

File S2. Econometric Specifications

GLM Model

We estimated the determinants of total patient-incurred costs using GLM with a gamma distribution and log link, which are well suited for modeling right-skewed, non-negative outcomes. The GLM specification is expressed as:

$$E(Y_i | X_i) = \exp(\beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_k X_{ki}) \quad (8)$$

where Y_i represents the monthly total cost incurred by patient i , X_i is a vector of explanatory variables, and β are parameters estimated using quasi-likelihood methods. To aid interpretation, we computed average marginal effects (AMEs), which quantify the expected change in mean costs associated with changes in each covariate.

OLS Sensitivity Models

As a robustness check, we also estimated ordinary least squares (OLS) models on log-transformed costs:

$$\ln(Y_i + 1) = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \dots + \alpha_k X_{ki} + \epsilon_i \quad (9)$$

where coefficients α are interpreted as semi-elasticities, representing the percentage change in costs associated with a one-unit change in the covariate. The OLS results are presented only as sensitivity analyses and are clearly distinguished from the primary GLM results. Model fit was assessed using Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), scaled deviance, and residual diagnostics.

Sensitivity Analysis

To assess robustness, we estimated OLS regressions on log-transformed cost variables:

$$\ln(C_i + 1) = X_i \beta + \epsilon_i \quad (10)$$

Where ϵ_i is the error term assumed to be normal. Both OLS models included the same predictors as the GLM specification and used heteroskedasticity-robust standard errors.

Cost Burden and Catastrophic Expenditure

We evaluated the cost burden as a share of income:

$$B_i = \left(\frac{C_{total_actual}}{I_i} \right) * 100 \quad (11)$$

Where B_i is the percent burden for individual i and I_i is their monthly income. Catastrophic expenditure was defined for each patient as:

$$Cat_i^{(t)} = \begin{cases} 1 & \text{if } B_i > t \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

for thresholds $t = \{10\%, 20\%, 25\%, 30\%, 40\%\}$. Proportions of individuals exceeding each threshold were tabulated