APPLICATION OF LOG-LINEAR MODELS TO CANCER PATIENTS: A CASE STUDY OF DATA FROM THE NATIONAL CANCER INSTITUTE

Montip Tiensuwan, Pornpis Yimprayoon and Yongwimon Lenbury

Department of Mathematics, Faculty of Science, Mahidol University, Bangkok, Thailand

Abstract. Cancer is a noninfectious disease which is on the increase throughout the world and has become a serious problem for public health in many countries, including Thailand. In Thailand, cancer has risen significantly to become a leading cause of death and most patients are admitted to the National Cancer Institute. The objective of this study is to identify the associated factors between personal, cancer/clinical variables of cancer patients using log-linear models. Tests of independence are used (chi-square and Cramer's V-value tests) to find out the relationships between any two variables. In addition two- and three-dimensional log-linear models are used to obtain estimated parameters and expected frequencies for these models. Amongst the models fitted, the best are chosen based on the analysis of deviance. The results of this study show that most paired variables of personal, cancer/clinical variables are significantly related at p-value <0.05. For both male and female patients, the variable site of the cancer is highly related to marital status, diagnostic evidence and treatment, which provide the highest Cramer's V value. Moreover, the site of cancer also affects the method of diagnostic evidence and treatment. Since the site of cancer in each sex is different, prevention for various sites of cancer should be considered for each specific sex. In addition, for male and female patients, treatment is related to the site of cancer. Consequently, physicians may consider these factors before selecting the appropriate method of treatment.

INTRODUCTION

Cancer is on the increase throughout the world and has become a serious public health problem in many countries. Cancer is not only a problem for patients but also affects their families and the community as a whole in terms of job loss, social isolation and family tension which may follow closely on the occurrence of cancer. The economic burden of cancer is most obvious in health care costs, such as for hospitals, drugs, and other health services.

In 2000, there were approximately 10.1 million new cancer cases, 4.7 million in more developed countries and 5.4 million in less developed countries. By the year 2020, these figures are likely to reach 15.3, 6.0, and 9.3 million, respectively (WHO, 2002) as shown in Table 1.

Correspondence: Montip Tiensuwan, Department of Mathematics, Faculty of Science, Mahidol University, Rama 6 Road, Bangkok 10400, Thailand.
Tel: 66 (0) 2201-5538, 66 (0) 2201-5340
E-mail: scmts@mahidol.ac.th

In Thailand, health problems have changed dramatically during the past two decades. Deaths due to communicable diseases, such as tuberculosis and pneumonia, have declined, while non-communicable diseases, such as cancer, have become of greater importance. Cancer has increased significantly to become a leading cause of death from non-infectious disease during the period 1993-2002 (Bureau of Policy and Strategy, 2002) as displayed in Fig 1. We found that since 1997, the mortality rate for cancer has continuously increased.

The public health statistics (Ministry of Public Health, 2002) show that since 1998, the mortality rate for cancer had increased continuously through the year 2002. Cancer has ranked first in the causes of death, as presented in Table 2.

The purpose of this study is to find associated factors between personal, cancer/clinical variables which effect the cancer patients of the National Cancer Institute by using the two- and three-dimensional log-liner models.

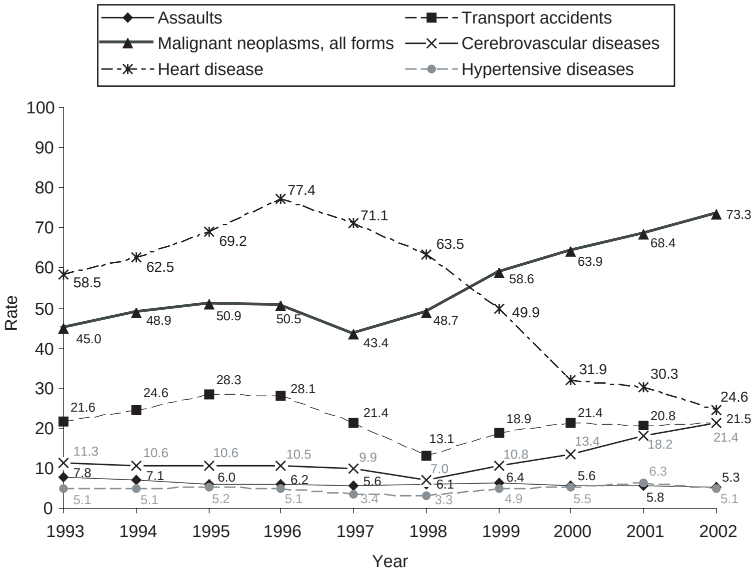

Fig 1–Causes of death from non-infectious disease, Thailand, 1993-2002.

cancer patients, who consisted of 1,248 male patients and 1,955 female patients.

Since the site of cancer for each sex is different, the cancer data was divided according to sex. Each sex was classified into personal data and cancer or clinical data.

Part 1: Personal data

The personal data records of the patients provided information for 5 variables: race, religion, marital status, age and region. If positive or negative, these were coded directly, other codings are as indicated in Table 3.

Table 1
Numbers of cancer deaths and new cancer cases in the world estimated for 2000 and predicted for 2020.

Year	Region	New cases (million)	Deaths (million)
2000	More developed countries	4.7	2.6
	Less developed countries	5.4	3.6
	All countries	10.1	6.2
2015	More developed countries	6.0	3.5
	Less developed countries	9.3	6.3
	All countries	15.3	9.8

MATERIALS AND METHODS

Data collection

The subjects were cancer patients treated at the National Cancer Institute. We collected cancer data using a cancer notification form of the National Cancer Institute. The classification and coding of primary site and morphology in cancer notification form used the International Classification of Diseases for Oncology (ICD-O) (WHO, 1990). These data were the number of new cancer patients treated between January 1999 and December 1999 at the National Cancer Institute. In this data set, there were 3,203

Part 2: Cancer/clinical data

The 5 variables were: diagnostic evidence, site of cancer, stage of diagnosis, treatment, and status of last contact. The groups for the categorical variables are summarized in Table 3.

For this study, a number of general statements can be made with respect to the patients with cancer. Within the personal data group the data show more female cancer patients than male cancer patients. For female and male patients, more than 90% of all cancer patients were Thai. Most of the cancer patients were Buddhist. Moreover, more than 80% of all patients were married/divorced/widowed patients. Most female patients were less than 50 years old, while most male patients were less than 50 years old or greater than 64 years old. Most cancer patients lived in central Thailand.

Within the cancer/clinical data group, the histology of primary was the diagnostic evidence that was used more than other methods for both the female and male patients. For females, a large number of cancer patients had breast cancer. In males, the trachea, bronchus and lungs had the highest occurrence. The localized stage was the group with the largest proportion of cancer patients. Most cancer patients were treated using radiation. Regarding status at last contact, for both female and male patients, more than

Table 2
Number of deaths and death rates per 100,000 population by leading causes of death, 1998-2002.

Cause of death	1998		1999		2000		2001		2002	
	Number	Rate	Number	Rate	Number	Rate	Number	Rate	Number	Rate
Total	310,534	507.8	362,607	589.0	365,741	592.1	369,493	595.1	380,364	608.1
Malignant neoplasm, all forms	29,812	48.7	36,091	58.6	39,480	63.9	42,497	68.4	45,834	73.3
Accident and poisonings	21,740	35.5	29,845	48.5	32,401	52.5	31,579	50.9	34,566	55.3
Disease of the heart and cerebrovascular	52,376	85.6	42,288	68.7	32,331	52.3	34,903	56.2	32,896	52.6
Pneumonia and other diseases of lung	6,099	10.0	9,444	15.3	9,286	15.0	11,163	18.0	13,185	21.1
Nephritis, nephrotic syndrome and nephrosis	5,981	9.8	6,745	11.0	9,091	14.7	10,139	16.3	10,587	16.9
Disease of liver and pancrease	5,742	9.4	6,395	10.4	6,736	10.9	7,761	12.5	8,025	12.8
Tubercolosis, all forms	4,252	7.0	5,265	8.6	6,246	10.1	6,284	10.1	6,751	10.8
Suicide	4,964	8.1	5,290	8.6	5,189	8.4	4,803	7.7	4,905	7.8
Hypertension and cerebrovascular disease	2,029	3.3	2,987	4.9	3,403	5.5	3,912	6.3	3,213	5.1
Dengue hemorrhagic fever	339	0.6	132	0.2	103	0.2	327	0.5	264	0.4
Others	177,200	289.8	218,125	354.3	221,475	358.5	216,125	348.1	220,138	351.9

Table 3
Divisions within each category variable.

Categories	Divisions
Personal	
Race	Thai, others
Religion	Buddhist, others
Marital status	Single, married/divorced/widow, unknown
Age(years)	≤49, 50-54, 55-59, 60-64, ≥65
Region	North, Northeast, Central, South, Bangkok
Cancer/clinical	
Diagnostic evidence	Clinical only, cytology and hematology (only for female), X rays/endoscopy/radiodiagnosis, histology of metastasis, histology of primary, others
Site of cancer	Female only: breast, cervix, uterus, ovary Male only: trachea/bronchus/lung, liver and intrahepatic, bile ducts, rectum/rectosigmoid and anus, others, unknown
Stage of diagnosis	Localized, local extension, regional lymph node involvement, others, unknown
Treatment	Surgery, radiation, chemotherapy, others, no treatment
Status of last contact	Alive with cancer, alive without cancer, death

90% of the cancer patients were alive with cancer.

Method of analysis

In this study, we used statistical analysis in three stages. A test of independence was applied to study the relationship between cancer variables incorporated with Cramer's V statistic (Bhattacharyya and Johnson, 1977; Marija, 1993) and the application of generalized linear models for use with two-and three-dimensional log-linear models (Nelder and Wedderburn, 1972; Nelder, 1974; Baker and Nelder, 1978; Fienberg, 1982; Aitkin *et al*, 1989; Dobson, 1996). These are employed to obtain estimates of parameter for the best models. These mod-

Table 4
Summary of the relationships between personal variables (Part 1) and cancer/clinical variables (Part 2) which give the maximum Cramer's V value.

Variable	Male		Female	
	Part 1	Part 2	Part 1	Part 2
Part 1				
Race	Region	Treatment	Region	Treatment
Religion	Race		Race	
Marital status	Religion	Site of cancer	Region	Site of cancer
Region	Race	Diagnostic evidence	Race	Site of cancer
Part 2				
Diagnostic evidence		Site of cancer		Site of cancer
Site of cancer		Diagnostic evidence		Treatment
Treatment		Site of cancer		Site of cancer
Status of last contact		Treatment		Site of cancer

Table 5
The frequencies (percent) of marital status and site of cancer for male patients.

Marital status	Site of cancer				Total
	Rectum, rectosigmoid and anus	Liver and intrahepatic bile ducts	Trachea, bronchus and lung	Others[a]	
Single	6 (4.8)	11 (8.8)	8 (6.4)	100 (80)	125 (100)
Married/divorced/widowed	46 (4.8)	178 (18.5)	238 (24.7)	502 (52.1)	964 (100)
Total	52 (4.8)	189 (17.4)	246 (22.6)	602 (55.3)	1,089 (100)

[a]such as brain, tongue, larynx, etc.

els indicate the associations between the personal and cancer/clinical variables.

RESULTS

In this section, we present the results from testing associations between two distinct cancer variables for males and females using the chi-square statistic and Cramer's V. Table 4 shows the summary of the relationships between the personal and cancer/clinical variables which gave a maximum Cramer's V for male and female cancer patients at the level of significant of $p<0.05$.

As Table 4 indicates, the site of cancer and treatment have a strong relationship with personal variables for males and females, while diagnostic evidence and region have a strong association only in male patients. Amongst clinical/cancer variables, site of cancer has a strong relationship with other cancer/clinical variables for both male and female patients while treatment has a strong relationship with status of last contract in male patients.

Two-dimensional log-linear models

From Table 4, we chose the two variables which had a relationship and gave the maximum Cramer's V-value to formulate the two-dimensional log-linear models, and obtained 8 models for each sex in the form $\eta_{jk} = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk}$. However, we present only one model corresponding to that of marital status and site of cancer for male patients (Table 5). Further, we fitted the model under H_0 and the maximal model under H_1 and obtained $\chi^2 = 37.95$, the log-likelihood ratio statistic (D) = 43.19, d.f. = 3, hence

Table 6
Estimates of parameters of the maximum model for marital status and site of cancer for male patients.

Parameter	Estimate	Parameter	Estimate	Parameter	Estimate
μ	3.95	β_3	-0.17	$(\alpha\beta)_{14}$	0.42
α_1	-1.23	β_4	1.46	$(\alpha\beta)_{21}$	-0.21
α_2	1.23	$(\alpha\beta)_{11}$	0.21	$(\alpha\beta)_{22}$	0.16
β_1	-1.14	$(\alpha\beta)_{12}$	-0.16	$(\alpha\beta)_{23}$	0.47
β_1	-0.16	$(\alpha\beta)_{13}$	-0.47	$(\alpha\beta)_{24}$	-0.42

Table 7
The log-likelihood ratio statistic, chi-square value and degree of freedom for associated variables for male and female patients.

Associated variables	D	χ^2	d.f.
Male			
Race and region	45.5706	57.5432	4
Religion and race	9.2347	17.1611	1
Race and treatment	12.8231	15.9829	3
Marital status and site of cancer	43.1869	37.9490	3
Region and diagnostic evidence	67.0251	73.9399	16
Diagnostic evidence and site of cancer	428.3895	467.4286	12
Treatment and site of cancer	146.2667	144.2966	9
Status of last contact and treatment	26.075	32.0758	6
Female			
Race and region	88.3077	117.7501	4
Religion and race	17.4603	35.3271	1
Race and treatment	11.8934	10.8908	3
Marital status and site of cancer	116.3535	95.3834	3
Region and site of cancer	56.6634	57.0435	12
Diagnostic evidence and site of cancer	226.8470	210.8422	12
Site of cancer and treatment	732.4459	680.4564	9
Status of last contact and site of cancer	35.6345	36.3682	6

at a significance level of 0.01, we rejected H_0 and chose the maximal model. The estimates of the main effects of interaction terms of the maximal model are shown in Table 6. The other models were obtained in the same manner. Table 7 presents the log-likelihood ratio statistic, chi-square value and degree of freedom for the associated variables for male and female patients.

Three-dimensional log-linear models

We constructed a three-variable contingency table to investigate whether an association between marital status and treatment was the same for both sexes. Further, we considered the data for each sex and constructed 2 different three variable contingency tables, since from the test of independence we found that the variable region was related to almost every other variable. We thus selected region to be one in each set of three variables. The other two variables for each set were as follows: marital status and site of cancer, and marital status and treatment. From these, three-dimensional log-linear models were formulated using the values of deviance difference and deviance difference degrees of freedom to analyze the models. The results are shown in Table 8, however, we present only one model; marital status, site of cancer and region for female patients.

According to the chi-square, the best model is model (2), that is, $\eta_{jk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + (\widehat{\alpha\beta})_{ij} + (\widehat{\alpha\gamma})_{ij} + (\widehat{\beta\gamma})_{jk}$, $i = 1,2$; $j = 1,\ldots,4$; $k = 1,\ldots,5$, which shows that, for female patients, region is related to marital status and site of cancer, while marital status is related to site of cancer. The observed expected frequency and standardized residual are shown in Table 9. The best model of the three-dimensional log-linear models for

Table 8
Analysis of deviance for marital status, site of cancer and region for female patients.

Model	Parameter	Deviance	d.f.	Deviance difference	d.f. difference
(1) Full model: μ, α_i, β_j, γ_k, $(\alpha\beta)_{ij}$, $(\alpha\gamma)_{ik}$, $(\beta\gamma)_{jk}$, $(\alpha\beta\gamma)_{ijk}$		0	0	-	-
(2) Delete $(\alpha\beta\gamma)_{ijk}$ from (1)		10.31	12	10.31	12
(3) Delete $(\alpha\beta)_{ij}$ from (2)		118.10	15	107.79[a]	3
(4) Delete $(\alpha\gamma)_{ik}$ from (2)		28.13	16	17.82[a]	4
(5) Delete $(\beta\gamma)_{jk}$ from (2)		58.98	24	48.67[a]	12

Note: [a]significant at $p<0.05$.

Table 9
Observed, expected frequency and standardized residual of marital status, site of cancer and region for female patients.

Cell i j k	Observed (Expected under H_1)	Expected under H_0	Standardized residual	Cell i j k	Observed (Expected under H_1)	Expected under H_0	Standardized residual
1 1 1	9	13.905	-1.315	2 1 3	210	261.838	-0.464
1 2 1	4	2.907	0.641	2 2 3	207	204.474	0.177
1 3 1	3	3.840	-0.429	2 3 3	30	29.941	0.011
1 4 1	23	18.348	1.086	2 4 3	239	234.746	0.278
2 1 1	66	61.095	0.627	1 1 4	3	2.891	0.064
2 2 1	108	109.093	-0.105	1 2 4	1	0.448	0.824
2 3 1	13	12.160	0.241	1 3 4	2	1.211	0.717
2 4 1	105	109.652	-0.444	1 4 4	1	2.450	-0.926
1 1 2	7	9.162	-0.714	2 1 4	29	29.109	-0.020
1 2 2	3	2.025	0.685	2 2 4	38	38.552	-0.089
1 3 2	2	2.400	-0.258	2 3 4	8	8.789	-0.266
1 4 2	12	10.413	0.492	2 4 4	35	33.550	0.250
2 1 2	53	50.838	0.303	1 1 5	38	37.881	0.019
2 2 2	95	95.975	-0.100	1 2 5	3	3.094	-0.053
2 3 2	10	9.600	0.129	1 3 5	4	3.490	0.273
2 4 2	77	78.587	-0.179	1 4 5	21	21.535	-0.115
1 1 3	75	68.162	0.828	2 1 5	113	113.119	-0.011
1 2 3	5	7.526	-0.921	2 2 5	79	78.906	0.011
1 3 3	13	13.059	-0.016	2 3 5	7	7.510	-0.186
1 4 3	50	54.254	-0.578	2 4 5	88	87.465	0.057

Table 10
Summary of the best model of three-dimensional log-linear models.

Variables		The best model
(1) Marital status (α_i), treatment (β_j), sex (γ_k)		$\hat{\eta}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\alpha\gamma)}_{ik} + \widehat{(\beta\gamma)}_{jk} + \widehat{(\alpha\beta\gamma)}_{ijk}$, $i = 1,2, j = 1,\ldots,4, k = 1,2$
(2) Marital status (α_i), site of cancer (β_j), region (γ_k)	Male:	$\hat{\eta}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\beta\gamma)}_{jk}$,
	Female:	$\hat{\eta}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\alpha\gamma)}_{ik} + \widehat{(\beta\gamma)}_{jk}$, $i = 1,2; j = 1,\ldots,4; k =1,\ldots,5$,
(3) Marital status (α_i), treatment (β_j), region (γ_k)	Male:	$\hat{\eta}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\beta)}_{ij}$,
	Female:	$\hat{\eta}_{ijk} = \hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\gamma}_k + \widehat{(\alpha\beta)}_{ij} + \widehat{(\alpha\gamma)}_{ik} + \widehat{(\beta\gamma)}_{jk}$, $i = 1,2; j = 1,\ldots,4; k =1,\ldots,5$,

each set of three variables is summarized in Table 10.

DISCUSSION

The test of independence can only show whether any two variables are related and cannot find the expected frequency of each parameter. This can be found from the estimates of parameters under the independence model (model under H_0) and the maximal model (model under H_1) by using the two-dimensional log-linear models. The expected frequencies of the maximal model are the observed values. In fact, any pair of relative or non-relative variables can be formulated in the log-linear model. In our study, we present only relative variables which give the maximum Cramer's V value in each table of relationship. Moreover, the test of independence cannot test whether any three or more variables are related but we can employ log-linear models to perform and test hypotheses.

From the numerical results of the two-dimensional log-linear models, Table 7, we see that most male patients were Thai and lived in the central region of Thailand, while other male patients such as Chinese, Japanese, Laotian, Vietnamese, etc lived in Bangkok.

A large number of male patients were Thai who were treated using radiation, while other male patients, such as Chinese, Japanese, Laotian, Vietnamese, were treated with surgery.

The greatest number of male patients who lived in the northeast region of Thailand were diagnosed by using the X-rays/endoscopy/ radiodiagnosis, while male patients who lived in the remaining regions of Thailand: the North, Central, South and Bangkok, were diagnosed using the histology of the primary.

Diagnostic evidence for liver and intrahepatic bile cancer were by X-ray/endoscopy/ radiodiagnostics. For the rectum, rectosigmoid, anus, trachea, bronchus, and lung, the diagnosis was made by histology of the primary.

Treatment for the liver and intrahepatic bile duct cancers was surgery, but for the trachea, bronchus and lung cancer, it was chemotherapy. For other sites of cancer in male patients, such as brain, tongue, and larynx, treatment was by radiation, but for the rectum, rectosigmoid and anus, other treatments, such as surgery with radiation, and surgery with chemotherapy were used.

Alive with cancer was the status of last contact for male patients who were treated using radiation, while alive without cancer and death were the status of last contact for male patients who were treated using surgery.

A large number of female patients were Thai who were treated using radiation, while other female patients, such as Chinese, Japanese, Laotian, and Vietnamese, were treated using other treatments, such as surgery with radiation, and surgery with chemotherapy.

Table 7 shows that most single female patients had breast cancer, while other sites of cancer, such as the mouth, thyroid gland, and stomach, occurred mostly in the female patients who were not single. The greatest number of female patients who lived in Bangkok, the Northeast and the South had breast cancer. The female patients living in the remaining regions, North and Central, had other sites of cancer, such as the mouth, thyroid gland, and stomach.

Clinical, cytology, and hematology were the diagnostic sources for female breast cancer. Histology of the metastases and other diagnostic evidences, such as X rays, endoscopy, radiodiagnosis, and surgery without histology, were used to detect other sites of cancer, such as the mouth, thyroid gland, and stomach. Histology of the primary was the source of diagnostic evidence for cervical/uterine cancer.

The treatments for cervical/uterine cancer was radiation, while for female breast cancer, ovarian cancer, and other sites of cancer, such as the mouth, thyroid gland, and stomach, surgery with radiation, and surgery with chemotherapy were used.

Alive without cancer was the most common status of last contact for female patients who had breast cancer, while alive with cancer and death were the status of last contact for female patients who had other sites of cancer.

In the case of the three-dimensional log-linear model, we obtained the best model, which shows interaction terms of variables that have a relationship. The expected frequencies of the best model can be found from the estimated parameters similar to those of the two-dimensional model. However, from Table 10 there is an interaction of three effects only in model (1). It follows from the best model, model (1) in Table 10, that the association between marital status and treatment is not the same for both sexes. Most single patients were treated by radiation or other methods for males, while female patients were treated by other methods. Furthermore, the best model form shows that for the married/divorced/widowed patients, the number of patients treated using radiation was more than the single patients in both sexes.

According to the best model, model (2) in males in Table 10, the distribution of the site of cancer is not the same for all regions. In each region there is an association between marital status and site of cancer. The relationship between marital status and site of cancer is the same for all regions.

The distribution of treatment was not the same for all marital status in Table 10 model (3) in males. Most male patients who were not single were treated using radiation, while single male patients were treated using other treatments, such as surgery with radiation, or surgery with chemotherapy.

From Table 10 model (2) in females, the best model shows that the distribution of the site of cancer was not the same not only for all the regions but also for marital status. In each region, there was an association between site of cancer and marital status. The relationship between site of cancer and marital status was the same for all regions.

According to the best model in Table 10, model (3) in females, we can explain that the distribution of treatment was not the same for either the regions or the marital status. In each region, there was an association between treatment and marital status and the relationship between treatment and marital status was the same for all the regions.

Finally, for both male and female patients, the variable site of cancer was highly related to personal variables, such as race, religion, marital status, and region, cancer and clinical variables, such as diagnostic evidence, treatment and status at last contact, and also provided the highest Cramer's V value. Marital status and region were the most influential factors affecting the site of cancer for male and female patients. The site of cancer also affected the method of diagnostic evidence and treatment. Since the site of cancer in each sex was different, for public health prevention regarding various sites of cancer, these should be considered for each sex. For both male and female patients, treatment was related to the site of cancer. Consequently, physicians should consider these factors in order to select the appropriate method of treatment.

There were other two variables which were quite important in cancer patients, age and stage of diagnosis. We could not find an association between these two variables and other cancer/clinical variables using log-linear models, since these two variables were ordinal in this study. Therefore, another interesting study would be to study these two variables with other personal and cancer/clinical variables using logistic regression models.

ACKNOWLEDGEMENTS

The authors would like to thank the officers of the National Cancer Institute for access to their complete data records and their helpful guidance. This work was supported by a grant RTA 4580005 from the Thailand Research Fund, Bangkok, Thailand.

REFERENCES

Aitkin M, Anderson D, Frances B, Hinde J. Statistical modelling in GLIM. Oxford: Oxford Science Publication UK, 1989: 67-112.

Baker RJ, Nelder JA. GLIM manual (Release 3). Oxford, UK: Oxford University Press, 1978.

Bhattacharyya GK, Johnson R. Statistical concepts and methods. New York, USA: Wiley, 1997: pp 434.

Burea of Policy and Strategy. Pulic health statistics AD 2002. Ministry of Public Health, Bangkok, 2003.

Dobson AJ. An Introduction to generalized linear models. London: Chapman and Hall, 1996: 123-38.

Fienberg SE. The analysis of cross-classified categorical data, 2nd ed. Massachusetts, USA: MIT Press, 1982.

GLM4: The statistical system for generalized linear modelling. New York, USA: Oxford University Press, 1994: 149-92.

Marija J. SPSS for Windows Base System user's guide release 7.5, Chicago, USA: SPSS, 1993.

Ministry of Public Health. Public health statistics AD 2000. Bangkok: MOTH, 2002.

Nelder JA. Log linear models for contingency tables: a generalization of classical least squares. *Appl Stat* 1974; 23: 323-9.

Nelder JA, Wedderburn RWM. Generalised linear models. *J R Stat Soc Ser A* 1972; 135: 370-84.

Word Health Organization. International classification of disease for oncology. 2nd ed. Geneva: WHO, 1990.

Word Health Organization National Cancer Control Programmes. Geneva: WHO, 2002.