

8.1 INTRODUCTION

Treatment selection based on average effects observed in potentially heterogeneous entire target populations (i.e., one-size-fits-all) masks variation among patients and may result in suboptimal decision making. Personalized medicine aims to improve outcomes by providing more tailored and individualized health care to patients. This however requires more extensive patient profiling and adds to costs. More elaborate profiling has a cost, and, moreover, treatment as such may become more costly given that the price of medication is likely to increase as target populations decrease. Evaluating whether the expected gain in clinical outcomes outweigh the cost of additional diagnostic or prognostic tests is therefore essential if one wants to optimize the use of the scarce health care resources. The novel modeling approaches introduced in this thesis can both be applied at late-stage to support health policy makers in determining whether a new biomarker test should be reimbursed, and at early-stage to assist technology developers in making considered investment decisions.
8.2 SUMMARY OF THE MAIN FINDINGS

8.2.1 Development of risk-stratified treatment recommendations

The first part of this thesis consists of two studies to illustrate how a risk-stratified treatment recommendation\textsuperscript{10} may be developed. Chapter 2 presents a specific example of how to develop a risk prediction model\textsuperscript{11-13} to estimate patient-level risk. The clinical outcomes of interest were the 10-year cumulative incidences of new onset heart failure (HF) with reduced and preserved ejection fraction (HFrEF and HFpEF). Twenty-one risk factors including several biomarkers were included as prognostic covariates. A graphical display of the mean predicted risk against observed risk within each decile of predicted risk indicated a good calibration for both HF outcomes. The discriminative accuracy of this model was moderate for the HFrEF outcome with a value of the c statistic of 0.70 and poor for the HFpEF outcome with a c statistic of 0.60. This study indicated that it is currently still very difficult to predict new onset of HF, even when using a multivariable risk prediction model that already incorporates several biomarkers.

In chapter 3, we introduced an approach to develop a risk-stratified treatment recommendation for a randomized controlled trial after synthesizing patient-level data from different sources of evidence. Such evidence includes the predicted risks captured from multivariable risk prediction model and the net monetary benefit (NMB) estimates.\textsuperscript{14} The risk threshold to generate the treatment recommendation can then be explored and determined using Subpopulation Treatment Effect Pattern Plot (STEPP)\textsuperscript{15,16}, a novel graphical methodology to explore the treatment cost-effectiveness modification along with the continuum of patient-level risk. We illustrated this methodology in a case study of HF disease management. Our proposed risk-stratified treatment recommendation significantly improved clinical outcome and reduced costs compared to the strategies to treat all patients with the same management programs. This finding indicates the potential value of personalized medicine in HF care, i.e., builds a case for optimal decision making.

8.2.2 Quantification of the added value of including biomarkers

The second part of this thesis consists of four studies in which the potential added value of biomarkers in personalized HF care was examined. First, chapter 4\textsuperscript{17} investigated the clinical determinants and added prognostic value of HE4, which is a novel biomarker that has not
previously been described in HF. This biomarker was found to be strongly associated with HF severity and outcome. The association was independent of other established risk factors for poor outcome in HF. The addition of HE4 to commonly used clinical parameters to predict HF-related outcomes resulted in a significant reclassification and improved stratification when estimating both continuous net reclassification improvement (NRI) and integrated discrimination improvement (IDI). This improvement was comparable to the improvement achieved when brain natriuretic peptide (BNP) was added to the same set of clinical parameters. It indicates that HE4 is a strong and independent prognostic biomarker for HF-related outcomes in our study population. A potential clinical utility and cost-effectiveness of HE4 or a panel of biomarkers includes HE4 to improve HF disease management needs to be further evaluated. Ideally, such evaluation of multiple biomarkers needs to be conducted when there is evidence available from the studies structured similar as this one to evaluate the added predicted value of each biomarker.

To support the quantification of the potential added value of biomarkers from a health economic perspective, chapter 5 introduced the reader to the continuous-time semi-Markov model. Compared to the generally applied Markov models, this type of model includes the time dependency between the elapsed time within the current health state and the future state transition. This inclusion results in a more parsimonious reflection of patients’ disease progression when suffering from complex medical procedure. Importantly, a first introduction of vertical modeling to specify health economic models was presented. The main advantage of this formulation compared to others is that it contains more interpretable model entities. Thus the parameter estimation and evaluation can be conducted in a more straightforward way while computational complexity remained limited. The use of vertical modeling may be recommended when patient-level data is available from clinical trials, and when the elapsed time within each health state is either observed exactly or subject to right-censoring.

A direct application of vertical modeling in early-stage health economic modeling was presented in chapter 6, where we set out to determine the commercial headroom available for a novel point-of-care testing device in the disease management of HF. Probability elicitation (PE) was adopted to capture the distributions of experts’ opinions with regard to the anticipated clinical performance of this novel medical technology. Such distributions contributed to generate the distribution of the commercial headroom available to inform further investment decision making. It was concluded that compared to the current care situation, in which the patients were followed by cardiologist and HF nurses, it may not
be economically profitable to further invest on this device prototype assuming an intended daily measurement frequency within five years.

Applying and extending a well-structured innovative approach\textsuperscript{32}, chapter 7 evaluated the potential cost-effectiveness of biomarkers in personalized HF disease management. Two multivariable risk prediction models were firstly developed to estimate the patient-level risk with or without the incorporation of biomarkers. A significantly ($P<0.001$) improved ability of risk classification contributed by these biomarkers was detected after evaluating the NRI. Our previously developed risk-stratified treatment recommendation (from chapter 3) was adopted to identify the personalized management strategies based on the risk predicted by both models. Early-stage health economic modeling with the same model structure as introduced in chapter 5 was used to evaluate the commercial headroom available of the biomarkers. Such headroom values were found to be considerable to advice further development of a practical test procedure incorporating these biomarkers. The results from different scenario analyses indicated that the estimates of the commercial headroom available were sensitive to different cut-off values assigned to stratify patients into different risk groups.
8.3 METHODOLOGICAL CONSIDERATIONS

8.3.1 Patient heterogeneity in economic evaluation

To support the development of risk-stratified treatment recommendations in a health economic setting, it is crucial to realize that patient heterogeneity, defined as the variation of patients explained by their clinical characteristics\textsuperscript{33-35}, cannot be examined and approached similar to how this is generally done in clinical studies.\textsuperscript{1} In clinical studies, it is common practice to investigate whether such variation among patients persists in treatment effect. Typically, this is expressed in terms of relative risks, odds ratios or hazard ratios.\textsuperscript{4,36} In health economic settings, however, patient heterogeneity may have profound impact on the additional parameters involved, such as resource utilization and health state utility. Ignoring or inadequately acknowledging these parameters may lead to health benefits foregone or suboptimal use of scarce resources.\textsuperscript{4,5} STEPP, the novel graphical method proposed in this thesis, uses the difference of patient-level net monetary benefit (NMB) to measure the cost-effectiveness of different treatment regimens. The threshold to generate the risk-stratified treatment recommendations can then be explored and determined after visualizing how the pattern of differences in NMB varies as a function of the patient-level risk. This therefore resulted in optimal treatment recommendations as such a development process takes into account how to properly examine and approach patient heterogeneity in health economic settings.

8.3.2 Early-stage health economic modeling

Worldwide, large amounts of resources are being invested in the research and development of new medical technology. This poses high and diverse demands from different stakeholders to be informed whether actual returns on these investments may be expected.\textsuperscript{37,38} Governments are particularly concerned about the benefits of spending public resources, while investors need to be informed about the commercial potential of such technologies given regulatory constraints. In line with the latter, academia is expected to focus on issues that would yield maximal societal benefit of public funding. Although more and more new medical technologies are developed, there are only a few launched to the market.\textsuperscript{38} This is because the development phases of these new technologies are quite costly and uncertain. Late-stage evaluations of their cost-effectiveness makes it impossible to rank order and prioritize
innovations and push forward the potentially valuable, i.e., promising ones for further development. Early-stage assessment of the potential cost-effectiveness during the development phases of the new medical technologies is therefore desired. Using such evaluations, it becomes less likely that new technologies ultimately become market failures and more likely to gain access to their intended markets.

A major complication in conducting early-stage health economic evaluations is that evidence regarding the clinical performance of the new technology is usually scarce or even missing. This evidence scarcity leads to high uncertainty in some of the model inputs to evaluate the commercial headroom available. Different approaches exist to deal with the lack of evidence in early-stage health economic modeling. Experts’ beliefs in a new technology’s intended clinical performance can be elicited through semi-structured interviews to quantitatively capture their opinions as probability distributions. Such distributions were then incorporated in the decision model to reflect the model parameter uncertainty and to propagate the probability distribution of the commercial headroom available. The evidence scarcity can also be dealt with by performing scenario analyses with different assumptions of the unknown model parameters. Deterministic sensitivity analysis was then used to investigate how the estimates of commercial headroom available can be influenced in each scenario.

The main advantage of using PE, although in a more complex way compared to the use of the deterministic approach, is that the resulting distribution of commercial headroom available can be included in some formal decision support methods to support product investment decision making. However, as this approach includes the elicitation of subjective opinions and the transformation of such opinions into probability distributions, we only recommend its use when the unknown parameters are relatively easy to interpret. For instance, we found in the try-out interview of our case study that the elicitation of the point estimate and its corresponding uncertainty of disease incidence can be conducted in a more straightforward way compared to the elicitation of such values of relative risk or hazard ratio. Another crucial aspect that needs to be taken into account is that the use of PE depends on to what extent the evidence that is already available can support the experts in speculating the unknown model parameters. For instance, a rather straightforward clinical setting was introduced in our case study that used PE. This includes the clinical evidence available from current practice and the biomarker-guided therapy selected as the comparative treatment regimen with unknown clinical performance. Experts were then provided with some reference
values to reflect the disease progression of the patients who received conventional care (i.e., how many patients have left the current health state within a specific amount of time, how many out of these patients ultimately died). It was shown that this helped them to speculate how the disease progresses when the new technology would be incorporated.\textsuperscript{27,41} On the other hand, instead of PE, the deterministic approach was used in our other case study. The unknown model parameter of that study was the treatment effect of a hypothetical intervention described in terms of hazard ratio. It is nearly impossible to recruit corresponding experts and ask them to make meaningful speculations regarding this parameter based on the reference value from the readily available evidence (i.e., NRI estimation of the added biomarkers, patient characteristics). The difficulty of such an elicitation can be specified when expecting answers on questions like: “In general, to what extent do you think a more intensive management strategy reduces the mortality of HF patients compared to a moderate strategy?”, and subsequently, “Can you quantify this reduction in terms of a hazard ratio?”.

8.4 LIMITATIONS AND FUTURE DIRECTIONS

8.4.1 Development of risk-stratified treatment recommendations

The main limitation of using STEPP to develop a risk-stratified treatment recommendation is that the selected treatment threshold may be data-driven. For instance, the results might be conditional on the value assumed for the willingness-to-pay threshold. This is also one of the debates among health economists regarding the use of NMB in general.\textsuperscript{42} To further verify the determination of the treatment threshold, we therefore recommend either our use of a formal quantification method\textsuperscript{2} or the use of sensitivity analysis, for instance, to repeat the graphical approach for different values of the willingness-to-pay threshold. In addition, patient preference was also suggested to be an important factor to be considered when developing the risk-stratified treatment recommendations.\textsuperscript{1,10} This was not considered in our application of the graphical approach. Future research should therefore be directed towards including formal techniques to elicit benefit-risk preference data from patients and subsequently incorporate such evidence in our proposed graphical approach or approaches alike to ensure more optimized and balanced decision making.

8.4.2 Serial biomarker measurements for dynamic risk prediction

One main limitation of the work presented in this thesis is that all the biomarkers considered were only measured at baseline. Nowadays more and more research suggest to investigate not only single but also serial measurements during follow-up to see if any improved prediction appears for disease occurrence and disease-related outcomes.\textsuperscript{43-47} For example, de Filippi et al.\textsuperscript{43} reported significant improvements in area under the receiver operating characteristic curve when adding follow-up N-terminal pro-B-type natriuretic peptide (NT-proBNP) measurements in the prediction model compared to only incorporating the baseline measurement (from 0.78 to 0.80 for incidence prediction of new HF onset & from 0.79 to 0.81 for prediction of cardiovascular death). Similarly, Motiwala et al.\textsuperscript{46} found a significantly improved prediction of time to first cardiovascular event when comparing the combination of a baseline galectin-3 value and a measurement at 6 months to the baseline measurement alone. A joint modeling framework for longitudinal and time-to-event data has been proposed to assess the added predictive value of serial biomarker measurements.\textsuperscript{48,49} Such a framework is able to display how the trajectory of the variable biomarker values is directly associated with clinical outcomes. This provides more comprehensive and reliable information to investigate the clinical value of a specific biomarker compared to only incorporating a fixed measurement.
8.4 LIMITATIONS AND FUTURE DIRECTIONS

8.4.1 Development of risk-stratified treatment recommendations

The main limitation of using STEPP to develop a risk-stratified treatment recommendation is that the selected treatment threshold may be data-driven. For instance, the results might be conditional on the value assumed for the willingness-to-pay threshold. This is also one of the debates among health economists regarding the use of NMB in general. To further verify the determination of the treatment threshold, we therefore recommend either our use of a formal quantification method or the use of sensitivity analysis, for instance, to repeat the graphical approach for different values of the willingness-to-pay threshold. In addition, patient preference was also suggested to be an important factor to be considered when developing the risk-stratified treatment recommendations. This was not considered in our application of the graphical approach. Future research should therefore be directed towards including formal techniques to elicit benefit-risk preference data from patients and subsequently incorporate such evidence in our proposed graphical approach or approaches alike to ensure more optimized and balanced decision making.

8.4.2 Serial biomarker measurements for dynamic risk prediction

One main limitation of the work presented in this thesis is that all the biomarkers considered were only measured at baseline. Nowadays more and more research suggest to investigate not only single but also serial measurements during follow-up to see if any improved prediction appears for disease occurrence and disease-related outcomes. For example, de Filippi et al. reported significant improvements in area under the receiver operating characteristic curve when adding follow-up N-terminal pro-B-type natriuretic peptide (NT-proBNP) measurements in the prediction model compared to only incorporating the baseline measurement (from 0.78 to 0.80 for incidence prediction of new HF onset & from 0.79 to 0.81 for prediction of cardiovascular death). Similarly, Motiwala et al. found a significantly improved prediction of time to first cardiovascular event when comparing the combination of a baseline galectin-3 value and a measurement at 6 months to the baseline measurement alone. A joint modeling framework for longitudinal and time-to-event data has been proposed to assess the added predictive value of serial biomarker measurements. Such a framework is able to display how the trajectory of the variable biomarker values is directly associated with clinical outcomes. This provides more comprehensive and reliable information to investigate the clinical value of a specific biomarker compared to only incorporating a fixed measurement.
as elevated biomarker levels are normally associated with increased risk of a specific disease or disease-related outcomes. As the serial measurement itself is a costly process, one of the open issues that requires further investigation is that what is the optimal measurement frequency and timing to enable dynamic risk prediction. In the meantime, challenges remain to further assess the potential clinical utility and economic value of serial biomarker measurements in personalized medicine. Future studies may therefore be performed to incorporate the joint modeling technique in the methods proposed in this thesis. However, to follow the previously suggested evaluation phases prior to biomarker’s clinical integration, it is important to firstly conduct a proof of concept study to evaluate the biomarker’s added predictive value. The main hypothesis of such a study is that the incorporation of the serial measurements of one or more biomarkers indeed improves the risk prediction of disease-related outcomes. The study needs to include at least two other comparative risk prediction models: a) a model containing only clinical predictors but with no biomarkers; b) a model containing clinical predictors and additional one or more biomarkers measured at baseline. Subsequently, the optimal model selected from a) and b) can be used together with the model which includes serial biomarker measurements to investigate whether such measurements indeed improves risk stratification, and whether this will result in a realistic gain in the commercial headroom available.

8.4.3 Concluding remarks

This thesis introduced innovative methods to assess the clinical value and commercial potential of biomarkers in personalized medicine. However, as personalized medicine is complex compared to traditional medicine, it may remain challenging to get it reimbursed. For instance, a specific hurdle to enable risk stratification might be the ethical issues that arise when incorporating genomics into screening. Such discussions are equally important compared to the consideration of cost-effectiveness. These other issues, however, go beyond the scope of this thesis.
8.5 REFERENCES
