

ROBUST STATISTICAL MODELING OF MEDICAL EXPENDITURES: NONLINEAR EFFECTS, INTERACTIONS, AND DISTRIBUTIONAL HETEROGENEITY

Dr. Emily Carter^{1*}, Dr. Daniel Müller², Dr. Sofia Rossi³

¹*Department of Health Economics, Example University, London, United Kingdom

²School of Applied Statistics, Example Institute, Berlin, Germany

³Department of Public Health, Example University, Milan, Italy

***Corresponding Author:- Dr. Emily Carter**

Abstract:-

The cost of medical services is frequently skewed, dispersed unevenly, and highly variable and the models that are commonly used to describe the cost structure are often based on the mean. The study looked at 1,338 individual insurance records, and compared the impact of age, sex, body mass index (BMI), number of children, smoking status and geographic region on medical charges. Descriptive analysis was followed by multiple linear regression, nonlinear specification, interaction modelling, diagnostic testing, heteroskedasticity-consistent inference, and quantile regression. The baseline model accounted for about 75% of the variation in charges; smoking, age, and BMI were most strongly associated with charges. A nonlinear age term and the smoking $\times$ BMI interaction term enhanced model fit, leading to a greater proportion of variance explained, of about 87%. There were no cases of non-normal residual behaviour and the level of multicollinearity was low but heteroskedasticity was high, according to the diagnostic testing. Standard errors were thus supplemented by robust HC3, to enhance the statistical inference. The additional results of the quantile regression showed that none of the effects of predictors was uniform over the expenditure distribution. The coefficient on BMI was especially large at 75th and 90th deciles and expenditure differences were especially large at higher expenditure levels, for smoking. These results illustrate that medical expenditure should be understood by using an integrated statistical framework that accounts for the nonlinear effects, interaction terms, robust inference and distribution sensitive modelling, rather than just average relationships.

Index Terms:- medical expenditure; robust inference; quantile regression; nonlinear effects; interaction effects; healthcare costs

I. INTRODUCTION

Healthcare utilization is determined by demographic, behavioral, anthropometric, and health-related factors, which means that the level of medical spending differs greatly among people. This type of expenditure tends to be highly positively skewed, with a relatively small proportion of people making a very high proportion of the expenditure. Body mass index (BMI) has been cited multiple times as being a factor for this variation. Previous studies in the United States have demonstrated that the level of healthcare spending varies by BMI, and that the association between BMI and spending varies by age and gender [1]. Similar findings were observed in Korea, where a positive relationship between overweight and obesity and medical costs was also found, and disease incidence increased over time [2].

Economic implications are of importance as spending could increase as body weight increases. There is evidence that direct medical costs due to obesity rise with BMI, suggesting that the relationship between BMI and cost of care is not linear and simple [3]. Similar estimates have been made at the international level suggesting that overweight and obesity create substantial health and economic costs at the international level in all countries with varying levels of income and health system [4]. These results confirm the importance of studying the quantity and form of the link between health-care costs and body weight.

Patterns of expenditure can also be influenced by behavioural characteristics. Smoking has been found to be associated with increased use of health services and healthcare costs [5] compared to non-smoking. The case for smoking as a major cost-related factor rather than a minor behavioural one becomes even stronger as evidence of Mendelian randomization has been found to associate smoking with higher hospital expenditure [6]. Importantly, smoking and BMI might not act independently. These two effects could result in expenditure differences which are not accounted for by additive regression models, and so interaction assessment may be useful in assessing heterogeneous costs of healthcare.

Other factors, such as age, household attributes and regional characteristics also affect medical expenditure. From experience in emerging economies it is known that several socio-economic and demographic factors affect healthcare expenditure, and that this influence differs according to the distribution of healthcare expenditure [7]. Studies of multimorbidity using quantile regression have shown that the effect of health-related factors may vary significantly between low and high cost groups [8]. Other similar methods which are distribution sensitive have been applied for determining factors associated with high healthcare expenditure; the factors observed at the centre of the distribution may be different from the ones observed in the high health expenditure cases [9]. The variation in expenditure determinants between the different regions and dependency groups also suggests that there are differences in the determinants of expenditure at the regional and level of expenditure level [10].

Another major component of expenditure variation is ageing as higher age is associated with higher rates of health services use and increased burden of chronic disease. Studies on health-system spending have demonstrated the potential for ageing and non-communicable diseases to be highly burdensome on health-systems [11]. Concurrently, studies conducted in recent years also have revealed the marked heterogeneity in healthcare expenditures, and the importance of analyzing the determinants at each conditional cost level, not just the average [12].

Despite these developments, many medical expenditure analyses still emphasize overall prediction or conventional mean-based regression without jointly assessing nonlinear effects, interaction structure, heteroskedasticity, robustness, and distributional heterogeneity within a single statistical framework. This leaves uncertainty regarding whether important predictors behave consistently across the expenditure range and whether conventional standard errors remain reliable when residual variance is unequal.

For this reason, the following analysis focused on medical charges by age, gender, BMI, number of children, smoking status and geographic region. The goals were to measure the adjusted associations with expenditure, to explore nonlinearity and interaction, to test the model assumptions and robustness of inferences, and to investigate whether the effect of the predictor was similar at lower, median, and upper levels of expenditure by means of quantile regression.

2. Methodology

2.1 Study Records and Variable Preparation

Data was analyzed on 1338 individual medical insurance records with the variables used consisting of demographic and anthropometric, behavioural, and expenditure data. Medical charges were considered as the outcome variable, and the explanatory variables, such as age, sex, body mass index, number of children, smoking status, and residential region, were considered. All records were examined for missing data, duplicate observations, unrealistic data and contradictory categories. For categorical variables, indicator terms were created for regression analysis. A binary BMI category was developed to separate BMI values $<30 \text{ kg/m}^2$ from BMI values $\geq 30 \text{ kg/m}^2$ for use in interactions assessment.

2.2 Descriptive and Exploratory Statistical Analysis

Means and standard deviations, medians and ranges were used to summarize continuous variables as appropriate. Frequencies and percentages were used to describe categorical characteristics. Asymmetry and dispersion of medical charges were analyzed. Prior to regression modelling descriptive patterns across the demographic and behavioural characteristics were examined for potential non-linear and non-homogeneous expenditure relationships.

2.3 Regression Specification, Nonlinearity, and Interaction Effects

Multiple linear regression was initially fitted to estimate the adjusted associations between medical charges and all the explanatory variables. The expanded specification added a squared age term to assess the curvature in the age–expenditure relationship and added BMI category and smoking status and their interaction to allow for the identification of whether there were differences in expenditure between the smoking status groups with BMI $<30 \text{ kg/m}^2$ and BMI $\geq 30 \text{ kg/m}^2$. R^2 , adjusted R^2 , residual standard error and significance tests were used to assess model improvement. The regression coefficients were presented along with the inferential statistics and probability values.

2.4 Model Diagnostics and Robust Statistical Estimation

After fitting the main models regression assumptions were checked. Homoskedasticity and normality were checked using the Breusch–Pagan test and the graphical inspection and Shapiro–Wilk test respectively. The Durbin–Watson statistic was used to investigate the presence of residual independence. Possible multicollinearity of the explanatory variables was explored by using variance inflation factors. Since unequal residual variance was found, HC3 heteroskedasticity-consistent standard errors were computed for the extended regression model, and statistical inferences were made using the robust standard errors and p-values. The main-effect model (with the variables age, sex, BMI, number of children, and smoking status) was tested for the presence of multicollinearity and for the diagnostic tests.

2.5 Quantile Regression and Comparative Evaluation

In order to investigate any change in the effects of the predictors based on different levels of medical expenditure, quantile regression was used. Models were estimated at the 25th, 50th, 75th and 90th conditional percentile. The quantile specifications were the same and included age, BMI, number of children, sex, smoking status and region. Changes in effect size as measured by coefficient estimates across percentiles were explored to determine whether the size of the effect varies throughout the expenditure distribution. In particular, smoking, BMI, and age, which had the greatest relationships in the previous analyses, were emphasised and were key factors when examining spending variation.

3. Results

3.1 Sample Characteristics and Distribution of Medical Expenditures

All 1,338 observations were retained in the analysis and there were no missing values in the seven variables. Participants had a mean age of 39.21 years (SD=14.05), mean BMI of 30.66 kg/m² (SD=6.10), and 1.09 children on average. Sixty-seven out of the 676 observations were male (49.5%), and 274 people (20.5%) were smokers. There was an even number of representatives by region: northeast 324, northwest 325, southeast 364, and southwest 325. Medical charges averaged \$13,270.42 (SD=\$12,110.01), exceeded the median of \$9,382.03, and ranged from \$1,121.87 to \$63,770.43, demonstrating substantial positive skewness and variability. The socio-economic and demographic, anthropometric, behavioural, and expenditure features of the analytical sample are summarized in Table 1.

Table 1. Descriptive Characteristics of the Analytical Sample

Variable/Category	n	Mean ± SD or n (%)	Median	Range
Age, years	1,338	39.21 ± 14.05	39.00	18–64
BMI, kg/m ²	1,338	30.66 ± 6.10	30.40	15.96–53.13
Children	1,338	1.09 ± 1.21	1.00	0–5
Medical charges, USD	1,338	13,270.42 ± 12,110.01	9,382.03	1,121.87–63,770.43
Male	676	676 (50.5%)	—	—
Female	662	662 (49.5%)	—	—
Smoker	274	274 (20.5%)	—	—
Non-smoker	1,064	1,064 (79.5%)	—	—
Northeast	324	324 (24.2%)	—	—
Northwest	325	325 (24.3%)	—	—
Southeast	364	364 (27.2%)	—	—
Southwest	325	325 (24.3%)	—	—

3.2 Baseline Associations and Nonlinear Effects

The model with the additive OLS method captured 75.09% of the variance of charges, or R²=0.7494, which was high. Using the same interacting effects, the adjusted increase in charge was found to be \$256.80 per year of age, \$339.30 per BMI unit and \$475.70 per additional child; smoking yielded the largest such increase (+\$23,847.50). While the sex of male was negligible (−\$131.30; p=0.693), the coefficients for south-east and south-west were modestly negative, when compared with north-east. In the extended specification, age² was significant (β=3.73, p<0.001), thus including the curvature effect in the age–expenditure relationship beyond linearity. The additive and expanded regression estimates are shown in Table 2.

3.3 Interaction Effects and Conditional Expenditure Patterns

Overall model fit was significantly enhanced by interaction modeling. The addition of a BMI≥30 indicator and its interaction with smoking, along with age² and the original covariates improved R² to 0.8664 (adjusted R²=0.8653). The smoking main effect was \$13,404.60 (p<0.001), while the BMI≥30×smoking interaction added \$19,810.15 (p<0.001). Hence, the difference in smoking costs was significantly greater in those who had a BMI of 30 or more. The specification was also expanded which reduced the residual standard error from around \$6,062 for the additive model to \$4,445, materially improving the explanatory adequacy. Table 2 summarizes the expanded interaction model as well as its improvement over the additive specification.

Table 2. Regression Estimates for the Additive and Expanded Models

Predictor	Additive model Estimate (SE)	p-value	Expanded model Estimate (HC3 SE)	HC3 p-value
Intercept	−11,941.6 (987.8)	<0.001	139.01 (1,363.14)	0.919
Age	256.8 (11.9)	<0.001	−32.62 (59.83)	0.586
Age ²	—	—	3.73 (0.75)	<0.001
BMI	339.3 (28.6)	<0.001	119.77 (34.28)	<0.001
BMI ≥30	—	—	−997.94 (422.96)	0.018
Children	475.7 (137.8)	<0.001	678.60 (105.89)	<0.001
Male sex	−131.3 (332.9)	0.693	−496.77 (244.37)	0.042
Smoker	23,847.5 (413.1)	<0.001	13,404.60 (439.96)	<0.001
Northwest	−352.8 (476.3)	0.459	−279.17 (349.28)	0.424
Southeast	−1,035.6 (478.7)	0.031	−828.03 (351.65)	0.019
Southwest	−959.3 (477.9)	0.045	−1,222.16 (350.53)	<0.001
BMI ≥30 × Smoker	—	—	19,810.15 (604.68)	<0.001
R ²	0.7509	—	0.8664	—
Adjusted R ²	0.7494	—	0.8653	—
Residual standard error	6,062	—	4,445	—

Note: Reference categories were female sex, non-smoking status, northeast region, and BMI <30 kg/m² where applicable. HC3 SE = heteroskedasticity-consistent HC3 standard error; BMI = body mass index. HC3-adjusted standard errors and p-values were used for inference in the expanded model.

The increase in explained variation following model expansion is illustrated in Figure 1.

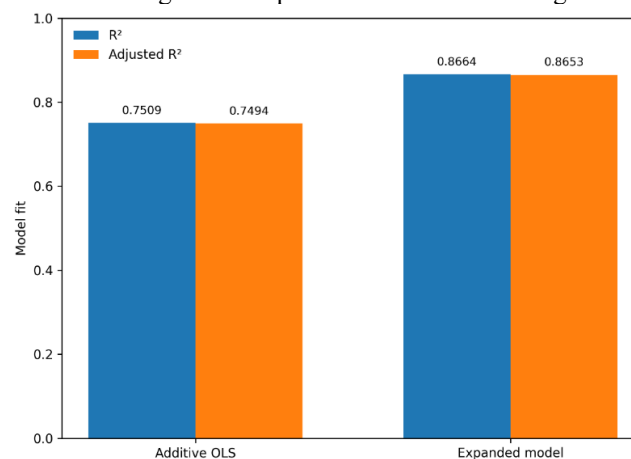


Figure 1. Improvement in model fit following incorporation of nonlinear age and BMI–smoking interaction terms.

The conditional smoking-associated expenditure difference was considerably greater among individuals with BMI ≥30 kg/m² (Figure 2).

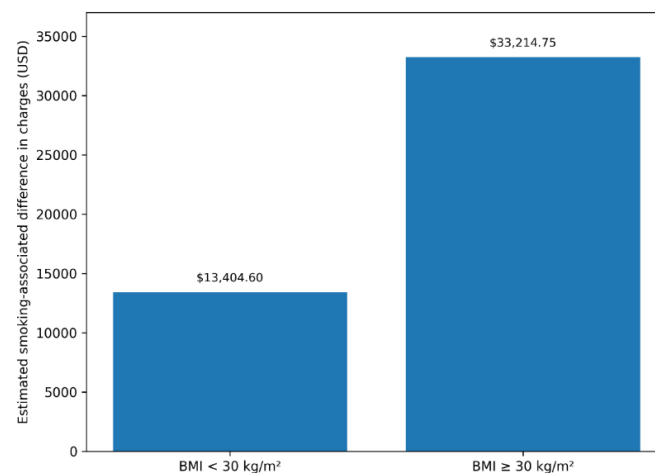


Figure 2. Model-implied smoking-associated difference in medical charges according to BMI category.

3.4 Regression Diagnostics and Robustness Analysis

Conventional OLS inference was found to be proceed with caution with regression diagnostics. Residual normality was

Volume-11| Issue-03|September 2025

strongly rejected ($W=0.8994$, $p<0.001$), and the Breusch–Pagan test showed that there was no heteroskedasticity ($BP=118.02$, $df=5$, $p<0.001$). The level of residual independence was acceptable (Durbin–Watson=2.0869, $p=0.944$) and no multicollinearity was observed as the main-effect VIFs were within the range of 1.00–1.02. Accordingly, HC3 heteroskedasticity-consistent covariance estimation was used for inference in the expanded model. These strong standard errors and p-values are presented with regression estimates in Table 2. A summary of regression diagnostic results and related considerations of robustness is presented in Table 3.

Table 3. Regression Diagnostic and Robustness Assessment

Diagnostic	Statistic	p-value	Interpretation
Shapiro–Wilk normality test	$W = 0.8994$	<0.001	Residual normality rejected
Breusch–Pagan test	$BP = 118.02$; $df = 5$	<0.001	Significant heteroskedasticity
Durbin–Watson test	$DW = 2.0869$	0.944	No evidence of positive residual autocorrelation
Variance inflation factor	1.002–1.015	—	Negligible multicollinearity
HC3 covariance estimation	Applied	—	Used to obtain heteroskedasticity-consistent inference

3.5 Distributional Heterogeneity from Quantile Regression

The effects of the predictors were not uniform across the conditional expenditure distribution as revealed by quantile regression. The smoking effect became more pronounced at the median and higher expenditure quantiles while the age remained relatively stable across the quantiles evaluated. BMI displayed especially strong upper-tail amplification: its quantile-specific slope rose from \$86.14 at $\tau=0.25$ to \$123.52 at $\tau=0.50$, \$424.19 at $\tau=0.75$, and \$1,412.55 at $\tau=0.90$. The children's positive association was found to drop to the median and rise again, being largest at $\tau=0.90$. These patterns indicate pronounced distributional heterogeneity in medical costs. The patterns with respect to quantiles are summarized in Table 4.

Table 4. Selected Quantile Regression Findings Across the Medical Expenditure Distribution

Predictor	$\tau = 0.25$	$\tau = 0.50$	$\tau = 0.75$	$\tau = 0.90$	Overall pattern
BMI coefficient, USD/unit	86.14	123.52	424.19	1,412.55	Strong upper-tail increase
Age coefficient	≈ 264	≈ 264	≈ 264	≈ 264	Comparatively stable
Smoking status	Positive	More pronounced	More pronounced	More pronounced	Greater effect toward upper expenditure levels
Children	Positive	Lower relative effect	Increased	Largest relative effect	Declined toward median, then increased

The progressive increase in the BMI coefficient toward the upper expenditure distribution is illustrated in Figure 3.

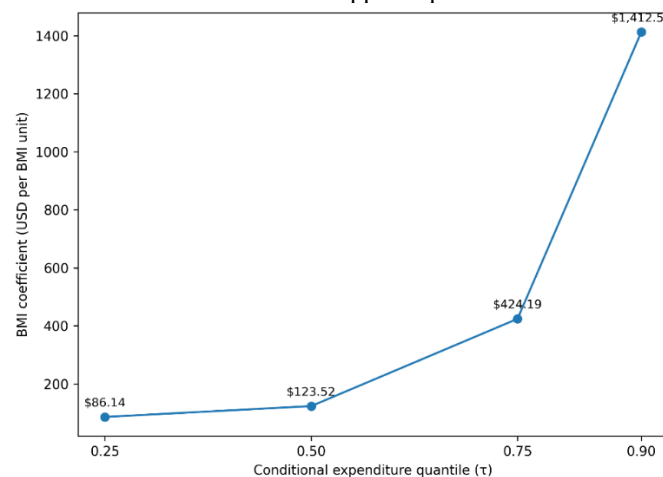


Figure 3. Variation in the BMI coefficient across conditional medical expenditure quantiles.

4. Discussion

The results demonstrated that the demographic, anthropometric and behavioural characteristics had varying effects on

the medical expenditure, with smoking status, age, and BMI the most important factors. The large variation in charges also indicated that expenditure could not be adequately understood through average values alone. The positive correlation between BMI and health care costs was similar to that found in a large UK cohort, where health care expenditure was higher at increasing BMI levels and higher if BMI was associated with complications of obesity [13]. This cost escalation has been seen among people with severe obesity, but especially where there are other obesity-related conditions [14].

The amount of contribution observed to medical expenditure is also consistent with the evidence that there is considerable direct and indirect economic burden resulting from excess body weight across healthcare systems [15]. Importantly, the present analysis did not indicate that BMI acted independently of other behavioural characteristics. There was an interaction between higher BMI and smoking that was related to a greater difference in expenditure for smoking, indicating that the combinations of risk characteristics may be more informative than main effects. Longitudinal studies have also found smoking and obesity to be separately correlated with increased health spending in the long-term [16]. The financial difference seen for adults who smoke and those who do not also has been shown with adults who have chronic respiratory disease [17].

The nonlinear age effect also suggested that spending was not linearly increasing over the age range observed. This result provides motivation for modelling healthcare expenditures in relation to continuous input variables with nonlinear terms instead of assuming a linear relationship between expenditure and continuous variables. Similar non-linear relationships have been found for BMI and catastrophic health expenditure, where adverse financial outcomes could vary with a graded level of the underlying risk factor [18]. These patterns are especially important to research into expenditure since expenses on medicine are frequently concentrated on a small share of people. Previous studies in the middle-aged and ageing population have shown that high medical costs tend to cluster in certain sub-populations [19].

Additional support for moving beyond the traditional ordinary least squares (OLS) inference was afforded by model diagnostics. Residual variation was found to be heteroskedastic, that is it varied with expenditure levels fitted. Thus, the standard errors computed with the assumption of constant variance might have led to incorrect inference. Medical expenditure studies of older people also have shown significant variations between population subgroups in spending on behaviour, and the specification of the model should be carefully done [20]. Out-of-pocket expenditure research has also indicated significant variations in determinants by expenditure levels and hence the need to use distribution-sensitive approaches [21].

This was confirmed by the results of quantile regression. The age displayed relatively stable associations while BMI and smoking-related effects increased toward the higher end of the expenditure distribution. Analyses by quantiles for dependent older adults have indicated that determinants of expenditures may be different for low and high spender groups [22]. There is also a strong BMI effect in the upper tail which is also pertinent to modelling work demonstrating that excess weight loss may affect the population level predicted healthcare spending [23]. Overall, clusters of chronic conditions have been linked to significant costs of health services [24], and could be an indicator of qualitatively distinct patterns of spending, not just of statistical extremes.

Therefore, as a statistical matter the difference between the mean based and the quantiles-specific estimates was meaningful. A predictor with a moderate level of association around the median might have a much higher association for high-expenditure people. More recently, using healthcare-cost analysis with quantile regression has been shown to strengthen the relationships between a selection of predictors at higher expenditure levels [25]. This distributional view can enhance the actuarial evaluation, modelling of insurance costs, and identification of high-cost structures. However, caution should be taken to overlook the findings as causes as the records were observational and the cross-sectional nature of the records did not support the assumption of causality. The scope of interpretation was also constrained by the relatively small number of explanatory variables, lack of detailed clinical diagnosis and treatment data, and the general regional categorizations.

5. Conclusion

Demographic, behavioural and anthropometric factors were highly correlated with medical expenses, the strongest associations being with smoking status, age and BMI and charges. These findings also revealed that these associations were not explained by a simple additive linear model. Non-linear age effect and interaction between high BMI and smoking enhanced model fit and showed more complex expenditure patterns. The results of the diagnostic tests showed heteroskedasticity and non-normal residual behaviour, which suggested that the use of the heteroskedasticity-consistent standard errors provided a more reliable inference. Quantile regression also revealed that associations with the predictors were more extreme at higher expenditure levels for BMI and smoking. These findings suggest that the analysis of the mean and distribution-specific effects should be taken into account in modelling medical costs. A combined framework using conventional regression, nonlinear terms, interaction analysis, robust inference, and quantile regression provided a more complete representation of expenditure variability than a single mean-based model. Results can be used for assessing healthcare costs, analyzing insurance expenditure and identifying high cost patterns. But it is important to note that the records were observational and that important clinical, socio-economic, and treatment-related variables were not available, thus making interpretation of the results associative. The framework could be extended in future by more detailed, longitudinal data sets and additional health indicators.

References

1. Ward ZJ, Bleich SN, Long MW, Gortmaker SL. Association of body mass index with health care expenditures in the United States by age and sex. *PLoS One*. 2021;16(3):e0247307. doi:10.1371/journal.pone.0247307.
2. Song HJ, Hwang J, Pi S, Ahn S, Heo Y, Park S, et al. The impact of obesity and overweight on medical

- expenditures and disease incidence in Korea from 2002 to 2013. *PLoS One*. 2018;13(5):e0197057. doi:10.1371/journal.pone.0197057.
3. Cawley J, Biener A, Meyerhoefer C, Ding Y, Zvenyach T, Smolarz BG, et al. Direct medical costs of obesity in the United States and the most populous states. *J Manag Care Spec Pharm*. 2021;27(3):354-366. doi:10.18553/jmcp.2021.20410.
4. Okunogbe A, Nugent R, Spencer G, Ralston J, Wilding J. Economic impacts of overweight and obesity: current and future estimates for eight countries. *BMJ Glob Health*. 2021;6(10):e006351. doi:10.1136/bmjgh-2021-006351.
5. Sweis NJ, Elfessi ZZ, Rubinstein I, Caskey R. Economic burden of self-reported tobacco smoking compared with non-smoking: systematic review and meta-analysis of direct health care costs. *Adv Ther*. 2025;42(10):5134-5147. doi:10.1007/s12325-025-03318-0.
6. Dixon P, Sallis H, Munafò M, Davey Smith G, Howe L. Mendelian randomization analysis of the causal effect of cigarette smoking on hospital costs. *Nicotine Tob Res*. 2024;26(11):1521-1529. doi:10.1093/ntr/ntae089.
7. Zhou L, Ampon-Wireko S, Antwi HA, Xu X, Salman M, Antwi MO, et al. An empirical study on the determinants of health care expenses in emerging economies. *BMC Health Serv Res*. 2020;20(1):774. doi:10.1186/s12913-020-05414-z.
8. Zhao Y, Atun R, Anindya K, McPake B, Marthias T, Pan T, et al. Medical costs and out-of-pocket expenditures associated with multimorbidity in China: quantile regression analysis. *BMJ Glob Health*. 2021;6(2):e004042. doi:10.1136/bmjgh-2020-004042.
9. Hu L, Li L, Ji J, Sanderson M. Identifying and understanding determinants of high healthcare costs for breast cancer: a quantile regression machine learning approach. *BMC Health Serv Res*. 2020;20(1):1066. doi:10.1186/s12913-020-05936-6.
10. Xu X, Wang Q, Li C. The impact of dependency burden on urban household health expenditure and its regional heterogeneity in China: based on quantile regression method. *Front Public Health*. 2022;10:876088. doi:10.3389/fpubh.2022.876088.
11. Hou X, Liu L, Cain J. Can higher spending on primary healthcare mitigate the impact of ageing and non-communicable diseases on health expenditure? *BMJ Glob Health*. 2022;7(12):e010513. doi:10.1136/bmjgh-2022-010513.
12. Lu M, Gao H, Shi C, Xiao Y, Li X, Li L, et al. Health care costs of cardiovascular disease in China: a machine learning-based cross-sectional study. *Front Public Health*. 2023;11:1301276. doi:10.3389/fpubh.2023.1301276.
13. Pearson-Stuttard J, Holloway S, Sommer Matthiessen K, Thompson A, Capucci S. Variations in healthcare costs by body mass index and obesity-related complications in a UK population: a retrospective open cohort study. *Diabetes Obes Metab*. 2024;26(11):5036-5045. doi:10.1111/dom.15785.
14. Butalia S, Luu H, Guigue A, Martins KJB, Williamson T, Klarenbach SW. Health care cost of severe obesity and obesity-related comorbidities: a retrospective cohort study from Alberta, Canada. *Obes Res Clin Pract*. 2023;17(5):421-427. doi:10.1016/j.orcp.2023.09.006.
15. Malkin JD, Baid D, Alsukait RF, Alghaith T, Alluhidan M, Alabdulkarim H, et al. The economic burden of overweight and obesity in Saudi Arabia. *PLoS One*. 2022;17(3):e0264993. doi:10.1371/journal.pone.0264993.
16. Kim Y. The effects of smoking, alcohol consumption, obesity, and physical inactivity on healthcare costs: a longitudinal cohort study. *BMC Public Health*. 2025;25:873. doi:10.1186/s12889-025-22133-4.
17. Gu D, Sung HY, Calfee CS, Wang Y, Yao T, Max W. Smoking-attributable health care expenditures for US adults with chronic lower respiratory disease. *JAMA Netw Open*. 2024;7(5):e2413869. doi:10.1001/jamanetworkopen.2024.13869.
18. Wang Y, Liu M, Liu J. Association of body mass index with risk of household catastrophic health expenditure in China: a population-based cohort study. *Nutrients*. 2022;14(19):4014. doi:10.3390/nu14194014.
19. Jiang L, Qiu Q, Zhu L, Wang Z. Identifying characteristics associated with the concentration and persistence of medical expenses among middle-aged and elderly adults: findings from the China Health and Retirement Longitudinal Survey. *Int J Environ Res Public Health*. 2022;19(19):12843. doi:10.3390/ijerph191912843.
20. Chen L, Xu X. Relationships between the physical activity intensity and the medical expenditure of middle-aged and elderly people: parsing from the CHARLS database. *Behav Sci (Basel)*. 2023;13(7):566. doi:10.3390/bs13070566.
21. Jeetoo J, Jaunky VC. An empirical analysis of income elasticity of out-of-pocket healthcare expenditure in Mauritius. *Healthcare (Basel)*. 2022;10(1):101. doi:10.3390/healthcare10010101.
22. Salinas-Rodríguez A, Manrique-Espinoza B, Torres Mussot I, Montañez-Hernández JC. Out-of-pocket healthcare expenditures in dependent older adults: results from an economic evaluation study in Mexico. *Front Public Health*. 2020;8:329. doi:10.3389/fpubh.2020.00329.
23. Yoong J, Schnecke V, Aekplakorn W, Bandgar T, Butt JR, Romano JGU, et al. Population-level impact of weight loss on predicted healthcare spending and the incidence of obesity-related outcomes in the Asia-Pacific region: a modelling study. *BMC Glob Public Health*. 2024;2(1):68. doi:10.1186/s44263-024-00094-x.
24. Hajat C, Siegal Y, Adler-Waxman A. Clustering and healthcare costs with multiple chronic conditions in a US study. *Front Public Health*. 2021;8:607528. doi:10.3389/fpubh.2020.607528.
25. Momenyan S, Chan H, Taylor JA, Staples JA, Bryan S, Harris DR, et al. Healthcare and productivity costs among Canadian road trauma survivors over the year following injury. *Sci Rep*. 2025;15:17723. doi:10.1038/s41598-025-01233-6.