Cost-Effective Optimization of Model-Based Prediction of Cardiovascular Disease (CVD)

Taehoon Ha (tah4002), Weill Cornell Medicine

ABSTRACT

Objective Determine the demographics and medical tests that help predict the likelihood and optimize in a cost-effective manner.
Method Two types of models were used to predict the CVD: proportional odds model and logistic regression model based on 920 patients from four sites.
Results Cost-effective model includes 7 covariates including age, sex, chest pain type, Exercise induced angina, ST depression induced by exercise relative to rest, Slope of the peak exercise ST segment, Number of major vessels (0-3) colored by fluoroscopy.

OBJECTIVES

The accuracy of the diagnosis of CVD relies on the individual doctor's knowledge and experiences. We want to increase the accuracy of the diagnosis of CVD based on models.

1) To determine the demographic factors and medical tests that help predict the likelihood of heart disease.
2) To find a combination of necessary medical tests that help predict the likelihood of heart disease in a cost-effective manner.

METHOD

- 920 patients were identified who were test in four sites: US1, US2, EU1, and EU2. Each patient received 11 different tests.
- Diagnoses were selected as an outcome variable.
- - 4 levels: No heart disease, 1, 2, and 3 narrow vessel(s)
- - Binary outcome:: No heart disease (0) and heart disease (1)
- Two modeling methods were used: proportional odds model and logistic regression model.
- Descriptive Statistics and Exploratory Data Analysis (EDA)
- - Kruskal-Wallis test were used to compare between groups of diagnoses for continuous variables. For categorical variables, Chi-square test and Fisher's exact test were used to compare between groups of diagnoses.
- - For normality assumption check, density plots and Shapiro - Wilk's normality test were used.
- - For homoscedasticity, Bartlett's test and residual plots were used.
- Based Model fitting
- - **Proportional odds model** Three model selection methods were used to choose predictors: forward, backward, and stepwise. The final model was selected based on AIC and hypothesis testing (ANOVA).
- - **Logistic regression model** Same method was used to choose predictors. The final model was selected based on AIC, AUC & ROC curves, non-nested model hypothesis test (Vuong), and VIF.
- Effects plot was used to summarize the results of each model.
- **Cost-effective analysis**
- - Based on logistic model as AIC is much lower compared to proportional odds model.
- - Determined by removing covariates with the highest ΔAUC/price starting from the logistic model determined earlier while the model AUC > 0.9.

RESULTS

[Base Model Fitting]

- **Proportional Odds Model**

Results showed that there is an increased chance to be observed in a lower severity heart disease for patients who are male, higher rest blood pressure, non-normal Resting electrocardiographic results, higher ST depression induced by exercise relative to rest, and non-increasing slope of the peak exercise ST segment. On the other hand, patients with lower maximum heart rate reached have are less likely to be observed with lower severity heart disease.

$$\log\frac{\pi_0}{\pi_1+\pi_2+\pi_3+\pi_4} = -0.3563 + \beta_1 x_{\text{sex}} + \beta_2 x_{\text{trestbps}} + \beta_3 x_{\text{restecgST-T}} + \beta_4 x_{\text{restecgLVH}} + \beta_5 x_{\text{thalach}} + \beta_6 x_{\text{oldpeak}} + \beta_7 x_{\text{slopeFlat}} + \beta_8 x_{\text{slopeDown}}$$

$$\log\frac{\pi_0+\pi_1}{\pi_2+\pi_3+\pi_4} = 0.9332 + \beta_1 x_{\text{sex}} + \beta_2 x_{\text{trestbps}} + \beta_3 x_{\text{restecgST-T}} + \beta_4 x_{\text{restecgLVH}} + \beta_5 x_{\text{thalach}} + \beta_6 x_{\text{oldpeak}} + \beta_7 x_{\text{slopeFlat}} + \beta_8 x_{\text{slopeDown}}$$

$$\log\frac{\pi_0+\pi_1+\pi_2}{\pi_3+\pi_4} = 1.9809 + \beta_1 x_{\text{sex}} + \beta_2 x_{\text{trestbps}} + \beta_3 x_{\text{restecgST-T}} + \beta_4 x_{\text{restecgLVH}} + \beta_5 x_{\text{thalach}} + \beta_6 x_{\text{oldpeak}} + \beta_7 x_{\text{slopeFlat}} + \beta_8 x_{\text{slopeDown}}$$

$$\log\frac{\pi_0+\pi_1+\pi_2+\pi_3}{\pi_4} = 3.8866 + \beta_1 x_{\text{sex}} + \beta_2 x_{\text{trestbps}} + \beta_3 x_{\text{restecgST-T}} + \beta_4 x_{\text{restecgLVH}} + \beta_5 x_{\text{thalach}} + \beta_6 x_{\text{oldpeak}} + \beta_7 x_{\text{slopeFlat}} + \beta_8 x_{\text{slopeDown}}$$

$$\text{, where } \beta_1 = 1.261, \beta_2 = 0.008, \beta_3 = 0.388, \beta_4 = 0.273, \beta_5 = -0.024, \beta_6 = 0.729, \beta_7 = 0.728, \beta_8 = 0.588.$$

- **Logistic Regression Model**

Results showed that there is an increased odds to have heart disease if a patient is older, male, with Asymptomatic chest pain, lower Serum cholesterol, Exercise induced angina, ST depression induced by exercise relative to rest, non-increasing slope of the peak exercise ST segment, higher number of major vessels colored, and non-normal defect.

$$\begin{aligned}\log\left(\frac{\pi}{1-\pi}\right) =& -3.977 + 0.027\cdot x_{\text{age}} + 0.956\cdot x_{\text{sexMale}} - 0.81\cdot x_{\text{cpAtypicalAngina}} - 0.316\cdot x_{\text{cpNon-aginalPain}} + 1.11\cdot x_{\text{cpAsymptomatic}}\\ &- 0.004\cdot x_{\text{chol}} + 0.969\cdot x_{\text{exangExAngina}} + 0.402\cdot x_{\text{oldpeak}} + 0.77\cdot x_{\text{slopeFlat}} + 0.259\cdot x_{\text{slopeDownsloping}} + 0.956\cdot x_{\text{ca}}\\ &+ 1.075\cdot x_{\text{thalFixed}} + 1.678\cdot x_{\text{thalReversable}}\end{aligned}$$

[Cost-effective Analysis]

- Akaike information criterion (AIC) of logistic model (636.76) is much lower than the AIC of proportional odds model (1899.38). Hence, logistic model will be the base to conduct the cost-effective analysis.
- Considering both the predictability and cost effectiveness, the necessary factors are age, sex, chest pain types, Exercise induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, and Number of major vessels (0-3) colored by fluoroscopy.

$$\begin{aligned}\log\left(\frac{\pi}{1-\pi}\right) =& -4.964 + 0.0332\cdot x_{\text{age}} + 1.483\cdot x_{\text{sexMale}} - 0.98\cdot x_{\text{cpAtypicalAngina}} - 0.324\cdot x_{\text{cpNonAginalPain}} + 1.176\cdot x_{\text{cpAsymptomatic}}\\ &+ 1.109\cdot x_{\text{exangExAngina}} + 0.4\cdot x_{\text{oldpeak}} + 0.837\cdot x_{\text{slopeFlat}} + 0.514\cdot x_{\text{slopeDownsloping}} + 1.023\cdot x_{\text{ca}}\end{aligned}$$

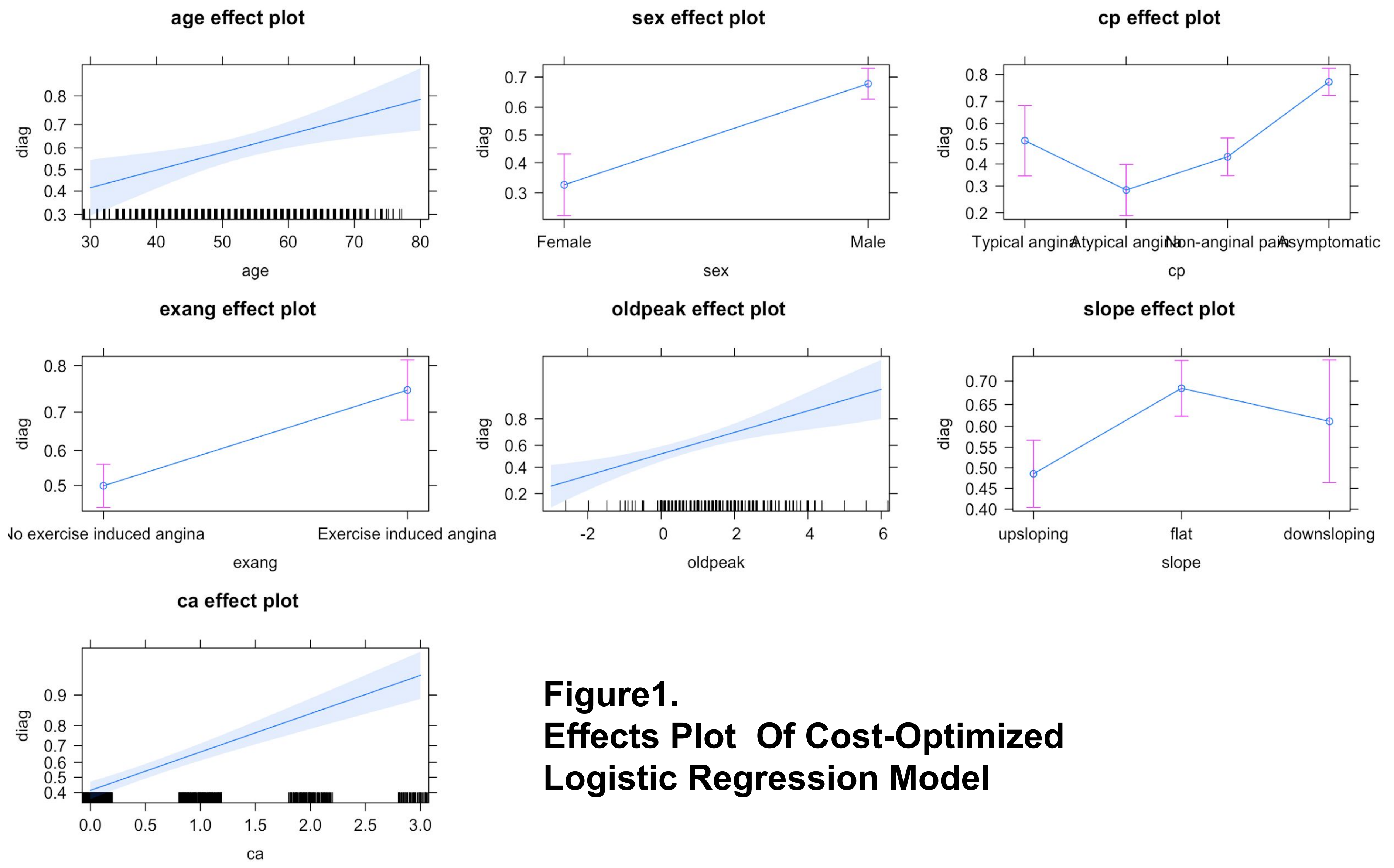
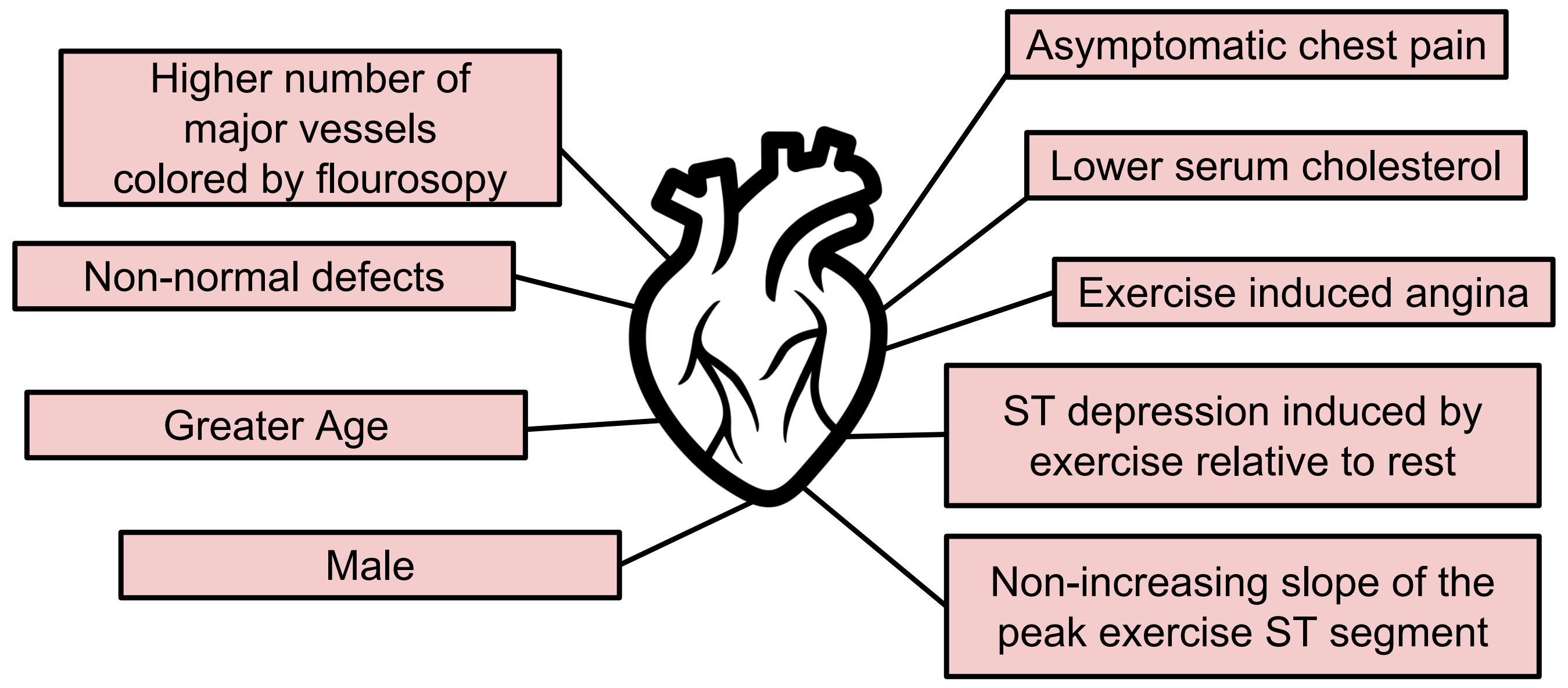


Figure1.
Effects Plot Of Cost-Optimized Logistic Regression Model

CONCLUSION

- To predict the patients at higher risk of heart disease, physicians should take into account of their age and sex. The risk of heart disease increase as patients get older and if the patient is male.
- Physicians should also test: chest pain types, serum cholesterol, exercise induced angina, ST depression induced by exercise relative to rest, slope of the peak exercise ST segment, number of major vessels colored by , and defect types.

Attributes of Patients with Higher Risk of Heart Disease



Higher number of major vessels colored by flourospy

Non-normal defects

Greater Age

Male

Asymptomatic chest pain

Lower serum cholesterol

Exercise induced angina

ST depression induced by exercise relative to rest

Non-increasing slope of the peak exercise ST segment

- When considering the cost of the medical exams, we should not only consider the actual cost of the exams but also the cost of each test to increase the predictability of heart disease.
- Hence, considering the unit costs of predictability, we may remove the examinations for cholesterol and defect types. Evenstill, the predictability is adequately high while the cost of the examination reduces by $110.17 per patient.

DISCUSSION

- Current results did not take into account the possible random effects caused by the study sites. Further study can be conducted using models such as generalized linear models where we can consider those effects.
- Machine learning techniques such as k-fold cross validation can be used to determine a better model or combination of demographics and medical tests in order to efficiently predict heart disease.

SELECTED BIBLIOGRAPHY

- Siadaty MS, Shu J. Proportional odds ratio model for comparison of diagnostic tests in meta-analysis. *BMC Med Res Methodol*. 2004;4(1):27. Published 2004 Dec 10. doi:10.1186/1471-2288-4-27
- van der Ende, YIdau & Hartman, et al (2016). The LifeLines Cohort Study: Prevalence and treatment of cardiovascular disease and risk factors. International Journal of Cardiology. 228. 10.1016/j.ijcard.2016.11.061.