# HERMES: A Health Resources Econometric Analysis Tool

**Kyungseon Choi, Sang Jun Park, Sola Han, Hae Sun Suh**

## Background

Estimating the economic burden through healthcare cost analysis is important in properly distributing limited healthcare resources. However, there are various huddles in estimating the unbiased precise healthcare costs. First, healthcare costs may have zero cost depending on the research topic and an asymmetry of the distribution due to patients with high medical cost.[1, 2, 3] Second, censored data in healthcare commonly present dependent on the event of interest or other covariates and are not informative.[3] Third, real-world data including non-reimbursement items, measurement results, and costs paid by payers and patients is needed to accurately estimate the economic impact on stakeholders. The claim data is well structured, but there is no information that is not necessary for the claim, including measurement results. Electronic Medical Records (EMR) is a data source that includes more data than claims data. It includes information derived from real-world clinical settings including healthcare resource use with non-reimbursement, measurement, and detailed cost. However, the structure and terminology of healthcare data are different for each institute. Observational Medical Outcomes Partnership-Common Data Model (OMOP-CDM) can structure EMR through standard terminology and common model. Various tools using OMOP-CDM are implemented for clinical analysis in ATLAS and Health Analytics Data to Evidence Suite (HADES). However, there is no cost-related analysis tool in ATLAS and HADES. Developing an econometric analysis tool can contribute to estimate the precise cost in experimental settings with various available sources.

To address these unmet needs, our objective was to develop the HERMES, a cost analysis tool using OMOP-CDM, and cross-validate through empirical study.

## Methods

This study was conducted according to the Observational Health Data Sciences and Informatics (OHDSI) guideline.[4] For the development of HERMES, we used OMOP-CDM 5.0 version, ATLAS, and HADES. To adjust positive skewness and zero cost by econometric model and estimate precise healthcare costs, we reviewed literature related to healthcare cost analysis. A model selection algorithm was structured based on well accepted method. Based on the algorithm, we implemented R functions to estimate the cost according to the characteristics of each cost data and compare health resource utilization.

To verify the algorithm and R functions, we conducted an empirical study on patients with exudative Age-related Macular Degeneration (AMD).[5] For cross-validation, we compared the cost analysis method and the estimate of the reimbursement cost with the previous study conducted from claims data in the experimental setting from South Korea with the similar period. During the empirical study and cross-validation, we received consultation from health economics experts and ophthalmologists.

We used Seoul National University Bundang Hospital OMOP-CDM data with approval from Institutional Review Board (X-2012-657-902).

## Results

We structured the algorithm based on well accepted method proposed by Manning and Mullahy in health-economics.[1] The algorithm proposed log transformed Ordinary Least Square (OLS) and the Exponential Conditional Model (ECM) to adjust the positive skewness in healthcare cost (Figure 1). Transformed OLS can adjust positive skewness by transforming the dependent variable to normal distribution. If the

transformed OLS meets the criteria for Best Linear Unbiased Estimator (BLUE), we can obtain an unbiased predictor for cost. However, there are few cases that satisfy BLUE criteria, and despite of retransformation by normality and heteroskedascity, we sometimes are not interested in estimating the log transformed outcome. As improving problems in transformed OLS, the ECM using GLM (Generalized Linear Model) is proposed for general analysis in healthcare cost. The ECM can predict untransformed cost with less bias and does not consider difficult reproducible processes like retransformation.

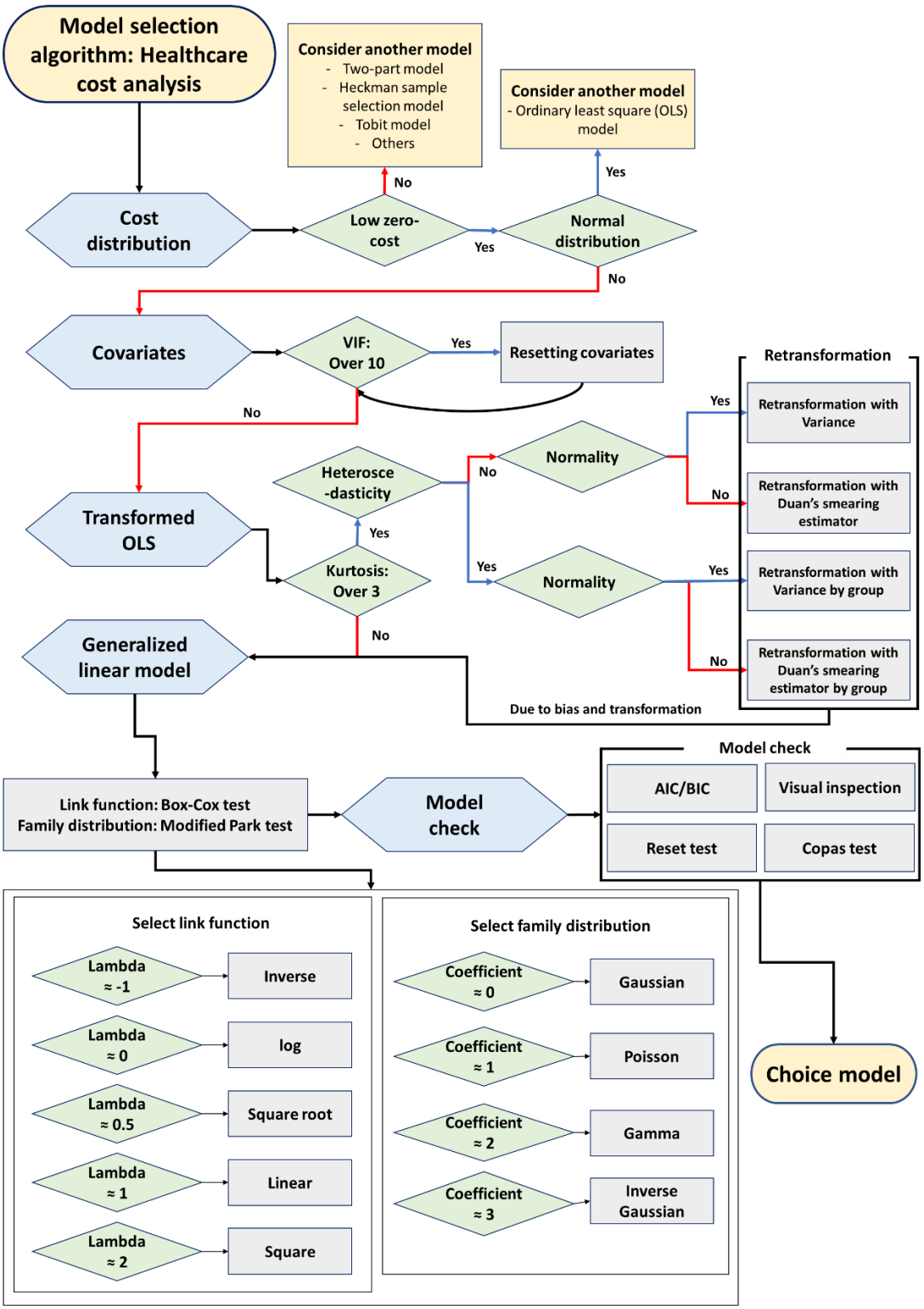


**Figure 1.** Model selection algorithm for healthcare cost analysis in OMOP-CDM

The HERMES mainly provides R functions to implement in the econometric model based on the algorithm (Table 1). R functions include eight domains for analyzing healthcare resource utilization and cost. Transformed OLS function estimates the mean and incremental cost based on the results of kurtosis test, Breush-Pegan test for heteroskedasticity, and Anderson-Darling or Shapiro-Wilk test for normality. The ECM estimates cost with various distributions based on the modified Park test. GLM with link function is also implemented.

**Table 1.** R functions for healthcare cost and resource utilization analysis in OMOP-CDM.

| Domain | Function | Description |
|---|---|---|
| Outlier | outlier_elimination | Eliminate cost outlier with inappropriate zero cost by group |
| Multicollinearity | covariate_vif | Check multicollinearity |
| Unadjusted observed cost | average_observation | Summary of mean, and incremental cost by type of cost and group |
| Transformed ordinary least square | average_ols | Transformed ordinary least square regression by type of cost and group |
| | bootavg_ols_trt | Bootstrap for estimating standard error in treatment group: ordinary least square |
| | bootavg_ols_con | Bootstrap for estimating standard error in control group: ordinary least square |
| Generalized linear model | recommend_average_glm | Generalized linear model as result of box-cox test and modified Park test by type of cost and group |
| | average_glm | The exponential conditional model with log link function and family distribution (Gaussian, Poisson, Gamma) by type of cost and group |
| | bootavg_glm_trt | Bootstrap for estimating standard error in treatment group: generalized linear model |
| | bootavg_glm_con | Bootstrap for estimating standard error in control group: generalized linear model |
| Time-series | year_cost | Time-series visualization |
| Difference in difference | period_cost | Data visualization for difference in difference analysis |
| Health resource utilization | health_resource_utilization | Summary of health resource utilization by type of visit and group |

As a result of an empirical study in exudative AMD (Table 2), we found that the method and results from an empirical study are similar to that of previous study.[6] Both studies used the ECM and OLS with transformation to estimate cost. A previous study used the ECM with gamma distribution commonly used in healthcare cost analysis. We performed the ECM with various distributions; gamma distribution was always the least AIC (Akaike information criterion) in all cost variables. As a result, the reimbursement cost of exudative AMD was significantly higher than that of non-exudative AMD in the empirical study (3.1K USD, 4.25 times per patient, $p<0.0001$) and a previous study (3.7K USD, 4.05 times per patient, $p<0.001$).

**Table 2.** Comparison of empirical study and previous study

| | Empirical study using HERMES | Kim et al., 2019 |
|---|---|---|
| Data source | EMR transformed to OMOP-CDM | Claims data |
| Experimental setting | Propensity score matching with econometric model | |
| Cost analysis method | OLS with a log transformation and heteroskedastic retransformation | |
| | ECM using GLM with a log link function and gamma distribution | |
| Outcomes | Annual mean of cost for two years: Total cost, reimbursement cost, non-reimbursement cost, out-of-pocket cost | Annual mean of reimbursement cost: Total health care, Pre-diagnosis, AMD treatment, Ophthalmology, Non-ophthalmology |
| | Incremental annual cost between exudative AMD and non-AMD | |
| Results for reimbursement cost in first year | | |
| Observed mean cost (USD) : Exudative AMD | 3,784 | 4,173 |
| Observed mean cost (USD) : Non-exudative AMD | 903 | 1,293 |
| Estimate of incremental cost (USD) : ECM | 3,130 | 3,699 |
| Estimate of incremental cost (USD) : OLS | 2,976 | 5,519 |
| Coefficient for group: ECM | 1.45 (p<0.001) | 1.40 (p<0.001) |
| Coefficient for group: OLS | 2.55 (p<0.001) | 2.13 (p<0.001) |

AMD, Age-related Macular Degeneration; ECM, Exponential Conditional Model; EMR, Electronic Medical Records; GLM, Generalized Linear Model; OLS, Ordinary Least Square; OMOP-CDM, Observational Medical Outcomes Partnership-Common Data Model; USD, United States Dollar.

## Conclusion

We developed the HERMES using OMOP-CDM. The HERMES has enabled estimating the economic burden of disease and we could not find any significant differences in methods and results between the previous study. We conclude that the HERMES can be used for analyzing the OMOP-CDM but needs to be verified with more studies. Also, the HERMES can analyze the estimate of costs by reimbursement, non-reimbursement, payers, and patients and lead to economic, clinical, and policy implications through figuring out the economic impact in stake holders. With the expectation that disease burden research through CDM will be further progressed in the future, we hope that the HERMES will contribute to healthcare cost research using OMOP-CDM.

## References/Citations


1. Manning WG, Mullahy J. Estimating log models: to transform or not to transform? J Health Econ. 2001;20(4):461-94.
2. Basu A, Manning WG. Issues for the next generation of health care cost analyses. Med Care. 2009 Jul;47(7 Suppl 1):S109-14.
3. Dario G, Michele P, Simona B, Alessandro D, Franco M, Eva P. Regression models for analyzing costs and their determinants in health care: an introductory review. IJQHC. 2011;23(3):331-341.
4. Observational Health Data Sciences and Informatics. The book of ohdsi. 2021. OHDSI.

https://ohdsi.github.io/Hades/index.html. [Accessed October 18, 2021].
5. Choi K, Park SJ, Han S, Suh HS. Economic Burden of Disease in Patients with Exudative Age-Related Macular Degeneration Using Common Data Model in South Korea. Value in Health. 2022;25(Supp2)
6. Kim S, Park SJ, Byun SJ, Park KH, Suh HS. Incremental economic burden associated with exudative age-related macular degeneration: a population-based study. BMC Health Serv Res. 2019;19:828