SUPPLEMENTAL MATERIALS

Supplemental Methods

Model building strategy

The assessment of the impact on the outcome (survival) of several potentially influential variables requires their simultaneous evaluation in a single model in order to explain di effect of each factor independently of the others. This avoid the misleading conclusion that multiple single factor (univariate) models may cause, even if these are a good starting point. Thus, a multivariate analysis was planned and a multivariable model-building strategy was carried out for the present study.

Aware that there is no consensus among researcher on the 'best' strategy to find a 'good' model, we chose a pragmatic approach as proposed by Royston and Sauerbrei (1). In accord with these Authors, "by 'good' we mean a model that is satisfactory from the subject-matter point of view, robust with respect to minor variation of the current data, predictive in new data, parsimonious and useful beyond the dataset on which it was created "(chap. 1.1.1 page 1 of ref 1).

The strategy used can be summarized as follow.

1. The initial step involves, first, the selection of the criteria employed to choose the pool of variables among which the subsequent analysis had to identify the factors relevant to the outcome and, second, the selection of the model class to be used, i.e., the Cox model, being the outcome of interest the survival time. Having in mind the main aim of our model (assess the effect of a new factor of interest, adjusting for some established factors in a multivariable model) we selected a pool of variables representing known factors affecting the outcome (1). *Parsimonious selection criteria* were used to avoid overfitting bias. The rule of "*at least 10 observed events for each tested variable*" (chap. 2.9.1 page 47 of ref 1) adopted at an early stage of the analysis was relaxed to "*at least 10 observed subjects for each tested variable*". This allows the assessment of a wider spectrum of candidate factors in a framework with an acceptable balance between the number of observations and the size of the model tested.
2. The multivariable analysis follow. The first aim of the analysis is the selection of the 'important' factors that independently affect the outcome picking them out as a subset of the initial pool by means of a stepwise selection algorithm. The second aim is the clarification of the functional form (linearity or non-linearity) of the continuous predictors since the linearity assumption "may prevent one from recognize a strong effect or lead one to mismodel the effect" (chap 1.2.1 pag.8 of ref 1). The Multivariable Fractional Polynomial (MFP) modelling algorithm developed by Sauerbrei and Royston (2) addresses these two key tasks in multivariable model building: elimination of 'unimportant' variables and selection of a 'reasonable' dose-response function for continuous variables (chap. 1.7.3 page 19 of ref 1). The MFP algorithm combines a backward variables elimination with a search for the best functional form (linear or not linear) of continuous variables (ref. 1, 2).
3. The weight of each significant factor in the model is evaluated by both the hazard ratio and by the contribution to the global explained variation (R^2). The partition of the global R^2 is accomplished by the Shapley-Owen decomposition algorithm (3). The hazard ratio of continuous variables is related to clinically meaningful variation (e.g. 10 years period for age, 5% units for left ventricular ejection fraction). For variables without a range of definite clinical meaning the hazard ratio relative to the interquartile range (75°-25° percentile difference) is employed. This allows comparison of the relative weight between factors and give a measure of the factor relevance on the studied population (e.g. NT-proBNP, GRK2, and norepinephrine). The interquartile range was preferred over the standard deviation given the non normal distribution of the variables.
4. Since the Cox model was adopted, it was mandatory to verify the proportionality assumption inherent with the basic formulation. To this end we used a modified version of the MFP (the MFPT) devoted

to comprehensively explore the time-variable(s) interaction in order to check the assumption (chap. 11.1.1 page 242-243 of ref 1).

5. The next step involve the computation of model performance measures, i.e., calibration, discrimination ability and internal validity.

a) The first is a goodness of fit assessment that we accomplished with the Gronnesby and Borgan calibration test (4). This test verifies the concordance between the observed survival (Kaplan Meier) and the survival estimated with the Cox model in five risk groups of the studied population. Five contiguous strata of the prognostic index (the linear combination of the factors with their Cox coefficients) identify the risk groups. Non-significant test indicates good calibration (4).

b) The discrimination ability refers to how well the model can distinguish between patient outcomes. We used two indices to quantify it, first the Harrel's C as a natural extension of the binary logistic C statistic (the area under the ROC curve), specifically we used a Harrel's C version corrected for the censoring bias as suggested by Gonen and Heller (5). Second, we used a measure of the explained variance in the natural scale of the Cox model (R2) as proposed by Royston and Sauerbrei (6).

c) Adhering to the suggestion expressed by Royston and Sauerbrei (chap 2.2 page 24 of ref 1) we measured the internal validity of the model by assessing the stability of the model characteristics with nonparametric bootstrap sampling (chap 8 page 183-186 of ref 1). Briefly, given the parameters of a model obtained applying the described model-building procedure, the stability of each factor tested in the model is measured by the frequency that this factor is selected as 'significant' in a series of bootstrap replications of the dataset by applying the same procedure. Each bootstrapped dataset may be considered as a random replicate of the original dataset.

External validity cannot be evaluated since it requires an independent dataset (i.e. an independent HF population) on which verify the model obtained in the studied population. Splitting the available dataset in 'test' and 'training' groups is also not feasible since it would require a greater dataset dimension. Therefore, as we pointed out in the limitations, the external validity had to be deferred to future studies.

6. The assessment of the clinical utility of a new marker is a mandatory step before its employment in the clinical practice and requires an 'ad hoc' study design oriented to the specific characteristics of the marker and of the clinical pathology involved. However, great interest has been raised by the possibility of gather measures of clinical utility from the same dataset used to assess the impact of a new biomarker. Several measures of usefulness have been suggested and gained popularity. These include the Net Reclassification Improvement (NRI) (7), weighted NRI (wNRI) (8), Net Benefit (NB) (9) and Relative Utility (RU) (10). A study comparing the performance of all these indices (11) concluded that the three utility measures that take into account misclassification cost (wNRI, NB and RU) are preferable over the NRI and, "being a mathematical transformations of each other, lead to equivalent information". Notably, the Authors that first introduced the named indices jointly conducted this study. The Authors concluded recommending the use and report of these decision-analytic measures for a range of risk thresholds, thus grounding the deduction over a meaningful range of risk.

We, therefore, adopted the NB plots of different models over a wide risk range to enlighten the utility of the new biomarker.

As regards the way to illustrate graphically the effect of the factor of interest on the outcome (survival), since the stratified Kaplan Meir graphs cannot take into account the presence of confounding, survival curves adjusted by influential covariate have to be employed.

We adopted the directly adjusted survival curve method (12), namely, for each subject in the data set a survival curve is computed using the estimated Cox model. Each curve is obtained using the covariate values specific of each subject except for the factor of interest that is set to a given value for all curves. An average curve is then computed. This curve represent an estimate of the Kaplan Meier survival curve that would be observed if all subjects in the study population had had the factor of interest at the chosen value. Actually, if the process is carried out using the values observed for each subject for all covariates (including the one of interest) the curve obtained indeed 'is' the overall Kaplan Meier.

Thus a plot of ‘directly adjusted curves’ at the appropriate factor of interest values along with the standard overall Kaplan Meier will give graphical view of the strength of the effect of the factor of interest independent from confounding.
STATA (v.13.0) was used to perform all analyses.

Supplemental Table I. Adjudicated causes of death and rates of death by cause

	N	% of pts	% of deaths
All Cause Deaths	131	51.0%	
Cardiovascular deaths	102	39.7%	77.9%
Sudden	42	16.3%	32.1%
Worsening HF	26	10.1%	19.8%
Fatal Stroke	5	1.9%	3.8%
Other Cardiovascular	29	11.3%	22.1%
Non Cardiovascular	27	10.5%	20.6%
Infection	14	5.4%	10.7%
Malignancy	11	4.3%	8.4%
Other	2	0.8%	1.5%
Unknown	2	0.8%	1.5%

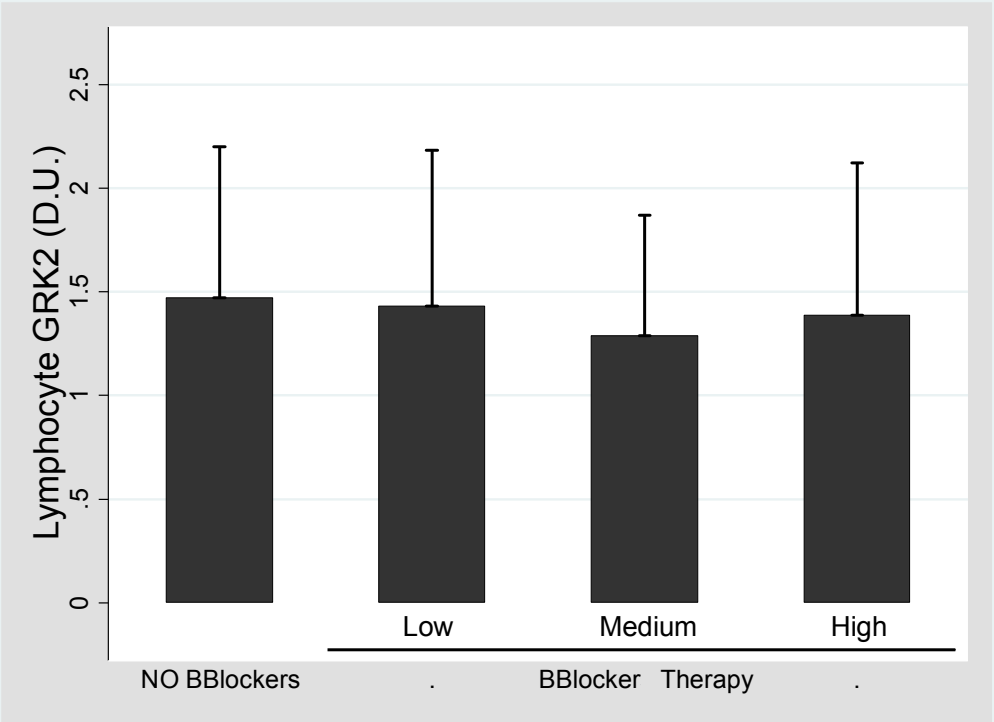
HF= heart failure; MI= myocardial infarction ; pts=patients.

Supplemental Table II. Beta blocker dose distribution in patients with lymphocyte GRK2 below and above the median value.

	Beta Blockers therapy dose		
	Low	Medium	High
Pts below lymphocyte GRK2 median value (≤ 1.31 D.U.), % (n)	44.8% (30)	43.3% (29)	11.9% (8)
Pts above lymphocyte GRK2 median value (> 1.30 D.U.), % (n)	50.0% (33)	39.4% (26)	10.6% (7)
	p=0.833		

GRK2= G protein-coupled receptor kinase 2.Pts= patients

Supplemental Figure I.



Lymphocyte GRK2 levels in patients not assuming beta-blocker therapy and in patients at low, medium and high doses of beta-blocker therapy. BBlocker= beta-blocker

Supplemental References

1. Royston P, Sauerbrei W. Multivariate model building. A pragmatic approach to regression analysis based on fractional polynomials for modeling continuous variables. Chichester, UK: Wiley;2008.

2. Royston P, Ambler G, Sauerbrei W. The use of fractional polynomials to model continuous risk variables in epidemiology. *Int J Epidemiol* 1999;28:964–974.

3. Shorrocks AF. Decomposition procedures for distributional analysis: a unified framework based on the Shapley value. *J Econ Inequal* 2013;11:99–126.

4. Grønnesby, J. K., & Borgan, O. A method for Checking Regression Models in Survival Analysis Based on the Risk Score. *Lifetime Data Analysis* 1996;2,315-328.

5. Gonen, M., G. Heller. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika* 2005;92:965–970.

6. Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Statistics in Medicine* 2004;23:723–748.

7. Pencina MJ, D'Agostino RB Sr, D'Agostino RB Jr, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med*. 2008; 27:157–72.

8. Pencina MJ, D'Agostino RB Sr, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med.* 2011;30:11–21.

9. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26:565–74.

10. Baker SG. Putting risk prediction in perspective: relative utility curves. *J Natl Cancer Inst.* 2009;101:1538–42.

11. Van Calster B, Vickers AJ, Pencina MJ, Baker SG, Timmerman D, Steyerberg EW. Elaluation of markers and risk prediction models: Overview of relationships between NRI and decision-analytic measures. *Med J Decis Making.* 2013;33:490-501.

12. Nieto FJ, Coresh J. Adjusting survival curves for confounders: a review and a new method. *Am J Epidemiol.* 1996;143:1059-68.