Towards optimal decision making in personalized medicine

Cao, Qi

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2016

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Cao, Q. (2016). Towards optimal decision making in personalized medicine: Potential value assessment of biomarkers in heart failure exemplars [Groningen]: University of Groningen

Copyright
Other than for strictly personal use, 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), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Probability elicitation to inform early health economic evaluations of new medical technologies: a case study in heart failure disease management

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

Value in Health 2013;16(4):529-35
ABSTRACT

Objectives: Early estimates of the commercial headroom available to a new medical device can assist producers of health technology in making appropriate product investment decisions. The purpose of this study was to illustrate how this quantity can be captured probabilistically by combining probability elicitation with early health economic modeling. The technology considered was a novel point-of-care testing device in heart failure disease management.

Methods: First, we developed a continuous-time Markov model to represent the patients’ disease progression under the current care setting. Next, we identified the model parameters that are likely to change after the introduction of the new device and interviewed three cardiologists to capture the probability distributions of these parameters. Finally, we obtained the probability distribution of the commercial headroom available per measurement by propagating the uncertainty in the model inputs to uncertainty in modeled outcomes.

Results: For a willingness-to-pay of €10,000 per life year, the median headroom available per measurement was €1.64 (IQR: €0.05–€3.16) when the measurement frequency was assumed to be daily. In the subsequently conducted sensitivity analysis, this median value increased to a maximum of €57.70 for different combinations of the willingness-to-pay threshold and the measurement frequency.

Conclusions: Probability elicitation can successfully be combined with early health economic modeling to obtain the probability distribution of the headroom available to a new medical technology. Subsequently feeding this distribution into a product investment evaluation method enables stakeholders to make more informed decisions regarding to which markets a currently available product prototype should being targeted.
6.1 INTRODUCTION

Investment in the research and development of new medical technology typically results in several promising product concepts. There is usually, however, insufficient funding to further develop each of these concepts into concrete products that can be brought to the market. This forces producers of medical technology and other stakeholders, such as venture capitalists and funding agencies, to already decide early during the product development process which of these concepts to abandon and which of them to push forward for further development.1-5

In the current practice of product investment decision making, such decisions seem often be based on potentially arbitrary representations of the expected improvements in outcomes and costs resulting from the use of the new technology. A factual representation of the current care situation and the specific changes that are likely to occur after the new technology has been fully adopted is generally not elaborated. Thus, decisions regarding the selection of suitable target markets for a currently available prototype technology are reached in a similarly arbitrary way. Early-stage health economic modeling has recently been suggested as a tool for supporting product investment decision making in a more formal way as it can provide insight into the maximum additional cost at which the intended clinical use of the new technology in a selected target market is still deemed cost-effective.3 This upper bound on the technology’s maximum cost, also known as the commercial headroom available4,6, can then be fed into an appropriate product investment evaluation method to determine whether further development of the prototype technology is likely to yield sufficient return on investment.7,8

In early-stage health economic evaluations, there is usually only a limited amount of data available with regard to the performance of the new technology, leading to high uncertainty in the values of some of the model inputs.9 Expert judgment therefore needs to be relied upon to obtain initial estimates of those parameters for which sufficient clinical evidence is not yet available. Probability elicitation (PE) refers to a set of techniques for formulating one or more experts’ beliefs about the unknown parameters into a probability distribution of those parameters.10,11 Previous work already described how PE can be applied to handle parameter uncertainty in health economic models.12-14 In this paper, we take the use of this approach one step further by illustrating how PE can be combined with early health economic modeling to obtain the probability distribution of the commercial headroom available to a novel point-of-care testing (POCT) device, which is defined as laboratory testing at or near the patient, in the disease management of patients with heart failure (HF).
6.2 METHODS

The approach that we used for combining PE with early health economic modeling is summarized in Figure 1. First, conceptual models of the current care setting (e.g., the health care setting in which the conventional disease management strategies are applied) and the new care setting (e.g., the health care setting in which the novel POCT device is introduced) were developed. Then, a continuous-time Markov model that appropriately reflects the disease progression in patients under the current care setting was developed. Next, the model parameters that are likely to change under the new care setting were identified. These served as the unknown Parameter(s) of Interest (uPoI) for which PE was subsequently conducted. Finally, the commercial headroom available was calculated and its uncertainty was captured probabilistically by propagating the uPoI distributions.

6.2.1 Conceptual models of the two health care settings

The current care comparator depends on the clinical setting in which the POCT device will be applied after it has been brought to the market. As this device could potentially be used in different clinical settings (e.g., outpatient clinic, home setting), we used semi-structured interviews to learn from the clinical experts in which setting the introduction of the POCT device was likely to generate the highest clinical impact. As none of the cardiologists could see any value in the use of this device in the outpatient setting, we decided to focus the early health economic assessment around the introduction of the POCT device in the home setting. More detailed assumptions on how the device would affect the care pathway when introduced in this setting were subsequently elicited from the same experts. Based on the results of these interviews, the current and new care settings were defined as follows:

![Diagram of the approach used in this paper](image)
The current care setting consisted of several follow-up visits to the cardiologist and of additional support provided by nurses with special education and training in HF management. During the first four months after discharge, patients on average have one visit to the cardiologist and four visits to one of the HF nurses. After this initial period, the number of outpatient visits reduces to a yearly follow-up visit to the cardiologist. After an HF-related readmission, the above visiting frequency was repeated.

The new care setting consisted of the introduction of the novel POCT device in the home setting to allow for more efficient monitoring of an HF patient’s disease progression. The home measuring itself was not expected to change the outpatient visiting frequency.

6.2.2 Markov model
To estimate the expected health outcomes and costs under the current care setting, we developed a continuous-time Markov model with three health states (Fig. 2): discharged alive from hospital, HF-related hospital readmission, and death. For practical purposes, the transition intensities were assumed to be constant over time and independent of patient-related risk factors. We used the data collected during the Coordinating Study Evaluating Outcomes of Advising and Counseling in Heart Failure (COACH) study, one of the largest multi-center, randomized controlled trials of nurse-led disease management programmes in HF\textsuperscript{15,16}, to

\begin{figure}[h]
\centering
\includegraphics[width=0.8\textwidth]{markov_model.png}
\caption{Structure of the continuous-time Markov model}
\end{figure}
estimate the current care model parameters. To be consistent with our description of the current care setting, we included all patients from the COACH study who received additional basic or intensive support from a nurse specialized in the disease management of HF patients. This resulted in a total sample size of 684. To make the model of the current care setting probabilistic, simple random sampling with replacement was conducted to obtain 10,000 resamples of equal size to the original sample. For each bootstrap resample, the current care transition intensities were subsequently estimated by using the msm package for R\(^{17}\), resulting in 10,000 realizations from the joint probability distribution of the current care transition intensities. Unit costs for outpatient visits and HF-related hospitalization were taken from Postmus et al.\(^{18}\) and set to be equal to €110/visit and €769/day, respectively.

### 6.2.3 Probability elicitation of the unknown model parameters

The transition intensities of the continuous-time Markov model were identified as the uPoI for the new care setting. The same three cardiologists who assisted in developing the conceptual models of the two health care settings were also invited to take part in the face-to-face PE interviews. We took the suggestions from Soares et al.\(^{13}\) and expressed the uPoI in terms of more directly observable quantities for which the experts’ beliefs were elicited. In particular, let \(T_i\) denote the amount of time spent in health state \(i\), let \(F_i(t_i)\) denote the proportion of patients that has left health state \(i\) by time \(t_i\), and let \(P_{ij}\) denote the probability that when leaving state \(i\) the next state will be state \(j\). For a continuous-time Markov model, it holds that \(T_i\) is exponentially distributed with mean value \(1/\sum_{i \neq j} q_{ij}\), where \(q_{ij}\) denotes the transition intensity between health states \(i\) and \(j\).\(^{17,19}\) This allows us to express \(F_i(t_i)\) as:

\[
F_i(t_i) = 1 - \exp(-\sum_{i \neq j} q_{ij} t_i)(i = 1,2; j = 1,2,3) 
\]

(1)

Similarly, the transition probabilities \(P_{ij}\) can be expressed in terms of the transition intensities as:

\[
P_{ij} = \frac{q_{ij}}{\sum_{i \neq j} q_{ij}}(i = 1,2; j = 1,2,3) 
\]

(2)

By combining (1) and (2), the transition intensities \(q_{ij}\) can be expressed as:

\[ q_{ij} = -p_{ij} \times \log(1 - F_i(t_j)) \quad (i = 1,2; j = 1,2,3; i \neq j) \]  

(3)

To obtain the probability distributions of \( q_{ij} \), we elicited point estimates of the transition probabilities \( p_{ij} \) and used the fixed interval elicitation method\(^{10,11}\) to capture the probability distributions of \( F_i(t_j)(i = 1,2) \) at the time points \( t_1 = 365 \) days and \( t_2 = 10 \) days. Equation (3) was then used to capture the \( q_{ij} \) distributions from the elicited \( F_i(t_j)(i = 1,2) \) distributions and \( p_{ij} \) values. A detailed description of the method of data capture and the distribution fitting is provided in Supplementary Material I.

6.2.4 Analysis of the commercial headroom available

Let \( e_0 \) and \( c_0 \) be the expected survival time and cost derived from the current care model, and let \( e_1 \) and \( c_r \) be the expected survival time and non-device-related cost derived from the new care model. The expected overall commercial headroom available to the new device can then be computed as:

\[ h = \lambda \Delta e - \Delta c \]  

(4)

where \( \lambda \) denotes the willingness-to-pay per life year, and \( \Delta c = c_r - c_0 \) and \( \Delta e = e_1 - e_0 \) denote the expected incremental non-device-related cost and the expected improvement in survival time due to the use of the new technology (also known as the effectiveness gap\(^{20}\)), respectively.

The POCT of circulating cardiac biomarkers involves placing a disposable containing a patient’s blood sample in an analyzer that is located at or near the patient. Denoting the cost of the analyzer by \( c_{ana} \) and the cost of a disposable by \( c_{dis} \), the total device-related cost can be expressed as

\[ c_{1d} = c_{ana} + c_{dis} \cdot t_m \cdot f_m \]  

(5)

where \( t_m \) and \( f_m \) are the expected out-of-hospital days and the out-of-hospital measurement frequency (expressed in number of measurements per day), respectively. Assuming that the fully developed product will only be applied in daily clinical practice if its use is considered to be cost-effective, the commercial headroom available reflects the maximum price at which
the POCT device can be sold on the market. Further development of the device into a tool for monitoring HF disease progression may therefore not be viable if the total device-related cost is likely to exceed this value. Equating (4) and (5) and dividing by the total number of measurements, this upper-bound on the device-related cost can formally be expressed as

\[
C_{\text{dis}} + \frac{c_{\text{ana}}}{t_m \cdot f_m} \leq \frac{\lambda \Delta e - \Delta c}{t_m \cdot f_m}
\]  

(6)

where the left-hand side denotes the average cost per measurement and the right-hand side the commercial headroom available per measurement.

Per expert, the uncertainty in the amount of headroom available per measurement was captured in a probabilistic way by propagating the uncertainty in the uPoI. This was achieved by repeatedly (10,000 times) sampling the current care transition intensities and the uPoI values from their corresponding probability distributions to obtain different values of \(\Delta e\), \(\Delta c\), and \(t_m\) and then computing \(\frac{\lambda \Delta e - \Delta c}{t_m \cdot f_m}\) for given values of \(\lambda\) and \(f_m\). For a given realization of the current care transition intensities and the uPoI, \(\Delta c\) was estimated by taking the difference in the expected values of \(c_{r1}\) and \(c_0\) and \(\Delta e\) was estimated by taking the difference in the expected values of \(e_1\) and \(e_0\). The expected values of \(c_0, e_0, c_{r1}, e_1,\) and \(t_m\) were obtained by taking the mean values of 10,000 first-order Monte Carlo simulation runs from the two continuous-time Markov models, using a 5-year time horizon. This relatively short time period was selected to ensure that the assumption of constant transition intensities for the continuous-time Markov model would still be reasonable. As up to 70% of all HF patients die within 5 years of their first hospital admission, however, this period should still be long enough to capture most of the health effects resulting from introducing the novel POCT device in this patient population. To combine multiple experts’ opinions into a single distribution of the commercial headroom available per measurement, we performed linear opinion pooling with equal experts’ weighting.\(^{10,11}\)
6.3 RESULTS

6.3.1 Current care model

After 18 months of follow up in the COACH study, 270 (39%) patients reached the combined endpoint of HF hospitalization and death. Out of these patients, 176 (65%) had been readmitted to the hospital and 94 (35%) had died. In total, 242 HF hospitalizations were recorded in the study cohort. The mean duration of hospital stay was 13 days and the incidence of in-hospital mortality was 14%.

The estimated values of the transition intensities under the current care setting are provided in Table 1. Using equation (2), the corresponding incidences of out-of-hospital and in-hospital mortality were 36% and 13%, respectively. The estimated mean hospital stay in the Markov model was 13 days. The event rates and mean hospital duration as predicted from the Markov model are close to the event rates and mean hospital duration as observed in the COACH study, showing that the current care model fitted the observed data well.

6.3.2 Probability elicitation

One cardiologist was unable to complete the PE interview because he found it was impossible to provide any reliable quantitative judgments with regard to this new device application. The probability distributions of in state ( =365 days) and state ( =10 days) and the values of that were elicited from the two remaining cardiologists are provided in Supplementary Material II. The resulting probability distributions of together with the corresponding probability distributions of these parameters under the current care setting are depicted in Figure 3. Both experts believed that the introduction of the POCT device would not have a profound impact on the number of transitions from the discharged alive to the death state, but they had different opinions on how the use of this device would affect the number of hospital readmissions. The first expert believed that individually tailored drug doses in response to the observed biomarker trajectories would result in fewer hospital readmissions, while the second expert believed that the repeated biomarker measurements would increase admission rates among patients with less severe clinical signs and symptoms. This explains why, compared to the current care setting, the mode of the distribution of is shifted to the left for expert 1 and to the right for expert 2. The first expert also believed that the introduction of the POCT device would not affect the transition intensities out of the hospital readmission state. For this
Table 1: Mean values (95% confidence intervals) of the transition intensities under the current care setting

<table>
<thead>
<tr>
<th></th>
<th>$q_1$</th>
<th>$q_2$</th>
<th>$q_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q_1$</td>
<td></td>
<td>0.00078 (0.00069, 0.00089)</td>
<td>0.00043 (0.00036, 0.00051)</td>
</tr>
<tr>
<td>$q_2$</td>
<td>0.06810 (0.05940, 0.07789)</td>
<td></td>
<td>0.01070 (0.00769, 0.01517)</td>
</tr>
<tr>
<td>$q_3$</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3: Comparison of the probability distributions of the current care transition intensities as obtained using bootstrapping and the probability distributions of the new care transition intensities as elicited from the two experts.

Figure 6.3.3 Headroom analysis

Figure 4 summarizes the distributions of the commercial headroom available per measurement based on the separate and pooled experts' opinions for the base-case scenario of a willingness-to-pay threshold of €10,000 per life year and a daily measurement frequency as suggested by one of the cardiologists. The corresponding pooled lower quartile, median, and upper quartile were €0.05, €1.64, and €3.16 per measurement, respectively. How this median value is affected by different assumptions for the willingness-to-pay threshold and the measurement frequency is depicted in Figure 5. In comparison to the base-case value of €1.64, the median commercial headroom available per measurement increased to €11.48 when the measurement frequency was decreased to weekly and increased to €8.24 when the willingness-to-pay threshold was increased to €80,000 per life year. When both of these parameters were simultaneously set to their most favorable values, the commercial headroom available per measurement further increased to €57.70.
expert, we therefore assumed that the probability distributions of the transition intensities out of the hospital readmission state under the new care setting were equal to the probability distributions of these intensities under the current care setting. The second expert, in contrast, believed that because of the increased admission rates amongst patients with less severe clinical signs and symptoms the average length-of-stay in the hospital would decrease and the proportion of patients discharged alive would increase. This explains why, compared to the current care setting, the mode of the distribution of $q_{21}$ is shifted to the right while the mode of the distribution of $q_{23}$ is shifted to the left.

### 6.3.3 Headroom analysis

Figure 4 summarizes the distributions of the commercial headroom available per measurement based on the separate and pooled experts’ opinions for the base-case scenario of a willingness-to-pay threshold of €10,000 per life year and a daily measurement frequency as suggested by one of the cardiologists. The corresponding pooled lower quartile, median, and upper quartile were €0.05, €1.64, and €3.16 per measurement, respectively. How this median value is affected by different assumptions for the willingness-to-pay threshold and the measurement frequency is depicted in Figure 5. In comparison to the base-case value of €1.64, the median commercial headroom available per measurement increased to €11.48 when the measurement frequency was decreased to weekly and increased to €8.24 when the willingness-to-pay threshold was increased to €80,000 per life year. When both of these parameters were simultaneously set to their most favorable values, the commercial headroom available per measurement further increased to €57.70.
Figure 4: Probability distribution of the commercial headroom available per measurement for $\lambda = €10,000$ per life year and $f_m = 1$

Figure 5: Median commercial headroom available per measurement against $\lambda$ for different values of the measurement frequency
6.4 Discussion

Early-stage health economic evaluations are characterized by evidence scarcity as data from clinical research is usually still missing. This lack of clinical evidence leads to high uncertainty in some of the model inputs, which can generally only be resolved by incorporating expert opinion. In this paper, experts’ beliefs were first elicited through semi-structured interviews to develop appropriate conceptual models of the healthcare settings without and with the use of a new medical technology. PE was then applied to quantitatively capture experts’ beliefs in a probabilistic way to handle parameter uncertainty in the subsequently constructed mathematical models. Finally, Monte Carlo simulation was applied to propagate the uncertainty in the model inputs to uncertainty in the modeled outcomes and to obtain a probability distribution of the amount of headroom available per measurement.

Technology-driven innovation consists of several phases, ranging from idea generation and application selection to the commercialization and launching of the developed products. Given the specific characteristics of such projects, the use of early-stage health economic evaluation as a tool for informing product investment decision making seems especially useful at the intersection between “investigation and technology transfer” and “development and validation”, where it has to be decided to which markets the currently available prototype technology is being targeted, if any. For the case study considered in this paper, the prototype technology was a novel POCT device for measuring one or more circulating biomarkers, and the market considered was the disease management of HF patients. For this application of the prototype technology, the median (IQR) of the commercial headroom available per measurement was found to be equal to €1.64 (€0.05–€3.16) for a willingness-to-pay threshold of €10,000 per life year and a daily measurement frequency. Further development of the device into a tool for monitoring HF disease progression may therefore not be viable if the average cost per measurement is expected to be much larger than these values, which can be assessed more formally by feeding the obtained distribution of the commercial headroom available into the product investment evaluation method proposed by Girling et al.. This involves comparing the projected post-market cash flows resulting from selling the analyzer and disposable at a price that puts the average cost per measurement just below the commercial headroom available per measurement to the expected development cost to determine whether further development of the prototype technology into a commercial product for HF disease management is likely to yield sufficient return on investment.
An important prerequisite of the use of elicitation in early health economic modeling is that experts need to be able or willing to provide values for the unknown model parameters. The method can therefore not be applied when experts do not want to speculate about the clinical impact of a not yet fully developed product. This problem was also encountered in our case study and led to the drop-out of one of the cardiologists from participating in the elicitation exercise. Attempts to support product investment decision making through early health economic modeling may therefore not be successful when the impact of introducing the new technology in a selected clinical setting is difficult to perceive by the participating experts. Also, at the start of the product development process, there are typically many technological solution principles and/or target markets left to choose from, giving rise to a large variety of potential products. Other, more qualitative approaches are then required to perform an initial screening of the generated product concepts.

In previous applications of early health economic modeling, expert beliefs were mostly incorporated by eliciting point estimates of the uPoI as well as their minimum and maximum values to facilitate a subsequent sensitivity analysis.\textsuperscript{24-27} Although this so-called deterministic value elicitation approach is easier to use than the PE approach applied in this paper, it does not allow the parameter uncertainty to be explicitly represented and assessed regarding its impact. Propagating the uncertainty in the model inputs to uncertainty in the modeled outcomes is nevertheless essential if one wants to support product investment decision making through the use of formal decision support methods, such as the one described previously. A downside of using PE is that attaching likelihoods to the values of unknown parameters is not straightforward. Care must therefore be taken to parameterize the uPoI in terms of quantities that still have a clear interpretation. For the Markov model considered in this paper, the transitions out of the two transient health states are subjected to competing risks, meaning that the overall effect of the POCT device on the number of transitions from state \( i \) to state \( j \) does not only depend on the effect of this device on \( q_{ij} \) but also on its effect on the other transition intensity out of state \( i \).\textsuperscript{28} We therefore decided not to use hazard ratios to parameterize the transition intensities of the new care model in terms of the transition intensities of the current care model. Instead, we followed the approach taken by Soares et al.\textsuperscript{13} and parameterized the transition intensities of the new care model in terms of two directly observable quantities and elicited their absolute values and their uncertainty conditional on the mean values of these quantities under the current care setting. The mean values of these quantities under the current setting therefore served as reference values based
on which the experts’ opinions were elicited. As the elicitation exercise was not repeated for
different reference values, the underlying assumption of our approach is that an expert’s
opinion on the new care transition intensities does not depend on the specific values of these
intensities under the current care setting. Our approach may therefore not be valid when
experts believe there is a large correlation between the current care and the new care transition
intensities.

In specifying the Markov model, we assumed that the transition intensities were
homogeneous (e.g., independent of time) and constant across patients. One way to relax these
assumptions would be to include time and patient characteristics as regressors within the same
model structure.\textsuperscript{29} This would also result, however, in many more parameters to elicit and
therefore seems less suitable for early-stage health economic modeling. In addition, there
were only two experts who participated in the PE interviews. The resulting pooled distribution
of the commercial headroom available may therefore not represent consensuses among a
larger group of experts. A final important decision that has to be taken when combining PE
with early health economic modeling is when the pooling of the different experts’ opinions
has to be conducted. In our study, the probability distributions of the uPoI served as
intermediate outcomes that were transformed into a probability distribution of the commercial
headroom available per measurement, the main outcome of interest. We therefore decided to
not directly pool the distributions of the uPoI but to postpone the pooling until the uncertainty
in the model inputs was propagated to uncertainty in the modeled outcomes, resulting in both
individual and aggregated distributions of the commercial headroom available per
measurement.

To conclude, this study illustrated by means of a case study how PE can be combined
with early health economic modeling to obtain the probability distribution of the commercial
headroom available to a new medical technology. By subsequently feeding this distribution in
a formal product investment evaluation method, the decision to which markets a currently
available prototype technology should being targeted, if any, can be taken in a more informed
way, which should ultimately result in higher return on investment for all stakeholders.
6.5 ACKNOWLEDGEMENTS
This research was performed within the framework of CTMM, the Center for Translational Molecular Medicine (www.ctmm.nl), project TRIUMPH (grant 01C-103), and supported by the Dutch Heart Foundation. We thank the three cardiologists who participated into the elicitation interviews.
6.6 REFERENCES

25. Luitjes SHE, Wouters MGAJ, Franx A, et al. Study protocol: Cost effectiveness of two strategies to implement the NVOG guidelines on hypertension in pregnancy: An innovative strategy including a computerised decision support system compared to a common strategy of professional audit and feedback, a randomized controlled trial. Implement Sci 2010;5:68.
Supplementary Material I: details of the elicitation exercise

Method of data capture

To capture the quantities of interest, two figures (Figure I & Figure II) describing how patients flow through the model’s health states under the current care setting were provided to the cardiologists. In particular, the following information was contained within these figures:

1. Patients’ transition to the composite endpoint of hospital readmission or death within a 1 year period;
2. Patients’ transition to each individual outcome (hospital readmission and death) conditional on having left the discharged alive state within 1 year;
3. Patients’ transition to the composite endpoint of discharged alive or death within a 10 day period;
4. Patients’ transition to each individual outcome (discharged alive and death) conditional on having left the hospital readmission state within 10 days.

For example, Figure I shows that out of the 684 patients who started in the discharged alive from hospital state, 246 reached the combined endpoint of HF-related hospital readmission or death within 1 year of follow-up. Out of these patients, 89 had died and 157 had been hospitalized. In producing these numbers, we only considered a patient’s first transition out of the discharged alive from hospital state (i.e., the 89 deaths do not include patients who were admitted to hospital, then discharged alive, and subsequently died within 1 year of follow-up). Figure II contains exactly the same information but then expressed in terms of proportions rather than absolute numbers. Depending on whether a cardiologist felt more confident to think in terms of absolute numbers or proportions, several questions were posed to adjust these numbers to the new care setting. These questions are listed in the subsections below.

Questions posed in terms of absolute numbers

- QI (a): In the current care setting, out of the 246 patients who left the discharged alive from hospital state within a 1 year period, 89 died and 157 were readmitted to the hospital. Can you adjust these numbers to reflect the amount of events that you expect to observe after 1 year of follow-up under the new care setting?
- QI (b): In the current care setting, out of the 86 patients who left the hospital readmission state within a 10 day period, 11 died and 75 were discharged alive. Can you adjust these numbers to reflect the amount of events that you expect to observe after 10 days of follow-up under the new care setting?
• QI (c): Suppose that for the new care setting you will conduct 100 different experiments (n=684) to establish the amount of patients having reached the combined endpoint of HF-related hospital readmission or death after 1 year of follow-up. How many times would you expect the observed value to be larger (smaller) than 246?

• QI (d): Suppose that for the new care setting you will conduct 100 different experiments (n=157) to establish the amount of patients having reached the combined endpoint of discharged alive or death after 10 days of follow-up. How many times would you expect the observed value to be larger (smaller) than 86?

Questions posed in terms of proportions

• QII (a): In the current care setting, out of all the patients who left the discharged alive from hospital state within a 1 year period, 36% died and 64% were readmitted to the hospital. Can you adjust these numbers to reflect the proportion of events that you expect to observe after 1 year of follow-up under the new care setting?

• QII (b): In the current care setting, out of all the patients who left the hospital readmission state within a 10 day period, 13% died and 87% were discharged alive. Can you adjust these numbers to reflect the proportion of events that you expect to observe after 10 days of follow-up under the new care setting?

• QII (c): In the current care setting, out of all the patients who started in the discharged alive from hospital state, 36% reached the combined endpoint of HF-related hospital readmission or death after 1 year of follow-up. In your opinion, what will be the most likely value of this proportion after 1 year of follow-up under the new care setting?

• QII (d): In the current care setting, out of all the patients who entered in the hospital readmission state, 55% reached the combined endpoint of discharged alive from hospital or death after 10 days of follow-up. In your opinion, what will be the most likely value of this proportion after 10 days of follow-up under the new care setting?

• QII (e): Suppose that for the new care setting you will conduct 100 different experiments to establish the proportion of patients having reached the combined endpoint of HF-related hospital readmission or death after 1 year of follow-up. How many times would you expect the observed value to be larger (smaller) than 36%?

• QII (f): Suppose that for the new care setting you will conduct 100 different experiments to establish the proportion of patients having reached the combined endpoint of discharged alive or death after 10 days of follow-up.
endpoint of discharged alive or death after 10 days of follow-up. How many times would you expect the observed value to be larger (smaller) than 55%?

Figure I: Current care model with the flow of patients expressed in absolute numbers

Figure II: Current care model with the flow of patients expressed in proportions
**Distribution fitting**

The answers provided to questions QI (a) and QI (b) or QII (a) and QII (b) directly resulted in point estimates of the transition probabilities \( p_{ij} \). To capture the uncertainty in the values of \( F_i(t_j) \), it was assumed that these parameters had a beta distribution. This allowed us to use the BetaBuster program (downloaded from http://www.epi.ucdavis.edu/diagnosticstests/betabuster.html) to translate a cardiologist’s answers to the questions above into corresponding values of the Beta parameters. In particular, the answers to questions QI (a) and QI (b) or QII (c) and QII (d) were used to obtain the distributions’ modes, whereas the answers to the questions QI (c) and QI (d) or QII (e) and QII (f) were used to obtain one of the distributions’ percentiles.
Supplementary Material II: results of the elicitation exercise

<table>
<thead>
<tr>
<th></th>
<th>$p_{12}$</th>
<th>$p_{13}$</th>
<th>$p_{21}$</th>
<th>$p_{23}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expert 1</td>
<td>0.625</td>
<td>0.375</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Expert 2</td>
<td>0.68</td>
<td>0.32</td>
<td>0.93</td>
<td>0.07</td>
</tr>
</tbody>
</table>