unable to comment in this respect. Among patients with serous cancer the impact of HRT on the risk of developing the disease was (RR 1.37). Hunn et al, who coordinated a multicenter study, reported similar results. The study enrolled 12110 patients, 55% of which received estrogen or estrogen-progesterone HRT. A significantly increased risk of developing ovarian cancer was revealed in patients taking HRT compared to patients without current or past HRT (RR 1.43). Administration of hormones mainly affected development of serous and endometrioid types of tumors, which are the most common forms of ovarian cancer. According to this retrospective study the increased risk of development of serous ovarian cancer resulting from HRT was RR 1.56 (serous), and 1.46 (endometrioid), whereas in our study the values amounted to RR 1.43 (serous) and 1.32 (endometrioid) [4]. In the Million Woman Study Collaborators study, which was coordinated by Valerie Beral, the researchers found that taking hormone replacement therapy, either in the past or currently has a significant impact on increased risk of ovarian cancer development [11]. However, Lacey et al. emphasize

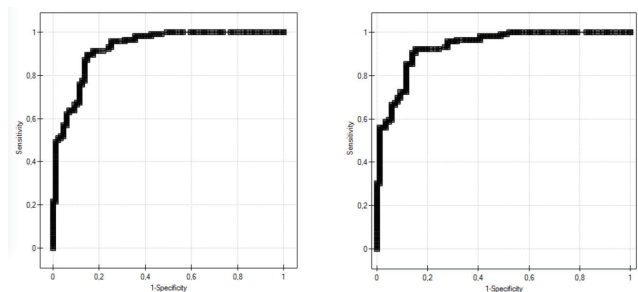


Figure 1: ROC curve for models 1 and 2.

Table 13: Evaluation of the parameters of the logit model.

Variable	Variable	Parameter estimate	p value	Quotient
	Constant	-9.8588	0.0000	
X_2	Age at last menstruation (years)	0.1372	0.0019	1.1471
X_3	Patient age	0.0661	0.0020	1.0683
X_8	HRT	4.1447	0.0001	13.1009
X_{10}	Higher education	-2.7135	0.0000	0.0663
X_{11}	Smoking	1.6559	0.0009	5.2377
X_{12}	Alcohol consumption	1.4831	0.0413	4.4066

Table 14: Accuracy of classification of the logit model.

Classification on the basis of the logit model	Observed classification		General accuracy of classification
	yi=1	yi=0	
$\hat{y}_i=1$	107	15	88.12%
$\hat{y}_i=0$	9	71	
Sensitivity	92.24%	79.9%	
Specificity	91.6%	82.56%	

Table 15: Goodness of fit of the logistic models to the empirical data.

Model	Coefficient of classification accuracy				Hosmer–Lemeshow test		Area under ROC curve
	R^2_{zlicz}	R^2_{Pseudo}	$R^2_{Nagelkerke}$	$R^2_{Cox-Snell}$	χ^2	p	
Model 1	87.19%	52.94%	69.11%	51.48%	5.64	0.6878	92.72%
Model 2	88.12%	54.38%	70.36%	52.38%	7.16	0.5192	93.39%

that the duration of estrogen-progesterone therapy correlated proportionally with an increased risk of serous or endometrioid ovarian cancer [12]. Other scientists in 2016 demonstrated that this estrogen therapy results in a significant increase in the risk of ovarian cancer in the form of serum and endometrial disease. The probability increases with the duration of hormone therapy [13]. Besides the effect of HRT, we should also consider the effect of oral contraceptives used by young women. It may be unequivocally concluded that taking oral contraceptive pills protects against development of ovarian cancer.

We found that taking OC reduced ovarian cancer morbidity to RR 0.74. Similar results were obtained by Dallal *et al.* They showed that administration of OC for 5 years reduces the risk to RR 0.54. This protective effect is extended to patients with genetic burden as well as those without it [14,15]. Faber *et al.* demonstrated that taking oral contraceptives protects against development of ovarian cancer regardless of its histopathological type. In our study, among patients taking OC for more than five years reduction in risk of ovarian cancer amounted to RR 0.78, compared to patients who have never used OC. These results do not differ from other reports. Faber *et al.* demonstrated that each year of OC use lowers the incidence of ovarian cancer by 6% (HR-0.94; 95% CI: 0.85, 1.02). Each year of ovulation accounts for a 2% increase in the chance of occurrence of ovarian cancer [16].

The situation is not clear when it comes to a risk factor, such as smoking.

In a multicenter study coordinated by Gram, analysis of patients' smoking history demonstrated an increased risk of ovarian cancer in patients who smoke or smoked in the past compared to non-smokers (RR 1.08) for all kinds of ovarian tumors. The risk of ovarian cancer is elevated for women who smoked in the past, but quit, as well as for those currently smoking. Furthermore, increased risk of mucinosum type of carcinoma was significantly higher in patients with malignancies borderline to mucinous – RR 2.25 vs. 1.49 [17]. Similar reports were published reports by La Vecchia (2016). He concluded that the risk of developing ovarian cancer is increased in patient smokers in general, but smoking particularly strongly increases the

incidence of borderline mucinous ovarian cancers [18]. In our research we found an increased risk of developing ovarian cancer among patients who are currently smoking or have smoked in the past compared to non-smokers (RR 1.16), without specifying the type of tumor. In recent studies by Jeon (2016), it was demonstrated that smoking patients were at increased risk of developing ovarian cancer. The mechanism of action was explained: nicotine increases the expression of metalloproteinase 9 and cathepsin D, initiating process of carcinogenesis [19].

In our study we found that low socioeconomic status predisposes to ovarian cancer. Patients with higher education are at 5-fold lower risk of developing ovarian cancer than patients with poor socio-economic status and who are less educated. Praestegard *et al.* obtained similar results, demonstrating that lower education level was associated with greater risk of ovarian cancer. Clinical staging of cancer also positively correlated with poorer education [20]. In 2016 Poole *et al.* examined 10000 patients, finding that low socioeconomic status and educational level not only influences the risk of ovarian cancer, but it is also a prognostic factor of poor outcome among patients already diagnosed with ovarian cancer. Since conclusive results for this factor have been obtained, we included it in our logit analysis [21].

We also found that the risk of ovarian cancer is increased in patients who have a history of cancer. We noted no statistically significant differences with regard to incidence of ovarian cancer among patients with other co-morbidities, such as: hypertension, type 2 diabetes or dyslipidemia. These results are comparable with the analysis of Teixeira *et al.*, who reported no increased risk for ovarian cancer among patients with metabolic diseases [22]. In our study, we demonstrated that the risk of ovarian cancer decreases with moderate alcohol consumption in comparison to women who abstained from alcohol (RR-0.78). Recent meta-analyzes based on 13 prospective studies have shown that alcohol consumption does not increase the risk of ovarian cancer [23]. Cook *et al.* found that consumption of wine reduced the incidence of ovarian cancer (RR 0.67), relative to patients who declared abstinence. Risk reduction is more pronounced in patients

who consumed red wine compared to those consuming white wine, (respectively, RR 0.44 vs. 0.79). There were no increased risk of the disease with consumption of beer or liquor regardless of age, or the amount of consumed alcohol compared to abstinent patients [24]. Similar results were published by Riman [25]. Based on our own studies and meta-analyses performed by other authors we concluded that alcohol consumption might be a factor in logit model. Taking into account the parity factor, our study showed that the chance of developing cancer is 2-fold higher in women who have never given birth compared with those who have. The likelihood of developing cancer is 2.2 times lower in women who have given birth multiple times. In a study published by Meritt et al. it was demonstrated that parity could be a protective factor. Possessing one child protects against development of serous cancer at a level of RR 0.68. Giving birth to three children protects at RR 0.68. Particularly strong risk reduction was found for endometrioid and clear-cell carcinoma [26]. Whiteman et al. presented results that corroborate these reports. High parity protects against the development of ovarian cancer, regardless of tumor type at a level of OR 0.71. Summarising our own research and the results obtained by other researchers we can assume that smoking women, aged around 60 years, who do not drink alcohol, who have not used contraception at youth, but are using hormone replacement therapy are at risk of developing ovarian cancer [27].

5 Conclusions

Logit, as simple mathematical model, can be a useful tool for identification of patients with increased risk of ovarian cancer.

Author contribution

A Cymbaluk-Płoska - Manuscript writing, participation in work methodology

A Chudecka-Głaz - Help with writing the manuscript

A Sompolska-Rzechuła - Assistance in the creation of the logit model

K Rasinska - Data collection

P Dubiel - Data collection

J Menkiszak - Help in the selection of references

List of abbreviations

OC - oral contraceptive

HRT - hormonal replacement therapy

Competing interests: The authors declare that there is no conflict of interest regarding the publication of this article.

References

- [1] Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J Clin.* 2015;65(2):87-108
- [2] Risch HA, McLaughlin JR, Cole DE, Rosen B, Bradley L, Kwan E, et al. Prevalence and penetrance of germline BRCA1 and BRCA2 mutations in a population series of 649 women with ovarian cancer. *Am J Hum Genet.* 2001;68(3):700-710
- [3] Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, Wozniak E, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet.* 2009;41(9):996-1000
- [4] Hunn J, Rodriguez GC. Ovarian cancer: etiology, risk factors, and epidemiology. *Clin Obstet Gynecol.* 2012;55(1):3-23
- [5] Cramer JS. *Logit models from economics and other fields.* Cambridge, UK ; New York: Cambridge University Press; 2003. x, 173 p. p
- [6] Kleinbaum DG, Klein M, SpringerLink (Online service). *Logistic regression a self-learning text.* New York, NY: Springer Science+Business Media, LLC; 2010. Available from: <http://dx.doi.org/10.1007/978-1-4419-1742-3>
- [7] Maddala GS. *Introduction to econometrics.* 3rd ed. Chichester; New York: John Wiley; 2001. xxvii, 636 p. p
- [8] Holschneider CH, Berek JS. *Ovarian cancer: epidemiology, biology, and prognostic factors.* *Semin Surg Oncol.* 2000;19(1):3-10
- [9] Rosner BA, Colditz GA, Webb PM, Hankinson SE. *Mathematical models of ovarian cancer incidence.* *Epidemiology.* 2005;16(4):508-515
- [10] Yang HP, Anderson WF, Rosenberg PS, Trabert B, Gierach GL, Wentzensen N, et al. Ovarian cancer incidence trends in relation to changing patterns of menopausal hormone therapy use in the United States. *J Clin Oncol.* 2013;31(17):2146-2151
- [11] Beral V, Banks E, Reeves G, Million Women Study C. Effects of estrogen-only treatment in postmenopausal women. *JAMA.* 2004;292(6):684; author reply 5-6
- [12] Lacey JV, Jr., Brinton LA, Leitzmann MF, Mouw T, Hollenbeck A, Schatzkin A, et al. Menopausal hormone therapy and ovarian cancer risk in the National Institutes of Health-AARP Diet and Health Study Cohort. *J Natl Cancer Inst.* 2006;98(19):1397-1405
- [13] Lee AW, Ness RB, Roman LD, Terry KL, Schildkraut JM, Chang-Claude J, et al. Association Between Menopausal Estrogen-Only Therapy and Ovarian Carcinoma Risk. *Obstet Gynecol.* 2016;127(5):828-836
- [14] Moorman PG, Havrilesky LJ, Gierisch JM, Coeytaux RR, Lowery WJ, Peragallo Urrutia R, et al. Oral contraceptives and risk of ovarian cancer and breast cancer among high-risk women: a systematic review and meta-analysis. *J Clin Oncol.* 2013;31(33):4188-4198
- [15] Beavis AL, Smith AJ, Fader AN. Lifestyle changes and the risk of developing endometrial and ovarian cancers: opportunities

- for prevention and management. *Int J Womens Health*. 2016;8:151-167
- [16] Faber MT, Jensen A, Frederiksen K, Glud E, Hogdall E, Hogdall C, et al. Oral contraceptive use and impact of cumulative intake of estrogen and progestin on risk of ovarian cancer. *Cancer Causes Control*. 2013;24(12):2197-2206
- [17] Gram IT, Lukanova A, Brill I, Braaten T, Lund E, Lundin E, et al. Cigarette smoking and risk of histological subtypes of epithelial ovarian cancer in the EPIC cohort study. *Int J Cancer*. 2012;130(9):2204-2210
- [18] La Vecchia C. Ovarian cancer: epidemiology and risk factors. *Eur J Cancer Prev*. 2016
- [19] Jeon SY, Go RE, Heo JR, Kim CW, Hwang KA, Choi KC. Effects of cigarette smoke extracts on the progression and metastasis of human ovarian cancer cells via regulating epithelial-mesenchymal transition. *Reprod Toxicol*. 2016;65:1-10
- [20] Praestegaard C, Kjaer SK, Nielsen TS, Jensen SM, Webb PM, Nagle CM, et al. The association between socioeconomic status and tumour stage at diagnosis of ovarian cancer: A pooled analysis of 18 case-control studies. *Cancer Epidemiol*. 2016;41:71-79
- [21] Poole EM, Merritt MA, Jordan SJ, Yang HP, Hankinson SE, Park Y, et al. Hormonal and reproductive risk factors for epithelial ovarian cancer by tumor aggressiveness. *Cancer Epidemiol Biomarkers Prev*. 2013;22(3):429-437
- [22] Teixeira N, Folgueira MA, Maistro S, Encinas G, Bock GH, Diz Mdel P. Association of family risk and lifestyle/comorbidities in ovarian cancer patients. *Rev Assoc Med Bras* (1992). 2015;61(3):234-239
- [23] Yan-Hong H, Jing L, Hong L, Shan-Shan H, Yan L, Ju L. Association between alcohol consumption and the risk of ovarian cancer: a meta-analysis of prospective observational studies. *BMC Public Health*. 2015;15:223
- [24] Cook LS, Leung AC, Swenerton K, Gallagher RP, Magliocco A, Steed H, et al. Adult lifetime alcohol consumption and invasive epithelial ovarian cancer risk in a population-based case-control study. *Gynecol Oncol*. 2016;140(2):277-284
- [25] Riman T, Dickman PW, Nilsson S, Nordlinder H, Magnusson CM, Persson IR. Some life-style factors and the risk of invasive epithelial ovarian cancer in Swedish women. *Eur J Epidemiol*. 2004;19(11):1011-1019
- [26] Merritt MA, De Pari M, Vitonis AF, Titus LJ, Cramer DW, Terry KL. Reproductive characteristics in relation to ovarian cancer risk by histologic pathways. *Hum Reprod*. 2013;28(5):1406-1417
- [27] Whiteman DC, Murphy MF, Cook LS, Cramer DW, Hartge P, Marchbanks PA, et al. Multiple births and risk of epithelial ovarian cancer. *J Natl Cancer Inst*. 2000;92(14):1172-1177