PhD thesis title: **Modelling of morbidity risk with the use of survival analysis**

Author: Jadwiga Ewa Borucka

The PhD thesis entitled 'Modelling of morbidity risk with the use of survival analysis' is dedicated to the use of survival analysis models for the morbidity risk forecasting purpose. The thesis consists of the introduction, 5 chapters and conclusions. The aims and content of the thesis are presented in two areas: theoretical and applied one. The theoretical part is focused on the comparison of semiparametric and parametric survival models which has not been sufficiently discussed in the subject literature so far. The applied part presents the example of usage of considered methods to analyse the morbidity risk for selected cardiovascular diseases (heart attack, stroke, hypertension, high blood cholesterol).

Research hypotheses:

Considering the motivation and the aims of the PhD thesis the following research hypotheses have been defined:

(1) Semiparametric and parametric survival models are useful tools to analyse the morbidity risk; they allow identifying morbidity risk factors among considered population.
(2) Methods of assessment and comparison of semiparametric and parametric survival models can lead to the selection of the best survival model with regard to: (a) discriminant power, (b) calibration, (c) overall goodness of fit.
(3) Selection of the best model is determined by the type of disease.

Research methods:

The research conducted for the PhD thesis purpose was based on the random data simulations as well as analysis of the real data collected in the frame of the Survey of Health, Ageing and Retirement in Europe (SHARE). Randomly generated database was used to assess methods considered for comparison of semiparametric and parametric survival models within the theoretical part of the thesis. Real database was used to present the example of usage of selected methods of comparison of semiparametric and parametric survival models within the applied part of the thesis.

Conclusions:

The most important conclusions of the thesis are those drawn from the analysis of randomly generated database performed within the theoretical part of the thesis. The results showed high sensitivity of considered methods of comparison between semiparametric and parametric survival models to the omission of significant covariates in the model. This conclusion proved the first research hypothesis is true – survival models are useful tools to identify morbidity risk factors as omission of significant risk factors in the model will be reflected in the values of discriminant power, calibration and goodness of fit measures. On the other hand, the results showed only slight sensitivity of considered methods of models comparison to the proper model selection: in most cases measures of discriminant power, calibration and goodness of fit worsened only for the models assuming proportional hazard in case this assumption was violated. If the assumption was satisfied, all models represented similar level of discriminant power, calibration and goodness of fit, the same was observed for all models which do not assume proportional hazards if the assumption was violated. These conclusions proved that the second research hypothesis is not fully true – measures of discriminant power, calibration and goodness of fit do not always allow selection of the best model; in particular, if proportional hazard assumption is satisfied, all considered survival models represent similar discriminant power, calibration and goodness of fit. The third research hypothesis was assessed on the basis of the SHARE data analysis which proved the hypothesis to be true – if there was a rationale to select the best model with regard to discriminant power, calibration or goodness of fit, the model selection was determined by the type of disease.

All in all, the thesis showed that survival models are useful tools to analyse the morbidity risk and allow morbidity risk factors identification. Proper model selection turned out not to be that important with regard to the discriminant power, calibration and goodness of fit (except for the cases when the proportional hazards assumption was violated) as the inclusion of all significant covariates in the model. Results obtained and conclusions drawn on the basis of the conducted research can be considered as the starting point for future analyses which might involve the following questions: how can efficient morbidity risk factors identification impact the actual morbidity rates (through preventive actions, diet or lifestyle changes)? Or: how can efficient morbidity risk factors identification impact the overall healthcare costs (including preventive actions costs and potentially reduced cure costs)? Answers to these questions would require more detailed studies, involving analysis of medical and economic data. However on the basis of some previous research it might be anticipated that there would be a positive effect of activities aimed at early identification of morbidity risk factors, concerning the actual morbidity rates and total health care costs. The more detailed analysis might provide precise estimation of the level of this impact.