MODERN APPLIED MEDICAL RESEARCH 1(6): 23-27
Corresponding author: Mututuvari Edington ISSN: 2582-9181

Multivariate analysis of factors associated with incidence of cervical cancer

Mututuvari Edington[1], Mashasha Maxwell[2]
Department of Applied Mathematics and Statistics, Midlands State University.
Corresponding author: mashashamn@staff.msu.ac.zw

REVIEW ARTICLE

Received: **12-01-2021**
Accepted: **30-01-2021**
Published: **03-06-2021**

Abstract: The global burden of cancer continues to increase largely because of the aging and growth of the world population alongside an increasing adoption of cancer-causing behaviours, particularly smoking, physical inactivity and eating habits. Cervical cancers are caused by infection with one of the high-risk Human Papilloma Virus (HPV) types, however not all people who are infected may develop cancer, it is likely that other factors also play a role in the development of cervical cancer. The burden of cervical cancer is still very high in Zimbabwe mainly as a result of late presentation of disease, poor screening, diagnosis and treatment facilities which is compounded by the very high HIV incidence. This study was necessitated by an ever-increasing number of incident cases of cervical cancer among the Zimbabwean population with particular interest placed on establishing and examining the key factors contributing to an increase in cervical cancer incidence in Zimbabwe. Factors of interest were classified as demographic characteristics, sexual and reproductive history as well as menstrual factors and were found responsible for 56.46 % of the variance in incidence. It was also noted that incidence of cervical cancer doubles in women with HIV as opposed to women without HIV. Since elder women tend to have more children there is a strong positive correlation between age and parity in cervical cancer cases.

Keywords: *Cancer, Multivariate, Principal component analysis, invasion*

Introduction

Cancer is a dynamic, complicated and diverse class of diseases that disregard the normal rules of cell division through unregulated cell growth, interactions and divisions (mutations), forming tumours which turn to invade other body parts and spread to other organs (metastasis). Cancer incidence increases with age, as a person grows older the cellular repair mechanism becomes less effective. The global burden of cancer continues to increase largely because of the aging and growth of the world population alongside. Age and use of tobacco, alcohol, early pregnancy, taking an unhealthy diet and lack of physical activity are some of the cancer risk factors in today's world. About 12.7 million cancer cases and 7.6 million cancer deaths were estimated in 2008, of which, 56% of the cases and 64% of the deaths occurred in the developing world (CA Cancer 2011,61:69-90, 2011 American Cancer Society Inc.)

There are different types of cancers including cervical cancer, breast cancer, prostate cancer, lung cancer among others. Cervical cancers may be caused by infection with the high-risk HPV types, but not all infected people may develop cancer. Cervical cancer is the most common HPV associated with cancer among women and second after breast cancer as the most common female malignancy in both incidence and mortality worldwide (Jemal A et al, 2013). Up to 90% of cervical cancers arise from squamous cells and are called squamous cell carcinomas, with most of the remainder coming from the glandular cells (adenocarcinomas). The biggest risk of cervical cancer is Human papilloma Virus (HPV) commonly detected in tumour specimens. (World Health Organisation, 2017).

Staging of cancer is an important part of determining the best treatment plan. Both the FIGO (International Federation of Gynaecology and Obstetrics) system and the AJCC (American Joint Committee on Cancer) have developed systems to stage cervical cancer. Cervical cancer is classified from stages 0 to 4, with many subcategories within each numerical stage. The systems are based on tumour extent, spread to any lymph nodes and distant spread. In the initial stage abnormal cells are on the cervix surface, the stages increase with the increase in the tumor spread and size such that in the last stage (stage 4) the cancer will be spread to other board parts such as the bladder, or rectum hence is the most advanced stage.

Cervical cancer symptoms may develop when the cancer cells start to invade surrounding tissue (metastasis). Symptoms include: vaginal bleeding, bleeding after sex, longer or heavier menstrual periods than usual among others. An estimated 90% of the globally recorded cervical cancer related deaths are in low- and middle-income countries (LMICs), for which 8 in 10 are recorded within the Sub Saharan African Region (Ferlay J et al, 2013). In LMICs there are challenges in affordability and availability of drugs, as well as access to treatment facilities (Kuguyo et al, 2017).

Global trends show that most high-risk cervical cancer countries are in Africa as it is estimated that 2270 women are diagnosed with cervical cancer in Zimbabwe annually and a mortality rate of 64% has been recorded (Globocan IARC, 2012).

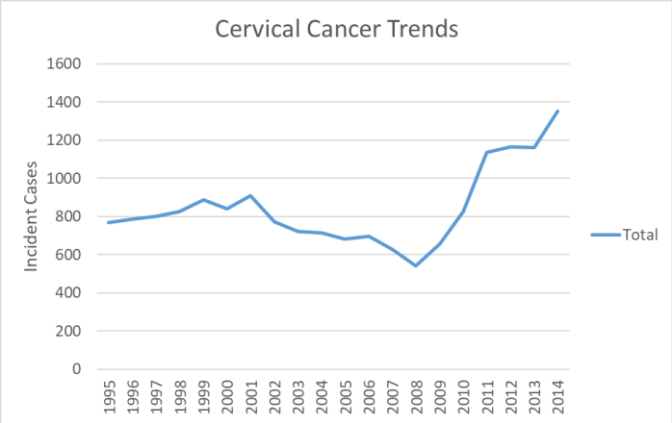

Figure 1: Cervical cancer trends in Zimbabwe. (Chokunonga E et al, 2014)

This study has been necessitated by an ever-increasing number of incident cases of cervical cancer among the Zimbabwean population. Awareness of the risk factors believed and known to lead to cervical cancer is quite significant for preventing the illness through detecting high risk groups and for early diagnosis as well as to establish and improve any programs and strategies to combat the disease. Hence the study wishes to establish and examine the key factors contributing to an increase in cervical cancer incidence. Although there has been a lot of epidemiological studies and publications in other countries, we are not aware of an earlier investigation into risk factors for cervical cancer carried out Zimbabwe. Identifying factors associated with the cervical precancerous lesions is important for planning more targeted screening programs to decrease the high morbidity and mortality of the disease in the country. The study seeks to analyse and examine the relationship between the factors and development of cervical cancer through the use of multivariate analysis techniques. These are techniques that allow more than two variables to be analysed at once. The ultimate goal of these analyses is either explanation or prediction i.e. more than just establishing an association (Rencher A C et al, 2012)

Multivariate analysis-Factor analysis

Multivariate analysis is based in observation and analysis of more than one statistical outcome variable at a time (Grimm L.G, 2011). Multivariate methods are designed to simultaneously analyse data sets and to find the cause-and-effect relationships between variables (Afifi et al 2011). Factor analysis is used to reveal the patterns of interrelationships among variables that are not readily apparent, for confirmation of hypothesis, and for reducing the number of variables to a manageable level. There are two main factor analysis methods: common factor analysis, which extracts factors based on the variance shared by the factors, and principal component analysis, which extracts factors based on the total variance of factors. The factor loadings are the correlations between the factor and the variables. Typically factor loading of 0.4 or higher is required to attribute a specific variable to a factor. The Kaiser-Meyer-Olkin (KMO) statistic, also called the measure of sampling adequacy (MSA), indicates whether the other variables in a dataset can explain the correlations between variables. The KMO is a statistic that indicates the proportion of variation caused by underlying factors. Kaiser (1974) introduced the statistic and recommends that a KMO value less than 0.5 is unacceptable. Kaiser criterion or latent root criterion is mainly used to determine the sampling adequacy of data to be used in factor analysis through determining the number of factors with eigen value greater than 1 which are considered significant (Russell 2002; Zwick and Velicer 1986). Another popular method is the scree plot which has a distinct break(elbow) in it, thereby showing the correct number of factors (Cattell 1966). This distinct break is called the "elbow" and all factors above the elbow should be retained. If different methods are used and they all give the same number of factors, it increases our confidence with the results. Factor loadings are then used which takes values between ± 1 representing correlation between the factors and variables. The higher the factor loading the greater the certainty that the factor represents a variable well. Odds ratios are also used to measure the association between exposure (HIV status) and outcome (cervical cancer incidences). They represent the odds that an outcome will occur given a particular exposure.

METHODOLOGY

The study was conducted in Gutu District at six VIAC implementing sites. Gutu District has a population of 203 533 (Census 2012 Report) and a population density of about $22.08/km^2$.The estimated population of women who are more than 15 years of age is 62 444 with 8755 being HIV positive. An unmatched case control study (participants are enrolled in the study based on whether they are VIAC positive or not) was used with those women with positive visual inspection treated as cases and those women with negative visual inspection as controls. Secondary data from medical records and National Cervical Cancer Screening Register for the year 2017 was reviewed.

The Fischer's sample size estimation (Fisher, 1973) of case control studies considering the parameters of 95% confidence interval, 80% power, taking 1:2 ratio of cases to controls was used to estimate the sampling size which was 724 for the case

Corresponding author: Mututuvari Edington
ISSN: 2582-9181

group and 1448 for the control group (Kasiulevicius et al , 2006). Data was collected from the six centres using the weighted size of cases and controls as shown in the Table1 and Figure 2 shows the proportion of clients tested per centre.

Table 1: Weighted sample size of cases and controls per health facility

Health Facility	Sample size of cases	Sample size of controls
Gutu Rural Hospital	261	522
Gutu Mission Hospital	65	130
Mukaro	87	174
Chimombe	123	246
Chinyika	87	174
Chitando	101	202
TOTAL	724	1448

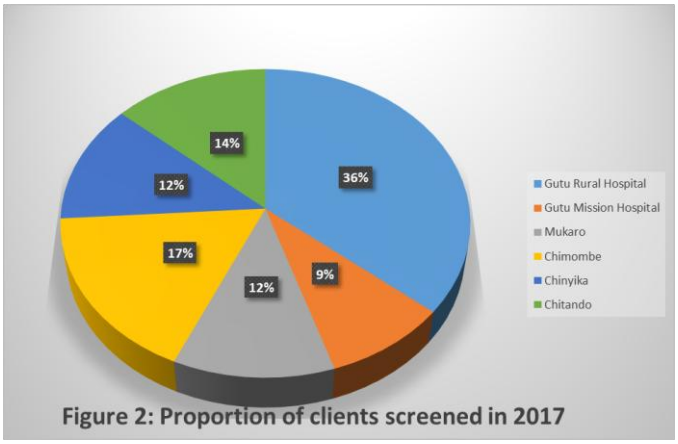

Figure 2: Proportion of clients screened in 2017

Plan for data analysis

The collected data was analysed using STATA Version 13. Simple frequency distributions, summary distributions and cross tabulations were also used in cleaning the data. Odds ratios with 95% confidence interval and two tailed p-values were calculated to identify the presence and strength of association. Statistical significance was declared at P-value less than 0.05.

Results

A total sample of 2172 women aged at least 15 years who were screened for cervical cancer at the six VIAC centres in Gutu were used. Variables of interest include age, age at menarche, age at first sexual encounter, HIV status, history of sexually transmitted disease, number of life partners, contraceptives commonly used, parity and marital status. Table 2 summarises the distribution of cases and controls according to the variables of interest.

Table 2: Distribution of cases and controls according to the variable characteristics

Factor	Cases (n=724)	Proportion of cases	Controls (1448)	Proportion of controls	Total (2172)
Age (years)					
≤49	619	28.50	1,111	51.15	1730(79.65)
50-59	54	2.49	171	7.87	225(10.36)
60+	51	2.35	166	7.64	217(9.99)
Parity					
≤2	341	15.70	568	26.15	909(41.85)
≥3	383	17.63	880	40.52	1263(58.15)
Number of life partners					
<2	445	20.49	1036	47.70	1481(68.19)
≥2	279	12.85	412	18.97	691(31.81)
Age at first sexual encounter					
<18	166	7.64	304	14.00	470(21.64)
≥18	558	25.69	1144	52.67	1702(78.36)
Age at menarche					
≤12	46	2.12	63	2.90	109(5.02)
13-14	265	12.20	682	31.40	947(3.60)
≥15	413	19.01	703	32.37	1116(51.38)
HIV Status					
Negative	440	20.26	1,141	52.53	1581(72.79)
Positive	284	13.08	307	14.13	591(27.21)
History of STI					
Yes	226	10.41	397	18.29	623(28.7)
No	498	22.94	1050	48.36	1548(71.3)
Marital status					
Married	539	24.82	1,109	51.06	1648(75.87)
Single	88	4.05	141	6.49	229(10.54)
Widowed	97	4.47	198	9.12	295(13.58)

Most of the cases were recorded in women who are less than 49 years of age, had less than 2 sexual partners and most of the participants were married, HIV negative and do not have any history of sexually transmitted diseases. Most participants had parity of at least three which then pauses a threat to most women as high parity increases the risk of squamous-cell carcinoma of the cervix among women.

Measures were done to determine whether the items correlate sufficiently using the Kaiser-Meyer-Olkin (KMO), was used to determine variable correlation and Table 3 shows the individual and overall KMO values.

Table 3: Kaiser-Meyer-Olkin measure of sampling adequacy

Variable	KMO
Age	0.5034
Parity	0.4906
Age at menarche	0.4519
History of STI	0.5718
Number of life partners	0.6679
Contraceptives	0.6442
HIV status	0.5746
Marital status	0.5361
Overall	0.5239

The overall KMO is 0.5239 which is above 0.5 suggesting that the data has sufficiently correlated variables, hence setting the tone for proceeding with factor analysis. Age at first sexual encounter had the least individual KMO of 0.3032 hence it was removed from the list of variables of interest. In the extraction and determination of factors both the Kaiser criterion and scree plot technique were used to determine significant factors. As the number of factors that the scree plot suggests is one factor less than the elbow indicates, we concluded that three factors were significant as shown in Figure 3 which was in agreement with the Kaiser criterion.

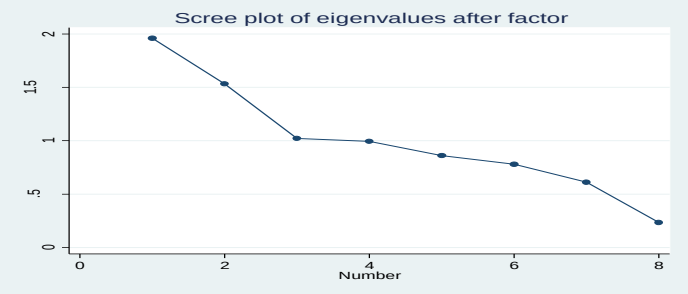

Figure 3: Scree plot of eigenvalues

Table 4: Rotated factor loadings and uniqueness

Variable	Factor 1	Factor 2	Factor 3	Uniqueness
Age	0.9165	0.0048	0.1352	0.1417
Parity	0.8810	-0.1505	-0.1156	0.1879
Age at menarche	-0.1927	-0.1355	0.5354	0.6578
History of STI	-0.0209	0.6576	-0.2893	0.4834
Number of life partners	-0.0802	0.5582	0.0194	0.6816
Contraceptives	0.1929	0.0476	0.6959	0.4762
HIV status	0.0149	0.7469	0.0900	0.4338
Marital status	0.4242	0.4392	0.4543	0.4248

Table 5: Factor rotation Matrix

	Factor 1	Factor 2	Factor 3
Factor 1	0.9461	0.1479	0.2882
Factor 2	-0.1695	0.9842	0.0515
Factor 3	-0.2760	-0.0976	0.9582

As its name implies, the rotated factor loadings block shows the factor loadings after rotation. Using Table 4, we observed that age and parity which we can combinedly call demographic characteristics have high loadings on factor 1, number of life partners, history of STI and HIV status load highly on factor 2 and we can name them sexual history characteristics while the reproductive characteristics that is age at menarche, contraceptive commonly used and marital status load highly on factor 3. We further assessed the goodness of fit for the analysis using the standardised residual covariance matrix as shown in Table 6.

Table 6: Residual matrix

	Age	Parity	Age at menarche	History of STI	Nmbr of life partners	Contracptv	HIV status	Marital status
Age	0.0000							
Parity	-0.0783	0.0000						
Age at menarche	0.0956	0.1784	0.0000					
History of STI	0.0297	0.0290	0.2114	0.0000				
Number of life partners	0.0248	0.0884	0.0130	-0.2079	0.0000			
Contracept ves	-0.0584	-0.0100	-0.3402	0.1289	-0.0083	0.0000		
HIV status	-0.0007	0.0499	0.0466	-0.1779	-0.2469	-0.0735	0.0000	
Marital status	-0.0673	-0.1191	-0.1124	-0.0961	-0.1001	-0.2168	-0.1181	0.0000

Note that the diagonal elements are equal to zero and there are few residuals with values greater than 0.05 implying that we have a better analysis that fit well to the available data as a good model must have low residual values (<0.05).

We further used odds ratios to establish the relationship between HIV status and cervical cancer incidence and found out that the odds of developing cervical cancer in HIV positive women is 2.39 times more than in HIV negative women. Hence HIV positive women are at a greater risk of developing cervical cancer than those who are HIV negative.

Conclusion

We can conclude that factors that we had classified as demographic characteristics, sexual history and reproductive history and menstrual factors are responsible for 56.46% of the variance. Cervical cancer incidences in HIV positive women are at least doubled compared to HIV negative women. We can also conclude that as women grow and have more children, they increase the risk of cervical cancer development as evidenced by the correlation between age and parity.

Corresponding author: Mututuvari Edington ISSN: 2582-9181

References

1. Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., ... & Bray, F. (2016). Cancer incidence and mortality worldwide: IARC CancerBase No. 11. Lyon, France: International Agency for Research on Cancer. GLOBOCAN 2012 v1. 0.
2. International Agency for Research on Cancer. (2012). Cervical cancer estimated incidence, mortality and prevalence worldwide in 2012. World Health Organization.
3. Kasiulevičius, V., Šapoka, V., & Filipavičiūtė, R. (2006). Sample size calculation in epidemiological studies. *Gerontologija*, *7*(4), 225-231.
4. Chokunonga, E., Borok, M. Z., Chirenje, Z. M., Makunike-Mutasa, R., Ndlovu, N., Nyakabau, A. M., & Vuma, S. (2017). Zimbabwe National Cancer registry: 2014 Annual Report: pattern of cancer in Zimbabwe. *Accessed on*, *5*.
5. Fisher R.A(1973),Statistical Methods and Scientific Inference, Third edition Hafner Press, London
6. Jemal, A., Center, M. M., DeSantis, C., & Ward, E. M. (2010). Global patterns of cancer incidence and mortality rates and trends. *Cancer Epidemiology and Prevention Biomarkers*, *19*(8), 1893-1907.
7. World Health Organization, Światowa Organizacja Zdrowia, Programme on Cancer Control (World Health Organization), & World Health Organization. Reproductive Health. (2002). *Cervical cancer screening in developing countries: report of a WHO consultation*. World Health Organization.
8. DeSantis, C. E., Ma, J., Goding Sauer, A., Newman, L. A., & Jemal, A. (2017). Breast cancer statistics, 2017, racial disparity in mortality by state. *CA: a cancer journal for clinicians*, *67*(6), 439-448.
9. Grimm LG, Yarnold PR. Eds (2011). Reading and understanding multivariate statistics. Washington, DC: American Psychological Association Washington, 2011
10. Rencher AC, Christensen WF. Eds (2012). Methods of Multivariate Analysis. New Jersey: Wiley
11. Afifi A, May S, Clark VA (2011). Practical Multivariate Analysis, Fifth Edition. Boca Raton, FL: CRC Press.
12. Kaiser, H.F (1958).The varimax criterion for factor analytic rotation in factor analysis.Educational and psychological measurement, page 770-773.
13. Kaiser, H. F. (1974). An index of factorial simplicity. Psychometrica, 39 (1), 31–36. *Rev Bras Epidemiol*.
14. Russell, D. W. (2002). In search of underlying dimensions: The use (and abuse) of factor analysis in Personality and Social Psychology Bulletin. *Personality and social psychology bulletin*, *28*(12), 1629-1646.
15. Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological bulletin*, *99*(3), 432.
16. Kuguyo, O., Matimba, A., Tsikai, N., Magwali, T., Madziyire, M., Gidiri, M., ... & Nhachi, C. (2017). Cervical cancer in Zimbabwe: a situation analysis. *The Pan African Medical Journal*, *27*.
17. World Health Organization (WHO) 2010. Human Papilloma virus laboratory: https://www.who.int/immunization/hpv/learn/hpv_laboratory_manual__who_ivb_2009_2010.pdf