

ISCHAEMIC HEART DISEASE TREATMENT COST ANALYSIS USING GENERALISED LINEAR MODEL (GLM)

*(Analisis Kos bagi Satu Episod Rawatan Penyakit Jantung Iskemik Menggunakan Model
Linear Teritlak)*

NORIZA MAJID*, HU ZHI YI & AZIMATUN NOOR AIZUDDIN

ABSTRACT

Ischaemic Heart Disease (IHD) is the most commonly diagnosed cardiovascular condition among patients with heart diseases. The global prevalence of IHD has risen significantly, from 112 million cases in 1990 to 254 million cases in 2021, imposing a substantial economic burden on healthcare systems worldwide. In Malaysia, the Ministry of Health (MOH) reported a total IHD treatment expenditure of RM389 million in 2017, equivalent to 38.4% of the total treatment cost of all cardiovascular diseases in public hospitals. Despite this burden, episode-level cost models derived from hospital administrative data, a prerequisite for specific resource planning, Diagnosis-Related Group (DRG) tariff setting and high-cost pathway identification, are still sparse in the Malaysian literature. This gap limits evidence-based healthcare planning and resource allocation. This study estimates the cost of single episode IHD treatment for patients admitted to Hospital Canselor Tuanku Muhriz (HCTM) using a Generalised Linear Model (GLM) framework. Using a retrospective dataset of 7,581 atherosclerotic cardiovascular disease cases, three regression models are compared: the Gamma GLM with a logarithm link function, Gaussian GLM with a logarithm link function, and log-transformed Multiple Linear Regression (MLR). Model selection is based on Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), deviance, Pearson chi-square statistics, and Quantile-Quantile (Q-Q) plots. The Gamma GLM demonstrated the best fit across all evaluation criteria, consistent with the theoretical appropriateness of the Gamma distribution for right-skewed, positive-valued cost data, where variance is proportional to the square of the mean. The key drivers of IHD treatment costs identified are the type of treatment received and the severity level of the patient. Patients undergoing Coronary Artery Bypass Grafting (CABG) incurred costs approximately 7.8 times higher than those managed without surgery, while severe patients incurred costs approximately 1.8 times higher than mild patients. Significant interactions between variables were found and led to a substantial differences in treatment costs. Most notably, the PCI x CABG interaction reduced the expected combined cost multiplier by 93.4% while PCI x Moderate severity level generated a significantly higher additional cost. This suggests that treatment type and patient severity level should be jointly modelled. The results provide an empirical evidence for policy interventions aimed at high-cost treatment pathways and patient risk classification.

Keywords: ischaemic heart disease (IHD); treatment cost; interaction terms; generalised linear model (GLM)

ABSTRAK

Penyakit Jantung Iskemik (IHD) merupakan penyakit kardiovaskular yang paling kerap didiagnosis dalam kalangan pesakit jantung. Prevalens IHD di peringkat global telah meningkat dengan ketara, daripada 112 juta kes pada tahun 1990 kepada 254 juta kes pada tahun 2021, sekali gus memberi beban ekonomi yang besar kepada sistem penjagaan kesihatan di seluruh dunia. Di Malaysia, Kementerian Kesihatan (KKM) melaporkan bahawa jumlah perbelanjaan rawatan IHD pada tahun 2017 adalah sebanyak RM389 juta, bersamaan 38.4% daripada jumlah kos rawatan bagi semua penyakit kardiovaskular di hospital awam. Walaupun beban ini sangat besar, kajian yang meneliti faktor penentu kos rawatan IHD serta menyediakan anggaran kos

pada peringkat satu episod rawatan menggunakan data pentadbiran hospital, yang merupakan prasyarat bagi perancangan sumber secara terperinci, penetapan tarif Kumpulan Berkaitan Diagnosis (DRG) dan pengenalpastian laluan kos-tinggi masih lagi terhad di Malaysia. Jurang ini menghadkan perancangan penjagaan kesihatan dan pengagihan sumber yang berasaskan bukti. Kajian ini bertujuan untuk menganggarkan kos rawatan bagi satu episod IHD untuk pesakit yang dimasukkan ke Hospital Canselor Tuanku Muhriz (HCTM) menggunakan kerangka model linear teritlak (GLM). Dengan menggunakan set data retrospektif yang merangkumi 7,581 kes penyakit kardiovaskular aterosklerotik, tiga model regresi dibandingkan: Gamma GLM dengan fungsi pautan logaritma, Gaussian GLM dengan fungsi pautan logaritma, dan regresi linear berganda (MLR) yang ditransformasi logaritma. Pemilihan model dibuat berdasarkan kriteria informasi Akaike (AIC), kriteria informasi Bayesian (BIC), devians, statistik khi kuasa dua Pearson, dan plot Kuantil-Kuantil (Q-Q). Model GLM Gamma menunjukkan padanan terbaik merentas semua kriteria penilaian, selaras dengan kesesuaian teori taburan Gamma bagi data kos yang bernilai positif dan pencong ke kanan, dengan varians yang berkadar dengan kuasa dua min. Faktor utama yang dikenal pasti mempengaruhi kos rawatan IHD ialah jenis rawatan yang diterima dan tahap keterukan pesakit. Pesakit yang menjalani pembedahan pentasan arteri koronari (CABG) menanggung kos lebih kurang 7.8 kali lebih tinggi berbanding mereka yang dirawat tanpa pembedahan, manakala pesakit yang berada dalam keadaan teruk menanggung kos lebih kurang 1.8 kali lebih tinggi berbanding pesakit dengan tahap ringan. Terdapat interaksi antara pemboleh ubah yang signifikan dan menyebabkan perubahan ketara dalam kos rawatan. Paling jelas adalah interaksi $PCI \times CABG$ yang telah mengurangkan pengganda kos gabungan yang dijangka sebanyak 93.4%, manakala $PCI \times$ tahap keterukan Sederhana pula menghasilkan kos tambahan yang tinggi. Ini mencadangkan bahawa jenis rawatan dan tahap keterukan pesakit perlu dimodelkan secara bersama. Dapatan ini memberikan sokongan empirikal terhadap intervensi dasar yang menyasarkan laluan rawatan kos-tinggi serta pengelasan risiko pesakit.

Kata kunci: penyakit jantung iskemik; kos rawatan; interaksi; model linear teritlak (GLM)

1. Introduction

Ischaemic Heart Disease (IHD), also known as coronary heart disease, is the most commonly diagnosed cardiovascular disease among patients with heart condition (British Heart Foundation 2025). IHD arose from the disruption of blood supply to the heart caused by the accumulation of fatty substances, known as obstructive atherosclerotic plaques, in the coronary arteries (National Health Service 2024 ; Severino *et al.* 2020). The risk of plaque formation increases with unhealthy lifestyle behaviours such as unhealthy diet, use of tobacco, physical inactivity and harmful use of alcohol; as well as medical conditions including hypertension, high cholesterol, elevated lipoprotein levels, diabetes, obesity, and a family history of IHD (National Health Service 2024 ; World Health Organization 2025). IHD usually progresses slowly, showing no symptoms early on, with signs only appearing when metabolic demand of the patient increases, such as during physical activity.

The global burden of IHD has grown substantially. The prevalence of IHD has doubled from 112 million in 1990 to 254 million in 2021 (Safiri *et al.* 2022), with projections reaching 510 million by 2050 (Shi *et al.* 2024). The incidence cases of IHD have similarly doubled from 16 million in 1990 to 32 million in 2021, with the highest burden found in South Asia, accounting for 8.4 million cases (Cheng *et al.* 2025). Although there is limited data on the prevalence of IHD in Malaysia, previous studies showed that 22.6% of atrial fibrillation patients and 12.8% of type-2 diabetes patients also have IHD. IHD is one of the most common causes of deaths worldwide, accounting for 13% of the deaths (World Health Organization 2024). In Malaysia,

IHD remains as the top cause of mortality for 20 consecutive years, with exceptions only in 2021 and 2023 (Department of Statistics Malaysia 2024).

IHD imposes a huge economic burden on the healthcare systems around the world (Rittiphairoj *et al.* 2025), annual healthcare spending of IHD patients are found to exceed national per-capita health expenditure in many countries (Shakya *et al.* 2025). Regional studies from Southeast Asia, including Thailand and Indonesia, have also reported rising IHD burden and associated cost pressures on national health budgets, highlighting a broader ASEAN-level challenge in managing this condition sustainably (Rittiphairoj *et al.* 2025). In Malaysia, IHD treatment accounted for RM389 million, or 38.4% from the total treatment cost of all cardiovascular diseases in the Ministry of Health (MOH) hospital in the year 2017 (Ministry of Health Malaysia 2022). A recent Malaysian study estimated a median treatment cost of RM8365 per IHD admission across three selected hospitals in Peninsular Malaysia (Puteh *et al.* 2024). Despite this burden, Malaysia lacks episode-specific cost models based on hospital administrative data. This gap has three specific policy implications. First, episode-level estimations are necessary to establish Diagnosis-Related Group (DRG) case-mix tariffs, which support activity-based hospital funding. Second, detailed cost modelling identifies high-cost treatment paths, such as CABG in severe patients, where focused clinical protocols can achieve the highest cost savings. Third, regression-based episode models generate covariate-adjusted cost estimates that account for changes in case mix among institutions, which aggregate cost studies cannot do. Existing Malaysian research is either disease-aggregate (Ministry of Health Malaysia 2022) or limited to certain patient sub-categories (Puteh *et al.* 2024), resulting in an evidence gap that this study addresses directly.

The literature identifies several factors of IHD treatment cost determinants. These factors are divided into two categories, namely the demographic factors and clinical factors. Demographic factors include age and gender, where older patients tend to incur lower costs due to receiving less invasive procedures (Fan *et al.* 2015; Krumholz *et al.* 1998), and male patients are generally incur higher costs (Kazemi *et al.* 2024; Shakya *et al.* 2025; Xu *et al.* 2024). Clinical factors also play an important role in determining IHD treatment costs. Clinical factors includes comorbidities such as hypertension (Yang *et al.* 2020), hyperlipidaemia (Venkitachalam *et al.* 2012), diabetes (Abdissa *et al.* 2020), chronic kidney disease (Jarrick *et al.* 2021) and complications such as heart failure (Abdissa 2020) which are positively associated with longer hospitalisation and higher episodic treatment costs (Puteh *et al.* 2024; Takroni *et al.* 2023; Xu *et al.* 2024). Severity level is another established predictor, with more severe patients incurring substantially higher costs (Xu *et al.* 2024; Puteh *et al.* 2024). The type of treatment received is among the strongest cost drivers. Patients who underwent coronary artery bypass graft (CABG) bear the highest costs, followed by those who received percutaneous coronary intervention (PCI), and finally those managed medically (Liu *et al.* 2022). Length of stay (LOS) and in-hospital mortality status are also associated with treatment costs (Kazemi *et al.* 2024; Mallow *et al.* 2023; Xu *et al.* 2024). Although this data base is present, the pairwise interaction effects between these cost determinants, i.e., the cost impact of a variable that is moderated by the presence of another variable, have rarely been explored in the IHD cost literature, particularly in the Malaysian setting. Modelling such interactions is vital as cost accumulation in complex multi-morbid patient populations may be fundamentally non-additive and main-effect only models risk systematic mis-estimation of spending for certain patient profiles. Based on this evidence, the variables included in the present study are age, LOS, gender, severity level (mild/moderate/severe), discharge status, comorbidities (hypertension, diabetes, hyperlipidaemia, unstable angina, chronic kidney disease), complications (acute renal failure, heart failure), and treatment type (anticoagulant injection, PCI, CABG).

Healthcare cost data exhibit recognised statistical problems that require solution by suitable modelling selections. Treatment cost distributions are typically right-skewed, strictly positive, and heteroskedastic, characteristics that make standard ordinary least squares (OLS) regression unsuitable without adjustment, as OLS assumes normally distributed, homoskedastic errors (Manning & Mullahy 2001; Blough & Ramsey 2000). Two primary methodologies have emerged in the health economics literature to tackle these characteristics: the log-transformation of the dependent variable within a multiple linear regression (MLR) framework, and the application of generalised linear models (GLM) with an appropriate distributional family and link function.

The log-transformation of the cost variable, specifically modelling $\log(Y)$ instead of Y directly, compresses the right tail of the distribution, diminishes heteroskedasticity, and yields residuals that more closely resemble normality (Jones 2011; Manning & Mullahy 2001). A logarithm link is used rather than the canonical reciprocal link because it ensures strictly positive predictions and yields coefficients interpretable as proportional effects on the mean cost. The logarithm link function is used because to the strictly positive nature of cost and the plausible multiplicative relationship between predictors and cost, indicating that a one-unit change in a predictor scales cost by a constant factor rather than altering it by a constant amount. On a logarithmic scale, this multiplicative connection transforms into a linear form, thus facilitating the application of conventional regression methods. Nonetheless, a significant limitation of log-transformed multiple linear regression (MLR) is the retransformation issue: back-transforming predictions to the original cost scale using the exponential function produces biased estimates unless a smearing correction is applied (Manning & Mullahy 2001).

The GLM framework resolves this by modelling cost directly on the original scale, accommodating non-normal error structures through specification of a distributional family and link function (Nelder & Wedderburn 1972). A logarithm link function is retained in the GLM to enforce positivity of predicted costs and to preserve the multiplicative interpretation of coefficients. Among candidate distributional families, the Gamma distribution is theoretically appropriate for cost data, as it assumes that variance is proportional to the square of the mean. It is consistent with the empirical observation that high-cost patients exhibit substantially greater absolute variability than low-cost patients (Blough & Ramsey 2000; Manning & Mullahy 2001). A logarithm link is used rather than the canonical reciprocal link because it ensures strictly positive predictions and yields coefficients interpretable as proportional effects on the mean cost.

This study fills a research gap by estimating the episode-level IHD treatment cost using hospital administrative data from HCTM and building a regression model appropriate for skewed cost data. This study identifies significant demographic and clinical determinants of IHD treatment costs; compares three regression modelling approaches for cost estimation: Gamma GLM, Gaussian GLM, and log-transformed MLR; examines clinically motivated pairwise interaction terms between treatment type, severity, and comorbidities to capture non-additive cost effects; and finally selects the most appropriate model based on statistical criteria to support evidence-based healthcare funding and hospital budgeting in Malaysia.

2. Methodology

2.1. Data

This study employed retrospective data from patients diagnosed with ischaemic heart disease (IHD) who were admitted to Hospital Canselor Tuanku Muhriz (HCTM), a public tertiary teaching hospital associated with Universiti Kebangsaan Malaysia (UKM), over a five-year

period from 1 January 2019 to 31 December 2023. It should be noted that the study period spans the COVID-19 pandemic years (2020–2021), during which cardiac care delivery, elective procedural volumes, and hospital costs in Malaysia were substantially disrupted. No pandemic-period correction or sensitivity analysis was performed; this represents a potential source of temporal confounding that future studies should address. The data were acquired from the International Casemix and Clinical Coding Centre (ITCC), HCTM, subsequent to ethical approval granted by the Secretariat of Research Ethics of Universiti Kebangsaan Malaysia (JEPUKM) under reference number JEP-2025-725 on 6 November 2025. This dataset has not been analysed in any prior published study. It contains episode-specific records of total treatment costs per hospital admission, patient demographic attributes (age and gender), clinical variables (comorbidities and complications recorded during admission), patient severity classification (mild, moderate, or severe), length of hospital stay (LOS), discharge status, and type of treatment administered. The use of routinely collected hospital administrative data enables assessment of actual treatment costs and their determinants within a tertiary care environment. The inclusion criterion encompassed all patients with a verified IHD diagnosis admitted to HCTM within the designated study period. Three exclusion criteria were applied to ensure comparability and statistical reliability: (1) admissions outside the study window (before 1 January 2019 or after 31 December 2023); (2) patients treated at the UKM Specialist Centre, which operates under a distinct administrative and billing framework yielding non-comparable cost structures; and (3) non-atherosclerotic IHD subtypes with insufficient sample sizes for reliable regression estimation, as limited subgroup representation undermines the stability and interpretability of regression coefficients. Following data cleaning and application of these criteria, the final analytical dataset comprised 7,581 atherosclerotic cardiovascular disease cases. Following the events per variable (EPV) heuristic commonly applied in regression modelling, a minimum of approximately 10–20 observations per predictor is recommended for stable coefficient estimation (Agresti 2015). With 23 parameters in the final model and 7,581 cases, this study substantially exceeds this threshold, supporting the stability and reliability of the regression estimates.

2.2. Variable Selection and Model Specification

The dependent variable, Y , represents the treatment cost, whereas the independent variable, X , comprises the demographic and clinical parameters. Independent variables were chosen based on their clinical significance as documented in previous literature and the availability of data. The initial collection of independent variables examined comprised: age (X_1), length of stay (LOS) (X_2), gender (X_3), severity level (moderate: X_4 ; severe: X_5 , with mild as the reference category), discharge status - death (X_6), comorbidities: hypertension (X_7), diabetes (X_8), hyperlipidaemia (X_9), unstable angina (X_{10}), chronic kidney disease (X_{11}), complications: acute renal failure (X_{12}), heart failure (X_{13}), and treatment type: anticoagulant injection (X_{14}), PCI (X_{15}), CABG (X_{16}). Indicator dummy variables were created for all categorical predictors. Clinically reasonable interaction terms were considered a priori, such as treatment type and severity level, treatment types, LOS and comorbidities, age and hypertension, and comorbidities and severity level. Interaction terms that significantly improved model fit, as evidenced by reduction in AIC, were maintained in the final model. Table 6 presents the ten interaction terms that passed the requirement. Variables that did not increase model fit were dropped from the final specification.

Backward elimination guided by the Akaike Information Criterion (AIC) was employed to reduce the candidate variable set to the final model. AIC penalises model complexity while rewarding goodness of fit, providing a more principled basis for variable selection than p-value

thresholds alone (Agresti 2015; Dunn & Smyth 2018). As a secondary check, variables with p -values exceeding 0.05 and a negligible AIC improvement upon removal were dropped. The final model therefore reflected a balance between model fit, parsimony, statistical significance, and clinical relevance. Plausible interaction terms, specifically between treatment type and severity level, and between age and severity level were assessed during model development. Those that produced statistically significant improvements in model fit (AIC reduction) were retained in the final model; the remainder were excluded. Multicollinearity was assessed using the Variance Inflation Factor (VIF) and Generalised VIF (GVIF) (Fox & Monette 1992). No significant multicollinearity issues were identified in the final model.

2.3. Regression Models

2.3.1. Multiple linear regression with log transformation

Multiple Linear Regression (MLR) is a widely used approach for healthcare cost modelling (Dwikasari *et al.* 2023; Taloba *et al.* 2022), expressing the dependent variable Y (treatment cost) as a linear combination of k independent variables. The general form is:

$$E[y_i] = \mu_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_k x_{ki} + \varepsilon_i \quad (1)$$

where y_i denotes the i -th observation of the dependent variable and X_{ki} represents the quantitative or categorical independent variables corresponding to the k -th predictor. β_0 is the intercept and β_j are the regression coefficients corresponding to the k -th predictor. The term ε_i represents the random error associated with the i -th observation. Standard MLR assumes normality and homoskedasticity of residuals, both of which are typically violated by cost data. Log-transformation of the dependent variable ($\log Y$) partially addresses skewness and heteroskedasticity (Jones 2011). A logarithm link is used because costs are strictly positive and the relationship between predictors and costs is plausibly multiplicative: a one-unit change in a predictor scales cost by a constant factor, which translates to a linear effect on the log scale. However, back-transformation to the original scale introduces retransformation bias, which must be corrected using methods such as the smearing estimator (Manning & Mullahy 2001). This correction was applied in the estimation phase.

2.3.2. Generalised linear model (GLM)

GLM is widely considered a more appropriate framework for healthcare cost data than standard MLR (Malehi *et al.* 2015). GLM explicitly accommodates non-normal response distributions and non-constant variance through the specification of a distributional family and a link function. Both the Gamma GLM and the Gaussian GLM retain the log link function and thus their model structures are directly comparable and so AIC and BIC comparisons between the two are valid. The log-transformed MLR is not considered for AIC and BIC comparisons as transforming the response variable changes the likelihood scale and hence comparison with GLMs is not appropriate. The log-transformed MLR is not included in the AIC and BIC comparisons as the transformation of the response variable alters the likelihood scale and hence the comparison with GLMs is not appropriate. In the GLM framework, the model is expressed as:

$$y_i \sim EDM\left(\mu_i, \frac{\phi}{w_i}\right)$$

$$g(\mu_i) = o_i + \beta_0 + \sum_{j=1}^p \beta_j x_{ij} \quad (2)$$

where $g(\cdot)$ is the link function, o_i is an optional offset, EDM denotes the exponential dispersion model family, and ϕ is the dispersion parameter. 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. The Gamma distribution with logarithm link function is theoretically appropriate for positive, right-skewed cost data because it assumes that the variance scales with the square of the mean ($\text{Var}(Y) \propto \mu^2$) relationship consistent with the empirical cost distributions, where high-cost patients exhibit substantially greater variability than low-cost patients (Blough & Ramsey 2000; Zhou *et al.* 2024). Unlike log-transformed MLR, Gamma GLM estimates costs directly on the original scale without requiring retransformation, thereby avoiding retransformation bias entirely (Chen *et al.* 2020). The logarithm link rather than the canonical reciprocal link is used to ensure predicted costs are strictly positive and to preserve multiplicative interpretation of coefficients. The Gaussian GLM with log link is estimated as an alternative to assess whether the distributional assumption influences model performance relative to the Gamma. Model parameters for both GLM specifications were estimated using Maximum Likelihood Estimation (MLE (Greene 2012; Montgomery & Runger 2014)).

2.4. Model Validation and Evaluation

Model performance was evaluated using multiple complementary criteria. For the two GLM models, penalised likelihood measures were computed:

$$AIC = -2 \times l(\hat{\beta}_0, \dots, \hat{\beta}_p, \phi; y) + 2p' \quad (3)$$

$$BIC = -2 \times l(\hat{\beta}_0, \dots, \hat{\beta}_p, \phi; y) + (\log n) p' \quad (4)$$

As both the Gamma and Gaussian GLMs use the same log link function, their likelihoods are evaluated on the same scale, making AIC and BIC comparisons valid between them. The log-transformed MLR was excluded from these comparisons because transformation of the dependent variable alters the likelihood scale, rendering direct comparison with GLMs inappropriate. Goodness-of-fit was assessed using deviance and Pearson chi-square statistics:

$$\text{Deviance}, D(y; \hat{\mu}) = -2[l(\hat{\mu}; y) - l(y; y)] \quad (5)$$

$$\text{Pearson Chi - Square Statistics}, \chi^2 = \sum \frac{(o-e)^2}{e} \quad (6)$$

Predictive accuracy was assessed using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (8)$$

MAE is less sensitive to extreme outliers than RMSE; both are reported to provide complementary perspectives on predictive accuracy (Hodson 2022). Quantile-Quantile (Q-Q) plots were used to visually assess agreement between the fitted distribution and empirical data (Dodge 2008). The dispersion parameter (ϕ) for each GLM was also inspected as an additional fit diagnostic: a Gamma GLM with lower dispersion relative to a Gaussian GLM indicates that the Gamma's variance assumption more closely matches the data's actual variance structure. The contrast between the Gamma GLM's dispersion of 0.336 and the Gaussian GLM's dispersion of 30,325,736 is particularly informative: the Gamma's near-unity dispersion indicates that its variance-mean squared relationship closely matches the empirical cost structure, whereas the Gaussian's extreme dispersion value confirms that the constant-variance assumption is fundamentally incompatible with the heteroskedastic nature of the data. Residual diagnostics, including plots of deviance residuals against fitted values and a normal probability plot of residuals were inspected to identify systematic deviations, heteroskedasticity patterns, and the influence of extreme observations.

3. Results and Discussion

3.1. Descriptive Analysis

Table 1: Frequency distribution of patients by gender, age group, severity, number of comorbidities and complications count

Number of Patient = 7581 (%)						
Gender	Male 5490 (72.42%)			Female 2091 (27.58%)		
Age	< 60 years 1502 (19.81%)	60 - 74 years 3036 (40.05%)	> 74 years 952 (12.56%)	< 60 years 449 (5.92%)	60 - 74 years 1189 (15.68%)	> 74 years 453 (5.98%)
Severity						
Mild	645 (8.51%)	903 (11.91%)	216 (2.85%)	164 (2.16%)	379 (5.00%)	97 (1.28%)
Moderate	415 (5.47%)	1051 (13.86%)	366 (4.83%)	157 (2.07%)	396 (5.22%)	179 (2.36%)
Severe	442 (5.83%)	1082 (14.27%)	370 (4.88%)	128 (1.69%)	414 (5.46%)	177 (2.33%)
Number of Comorbidities and Complications						
0	228 (3.01%)	177 (2.33%)	62 (0.82%)	48 (0.63%)	52 (0.69%)	26 (0.34%)
1	308 (4.06%)	547 (7.22%)	170 (2.24%)	97 (1.28%)	215 (2.84%)	82 (1.08%)
2	445 (5.87%)	986 (13.01%)	264 (3.48%)	144 (1.90%)	354 (4.67%)	147 (1.94%)
3	366 (4.83%)	875 (11.54%)	260 (3.43%)	112 (1.48%)	362 (4.78%)	110 (1.45%)
4	121 (1.60%)	344 (4.54%)	139 (1.83%)	39 (0.51%)	167 (2.20%)	70 (0.92%)
5 and above	34 (0.45%)	107 (1.41%)	57 (0.75%)	9 (0.12%)	39 (0.51%)	18 (0.24%)

Table 1 shows the distribution of the 7,581 IHD patients according to gender, age group, severity level, and number of comorbidities and complications. Male patients aged between 60 to 74 years form the dominant demographic group among IHD patients in HCTM consistent with epidemiological patterns reported in other Malaysian and Asian studies. All three severity categories (mild, moderate, severe) are represented in broadly similar proportions. Patients aged 60 to 74 tend to present with higher severity levels compared to younger cohorts, likely reflecting cumulative cardiovascular risk factor accrual over time, a trend that aligns with the observed age-comorbidity relationship (Barnett *et al.* 2012; Salive 2013). Five comorbidities (hypertension, diabetes, hyperlipidaemia, unstable angina, and chronic kidney disease) and two complications (acute kidney failure and heart failure) were selected based on their established clinical association with IHD in the published literature and their prevalence in the study population at HCTM. Most patients were diagnosed with two or three comorbidities and complications (30.87% and 27.50% respectively). No significant gender differences were observed in comorbidities distribution, although female patients tended to have slightly higher comorbidities and complications count. The literature suggests that while male and female IHD patients may have different types of comorbidities, the overall number of conditions tends to be similar (Mehta *et al.* 2016; Majidi *et al.* 2021). Insignificantly greater comorbidity rates in female patients are clinically feasible. Female IHD patients typically present at an older age than males, as oestrogen offers cardiovascular protection prior to menopause (Mehta *et al.* 2016). When IHD develops in women, they have had longer exposure to risk factors, allowing more comorbid illnesses to accrue, which may explain the slightly higher counts. Age-based study revealed an association between age and the prevalence of comorbidities and complications. Patients aged 60 years and above were more frequently represented in the higher comorbidity categories compared with younger patients. The association between age and number of comorbidities was also further assessed by using Spearman's correlation. A statistically significant positive correlation was observed between the two variables ($\rho = 0.111, p < 0.001$). However, the strength of correlation was weak. This finding is consistent with previous studies reporting that multimorbidity increases with advancing age (Barnett *et al.* 2012; Salive 2013). Prolonged exposure to cardiovascular risk factors, including hypertension, dyslipidaemia, and insulin resistance, facilitates the emergence of more chronic diseases over time. As a result, elderly people with ischaemic heart disease (IHD) typically have a higher prevalence of comorbidities. Considering that age might directly and indirectly affect treatment costs via multimorbidity, the comorbidities ought to be incorporated in the regression model.

Table 2: Descriptive statistics of IHD treatment costs (RM)

Statistic	Value (RM)
Mean (standard deviation)	8,164.37 (9,941.60)
Median	5,310.00
Mode	4,790.00
Minimum	96.00
Maximum	96,104.00
Range	96,008.00
Interquartile Range (IQR)	7,126.00
Skewness (α_3)	5.00
Kurtosis (α_4)	34.85

Table 2 summarises the distributional properties of IHD treatment cost. The mean treatment cost of RM8,164.37 with a considerable standard deviation of RM9,941.60 indicating large variation in treatment costs amongst patients. The median was RM 5,310.00. The mean is much greater than the median and the mode, showing that the distribution of treatment costs is positively skewed, with some patients having very high treatment expenses while most patients incurred moderate treatment costs. The minimum treatment cost recorded was RM 96.00 while the maximum was RM 96,104.00 which gave a very wide range of RM 96,008.00. This implies a considerable variation in treatment costs from the minimum to the maximum. The interquartile range (IQR) of RM 7,126.00 also demonstrates a reasonable spread of the middle 50% of the data. Skewness score is 5.00 which indicates a substantial positive right-skewness of the distribution. Moreover, the kurtosis value of 34.85 is far above the normal distribution standard, indicating a leptokurtic distribution with a long right tail driven by a small proportion of patients with extremely high costs, mostly those undergoing CABG at high severity. The results generally show that the costs of IHD treatment are rather diverse in patients. Most patients are characterised by moderate costs, and a minority number of patients have very high expenditures of medical care. This strong right-skewness has direct implications for model selection. It provides empirical support for the Gamma GLM over OLS-based approaches. Gamma GLMs are appropriate for positive and highly skewed cost data since the Gamma distribution assumes that variance is proportional to the square of the mean, allowing for heteroskedasticity in which the spread of treatment costs grows as the mean cost increases. OLS regression, on the other hand, assumes a Gaussian distribution with constant variance at all levels of the outcome variable, an assumption that is obviously contradicted by the observed cost distribution. Overall, the results reveal heterogeneity in treatment costs for IHD and a skewed distribution of costs among a small group of patients who account for disproportionately large expenditures. The Gamma GLM is thus a more appropriate and statistically sound framework for modelling healthcare cost data of this sort.

Table 3 shows that treatment costs for male patients are higher than for female patients, consistent with prior studies (Kazemi *et al.* 2024; Shakya *et al.* 2025; Xu *et al.* 2024). Among males, costs decrease with age, with the highest mean costs for patients with an age below 60 years, and lowest for those aged 75 and above. Probably because invasive procedures such as PCI are more frequently offered to younger patients for whom the benefits of procedures are better documented (Fan *et al.* 2015). An increasing cost gradient across severity levels is consistently observed for both genders and all age groups, providing descriptive support for severity as a key regression covariate. No clear monotonic relationship is observed between comorbidity count and treatment cost in the bivariate analysis; this is expected because comorbidity burden interacts with treatment decisions, severity classification, and LOS in ways that can only be separated through multivariate regression.

Table 4 illustrates the dominant role of treatment type as a cost determinant at the descriptive level. CABG, as an open-heart surgical procedure, involves substantially higher costs relative to PCI or anticoagulant therapy due to the utilization of operating room resources, intensive care unit requirements, prolonged hospital stays, and post-operative monitoring. The difference between CABG (RM60,970) and the no-specified-treatment group (RM6,108), a tenfold gap suggests the strong influence of treatment type that emerges from the regression analysis.

Table 3: IHD treatment cost by gender, age group , severity level and number of comorbidities and complications

Mean Cost ± Standard Deviation (RM)						
Gender	Male			Female		
	8363.42 ± 10588.38			7641.73 ± 7976.96		
Age	< 60 years	60 - 74 years	> 74 years	< 60 years	60 - 74 years	> 74 years
	9387.06 ± 12862.85	8297.69 ± 10164.07	6958.01 ± 7253.81	7565.09 ± 7222.51	7973.21 ± 8796.21	6847.64 ± 6183.52
Severity						
Mild	6167.74 ± 5711.56	5792.92 ± 5277.06	5068.08 ± 5274.63	5340.89 ± 4723.26	5175.44 ± 3453.52	4837.92 ± 3620.39
Moderate	8477.72 ± 9965.63	7672.60 ± 7966.05	6527.78 ± 6176.50	7967.23 ± 6839.99	7696.88 ± 7010.24	6628.71 ± 5852.40
Severe	14938.75 ± 19376.30	10995.28 ± 13866.08	8486.91 ± 8777.46	9921.62 ± 9258.90	10798.78 ± 12222.44	8170.42 ± 7236.92
Number of Comorbidities and Complications						
0	12599.70 ± 17976.32	8373.69 ± 9566.04	6433.40 ± 7021.39	6740.37 ± 4473.04	9952.10 ± 13935.42	6005.05 ± 4485.85
1	8798.49 ± 10501.11	7194.78 ± 8219.32	6894.98 ± 6466.32	7613.78 ± 6388.64	7334.09 ± 9577.48	7279.30 ± 7151.46
2	9212.82 ± 12809.91	8675.85 ± 10958.37	6780.65 ± 6430.45	8051.57 ± 7806.65	7389.23 ± 8111.38	6387.47 ± 5539.95
3	8690.33 ± 12379.71	8224.86 ± 10024.23	6965.13 ± 7329.80	7090.27 ± 7669.38	8101.43 ± 8282.37	7061.54 ± 6519.23
4	7303.79 ± 6924.31	9105.10 ± 11706.75	7776.09 ± 9928.22	6778.15 ± 6672.81	9177.15 ± 8226.34	7532.01 ± 6942.72
5 and above	10369.92 ± 10116.37	8325.33 ± 7768.09	6510.72 ± 5075.02	12974.43 ± 12224.84	7813.47 ± 8113.59	5887.73 ± 2146.68

Table 4: IHD treatment cost by type of treatment offered

Type of Treatment	Mean Cost ± Standard Deviation (RM)
Anticoagulant Injection	15005.07 ± 15849.95
Percutaneous Coronary Intervention (PCI)	18567.95 ± 15012.70
Coronary Artery Bypass Grafting (CABG)	60970.05 ± 17147.44
Not offered any treatment stated above	6107.79 ± 4717.10

3.2. Model and Treatment Cost Estimation

Comparisons are made across three regression models built, Gamma GLM with log link function, Gaussian GLM with log link function and MLR model with log transformation in order to select the best model fulfilling the purpose of estimating the IHD treatment costs of patients.

The comparison of statistical aspects between the models are recorded in Table 5. The log-transformed MLR model was excluded from AIC and BIC comparisons because transformation of the dependent variable leads to artificially lower AIC and BIC values, making the comparison with GLM models inappropriate. The best model is the Gamma GLM. The Gamma records a substantially lower AIC and BIC, indicating better penalised fit despite using fewer parameters. Most importantly, the Gamma GLM has deviance and Pearson chi-square that are

orders of magnitude lower than those of the Gaussian GLM, indicating that the constant-variance assumption in the Gaussian is fundamentally incompatible with the heteroskedastic structure of the cost data. The dispersion parameter further reinforces this conclusion: the Gamma GLM's dispersion of 0.336 indicates that its variance-mean squared relationship closely matches the empirical cost data, whereas the Gaussian GLM's extreme dispersion of 30,325,736 confirms a severe distributional misspecification. The Gaussian GLM's lower RMSE (RM5,491 vs RM6,892) and higher pseudo- R^2 (0.695 vs 0.579) reflect its better performance on smooth, central-distribution predictions but these advantages do not compensate for its distributional misspecification, particularly given that the cost data's heavy right tail is precisely where accurate estimation matters most for budgeting purposes.

Table 5: Model Comparison between Gamma GLM, Gaussian GLM and log-transformed MLR

Aspects	Model		
	GLM - Gamma (Log Link Function)	GLM - Gaussian (Log Link Function)	MLR (Log Transformation)
Number of Variables	23	45	26
AIC	143845	152163	12449
BIC	144012	152482	12636
Adjusted R^2	-	-	0.471
Pseudo- R^2 (Deviance R^2)	0.579	0.695	-
Deviance	2158.856	2.28535E+11	2276.859
Pearson Value	2542.265	2.28535E+11	2276.859
Dispersion	0.336	30325736	0.301
MAE	3105.107	3043.295	3075.351
RMSE	6891.558	5490.512	5975.077

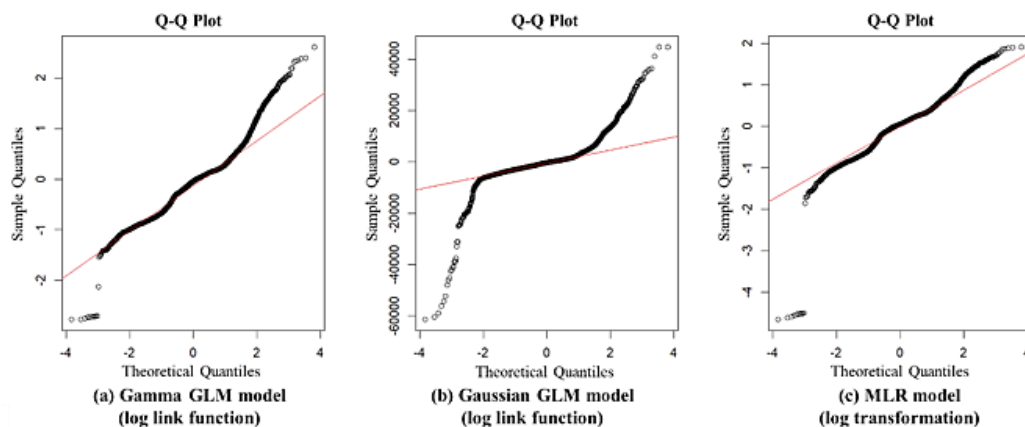


Figure 1: Q-Q Plots for gamma GLM, Gaussian GLM and lognormal MLR model

Besides the statistical measures in Table 5, the Q-Q plots (Figure 1) corroborate these findings. In contrast, both the Gaussian GLM and log-transformed MLR exhibit systematic deviation in the upper tail, with the Gaussian plot showing significant upward curvature and the MLR plot showing downward departure, confirming that neither model adequately captures the extreme high-cost observations of most significance for healthcare budgeting. The final Gamma GLM equation (on the original cost scale) is:

$$E[Y|X] = e^{8.301} \cdot e^{-0.003X_1} \cdot e^{0.021X_2} \cdot e^{0.359X_4} \cdot e^{0.586X_5} \cdot e^{0.080X_6} \cdot e^{0.081X_7} \cdot e^{-0.148X_{10}} \cdot e^{0.113X_{11}} \cdot e^{-0.178X_{13}} \cdot e^{0.127X_{14}} \cdot e^{0.885X_{15}} \cdot e^{2.055X_{16}} \cdot e^{0.003X_{17}} \cdot e^{-0.018X_{18}} \cdot e^{-0.006X_{19}} \cdot e^{-0.174X_{20}} \cdot e^{-0.101X_{21}} \cdot e^{-0.104X_{22}} \cdot e^{0.206X_{23}} \cdot e^{-2.716X_{24}} \cdot e^{0.507X_{25}} \cdot e^{0.264X_{26}} \quad (9)$$

Table 6 presents the parameter estimates from the Gamma Generalized Linear Model (GLM) with a log-link function for ischemic heart disease (IHD) treatment costs. In a Gamma GLM with a log-link, the exponentiated coefficients, $\text{Exp}(\beta_i)$, represent multiplicative effects on the expected treatment cost. Parameters with positive coefficient increases treatment cost mean for e^{β_i} times when X_i increase for 1 unit; while parameters with negative coefficient decreases treatment cost mean for e^{β_i} times when X_i increase for 1 unit, assuming no change for other variables. Thus the change on $E[Y|X]$ column in Table 6 indicates the percentage increase or decrease in expected treatment costs associated with each predictor.

The intercept coefficient was statistically significant ($p < 0.001$), representing the baseline expected treatment cost when all predictors were at their reference categories. Among the demographic variables, age showed a small but statistically significant negative association with treatment costs ($\beta = -0.003$, $p = 0.001$). The exponentiated coefficient of 0.997 indicates that for every additional year of age, the expected treatment cost decreased by approximately 0.3%, holding other variables constant.

Treatment type exerted the strongest influence on mean IHD treatment costs, consistent with findings from China (Xie *et al.* 2023) and other international settings (Liu *et al.* 2022). Patients undergoing CABG incurred costs approximately 7.8 times higher than patients who did not receive surgery, corresponding to a 680.5% increase in expected treatment costs ($\text{Exp}(\beta) = 7.805$, $p < 0.001$). Similarly, patients undergoing PCI experienced treatment costs approximately 2.4 times higher than the no-specified-treatment reference group, representing a 142.4% increase in expected costs ($\text{Exp}(\beta) = 2.424$, $p < 0.001$). These findings confirm that invasive cardiac revascularisation procedures are the dominant drivers of IHD treatment expenditure. From a healthcare policy perspective, these findings have important implications for hospital budgeting and Diagnosis-Related Group (DRG) tariff calibration. CABG cases should receive substantially higher case weights within any activity-based funding system.

Severity was also identified as a major determinant of treatment expenditure. Compared with patients in the mild severity category, moderate severity cases incurred 43.2% higher expected costs ($\text{Exp}(\beta) = 1.432$, $p < 0.001$), while severe cases incurred 79.6% higher costs ($\text{Exp}(\beta) = 1.796$, $p < 0.001$). This positive dose-response relationship between disease severity and treatment cost is consistent with previous healthcare expenditure studies (Xu *et al.* 2024; Puteh *et al.* 2024). The severity gradient also highlights the economic importance of early secondary prevention and timely clinical intervention. Based on the model estimates, avoiding escalation from mild to severe severity corresponds to an estimated cost avoidance of approximately RM 4,780 per treatment episode, assuming other covariates remain at their mean values.

Length of stay (LOS) showed a positive relationship with treatment costs ($\beta = 0.021$, $p < 0.001$). The exponentiated coefficient of 1.021 suggests that each additional inpatient day increased expected treatment costs by approximately 2.1%. This finding is consistent with those of Kazemi *et al.* 2024 and Xu *et al.* 2024. Several comorbidities significantly influenced treatment costs. Chronic kidney disease (CKD) increased expected treatment costs by 12.0% ($\text{Exp}(\beta) = 1.120$, $p = 0.002$). The negative coefficient for heart complications ($\beta = -0.178$, $\text{Exp}(\beta) = 0.837$, $p < 0.001$), corresponding to a 16.3% reduction in adjusted treatment costs, likely reflects treatment selection effects. Patients with heart failure or severe cardiac

complications may have been managed conservatively with medical therapy rather than undergoing costly surgical interventions. Heart failure often increases operative risk beyond acceptable thresholds. Similarly, unstable angina was associated with a 13.7% reduction in expected treatment costs ($\text{Exp}(\beta) = 0.863$, $p < 0.001$) reflecting patient inclination towards protocolised early discharge pathways for low-to-intermediate risk presentations rather than prolonged inpatient monitoring or catheterisation (Ibanez *et al.* 2018).

The interaction terms further demonstrate that the effects of clinical conditions and treatments vary across patient subgroups. The Age $\times$ Hypertension interaction (X_{17} , $\beta = 0.003$, $p = 0.010$) partially offsets the negative main effect of age, indicating that the cost-saving effect of increasing age is decreased in hypertensive patients. This reflects the additional pharmacological complexity and monitoring burden of managing long-standing hypertension during acute cardiac admission. The interactions between LOS and unstable angina (X_{18} , $\beta = -0.018$, $p = 0.001$) and LOS and CKD (X_{19} , $\beta = -0.006$, $p = 0.003$) show that the marginal daily cost of admission is lower in these patient groups compared to the general LOS main effect. This is likely due to the fact that extended stays in these patients involve ward-level medical management rather than high-dependency procedural care. Anticoagulant injections combined with severe disease severity increased expected treatment costs by 22.9% (X_{23} , $\beta = 0.206$, $p = 0.001$). This reflects intensified treatment requirements among critically ill patients receiving anticoagulation, including more frequent laboratory monitoring, dose adjustment, and management of bleeding complications. The interaction between unstable angina and severity (X_{20} , $\beta = -0.174$, $p < 0.001$) resulted in a 15.9% cost reduction, in addition to the already negative main effect of unstable angina. This is consistent with clinical risk channelling, where the co-occurrence of unstable angina with established comorbidity reinforces conservative management strategies that avoid costly procedural pathways.

The interaction between PCI and CABG produced a very large negative coefficient ($\beta = -2.716$, $\text{Exp}(\beta) = 0.066$, $p < 0.001$), attenuated the product of the two main effects by 93.4%. This is the most notable result in the model. Without the interaction, the overall cost multiplier would be around 19-fold (2.424×7.805). However, the interaction term reduces this to about 1.25-fold. This nearly-complete cancellation most plausibly reflects the clinical and administrative characteristics of hybrid revascularisation, a strategy combining minimally invasive surgical bypass with percutaneous stenting in selected multi-vessel disease patients where the surgical component is intentionally limited, the admission is shorter than a full CABG stay, and DRG coding consolidates both procedures under a single classification that prevents multiplicative cost stacking (Harskamp *et al.* 2013).

Both PCI $\times$ Moderate (X_{25} , $\beta = 0.507$, $p < 0.001$) and PCI $\times$ Severe (X_{26} , $\beta = 0.264$, $p < 0.001$) interactions suggests superadditive cost implications for PCI patients at higher severity levels. For moderate-severity patients, the combined log-cost contribution ($0.885 + 0.359 + 0.507 = 1.751$) gives a total multiplier of about 5.76 times baseline, which is significantly higher than the product of the two main effects alone ($2.424 \times 1.432 \approx 3.47$), illustrating the disproportionate pre- and post-procedural resource demands of intervening in haemodynamically unstable patients, including advanced physiological lesion assessment, more complex procedural requirements, and prolonged high-dependency monitoring (De Bruyne *et al.* 2014). It is to be noted that the PCI $\times$ Severe interaction is smaller than PCI $\times$ Moderate, providing a total multiplier of about 5.67 times baseline for severe PCI patients, slightly smaller than its moderate counterpart despite higher severity designation. This clinically meaningful inversion suggests that critically ill patients undergoing primary PCI follow time-driven, protocolised pathways (such as STEMI door-to-balloon protocols) that

limit the diagnostic and planning phase that inflates cost in the elective setting, and may involve single-vessel culprit revascularisation with deferred multi-vessel treatment (Ibanez et al. 2018).

Table 6 also shows the VIF value for each predictor variables selected. VIF is accounted in this study instead of GVIF since all variables are either continuous or binary, which their interaction terms are all subjected to degree of freedoms of 1, allowing the use of VIF in determining the multicollinearity among the predictor variables. The VIF values ranged from 1.048 to 5.030, where only the variable Chronic Kidney Disease (X11, VIF=5.030) has a VIF value higher than 5, displaying potential risk for multicollinearity. However, the variable is not removed from the model since its intraction terms with Length of Stay (X19) and Severity Levels (X21 and X22) are shown significant in the model. Whilst for all other variables, the values were all below the commonly used threshold of 5, indicating that multicollinearity was not a concern in the final model.

Table 6: Parameters for gamma GLM with log link function

Parameters	X_i	Coefficient Estimation, β_i	Standard Deviation	p -Value	VIF	Exponential (β_i)	Change on $E(Y X)$ (%)
Intercept	X_0	8.301	0.019	<0.001***	-	4027.729	-
Age	X_1	-0.003	0.001	0.001**	2.751	0.997	-0.3%
Length of Stay (LOS)	X_2	0.021	0.001	<0.001***	1.643	1.021	2.1%
Severity							
Moderate	X_4	0.359	0.021	<0.001***	2.188	1.432	43.2%
Severe	X_5	0.586	0.021	<0.001***	2.319	1.796	79.6%
Discharged Status							
Death	X_6	0.080	0.037	0.030*	1.065	1.083	8.3%
Comorbidity							
Hypertension	X_7	0.081	0.015	<0.001***	1.076	1.085	8.5%
Unstable Angina	X_{10}	-0.148	0.031	<0.001***	1.994	0.863	-13.7%
Chronic Kidney Disease	X_{11}	0.113	0.036	0.002**	5.030	1.120	12.0%
Complication							
Heart Failure	X_{13}	-0.178	0.023	<0.001***	1.071	0.837	-16.3%
Treatment							
Anticoagulant Injection	X_{14}	0.127	0.030	<0.001***	1.963	1.136	13.6%
PCI	X_{15}	0.885	0.035	<0.001***	2.568	2.424	142.4%
CABG	X_{16}	2.055	0.060	<0.001***	1.048	7.805	680.5%
Interaction Terms							
Age × Hypertension	X_{17}	0.003	0.001	0.010**	2.657	1.003	0.3%
LOS × Unstable Angina	X_{18}	-0.018	0.005	0.001**	1.293	0.982	-1.8%
LOS × Chronic Kidney Disease	X_{19}	-0.006	0.002	0.003**	1.562	0.994	-0.6%
Unstable Angina × Moderate	X_{20}	-0.174	0.045	<0.001***	1.931	0.841	-15.9%
Chronic Kidney Disease × Moderate	X_{21}	-0.101	0.044	0.024*	3.405	0.904	-9.6%
Chronic Kidney Disease × Severe	X_{22}	-0.104	0.045	0.021*	4.144	0.902	-9.8%
Anticoagulant Injection × Severe	X_{23}	0.206	0.063	0.001**	2.145	1.229	22.9%
PCI × CABG	X_{24}	-2.716	0.584	<0.001***	1.014	0.066	-93.4%
PCI × Moderate	X_{25}	0.507	0.054	<0.001***	1.609	1.661	66.1%
PCI × Severe	X_{26}	0.264	0.065	<0.001***	2.303	1.302	30.2%

***, **, * and · denote that the parameters are significant at the 0.001, 0.01, 0.05 and 0.10 significance levels, respectively.

Overall, the Gamma GLM results demonstrate that treatment type, disease severity, hospitalization duration, and comorbid conditions are the principal determinants of IHD treatment costs. Among these factors, invasive procedures such as CABG and PCI contributed

the largest increases in expenditure, while disease severity showed a strong dose-response relationship with healthcare costs. The significant interaction effects further highlight the complex relationships between treatment pathways, patient severity, and clinical comorbidities. Importantly, the model-estimated mean treatment cost of approximately RM8,164 per episode is broadly consistent with findings reported in previous Malaysian studies, supporting the external plausibility of the estimates despite the single-centre study design. It is important to note that the previous estimate by Puteh *et al.* (2024) for Malaysia is reported as a median of RM8,365, whereas the current estimate is a mean. Due to the right-skewed cost distribution, these two measures of central tendency are not directly comparable and the comparison should be understood accordingly. The Gamma GLM therefore provides a statistically robust and policy-relevant framework for modelling highly skewed healthcare cost data and generating covariate-adjusted estimates suitable for DRG tariff development, hospital budgeting, and healthcare resource planning.

Table 7 illustrates individual-level predictions for a subset of patients from the held-out validation set. Overall, the percentage errors varied from -28.7% to +33.4%. In the context of predictive cost modelling, error ranges of this magnitude are regarded as acceptable for treatment cost prediction applications (Akinbode & Taiwo 2025), especially considering the significant variance of the cost distribution. The commonly accepted benchmark of a 15% - 45% tolerable error range in health economic modelling (Akinbode & Taiwo 2025) is utilised as a practical reference, while more severe thresholds may be useful for future budgeting applications.

The largest positive error (+33.4%) was observed for patient's ID 423183, whose actual cost exceeded the model's prediction, most likely due to the combination of extended LOS and anticoagulant injection treatment, resulting in an unusually high cost profile relative to other patients with similar observable characteristics. Patient's ID 410281 had the highest negative error (-28.7%), indicating an extreme outlier that the model's central tendency framework failed to capture. This is an inherent limitation of parametric regression models applied to skewed data, while the Gamma GLM performs well across the central mass of the distribution, it systematically underestimates costs for the highest-cost cases.

Several study limitations should be noted. First, the data came from a single tertiary care facility (HCTM), limiting their applicability to other hospitals, healthcare settings, or Malaysian areas. Second, the model was not verified on an external dataset from another institution, therefore external validity is unknown. Third, the dataset lacked socioeconomic characteristics such as household income and insurance status, which may affect costs and be crucial for equity analysis. Fourth, like other parametric regression models, the Gamma GLM underperforms for patients at the extreme upper end of the cost distribution. Two-part models, quantile regression, or machine learning could enhance prediction for high-cost outliers. To better evaluate IHD in Malaysia economically, future research should use multi-center data, patient-reported outcomes, and indirect costs.

4. Conclusion

This study used a Gamma Generalised Linear Model with a log link function, which was chosen as the best framework above Gaussian GLM and log-transformed MLR, to estimate episode-level IHD treatment costs among 7,581 patients admitted to HCTM between 2019 and 2023. Treatment type and disease severity emerged as the primary cost determinants, with CABG patients incurring costs approximately 7.8 times higher than the non-surgical reference group and severe patients incurring costs 79.6% higher than mild patients, with length of stay, chronic kidney disease, and other comorbidities serving as secondary predictors.

Table 7: Predicted versus actual treatment costs for a selected sample of individual patients using the Gamma GLM (held-out validation subsample)

Patient's ID	Age	Length of Stay (LOS)	Discharge Status	Diagnosis of Patients			Treatment Type	Severity	Estimated Cost (RM)	Actual Cost (RM)	Error (RM)	Percentage Error (RM)
				Comorbidities	Complications							
299974	34	3	Alive	None	None		None	Mild	4155.82	3517.34	638.48	18.2%
424860	63	5	Alive	Hypertension	None		None	Mild	4230.05	4790.00	-559.95	-11.7%
410281	71	12	Alive	Hypertension	None		CABG	Severe	68511.88	96046.00	-27534.12	-28.7%
423183	45	22	Alive	Hypertension, Chronic Kidney Disease	None		Anticoagulant Injection, PCI	Moderate	36183.93	27129.00	9054.93	33.4%
289128	73	3	Alive	Hypertension	None		PCI, CABG	Mild	5075.52	5075.52	0.00	0.0%
323045	71	15	Dead	Hypertension, Chronic Kidney Disease	None		PCI	Moderate	31198.15	27129.00	4069.15	15.0%
425552	71	16	Alive	None	Heart Failure		Anticoagulant Injection, PCI	Moderate	26301.78	27129.00	-827.22	-3.0%
363974	79	15	Alive	Hypertension, Chronic Kidney Disease, Unstable Angina	Heart Failure		None	Severe	5559.65	4996.00	563.65	11.3%
330089	64	21	Alive	Hypertension	Heart Failure, Acute Kidney Failure		PCI	Severe	27922.98	36235.00	-8312.02	-22.9%
402716	75	6	Dead	Hypertension, Chronic Kidney Disease	Heart Failure, Acute Kidney Failure		Anticoagulant Injection, PCI	Severe	28965.84	36235.00	-7269.16	-20.1%

The inclusion of interaction terms revealed that cost accumulation is fundamentally non-additive. Specifically, the PCI $\times$ CABG interaction collapsed the expected combined cost multiplier by 93.4%, while PCI $\times$ Moderate severity generated a superadditive cost premium exceeding the product of either main effect alone. This demonstrates that treatment type and patient severity must be considered jointly rather than independently in any cost-prediction or DRG-based reimbursement. The model-estimated mean treatment cost of around RM8,164 per episode is largely consistent with previous Malaysian estimates, indicating external plausibility, while individual-level projections came within acceptable tolerance ranges for health economic modelling.

These findings establish an empirical foundation for activity-based hospital finance, early secondary prevention investment, and organised high-cost pathway management in Malaysian cardiac care. Future research could broaden this technique to multi-center datasets, integrate socioeconomic factors, and investigate two-part models using quantile regression to increase predicted accuracy at the higher end of the cost distribution.

Acknowledgments

We thank the UKM Research Ethics Committee for their ethical approval of the research, as well as the International Casemix and Clinical Coding Centre (ITCC), Hospital Canselor Tuanku Muhriz (HCTM) for data assistance.

References

- Abdissa S.G. 2020. Predictors of incident heart failure in a cohort of patients with ischemic heart disease. *Pan African Medical Journal* **35**(1): 1-12.
- Abdissa S.G., Deressa W. & Shah A.J. 2020. Incidence of heart failure among diabetic patients with ischemic heart disease: A cohort study. *BMC Cardiovascular Disorders* **20**(1): 1-9.
- Agresti A. 2015. *Foundations of Linear and Generalized Linear Models*. Hoboken, NJ: John Wiley & Sons, Inc.
- Akinbode A. & Taiwo K. 2025. Predictive modeling for healthcare cost analysis in the United States: A comprehensive review and future directions. *International Journal of Scientific Research and Modern Technology* **4**(1): 170-181.
- Barnett K., Mercer S.W., Norbury M., Watt G., Wyke S. & Guthrie B. 2012. Epidemiology of multimorbidity and implications for health care, research, and medical education: A cross-sectional study. *The Lancet* **380**(9836): 37-43.
- Blough D.K. & Ramsey S.D. 2000. Using generalized linear models to assess medical care costs. *Health Services and Outcomes Research Methodology* **1**(2): 185-202.
- British Heart Foundation. 2025. BHF UK CVD Factsheet. <https://www.bhf.org.uk/what-we-do/our-research/heart-statistics/heart-statistics-publications/cardiovascular-disease-statistics-2025> (5 January 2026).
- Chen J., Gu Y., Jones A.M. & Peng B. 2020. Modelling healthcare costs: A semiparametric extension of generalised linear models. *Macquarie Business School Research Paper*.
- Cheng Q., Zhou S., Zhong H., Wang Z., Liu C., Sun J. & Deng J. 2025. Global, regional, and national burden and risk factors of ischemic heart disease, 1990–2021: an analysis of the global burden of disease study. *Frontiers in Public Health* **13**:1563631.
- De Bruyne B., Fearon W.F., Pijls N.H., Barbato E., Tonino P., Piroth Z., Jagic N., Mobius-Winckler S., Rioufol G., Witt N., Kala P., MacCarthy P., Engström T., Oldroyd K., Mavromatis K., Manoharan G., Verlee P., Frobert O., Curzen N., Johnson J.B., Limacher A., Nüesch E., Jüni P. 2014. Fractional flow reserve-guided PCI for stable coronary artery disease. *New England Journal of Medicine* **371**(13): 1208-1217.
- Department of Statistics Malaysia. 2024. Statistics on causes of death, Malaysia, 2024. <https://www.dosm.gov.my/portal-main/release-content/statistics-on-causes-of-death-malaysia-2024> (5 January 2026).
- Dodge Y. 2008. *The Concise Encyclopedia of Statistics*. New York: Springer.
- Dunn P.K. & Smyth G.K. 2018. *Generalized Linear Models with Examples in R*. New York: Springer.
- Dwikasari N.M.D., Sutramiani N.P., Putri K.S.Y., Kusuma N.T.R., Pramana M.D.A.D. & Darma I.W.A.S. 2023. Medical costs estimation using linear regression method. *Jurnal Ilmiah Merpati* **11**(3): 171-179.

- Fan G.Q., Fu K.L., Jin C.W., Wang X.Z., Han L., Wang H., Zhong M., Zhang Y., Zhang W. & Wang Z.H. 2015. A medical costs study of older patients with acute myocardial infarction and metabolic syndrome in hospital. *Clinical Interventions in Aging* **10**: 329-337.
- Fox J. & Monette G. 1992. Generalized collinearity diagnostics. *Journal of the American Statistical Association* **87**(417): 178-183.
- Greene W.H. 2012. *Econometric Analysis*. 7th Ed. Boston, MA: Pearson Education Limited.
- Harskamp R.E., Lopes R.D., Baisden C.E., de Winter R.J., Alexander J.H. 2013. Saphenous vein graft failure after coronary artery bypass surgery: pathophysiology, management, and future directions. *Annals of Surgery*. **257**(5):824-33.
- Hodson T.O. 2022. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. *Geoscientific Model Development Discussions*. **15**(14): 5481-5487.
- Ibanez B., James S., Agewall S., Antunes M.J., Bucciarelli-Ducci C., Bueno H., Caforio A.L.P., Crea F., Goudevenos J.A., Halvorsen S., Hindricks G., Kastrati A., Lenzen M.J., Prescott E., Roffi M., Valgimigli M., Varenhorst C., Vranckx P., Widimský P. 2018. 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *European Heart Journal* **39**(2):119-177.
- Jarrick S., Lundberg S., Sundström J., Symreng A., Warnqvist A. & Ludvigsson J.F. 2021. Immunoglobulin A nephropathy and ischemic heart disease: A nationwide population-based cohort study. *BMC Nephrology* **22**(1): 165.
- Jones A.M. 2011. 625 Models for health care. In Clements M.P. & Hendry D.F. (eds.). *The Oxford Handbook of Economic Forecasting*: 625–654. New York: Oxford University Press.
- Kazemi Z., Emamgholipour S., Daroudi R., Yunesian M. & Hassanvand M.S. 2024. Estimation and determinants of direct hospitalisation cost for coronary heart disease in a low-middle-income country: evidence from a nationwide study in Iranian hospitals. *BMJ Open* **14**(8): e074711.
- Krumholz H.M., Chen J., Murillo J.E., Cohen D.J. & Radford M.J. 1998. Clinical correlates of in-hospital costs for acute myocardial infarction in patients 65 years of age and older. *American Heart Journal* **135**(3): 523-531.
- Liu L., Xu Y., Yu J., Man X., Jiang Y., Zhao L. & Cheng W. 2022. The impact of comprehensive public hospital reforms on the direct medical cost of inpatients with coronary heart disease. *Frontiers in Public Health* **10**: 891186.
- Majidi M, Eslami V, Ghorbani P, Foroughi M. 2021. Are women more susceptible to ischemic heart disease compared to men? A literature overview. *Journal of Geriatric Cardiology* **18**(4):289-296.
- Malehi A.S., Pourmohammadi F. & Angali K.A. 2015. Statistical models for the analysis of skewed healthcare cost data: A simulation study. *Health Economics Review* **5**(1): 11.
- Mallow P.J., Browne F. & Shemisa K. 2023. The high cost of death after acute myocardial infarctions: results from a national US hospital database. *ClinicoEconomics and Outcomes Research* **15**: 63-68.
- Manning W.G. & Mullahy J. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics* **20**(4): 461-494.
- Mehta L.S., Beckie T.M., DeVon H.A., Grines C.L., Krumholz H.M., Johnson M.N., Lindley K.J., Vaccarino V., Wang T.Y., Watson K.E. & Wenger N.K. 2016. Acute myocardial infarction in women: a scientific statement from the American Heart Association. *Circulation* **133**(9): 916-947.
- Ministry of Health Malaysia. 2022. *The Direct Health-Care Cost of Noncommunicable Diseases in Malaysia*. Putrajaya: Ministry of Health Malaysia.
- Montgomery D. & Runger G. 2014. *Applied Statistics And Probability For Engineers*. New York, NY: John Wiley & Sons
- National Health Service. 2024. Coronary heart disease. <https://www.nhs.uk/conditions/coronary-heart-disease> (5 Jun 2025).
- Nelder J.A. & Wedderburn R.W.M. 1972. Generalized linear models. *Journal of the Royal Statistical Society Series A* **135**(3): 370-384.
- Puteh S.E.W., Kamarudin N.A., Hussein Z., Adam N. & Shahari M.R. 2024. Cost of cardiovascular disease events in patients with and without type 2 diabetes and factors influencing cost: A retrospective cohort study. *BMC Public Health* **24**(1): 2003.
- Rittiphairoj T., Bulstra C., Ruampatana C., Stavridou M., Grewal S., Reddy C.L. & Atun R. 2025. The economic burden of ischaemic heart diseases on health systems: A systematic review. *BMJ Glob Health* **10**(2): e015043.
- Safiri S., Karamzad N., Singh K., Carson-Chahhoud K., Adams C., Nejadghaderi S.A., Almasi-Hashiani A., Sullman M.J.M., Mansournia M.A., Bragazzi N.L., Kaufman J.S., Collins G.S. & Kolahi A.A. 2022. Burden of ischemic heart disease and its attributable risk factors in 204 countries and territories, 1990-2019. *European Journal of Preventive Cardiology* **29**(2): 420-431.
- Salive M.E. 2013. Multimorbidity in older adults. *Epidemiologic Reviews* **35**(1): 75-83.

- Severino P., D'amato A., Pucci M., Infusino F., Birtolo L.I., Mariani M.V., Lavalle C., Maestrini V., Mancone M. & Fedele F. 2020. Ischemic heart disease and heart failure: role of coronary ion channels. *International Journal of Molecular Sciences* **21**(9): 3167.
- Shakya S., Shrestha A., Robinson S., Randall S., Mnataganian G., Brown H., Boyd J., Xu D., Lee C.M.Y., Brumby S., Peeters A., Lucas J., Gauci S., Huxley R., O'neil A. & Gao L. 2025. Global comparison of the economic costs of coronary heart disease: a systematic review and meta-analysis. *BMJ Open* **15**(1): e084917.
- Shi H., Xia Y., Cheng Y., Liang P., Cheng M., Zhang B., Liang Z., Wang Y. & Xie W. 2024. Global burden of ischaemic heart disease from 2022 to 2050: projections of incidence, prevalence, deaths, and disability-adjusted life years. *European Heart Journal-Quality of Care and Clinical Outcomes* **11**(4):355-366.
- Takroni M., Ogbeivor C., Mirza N., Al-Subaie M. & Al-Zahrani M. 2023. The impact of common comorbidities on the cardiopulmonary rehabilitation outcomes and cost of care for post-CABG surgery patients. *Cardiology and Cardiovascular Research* **7**(4): 75-81.
- Taloba A.I., Abd El-Aziz R.M., Alshanbari H.M. & El-Bagoury A.H. 2022. Estimation and prediction of hospitalization and medical care costs using regression in machine learning. *Journal of Healthcare Engineering* **2022**(1): 7969220.
- Venkitachalam L., Wang K., Porath A., Corbalan R., Hirsch A.T., Cohen D.J., Smith S.C., Ohman E.M., Steg P.G., Bhatt D.L. & Magnuson E.A. 2012. Global variation in the prevalence of elevated cholesterol in outpatients with established vascular disease or 3 cardiovascular risk factors according to national indices of economic development and health system performance. *Circulation* **125**(15):1858-1869.
- World Health Organization. 2025. Cardiovascular diseases (CVDs). <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-cvds> (5 January 2026).
- World Health Organization 2024. The top 10 causes of death. <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> (5 January 2026).
- Xie P., Li X., Guo F., Zhang D. & Zhang H. 2023. Direct medical costs of ischemic heart disease in urban Southern China: a 5-year retrospective analysis of an all-payer health claims database in Guangzhou City. *Frontiers in Public Health* **11**:1146914.
- Xu C., Liu R., Wang J. & Nicholas S. 2024. Hospitalization expenses of coronary heart disease inpatients in China: evidence from two hospitals in Ningxia Hui autonomous region. *Frontiers in Public Health* **12**:1266456.
- Yang G.-R., Li D. & Li L. 2020. Comparison of coronary heart disease and stroke in association with diabetic retinopathy in adults with diabetes using a national survey. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy* **13**:5079-5084.
- Zhou J., Williams C., Keng M.J., Wu R. & Mihaylova B. 2024. Estimating costs associated with disease model states using generalized linear models: A tutorial. *PharmacoEconomics* **42**(3):261-273.

Department of Mathematical Sciences
Faculty of Science and Technology
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor, MALAYSIA
E-mail: nm@ukm.edu.my*, a196023@siswa.ukm.edu.my

Department of Public Health Medicine
Faculty of Medicine
Universiti Kebangsaan Malaysia
56000 Cheras
Kuala Lumpur, MALAYSIA
E-mail: azimatunnoor@hctm.ukm.edu.my

Received: 8 April 2026
Accepted: 26 August 2026

*Corresponding author