Proteomic signature for early diagnosis of left ventricular remodeling after myocardial infarction

Wilfried Heyse, Vincent Vandewalle, Philippe Amouyel, Guillemette Marot, Christophe Bauters, Florence Pinet

Proteomic signature for early diagnosis of heart failure or death after myocardial infarction

W. Heyse[1-2], V. Vandewalle[3-2], P. Amouyel[1], G. Marot[3-2], C. Bauters[1], F. Pinet[1]

[1]U1167, Inserm, Université de Lille, CHU Lille, Institut Pasteur de Lille; [2]Équipe Projet Modal, Inria Lille - Nord Europe; [3]ULR2694, Université de Lille

Introduction

Heart failure (HF) remains a main cause of mortality worldwide. The most common cause of HF is coronary artery disease and particularly myocardial infarction (MI). The aim here is to identify plasmatic proteins that could enhance the prediction of the occurrence HF after MI and provide a better understanding of the proteins involved in that phenomenon. In order to do that a clinical only prediction model will be build, this model will be used as a reference. Then a proteomic model will be build by selecting proteins that can enhance the clinical variables to build a better model.

Material and methods

Study Population :

This study focuses on the REVE-1 and REVE-2 cohorts which have been previously reported (Christine Savoye 2006) and (Fertin et al. 2010). Those two cohorts were designed to analyze the association of circulating biomarkers with left ventricular remodeling after a myocardial infarction. Patients were included in the cohort after a myocardial infarction for which the infarct zone comprised at least three left ventricular segments that were akinetic at predischarge echocardiography. Patients included in these cohorts are described in *Table 1.* Blood samples were collected for each patient after the myocardial infarction.

In this study, we focus on the longterm follow-up, the event observed was the death of a patient or his hospitalization for heart failure. Since both cohorts have the same design and include same kind of patients so we will use REVE-1 as the derivation cohort and REVE-2 as the validation cohort because REVE-1 has a larger number of events.

	REVE 1 (n=255)	REVE 2 (n=238)
Age (years)	58 (48 to 72)	56 (46 to 69)
Women (no.)	67	46
Follow up (years)	11 (4 to 13)	9 (7 to 10)
Nb. Observed Events	77	41
Time to event (years)	5 (1 to 9)	4 (1 to 6)

Table 1. Characteristics of the study cohorts. Results are presented as median (interquartile range).

Quantification method :

The blood sample collected on every patient was used to perform protein quantification. The protein quantification was performed by SOMALogic with a method consisting in the use of SOMAmer (modified aptamers) to quantify 5284 proteins per sample hence per patient.

Data pre-processing :

The firsts analysis showed missing values in the clinical variables as show in *Figure 1.* In order to keep as much data as possible, imputation of missing values was performed using the missMDA method (Josse and Husson 2016). This method aims to impute values by minimizing their impact on the principal component analysis of all the data. Morally, with this method, data is imputed to look as much as possible as the complete data available.

The proteomic data was transformed to a $\log_2$ scale and 402 proteins were removed due to unusable measurements.

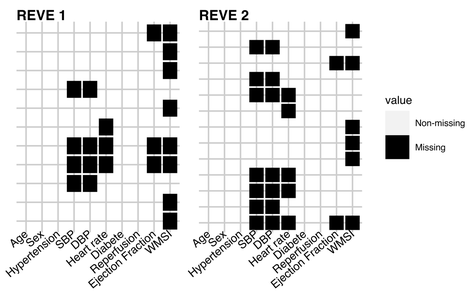

Figure 1. Missing data distribution across both cohorts.

Models and results

The goal of this study is to find proteins which could enhance the performance a risk model based only on the available clinical data on the day of inclusion with the proteomic data available on the day of inclusion. Due to the nature of the data, Cox proportional hazards regression models are used to model the risk of death or heart failure after myocardial infarction. In order to avoid overfitting of the model, the number of variables included in the model was limited to 8 to keep a ratio of 1 variable for 10 events as recommended in (Peduzzi et al. 1996).

Clinical model :

The first step was to develop the reference risk model using clinical data only. A stepwise backward procedure was used to select variables of the clinical model among the available clinical variables (age, sex, history of hypertension, systolic blood pressure, diastolic blood pressure, heart rate, diabetes, reperfusion or not after the myocardial infarction, ejection fraction and wall motion score index). The clinical model include 4 clinical variables which are :

Age, heart rate, diabetes and **ejection fraction**

This model has a Harell's C-index of **0.759** ([0.63;0.85]) on the derivation cohort and **0.771** ([0.60; 0.88]) on the validation cohort.

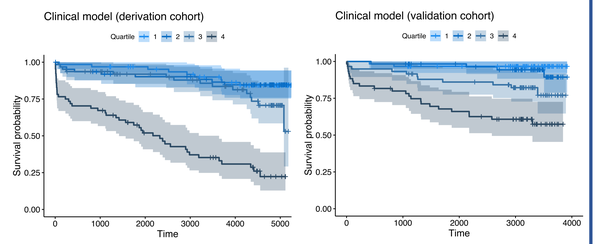

Figure 2. Survival probability curves for endpoints of HF and death stratified by quartiles of the clinical model

Clinical and Proteomic model

To perform protein selection, the Least Absolute Shrinkage and Selection Operator (LASSO) (Tibshirani 1996) model was used. This model was design to select only 4 proteins in order to respect the limitation on the total number of variables. In addition of the 4 previous clinical variables, 4 proteins were selected :

Regenerating islet-derived protein 3-alpha, BNP, Beta-2-microglobulin, Insulin-like growth factor-binding protein complex acid labile subunit.

This model has a Harell's C-index of **0.771** ([0.65;0.86]) on the derivation cohort and **0.787** ([0.62; 0.89]) on the validation cohort. The Harell's C-index is better on this model than the only clinical model and survival probability curves are more separated for this model (*Figure 3*) than for the clinical model (*Figure 2*).

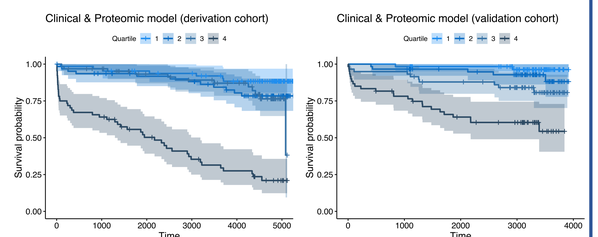

Figure 3. Survival probability curves for endpoints of HF and death stratified by quartiles of the clinical and proteomic model

Conclusion

On both cohorts, the addition of four wisely selected proteins to the clinical model contributed to enhance the performances of the model and the accuracy of the predictions. Both the clinical and clinical and proteomic models have higher Harell's C-index in the validation cohort than in the derivation cohort which shows the robustness of our models.

With the support of Fondation pour la Recherche Médicale : FRM PBR201903008963