A systems biology study to tailored treatment in chronic heart failure
Ouwerkerk, W.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 10

GENERAL DISCUSSION AND FUTURE RESEARCH
Introduction

Part I

The aim of the first part of this thesis was to look at individualized treatment of heart failure patients. We tried to select patients at low and high risk of mortality and/or heart failure-related hospitalization, and patients who were likely to achieve recommended pharmaceutical treatment doses or not. First, in Chapter 2 we found that recently published prediction models exhibited only moderate performance on mortality, and performed poorly on heart failure hospitalization, using a great variety of variables. We therefore developed our own prediction models in Chapter 3, predicting mortality, heart failure hospitalization and the first occurrence of heart failure hospitalization or death. We developed three models for each outcome parameter, namely: a full model consisting of variables selected in more than 40% of 5,000 bootstrap samples; a reduced model with the five best predictor variables from the full model (ten for the first occurrence of heart failure hospitalization or death) and a simplified risk-score model based on the reduced model. The risk score counts the variables that meet a pre-specified value. We validated our models in a separate independent cohort. The validation found to have similar performance as those found during model development. Our model performances were comparable to other published models.

We also developed two models to select patients who were likely to achieve higher doses of angiotensin-converting-enzyme inhibitors (ACE-inhibitors) and angiotensin II receptor blockers (ARBs) and beta-blockers (Chapter 4). We were able to confirm that achieving higher ACE-inhibitor, ARB and beta-blocker doses resulted in better survival. However, despite the fact that physicians in the BIOSTAT-CHF project were encouraged to treat patients according to European Society of Cardiology (ESC) guidelines, the number of patients reaching ESC-recommended ACE-inhibitor, ARB and beta-blocker doses was found to be lower than reported in large randomized controlled trials. In addition, we found that patients not achieving recommended doses because of drug intolerance had worse survival than patients who did not reach recommended doses for other reasons.

In Chapter 4 we saw that the number of patients achieving recommended ACE-inhibitor/ARB and beta-blocker doses was low. This implied that only few patients fully benefited from ACE-inhibitor/ARB and beta-blocker treatment. In Chapter 5 we found that the decision to start or desist from up-titration - based on a biomarker treatment-selection model - was more successful than the scenario to start up-titrating all patients. However, this benefit was a minor one.

Part II

The purpose of the second part of this thesis was to look at the pathophysiology of heart failure. Heart failure is a heterogeneous symptomatic disease, with several underlying causes and involving a multitude of pathologies. The human body has the ability to counter many of the processes involved in heart failure progression. Therefore, it remains difficult to gain an understanding of the full pathobiology of heart failure.

Cluster methods are commonly used to group heart failure patients, based on several, mostly clinical, characteristics. These methods can serve to create clinically distinct groups of patients, hopefully with a comparable pathogenetic background of heart failure. In Chapter 6 however, we found that these cluster methods lacked robustness. When using cluster methods to search for different patterns in a population, one must keep in mind that different cluster methods are likely to produce very different results. We could see this in the number of clusters and the reproducibility of clusters in a different patient cohort. In Chapter 7 we conducted our
own cluster analysis. We clustered not on clinical phenotypes as in previous published and criticized papers, but on pathophysiological determinants. Clusters based on this data were found to be clinical relevant, could be reproduced in a separate independent cohort, and could be characterized by a small number of parameters.

In Chapter 8 and 9 we developed two methods for analyzing data to help in the discovery of new mechanisms involved in heart failure that could result in new targets for heart failure treatment. First, we developed a probabilistic formulation for the alternative splicing problem in Chapter 8. Secondly, we developed a penalized canonical correlation analysis method in Chapter 9, which enabled us to analyze multiple high-dimensional data sets and present the results in an elegant and relatively simple manner.

This chapter discusses the results of both part I and II separately. First, the discussion of part I starts with a deliberation on the differences and similarities between the models predicting mortality and/or heart failure-related hospitalization and those models dealing with recommended ACE-inhibitor/ARB and beta-blocker doses. We also look at variables found in the meta-analysis and used in our prediction models. Secondly, we discuss the implications and opportunities of the developed methods and results found in the chapters of part II. Finally, we propose potential scenarios for future research and possibilities to improve our developed methods.

Part I

Prediction of mortality and heart failure hospitalization

In Chapter 2, we reviewed a large number (117 models in 55 papers) of prediction models. These models used a wide variety of variables (249 different variables in total). In the meta-analysis of this chapter, we found similar predictor variables as previously published reviews [147–150], and confirmed that the performance of mortality prediction was better than that for heart failure hospitalization, or the combined endpoint of heart failure hospitalization and death. Previous reviews focused on heart failure hospitalization and, in general, only compared the differences between the characteristics of prediction models, while in Chapter 2 we focused on quantifying the performance of prediction variables and models.

The predictive performance for heart failure hospitalization was lower than that for mortality. It is likely that, despite the use of different predictors, either relevant predictors of heart failure hospitalization are lacking in prediction models, or that non-medical factors play a major role in heart failure hospitalization risk. Co-morbidity, frailty, community heart failure services, ability to manage lifestyle and medications, social support networks and cultural factors may also all play an important role in becoming hospitalized for heart failure [136,212–215].

In Chapter 3 we developed new prediction models. These models performed as well as previously published models. In contrast to those existing models, however, we followed the Transparency Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement for transparent reporting. We also validated our results in a separate, independent cohort. Our models were also applied to heart failure with reduced ejection fraction (HFrEF) and heart failure with preserved ejection fraction (HFP EF) sub-populations. The prediction models performed similarly in patients with either HFP EF or HFrEF.

The multitude of countries in the BIOSTAT-CHF study, all with their own health-care systems, renders the patient population more heterogeneous. We think that our results are therefore sufficiently generalizable. This perception of generalizability was strengthened by the comparable prediction performance in the validation cohort and the lack of difference in prediction performance between HFP EF and HFrEF patients. The advantage of generalizability outweighed the advantages that would be achieved when we corrected for the effect of country-specific differences
in our models.

We think that our risk-prediction models are useful to predict prognosis in a generic heart failure population, as typically found in daily clinical practice throughout Europe. We followed TRIPOD for transparent reporting. We recruited patients from cardiology clinics in Europe. We had comparable results between index and validation cohorts and in HFpEF and HFrEF patients. This was not the case in the majority of previously published models.

In our model development procedure, we did not include all variables measured in BIOSTAT-CHF. Some variables, despite being good predictors, were not available in the validation cohort (e.g. troponin, B-type natriuretic peptide (BNP) \[88\]–\[380\]), others were new biomarkers in heart failure and not used in most laboratory tests in daily clinical practice. We concentrated only on the variables that would be easily available in the standard clinical setting. Furthermore, our meta-analysis in Chapter 2 found the presence of cancer and race to be important predictor variables. These variables were not selected in our models, however, because they were only infrequently observed in BIOSTAT-CHF: 3.8% of patients had current malignancy, and 1.1% was of non-Caucasian origin. This was also the case with New York Heart Association class (NYHA class), where 2% and 12% had NYHA class I and IV, respectively. Left ventricular ejection fraction (LVEF) was another good predictor variable in the meta-analysis, but was not selected in any of the developed models. This is due most likely to the strict inclusion criteria of the index cohort, where patients were required to have LVEF ≤40% or BNP or N-terminal pro-B-type natriuretic peptide (NT-proBNP) levels >400 pg/mL and/or >2,000 pg/mL, respectively.

Blood urea nitrogen (BUN), on the other hand, was selected in the mortality models. BUN is known to be a good predictor for outcome in heart failure and could be more than a predictor solely representing renal function. Unfortunately, BUN was not selected for predicting heart failure hospitalization. In these models, estimated glomerular filtration rate (eGFR) was selected as predictor. eGFR was also selected in the up-titration models of Chapter 4. Because BUN and eGFR were highly correlated, the inclusion of eGFR in the heart failure hospitalization models could explain the absence of BUN, and vice versa. The relationship between eGFR and prognosis has been previously reported. For calculating eGFR, we used the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation. In this equation, creatinine has a central position and it is therefore highly correlated with eGFR. Creatinine itself is one of the predictors that had good performance in the meta-analysis, and was also selected as a good predictor for the treatment-selection models addressed in Chapter 5.

While peripheral edema was selected in our hospitalization and combined endpoint models, it was not, according to our meta-analysis, a frequently used variable. Finding edema as a marker of increased risk of hospitalization but not of mortality supports the idea that their underlying pathology may differ significantly.

Since our models were primarily developed for use in daily clinical practice, our aim was to make sure the variables were easily obtainable. We also limited the number of predictor variables in our models, thereby possibly limiting the performance of those models. The reduction in performance from full to compact models (mean difference = 0.02) was minor, suggesting that the model performance was primarily based on the compact models’ small set of predictor variables. With the use of new heart failure biomarkers, however, performance might improve. Adding co-morbidity, frailty and markers predictive for community heart failure services, lifestyle management and social support networks might improve prediction as well.

**Treatment**

As previously reported, survival did improve significantly when patients achieved recommended treatment, and higher doses also resulted in better survival. Achieving recommended doses for both ACE-inhibitor/ARB and beta-blocker yielded
better survival than did reaching recommended doses for either of the two. However, the number of patients achieving these recommended doses for either ACE-inhibitor/ARB or beta-blocker was lower than expected, based on previous studies. Patients in these studies had more mild and moderate heart failure than BIOSTAT-CHF patients, and studies were conducted frequently in a clinical trial setting. This may have resulted in an overestimation of the success rate of achieving recommended treatment. Kalra et al.\cite{155} found that 42.8% patients achieved \( \geq 50\% \), and 17.8% \( \geq 100\% \) of the recommended beta-blocker dose in a UK primary care cohort. These results are similar to those we found in our study, and might prove more realistic in daily clinical practice.

Because BIOSTAT-CHF was not a randomized controlled trial but a prospective cohort study, we tried to correct for indication bias in Chapter 4 using four well-established methods.\cite{235, 236, 237, 238} In Chapter 5, we used only two of these methods to correct for indication bias. All methods gave similar results, which encouraged us in the belief that we had adequately corrected for indication bias. Although the methods to correct for indication bias have been well-defined and regularly used in such studies, there was as yet no method to test if the correction was sufficient.

Differences found between European countries in terms of up-titration could reflect differences in national health systems, local practice or differences in patient characteristics. Geographic differences play an important role in both ACE-inhibitor/ARB and beta-blocker dose up-titration, regardless of BIOSTAT-CHF encouragement. It is known that there are differences in mortality and hospitalization between European countries.\cite{381, 382} Apparently, there are also differences in medication prescription and adherence, despite ESC treatment recommendations.

Alkaline phosphatase, as predictor for ACE-inhibitor/ARB dose and mortality, was not previously associated with ACE-inhibitors/ARBs. Elevated alkaline phosphatase could be a sign of liver dysfunction. Some of the ACE-inhibitors/ARBs (enalapril, ramipril, fosinopril,trandolapril, quinapril, benazepril, moexipril, and losartan) are prodrugs, and require transformation by the liver into active metabolites. With liver dysfunction, decreases in prodrug transformation and inactivation of active drugs may occur.\cite{254, 255} Another characteristic of alkaline phosphate might be as a marker for higher abdominal pressure, which in patients with liver dysfunction can reduce tolerance of ACE-inhibitor/ARB. Both of these conjectures, however, are speculative, and the literature provides little evidence.

Our meta-analysis in Chapter 2 showed both systolic blood pressure (SBP) and diastolic blood pressure (DBP) to be good predictors of mortality and hospitalization. Blood pressure is a well-known predictor for heart failure.\cite{101, 383} Only DBP was selected in the final beta-blocker dose model. Heart rate was another predictor of beta-blocker dose, which was probably related to the ESC guidelines which recommend a reduced beta-blocker dose when patients have a low heart rate (<50 beats/min) or asymptomatic low blood pressure and increasing congestion.\cite{101}

BIOSTAT-CHF was specifically designed to investigate why patients do or do not achieve recommended doses. We compared survival of patients divided into three groups:

a) those who reached the recommended dose

b) those who did not reach the recommended dose because of symptoms, side effects or non-cardiac organ dysfunction

c) those who did not reach the recommended dose because of other/unknown/not specified reasons

In only a minority of patients were recommended doses not reached due to intolerance to the drug, either because of organ dysfunction or for reasons related to symptoms and/or side effects. Patients who did not achieve recommended doses because of these reasons had the highest mortality rate; this result was supported by other studies.\cite{259}
It was remarkable that, despite all efforts, no clear explanation was given for the majority of patients who did not achieve recommended dose. There could be many reasons why this was not reported; perhaps the up-titration period was too short, perhaps patients were not compliant with treatment, or perhaps physicians did not comply with recommended guidelines. The only certainty is that the exact reason is lacking.

In Chapter 3 we found that there were some biomarkers predictive for mortality and/or hospitalization in both patients who were successfully and unsuccessfully up-titrated to ≥50% of recommended ACE-inhibitor/ARB or beta-blocker dose (BUN, fibroblast growth factor 23 (FGF-23), NT-proBNP and pro-enkephalin (pro-ENK)). Other biomarkers were predictive only in either successfully or unsuccessfully up-titrated patients to ≥50% of recommended ACE-inhibitor/ARB or beta-blocker dose. This seems to imply that predicting mortality and/or hospitalization in up-titrated patient could be different than predicting mortality and/or hospitalization in not up-titrated patients: this may not be entirely unexpected, seeing as biomarkers related to ACE inhibition/ARB and beta-blocking pathways are likely to change substantially as a result of up-titration.

With c-statistic values of 0.72 and 0.71 in up-titrated ACE-inhibitor/ARB and beta-blocker patients and 0.70 and 0.72 in not up-titrated patients, respectively, we found high c-statistic values for predicting death and/or heart failure hospitalization. In our meta-analysis of Chapter 2, we found an average c-statistic of 0.68. Our multivariate prediction models of Chapter 3 achieved c-statistic values of 0.71 and 0.69 for the full and compact models. Our biomarker based treatment-selection model in Chapter 5 did perform as well as our mortality and/or heart failure hospitalization prediction model in Chapter 3. In addition, it showed that survival in the biomarker scenario was favorable to the treat-all patients scenario. However, the gain was limited. Another important finding was that only one in 5 patients did not benefit from ACE-inhibitor/ARB up-titration, but half the patients from beta-blocker up-titration.

**Part II**

Heart failure is known to be heterogeneous in nature, in terms of etiology, prognosis and response to therapy. Cluster methods have been frequently and successfully applied to patient groups, based on clinically meaningful phenotypes. The use of cluster algorithms has been criticized for its lack of robustness and reproducibility. In Chapter 6, we compared four established cluster methods (gaussian mixture for model-based clustering (Mclust), polynomous latent class analysis (poLCA), partitioning around k-medoids (PAM) and hierarchical cluster analysis (Hclust)). All methods found very different clusters in the same patient population. In contrast to previous published results, inclusion criteria in index and validation cohorts of BIOSTAT-CHF were not totally identical. This resulted in cohorts that were not fully comparable. Reproducibility, in both the index and validation cohort, of these clusters was more difficult in gaussian mixture for model-based clustering (Mclust) and partitioning around k-medoids (PAM), polynomous latent class analysis (poLCA) and hierarchical cluster analysis (Hclust) were best able to select patients in most often in the same cluster with comparable patients. Based on the given clinical variables, all methods were able to groups patients with differences in survival. Patients with the worst and best prognosis in all methods were older, had higher NYHA classes, smoked or used to smoke more often, had more often heart failure with ischemic etiology and comorbidities. They also had lower hemoglobin, sodium, eGFR, albumin, and aspartate amino transferase (ASAT) levels, and higher NT-proBNP levels. There was no significant difference in other biomarker values. Most of these variables were also found to be good predictor in heart failure prediction models as was shown in Chapters 2 and 3. With the given variables all cluster methods found at least one cluster with high NT-proBNP, high NYHA class, older age
and disturbed kidney function. This observation confirmed the observation of previous published data [154, 159, 163, 312, 313] that different clustering methods will give different results and clustering based on phenotypical data will find similar clusters across the heart failure phenotype. Given these results we proposed that clinical relevant and robust clusters can only be identified by a step-wise approach:

- The amount of input data and missing data should be considered together with imputation
- A robust approach to deal with data redundancy should be used
- Extending on recommendations by Wiwie et al, key clustering algorithms including Hclust, PAM and poLCA should be compared for robustness, independent reproducibility as well as clinical meaning of the results

Clustering based on phenotypical data will find similar clusters across the heart failure phenotype. This suggests that clustering on phenotypical data might actually show clusters of signs and symptoms and not distinct disease subtypes.

We applied this approach in Chapter 7, where we used a panel of 92 biomarkers without missing values. We dealt with redundancy by doing a principal component analysis (PCA). We used only the first 10 principal components (all eigenvalue >1). Clustering on these components resulted in 8 distinctly different endotypes, with differences in biomarker levels; survival; up-titration rates; and interaction between up-titration and survival. We also found these endotypes in an independent validation cohort. These endotypes, however, did not differ greatly on classical heart failure status (HF rEF/heart failure with mid-range ejection fraction (HFmrEF)/HFpEF), suggesting that using LVEF for heart failure classification might not provide for the best distinction between heart failure subgroups. We also found two endotypes that had better ACE-inhibitor/ARB up-titration rate. This was probably due to lower rates of chronic kidney disease and diabetes mellitus, which are known to limit up-titration of ACE-inhibitors/ARBs. Interestingly, one endotype, with high levels of chronic kidney disease and diabetes mellitus, did benefit more from ACE-inhibitor/ARB up-titration compared to other endotypes. For beta-blockers, one endotype was of particular interest. These patients did not derive any benefit from beta-blocker up-titration. This suggests that the biomarker identifying this endotype (chitotriosidase-1 (CHIT1)), might play a role in identifying patients not responding to recommended beta-blocker treatment.

In Chapter 8 we studied gene expression of specific genes relevant for Marfan syndrome. We developed a restricted model to predict splicing variants. In particular, we assumed that exons present in a specific splice variant have the same expected expression level. Our model differs from the less restrictive normal mixture model (identical to the Mclust method in Chapter 6) that we used as positive control. The restrictions we applied were based on the biological process of synthesis of a ribonucleic acid (RNA)-molecule. We assumed that the frequency in which premRNA is transcribed from the deoxyribonucleic acid (DNA) is identical for each pre-messenger ribonucleic acid (mRNA), regardless of the splice variant. Pre-mRNAs still include both introns and exons and are identical for each splice variant. The exons to be retained in the mRNA are determined during the splicing process. The expression of each exon is determined by the number of times the exon is included in each splice variant, and the amount of splice variant produced. In our model, we tried to estimate these parameters.

For small genes, we analyzed all possible combinations of exons. This resulted in a set of $2^E$ possibilities, with $E$ being the number of exons of the gene. The set of splice variants included biologically highly improbable splice variants. The entire set of splice variants was relatively small; the computational implications were therefore minor and the biologically unlikely splice variants were given a low prevalence in the simulations and real data analyses. When the gene
size increased, and the number of possible combinations increased to a non-computational size, we turned to a scenario-based method. In this method, we systematically searched for splice variants based on predefined rules. Using this method, it was not possible for us to analyze all possible splice variants; we could, however, artificially reduce the number of variants by removing the biologically improbable splice variants.

At this moment, we do not know the impact of splice variants on the function and/or structure of proteins. We hypothesize that the splice variants most common in SLC2A10 and TGFβR2 are transcribed in non-functional proteins, and therefore alter expression of the TGFβR2 pathway. The most common splice variant for FBN1 in our sample was a variant without exon 25. Mutations in this specific exon are well known to be associated with neonatal Marfan syndrome, which is the most severe Marfan syndrome form, but the lack of exon 25 expression would seem to be important for adult Marfan syndrome as well.

In Chapter 9 we developed a penalized canonical correlation analysis (pCCA) method that is capable of dealing with multiple high-dimensional data sets. Canonical weights are calculated in such a way that maximal correlation is found between sets or canonical variates representing pathogenic pathways. Variables with high weights represent clusters of strongly associated variables. Because we were able to incorporate the elastic net penalty, as mentioned by Friedman et al. and van der Kooij, which enabled us to do the estimations in parallel, we were also able to overcome the computational calculation time with the sequential method proposed by Zou and Hastie, and thereby also reduce memory load.

One of the disadvantages of canonical correlation analysis (CCA), and of our penalized canonical correlation analysis (pCCA), is that it estimates linear relations between variables and its canonical variate, and is specifically developed for quantitative variables. Genomic markers are not quantitative, but non-linear canonical correlation analysis (CCA) approaches do exist for qualitative variables, and the implementation of non-linear genomic data in CCA has already been proposed. In our pCCA method, the properties of genomic data are ignored. Each genomic marker (single nucleotide polymorphism (SNP)) has three possible values or genotypes: a) the common allele (wild type); b) heterozygous; and c) the less common allele (homozygous). These genotypes can have an additive, dominant, recessive or constant effect; this knowledge is important in the ordering and transformation into a quantitative scaled version of the marker. This transformation procedure has to be repeated in each iteration step of the optimization algorithm. This makes the algorithm in its current form impracticable, because of the excessive computation time it would require.

**Future research**

This thesis developed models for predicting mortality, heart failure-related hospitalization and the combined outcome of heart failure hospitalization or death, using data measured in sub-optimally treated patients, with an anticipated initiation or up-titration to pharmaceutical ESC-recommended treatment. In these models we used only a relatively small set of readily available baseline variables. Model performance could be improved if the models were extended with new cardio-vascular markers, measured in BIOSTAT-CHF. We can use the prediction models as a baseline model, and see which novel cardio-vascular markers may improve risk prediction.

BIOSTAT-CHF was specifically designed to use a systems biology approach to combining clinical, laboratory, genomic and proteomic data, in order to find different pathologies involved in heart failure with respect to treatment response and outcome. We demonstrated our pCCA method with only 7 phenotypic variables. In future analyses, we are planning to extend the analyses towards other phenotypic variables. For now, however, we have only looked at mechanisms involved in mortality, heart failure-related hospitalization and response to treatment. Because
heart failure is a heterogeneous disease with many pathologies involved, we want to determine these different pathologies using the data of the BIOSTAT-CHF project. We are therefore interested in other phenotypic outcome parameters. Current ESC guidelines draw a distinction between HFrEF, HFrEF, and an intermediate HFmrEF. It would be interesting to locate differences in pathologies between these groups. We would also like to investigate gender differences and other phenotypes.

Until now, BIOSTAT-CHF has only measured genetic markers (SNPs), proteomic, metabolomic and phenotypic variables. In the future, gene-expression in BIOSTAT-CHF patients will also be measured. This would enable us to calculate new splice variants involved in heart failure patients using our splice variant estimation model. But it would also complement pathway analysis, and enable us to add these markers in our pCCA analyses. If gene-expression is not yet measured, we could also estimate gene-expression values using PrediXcan. PrediXcan is able to predict genes most likely to effect heart failure outcomes, by establishing the relationship between genotype, gene-expression levels from large-scale transcriptome studies, and its associations from genome wide association study (GWAS) studies. It is a method that is able to account for gene-regulation mechanisms.

We have optimized calculations in the pCCA method for linear relations between multiple data sets. Currently, we are not taking into account the characteristics of genetic markers (wild type, heterozygous and homozygous). We will implement a non-linear version of our pCCA method, using an adaption of the method proposed by Waaijenborg and Zwinderman. This method transforms each SNP into a quantitative variable, while retaining the knowledge of SNPs with an additive, dominant, recessive, or constant effect. This scaling procedure is computationally intensive, and could not yet be implemented in the current version of the pCCA method.

Another feature we hope to implement into this method is the use of longitudinal data to analyze disease progression in relation to other factors (e.g. genetic, genomic, proteomic). Waaijenborg and Zwinderman have proposed a method for incorporating longitudinal data that could be implemented in our pCCA method. They proposed the use of mixed-effects models, which deals with intra-subject correlation by allowing random effects in the models, and which focuses on both population-average and individual profiles.