Towards optimal decision making in personalized medicine
Cao, Qi
Appendix II. List of candidate predictor variables

The following variables were considered as candidate predictors during model building:

- Sex
- Body mass index
- Current smoker or quit smoking <1 year
- Systolic blood pressure
- Diastolic blood pressure
- Heart rate
- History of myocardial infarction
- History of stroke
- History of atrial fibrillation
- Glucose
- Total cholesterol
- High-density lipoprotein cholesterol
- Low-density lipoprotein cholesterol
- Triglycerides
- Cystatin C
- Serum creatinine
- 24-hours urinary albumin excretion
- High-sensitivity c-reactive protein
- N-terminal pro-B-type natriuretic peptide
- High-sensitivity troponin T

Using Subpopulation Treatment Effect Pattern Plot to identify more efficient resource allocation policies: A case study in disease management of heart failure

Qi Cao, Hans L. Hillege, Tiny Jaarsma, Erik Buskens, Douwe Postmus

Submitted
ABSTRACT

Objectives: When cost-effectiveness analyses are conducted alongside randomized controlled trials it is important to acknowledge patient heterogeneity as this may result in more efficient resource allocation policies. In this paper, we sought to explore to what extent the use of Subpopulation Treatment Effect Pattern Plot (STEPP) may facilitate such subgroup analysis strategies.

Methods: The analysis was based on data from the COACH study, in which 1,023 patients with heart failure were randomly assigned to three treatments: care-as-usual, basic support, and intensive support. First, using predicted 18-month mortality risk as the stratification basis, a suitable strategy for assigning different treatments to different risk groups of patient was developed. To that end a graphical exploration of the difference in net monetary benefit (NMB) across treatment regimens and baseline risk was used. Next, the efficiency gains resulting from this proposed subgroup strategy were quantified by computing the difference in NMB between our stratified approach and the best performing population-wide strategy.

Results: The STEPP approach allowed distinguishing between subgroups, i.e., intensive support appeared optimal for low-risk patients (18-month mortality risk ≤ 0.16), while basic support appeared optimal for intermediate to high-risk patients (18-month mortality risk > 0.16). The average gain in NMB resulting from a stratified approach compared to basic support for all was €1,312 (95% CI: €390-€2,346).

Conclusions: A risk-based analysis using STEPP seems promising to explore the impact of baseline risk for the relative cost-effectiveness in optimizing treatment trade-off and subsequently in the quest for more efficient reimbursement policies.
3.1 INTRODUCTION

Cost-effectiveness analysis supports resource allocation decision making by comparing the differences in costs and effects of alternative treatment regimens.\(^1\)\(^-\)\(^4\) When such analyses are conducted alongside randomized controlled trials (RCTs), the cost-effectiveness of the evaluated treatments is generally expressed in terms of population averages. This provides insight into which of the available treatments performs best for the patient population considered. However, when these patients are characterized by a heterogeneous clinical condition, and their risk profiles are determined by factors like demographic variations, biometric variations, and co-morbidities, there may be considerable variation in response. In fact the likelihood of subpopulations for whom response to one or the other treatment is obscured may be substantial.\(^5\)\(^-\)\(^8\) Such differences among patients may also lead to systematic variation in resource use and costs, which could be another reason why one of the other treatments performs better in specific subpopulations.\(^2\)\(^,\)\(^6\) Acknowledging patient heterogeneity in health economic evaluation has therefore considerable potential to contribute to more efficient resource allocation decisions.\(^5\)\(^,\)\(^8\)-\(^10\)

A recently conducted systematic review\(^5\) identified baseline risk, treatment effect, health state utility, and resource utilization as the four input parameters of a health economic evaluation that may be prone to patient heterogeneity. However, as the cost-effectiveness of one treatment compared to another is ultimately determined by the net effect on all these parameters, it is essential that the impact of patient heterogeneity on each of these parameters is considered conjointly rather than in isolation, especially when the purpose is to identify more efficient reimbursement policies. For health economic evaluations conducted alongside an RCT, this can be achieved by conducting such analyses directly in terms of net monetary benefit (NMB).\(^11\)\(^-\)\(^13\)

Hoch et al.\(^14\) have previously proposed to assess the impact that different sources of patient heterogeneity may have on a treatment’s NMB by means of regression analysis. For example, suppose that one wants to explore whether the cost-effectiveness of a new treatment compared to the current standard treatment is affected by the age of the patient. Using regression analysis, this can be achieved by fitting a regression model with NMB as the dependent variable and the treatment indicator, age, and the interaction between age and the treatment indicator as the independent variables. A low p-value for the regression coefficient
corresponding to the interaction term then shows that age has a relatively strong influence on the new treatment’s relative cost-effectiveness.

While the use of multivariable regression models may provide insight into which sources of patient heterogeneity potentially have an impact on the relative cost-effectiveness of the evaluated treatments, the statistical power to detect such interaction effects is usually low. Moreover, actually being able to verify relevant heterogeneity using such models strongly depends on whether the assumed multiplicative structure of interaction fits reality. This may lead to missing or over-interpretation of the detected significant interaction terms.

An alternative approach for studying treatment-covariate interaction that makes no assumptions about the nature of the relationship between the outcome and the covariate in each treatment group is the Subpopulation Treatment Effect Pattern Plot (STEPP) methodology.\textsuperscript{15-17} This is based on a graphical exploration of the fluctuation in treatment effect across different but overlapping subpopulations defined with respect to increasing levels of the covariate of interest. Although the ability to graphically explore how the difference in NMB between two treatments varies as a function of one or more sources of patient heterogeneity could potentially be very useful in identifying more efficient reimbursement policies, such applications of the STEPP have not yet been considered. Using the difference in NMB as the measure of treatment benefit and a patient’s predicted 18-month mortality risk as the covariate of interest, the objective of this paper is to evaluate the potential of using the STEPP to derive clinical decision rules that reflect selection and reimbursement policies based on cost-effectiveness. Specifically, a case study in heart failure (HF) disease management was elaborated.
3.2 METHODS

3.2.1 Study cohort

The data that we used to conduct our analysis was taken from the Coordinating study evaluating Outcomes of Advising and Counseling in Heart failure (COACH), a multicenter, randomized controlled trial in which 1,023 patients were randomly assigned to one of three disease management programs (DMPs). \(^{18,19}\) Patients in the care-as-usual group received routine follow-up management by a cardiologist. Along with routine management by a cardiologist, patients in the basic and intensive support groups received additional care from an HF nurse. In addition, patients in the intensive support group received multidisciplinary advice and 2 or more home visits by the HF nurse. The total follow-up time of the trial was 18 months.

3.2.2 Baseline risk assessment

The patients’ predicted 18-month all-cause mortality risk was obtained from a previously developed multivariable risk prediction model. \(^{20}\) This model included the following 14 predictor variables: age, gender, diastolic blood pressure, systolic blood pressure, history of stroke, history of myocardial infarction, atrial fibrillation, peripheral arterial disease, diabetes, left ventricular ejection fraction, previous HF hospitalization, serum sodium, estimated glomerular filtration rate (eGFR), and N-terminal pro brain natriuretic peptide (NT-proBNP). Missing values on these predictor variables were dealt with using multiple imputation. \(^{21}\) Mortality risk values were then computed by taking the average of the risk values obtained from each of the ten imputed datasets.

3.2.3 Patient-level NMB assessment

The patient-level NMB was calculated as \( NMB_i = \lambda \times e_i - c_i \), where \( e_i \) and \( c_i \) denote the observed effect and cost for patient \( i \) and where \( \lambda \) denotes the willingness-to-pay threshold. \(^{14}\) The patients’ observed survival time, which was censored at 562 days for those who were still alive at the end of the study’s follow-up, was taken as the measure of effectiveness. Costs were calculated at patient level by multiplying the patients’ volumes of resource use with their respective unit costs as described in more detail in our previously conducted economic
Chapter 3

48

evaluation in this patient population.\textsuperscript{22} The willingness-to-pay threshold was set equal to €20,000 per life year.

3.2.4 Exploration of treatment-predicted risk interaction and determination of subgroup strategy

To explore whether an interaction existed between the COACH DMPs and predicted 18-month mortality risk, we applied the STEPP methodology\textsuperscript{15,16} using the difference in NMB as the outcome of interest. STEPP is a novel graphical method for assessing treatment-covariate interaction on different but overlapping subpopulations defined with respect to the covariate of interest. The subpopulations are defined on the basis of two parameters: the number of patients belonging to two adjacent subgroups ($n_1$), and the sample size of each subpopulation ($n_2$). For the analysis conducted in this study, $n_1$ and $n_2$ were set equal to 120 and 150, respectively. Separate STEPPs were created for the difference in NMB between (i) care-as-usual and basic support and (ii) basic support and intensive support. Basic support, which was previously shown to be the optimal population-wide strategy\textsuperscript{22}, was selected as the reference group for both of these STEPPs. Based on the observed patterns of treatment-predicted risk interaction, a suitable strategy for assigning different DMPs to different risk groups of patient was subsequently identified.

3.2.5 Quantification of the efficiency gains resulting from the subgroup strategy

To evaluate the optimality of our proposed subgroup strategy, we quantified its efficiency gains as suggested by Coyle et al..\textsuperscript{8} First, the average NMB was evaluated separately per DMP for each of the established risk categories. Subsequently, the average gain in NMB resulting from stratification compared to the best performing population-wide strategy was calculated as $\overline{NMB} = \sum_{j} \frac{\Delta NMB_j \times n_j}{N}$, where $\Delta NMB_j$ denotes the difference in average NMB between the proposed treatment for subgroup $j$ and basic support, $n_j$ denotes the sample size of subgroup $j$, and $N$ denotes the sample size of the overall study population. In our previously conducted economic evaluation, we suggested that NYHA class could be a suitable basis for offering different treatments to different subgroups of patient. For comparative purposes, the
average gain in NMB resulting from using NYHA class as the stratification basis was computed as well. The 95% confidence intervals (CIs) for all the average estimates were captured from their corresponding 1,000 bootstrap resamples.
3.3 RESULTS

3.3.1 Exploration of treatment-predicted risk interaction and determination of subgroup strategy

The estimated difference in NMB across the overlapping patient subpopulations are depicted in Figures 1 and 2. The difference in NMB between care-as-usual and basic support never reached the 5% significance level (Figure 1), showing that patient heterogeneity did not have a clear impact on the difference in NMB between these two treatments. However, a significant pattern of treatment-covariate interaction was found when comparing the difference in NMB between intensive support and basic support (Figure 2). A risk value of 0.16 was found to be the zero point at which the difference in NMB between these two treatments started to change signs. Based on this finding, our proposed subgroup strategy was to assign intensive support to low-risk patients (patients with predicted risk value ≤ 0.16) and basic support to intermediate to high-risk patients (patients with predicted risk value > 0.16).

3.3.2 Quantification of the efficiency gains resulting from the subgroup strategy

Table 1 depicts the results of the cost-effectiveness analysis within each risk stratum. For the low-risk patients, intensive support was found to be the best performing strategy while basic support performed best in the intermediate to high-risk patients. When NYHA class was used as the stratification basis, basic support was found to be optimal for patients belonging to NYHA class II while care-as-usual was found to be optimal for patients belonging to NYHA class III&IV. Table 2 depicts the average gains in NMB (95% CI) resulting from each subgroup strategy. Both strategies were found to be more cost-effective compared to assigning basic support to the whole patient population. However, the subgroup strategy proposed in this study outperformed the one proposed previously, with an average gain in NMB of €1,174 (95% CI: €-1,146-€3,284).
3.3 RESULTS

3.3.1 Exploration of treatment - predicted risk interaction and determination of subgroup strategy

The estimated difference in NMB across the overlapping patient subpopulations are depicted in Figures 1 and 2. The difference in NMB between care-as-usual and basic support never reached the 5% significance level (Figure 1), showing that patient heterogeneity did not have a clear impact on the difference in NMB between these two treatments. However, a significant pattern of treatment-covariate interaction was found when comparing the difference in NMB between intensive support and basic support (Figure 2). A risk value of 0.16 was found to be the zero point at which the difference in NMB between these two treatments started to change signs. Based on this finding, our proposed subgroup strategy was to assign intensive support to low-risk patients (patients with predicted risk value ≤ 0.16) and basic support to intermediate to high-risk patients (patients with predicted risk value > 0.16).

3.3.2 Quantification of the efficiency gains resulting from the subgroup strategy

Table 1 depicts the results of the cost-effectiveness analysis within each risk stratum. For the low-risk patients, intensive support was found to be the best performing strategy while basic support performed best in the intermediate to high-risk patients. When NYHA class was used as the stratification basis, basic support was found to be optimal for patients belonging to NYHA class II while care-as-usual was found to be optimal for patients belonging to NYHA class III&IV. Table 2 depicts the average gains in NMB (95% CI) resulting from each subgroup strategy. Both strategies were found to be more cost-effective compared to assigning basic support to the whole patient population. However, the subgroup strategy proposed in this study outperformed the one proposed previously, with an average gain in NMB of €1,174 (95% CI: €-1,146-€3,284).

Figure 1: STEPP comparing the difference in NMB between care-as-usual and basic support across different but overlapping subpopulations with increased mortality risk; a difference in NMB > 0 indicates that care-as-usual is the preferred strategy.

Figure 2: STEPP comparing the difference in NMB between intensive support and basic support across different but overlapping subpopulations with increased mortality risk; a difference in NMB > 0 indicates that intensive support is the preferred strategy.
**Table 1: Results of the cost-effectiveness analysis**

<table>
<thead>
<tr>
<th>Patient subgroup (Sample size)</th>
<th>Mean (95% CIs)</th>
<th>Mean (95% CIs)</th>
<th>Mean (95% CIs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>survival time (days)</td>
<td>cost (€)</td>
<td>NMB (€)</td>
</tr>
<tr>
<td>Predicted 18-month mortality</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk ≤ 0.16 (N=321)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Care-as-usual</td>
<td>521.2 (497.5-543.6)</td>
<td>6151 (3961-8826)</td>
<td>22389 (19200-25101)</td>
</tr>
<tr>
<td>Basic support</td>
<td>525.6 (505.2-545.5)</td>
<td>8653 (6117-11664)</td>
<td>20127 (17107-22844)</td>
</tr>
<tr>
<td>Intensive support</td>
<td>557.2 (547.7-562.0)</td>
<td>6213 (4950-7624)</td>
<td>24307 (22857-25577)</td>
</tr>
<tr>
<td>Risk &gt; 0.16 (N=702)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Care-as-usual</td>
<td>428.2 (403.2-451.6)</td>
<td>11175 (9348-13226)</td>
<td>12265 (9791-14496)</td>
</tr>
<tr>
<td>Basic support</td>
<td>454.0 (429.6-480.5)</td>
<td>10041 (8257-11935)</td>
<td>14819 (12608-17071)</td>
</tr>
<tr>
<td>Intensive support</td>
<td>432.3 (406.1-456.9)</td>
<td>13155 (11221-15142)</td>
<td>10525 (7935-12803)</td>
</tr>
<tr>
<td>NYHA class</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYHA II (N=513)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Care-as-usual</td>
<td>481.4 (455.9-504.5)</td>
<td>8955 (6884-11522)</td>
<td>17405 (14318-20026)</td>
</tr>
<tr>
<td>Basic support</td>
<td>506.6 (486.0-528.6)</td>
<td>7170 (5788-8898)</td>
<td>20570 (18607-22250)</td>
</tr>
<tr>
<td>Intensive support</td>
<td>505.7 (484.8-527.0)</td>
<td>9099 (7256-11220)</td>
<td>18581 (16087-20758)</td>
</tr>
<tr>
<td>NYHA III&amp;IV (N=495)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Care-as-usual</td>
<td>428.7 (397.0-462.3)</td>
<td>10692 (8279-13206)</td>
<td>12788 (10112-15948)</td>
</tr>
<tr>
<td>Basic support</td>
<td>443.7 (414.7-471.1)</td>
<td>11793 (9435-14403)</td>
<td>12507 (9465-15219)</td>
</tr>
<tr>
<td>Intensive support</td>
<td>448.9 (422.2-474.8)</td>
<td>12462 (10279-14779)</td>
<td>12118 (9304-14707)</td>
</tr>
</tbody>
</table>
Table 2: Average gains in NMB (95% CIs) resulting from each subgroup strategy

<table>
<thead>
<tr>
<th>Stratification basis</th>
<th>Subgroup strategy</th>
<th>Average (95% CIs) gain in NMB (€)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predicted 18-month mortality</td>
<td>Intensive support to low-risk group; Basic support to intermediate to high-risk group</td>
<td>1312 (390-2346)</td>
</tr>
<tr>
<td>NYHA class</td>
<td>Basic support to NYHA II group; Care-as-usual to NYHA III&amp;IV group</td>
<td>138 (-1854-2246)</td>
</tr>
</tbody>
</table>
3.4 DISCUSSION

We have shown that a risk-based analysis strategy, acknowledging patient heterogeneity, is a promising tool for reimbursement policies to optimize NMB. By using STEPP to graphically explore treatment-covariate interaction, we found that the difference in NMB between intensive support and basic support varied greatly across different but overlapping subpopulations defined with respect to increasing levels of predicted 18-month mortality risk. The difference in NMB between care-as-usual and basic support, in contrast, never reached the 5% significance level. By subsequently selecting the 18-month mortality risk at which the difference in NMB between intensive support and basic support started to change signs as the cut-off to stratify patients into two risk categories, we found that compared to applying basic support to all patients, the use of a stratified approach based on offering intensive support to low-risk patients and basic support to intermediate to high-risk patients would result in an average gain in NMB of €1,312 (95% CI: €390-€2,346).

In our analysis, we first followed recommendations in applying a previously developed multivariable risk prediction model to combine the information captured within several covariates into a single prognostic index to represent baseline risk. We subsequently used this index to explore for heterogeneity in treatment effect across different subgroups of patient. Compared to conventional subgroup analysis based on a single prognostic covariate, integrating multiple independent patient characteristics associated with the outcome parameters of interest in a multivariate risk prediction model improves risk stratification. This, in turn, can greatly enhance the statistical power to detect variations in treatment benefit as was shown in a previously conducted simulation study. Moreover, the use of such a multivariable approach avoids the problem of multiple testing resulting from the need to repeat the subgroup analysis for different individual risk factors. Thus the chances of obtaining false positive findings are reduced.

While treatment-predicted risk interaction can best be assessed on a continuous scale, discretization of the predicted risks into two or more ordinal categories becomes essential if we want to use the underlying risk prediction model to guide the selection of therapy. By deriving the cut-off of 0.16 from the treatment effect pattern observed in a STEPP, we were still able to make effective use of the discriminative power of a continuous prognostic index in our quest for an efficient reimbursement policy. This does not hold when applying conventional subgroup analysis based on a single prognostic covariate as we did as part of our
previously published economic evaluation in this patient population.\textsuperscript{22} When quantifying the net benefit gains of one over the other stratification basis, the subgroup strategy proposed in this study was found to outperform the previous one with an average gain in NMB of €1,174 (95% CI: €-1,146-€3,284).

A limitation of this study is that the cut-off of 0.16 may be specific for the data analyzed in this paper. It was selected by taking into account the pattern of treatment-risk interaction in a single clinical trial. Future research is thus required to determine to what extent this cut-off also shows up in other studies. In addition, rather than using an external model (i.e., a risk prediction model developed on another dataset), we used an internally developed risk prediction model to assess the treatment effect across different subpopulations of predicted risk. The validity of this approach was recently assessed by Burke et al.\textsuperscript{23}, who concluded that “appropriately developed internal models produce relatively unbiased estimates of treatment effect across the spectrum of risk”. In addition, these authors also found that “when estimating treatment effect, internally developed risk models using both treatment arms should, in general, be preferred to models developed on the control population”. As all treatment groups of COACH were included in the development of the COACH risk prediction model, this is exactly the strategy that we have followed in the current paper. Finally, because we selected the difference in NMB as the measure of treatment benefit, our results are conditional on the value assumed for the willingness-to-pay threshold. For actual decision making purposes, it would therefore be recommendable to repeat the STEPP analysis for different values of the willingness-to-pay threshold.

To conclude, a risk-based analysis using STEPP seems promising to explore the impact of baseline risk for the relative cost-effectiveness in optimizing treatment trade-off and subsequently in the quest for more efficient reimbursement policies. Our finding highlights the importance to acknowledge patient heterogeneity in a proper manner when conducting cost-effectiveness analysis alongside RCTs and it underscores the clinical and health economic potential of personalized medicine.
3.5 ACKNOWLEDGEMENTS

This research was performed within the framework of CTMM, the Center for Translational Molecular Medicine (www.ctmm.nl), projects TRIUMPH (grant 01C-103) and ENGINE (grant 01C-401), and supported by the Dutch Heart Foundation. The COACH study was supported by grant 2000Z003 from the Dutch Heart Foundation and by additional unrestricted grants from Biosite France SAS, Jouy-en-Josas, France (brain natriuretic peptide), Roche Diagnostics Nederland BV, Venlo, the Netherlands (N-terminal prohormone brain natriuretic peptide), and Novartis Pharma BV, Arnhem, the Netherlands.
3.6 REFERENCES

PART II

QUANTIFICATION OF THE ADDED VALUE OF INCLUDING BIOMARKERS