

A Comparison Between Probit And Logit Regression In Identifying Risk Factors Of Cervical Cancer In Public Hospital

Tan Li Jun, Suliadi Firdaus Sufahani

Faculty of Applied Sciences and Technology, University Tun Hussein Onn Malaysia, 84600 Pagih, Johor, Malaysia

* Corresponding author: aw170237@siswa.uthm.edu.my;suliadi@uthm.edu.my

ARTICLE HISTORY

Received: 3 March 2020
Accepted: : 28 March 2020
Published Online: 28 April 2020

KEYWORDS

Cervical cancer
Risk factor
Generalized linear model
Pap smear test

ABSTRACT

Cervical cancer(CC) is one of the cancer that easily to get in female. Globally around 569,847 new cervical cancer cases and 311,365 deaths in every year for 2018 are estimated by the Global Cancer Observatory (GLOBOCAN). It is the third leading cause of female cancer in Malaysia. In this review paper, the risk factor of CC was discussed. The related research of method was discussed and carried out which include Generalized linear model (GLMs), Case-control technique, Multiple linear regression (MLR). Previous researchers of research will be summaries into an overview form of tabular for making more easier to study.

1.0 INTRODUCTION

Every female has an organ of reproductive system in their body. In the organ, it mainly includes vulva, vagina, cervix, uterus, ovaries, endometrium, and fallopian tubes to form a complete function of reproductive system. Female reproductive system can produce ovum through ovaries. It is also a place where baby live from fertilization to mature fetus. Cervix is one of the parts in female reproductive system. It has placed the lower part of uterus which is in the middle of uterus and vagina. In the middle of cervix has a small hole, it can expand 10cm to allow deliver baby successfully. The function of cervix is to allow the sperm enter into fallopian tubes (F. Martyn, 2014), and the second function is to allow the blood flow from uterus to vagina.

Cancer is an abnormal cell which can spread to every parts of the body. Cell made up carbon, hydrogen and oxygen. Although it is tiny, it can bring huge effect to human. A cancer cell is from a normal cell turn into an abnormal cell. Cancer cell is developed when body control mechanism is stopped working. When cancer cells are gathered, it forms a mass of tissue which called tumor. Moreover, breast cancer and cervical cancer are the common cancer to women. In men, prostate cancer can be easily get. Based on statistics in Malaysia, around 37,000 new cases of cancer in every year and estimated 55,000 cases in 2030 (Malaymail, 2018). In underdeveloping countries, taking alcohol, smoking, lack of taking fruit and vegetables can increase the risk of getting cancer. Moreover, obesity and overweight are the risk factor of getting cancer in high income countries (Danaei et al.,

2005). It is very important to prevent cancer from daily lifestyle. Eating healthily and taking enough nutrition is very important to our health. It can reduce the risk of getting cancer.

However, an unhealthy cervix can easily get cervical cancer. Cervical cancer is the major cause of death for women in worldwide (Shin-je G., 2002). It is also a second largest cancer for women in worldwide. Human papillomavirus also known as HPV is virus that usually grown in cervix. In other word, cervical cancer is caused by HPV. HPV infection usually play a main role to cause most of the time in cervical cancer. Cervical cancer occur in the cell of cervix. Thus, it starts to occur when the cell of cervix change (American Cancer Society, 2016). When cervix start to have virus in the cell of cervix, the cell will change and getting serious through a period of time. When the infection is small, it will be more easier to treat. Due to the body of immune system, infection will not go long as immune system can fight for it. However, a long-term infection may lead to cervical cancer that immune system are not able to fight for it. There are two main type of cervical cancer for women. To know the type of cervical cancer can help doctor to give a treatment accurately toward the patient. The two mainly type of cervical cancer are squamous cell carcinoma and adenocarcinoma.

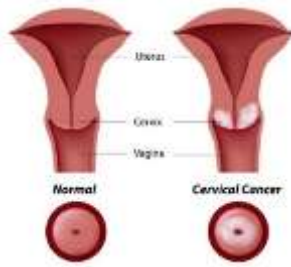


Figure 1. Cervix Normal and Cervical Cancer in Female Reproductive System

1.1 Research Objectives

The following are the objectives of this research:

- i. to identify the risk factors of female patients in Hospital Pakar Sultanah Fatimah (HPSF), Muar getting cervical cancer by using probit and logit regression.
- ii. to build a regression model that is able to be used to predict the probability of women getting cervical cancer by using the best method.
- iii. to investigate the relationship between cervical cancer and its risk factors by using the best method.

2.0 REVIEW OF LITERATURE

2.1. Risk factors of cervical cancer

Different categories of people will have different risk factors in each countries. Risk of CC occur in every races of woman. The risk of getting CC was related to economics, sexual different and social in different country (Zaher K., 2019). Some of direct factor could have association with CC such as (1)low level of education, (2) age group, (3) marital status, (4) scare for screening which include fear, embarrass (Patricia B., 2007). There were still a lot of factor can bring CC to women. CC occurred in underdeveloping and developing countries. CC was commonly occur in the age group of 35-65 years which the risk of getting CC increased according to its risk factor.

2.1.1. Having many sexual partners

In a survey of Low L. E. et al., 2012, the result showed that 884 out of 1392 (63.5%) respondents recognised who had many sexual partner will increase the risk of getting CC. When the number of sexual partner increase, women will increase the risk. Moreover, Agarwal et al., 1993 stated husband had sexual relationship before or after marriage are also a one of the risk to get CC toward women. When husband had three or more sexual partner, the risk of wife getting CC will also increase.

2.1.2. Smoking

Smoking was identified in many research as the one of the risk factor of getting CC. Based on the case-control study of Shields et al., 2004 stated that smoking was a risk factor which had a relationship between the increasing of taking cigarette and the duration of smoking. Over 50% of respondents were recognised smoking was the risk factor in United Kingdom of women (Low L. E. et al., 2012). However, one of the researcher found that only a few of women who were smoking had CC. Social and culture difference may affect the risk of getting CC (Abduljabbar et al., 2014).

2.1.3. Age group

Different age group affected women to get CC. CC commonly occurred the age group around 35 to 65 years old of women (Abduljabbar et al., 2014). Some of the research stated that the most frequency of age was above 50 years old (Sharma & Pattanshetty, 2018, Lai et al., 2009). Sharma & Pattanshetty, 2018 stated that the average age group of women to get CC was 51 to 60 year old. Low L. E. et al., 2012 found that British women who were younger age had low awareness of risk factor for CC compare to older women.

2.1.4. Family historical

Family historical was one of the factors to get CC. It had found significantly associated with CC (Sharma & Pattanshetty, 2018). It had more chance to get the disease compare with no family historical. Genes from family can cause family member to get CC.

2.2. Previous research method

Some research were carried out with different method of various areas. Risk factor of CC has been studied in different countries of various method. Some of the research using a few type of method to carried out an analysis which include Generalized linear model (GLMs), Case-control technique, Multiple linear regression (MLR).

2.2.1. Generalized linear model (GLMs)

Generalized linear model also known as probit model. It is an extension of general linear model (Turner H., 2008). General linear model is a linear regression model using in continuous respond variables and the distribution of residual need to be normal. However, probit model can be used for binary dependent variables with two or more independent variables which can consist of categorical and continuous variables (Dobson A., 2002). Distribution of residual can be non-normal distribution, such as binomial, poisson, exponential. The GLMs model is

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_n \quad (1)$$

Sabates & Feinstein, 2006 was used the techniques of GLMs to estimate the parameter of the model. It was to study the relationship between education and screening of CC. The qualification increased, the likelihood of women

did Pap smear test will also increase. This research also found that female who were smoking, the likelihood of doing Pap smear test were decreased. The chance of getting CC decrease to women who had higher education (Nesrin, 2011).

The research of Harmon et al., 1996 was studied the culture level of knowledge in Pap smear test, fear towards CC and screening. Ordinary least squares (OLS) regression and probit regression were conducted. Result of OLS regression found that Hispanic women were more educated, the fear of facing cancer was declined. While probit regression also found a significant variable of knowledge ($\beta=0.24$) towards Pap smear test. The increment of acculturation level was affected of knowledge level for Pap smear test.

Probit and logit models were used in Pourhoseingholi et al., 2008 to investigate the association between demographic factor and the type of Gastrointestinal Cancer (GC) and to compare GC with probit and logit models. It used cross-sectional study in selected the sample. The result showed gender and age gave an effect of colorectal cancer in logit model and probit model. This research concluded logit and probit model were similar from the result. Moreover, male had higher percentage to get colorectal cancer compare to female.

2.2.2. Case-control technique

Case-control study is a technique which used to evaluate the patient with or without disease to determine the risk factor of disease. The method is divided into two group which are cases (patient with disease) and controls (patient with no disease) (Maninder S., 2016). The research of Sharma P., 2017 was using case-control study to carried out. Univariate and multivariate were conducted to study the relationship between CC and other variables. In figure 2, the cases was patients of CC who were willing to participant, and controls was patients who were not having CC. The result found female who were married, at least one abortion, alcohol intake, first intercourse below 18 years old were had high chance to get CC.

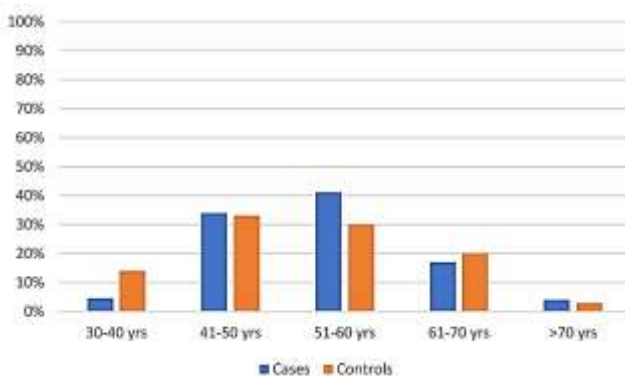


Figure 2. Age Distribution of Cases and Control

A research of Nesrin, 2011 was carried out the risk factor of cervical cancer in Turkish women. It used the method of case-control study and odd-ratio and 95% confident interval were obtained from multivariate logistic regression. The result showed early age of first sexual and history of reproduction infection were the risk factor to Turkish women.

Another research of Flores Y., 2008 had found the risk factor of cervical intraepithelial neoplasia (CIN) among HIV positive women in Mexico. They used case-control method to carry out the research. It had 94 cases and 501 controls by using 95% of confident interval. The research found out increase age, young age of first intercourse and low social-economic were affected women easier to get CC in Mexico.

2.2.3. Multiple linear regression (MLR)

MLR method is used to study the relationship between one dependent variable and two or more continuous independent variables. The MLR linear model is

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_pX_n \tag{2}$$

where

a = y-intercept

b_p = slope of coefficient in each independent variables

n = number of observations

It has a few of assumption to meet before using the method of MLR. First, the residual must be in normally distributed with mean is zero and variance σ . It is a linear relationship between dependent and independent variables. Coefficient of determination (R^2) is used to measure the goodness-of-fit in the model. The range of R^2 is $0 \leq R^2 \leq 1$. The higher the R^2 , the more fit the model. Whereas R^2 is 1 means the model is overfitting. A research of Dascălu & Cozma, 2009 using MLR and principle component analysis (PCA) to identify the correlation between variables and to reduce the predictor number to build model. The result showed that although the variables is reduced by using PCA, the most advantages it was more suitable and provided more weight to the model.

Huang et al., 2017 was carried out a research of to validate the predict 5-year mortality of breast cancer patient in Taiwan by using artificial neural network (ANN) model. It also compared predictive accuracy between ANN, MLR and Cox regression model. The comparison of training, testing and validation dataset of three models, the result showed that ANN model was more accurate and highest sensitivity among the models.

Wang et al., 2016 was using MLR method to study the correlation between blood glucose and the parameters.

It was also developed a model to over-weight patients. The six parameters were found by using MLR which were age, fasting alanine transaminase, blood urea nitrogen, total

protein, uric acid, and BMI. It showed a significant correlation to overweight patient.

3.0 SUMMARIES OF RESEARCH PAPER

Table 1. Summaries of Journals

Authors/ Year	Title	Technique	Objective	Findings/Results
Abduljabbar et al., 2014	Types and risk factors of cervical cancer	Retrospective, Descriptive cross-sectional study	- Determine the type of CC and risk factors for women. - Studied the relationship between risk factors and each type of CC.	- family member had history if any cancer was included in risk factor of CC.
Sharma & Pattanshetty, 2018	A study on risk factors of cervical cancer among patients attending a tertiary care hospital: A case-control study	Case-control study	- To study the relationship between CC and parity. - To study the relationship between CC and selected variables.	- risk factor were: marriage, first intercourse before 18 years old and history of alcohol intake.
Agarwal et al., 1993	Role of male behavior in cervical carcinogenesis among women with one lifetime sexual partner	Case-control study	- To investigate the role of male behavior in CC among Indian women who had one lifetime sexual partner.	- Male had sexual partner with 3 or more will increase the risk of getting CC to women.
Shields et al., 2004	A case-control study of risk factors for invasive cervical cancer among U.S. women exposed to oncogenic types of human papillomavirus	Case-control study	- To study the risk factors for invasive CC among U.S. Women.	- risk factor were: number of sexual partner, black race and taking birth control pills.
Lai et al., 2009	Age-specific correlates of quality of life in Chinese women with cervical cancer	Multiple linear regression	- To examine quality of life (QOL) and its age-specific factors in CC patients.	- young adulthood and education level were associated with quality of life.
Pourhoseingholi et al., 2008	Relation between Demographic Factors and Type of Gastrointestinal Cancer(GC) using Probit and Logit Regression Asma	Probit and logit regression	- to investigate the association between demographic factor and the type of GC. - to compare GC with probit and logit models.	-for logit model, gender gave an effect of colorectal cancer. - for probit model, age gave an effect of colorectal cancer. - result of probit and logit models were similar.
Sabates & Feinstein, 2006	The role of education in the uptake of preventative health care: The case of cervical screening in Britain	Probit regression	- to study the relationship between education and screening of CC.	- education was an importance factor for CC. - women who smoking were less went to screening.
Harmon et al., 1996	Acculturation and Cervical Cancer: Knowledge, Beliefs, and Behaviors of Hispanic Women	OLS and probit regression	- to study the culture level of knowledge in smear test, fear towards CC and screening.	- 90% of female were had Pap smear test in Phoenix metropolitan area.
Flores et al., 2008	Risk Factor of Cervical Cancer among HPV Positive Women in Mexico	Case-control study	- to identify the factor of association of cervical intraepithelial neoplasia(CIN) among HPV women in Mexico.	- factor were increase age, young age of first intercourse, low social-economic.
Reis et al., 2011	Risk Factors for Cervical Cancer: Results from a Hospital-Based Case-Control Study	Case-control study	- to identify the risk factor of CC in Turkish women.	- Early age of intercourse and history of infection reproduction system were the factors to get CC.

4.0 METHODOLOGY

4.1. Test of Hypothesis

Hypothesis testing is a method that can be used to make statistical decision about the hypothesis. It is used to test an assumption regarding a parameter. Hypothesis testing consists of two hypotheses, which are null hypothesis, H_0 and alternate hypothesis, H_a/ H_1 . Based on these two hypotheses, only one of it can be chose to make a right decision. Null hypothesis means that there is no difference between parameters while alternate hypothesis means that there is difference between the parameters.

A decision of hypothesis testing is referring to the level of significance. Level of significance also known as alpha or α and it is the probability of rejecting the null hypothesis when it is true. For example, 0.05 of significant level indicates that there is a 5% of error of rejecting null hypothesis. The reason is that there is no 100% accuracy in making decision of rejecting or accepting the null hypothesis.

The hypothesis is made based on the objective that have been discussed in Chapter 1. The assumption of hypothesis test is:

H_0 : There is no significant relationship between selected factors and cervical cancer.

H_1 : There is a significant relationship between selected factors and cervical cancer.

4.2. Data Collection Process

At this stage, the registration of data was done in the government website of National Medical Research Register (NMRR). After getting the approval from NMRR, the director of Hospital Pakar Sultanah Fatimah (HPSF) will proceed to do the approval for this research. The data for doing this research will be collected from HPSF Muar. The data is a secondary data which collected from department of Obstetrics and Gynaecology of HPSF. Secondary data is a kind of data that is collected by other parties. The data may be found on internet, books, journals or obtained from other organisations.

4.3. Data Analysis

This section will be discussing about the methods applied to analyse the data. The analysis of data is very important in order to achieve the objectives set in Chapter 1. The other subtopics that will be discussed are probit and logit regression model, assumptions of GLMs, Chi-squared Wald's test and Pearson goodness-of-fit test.

4.3.1 Generalized Linear Model (GLMs)

Generalized linear model is an extension of general linear model (Turner H., 2008). General linear model is a linear regression model used in continuous response variables and the distribution of residual need to be normal. Generalized linear model is a linear regression that allows the response variables have error distribution. The dependent variable does not need to be normally distributed, but it assumes a distribution from an exponential family such as binomial, Poisson and exponential. There are three components in GLMs, which

are random component, linear predictor and link function. Random components refer to probability distribution of response variables (Y) which come from an exponential family distribution. Family of exponential distribution include gaussian or normal distribution, Poisson distribution, binomial distribution and gamma distribution. For linear predictor, it is a linear combination of parameter and explanatory variable (X_i) such as $\beta_0 + \beta_1 X_1 + \beta_2 X_2$. The third component is link function. Link function is used to link between probability distribution and linear predictor. Table 2 shows the link function of exponential distributions families.

Table 2: Link Function of Exponential Distributions Families.

Distributions	Link name	Link function, $X\beta$
Normal	Identity	μ
Exponential	Negative inverse	$-\mu^{-1}$
Gamma	Negative inverse	$-\mu^{-1}$
Poisson	Log	$\ln(\mu)$
Binomial	Logit	$\ln\left(\frac{\mu}{1-\mu}\right)$
Binomial	Probit	μ^{-1}
Bernoulli	Logit	$\ln\left(\frac{\mu}{1-\mu}\right)$

4.3.1.1. Assumptions of Generalized Linear Model

There are a few assumptions need to be checked before the analysis can be conducted. All the assumptions are listed below:

The dependent variables (Y_i) not need to be normally distributed. The distribution needs to assume from exponential family like normal distribution, poisson distribution and binomial distribution.

The relationship between dependent and independent variable does not assumed to be linear, but it assumes linear relationship of transform response variable through link function.

The error needs to be independent, but no need to be normally distributed.

4.3.1.2. Probit Regression Model

Probit regression model is one of the binary link functions in GLMs. Binary probit model is a regression model in which its dependent variables are dichotomous. Binary dependent variables consist of two outcomes variables, it is either yes or no or with disease or no disease. Binary probit model is used in this research and the dependent variables in this research are:

$$y = \begin{cases} 1, & y > 0 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where 1 indicated having cervical cancer while 0 indicates no cervical cancer.

The dependent variable, Y of dichotomous needed to transform into a continuous variables $Y' \in (-\infty, \infty)$, the transformed Y' is

$$F(Y) = Y' = X\beta \quad (4)$$

Since the link function is follow standard normal distribution, it can say that

$$Y = \Phi(X\beta) \quad (5)$$

The probit model is thus

$$\Phi^{-1}(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_n \quad (6)$$

where

β_0 = intercept of the model

β_p = slope of independent variables

X_n = variables of independent

4.3.1.3 Logit Regression Model

Logit regression is known as logistic regression model and it is one of the binary link functions in Generalized Linear Model (GLMs). Logit regression and probit regression are the most frequent methods used in the relationship between binary dependent variables and covariates (Pourhoseingholi et al., 2010). It is used to make prediction when the dependent variables are binary while independent variables are continuous and/ or discrete (Pourhoseingholi et al., 2008). Since the dependent variable, Y in this research is dichotomous, link function is used to transform the probability of cervical cancer into log odds of cervical cancer. The link function of the binary logit model is:

$$\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) \quad (7)$$

where

π = the probability of CC

$(1-\pi)$ = the probability of no CC

The logit regression model is

$$\log\left(\frac{\pi}{1-\pi}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_n \quad (8)$$

4.3.1.3.1 Odds, Log Odds and Odds Ratio

Odds is a probability of success relative to probability of failure.

$$\text{Odds} = \frac{p}{1-p} \quad (9)$$

where

p = probability of success

(1-p) = probability of failure

Log of odds is used to link the probability and linear regression which is from [0,1] to $[-\infty, \infty]$. The probability of s-curve is transformed into a linear regression by using log of odds. Odds ratio is an odds of group divided by another odds of group. The formula of odds ratio is

$$\text{OR} = \frac{\text{odds}(\text{event A})}{\text{odds}(\text{event B})} \quad (10)$$

If odds ratio is below 1, it indicates that the event is more likely associated with lower odds of outcome. When odds ratio is equal to 1, it indicates the event is exactly exposed in two groups while the odds ratio greater than 1 indicates the event is more likely associated with higher odds of outcomes (Sedgwick & Marston, 2010).

4.3.2 Chi-squared Wald's Test

Wald's test is a chi-squared non-parametric test used to test the significance difference between independent and dependent variables. The aim of this test in this research is to identify the risk factors of cervical cancer. This test used the hypothesis testing to test the significant difference of the binary dependent variable and some factors of cervical cancer. If the p-value of the variables is less than 0.05, the variable is assumed to be significant. In other word, if the p-value is greater than 0.05, it does not have enough evidence to reject the null hypothesis (Fillon, 2016).

4.3.3. Pearson Goodness-of-Fit Test

Pearson goodness-of-fit test is a non-parametric test used to measure the fitness of the model in the data. The model is good if the model fit the observed data. The formula of Pearson goodness-of-fit is

$$\chi^2 = \sum_{j=1}^n \frac{(O_j - E_j)^2}{E_j} \quad (11)$$

where

χ^2 = Pearson's test statistic

n = number of cells in the table

O_j = Number of observations in cell j

E_j = Expected count in cell j

The hypothesis testing is used to make the decision.

H0: There is no significant difference between the observed and expected value or the model is fitted.

H1: There is a significant difference between the observed and expected value or the model is not fitted.

5.0 EXPECTED RESULT

Although there is no data collected yet at this stage, but we can expect some results based on the objectives of this research. The first expected result based on the objectives is that the risk factors of female patients in HPSF, Muar getting cervical cancer can be identified and can be more specified. For example, the age range or habits of women that make women get cervical cancer easier can be specified.

Next, a regression model will be built by using probit regression model after the risk factors of cervical cancer in female patients in HPSF, Muar is identified. The regression model can be used in future to help in predicting the probability of a women getting cervical cancer. The woman just needs to fill in the information or variables required by the model and the probabilities can be predicted.

Lastly, the relationship the cervical cancer and its risk factors will be identified after the risk factors are specified.

References

- Abduljabbar, D., Al-Rawahi, F., Faqih, F., Al-Khayat, M., Al-Mahmeed, M., Al-Khazali, M., Al-Sayed, N., AlGhaffar, S., & AlNasir, F. (2014). Types and risk factors of cervical cancer. *Bahrain Medical Bulletin*, 36(2), 94–96.
- Agarwal, S. S., Sehgal, A., Sardana, S., Kumar, A., & Luthra, U. K. (1993). Role of male behavior in cervical carcinogenesis among women with one lifetime sexual partner. *Cancer*, 72(5), 1666–1669.
- Bessler, P., Aung, M., & Jolly, P. (2007). Factors affecting uptake of cervical cancer screening among clinic attendees in Trelawny, Jamaica. *Cancer Control*, 14(4), 396–404.
- Bruni L. (2016). Human Papillomavirus and Related Diseases Report. HPV Information Centre. Retrieved from <https://www.hpvcentre.net/statistics/reports/XWX.pdf>
- Danaei, G., Vander Hoorn, S., Lopez, A. D., Murray, C. J. L., & Ezzati, M. (2005). Causes of cancer in the world: Comparative risk assessment of nine behavioural and environmental risk factors. *Lancet*, 366(9499), 1784–1793.
- Dascălu, C. G., & Cozma, C. D. (2009). The principal components analysis – method to reduce the collinearity in multiple linear regression model; application in medical studies. *International Conference on Multivariate Analysis and Its Application in Science and Engineering*, 140–145.
- Fillon, M. (2016). Engineered T Cells Improve Pancreatic Cancer Outcomes in Mice. *Journal of the National Cancer Institute*, 108(8), 8–9.
- Ghim, S. J., Basu, P. S., & Jenson, A. B. (2002). Cervical cancer: Etiology, pathogenesis, treatment, and future vaccines. *Asian Pacific Journal of Cancer Prevention*, 3(3), 207–214.
- Harmon, M. P., Castro, F. G., & Coe, K. (1996). Acculturation and cervical cancer: Knowledge, beliefs, and behaviors of Hispanic women. *Women and Health*, 24(3), 37–57.
- Hu, K., Wang, W., Liu, X., Meng, Q., & Zhang, F. (2018). Comparison of treatment outcomes between squamous cell carcinoma and adenocarcinoma of cervix after definitive radiotherapy or concurrent chemoradiotherapy. *Radiation Oncology*, 13(1), 1–7.
- Huang, S. H., Loh, J. K., Tsai, J. T., Houg, M. F., & Shi, H. Y. (2017). Predictive model for 5-year mortality after breast cancer surgery in Taiwan residents. *Chinese Journal of Cancer*, 36(1), 1–9.
- Kang, H. (2013). The prevention and handling of the missing data. *Korean Journal of Anesthesiology*, 64(5), 402–406.
- Lai, B. P. Y., Tang, C. S. K., & Chung, T. K. H. (2009). Age-specific correlates of quality of life in Chinese women with cervical cancer. *Supportive Care in Cancer*, 17(3), 271–278.
- Li, W., Mo, W., Zhang, X., Squiers, J. J., Lu, Y., Sellke, E. W., Fan, W., DiMaio, J. M., & Thatcher, J. E. (2015). Outlier detection and removal improves accuracy of machine learning approach to multispectral burn diagnostic imaging. *Journal of Biomedical Optics*, 20(12), 121305.
- Liu, S., Semenciw, R., & Mao, Y. (2001). Cervical cancer: The increasing incidence of adenocarcinoma and adenosquamous carcinoma in younger women. *Cmaj*, 164(8), 1151–1152.
- Low, E. L., Simon, A. E., Lyons, J., Romney-Alexander, D., & Waller, J. (2012). What do British women know about cervical cancer symptoms and risk factors? *European Journal of Cancer*, 48(16), 3001–3008.
- Pourhoseingholi, A., Pourhoseingholi, M. A., Rostami-Nejad, M., Rostami, K., Mirsatari, D., Zojaji, H., Solhpour, A., & Zali, M. R. (2010). Implementation of statistical analysis in the clinical research of coeliac disease: use of probit and logit analysis. *East African Journal of Public Health*, 7(2), 168–170.
- Pourhoseingholi, A., Pourhoseingholi, M. A., Vahedi, M., Safaee, A., Moghimi-Dehkordi, B., Ghafarnejad, F., & Zali, M. R. (2008). Relation between demographic factors and type of gastrointestinal cancer using probit and logit regression. *Asian Pacific Journal of Cancer Prevention*, 9(4), 753–755.
- Sabates, R., & Feinstein, L. (2006). The role of education in the uptake of preventative health care: The case of cervical screening in Britain. *Social Science and Medicine*, 62(12), 2998–3010.
- Sedgwick, P., & Marston, L. (2010). Statistical question: Odds ratios. *BMJ (Online)*, 341(7769), 407.
- Sharma, P., & Pattanshetty, S. M. (2018). A study on risk factors of cervical cancer among patients attending a tertiary care hospital: A case-control study. *Clinical Epidemiology and Global Health*, 6(2), 83–87.

Shields, T. S., Brinton, L. A., Burk, R. D., Wang, S. S., Weinstein, S. J., Ziegler, R. G., Studentsov, Y. Y., McAdams, M., & Schiffman, M. (2004). A case-control study of risk factors for invasive cervical cancer among U.S. women exposed to oncogenic types of human papillomavirus. *Cancer Epidemiology Biomarkers and Prevention*, 13(10), 1574–1582.

Suarez, S. S., & Pacey, A. A. (2006). Sperm transport in the female reproductive tract. *Human Reproduction Update*, 12(1), 23–37.

Wang, J., Wang, F., Liu, Y., Xu, J., Lin, H., Jia, B., Zuo, W., Jiang, Y., Hu, L., & Lin, F. (2016). Multiple Linear Regression and Artificial Neural Network to Predict Blood Glucose in Overweight Patients. *Experimental and Clinical Endocrinology and Diabetes*, 124(1), 34–38.