

Prediction of Medical Costs in a Health Insurance Carrier according to Risk Profiles and Uses by its Affiliates

Predicción del Costo Médico de una Empresa Administradora de Planes de Beneficio en Salud de acuerdo a los Perfiles de Riesgo y Uso de sus Afiliados

Rodolfo Herrera Medina^{1,3}, Jaime Lee^{2,3}, Ferney Herrera Cruz³

¹ Universidad Libre, Grupo de Investigación en Epidemiología y Servicios GRIEPIS and EPICRITICA. Cali, Colombia.

² Universidad del Valle. Facultad de Salud, Departamento de Medicina Familiar. Cali, Colombia.

³ COOMEVA group, Cali-Colombia.

Correspondence to: rodolfo-herreram@unilibre.edu.co

Received: July 15, 2018

Accepted: September 27, 2018

Keywords: Health care costs, Health Programs and Plans, Two-part model, Generalized linear model.

Palabras clave: Costos de la Atención en Salud, Planes y Programas de Salud, modelo a dos partes, modelos lineales generalizados

Cite this as: Herrera MR, Lee J, Herrera CF. Prediction of Medical Costs in a Health Insurance Carrier according to Risk Profiles and Uses by its Affiliates. Colombia. *ijepH*. 2018; 1(1): e0004. Doi: [10.18041/2665-427X/ijepH.1.3872](https://doi.org/10.18041/2665-427X/ijepH.1.3872)

Abstract

Objective: To find a model of prediction of the medical cost of a Health Benefits Management Company (EAPB) with adequate statistical criteria.

Methods: A Cross-sectional study with retrospective follow-up of the use of health services in an EAPB during a one-year period. The sampling frame consisted of a population of 1,529,188 affiliates who were assigned to a primary care IPS group. By simple random sampling size was estimated at 190,917 users. The dependent variable was the cost of the services used deflated to the year 2013. As independent variables besides the traditional sociodemographic variables chosen in this type of prediction models, variables of the insurance were added; Variables of risk management (inclusion or not in promotion and prevention program) and of comorbidities.

Results: Simple Linear Regression modeling showed errors of inappropriate statistical criteria such as violating the principle of normality in cost errors. The Generalized Linear Models, proposed to estimate POS average costs, have an appropriate goodness of fit and evaluated with small Deviations and minimum Akaike criterion (AIC) compared to other models of the exponential family

Conclusions: The appropriate statistical model to predict medical costs was the Generalized Linear Model with two parts segmented by age groups and gender. This research suggests that to estimate the benefit premium of any EAPB, besides socio-demographic variables, insurance variables, membership or not in promotion programs and risk prevention and/or management and the burden of disease of that population should be used.

Resumen

Objetivo: Diseñar un modelo con criterios estadísticos adecuados para la predicción del costo médico de una Empresa Administradora de Planes de Beneficio (EAPB) en Salud.

Métodos: Estudio de corte transversal con seguimiento retrospectivo de la utilización de los servicios de salud en una EAPB durante un periodo de un año. El marco muestral lo constituyó una población de 1.529.188 afiliados que estaban asignados a un grupo de IPS de atención primaria. Por muestreo aleatorio simple, se estimó el tamaño en 190,917 usuarios. La variable dependiente fue el costo de los servicios utilizados deflactados al año 2014. Como variables independientes, además de las tradicionales variables sociodemográficas escogidas en estos tipos de modelos de predicción, se agregaron variables del aseguramiento; variables de gestión de riesgo (inclusión o no en programa de promoción y prevención) y de comorbididades.

Resultados: La modelación con Regresión Lineal Simple mostró errores de criterios del modelo inapropiados como el de violar el principio de la normalidad. Los Modelos Lineales Generalizados, propuestos para estimar los costos medios POS, tienen bondad de ajuste apropiadas y evaluadas con Devianzas pequeñas y criterio Akaike (AIC) mínimo comparado con otros modelos de la familia exponencial.

Conclusiones: El modelo estadístico apropiado para predecir costos médicos fue el Modelo Lineal Generalizado a dos partes segmentados por grupos de edad y género. La presente investigación sugiere que para estimar la prima de beneficios de cualquier EAPB se utilicen además de las variables sociodemográficas, variables de aseguramiento, la pertenencia o no a programas de promoción y Prevención y/o gestión de riesgo y la carga de enfermedad de esa población.

Key Study Facts

Objective	To find a model of prediction of the medical cost of a Health Benefits Management Company (EAPB) with adequate statistical criteria.
Study design	Cross-sectional with retrospective follow-up
Source of data	Obligatory Health Plan health services of affiliates to an EPS during a one-year period
Population/Sample	1,529,188 affiliates who were assigned to a primary care IPS group. By simple random sampling size was estimated at 190,917 users
Statistical analysis	Generalized Linear Regression modeling The dependent variable was the cost of the services used deflated to the year 2013. As independent variables besides the traditional sociodemographic variables chosen in this type of prediction models, variables of the insurance were added; Variables of risk management (inclusion or not in promotion and prevention program) and of comorbidities
Main finding	To estimate the benefit premium of any EAPB, besides socio-demographic variables, insurance variables, membership or not in promotion programs and risk prevention and/or management and the burden of disease of that population should be used



Introduction

Estimating medical costs has been a concern in countries where services are offered through a mean fee. However, many of the estimations have been imprecise. Some countries, like Peru, have undertaken methodologies to try to estimate the standard costs of the medical procedures that will then be incorporated into the Universal Health Benefits Plan (1).

In Colombia, estimates have been made since 1993 of the medical cost that will be incurred during the following year by a population ensured with the General Social Security System (SGSS, for the term in Spanish), an estimation denominated: "ex ante approximation", and legislatively called Capitation Payment Unit "CPU", articles 156 and 179 of Legislation 100 (2) and with which the health promoter entities (EPS or HMO) must manage the offer of the health services hiring the network that provides these services to the population ensured, article 205 of Legislation 100 (2). The Ministry of Health traditionally, since 1994 to date, has calculated this fee by using simple linear regression models (SLR).

This work proposed the search for cost estimation models that include independent variables, like the age and gender of the affiliate, as has usually been done by the Colombian Ministry of Health, as well as the inclusion of variables associated with the health-disease status of each affiliate cared for. Likewise, it sought to propose modelling techniques different from the traditional, like the SLR models to estimate costs, and included other models until achieving an adequate one with satisfactory performance in the valuation of statistical goodness of fit indicators, like the case of generalized linear models (GLM). This type of model measures goodness of fit with a couple of indicators: the Deviance indicator, which may be understood as a measurement of the lack of information available in the model to explain the phenomenon of interest in the response variable (the smaller the better), and the Akaike Information Criterion (AIC), an indicator that in its calculation involves the Deviance penalized by the number of parameters involved in the model (the smaller the better). The first model estimated, SLR, did not fulfill the basic validation assumptions, like normality and homoscedasticity in the dependent variable, additionally, no adequate goodness of fit indicators were obtained, given that the R^2 at most reached 6%. Thereafter, more robust estimation models were proposed, like the GLM applied in modelling two parts (estimating in the first part the probability of using the health services, and in the second part estimating the costs implied in said use), which has found considerable acceptance in the methodology of evaluating risks, thus, obtaining a general expenditure prediction for an individual, upon multiplying both predictions obtained from each part of the model (3,4).

This work sought to find an ex ante prediction model that, besides estimating the POS medical cost, complies with more robust or adequate statistical criteria to the traditional, as well as a search for predictive methodologies of the medical costs that can serve to monitor the costs of health services required by the users of a health insurer without having the prior invoicing and only having demographic factors, variables appertaining to the insurance business, and the user risk classification variable.

Methods

Epidemiological design

This was a cross-sectional study with a retrospective follow up of the POS health services of affiliates to an EPS during a one-year period.

Probabilistic models were used to analyze the estimate of medical costs defined in the Obligatory Health Plan (POS, for the term in Spanish) (5). A probabilistic sample was extracted from several geographic- administrative regions of an EPS from the contributive regime with which the model's parameters were tested and, according to their health risk characteristics costs were estimated by regions and age range.

Sample design

The simple framework was the population cared for as a primary-care IPS and who are managed by the same EPS to which they belong. This guaranteed the quality of the diagnoses, and the totality of the orders to provide medical services.

Inclusion and exclusion criteria

The study included affiliates who were compensated, see article 205 of Legislation 100 of 1993 (1), at some moment for the period comprised between 01 October 2013 and 30 September 2014. The work excluded those who did not have complete information or who had costs registered but not invoiced, or with evidence of the service not having been provided.

Calculation of simple size

The type of sampling used was stratified with proportional assignment through six geographic regions established to manage the affiliate population. The global sample was estimated at 190,917 people whose simple framework was constituted by 1,529,188 users affiliated to the IPS which guaranteed the data on authorizations, clinical histories of the services provided and the costs of invoicing with transactional systems arranged by the EPS under its technological platform. Using for its estimation a standard deviation of the cost, according to historical data, at cop \$ 5,900,000, 95% CI and a precision of cop\$ 31,000 with an effect of the design of 1.50.

Statistical design

As dependent variable, it was available from the POS cost in Colombian pesos (cop\$). The costs of 2013 were deflated to costs of 2014. Every user will have as constant monthly cost the cost paid according to hiring modality with capita, according to the CPU.

There were three types of independent variables: a. demographic factors; insurance factors: new in the EPS (less than six months affiliated), wage range ($1 \leq 2$ SMLV; $2 =$ between 2 and 4 SMLV; $3 \leq 4$ SMLV), type of affiliate ($1 =$ Main contributor, $2 =$ Beneficiary, $3 =$ Additional or Secondary contributor), type of contributor ($1 =$ Pensioner, $2 =$ Independent, $3 =$ dependent); c. factors of the offer: affiliates reported in Resolution 4505 by the Ministry of Health (6), and with management of the health risk, users reported in the High-Cost Account - HCA (7), complementary plan (affiliated to a voluntary insurance plan or

pre-paid medical care), poly-consultant (having seven or more external consultations per year); d. factors of disease burden: classification of risks according to diagnoses from the CIE- 10 (8) in nine groups (Low prevalence and high cost, oncological (excludes those identified in HCA as patients with cancer), chronic obstructive pulmonary disease (COPD), mental disease, maternal perinatal, cardiocerebral-renovascular (CCRV), chronic disease, acute disease, avoidable disease. The classification of Diagnostic Relational Groups (DRG) classifies users into one or several of the groups defined according to diagnoses registered during any contact with the health services for a given year.

Modelling began with an SLR and, thereafter, to a multiple linear regression with regressors, like age and gender. The regressor variables (independent variables) were included, with the "POS cost" being the dependent variable or response variable with the following equation:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \epsilon \quad [1]$$

Compliance was verified of the properties of homogeneity and normality in errors. However, due to prior and descriptive knowledge on the asymmetry in the distribution of costs (Figure 1), the large amount of zeros (affiliates who incur no costs) and the long tails in the distribution, the asymmetry in the distribution of costs, lead to proposing models that explain this behavior (3), herein emerges the concept of two-part models (3,9). The final model is the product of both models, which conjugates frequencies and mean costs. Where the first part estimates the probability of an individual requesting a service, given their sociodemographic factors and disease burden where logistic regression models were used, one per segment of age and gender.

The second part estimates the value of the expected cost of the service given that a request for such is made (users), according to the sociodemographic factors of the affiliates and their disease burden; this last part of the model was conducted through generalized linear models (10-12). This second part of the model will be used to determine factors associated to costs. Although cross-sectional studies, the principal objective is to estimate the frequency of the health event; furthermore, the OR can be

calculated as if it were a cases and controls study (13). Cross-sectional studies have been considered by some authors as cases and controls studies in which it is assumed that the time between the exposure and onset of the disease is null (14)

These models were evaluated according to the goodness of fit indicators that fulfill the statistical conditions required, so that interpretative value is added to the information and robustness upon changes of population structure, as well as with the sufficient financial conditions for a satisfactory application in the EPS.

Explanation of the GLM with Gamma and logarithmic link

$$fT(t) = \frac{r^r t^{r-1} e^{-\frac{t}{\mu}}}{\Gamma(r)\mu^r}, \text{ for } t, r > 0 \quad [2]$$

Results

Behavior of costs

To model the cost of service, a database was available with 190,693 records of affiliates to the EPS from which 63,000 had no costs associated with health services (33%), 67% of the users had annual costs, only POS services, ranging from cop\$ 26,000 to cop\$ 284,249,547, which shows a highly variable behavior with a positive strongly asymmetric distribution, as noted in Figure 1.

The difference of the mean cost between men and women is not statistically significant ($p = 0.066$), but significant differences are noted by gender for groups from 15 to 18 years of age and from 19 to 44 years of age ($p = 0.000$) (Figure 2). Additionally, it is observed how the mean cost, as well as its variance increases as the age of the population group increases.

In spite of the asymmetry of the costs, the simple linear regression model was applied with age as independent variable, giving the following results: Standard deviation = 3.238.940; $R^2 = 16.4\%$; R^2 (adj) = 16.3%; and the errors are not distributed normally (Figure 3A). These results contradict the use of these types of statistical models when the dependent variable is the health costs. From the aforementioned, the need emerged to search for other statistical models that permit statistical validity to any

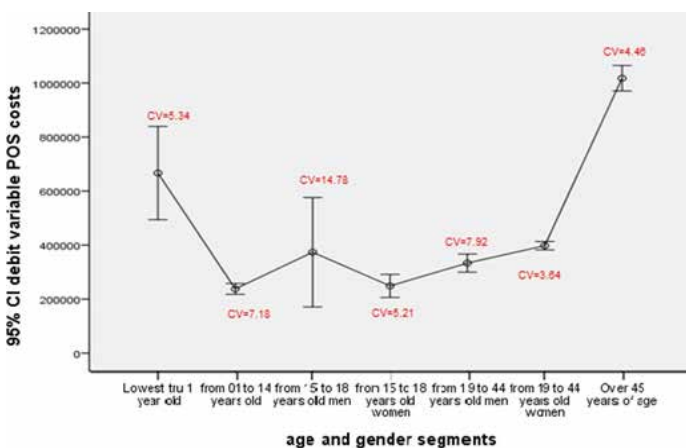


Figure 1. POS cost slanted to the left

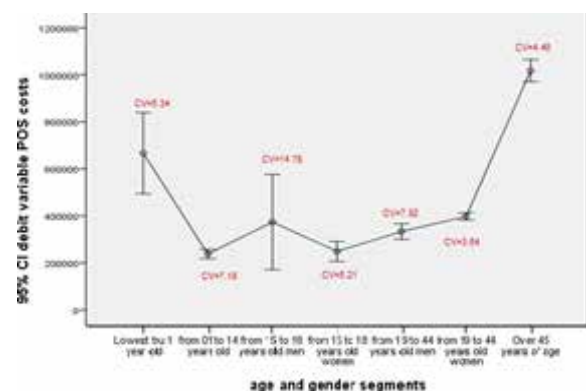


Figure 2. Differential behavior of POS costs, summarized through variation coefficients (VC) by age and gender segments

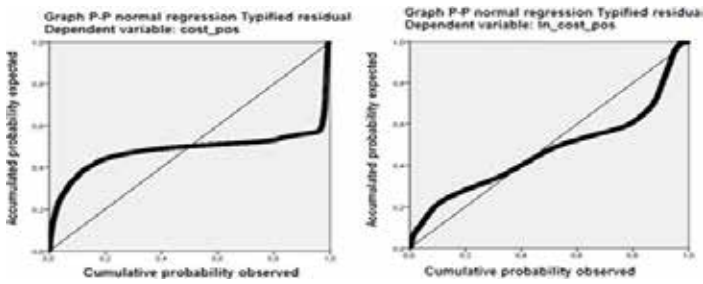


Figure 3. Graphic of residual normality in the simple regression model for the POS cost (A), as well as for the logarithmically transformed model (B).

study where costs behave asymmetrically, as is the case and as is manifested by Mihaylova (9) in the form of “inconsistent or inefficient results” (15) under this analysis category of the costs of health care.

The behavior described led to testing some transformations on the dependent variable to try to stabilize the variance. After diverse transformation explorations, the logarithmic transformation was chosen, which helps to smooth the variance of the residuals. The multiple regression model was then used with all the variables, eliminating those that were highly correlated, and those that did not contribute greatly to the model, and considering as dependent variable the Natural Logarithm of the POS cost. The R^2 value indicated that the predictors explain 68.91% (68.87% adjusted) of the variance of the POS costs, which suggests that the model is adjusted “moderately” to the data. Other results seeking the transformation of the dependent variable: analyzing the diagram of normal probability of the residuals for the model transformed logarithmically (Figure 3B), it is noted that “tail” still persist, although these are now thin. This shows that the principal premises of normality and constant variance in the errors are also not fulfilled. This led to seeking alternative models that permit better adjusting the data. According to the distribution, these suggest adjusting the exponential family, a necessary condition to explore generalized linear models.

“Statistical” results of the generalized linear model. Upon observing Table 1, from the previous model from formula [2], it presents for its estimations acceptable goodness-of-fit indicators, small deviances, and minor Akaike Information Criterion (AIC), compared with the null model. However, we also have a deterministic indicator of 105% financial effectivity, that is, the model overestimates the cost observed by 5%.

Observing Figure 4 on residuals, it may be determined that there is a high percentage of affiliates to whom the model proposed overestimates the cost above cop\$500,000.

Table 1. Goodness-of-fit indicators for GLM models, considered globally for age with aggregation of factors.

Model name	Parameter	Novelty	Deviance	AIC	DF
Null model	0	It’s the null model	508,977.51	3,457,355.56	127,692.00
Quantitative with regions	14	The regions are included	257,628.52	3,345,237.08	127,658.00
Basics	8	It is the model with age, sex, regions	458,974.52	3,439,650.19	127,684.00
Basics + DRG	24	It is model with age, sex, DRG risk, more risk interactions	347,014.16	3,393,017.65	127,668.00
Qualitative with regions	34	The regions are included	295,191.77	3,366,847.28	127,658.00

AIC: Akaike Criterion, DF: Degrees of freedom; Null_model = Null model, DRG = Diagnostic Risk Groups.

Final results applicable to the health model

By analyzing the cost against the age variable, it was possible to identify a differential behavior when grouping the data by age ranges and by gender; this suggests a segmentation of the models, observed Figure 2, where the segments are constructed according to the variation coefficients estimated for each of them, whose behavior is quite similar or homogeneous.

Due to the high variability of costs by age groups, it was necessary to elaborate the segments according to the variation coefficients, which permitted improving the goodness of fit of the models by segments, compared with the AIC of the global model. Besides, significant improvement was observed in the effectivity of the mean cost (non- statistical indicator), which is 14.5% above that observed, with data from a subsequent period; estimating 95% CI of 2.5% and 37.9%, as average effectivity of the model with respect to the real cost.

In Table 2, and applying the antilogarithm to the coefficients of the model part one, whose practical interpretation may be assimilated to an association measurement, like the OR between the independent variables and costs per segments of age and gender, the GLM part two does not have a fit or adjustment test, which is why residual analysis is suggested (Figure 4).

Reference category as co-variable to explain the OR: for the qualitative variables, the reference category was coded as “1” in the database (last category) and in the table, it is described in parenthesis when it is not explicit in the name of the variable, as it is for risks where the reference category is having the risk.

Predictors increasing the medical cost of services are those that at the end of 12 months caused some variable cost due to the use of the service, showing in Table 2 those factors impacting positively upon the cost, like being a patient listed in the High-Cost Account (7) HCA (VIH, CKD, rare, cancer) increased the cost between 4.9 and 6.5 times, compared with the non-HCA affiliates >19 years of age, and 2.0 times for those from 1 to 14 years of age. Having been classified with oncological risk increased the cost between 2.4 and 4.7 times for those from 1 to 14 years of age or who are >19 years of age, with the cost being higher for men from 15 to 18 years of age, with a weight of 31.6 times (here we basically find patients with leukemia and non-Hodgkin lymphoma). Being poly-consultant increased the cost by 10.6 times for those <1 year of age, by 7.0 times for men from 19 to 44 years of age, and by 5.1 times for those >45 years of age. The cost is higher in the Andean and Northeast regions, compared with the Pacific and Southern regions, for men between 15 and 18 years of age by 1.6 and 1.3 times, respectively.

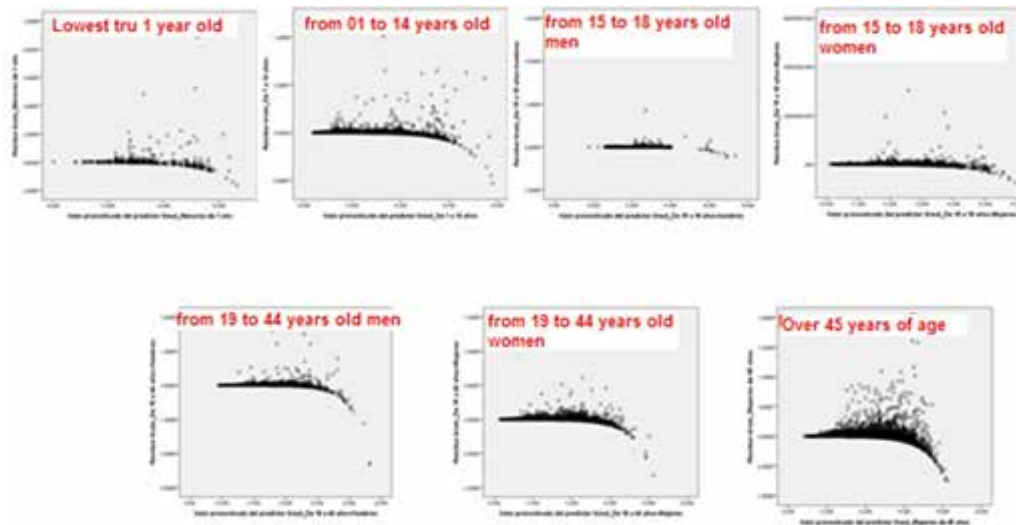


Figure 4. Model part 2. Residual vs. estimated values of the linear predictor by segments of age and gender.

Factors that diminish the cost, shown in Table 2 in green: being a woman between 1 and 14 years of age diminishes the cost by 17%, and for those >45 years of age on average diminishes by 13%, compared with the cost of being a man for the respective group. Being new in the EPS, <1 year of age, from 19 to 44 years of age, and >45 years of age, age may be defined as a factor that favors the cost for beneficiaries from 1 to 18 years of age, given that this diminishes between 3% and up to 16%, but end up impacting the cost unfavorably in those >19 years of age. Another behavior highlighted in the reduction of the cost is that for the segments of affiliates >19 years of age, the medical cost reduces on average by 20% compared with the costs from the Pacific and Southern zone (departments from the Colombian Pacific Coast).

Factors with interaction

Being listed in the HCA and having a low- prevalence and high-cost risk diminishes costs in half; it must be considered that, separately, these pathologies increase costs in those >45 years of age by 4.9 and 2.1 times, respectively. This same interaction effect occurs for patients listed with identification of health risks, according to that disposed in Resolution 4505 (6), that is, they are being intervened, once the risks are identified, in some health promotion and prevention program and, in turn, have been classified as poly-consultants; this junction diminishes costs by 30% for women from 19 to 44 years of age and for those >45 years of age. Again, it was observed that these factors, in separate and in added manner, increased costs by 1.4 and 5.1 times, respectively. Thus, we have that in general where the interaction occurred a cost diminishing effect emerges; however, the opposite occurs for patients with COPD and cardio-cerebrovascular risk, where a multiplicative effect was achieved, that is, besides the added increase of each of the pathologies, their interaction in a patient generates medical costs increase by 4.2 times.

Discussion

This research contributes to and enhances the theory de

the models to explain costs in health insurance, offers “methodological changes” gathering different categories of analytic approaches, like that expressed in the work on the strategy to estimate costs in EuroHOPE (16). The design of policies focused on implementing the use of these models is recommended in the insurance carriers and state entities.

The Ministry of Health traditionally, since 1994 to date, has calculated this fee by using SLR models, which reach R^2 determination coefficients that range from 2% to 12% of the medical care cost, besides, does not fulfill the statistical criteria of model fit, that is, the regressor variables do not contribute significantly in the explanation of the cost. The situation of estimating a CPU with such a low determination coefficient and with so few variables included from a health – multi-systemic disease model, like that proposed by Lalonde (17), gives rise to underestimated or overestimated cost values with the subsequent economic implications and of viability of the model of health insurance in Colombia.

In specific modelling exercises to estimate medical costs, it is noted that the linear regression models underestimate medical costs and do not fulfill statistic criteria to estimate from these; this is the case of the CPU calculation, or payment fee to the EPS for the health expense. Due to this, the major strength in this study is the application of generalized linear models segmented by age groups, given that these shows better goodness of fit to the estimation of the costs, according to a user’s average accident rate.

Guaranteeing the EPS’s solvency is fundamental for each of the SGGSS sectors (15) because these entities are in charge of managing health risk. This is one of the intrinsic premises of most studies on the economy of health conducted in Colombia on its SGGSS. Many of these studies segment the population into specific risk groups with a combination of gender, age group, geographic location, and diagnostic group. Including the diagnosis permits estimating more precisely the distribution of costs of the EPS (18). The results of applying the model in two

Table 2. Significant parameters of the Model according to Segments (the antilogarithm was applied to the coefficients of the original model).

Components	<1 year	From 1 to 14 years	15 to 18 Male	15 to 18 Female	19 to 44 Male	19 to 44 Female	>45
Intercept	\$ 422,487	\$ 85,570	\$ 1,414,096	\$ 868,184	\$ 30,437	\$ 34,898	\$ 30,498
Gender (Fem)	1.0	0.83	1.0	1.0	1.0	1.0	0.87
New (< 6 months)	0.59	1.0	1.0	1.0	0.51	1.0	0.72
High-cost patient	1.0	2.0	1.0	1.0	6.5	5.6	4.9
Registry Res.4505	2.8	1.2	1.0	1.0	1.0	1.1	1.4
Low-prevalence risk and High cost	1.0	3.6	1.0	1.0	4.0	1.9	2.1
Oncological risk	1.0	4.7	31.6	1.0	5.0	2.8	2.4
COPD	1.0	7.1	0.09	1.0	3.9	1.0	1.5
Mental risk	1.0	1.4	1.0	1.0	2.3	2.0	2.4
Maternal Perinatal	1.0	5.3	1.0	4.4	1.0	4.9	1.0
CCRV risk	1.0	1.6	1.0	1.8	1.9	1.9	1.4
Chronic risk	1.0	2.0	1.0	1.0	2.0	1.8	1.8
Acute risk	0.47	2.1	3.5	2.8	2.6	2.5	2.1
Avoidable risk	0.76	1.3	0.8	1.0	1.0	1.2	1.1
Poly-consultant (>7 cons/year)	10.6	4.8	1.0	4.6	7.0	4.4	5.1
Wage range: between 2 and 4 SMLV	1.0	1.0	1.0	0.8	0.9	1.1	1.0
Wage range: >4 SMLV	1.0	1.2	1.6	1.0	1.2	1.0	1.0
Beneficiary	1.0	1.0	1.0	1.4	5.1	1.4	1.0
Additional affiliate	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Contributor type: Independent	3.3	0.8	2.2	1.0	1.0	0.8	1.2
Contributor type: Dependent	2.9	0.8	1.0	0.6	1.0	0.7	1.0
Age (years)	1.0	0.97	0.84	0.91	1.02	1.02	1.03
Andean region	1.0	1.0	1.6	1.0	0.8	0.9	0.8
Northwest region	0.7	1.0	1.0	0.7	0.8	0.9	0.9
Caribbean region	0.6	1.0	1.0	0.7	1.0	1.0	0.8
Coffee axis region	1.0	1.0	1.0	0.5	0.8	0.8	0.8
Northeast region	0.2	1.0	1.3	0.5	0.7	0.7	0.7
COPD X CCRV	1.0	0.1	1.0	1.0	4.2	1.0	1.2
Oncological X Mental risk	1.0	0.4	1.0	1.0	0.3	1.0	0.7
COPD X Mental risk	1.0	0.1	1.0	1.0	1.0	1.0	1.0
Mental X Maternal Perinatal risk	1.0	0.5	1.0	1.0	1.0	0.5	1.0
Mental X CCRV risk	1.0	1.0	1.0	1.0	0.4	1.0	0.5
Mental X Avoidable risk	1.0	1.0	1.0	1.0	1.0	0.8	1.0
High cost X BP and AC	1.0	1.0	1.0	1.0	1.0	1.0	0.5
Res.4505 X Poly-consultant	1.0	1.0	1.0	1.0	1.0	0.7	0.7
Total components	11	24	10	15	25	28	27

Color code. Red: factors impacting upon the cost positively; Green: factors diminishing the cost; Orange: factors impacting minimally upon the cost in positive manner; and White: factors with no significant incidence on the POS cost.

parts, with one of them using GLM, shows an administrative plausibility in the majority of the coefficients of the independent variables. Understanding this term as an homologation to the term of causality in epidemiology, and which could be expressed herein in terms of the cultural acceptance within the health insurance of determinants of costs, for example, gender impacts

upon the cost, exemplified in that women who use less services during non-reproductive years and that during these years, upon using more services (prenatal controls, hospitalizations due to deliveries) they begin to cost more than the men. Likewise, the model yielded plausible results when costs increased 4.1 times with comorbidities, like a patient with COPD risk + pathologies

that increase cardio-cerebrovascular risk, like high blood pressure, diabetes, obesity, etc.

Another demonstration of the applicability of the model is constituted by the result showing that health costs have a “U” shaped behavior (Figure 3); high during the first year of life (secondary to the management of neonatal pathologies), descending until 15 years of age (school group and initial adolescent characterized by their scarce use of health services, and if these are used such are of low cost: general medicine consultation, vaccines, anticonception, others), and costs gradually start to increase until 65 years of age, as of which costs have an accelerated growth (onset of chronic, non-transmissible disease and – in many cases – of their chronic sequelae). This behavior is demonstrated by most insurance carriers globally.

For example, to calculate the CPU, almost all the results show that they vary significantly when including the health diagnosis (18). For this reason, not bearing in mind the epidemiological composition of the benefit plan insurers – EPS - for these calculations, may lead to defining economic results with very high insurance fees for the “healthier” population and very low for other EPSs with more ill-ridden population. This may occur because of the practice denominated skimming the market or “selection of risks” (18), that is, some insurance carriers limit affiliations of populations presumably of high demand of resources and, hence, in costs. Thereby, the authors consider that the disease burden should be included in modelling costs or in calculations for the CPU for the benefit plan insurers (EAPB) so that economic imbalance is avoided among them, which leads to confusing economic results for the SGSSS; some with large gains and others with large losses.

Since its beginning, the POS has been organized around a list of interventions and medications, and with respect to the estimation of the basic fee, which has been calculated annually in function of the records of activities, is adjusted with inflation, and since a few years back it has been adjusted by cities, where service offer problems could exist; in addition, it is adjusted ex-post by the High-Cost Account, given the presence of some catastrophic diseases.

Although we mention an insurance model, the fee calculated is based on the record of historical consumption of the previous year and does not effectively include any risk model or disease burden of the population of each EPS, given that beyond the adjustment made, a greater offer is obtained of services present in some cities, or in some catastrophic diseases.

The insurance carrier sector and the other sectors that construct the country’s economy need a model that explains better the population’s behavior and permits estimating adequately the fee to be paid into the system by each affiliate, based on epidemiological variables and not simply on the demographic variables. Thereby, classification models must be included according to the risks of each individual, and from these, it has been possible to demonstrate how the majority of the costs of an EPS concentrate on a small group of the population, Kaiser pyramid (19), as a scheme to explain the relationship between the

number of affiliates and the total cost (19-21). The Kaiser pyramid is the model most-often used to classify the population into categories of intervention, depending on its level of complexity. It was developed by Kaiser Permanente in the United States in 1945. On the base of the pyramid, Kaiser locates the healthy users of the population for which the promotion, prevention, and early and timely diagnosis of the disease are primordial. The second level classifies users who have some type of chronic disease, the achievement focuses on self-care, appropriate administration of medications, and education in healthy life habits. The third level has users classified as high cost or complex, who are set with risk programs guided by disease case management plans designed to mitigate inadequate use of specialized procedures and avoid hospital admissions and readmissions (20).

This type of risk-stratification model emerged because of imbalance in economic conditions. To the extent the insurance carriers began using it to develop different types of products and tariff structures according to its users’ risk profiles, these economic imbalances began to disappear. In health systems, risk adjustment and stratification permits the differential retribution of medical care services and activities and, above all, of the resources, this is done to avoid an unfeasibility of the system (21). In other words, the models of risk stratification permit identifying and managing patients who merit more specialized interventions, for example, elderly individuals with more than two diseases diagnosed. In these particular cases, the stratification seeks to avoid hospital admissions not programmed, optimize the assignment of resources, promote the patient’s self-care, prioritize the intensity of interventions at all levels.

Additionally, this methodological proposal could serve as the base to construct tools or instruments that permit continuous monitoring of the costs of health services demanded by the users of a health insurance carrier, without having the prior invoicing and only having the demographic factors, variables appertaining to the insurance business, and risk classification variable that aims to comply with one of the “Standards of monitoring and evaluation of the comprehensive networks of health services offer” that EPSs in Colombia must comply according to that disposed in Resolution 1441 of 2016 (22).

The accumulated distribution of health costs has a ‘peak’ at zero. This corresponds to an important percentage of individuals who have no variable medical costs in the POS; in the case of this study, this cost is constant through the figure of hiring through capitation of the basic level of care, that is, these people are managed by a third party at a fixed cost. For those people who do have medical costs, the distribution of costs is quite slanted to the right with the variance not constant. That is, there tends to be greater variability among the medical costs of the individuals when these are large than when these are small costs.

From the statistics point of view, we suggest modelling with other non- parametric alternatives, like the Generalized Additive Models – GAM (23,24), which would permit seeing the model integrated as a single one, not by segments, as well as without the possibility of estimating the magnitude with which the factors contribute to explaining costs.

As an opportunity for future improvement of the present study, or as a starting point for future research, we could think of a classification that would consider the classification in the diagnostic-related groups (DRG), according to the number of contacts per types of services, as suggested by Vivas et al (25).

For improvement purposes, the model was carried out with data from 2012 to 2013, during the transition of the modification of the tariff plan in which some procedures and activities that were then considered outside the obligatory health plan (POS), as of 2014 these were included in the POS. Similarly, other updates of this model must consider the inclusion of other independent variables, like time affiliated to the EPS, type of affiliation regime (contributive, subsidized), especially to recognize the theme of mobility, the users who go from being subsidized by the state (subsidized regime) when they are unemployed and which began in effect for Colombia as of 2014, according to Decree 3047 of 2013 (26).

For pathologies whose risk coefficient is beneficial to the cost, it may be explained by an efficiency problem in the care opportunity of the patient and in the monitoring of the cost, which in this study was of one year. It may be that the intervention is late, and although the risk is present, affiliates initially do not consume, but in the long term we could evidence the effect of consumption.

Conclusions

Statistics: the “medical POS cost” is a variable not distributed formally and, although the transformation was made with Natural Logarithm, normality was not achieved, nor stability in the residual variance, which induces the search for other alternative methods to transform data, whose exploration and way of distributing these suggest a transformation of the exponential family; a necessary condition to model or treat the data through generalized linear models (GLM) segmented by age groups.

Due to the aforementioned, a good statistical model adjusted to the behavior of an insurance carrier’s cost is a GLM, which best fits to explain an abnormal variable, like the medical cost. Additionally, it included independent variables, like comorbidities and insurance variables, as well as the traditional sociodemographic variables (age, gender, and zone).

Applicable to a health model: the variables “favoring medical costs”, depending on the geographic region where the user is cared for are: being a woman in extreme groups of the vital cycle; having less than six months of affiliation to the EPS; being in wage range 2; belonging to regions different from the Pacific and South (except for males in age groups 15 to 18 years in the Andean and Northeast regions).

The variables associated most with increased costs are:

Patients with high-cost pathologies (CKD, HIV-AIDS, Cancer, rare diseases), being poly-consultant, having been reported as user with ex-ante risks and with current comorbidities, affiliates whose contributor is in wage range 3, being beneficiary for ages

between 19 and 44 years, having COPD risks, risks of neoplastic origin, risk of mental illness, maternal perinatal users, CCRV risk, risk of chronic pathologies.

To validate the model, with 2014 budget data, and using the statistical model proposed for 2013, we have that the increase on the traditional model to estimate the mean cost for 2014, after indexing, was 12% above the value observed with an interval going from 2% above and 23% for the period comprised between October 2013 and September 2014.

Likewise, with a modelling exercise during a subsequent period and with data from 2015, effectiveness is achieved on the estimation performed by the model of 4.8% above the value observed.

Acknowledgments

Our sincere gratitude to Professor Jaime Mosquera from the School of Statistics at Universidad del Valle who provided significant support in the modelling process from the statistics point of view. Likewise, we thank all the members of the area of knowledge management in the EPS, especially to Viviana Benavidez for her accompaniment and suggestions during the process. Gratitude is also expressed to physician Carlos Marino Escobar for his contributions of this model for public health in practical terms aimed at health services.

References

1. Escobedo S. Metodología para la estimación de costos estándar de los procedimientos médicos Incluidos en el plan universal de beneficios de salud. Lima: PRAES – Promoviendo Alianzas y Estrategias; 2007. Pp. 38.
2. Ministerio de la Protección Social. Ley 100 de 1993 por la cual se crea el sistema de seguridad social integral y se dictan otras disposiciones. Bogotá: Ministerio de la Protección Social; 1993.
3. Mullahy J. Much ado about two reconsidering retransformation and the two-part model in health econometrics. *J Health Economics*. 1998; 17(3): 247–281.
4. Deb P, Munkin MK, Trivedic PK. Bayesian analysis of the two-part model with endogeneity application to health care expenditure. *J Appl Econometrics*. 2006; 21 (7): 1081–1099. Doi: doi.org/10.1002/jae.891
5. Comisión de Regulacion en Salud. Acuerdo 029 de 2011 Por el cual se sustituye el Acuerdo 028 de 2011 que define, aclara y actualiza integralmente el Plan Obligatorio de Salud”. Bogotá: Comisión de Regulacion en Salud; 2011.
6. Ministerio de Salud y Protección Social. Resolución 4505 de 2012. Por la cual se establece el reporte relacionado con el registro de las actividades de protección específica, detección temprana y la aplicación de las guías de atención integral para las enfermedades de interés en salud pública de obligatorio cumplimiento. Bogotá: Ministerio de Salud y Protección Social; 2012.

7. Ministerio de Salud y Protección Social. Resolución 4700 de 2008 Por la cual se definen la periodicidad, la forma y el contenido de la información que deben reportar las Entidades Promotoras de Salud y las demás Entidades Obligadas a compensar para la operación de la Cuenta de Alto Costo. Diario Oficial No. 47.190. Bogota: Ministerio de Salud y Protección Social; 2008.
8. WHO. Classifications. International Classification of Diseases (ICD). Consulted on 30/06/2016.
9. Mihaylova B, Briggs A, O'Hagan A, Thompson SG. Review of statistical methods for analysing healthcare resources and costs. *Health Economics*. 2011; 20 (8): 897–916.
10. Cayuela L. Modelos lineales generalizados (GLM). 1a ed. Granada – España: EcoLab, Centro Andaluz de Medio Ambiente, Universidad de Granada - Junta de Andalucía, Avenida del Mediterráneo s/n, E-18006; 2010, 57-87 p.
11. Dobson AJ, Barnett AG. An introduction to generalized linear Models. 3a ed. British Columbia, Canada: Chapman and Hall/CRC Textbook 320. 2008. 221 p.
12. Blough DK, Madden CW, Hornbrook MC. Modeling risk using generalized linear models. *J Health Economics*. 1999; 18(1999): 153–171.
13. Roca AJ, Muñoz VA. Los estudios de prevalencia. En : Martínez, NJF, Antó, BJM, Castellanos PL, Gili MM, Marset CP, Navarro LV, et al. *Salud Pública*, Madrid: McGrawHill; 1998.
14. Rothman, KJ, Lash TL, Greenland S. *Modern epidemiology*. Little Brown & Co. Boston, MA; 1986, pag 84.
15. Peñaloza QRE, García CM, Orozco RN, Puerto GS, Ríos MAM. Financiamiento del sistema general de seguridad social en salud. Seguimiento y control preventivo a las políticas públicas. ECOE Ediciones Ltda. 2012. Procuraduría General de la Nación, Bogotá-Colombia.
16. euroHOPE. European health care outcomes, performance and efficiency. Cost measurement and estimation of cost functions. *euroHOPE Discussion Papers No 2*; 2012. [cited 10 Aug 2015]. ver. 29th. Available in: http://www.eurohope.info/doc/EHDP2_Cost.pdf.
17. Lalonde M. Lalonde and beyond: Looking back at “A New Perspective on the Health of Canadians”. *Health Promotion International* 1986; 1 (1): 93-100. doi: 10.1093/heapro/1.1.93.
18. Vargas JM, Giraldo JA. Modelo de predicción de costos en servicios de salud soportado en simulación discreta. *Información Tecnológica*. 2014; 25 (4): 175-184.
19. Department of Health. Supporting People with LTCs. An NHS and Social Care Model to support local innovation and integration. Long Term Conditions Team Primary Care, Department of Health: Leeds; 2005.
20. Feacham R, Sekhri N, White K, Getting. More for their dollar: a comparison of the NHS with California's Kaiser Permanente. *BMJ*. 2002; 324: 135-143. doi: 10.1136/bmj.324.73.
21. Ham C, York N, Sutch S, Shaw R. Hospital bed utilisation in the NHS, Kaiser Permanente, and the US Medicare programme; analysis of routine data. *BMJ*. 2003; 327(7426): 1257.
22. Ministerio de Salud y Protección Social. Resolución 1441 de 2016. Por la cual se definen los procedimientos y condiciones que deben cumplir los prestadores de servicios de salud para habilitar los servicios y se dictan otras disposiciones. Diario Oficial, No. 49.851. Bogotá: Ministerio de Salud y Protección Social; 2016.
23. Eubank RL. *Nonparametric regression and spline smoothing*. 2nd ed. *Statistic: Text book and monographs*, vol 157. New York: Marcel Dekker; 1999.
24. Buja A, Hastie T, Tibshirani R. Linear smoothers and additive models. *Ann Statistics*. 1989; 17 (2): 453-510.
25. Vivas D, Guadalajara N, Barrachina I, Trillo JL. Explaining primary healthcare pharmacy expenditure using classification of medications for chronic conditions. *Health Policy*. 103(1): 9-15.
26. Ministerio de Salud y Protección Social. Decreto 3047. Por el cual se establecen reglas sobre movilidad entre regímenes para afiliados focalizados en los niveles I y II del Sisbén. Diario Oficial, No. 49.016. Bogota: Ministerio de Salud y Protección Social; 2013.

© Universidad Libre. 2018. Licence Creative Commons CC-BY-NC-ND-4.0. <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>.

