Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Predictive Model for the Average Cost per Treatment Episode of Type 2 Diabetes Using Multivariate Regression Based on Health Insurance Database in Hanoi, Vietnam

Linh Viet Nguyen
Hanoi University of Pharmacy
Ngoc Yen Nhi Vo
Hong Bang International University
Cao Duc Huy Nguyen
University of Science, VNUHCM
Thi Thanh Huong Nguyen
Hanoi University of Pharmacy
Thi Thu Thuy Nguyen
thuyntt1@hiu.vn

Hong Bang International University

Research Article

Keywords: Type 2 diabetes mellitus, Social Health Insurance, direct medical cost

Posted Date: January 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8454323/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background: Type 2 diabetes mellitus (T2DM) is a chronic condition requiring lifelong treatment; therefore, the development of a multivariate regression model to support cost prediction is essential.

Objective: To develop a multivariate regression model for predicting the average treatment episode cost among patients with T2DM in Hanoi, Vietnam.

Methods: A cross-sectional descriptive study was conducted using retrospective data from the Hanoi Social Health Insurance database, covering T2DM patients from January 1, 2018, to December 31, 2022.

Results: The study included 1,397,287 patients across 6,940,296 treatment episodes. The male-to-female ratio ranged from 1:1.07 to 1:1.08, with a mean age ranging from 64.59 ± 11.16 to 64.95 ± 11.78 years. The average direct medical cost per treatment episode showed an increasing trend over the years, from 115.93 USD (95% CI: 115.33 – 116.54 USD) in 2018 to 198.10 USD (95% CI: 177.40 – 179.74 USD) in 2022. The most significant increase was observed during 2019–2020 (22.44%). Medication costs accounted for the largest proportion of the total cost and showed a decreasing trend, from 57.92% to 52.01%. Notably, 90% of the treatment cost was reimbursed by health insurance.

Conclusion: The average direct medical cost per treatment episode accounted for approximately 8.54% to 14.61% of the average monthly GDP per capita in Vietnam in 2024. The developed model provides a valuable tool for policymakers to forecast costs and implement cost-saving strategies.

Introduction

Diabetes mellitus (DM), particularly T2DM, is a major global health issue. Worldwide statistics indicate prevalence rates of approximately 5% among individuals aged 35–39, 10% in those aged 45–49, 15% among 55–59-year-olds, and nearly 20% in the 65–69 age group. DM is among the leading causes of death globally and can reduce life expectancy by up to 30% [1]. In Vietnam, T2DM is projected to become one of the top seven leading causes of mortality by 2030 [2]. As a chronic condition requiring continuous management, including daily medication and frequent monitoring, diabetes often results in hospitalizations and complications. Poor households may spend up to 47.12% of their total income on diabetes-related treatment [3]. Given the substantial economic burden posed by the disease, analyzing healthcare costs and developing cost prediction models are critically important.

In Vietnam, the Social Health Insurance (SHI) system plays a central role in financing diabetes care and mitigating out-of-pocket expenditures. SHI beneficiaries are classified into five main entitlement groups: (1) household-contributed enrollees; (2) individuals fully subsidized by the state; (3) partially subsidized groups such as near-poor households; (4) employees and employers with joint contributions; and (5) individuals covered by the Social Insurance Fund, such as retirees. Reimbursement rates under SHI vary by group and are commonly set at 80%, 95%, or 100%. Insured individuals are also registered under one of seven geographic or administrative regions (Region 1 through Region 7), each with distinct

reimbursement and administrative policies. Additionally, Vietnam's healthcare system is stratified by facility level, including commune-level health stations, district-level facilities, provincial hospitals, central-level hospitals, and institutional units such as military hospitals. The cost of treatment may differ significantly depending on the level of care. Moreover, hospital admissions are categorized into several types that impact SHI reimbursement: referred admissions, emergency admissions, out-of-line referral admissions (bypassing referral protocols), and cross-level accessible admissions, which allow patients to access higher-level facilities without referral under specific eligibility conditions. These structural features of the SHI system are important determinants of direct medical costs and should be incorporated into predictive models.

While several studies on the direct medical cost (DMC) of diabetes have been conducted in Vietnam, those focusing on Hanoi have mostly been limited to single-hospital settings with relatively small sample sizes. Meanwhile, the national Social Health Insurance (SHI) database represents a large and rich data source. However, to date, it has only been utilized in a single study in 2017 [4], indicating a gap in more extensive and in-depth research in this region. Therefore, this study titled “Multivariate Regression Model of Average Treatment Episode Cost for Type 2 Diabetes in Hanoi” was conducted with two primary objectives: (1) To analyze the direct medical cost per treatment episode for type 2 diabetes in Hanoi, Vietnam during the period 2018–2022; and (2) To develop a multivariate regression model for predicting direct medical costs in the treatment of type 2 diabetes in Hanoi, Vietnam.

Material and methods

2.1. Study setting and design

A cross-sectional descriptive study was conducted based on retrospective data obtained from the Hanoi Social Health Insurance database for the period 2018–2022. The processed data were incorporated into a generalized linear regression model to forecast the average DMC per treatment episode for patients with T2DM. Data collection took place from December, 2021 to March, 2023.

2.2. Sample size and sampling type

All patients who met the inclusion criteria and did not violate any exclusion criteria during the data extraction period.

2.3. Inclusion criteria and exclusion criteria

Inclusion criteria

- Patients diagnosed with T2DM (ICD-10 code: E11).

- Patients aged ≥18 years.

- Patients enrolling in and utilizing SHI for treatment.

Exclusion criteria

- Patients with insufficient data for analysis (demographics or cost data).

2.4. Study variables

The study collected data on patient characteristics and treatment costs, with the following key variables:

- Patient characteristics including demographic profile (number of patients, number of treatment episodes), sociodemographic characteristics (gender, age group, age), SHI-related characteristics (SHI entitlement group, reimbursement rate, region of SHI registration), clinical characteristics (comorbidities, types of comorbidities, complications, number and types of complications, complication scores), treatment characteristics (level of healthcare facility, reason for hospitalization, type of visit, average number of inpatient days per episode).

- Cost analysis including average DMC per treatment episode, categorized by cost components and payer source.

- Regression modeling: all variables related to demographics, SHI, clinical, and treatment characteristics as well as the summarized outcomes from cost analysis (Objective 1) were used as independent variables in the predictive model for Objective 2.

2.5. Statistical analysis

The DMCs are converted to USD in 2024 by using the CCEMG - EPPI-Centre Cost Converter tool.

Data were processed and analyzed using SQL, R, and Python programming languages. Statistical analyses were conducted with a 95% confidence level. Correlations between variables were assessed using appropriate statistical hypothesis tests (H_0 and H_1).

DMCs were calculated for T2DM patients utilizing SHI-covered healthcare services between January 1, 2018, and December 31, 2022, based on data from the Hanoi Social Health Insurance database. The regression analysis employed a Generalized Linear Model (GLM) with a Gamma distribution to estimate the direct medical costs.

Results

3.1. Analysis of direct medical cost per treatment episode for type 2 diabetes patient in Hanoi during the period 2018–2022

Patient characteristics

The demographic characteristics and SHI participation status of the study population are presented in Table 1.

Table 1. The demographic characteristics and social health insurance participation status (n=1,396,287)

Variables	Category	2018	2019	2020	2021	2022
Population characteristics						
Number of patients		239,863	252,108	270,294	291,876	343,146
Number of treatment episodes		1,057,122	1,432,415	1,409,538	1,323,610	1,717,611
Demographic characteristics						
Gender	Male	115,111 (47.99%)	121,319 (48.12%)	129,975 (48.09%)	141,038 (48.32%)	165,577 (48.25%)
	Female	124,752 (52.01%)	130,789 (51.88%)	140,319 (51.91%)	150,838 (51.68%)	177,569 (51.75%)
Age group	>=60	168,934 (70.43%)	167,501 (66.44%)	185,326 (68.56%)	204,348 (70.01%)	246,493 (71.83%)
	40-59	64,840 (27.03%)	75,026 (29.76%)	75,160 (27.81%)	78,439 (26.87%)	86,446 (25.19%)
	18-39	6,089 (2.54%)	9,581 (3.80%)	9,808 (3.63%)	9,089 (3.11%)	10,207 (2.97%)
Age	Mean ± SD	64.59 ± 11.16	63.58 ± 12.01	63.99 ± 12.00	64.51 ± 11.75	64.95 ± 11.78
SHI participation status						
HI Reimbursement rate	80%	76,480 (31.88%)	100,677 (39.93%)	109,090 (40.36%)	125,497 (43.00%)	150,108 (43.74%)
	95%	54,578 (22.75%)	49,300 (19.56%)	53,269 (19.71%)	54,346 (18.62%)	65,092 (18.97%)
	100%	108,805 (45.36%)	102,131 (40.51%)	107,935 (39.93%)	112,033 (38.38%)	127,946 (37.29%)
HI entitlement group	Group 1	56,254 (23.45%)	68,635 (27.22%)	77,729 (28.76%)	94,344 (32.32%)	112,631 (32.82%)
	Group 2	42,423 (17.69%)	55,431 (21.99%)	57,696 (21.35%)	61,124 (20.94%)	71,019 (20.70%)
	Group 3	3,656 (1.52%)	5,181 (2.06%)	5,623 (2.08%)	4,424 (1.52%)	5,847 (1.70%)

	Group 4	11,496 (4.79%)	16,762 (6.65%)	17,504 (6.48%)	17,064 (5.85%)	20,559 (5.99%)
	Group 5	126,034 (52.54%)	106,099 (42.08%)	111,742 (41.34%)	114,920 (39.37%)	133,090 (38.79%)
Region / Area of HI registration	Region 1	1,164 (0.49%)	2,051 (0.81%)	2,157 (0.80%)	1,848 (0.63%)	3,555 (1.04%)
	Region 2	3,494 (1.46%)	5,013 (1.99%)	5,608 (2.07%)	5,692 (1.95%)	4,712 (1.37%)
	Region 3	20 (0.01%)	37 (0.01%)	51 (0.02%)	45 (0.02%)	89 (0.03%)
	Region 4	231,222 (96.4%)	239,230 (94.89%)	256,472 (94.89%)	278,904 (95.56%)	326,330 (95.1%)
	Region 5	201 (0.08%)	366 (0.15%)	450 (0.17%)	450 (0.15%)	738 (0.22%)
	Region 6	55 (0.02%)	106 (0.04%)	114 (0.04%)	111 (0.04%)	227 (0.07%)
	Region 7	3,707 (1.55%)	5,305 (2.10%)	5,442 (2.01%)	4,826 (1.65%)	7,495 (2.18%)

Group 1- household-contributed, Group 2 – State budget fully subsidized, Group 3 – State budget partially subsidized, Group 4 – Employer and employee-contributed, Group 5 – Social Insurance Fund-contributed, HI – health insurance, Region 1 – North Central and Central Coastal Region, Region 2 – Ministry of National Defense and Government Agencies, Region 3 – Mekong River Delta, Region 4: Red River Delta, Region 5: Southeast Region, Region 6 – Central Highlands, Region 7 – Northern Midland and Mountainous Region, SD – standard deviation.

According to Table 1, the number of patients increased from 239,863 in 2018 to 343,146 in 2022, with females consistently accounting for a higher proportion (~52%). The ≥60 age group represented the highest proportion (over 66%) and showed an increasing trend over the study period. The average age across the entire sample ranged from 64.59 ± 11.16 to 64.95 ± 11.78 years. The proportion of patients entitled to full (100%) reimbursement slightly declined over time, while the proportion receiving 80% reimbursement increased markedly, from 31.88% to 43.74%. The group covered by contributions from the Social Insurance Fund (Group 5) accounted for the largest share of insured individuals but showed a

decreasing trend—from 52.54% in 2018 to 38.79% in 2022. Most patients were registered for SHI coverage in the Red River Delta region (>94%).

The clinical and treatment characteristics of the study population are presented in Table 2.

Table 2. The clinical and treatment characteristics (n=1,396,287)

Variables	Category	2018	2019	2020	2021	2022
The clinical characteristics						
Comorbidities	No	29,329 (12.23%)	7,200 (2.86%)	4,191 (1.55%)	3,393 (1.16%)	4,233 (1.23%)
	Yes	210,534 (87.77%)	244,908 (97.14%)	266,103 (98.45%)	288,483 (98.84%)	338,913 (98.77%)
Type of Comorbidities	Other	57,606 (24.02%)	35,161 (13.95%)	32,810 (12.14%)	31,016 (10.63%)	43,531 (12.69%)
	Hyperlipidemia	39,377 (16.42%)	50,651 (20.09%)	55,679 (20.6%)	58,228 (19.95%)	69,610 (20.29%)
	Hyperlipidemia + Hypertension	104,031 (43.37%)	135,009 (53.55%)	150,075 (55.52%)	167,401 (57.35%)	186,343 (54.3%)
	Hypertension	38,849 (16.2%)	31,287 (12.41%)	31,730 (11.74%)	35,231 (12.07%)	43,662 (12.72%)
Complications	No	116,191 (48.44%)	88,571 (35.13%)	86,799 (32.11%)	91,688 (31.41%)	117,083 (34.12%)
	Yes	123,672 (51.56%)	163,537 (64.87%)	183,495 (67.89%)	200,188 (68.59%)	226,063 (65.88%)
Type of complications	MC	439 (0.18%)	582 (0.23%)	632 (0.23%)	727 (0.25%)	855 (0.25%)
	No	116,191 (48.44%)	88,571 (35.13%)	86,799 (32.11%)	91,688 (31.41%)	117,083 (34.12%)
	MacroVC	59,483 (24.8%)	76,124 (30.19%)	87,404 (32.34%)	96,879 (33.19%)	111,062 (32.37%)
	MacroVC+ MC	729 (0.3%)	969 (0.38%)	1,137 (0.42%)	1,265 (0.43%)	1,278 (0.37%)
	MacroVC + MicroVC	40,303 (16.8%)	56,242 (22.31%)	61,897 (22.9%)	67,804 (23.23%)	73,611 (21.45%)
	MacroVC + MicroVC + MC	1,494 (0.62%)	2,022 (0.8%)	2,153 (0.8%)	2,314 (0.79%)	2,481 (0.72%)
	MicroVC	20,775 (8.66%)	26,974 (10.7%)	29,574 (10.94%)	30,424 (10.42%)	35,909 (10.46%)
	MicroVC + MC	449 (0.19%)	624 (0.25%)	698 (0.26%)	775 (0.27%)	867 (0.25%)
Number of complications	0	57,606 (24.02%)	35,161 (13.95%)	32,810 (12.14%)	31,016 (10.63%)	43,531 (12.69%)

	1	39,377 (16.42%)	50,651 (20.09%)	55,679 (20.6%)	58,228 (19.95%)	69,610 (20.29%)
	2	104,031 (43.37%)	135,009 (53.55%)	150,075 (55.52%)	167,401 (57.35%)	186,343 (54.30%)
	≥3	38,849 (16.2%)	31,287 (12.41%)	31,730 (11.74%)	35,231 (12.07%)	43,662 (12.72%)
The treatment characteristics						
Level of healthcare facility	District-level facility or equivalent	237,792 (22.49%)	319,168 (22.28%)	318,303 (22.58%)	321,226 (24.27%)	444,065 (25.85%)
	Provincial-level facility or equivalent	106,167 (10.04%)	144,865 (10.11%)	148,747 (10.55%)	141,304 (10.68%)	206,135 (12%)
	Central-level hospital	704,660 (66.66%)	941,885 (65.76%)	900,723 (63.9%)	801,753 (60.57%)	996,409 (58.01%)
	Commune-level or equivalent	7,934 (0.75%)	25,965 (1.81%)	41,375 (2.94%)	59,042 (4.46%)	70,737 (4.12%)
	Institutional healthcare unit	569 (0.05%)	532 (0.04%)	390 (0.03%)	285 (0.02%)	265 (0.02%)
Reason for hospital admission	Referred admission	1,364,235 (95.24%)	1,333,928 (94.64%)	1,205,412 (91.07%)	1,564,675 (91.1%)	1,364,235 (95.24%)
	Emergency	11,117 (0.78%)	10,081 (0.72%)	9,580 (0.72%)	12,631 (0.74%)	11,117 (0.78%)
	Out-of-line referral admission	9,839 (0.69%)	9,856 (0.7%)	15,491 (1.17%)	30,952 (1.8%)	9,839 (0.69%)
	Cross-level accessible admission	47,224 (3.3%)	55,673 (3.95%)	93,127 (7.04%)	109,353 (6.37%)	47,224 (3.3%)
Type of medical service received	Medical consultation	839,710 (79.43%)	1,171,900 (81.81%)	1,225,030 (86.91%)	1,148,647 (86.78%)	1,312,665 (76.42%)
	Outpatient	174,779 (16.53%)	190,186 (13.28%)	111,087 (7.88%)	102,528 (7.75%)	299,897 (17.46%)
	Inpatient	42,633 (4.03%)	70,329 (4.91%)	73,421 (5.21%)	69,135 (5.22%)	103,297 (6.01%)
	Short-stay inpatient (< 4 hours)	0 (0%)	0 (0%)	0 (0%)	3,300 (0.25%)	1,738 (0.1%)
	Scheduled medication	0 (0%)	0 (0%)	0 (0%)	0 (0%)	14 (0%)

	pick-up					
Number of inpatient days	Mean ± SD	10.94 ±6.56	10.25 ±6.84	10.33 ±6.95	10.13 ±7.31	9.93 ±7.05
Number of treatment episodes	Mean ± SD	4.41± 3.51	5.68± 4.89	5.21± 4.27	4.53± 3.64	5.01± 4.27

Note: MacroVC – Macrovascular complications, MC – Metabolic, MicroVC – Microvascular complications, SD – standard deviation.

According to Table 2, the study recorded a very high and increasing proportion of patients with comorbidities over time (from 87.77% to 98.84%). Among these, the group with coexisting dyslipidemia and hypertension was predominant (>50%). The proportion of patients with complications ranged from 51.56% in 2018 to 68.59% in 2021, with macrovascular complications being the most common. The majority of patients had two complications (ranging from 43.37% to 57.35%). Regarding treatment facility level, most patients were treated at central-level hospitals (>58%), though this showed a slight downward trend. The majority of admissions followed appropriate referral pathways (>91%), and outpatient visits accounted for the highest proportion of medical services (>76%). The average number of inpatient days slightly decreased from 10.94 ± 6.56 days to 9.93 ± 7.05 days, while the number of treatment episodes per year ranged from 4.41 ± 3.51 to 5.68 ± 4.89 episodes.

The analysis of the average DMC per treatment episode by cost components is presented in Table 3**.**

Table 3. Average DMC per treatment episode for T2DM patients, categorized by cost components (unit: USD)

	2018	2019	2020	2021	2022
Mean (95% CI)					
Medications	67.15 (66.88-67.43)	73.50 (73.18-73.83)	87.66 (87.24-88.08)	99.33 (98.86-99.81)	92.88 (92.42-93.33)
Laboratory tests	20.24 (20.14-20.14)	20.19 (20.09 -20.29)	21.91 (21.79-22.03)	22.94 (22.79-23.08)	22.66 (22.53-22.79)
Medical consultation	4.92 (4.92-4.92)	4.94 (4.94-4.94)	4.97 (4.96-4.97)	4.77 (4.77-4.77)	4.57 (4.56-4.57)
Hospital bed	12.46 (12.32-12.61)	15.54 (15.39-15.70)	16.54 (16.38-16.69)	15.60 (15.44-15.76)	16.30 (16.15-16.44)
Imaging diagnostics	5.23 (5.17-5.28)	6.13 (6.07-6.18)	7.88 (7.81-7.95)	8.97 (8.89-9.05)	9.63 (9.56-9.71)
Procedures and surgeries	3.65 (3.55-3.75)	6.01 (5.88-6.13)	8.01 (7.86-8.17)	9.66 (9.49-9.84)	12.04 (11.86-12.22)
Medical supplies	1.63 (1.46-1.80)	4.71 (4.71-5.05)	10.78 (10.20-11.36)	14.65 (13.95-15.35)	18.43 (17.78-19.09)
Other*	0.65 (0.57-0.73)	1.23 (1.14-1.31)	1.84 (1.71-1.97)	2.27 (2.14-2.40)	2.06 (1.96-2.17)
Total	115.93 (115.33-116.54)	132.25 (131.47-133.03)	159.58 (158.51-160.65)	178.19 (176.94-179.45)	178.57 (177.40-179.74)

*Note: CI – Confidence interval; *Other includes costs of technical services, blood products, and transportation.*

The study recorded an increasing trend in average DMC per treatment episode, rising from 115.93 USD (95% CI: 115.33 USD – 116.54 USD) in 2018 to 178.57 USD (95% CI: 177.40 USD – 179.74 USD) in 2022, with the largest increase observed during 2019–2020 (+22.44%). Among the cost components, medications accounted for the highest share, showing a cumulative increase of 58.01% over the first four years, followed by a slight decrease of 2.89% in the final year.

The cost structure of average DMC for T2DM patients by cost components is presented in Table 4.

Table 4. Cost structure of average DMC for T2DM patients by cost components

Component	2018	2019	2020	2021	2022
Medications	57.92%	55.58%	54.93%	55.74%	52.01%
Laboratory tests	17.45%	15.27%	13.73%	12.87%	12.69%
Medical consultation	4.24%	3.74%	3.11%	2.68%	2.56%
Hospital bed	10.75%	11.75%	10.36%	8.75%	9.13%
Imaging diagnostics	4.51%	4.63%	4.94%	5.03%	5.39%
Procedures and surgeries	3.15%	4.54%	5.02%	5.42%	6.74%
Medical supplies	1.41%	3.56%	6.75%	8.22%	10.32%
Other*	0.56%	0.93%	1.15%	1.28%	1.16%
Total	100.00%	100.00%	100.00%	100.00%	100.00%

**Other includes costs of technical services, blood products, and transportation.*

The DMC of T2DM treatment by source of payment is presented in Table 5.

Table 5. Average DMC per treatment episode for T2DM patients by source of payment (unit: USD)

	2018	2019	2020	2021	2022
		Mean (95% CI)			
Paid by HI	106.52 (105.96-107.08)	118.97 (118.29-119.66)	142.13 (141.22-143.05)	159.33 (158.26-160.39)	157.63 (156.66-158.61)
Co-payment by patient	6.23 (6.18-6.29)	8.42 (8.33-8.51)	10.25 (10.14-10.36)	12.00 (11.86-12.13)	11.88 (11.76-12.00)
Out-of-pocket payment by patient	3.18 (3.09-3.27)	4.85 (4.70-5.00)	7.20 (6.99-7.41)	6.87 (6.61-7.13)	9.06 (8.78-9.33)
Total	**115.93 (115.33-116.54)**	**132.25 (131.47-133.03)**	**159.58 (158.51-160.65)**	**178.19 (176.94-179.45)**	**178.57 (177.40-179.74)**

The study found that health insurance payments accounted for the largest share of total direct medical costs, although this proportion slightly declined over time, from 91.88% in 2018 to 88.28% in 2022. Despite this decline, the absolute value of insurance-covered payments increased, from 106.52 USD (95% CI: 105.96 USD – 107.08 USD) in 2018 to 157.63 USD (95% CI: 156.66 USD – 158.61 USD) in 2022.Meanwhile, co-payment and out-of-pocket expenses by patients also increased in both value and proportion during the study period.

The cost structure of average DMC per treatment episode for T2DM patients by source of payment is presented in Table 6.

Table 6. Cost structure of average DMC per treatment episode for T2DM patient, by source of payment

Source of payment	2018	2019	2020	2021	2022
Paid by HI	91.88%	89.96%	89.07%	89.07%	88.28%
Co-payment by patient	5.38%	6.37%	6.42%	6.73%	6.65%
Out-of-pocket payment by patient	2.74%	3.67%	4.51%	3.86%	5.07%
Total	**100.00%**	**100.00%**	**100.00%**	**100.00%**	**100.00%**

3.2. Development of a multivariate regression model to predict direct medical costs for type 2 diabetes treatment in Hanoi

The study examined the normality assumption of the cost variable and found that cost data exhibited a right-skewed distribution, violating the assumption of normality. Therefore, the cost variable was log-transformed, resulting in a distribution that approximates normality. The processed dataset comprised 6,940,296 observations, which were randomly split into two subsets: 80% for the training set and 20% for the testing set.

The model performance metrics are presented in Table 7.

Table 7. Model performance metrics for the cost prediction model

	R^2 deviance	RMSE	Min	Max
Training set	0.6754	439.43	0.02	85,616.59
Test set	0.6743	423.02	0.01	62,466.92

Note: RMSE – Root mean square errors

Results from the multivariable regression analysis indicated good model performance, with R^2 values of 0.6754 for the training set and 0.6743 for the testing set. The corresponding root mean square errors (RMSE) were 439.43 USD and 423.02 USD, respectively.

The multivariate regression results are presented in Table 8.

Table 8. Multivariate regression model for direct medical cost per treatment episode

Variable name	Estimated coefficient	Standard error	Test statistic	p-value
Intercept	-0.02	0.515	-346.1	< 0.001
Year of treatment	0.095	0.000	372.88	< 0.001
Age	0.001	0.000	20.26	< 0.001
Number of inpatient days	0.072	0.000	286.78	< 0.001
Gender: Female	-0.044	0.001	-60.70	< 0.001
Age group: 40-59	0.012	0.003	3.94	< 0.001
Age group: ≥60	0.010	0.003	3.05	0.002
Reason for admission: Cross-level accessible	0.323	0.004	71.95	< 0.001
Reason for admission: Out-of-line referral	-0.017	0.004	-4.44	< 0.001
Reason for admission: Emergency	-0.149	0.002	-82.89	< 0.001
Type of medical service: Outpatient	0.050	0.001	44.33	< 0.001
Type of medical service: Inpatient	1.663	0.003	579.33	< 0.001
Insurance registration region: Southeast	0.091	0.012	7.59	< 0.001
Insurance registration region: North Central and South Central Coast	0.402	0.005	75.98	< 0.001
Insurance registration region: Mekong River Delta	0.293	0.035	8.39	< 0.001
Insurance registration region: Northern Midlands and Mountains	0.260	0.003	87.62	< 0.001
Insurance registration region: Central Highlands	0.240	0.024	10.16	< 0.001
Insurance registration region: Ministry of Defense/Government Agencies	0.246	0.003	80.32	< 0.001
Insurance eligibility: Fully subsidized by the state budget	0.069	0.001	61.39	< 0.001
Insurance eligibility: Partially subsidized by the state budget	0.110	0.003	39.01	< 0.001
Insurance eligibility: Employee and employer contributions	0.013	0.002	6.89	< 0.001
Insurance eligibility: Paid by Social Insurance Fund	0.099	0.001	93.54	< 0.001
Level of healthcare facility: Central	-0.838	0.001	-654.06	< 0.001

Level of healthcare facility: District	-1.363	0.001	-965.65	< 0.001
Level of healthcare facility: Commune	-2.142	0.002	-866.88	< 0.001
Level of healthcare facility: Institutional facility	-2.092	0.021	-101.79	< 0.001
Comorbidity: Yes	0.110	0.003	37.64	< 0.001
Type of complication: MC	0.086	0.007	11.84	< 0.001
Type of complication: MacroVC	0.026	0.001	29.51	< 0.001
Type of complication: MacroVC + MicroVC	0.068	0.001	74.31	< 0.001
Type of complication: MacroVC + MC	0.077	0.005	15.07	< 0.001
Type of complication: MicroVC + MC	-0.016	0.006	-2.54	0.0111
Type of complication: MacroVC + MicroVC + MC	0.039	0.003	11.54	< 0.001

Note: MacroVC – Macrovascular complications, MC – Metabolic, MicroVC – Microvascular complications

Variables associated with increased treatment costs included: year of treatment, age, age groups 40–59 and ≥60, number of inpatient days, type of care (outpatient and inpatient), reason for admission (cross-level accessible admission), insurance registration region (Southeast, North Central and South Central Coast, Mekong River Delta, Northern Midlands and Mountains, Central Highlands, and the Ministry of Defense/Government-affiliated group), insurance eligibility group (state budget or social insurance fund coverage), presence of comorbidities, and the presence of complications—particularly combinations such as MacroVC + MicroVC and MacroVC + MC. Among these, the factor with the strongest positive association was “type of care: inpatient”, with an estimated coefficient of 1.663, corresponding to an increase in cost by approximately 5.27 times compared to the reference group (medical consultation), calculated as exp (1.663) ≈ 5.27.

In contrast, variables associated with lower treatment costs included: female gender, reason for admission (out-of-line referral or emergency), treatment facility level (central, district, commune, and institutional facilities), and certain rare complication combinations such as MicroVC + MC. Among these, the factor with the strongest negative association was “treatment facility level: commune”, with an estimated coefficient of –2.142, indicating that treatment cost at commune-level facilities was only about 11.8% of that at provincial-level facilities (exp (–2.142) ≈ 0.118).

Discussion

This study aimed to estimate the average DMC per treatment episode for patients with T2DM in Hanoi from 2018 to 2022, utilizing real-world data from Hanoi’s SHI database. A total of 1,092,178 patients and

over 6 million treatment episodes were analyzed. The average DMC per episode ranged from USD 90.13 to USD 101.76, with pharmaceutical costs accounting for approximately 72.77% of total expenditure. Notably, over 85% of total costs were reimbursed by the national health insurance scheme. A GLM with a Gamma distribution was applied to identify key cost drivers. Results showed that inpatient care, longer length of stay, higher-level hospital treatment, and the presence of vascular complications were significantly associated with increased costs. In contrast, outpatient care, treatment at primary healthcare facilities, and the absence of complications were linked to lower costs.

This study analyzed DMC of 1,397,287 patients with T2DM across 6,940,296 treatment episodes in Hanoi from 2018 to 2022. The average age of patients ranged from 64.59 ± 11.16 to 64.95 ± 11.78 years, and the male-to-female ratio ranged from 1:1.07 to 1:1.08. These findings are consistent with a study by Tran Hoang Kim et al. (2024), which reported a mean age of 65.3 ± 9.8 years and a male-to-female ratio of 1:1.2 [5]. The average DMC per treatment episode in Hanoi showed an upward trend over time, increasing from 115.93 USD (95% CI: 115.33–116.54 USD) in 2018 to 178.57 USD (95% CI: 177.40–179.74 USD) in 2022. Medication costs accounted for the highest proportion of expenses (ranging from 52.01% to 57.92%) and showed a marked increase between 2018 and 2021, with a peak growth rate of nearly 41% during 2019–2021, before slightly declining in 2022. This pattern aligns with the findings of Le Thi Thuy Hien et al. (2020), who reported an average DMC of 169.35±36.80 USD, with medications accounting for 78.04% of the total cost [6] and Nguyen Thi Thuy Trang et al. (2022), who found that medication represented 69.4% of treatment costs at District 8 Hospital in Ho Chi Minh City [7]. A significant portion of DMC was covered by SHI, ranging from 88.28% to 91.88% of total costs. This is in line with findings by Le Van Pho et al. (2022), which indicated that SHI reimbursed 91.94% of treatment costs [8].

A GLM using a Gamma distribution with a log link function was applied specifically for the Hanoi context to investigate cost-driving factors. The model demonstrated good fit (R^2 deviance ≈ 0.675), indicating a strong explanatory power for variations in treatment costs. The most influential predictors included the number of treatment days, inpatient status, level of care, and presence of complications. Notably, treatment at central-level hospitals was associated with lower costs compared to provincial-level care, while the presence of comorbidities significantly increased total costs. These findings highlight cost structure variations across levels of care and clinical characteristics and provide quantitative evidence to inform cost-containment policy development tailored to local contexts.

This study presents several notable strengths. It leveraged a large-scale, real-world dataset derived from the SHI system in Hanoi, encompassing more than one million patients over a five-year period. This enhances the generalizability and validity of the findings, capturing actual clinical practice rather than controlled trial settings. Additionally, the application of a GLM with a Gamma distribution was methodologically appropriate, as it accounts for the right-skewed nature of healthcare cost data. Unlike conventional descriptive approaches, the model not only quantified average costs but also identified statistically significant predictors of cost variation, including care settings, healthcare facility levels, and complication status

However, as a retrospective cross-sectional study limited to Hanoi, it cannot fully capture associations between costs and important clinical factors such as disease control level, complication severity, or individualized treatment regimens. It also may not reflect the national picture. Future studies should adopt a prospective design, incorporate more detailed clinical variables such as HbA1c levels, specific medication types, and target organ damage severity, and extend coverage to other regions to better identify key cost drivers and inform resource allocation policies that reflect the reality of care delivery.

The findings provide a practical tool for health insurers and hospital managers to forecast diabetes treatment costs and allocate budgets more effectively. The model supports evidence-based reimbursement decisions and enhances cost control efforts within Vietnam's SHI framework.

Conclusion

The average DMC per treatment episode for T2DM patients in Hanoi showed a modest upward trend during the 2018–2022 period. These costs accounted for approximately 8.54% to 14.61% of the average monthly GDP per capita in Vietnam in 2024. The developed predictive model provides a valuable tool for policymakers to forecast treatment expenses and support the implementation of cost-saving measures.

Abbreviations

T2DM – Type 2 diabetes mellitus

DM – Diabetes mellitus

SHI – Social Health Insurance

DMC – Direct medical cost

GLM – Generalized Linear Model

MacroVC – Macrovascular complications

MicroVC – Microvascular complications

MC – Metabolic

RMSE – Root mean square errors

Declarations

Acknowledgements

Not applicable.

Clinical trial number

Not applicable.

Funding

This research received no external funding.

Competing interest

The authors have no conflicts of interest to declare for this study.

Ethics approval and consent to participate

The collection of electronic payment data on diabetes treatment costs for the period 2018–2022 was authorized by the Boards of Directors of the Social Health Insurance Agencies in Hanoi, following the official procedures. The study was conducted in accordance with ethical principles of medical research. No personal identifiers such as patient names or identification codes were collected, and all research data were kept confidential. Due to the retrospective nature and the anonymity of the patient data, the requirement of individual informed consent was waived. The findings of this study are intended solely for the benefit of the community.

Consent for publication

Not applicable.

Availability of data and material

The datasets generated by or analyzed during the current study are available from the author, Thi Thu Thuy Nguyen (thuyntt1@hiu.vn), upon reasonable request.

Author contributions

LVN, NYNV, CDHN, TTHN, and TTTN conceptualized the manuscript. LVN wrote the original draft. NYNV, CDHN, TTHN, and TTTN supervised the process. NYNV, CDHN, TTHN, and TTTN reviewed and edited the manuscript. All authors read and approved the final manuscript.

References

1. Alberti, Kurt George Matthew Mayer, Paul Zimmet, and Johnathon Shaw. "International Diabetes Federation: a consensus on Type 2 diabetes prevention." Diabetic Medicine 24.5 (2007): 451-463. DOI: https://doi.org/10.1111/j.1464-5491.2007.02157.x.
2. Mathers, Colin D., and Dejan Loncar. "Projections of global mortality and burden of disease from 2002 to 2030." PLoS medicine 3.11 (2006): e442. DOI: https://doi.org/10.1371/journal.pmed.0030442.

3. Ngoc, Nguyen Bich, Zhou Lu Lin, and Waqas Ahmed. "Diabetes: What challenges lie ahead for Vietnam?" Annals of Global Health 86.1 (2020): 1. DOI: 10.5334/aogh.2526.
4. Tuan Kiet Pham, Huy et al. (2020), "Direct medical costs of diabetes and its complications in Vietnam: A national health insurance database study", *Diabetes Research and Clinical Practice*. 162, pp. 1-8. DOI: https://doi.org/10.1016/j.diabres.2020.108051.
5. Trần, H. K., Trần, H. K., Nguyễn, T. H., Nguyễn, T. T., Nguyễn, T. H., Nguyễn, H. U., Nguyễn, T. T. H., & Đỗ, N. K. (2024). "Thừa Cân Béo Phì Và Một Số Yếu Tố Liên Quan Của Người Bệnh Đái Tháo Đường Type 2 Tại Bệnh Viện 19-8." Tạp Chí Y học Việt Nam, 545(2). DOI: https://doi.org/10.51298/vmj.v545i2.12261
6. N. T. B. Y. Lê Thị Thuý Hiền, Dương Hồng Lý, Hồ Thị Kim Thanh, "Chi phí trực tiếp trong một tháng điều trị của người bệnh đái tháo đường typ 2 cao tuổi ngoại trú," Tạp chí nghiên cứu y học, vol. 133, 9, p. 6, 2020.
7. Nguyễn Thị Thùy Trang et al. (2022), "Phân tích chi phí trực tiếp y tế trong điều trị đái tháo đường type 2 tại Bệnh viện Quận 8 thành phố Hồ Chí Minh giai đoạn 2019-2021", *Tạp chí Y học Việt Nam*. 514(1). DOI: https://doi.org/10.51298/vmj.v514i1.2530.
8. Văn Phó, L., Thị Thu Hà, N., & Văn Trầm, T. (2022). "Chi Phí Điều Trị Trực Tiếp Cho Người Bệnh Đái Tháo Đường Ngoại Trú Tại Phòng Khám Nội Tiết Bệnh Viện Đa Khoa Trung Tâm Tiền Giang Năm 2021." Tạp Chí Y học Việt Nam, 508(2). DOI: https://doi.org/10.51298/vmj.v508i2.1618.