

Cost Estimation for a Single Episode of Lung Cancer Treatment in a Malaysian Tertiary Hospital: A Generalised Linear Model Approach

(Anggaran Kos bagi Satu Episod Rawatan Kanser Paru-paru di Hospital Tertier Malaysia: Pendekatan Model Linear Teritlak)

NORIZA MAJID^{1,*}, AZIMATUN NOOR AIZUDDIN² & YI HUI CHENG¹

¹*Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor, Malaysia*

²*Department of Community Health, Faculty of Medicine, Hospital Canselor Tuanku Muhriz, Universiti Kebangsaan Malaysia, 56000 Kuala Lumpur, Malaysia*

Received: 13 April 2026/Accepted: 8 September 2026

*Corresponding author; email: nm@ukm.edu.my

ABSTRACT

Lung cancer is a leading cause of cancer-related mortality in Malaysia and imposes a substantial financial burden on patients and the healthcare system. This study identifies key determinants of lung cancer treatment costs and develops cost estimates for a single episode of care using Generalised Linear Models (GLM). A total of 1,132 lung cancer patient records from the International Casemix and Clinical Coding Centre (ITCC), Hospital Canselor Tuanku Muhriz (HCTM), spanning 2019 to 2023, were analysed. Patient demographics (age, gender, smoking status) and clinical variables (cancer severity, length of hospitalisation, treatment type, metastatic status, comorbidities, and discharge status) were examined. The GLM with Gamma distribution and logarithmic link function showed the best goodness-of-fit when compared to the Gamma-Log, Gaussian-Log, and Lognormal models. This was based on the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), deviance, and predictive accuracy metrics. The final model identified 17 significant predictors, including length of hospitalisation, cancer severity, brain metastasis, surgical and biopsy procedures, age group, and discharge status, together with seven significant interaction variables. These interaction effects indicate that the co-occurrence of specific clinical factors amplifies treatment costs beyond their individual contributions. The Gamma-Log model provided accurate cost estimates for mild-to-moderate cases but exhibited underestimation for patients with extensive comorbidities who did not receive primary cancer treatment, suggesting that future models should incorporate palliative and supportive care variables. These results offer practical data for the allocation of healthcare resources and cost assessment in Malaysia.

Keywords: Gamma distribution; generalised linear model; lung cancer; treatment costs

ABSTRAK

Kanser paru-paru merupakan punca utama kematian akibat kanser di Malaysia dan memberi tekanan ekonomi yang ketara kepada pesakit serta sistem kesihatan negara. Kajian ini mengenal pasti faktor penentu utama kos rawatan kanser paru-paru dan membangunkan anggaran kos bagi satu episod rawatan menggunakan Model Linear Teritlak (GLM). Sebanyak 1,132 rekod pesakit kanser paru-paru dari Pusat Antarabangsa Casemix dan Pengekodan Klinikal (ITCC), Hospital Canselor Tuanku Muhriz (HCTM) bagi tahun 2019 hingga 2023 telah dianalisis. Pemboleh ubah demografi pesakit (umur, jantina, tabiat merokok) dan klinikal (tahap ketenatan kanser, tempoh penghospitalan, jenis rawatan, status metastasis, komorbid dan status discaj) dikaji. GLM dengan taburan Gamma dan fungsi pautan logaritma menunjukkan kesesuaian terbaik berbanding model Gamma-Log, Gaussian-Log dan Lognormal. Ini berdasarkan Kriteria Maklumat Akaike (AIC), Kriteria Maklumat Bayesian (BIC), devians dan metrik ketepatan ramalan. Model akhir mengenal pasti 17 peramal signifikan termasuk tempoh penghospitalan, tahap ketenatan kanser, metastasis otak, pembedahan dan biopsi, kumpulan umur serta status discaj, bersama-sama tujuh pemboleh ubah interaksi yang signifikan. Model Gamma-Log menghasilkan anggaran kos yang tepat bagi kes ringan hingga sederhana, tetapi cenderung merendah anggaran bagi pesakit dengan komorbid meluas yang tidak menerima rawatan kanser utama. Keputusan ini menawarkan data praktikal untuk peruntukan sumber penjagaan kesihatan dan penilaian kos di Malaysia.

Kata kunci: Kanser paru-paru; kos rawatan; model linear teritlak; taburan Gamma

INTRODUCTION

Lung cancer is a malignant cancer that starts with a genetic mutation in the process of cell division in the cells lining the airways or lung tissue, causing uncontrolled cell growth that can invade surrounding tissue and spread to other organs. According to the American Cancer Society (2024), there are two types of lung cancer, namely non-small cell lung cancer (NSCLC) and small cell lung cancer (SCLC). NSCLC accounts for approximately 85% of all lung cancer cases and can be further divided into three common subtypes, including adenocarcinomas, squamous cell carcinoma and large cell carcinoma. On the other hand, SCLC takes up 15% of all lung cancer cases and spreads more rapidly than NSCLC (Cancer Council 2019; Cancer Research UK 2022).

In Malaysia, lung cancer ranks among the three most frequently diagnosed malignancies and represents the leading cause of cancer-related mortality (Ab Manan et al. 2019). This pattern mirrors global trends: the Global Cancer Observatory (GLOBOCAN 2022) reported 2.5 million new lung cancer cases in 2022, representing 12.4% of all cancer incidences worldwide. The World Health Organization (2024) attributes 1.8 million deaths annually to lung cancer, corresponding to 18.7% of all cancer deaths globally. The disproportionate mortality burden is partly explained by late-stage detection; approximately 90% of Malaysian cases are diagnosed at Stage III or IV, where the five-year relative survival rate is only 11% (Mellor 2019).

Epidemiologically, lung cancer in Malaysia predominantly affects adults aged 55 to 74 years (Khalib et al. 2018). Established risk factors that contribute to the development of lung cancer include tobacco smoking (Schabath & Cote 2019), occupational and ambient air pollutant exposure, chronic pulmonary diseases, respiratory infections, and lifestyle-related behaviours (Bade & Dela Cruz 2020). Common comorbidities in this population include chronic obstructive pulmonary disease (COPD) (36.7%), peripheral vascular disease (PVD) (18.7%), diabetes without chronic complications (17.4%), and congestive heart failure (CHF) (16.5%) (Hernandez et al. 2023; Park et al. 2024).

Lung cancer, in general, imposes a significant financial strain on patients, their families, society and healthcare systems worldwide (Yousefi et al. 2023). Treatment costs vary according to patient age, cancer cell type, disease stage and treatment modality (Kennedy, Hall & Callister 2016). Age, smoking habits, cancer severity, metastatic status, treatment modalities and length of hospitalisation (LOS) are the primary cost drivers of lung cancer care, as indicated by analyses of medical expenses for affected patients (Corral et al. 2015; Fox, Brooks & Kim 2008; Kim et al. 2018; Sari et al. 2016; Yu et al. 2015). Despite a growing number of health economic research in Malaysia, a systematic review by Ku Abdul Rahim and Hanin Farhana (2018) identified only eight

cancer-related economic studies among 39 health economic analyses conducted in the country, none of which focused specifically on lung cancer. This evidence gap limits the capacity for evidence-informed resource allocation and healthcare planning. The present study addresses this gap by examining the demographic and clinical determinants of single-episode lung cancer treatment costs among patients treated at Hospital Canselor Tuanku Muhriz (HCTM) between 2019 and 2023, and by developing a GLM-based cost estimation framework suitable for application in the Malaysian healthcare context.

Beyond the Malaysian context, cancer cost modelling internationally has drawn on several complementary analytical approaches. Two-part or hurdle models have been used where a substantial proportion of patients incur zero cost for a given service thereby addressing the excess-zero problem in cost data (Deb & Norton 2018). Quantile regression has been applied to characterise cost determinants across the full cost distribution rather than focusing solely on the conditional mean, providing valuable insights into heterogeneous cost effects across different expenditure levels (Buchinsky 1998). More recently, machine-learning methods such as random forests and gradient boosting have been explored for healthcare cost prediction because of their ability to model complex nonlinear relationships and improve predictive accuracy in heterogeneous cost data (Breiman 2001; Chen & Guestrin 2016; Rajkomar, Dean & Kohane 2019). These alternative approaches were, however, considered less suitable for the present study. A two-part specification is unnecessary here because every patient in the cohort incurred a non-zero treatment cost, so there is no excess-zero problem to address. Quantile regression, while informative for distributional analysis, does not readily decompose cost drivers into interpretable multiplicative effects on mean cost, limiting its direct use for resource-planning purposes. GLMs were therefore selected because they model the mean of a skewed, strictly positive cost distribution directly through an appropriate link function, they produce readily interpretable multiplicative cost effects, and they are well established in the health-economics literature for analyses of this kind (Buja et al. 2021; Kennedy, Hall & Callister 2016; Manning & Mullahy 2001).

MATERIALS AND METHODS

DATA SOURCE AND VARIABLES

This study employed a retrospective cohort design using routinely collected administrative and clinical data from lung cancer patients treated at the Respiratory Unit of HCTM over a five-year period (1 January 2019 to 31 December 2023). Ethical approval was granted by

the UKM Research Ethics Committee on 18 December 2025 (Reference: UKM PPI/111/8/JEP-2025-726) prior to data retrieval. Patient records were obtained from the International Casemix and Clinical Coding Centre (ITCC) database, with lung cancer admissions identified using the International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) diagnostic codes. The extracted dataset comprised 1,132 patient records. Records with missing values on the dependent variable (total treatment cost) or on any of the core demographic and clinical predictors were excluded at source during data extraction from the ITCC database; consequently, the analytic sample of 1,132 records represents complete cases only, and no imputation was performed. The dependent variable was total lung cancer treatment cost per episode (Malaysian Ringgit, RM), which is continuous, strictly positive, and right skewed. Independent variables included: (i) demographic factors such as age, gender, and history of smoking (nicotine use); and (ii) clinical factors including cancer severity (mild, moderate, severe), LOS (in days, treated as a continuous variable), treatment type (surgery, chemotherapy, biopsy, or combinations thereof), presence and location of metastases, comorbidities (hypertension, pleural effusion, diabetes, hyperlipidaemia, pneumonia), and discharge status. For modelling purposes, patients were classified into two age groups (younger than 65 years versus 65 years and above) based on the sample median, following best-practice recommendations for categorical age groupings in clinical cost analyses (Gallucci 2011). This dichotomisation facilitated interpretation and enhanced the stability of parameter estimates.

METHODOLOGY

Given that treatment cost data are continuous, strictly positive, and exhibit right-skewed distributions characteristic of healthcare expenditure, Generalised Linear Models (GLMs) were employed as the primary analytical framework (Dunn & Smyth 2018; Manning & Mullahy 2001). The GLM comprises three components: (i) the random component, specifying the conditional distribution of the dependent variable Y ; (ii) the systematic component, defining the linear predictor as a function of the independent variables X ; and (iii) the link function $g(\cdot)$, which relates the expected cost of the i -th observation, $E(Y_i)$, to the linear predictor, η_i , expressed as:

$$g[E(Y_i)] = \eta_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + \varepsilon_i, \quad i = 1, \dots, n \quad (1)$$

where Y_i denotes the i -th observation of the dependent variable and X_{ij} represents the quantitative or categorical independent variables corresponding to the j -th predictor. β_0 is the intercept and β_j are the regression coefficients corresponding to the j -th predictor, estimated using the

maximum likelihood estimation (MLE) method. The term ε_i represents the random error associated with the i -th observation. In this study, the dependent variable is lung cancer treatment cost, which is continuous and strictly positive. Three candidate distributions within the exponential dispersion family were evaluated: Gamma, Inverse Gaussian, and Gaussian, each paired with a logarithmic link function (Agresti 2015; Dunn & Smyth 2018). A Lognormal model was additionally assessed as a comparator.

Predictor selection followed a backward elimination procedure. An initial saturated model was specified containing all a priori candidate predictors. At each iteration, the predictor with the largest p -value (least statistically significant) was removed, provided its exclusion did not meaningfully degrade model fit. Interaction terms were considered for clinically plausible predictor combinations and retained in accordance with the principle of model hierarchy (i.e., main effects were preserved whenever their associated interaction terms were significant).

Model adequacy was assessed using a comprehensive set of criteria: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for parsimony; deviance and Pearson dispersion statistics for goodness-of-fit; deviance R^2 (for Gamma and Gaussian models) and conventional R^2 and adjusted- R^2 (for the Lognormal model) to quantify explained variation; Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) for predictive accuracy; quantile-quantile (Q-Q) plots for residual distributional assessment; and generalised variance inflation factors (GVIF) for multicollinearity detection. Note that AIC and BIC are not directly comparable between Lognormal and GLM specifications due to differing likelihood formulations; model comparisons across these two classes therefore relied on deviance R^2 , MAE, and RMSE.

RESULTS AND DISCUSSION

DESCRIPTIVE STATISTICS

Table 1 summarises the frequency distribution of the 1,132 lung cancer patients by gender, age group, cancer severity, and number of comorbidities. There were notably more male patients than female patients, with 755 (66.70%) males compared to 377 (33.30%) females. Lung cancer patients were categorised into three groups, namely younger than 55 years, between 55 and 74 years, and 75 years or older, following Khalib et al. (2018). Initial analysis showed that most patients were between the ages of 55 and 74, with 737 (65.11%) individuals in this category. This finding is consistent with the established epidemiological profile of lung cancer in Malaysia as described by Khalib et al. (2018) indicating that the disease mostly affects individuals aged between 55 and 74 years.

TABLE 1. Frequency distribution of lung cancer patients by gender, age group, cancer severity, and comorbidity burden ($n = 1,132$)

[illegible]

With respect to cancer severity, more than half of patients ($n = 587$, 51.86%) presented with severe disease, while only 126 (11.13%) were classified as mild, indicating that delayed diagnosis remains a significant clinical challenge. Comorbidity analysis demonstrated that 33.30% of patients had no comorbidities, while 24.91%, 23.41%, and 18.37% had one, two, and three or more comorbidities, respectively. Increasing comorbidity burden was associated with older age (≥ 55 years), consistent with established age-related patterns in the literature (Ding et al. 2020).

Table 2 summarises mean treatment costs ($\pm$ standard deviation) stratified by cancer severity, age group, gender, and LOS. Mean costs increased monotonically with severity: RM4,895.54 $\pm$ 2,654.22 (mild), RM7,261.82 $\pm$ 3,495.52 (moderate), and RM10,073.20 $\pm$ 7,272.78 (severe). Patients aged between 55 and 74 years with severe disease incurred the highest mean costs (RM10,195.53 $\pm$ 7,742.66). Cost distributions by gender were broadly comparable across age strata. Extended LOS was strongly associated with higher costs. For example, patients aged 75 years and above with severe disease incurred mean costs of RM8,446.13 $\pm$ 2,302.07 for LOS less than six days versus RM10,613.67 $\pm$ 7,467.86 for LOS extended beyond six days. The finding is in line with previous research connecting extended hospitalisation to increased costs (Umeda et al. 2024; Zhao & Tang 2024).

Table 3 describes treatment cost distributions by treatment type. The majority of patients ($n = 672$, 59.36%) received none of the specified treatment modalities. Patients undergoing both lung surgery and biopsy incurred the highest mean costs (RM11,368.82 $\pm$ 9,731.17), while those treated with surgery and chemotherapy exhibited substantial cost variability (RM9,417.80 $\pm$ 8,186.25). Only one patient received all three treatment types, incurring a total cost of RM9,242.00.

Tables 4 and 5 present cost distributions by metastatic status and comorbidity profile, respectively. Among patients with brain metastasis only, mean costs were RM7,606.91 $\pm$ 5,234.68 (moderate severity, $n = 50$) and RM11,739.40 $\pm$ 7,883.49 (severe severity, $n = 39$). Multi-site metastasis was associated with higher mean costs despite smaller patient numbers; for example, the bone-brain-liver combination yielded a mean cost of RM16,291.61 $\pm$ 16,899.64. Among single-comorbidity patients, those with pleural effusion alone ($n = 119$) incurred the highest mean costs (RM11,266.79 $\pm$ 9,224.37), substantially exceeding those with diabetes alone (RM7,129.81 $\pm$ 2,323.94) or hypertension alone (RM7,682.99 $\pm$ 3,600.60). Multi-comorbidity combinations further elevated costs, with pleural effusion combined with diabetes and hyperlipidaemia yielding the highest observed mean (RM19,601.33 $\pm$ 17,942.89).

MODEL SELECTION

Table 6 presents fit statistics for the three candidate models. The Gamma-Log model attained the lowest AIC (21,659.45) and BIC (21,755.05), outperforming the Gaussian-Log model (AIC = 22,635.87; BIC = 22,726.44). Deviance R^2 was highest for the Gamma-Log model (0.315) compared with the Gaussian-Log (0.238) and Lognormal ($R^2 = 0.243$; adjusted- $R^2 = 0.232$). The Gamma-Log model also exhibited the lowest deviance (240.81) and Pearson dispersion (0.256), contrasting with the Gaussian-Log model which showed substantially inflated deviance values indicative of poor fit. These figures were verified and are correctly reported: because Gaussian deviance is computed as the residual sum of squares on the raw cost scale (RM), and cost values range up to several tens of thousands of Ringgit, squaring these residuals across 1,132 observations produces very large absolute totals; the Gaussian Pearson dispersion statistic (the deviance divided by the residual degrees of freedom) is correspondingly inflated. This scale-dependence is itself a symptom of the Gaussian model's poor suitability for right-skewed cost data, and is not directly comparable to the unitless Gamma deviance, which is computed on the relative (percentage) error scale. In terms of predictive accuracy, the Gamma-Log model achieved a lower RMSE (5,233.29), though marginally higher MAE relative to the Lognormal model. Collectively, these diagnostics supported the Gamma-Log model as the most appropriate specification for the right-skewed, strictly positive treatment cost data.

Although the Gamma-Log model attained the highest deviance R^2 of the three specifications (0.315), this value indicates that the model explains approximately 31% of the deviance in treatment cost, meaning that a substantial share of cost variation remains unexplained by the observed demographic and clinical predictors. This is not unusual for individual-level healthcare cost data, where costs are driven by many unmeasured or unmeasurable factors (e.g., physician-specific practice patterns, exact treatment protocols, complication timing) that are not captured in routinely coded administrative datasets; comparable studies of hospital cost drivers have reported deviance or pseudo- R^2 values in a similar range (Ding et al. 2020; Kennedy, Hall & Callister 2016). A low R^2 does not imply that the model is not statistically significant: the significance of individual predictors and interaction terms in Table 7 is assessed via Wald tests on the estimated coefficients, which is a distinct question from how much of the total deviance those predictors jointly explain. Twelve of the seventeen retained terms in the final model were significant at the 5% level, indicating that the identified predictors have genuine, non-random associations with cost even though other, unobserved sources of cost variation remain.

Figure 1 presents the deviance-residual quantile-quantile (Q-Q) plot for the final Gamma-Log model as a model diagnostic check to assess the adequacy of the fitted model. The residuals generally follow the theoretical reference line, with only moderate deviations observed in the upper tail. These deviations reflect the presence of a few extreme observations, which are common in highly right-skewed treatment cost data. Overall, the

Q-Q plot provides diagnostic evidence that supports the adequacy of the Gamma-Log model and indicates that the model provides a satisfactory fit to the treatment cost data.

Table 7 presents the estimated regression coefficients, exponentiated coefficients $\exp(\beta_j)$, standard errors, and significance tests for the final Gamma-Log model. Exponentiated coefficients are interpreted as

TABLE 2. Mean treatment costs (RM) by cancer severity, age group, gender, and length of hospitalisation

Cost (RM)									
Mean $\pm$ Standard Deviation									
Severity level		Mild			Moderate			Severe	
		4895.54			7261.82			10073.20	
		$\pm$			$\pm$			$\pm$	
		2654.22			3495.52			7272.78	
Age Group	Age < 55	55 $\leq$ Age < 74	Age $\geq$ 75	Age < 55	55 $\leq$ Age < 74	Age $\geq$ 75	Age < 55	55 $\leq$ Age < 74	Age $\geq$ 75
	4995.09	4772.40	5155.96	7489.90	7190.93	7256.88	9731.40	10195.53	9945.18
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$
	3123.75	1922.10	3836.57	3822.14	3262.70	4084.79	6231.87	7742.66	6404.55
<i>Gender</i>									
Male	5252.17	4601.98	4082.06	7441.06	7219.17	6893.14	9793.15	10056.67	10806.92
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$
	4234.45	1615.56	1080.58	2761.92	3177.11	3528.04	7013.82	7586.63	7703.84
Female	4738.01	5179.93	7303.76	7587.59	7126.39	7778.78	9586.60	10468.95	8557.99
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$
	1629.73	2505.58	6105.36	5415.36	3469.27	4808.92	3925.33	8064.20	3009.70
<i>Length of Hospitalisation</i>									
< 6 days	5098.85	4382.54	4611.12	6835.55	7115.10	6809.38	8789.09	8950.59	8446.13
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$
	3438.98	933.38	3012.20	2049.96	3207.22	2777.67	3319.23	5434.60	2302.07
$\geq$ 6 days	4725.31	5704.68	6097.06	8509.93	7312.55	7773.23	10058.59	10777.46	10613.67
	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$	$\pm$
	2421.74	3082.79	4981.53	5448.84	3361.55	5219.25	6954.54	8559.58	7467.86

TABLE 3. Frequency and mean treatment costs (RM) by treatment type combination

Type of treatment	Number of patients (%)	Cost (RM)
		Mean $\pm$ Standard Deviation
Surgery and Biopsy only	258 (22.79%)	11368.82 $\pm$ 9731.17
Chemotherapy only	143 (12.63%)	7406.15 $\pm$ 2273.88
Radiotherapy only	32 (2.83%)	7097.58 $\pm$ 2330.90
Surgery and Biopsy + Chemotherapy	18 (1.59%)	9417.80 $\pm$ 8186.25
Surgery and Biopsy + Radiotherapy	8 (0.71%)	7849.46 $\pm$ 3962.98
Surgery and Biopsy + Chemotherapy + Radiotherapy	1 (0.09%)	9242.00 $\pm$ 0.00
No treatment mentioned above has been received	672 (59.36%)	7606.54 $\pm$ 4106.29

TABLE 4. Frequency and mean treatment costs (RM) by metastatic status and cancer severity

Metastasis	Severity Level			
	Number	Moderate Mean (RM) $\pm$ Standard Deviation	Number	Severe Mean (RM) $\pm$ Standard Deviation
Bone only	41	7389.53 $\pm$ 3649.58	52	8999.84 $\pm$ 4343.84
Brain only	50	7606.91 $\pm$ 5234.68	39	11739.44 $\pm$ 7883.49
Liver only	21	6322.89 $\pm$ 1820.84	31	8843.12 $\pm$ 3810.16
Lymph Node only	5	7111.52 $\pm$ 2483.38	15	13525.68 $\pm$ 11062.59
Bone + Brain only	13	7789.33 $\pm$ 4321.91	13	8922.18 $\pm$ 3630.25
Bone + Liver only	16	6265.17 $\pm$ 1902.31	16	13010.89 $\pm$ 10853.09
Bone + Lymph Node only	17	7396.50 $\pm$ 1172.92	11	12346.65 $\pm$ 9643.09
Brain + Liver only	8	6842.90 $\pm$ 1247.81	14	10866.43 $\pm$ 8640.49
Brain + Lymph Node only	1	7812.00 $\pm$ 0.00	0	0
Liver + Lymph Node only	4	6046.12 $\pm$ 2039.07	3	12138.39 $\pm$ 6130.13
Bone + Brain + Liver only	5	5561.85 $\pm$ 2246.80	3	16291.61 $\pm$ 16899.64
Bone + Brain + Lymph Node only	1	7812.00 $\pm$ 0.00	1	5062.20 $\pm$ 0.00
Bone + Liver + Lymph Node only	4	9771.39 $\pm$ 3918.77	9	12695.11 $\pm$ 10359.33
Brain + Liver + Lymph Node only	2	6046.12 $\pm$ 2497.34	2	24781.00 $\pm$ 21975.46

multiplicative changes in expected cost associated with a one-unit change in the corresponding predictor, holding other variables constant. The final model retained 17 predictor terms in total: one continuous variable (LOS), nine categorical main effects (age group, nicotine use, pleural effusion, pneumonia, brain metastasis, treatment type, cancer severity, discharge status), and seven interaction terms. Of these 17 terms, 12 were statistically significant at the conventional 5% level; the remaining five (age group, nicotine use, pleural effusion, lung surgery and biopsy, and discharge status as main effects) were retained despite not individually reaching $p < 0.05$ because each participates in at least one statistically significant interaction term, and the principle of model hierarchy requires that constituent main effects be kept in the model whenever their associated interaction is significant (Dunn & Smyth 2018). This distinction between the 17 retained model terms and the 12 individually significant terms is clarified here to avoid ambiguity in Table 7.

Length of hospitalisation was the strongest continuous predictor, with each additional inpatient day associated with an estimated 1.1% increase in mean treatment cost ($\exp(\beta) = 1.011$; $p < 0.001$), consistent with findings by Yu et al. (2015). Brain metastasis was the most influential categorical predictor, associated with a 44.7% increase in mean cost ($\exp(\beta) = 1.447$; $p < 0.001$), attributable to the intensive therapeutic interventions, specialised pharmacotherapy, and prolonged hospitalisation required. This study coincides with Girard et al. (2018), who found that the monthly treatment expenses for lung

cancer patients with brain metastasis averaged €2979, which is €553 greater than the expenditures for those without brain metastasis (€2426). The disparity in costs is primarily due to the use of radiation and palliative care. Conversely, the presence of pneumonia as a comorbidity was associated with a 10.5% reduction in mean cost ($\exp(\beta) = 0.895$; $p < 0.05$), a finding that may reflect shortened treatment episodes due to a lower survival rate in this sub-group. Cancer severity exhibited a clear stepwise cost escalation: patients with moderate and severe disease incurred mean costs 49.7% ($\exp(\beta) = 1.497$) and 87.7% ($\exp(\beta) = 1.877$) higher than those with mild disease, respectively (both $p < 0.001$). This pattern aligns with evidence of stage-dependent cost increases in NSCLC (Buja et al. 2021; Jovanoski et al. 2023).

Seven statistically significant interaction terms were retained in the final model. The age-by-surgery interaction indicated that surgical and biopsy procedures incurred an additional 36.4% cost premium for patients aged 65 years and above, relative to younger counterparts. The LOS-by-nicotine-use interaction showed that each additional hospitalisation day was associated with a 1.9% further cost increment for patients with a smoking history. The LOS-by-pleural effusion interaction demonstrated a small negative effect (-0.8%), suggesting that extended hospitalisation does not proportionally increase costs for this subgroup, potentially because pleural effusion management relies predominantly on procedural (rather than pharmacological) interventions; this interpretation is supported by the positive pleural effusion-by-surgery interaction ($\exp(\beta) = 1.481$, $+48.1\%$).

TABLE 5. Frequency and mean treatment costs (RM) by comorbidity combination

Comorbidities	Number of patients (%)	Cost (RM)
		Mean $\pm$ Standard deviation
Hypertension only	49(4.3%)	7682.99 $\pm$ 3600.60
Pleural effusion only	119(10.5%)	11266.79 $\pm$ 9224.37
Diabetes only	20(1.8%)	7129.81 $\pm$ 2323.94
Hyperlipidemia only	7(0.6%)	8561.81 $\pm$ 7958.48
Pneumonia only	87(7.7%)	8224.28 $\pm$ 4246.20
Hypertension + Pleural effusion only	20(1.8%)	9085.23 $\pm$ 3449.63
Hypertension + Diabetes only	86(7.6%)	6963.07 $\pm$ 3500.12
Hypertension + Hyperlipidemia only	61(5.4%)	8849.95 $\pm$ 6573.61
Hypertension + Pneumonia only	15(1.3%)	10278.58 $\pm$ 6324.77
Pleural effusion+ Diabetes only	11(1.0%)	11209.67 $\pm$ 9906.51
Pleural effusion + Hyperlipidemia only	5(0.4%)	15492.37 $\pm$ 12792.85
Pleural effusion + Pneumonia only	48(4.2%)	9052.89 $\pm$ 6417.88
Diabetes + Hyperlipidemia only	9(0.8%)	5697.54 $\pm$ 2514.19
Diabetes + Pneumonia only	8(0.7%)	9919.66 $\pm$ 12514.44
Hyperlipidemia + Pneumonia only	2(0.2%)	6115.00 $\pm$ 0.00
Hypertension + Pleural effusion+ Diabetes only	17(1.5%)	11060.99 $\pm$ 11247.28
Hypertension + Pleural effusion+ Hyperlipidemia only	17(1.5%)	10366.48 8529.18 $\pm$
Hypertension + Pleural effusion+ Pneumonia only	6(0.5%)	7690.75 $\pm$ 1810.03
Hypertension + Diabetes + Hyperlipidemia only	49(4.3%)	7715.29 $\pm$ 3410.56
Hypertension + Diabetes + Pneumonia only	25(2.2%)	7359.24 $\pm$ 2366.42
Hypertension + Hyperlipidemia + Pneumonia only	15(1.3%)	7451.02 $\pm$ 2439.31
Pleural effusion+ Diabetes + Hyperlipidemia only	3(0.3%)	19601.33 $\pm$ 17942.89
Pleural effusion+ Diabetes + Pneumonia only	7(0.6%)	13079.22 $\pm$ 12646.27
Pleural effusion+ Hyperlipidemia + Pneumonia only	2(0.2%)	9242.00 $\pm$ 0.00
Diabetes + Hyperlipidemia + Pneumonia only	1(0.1%)	5062.20 $\pm$ 0.00
Hypertension + Pleural effusion+ Diabetes + Hyperlipidemia only	18(1.6%)	9729.58 $\pm$ 8157.06
Hypertension + Pleural effusion+ Diabetes + Pneumonia only	15(1.3%)	11496.31 $\pm$ 9553.41
Hypertension + Pleural effusion+ Hyperlipidemia + Pneumonia only	7(0.6%)	16113.70 $\pm$ 11061.16
Hypertension + Diabetes + Hyperlipidemia + Pneumonia only	18(1.6%)	8308.38 $\pm$ 3919.02
Pleural effusion+ Diabetes + Hyperlipidemia + Pneumonia only	2(0.2%)	11145.98 $\pm$ 6187.16
Hypertension + Pleural effusion+ Diabetes + Hyperlipidemia + Pneumonia	6(0.5%)	9648.37 $\pm$ 5362.35

All interaction terms involving brain metastasis yielded negative coefficients, indicating that the presence of brain metastasis reduces the additional cost contributions of pneumonia, cancer severity, and death-related discharge status. This pattern is consistent with the possibility that patients with brain metastasis follow distinct, intensive care pathways that level out other cost-influencing factors. In accordance with the principle of model hierarchy, all corresponding main effects were retained to ensure proper interaction and interpretation of

both the individual and combined effects of these factors (Dunn & Smyth 2018). The fitted regression equation for the log-transformed expected cost of a single episode is:

$$\ln E[Y|X] = 8.413 - 0.036x_1 + 0.011x_2 - 0.027x_3 - 0.045x_4 - 0.369x_6 - 0.111x_7 + 0.403x_8 + 0.630x_9 - 0.030x_{10} + 0.310x_{11} + 0.019x_{12} - 0.008x_{13} + 0.393x_{14} - 0.294x_{15} - 0.357x_{16} - 0.314x_{17} \quad (2)$$

TABLE 6. Goodness-of-fit statistics and model selection criteria for candidate GLM specifications

Distribution Link function	Gamma Logarithmic	Gaussian Logarithmic	Lognormal N/A
AIC	21659.45	22635.87	1532.48
BIC	21755.05	22726.44	1623.06
Deviance R^2	0.315	0.238	N/A
R^2 / Adjusted R^2	N/A	N/A	0.243 / 0.232
Deviance	240.81	31035983519	N/A
Pearson Dispersion Statistic	0.256	27835171	N/A
MAE	3055.15	3135.76	2942.59
RMSE	5233.29	5236.12	5429.12

N/A refers to 'Not Applicable'

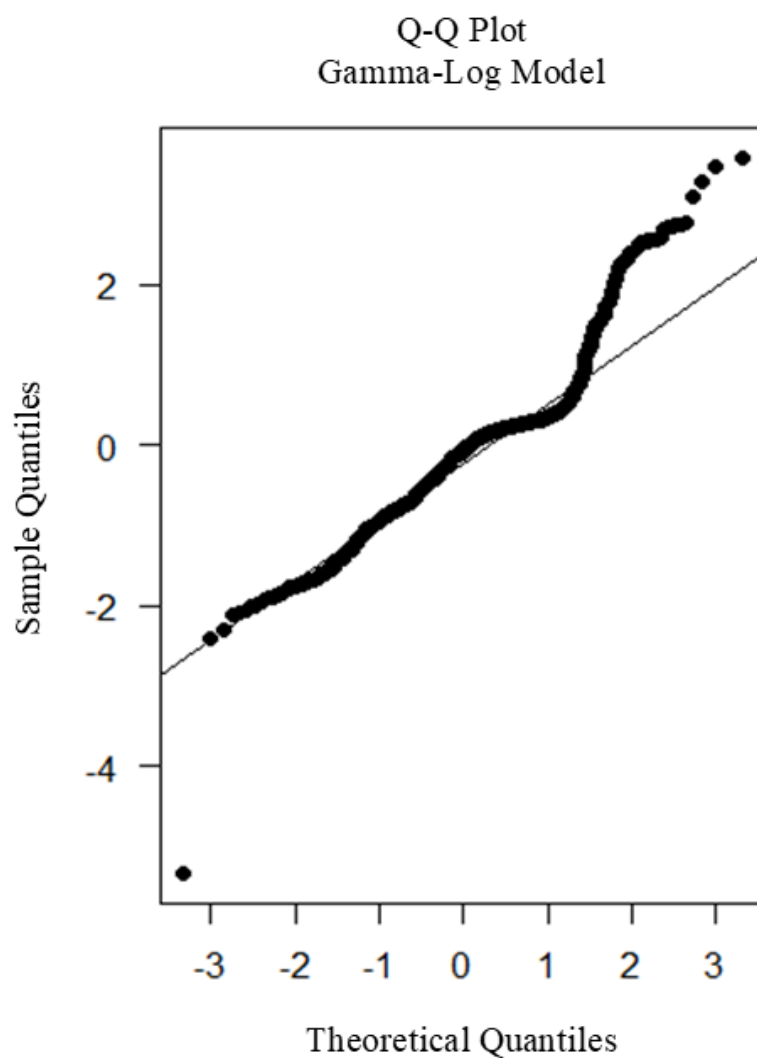


FIGURE 1. Q-Q plot for Gamma-Log model

TABLE 7. Estimation of regression coefficients for Gamma-Log GLM

Predictor	β_i	Estimation of coefficient	Exp(β_i)	Confidence Interval	GVIF value	p-value	Changes to $E(Y X)$ (%)
Intercept	β_0	8.413	4503.846	(8.3080, 8.5174)	-	<0.001 ***	-
Age (Reference – Age less than 65)							
65 years and above	β_1	-0.036	0.964	(-0.1063, 0.0337)	1.3843	0.3093	-3.57
Length of hospitalisation (LOS)	β_2	0.011	1.011	(0.0061, 0.0164)	1.9789	<0.001 ***	1.13
Nicotine use	β_3	-0.027	0.973	(-0.2153, 0.1607)	3.0392	0.7758	-2.69
Comorbidity							
Pleural effusion	β_4	-0.045	0.956	(-0.1698, 0.0797)	3.5132	0.4791	-4.41
Pneumonia	β_5	-0.111	0.895	(-0.1969, -0.0248)	1.5259	0.0117 *	-10.49
Metastasis							
Brain	β_6	0.369	1.447	(0.2240, 0.5144)	3.1874	<0.001 ***	44.66
Treatment							
Lung surgery and biopsy	β_7	-0.111	0.895	(-0.2394, 0.0177)	3.5842	0.0912 .	-10.49
Severity level (Reference – Mild)							
Moderate	β_8	0.403	1.497	(0.2987, 0.5081)	2.9435	<0.001 ***	49.68
Severe	β_9	0.630	1.877	(0.5141, 0.7455)	3.8483	<0.001 ***	87.73
Discharge status (Reference – Discharge to home)							
Died	β_{10}	-0.030	0.970	(-0.1242, 0.0639)	1.2869	0.5300	-2.97
Interaction variables							
Age $\times$ Surgery	β_{11}	0.310	1.364	(0.1703, 0.4502)	2.9492	<0.001 ***	36.37
LOS $\times$ Nicotine use	β_{12}	0.019	1.019	(0.0019, 0.0362)	3.0973	0.0299 *	1.92
LOS $\times$ Pleural effusion	β_{13}	-0.008	0.992	(-0.0164, -0.0002)	3.7501	0.0455 *	-0.82
Pleural effusion $\times$ Surgery	β_{14}	0.393	1.481	(0.2395, 0.5461)	3.1130	<0.001 ***	48.11
Pneumonia $\times$ Brain metastatic	β_{15}	-0.294	0.745	(-0.5376, -0.0510)	1.5298	0.0179 *	-25.50
Brain metastatic $\times$ Moderate severity	β_{16}	-0.357	0.700	(-0.5401, -0.1733)	2.7475	<0.001 ***	-30.00
Brain metastatic $\times$ Died	β_{17}	-0.314	0.731	(-0.5429, -0.0847)	1.5598	0.0074 **	-26.93

***, **, * and $\times$ indicate statistical significance at the 0.001, 0.01, 0.05 and 0.10 levels respectively

DISCUSSION

This study developed and validated a GLM-based framework for estimating single-episode lung cancer treatment costs in a Malaysian tertiary hospital setting. The Gamma distribution with a logarithmic link function was identified as the most appropriate model for right-skewed, strictly positive cost data, a finding that is consistent with the broader health economics literature (Duan et al. 1983; Manning & Mullahy 2001). The identified cost drivers which include LOS, cancer severity, brain metastasis, treatment type, age, and discharge status are consistent with findings from worldwide studies, confirming the external validity of these results within the Malaysian context.

The strong effect of brain metastasis on treatment costs (+44.7%) highlights the need for healthcare

resource planning to account for patients at high neurological risk. The stepwise cost escalation with increasing cancer severity reinforces the economic case for investment in early detection programmes, which could substantially reduce per-episode treatment expenditure by shifting the case-mix towards earlier stages. The interaction effects identified in this study carry practical implications: for example, the amplified cost of surgery in older patients (≥ 65 years) suggests that age-stratified cost models may better reflect actual resource utilisation than models treating age as a linear covariate.

There are several limitations that should be acknowledged. First, the retrospective single-centre design limits generalisability to other Malaysian public or private hospitals. Second, the current model does not explicitly capture costs of supportive and palliative care,

TABLE 8. Observed treatment costs, model-predicted costs, and percentage prediction errors for selected patients by severity group

Patient	Age	LOS	Discharge status	Severity level	Comorbidity	Metastasis	Treatment	Cost (RM)	Estimated Cost (RM)	Percentage difference
A	61	3	Home	Mild	-	-	-	4656.00	4658.10	0.05%
B	59	9	Home	Mild	-	-	-	4656.00	4982.63	7.02%
C	53	4	Home	Mild	Hypertension	-	-	4578.00	4710.68	2.90%
D	65	3	Home	Moderate	Hypertension Diabetes	Bone Liver	Surgery Chemotherapy	7812.00	8207.29	5.06%
E	73	13	Home	Moderate			Surgery	11050.26	9182.21	-16.91%
F	63	11	Home	Moderate		Bone	Surgery	13969.00	6827.10	-51.13%
G	71	27	Died	Severe	Hypertension Diabetes Hyperlipidemia Pneumonia	Brain Liver	Surgery Chemotherapy	9242.00	9216.94	-0.27%
H	60	10	Died	Severe	Hypertension Diabetes Hyperlipidemia Pneumonia	Bone Liver Lymph Node		9242.00	9671.96	4.65%
I	75	9	Home	Severe	Hypertension Hyperlipidemia Pneumonia	Liver Lymph Node		9242.00	8073.92	-12.64%
J	53	10	Died	Severe	Hypertension Diabetes	Bone Brain Liver Lymph Node		7812.00	5414.80	-30.69%

where $x_1 - x_{10}$ denote main effects and $x_{11} - x_{17}$ denote interaction terms, as defined in Table 7.

To illustrate model performance at the individual-patient level, ten patients (Patients A-J) were purposively selected from the dataset rather than drawn at random: three patients from each of the mild and severe strata and four from the moderate stratum, chosen to span a range of comorbidity burden, metastatic involvement, and treatment combinations within each severity level. This purposive approach was adopted so that the illustrative comparison would capture both straightforward cases (e.g., mild disease with no comorbidities) and clinically complex cases (e.g., severe disease with multiple comorbidities and multi-site metastasis) within the same table, rather than relying on chance to include such variation. Table 8 compares observed and model-predicted treatment costs for these patients across severity strata, with percentage errors as the primary accuracy metric. For mild-severity patients, the Gamma-Log model produced stable, low-error estimates with percentage differences of just 0.05%, 7.02%, and 2.90% respectively, indicating

that predicted costs closely align with actual costs in less complex cases. Estimation errors were larger and more variable in the moderate-severity group. Patient D shows a reasonable 5.06% deviation, while Patients E and F exhibit larger discrepancies of -16.91% and -51.13%, suggesting the model underestimates costs when surgery is involved without accompanying comorbidity data. This may suggest the presence of cost-relevant heterogeneity not fully captured by the observed predictors. For severe-severity patients, the errors vary in direction. Patient G is nearly perfect at -0.27%, Patient H shows a modest 4.65% overestimate, while Patients I and J reflect notable underestimations of -12.64% and -30.69%, respectively. Overall, the model performs most reliably for mild cases and becomes more inconsistent with increasing clinical complexity. This is especially true when dealing with situations where there is a lot of uncertainty in cost prediction due to factors like multiple comorbidities, significant metastases, or other missing components that are not considered in the current model specification.

which likely explains the systematic underestimation observed for severe cases. Future work should incorporate these cost components, potentially through additional data linkage with pharmacy and palliative care records. Third, some clinical variables potentially relevant to cost variation such as specific chemotherapy regimens, radiation therapy doses, or molecular targeted therapy usage. Fourth, while backward elimination is a widely used variable selection strategy, it is susceptible to model instability; confirmatory analyses using regularised regression or cross-validation are recommended for future replication studies. Finally, the binary age grouping, while analytically convenient, inevitably sacrifices some precision; a continuous or finer categorical treatment of age may be explored in future analyses. Despite these limitations, this study represents one of the first rigorous statistical analyses of lung cancer treatment costs in Malaysia using GLMs, and provides a reproducible methodological framework applicable to other cancer types or chronic disease cost analyses in the region.

CONCLUSION

This study demonstrates that lung cancer treatment costs in Malaysia are driven by a multifactorial combination of demographic and clinical variables. Length of hospitalisation, cancer severity, brain metastasis, surgical and biopsy procedures, patient age (particularly in interaction with surgery), and discharge status emerged as the principal cost determinants. Significant interaction effects further reveal that the joint presence of specific clinical factors amplifies treatment expenditure beyond the sum of individual contributions, underscoring the complexity of cost modelling in oncology. The Gamma-Log GLM was identified as the most appropriate modelling framework for right-skewed treatment cost data, outperforming Gaussian-Log and Lognormal alternatives across multiple diagnostic criteria. The model demonstrated reliable estimation accuracy for mild-to-moderate severity cases, while systematic underestimation for complex severe cases signals a need to incorporate palliative and supportive care cost components in future model extensions. These findings provide foundational evidence for healthcare resource allocation, cost projections, and economic evaluation of lung cancer care in Malaysia. Policymakers and hospital administrators may use these estimates to anticipate budget requirements across patient subgroups and to prioritise interventions that address the primary drivers of high treatment costs.

ACKNOWLEDGEMENTS

We thank the UKM Research Ethics Committee for granting ethical approval for this study. The support

of the International Casemix and Clinical Coding Centre (ITCC) and the staff of Hospital Canselor Tuanku Muhriz (HCTM) in providing access to the research data is gratefully acknowledged.

REFERENCES

- Ab Manan, A., Basri, H., Kaur, N., Abd Rahman, S.Z., Amir, P.N., Ali, N., Raman, S. & Bahtiar, B. 2019. *Malaysia National Cancer Registry Report* (MNCR) 2012-2016.
- Agresti, A. 2015. *Foundation of Linear and Generalized Linear Models*. 1st ed. New Jersey: John Wiley & Sons.
- American Cancer Society. 2024. *What is Lung Cancer?* <https://www.cancer.org/cancer/types/lung-cancer/about/what-is.html> Accessed on 7 July 2025.
- Bade, B.C. & Dela Cruz, C.S. 2020. Lung cancer 2020: Epidemiology, etiology, and prevention. *Clin. Chest Med.* 41(1): 1-24. DOI: 10.1016/j.ccm.2019.10.001
- Breiman, L. 2001. Random forests. *Machine Learning* 45(1): 5-32. DOI: 10.1023/A:1010933404324
- Buchinsky, M. 1998. Recent advances in quantile regression models: A practical guideline for empirical research. *Journal of Human Resources* 33(1): 88-126. DOI: 10.2307/146316
- Buja, A., Rivera, M., De Polo, A., Brino, E.D., Marchetti, M., Scioni, M., Pasello, G., Bortolami, A., Rebba, V., Schiavon, M., Calabrese, F., Mandoliti, G., Baldo, V. & Conte, P. 2021. Estimated direct costs of non-small cell lung cancer by stage at diagnosis and disease management phase: A whole-disease model. *Thorac Cancer* 12(1): 13-20. DOI: 10.1111/1759-7714.13616
- Cancer Council. 2019. *What is Lung Cancer?* <https://www.cancer.org.au/cancer-information/types-of-cancer/lung-cancer> Accessed on 21 May 2025.
- Cancer Research UK. 2022. *Lung Cancer*. <https://www.cancerresearchuk.org/about-cancer/lung-cancer> Accessed on 21 May 2025.
- Chen, T. & Guestrin, C. 2016. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. pp. 785-794. DOI:10.1145/2939672.2939785
- Corral, J., Espinàs, J.A., Cots, F., Pareja, L., Solà, J., Font, R. & Borràs, J.M. 2015. Estimation of lung cancer diagnosis and treatment costs based on a patient-level analysis in Catalonia (Spain). *BMC Health Services Research* 15(1): 70. DOI: 10.1186/s12913-015-0725-3
- Deb, P. & Norton, E.C. 2018. Modeling health care expenditures and use. *Annual Review of Public Health* 39: 489-505. DOI: 10.1146/annurev-publhealth-040617-013517

- Ding, R., Zhu, D., He, P., Ma, Y., Chen, Z. & Shi, X. 2020. Comorbidity in lung cancer patients and its association with medical service cost and treatment choice in China. *BMC Cancer* 20(1): 250. DOI: 10.1186/s12885-020-06759-8
- Duan, N., Manning Jr. W.G., Morris, C.N. & Newhouse, J.P. 1983. A comparison of alternative models for the demand for medical care. *Journal of Business & Economic Statistics* 1(2): 115-126.
- Dunn, P.K. & Smyth, G.K. 2018. *Generalized Linear Models with Examples in R*. New York/Berlin: Springer Nature.
- Fox, K.M., Brooks, J.M. & Kim, J. 2008. Metastatic non-small cell lung cancer: Costs associated with disease progression. *Am. J. Manag. Care* 14(9): 565-571.
- Gallucci, M. 2011. Best practices for using median splits, artificial categorization, and their continuous alternatives. *Journal of Experimental Psychopathology* 2(2): 197-209. DOI: 10.5127/jep.008310
- Girard, N., Cozzone, D., De Leotoing, L., Tournier, C., Vainchtock, A., Tehard, B. & Cortot, A.B. 2018. Extra cost of brain metastases (BM) in patients with non-squamous non-small cell lung cancer (NSCLC): A French national hospital database analysis. *ESMO Open* 3(6): e000414.
- Global Cancer Observatory (GLOBOCAN). 2022. *Global Cancer Observatory: Cancer Today: Lung*.
- Hernandez, D., Cheng, C.Y., Hernandez-Villafuerte, K. & Schlander, M. 2023. Survival and comorbidities in lung cancer patients: Evidence from administrative claims data in Germany. *Oncol. Res.* 30(4): 173-185. DOI: 10.32604/or.2022.027262
- Jovanoski, N., Abogunrin, S., Di Maio, D., Belleli, R., Hudson, P., Bhadti, S. & Jones, L.G. 2023. Systematic literature review to identify cost and resource use data in patients with early-stage non-small cell lung cancer (NSCLC). *Pharmacoeconomics* 41(11): 1437-1452.
- Kennedy, M.P., Hall, P.S. & Callister, M.E. 2016. Factors affecting hospital costs in lung cancer patients in the United Kingdom. *Lung Cancer* 97: 8-14. DOI: 10.1016/j.lungcan.2016.04.009
- Khalib, A., Ab Manan, A., Amir, P.N., Ali, N., Raman, S., Bahtiar, S. & Raman, R.K. 2018. *Malaysian Study on Cancer Survival (MySCan)*. Putrajaya: Ministry of Health.
- Kim, D., Kim, S.Y., Suh, B. & Park, J.H. 2018. Trend analysis for the choice and cost of lung cancer treatment in South Korea, 2003-2013. *Cancer Res. Treat.* 50(3): 757-767. DOI: 10.4143/crt.2017.050
- Ku Abdul Rahim, K.N. & Hanin Farhana, K. 2018. Systematic review of healthcare economic evaluation on Malaysia: An overview on cancer-related economic studies. *Journal of Global Oncology* 4(Supplement 2): 76s-76s. DOI: 10.1200/jgo.18.20800
- Manning, W.G. & Mullahy, J. 2001. Estimating log models: To transform or not to transform? *Journal of Health Economics* 20(4): 461-494. DOI: 10.1016/s0167-6296(01)00086-8
- Mellor, M. 2019. *Kanser Paru-Paru Memiliki Kadar 'Survival' Terendah*. <https://subangjayamedicalcentre.com/blog-content/kanser-paru-paru-memiliki-kadar-survival-terendah-sjmc-dr-matin-mellor> Accessed on 21 May 2025.
- Park, J.H., Kim, S.J., Medina, M., Prochnow, T., Min, K. & Chang, J. 2024. Are comorbidities associated with differences in healthcare charges among lung cancer patients in US hospitals? Focusing on variances by patient and socioeconomic factors. *Chronic Illn.* 20(3): 434-444. DOI: 10.1177/17423953241241759
- Rajkomar, A., Dean, J. & Kohane, I. 2019. Machine learning in medicine. *New England Journal of Medicine* 380(14): 1347-1358. DOI: 10.1056/NEJMr1814259
- Sari, A.A., Rezaei, S., Arab, M., Majdzadeh, R., Matin, B.K. & Zandian, H. 2016. Effects of smoking on cost of hospitalization and length of stay among patients with lung cancer in Iran: A hospital-based study. *Asian Pac. J. Cancer Prev.* 17(9): 4421-4426. DOI: 10.7314/APJCP.2016.17.9.4421
- Schabath, M.B. & Cote, M.L. 2019. Cancer progress and priorities: Lung cancer. *Cancer Epidemiol. Biomarkers Prev.* 28(10): 1563-1579. DOI: 10.1158/1055-9965.EPI-19-0221
- Umeda, Y., Matsumoto, S., Hagiwara, K., Yoshikawa, S. & Chen, A. 2024. Evaluation of the factors affecting higher hospitalization cost of lung resection for primary lung cancer: A retrospective cohort study. *Ann. Thorac Cardiovasc. Surg.* 30(1): 23-00180. DOI: 10.5761/atcs.0a.23-00180
- World Health Organization (WHO). 2024. Global cancer burden growing, amidst mounting need for services. <https://www.who.int/news/item/01-02-2024-global-cancer-burden-growing--amidst-mounting-need-for-services> Accessed on 21 May 2025.
- Yousefi, M., Jalilian, H., Heydari, S., Seyednejad, F. & Mir, N. 2023. Cost of lung cancer: A systematic review. *Value Health Reg. Issues* 33: 17-26. DOI: 10.1016/j.vhri.2022.07.007
- Yu, T., He, Z., Zhou, Q., Ma, J. & Wei, L. 2015. Analysis of the factors influencing lung cancer hospitalization expenses using data mining. *Thorac Cancer* 6(3): 338-345. DOI: 10.1111/1759-7714.12147
- Zhao, L. & Tang, J. 2024. Analysis of the composition and influencing factors of hospitalization expenses for 1517 patients with lung malignant tumors in Beijing. *Medicine (Baltimore)* 103(11): e37385. DOI: 10.1097/MD.00000000000037385