

Assessing Efficient Risk Ratios: An Application to Surgical Stage Prediction in Cervical Cancer

Jean C. Jesang, Collins O. Odhiambo*

Institute of Mathematical Sciences, Strathmore University, Nairobi, Kenya

Email: *codhiambo@strathmore.edu

How to cite this paper: Jesang, J.C. and Odhiambo, C.O. (2020) Assessing Efficient Risk Ratios: An Application to Surgical Stage Prediction in Cervical Cancer. *Open Journal of Statistics*, 10, 274-302. <https://doi.org/10.4236/ojs.2020.102020>

Received: January 30, 2020

Accepted: April 19, 2020

Published: April 22, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Background: Cervical cancer remains the second most commonly diagnosed cancer and the third leading cause of cancer death in developing countries. Improving clinicians' knowledge and understanding of surgical staging is critical in the fight against the disease. However, a systematic evaluation of different ordinal regression models based on diverse predicted outcomes has not been given its due share in literature. **Objective:** To systematically assess the flexibility of odds ratios for three popular ordinal regression models *i.e.* the Multinomial Logistic (ML) model, the Continuation Ratio (CR) model and Adjacent Category Logistic (ACL) model when applying cervical cancer data in surgical stage prediction. **Method:** We systematically, compared the performance of CR, ML and the ACL as the predictive mechanisms, and evaluate the most appropriate model in the cervical cancer setting. The study considered women who visited the Oncology department at the Moi Teaching and Referral Hospital's Chandaria Cancer and Chronic Diseases Center and were diagnosed and surgically treated for cervical cancer from January 2014 to December 2018. **Results and Conclusion:** We presented the comparison between 3 different regression models for ordinal data within the cervical cancer setting. We found that the CR model without proportional odds yielded better results comparing Akaike Information Criterion (AIC), log likelihood ratio and residual deviance. In addition, the key prognostic factor associated with invasive cervical cancer was the (International Federation of Gynecology and Obstetrics) FIGO clinical stage which in particular, had a higher influence on the surgical Stage 2 outcomes compared to the lesser surgical stage categories. All the 5 independent features selected for classifying the patients into surgical stages were the FIGO clinical stage and partly, the presence or absence of symptomatic vaginal discharge.

Keywords

Surgical Stage, Ordinal Regression, Cervical Cancer, Odds Ratio, Predictive Variables

1. Introduction

Cervical cancer remains the leading type of malignant growth in Kenya among women of all ages with a crude incidence rate of 22.4 per 100,000 persons and a crude mortality rate of 11.5 per 100,000 in the year 2017 [1]. Cervical cancer is caused by infection of the cervix by the human papilloma virus (HPV). The persistence of the HPV infection on the cervix causes oncogenic cell transformation at the squamous columnar junction [2]. HPV types 16 and 18 are the most prevalent among women with a normal cytology, low and high grade cervical lesions and those who progress to cervical cancer [1]. Nonetheless, cervical cancer is the best preventable malignancy of all relevant human cancers with an increase in the establishment of cervical cancer screening centers in middle and low income countries. The introduction of screen and treatment strategies for patients with abnormal Visual Inspection with Acetic Acid (VIA) of the cervix has increased the number of women screened and treated for cervical cancer in Kenya [3]. However, with the availability of HPV vaccines, the high cost limits their implementation in middle and low income countries leading to more access to surgical care than chemotherapy and radiology [4].

Surgical treatment is among the curative options given to women diagnosed with cervical cancer in the middle and low income countries. The extracted specimen undergoes pathological assessment to determine the full extent of the disease thus classifying the specimen into a surgical stage. Allanson *et al.* [5] found that systematic evaluation of surgical treatment outcomes such as adverse effects and complications vitally improve patient health outcomes.

Authors who have looked at statistical and mathematical models that are applied in cancer setting include [6] [7] [8] [9]. However, medical studies with ordinal data, [10] have generally been dichotomized prior to analysis. According to Javali *et al.* [10], estimating the risk of adverse effects, often measured on interval scales remains critical interest of epidemiologists and statisticians. Ordinal regression models have been underutilized despite being applicable in many fields. In support of risk estimation, Freedman [11] reports that the National Cancer Institute had identified risk prediction as an area of extra-ordinary opportunity in the “Nation’s Investment in Cancer Research”. The relevance of risk prediction today in cervical cancer care is best summarized by Dr. Micheal Rothberg [12]:

While HPV tests are very helpful in predicting cancer risk, other factors are just as powerful at predicting cervical cancer risk. The more that we can personalize risk prediction, the more efficient our screening efforts will become.

Globally, the development and use of predictive models today is growing rapidly and highly applicable in the health care sector for the provision of efficient care and resources to patients. Predictive models are developed from statistically significant factors associated with the outcome of interest and the models can range from complex to simple. The application of predictive modeling techniques in the early diagnosis and prognosis of cancer has become a requisite to facilitate effective clinical management of patients. More so, machine learning techniques aim to model the progression and treatment outcomes of the cancer and improve our understanding of the disease thus resulting in accurate and effective management of cancer patients. The techniques could improve the accuracy of cancer susceptibility, recurrence and survival prediction.

Further, predictive models can be used to risk-stratify patients and appropriately distribute resources such as caregivers and treatment combinations to the women and also, identify women who are at high risk of progression to clinical disease for disease management programs. Notably, predictive modeling in the health sector has the potential to impact clinical and therapeutic decision making.

This article gives an overview of 3 regression models developed for ranked data. It is clear that the most popular model for the analysis of ordinal data is the CPO model. However, the inflexibility of the proportional odds assumption brought about the development of other regression models for ordinal data that would ease on the proportional odds assumption. Generally, regression analysis investigates the influence of multiple predictors or independent variables on a dependent variable or outcome. The assumption of proportional odds in ordinal regression is that the effects of any explanatory variables are consistent or proportional across the different categories. One of the major shortcomings of the CPO model is the relationship between the predictors and the response variables that can be greatly misleading when assumptions are violated. Theoretically a more recommended model for ordinal data would take into account the categorical nature of the response since more information is contained within the ordered structure of the categories [13]. Ordinal data is non-separable, independent, strictly increasing (decreasing) with arbitrary cut-points of some underlying continuum [13].

Based on the pathologist's point of view *i.e.* the surgical stage in this study, the most vital prognostic factors were presented and existing dissensions in the classification and diagnosis of the extracted tumors clarified by 3 types of regression models. In this study, we seek to assess 3 types of regression models for ordinal responses to predict the surgical stage of HIV infected and uninfected women surgically treated upon being diagnosed with cervical cancer. The 3 predictive mechanisms covered here have previously been looked by [14] [15] [16] [17]. Brant sought to assess the goodness of fit of the proportional odds model for ordinal logistic regression. This particular model represented a series of logistic regressions for dependent binary variables utilizing common regression parameters (with the proportional odds assumption) [16]. Methods for ordinal variables

are considered natural extensions of probit and logit models for dichotomous variables [17]. Such models explicitly recognize ordinality, avoid arbitrary assumptions concerning the ordinal scales and allow for analysis of continuous, dichotomous and ordinal variables within a common statistical framework [17]. Statistical packages such as lme4, nnet were developed to allow for the implementation of cumulative link (mixed) models which are also known as ordered regression models, proportional odds models, proportional hazards models for group survival times and ordered logit/probit model. Estimation techniques were mainly via maximum likelihood [15]. Through extensions to non-linear models, McCullagh reports that the method of iteratively reweighted least squares converged to the maximum likelihood estimate which greatly simplifies the necessary computation of regression models for ordinal data [14]. Excellent summary can be found in [18]. Statistical methods for modeling ordinal response data such as the continuation ratio model, the polytomous logistic model among others are fully described with application to perinatal health programme data [18].

The rest of the paper is organized as follows. The methods and materials are covered in Section 2. In Section 3, we give an elaborate description of analysis and results. In Section 4, we discuss and describe the results. We compare the three models and show application of these methods to the cervical cancer data.

2. Materials and Methods

2.1. Ordinary Regression Models

2.1.1. Multinomial Logistic Regression (ML) Model

Let Ψ be a multinomial response variable with categorical outcome $1, 2, \dots, n$ and let ψ_i denote a p -dimensional vector of exploratory variables. The dependence of Ψ on ψ_i can be expressed as [18]:

$$Pr(\Psi = \Psi_j | \psi) = \frac{\exp(\alpha_j + \psi' \beta_j)}{1 + \exp(\alpha_j + \psi' \beta_j)}, j = 1, 2, \dots, n. \quad (1)$$

The logit form of Equation (1) yields:

$$\text{logit}(\Pi_j) = \log \frac{\Pi_j}{1 - \Pi_j} = \log \left[\frac{Pr(\Psi = \Psi_j | \psi)}{Pr(\Psi = \Psi_k | \psi)} \right]. \quad (2)$$

The parameter α_j is the unknown intercept and $\beta = (\beta_1, \beta_2, \dots, \beta_n)$ is a vector of unknown coefficients responding to ψ . Extensive coverage of the properties of β and α can be found in [14] [18].

The odds ratio, Θ_i of the k^{th} covariate ψ_k is expressed as:

$$\Theta_p = \frac{Pr(\Psi = \Psi_j | \psi_k^{(1)})}{Pr(\Psi = \Psi_j | \psi_k^{(0)})} = \exp(-\beta(\psi_k^{(1)} - \psi_k^{(0)})). \quad (3)$$

2.1.2. Continuation Ratio (CR) Model

Here we replace $\Pi_j = Pr(\Psi \leq \Psi_j)$ with one of the j^{th} of the CR model with

probability of being in category j ($\theta_j = Pr(\Psi = \Psi_j)$) conditional on being in category greater than j . Let $\Omega_j = \frac{\theta_j}{1 - \Pi_j}$. The CR model can be expressed as:

$$\text{logit}(\Omega_j) = \log \left[\frac{\Omega_j}{1 - \Omega_j} \right] = \log \frac{Pr(\Psi = \Psi_j | \psi)}{Pr(\Psi > \Psi_j | \psi)} = \alpha_j - \psi'\beta, j = 1, 2, \dots, n. \quad (4)$$

The odds ratio of CR model is then obtained as:

$$\Theta_C = \frac{Pr(\Psi = \Psi_j | \psi_k^{(1)}) / Pr(\Psi > \Psi_j | \psi_k^{(1)})}{Pr(Pr(\Psi = \Psi_j | \psi_k^{(0)}) / Pr(\Psi > \Psi_j | \psi_k^{(0)}))} = \exp(-\beta(\psi_k^{(1)} - \psi_k^{(0)})). \quad (5)$$

2.1.3. Adjacent-Category Logistic (ACL) Model

The ACL model involves the ratio of two probabilities *i.e.* $Pr(\Psi = \Psi_j)$ and $Pr(\Psi = \Psi_{j+1})$ for $j = 1, 2, \dots, n$. The model is expressed as

$$\log \left[\frac{Pr(\Psi = \Psi_j)}{Pr(\Psi = \Psi_{j+1})} \right] = \alpha_j - \psi'\beta, j = 1, 2, \dots, n. \quad (6)$$

The parameter β_1 corresponds to the coefficients of the log-odds of $(\Psi = \Psi_1)$ relative to $(\Psi = \Psi_2)$ if $\alpha_k = 0$ and $\beta_k = 0$. Consequent odds follow the same pattern.

2.2. Study Subjects

We adopted a cross-sectional design which utilized the retrospectively maintained database to identify all the women with International FIGO Stage 0-IVB cervical cancer managed by the Oncology department as outpatients at the MTRH's CCCDC from January 2014 to December 2018. Staging occurred according to the guidelines of the FIGO system; these did not change during the inclusion period. Eighty seven women were diagnosed and confirmed to have invasive cervical cancer between January 2014 and December 2018. These women were found to be eligible for surgical treatment and underwent surgery at the Chandaria Cancer and Chronic Diseases center. Women whose HIV status was unknown and had incomplete follow up data within the stipulated period were excluded from the study. Thus, only women who were either HIV positive or HIV negative were eligible to take part in the study. Most women had experienced the symptoms associated with cervical cancer such as abnormal bleeding and unusual vaginal discharge with possible foul smell thus prompting the women to seek cervical cancer screening services.

2.3. Demographics

The overall mean and median age at first contact with the oncology team were 46.61 and 46.00 years. The overall mean and median age at surgery were 46.76 and 47.00. For the HIV status, 77.6%, 16.5% and 5.88% were found to be HIV

negative, HIV positive and unknown HIV status therefore all the patients with unknown HIV status were dropped leaving 82.28% and 17.72% being HIV negative and HIV positive. The marital status was classified as either single or married. The single patients comprised of the singles, widows, divorced and those who did not state their marital status. The clinical stages were merged to clinical stages 1 and 2 at 78.67% and 21.33% respectively with the clinical stage stated as “others” were dropped. The clinical stages were dichotomized into 2 categories and it became clear that on categorizing the Clinical stage to 1 and 2 only, it is found to be statistically significant with a p -value of <0.001 . Whether there was vaginal involvement and parametrial involvement during diagnosis of cervical cancer were found to be statistically significant with p -values of <0.001 and 0.008. The symptoms of vaginal discharge and lower abdominal pain during diagnosis of cervical cancer were found to be statistically significant with p -values of 0.029 and 0.048. All other variables were found to be statistically insignificant. The median number of child births per woman was 4 with the majority of women stating to be married. Majority of the women were clearly non-smokers and non-alcoholics. The main method of contraceptive used was the injectable form known as depo provera. However, most women seemed to not be using any form of birth control contraceptive. For the response variable, only 1 individual was found to be classified under surgical Stage 4. This particular individual was dropped leaving 55 under surgical Stage 0, 19 under surgical Stage 1 and 13 classified under surgical Stage 2.

2.4. Materials

Upon visiting the cervical cancer clinics for screening within the Western and Rift region of Kenya, women with suspicious lesions on the cervix would undergo colposcopic biopsy whereby a colposcope was used to examine the cervix for any abnormal tissue. A biopsy punch forceps was utilized to remove a small fragment of the abnormal area or suspicious lesion which was taken for pathological evaluation to determine the type of invasive cancer (squamous cell carcinoma or adenocarcinoma). In addition, a physical examination of the cervix was done to determine the clinical stage of the cancer, blood tests, CT scans and chest x-rays. The pathology result was received after two weeks and the women underwent gynecological review. The women were asked standardized questions concerning their social behaviors, demographic details and past treatments assigned which determined the new treatment given at that particular time. Women assigned to have surgical treatment were scheduled and surgery carried out. The specimens were taken to the pathologist for surgical pathological evaluation to clearly assess the extent of the disease and determine the direction of treatment. The pathologists carried out physical and microscopic examination of the extracted tissues. The specimens were classified under surgical stages that state the involvement of the lymph nodes, the parametrium and also, determine whether surgery was the only treatment necessary or whether alternative treatment would be needed.

2.5. Procedure

The research design for this study was cross-sectional. The data for the study was retrospectively retrieved from the gynecological cervical cancer database. The data had been collected previously and was parallel to the patients' record files. The women who attend the gynecology clinic usually return for follow ups weekly, monthly and after 3 months. The gynecologists use files to record patient information at every visit and research assistants key in the recorded data into an MS access database at the close of the clinic sessions.

690 women with complete records sought treatment at the oncology clinic with only 75 women found to be eligible and their data utilized in the building of the predictive models. Moreover, data was simulated to test the performance of the developed models as the original data of 75 women was small to allow for partitioning. The independent variables in this study were age at first contact with the oncology team ranging from approximately 22 - 81, parity of at least 2 live births per woman, international FIGO clinical stage which was dichotomized to clinical stages 1 and 2, HIV status of patient limited to either being HIV positive or HIV negative, vaginal involvement, parametrial involvement, marital status, weight of patient, smoking status, contraceptive use, method of cancer detection, biopsy pathology result, type of surgery done, symptoms which included bleeding, vaginal discharge or lower abdominal pain, location of the cervical cancer tumour, grade of the tumour, the duration of the symptoms prior to diagnosis with the options being <1 month, <6 months, <1 year, >1 year and not stated with the dependent variable being surgical stage with the 3 categories being surgical stages 0,1 and 2.

Figure 1 shows a flowchart displaying the surgical treatments that were availed and the surgical stage outcomes.

Figure 2 is a flowchart displaying the surgical stage outcomes based on the colposcopic biopsy results.

2.6. Statistical Tests

In this study, regression models were used to explore the relationship between the response variable (surgical stage) and the explanatory variables. The data was analyzed using R studio version 3.6.1. Chi-square tests and analysis of variance tests were carried out for categorical and numerical variables. The ANOVA test allowed us to examine the variation in the frequencies within each surgical stage (the response variable). Three regression models for ordinal data were developed and their predictive performance evaluated by comparing the odds ratios. These models were adapted because the response variable was an ordered variable. The 3 models were the multinomial (polytomous) logistic model, the continuation-ratio model and the adjacent-category logistic model for which the later 2 were developed with and without the proportional odds assumption.

We utilized R command multinom (Package: nnet) to fit 2 multinomial log-linear models via neural networks. For the ACL model, we utilized the R vgam package

Flowchart displaying surgical stage outcomes and surgical treatment

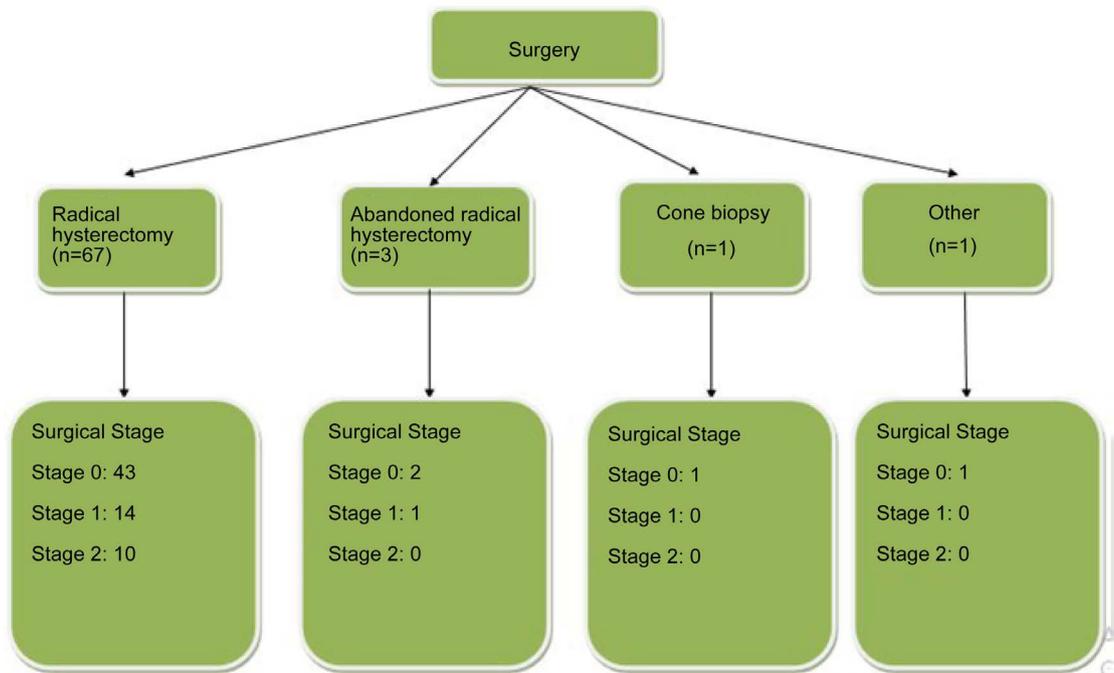


Figure 1. The surgical treatments availed and the surgical stage outcomes.

Flowchart displaying cervical biopsy outcomes and FIGO clinical stages

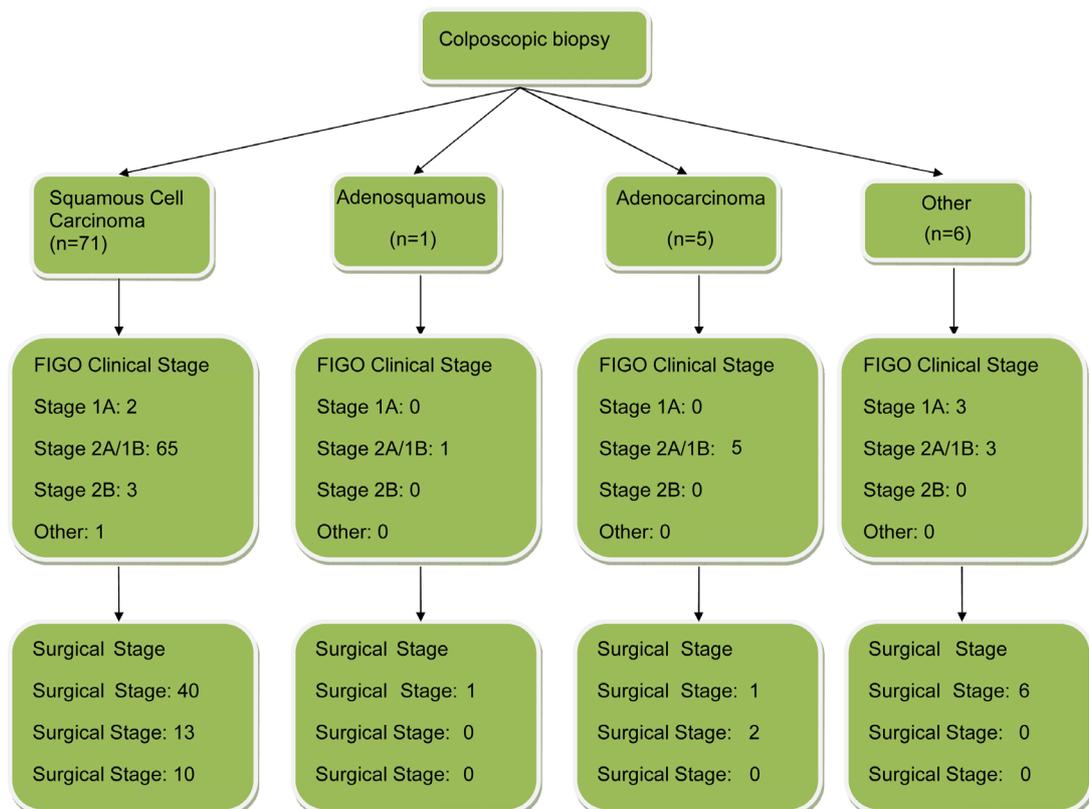


Figure 2. The colposcopic biopsy results and the surgical stage outcomes.

that fits vector generalized and linear additive models appropriate to build the 2 adjacent-category models and the continuation ratio models both with and without proportional odds. We focused on the AIC goodness of fit statistic and the log likelihood ratios to compare the models. The response variable was coded as 0 for surgical Stage 0, 1 for surgical Stage 1 and 2 for surgical Stage 2.

3. Results

3.1. Descriptive Analysis

The data from patients with confirmed invasive cervical cancer was analyzed. The entire dataset had 690 women with confirmed invasive cervical cancer. **Table 1** presents the descriptive statistics for the predictor variables with a chi-square test and an anova test carried out for each categorical and numerical predictor respectively.

The overall mean and median age at first contact with the gynecologists were 46.61 and 46.00 years. The overall mean and median age at the time of surgery were 46.76 and 47.00 years. For the HIV status, 77.6%, 16.5% and 5.88% were found to be HIV negative, HIV positive and unknown HIV status and therefore, all the patients with unknown HIV status were dropped leaving 82.28% and 17.72% being HIV negative and HIV positive respectively. The marital status was classified as either single or married. The single patients comprised of the singles, widows, divorced and those who did not state their marital status. The international FIGO clinical stages were merged into clinical stages 1 and 2 and found to be 78.67% and 21.33% respectively with the clinical stage stated as “others” being dropped. It became clear that on categorizing the FIGO clinical stages as 1 and 2 only, it was found to be statistically significant with a p -value of <0.001 . Whether there was vaginal involvement and parametrial involvement during diagnosis of cervical cancer were found to be statistically significant with p -values of <0.001 and 0.008. The symptoms of vaginal discharge and lower abdominal pain during diagnosis of cervical cancer were found to be statistically significant with p -values of 0.029 and 0.048. All other variables were found to be statistically insignificant.

3.2. Regression Analysis

Comparisons were made based on parameter estimates, log likelihood, residual deviance and AIC for the 3 regression models for ordinal data. Only 5 predictor variables were significantly associated with the response variable: Surgical stage.

3.3. Multinomial Logistic Regression Model

During the analysis, the baseline category was surgical Stage 0. The 3 models fitted were the null model, univariate model and the multivariate model. The aim of the null model was to better understand the marginal distribution of the response variable in the absence of predictors.

Table 1. Descriptive statistics for the predictor variables with a chi-square test and an anova test carried out for each categorical and numerical predictor respectively.

| Surgical Stage | 0 (N = 50) | 1 (N = 15) | 2 (N = 10) | Total (N = 75) | <i>p</i> -value |
|--------------------------------------|----------------------------|----------------------------|----------------------------|----------------------------|-----------------|
| Parity | | | | | 0.615 |
| N-Miss | 4 | 1 | 0 | 5 | |
| Mean (SD) | 4.652 (2.282) | 5.357 (2.274) | 4.800 (2.658) | 4.814 (2.318) | |
| Median (Q1, Q3) | 4.000 (3.000, 6.000) | 5.500 (4.000, 6.000) | 4.000 (3.000, 5.750) | 4.000 (3.000, 6.000) | |
| Range | 0.000 - 10.000 | 2.000 - 10.000 | 2.000 - 11.000 | 0.000 - 11.000 | |
| Clinical Stage | | | | | <0.001 |
| 1 | 43 (86.0%) | 13 (86.7%) | 3 (30.0%) | 59 (78.7%) | |
| 2 | 7 (14.0%) | 2 (13.3%) | 7 (70.0%) | 16 (21.3%) | |
| Age at first clinical contact | | | | | 0.595 |
| N-Miss | 3 | 0 | 0 | 3 | |
| Mean (SD) | 46.979 (11.709) | 45.533 (8.790) | 43.200 (9.762) | 46.153 (10.858) | |
| Median (Q1, Q3) | 46.000 (40.000, 53.500) | 46.000 (43.000, 50.500) | 41.500 (37.250, 49.000) | 46.000 (40.000, 53.000) | |
| Range | 22.000 - 81.000 | 24.000 - 59.000 | 27.000 - 59.000 | 22.000 - 81.000 | |
| HIV Status | | | | | 0.509 |
| Negative | 42 (84.0%) | 13 (86.7%) | 7 (70.0%) | 62 (82.7%) | |
| Positive | 8 (16.0%) | 2 (13.3%) | 3 (30.0%) | 13 (17.3%) | |
| Vaginal Involvement | | | | | <0.001 |
| No | 48 (96.0%) | 14 (93.3%) | 5 (50.0%) | 67 (89.3%) | |
| Yes | 2 (4.0%) | 1 (6.7%) | 5 (50.0%) | 8 (10.7%) | |
| Parametrial Involvement | | | | | 0.008 |
| No | 49 (98.0%) | 13 (86.7%) | 7 (70.0%) | 69 (92.0%) | |
| Yes | 1 (2.0%) | 2 (13.3%) | 3 (30.0%) | 6 (8.0%) | |
| Marital Status | | | | | 0.757 |
| N-Miss | 5 | 0 | 0 | 5 | |
| Married | 37 (82.2%) | 11 (73.3%) | 8 (80.0%) | 56 (80.0%) | |
| Single | 8 (17.8%) | 4 (26.7%) | 2 (20.0%) | 14 (20.0%) | |
| Weight | | | | | 0.690 |
| Mean (SD) | 66.200 (27.117) | 63.867 (22.624) | 58.900 (10.888) | 64.760 (24.584) | |
| Median (Q1, Q3) | 68.500 (56.500, 80.000) | 67.000 (62.000, 75.500) | 55.500 (51.750, 64.500) | 67.000 (55.500, 77.500) | |
| Range | 0.000 - 163.000 | 0.000 - 102.000 | 45.000 - 79.000 | 0.000 - 163.000 | |

Continued

| | | | | | |
|-------------------------------|-------------|-------------|-------------|-------------|-------|
| Smoker | | | | | 0.526 |
| No | 49 (98.0%) | 14 (93.3%) | 10 (100.0%) | 73 (97.3%) | |
| Yes | 1 (2.0%) | 1 (6.7%) | 0 (0.0%) | 2 (2.7%) | |
| Contraception: None | | | | | 0.725 |
| No | 36 (72.0%) | 11 (73.3%) | 6 (60.0%) | 53 (70.7%) | |
| Yes | 14 (28.0%) | 4 (26.7%) | 4 (40.0%) | 22 (29.3%) | |
| Contraception: Condoms | | | | | |
| No | 50 (100.0%) | 15 (100.0%) | 10 (100.0%) | 75 (100.0%) | |
| Yes | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | |
| Contraception: | | | | | 0.274 |
| Intrauterine Device | | | | | |
| No | 45 (90.0%) | 12 (80.0%) | 10 (100.0%) | 67 (89.3%) | |
| Yes | 5 (10.0%) | 3 (20.0%) | 0 (0.0%) | 8 (10.7%) | |
| Contraception: | | | | | 0.122 |
| Oral Pill | | | | | |
| No | 46 (92.0%) | 12 (80.0%) | 7 (70.0%) | 65 (86.7%) | |
| Yes | 4 (8.0%) | 3 (20.0%) | 3 (30.0%) | 10 (13.3%) | |
| Contraception: | | | | | 0.838 |
| Depo Provera | | | | | |
| No | 37 (74.0%) | 12 (80.0%) | 7 (70.0%) | 56 (74.7%) | |
| Yes | 13 (26.0%) | 3 (20.0%) | 3 (30.0%) | 19 (25.3%) | |
| Method of Cancer | | | | | 0.146 |
| Detection | | | | | |
| N-Miss | 1 | 1 | 1 | 3 | |
| Incidental | 2 (4.1%) | 1 (7.1%) | 0 (0.0%) | 3 (4.2%) | |
| Screening | 18 (36.7%) | 2 (14.3%) | 0 (0.0%) | 20 (27.8%) | |
| Symptoms | 25 (51.0%) | 10 (71.4%) | 9 (100.0%) | 44 (61.1%) | |
| Via | 4 (8.2%) | 1 (7.1%) | 0 (0.0%) | 5 (6.9%) | |
| Cervical Biopsy | | | | | 0.245 |
| Pathology Result | | | | | |
| N-Miss | 2 | 0 | 0 | 2 | |
| Adeno Carcinoma | 1 (2.1%) | 2 (13.3%) | 0 (0.0%) | 3 (4.1%) | |
| Adeno Squamous | 1 (2.1%) | 0 (0.0%) | 0 (0.0%) | 1 (1.4%) | |
| Other | 6 (12.5%) | 0 (0.0%) | 0 (0.0%) | 6 (8.2%) | |
| Squamous Cell | 40 (83.3%) | 13 (86.7%) | 10 (100.0%) | 63 (86.3%) | |

Continued

| | | | | | |
|--------------------------------|------------|------------|-------------|------------|-------|
| Surgery Done | | | | | 0.940 |
| N-Miss | 3 | 0 | 0 | 3 | |
| Abandoned Radical Hysterectomy | 2 (4.3%) | 1 (6.7%) | 0 (0.0%) | 3 (4.2%) | |
| Cone Biopsy | 1 (2.1%) | 0 (0.0%) | 0 (0.0%) | 1 (1.4%) | |
| Other | 1 (2.1%) | 0 (0.0%) | 0 (0.0%) | 1 (1.4%) | |
| Radical Hysterectomy | 43 (91.5%) | 14 (93.3%) | 10 (100.0%) | 67 (93.1%) | |
| Symptom: Bleeding | | | | | 0.169 |
| No | 21 (42.0%) | 3 (20.0%) | 2 (20.0%) | 26 (34.7%) | |
| Yes | 29 (58.0%) | 12 (80.0%) | 8 (80.0%) | 49 (65.3%) | |
| Symptom: Discharge | | | | | 0.029 |
| No | 34 (68.0%) | 5 (33.3%) | 4 (40.0%) | 43 (57.3%) | |
| Yes | 16 (32.0%) | 10 (66.7%) | 6 (60.0%) | 32 (42.7%) | |
| Symptom: Pain | | | | | 0.048 |
| No | 34 (68.0%) | 5 (33.3%) | 5 (50.0%) | 44 (58.7%) | |
| Yes | 16 (32.0%) | 10 (66.7%) | 5 (50.0%) | 31 (41.3%) | |
| Tumour Location | | | | | 0.347 |
| N-Miss | 31 | 6 | 5 | 42 | |
| Both | 7 (36.8%) | 6 (66.7%) | 3 (60.0%) | 16 (48.5%) | |
| Endo-cervix | 6 (31.6%) | 0 (0.0%) | 0 (0.0%) | 6 (18.2%) | |
| Exo-cervix | 5 (26.3%) | 3 (33.3%) | 2 (40.0%) | 10 (30.3%) | |
| None | 1 (5.3%) | 0 (0.0%) | 0 (0.0%) | 1 (3.0%) | |
| Grade | | | | | 0.576 |
| N-Miss | 2 | 0 | 0 | 2 | |
| Grade 1 | 5 (10.4%) | 0 (0.0%) | 0 (0.0%) | 5 (6.8%) | |
| Grade 2 | 12 (25.0%) | 4 (26.7%) | 2 (20.0%) | 18 (24.7%) | |
| Grade 3 | 11 (22.9%) | 4 (26.7%) | 2 (20.0%) | 17 (23.3%) | |
| Grade Not Stated | 16 (33.3%) | 7 (46.7%) | 6 (60.0%) | 29 (39.7%) | |
| N/A | 4 (8.3%) | 0 (0.0%) | 0 (0.0%) | 4 (5.5%) | |
| Symptoms Duration | | | | | 0.322 |
| N-Miss | 23 | 3 | 2 | 28 | |
| <1 Months | 2 (7.4%) | 0 (0.0%) | 0 (0.0%) | 2 (4.3%) | |
| <1 Year | 3 (11.1%) | 1 (8.3%) | 3 (37.5%) | 7 (14.9%) | |
| <6 Months | 8 (29.6%) | 2 (16.7%) | 3 (37.5%) | 13 (27.7%) | |
| >1 Year | 1 (3.7%) | 2 (16.7%) | 1 (12.5%) | 4 (8.5%) | |
| N/A | 2 (7.4%) | 0 (0.0%) | 0 (0.0%) | 2 (4.3%) | |
| Not Stated | 11 (40.7%) | 7 (58.3%) | 1 (12.5%) | 19 (40.4%) | |

Table 2 shows a univariate model which was fitted with only the international FIGO clinical stage as the predictor. There were 2 distinct rows of output representing surgical Stage 1 and surgical Stage 2. Each row corresponds 1 model equation. The output shows that the log odds of being in surgical Stage 1 compared to the baseline category surgical Stage 0 decreased by 0.06 if moving from clinical Stage 1 to clinical Stage 2 and the log odds of being in surgical Stage 2 compared to the baseline category surgical Stage 0 increased by 2.66 if moving from clinical Stage 1 to clinical Stage 2. The intercepts show the log odds for the baseline category surgical Stage 0.

Table 3 shows a multivariate ML model which was fitted with 5 statistically significant predictors.

Table 2. Summary for a univariate ML model built with the inclusion of the FIGO clinical stage predictor.

| | Coefficient | Standard Errors | z-statistic | p-value |
|------------------|-------------|-----------------|-------------|---------|
| Surgical Stage 1 | | | | |
| Intercept | -1.20 | 0.32 | -3.78 | 0.00016 |
| Clinical Stage 2 | -0.06 | 0.86 | -0.07 | 0.94757 |
| Surgical Stage 2 | | | | |
| Intercept | -2.66 | 0.60 | -4.46 | 0.00001 |
| Clinical Stage 2 | 2.66 | 0.80 | 3.32 | 0.00089 |

Table 3. Summary for a multivariate ML model built with the inclusion of the 5 statistically significant predictors.

| | Coefficient | Standard Errors | z-statistic | p-value |
|------------------------------|-------------|-----------------|-------------|---------|
| Surgical Stage 1 | | | | |
| Intercept | -1.20 | 0.32 | -3.78 | 0.00016 |
| Clinical Stage 2 | -0.06 | 0.86 | -0.07 | 0.94757 |
| Vaginal Involvement: Yes | -0.359 | 1.623 | -0.221 | 0.82509 |
| Parametrial Involvement: Yes | 1.509 | 1.354 | 1.114 | 0.26528 |
| Symptoms Discharge: Yes | 1.261 | 0.657 | 1.919 | 0.05498 |
| Symptoms Pain: Yes | 1.209 | 0.659 | 1.835 | 0.06651 |
| Surgical Stage 2 | | | | |
| Intercept | -3.494 | 0.901 | -3.878 | 0.00011 |
| Clinical Stage 2 | 2.401 | 1.056 | 2.274 | 0.02297 |
| Vaginal Involvement: Yes | 1.061 | 1.183 | 0.897 | 0.36972 |
| Parametrial Involvement: Yes | 2.911 | 1.538 | 1.893 | 0.05836 |
| Symptoms Discharge: Yes | 0.934 | 0.910 | 1.026 | 0.30489 |
| Symptoms Pain: Yes | -0.155 | 0.975 | -0.159 | 0.89367 |

The first group compares surgical Stage 1 to the reference category which is surgical Stage 0. Based on the p -values of 0.05498, only the women displaying symptomatic vaginal discharge had an effect on the surgical stage outcome. The second group compares the surgical Stage 2 to the reference category whereby only the FIGO clinical stage had a statistically significant effect based on the p -value of 0.02297.

The log odds of being in surgical Stage 1 compared to the surgical Stage 0 will increase by 0.006 if moving from clinical Stage 1 to clinical Stage 2 and the log odds of being in surgical Stage 2 compared to surgical Stage 0 will increase by 2.401 if moving from clinical Stage 1 to clinical Stage 2. Thus, FIGO clinical stage exhibited positive regression coefficients and likely to increase with the higher categories of surgical stage.

The log odds of being in surgical Stage 1 compared to surgical Stage 0 decreased by 0.359 if there was vaginal involvement observed during diagnosis and the log odds of being in surgical Stage 2 compared to surgical Stage 0 increased by 1.061 if there was vaginal involvement observed during diagnosis.

The log odds of being in surgical Stage 1 compared to surgical Stage 0 increased by 1.509 and the log odds of being in surgical Stage 2 compared to surgical Stage 0 increased by 2.911 if the parametrium region was affected by the cervical cancer. The positive regression coefficients indicated that observed parametrial involvement was likely to lead to a higher category of surgical stage.

The log odds of being in surgical Stage 1 compared to surgical Stage 0 increased by 1.261 and by 1.209 when a patient displayed symptomatic vaginal discharge and lower abdominal pain respectively. The log odds of being in surgical Stage 2 compared to surgical Stage 0 decreased by 0.934 and decreased by 0.155 when a patient displayed symptomatic vaginal discharge and lower abdominal pain respectively.

Table 4 clearly indicates that the full model with the 5 statistically significant predictor variables had the lowest AIC and residual deviance of 121.72 and 97.72 respectively with the highest log likelihood ratio of -48.860 . Thus, the multivariate ML model was a better fit for the cervical cancer data compared to the univariate and null ML models.

With reference to **Table 5**, the odds of being classified into surgical Stage 1 over surgical Stage 0 was 1.01 [CI: 0.13 - 7.61] higher for patients diagnosed with FIGO clinical Stage 2 versus those diagnosed with FIGO clinical Stage 1 while holding all other predictors constant. In contrast, the odds of being classified into surgical

Table 4. Summary of deviance, AIC and log-likelihood outputs.

| | Deviance | Log Likelihood | AIC |
|-----------------------|----------|----------------|--------|
| Null ML model | 115.87 | -57.94 | 123.87 |
| Univariate ML model | 129.13 | -64.56 | 133.13 |
| Multivariate ML model | 97.72 | -48.86 | 121.72 |

Table 5. The table of the odds ratios extracted from the multivariate multinomial logistic model which displayed the best fit model for the cervical cancer data.

| | Intercept | Clinical Stage2 | Vaginal Involvement | Parametrial Involvement | Symptoms Discharge | Symptoms Pain |
|-------------|-------------------|--------------------|---------------------|-------------------------|--------------------|--------------------|
| | Surgical Stage 1 | | | | | |
| Coefficient | -2.499 | 0.007 | -0.359 | 1.509 | 1.261 | 1.209 |
| Std Error | 0.622 | 1.032 | 1.623 | 1.354 | 0.657 | 0.659 |
| z-statistic | -4.018 | 0.007 | -0.221 | 1.114 | 1.919 | 1.835 |
| p-value | <0.01 | 0.9944 | 0.82509 | 0.26528 | 0.05498 | 0.06651 |
| OR (95% CI) | 0.08 (0.02, 0.28) | 1.01 (0.13, 7.61) | 0.07 (0.03, 16.83) | 4.52 (0.32, 64.28) | 3.53 (0.97, 12.78) | 3.35 (0.92, 12.18) |
| | Surgical Stage 2 | | | | | |
| Coefficient | -3.494 | 2.401 | 1.061 | 2.911 | 0.934 | -0.155 |
| Std Error | 0.901 | 1.056 | 1.183 | 1.538 | 0.91 | 0.975 |
| z-statistic | -3.878 | 2.274 | 0.897 | 1.893 | 1.026 | -0.159 |
| p-value | <0.001 | 0.02297 | 0.36972 | 0.05836 | 0.30489 | 0.87367 |
| OR (95% CI) | 0.03 (0.01,0.18) | 11.03 (1.39,87.36) | 2.89 (0.28,29.35) | 18.38 (0.9,374.72) | 2.54 (0.43,15.13) | 0.86 (0.13, 5.79) |

Stage 2 over surgical Stage 0 was 11.03 CI: [1.39 - 87.36] times lower for patients diagnosed with FIGO clinical Stage 2 versus those diagnosed with FIGO clinical Stage 1 while holding all other predictors constant.

The odds of being classified into surgical Stage 1 over surgical Stage 0 was 0.70 CI: [0.03 - 16.83] times lower and in contrast, the odds of being in surgical Stage 2 over surgical Stage 0 was 2.89 CI: [0.28 - 29.35] times higher for patients with the vaginal region observed to be affected by the cancer during diagnosis versus those without any vaginal involvement while holding all other predictors constant.

The odds of being classified into surgical Stage 1 over surgical Stage 0 was 4.52 CI: [0.32 - 64.28] times higher and the odds of being classified into surgical Stage 2 over surgical Stage 0 was 18.38 CI: [0.9 - 374.72] times higher for patients with the parametrial region affected by the cancer versus those without any parametrial involvement while holding other predictors constant.

The odds of being classified into surgical Stage 1 over surgical Stage 0 was 3.53 CI: [0.97 - 12.78] times higher and the odds of being classified into surgical Stage 2 over surgical Stage 0 was 2.54 CI: [0.43 - 15.13] times higher for the patients with symptomatic vaginal discharge during diagnosis versus those without the symptomatic vaginal discharge, holding all other predictors constant.

The odds of being classified into surgical Stage 1 over surgical Stage 0 was 3.35 CI: [0.92 - 12.18] times higher and in contrast, the odds of being into surgical Stage 2 over surgical Stage 0 was 0.86 CI: [0.13 - 5.79] times lower for the patients with symptomatic lower abdominal pain versus those without any pain, holding all other predictors constant.

3.4. The Continuation Ratio Model

When the focus is on a particular category given that a patient must pass through a lower surgical stage category before achieving a higher category, the continuation ratio model is considered a more appropriate choice. The proportional odds assumption was tested by fitting this particular model with and without the proportional odds assumption.

Table 6 shows the univariate CR model with and without proportional odds output whereby the FIGO clinical stage was the only predictor. For the CR model with proportional odds, it is clear that the FIGO clinical stage had a significant effect on the surgical stage response with a p -value of 0.000892. The estimated log regression coefficient for the FIGO clinical stage, $\beta = 1.649(0.496)$, $z = 3.322$ and $p < 0.05$ showed that the FIGO clinical stage upon diagnosis had a positive effect on the surgical stage responses.

The CR model without proportional odds gave separate effects. The FIGO clinical stage predictor variable was found to be statistically significant. For surgical Stage 1, the estimated logit regression coefficient for FIGO clinical stage was $\beta = 1.240(0.583)$, z -value = 2.127 and a p -value of 0.0334 indicating that the FIGO clinical stage had a significant positive effect on the surgical Stage 1 responses. In addition, for surgical Stage 2, the estimated logit regression coefficient for FIGO clinical stage was $\beta = 2.719(1.026)$, z -value = 2.650 and a p -value of 0.00806 indicating that the FIGO clinical stage had a significant positive effect on surgical Stage 2 responses. **Table 7** below shows the multivariate CR model with and without proportional odds output when the 5 predictors that were found to be statistically significant were included. For the CR multivariate model with proportional odds, only the FIGO clinical stage estimated log regression coefficient, $\beta = 1.449(0.632)$, z -value = 2.293 had a positive effect on the surgical stage responses. In addition, a p -value of 0.02182 showed that the FIGO clinical stage is a statistically significant predictor. The remaining 4 predictors that were not statistically significant to the surgical stage responses were vaginal involvement, parametrial involvement, symptomatic vaginal discharge and lower abdominal pain.

Table 6. The Summary for a CR univariate model with the inclusion of the FIGO clinical stage predictor.

| | Coefficient | Standard Errors | z -statistic | p -value |
|------------------------------------|-------------|-----------------|----------------|------------|
| CR Model with Proportional Odds | | | | |
| Intercept 1 | -1.092 | 0.288 | -3.790 | <0.00015 |
| Intercept 2 | -1.042 | 0.492 | -2.115 | <0.03443 |
| Clinical Stage 2 | 1.649 | 0.496 | 3.322 | 0.00089 |
| CR Model without Proportional Odds | | | | |
| Intercept 1 | -0.989 | 0.293 | -3.376 | <0.00074 |
| Intercept 2 | -1.466 | 0.641 | -2.289 | 0.02206 |
| Clinical Stage 1 | 1.240 | 0.583 | 2.127 | 0.03339 |
| Clinical Stage 2 | 2.719 | 1.026 | 2.650 | 0.00806 |

Table 7. The Summary for a CR multivariate model with the inclusion of the 5 predictors for Surgical Stage 1 and Stage 2.

| | Coefficient | Standard Errors | z-statistic | p-value |
|------------------------------------|-------------|-----------------|-------------|---------|
| CR Model with Proportional Odds | | | | |
| Intercept 1 | -1.773 | 0.441 | -4.021 | < 0.01 |
| Intercept 2 | -2.254 | 0.698 | -3.229 | < 0.01 |
| Clinical Stage 2 | 1.449 | 0.632 | 2.293 | 0.022 |
| Vaginal Involvement: Yes | 0.982 | 0.868 | 1.131 | 0.258 |
| Parametrial Involvement: Yes | 1.61 | 0.882 | 1.825 | 0.068 |
| Symptoms Discharge: Yes | 0.717 | 0.498 | 1.438 | 0.1504 |
| Symptoms Pain: Yes | 0.382 | 0.507 | 0.754 | 0.4508 |
| CR Model without Proportional Odds | | | | |
| Intercept 1 | -2.05 | 0.512 | -4.001 | <0.01 |
| Intercept 2 | -1.195 | 1.239 | -0.965 | 0.335 |
| Clinical Stage 1 | 1.001 | 0.787 | 1.271 | 0.204 |
| Clinical Stage 2 | 3.833 | 1.817 | 2.109 | 0.035 |
| Vaginal Involvement: Yes 1 | 0.736 | 1.161 | 0.634 | 0.526 |
| Vaginal Involvement: Yes 2 | 1.584 | 1.699 | 0.933 | 0.351 |
| Parametrial Involvement: Yes 1 | 1.829 | 1.223 | 1.495 | 0.131 |
| Parametrial Involvement: Yes 2 | 3.220 | 1.918 | 1.679 | 0.093 |
| Symptoms Discharge: Yes 1 | 1.103 | 0.568 | 1.943 | 0.052 |
| Symptoms Discharge: Yes 2 | -1.412 | 1.542 | -0.916 | 0.360 |
| Symptoms Pain: Yes 1 | 0.823 | 0.577 | 1.427 | 0.154 |
| Symptoms Pain: Yes 2 | -0.832 | 1.515 | -0.209 | 0.227 |

For the CR multivariate model without proportional odds, we got separate effects for the surgical stage responses. For surgical Stage 1, the symptomatic vaginal discharge was found to have a positive effect on surgical Stage 2 response and was significant with estimated logit coefficient $\beta = 1.103(0.568)$, z -value = 1.943 and a p -value of 0.052. The estimated logit coefficients for FIGO clinical stage, vaginal involvement, parametrial involvement and symptomatic lower abdominal pain were not statistically significant and thus, had no effect on the surgical Stage 1 responses.

Also, it was clear that for surgical Stage 2 response, the FIGO clinical stage had a positive effect and was statistically significant with an estimated logit coefficient $\beta = 3.833(1.817)$, z -value = 2.109 and p -value of 0.0349. The estimated logit coefficients for vaginal involvement, parametrial involvement, symptomatic vaginal discharge and symptomatic lower abdominal pain were not statistically significant and thus, had no effect on the surgical Stage 2 responses.

The 2 CR models with and without proportional odds were compared to determine the model best fit for the cervical cancer data. The fitted multivariate CR model with proportional odds had a misclassification rate of 29.33% and 37.74% whereas the fitted multivariate CR model without proportional odds had a misclassification rate of 30.67% and 39.09% when the train and validation data sets were utilized respectively. An AIC of 118.899 shows that the multivariate CR model without proportional odds gave the best fit for the cervical cancer data with further confirmation based on a residual deviance and log likelihood ratio of 94.89 and -47.44 respectively. Moreover, the VGAM likelihood ratio test was carried out for the 2 CR multivariate models and a chi-square p -value of 0.08023 showed that the fit was not significantly different and thus, the multivariate CR model without proportional odds was found to be adequate.

Table 8 and **Table 9** show the goodness of fit statistics for the 2 CR multivariate models.

Equation (7) and (8) shows the multivariate CR model without proportional odds assumptions for surgical Stage 1 and surgical Stage 2.

$$\begin{aligned} \log [P(SS = 1 | SS \geq 1)] &= -2.050(0.512) + 1.001(0.787)\text{clinicalstage} \\ &+ 0.736(1.161)\text{Vaginalinvolvement} \\ &+ 1.829(1.223)\text{Parametrialinvolvement} \\ &+ 1.103(0.568)\text{Symptom:Discharge} \\ &+ 0.823(0.577)\text{Symptom : Pain} \end{aligned} \tag{7}$$

$$\begin{aligned} \log [P(SS = 2 | SS \geq 2)] &= -1.1950(1.239) + 3.833(1.817)\text{clinicalstage} \\ &+ 1.584(1.699)\text{Vaginalinvolvement} \\ &+ 3.220(1.918)\text{Parametrialinvolvement} \\ &- 1.412(1.542)\text{Symptom : Discharge} \\ &- 0.832(1.515)\text{Symptom : Pain} \end{aligned} \tag{8}$$

Table 8. Goodness of fit statistics for the CR models with proportional odds.

| | Deviance | Log Likelihood | AIC |
|-----------------------|----------|----------------|--------|
| Null CR model | 129.13 | -64.56 | 133.13 |
| Univariate CR model | 117.56 | -58.78 | 123.56 |
| Multivariate CR model | 104.72 | -52.36 | 118.72 |

Table 9. Goodness of fit statistics for the CR models without proportional odds.

| | Deviance | Log Likelihood | AIC |
|-----------------------|----------|----------------|--------|
| Null CR model | 129.13 | -64.56 | 133.13 |
| Univariate CR model | 115.87 | -57.94 | 123.87 |
| Multivariate CR model | 94.89 | -47.44 | 118.89 |

Table 10 shows the odds ratios extracted from the CR multivariate model without proportional odds.

A brief summary of the odds ratios for the model $\text{logit}[P(SS = 1 | SS \geq 1)]$ is given below:

The odds of having an outcome greater than surgical Stage 1 relative to being in surgical Stage 1 was 2.72 times higher among the patients diagnosed with FIGO clinical Stage 2 compared to the patients diagnosed with FIGO clinical Stage 1, after controlling for the effects of other predictors in the model. The odds of having an outcome greater than surgical Stage 1 relative to being in surgical Stage 1 was 2.09 times higher among the patients considered to have the vaginal region affected by the cancer (vaginal involvement) compared to the patients without any vaginal involvement after controlling for the effects of other predictors in the model. The odds of having an outcome greater than surgical Stage 1 relative to being in surgical Stage 1 was 6.23 times higher among the patients considered to have the parametrium region affected by the cervical cancer (parametrial involvement) compared to the patients without any parametrial involvement after controlling for the effects of other predictors in the model. The odds of having an outcome greater than surgical Stage 1 relative to being in surgical Stage 1 was 3.01 times higher among the patients with symptomatic vaginal discharge (Symptoms: Discharge) compared to the patients who did not have symptomatic vaginal discharge after controlling for the effects of other predictors in the model. The odds of having an outcome greater than surgical Stage 1 relative to being in surgical Stage 1 was 2.28 times higher among the patients displaying symptomatic lower abdominal pain (Symptoms: Pain) compared to the patients without symptomatic lower abdominal pain after controlling for the effects of other predictors in the model.

Table 10. The odds ratios for the CR multivariate model without proportional odds were extracted.

| | Odds ratio | 2.5% | 97.5% |
|--------------------------------|------------|------|---------|
| (Intercept): 1 | 0.13 | 0.05 | 0.35 |
| (Intercept): 2 | 0.30 | 0.03 | 3.43 |
| Clinical Stage 2: 1 | 2.72 | 0.58 | 12.72 |
| Clinical Stage 2: 2 | 46.20 | 1.31 | 1627.54 |
| Vaginal Involvement Yes: 1 | 2.09 | 0.21 | 20.30 |
| Vaginal Involvement Yes: 2 | 4.88 | 0.17 | 136.11 |
| Parametrial Involvement Yes: 1 | 6.23 | 0.57 | 68.45 |
| Parametrial Involvement Yes: 2 | 25.02 | 0.58 | 1073.25 |
| Symptoms Discharge Yes: 1 | 3.01 | 0.99 | 9.16 |
| Symptoms Discharge Yes: 2 | 0.24 | 0.01 | 5.00 |
| Symptoms Pain Yes: 1 | 2.28 | 0.74 | 7.06 |
| Symptoms Pain Yes: 2 | 0.16 | 0.01 | 3.12 |

A brief summary of the odds ratios for the model $\text{logit}[P(SS = 2 | SS \geq 2)]$ is given in **Table 11**. The odds of having an outcome greater than surgical Stage 2 relative to being in surgical Stage 2 was 46.20 times higher among the patients diagnosed with FIGO clinical Stage 2 compared to the patients diagnosed with FIGO clinical Stage 1, after controlling for the effects of other predictors in the model. The odds of having an outcome greater than surgical Stage 2 relative to being in surgical Stage 2 was 4.88 times higher among the patients considered to have the vaginal region affected by the cancer (vaginal involvement) compared to the patients without any vaginal involvement after controlling for the effects of other predictors in the model. The odds of having an outcome greater than surgical Stage 2 relative to being in surgical Stage 2 was 25.02 times higher among the patients considered to have the parametrium region affected by the cervical cancer (parametrial involvement) compared to the patients without any parametrial involvement after controlling for the effects of other predictors in the model. The odds of having an outcome greater than surgical Stage 2 relative to being in surgical Stage 2 was 0.24 times lower among the patients with symptomatic vaginal discharge (Symptoms: Discharge) compared to the patients who did not have symptomatic vaginal discharge after controlling for the effects of other predictors in the model. The odds of having an outcome greater than surgical Stage 2 relative to being in surgical Stage 2 was 0.16 times lower among the patients displaying symptomatic lower abdominal pain (Symptoms: Pain) compared to the patients without symptomatic lower abdominal pain after controlling for the effects of other predictors in the model.

3.5. Adjacent Category Logistic Model

The Adjacent Category Logit model is a special form of generalized logit models that involves the simultaneous estimation of the effects of predictor variables in pairs of adjacent categories. The ACL model involves the ratio of two probabilities $P[Y = y_j]$ and $P[Y = y_{j+1}]$. The proportional odds assumption was tested by fitting the ACL model with and without the proportional odds assumption. **Table 12** and **Table 13** show the summary of the ACL univariate and multivariate model with and without proportional odds respectively.

For the ACL model with proportional odds, we found that the FIGO clinical stage had a statistically significant effect on the surgical stage response with a p -value of 0.00207. The estimated logit regression coefficient for the FIGO clinical stage, $\beta = -1.1740$, z -value = -3.080 and a p -value < 0.05 showed that the FIGO clinical stage upon diagnosis had a negative effect on each adjacent surgical stage response category. The ACL model without proportional odds gave separate effects. We found that the estimated logit regression coefficient $\beta = -2.719$, z -value = -2.650 and the p -value = 0.008057 indicated that the log odds of being in surgical Stage 2 versus surgical Stage 1 was -2.719 when the FIGO clinical stage increased by 1 unit, holding all other predictors constant. Thus, the FIGO clinical stage had a significant effect on the probability of being in surgical Stage 2 versus surgical Stage 1. However, the FIGO clinical stage had no effect on the probability of being in surgical Stage 1 versus surgical Stage 0.

Table 11. The Summary for an ACL null model.

| | Coefficient | Standard Errors | z-statistic | p-value |
|-------------------------------------|-------------|-----------------|-------------|---------|
| ACL Model with Proportional Odds | | | | |
| Intercept1 | 1.204 | 0.294 | 4.090 | <0.0000 |
| Intercept2 | 0.405 | 0.408 | 0.993 | <0.3206 |
| ACL Model without Proportional Odds | | | | |
| Intercept1 | 1.204 | 0.294 | 4.090 | <0.0000 |
| Intercept2 | 0.405 | 0.408 | 0.993 | <0.3206 |

Table 12. The Summary for an ACL univariate model with the inclusion of the FIGO clinical stage predictor.

| | Coefficient | Standard Errors | z-statistic | p-value |
|-------------------------------------|-------------|-----------------|-------------|---------|
| ACL Model with Proportional Odds | | | | |
| Intercept1 | 1.434 | 0.306 | 4.688 | <0.0000 |
| Intercept2 | 0.914 | 0.485 | 1.887 | <0.0592 |
| Clinical Stage 2 | -1.174 | 0.381 | -3.080 | 0.0021 |
| ACL Model without Proportional Odds | | | | |
| Intercept1 | 1.196 | 0.317 | 3.779 | <0.0002 |
| Intercept2 | 1.466 | 0.640 | 2.289 | 0.0221 |
| Clinical Stage 1 | 0.057 | 0.862 | 0.066 | 0.9478 |
| Clinical Stage 2 | -2.719 | 1.026 | -2.650 | 0.0081 |

For the ACL multivariate model with proportional odds, only the FIGO clinical stage estimated logit regression coefficient $\beta = -1.044(0.509)$, z -value = -2.05 had a negative effect on the surgical stage responses. Moreover, a p -value of 0.04036 is a statistically significant predictor. As with the continuation ratio model, the remaining 4 predictors that were not statistically significant to the surgical stage responses were vaginal involvement, parametrial involvement, symptomatic vaginal discharge and lower abdominal pain.

For the ACL multivariate model without proportional odds, we get separate effects for the surgical stage responses. The FIGO clinical stage had a negative effect on the probability of being classified under surgical Stage 2 versus surgical Stage 1. The estimated logit regression coefficient $\beta = -2.349(1.258)$, z -value = -1.903 and a p -value of 0.057 indicates that it is an insignificant predictor and the log-odds of being classified under surgical Stage 2 versus surgical Stage 1 was -2.394 when the FIGO clinical stage increased by 1 unit, holding all other predictors constant. In addition, the symptomatic vaginal discharge predictor had a negative effect on the on the probability of being classified under surgical Stage 1 versus surgical Stage 0. The 3 ACL models with and without proportional odds

Table 13. The Summary for an ACL multivariate model with the inclusion of the 5 predictors for Surgical Stage 1 and Stage 2.

| | Coefficient | Standard Errors | z-statistic | p-value |
|-------------------------------------|-------------|-----------------|-------------|----------|
| ACL Model with Proportional Odds | | | | |
| Intercept 1 | 1.937 | 0.397 | 4.874 | < 0.0000 |
| Intercept 2 | 1.861 | 0.674 | 2.760 | < 0.0058 |
| Clinical Stage 2 | -1.044 | 0.509 | -2.050 | 0.0404 |
| Vaginal Involvement: Yes | -0.624 | 0.661 | -0.943 | 0.3456 |
| Parametrial Involvement: Yes | -1.192 | 0.676 | -1.762 | 0.078 |
| Symptoms Discharge: Yes | -0.668 | 0.409 | -1.635 | 0.1021 |
| Symptoms Pain: Yes | -0.330 | 0.410 | -0.806 | 0.4200 |
| ACL Model without Proportional Odds | | | | |
| Intercept 1 | 2.499 | 0.622 | 4.017 | < 0.0001 |
| Intercept 2 | 0.994 | 1.044 | 0.953 | 0.3407 |
| Clinical Stage 1 | -0.006 | 1.032 | -0.006 | 0.9950 |
| Clinical Stage 2 | -2.394 | 1.258 | -1.903 | 0.057 |
| Vaginal Involvement: Yes 1 | 0.358 | 1.623 | 0.221 | 0.8253 |
| Vaginal Involvement: Yes 2 | -1.420 | 1.552 | -0.915 | 0.360 |
| Parametrial Involvement: Yes 1 | -1.509 | 1.354 | -1.114 | 0.2652 |
| Parametrial Involvement: Yes 2 | -1.403 | 1.444 | -0.972 | 0.3313 |
| Symptoms Discharge: Yes 1 | -1.261 | 0.657 | -1.920 | 0.0549 |
| Symptoms Discharge: Yes 2 | 0.327 | 1.027 | 0.319 | 0.7501 |
| Symptoms Pain: Yes 1 | -1.209 | 0.659 | -1.835 | 0.067 |
| Symptoms Pain: Yes 2 | 1.364 | 1.063 | 1.283 | 0.1996 |

were compared to determine the model best fit for the cervical cancer data. The multivariate ACL model with proportional odds had a misclassification rate of 32.00% and 37.32% whereas the multivariate ACL model without proportional odds had a misclassification rate of 29.33% and 37.03% when the train and validation data sets were utilized respectively. There was an increase in misclassification by 5.32% and 7.70%.

Table 14 and **Table 15** show the goodness of fit statistics for the ACL multivariate models with and without proportional odds respectively.

An AIC of 121.72 indicates that the multivariate ACL model without proportional odds gave the best fit for the cervical cancer data with further confirmation based on a residual deviance and log likelihood ratio of 97.72 and -48.86 respectively. We carried out the likelihood ratio test for the 2 multivariate ACL models and a chi-square *p*-value of 0.002981 indicating that both fits were significantly different from each other.

Table 14. Goodness of fit statistics for the ACL models with proportional odds.

| | Deviance | Log Likelihood | AIC |
|------------------------|----------|----------------|--------|
| Null ACL model | 129.13 | -64.56 | 133.13 |
| Univariate ACL model | 119.03 | -59.51 | 125.03 |
| Multivariate ACL model | 106.13 | -53.07 | 120.13 |

Table 15. Goodness of fit statistics for the ACL models without proportional odds.

| | Deviance | Log Likelihood | AIC |
|------------------------|----------|----------------|--------|
| Null ACR model | 129.13 | -64.56 | 133.13 |
| Univariate ACL model | 115.87 | -57.94 | 123.87 |
| Multivariate ACL model | 97.72 | -48.86 | 121.72 |

Equation (9) and (10) shows the multivariate ACL model without proportional odds assumptions for surgical Stage 1 versus surgical Stage 0 and surgical Stage 2 versus surgical Stage 1.

$$\begin{aligned} \log [P(SS = 0 | SS = 1)] &= 2.499(0.622) - 0.0065(1.032)clinicalstage \\ &+ 0.358(1.632)Vaginalinvolvement \\ &- 1.509(1.354)Parametrialinvolvement \\ &- 1.261(0.657)Symptom : Discharge \\ &- 1.209(0.659)Symptom : Pain \end{aligned} \tag{9}$$

$$\begin{aligned} \log [P(SS = 1 | SS = 2)] &= 0.994(1.044) - 2.394(1.258)clinicalstage \\ &- 1.4199(1.551)Vaginalinvolvement \\ &- 1.403(1.444)Parametrialinvolvement \\ &+ 0.327(1.027)Symptom : Discharge \\ &+ 1.364(1.063)Symptom : Pain \end{aligned} \tag{10}$$

The 3 ACL models with and without proportional odds were compared to determine the model best fit for the cervical cancer data. The multivariate ACL model with proportional odds had a misclassification rate of 32.00% and 37.32% whereas the multivariate ACL model without proportional odds had a misclassification rate of 29.33% and 37.03% when the train and validation datasets were utilized respectively. Clearly, there was an increase in misclassification by 5.32% and 7.70% respectively.

Table 16 shows the summary of the odds ratios for the ACL model without proportional odds. For the model 9 $\log [P(SS = 0 / SS = 1)]$ and model 10 $\log [P(SS = 1 / SS = 2)]$ respectively. For the patients diagnosed with FIGO clinical Stage 2, the odds of being classified into surgical Stage 1 versus surgical Stage 0 was 0.99 [0.13 - 7.51] times lower and the odds of being classified into surgical Stage 2 versus surgical Stage 1 was 0.09 [0.01 - 1.07] times lower than for the patients with FIGO clinical Stage 1, holding all other predictors constant. Additionally,

Table 16. Odds ratios for the ACL model without proportional odds.

| Predictor variables | Odds ratio | CI: 2.5% | CI: 97.5% |
|--------------------------------|------------|----------|-----------|
| (Intercept): 1 | 12.17 | 3.60 | 41.20 |
| (Intercept): 2 | 2.70 | 0.35 | 20.91 |
| Clinical Stage 2: 1 | 0.99 | 0.13 | 7.51 |
| Clinical Stage 2: 2 | 0.09 | 0.01 | 1.07 |
| Vaginal Involvement Yes: 1 | 1.43 | 0.06 | 34.47 |
| Vaginal Involvement Yes: 2 | 0.24 | 0.01 | 5.06 |
| Parametrial Involvement Yes: 1 | 0.22 | 0.02 | 3.14 |
| Parametrial Involvement Yes: 2 | 0.25 | 0.01 | 4.17 |
| Symptoms Discharge Yes: 1 | 0.28 | 0.08 | 1.03 |
| Symptoms Discharge Yes: 2 | 1.39 | 0.19 | 10.39 |
| Symptoms Pain Yes: 1 | 0.30 | 0.08 | 1.09 |
| Symptoms Pain Yes: 2 | 3.91 | 0.49 | 31.42 |

for the patients whose vaginal region was affected by the cervical cancer, the odds of being classified into surgical Stage 1 versus surgical Stage 0 was 1.43 [0.06 - 34.47] times higher and the odds of being classified into surgical Stage 2 versus surgical Stage 1 was 0.24 [0.01 - 5.06] times lower than for the patients without vaginal involvement, holding all other predictors constant. For the patients who had the parametrium affected by the cervical cancer, the odds of being classified under surgical Stage 1 versus surgical Stage 0 was 0.22 [0.02 - 3.14] times lower and the odds of being classified under surgical Stage 2 versus surgical Stage 1 was 0.25 [0.01 - 4.17] times lower than for patients without parametrial involvement, holding other predictors constant. For the patients with symptomatic vaginal discharge, the odds of being classified under surgical Stage 1 versus surgical Stage 0 was 0.28 [0.08 - 1.03] times lower and the odds of being classified under surgical Stage 2 versus surgical Stage 1 was 1.39 [0.19 - 10.39] times higher than for the patients without vaginal discharge, whilst holding other predictors constant. For the patients with symptomatic abdominal pain, the odds of being classified into surgical Stage 1 versus surgical Stage 0 was 0.30 [0.08 - 1.09] times lower and the odds of being classified into surgical Stage 1 versus surgical Stage 0 was 3.91 [0.49 - 31.42] times higher than for the patients without abdominal pain, whilst holding all other predictors constant.

4. Discussion

The aim of cervical cancer screening is to detect the pre-cancerous changes on the cervix which may lead to cancer. The objective of this study was to evaluate the predictive performance of 3 regression models for ordinal responses on the surgical stage of women treated surgically for invasive cervical cancer. The results provide an understanding of the future possibilities of using predictive algorithms in the Kenyan oncology setting. The relationships between the surgical

stage and 5 statistically significant variables were investigated by applying regression models and comparing the odds ratios. The findings showed that the FIGO clinical stage, parametrial involvement, vaginal involvement, symptomatic vaginal discharge and lower abdominal pains are independently associated with the surgical stage.

4.1. Application to Surgically-Treated Cervical Cancer Patients

Results show that among the 3 ordinal regression models, the CR model without proportional odds was found to best classify the surgical stages of the patients with a misclassification rate of 30.67% and 39.09% for the train(original) and test (simulation) set. Although the 3 models are similar in that they fit multiple simultaneous binary logits, there were some restructuring of categories. The CR model fits 2 logits on each consecutive step; in terms of dummy variables, with the increasing “0” category, the “1” category is considered the higher category. The ML model compares each of the surgical Stages 1 to 2 with surgical Stage 0 (the reference category) in 2 simultaneous logit models and the ACL model fits logit models to 2 adjacent pairs of surgical stage categories. The results showed that for each model, the multivariate models took precedence which indicated that a combination of predictors could best determine the surgical stage outcome of a patient prior to surgery. The multivariate CR model without proportional odds presented the lowest AIC value of 118.89 indicating that it would be the best model to select for the cervical cancer data. The study demonstrated a similarity between the ML and ACL model. The multivariate ML and ACL model without proportional odds had similar likelihood ratios of -48.86 whilst the CR model without proportional odds had a likelihood ratio of -47.44 showing that the later model was statistically different from the 2 models. The goodness-of-fit statistics showed that the CR model without proportional odds gave the lowest deviance of 94.89 and a low AIC statistic of 118.72. On analyzing the results, the CR null models with and without proportional odds gave similar coefficients and negligible differences were observed. The univariate and multivariate CR models without proportional odds gave separate effects for each independent variable. Both univariate CR models supported that the FIGO clinical stage did have a significant positive influence on the surgical stage outcomes. Although the CR model without proportional odds gave the lowest deviance and a low AIC statistic, this particular model showed that information on the FIGO clinical stage had a higher predictive influence on the patients with surgical Stage 2 compared to those with FIGO clinical Stage 2.

We compared the odds ratios of the 3 models. The odds ratio is not an absolute number [19]. In addition, odds ratios are simple to compute and can be applied to discrete and continuous explanatory variables [19]. The odds ratios compare the relative odds of the response (in our case, surgical stage), given exposure to explanatory variables of interest. She further expounds that the odds ratios can ascertain whether a particular exposure is a risk factor and compare the magnitude of various risk factors for the specific response. The 95% odds ratios confidence

intervals estimate the precision of the odds ratios and are considered a substitute for the presence of statistical significance when the null value ($OR = 1$) is not overlapped. Low levels of precision are indicated by large confidence intervals whereas high levels of precision are indicated by small confidence intervals. Specifically, the FIGO clinical stage had a higher influence on women whose odds of having a surgical stage greater than surgical Stage 2 relative to being in surgical Stage 2. Though the results gave large confidence intervals indicative of low precision, a statistically significant p -value (0.0349) and confidence intervals that did not span the null value ($OR = 1$) confirmed the result. The OR for the other 4 predictors showed decreased odds of having a surgical stage greater than surgical Stage 2 relative to being in surgical Stage 2. Also, there was decreased odds of having a surgical stage greater than surgical Stage 1 relative to being in surgical Stage 1 with the confidence intervals for the 5 statistically significant predictors spanning the null value ($OR = 1$). Clearly, there was no statistical significance with the regression coefficients having p -values at >0.05 . The likelihood chi-square ratio test showed that the CR model without proportional odds (chi-square p -value = 0.0823) is adequate compared to the CR model with proportional odds.

In our study, the CR model without the proportional odds assumption was the best fit compared to the CR model with proportional odds. Based on the comparison of models, the continuation ratio model, the adjacent category model, the multinomial model and two other models on the ordinal response of hospital length of stay with patient characteristics as covariates were compared. The ordinal regression model, the CR model and the ACL model violated the proportional odds assumption. Moreover, the estimated relative risks of the multinomial model, the cumulative ratio model and the continuation ratio model on blood cancer ordinal responses were compared [20]. The authors determined through the goodness-of-fit statistics, the regression diagnostic analysis, small standard errors and smaller 95% confidence intervals that the CR model was the best fit model for the ordinal responses. The CR model as compared to the ACL model and the baseline category model, the CR model is recognized for being a simple decomposition of a multinomial distribution, its possession of the property of conditional independence between categories and the model's significance levels capability of being affected by a reversal in the order of the categories. A prior study compared the fit of the baseline category model, the proportional odds model and the adjacent category model in determining the prostate cancer stage and found the baseline category model to have the highest DIC [21]. The authors took the investigation further by comparing the baseline category model to a logistic regression model fitted to dichotomized ordinal responses which demonstrated that the baseline category model was a superior fit. At least 50 multinomial events per variable was recommended leading to the MLR predictive performance gradually improving as the number of multinomial events per variable increases [22]. Our study results show that this could be the possible reason for the MLR model estimated by maximum likelihood being the most unlikely choice among the 3 predictive mechanisms.

4.2. Conclusion and Recommendations

This article presented the comparison between 3 different regression models for ordinal data with respect to the best fit model for our cervical cancer dataset. We found that the CR model without proportional odds yielded better results due to the highest AIC and log likelihood ratio and the lowest residual deviance. In addition, it is clear that with our cervical cancer data, the key prognostic factor associated with invasive cervical cancer was the FIGO clinical stage which particularly, had a higher influence on the surgical Stage 2 outcomes compared to the lesser surgical stage categories. All the 5 independent features selected for classifying the patients into surgical stages that made sense were the FIGO clinical stage and partly, the presence or absence of cancer of symptomatic vaginal discharge. The study was limited by the fact that the cervical cancer data was not created for the purpose of building statistical models thus was not sufficient and probably lacked key predictors for the type of analysis carried out in our study. Thus, our study demonstrates the need of databases with additional variables that could be significant to determining the suitability of surgical treatment such as molecular data, CT/MRI imaging information and HPV-DNA types. Moreover, research and data collection for predictive algorithms could introduce practical learning tools for the medical students who undergo medical training at the Moi Teaching and Referral hospital. The data was biased due to the dropping of incomplete records which left a small sample for building the models. Also, data was simulated to test the predictive capabilities of the models and statistical techniques were not utilized to address the imbalanced nature of the data as well as missing data. Although 4 predictors were not found to be key prognostic factors for highly accurate classifications in our models, future research utilizing data structured for developing predictive models in the cervical cancer setting could yield better results that could be integrated into the oncology system. A strict and validated ordinal classifier can more accurately predict the cancer stages (ordinal scales) compared to non-ordinal classifiers as noted by the polytomous logistic regression model [23].

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Wilson, K.L., Cowart, C.J., Rosen, B.L., Pulczynski, J.C., Solari, K.D., Ory, M.G. and Smith, M.L. (2018) Characteristics Associated with HPV Diagnosis and Perceived Risk for Cervical Cancer among Unmarried, Sexually Active College Women. *Journal of Cancer Education*, **33**, 404-416. <https://doi.org/10.1007/s13187-016-1131-1>
- [2] Petry, K.U. (2014) HPV and Cervical Cancer. *Scandinavian Journal of Clinical and Laboratory Investigation*, **74**, 59-62. <https://doi.org/10.3109/00365513.2014.936683>
- [3] Viviano, M., DeBeaudrap, P., Tebeu, P.M., Fouogue, J.T., Vassilakos, P. and Petignat, P. (2017) A Review of Screening Strategies for Cervical Cancer in Human Im-

- munodeficiency Virus-Positive Women in Sub-Saharan Africa. *International Journal of Women's Health*, **9**, 69. <https://doi.org/10.2147/IJWH.S103868>
- [4] Arbyn, M., Castellsagu'e, X., de Sanjos'e, S., Bruni, L., Saraiya, M., Bray, F. and Ferlay, J. (2011) Worldwide Burden of Cervical Cancer in 2008. *Annals of Oncology*, **22**, 2675-2686. <https://doi.org/10.1093/annonc/mdr015>
- [5] Allanson, E.R., Powell, A., Bulsara, M., Lee, H.L., Denny, L., Leung, Y. and Cohen, P. (2019) Morbidity after Surgical Management of Cervical Cancer in Low and Middle Income Countries: A Systematic Review and Meta-Analysis. *PLoS ONE*, **14**, e0217775. <https://doi.org/10.1371/journal.pone.0217775>
- [6] Nascimento-Goncalves, E., Faustino-Rocha, A.I., Seixas, F., Ginja, M., Colao, B., Ferreira, R., Oliveira, P.A., et al. (2018) Modelling Human Prostate Cancer: Rat Models. *Life Sciences*, **203**, 210-224. <https://doi.org/10.1016/j.lfs.2018.04.014>
- [7] Ji, J., Liu, J., Liu, H. and Wang, Y. (2014) Comparison of Serum and Tissue Levels of Trace Elements in Different Models of Cervical Cancer. *Biological Trace Element Research*, **159**, 346-350. <https://doi.org/10.1007/s12011-014-9981-z>
- [8] Odhiambo, C., Odhiambo, J. and Omolo, B. (2017) Validation of the Smooth Test of Goodness-of-Fit for Proportional Hazards in Cancer Survival Studies. *International Journal of Statistics in Medical Research*, **6**, 49-67. <https://doi.org/10.6000/1929-6029.2017.06.02.1>
- [9] Heyer, J., Yang, K., Lipkin, M., Edelmann, W. and Kucherlapati, R. (1999) Mouse Models for Colorectal Cancer. *Oncogene*, **18**, 5325-5333. <https://doi.org/10.1038/sj.onc.1203036>
- [10] Javali, S.B. and Pandit, P.V. (2010) A Comparison of Ordinal Regression Models in an Analysis of Factors Associated with Periodontal Disease. *Journal of Indian Society of Periodontology*, **14**, 155. <https://doi.org/10.4103/0972-124X.75909>
- [11] Freedman, A.N., Seminara, D., Gail, M.H., Hartge, P., Colditz, G.A., Ballard-Barbash, R. and Pfeiffer, R.M. (2005) Cancer Risk Prediction Models: A Workshop on Development, Evaluation, and Application. *Journal of the National Cancer Institute*, **97**, 715-723. <https://doi.org/10.1093/jnci/dji128>
- [12] Rothberg, M. (2018) Cleveland Clinic Develops First Overall Risk Prediction Model for Cervical Cancer.
- [13] Pang, W.K. (2018) Continuation-Ratio Model for Categorical Data: A Gibbs Sampling Approach. *Proceedings of the International Multi-Conference of Engineers and Computer Scientists*, 14-16 March 2018, Hong Kong, 1-6.
- [14] McCullagh, P. (1980) Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological)*, **42**, 109-127. <https://doi.org/10.1111/j.2517-6161.1980.tb01109.x>
- [15] Christensen, R.H.B. (2015) Ordinal-Regression Models for Ordinal Data. R Package Version.
- [16] Brant, R. (1990) Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics*, **46**, 1171-1178. <https://doi.org/10.2307/2532457>
- [17] Winship, C. and Mare, R.D. (1984) Regression Models with Ordinal Variables. *American Sociological Review*, **49**, 512-525. <https://doi.org/10.2307/2095465>
- [18] Ananth, C.V. and Kleinbaum, D.G. (1997) Regression Models for Ordinal Response: A Review of Methods and Application. *International Journal of Epidemiology*, **26**, 1323-1333. <https://doi.org/10.1093/ije/26.6.1323>
- [19] Norton, E.C., Cowart and Dowd, B.E. (2018) Log Odds and the Interpretation of Logit Models. *Health Services Research*, **53**, 859-878.

<https://doi.org/10.1111/1475-6773.12712>

- [20] Buch, S.C., Branch, R.A., Jones, K., Arena, V.C., Persad, R. and Romkes, M. (2005) The Effect of Model Choice on the Estimates of Relative Risk for Bladder Cancer. 955-956.
- [21] Zhou, H., Lawson, A.B., Hebert, J.R., Slate, E.H. and Hill, E.G. (2008) A Bayesian Hierarchical Modeling Approach for Studying the Factors Affecting the Stage at Diagnosis of Prostate Cancer. *Statistics in Medicine*, **27**, 1468-1489. <https://doi.org/10.1002/sim.3024>
- [22] de Jong, V.M.T., Eijkemans, M.J.C., van Calster, B., Timmerman, D., Moons, K.G.M., Steyerberg, E.W. and van Smeden, M. (2019) Sample Size Considerations and Predictive Performance of Multinomial Logistic Prediction Models. *Statistics in Medicine*, **38**, 1601-1619. <https://doi.org/10.1002/sim.8063>
- [23] Chen, C.-K. (2012) The Classification of Cancer Stage Microarray Data. *Computer Methods and Programs in Biomedicine*, **108**, 1070-1077. <https://doi.org/10.1016/j.cmpb.2012.07.001>

Abbreviation

ACL: Adjacent Category Logistic

AIC: Akaike Information Criterion

CPO: Cumulative Proportional Odds

CR: Continuation Ratio

FIGO: The International Federation of Gynecology and Obstetrics

HIV: Human Immunodeficiency Virus

HPV: Human Papilloma Virus

ML: Multinomial Logistic

OR: Odds Ratio

VIA: Visual Inspection with Acetic Acid