Application of logistic regression models to cancer patients: a case study of data from Jigme Dorji Wangchuck National Referral Hospital (JDWNRH) in Bhutan

Choney Zangmo and Montip Tiensuwan
Department of mathematics, Faculty of Science, Mahidol University, Bangkok 10400, Thailand
Email: zangmo_c@yahoo.com

Abstract: Cancer is an uncontrolled disease caused by a damage to the cell’s DNA. Cancer rates are increasing every year and is the biggest concern in many countries including Bhutan. Bhutan being a very small country loses around 100 to 200 patients every year due to cancer. This study involved 3013 cancer patients who are treated at Jigme Dorji Wangchuck National Referral Hospital (JDWNRH) from 2010 to 2016. Those patients who cannot be treated in the country are referred out to other countries for treatment. This study is to identify the factors that affect the survival of all cancer patients as well as male and female patients separately. This has been done using logistic regression model where the dependent variable is the patient’s last status. The best fitted model was obtained from the analysis of deviance. The test showed that the last status of the patient and the variables of personal and clinical data are mostly significant at $p$-value less than 0.05. The factors affecting the last status of cancer are age, length of stay and the cancer site. Cancer site is one of the main factors that affect the survival of both male and female patients.

1. Introduction
Cancer is a disease caused by an uncontrolled division of abnormal cells in the parts of the body. Cancer has become more than just a global health problem and is also a cause of deep suffering to the individual patients, their families and to the community at large. Cancer affects everyone which represents a tremendous burden to the family and country. According to world Health Organization, cancer has found to be the second leading cause of death throughout the globe and is accounted for death of 8.8 million people in 2015 worldwide [1].

In Taiwan, pancreatic cancer is the eighth leading cause of death. Approximately 800 people die of pancreatic cancer yearly [2] and prostate cancer is the second leading cause of death in the world [3]. The exact cause of cancer is unknown but there are some preventable risk factors which can be avoided such as using tobacco, alcohol, unhealthy foods and unhealthy lifestyles [4]. Cancer of the oral cavity, pharynx, larynx, and esophagus in Italy showed smokers have higher risk to cancer. However, for cancer of esophagus, alcohol consumption showed higher risk to cancer than smoking [5].

Cancer can be very expensive, devastating, worrisome and could possibly lead to death. However, early detection increases the survival figures [6-7]. It has become a major concern around the world. In Bhutan, the number of cancer cases has been increasing over the years and there were 473 cases reported in 2016. Female patients are mostly affected by cancer than male patients. To create awareness of such
disease, Bhutan has come up with programs such as health campaign, banning of tobacco, pap-smear screenings among women and vaccination to minimize cancer risk [8]. Therefore, the purpose of this study is to apply logistic regression models to identify the cancer variables associated with patient’s last status for all cancer patients and for individual gender.

2. Methodology

2.1 Data collection
The subjects on this study were cancer patients treated at Jigme Dorji Wangchuck National Referral Hospital (JDWNRH) in Thimphu, Bhutan from 2010 to 2016. The hospital is located in the capital of the country. A total of 3013 cancer patients with 1657 females and 1356 males were collected from the record section of JDWNRH through the application form. The cancer data was classified into two categories as follows:

Part I Personal data: This part consists of only two variables i.e. the age and gender
Part II Cancer / clinical data: This part consists of four variables including visits, length of stay, cancer site, and patient’s last status.

Table 1 provides the categories and levels of personal and clinical variables in the models.

| Category          | Level                                                                 |
|-------------------|----------------------------------------------------------------------|
| **Personal**      |                                                                        |
| Gender            | male, female                                                         |
| Age (years)       | 0-25, 26-50, 51-75, >75                                             |
| **Clinical**      |                                                                        |
| Length of stay (days) | ≤ 5, >5                                                              |
| Visits (times)    | ≤ 5, >5                                                              |
| Cancer site       | lip, oral cavity and pharynx, digestive organs, respiratory organs, female genital organs, male genital organs, breast, bone and skin, others (brain, throat, neuroendocrine tumors and urinary tract) |
| Patients last status | alive, death (binary logistic regression)                          |
|                   | alive, death, uncertain (multinomial logistic regression)            |

For the binary logistic regression, we use the dependent variable as death and alive. For multinomial logistic regression, we used the dependent variables as alive, death or uncertain.

Data were entered into excel and has been analyzed using IBM SPSS 18 (SPSS Ins).

2.2 Statistical Analysis
In this study, the statistical analyses are given as follows.

2.2.1 Probability distribution
In logistic regression models we have to construct design variables. The partial method involves setting all of the design variables equal to zero for the reference group, and then setting a simple design variable equal to one for each of the other groups. If the independent variable is declared to be in \(k\) categories, then the number of the design variables is equal to \(k-1\). This is illustrated in Table 2.

Next, We define \(X = 1\) if the outcome is a success and \(X = 0\) if the outcome is a failure, then the probability of \(X\) is given as \(Pr(X = 1) = \pi\) and \(Pr(X = 0) = 1 - \pi\) where \(0 \leq \pi \leq 1\).
Table 2 Illustration of the coding of the design variables using the partial method

| Level | X₁ | X₂ | ... | Xₖ₋₁ |
|-------|----|----|-----|------|
| 1     | 0  | 0  | ... | 0    |
| 2     | 1  | 0  | ... | 0    |
| 3     | 0  | 1  | ... | 0    |
| ...   | ...| ...| ... | ...  |
| k     | 0  | 0  | ... | 1    |

The logit is defined as the logarithm of the odds ratio \( \log \left( \frac{\pi_i}{1-\pi_i} \right) \) which is given by an equation

\[
\text{Logit}(\pi_i) = \log \left( \frac{\pi_i}{1-\pi_i} \right) = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k
\]

The fitted value is \( \hat{\pi}_i = \frac{e^{\text{logit}(\pi_i)}}{1 + e^{\text{logit}(\pi_i)}} \).

2.2.2 Goodness of fit
To measure the goodness of fit model, we use log likelihood ratio statistic (deviance).

\[
D = 2 \sum o \log \frac{o}{e}
\]

where \( o \) denotes the observed frequencies and \( e \) denotes the expected frequencies.

From the value of \( D \), we can test the hypothesis of the model which are

- \( H_0: \) The modified model fits the data well.
- \( H_1: \) The modified model does not fit the data well.

At significant level \( \alpha = 0.05 \), we reject the \( H_0 \) if \( D > \chi^2 \), (degrees of freedom) or accept \( H_0 \) otherwise.

The standard procedure of choosing the significant value is \( \alpha = 0.05 \).

3. Results and Discussion

3.1 Binary logistic regression
Binary logistic regression is used when the dependent variable is dichotomous. The number and percentage of clinical and personal data by gender are shown in Table 3. From Table 3, according to the age groups, most of the male and female patients affected by cancer ranges from 51-75 years of age. Most of the hospital visits were less than or equal to five times which is about 47.2% for male and 52.8% for female. Majority of the patients were admitted in the hospital for more than five days. Regarding the patients last status, more patients recovered from cancer. Death occurred more to male cancer patients (56%) than female patients (44%). In contrast majority of the female patients are diagnosed with cancer. The death percentage of the male patients was higher than the female patients although females were the ones who were mostly diagnosed with cancers. For the cancer site, most of the male patients suffered from digestive organ cancer. For female, the genital organs cancer had the highest number of occurrences. There are few numbers of male patients who were diagnosed with genital organs cancer.
First we describe the relationship between the dependent and the independent variables. We find the significant values and the values for each variable is recorded. Backward elimination is used to delete the variables that are not significant and only keep the variables that are significant. The \( p \)-values for gender \( (X_1) \), age \( (X_2) \), length of stay \( (X_3) \), and cancer site \( (X_5-X_{11}) \) are 0.046, 0.007, 0.000 and 0.000, respectively and they are all significant. For the variable visits \( (X_4) \), the value is not significant with \( p \)-value 0.563. Through the backward elimination method, the variable visit is removed. The result is shown in Table 4. It shows the significant values for all cancer variables. For cancer site, some categories are not significant. We cannot delete the categories which are not significant because the cancer site is a design variable.

Using the variables that are significant, we conducted an interaction test. All the interaction conducted was found to be significant and the result is shown in Table 5. Thus, these variables were considered in the logistic regression analysis.

Table 3 The number and percentage of personal and clinical data by gender

| Variables                  | Male          | Female        |
|----------------------------|---------------|---------------|
|                            | Number | Percent | Number | Percent |
| Age Groups (Years)         |         |         |         |         |
| 0-25                       | 55     | 38.2    | 89     | 61.8    |
| 26-50                      | 165    | 26.8    | 450    | 73.2    |
| 51-75                      | 715    | 46.6    | 819    | 53.4    |
| >75                        | 283    | 60.9    | 182    | 39.1    |
| Visits                     |         |         |         |         |
| \( \leq 5 \)               | 1186   | 44.1    | 1505   | 55.9    |
| \( > 5 \)                  | 32     | 47.8    | 35     | 52.2    |
| Length of stay (Days)      |         |         |         |         |
| \( \leq 5 \)               | 490    | 40.3    | 726    | 59.7    |
| \( > 5 \)                  | 728    | 47.2    | 814    | 52.8    |
| Patient status             |         |         |         |         |
| Alive                      | 881    | 40.9    | 1275   | 59.1    |
| Death                      | 337    | 56      | 265    | 44      |
| Cancer Site                |         |         |         |         |
| Lip, oral cavity and pharynx | 79   | 62.7    | 47     | 37.3    |
| Digestive organs           | 707    | 58.9    | 493    | 41.1    |
| Respiratory organs         | 134    | 59.8    | 90     | 40.2    |
| Female genital organs      | 0      | 0       | 511    | 100     |
| Male genital organs        | 53     | 100     | 0      | 0       |
| Breast                     | 0      | 0       | 109    | 100     |
| Bones and skin             | 54     | 54.5    | 45     | 45.5    |
| Others*                    | 191    | 43.8    | 245    | 56.2    |
| Total                      | 1218   | 44.2    | 1540   | 55.8    |

Note: *such as brain, throat, neuroendocrine tumor and urinary tract.
Table 4 Significant values of all cancer patients

| Variables                  | p-value |
|----------------------------|---------|
| **Gender**                 |         |
| Male                       | 0.046   |
| Female                     | 0.007   |
| **Age**                    |         |
| ≤ 5                        | -       |
| > 5                        | 0.000   |
| **Length of Stay**         |         |
| ≤ 5                        | -       |
| > 5                        | 0.000   |
| **Cancer site**            |         |
| Lip, oral cavity and pharynx | 0.000  |
| Digestive organs           | 0.059   |
| Respiratory organs         | 0.083   |
| Female genital organs      | 0.007   |
| Male genital organs        | 0.113   |
| Breast                     | 0.026   |
| Bones and skin             | 0.227   |
| Others*                    | 0.506   |
| **Constant**               | 0.000   |

Note: *such as brain, throat, neuroendocrine tumors and urinary tract.

Table 6 Significant values of cancer data for the individual gender

| Variables                  | Male | Female |
|----------------------------|------|--------|
| **Age**                    | 0.541| 0.001  |
| **Length of Stay**         |      |        |
| ≤ 5                        | -    | -      |
| > 5                        | 0.000| 0.040  |
| **Cancer site**            |      |        |
| Lip, oral cavity and pharynx | 0.002| 0.000  |
| Digestive organs           | 0.065| 0.494  |
| Respiratory organs         | 0.177| 0.281  |
| Female genital organs      | -    | 0.037  |
| Male genital organs        | 0.210| -      |
| Breast                     | -    | 0.041  |
| Bones and skin cancer      | 0.336| 0.338  |
| Others*                    | 0.845| 0.643  |
| **Constant**               | 0.000| 0.000  |

Note: *such as brain, throat, neuroendocrine tumors and urinary tract.

Table 5 Analysis of deviance for the determined variables and p-value for possible interaction of interest to be added to the main effects only model for all cancer patients

| Interaction                    | Deviance | Deviance difference | d.f  | d.f. difference | p-value |
|--------------------------------|----------|---------------------|------|-----------------|---------|
| **Main effects only**          | 1507.006 | -                   | 1470 | -               | -       |
| Gender × Age                   | 1504.758 | 2.248**             | 1469 | 1               | 0.000   |
| Gender × Length of stay        | 1505.378 | 1.628**             | 1469 | 1               | 0.000   |
| Gender × Cancer site           | 1505.805 | 1.201**             | 1466 | 4               | 0.000   |
| Age × length of stay           | 1505.243 | 1.763**             | 1469 | 1               | 0.000   |
| Age × Cancer site              | 1501.948 | 5.058**             | 1463 | 7               | 0.000   |
| Length of stay × Cancer site   | 1490.064 | 16.942**            | 1463 | 7               | 0.000   |

Note: ** Significant at p-value < .01
Main effects only: Age, Gender, Length of stay and Cancer site

Since the variable gender is significant, we consider the model for individual gender. For the male cancer patients, length of stay (X₃) and cancer site (X₄-X₉) are significant with p-value < 0.01. The variable age (X₁) and visits (X₃) has p-value 0.541 and 0.730 and they are not significant but age being one of the most biologically important variables and interacts with other variable, we have to make a decision concerning variable age in the model. Thus, we have included the variable age in the model.

For the female cancer patients, the results of fitting the model of female cancer patients with the analysis of deviance for the patient’s last status contains the age (X₁), length of stay (X₃) and the cancer site (X₄-X₉) which has the p-value 0.001, 0.040 and 0.000, respectively. Again the variable visits(X₃) is not significant with p-value 0.746. This variable has been removed using backward elimination which is shown in Table 6.
Table 7 shows the deviance and p-value for possible interaction added in the main effects only model for male cancer patients.

**Table 7** Analysis of deviance for the determined variables and p-value for possible interaction of interest to be added to the main effects only model for male cancer patients

| Interaction                               | Deviance | Deviance difference | d.f  | d.f. difference | p-value |
|-------------------------------------------|----------|---------------------|------|-----------------|---------|
| **Main effects only**                     | 698.103  | -                   | 622  | -               | -       |
| (Age × Length of stay) + Age              | 697.290  | 0.813**             | 620  | 2               | 0.000   |
| (Age × Cancer site) + Age                 | 693.806  | 4.297**             | 616  | 6               | 0.000   |
| Length of stay × Cancer site              | 692.270  | 5.833**             | 616  | 6               | 0.000   |

Note: ** Significant at p-value < .01

Main effects only: Age, Length of stay and Cancer site

Table 8 shows the deviance and p-value for possible interaction added in the main effects only model for female cancer patients.

**Table 8** Analysis of deviance for the determined variables and p-value for possible interaction of interest to be added to the main effects only model for female cancer patients

| Interaction                               | Deviance | Deviance difference | d.f  | d.f. difference | p-value |
|-------------------------------------------|----------|---------------------|------|-----------------|---------|
| **Main effects only**                     | 803.047  | -                   | 843  | -               | -       |
| Age × length of stay                      | 795.860  | 7.187**             | 842  | 1               | 0.000   |
| Age × Cancer site                         | 797.003  | 6.044**             | 837  | 6               | 0.000   |
| length of stay × Cancer site              | 786.509  | 16.538**            | 837  | 6               | 0.000   |

Note: ** Significant at p-value < .01

Main effects only: Age, length of stay and Cancer site

The interaction terms included into this model shows the possibility to describe the effects of the associated variables to the patient’s last status. Table 8 shows the individual interaction added to the main effects only model. It shows all the interaction tests to be significant. When all the main effects only model and interaction terms are added together, we get the result which is shown in Table 9. We see that the variables age, length of stay, cancer site and interaction terms between cancer site and length of stay are significant at p-value < 0.05. Since the interaction term between cancer site and length of stay is significant, the variable cancer site is included in the model.

The odd of age covariate is 1.064. This shows that with every increase in age, the odds of death increases by 1.064 times. The female patients admitted in hospital for more than 5 days have about 29.067 times higher odds of death than those patients who are not admitted and the odds ratio is relatively very high. The group of female patients who has cancer located in the digestive organs, respiratory organs, female genital organs, breast, bone and skin and some other sites have higher odds to death than the female patients who are affected by lip, oral cavity and pharynx cancer (reference group). That is because the odds of death of these groups are greater than the groups of patients who have cancer located in lip, oral cavity and pharynx. The group of female patients who are admitted in hospital for more than 5 days and has digestive organs cancer has 0.081 times less odds to death than the patients with the cancer located in lip, oral cavity and pharynx. A female patient with cancer located in female reproductive organs who is admitted in the hospital for more than 5 days has about 0.142 times less odds to death.
### Table 9 Estimated coefficients, standard errors, Wald's test and odds ratio with 95% C.I. of the best model for female cancer patients

| Variable | Estimated coefficients | Estimated Std. Error | Wald's test | p-value | Exp(B) | 95% C.I for EXP(B) |
|----------|------------------------|----------------------|-------------|---------|--------|--------------------|
|          |                        |                      |             |         |        |        |
| X_1      | 0.062                  | 0.031                | 3.945       | 0.047   | 1.064  | 1.001 - 1.13      |
| X_2      | 3.377                  | 1.341                | 6.341       | 0.012   | 29.278 | 2.114 - 405.515   |
| X_3      | 4.643                  | 2.431                | 3.648       | 0.056   | 103.865| 0.886 - 12178.027 |
| X_4      | 2.313                  | 2.716                | 0.725       | 0.394   | 10.106 | 0.049 - 2073.456  |
| X_5      | 2.791                  | 2.461                | 1.286       | 0.257   | 16.292 | 0.131 - 2025.326  |
| X_6      | 3.540                  | 2.877                | 1.514       | 0.219   | 34.453 | 0.123 - 9676.525  |
| X_7      | 4.198                  | 2.768                | 2.301       | 0.129   | 66.558 | 0.293 - 15101.951 |
| X_8      | 3.881                  | 2.426                | 2.559       | 0.11    | 48.454 | 0.417 - 5625.991  |
| X_9      | -0.017                 | 0.009                | 3.404       | 0.065   | 0.983  | 0.965 - 1.001     |
| X_1 × X_2| -0.042                 | 0.031                | 1.783       | 0.182   | 0.959  | 0.903 - 1.02      |
| X_1 × X_3| -0.009                 | 0.035                | 0.066       | 0.797   | 0.991  | 0.925 - 1.062     |
| X_2 × X_1| -0.035                 | 0.032                | 1.185       | 0.276   | 0.966  | 0.907 - 1.028     |
| X_2 × X_4| -0.065                 | 0.041                | 2.484       | 0.115   | 0.937  | 0.864 - 1.016     |
| X_2 × X_5| -0.060                 | 0.037                | 2.633       | 0.105   | 0.942  | 0.876 - 1.013     |
| X_2 × X_6| -0.039                 | 0.031                | 1.575       | 0.210   | 0.962  | 0.905 - 1.022     |
| X_2 × X_7| -2.513                 | 1.199                | 4.392       | 0.036   | 0.081  | 0.008 - 0.85      |
| X_2 × X_8| -1.860                 | 1.295                | 2.062       | 0.151   | 0.156  | 0.012 - 1.971     |
| X_2 × X_9| -1.955                 | 1.228                | 2.536       | 0.111   | 0.142  | 0.013 - 1.57      |
| X_7 × X_2| -0.931                 | 1.476                | 0.398       | 0.528   | 0.394  | 0.022 - 7.116     |
| X_7 × X_3| -1.319                 | 1.542                | 0.732       | 0.392   | 0.267  | 0.013 - 5.488     |
| X_7 × X_4| -1.685                 | 1.244                | 1.835       | 0.176   | 0.185  | 0.016 - 2.123     |
| X_7 × X_5| -1.685                 | 2.426                | 8.042       | 0.005   | 0.001  |                    |

Note: X_1: Age, X_2: Length of stay, X_4-X_9: Cancer site

The estimated coefficients, standard errors, Wald’s test and odds ratio (95% confidence interval) of the best model for all cancer patients and male cancer patients were obtained in the similar manner as obtained for female cancer patients. The patient’s visits to hospital has found to be not significant for all cancer patients and for the individual gender. The factors affecting the patient’s last status of cancer are age, length of stay and the cancer site. Cancer site and length of stay are the main factors that affect the survival of all cancer patients as well as the individual gender.

### 3.2 Multinomial logistic regression

Multinomial logistic regression is used when the dependent or the outcome variable has more than two categories.

According to Table 10, more than 70% of the patients recovered from cancer. Death occurred more to male cancer patients than female patients but majority of the females are diagnosed with cancer. Very few patients have the status uncertain and we do not know whether the patients are dead or alive. The patients are either referred out or not recovered in the uncertain category. The patients with digestive organs cancer has the highest number of death and about 17.5% of the patients with bone and skin cancer is uncertain to death.
### Table 10 The number and percentage of personal and clinical data by patient’s last status

| Variables            | Patients last status | Total |
|----------------------|----------------------|-------|
|                      | Alive | Death | Uncertain |       |
| Gender               |        |       |           |       |
| Male                 | 881   | 337   | 138       | 1356  |
| Female               | 1275  | 265   | 117       | 1657  |
| Age group            |        |       |           |       |
| 0-25                 | 123   | 21    | 41        | 185   |
| 26-50                | 518   | 97    | 47        | 662   |
| 51-75                | 1183  | 351   | 119       | 1653  |
| >75                  | 332   | 133   | 48        | 513   |
| Length of stay       |        |       |           |       |
| ≤ 5                  | 1015  | 201   | 60        | 1276  |
| > 5                  | 1141  | 401   | 195       | 1737  |
| Visits               |        |       |           |       |
| ≤ 5                  | 2099  | 592   | 254       | 2945  |
| > 5                  | 57    | 10    | 1         | 68    |
| Cancer site          |        |       |           |       |
| Lip, oral cavity and pharynx | 101 | 25 | 24 | 150 |
| Digestive organs     | 851   | 349   | 100       | 1300  |
| Respiratory organs   | 156   | 68    | 27        | 251   |
| Female genital organs| 467   | 44    | 18        | 529   |
| Male genital organs  | 47    | 6     | 2         | 55    |
| Breast               | 102   | 7     | 7         | 116   |
| Bones and skin       | 86    | 13    | 21        | 120   |
| Others*              | 346   | 90    | 56        | 492   |

Note: *such as brain, throat, neuroendocrine tumor and urinary tract. The number in parenthesis are in percentage.

Table 11 represents the parameter estimates for the patient’s last status and cancer data of all cancer patients. For this analysis, the reference group is death of patients due to cancer. Alive of patient’s status relative to death of patient’s status and uncertain of patient’s status relative to death of patient’s status are used to estimate the parameters of the model. For the patients last status alive, the variables age, gender, length of stay and cancer site are significant with p-values 0.010, 0.040, 0.000 and (0.490, 0.060, 0.140, 0.000, 0.040, 0.000, 0.050), respectively. Similarly, for the patients last status uncertain, the variables age, visits, length of stay and cancer site are significant with p-values 0.000, 0.000, 0.010 and (0.020, 0.130, 0.790, 0.790, 0.950, 0.220, 0.010), respectively. Gender and visits are not significant, so we do not consider the model for individual gender. Thus, the model considers the factors age, length of stay and cancer site as the best fit model. Moreover, cancer mostly affected the elder patients.

The first half of the table describes the factors associated with the alive of patients. When looking at the age, the odds ratio of age is 0.992. For the patients admitted in the hospital with less than or equal to 5 days, the odds for the alive of patients relative to death of patients is 1.535. For the cancer site, the patients diagnosed with lip, oral cavity and pharynx cancer has 1.196 times chances to live. Patients with digestive organs cancer has the log odd value decreased by 0.27 when comparing the alive of patients to the death of patients and the odds ratio is 0.764. For patients with female genital organs cancer, male genital organs cancer, breast cancer and bone and skin cancer, the odds of being alive is 2.540, 2.614, 3.328 and 1.905, respectively when compared to the death of patients.
The second half of the table describes the factors associated with the uncertain category of patient’s last status. For the factor age, the log odds value for the uncertain of patients compared to the death of patients is decreased by 0.02. The odds ratio is 0.979. The log odds value for the patients visiting the hospital is decreased by 0.51. So the odds for uncertain of patients compared to death of patients is 0.603. For the patients admitted in the hospital with less than or equal to 5 days, the odds for the uncertain of patients is 0.633 times less than the death of patients. For the patients diagnosed with lip, oral cavity and pharynx cancer, breast cancer and bone and skin cancer, the odds of uncertain relatively to death is 2.232, 2.023 and 2.997, respectively.

**Table 11 Parameter Estimates for patients last status and cancer data of all cancer patients**

| last status | Variables                  | B    | Std. Error | Wald   | *p*-value | Exp(B) | 95% C I for Exp(B) |
|-------------|----------------------------|------|------------|--------|-----------|--------|--------------------|
| Alive       | Intercept                  | 1.603| 0.208      | 59.55  | 0.000     | 0.986  | 0.986              |
|             | Age                        | -0.01| 0.003      | 6.624  | 0.010     | 0.992  | 0.986              |
|             | Visits                     | 0.024| 0.043      | 0.301  | 0.580     | 1.024  | 0.941              |
|             | Gender                     |      |            |        |           |        |                    |
|             | Male                       | -0.22| 0.104      | 4.329  | 0.040     | 0.806  | 0.658              |
|             | Female                     |      |            |        |           |        | 0.988              |
|             | length of stay             |      |            |        |           |        |                    |
|             | ≤ 5                        | 0.428| 0.102      | 17.73  | 0.000     | 1.535  | 1.257              |
|             | > 5                        |      |            |        |           |        | 1.873              |
|             | Cancer Site                |      |            |        |           |        |                    |
|             | Lip, oral cavity and pharynx| 0.179| 0.257      | 0.484  | 0.490     | 1.196  | 0.722              |
|             | Digestive organs           | -0.27| 0.145      | 3.459  | 0.060     | 0.764  | 0.575              |
|             | Respiratory organs         | -0.29| 0.196      | 2.213  | 0.140     | 0.747  | 0.509              |
|             | Female genital organs      | 0.932| 0.206      | 20.57  | 0.000     | 2.54   | 1.698              |
|             | Male genital organs        | 0.961| 0.459      | 4.39   | 0.040     | 2.614  | 1.064              |
|             | Breast                     | 1.202| 0.413      | 8.487  | 0.000     | 3.328  | 1.482              |
|             | Bones and skin             | 0.645| 0.323      | 3.979  | 0.050     | 1.905  | 1.011              |
|             | Others*                    |      |            |        |           |        | 3.589              |
| Uncertain   | Intercept                  | 1.218| 0.315      | 14.9   | 0.000     | 0.97   | 0.97               |
|             | Age                        | -0.02| 0.004      | 23.7   | 0.000     | 0.997  | 0.97               |
|             | Visits                     | -0.51| 0.135      | 14.09  | 0.000     | 0.603  | 0.463              |
|             | Gender                     |      |            |        |           |        |                    |
|             | Male                       | 0.068| 0.166      | 0.166  | 0.680     | 1.07   | 0.773              |
|             | Female                     |      |            |        |           |        | 1.481              |
|             | length of stay             |      |            |        |           |        |                    |
|             | ≤ 5                        | -0.46| 0.177      | 6.638  | 0.010     | 0.633  | 0.448              |
|             | > 5                        |      |            |        |           |        | 0.897              |
|             | Cancer Site                |      |            |        |           |        |                    |
|             | Lip, oral cavity and pharynx| 0.803| 0.345      | 5.405  | 0.020     | 2.232  | 1.134              |
|             | Digestive organs           | -0.35| 0.229      | 2.331  | 0.130     | 0.705  | 0.451              |
|             | Respiratory organs         | -0.08| 0.303      | 0.072  | 0.790     | 0.922  | 0.509              |
|             | Female genital organs      | -0.09| 0.343      | 0.068  | 0.790     | 0.914  | 0.467              |
|             | Male genital organs        | -0.05| 0.848      | 0.004  | 0.950     | 0.948  | 0.18               |
|             | Breast                     | 0.705| 0.571      | 1.523  | 0.220     | 2.023  | 0.661              |
|             | Bones and skin             | 1.098| 0.400      | 7.537  | 0.010     | 2.997  | 1.369              |
|             | Others*                    |      |            |        |           |        | 6.561              |

The reference category is: Death.

Note: *such as brain, throat, neuroendocrine tumors and urinary tract.
4. Conclusion
According to the results, we obtained the best model for all cancer, male and female cancer patients with the dependent variable as the patient’s last status. If we randomly sample one patient and get information for the factors in the model, we can predict the status of last contact of the patient. From each model we obtain the estimated logit to get the fitted value.

In the binary logistic regression, for all cancer patients, the factors which effect the patient’s last status are age, length of stay and cancer site. We perform in the same manner to the female and male cancer patients. Age being one of the most important factors, we considered the variable age in the model. Interaction terms were included in the model using the analysis of deviance to assess their significance. For multinomial logistic regression also, the factors age, length of stay and cancer site were highly related to the patient’s last status. Therefore, these factors were considered into the model.

Finally, we conclude that, the logistic regression models that can identify the factors which affect the status of last contact of the cancer patients are also used to predict the status of cancer patients. Thus, further interesting studies could be applying this method to be used with other diseases such as diabetes, cardiovascular disease, HIV, etc. and also applying other statistical method to analyze the data and compare the results.

References
[1] World Health Organization. Available from http://www.who.int/mediacentre/factsheet/fs297/en/. [Assessed 5th, October, 2017].
[2] Chang Chun-Lang, Hsu Ming-Yuan. The study that applies artificial intelligence and logistic regression for assistance in differential diagnostic of pancreatic cancer. Expert Systems with Applications 36 2009; 10663-10672
[3] Yuri Prahara, Rochadi Sungsang, Danarto Raden. A Device for Predicting Prostate Cancer Risk: A Logistic Regression. Journal of Prostate Cancer 2016; 1-5
[4] Cancer facts in Bhutan-Bhutan cancer society. Available from http://www.bhutancancer.org/cancer-facts-in-bhutan/. [Assessed 11th, October, 2017]
[5] Franceschi Silvia, Talamini Renato, Barra Salvatore, E.Baron Anna, et al. Smoking and Drinking in Relation to Cancer of the Oral Cavity, Pharynx, Larynx, and Esophagus in Northern Italy. Cancer Research 1990; 6502-6507
[6] Yusuff H, Mohamad N, Ngah KU, Yahaya SA. Breast cancer analysis using logistic Regression. 2012; 14-22
[7] Madhu B et al. A Multinomial Logistic Regression Analysis to Study The Influence Of Residence And Socio-Economic Status On Breast Cancer Incidence In Southern Karnataka. International Journal of Mathematics and Statistics Invention (IJMSI).E-ISSN: 2321 – 4767 P-ISSN: 2321 – 4759. Volume 2 Issue 5 2014
[8] Stomach cancer in Bhutan- world life expectancy. Available from http://www.worldlifeexpectancy.com/bhutan-stomach--cancer [Assessed 7th, October, 2017]

Acknowledgements
We would like to thank Thailand International Cooperation Agency (TICA) for the scholarship and the record section of Jigme Dorji Wangchuck National Referral Hospital (JDWNRH) in Bhutan for providing and supporting us with cancer data.