

Modelling Benign Ovarian Cyst Risk Factors and Symptoms via Log-Linear Model

Siti Zulaikha Mohd Jamaludin¹, Mohd Tahir Ismail¹, Mohd Shareduwan Mohd Kasihmuddin¹, Mohd. Asyraf Mansor², Siti Noor Farwina Mohamad Anwar Antony^{1*} and Adnin Adawiyah Makhul³

¹*School of Mathematical Sciences, Universiti Sains Malaysia, 11800, Penang, Malaysia*

²*School of Distance Education, Universiti Sains Malaysia, 11800, Penang, Malaysia*

³*Pusat Penyelidikan Klinikal, Hospital Sultan Ismail, Taman Mount Austin, 81100 Johor Bahru, Johor, Malaysia*

ABSTRACT

Ovarian cancer among women is known as “The Silent Killer”. It is caused by the malignant ovarian cyst, which can spread to other organs if it is not treated at an early stage. Some are benign ovarian cyst which can be treated through medical procedures such as laparoscopic and laparotomy. The type of medical procedure that the patients have to undergo depends on the size of cyst. A few risk factors that can cause benign ovarian cyst are age, pregnancy, menopause and menstrual cycle. Apart from that, there are a few symptoms of benign ovarian cyst which are fever, nausea and abdominal pain, abdominal distension, dysmenorrhea and intermenstrual bleeding. The association between these 12 discrete categorical data variables (factors, symptoms, treatment and size) are measured using the log-linear analysis in this study. According to the analysis, the patients who have large benign ovarian cyst need laparoscopic procedure, while those with smaller cyst need

either laparotomy procedure or they do not have to undergo any surgery at all. Among all of the factors, menopause gives the highest risk factor of benign ovarian cyst, followed by age, pregnancy and menstrual cycle. Meanwhile, the interaction between nausea, abdominal pain and intermenstrual bleeding give the highest symptom rate to the benign ovarian cyst.

Keywords: Abdominal pain, benign ovarian cyst, fever, log-linear analysis, menopause, pregnancy

ARTICLE INFO

Article history:

Received: 02 February 2021

Accepted: 24 May 2021

Published: 31 July 2021

DOI: <https://doi.org/10.47836/pjst.29.3.26>

E-mail addresses:

szulaikha.szmj@usm.my (Siti Zulaikha Mohd Jamaludin)

m.tahir@usm.my (Mohd Tahir Ismail)

shareduwan@usm.my (Mohd Shareduwan Mohd Kasihmuddin)

asyrafiman@usm.my (Mohd. Asyraf Mansor)

farwina@usm.my (Siti Noor Farwina Mohamad Anwar Antony)

adninadawiyah@gmail.com (Adnin Adawiyah Makhul)

* Corresponding author

INTRODUCTION

Around the world, ovarian cyst causes the most hospitalizations and surgical procedures among women (Farahani & Datta, 2016). It can cause cancer and has been coined as a “The Silent Killer”, as it can lead to more deaths than any other gynaecological cancers (Ahmad & Arslan, 2015; Al-Azri et al., 2018). A cyst is a pocket or fluid-filled sac that forms in female sex cell-producing organs called ovaries. Normally, the size of a cyst would be smaller than 1 centimeter in diameter, which is it could be as small as a pea. However, the size of the cyst can also be as large as an orange which may need to be removed (Shiota et al., 2012; Jha et al., 2019). The ovarian cyst is categorized into three main types: malignant, functional and benign. Benign cases can be defined as the cyst that consists of is non-cancerous cells and is not harmful to the body (Telli et al., 2013). Sanersak et al. (2006) found that majority of the ovarian cyst cases are benign ovarian cyst. It has been reported that around 70% of the total ovarian cyst masses are benign ovarian cyst, 24% are functional cyst and 6% are malignant cyst (Rofe et al., 2013).

In gynaecology, a benign ovarian cyst is regularly viewed with the most precise surgical excision (Mohamed et al., 2016). Most of the ovarian cyst patients need to go through primary surgical procedure either laparoscopic or laparotomy (Wu et al., 2013). Cyst with 5 diameter centimetres can be removed with recent advances in minimal surgery procedure known as laparoscopy. Hizkiyahu et al. (2019) considered laparoscopic as minimal surgery due to less pain, shorter duration of hospitalization, lower infection rates and shorter recovery period. However, in some cases, the patient can recover without any surgical treatment throughout their annual ultrasound scan (Akkoyun & Gülen, 2012). The use of medications can help to control the cyst’s growth and lower the pain especially when the cysts are small in size. According to Medeiros et al. (2009), if the cyst has a diameter that is larger than 6 centimetres and it is considered cancerous, then laparotomy surgery will be proposed. However, there are cases where the large size of ovarian cyst can be removed via laparoscopic (Hizkiyahu et al., 2019).

The occurrence of ovarian cyst among female is associated with numerous symptoms and risk factors. Over the years, medical practitioners such as doctors and specialists have relied on the late symptoms or factors to give an accurate diagnosis. According to Udomsikul et al. (2020), potential symptoms such as abdominal pain, breast tenderness, abnormal bleeding, dysmenorrhea, nausea or vomiting and abdominal bloating lead to benign ovarian cyst. However, the symptoms of ovarian cyst vary according to the risk factors (Mandai et al., 2012). Mukhopadhyay et al. (2016) and Gameraddin et al. (2018) evaluated the risk factors of benign ovarian cyst using chi-square analysis. They found that the potential risk factors associated with benign ovarian cyst were age, menstrual cycle, pregnancy, and menopause stage. Therefore, all these potential risk factors and symptoms could help doctors to predict the size of ovarian cyst and decide on the type of treatment

a patient might need. A specific data-adaptive model is required to extract the relationship of these factors that contribute to ovarian cyst. Thus, this study employs the log-linear model due to the effective feature selection mechanism to analyze specific symptoms and risk factors that contribute to benign ovarian cyst.

Numerous studies have employed the log-linear model to capture the behaviour of the medical data. Milewska et al. (2018) proposed log-linear analysis to estimate the chance of pregnancy among couples who implemented Assisted Reproductive Technology (ART) insemination. The proposed model provided a promising interaction extraction that predicts the chance of achieving a clinical pregnancy. In another context, Vilsen et al. (2019) utilized log-linear analysis to predict Lithium-ion resistance. In this model, log-linear is modelled as a function of State of the Art Charge (SOC) for lithium battery. The model was fitted to the extracted internal resistance and was reported to be consistent with the conventional internal resistance. The log-linear approach is also beneficial in the field of macroeconomics. Eggertsson and Singh (2019) utilize log-linear as a prediction model for the Keynesian model. Although their work revealed some limitations in log-linear model, the proposed exact non-linear model revealed the same predictions capability as in several policies such as government spending and tax multipliers. In a recent development, the log-linear model has been used to estimate the progression of the COVID-19 infection in Tamil Nadu (Bhaskar et al., 2020). Two separate log-linear models were fitted to model the growth phase before the peak and the decay phase after the peak. As a result, both models revealed the effectiveness of control measures by the government. It is to be noted that the common contribution of all the above studies emphasize the log-linear model as a good prediction model in various perspectives. Unfortunately, the application of log-linear with logic mining in predicting specific cancer disease is not properly studied. The closest model that is associated with cancer disease is the primary liver cancer where the log-linear capture and recapture model has been utilized to estimate a more correct incidence of underreporting for hepatocellular carcinoma (Törner et al., 2017). In addition, there is a work by Jamaludin et al. (2020), emphasizing the logistic regression and log-linear analysis in benign ovarian cyst. However, there is no recent effort to diagnose patients with benign ovarian cyst using log-linear model and logic mining approach. Moreover, the standard prediction model such as machine learning approach may become infeasible in terms of risk factors selection; thus, we proposed the standalone log-linear model to cope with the benign ovarian cyst detection task on risk factors and symptoms, separately. From the log-linear model, assumptions can be obtained for possible prediction among all the contributing attributes.

Hence, the objectives of this paper are as follows:

1. To measure the association between the types of operation and the size of a benign ovarian cyst.
2. To build a suitable model for risk factors and symptoms of a benign ovarian cyst.

The remaining part of this paper is as follows: Section 1 explores the uses of log-linear model in studies of benign ovarian cyst; Section 2 explains the material and method in detail; Section 3 discusses the result and discussion thoroughly and Section 4 gives the concluding remarks.

MATERIALS AND METHOD

General Overview of Log-linear Analysis

Log-linear analysis is a multi-dimensional data analysis model that is commonly known as a log-linear model (Zhu et al., 2006). According to McCullagh and Nelder (1989), the log-linear model is defined by assuming a Poisson distribution (with mean μ). This method is favoured when all variables of interest are presented on a qualitative scale and discrete in nature, while their relationships are presented in the form of a contingency table. It is to be noted that, the significance testing of main and interaction effects can be conducted using either the Wald test or the likelihood ratio test (Wiedermann & von Eye, 2020). According to Cox and Hinkley (1979), the Wald test and the likelihood ratio test are asymptotically equivalent as the sample size approaches infinity. An exception for small sample size where the likelihood ratio test may be preferred (Agresti, 2003).

Formulation of Log-linear Model

Analytically, the log-linear model is defined as an expression of expected frequencies (μ_{ij}) in the form of a function of parameters that represent the characteristics of discrete variables and the interactions taking place between them. According to Milewska et al. (2018), log-linear will choose the model with the lowest possible number of parameters. The model that is usually applied in the log-linear analysis is a saturated model, which includes model components. Model components are the number of main effects and interactions in the model. A saturated model always best represents the data and is the least parsimonious as every component is included. In this perspective, log-linear is also characterized by a good fit to data (Agresti, 2003). Consider two discrete variables A and J with their empirical frequencies n_{ij} in a contingency table with I row and J columns. According to Agresti (2003), the additive model for A and B is given as Equation (1):

$$\ln(\mu_{ij}) = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB} \quad (1)$$

where μ_{ij} is an expected values with regards to variable A and B . The symbol λ_i^A and λ_j^B represent the effect of i -th category of variable A and j -th category of variable B , respectively. Meanwhile, the symbol λ_{ij}^{AB} represents the effect of both i -th category of

variable A and j -th category of variable B . It is to be noted that, logarithmic mean μ for all cells is given as Equation (2):

$$\lambda = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \ln \mu_{ij} \quad (2)$$

It is worth mentioning that the parameters of any log-linear model must satisfy the following Equation (3):

$$\sum_{i=1}^I \lambda_i^A = \sum_{j=1}^J \lambda_j^B = \sum_{i=1}^I \lambda_{ij}^{AB} = \sum_{j=1}^J \lambda_{ij}^{AB} = 0 \quad (3)$$

where $\mu_{ij} > 0$. The log-linear model is utilized in this paper due to the nature of the categorical data. The categorical data analyzed by the log-linear model can provide us the information on association among the variables of the dataset. Log-linear is also a useful technique for modelling the count of μ_{ij} in the tables of categorical data with three and more dimension. Since more than two variables involve in this study, log-linear is considered practical for the medical dataset where the associations between discrete categorical data are to be analysed.

Data Analysis Setup

A cross-sectional study was conducted in which the data was collected based on 201 patients who were detected with ovarian cyst from January 2016 to December 2016 in Hospital Sultan Ismail, Johor Bahru, Malaysia. However, the inclusion criteria in this study identified 108 patients, who were diagnosed with benign ovarian cyst. The exclusion criteria eliminated 93 patients who were not diagnosed with benign ovarian cyst. The data was analysed by using IBM SPSS Statistic Version 24 in Windows 10 with computer specifications Intel CORE i5 processor, RAM 16GB. Each data simulation was conducted in single simulation rather than parallel simulation. Single simulation was chosen due to parallel simulation can lead to bad sector which will affect the final result (Kasihmuddin et al., 2019). In this study, we employed 95% confidence interval in the simulation where the value is ideally selected based on previous study by Milewska et al. (2018). This dataset has no missing entries and must be represented in the form of binary. In addition, log-linear model adopted unlimited ways of interaction among the risk factors and symptoms of benign ovarian cyst.

Benign Ovarian Cyst Dataset

They were 12 factors of interest used in this study: size of benign ovarian cyst, type of operation, risk factors and symptoms. For the risk factors, we emphasized several factors such as age, pregnancy, menopause and menstrual cycle. Meanwhile, as of factors for symptoms, we emphasized on fever, nausea, abdominal pain, abdominal distension, dysmenorrhea and intermenstrual bleeding. These variables were selected because the previous literature (Ahmad & Arslan, 2015; Al-Azri et al., 2018) utilizes the same variables. All these variables are important to be assessed statistically based on the clinical reports, which are obtained from the Hospital Sultan Ismail, Johor Bahru, Malaysia for the year 2016. This study is an extension of the study by Jamaludin et al. (2020) where all the variables interest in this study is critically analyzed by using log-linear analysis described in the previous section. Table 1 describes the information of the dataset used in this study.

Table 1
Summary of dataset for risk factor

Factor	Variable	Description	Scale	Unit
	size	Size of benign ovarian cyst	Nominal	0 = Small 1 = Large
	operation	Type of operation	Nominal	0 = No operation 1 = Laparotomy 2 = Laparoscopic
	age	Age of patient	Nominal	0 = 41 years old and above 1 = 40 years old and below (Zhu et al., 2019)
Risk Factor	pregnancy	The patient is pregnant or not	Nominal	0 = Yes 1 = No
	menopause	The menopause stage of patient	Nominal	0 = Yes 1 = No
	menscycle	The normality of menstrual cycle of patient	Nominal	0 = Not normal 1 = Normal
Symptom	fever	Symptoms of fever absent or not	Nominal	0 = Yes 1 = No
	nausea	Symptom of nausea is absent or not	Nominal	0 = Yes 1 = No
	abdopain	The patient have abdominal pain or not	Nominal	0 = No 1 = Yes
	abdodisten	Symptom of abdominal distension is absent or not	Nominal	0 = Yes 1 = No
	dysmenorrhoea	Symptoms of dysmenorrhoea is absent or not	Nominal	0 = Yes 1 = No
	bleeding	The patient have intermenstrual bleeding or	Nominal	0 = Yes 1 = No

RESULT AND DISCUSSION

This section presents results of log-linear analysis for two key components of the datasets which are risk factors and symptoms. Both analysis are made based on K-way effects, partial association and odds ratio. In order to visualize the datasets discussed in the previous session, this study proposed mosaic plot to demonstrate the composition of the type of treatment with respect to the size of the benign ovarian cyst. In terms of the frequency of the treatment, contingency table has been proposed to reveal the total number of patients that should undergo treatment. Note that the chi-square test will be utilized to test the independence of the variables in contingency table. By analysing the partial association and odds ratio for each risk factors and symptoms, the study can determine the association of the variables for main effects and interaction. This information is required to build a suitable model from the log-linear assumption for risk factors and symptoms of a benign ovarian cyst.

Mosaic Plot

According to Figure 1, the red section refers to the patients with large benign ovarian cyst, while the green section refers to the patients with small benign ovarian cyst. Based on Figure 1, most of the patient with small benign ovarian cyst have to undergo laparoscopic procedure or they do not have to undergo any operation procedure. Meanwhile, patients with large benign ovarian cyst will undergo laparotomy procedure.

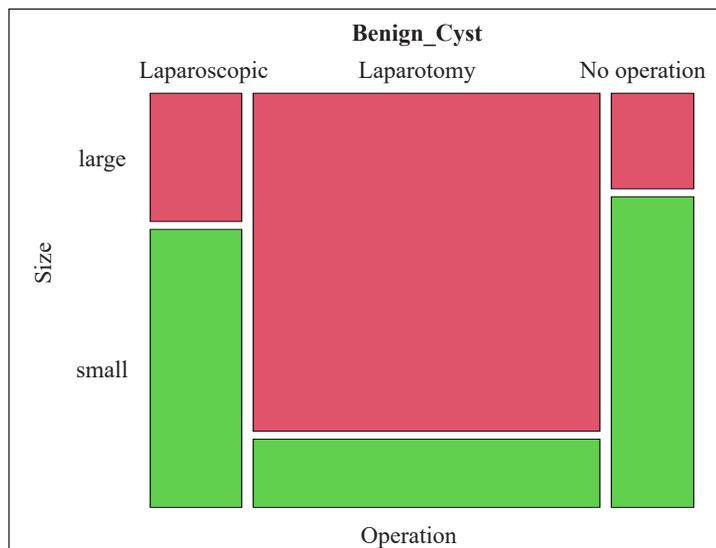


Figure 1. Mosaic plot of type of treatment by size of benign ovarian cyst

Contingency Table

From cross tabulation in Table 2, patients with small cyst had about 31.6% tendency of having laparoscopic surgery and about 34.2% for not having any operation or laparotomy surgery. For patient with large benign ovarian cyst, the probability to undergo laparoscopic was quite large which was 46.7% and only 11.0% of patient does not have to undergo any operation.

Table 2
Contingency table of type of treatment by size of benign cyst

Size		Type of Treatment			TOTAL
		No Operation	Laparoscopic	Laparotomy	
Small	Count	13	12	13	38
	% within Size	34.2	31.6	34.2	100.0
Large	Count	4	60	6	70
	% within Size	11.0	46.7	12.3	100.0
TOTAL	Count	17	72	19	108
	% within Size	15.7	66.7	17.6	100.0

Chi-Square Test

Based on the Table 3, the Chi-square value was 34.214 and the significance value was less than 0.05. It was concluded that there was an association between the size of benign ovarian cyst and the type of treatment. In other words, if the patient had a large size of cyst, they would have to undergo laparotomy surgery.

Table 3
Table of Chi-Square test of association

	Value	Exact Sig.
Pearson Chi-Square	34.214	0.000

Log-linear Analysis for Risk Factors

There are four variables in consideration when conducting log-linear analysis for demographic data which are age, pregnancy, menopause and menstrual cycle.

K-way Effects. The K-way effect was used to analysed which components of the model can be removed.

Table 4 manifests the K-way effects component, where the possible effect components obtained by the model can be filtered. Based on the first row ($K = 1$) with degree freedom of 4, the main effect terms for both likelihood-ratio and Pearson test are statistically significant at 0.05 level of significance. This implies that the main effect terms are included in this model. The significant p -value for the two-way effect ($K = 2$) indicates that two-factors

Table 4
K-Way effects component of the model for risk factor

K	df	Likelihood Ratio		Pearson	
		χ^2	<i>p</i> -value	χ^2	<i>p</i> -value
1	4	108.507	0.000	172.975	0.000
2	6	64.440	0.000	107.633	0.000
3	4	5.327	0.255	5.350	0.253
4	1	0.206	0.650	0.116	0.733

interaction would be included in the model. In conclusion, a good fitted model for risk factor can have up to two interaction terms, as supported by the work of Fatin et al. (2020).

Partial Association. Although the two-way interactions significantly affect the model, the specific two-way interaction among the risk factors need to be determined statistically via log-linear model.

Table 5 manifested the partial association obtained by assessing all risk factor variables. Consequently, based on Table 5, the Pearson χ^2 test were reported to be statistically significant for several interactions such as age*pregnancy and age*menopause. This is due to the optimal *p*-value for these two interactions ($p \leq 0.05$). Furthermore, the similar *p*-value have been reported in other parameters ($p \leq 0.05$) such as the age, pregnancy and menopause that statistically justify the significance of the effects.

Table 5
Table of partial association for risk factor

Effect	df	Partial χ^2	<i>p</i> -value
age*pregnancy*menopause	1	.080	.777
age*pregnancy*menscycle	1	1.256	.262
age*menopause*menscycle	1	.004	.949
pregnancy*menopause*menscycle	1	3.695	.055
age*pregnancy	1	25.032	.000
age*menopause	1	11.904	.001
pregnancy*menopause	1	.366	.545
age*menscycle	1	.517	.472
pregnancy*menscycle	1	.536	.464
menopause*menscycle	1	1.234	.267
age	1	35.303	.000
pregnancy	1	49.245	.000
menopause	1	20.233	.000
menscycle	1	3.725	.054

General Log-linear. Table 6 shows the summary of parameter estimate of variables that significantly effects the model. Note that, the parameters in Table 6 will be estimated by using z -score instead of χ^2 test. This is due to the usefulness of z -score in comparing the effect of variables (Field, 2013). High magnitude of z -score indicates that the effect of the parameters is more significant compared to low value of z -score. Based on Table 6, menopause ($z = 4.626$) is the most important effect in the model and followed by age, pregnancy, menscycle and age*menopause.

Note that, the obtained log-linear model that evaluate the risk factors for benign ovarian cyst is given as Equation (4):

$$\ln \mu_{hijk} = \lambda + \lambda_{A(k)} + \lambda_{B(l)} + \lambda_{C(m)} + \lambda_{D(n)} + \lambda_{AC(km)} \quad (4)$$

where,

A : Age (0 = 41 years old and above, 1 = 40 years old and below); B : Pregnancy (0 = yes, 1 = no); C : Menopause (0 = yes, 1 = no); D : Menstrual cycle (0 = not normal, 1 = normal).

Table 6
Summary of parameter estimate for risk factor

Parameter	Estimate	Std. Error	z	p-value	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	3.795	.150	25.319	.000	3.502	4.089
[age = 0]	-2.879	.650	-4.430	.000	-4.153	-1.605
[pregnancy = 0]	-2.879	.650	-4.430	.000	-4.153	-1.605
[menopause = 0]	-2.091	.452	-4.626	.000	-2.977	-1.205
[menscycle = 0]	-0.638	.255	-2.504	.012	-1.138	-.139
[age = 0] * [menopause = 0]	2.091	1.002	2.086	.037	.127	4.055

Odds Ratio. Odds ratios are used to determine the association of the variables for main effects and two interaction, since the the model for risk factor is obtained with two-way effect.

According to Table 7, patients with the age of 41 and older has a higher odds of benign ovarian cyst by 0.06 times compared to 40 years old and below. Similar odds ratio is reported for patient that is pregnant compared to non-pregnant women. As for other parameters such as menopause and menscycle, the odds ratio for benign ovarian cyst with age 40 years old is 0.12 and 0.52 respectively. Meanwhile, that the odds of benign ovarian cyst patients with age 41 years old and above and menopause are 8.09 times higher than the odds of patients with age 40 years old and below and not menopause.

Table 7
Summary of odds ratio for risk factor

Parameter	Estimate	Odds Ratio
[age = 0]	-2.879	0.0562
[pregnancy = 0]	-2.879	0.0562
[menopause = 0]	-2.091	0.1237
[mencycle = 0]	-0.638	0.5283
[age = 0] * [menopause = 0]	2.091	8.0930

Log-linear Analysis for Symptoms

There are six variables in conducting log-linear analysis for symptoms data which are fever, nausea, abdominal pain, abdominal distension, dysmenorrhea and intermenstrual bleeding.

K-way Effects. The K-way effect was used to analyze which components of the model can be removed.

Table 8
K-Way Effects Component of the Model for Symptom

K	df	Likelihood Ratio		Pearson	
		χ^2	p-value	χ^2	p-value
1	6	197.876	.000	471.381	.000
2	15	127.358	.000	263.375	.000
3	20	58.437	.000	55.269	.000
4	15	0.664	.917	0.345	.988
5	6	0.000	.960	0.000	.993
6	1	0.000	1.000	0.000	1.000

Based on Table 8, for the case $K = 1$, the main effect terms for both likelihood-ratio and Pearson χ^2 test are statistically significant at 0.05 level of significant. This implies that the main effect terms are included in this model. The significant p -value for the two-way effects ($K = 2$) and three-way effects ($K = 3$) indicate that both two-factors and three-factors interaction would be included in the model. In conclusion, a good fitted model can have up to three interaction terms.

Partial Association. As mentioned in Table 8, removing all three-way interactions significantly affects the model. Table 9 shows the partial association among all symptom variables.

Table 9
Table of partial association for symptom

Effect	df	Partial χ^2	<i>p</i> -value
fever*abdopain*dysmenorrhea	1	6.113	.013
nausea*abdopain*bleeding	1	37.073	.000
nausea*abdodisten	1	17.661	.000
abdopain*dysmenorrhea	1	25.022	.000
abdodisten*bleeding	1	14.379	.000
dysmenorrhea*bleeding	1	14.598	.000
fever	1	83.085	.000
nausea	1	18.458	.000
abdopain	1	9.625	.002
abdodisten	1	12.233	.000
dysmenorrhea	1	46.220	.000
bleeding	1	28.255	.000

Pearson χ^2 as shown in Table 9 indicated that test of partial association for symptoms were significant for several interactions such as nausea*abdopain*bleeding, nausea*abdodisten, and dysmenorrhea*bleeding interactions since the *p*-value for these two and three-interactions were less than 0.05. The main effects of fever, nausea, abdominal pain, abdominal distension, dysmenorrhea and intermenstrual bleeding were significant.

General Log-linear. The summary of parameter estimate of variables that significantly effects the model can be viewed in Table 10. It can be concluded that the nausea*abdopain*bleeding ($z = 4.178$) was the most important effect in the model and followed by nausea*bleeding, nausea, fever, nausea*abdopain, dysmenorrhea, bleeding, abdopain*bleeding, abdopain, nausea*abdopain*dysmenorrhea, fever*nausea, abdopain*dysmenorrhea and fever*nausea*abdopain.

Hence, the final log-linear model for symptom of benign ovarian cyst is defined as Equation (5):

$$\begin{aligned} \ln \mu_{opqst} = & \lambda + \lambda_{E(o)} + \lambda_{F(p)} + \lambda_{G(q)} + \lambda_{I(s)} + \lambda_{J(t)} \\ & + \lambda_{FJ(pt)} + \lambda_{FG(pq)} + \lambda_{GJ(qt)} + \lambda_{EF(op)} + \lambda_{GI(qs)} \quad (5) \\ & + \lambda_{FGJ(pqt)} + \lambda_{EFG(opq)} + \lambda_{FGI(pqs)} + \lambda_{EFG(opq)} \end{aligned}$$

where,

E: Fever (0 = yes, 1 = no); *F*: Nausea (0 = yes, 1 = no); *G*: Abdominal pain (0 = no, 1 = yes); *I*: Dysmenorrhea (0 = yes, 1 = no); *J*: Intermenstrual bleeding (0 = yes, 1 = no).

Table 10
Summary of parameter estimate for symptom

Parameter	Estimate	Std. Error	z	p-value	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	3.068	.216	14.226	.000	2.645	3.491
[fever = 0]	-2.663	.844	-3.153	.002	-4.318	-1.007
[abdpain = 0]	-3.761	1.431	-2.629	.009	-6.565	-.957
[dysmenorrhea = 0]	-1.815	.576	-3.149	.002	-2.945	-.686
[bleeding = 0]	-1.815	.576	-3.149	.002	-2.945	-.686
[fever = 0] * [nausea = 0]	2.663	1.230	2.165	.030	.252	5.074
[nausea = 0] * [abdpain = 0]	5.096	1.597	3.190	.001	1.965	8.227
[nausea = 0] * [bleeding = 0]	3.250	.910	3.573	.000	1.467	5.033
[abdpain = 0] * [dysmenorrhea = 0]	3.425	1.653	2.072	.038	.185	6.664
[abdpain = 0] * [bleeding = 0]	4.760	1.561	3.049	.002	1.700	7.820
[fever = 0] * [nausea = 0] * [abdpain = 0]	-5.607	2.760	-2.031	.042	-11.017	-.197
[nausea = 0] * [abdpain = 0] * [dysmenorrhea = 0]	-4.760	2.138	-2.226	.026	-8.950	-.569
[nausea = 0] * [abdpain = 0] * [bleeding = 0]	-8.041	1.925	-4.178	.000	-11.813	-4.268

Odds Ratio. Odds ratios are used to determine the association of the variables for main effects, two interaction and three interaction since the best model obtained is the model with three-way effect.

The summary of odds ratio as shown in Table 11 indicates that the odds of benign ovarian cyst patients who have fever are 0.07 times higher than the odds of patients who not have fever. This odds ratio also indicates that the odds of benign ovarian cyst patients who have no abdominal pain symptom are 0.02 times higher than the odds of patients who have abdominal pain. Other than that, the odds ratio indicates that the odds of benign ovarian cyst patients who experience dysmenorrhea and intermenstrual bleeding are 0.16 times higher than the odds of patients who did not experience it. It also indicate the odds of benign ovarian cyst patients who have fever and nausea are 14.33 times higher than the odds of patients who did not experience any fever and nausea. Lastly, the odds ratio able to indicate that the odds of benign ovarian cyst patients who have fever, abdominal pain and intermenstrual bleeding are 0.0003 times higher than the odds of patients who not have fever, abdominal pain and intermenstrual bleeding.

A study done by Al-Azri et al. (2018) which explored knowledge, risk factors, symptoms and time taken to seek early medications for ovarian cancer has agreed that abdominal pain and dysmenorrhea are the important factors for an ovarian cancer. Another interesting note, the finding of this study has a good agreement with the previous study of

Table 11
Summary of odds ratio for symptom

Parameter	Estimate	Odds Ratio
[fever = 0]	-2.663	0.0674
[abdominal pain = 0]	-3.761	0.0233
[dysmenorrhea = 0]	-1.815	0.1628
[bleeding = 0]	-1.815	0.1628
[fever = 0] * [nausea = 0]	2.663	14.3392
[nausea = 0] * [abdominal pain = 0]	5.096	163.3671
[nausea = 0] * [bleeding = 0]	3.250	25.7903
[abdominal pain = 0] * [dysmenorrhea = 0]	3.425	30.7226
[abdominal pain = 0] * [bleeding = 0]	4.760	116.7459
[fever = 0] * [nausea = 0] * [abdominal pain = 0]	-5.607	3.672×10^{-3}
[nausea = 0] * [abdominal pain = 0] * [dysmenorrhea = 0]	-4.760	8.566×10^{-3}
[nausea = 0] * [abdominal pain = 0] * [bleeding = 0]	-8.041	3.22×10^{-4}

Udumsinkul et al. (2020) where they implemented logistic regression to identify risk factors for endometriotic-cyst associated with ovarian cancer. In a good agreement by Huang et al. (2017), the most common symptom of ovarian torsion was pelvic pain and nausea. Although the mentioned studies supported our findings, the result from this study cannot be compared with the mentioned study due to properties of the datasets. Robust comparison can be made if all the dataset has been taken from the same hospital (Jamaludin et al., 2020).

CONCLUSION

Among 108 patients who were diagnosed benign ovarian cyst, about 38% and 70% of patients were diagnosed with benign ovarian cyst of small and large sizes, respectively. The contingency table was presented to measure the association between types of operation and the size of benign ovarian cyst. The results found that patients who had large size of benign ovarian cyst has to undergo laparoscopic surgery, while patients who had small size has to undergo either no operation or laparotomy surgery. Analyzing the odds ratio from each symptoms and risk factors are required to build a suitable model for risk factors and symptoms of a benign ovarian cyst. From log-linear analysis, it is concluded that menopause was the most dominant risk factor of benign ovarian cyst. Whereas, interaction effect of nausea, abdominal pain and intermenstrual bleeding were the most important effect for symptoms of benign ovarian cyst. By considering more criteria of risk factors and symptoms, our proposed model could generate a more generalized rule for classification that have been highlighted in several studies. Therefore, it can be concluded that this study achieved all its objectives.

The limitations of this study can be seen in the cross-sectional design, in which the data was collected at one time. Furthermore, this study employed the retrospective observational study where the data was extracted and limited from the patient's history. In addition, there was a possibility of sampling bias, where the data of patients might be redundant, so the system of patient's database should be enhanced via better technology. The documentation of the patient's history may not be too accurate since this study utilized the secondary data. The secondary data is limited and prone to the possible errors and bias of the data entries. The aforementioned limitation might originate from the medical practitioners during data collection and interpretation. Moreover, the accuracy of the symptoms or risks factors possessed by the patients can be limited in secondary data.

For future work, the simulation would include more patients that have other gynecologic problems since different problems may bring different risk factors and symptoms. This perspective will improve the accuracy of diagnosis and provide holistic findings that will benefit the medical practitioners. Moreover, a larger sample size (more than 1000) is required to investigate the more significant risk factors and symptoms of benign ovarian cyst (Jamaludin et al., 2020). By considering more criteria of risk factors and symptoms, our proposed model can generate more generalized rule for classification task that has been highlighted in several studies such as Zamri et al. (2020) and Alway et al. (2020).

ACKNOWLEDGEMENT

The authors would like to express special dedication to all of the researchers from AI Research Development Group (AIRDG) for the continuous support. This research was funded by Univeriti Sains Malaysia (USM), grant number 304/PMATHS/6315390.

REFERENCES

- Ahmad, G., & Arslan, M. (2015). Unique expression of 35 KDa protein in serum and cystic fluid of women with malignant ovarian cyst substantiates its role in disease progression. *Asian Pacific Journal of Reproduction*, 4(4), 298-304. <https://doi.org/10.1016/j.apjr.2015.07.007>
- Al-Azri, M., Al-Saidi, M., Eman, A. M., & Panchatcharam, S. M. (2018). Awareness of risk factors, symptoms and time to seek medical help of ovarian cancer amongst Omani women attending teaching hospital in Muscat Governorate, Oman. *Asian Pacific journal of cancer prevention: APJCP*, 19(7), Article 1833. <https://doi.org/10.22034/APJCP.2018.19.7.1833>
- Alway, A., Zamri, N. E., Kasihmuddin, M. S. M., Mansor, A., & Sathasivam, S. (2020). Palm oil trend analysis via logic mining with discrete Hopfield neural network. *Pertanika Journal of Science & Technology*, 28(3), 967-981.
- Agresti, A. (2003). *Categorical data analysis* (Vol. 482). John Wiley & Sons.

- Akkoyun, İ., & Gülen, S. (2012). Laparoscopic cystectomy for the treatment of benign ovarian cyst in children: An analysis of 21 cases. *Journal of Pediatric and Adolescent Gynecology*, 25(6), 364-366. <https://doi.org/10.1016/j.jpjg.2012.06.007>
- Bhaskar, A., Ponnuraja, C., Srinivasan, R., & Padmanaban, S. (2020). Distribution and growth rate of COVID-19 outbreak in Tamil Nadu: A log-linear regression approach. *Indian Journal of Public Health*, 64(6), 188-191. https://doi.org/10.4103/ijph.IJPH_502_20
- Cox, D. R., & Hinkley, D. V. (1979). *Theoretical statistics*. CRC Press.
- Eggertsson, G. B., & Singh, S. R. (2019). Log-linear approximation versus an exact solution at the ZLB in the New Keynesian Model. *Journal of Economic Dynamics and Control*, 105, 21-43. <https://doi.org/10.1016/j.jedc.2018.09.011>
- Farahani, L., & Datta, S. (2016). Benign ovarian cyst. *Obstetrics, Gynaecology & Reproductive Medicine*, 26(9), 271-275. <https://doi.org/10.1016/j.ogrm.2016.06.003>
- Fatin N. S. A., Norlida, M. N., & Siti, Z. M. J. (2020). The application of log-linear model to selected poison patients. *ASM Science Journal*, 13, 1-7. [https://doi.org/10.32802/asmscj.2020.sm26\(1.21\)](https://doi.org/10.32802/asmscj.2020.sm26(1.21))
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.
- Gameraddin, M. B., & Bashab, N. K. (2018). Characterisation of benign ovarian lesions among Sudanese women undergoing pelvic ultrasound scans: The impact of parity and age. *Journal of Clinical & Diagnostic Research*, 12(5), 6-10. <https://doi.org/10.7860/JCDR/2018/35107.11459>
- Hizkiyahu, R., Yahav, L., Yakovi, S., Davidesko, S., Abecassis, A., & Weintraub, A. Y. (2019). Short-and long-term outcomes of intraoperative spillage during laparoscopic removal of benign ovarian cyst. *Surgical Endoscopy*, 34(9), 3883-3887. <https://doi.org/10.1007/s00464-019-07154-6>
- Huang, C., Hong, M. K., & Ding, D. C. (2017). A review of ovary torsion. *Tzu-chi Medical Journal*, 29(3), 143-147. https://doi.org/10.4103/tcmj.tcmj_55_17
- Jamaludin, S. Z. M., Makhul, A. A., Kasihmuddin, M. S. M., Kustiani, A., Rahim, S. A., & Ismail, M. T. (2020). Modeling symptoms and risk factors towards size of benign ovarian cyst. *Proceedings of The 27th National Symposium On Mathematical Sciences (SKSM27), Malaysia*, 2266(1), Article 090010. <https://doi.org/10.1063/5.0018108>
- Jha, P., Shekhar, M., Goldstein, R., Morgan, T., & Poder, L. (2019). Size threshold for follow-up of postmenopausal adnexal cyst: 1 cm versus 3 cm. *Abdominal Radiology*, 45, 3213-3217. <https://doi.org/10.1007/s00261-019-02176-z>
- Kasihmuddin, M. S. M., Mansor, M., Basir, M. F. M., & Sathasivam, S. (2019). Discrete mutation Hopfield neural network in propositional satisfiability. *Mathematics*, 7(11), Article 1133. <https://doi.org/10.3390/math7111133>
- Mandai, M., Suzuki, A., Matsumura, N., Baba, T., Yamaguchi, K., Hamanishi, J., Yoshioka, Y., Kosaka, K., & Konishi, I. (2012). Clinical management of ovarian endometriotic cyst (chocolate cyst): Diagnosis, medical treatment, and minimally invasive surgery. *Current Obstetrics and Gynecology Reports*, 1(1), 16-24. <https://doi.org/10.1007/s13669-011-0002-3>

- McCullagh, P., & Nelder, J. A. (1989). Binary data. In *Generalized linear models* (pp. 98-148). Springer. <http://dx.doi.org/10.1007/978-1-4899-3242-6>
- Medeiros, L. R., Rosa, D. D., Bozzetti, M. C., Fachel, J. M., Furness, S., Garry, R., Rosa, M. I., & Stein, A. T. (2009). *Laparoscopy versus laparotomy for benign ovarian tumour*. John Wiley & Sons, Ltd. <https://doi.org/10.1002/14651858.CD004751.pub3>
- Milewska, A. J., Citko, D., Jankowska, D., Milewski, R., Konończuk, K., Więsak, T., Morgan, A., & Milewski, R. (2018). The use of log-linear analysis for pregnancy prediction. *Studies in Logic, Grammar and Rhetoric*, 56(1), 7-18. <https://doi.org/10.2478/slgr-2018-0037>
- Mohamed, A. A., Al-Hussaini, T. K., Fathalla, M. M., El Shamy, T. T., Abdelaal, I. I., & Amer, S. A. (2016). The impact of excision of benign nonendometriotic ovarian cyst on ovarian reserve: A systematic review. *American journal of obstetrics and gynecology*, 215(2), 169-176. <https://doi.org/10.1016/j.ajog.2016.03.045>
- Mukhopadhyay, A., Shinde, A., & Naik, R. (2016). Ovarian cyst and cancer in pregnancy. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 33, 58-72. <https://doi.org/10.1016/j.bpobgyn.2015.10.015>
- Rofe, G., Auslender, R., & Dirnfeld, M. (2013). Benign ovarian cyst in reproductive-age women undergoing assisted reproductive technology treatment. *Open Journal of Obstetrics and Gynecology*, 3(7A), Article 36929. <https://doi.org/10.4236/ojog.2013.37A1005>
- Sanersak, S., Wattanakumtornkul, S., & Korsakul, C. (2006). Comparison of low-dose monophasic oral contraceptive pills and expectant management in treatment of functional ovarian cyst. *Journal-Medical Association of Thailand*, 89(6), 741-747.
- Shiota, M., Kotani, Y., Umemoto, M., Tobiume, T., & Hoshiai, H. (2012). Study of the correlation between tumor size and cyst rupture in laparotomy and laparoscopy for benign ovarian tumor: Is 10 cm the limit for laparoscopy? *Journal of Obstetrics and Gynaecology Research*, 38(3), 531-534. <https://doi.org/10.1111/j.1447-0756.2011.01748.x>
- Telli, E., Oge, T., Ozalp, S. S., & Yalcin, O. T. (2013). Giant peritoneal inclusion cyst mimicking ovarian cyst. *Asian Pacific Journal of Reproduction*, 2(4), 333-334. [https://doi.org/10.1016/S2305-0500\(13\)60174-8](https://doi.org/10.1016/S2305-0500(13)60174-8)
- Törner, A., Stokkeland, K., Svensson, Å., Dickman, P. W., Hultcrantz, R., Montgomery, S., & Duberg, A. S. (2017). The underreporting of hepatocellular carcinoma to the cancer register and a log-linear model to estimate a more correct incidence. *Hepatology*, 65(3), 885-892. <https://doi.org/10.1002/hep.28775>
- Udomsinklul, P., Triratanachart, S., & Oranratanaphan, S. (2020). Risk factors for endometriotic-cyst associated ovarian cancer: A case controlled study. *Taiwanese Journal of Obstetrics and Gynecology*, 59(2), 269-274. <https://doi.org/10.1016/j.tjog.2020.01.016>
- Vilsen, S. B., Kaer, S. K., & Stroe, D. L. (2019). Predicting Lithium-ion battery resistance degradation using a log-linear model. In *2019 IEEE Energy Conversion Congress and Exposition (ECCE)* (pp. 1136-1143). IEEE Conference Publication. <https://doi.org/10.1109/ECCE.2019.8912770>
- Wiedermann, W., & von Eye, A. (2020). Log-linear models to evaluate direction of effect in binary variables. *Statistical Papers*, 61(1), 317-346. <https://doi.org/10.1007/s00362-017-0936-2>

- Wu, M. P., Wu, C. J., Long, C. Y., Ho, C. H., Huang, K. H., Chu, C. C., & Chou, C. Y. (2013). Surgical trends for benign ovarian tumors among hospitals of different accreditation levels: An 11-year nationwide population-based descriptive study in Taiwan. *Taiwanese Journal of Obstetrics and Gynecology*, *52*(4), 498-504. <https://doi.org/10.1016/j.tjog.2013.10.008>
- Zamri, N. E., Mansor, M., Kasihmuddin, M. S. M., Alway, A., Jamaludin, S. Z. M., & Alzaemi, S. A. (2020). Amazon employees resources access data extraction via clonal selection algorithm and logic mining approach. *Entropy*, *22*(6), Article 596. <https://doi.org/10.3390/e22060596>
- Zhu, B., Walter, S. D., Rosenbaum, P. L., Russell, D. J., & Raina, P. (2006). Structural equation and log-linear modeling: a comparison of methods in the analysis of a study on caregivers' health. *BMC Medical Research Methodology*, *6*(1), 1-14. <https://doi.org/10.1186/1471-2288-6-49>
- Zhu, D., Chung, H. F., Dobson, A. J., Pandeya, N., Giles, G. G., Bruinsma, F., Brunner, E., Kuh, D., Hardy, R., Avis, N. E., Gold, E. B., Derby, C. A., Matthews, K. A., Cade, J. E., Greenwood, D. C., Demakakos, P., Brown, D. E., Sievert, L. L., Anderson, D., ... & Mishra, G. D. (2019). Age at natural menopause and risk of incident cardiovascular disease: A pooled analysis of individual patient data. *The Lancet Public Health*, *4*(11), e553-e564. [https://doi.org/10.1016/S2468-2667\(19\)30155-0](https://doi.org/10.1016/S2468-2667(19)30155-0)