ORIGINAL RESEARCH

Type 2 diabetes in Vietnam: a cross-sectional, prevalence-based cost-of-illness study

Nguyen Tu Dang Le
Luyen Dinh Pham
Trung Quang Vo

Department of Pharmacy Administration, Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam

Background: According to the International Diabetes Federation, total global health care expenditures for diabetes tripled between 2003 and 2013 because of increases in the number of people with diabetes as well as in the average expenditures per patient. This study aims to provide accurate and timely information about the economic impacts of type 2 diabetes mellitus (T2DM) in Vietnam.

Method: The cost-of-illness estimates followed a prospective, prevalence-based approach from the societal perspective of T2DM with 392 selected diabetic patients who received treatment from a public hospital in Ho Chi Minh City, Vietnam, during the 2016 fiscal year.

Results: In this study, the annual cost per patient estimate was US $246.10 (95% CI 228.3, 267.2) for 392 patients, which accounted for about 12% (95% CI 11, 13) of the gross domestic product per capita in 2017. That includes US $127.30, US $34.40 and US $84.40 for direct medical costs, direct nonmedical expenditures, and indirect costs, respectively. The cost of pharmaceuticals accounted for the bulk of total expenditures in our study (27.5% of total costs and 53.2% of direct medical costs). A bootstrap analysis showed that female patients had a higher cost of treatment than men at US $48.90 (95% CI 3.1, 95.0); those who received insulin and oral antidiabetics (OAD) also had a statistically significant higher cost of treatment compared to those receiving OAD, US $445.90 (95% CI 181.2, 690.6). The Gradient Boosting Regression (Ensemble method) and Lasso Regression (Generalized Linear Models) were determined to be the best models to predict the cost of T2DM (R^2=65.3, mean square error [MSE]=0.94; and R^2=64.75, MSE=0.96, respectively).

Conclusion: The findings of this study serve as a reference for policy decision making in diabetes management as well as adjustment of costs for patients in order to reduce the economic impact of the disease.

Keywords: cost-of-illness, diabetes mellitus, hospital, Vietnam

Introduction

Diabetes mellitus is a rapid-onset, chronic, noncommunicable disease that causes many serious complications, such as a decrease in quality of life, and is a major cause of death in most countries.[1,2] In 1980, there were 108 million people living with diabetes, and this number had quadrupled by 2014.[2,3] The World Health Organization estimated in 2014 that there were 422 million people over the age of 18 (8.5%) with diabetes mellitus worldwide, putting the risk of death from diabetes 3.3 times higher than human immunodeficiency virus/acquired immune deficiency syndrome and tuberculosis and 8.3 times higher than malaria.[1,2,4] In 2015, 5 million people had died of diabetes or complications from diabetes worldwide, an alarming 14% increase from just 3 years

Correspondence: Trung Quang Vo
Department of Pharmacy Administration, Faculty of Pharmacy, University of Medicine and Pharmacy at Ho Chi Minh City, Vietnam
Email voquangtrungdk@gmail.com

earlier.[1,4] Based upon this rate, the International Diabetes Federation (IDF) has predicted that by 2040, one in 10 people will have diabetes, with one in 11 people contracting it during that time period.[1]

Diabetes mellitus is a chronic disease that occurs when the body is unable to produce enough insulin or cannot use insulin.[5] The diagnosis of diabetes mellitus is clinical and is based on multiple measurements of blood glucose; the condition is further classified into one of the three major types of diabetes, including type 1, type 2, and gestational diabetes.[6,7] The most common form of diabetes is type 2 diabetes mellitus (T2DM), with 91% in high-income countries being diagnosed with it.[8–11] Although manifestations of the disease are difficult to identify in their early stages, with many cultural and social changes, the indications of the disease are more conspicuous.[1,12–14] Individuals diagnosed with diabetes mellitus suffer not only from the dangerous health-related complications but also from significant financial burdens associated with the treatment and lifelong management of the disease progression.[15] Based on a cost estimate from the 2015 statistics with 5 million deaths from the disease, around 673–1197 billion US dollars (USD) were spent on medical expenses from treatment of illness to death and with 827 billion USD as the annual direct medical fees.[1,3,16]

According to the IDF, total global health care expenditures for diabetes have tripled between 2003 and 2013 as a result of an increase in the number of people with diabetes as well as the average health care expenditure for the treatment of diabetes.[17] This increase is expected to grow exponentially in the coming years, especially for those with low and middle incomes, and the cost of treatment for diabetes in low and middle countries will be more burdensome than high-income countries.[1] In the USA, the estimated total cost of treatment for diabetes in 2012 was $245 billion USD, including $176 billion USD for direct medical expenses and $69 billion USD for losses because of reduced productivity.[18] As Vietnam is a low-income developing country,[19] its incidence of diabetes is increasing at an alarming rate, nearly doubling in the past 10 years.[20] Pham estimated that at present, one in every 20 people is diagnosed with diabetes.[21] In addition, the number of people with prediabetes has also increased significantly to three times as many as for people with diabetes.[20] By the end of 2015, Vietnam had 63,021 cases of diabetes, representing 5.6% of people aged 20–79 years, of whom 53,457 died.[22] Pham also estimated that by 2035, the incidence of diabetes and prediabetes will increase to 7% and 15.7%, respectively.[21] In Vietnam, the average treatment cost per diabetic patient was $162.7 USD in 2015, which is more than the $150 USD

per capita monthly average.[22] However, no studies on the cost of treatment of diabetes in Vietnam are available.[15] This study endeavors to present treatment costs from a societal perspective and also provide accurate and timely information about the economic impacts of T2DM in Vietnam.

Methods
Study design
In this cross-sectional study, the cost-of-illness estimates follow the sum-all cost and prevalence-based approach from a societal perspective, with selected diabetic patients who received treatment from a public hospital in Ho Chi Minh City of Vietnam in the 2016 fiscal year.[23,24] The sum-all medical cost method involves adding all the medical costs of providing care to a patient, regardless of whether the cost was directly related to the disease or not.[25] Incorporation with a prevalence-based approach yields a general picture of the economic burden as well as component costs to provide better management reference for policymakers to manage the disease.[26]

Data collection
This was a cross-sectional study of T2DM patients who had received care in a public government hospital in 2016. The data are taken from the patient's medical records, which are located in the hospital's electronic database along with the demographic characteristics from the face-to-face interview. From this, the necessary data are synthesized to develop a model of treatment costs. Patients with T2DM are identified using the International Classification of Diseases Revision, tenth revision (ICD-10 code) with disease diagnosis code E11. Diabetes complications and comorbidities are also identified using ICD-10.

As the hospital examination period is once a month, a simple random sampling will be performed within a month (i.e., the number of patients in the day will be randomly selected with generating random numbers following a code line in Python random.choice([0, 1]), with the output equal 1 being selected and vice versa). This study included patients who qualified for all of the following two criteria such as 1) assigned ICD-10 code of E11; 2) attended treatment for T2DM for 1 year in this facility. Patients with type1 diabetes and women with gestational diabetes were excluded.

Estimation of costs
Cost-of-illness estimates in this study were incidence-based from a societal perspective, and hence considered all costs, irrespective of who incurred them. The costs consisted of direct costs and indirect costs.[27]

Direct costs

Direct costs include direct medical cost and direct nonmedical cost. In this study, direct medical costs are classified as medical care costs (visits and emergency care as well as procedures performed by medical personnel), diagnosis, pharmaceuticals, operation, and medical supplies. To standardize direct medical costs, the study uses reference unit costs of medical services in Vietnam from the study of Vo et al.[28] Because Vo et al's investigation was conducted in 2014, all unit costs will be changed to their 2017 values $\left(\dfrac{CPI_{2017}}{CPI_{2014}} = 1.075 \right)$.[29] Self-reported illness-related cost

items, such as meals (special food or diabetic supplements), transportation, accommodation, and informal care costs, were collected from patients and their family members and were considered direct nonmedical costs.

Indirect costs

The indirect costs associated with diabetes include workdays missed because of health conditions (absenteeism), reduced work productivity while working because of health conditions (presenteeism), reduced workforce participation due to disability, and productivity lost because of premature mortality.[30–32] To estimate the indirect costs, this study uses the human capital method,[26] as follows:

- Absenteeism: the number of workdays missed because of poor health.[33] The study recorded the total number of days off for routine medical visits as well as days off because of sickness based upon information provided by patients.
- Presenteeism: reduced productivity while at work,[33] which is generally measured through worker responses to surveys. However, their responses are not entirely accurate because workers, especially laborers, depend on daily productivity for their income. Consequently, the productivity loss associated with diabetes-attributed presenteeism was determined using an estimate (14 days per worker with diabetes per year).[34]
- Unemployment: reduced productivity for those not in the workforce, which was defined as the reduction in people not in the workforce in personal activities, includes time spent providing child care, household activities, and other tasks, such as volunteering in the community.[33] This population includes all adults under 65 years of age who are not employed. The earnings of patients were multiplied by a factor of 75% to account for those not in the workforce.[34]

Data analyses and statistical methods

Cost data are typically highly skewed because a few patients incur particularly high costs.[35,36] Such problems were handled using the nonparametric double bootstrap-t method to calculate the 95% CIs as well as the significant differences between means (with a 95% significance level).[37,38] Consequently, cost data are represented as an arithmetic mean (bootstrap 95% CI) cost to make that information available for decision makers.[36,39] All costs were measured in terms of Vietnamese currency (dong [₫]) and converted to USD using the exchange rate of $1 USD = ₫22,708.6.[29]

Microvascular complications include diabetic retinopathy, diabetic nephropathy, and diabetic neuropathy, while macrovascular complications include cardiovascular and coronary artery diseases.[40] The study excluded comorbidities with a prevalence of <5%. The Chi-square test was performed to compare statistically significant differences in the prevalence of complications and comorbidities among men and women. Statistically significant correlations between the prevalence of comorbidities and complications associated with patient age were assessed with the Spearman's rank correlation coefficient (ρ). Descriptive statistics (frequency, percentage, mean, median, 25th–75th interquartile) were used to summarize demographic information, clinical status, and component expenditures of the data.

A two-way sensitivity analysis was conducted to find the variance of total treatment costs when the two hypothetical values (cost lost because of "Presenteeism" and "Unemployment") changed ("Presenteeism" is between 5.5 and 33.5 days,[41] and the "Unemployment" rate of the average earnings is 65%–85%).

This study applied machine learning algorithms to determine the fittest model for prediction and also to identify the factors affecting the direct medical costs of T2DM. Those algorithms compared the precision using the R^2-score and the mean square error (MSE). Direct medical costs and continuous value features (age, blood glucose level, number of comorbidities, body mass index [BMI], and disease duration), which did not have a Gaussian distribution, were transformed using the Box–Cox method,[42,43] and then all continuous value features were scaled with a standardization method to convert the data from differences in measured units to the scaling factor (z-score).[44,45] This study labeled the category variables (gender, health insurance status, marriage status, education, occupation, monthly income, drinking, smoking, exercising, relatives' T2DM history, treatment, and type of complications) with a value between 0 and n-1 (Supplementary material).

All statistical analyses were performed using Python version 3.5 (Anaconda platform 4.4.0).[46,47]

Ethical consideration

Written informed consent was obtained from adult participants and their caregivers who were involved in the study, before the specimen collection and the questionnaire survey began. The research protocol was approved by the Biomedical Research Ethics Council at the District 9 Hospital in Ho Chi Minh City, Vietnam, and also the University of Medicine and Pharmacy in Ho Chi Minh City, Vietnam (IORG0007145).

Results

Patient characteristics

Data collected from 392 patients with T2DM were statistically analyzed according to the most common characteristic. The mean (SD) age of the patients was 65.14 (13.45) years, and participants were divided into six groups according to age. Those from 60 to 69 years of age made up the largest group, accounting for 30.4% of the sample. In addition, female patients were nearly twice as likely as males (66.6% and 33.4%, respectively) to have T2DM. Most patients reported having been diagnosed with T2DM for at least 6–10 years (36.7%); the rate of patients with disease duration from 1–5 years or ≥21 years is very low (2.6% and 2.0%, respectively).

The proportion of patients with blood glucose levels 131–179 mg/dL was the highest at 42.9%. Moreover, 69.6% of participants reported having diabetes complications, while those with comorbidities accounted for 97.4% of the total number of patients included in the survey. The vast majority of inpatients (88.0%) had different levels of health insurance coverage; the most common (80.4% of participants) had 80% health insurance coverage. A middle-school education level was most prevalent (60.2%), as was a monthly income of $200 USD (53%) (Table 1).

Prevalence of complications and comorbidities

The proportion of types of complications and number of comorbidities included were significantly different between men and women with T2DM (p=0.002; p=0.006, respectively). Similarly, the majority of comorbidities were statistically significantly different between males and females with T2DM, with a higher proportion of females reporting comorbidities than males, except for those with hypertension (p=0.434), liver diseases (p=0.058), or spondylitis (p=0.057).

Furthermore, in terms of the correlation between the prevalence of complications, with the exception of microvascular ones, patients with other complications had statistically significant correlations with patient age, such as I10-hypertension and J45-asthma (ρ=0.68, p=0.029; ρ=0.75, p=0.013; respectively) (Table 2).

Spearman's analysis also indicates that the association between comorbidities and complications is associated with hypertension and vestibular disorders with increasing age (ρ=0.79). The same relationship was observed between vestibular disorders and gastritis/duodenitis (ρ=0.93) and hypertension and macrovascular complications (ρ=0.88) (Figure 1). Figure 1 shows the correlation coefficient matrix of the Spearman test between the prevalence of complications and comorbidities associated with patient age.

Annual treatment costs associated with sociodemographic characteristics, complications, and comorbidities

The proportion of cost components varies depending on groups of each characteristic. However, it can be seen that direct costs (including direct medical costs and direct nonmedical costs) is almost higher compared to indirect costs. Moreover, in most cases, the direct medical cost is about three to four times higher than direct nonmedical cost. Intuitively, the direct medical cost of patients with treatment of oral antidiabetics (OAD) is US $ 117.4 (95% CI 102.1, 135.7), while the latter is US $42.7 (95% CI 20.7, 88.3). Direct medical costs of macrovascular complications have the smallest arithmetic mean at US $ 81.8 (95% CI 50.9, 131.1), while the largest arithmetic mean is M47-spondylosis at US $ 240.2 (95% CI 199.1, 292.5). Direct nonmedical costs have an arithmetic mean ranging from 31.0 to 36.0. The indirect cost for treating macrovascular disease has the smallest arithmetic mean US $ 80.3 (95% CI 67.1, 95.1). Indirect costs for patients with liver diseases have the largest arithmetic mean value (US $ 91.0, bootstrap 95% CI 80.5, 105.4) (Table 3).

The bootstrap analysis was conducted to identify statistically significant differences in treatment costs associated with the characteristics of patients. Of all the characteristics described, only BMI and lifestyle had no difference in total cost between each group. For age group, the average cost difference was statistically significant; specifically, the average cost of patients aged 60–69 years was higher than the 40–49 years group, the 50–59 years group, and the ≥80 years group. The average cost of the 40–49-year-old age group was higher than that of the 70–79 age group at US $94.1 (95% CI 19.8–165.9). Female patients had a higher cost of treatment than men with

Table 1 Socio-demographic and clinical characteristics of patients with diabetes mellitus in Vietnam (2016, n=392)

Patients' characteristics	n (%)	Patients' characteristics	n (%)
Age, years		**Marriage status**	
Mean (SD)	65.14 (13.45)	Single	33 (8.4)
Median (Q1–Q3)	64.0 (56–72.7)	Married	315 (80.4)
Age groups (years)		Divorced/Widow	44 (11.2)
<40	11 (2.8)	**Education***	
40–49	38 (9.7)	Low	84 (21.4)
50–59	83 (21.2)	Middle	236 (60.2)
60–69	119 (30.4)	High	72 (18.4)
70–79	73 (18.6)	**Occupation**	
≥80	68 (17.3)	Labor	74 (18.9)
Gender		Govt. officials	16 (4.1)
Male	131 (33.4)	Business	22 (5.6)
Female	261 (66.6)	Housewife	73 (18.6)
Duration of diabetes in years		Unemployed	39 (9.9)
1–5	10 (2.6)	Retired	52 (13.3)
6–10	144 (36.7)	Others	116 (29.6)
11–15	115 (29.3)	**Monthly income**	
16–20	40 (10.2)	<$200	208 (53.0)
≥21	8 (2.0)	$200–$399	23 (5.9)
Glycemic control (mg/dL)		$400–$499	27 (6.8)
70–130	152 (33.8)	$500–$599	10 (2.5)
131–179	168 (42.9)	$600–$699	12 (3.1)
≥180	72 (18.4)	No income	112 (28.6)
Complications		**Impact on income**	
Without complication	133 (38.4)	Yes	146 (37.2)
With complication	259 (69.6)	Loan	67 (17.1)
Number of complications (median, Q1–Q3)	2 (1–4)	Sell property	2 (0.5)
Comorbidities		Others	77 (19.6)
Without comorbidity	10 (2.6)	No	246 (62.8)
With comorbidity	382 (97.4)	Curtailment of expenditure	185 (47.2)
Number of comorbidities (median, Q1–Q3)	7 (3–11)	Saving money	61 (15.5)
Department		**Lifestyle**	
OPD	345 (88.0)	Drinking	22 (5.6)
IPD	47 (12.0)	Smoking	16 (4.1)
Health insurance status		Exercising	268 (68.4)
80%	315 (80.4)	**Relatives' T2DM history**	
95%	7 (1.8)	Yes	111 (28.3)
100%	70 (17.9)	No	281 (71.7)
BMI		**Transportation**	
<18.5	49 (12.5)	Motorcycle	266 (67.9)
18.5–22.9	105 (26.8)	Car	5 (1.2)
23–26.9	127 (32.4)	Taxi	3 (0.8)
≥27	102 (26.0)	Motorcycle rider	9 (2.3)
Treatment drug		Bus	109 (27.8)
Insulin and OAD	8 (2.0)		
OAD	384 (98.0)		

Note: *Low: none and primary school; Middle: lower and upper secondary school; High: diploma or Bachelor's degree or postgraduation. Data presented as number (percentage) unless stated otherwise.
Abbreviations: BMI, body mass index; IPD: inpatients department; OAD, oral antidiabetics; OPD: outpatients department; SD, standard deviation; T2DM, type 2 diabetes mellitus.
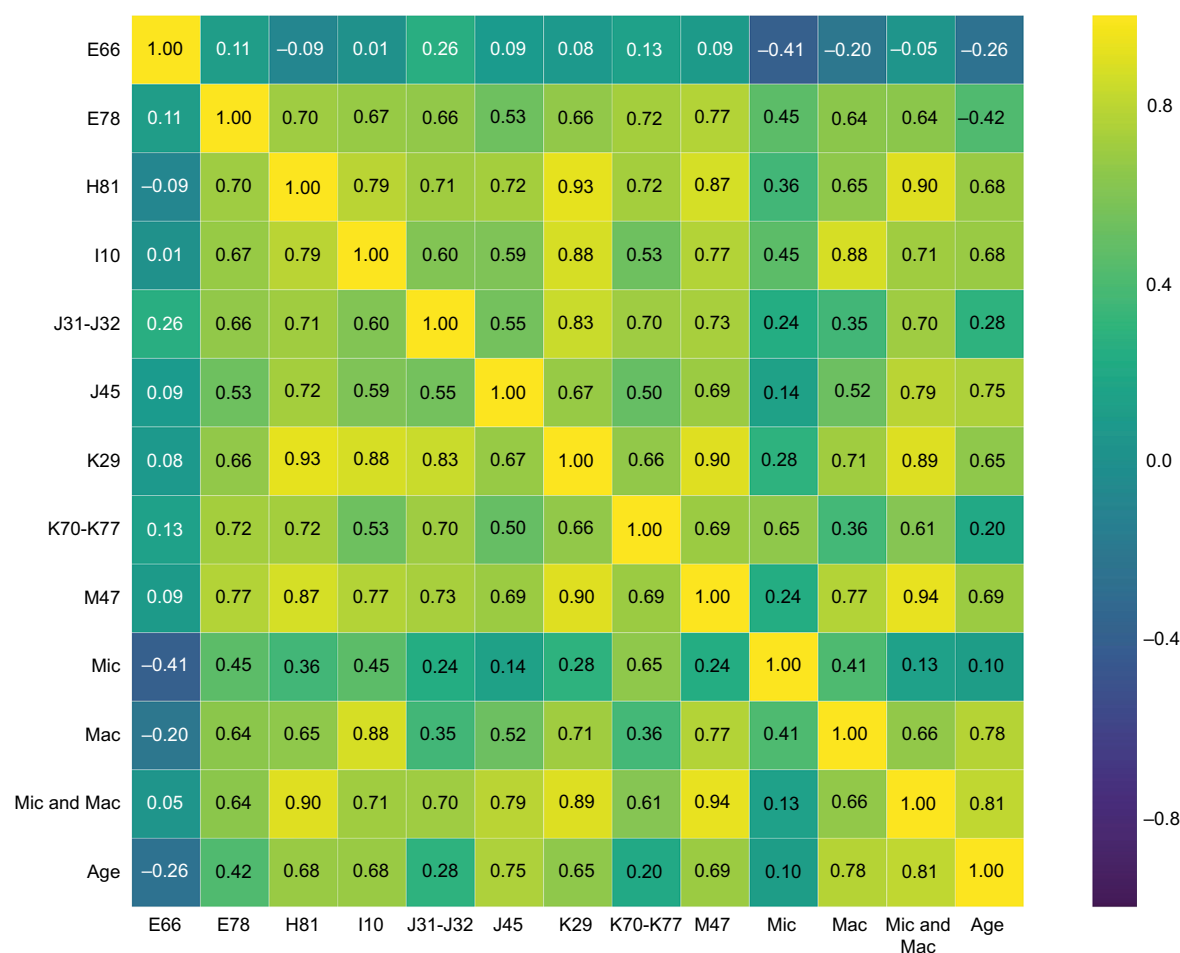
$48.9 USD (3.1, 95.0). Patients with a T2DM duration of ≥21 years had significantly higher treatment costs than those who had a T2DM duration between 6 and 10 years and 11–15 years, which were US $81.0 (95% CI 24.2, 135.3) and US $78.9 (95% CI 23.3, 140.3), respectively. When classified according to blood glucose level, the treatment cost of the 70–130 mg/dL group was lower than the other two groups and is statistically significant. Patients who used insulin and OAD had a statistically significant higher cost of treatment compared to those receiving OAD at US $445.9 (95% CI 181.2, 690.6) (Table 4).

Table 2 Prevalence of complications and comorbidities among patients with T2DM in Vietnam (2016, n=392)

	Whole sample (n=392)	Gender		p-value*	Spearman's rho**	
		Female (n=261)	Male (n=131)		ρ	p-value
Complications						
Microvascular	79 (18.2)	52 (19.9)	27 (20.6)	0.002	0.10	0.777
Macrovascular	37 (8.5)	19 (7.3)	18 (13.7)		0.78	0.008
Microvascular and macrovascular	143 (32.9)	109 (41.8)	34 (26.0)		0.81	0.005
Number of comorbidities						
None	10 (2.6)	5 (1.9)	5 (3.8)	0.006		
One	29 (7.4)	18 (6.9)	11 (8.4)			
Two	41 (10.5)	19 (7.3)	22 (16.8)			
Three	37 (9.4)	30 (11.5)	7 (5.3)			
Four or more	275 (70.2)	189 (72.4)	86 (65.6)			
(ICD-10) Comorbidities						
(E66) Obesity	102 (26.0)	64 (16.3)	38 (9.7)	0.026	−0.22	0.533
(E78) Disorders of lipoprotein metabolism	295 (75.3)	205 (78.5)	90 (68.7)	0.035	0.42	0.229
(H81) Disorders of vestibular function	168 (42.9)	129 (49.4)	39 (29.8)	<0.001	0.68	0.029
(I10) Hypertension	302 (77.0)	207 (79.3)	95 (72.5)	0.128	0.68	0.029
(J31–J32) Chronic diseases of upper respiratory system	111 (28.3)	83 (31.8)	28 (21.4)	0.011	0.28	0.425
(J45) Asthma	129 (32.9)	89 (34.1)	40 (30.5)	0.434	0.75	0.013
(K29) Gastritis/duodenitis	180 (45.9)	129 (49.4)	51 (38.9)	0.031	0.65	0.043
(K70–K77) Diseases of liver	149 (38.0)	91 (34.9)	58 (44.3)	0.058	0.20	0.579

Notes: *p-value of the Chi-square test; **Spearman's test value (ρ is the correlation and p is the p-value of ρ). Data presented as number (percentage).

Figure 1 Spearman's rank correlation coefficient of complications and comorbidity associated with age
Abbreviation: Mic, microvascular; Mac, macrovascular.

Table 3 Details of total cost per annum in USD for management of diabetes in Vietnam (2016, USD)

Patients' characteristics	Direct cost		Indirect cost	Total cost	
	Direct medical cost	Direct nonmedical cost			
	Arithmetic mean (bootstrap 95% CI)			Median	Arithmetic mean (bootstrap 95% CI)
Age groups (years)					
≤39	67.9 (19.4, 162.6)	36.1 (21.7, 54.6)	103.7 (71.4, 155.9)	136.4	207.7 (130.1, 354.9)
40–49	57.1 (31.9, 115.2)	34.1 (25.0, 47.3)	81.1 (66.8, 98.7)	145.7	172.3 (139.4, 216.9)
50–59	97.0 (66.3, 147.4)	33.5 (25.9, 43.2)	89.2 (77.6, 103.8)	172.2	219.7 (189.6, 262.3)
60–69	180.1 (145.7, 228.2)	33.0 (26.9, 40.7)	84.6 (73.7, 100.7)	216.5	297.7 (257.6, 350.7)
70–79	149.4 (109.4, 194.7)	37.1 (29.5, 47.3)	79.9 (69.2, 92.6)	233.8	266.4 (227.6, 314.2)
≥80	97.0 (76.2, 127.8)	34.7 (26.8, 44.6)	81.9 (69.5, 100.2)	172.8	213.6 (182.1, 254.5)
Gender					
Male	94.5 (73.6, 128.2)	29.5 (24.4, 35.9)	89.5 (80.3, 101.5)	162.3	213.5 (188.2, 246.7)
Female	143.8 (123.1, 167.9)	36.8 (32.3, 41.9)	81.9 (75.0, 90.6)	189.8	262.5 (238.6, 290.9)
Duration of diabetes in years					
1–5	112.0 (57.6, 168.1)	35.7 (17.7, 63.2)	87.0 (56.8, 132.9)	218.9	234.7 (166.7, 303.8)
6–10	125.7 (100.3, 160.2)	32.5 (27.3, 38.6)	85.2 (75.8, 97.8)	176.4	243.4 (214.9, 281.0)
11–15	128.7 (103.7, 160.1)	36.1 (29.5, 44.3)	80.5 (71.4, 91.1)	194.9	245.3 (215.7, 280.1)
16–20	115.2 (71.5, 207.6)	35.2 (24.3, 49.7)	84.5 (69.4, 104.0)	176.5	234.9 (187.0, 307.4)
≥21	174.4 (64.3, 300.9)	35.9 (13.3, 91.0)	113.9 (58.9, 268.7)	204.3	324.2 (154.4, 535.1)
Glycemic control (mg/dL)					
70–130	90.3 (68.8, 117.9)	36.1 (30.4, 42.6)	80.6 (73.1, 90.0)	151.1	207.0 (182.6, 241.4)
131–179	141.7 (118.7, 174.6)	35.0 (29.6, 41.2)	81.1 (72.2, 93.1)	204.8	257.8 (232.1, 289.9)
≥180	171.8 (129.4, 262.8)	29.3 (22.8, 38.6)	100.3 (86.6, 115.3)	242.5	301.4 (255.0, 370.1)
BMI					
<18.5	91.1 (64.4, 147.0)	29.3 (20.1, 43.1)	87.8 (69.9, 115.7)	170.7	208.2 (175.9, 261.7)
18.5–22.9	144.1 (111.0, 198.6)	39.3 (32.4, 47.2)	82.8 (72.7, 96.1)	204.4	266.2 (231.0, 314.8)
23–26.9	124.7 (100.1, 158.1)	33.7 (27.9, 40.6)	85.3 (76.0, 97.8)	168.2	243.7 (211.5, 284.5)
≥27	132.6 (100.9, 176.4)	32.6 (26.0, 40.5)	82.9 (72.8, 95.6)	191.8	248.1 (215.9, 290.4)
Treatment type					
OAD	117.4 (102.1, 135.7)	34.4 (30.8, 38.2)	84.4 (78.8, 91.3)	187.0	236.2 (227.9, 268.0)
Insulin and OAD	548.3 (354.3, 845.0)	42.7 (20.7, 88.3)	90.8 (55.5, 149.9)	557.9	681.8 (484.2, 954.2)
Education					
Low	123.1 (88.9, 189.1)	33.3 (25.9, 43.3)	80.5 (70.1, 95.4)	186.5	236.9 (201.4, 292.7)
Middle	137.0 (115.4, 165.1)	35.2 (30.7, 40.3)	87.2 (79.8, 96.0)	198.4	259.4 (234.9, 287.4)
High	100.3 (76.1, 130.3)	32.9 (25.8, 42.3)	79.8 (68.6, 100.3)	150.2	213.0 (180.0, 254.6)
Complications					
Microvascular	143.1 (107.5, 204.6)	36.0 (29.2, 44.1)	81.3 (71.3, 92.8)	209.3	260.4 (220.9, 322.7)
Macrovascular	81.8 (50.9, 131.1)	35.4 (23.6, 52.9)	80.3 (67.1, 95.1)	166.6	197.5 (156.0, 246.1)
Microvascular and macrovascular	227.8 (199.1, 262.8)	34.8 (29.0, 42.0)	86.8 (76.3, 100.3)	319.7	349.4 (317.0, 389.7)
ICD-10					
(E66) Obesity	132.6 (103.7, 175.1)	32.6 (26.1, 41.0)	82.9 (73.4, 96.2)	191.8	248.1 (215.4, 290.6)
(E78) Disorders of lipoprotein metabolism	157.8 (138.0, 182.6)	33.7 (29.9, 38.2)	85.2 (78.2, 93.5)	216.5	276.7 (254.4, 302.6)
(H81) Disorders of vestibular function	214.0 (185.0, 251.0)	33.8 (28.7, 39.7)	85.0 (76.1, 97.2)	275.6	332.8 (299.6, 371.8)
(I10) Hypertension	153.6 (133.6, 178.9)	35.4 (31.2, 40.0)	82.8 (76.5, 90.2)	213.0	271.8 (250.4, 298.0)
(J31–J32) Chronic diseases of upper respiratory system	231.2 (194.9, 276.6)	32.6 (26.8, 39.8)	84.6 (73.9, 99.2)	293.4	348.4 (308.9, 396.3)
(J45) Asthma	198.5 (167.4, 238.6)	31.0 (25.7, 37.3)	90.8 (80.2, 103.9)	281.9	320.3 (284.8, 362.5)
(K29) Gastritis/duodenitis	187.1 (160.4, 220.1)	31.5 (26.9, 36.7)	87.0 (77.8, 98.0)	256.3	305.6 (276.2, 341.1)
(K70–K77) Diseases of liver	208.8 (179.3, 245.8)	35.9 (30.3, 42.8)	91.0 (80.5, 105.4)	289.6	335.7 (302.9, 375.6)
(M47) Spondylosis	240.2 (199.1, 292.5)	35.4 (29.3, 42.8)	80.9 (71.2, 94.6)	298.2	356.5 (311.9, 411.5)

Abbreviations: BMI, body mass index; OAD, oral antidiabetics.

Table 4 Mean cost statistically significantly difference between groups in Vietnam (2016, USD)

Patients' characteristics			Mean difference between 2 groups	95% CI
Age groups (years)				
40–49	60–69		−125.4	(−211.7, −24.1)
	70–79		−94.1	(−165.9, −19.8)
50–59	60–69		−78.0	(−142.9, −15.3)
60–69	≥80		84.1	(29.3, 136.7)
Gender				
Male	Female		−48.9	(−95.0, −3.1)
Duration of diabetes in years				
6–10	≥21		−81.0	(−135.3, −24.2)
11–15	≥21		−78.9	(−140.3, −23.3)
Glycemic control (mg/dL)				
70–130	131–179		−50.8	(−88.3, −8.2)
	≥180		−94.4	(−137.0, −44.2)
Treatment type				
Insulin and OAD	OAD		445.9	(181.2, 690.6)
Education				
Middle	High		46.4	(12.3, 79.2)
Complications				
Microvascular	Macrovascular		62.9	(4.8, 117.0)
	Microvascular and macrovascular		−89.0	(−155.8, −19.0)
Macrovascular	Microvascular and macrovascular		−151.9	(−230.2, −64.3)
(ICD-10) Comorbidities				
(E66) Obesity	(H81) Disorders of vestibular function		−84.7	(−142.7, −29.1)
	(J31–J31) Chronic diseases of upper respiratory system		−100.3	(−156.0, −41.4)
	(J45) Asthma		−72.3	(−125.7, −17.4)
	(K29) Gastritis/duodenitis		−57.5	(−113.6, −1.9)
	(K70–K77) Diseases of liver		−87.6	(−147.4, −30.7)
	(M47) Spondylosis		−108.5	(−169.1, −44.1)
(E78) Disorders of lipoprotein metabolism	(H81) Disorders of vestibular function		−56.1	(−94.1, −19.3)
	(J31–J31) Chronic diseases of upper respiratory system		−71.8	(−107.0, −33.0)
	(J45) Asthma		−43.8	(−79.6, −10.0)
	(K70–K77) Diseases of liver		−59.1	(−94.8, −23.3)
	(M47) Spondylosis		−80.0	(−117.1, −41.2)
(I10) Hypertension	(J31–J31) Chronic diseases of upper respiratory system		−76.6	(−110.9, −42.9)
	(J45) Asthma		−48.6	(−84.9, −15.5)
	(K70–K77) Diseases of liver		−63.9	(−98.3, −30.2)
	(M47) Spondylosis		−84.8	(−122.4, −43.7)

Abbreviation: OAD, oral antidiabetics.

Annual costs of diabetes

Direct medical costs accounted for the highest proportion of total costs at 51.7%, while direct nonmedical costs made up only 14.0%; indirect costs accounted for 34.3% of the total. Of all the costs of direct medical costs, the cost of pharmaceuticals accounted for the highest proportion (27.5%), the cost of operation was lower (12.5%), and medical care and diagnoses and medical supplies (4.6%, 4.9%, and 2%, respectively) comprised a very low proportion. Similarly, in indirect cost categories, "Absenteeism" costs accounted for the highest rate of 20.9%, while "Presenteeism" cost and "Unemployment" costs were low (8.7% and 4.7%, respectively). In general, the average annual cost of treatment for a person with T2DM is US $246.1 (228.3, 267.2), while average annual direct medical costs total US $127.3 (111.3, 147.1); expenditures are US $34.4 (30.9, 8.3) for average annual direct nonmedical costs and US $84.4 (78.8, 91.1) for indirect costs (Table 5).

Sensitivity analysis

The results of the two-way sensitivity analysis revealed that the total indirect costs ranged from US $27,388.80 (−17.3%) to US $45,407.20 (+37.2%). In addition, when the average indirect costs varied from US $69.90 to US $115.80, the average total cost of treatment ranged from US $231.60 to US $277.50. Thus, the change in the average total cost of treatment when the two variables of indirect cost estimates

Table 5

Cost components	Total treatment (% of total cost of illness)	Arithmetic mean	Bootstrap 95% CI
Direct medical cost			
Medical Care	4460.6 (4.6)	11.4	(8.5, 15.6)
Diagnosis	4680.2 (4.9)	15.1	(13.4, 17.1)
Operation	12,323.3 (12.7)	77.0	(59.9, 103.6)
Pharmaceutical	26506.4 (27.5)	67.6	(58.9, 77.9)
Medical supplement	1925.9 (2.0)	5.6	(4.2, 10.7)
Total direct medical cost	**49,896.4 (51.7)**	**127.3**	**(111.3, 147.1)**
Direct non-medical cost			
Transportation	13,470.0 (14.0)	34.4	(30.9, 38.3)
Indirect cost			
Absenteeism cost	20,179.0 (20.9)	67.0	(63.4, 71.1)
Presenteeism	8408.2 (8.7)	113.6	(107.5, 123.1)
Reduced productivity for those not in the workforce	4507.5 (4.7)	49.5	(44.7, 55.9)
Total indirect cost	**33,094.8 (34.3)**	**84.4**	**(78.8, 91.1)**
Total cost of illness	**96,461.1 (100.0)**	**246.1**	**(228.3, 267.2)**

Note: Bold data is the total cost of each cost type

fluctuated remained within the 95% CI of the mean total cost of treatment estimated by the study.

Machine learning algorithms to predict T2DM treatment costs

The Gradient Boosting Regression and the Lasso Regression are the two best machine learning regression algorithms for predicting annual direct medical costs (R^2=65.42, MSE=0.938; and R^2=64.32, MSE=0.968, respectively) with a Box–Cox transformation and the total direct medical costs (with lambda = 0.024) (Table 6). The optimal model (Gradient Boosting Regression) identified seven factors that contribute to the cost of diabetes: number of comorbidities (0.42), age (0.23), health insurance status (0.1), duration of illness (0.09),

Table 6

Machine learning algorithms	R^2 score	MSE
Generalized linear models		
Ordinary least squares	63.71	0.985
Ridge Regression	63.83	0.982
Lasso Regression	**64.32**	**0.968**
Ensemble methods		
Random forest	63.58	0.988
Gradient boosting	**65.42**	**0.938**
ADA boosting	64.86	0.954
Extra trees	63.38	0.994

Note: R^2 score: Higher is better. MSE (mean square error): Lower is better. Bold data is the fittest model in each models.
Abbreviation: ADA, adaptive.

type of treatment (0.07), blood glucose level (0.07), and type of complication (0.02).

Discussion

This prevalence-based cost-of-illness study involved a targeted-group population that was generated from cost and administrative data retrieved from the Government Public Hospital on patients with T2DM in Vietnam. The annual cost per patient estimate in this study was US $246.10 (95% CI 228.3, 267.2), which accounted for about 12% (95% CI 11, 13) of the gross domestic product (GDP) per capita in 2017. Of this cost per patient, US $127.30, US $34.40, and US $84.40 were spent on direct medical costs, direct nonmedical costs, and indirect costs, respectively. These expenses seem lower than those reported in other Asian and Southeast Asian countries. According to a 2008 study in Thailand, the average cost treating a diabetes patient was estimated at US $881.47 (20% of the GDP per capita in 2008),[48] while a study of the economic costs of diabetes in Iran reported a cost of US $544.00 (accounting for 17% of the GDP per capita in 2005).[49] A Chinese study reported costs of US $1,707.80 (45% of the GDP per capita in 2009) to manage each patient with T2DM.[50] It is notable that the direct medical costs reported by high-income countries, such as Singapore or nations in the Americas and Europe, are much higher.[33,35,51,52] In addition, the indirect cost in this study was recorded as US $84.40 (78.80, 91.10); absenteeism costs accounted for 60.9% of this cost. This finding indicated that the correlation between the number of work-days missed significantly influences the indirect costs; one possible explanation for this is that the indirect costs in the study were lower than those reported by other studies[15] because the GDP per capita in Vietnam is lower than that of most other countries.

In terms of cost components, pharmaceuticals accounted for the bulk of the total expenditures in our study (27.5% of the total costs and 53.2% of the direct medical costs), which suggests that patients with T2DM are aware of the importance of treating the disease, as demonstrated by their compliance with prescribed prescription medications. In contrast with the findings of other studies, the second most expensive component of our total costs was medications.[33,35,49,52] The present study also found that patients in the insulin and OAD group had higher treatment costs; some other studies also reported similar results.

Patients with poor glycemic control and associated diseases (hypertension, hyperlipidemia, and obesity) had higher treatment costs than those in the other groups, which

was in agreement with the findings of another study.[48] In addition, more highly educated people have lower treatment costs, which may be because more educated people are likely to maintain a healthy lifestyle while monitoring and treating their disease from its onset, while the average and less educated people are often uninterested in the disease until the onset of complications, so that the cost of treatment for severe illness is high.

Machine learning algorithms generally provide almost the same results in predicting the cost of treatment, but only two algorithms can eliminate factors that do not directly affect the cost of treatment, namely the Gradient Boost Regression and the Lasso Regression. The former algorithm yields the highest accuracy and identifies important features in estimating the costs, such as the number of comorbidities (0.42), age (0.23), health insurance status (0.1), the duration of illness (0.09), type of treatment (0.07), blood glucose level (0.07), and type of complications (0.02) (Figure 2). However, this algorithm is difficult to explain and is complex for nonexperts.[53] Therefore, the Lasso Regression algorithm is a better choice when presenting at nonprofessional seminars. It selects 15 of 17 features to estimate the annual direct medical cost; the regression is shown in Figure 3 with the significant factors that influence the costs, and the regression formula is shown below:

$$\begin{aligned} y = {}& 4.065 + 1.288x_1 + 0.680x_2 + 0.131x_3 \\ & + 0.107x_4 + 0.063x_5 - 0.002x_6 - 0.008x_7 \\ & - 0.009x_8 - 0.012x_9 - 0.031x_{10} - 0.036x_{11} \\ & - 0.048x_{12} - 0.052x_{13} - 0.221x_{14} - 0.237x_{15} \end{aligned}$$

As a prevalence-based cost-of-illness study from a social perspective, the strength of this study was to estimate the average annual T2DM treatment cost per case, as well as the cost of treatment for specific disease groups, especially those with different comorbidities. Moreover, the study used a variety of machine learning algorithms for a more in-depth and accurate analysis, especially in the era of technological development science, which many developing countries are using step by step to apply management data to electronics. The application of machine learning algorithms and data science knowledge to large data sources will help decision makers adopt more appropriate and effective policies for the management of diabetes. Nevertheless, this research study has several limitations that must be pointed out. First, the direct nonmedical costs and indirect costs were estimated using the subjective opinion of the patient. For this reason, they may have been inaccurate. Second, the direct medical costs were estimated using a source of unit cost reference that led to bias. Finally, the sample was small and, therefore,
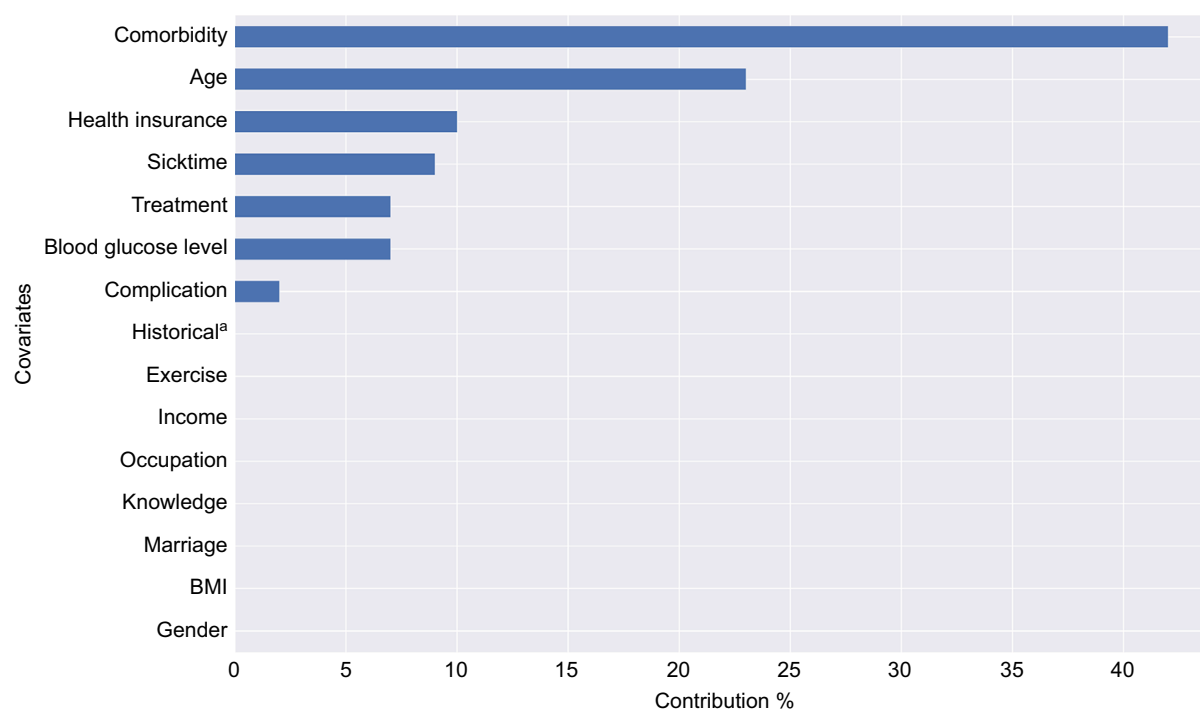

Figure 2 The percentage of covariances that contribute to the cost of treatment. [a]History of Type 2 diabetes of relative.
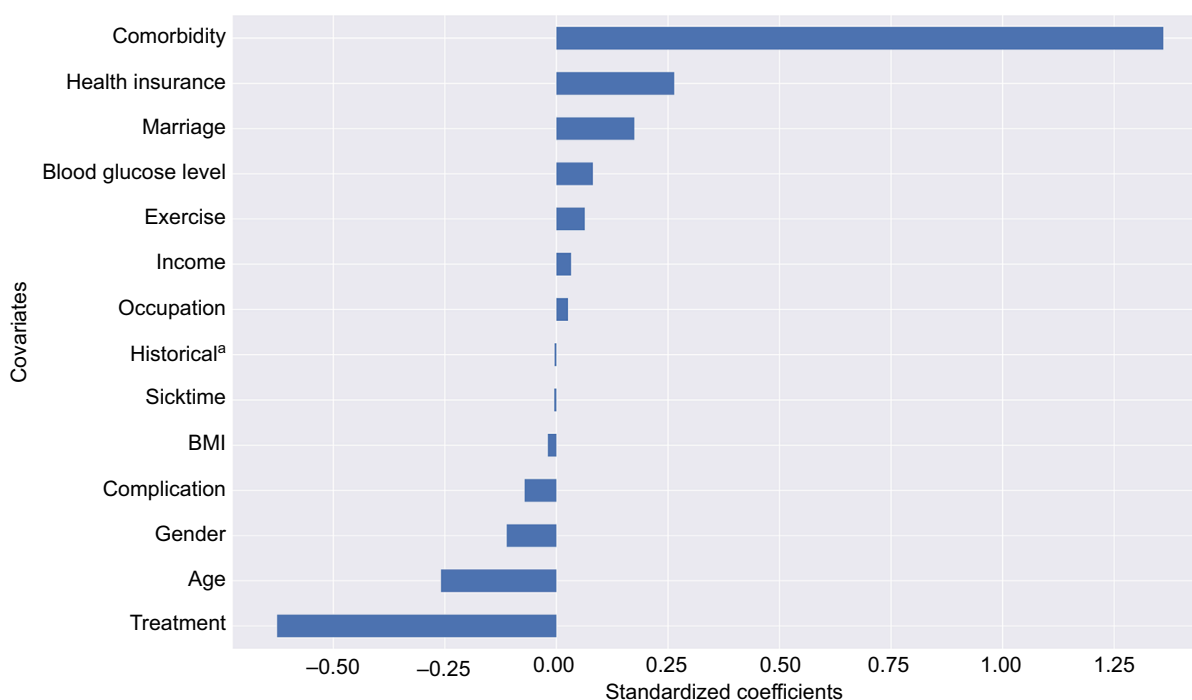**Abbreviation:** BMI, body mass index.

Figure 3 Standardized coefficient of Lasso Regression covariates
Note: [a]History of Type 2 diabetes of relative.
Abbreviation: BMI, body mass index.

cannot represent the entire population of Vietnam. Future research on public health facilities will help to minimize these shortcomings.

Conclusion

This is the first study to provide a comprehensive cost analysis of the cost of treating diabetes in Vietnam. The results indicate that the numbers of comorbidities and also the selected treatment type significantly affect the cost of treatment. With the increase in the incidence of diabetes along with obesity, diabetes will continue to be a burden on health budgets. In addition, new technology has simplified data management and analysis, and the application of data analysis techniques will make disease research more effective and accurate. Research is, therefore, an important step for other studies in Vietnam to help policymakers better manage diabetes as well as adjust the costs of the disease to reduce the long-term economic burden of the increasing pandemic.

Acknowledgments

Disclosure

The authors report no conflicts of interest in this work.

References

1. International Diabetes Federation. *IDF diabetes atlas.* 2015.
2. World Health Organization. *Global report on diabetes.* 2016.
3. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet.* 2016;387(10027):1513–1530.
4. World Health Organization. Global Health Observatory data repository; 2015. Available from: http://apps.who.int/gho/data/node.main. Accessed February 2, 2017.
5. Alberti KGMM, DeFronzo RA, Zimmet P, editors. *International textbook of diabetes mellitus.* New York, NY: J. Wiley; 1997.
6. American Diabetes Association. *Standards of medical care in diabetes.* Vol 39: Diabetes Care; 2016. Available from: http://care.diabetes-journals.org/content/suppl/2015/12/21/39.Supplement_1.DC2/2016-Standards-of-Care.pdf.
7. World Health Organization. *Definition and diagnosis of diabetes mellitus and intermediate hyperglycemia.* 2006. Available from: http://www.who.int/diabetes/publications/Definition%20and%20diagnosis%20of%20diabetes_new.pdf.
8. Bruno Graziella RC, Cavallo-Perin P, Merletti F, et al. Incidence of type 1 and type 2 diabetes in adults aged 30–49 years. *Diabetes Care.* 2005;28(11):2613–2619.
9. Evans JM, Newton RW, Ruta DA, MacDonald TM, Morris AD. Socio-economic status, obesity and prevalence of Type 1 and Type 2 diabetes mellitus. *Diabet Med.* 2000;17(6):478–480.
10. Holman N, Young B, Gadsby R. Current prevalence of Type 1 and Type 2 diabetes in adults and children in the UK. *Diabet Med.* 2015;32(9):1119–1120.
11. Boyle JP, Thompson TJ, Goldschmid MG, et al. Estimating prevalence of type 1 and type 2 diabetes in a population of African Americans with diabetes mellitus. *Am J Epidemiol.* 1999;149(1):55–63.

12. World Health Organization. *Global action plan for the prevention and control of noncommunicable diseases 2013–2020.* 2013. Available from: http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng. pdf.

13. World Health Organization. *Global status report on noncommunicable diseases 2015.* 2015. Available from: http://apps.who.int/iris/bitstream/10665/148114/1/9789241564854_eng.pdf.

14. Srinivasan B, Taub N, Khunti K, Davies M. Diabetes: glycaemic control in type 2. *BMJ Clin Evid.* 2008;6(09).

15. Seuring Till AO, Suhrcke M. The economic costs of type 2 diabetes: a global systematic review. *Pharmacoeconomics.* 2015;33(8): 811–831.

16. Wong Evelyn, Backholer K, Gearon E, Harding J, Freak-Poli R, Stevenson C, Peeters A. Diabetes and risk of physical disability in adults: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol.* 2013;1(2):106–114.

17. Florencia A, Alex B, Nam HC, et al. IDF diabetes atlas. In: Leonor G, Tim N, Jessica B, Ute L, Olivier J, editors. Basel, Switzerland: International Diabetes Federation; 2013.

18. American Diabetes Association. Economic costs of diabetes in the US in 2012. *Diabetes Care.* 2013;36(4):1033–1046.

19. The World Bank. Vietnam Data. Available from: http://www.worldbank. org/vi/country/vietnam. Accessed February 12, 2017.

20. World Health Organization/Western Pacific Region. *The growing burden of diabetes in Viet Nam.* 2016; Available from: http://www.wpro.who.int/vietnam/mediacentre/features/feature_world_health_day_2016_vietnam/en/. Accessed February 12, 2017.

21. Pham NM. Prevalence and determinants of diabetes and prediabetes among Vietnamese adults. *Diabetes Res Clin Pract.* 2016;113:116–124.

22. International Diabetes Federation Western Pacific. *Diabetes in Vietnam 2015.* 2015; Available from: http://www.idf.org/membership/wp/vietnam.

23. World Health Organization. *WHO guide to identifying the economic consequences of disease and injury;* 2009. Available from: http://www. who.int/choice/publications/d_economic_impact_guide.pdf.

24. Tarricone R. Cost-of-illness analysis: what room in health economics? *Health Policy.* 2006;77(1):51–63.

25. Onukwugha E, McRae J, Kravetz A, Varga S, Khairnar R, Mullins CD. Cost-of-illness studies: an updated review of current methods. *Pharmacoeconomics.* 2016;34(1):43–58.

26. Changik J. Cost-of-illness studies: concepts, scopes, and methods. *Clin Mol Hepatol.* 2014;20(4):327–337.

27. Kobelt G. Health economics, economic evaluation, and glaucoma. *J Glaucoma.* 2002;11(6):531–539.

28. Trung QV, Huong TN, Riewpaiboon A. Hospital service cost analysis in developing countries: a method comparison in Vietnam. In press 2016.

29. Bank TW. *Vietnam;* 2017. Available from: http://data.worldbank.org/country/vietnam. Accessed June 27, 2017.

30. Cawley J, Rizzo JA, Haas K. The association of diabetes with job absenteeism costs among obese and morbidly obese workers. *J Occup Environ Med.* 2008;50(5):527–534.

31. Fu AZ, Qiu Y, Radican L, Wells BJ. Health care and productivity costs associated with diabetic patients with macrovascular comorbid conditions. *Diabetes Care.* 2009;32(12):2187–2192.

32. Lee LJ, Yu AP, Cahill KE, et al. Direct and indirect costs among employees with diabetic retinopathy in the United States. *Curr Med Res Opin.* 2008;24(5):1549–1559.

33. American Diabetes Association. Economic costs of diabetes in the US in 2012. *Diabetes Care.* 2013;36(4):1033–1046.

34. American Diabetes Association. Economic costs of diabetes in the US in 2007. *Diabetes Care.* 2008;31(3):596–615.

35. Ng CS, Toh MPHS, Ko Y, Joyce YCL. Direct medical cost of type 2 diabetes in Singapore. *PLoS One.* 2015;10(3):e0122795.

36. Barber JA. Analysis and interpretation of cost data in randomised controlled trials: review of published studies. *BMJ.* 1998;317(7167): 1195–1200.

37. Desgagne A, Castilloux AM, Angers JF, LeLorier J. The use of the bootstrap statistical method for the pharmacoeconomic cost analysis of skewed data. *Pharmacoeconomics.* 1998;13(5):487–497.

38. Efron B, Tibshirani RJ. *An introduction to the bootstrap.* USA, CRC Press; 1994.

39. Thompson SG. How should cost data in pragmatic randomised trials be analysed? *BMJ.* 2000;320(7243):1197.

40. Fowler MJ. Microvascular and macrovascular complications of diabetes. *Clin Diabetes.* 2008;26(2):77–82.

41. Png ME, Yoong J, Phan TP, Wee HL. Current and future economic burden of diabetes among working-age adults in Asia: conservative estimates for Singapore from 2010–2050. *BMC Public Health.* 2016;16(1):153.

42. Chaze JP. Assessing household health expenditure with Box–Cox censoring models. *Health econ.* 2005;14(9):893–907.

43. Gregori D, Petrinco M, Bo S, Desideri A, Merletti F, Pagano E. Regression models for analyzing costs and their determinants in health care: an introductory review. *Int J Qual Health Care.* 2011;23(3):331–341.

44. Edward JJ. *A user's guide to principal components.* Vol 587: USA, John Wiley & Sons; 2005.

45. Mohamad IB, Usman D. Standardization and its effects on K-means clustering algorithm. *Res J Appl Sci Engineer Technol.* 2013;6(17): 3299–3303.

46. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. *J Mach Learn Res.* 2011;12:2825–2830.

47. Van Rossum G, Drake Jr, FL. *Python tutorial.* Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands; 1995.

48. Susmita C, Arthorn R, Piyanuch P, et al. Cost of diabetes and its complications in Thailand: a complete picture of economic burden. *Health Soc Care Community.* 2011;19(3):289–298.

49. Esteghamati A, Khalilzadeh O, Anvari M, et al. The economic costs of diabetes: a population-based study in Tehran, Iran. *Diabetologia.* 2009;52(8):1520–1527.

50. Wang W, McGreevey WP, Fu C, et al. Type 2 diabetes mellitus in China: a preventable economic burden. *Am J Manag Care.* 2009;15(9):593–601.

51. Leśniowska J, Schubert A, Wojna M, Skrzekowska-Baran I, Fedyna M. Costs of diabetes and its complications in Poland. *Eur J Health Econ.* 2014;15(6):653–660.

52. Ingrid K, Eduard H, Hans H, Ingrid S. Direct costs of diabetes mellitus in Germany–CoDiM 2000–2007. *Exp Clin Endocrinol Diabetes.* 2011;119(06):377–385.

53. Guido S, Muller AC. Introduction to machine learning with Python: a guide for data scientists. USA, O'Reilly Media; 2017.