Assessing the Transportability of a Cardiovascular Disease Event Model Constructed from the Framingham Heart Study to the NHANES Dataset

Haiyue Song[1], Alice Paul[1]
[1]Brown University, RI

Overview

In this project, we analyzed the transportability of a cardiovascular disease (CVD) event model, which was fitted using Framingham data, to the extensive national NHANES dataset. We assessed this by using the Brier Score, both when individual-level data were available and when they were unavailable, through simulation.

Background

- Users of prediction models usually aim to apply model-derived predictions to some target populations. Many are interested in using the risk prediction model for CVD events built from the Framingham Heart Study within the extensive population (e.g., NHANES).
- However, there are differences between the target population and the population underlying the source study data.
- We evaluated the transportability of the Framingham model and a tailored model by calculating the Brier score in NHANES data (Steingrimsson et al., 2023), when individual-level data are available.
- Additionally, we illustrated the transportability using simulated data when only summary data are accessible.

Methods

- **Source population:** Patients from Framingham Heart Study (eligible criteria: 30-74 years old without history of CVD)
- **Target population:** Patients in the National Health and Nutrition Examination Survey (NHANES) study. We subset NHANES data to meet the eligible criteria of the source population.
- **Train-test split:** 80% in train, 20% in test. When we evaluate the Framingham model, we only split the Framingham data; when we evaluate the tailored model, we split both the Framingham and the NHANES.
- **Missing Data:** We only use complete data in Framingham study and use multiple imputation to impute five complete data set for NHANES.
- **Model:** $logit\big(Prob(CVD)\big) = \beta X$ on Framingham's train.
  X include age, serum total cholesterol (Total-C), high-density lipoprotein cholesterol (HDL-C), systolic blood pressure (Systolic BP), use of anti-hypertensive medication during the exam (BP treatment), current cigarette smoking status at the time of the exam, and diabetes status. First four features are continuous variables with log-transformation.
- **Model Tailoring**: Fit the model on composite training data from Framingham and NHANES with inverse-odds weights, $\frac{\Pr(S=0|X,D=0)}{\Pr(S=1|X,D=0)}$, where S is the source indicator and D is the test indicator.
- **Brier Score in Target Population:**

$$\hat{\psi}_{\hat{\beta}} = \frac{\sum_{i=1}^{n} I(S_i = 1, D_i = 1)\hat{o}(X_i)\left(Y_i - g_{\hat{\beta}}(X_i)\right)^2}{\sum_{i=1}^{n} I(S_i = 0, D_i = 1)},$$

where $\hat{o}(X_i) = \frac{\Pr(S=0|X,D=1)}{\Pr(S=1|X,D=1)}$, and $g_{\hat{\beta}}(X_i)$ is the prediction of outcome.

Simulation Studies

- We simulated NHANES data when only summary data are available.

Simulation Studies

Aims

To investigate whether transportability analysis can be applied to simulated data only summary data is available. We used transporting CVD model to simulated NHANES data as an example.

Data-generating Mechanisms

DGM1. From a multivariate normal distribution, using the summary of NHANES after log-transformation of continuous variables.

DGM2. From a multivariate normal distribution, using only the summary of raw NHANES data.

DGM3. From marginal distributions with a Cholesky transformation applied.

Estimands: Brier Score in target population

Methods: As the "Methods" section

Performance Measures: Bias, empirical SE, MSE, comparing with Brier Score from individual-level NHANES data

Results

The Brier score on Framingham's test data is 0.2134 for men and 0.1218 for women.

Model	Framingham Model		Tailored Model	
Individual-level Data: Brier Score				
Imputation	Men	Women	Men	Women
NHANES 1	0.1074	0.0408	0.1044	0.0444
NHANES 2	0.1092	0.0421	0.1054	0.0463
NHANES 3	0.1089	0.0407	0.1042	0.0440
NHANES 4	0.1089	0.0405	0.1081	0.0419
NHANES 5	0.1079	0.0420	0.1061	0.0437
Average	0.1085	0.0412	0.1056	0.0441
SD	0.0008 (0.73%)	0.0008 (1.94%)	0.0016 (0.15%)	0.0016 (0.36%)
Simulated Data: Brier Score (SD in Parentheses)				
DGM	Men	Women	Men	Women
DGM1	0.097 (0.004)	0.032 (0.002)	0.096 (0.008)	0.034 (0.004)
DGM2	0.106 (0.006)	0.035 (0.002)	0.109 (0.010)	0.037 (0.004)
DGM3	0.032 (0.008)	0.001 (0.001)	0.009 (0.003)	0.003 (0.002)

Results (Cont'd)

DGM	Framingham Model		Tailored Model	
	Men	Women	Men	Women
Bias (Monte Carlo SEs in parentheses)				
DGM 1	-0.01128 (<0.00001)	-0.00905 (<0.00001)	-0.00998 (<0.00001)	-0.01031 (<0.00001)
DGM 2	-0.00255 (<0.00001)	-0.00641 (<0.00001)	0.0033 (0.00032)	-0.00711 (<0.00001)
DGM 3	-0.0767 (<0.00001)	-0.04019 (<0.00001)	-0.09642 (<0.00001)	-0.04137 (<0.00001)
MSE (Monte Carlo SEs in parentheses)				
DGM 1	0.00014 (<0.00001)	0.00009 (<0.00001)	0.00017 (0.00001)	0.00012 (<0.00001)
DGM 2	0.00005 (<0.00001)	0.00005 (<0.00001)	0.00012 (0.00001)	0.00007 (<0.00001)
DGM 3	0.00595 (0.00004)	0.00162 (<0.00001)	0.00931 (0.00002)	0.00172 (0.00001)
Empirical SE (Monte Carlo SEs in parentheses)				
DGM 1	0.00413 (0.00009)	0.00183 (0.00004)	0.00826 (0.00018)	0.00377 (0.00008)
DGM 2	0.00554 (0.00012)	0.00213 (0.00005)	0.01027 (0.00023)	0.00434 (0.0001)
DGM 3	0.00822 (0.00018)	0.0008 (0.00002)	0.00301 (0.00007)	0.00232 (0.00006)

Conclusions and Limitations

- **When individual-level NHANES data are available:**
  1. The transportability of the predictive model appears to be potentially better for women than for men.
  2. The Brier scores for both women and men are lower than those from the Framingham data before transportation, suggesting that the models are robust and appropriate for use in the NHANES population. This may also indicate a significant imbalance in the NHANES population between patients with or without cardiovascular disease (CVD) events.
  3. The Brier score for the NHANES population shows improved performance with the tailored model compared to the Framingham model, indicating the effectiveness of model tailoring.
  4. The small standard deviation of the Brier score estimates indicates consistent transportability across multiple imputation datasets.
- **From simulation studies:**
  1. DGM1 and DGM2 effectively generated data, with Brier Scores very close to those calculated from individual-level NHANES data, supporting the potential for transportability analysis when only summary data are available. However, DGM3 did not generate data effectively, leading to an overestimation of transportability.
  2. Performance measures indicate that DGM2 generates data more effectively than DGM1, demonstrating that even when only a summary of raw data without any transformation is available, the covariance matrix after transformation can be effectively estimated.
  3. The simulation study demonstrates that using simulated data from summary statistics is workable and feasible for assessing the transportability of the prediction model to a new target population.
- **Limitations:** 1. Multiple imputation assumes that the data are missing at random (MAR). 2. We generated categorical variables continuously and then categorized them, which may induce additional statistical association. 3. The summary data for NHANES might be based on complete data, which could differ from the summary after multiple imputation. 4. The outcome in NHANES data may be significantly imbalanced, potentially simplifying prediction. 5. To better generate correlated data, the use of Copulas could be considered. 6. The study only used the Brier Score; further research could include the Area Under the Curve (AUC) in the target population.

This project is a collaboration with Dr. Jon Steingrimsson in the Department of Biostatistics, Brown University School of Public Health.