ORIGINAL RESEARCH ARTICLE

Medical Costs of Patients with Type 2 Diabetes in a Single Payer System: A Classification and Regression Tree Analysis

Paola Rucci[1] · Vera Maria Avaldi[2] · Claudio Travaglini[3] · Cristina Ugolini[4] · Elena Berti[5] · Maria Luisa Moro[5] · Maria Pia Fantini[1]

Abstract

Background and Objectives Many studies and systematic reviews have estimated the healthcare costs of diabetes using a cost-of-illness approach. However, in the studies based on this approach patients' heterogeneity is rarely taken into account. The aim of this study was to stratify patients with type 2 diabetes into homogeneous cost groups based on demographic and clinical characteristics.

Methods We conducted a retrospective cost-of-illness study by linking individual data on health services utilization retrieved from the administrative databases of Emilia-Romagna Region (Italy). Direct medical costs (either all-cause or diabetes-related) were calculated from the perspective of the regional health service, using tariffs for hospitalizations and outpatient services and the unit costs of prescriptions for drugs. The determinants of costs identified in a generalized linear regression model were used to characterize subgroups of patients with homogeneous costs in a classification and regression tree analysis.

Results The study population consisted of a cohort of 101,334 patients with type 2 diabetes, followed up for 1 year, with a mean age of 70.9 years. Age, gender, complications, comorbidities and living area accounted significantly for cost variability. The classification tree identified ten patient subgroups with different costs, ranging from a median of €483 to €39,578. The two subgroups with highest costs comprised dialysis patients, and the largest subgroup (57.9%) comprised patients aged ≥ 65 years without renal, cardiovascular and cerebrovascular complications.

Conclusions Classification of patients into homogeneous cost subgroups can be used to improve the management of, and budget allocation for, patients with type 2 diabetes.

Key Points for Decision Makers

Patients with type 2 diabetes differ on comorbid conditions and complications.

There is no 'average' patient with diabetes.

The stratification into ten subgroups proposed in this paper may support clinicians in implementing effective preventive interventions and policy makers in designing resource allocation tailored to patients' needs.

1 Introduction

Diabetes mellitus (DM) is a major public health problem with a high morbidity and mortality burden and high healthcare costs [1–3]. The International Diabetes Federation estimated that, in 2017, 451 million people were living with diabetes worldwide, and this figure is expected to increase to 693 million by 2045 [4] due to population growth, population aging, fast urbanization, excessive caloric intake and sedentary behaviors [5].

✉ Vera Maria Avaldi
  veramaria.avaldi@studio.unibo.it

1 Department of Biomedical and Neuromotor Sciences, Alma Mater Studiorum University of Bologna, via san Giacomo 12, 40126 Bologna, Italy

2 Advanced School for Healthcare Policies, Alma Mater Studiorum University of Bologna, via San Giacomo 12, 40126 Bologna, Italy

3 Department of Management, Alma Mater Studiorum University of Bologna, via Capo di Lucca 34, Bologna, Italy

4 Department of Economics and Advanced School for Healthcare Policies, Alma Mater Studiorum University of Bologna, Piazza Scaravilli 2, 40126 Bologna, Italy

5 Regional Agency for Health and Social Care, Viale Aldo Moro 21, 40127 Bologna, Italy

△ Adis

Type 2 diabetes (T2DM) is the most prevalent type of diabetes, affecting almost 90% of people with diabetes [6].

In Italy, it has been estimated that the prevalence of DM in 2016 was 6.3%, and that one in five people were not diagnosed [7].

The total global healthcare expenditure for people with diabetes aged 18–99 years was estimated at US$850 billion in 2017 and is expected to increase by 7% by 2045 [4]. In 2017, the North American and Caribbean Region accounted for 52% of the total amount spent worldwide on diabetes and the Europe Region accounted for a large share of the total global spending (26%), with a mean expenditure of 3432 international dollars [4]. In 2045, it is expected that healthcare costs for diabetes will remain stable for the population under the age of 50 years, but will increase by 37% for the population above 70 years [8]. The principal cost components are hospital and outpatient care, even if drug costs are becoming more relevant since the introduction of the expensive analog insulins and incretin-based agents and sodium–glucose cotransporter 2 (SGLT-2) inhibitors [9]. Many studies and systematic reviews [8, 10–17] have evaluated healthcare costs incurred by a population with diabetes using a cost-of-illness approach that includes all direct and indirect costs related to lost productivity. A recent systematic review [3] showed that the annual direct costs of diabetes per patient in international dollars, regardless of the costing method applied and the cost components included, range from $242 in Mexico in 2010 [18] to $11,917 in a study conducted in the USA in 2007 [19], while indirect costs range from $45 for Pakistan in 2006 [20] to $16,914 for the Bahamas in 2000 [21].

A recent study by Marcellusi et al. [22] reported that the total economic burden of patients with diabetes in Italy accounts for €20.3 billion/year, of which 54% is attributable to indirect costs and 46% to direct costs. It has been estimated that in Italy the mean annual healthcare cost for patients with diabetes is about €3000 [7, 21, 23, 24]; that is more than twice that of patients without diabetes.

Converging evidence from different studies suggests that few patients with great clinical complexity account for a large part of diabetes-related costs [10, 11]. In fact, diabetes complications, in particular, those due to microvascular and macrovascular damage (eye, renal and cardiovascular complications) account for a large proportion of healthcare expenditure [6, 10, 11, 25, 26], and comorbidities not directly related to diabetes also contribute to the economic burden of this condition [25, 27, 28].

Given the heterogeneity of patients with T2DM in terms of age, complications and comorbid conditions, the aim of the present study was to estimate, in a geographical area of North-Eastern Italy, the annual medical costs of T2DM, either all-cause or diabetes-related, and to stratify patients into different cost groups based on demographic and clinical characteristics. The Italian National Health System (NHS) provides universal coverage for all its citizens and funds each Regional Health System, which pays for almost all medical costs, except for the small part of the costs of drugs and services that are co-paid by patients. People with diabetes are exempt from co-paying. In this context, the identification of patient subgroups with similar healthcare needs and clinical complexity could be useful to improve quality and cost control and to provide effective and cost-effective healthcare interventions and implement a payment system based on a tailored budget perspective, as applied in other countries.

2 Methods

We conducted a prevalence-based cost-of-illness study using a comprehensive global bottom-up approach to calculate the direct medical costs incurred by the NHS for patients with T2DM over 1 year.

The comprehensive approach, which incorporates costs attributable and non-attributable to T2DM diabetes, provides an accurate description of medical expenditure because diabetes is a chronic disease frequently affected by complications and comorbid conditions.

2.1 Data Source

Patient data including healthcare use (outpatient and inpatient care) and drugs were obtained by linking data from different regional administrative databases using a unique anonymized patient identifier.

The hospital discharge record (HDR) database includes demographic characteristics, admissions and discharge dates, the main and up to five secondary diagnoses and up to six interventions [identified using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) coding system]. In Italy, HDRs are a reliable and accurate source of hospital data based on the diagnosis-related group (DRG) system that is used to allocate funds to hospitals and to monitor quality of care and outcomes at the national level.

The outpatient pharmaceutical database includes information on patients' gender and age, prescriptions [substance name, Anatomical Therapeutic Chemical (ATC) classification system code-V.2013, trade name, date of prescription filling, and number of packages]. This database includes drugs reimbursed by the healthcare system that are prescribed by the general practitioner (GP) or a specialist. Drug prescriptions are tracked in the AFT (outpatient pharmaceutical supply) and FED (direct supply drugs) databases.

The outpatient services database (ASA) includes laboratory tests, diagnostic, therapeutic and rehabilitation services and specialty visits, but does not include primary care visits, that cannot be traced.

Demographic information was retrieved from the different sources (HDR, ASA, AFT, and FED).

2.2 Study Population

The study population included adult patients with T2DM living in three Local Health Authorities (LHAs) of Emilia-Romagna Region (Italy), comprising about 2 million people. Patients were identified in 2014 through a regional algorithm [24] that combined information from different administrative databases (hospital admissions, outpatient care, drug prescriptions, exemption for disease) and LHA data. In particular, we selected patients who in 2014 met at least one of the following criteria:

- Hospital discharge with ICD-9 CM primary or secondary diagnosis of diabetes (code 250.XX).
- At least two consecutive prescriptions of drugs for diabetes classified with the ATC classification system (code A10).
- Exemption from co-payment healthcare costs for DM.

Patients alive on 1/1/2015 were followed up to 31/12/2015 to collect health services use and calculate the related annual direct medical costs. Subjects with missing data were excluded. Comorbidities were tracked from drug prescriptions in the previous year using an algorithm developed by Maio et al. [29], and complications in the previous 3 years were retrieved from the HDR database and classified according to a regional dossier [24].

2.3 Outcomes

The outcomes of interest were the annual healthcare service use (outpatient and inpatient care and drug treatment) and costs of health services provided to patients with T2DM over the year 2015.

2.4 Costs

To calculate the annual medical costs for each patient with T2DM, we multiplied the number of services used by the respective unit cost. We used DRG tariffs as a proxy of costs for hospital admissions (http://salute.regione.emilia-romagna.it/siseps/sanita/sdo/files/DGR_1673_2014.pdf/view), the regional nomenclator (http://salute.regione.emilia-romagna.it/documentazione/nomenclatore-tariffario-rer)

for specialty visits and the unit cost of prescriptions for medications.

2.5 Statistical Analysis

Categorical variables were summarized using frequencies and percentages, and continuous variables using mean, standard deviation, median and range.

A generalized linear model (GLM) with gamma probability distribution and log link was used to investigate the determinants of costs. In this model, costs were regressed on gender, age group (under 65 years, between 65 and 75 years, over 75 years), LHA of residence (Bologna, Parma, Modena), duration of illness (less than 1 year, between 1 and < 5 years, 5 years or more), complications in the previous 3 years (coma, ischemic heart disease, stroke, peripheral revascularization, amputation, eye complications, renal complications and dialysis) and number of comorbidities (none, one, two, three or more), detected through drug prescriptions. The goodness of fit of the model was measured using the predictive ratio, which is calculated as the ratio of predicted costs to observed costs. The value of the predictive ratio is always positive, with a value closer to one indicating higher predictive performance.

Then a classification and regression tree (CART) model [30] was used to stratify patients into homogeneous subgroups of costs based on demographic and clinical characteristics. This method offers advantages with respect to other classification methods such as logistic regression or cluster analysis because it relies on assumptions about the type of relationship between the outcome and the explanatory variables, can handle multicollinearity among explanatory variables by selecting the best splitter and deals easily with interactions among variables [31].

The independent variables used in the CART were the same as those included in the GLM, except for the number of comorbidities. This was done because more than 90% of patients had at least one comorbidity derived from drug prescriptions, and some of these (gastrointestinal and musculoskeletal) were very common and non-specific, thereby hindering the creation of distinct subgroups.

The CART selects in an automated way from the set of variables the one most associated with costs, and uses it to divide the population into two subgroups. If the independent variable is continuous (such as age), the procedure identifies a specific cut-off that maximizes the cost difference between two subgroups. If the independent variables consist of more than two categories, the categories are merged to optimize the distinction into subgroups. Dichotomous variables are used as such by the procedure. The procedure continues recursively, selecting the second best variable and subdividing each subgroup recursively into two further subgroups and so on, until no significant improvement in

the classification of participants is possible. The maximum depth that can be achieved in the tree, i.e., maximum number of subsequent divisions, is five. An internal validation of the CART was conducted using a split-sample approach, in which the tree was first generated using a training sample, including 50% randomly selected cases, and then tested for classification accuracy in the remaining 50% of cases. Moreover, the tree was pruned to avoid overfitting the model. The tree's predictive accuracy was measured using the risk estimate that denotes the within-node variance. A low risk estimate indicates node homogeneity.

Statistical analyses were performed using IBM SPSS, version 25.

3 Results

Through the database linkage we identified 102,638 patients with T2DM at 1/1/2015, of which 1304 were excluded, because of missing data. The final study population therefore includes 101,334 patients

Table 1 describes the characteristics of the study population. Patients had a mean age of 70.9 years and a slight predominance of males (54.1%); 72.8% had diabetes for at least 5 years; 65.6% had two or more comorbidities and 77.0% had no complications in the previous 3 years. Almost all patients were taking medications or had received outpatient care, and one in five was hospitalized during the follow-up.

The total costs for diabetic patients incurred in 2015 by the three LHAs (Table 2) show a skewed frequency distribution, with a median value of €1012 and a mean value of €3086. Hospitalizations, drugs and outpatient care accounted for 53.1%, 29.5% and 17.4% of the total cost, respectively.

In the GLM estimating the total costs as a function of sociodemographic and clinical variables (Table 3), all independent variables were statistically significant ($p < 0.001$). Annual costs were higher in males, in patients aged more than 65 years compared with younger patients, and increased with the number of comorbidities. In addition, the complications that were the most relevant drivers of costs were dialysis ($b = 2.165$, 95% CI 2.045–2.286; $p < 0.001$), peripheral revascularization ($b = 0.494$, 95% CI 0.387–0.600; $p < 0.001$), renal complications ($b = 0.470$, 95% CI 0.427–0.514; $p < 0.001$), coma ($b = 0.369$, 95% CI 0.304–0.434; $p < 0.001$), and ischemic heart disease ($b = 0.317$, 95% CI 0.289–0.344; $p < 0.001$).

Notably, costs were also higher for newly diagnosed patients compared with those of patients with an established diagnosis and for patients living in the Parma LHA as compared with those living in Bologna or Modena LHAs.

In the CART model, the costs estimated by the GLM were used as the dependent variable. This procedure initially split the study population into two subgroups based on age

Table 1 Characteristics of the study population ($N = 101,334$)

	N	%
Gender		
Female	46,538	45.9
Male	54,796	54.1
Age (mean ± SD)	70.9 ± 12.8	
Age group		
< 65	29,937	29.5
65–74	29,351	29.0
≥ 75	42,046	41.5
Length of illness		
< 1 year	7370	7.3
≥ 1 year to < 5 years	20,147	19.9
≥ 5 years	73,817	72.8
No. of complications in 2012–2014		
0	78,011	77.0
1	17,849	17.6
2	4225	4.2
3 or more	1249	1.2
Coma	1319	1.3
Ischemic heart disease	8676	8.6
Stroke	5952	5.9
Peripheral revascularization	495	0.5
Amputation	384	0.4
Dialysis	381	0.4
Other renal complications	3153	3.1
Eye complications	9936	9.8
No. of comorbidities in 2012–2014[a]		
0	9824	9.7
1	25,002	24.7
2	26,992	26.6
3 or more	39,470	39.0
Comorbidities in the last 3 years		
Cancer	4907	4.8
Eye	9146	9.0
Gastrointestinal	43,592	43.0
Genitourinary	1314	1.3
Hematological	3345	3.3
Hepatobiliary	267	0.3
Immunologic	281	0.3
Male genital	11,396	11.2
Musculoskeletal	30,291	29.9
Neurologic	10,072	9.9
Psychological	18,149	17.9
Respiratory	11,562	11.4
Skin	1181	1.2
Cardiovascular	81,253	80.2
Death	4116	4.1

SD standard deviation

[a] 116 missing values

Table 2 Summary of direct medical costs

	Patients		Median (€)	Mean (€)	SD (€)	Min (€)	Max (€)	Sum (€)	%
	N	%							
Hospitalizations (HDR)	22,937	22.6	4221.4	7236.3	8884.2	70.0	193,830.7	165,980,102	53.1
Outpatient care[a] (ASA)	94,915	93.7	262.0	572.3	2253.6	1.1	64,496.0	54,321,811	17.4
Diabetes drugs (ATC A10)	87,613	86.5	71.7	183.2	265.4	0.0	3419.4	16,047,912	5.1
All drugs (AFT + FED)[b]	99,845	98.5	514.4	925.7	3453.8	0.0	246,620.9	92,424,247	29.5
Total	101,334	100	1012.6	3086.1	7085.9	1.3	252,361.7	312,726,161	100

AFT outpatient pharmaceutical supply database, *ASA* outpatient services database, *ATC* Anatomical Therapeutic Chemical, *FED* direct supply drugs database, *HDR* hospital discharge record, *SD* standard deviation

[a]Lab tests and specialty visits

[b]Including diabetes drugs

Table 3 Predictors of overall costs. Results of the generalized linear model. The predictive ratio is 1.011, denoting a good concordance between predicted and observed costs

	b	SE (b)	P	95% CI	
				Lower limit	Upper limit
Constant	7.180	0.0191	< 0.001	7.143	7.218
Male gender	0.136	0.0076	< 0.001	0.121	0.151
Age groups					
< 65 years	(Reference category)				
65–74 years	0.190	0.0100	< 0.001	0.171	0.210
≥ 75 years	0.177	0.0096	< 0.001	0.159	0.196
LHA of residence					
Parma	*(Reference category)*				
Bologna	−0.044	0.0103	0.001	−0.064	−0.024
Modena	−0.056	0.099	< 0.001	−0.075	−0.036
Length of illness					
< 1 years	(Reference category)				
≥ 1 and < 5 years	−0.228	0.0164	< 0.001	−0.260	−0.196
≥ 5 years	−0.139	0.0149	< 0.001	−0.169	−0.110
Number of comorbidities in the last 3 years					
0	(Reference category)				
1	0.270	0.0145	< 0.001	0.242	0.298
2	0.584	0.0146	< 0.001	0.555	0.613
≥ 3	1.000	0.0143	< 0.001	0.972	1.028
Complications in the last 3 years					
Coma	0.369	0.0333	< 0.001	0.304	0.434
Ischemic heart disease	0.317	0.0140	< 0.001	0.289	0.344
Stroke	0.156	0.0164	< 0.001	0.124	0.188
Peripheral revascularization	0.494	0.0544	< 0.001	0.387	0.600
Amputation	0.419	0.0617	< 0.001	0.298	0.540
Eye complications	0.172	0.0127	< 0.001	0.147	0.197
Dialysis	2.165	0.0614	< 0.001	2.045	2.286
Renal complications	0.470	0.0220	< 0.001	0.427	0.514

CI confidence interval, *LHA* Local Health Authority, *SE* standard error

(< 65 years and ≥ 65 years) and created further subdivisions using five variables (dialysis, ischemic heart disease, renal complications, stroke and duration of illness) to reach ten final homogeneous cost subgroups. The other complications, including coma, peripheral revascularization, amputation and eye complications, did not enter the model. The training and the test sample yielded overlapping results. The risk coefficient was low (0.126, with a standard error of 0.001)

both in the training and in the test sample, denoting a good fit to the data. We report here the findings related to the test sample, which comprises 50% of cases (Fig. 1).

The empirical costs in euro in the final ten subgroups of the test sample ranged from a median of €483 to €39,578 (Table 4). The two subgroups with the highest cost (> €31,000), which comprise 0.3% of the population, consisted of dialysis patients, regardless of age and the presence of ischemic heart disease. Groups B, C, E, F and L, comprising 14% of cases, with median cost ranging between €1415 and €3303, included patients with ischemic heart disease, renal disease or stroke. Group I, which was the largest (57.9%), included subjects aged 65 years or over without complications, with a median cost of €1054. The last two groups, which together comprised about one fourth of the cases (27.8%), consisted of subjects aged less than 65 years and with a duration of illness ≥ 5 years or < 1 year (median cost €720) or a duration of illness between 1 and 4 years (median cost €483).

4 Discussion

In this cost-of-illness study we analyzed the annual direct medical costs of a population of patients with T2DM comprising over 70% elderly, a large majority (72.9%) with a duration of illness > 5 years, and 65.6% with two or more chronic conditions.

The mean annual medical cost per capita was €3086, consistent with the results of other studies conducted in Italy or in Europe [7, 8, 11, 23, 32]. However, the median annual medical cost (€1012) is a more accurate summary measure because the distribution of costs is skewed, with a few patients accounting for a large proportion of costs, in line with other studies [10, 12]. We confirmed that hospitalization accounts for the large majority of costs (53.1%) followed by drugs (29.5%) and outpatient services (17.4%) [7, 8, 11, 23].

Our results support the evidence [6, 10–12, 25, 33] that the most relevant cost drivers are age, complications and comorbidities. Duration of illness had an independent association with costs. In fact, we found that medical costs are higher in the first year of diagnosis. It is reasonable to assume that in the first year of illness, resource consumption and access to healthcare services are more relevant [14] for diagnosis and staging in order to identify the best therapeutic strategies. Notably, the area of residence was associated with different costs, after adjusting for patients' demographic and clinical characteristics, suggesting a role of organizational factors that warrants further investigation.

Through CART analysis we identified ten homogeneous subgroups of patients, five aged < 65 years and five aged 65 years or more. The large majority (85.7% of the study population) fell into three subgroups including elderly or younger patients without ischemic heart disease, renal complications, stroke and not on dialysis, with annual median costs < €1055. Five patient groups with higher costs, up to €3303, were characterized by the presence of at least one complication. Dialysis, which represents the end stage of chronic kidney disease, was the most expensive intervention and was delivered to a small minority of patients (0.3%) with annual costs amounting to more than €31,000.

These results suggest that preventive interventions are required to control the increasing prevalence of the disease, due to the aging population, and the rate of complications.

In this way, we could improve not only the health status of T2DM patients but also the amount of their medical costs [34, 35]. Because the natural process of ageing cannot be controlled, the focus of interventions should be mainly on correcting unhealthy behaviors related to poor diet or sedentary lifestyles to prevent both the onset of diabetes and its related complications [36, 37]. Concerning hospitalizations, which accounted for the majority (53.1%) of costs, a possible strategy would be to reduce those potentially preventable through more appropriate management of patients' follow-up [6] and of drug treatments. In particular, some studies highlighted that the use of some SGLT-2 inhibitors compared with placebo could lead to a significant reduction in the hospitalization for heart failure, which is one of the major causes of admission to hospital in patients with T2DM [38, 39]. Our results concerning the costs of chronic kidney disease are consistent with other reports showing that, starting from stage 3A, the financial burden of chronic kidney disease increases as renal function declines [40, 41]. This supports early identification and clinical interventions for chronic kidney disease in patients with diabetes to delay progression towards end-stage renal failure through optimal management strategies [42].

Given the budget constraints and the need to allocate healthcare resources in an equitable way, our stratification of T2DM patients into cost subgroups could inform treatment decisions and the allocation of related resources in a way that matches patients' needs. Our results are useful especially because of the growing importance of the budget allocation from the perspective of activity-based provider payment systems [43], bundle payments and value-based bundle payments [44].

The results of the present study should be interpreted in light of strengths and limitations. To the best of our knowledge this is the first study to stratify T2DM patients into cost subgroups based on demographic and clinical characteristics using a classification tree approach. This procedure allows us to define subgroups of patients with combinations of demographic and clinical characteristics, while in general, stratification is performed using only one criterion. For instance, another Italian study applied a cost-of-illness

Fig. 1 Classification and regression tree based on the testing sample that comprises 50% of the study population (*N* = 50,799). Boxes include the absolute and percentage frequency of subgroups and the histogram of the cost distribution. The final nodes are marked in red
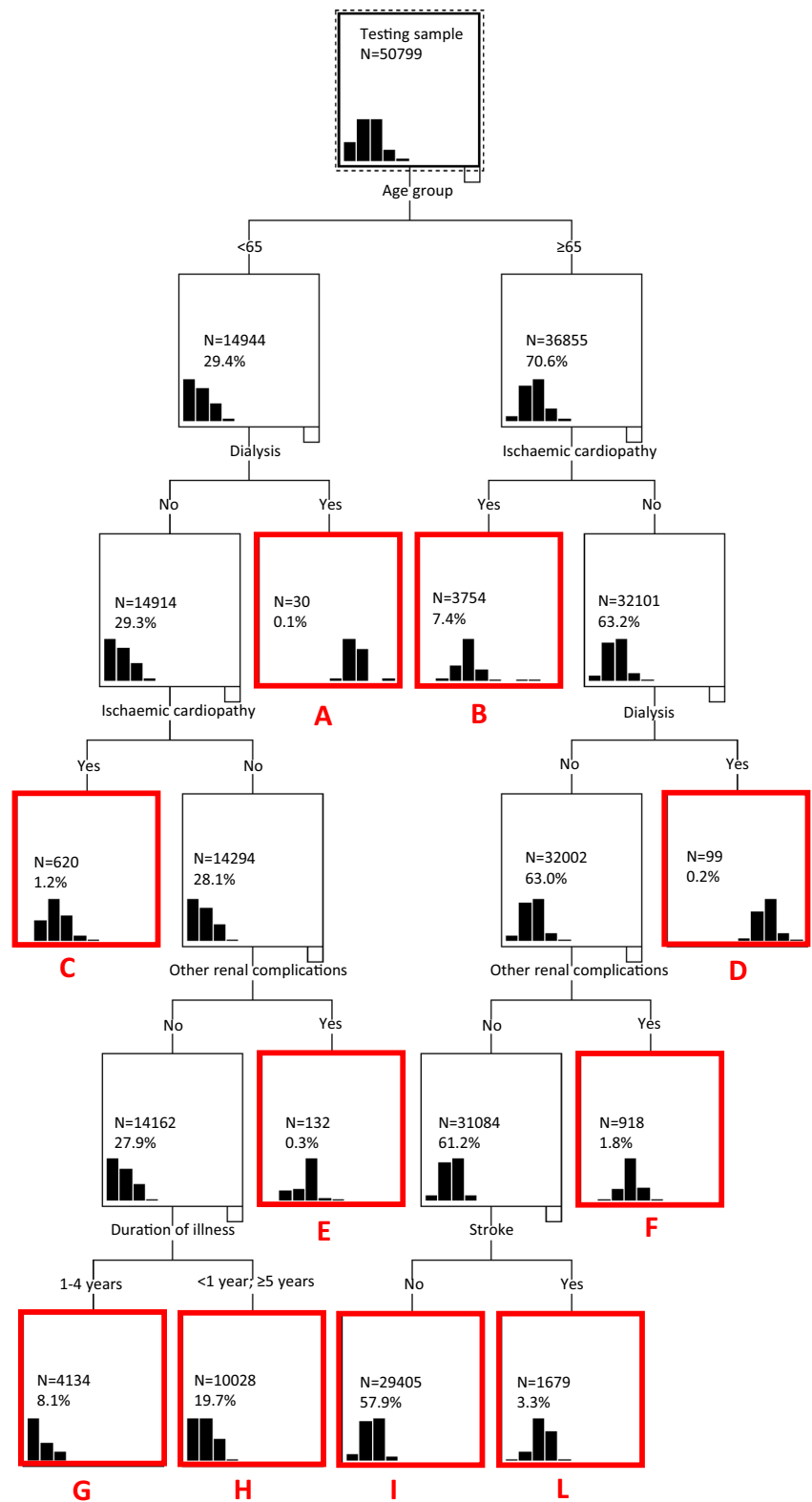
approach to administrative data and stratified the patient population according to the number of comorbidities [22]. Thus, our results can provide useful insights as regards the

implementation of healthcare policies and the organization of healthcare services.

Table 4 Median empirical costs of the 10 subgroups identified using classification and regression tree analysis. Subgroups are arranged in decreasing order of costs

		Patients in the subgroup		Age		Costs per patient (€)		
		N	%	Mean	SD	Median	95% CI lower limit	95% CI upper limit
D	≥ 65 years, dialysis, no ischemic cardiopathy	99	0.2	76.2	7.2	39,578.66	34,247.69	43,757.97
A	< 65 years, dialysis	30	0.1	56.2	5.7	31,490.53	22,852.99	52,325.34
E	< 65 years, renal complications, no ischemic cardiopathy	132	0.3	55.8	7.2	3303.67	2031.39	4605.30
F	≥ 65 years, renal complications, no dialysis, no ischemic cardiopathy	918	1.8	80.7	7.8	2465.96	2163.13	2930.63
B	≥ 65 years ischemic cardiopathy	3754	7.4	78.7	7.5	2461.89	2282.93	2647.08
L	≥ 65 years stroke	1679	3.3	80.8	7.6	1545.06	1441.40	1670.65
C	< 65 years ischemic cardiopathy no dialysis	620	1.2	58.6	5.3	1415.44	1303.13	1573.01
I	≥ 65 years no complications	29,405	57.9	77.0	7.5	1054.64	1040.68	1068.97
H	< 65 years, no complications, < 1 year or ≥ 5 years of diabetes	10,028	19.7	55.8	7.7	720.93	700.40	739.74
G	< 65 no complications, 1–4 years of diabetes	4134	8.1	53.2	9.1	483.75	461.82	502.18

CI confidence interval, *SD* standard deviation

In addition, the study is population-based, so it has no selection bias or external validity limitations. Moreover, it takes advantage of the linkage of individual data from high-quality administrative databases [45]. The inclusion criteria for this study have been validated in other studies [24, 46].

The implications of our study need to be interpreted also in light of some limitations. First, microcosting was not possible. In fact, we used tariffs that do not represent the real cost for the local health authority, but only the amount of money the LHA reimburses to providers. Second, in this study, we did not estimate indirect costs for patients and caregivers. In Italy, information on productivity losses and out-of-pocket expenditure is not available at the patient level, but more research should be conducted to acquire these data, as social costs associated with diabetes are substantial. Third, it is possible that people suffering from T2DM are untreated and therefore are not captured by administrative databases.

Our results can be generalized to the other Italian regions because people with diabetes are exempt from co-payment of health services across the country and costs of drugs and outpatient and inpatient care are set at the national level, with small regional variations related to local agreements. Moreover, the methodology is replicable in other countries where the availability of administrative healthcare databases allows the reconstruction of the care pathway of patients with T2DM.

In conclusion, this study provides policy makers and clinicians relevant information about the cost of T2DM care based on patients' demographic and clinical characteristics. This information, combined with available data on epidemiological trends, may support clinicians in implementing effective healthcare interventions (e.g., prevention of complications and associated comorbidities) and policymakers in designing resource allocation tailored to patients' needs.

Author Contributions PR drafted the article, gave substantial contributions to the conception and design of the study, performed statistical analyses and gave final approval of the version to be published. VMA drafted the article, gave substantial contributions to the conception of the study and data interpretation, and gave final approval of the version to be published. CT, CU, EB, MLM gave substantial contributions to data interpretation, revised critically the article for important intellectual content and gave final approval of the version to be published. MPF gave substantial contributions to the conception of the study, contributed to data interpretation, revised critically the article for important intellectual content and gave final approval of the version to be published.

Compliance with Ethical Standards

Conflict of interest Paola Rucci, Vera Maria Avaldi, Claudio Travaglini, Cristina Ugolini, Elena Berti, Maria Luisa Moro, and Maria Pia Fantini declare that they have no conflict of interest in relation to the present work.

Ethical Approval The retrospective study was carried out in conformity with the regulations on data management with the Italian law on privacy (Legislation Decree 196/2003 amended by Legislation Decree 101/2018). Data were anonymized prior to the analysis at the regional statistical office, where each patient is assigned a unique identifier. This identifier does not allow others to trace the patient's identity and other sensitive data. Anonymized regional administrative data can be used without a specific written informed consent when patient information is collected for healthcare management and healthcare quality evaluation and improvement (according to art. 110 on medical and biomedical and epidemiological research, Legislation Decree 101/2018).

References

1. 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):1513–30.
2. Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration. Cardiovascular disease, chronic kidney disease, and diabetes mortality burden of cardiometabolic risk factors from 1980 to 2010: a comparative risk assessment. Lancet Diabetes Endocrinol. 2014;2:63–447.
3. Seuring T, Archangelidi O, Suhrcke M. The economic costs of type 2 diabetes: a global systematic review. Pharmacoeconomics. 2015;33(8):811–31.
4. Cho NH, Shaw JE, Karuranga S, Huang Y, da Rocha Fernandes JD, Ohlrogge AW, Malanda B. IDF Diabetes Atlas: global estimates of diabetes prevalence for 2017 and projections for 2045. Diabetes Res Clin Pract. 2018;138:271–81.
5. Hu FB. Globalization of diabetes: the role of diet, lifestyle, and genes. Diabetes Care. 2011;34(6):1249–57.
6. American Diabetes Association. Standards of care in diabetes-2018. Diabetes Care. 2018;41(suppl 1):S1–159.
7. ARNO observatory diabetes. The profile of the care population with diabetes. Report 2017 Volume XXX—Series ''Reports ARNO'' Cineca 2017.
8. Nuño-Solinís R, Alonso-Morán E, Arteagoitia Axpe JM, Ezkurra Loiola P, Orueta JF, Gaztambide S. Healthcare costs of people with type 2 diabetes mellitus in the Basque Country (Spain). Endocrinol. Nutr. 2016;63(10):543–50.
9. World Health Organization. Global report on diabetes. 2016. http://www.who.int/diabetes/global-report/en/. Accessed Mar 2019.
10. Bron M, Guerin A, Latremouille-Viau D, Ionescu-Ittu R, Viswanathan P, Lopez C, Wu EQ. Distribution and drivers of costs in type 2 diabetes mellitus treated with oral hypoglycemic agents: a retrospective claims data analysis. J Med Econ. 2014;17(9):646–57.
11. Kähm K, Laxy M, Schneider U, Rogowski WH, Lhachimi SK, Holle R. Health care costs associated with incident complications in patients with type 2 diabetes in Germany. Diabetes Care. 2018;41(5):971–8.
12. de Lagasnerie G, Aguadé AS, Denis P, Fagot-Campagna A, Gastaldi-Menager C. The economic burden of diabetes to French national health insurance: a new cost-of-illness method based on a combined medicalized and incremental approach. Eur J Health Econ. 2018;19(2):189–201.
13. Sørensen M, Arneberg F, Line TM, Berg TJ. Cost of diabetes in Norway 2011. Diabetes Res Clin Pract. 2016;122:124–32.
14. Rosella LC, Lebenbaum M, Fitzpatrick T, O'Reilly D, Wang J, Booth GL, Stukel TA, Wodchis WP. Impact of diabetes on healthcare costs in a population-based cohort: a cost analysis. Diabet Med. 2016;33(3):395–403.
15. Marcellusi A, Viti R, Sciattella P, Aimaretti G, De Cosmo S, Provenzano V, Tonolo G, Mennini FS. Economic aspects in the management of diabetes in Italy. BMJ Open Diabetes Res Care. 2016;4(1):e000197.
16. Bommer C, Heesemann E, Sagalova V, Manne-Goehler J, Atun R, Bärnighausen T, Vollmer S. The global economic burden of diabetes in adults aged 20-79 years: a cost-of-illness study. Lancet Diabetes Endocrinol. 2017;5(6):423–30.
17. Grimaccia F, Kanavos P. Cost, outcomes, treatment pathways and challenges for diabetes care in Italy. Global Health. 2014;10:58.
18. Arredondo A, Zúñiga A, Parada I. Health care costs and financial consequences of epidemiological changes in chronic diseases in Latin America: evidence from Mexico. Public Health. 2005;119(8):711–20.
19. Condliffe S, Link CR, Parasuraman S, Pollack MF. The effects of hypertension and obesity on total health-care expenditures of diabetes patients in the United States. Appl Econ Lett. 2013;20:649–52.
20. Khowaja LA, Khuwaja AK, Cosgrove P. Cost of diabetes care in out-patient clinics of Karachi, Pakistan. BMC Health Serv Res. 2007;7:189.
21. Barceló A, Aedo C, Rajpathak S, Robles S. The cost of diabetes in Latin America and the Caribbean. Bull World Health Organ. 2003;81(1):19–27.
22. Marcellusi A, Viti R, Mecozzi A, Mennini FS. The direct and indirect cost of diabetes in Italy: a prevalence probabilistic approach. Eur J Health Econ. 2016;17(2):139–47.
23. Scalone L, Cesana G, Furneri G, Ciampichini R, Beck-Peccoz P, Chiodini V, Mangioni S, Orsi E, Fornari C, Mantovani LG. Burden of diabetes mellitus estimated with a longitudinal population-based study using administrative databases. PLoS ONE. 2014;9(12):e113741.
24. Nobilio L, Leucci AC, Ugolini C, Avaldi V, Berti E, Moro ML. Dossier n. 264/2018—Profili di assistenza e costi del diabete in Emilia-Romagna. Analisi empirica attraverso dati amministrativi 2010–2016. Bologna (2018). http://assr.regione.emilia-romagna.it/it/servizi/pubblicazioni/dossier/doss264. Accessed Mar 2019.

25. Lin PJ, Pope E, Zhou FL. Comorbidity type and health care costs in type 2 diabetes: a retrospective claims database analysis. Diabetes Ther. 2018. https://doi.org/10.1007/s13300-018-0477-2.

26. Ng CS, Lee JY, Toh MP, Ko Y. Cost-of-illness studies of diabetes mellitus: a systematic review. Diabetes Res Clin Pract. 2014;105(2):151–63.

27. Struijs JN, Baan CA, Schellevis FG, Westert GP, van den Bos GAM. Comorbidity in patients with diabetes mellitus: impact on medical health care utilization. BMC Health Serv Res. 2006;6:84.

28. Sancho-Mestre C, Vivas-Consuelo D, Alvis-Estrada L, Romero M, Usó-Talamantes R, Caballer-Tarazona V. Pharmaceutical cost and multimorbidity with type 2 diabetes mellitus using electronic health record data. BMC Health Serv Res. 2016;16(1):394.

29. Maio V, Yuen E, Rabinowitz C, Louis D, Jimbo M, Donatini A, Mall S, Taroni F. Using pharmacy data to identify those with chronic conditions in Emilia Romagna, Italy. J Health Serv Res Policy. 2005;10(4):232–8.

30. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and regression trees. Pacific Grove: Wadsworth; 1984.

31. Henrard S, Speybroeck N, Hermans C. Classification and regression tree analysis vs. multivariable linear and logistic regression methods as statistical tools for studying haemophilia. Haemophilia. 2015;21(6):715–22.

32. Mata-Cases M, Casajuana M, Franch-Nadal J, Casellas A, Castell C, Vinagre I, Mauricio D, Bolíbar B. Direct medical costs attributable to type 2 diabetes mellitus: a population-based study in Catalonia, Spain. Eur J Health Econ. 2016;17(8):1001–10.

33. Clarke PM, Glasziou P, Patel A, Chalmers J, Woodward M, Harrap SB, Salomon JA. ADVANCE Collaborative Group: Event rates, hospital utilization, and costs associated with major complications of diabetes: a multicountry comparative analysis. PLoS Med. 2010;7(2):e1000236.

34. Herman WH, Edelstein SL, Ratner RE, Montez MG, Ackermann RT, Orchard TJ, Foulkes MA, Zhang P, Saudek CD, Brown MB. Diabetes Prevention Program Research Group: effectiveness and cost-effectiveness of diabetes prevention among adherent participants. Am J Manag Care. 2013;19(3):194–202.

35. Diabetes Prevention Program Research Group. The 10-year cost-effectiveness of lifestyle intervention or metformin for diabetes prevention: an intent-to-treat analysis of the DPP/DPPOS. Diabetes Care. 2012;35:723–30.

36. Harwell TS, Vanderwood KK, Hall TO, Butcher MK, Helgerson SD. Montana Cardiovascular Disease and Diabetes Prevention Workgroup: factors associated with achieving a weight loss goal among participants in an adapted Diabetes Prevention Program. Prim Care Diabetes. 2011;5:125–9.

37. Florez H, Pan Q, Ackermann RT, Marrero DG, Barrett-Connor E, Delahanty L, Kriska A, Saudek CD, Goldberg RB, Rubin RR. Diabetes Prevention Program Research Group: impact of lifestyle intervention and metformin on health-related quality of life: the diabetes prevention program randomized trial. J Gen Intern Med. 2012;27(12):1594–601.

38. Zinman B, Wanner C, Lachin JM, Fitchett D, Bluhmki E, Hantel S, Mattheus M, Devins T, Johansen OE, Woerle HJ, Broedl UC, Inzucchi SE. EMPA-REG OUTCOME Investigators: empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med. 2015;373(22):2117–28.

39. Fitchett D, Butler J, van de Borne P, Zinman B, Lachin JM, Wanner C, Woerle HJ, Hantel S, George JT, Johansen OE, Inzucchi SE. EMPA-REG OUTCOME® trial investigators: effects of empagliflozin on risk for cardiovascular death and heart failure hospitalization across the spectrum of heart failure risk in the EMPA-REG OUTCOME® trial. Eur Heart J. 2018;39(5):363–70.

40. Yeaw J, Halinan S, Hines D, Delozier A, Perez M, Boye M, Boye KS, Blanchette CM. Direct medical costs for complications among children and adults with diabetes in the US commercial payer setting. Appl Health Econ Health Policy. 2014;12(2):219–30.

41. McQueen RB, Farahbakhshian S, Bell KF, Nair KV, Saseen JJ. Economic burden of comorbid chronic kidney disease and diabetes. J Med Econ. 2017;20(6):585–91.

42. Santoro A, Gibertoni D, Rucci P, Mancini E, Bonucchi D, Buscaroli A, Campagna A, Cappelli G, David S, Gregorini MC, La Manna G, Mosconi G, Rigotti A, Scarpioni R, Storari A, Mandreoli M. The PIRP project (Prevenzione Insufficienza Renale Progressiva): how to integrate hospital and community maintenance treatment for chronic kidney disease. J Nephrol. 2019.

43. Moran V, Jacobs R. Investigating the relationship between costs and outcomes for English mental health providers: a bivariate multi-level regression analysis. Eur. J. Health Econ. 2018;19(5):709–18.

44. Porter, M.E., Kaplan, R.S.: How should we pay for health care? HBS Working Paper Series. Harvard Business School. 2015. http://hbswk.hbs.edu/item/how-should-we-pay-for-health-care. Accessed Mar 2019.

45. Stivanello E, Rucci P, Carretta E, Pieri G, Seghieri C, Nuti S, Declercq E, Taglioni M, Fantini MP. Risk adjustment for inter-hospital comparison of caesarean delivery rates in low-risk deliveries. PLoS ONE. 2011;6(11):e28060.

46. Gini R, Francesconi P, Mazzaglia G, Cricelli I, Pasqua A, Gallina P, Brugaletta S, Donato D, Donatini A, Marini A, Zocchetti C, Cricelli C, Damiani G, Bellentani M, Sturkenboom MC, Schuemie MJ. Chronic disease prevalence from Italian administrative databases in the VALORE project: a validation through comparison of population estimates with general practice databases and national survey. BMC Public Health. 2013;13:15.