The use of metamodels for updating cost-effectiveness estimates over time: a review of methods and a case study

Thesis

MSc Health Economics, Policy and Law

Jonas Esser

June 23rd 2021, Rotterdam, Netherlands

Student number: 575740

Thesis supervisor: Naomi van der Linden

First reader: Frederik T. Schut

Word count (excluding appendix): 14010

1. Contents

2. Summary

This thesis was written in combination with an internship at AstraZeneca Netherlands, Market Access department. The primary goal was to develop a method to forecast the cost-effectiveness of the drug dapagliflozin as the conditions determining its cost-effectiveness change. Dapagliflozin is a SGLT2 inhibitor used in treatment of diabetes type 2 treatment for patients whose blood sugar levels are not sufficiently controlled on metformin alone. The main competitor of dapagliflozin, and the comparator used in this study, is the drug group of sulfonylurea (SU). Dapagliflozin is associated with weight loss and lowered rates of hypoglycemia and heart failures compared to SU. The effect on patient's blood sugar is similar across both treatment options. SU is a lot cheaper and currently more widely used than dapagliflozin in the Netherlands.

AstraZeneca uses a stochastic simulation model, called the Cardiff diabetes model, to estimate the cost-effectiveness of dapagliflozin in various European countries. Since the Cardiff model does not offer sufficient functionality for the aims of the study, a second model was developed instead, which approximates the incremental costs and effects of dapagliflozin treatment as a function of selected input parameters. In order to precisely estimate the uncertainty involved in the estimates, all statistical analyses in this study are performed from a Bayesian perspective. Incremental costs and effects are modeled as a bivariate normal distribution, with the two means being linear functions of the selected input parameters.

The so-called metamodel requires training data which is generated from the original model. This is achieved by performing multiple probabilistic sensitivity analyses (PSAs) within the Cardiff model while systematically varying the selected inputs. The results are stored together with the inputs. Four inputs parameters are included in the analysis: prices of dapagliflozin, sulfonylurea and metformin, as well as the yearly weight loss that dapagliflozin is associated with.

When validated against test data which was not included in the training data, the metamodel produced approximately equal estimates to those of the Cardiff model. The weight loss parameter had a moderate negative impact on the incremental effects and a small positive impact on the incremental costs. Out of the investigated prices, the price of dapagliflozin had a strong positive effect on the incremental costs. The SU price had a small but credible negative effect, while the price of metformin had no credible impact. A forecast based on an assumed price drop after patent loss indicates that dapagliflozin treatment will become cost-saving once the drug's costs fall below € 360 per year per patient.

The study's methodology makes it possible to model the development of a treatments cost-effectiveness in the future, provided that there is information of how the prices involved will develop. Such forecasts are of course very much dependent on and sensitive to assumptions, but could nevertheless be a valuable add-on to health economic evaluations. A medicine may not be cost-effective right now, but might become cost-effective in the near future, and would thus be the most optimal long-term choice for reimbursement. The major unresolved challenge in applying the metamodeling method lies in the need to generate a large amount of training data from the original model. For most health economic models which are currently in use, this will require a lot of time-intensive and error-prone work.

3. Introduction

The conditions surrounding pharmaceutical cost-effectiveness are in constant flux: prices change, new clinical evidence regarding existing medicines becomes available and completely new medicines are launched onto the market.

In order to facilitate the efficient allocation of the health care system's limited resources, it is of utmost importance that cost-effectiveness evaluations are based on correct assumptions and input parameters. This can be challenging, especially when multiple treatments in a therapeutic area are subject to changes in costs or other attributes. Performing a scenario analysis based on the expected new parameters can provide information on the expected future cost-effectiveness. But each analysis performed is limited to a single scenario. It remains difficult to understand how the changing parameters affect the cost-effectiveness in terms of effect size, uncertainty, development over time and potential correlations. Particularly when cost-effectiveness evaluations are performed with a long time horizon, it may be relevant to anticipate changes in the market conditions (e.g. price decreases) and incorporate these into the evaluation.

Forecasting the effects of changing inputs is also important in the context of so-called value-based payments or innovative contracts [1]. These arrangements are attractive when healthcare payers and providers cannot find an agreement through traditional fixed-pay per unit contracts. Providers are instead paid "per value", for example per patient successfully treated, as measured by clinical response indicators. In order to have an appropriate basis for negotiations between providers and payers, value-based arrangements require that the risk (clinical or financial) associated with a medical treatment is precisely estimated and updated as soon as new relevant information becomes available. This necessitates the incorporation of (uncertain) changes into cost-effectiveness analyses, which is simply not possible within standard health economic models.

Neumann et al [2] expressed the need for the development of metamodeling techniques in order to approach problems like the ones discussed above. Metamodeling can be used to estimate the impact that certain key factors have on a medicine's costs and benefits, and create useful forecasts as these factors change. Within this study I seek to answer this call and provide ideas of how to apply metamodels for health economic evaluations.

The research goal is two-fold:

- Develop an effective and widely applicable metamodeling method, which can be used to better understand the determinants of a drug's cost-effectiveness. The method should be reasonably simple to execute and produce demonstrably accurate estimates.
- Apply the proposed method in a suitable case study: the diabetes drug dapagliflozin and its cost-effectiveness in the Netherlands.

Background and methods are closely intertwined in this study, and separating them would have lead to many repetitions and cross-references. Thus they are instead dealt with in a joint section, for the sake of readability and conciseness. The section begins by motivating why all statistical modelling in the study is performed under the Bayesian paradigm. It goes on to explain what metamodeling is and which viable metamodeling methods could be identified in the literature. Section 4.3 introduces the case of diabetes type 2 in the Netherlands, as well the Cardiff model and the clinical evidence it is based on. The two are brought together by constructing a metamodel based on the literature research and case study. Section 5 presents the results of the conducted analysis. This includes both the estimated parameters of the metamodel and simulations based on these parameters. I will investigate the impact of the prices involved in the cost-effectiveness of dapagliflozin, as well as the

impact of one selected clinical parameter. Finally, I put the key findings into context and discuss strengths and weaknesses of the study design, as well as possible applications of my methods in value-based pricing of pharmaceuticals.

The appendix provides the inputs used for the Cardiff model, a summary of the training data, references as well as all R and Stan code used within this project.

4. Background and methods

4.1. A short introduction to Bayesian statistics

It is anticipated that most readers will have had limited exposure to Bayesian statistics. In order to make sense of the methodology of this study, a minimal understanding of theoretical foundations is needed. Thus this section quickly summarizes the most important concepts as well as differences to the more common frequentist paradigm. Some mathematical details are necessary; they will become clearer when they are applied to the case study in later sections. For a more sophisticated treatment of mathematical statistics and the differences between frequentist and Bayesian inference, see Betancourt [3].

Fundamentally, Bayesian statistics is about the quantification of belief. It assumes that a given set of observed data is caused by some data-generating process, which cannot be observed but is approximated through an adequate statistical model. The components of a model are the *likelihood* (probability of observing the observed data, given a set of parameters and their mathematical relationship) and the prior probability of these parameters, which is based on knowledge available before seeing the data. This can be formally expressed through Bayes' theorem:

$$p(\theta|y) = \frac{p(y|\theta) * p(\theta)}{p(y)} = \frac{p(y|\theta) * p(\theta)}{\int p(y|\theta) * p(\theta) d\theta}$$

Where

- θ is a parameter of interest (or, more commonly, a vector of multiple parameters)
- y is the observed data
- $p(\theta)$ is the prior distribution of the parameter(s)
- $p(y|\theta)$ is the likelihood
- $p(y)$ is the average likelihood
- $p(\theta|y)$ is the posterior distribution

The quantity $p(y)$ serves as a normalization constant which ensures that the posterior distribution sums to 1 and comprises a valid probability distribution. For all but the most simple models, it will be difficult or impossible to compute. It is thus more common to work with the unnormalized posterior distribution (more on this at the end of the section). Bayes' theorem can then be rewritten as:

$$p(\theta|y) \propto p(y|\theta) * p(\theta)^{1}$$

Any outcome of a Bayesian model will always, by definition, be a probabilistic *distribution* of parameter values, rather than just a single point estimate and confidence interval for each parameter. It is possible and common to summarize these distributions into a single value such as the mean, median or mode of the distribution, or to calculate the probability that an outcome falls within a certain interval[2] or is larger or smaller than a specific value, such as zero. However, these values are merely *summaries* of the estimated parameter distribution, not the estimate itself.

The aforementioned concepts can be illustrated through a simple example of a model with only one parameter. Let us imagine a newly discovered rare disease about which not much is known yet. An analyst wants to estimate the sex-proportion p of people affected by the disease, where $p = 0$ and $p = 1$ imply that only women or only men are affected, respectively. The likelihood of the model is

[1] The symbol $\propto$ should be understood as "proportional to".
[2] Commonly called a *credibility interval*, in order to avoid confusion with frequentist confidence intervals.

$$y \sim binomial(1, p)$$

where y is the sex of a patient. There is no *a priori* reason to believe that either men or women are more likely to be affected and most diseases are roughly equally distributed across the sexes. Thus the analyst chooses a prior distribution which expresses this belief:

$$p \sim beta(8, 8)$$

This distribution puts most of the probability density around 0.5 and little on the extreme ends of the spectrum (see Figure 1, blue curve). Out of the 20 cases documented thus far, 15 were male. The model estimates the posterior distribution for the parameter p:

Table 1: summary of poesterior distribution for the example case

	mean	**CI lower bound – 5%**	**CI upper bound – 5%**
$\boldsymbol{p}$	0.61	0.48	0.74

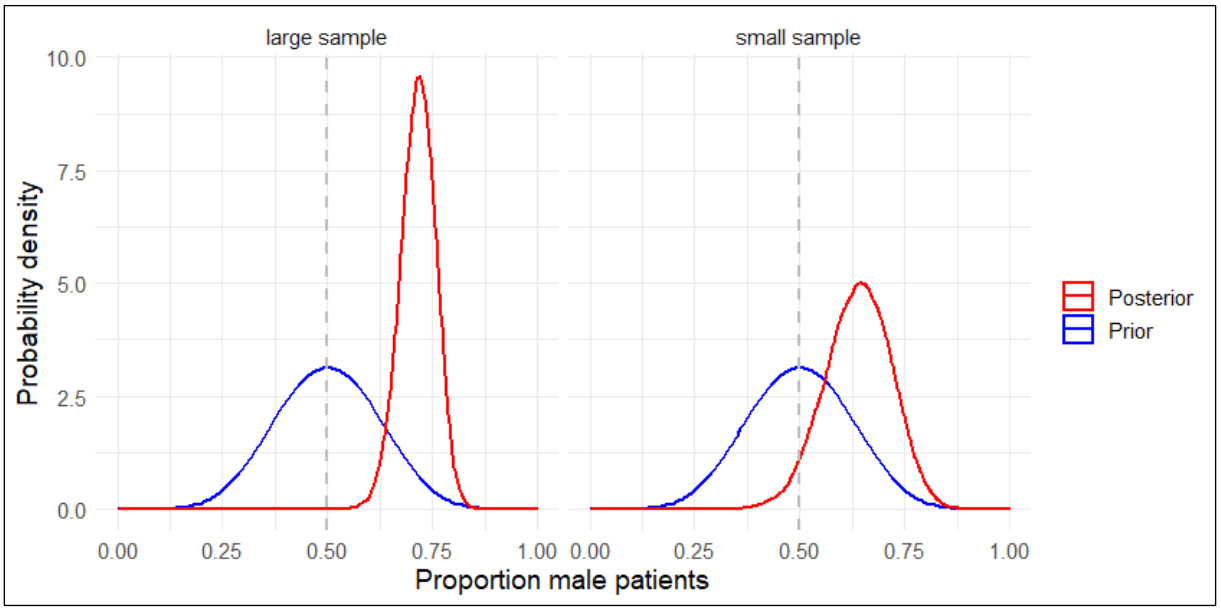


Figure 1: illustration of Bayesian inference

The expected proportion of men affected by the disease, based on the prior knowledge and the available data, is 61%. The credibility interval (CI) in Table 1 is a *90% highest posterior density interval* (HPDI). This is the most narrow interval possible that contains 90% of the distribution's probability mass. And 90% of samples drawn from this distribution will fall within the interval. Thus there is a 5% chance a realized value from this distribution is smaller than the lower bound of the interval, and a 5% chance that it is larger than the upper bound.

We see that there is still a probability of more than 5% that the proportion is 0.5 or lower, and the observed sample just happened to contain a lot more men by chance (Figure 1, small sample). Thus we can say that the proportion is not credibly different from 0.5, based on a 90% credibility level. Now suppose that instead 100 cases of the disease have been observed, 75 of whom were male. Even though the proportion of male patients is the same as in the previous example, the estimated parameter distribution of the parameter p is different: it has a larger mean and a smaller spread (Figure 1, large sample). In this case, the proportion is in fact credibly higher than 0.5: there is a

strong indication that the disease does affect more men than women. This kind of result is to be expected, since more data generally constitutes stronger evidence, and as such the prior belief has a smaller impact on the posterior belief.

The obtained posterior distribution can be implemented into a probabilistic model and used to generate outcome samples (or, technically incorrect but perhaps more intuitive: new "observations"). This latter aspect is extremely important, since it means that a Bayesian model is *generative*: it allows us to generate predictions which are similar to the original observations (data) the model is based on.

Let y^* represent a future occurrence of the outcome variable y. The posterior predictive distribution is then defined as

$$p(y^*|y) = \int p(y^*|\theta) * p(\theta|y) d\theta$$

In the context of the discussed example, this might come into play when building a patient simulation model involving the example disease: the posterior distribution can be used to randomly generate the sex of simulated patients. Most models contain more than one parameter, but the procedure is analogous to the univariate case.

For most models in applied Bayesian statistics, the integrals in the above formulas above will be difficult or impossible to solve, and thus no closed mathematical form of the posterior distribution can be found. It is however possible to generate exact samples from the posterior distribution, through the use of Markov chain Monte Carlo sampling [4]. When combined, these samples will converge to an increasingly accurate representation of the posterior distribution as more samples are generated. This is how the posterior distributions in the example (Figure 1) were obtained (even though in this simple case, an analytical solution would have been possible).

For the modeling in this study, I use the probabilistic programming software Stan [5]. Stan estimates the posterior distribution through the No-U-Turn Sampler (NUTS, a variant of Hamiltonian Monte Carlo, which in turn is a variant of Markov chain Monte Carlo) as described by Hoffman and Gelman [6].

4.1.1. Difference from frequentist statistics

In frequentist statistics, it is not possible to calculate the probability of either parameters or outcome taking any specific value or falling within some range of values; any interpretations of this sort are always wrong [7]. A 95% confidence interval does *not* contain 95% of the parameter values, since there is only one "true" parameter value which is fixed and therefore not subject to stochastic variation. The confidence interval also does *not* have a 95% chance to contain the true parameter. Rather, if the experiment were repeated an infinitive amount of times under the same conditions, 95% of the estimated confidence intervals would contain the (still unknown) true parameter. See Greenland et al [8] for a detailed treatment of these issues.

The correct interpretation of a confidence interval is generally not very useful for practical analyses, and as such it is extremely common to misinterpret frequentist estimates as if they were Bayesian. In order to avoid such inconsistencies, this study was conducted under a strict Bayesian paradigm. Consequently, all intervals presented are credibility intervals, *not* confidence intervals.

4.1.2. Bayesian statistics and health economics

The use of Bayesian statistics in health economics was strongly advocated by Baio [9], who emphasized the uncertainty quantification inherent in the Bayesian paradigm. This has obvious

advantages in health economic evaluation, where a "wrong" decision can lead to huge opportunity costs in terms of money and human well-being. In their standard work *Decision Modelling for Health Economic Evaluation*, Briggs et al argue that decision modelling is always essentially Bayesian [10]:

"In decision analysis, a probability is taken as a number indicating the likelihood of an event taking place in the future. As such, decision analysis shares the same perspective as Bayesian statistics. This concept of probability can be generalized to represent a strength of belief which, for a given individual, is based on their previous knowledge and experience. This more 'subjective' conceptualization of probabilities is consistent with the philosophy of decision analysis, which recognizes that decisions cannot be avoided just because data are unavailable to inform them, and 'expert judgement' will frequently be necessary."

Nevertheless, most health economic analyses have been and are performed using frequentist methods (almost always suffering from the aforementioned misinterpretations). With the recent development of fast and simple tools like Stan and Turing [11] and an abundance of information about their usage [5, 7], there is nothing stopping researchers from familiarizing themselves with Bayesian methods.

4.2. Metamodelling

As indicated in the introduction, conventional health economic models have certain severe limitations. They tend to be "black boxes": it is not clear how a specific results comes to be and which impact the inputs have on the result. Simply simulating different scenarios provides a rough idea at best about the effect of inputs and does not uncover any precise mathematical relationships. Furthermore, simulating a large number of scenarios is often unfeasible, since complex simulation models can take many hours to sample a single scenario [12]. These are not good preconditions for a sophisticated analysis.

A possible solution to these issues is to create another model which is trained on data generated by the original health economic model. Here the original model is the *data-generating process* as described in the previous section, and its mathematical properties can be approximated through Bayesian inference. This second model is called a metamodel or surrogate model. **Error! Reference s ource not found.** provides an illustration of the idea. A Bayesian model is generally preferable to a frequentist one because it retains the uncertainty inherent in the original simulation, instead of producing only a single point estimate.

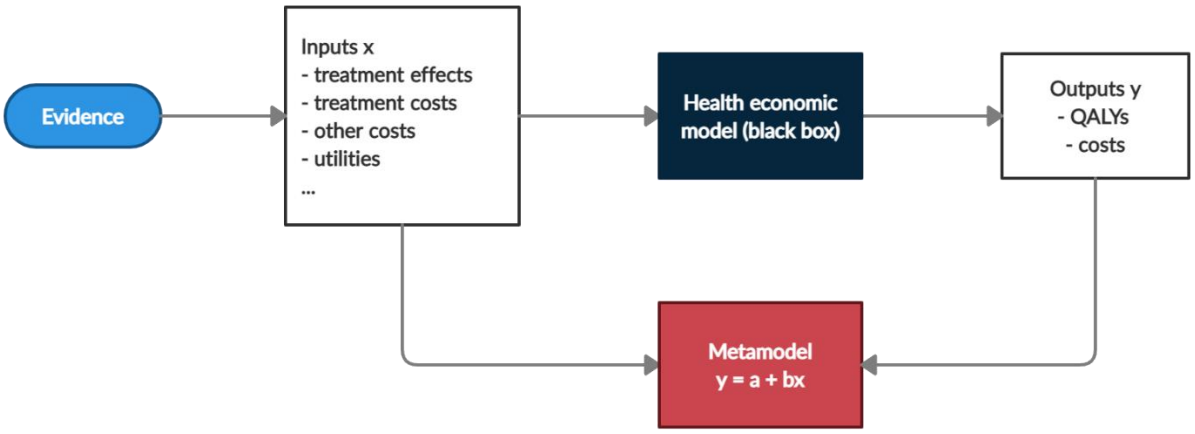


Figure 2: illustration of the metamodel concept3

In other words: assuming that a chosen health economic model provides a reasonable approximation of costs and effects for a defined scenario, a metamodel can be utilized to extrapolate the outcomes of the model as a function of specific inputs. Jalal et al [13] define a metamodel as "a second model that simplifies the relationship between the inputs and outputs of a simulation model". The purpose of metamodels is twofold: they can be used to

- Reduce computing time and replace the original model for most purposes
- Increase transparency by uncovering the relationship of specific inputs and outputs in the original model

The latter aspect is more important in the context of this study, but the former will be investigated as well and briefly commented on. While metamodeling is common in fields such as biomedicine and engineering, it is not very widespread yet in health economics. A comprehensive literature review from 2019 found only 13 studies utilizing them [12]. The available research has been concerned with expected value of information (EVPI) analyses [15, 16], sensitivity analysis [17], parameter calibration [18], and simplification of a complex model [13]. Using a metamodel to forecast cost-effectiveness in the future has, to my knowledge, not been done yet.

Degeling et al have written a valuable article on the application of metamodeling in health technology assessment [14], which will serve as a guide for my research. They recommend a five step process to build a metamodel:

- Identifying candidate metamodel techniques
- Simulating data sets
- Fitting metamodels
- Assessing metamodel performance (validation)
- Applying the metamodel

My methodology is based on this framework.

4.2.1. Modelling techniques

When picking a modelling method, there is generally a tradeoff between interpretability and predictive accuracy: simple parametric models tend to be more interpretable than complex machine learning algorithms but produce less accurate predictions, and vice versa [14]. I therefore had to

[3] inspired by Jalal et al [13] and Degeling et al [14]

determine which of the three techniques provided the best compromise in that regard. In the early stages of the study, a literature review was conducted in order to identify suitable metamodeling techniques. Three stood out:

- Linear Model [13]
- Gaussian Process [15, 16]
- Artificial Neural Network (ANN) [17, 18]

I discarded the ANN-approach due to the high computational burden and the "black box" nature of neural networks, which makes them unfeasible for interpretation [19]. Non-parametric regression using a Gaussian process was attempted and worked appropriately on synthetic test-data. However, with actual data the computation time quickly became unmanageable. One study using a Gaussian process for health economic metamodeling reported a computation time of 260 hours [20]. Therefore this was discarded as well. I still consider the Gaussian process a promising candidate for health economic metamodeling, but it requires the use of approximation methods instead of full inference. Further research on this would be valuable.

Ultimately I decided on modelling incremental cost and effects through a multivariate normal distribution. My methodology thus expands on the linear model approach used by Jalal et al [13]. Through testing it could be determined that this model produces accurate approximations of the original model's estimates, despite the simplifying assumption that the mean outcomes are a linear function of the inputs. The concrete model definition is presented in section 4.4.2.

There are two main reasons to assume a multivariate normal distribution:

- *Interpretability:* when given the model parameters, one can easily get an idea of what the resulting distribution looks like and which impact the individual parameters have. Each β or slope-parameter is a linear effect of the input it is assigned to, while the α or intercept parameters are the mean outcome when all inputs are zero. The correlation coefficient ρ describes the correlation between the two outcomes, incremental costs and effects.
- *Practical considerations:* the normal distribution is easily extended to multiple dimensions and included in all common statistical programming libraries. Multivariate versions of other distributions are often not readily available and have to be programmed manually, which requires advanced knowledge of mathematical statistics. Since the method presented in this study is supposed to be accessible to a large number of people, I decided to stick with a normal distribution. And indeed, despite their simplicity, normal distributions work very well for many applied problems. However, if another distribution seems more appropriate for a problem at hand, the methodology of this study can easily be adapted to use that instead.

A disadvantage of the method is that it assumes the outcome distribution of the original simulation model to take a multivariate Gaussian shape. The more the actual distribution deviates from this, the higher the risk for inaccurate results becomes. For further discussion of this point see section 6.2.4.

4.3. Case study

Diabetes type 2 (TDM) is the most prevalent chronic disease in the Netherlands, with an estimated patient count of more than one million in 2019 [21] and a yearly economic burden exceeding one and a half billion euros [22]. The patient count is expected to grow even larger in the coming years [23], putting more pressure on Dutch healthcare payers.

Drug manufacturers in turn have to deal with a lot of competing treatments and the need to demonstrate the long-term value of their own product to payers. These factors make diabetes drugs

a prime candidate for innovative contracts [1]. Within this thesis, I focus on the drug dapagliflozin [24] as a case study. Reimbursement of dapagliflozin in the Netherlands is currently restricted to patients who do not use insulin and cannot be treated with a combination of metformin and sulfonylurea. Patients receive dapagliflozin in combination therapy with either or both of those [25]. Both dapagliflozin and SU are considered to be equally effective at controlling the levels of glycated hemoglobin (commonly abbreviated to hemoglobin A1c or HbA1c). Dapagliflozin is associated with a higher quality of life and slightly improved survival for patients, due to a decreased risk of hypoglycemic episodes and heart attacks as well a reduction in patient's bodyweight. SU treatment is associated with higher risks of adverse events and weight gain [26], but is currently a lot cheaper.

The price of dapagliflozin is subject to change in the near future, due to a recently decided change to the Dutch reference pricing system [27] as well as the loss of patent protection of dapagliflozin in 2028 [28, 29], which will allow generic medicines to enter the market. It can also not be ruled out that newly released clinical evidence will expand the knowledge about dapagliflozin's (relative) efficacy. All of these factors will affect the cost-effectiveness of dapagliflozin against alternative treatment options. If we were able to forecast the potential impact of these changes on dapagliflozin's cost-effectiveness, healthcare payers would be able to anticipate these changes in their decisions.

The study is focused on the following comparison:

Table 2: overview of treatment arms in the case study

	Treatment arm:	**Control arm:**
First line	Metformin + dapagliflozin	Metformin + SU
Second line	Metformin + insulin	Metformin + insulin
Third line	Metformin + insulin + bolus	Metformin + insulin + bolus

The cost-effectiveness of dapagliflozin in this specific comparison has been investigated for the UK [30] and the Nordic countries [31]. However, these prior studies were mostly based on results of the UKPDS68 study [32], while the current version of the Cardiff model also includes evidence from the more recent DECLARE study [33], supplemented with data from UKPDS68 where necessary. A health economic study based on this setup has been published in 2020, albeit with a slightly different scenario focused on high-risk patients [34]. None of these studies were performed from the Dutch perspective. All of them made use of the Cardiff type 2 diabetes model, which I will explain shortly in the next section.

Unlike the aforementioned studies, I will not investigate any specific clinical outcomes or adverse events in the main analysis and only model incremental costs and effects. While the metamodel approach can be extended to include any outcome which is estimated in the base model, the complexity and effort involved increases with any input and outcomes included, especially regarding the generation of training data. It is therefore recommended to decide which parameters are the most relevant and focus on those.

4.3.1. The Cardiff model

The Cardiff type 2 diabetes model provides the training data which the metamodel is trained on. It is a microsimulation model which simulates individual patients and their clinical development over the course of fixed half-yearly increments. Patients switch from first to second and second to third line treatment when their HbA1c levels exceed 8% and 9%, respectively. The model is accessed through an interface in Microsoft Excel and Visual Basic for Applications (VBA), while most of the calculations happen in an external engine coded in C++. For a detailed explanation and validation of the Cardiff

model see McEwan et al [35]. The risk equations put into the model are based on results of the DECLARE study [33, 34].

Like many health economic models, the Cardiff model has the option to perform a *probabilistic sensitivity analysis* (PSA) where all model inputs are varied according to predefined probability distributions. For example, one can specify to simulate a cohort of 1000 patients, 1000 times. For each of these simulation runs, the inputs of the simulation (patients' clinical history, treatment effects, costs and utilities) are slightly different, since they are drawn from a probability distribution which is parameterized by a mean and standard error for each input. The rationale behind the PSA is that the model inputs are *uncertain*: they are based on limited available knowledge and evidence. This uncertainty should be reflected in the results as well, in order to avoid overconfidence in the results. See Briggs et al [10] for a detailed exposition.

Figure 3 shows a screenshot from the Cardiff model. Based on this simulation, the treatment is on average less costly and more effective than the comparator, but there is a lot of uncertainty surrounding both of these.

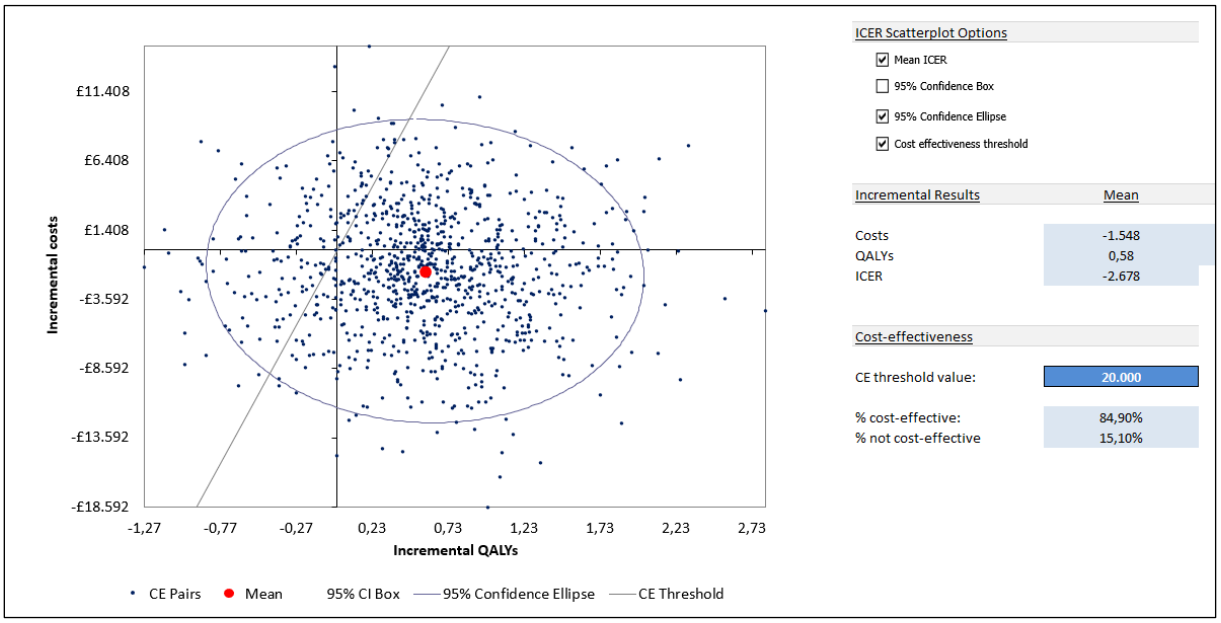


Figure 3: screenshot from the Cardiff model. In health economic evaluation, quality-adjusted life years *(QALYs) are a common measure for the effectiveness of medical treatments.*

4.3.2. Prices

A short note on terminology: throughout this paper, I frequently refer to drug prices, when what is meant is strictly speaking the yearly cost per patient associated with the use of a particular drug (which is determined by the price per drug unit), i.e. drug costs. The reason for this is to avoid the confusion of drug costs with the total treatment costs per patient simulated by the model.

Table 3 shows the three drug prices which are included in the metamodel and their current level as of April 2021. They will be varied when simulating training data for the metamodel (section 4.5.1). Dapagliflozin is subject to the Dutch reference pricing system and as such, its maximum price per pack is defined as the average price in the four reference countries [27].

Table 3: yearly drug costs

Drug	Costs per day	Cost per year
Dapagliflozin	€ 1.26 [36]	€ 459.85

SU[4]	€ 0.09 [36]	€ 33.46
Metformin	€ 0.05 [36]	€ 18.26

The yearly treatment costs put into the Cardiff model (Table 4, far right) are calculated by taking the daily costs times the number of days in a year, plus additional pharmacy dispensing fees of € 28 per drug [37], plus consumables in the treatment profiles where insulin is involved. Insulin costs are calculated separately by the Cardiff model, based on the weight and weight change of the individual simulated patient (€0.014/kg/day for insulin, €0.033/kg/day for insulin + insulin bolus [38]). Since insulin costs are equal for both treatment arms, they are not further discussed here.

Table 4: yearly treatment costs by treatment arm (excluding insulin)

Treatment	Drug costs (excluding insulin)	Consumables	Pharmacy costs	Treatment costs (excluding insulin)
Metformin + dapagliflozin	€ 478.11	-	€ 56	€ 534.11
Metformin + SU	€ 51.72	-	€ 56	€ 107.72
Metformin + insulin	€ 18.26	€ 284.90	€ 56	€ 359.16
Metformin + insulin + bolus	€ 18.26	€ 1139.50	€ 56	€ 1213.76

For the prices put into the metamodel, I decided to use only the yearly drug costs themselves, without the overhead costs. For the prices I will discuss later, the daily price can thus be recovered by dividing by 365.25. The reason for doing this is that the drug prices are the interesting part of the costs, while the overhead costs are constant across treatment profiles.

4.3.3. Clinical efficacy

In addition to the prices of the drugs involved in the treatment of type 2 diabetes, the impact of changes in clinical efficacy of dapagliflozin is to investigated as well. The goal is not to anticipate any specific scenario, but rather to provide an idea how such an anticipatory analysis can be undertaken. The impact of dapagliflozin on body weight has been chosen as an example, since it is typically cited as a major benefit of the drug [39].

According to clinical evidence [40], dapagliflozin leads to an average reduction in bodyweight of 3.3 kilograms per year used. Reduced bodyweight has a direct benefit to a patient's quality of life and also decreases the risk of adverse clinical events, most notably heart disease. It is reasonable to assume that the impact on adverse events and survival will also affect the treatment costs. The release of hypothetical novel trial or registry data, which shows that the effect of dapagliflozin on bodyweight is different from what is currently known, could therefore change both the expected costs and benefits of treatment with dapagliflozin.

[4] In the case of sulfonylurea, there are multiple generic drugs which are used in diabetes therapy, therefore a weighted average is used to determine the cost. See Table 16 in the appendix for precise calculations.

4.3.4. Other inputs to the Cardiff model

As discussed, three price parameters and one clinical parameter are varied, while the rest of the model setup stays constant. Since cost-effectiveness analyses dealing with the dapagliflozin-SU comparison using the Cardiff model have been performed before, with similar results to the base-case in my analysis, I will not discuss the input choices in depth and only point out some important details. A detailed list of all inputs used is provided in the appendix.

Means and standard errors for all inputs were sourced from the literature where possible. When there was no standard error reported and no option to calculate it, the standard error was assumed to be 10% of the mean. An important exception to this are the patient characteristics and clinical risk factors: for the inputs where the standard error is known, it is small, mostly below 1% of the mean. Setting the unknown ones to 10% resulted in a massive spread of the resulting cost-effectiveness pairs which were not at all in line with the available literature. It also made the model very unresponsive to changes in the treatment profiles. This was considered to be both unrealistic and unsuitable for analysis, and I therefore set all unknown standard errors in the patient characteristics and clinical risk factors to 1% instead of 10% of the mean.

4.4. Statistical metamodel

4.4.1. Causal structure of the case

The main outcomes of a health economic model are incremental costs and effects. Costs are directly affected by drug prices, since they comprise parts of the treatment costs. Clinical efficacy parameters impact the effects and, as motivated earlier, potentially also the costs associated with a treatment.

In each model run in a PSA, there are lots of factors which have an effect on the incremental costs and effects of the respective simulation run:

- Sampled risk equations
- Sampled treatment effects
- Sampled costs, utilities and patient population

These randomized factors induce covariance between clinical costs and effects: for a simulation run where the treatment leads to far fewer adverse events relative to the comparator, incremental costs will be low and QALYs high, and vice versa (negative correlation between costs and effects). The reverse can be reasonable as well: for simulation runs where patients in the treatment group live much longer and enjoy more QALYs, costs will also be higher (positive correlation). These are just two examples of correlation effects; many more are imaginable. It is thus very important to enable the model to capture and reproduce this correlation between costs and effects. Since the correlation-inducing factors are not included in the training data, they are treated as unobserved variables (U).

This causal structure of an analysis can be represented through a directed acyclical graph (DAG) [7]. The DAG is important because it helps us decide how to interpret the model's parameters. A model can only report statistical associations. If we are investigating some association between two variables A and B, the model cannot determine whether A causes B, B causes A, A and B are both caused by some third variable C, or a combination of those. The causal interpretation depends on domain knowledge outside of the model, which can be expressed through the DAG. Causal interpretations within this study are conditional on the DAG and should be understood in the following manner:

A has an effect of β on B, assuming that the DAG is correct

The DAG for the case study is shown in Figure 4 and Table 5. P and W are experimentally controlled; therefore they are not caused by anything and the association they have with their respective outcomes can be interpreted as a causal effect. While U has a causal effect on both outcomes, it does not affect any of the exposures and is thus not a confounder with respect to E and C. The parameters of E and C are unbiased and represent full causal effects. The correlation parameter of E and C has no causal interpretation, it is purely a measure of correlation. E does not cause C and vice versa; they are both caused by the unobserved variables U as well the exposures P and W.

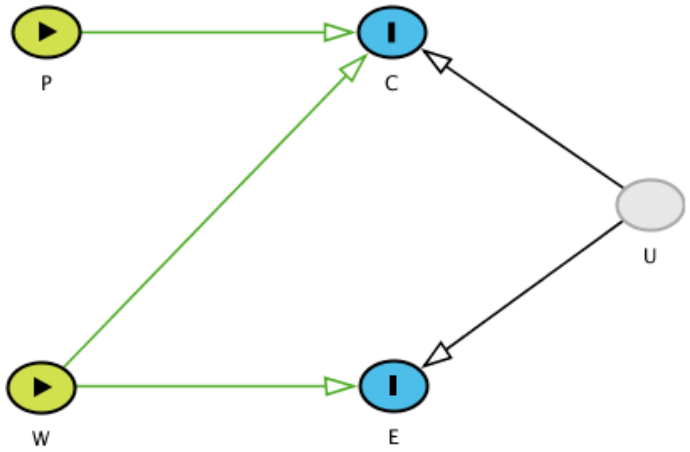


Figure 4: Directed acyclical graph (DAG) illustrating the causal structure of the case

Table 5: Legend of DAG variables

Variable	Description	Type
C	Incremental costs	Outcome
E	Incremental effects	Outcome
P	Drug prices	Exposure
W	Weight change in dapagliflozin patients	Exposure
U	Unobserved variables	Affect the outcomes but not the exposures

4.4.2. Model definition (likelihood)

A multivariate normal distribution is parameterized by a vector of means and a covariance matrix, which is in turn parameterized by a vector of standard deviations and a correlation matrix. The following model includes two outcomes, incremental costs and effects, and thus has a *bivariate* normal distribution as its outcome. In line with the DAG presented in the previous section, the mean effects are a (linear) function of the weight change parameter, while the mean costs are a linear function of the weight change parameter and the drug prices. The method is a generalization of ordinary linear regression to multiple outcomes.

$$\begin{pmatrix} E_i \\ C_i \end{pmatrix} \sim Multivariate\ Normal\left(\begin{pmatrix} \mu_{E,i} \\ \mu_{C,i} \end{pmatrix}, S\right)$$

$$\mu_{E,i} = \alpha_E + \beta_{WE} * W_i$$

$$\mu_{C,i} = \alpha_C + \beta_{WC} * W_i + \sum_{j=1}^{3} \beta_{P_j} * P_{ij}$$

$$S = \begin{pmatrix} \sigma_E & 0 \\ 0 & \sigma_C \end{pmatrix} \times R \times \begin{pmatrix} \sigma_E & 0 \\ 0 & \sigma_C \end{pmatrix} = \begin{pmatrix} \sigma_E^2 & \sigma_E \sigma_C \rho \\ \sigma_E \sigma_C \rho & \sigma_C^2 \end{pmatrix}$$

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

Table 6: legend of model variables and parameters

	Name in DAG and equations	Name in code	Description
Variables (data)	E	effects	Incremental effects
	C	costs	Incremental costs
	W	weight	Weight change parameter
	$P_{dapagliflozin}$, $P_{sulfonylurea}$, $P_{metformin}$	p_dapa, p_SU, p_met	Drug prices
Parameters	α_E, α_C	alphaE, alphaC	Intercepts
	β_{WE}	betaE_weight	Effect of weight change on incremental effects
	β_{WC}	betaC_weight	Effect of weight change on incremental costs
	$\beta_{dapagliflozin}$, $\beta_{sulfonylurea}$, $\beta_{metformin}$	betaC_dapa, betaC_SU, betaC_met	Effects of prices on incremental costs
	σ_E, σ_C	sigma[1], sigma[2]	Standard deviations
	R	Rho	Correlation matrix
Transformed parameters[5]	μ_E, μ_C	muE, muC	Means
	ρ	-	Correlation coefficient
	S	-	Covariance matrix

The mathematical model definition comprises the likelihood of the Bayesian model (see section 4.1). Each of the parameters is assigned a prior distribution. Bayesian inference is about estimating the joint distribution of parameter values that is most consistent with the prior distributions, the likelihood and the observed data. The specific prior distributions are discussed in section 4.6.

4.4.3. Technical implementation

The data processing and model specifications are done in R. Model and data input are then passed to Stan via the RStan package [41]. Stan will produce a distribution of samples for each parameter in the defined surrogate model; all of these sample distributions combined comprise the posterior distribution. The sampled parameters can be used to simulate new cost-effectiveness pairs,

[5] Transformed parameters are a deterministic implication of other parameters. Therefore they do not have a prior distribution.

dependent on the defined inputs for costs and clinical affects. Processing and visualization of the obtained samples happen again in R.

As recommended in the Stan User's Guide, the data are standardized before the parameters are estimated [42]. The posterior distribution therefore includes both standardized and unstandardized parameters. The primary reason for doing this is that it that even though the standardized parameters are not used for any of the further analyses, they are helpful for interpretation. Additionally, it makes the Stan program run faster.

Since the correlation coefficient ρ is invariant to scale, it can be ignored when rescaling the other parameters. The two mean equations in the previous sections are of the form

$$y_i = \alpha + \beta * x_i + e$$

Where y is the outcome, x is the predictor and e is the stochastic deviation from the mean:

$$e \sim normal(0, \sigma)$$

For brevity I will use this generic form instead of the full form of the equations. It can be transformed by inserting the respective outcomes, predictors and parameters. The standardized value of any observed outcome y_i is defined as

$$z_y(y_i) = \frac{y_i - \bar{y}}{sd(y)}$$

The sample mean $\bar{y}$ and standard deviation $sd(y)$ are calculated as follows:

$$\bar{y} = \frac{1}{N}\sum_{i=1}^{n} y_i$$

$$sd(y) = \left(\frac{1}{N}\sum_{i=1}^{n}(y_i - \bar{y})^2\right)^{\frac{1}{2}}$$

The calculations for the predictor x are analogous. Since the data are standardized, the parameters are estimated on the standardized scale as well, and the regression equation turns to

$$z_y(y_i) = \alpha' + \beta' * z_{x(}x_{i)} + e'$$

The unstandardized parameters are recovered in the following manner:

$$\alpha = sd(y)\left(\alpha - \beta'\frac{\bar{x}}{sd(x)}\right) + \bar{y}$$

$$\beta = \beta'\frac{sd(y)}{sd(x)}$$

$$\sigma = sd(y)\sigma'$$

For a derivation of these calculations see the Stan User's guide [42].

4.5. Simulating data sets

This section deals with the creation of adequate training data which the metamodel can be trained on. As explained earlier, the Cardiff model has the option to perform a probabilistic sensitivity analysis (PSA). In order to determine the impact that the input parameters have, they have to be varied and the results stored together with the input values used. The basic workflow looks like this:

- Randomize parameters
- Run health economic model
- Store results together with inputs used

The first step is to randomize the input values put into the original model. Most parameters in a model have no logical boundaries: a price for instance, can range from 0 to infinity. Thus we have to decide on a range for each input which constraints the simulation to meaningful values. Furthermore, in order to produce useful training data, it has to be ensured that each region in the *parameter space* defined by these boundaries is adequately represented in the simulation.

If the model is directly coded in a high-level programming language like Python or R, it is usually easy to define a loop which runs the model many times with randomized parameters and store the results together with the inputs used for each simulation run. In that case, for each input a random value within the defined range is drawn for each simulation run in the loop, independently of the other inputs. Given enough runs, this is guaranteed to cover the entire parameter space well. Figure 5 (right) shows an example for two generic parameters that are varied between 0 and 1, 2000 times.

However, if the model is accessed through a visual interface (e.g. spreadsheet software like Microsoft Excel or a dashboard of some sort), the process is more challenging and can usually not be automated. We have to manually change the input parameters and store the results for each simulation. Doing this for thousands of times is obviously not feasible. Instead, a limited number of input scenarios has to be defined; then the model is run multiple times for each of these. In order to minimize the time spent on this, we need to reach an accurate representation of the model with as few experiments as possible. Degeling et al [12] recommend 10 experiments per input parameter involved, which leads to 20 experiments for the two-parameter example. The actual case study includes more parameters and experiments (see next section).

Simply sampling random values within the parameter ranges is not a good option anymore, since it is inevitable that certain regions of the parameter space will be underexplored while others will be overrepresented and therefore redundant. The larger the number of dimensions (parameters), the more severe this problem becomes. See Figure 5, left.

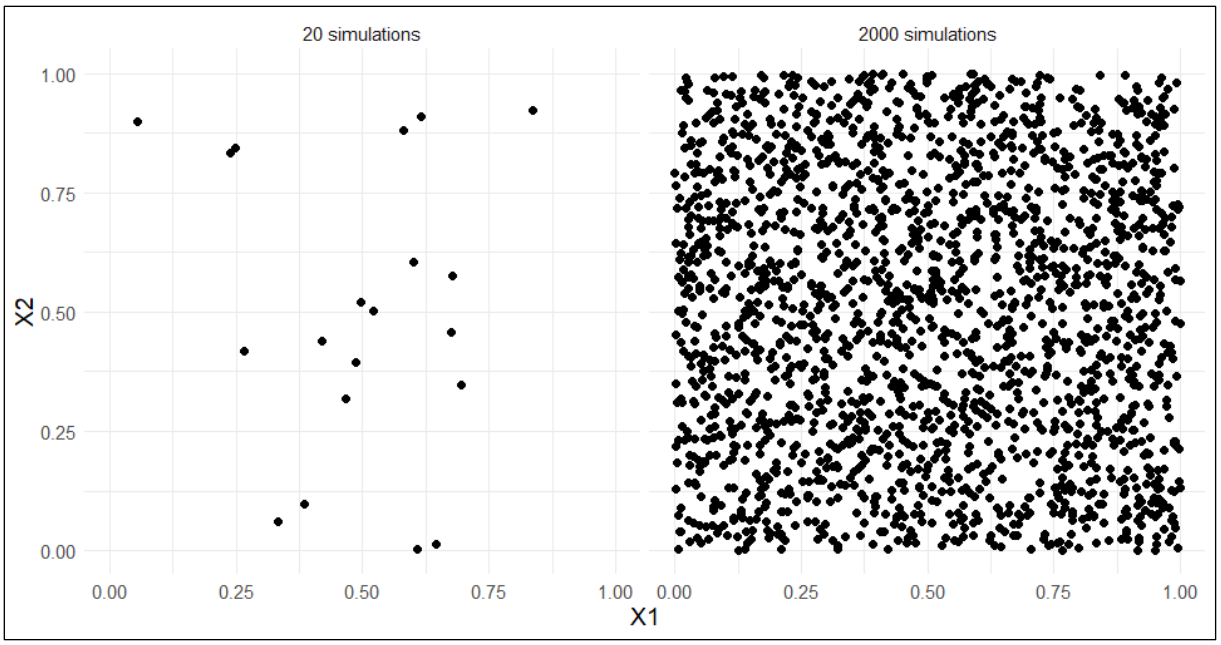


Figure 5: 20 and 2000 parameter combinations, obtained through random sampling using uniform distributions

The input values thus have to varied in a manner which covers the parameter space as well as possible, given the limited number of experiments. A useful method to achieve this is Latin Hypercube Sampling (LHS) [43]. The LHS package for the R language provides an interface. LHS requires the user to define a number of N experiments (combinations of parameter values) and a range for each parameter, within which the parameter will be varied. The algorithm then returns N combinations of all parameters. In the aggregate, these parameter combinations cover the parameter space as well as possible, given the number of experiments. Figure 6 shows 20 random samples for two parameters obtained through LHS.

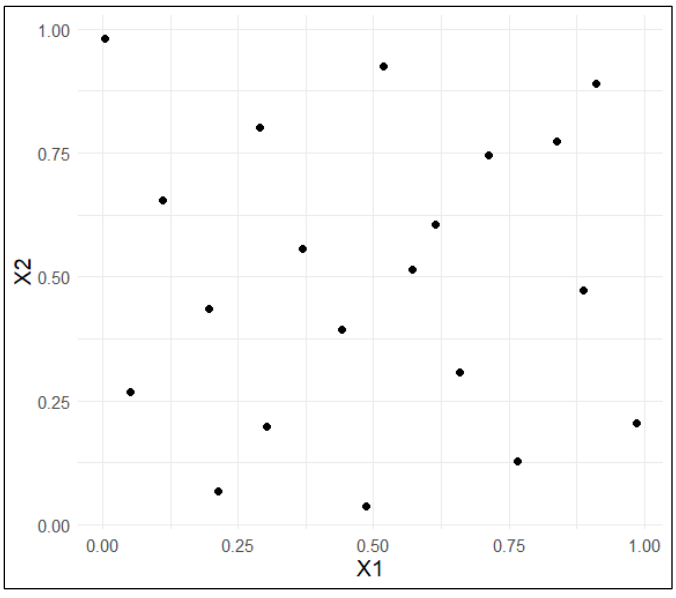


Figure 6: 20 parameter combinations obtained through LHS sampling

The method can be extended to more than two dimensions (parameters). For three dimensions we would have a cube instead of a square, and for more dimensions a hypercube (multidimensional

generalization of a cube). This is where the name of the method is derived from. Dimensions higher than two are difficult to visualize, but work analogously to the two-dimensional example.

4.5.1. Application to the case study

Four input parameters and thus four dimensions are involved in the analysis:

- Annual price of dapagliflozin
- Annual price of sulfonylurea (SU)
- Annual price of metformin
- Annual weight change associated with dapagliflozin

For the prices, I defined the range to be between 0 and the current price times 2. Even if substantial cost increases seem unlikely in the current situation, the model should nevertheless be able to provide meaningful approximations for situations with higher prices. The clinical parameter (weight change of patients receiving dapagliflozin) was varied between 0 and the currently estimated mean value times 2.

In line with the recommendations by Degeling et al, I used 40 experiments, so 40 different input combinations (10 per parameter involved). For each experiment I simulated 250 cohorts of 1000 patients from the Cardiff model, resulting in a total of 10000 observations. The appendix includes a detailed list of the input parameters used, as well as a plot of the obtained training data.

4.6. Priors

The prior parameters are unstandardized according to the formulas introduced in section 4.4.3, which can only happen after the data is collected. This is why the priors are discussed here and not in the model definition section (4.4.2). The following priors (on the standardized scale) were chosen for the model parameters:

$$\alpha_E \sim Normal(0, 0.5)$$

$$\alpha_C \sim Normal(0, 0.5)$$

$$\beta_{WE} \sim Normal(0, 0.5)$$

$$\beta_{WC} \sim Normal(0, 0.5)$$

$$\beta_{P_j} \sim Normal(0, 0.5)$$

$$\sigma_E \sim Exponential(1)$$

$$\sigma_C \sim Exponential(1)$$

$$R \sim LKJCorr(2)$$

Since the data set used is large relative to the model (10000 observations versus 10 parameters), the Bayesian inference is highly insensitive to the choice of prior distributions (see the example in section 4.1). Nevertheless, it is considered good practice to set sensible prior distributions which assign more probability mass to more reasonable parameter values, and less to extreme and unrealistic ones [7]. Thus, the prior distributions for all parameters are *mildly informative*, which means that they are skeptical of extreme and unrealistic values. For instance, it seems unlikely that an increase of one standard deviation in the price of metformin leads to a change of more than one standard deviation in the incremental treatment costs, in either direction. The prior distribution for the metformin β-parameter expresses this belief.

The slightly exotic "LKJCorr" distribution (proposed by Lewandowski et al [44]) deserves a short explanation: it serves as a prior for the correlation matrix R, parameterized by a set of correlation coefficients. For the given model, there are two outcomes whose correlation is to be estimated, and therefore is only one coefficient ρ. For values higher than 1, the distribution puts more probability mass on small correlations near 0, and vice versa. $LKJCorr(2)$ is thus a mildly informative prior in favor of smaller correlations.

Figure 7 visualizes the prior distribution of the slope parameters. This plot by itself is not very interesting, but it allows to directly compare the prior distribution to the posterior distribution in Figure 9.

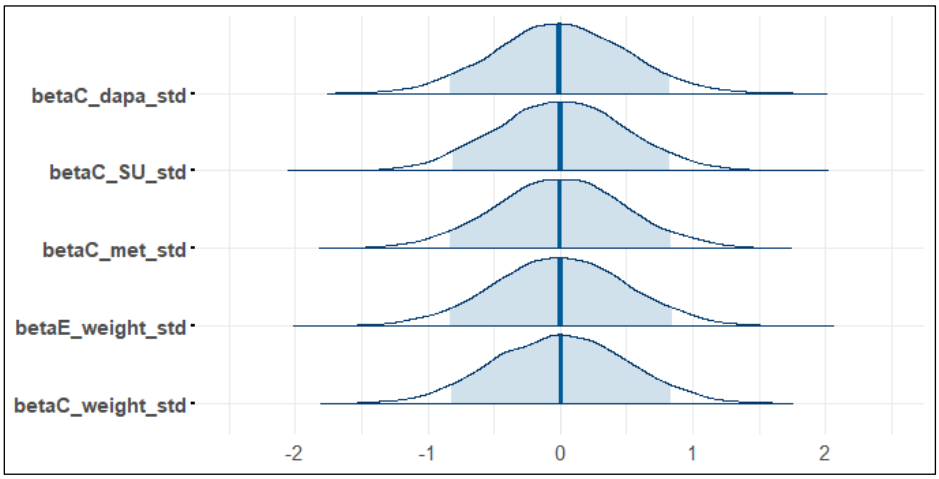


Figure 7: standardized prior distribution of beta-parameters.

Using the model definition and the prior parameters, we can simulate outcomes. This is helpful to determine whether the priors are sensible: the simulated values should focus on areas which are deemed to be realistic based on prior knowledge, but also not totally exclude extreme outcomes. Figure 8 shows 4000 simulated cost-effectiveness pairs based on the unstandardized prior parameters. Because the parameters are unstandardized using the collected data, the alpha parameters are centered around the sample mean of the simulated data. The parameters are however not conditioned on the data yet, they are random draws from the prior distributions.

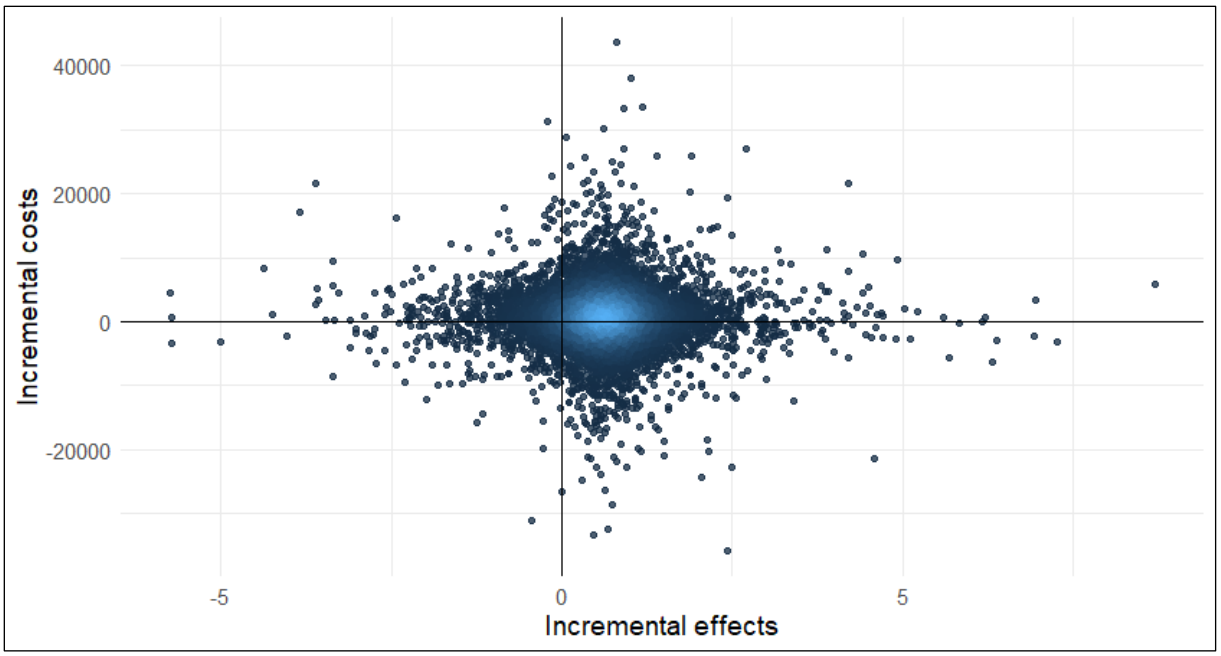


Figure 8: prior predictive distribution

Again, Figure 8 should be compared to the posterior predictive distribution in Figure 10.

5. Results

5.1. Parameter estimates

As motivated earlier, one of the motivations for building a metamodel is to uncover the precise impact that individual inputs in the original model have on the results. One can get an idea about this by investigating the regression coefficient estimates of the model (see model definition).

Table 7 and Table 8 summarize the unstandardized and standardized estimates, respectively. They are summarized in terms of mean, standard deviation and 90% credibility interval. All these reported values are merely descriptive statistics summarizing the distribution of values per coefficient, intended to ease interpretation. When simulating new observations from the metamodel, all of the coefficient samples are used. The correlation between the two outcomes, incremental costs and effects, is reliably negative.

Unstandardized coefficients are always sensitive to scale, which can make them difficult to interpret if they relate to variables of different magnitude. This becomes obvious when investigating the coefficients for the prices of dapagliflozin and the comparator, SU. Since the price of dapagliflozin is generally much higher, the coefficient will be smaller relative to the price, compared to SU. This is why the coefficient of SU has a higher absolute value. When investigating the standardized coefficients (Figure 9) it is immediately obvious that the dapagliflozin price has a much larger impact on the incremental costs.

Table 7: unstandardized joint posterior distribution of all parameters

Coefficient	Mean	Standard deviation	HPDI – 5% bound	HPDI – 95% bound
alphaC	-1233.02	91.90	-1383.94	-1082.27
betaC_dapa	4.55	0.09	4.39	4.70
betaC_SU	-6.51	2.19	-10.09	-2.90

betaC_met	1.75	2.21	-1.83	5.38
alphaE	0.51	0.01	0.50	0.53
betaE_weight	-0.03	0.00	-0.03	-0.03
betaC_weight	20.06	12.77	-1.02	41.06
sigma[1]	0.51	0.00	0.50	0.51
sigma[2]	2389.32	16.72	2361.58	2416.78
Rho	-0.14	0.01	-0.15	-0.12

Table 8: standardized joint posterior distribution of all parameters

Coefficient	Mean	Standard deviation	HPDI – 5% bound	HPDI – 95% bound
alphaC_std	0.00	0.01	-0.01	0.01
betaC_dapa_std	0.45	0.01	0.43	0.46
betaC_SU_std	-0.03	0.01	-0.04	-0.01
betaC_met_std	0.01	0.01	-0.01	0.02
alphaE_std	0.00	0.01	-0.02	0.02
betaE_weight_std	-0.11	0.01	-0.13	-0.09
betaC_weight_std	0.01	0.01	0.00	0.03
sigma_std[1]	0.99	0.01	0.98	1.01
sigma_std[2]	0.89	0.01	0.88	0.90
Rho	-0.14	0.01	-0.15	-0.12

We can investigate the standardized estimates visually, in order to compare the relative impact that each parameter has. Figure 9 illustrates the posterior density of the standardized beta-parameters, with the credibility interval indicated in light blue. The price of dapagliflozin has by far the strongest impact on the resulting treatment costs, while the impact of the other two prices is small. The price of SU has a small but credible negative effect on the incremental costs, as is to be expected. The price effect of metformin is not credibly different from 0.

The weight change parameter has a credible negative impact on the incremental effects and a very small positive impact on costs. The weight change increment is expressed in kilogram per year, and weight loss is associated with utility gain, while weight gain is associated with utility loss. Since dapagliflozin leads to a reduction in bodyweight, the weight increment is negative, leading to a *positive* utility increment when multiplied with the negative parameter. The mean unstandardized

effect of weight change is -0.03 QALYs per kg/year. This means that an increase of the mean weight loss per year of one kilogram leads to an increase in incremental QALYs of 0.03.

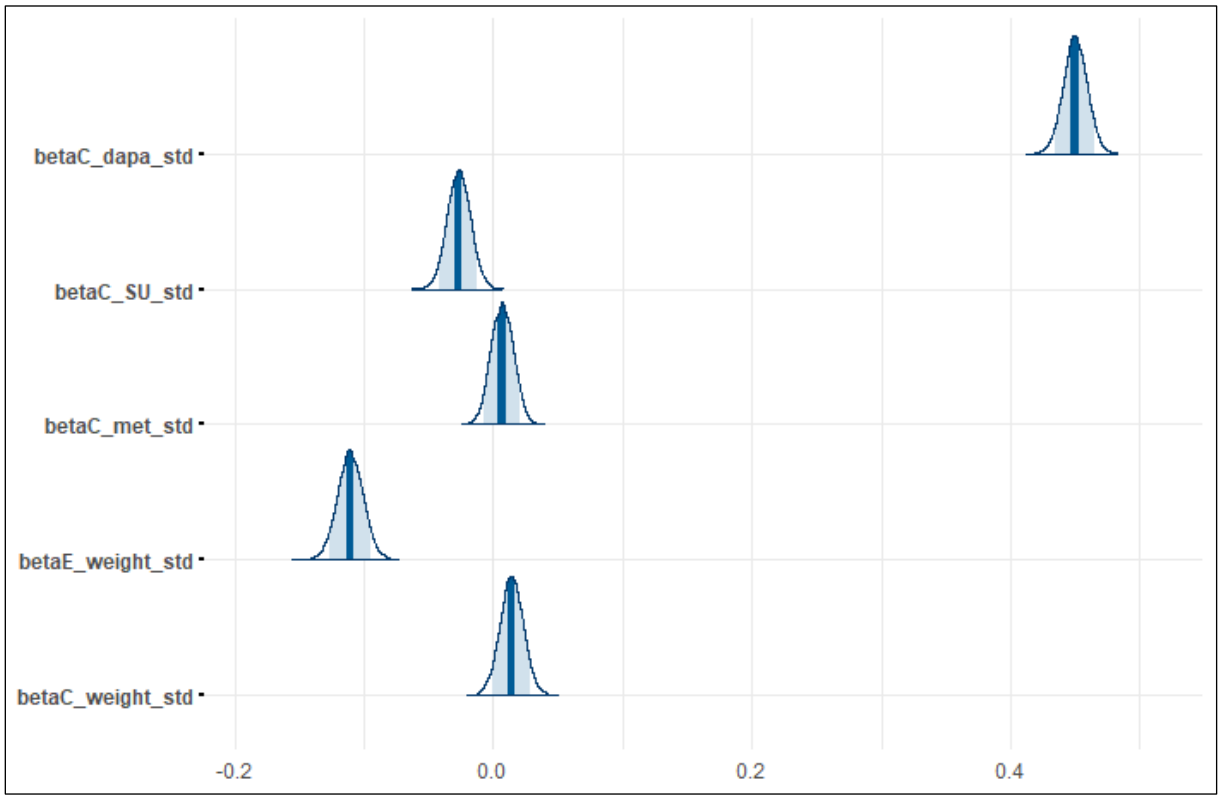


Figure 9: standardized posterior distribution of beta-parameters

5.2. Posterior predictive simulation and model validation

After the metamodel has been trained, the predictions it makes can be compared to output samples from the original model which the surrogate has not been trained on (Figure 10). For demonstration, 10000 new samples are drawn from the Cardiff as well as from the surrogate model and visually compared. Figure 10 shows the draws from the Cardiff model on the left and those from the metamodel on the right. Lighter areas indicate higher density. The inputs used represent the Dutch situation as of April 2021 (see Table 9).

Table 9: Input values used for validation of the metamodel's performance

Input	Value
Price dapagliflozin	459.85€
Price metformin	18.26€
Price sulfonylurea	33.46€
Effect of dapagliflozin on patients' bodyweight	-3.3 kg / year

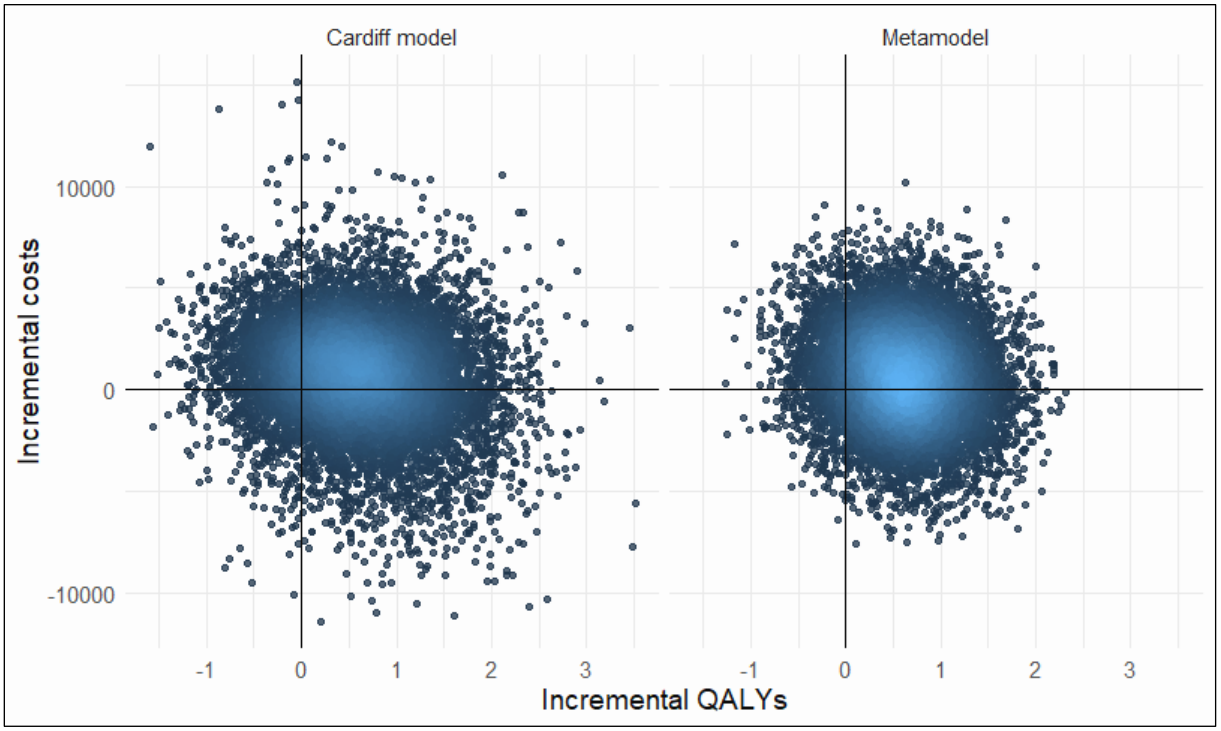


Figure 10: comparison of the metamodel's predictions with a new dataset obtained from the Cardiff model, which the metamodel has not been trained on

Table 10: mean predictions by the Cardiff model and the metamodel

	Mean incremental costs	**Mean incremental effects**	**Incremental costs – standard deviation**	**Incremental effects – standard deviation**
Cardiff model	694.52	0.60	2655.10	0.68
Metamodel	645.27	0.61	2414.53	0.51

Table 10 compares the predictions made by the Cardiff model with those of the metamodel. The metamodel predicts slightly lower mean incremental costs. Furthermore, it underestimates the uncertainty of predicted costs and effects compared to the original model. The two results are certainly not identical. However, this is not necessarily a problem: the goal of a metamodel is not to reconstruct the original model, but rather to provide an approximation which is good enough to answer all questions of interest. In the context of the study, the two main questions are::

- What are the expected incremental costs caused and QALYs gained by a patient starting dapagliflozin treatment?
- What is the probability that the treatment is cost-saving and/or cost-effective?

The first question is adequately answered by the fact that the metamodel and the Cardiff model predict near identical mean values. We can investigate question 2 by looking at the cost-effectiveness acceptability curve (CEAC), shown in Figure 11. The curve for the predictions of the Cardiff model is indicated in black, and that for the metamodel in blue. I would like to draw special attention to two points in the chart:

- X = 0, which shows the probability that the treatment is cost-saving

- X = 20000, which shows the probability that the treatment is cost-effective, given the customary ICER-threshold of 20000€ per incremental QALY

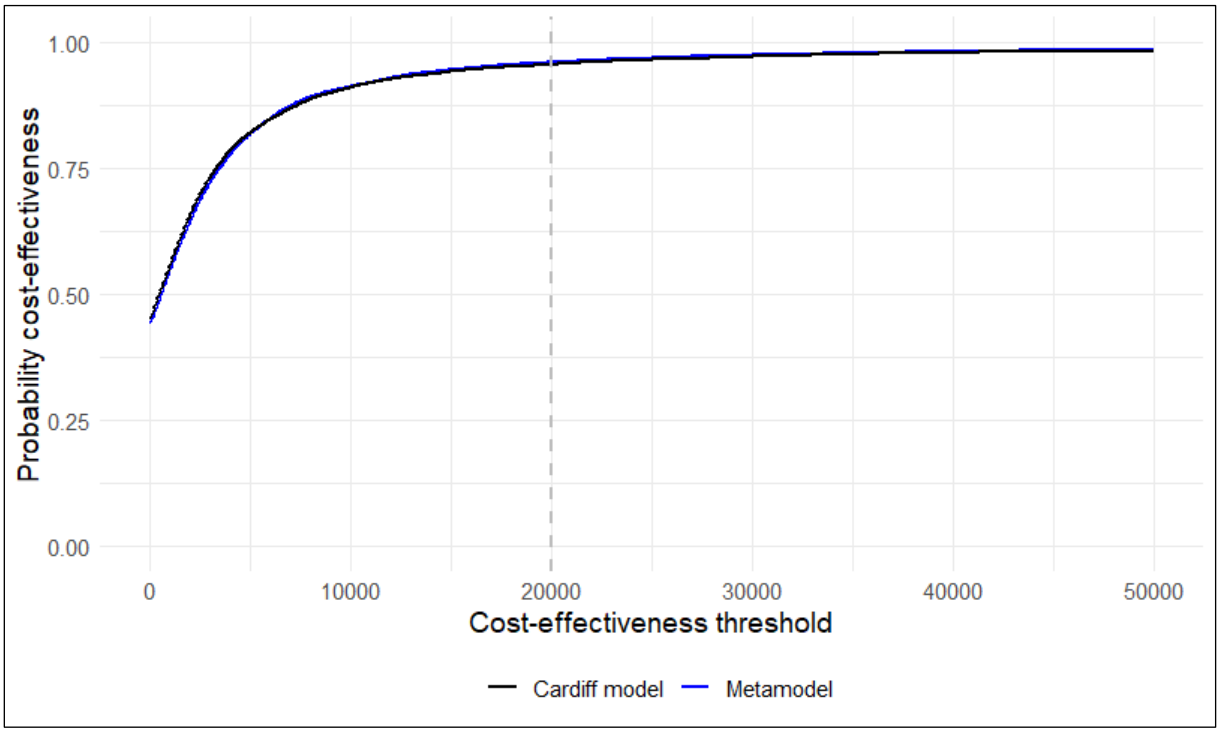


Figure 11: cost-effectiveness acceptability curve (CEAC) for the metamodel (blue) and the Cardiff model (black)

The curves are almost identical, which indicates that the metamodel captures the behavior of the original model well. Thus, the minor differences in the estimates have no distinct impact on the probability that dapagliflozin treatment is cost-saving or cost-effective at any threshold. The Cardiff model and metamodel are interchangeable when it comes to decisions based on estimated costs and effects.

5.3. Model application

In the previous section, one simulation of 10000 cost-effectiveness pairs was performed, using a single set of input parameters. The following analyses are of the same nature, but instead of only one simulation, a large number of different simulations is performed with different input values.

5.3.1. Impact of changes in drug prices and effects

5.3.1.1. Prices

In the dapagliflozin case study, one of the main interests is what the consequences of price changes are on the treatment's costs and cost-effectiveness. To investigate the impact of the dapagliflozin price in isolation, a grid is defined, with evenly spaces input values from € 0 to € 1000. The other input parameters are kept constant at the level defined in Table 9.

Table 11: illustration of input grid used for price variations

	Simulation 1	**Simulation 2**	**Simulation 3**	**…**	**Simulation 999**	**Simulation 1000**
Price dapagliflozin	€ 1.00	€ 2.00	€ 3.00	…	€ 999.00	€ 1000.00
Price metformin	18.26€	18.26€	18.26€	…	18.26€	18.26€

Price sulfonylurea	33.46€	33.46€	33.46€	…	33.46€	33.46€
Effect of dapagliflozin on patients' bodyweight	-3.3 kg / year	-3.3 kg / year	-3.3 kg / year	…	-3.3 kg / year	-3.3 kg / year

For each of these input combinations, 10000 samples are drawn and summarized in terms of mean, 50% credibility interval and 90% credibility interval (Figure 12). Due to stochastic variance in the simulations, the graphs are slightly wiggly. As is to be expected, the price of dapagliflozin has a strong positive impact on the incremental costs of the dapagliflozin treatment arm. Figure 12 shows the predictive distribution of incremental costs at each value of the price grid. The current price of 459.85€ is indicated by the vertical dashed line.

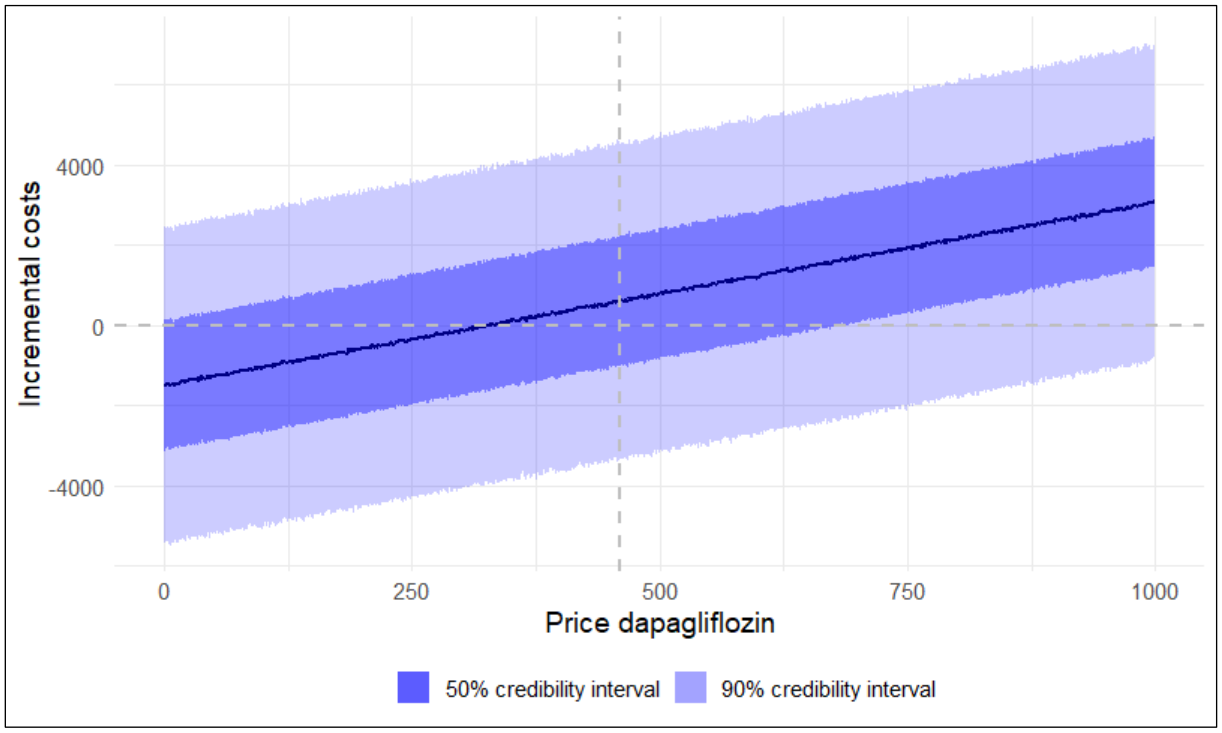


Figure 12: predictive cost distribution for different dapagliflozin prices

Furthermore we can calculate the probability that the treatment is cost-effective and cost-saving at each value of the defined input-grid (Figure 13). The impact on the treatment's cost-effectiveness at a threshold of € 20000 per QALY is small: even at a yearly price of € 1000, dapagliflozin would still have a chance 92% to be cost-effective. This probability would drop if the price was further increased, but such large price increases are unrealistic. The price does have a strong impact on the

probability that the treatment is cost-saving. In order for it to be cost-saving on average, the dapagliflozin price would need to fall below € 360 per year.

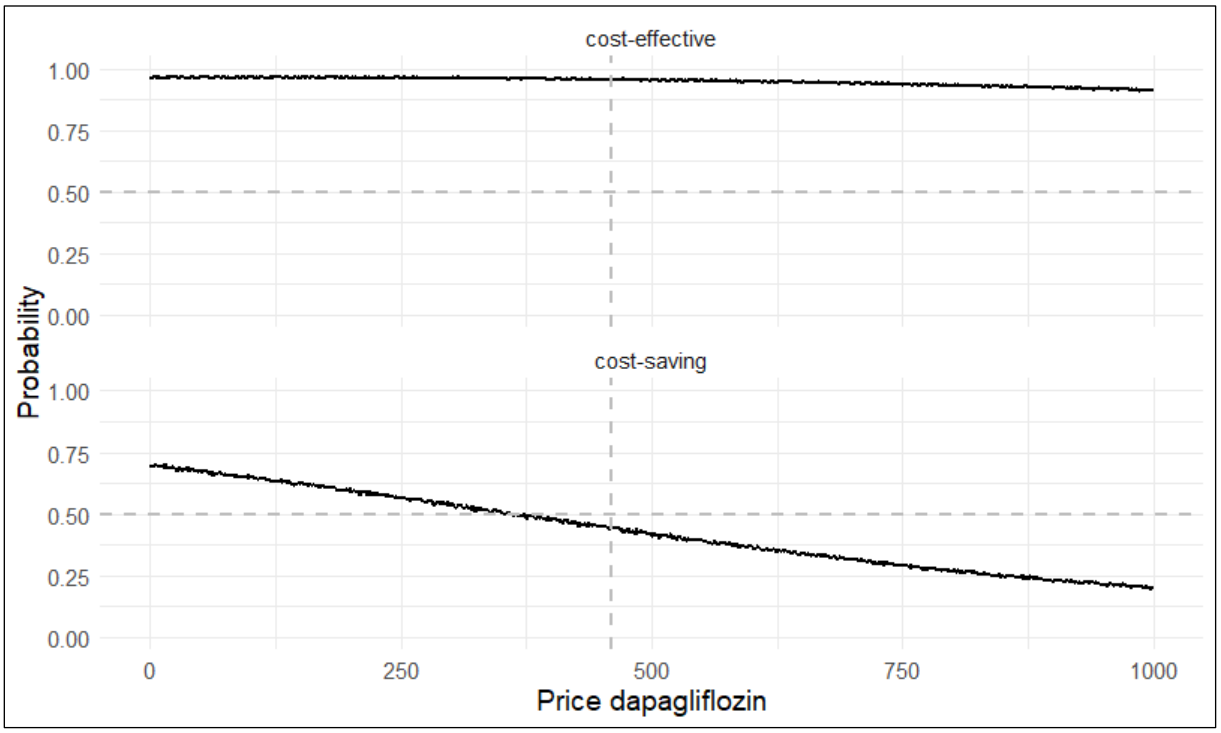


Figure 13: probability that dapagliflozin is cost-effective and cost-saving at different price levels

Analogous plots could be created for the other price parameters involved. However, as can be seen from the parameter distributions in Figure 9, their impact on the incremental costs is small and the mean lines and credibility intervals would be approximately horizontal. Therefore they are omitted for the sake of brevity.

5.3.1.2. Clinical effects

According to the model definition, the weight change parameter is assumed to affect the incremental effects as well as the incremental costs. Similarly to the previous section, I defined an evenly spaced grid of 1000 values from -6.6 (double the effect that is currently known) and 0 (dapagliflozin has no effect on bodyweight), while keeping the other inputs constant.

From the parameter estimates in section 5.1 we know that the weight effect has virtually no impact on the incremental costs, which is why the graph regarding costs (Figure 14, top) is almost horizontal. More surprisingly, the impact on quality of life is fairly small as well: if the yearly weight loss associated with dapagliflozin were double of what is currently known, the incremental QALYs would only be 0.1 higher than they are now. If it had no effect on bodyweight at all, dapagliflozin would still have chance of more than 99% to be more effective than SU. This is surprising and counter-intuitive, since according to earlier research [30, 31], weight loss is one of the main drivers of improved quality of life in dapagliflozin patients. According to the predictions of the Cardiff model, this is not the case. As such, the clinical parameter investigated in this study has a negligible effect on the probability of the treatment being cost-effective or cost-saving (Figure 15). The issue is further discussed in section 6.1.2.

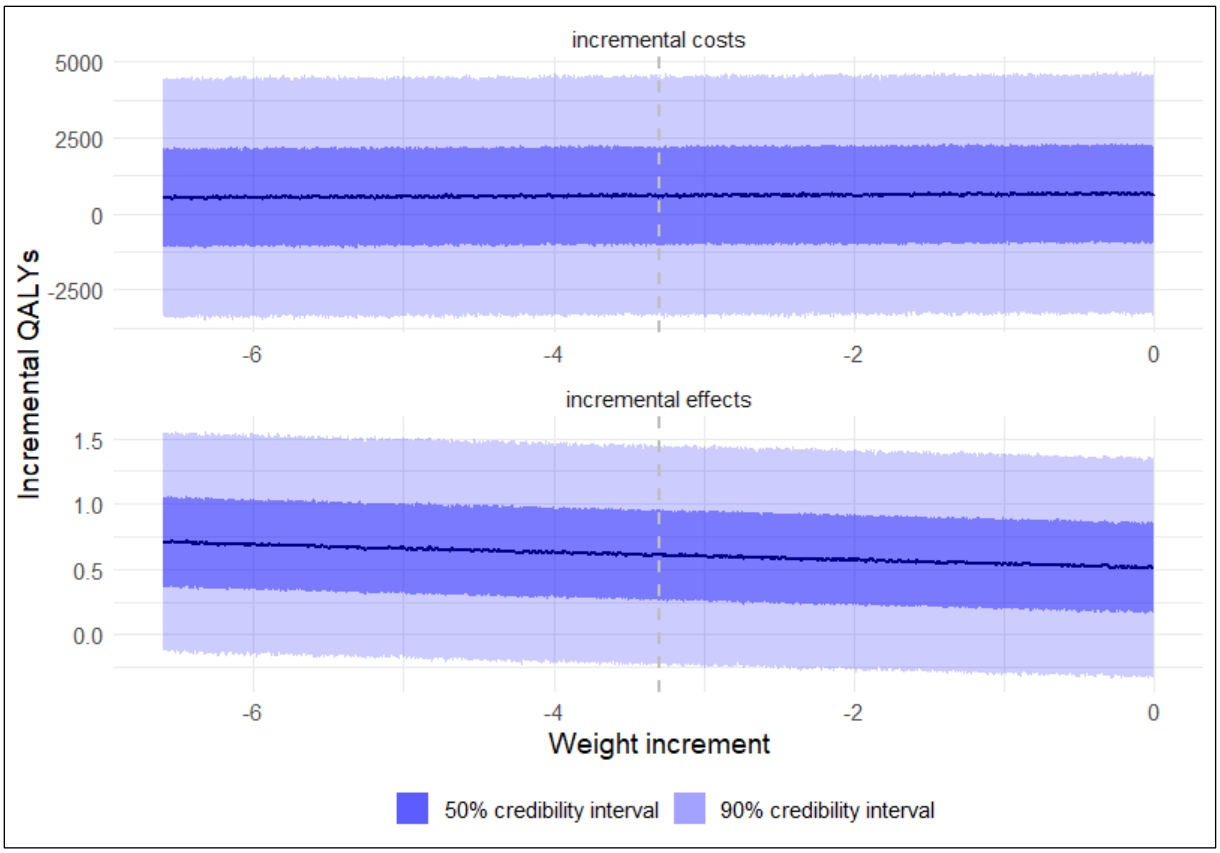


Figure 14: predictive distribution of incremental costs and effects for different values of the weight increment associated with dapagliflozin treatment

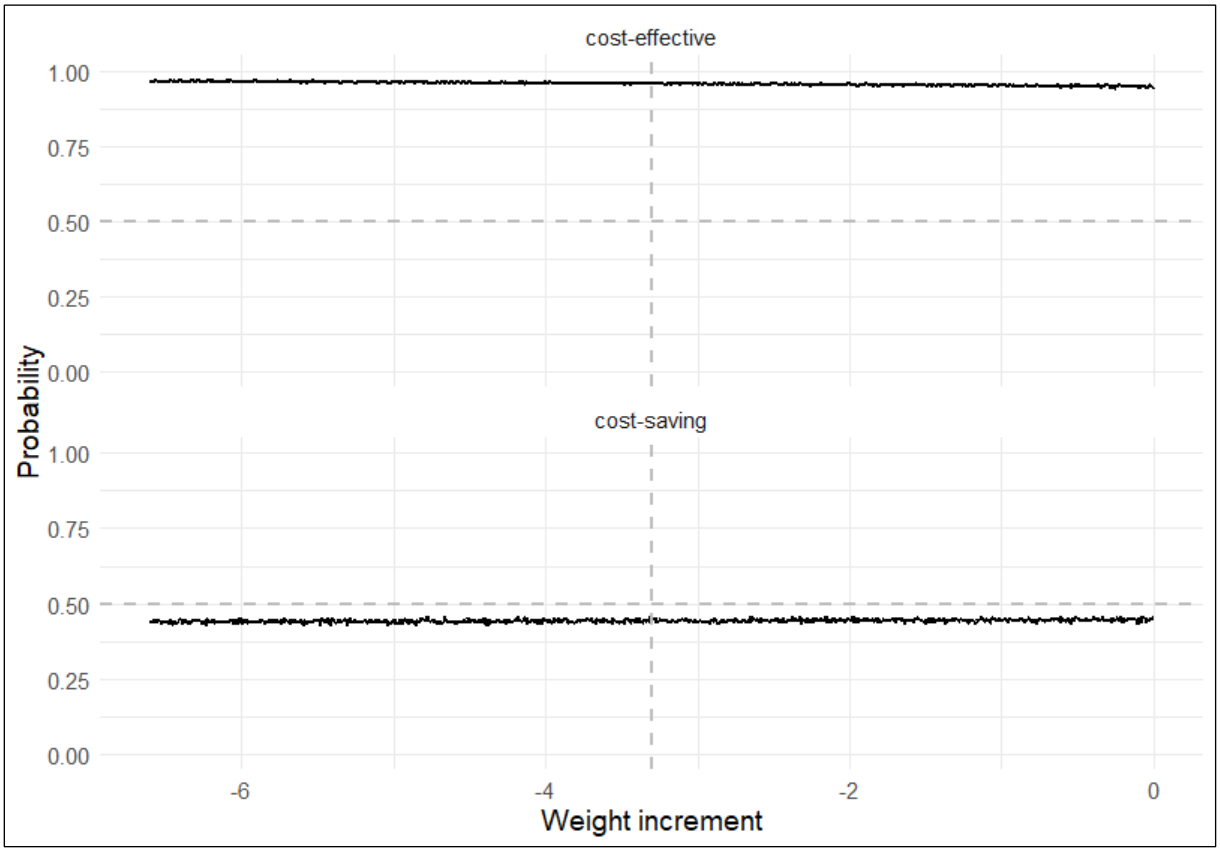


Figure 15: probability that dapagliflozin is cost-effective and cost-saving at different values of the weight increment

5.3.2. Timeline

The prices in the reference countries in the near future are known [45-47] and can be used to anticipate the price of dapagliflozin in the Netherlands (see Table 12, until 2027). Additionally, dapagliflozin's patent was filed in 2008 and will therefore expire in 2028 [29], which will allow generics to enter the market. Unfortunately, the literature provides little information regarding the price development in European countries for drugs which lose their patent protection. In a comprehensive literature review from 2018, Vondeling et al [48] found that while prices generally tend to drop after patent loss, the estimated amount varies a lot both by country and by investigating study. The price development of dapagliflozin is therefore highly speculative. Based on internal forecasts at AstraZeneca, the following price development scheme was assumed. It assumes a sharp initial price drop after patent loss, which then slows down over the course of five years (column 4 in Table 12).

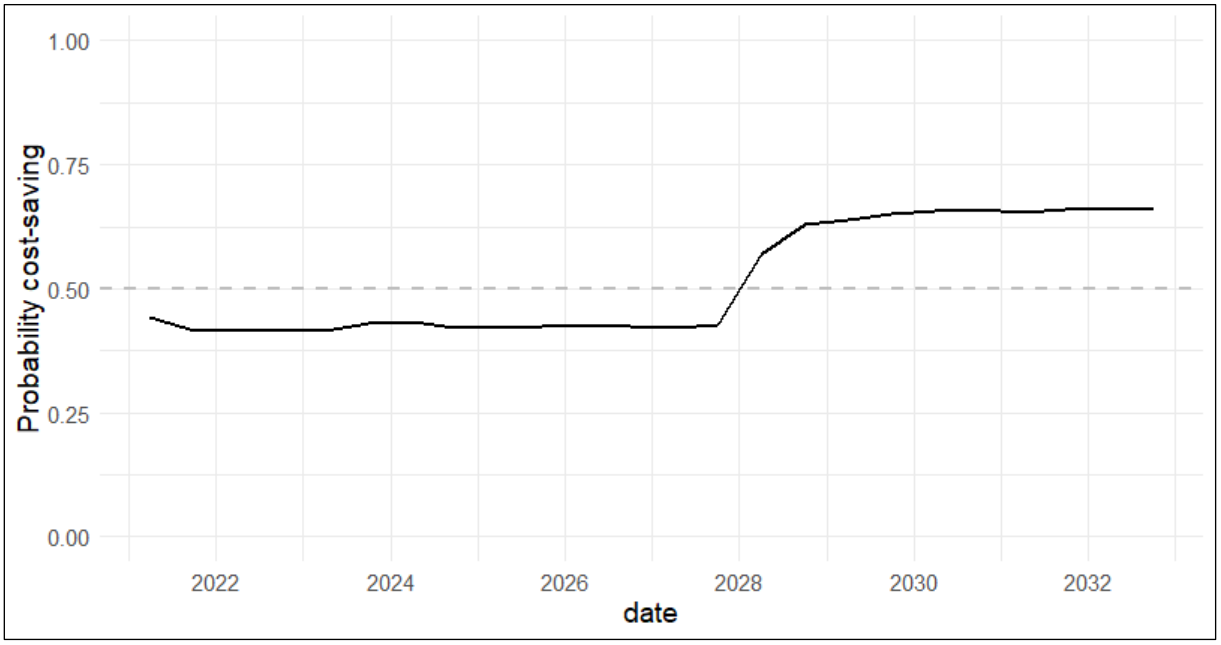
The patents of metformin and the commonly used sulfonylurea have expired a long time ago [49], and therefore no significant price changes are to expected in the coming years. As previously discussed, small price changes of both drugs have a negligible impact on the incremental costs of dapagliflozin treatment. Thus the prices of both are assumed to stay constant.

Table 12: assumed price development of dapagliflozin

date	price_dapa	price_met	price_SU	Price drop due to patent loss, relative to previous year
1-apr-21	€ 487.85	€ 18.26	€ 33.46	-
1-okt-21	€ 527.18	€ 18.26	€ 33.46	-
1-apr-22	€ 527.18	€ 18.26	€ 33.46	-
1-okt-22	€ 527.18	€ 18.26	€ 33.46	-
1-apr-23	€ 522.31	€ 18.26	€ 33.46	-
1-okt-23	€ 522.31	€ 18.26	€ 33.46	-
1-apr-24	€ 517.44	€ 18.26	€ 33.46	-
1-okt-24	€ 517.44	€ 18.26	€ 33.46	-
1-apr-25	€ 517.44	€ 18.26	€ 33.46	-
1-okt-25	€ 517.44	€ 18.26	€ 33.46	-
1-apr-26	€ 517.44	€ 18.26	€ 33.46	-
1-okt-26	€ 517.44	€ 18.26	€ 33.46	-
1-apr-27	€ 517.44	€ 18.26	€ 33.46	-
1-okt-27	€ 517.44	€ 18.26	€ 33.46	-
1-apr-28	€ 258.72	€ 18.26	€ 33.46	-50%
1-okt-28	€ 129.36	€ 18.26	€ 33.46	-50%
1-apr-29	€ 109.95	€ 18.26	€ 33.46	-15%
1-okt-29	€ 101.16	€ 18.26	€ 33.46	-8%
1-apr-30	€ 96.10	€ 18.26	€ 33.46	-5%
1-okt-30	€ 91.30	€ 18.26	€ 33.46	-5%
1-apr-31	€ 86.73	€ 18.26	€ 33.46	-5%
1-okt-31	€ 84.13	€ 18.26	€ 33.46	-3%
1-apr-32	€ 81.61	€ 18.26	€ 33.46	-3%
1-okt-32	€ 79.16	€ 18.26	€ 33.46	-3%

In order to analyze which impact this price development will have on the estimated costs in the coming years, the methodology is analogous to the previous sections. For each point in time, 10000

cost-effectiveness pairs are simulated and then summarized in terms of mean, 50% credibility interval and 90% credibility interval (Figure 16



). The minor price changes up until 2027 have only a negligible impact: the expected incremental costs hover around € 1100 and the chance of the treatment being cost-saving is approximately 30%. After the loss of exclusivity, the costs associated with the treatment decline, culminating in mean incremental savings of € 510 and a probability of 59% of being cost-saving.

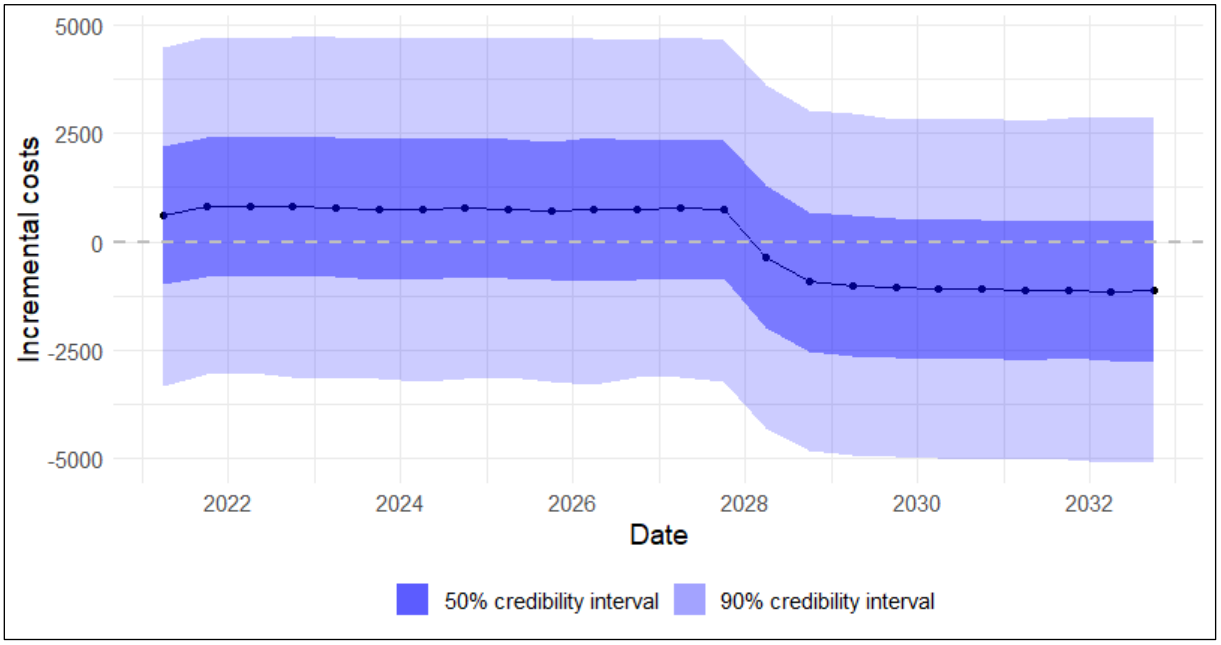


Figure 16: distribution of incremental costs as time passes

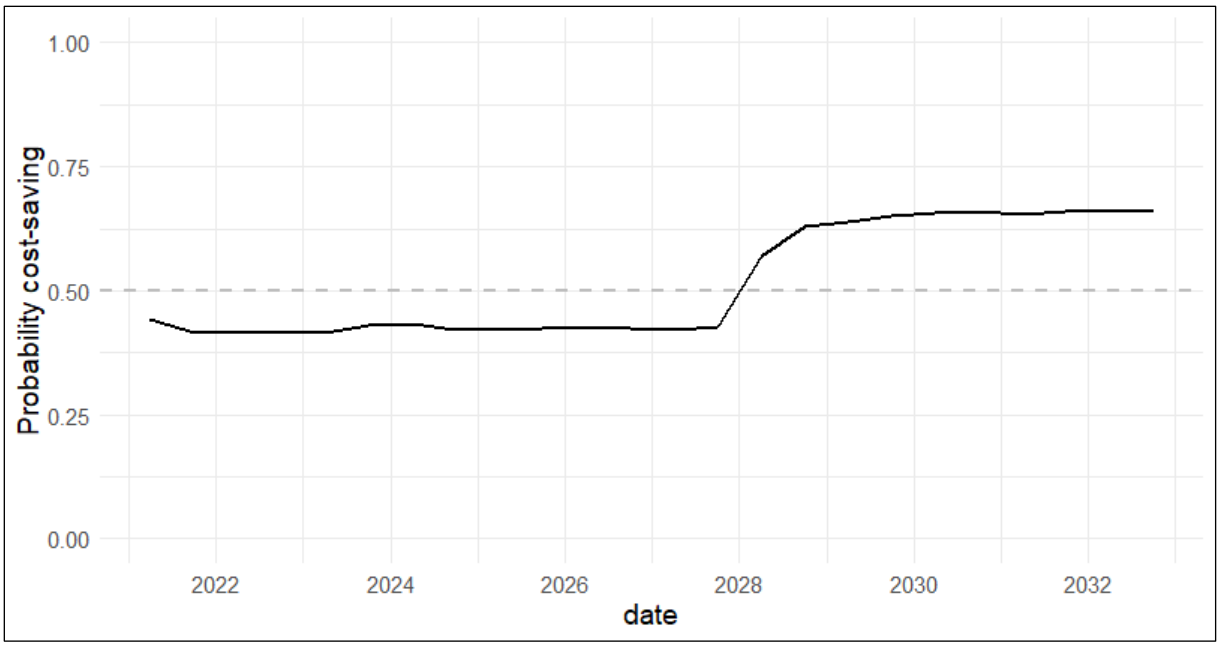


Figure 17: probability that dapagliflozin is cost-saving as time passes

As we saw in the previous section, dapagliflozin's price has very little impact on the treatment's cost-effectiveness. Versions of the above plots for cost-effectiveness are therefore not shown.

Obviously the forecasts are highly dependent on the assumptions made about the price development. However, once more concrete information becomes available, it can very easily be incorporated in the analysis. The price forecasts are not part of the metamodel, but are simply passed to the model as inputs for the simulation. Thus the same model can be used for different price developments.

5.4. Computation/running time

As indicated in the methods section, one of the advantages of a metamodel is the greatly increased computation speed compared to the original model. Simulating the validation dataset from the Cardiff model (10000 observations, Figure 10, left) took roughly one hour. The metamodel took less than five seconds to simulate and graph the 10000 observations on the left. The metamodel thus makes it possible to simulate outcomes for arbitrary input values in a fraction of the time it takes the original model. Of course this comes at the expense of being limited to the inputs that the surrogate has been trained on, while the original model allows for adjustment of all input factors.

The further simulations in section 5.3.1 took between one and five minutes each. The simulations are of the same type as the one illustrated in Figure 10, but then applied iteratively over the defined grid of different input values. This means that for Figure 12, 10000 cost-effectiveness pairs were simulated 1000 times, resulting in a total of 10 million simulated pairs.

6. Discussion

After presenting the insights that can be produced through the use of a statistical metamodel, I would like to critically discuss how these results can be used, as well as the trade-off between the additional value and additional workload associated with the creation of a metamodel.

6.1. Applications

6.1.1. Incorporation of price-changes into health economic evaluations

In section 5.3.2 I forecasted the development of dapagliflozin's incremental costs compared to SU as the price decreases over time. According to the estimates, dapagliflozin is cost-effective with a very high probability, even if the price were much higher that it is now. Price drops therefore hardly affect the probability of cost-effectiveness. I thus focused on the probability that the treatment is cost-saving. An equivalent analysis can be performed for cost-effectiveness (it would just not be very interesting in the context of the case study).

Such a forecast could be included in health technology appraisal submissions and their evaluation by the respective authorities. A drug might not be cost-effective or cost-saving yet, but may be in the near future with a high probability. In that case the drug should be reimbursed right away, so that when it becomes cost-effective, it is already established on the market. Not incorporating a forecast of this sort will lead to a time-lag between market changes and efficient reimbursement: a formerly not cost-effective drug may become cost-effective, but is only reimbursed years later or not at all, because the decision-maker's policy is based on old, outdated cost-effectiveness evaluation.

An obvious implication of such a procedure is that drugs which are closer to patent expiry have a higher chance of reimbursement. The incentives in pharmaceutical lifecycle management would thus be affected: under usual circumstances, a manufacturer wants to introduce a drug as quickly as possible, in order to sell the drug under patent protection [50]. However, for drugs whose cost-effectiveness is doubtful, the manufacturer might delay the market entry on purpose, in order to have a stronger claim on cost-effectiveness in the future (as the loss of exclusivity and anticipated price drop will be closer).

We see that the incorporation of pricing forecasts has potential benefits, but might also lead to perverse incentives and manufacturers withholding viable drugs from the market. How this will play out in reality cannot be answered from this study alone, and further research is necessary.

6.1.2. Incorporation of new clinical studies

Compared to drug prices, the incorporation of clinical data is not as straightforward: any published study will provide information not only about a specific input that is included in the metamodel (e.g. weight change), but will also potentially differ in patient population and other factors. Selectively taking certain inputs from a new study while ignoring those which the metamodel has not been trained on will potentially lead to biased results.

Thus, it is vital that the new study results are incorporated into the original model and (as far as possible) into the metamodel and then compared (i.e. the metamodel has to be validated again, see section 5.2). If the metamodel continues to make accurate predictions, we can continue to use it and analyze the impact of the new clinical data. However, if the predictions of the two models are different, the metamodel has to be trained on new simulated data by the original model. The number and constellation of inputs do not necessarily have to change.

We see that the metamodel can never completely replace the original model: the latter is needed to validate outputs of the former and produce new training data if necessary. The more different inputs are incorporated into the metamodel, the more flexible the model becomes, and thus the probability that will be able to usefully incorporate future information increases. However, increasing the number of inputs is not trivial; see section 6.2.2 for further discussion.

Note that this section was strictly concerned with whether the metamodel accurately approximates the estimates of the original model when incorporating new evidence. Whether these estimates make sense or not is a different question, which is discussed in section 6.2.1.

6.1.3. Value-based payments

In the introduction, I alluded to the possibility of applying the estimates of a metamodel in the design of value-based payments and risk-sharing agreements. Out of the factors I investigated, the price of dapagliflozin was by far the most impactful, while the one clinical parameter and the prices of other drugs involved turned out to not have a strong association with incremental costs or effects. Thus I will focus on financial-based risk-sharing agreements, and how the price of dapagliflozin can be used in the context of these.

Using the analysis shown in Figure 12, we can quantify the expected distribution of costs per patient at any price. If a payer is trying to contain costs, they might be hesitant to have a large number of patients switch from SU to dapagliflozin, even though the latter is more effective and leads to a higher quality of life. Payer and manufacturer could then agree on a price at which there is an 80% chance that the incremental costs are lower than 2000€. If the costs are higher than that, the manufacturer pays the exceeding amount. Such an agreement allows the payer to contain costs, and the manufacturer will still generate profit in the long run, assuming that the profit margin of dapagliflozin is high enough to cover for the cases where costs exceed the 2000€ threshold.

There are two main challenges to this kind of approach: first off, incremental costs cannot be directly measured in the real world, since patients either receive dapagliflozin or SU, and there is no reliable way to estimate what the costs would have been if the patient had received the other medicine instead. The costs instead have to be averaged out over the entire Dutch patient population. I am assuming that health insurers have access to databases which make this possible in principle. The method only works if in the future both dapagliflozin and SU are reimbursed at a large scale, as this is a precondition to calculate incremental costs between the two groups. Instead of using a threshold for incremental costs, it would in theory also be possible to use a threshold for absolute costs that dapagliflozin patients cannot exceed. This would make it unnecessary to track SU patients. However, this approach brings its own challenges, since the market as a whole changes over time. Thus the absolute cost distribution of dapagliflozin patients will likely be different in five years from what it is today. The incremental costs (after adjusting for the prices of dapagliflozin and SU themselves) can be assumed to stay roughly constant, since most market changes will likely affect both treatment lines in a similar manner.

Secondly, the costs estimated by the Cardiff model (and therefore also the metamodel) are based on a lifetime horizon. Any threshold that could be applied to these costs will also have to be based on a lifetime horizon (i.e. the average costs of a patient over the entire treatment duration). If the costs for any patient exceed the threshold, it will likely only happen after many years. As a consequence, the threshold mechanism provides little protection against high short-term costs: a patient might experience many expensive complications early on; the expenses will be much higher than average, but not high enough to reach the threshold right away. The payer thus has to bear the costs and will only get them paid back in the future by the manufacturer.

A possible solution to this second challenge is to not use a lifetime horizon when simulating training data from the original model, but instead set a short time horizon, like five years. This would allow better use of the results for financial value-based agreements, since spending caps or rebates can be better set for a shorter timeframe. On the other hand, many medicines are valuable because of their long-term effects and only become cost-effective when using a lifetime horizon. NICE and other health technology assessment agencies would likely not accept studies based on such short time-horizons (at least not by themselves), since this contradicts the requirements for cost-effectiveness studies [51].

Of course one could also train two versions of the metamodel, one with a short-term and one with a lifetime horizon. However, since the generation of training data is inherently problematic with many health economic models (see section 6.2.2), this will often be unfeasible.

6.2. Limitations and unresolved issues

6.2.1. Manipulation of clinical parameters

The estimates of any model are only as valid as the data it was trained on. Therefore, any insights gained from using a metamodel are conditional on the assumption that the original model produces valid outputs. If the original model produces unrealistic estimates (at least partly depending on the precise inputs used for producing the training data), the metamodel will replicate these errors. Worse, since the metamodel omits most of the more detailed outputs included in the original model (e.g. rates of adverse events, in the case of this study), obviously false estimates are more difficult to identify.

To illustrate this, I would like to critically discuss the results in section 5.3.1.2. It turned out that modifying the bodyweight change parameter of dapagliflozin had a comparably small impact on the incremental QALYs in the dapagliflozin-SU comparison. This is surprising; the Cardiff model includes a sophisticated mechanism to simulate the development of the patients' bodyweight and its impact on the risk for heart failures (see e.g. McEwan et al. for an explanation [30]). According to the DECLARE study, which the risk equations in the current version of the Cardiff model are based on, dapagliflozin is directly associated with lower heart failure rates. It also decreases patients' bodyweight, which indirectly leads to lower rates of heart failure in the long-term. However, the latter effect is actually very small.

The incremental QALYs predicted by the Cardiff model are driven primarily by differences in the number of hypoglycemic episodes, utility associated with the weight loss and temporarily lower weight of patients in the dapagliflozin group, and finally differences in the occurrence of heart failures *unrelated* to the patients' bodyweight.

This was confirmed by a manual sensitivity analysis, using the Cardiff model instead of the metamodel (results in Table 13). Even when setting the weight change associated with dapagliflozin to 0 (i.e. it does not lead to weight loss at all), the difference in the incremental number of heart failures is minimal compared to the base scenario. The miniscule effect of the weight increment on rates of heart failure can potentially be explained by the fact that the weight difference between dapagliflozin and SU patients is only temporary and patients gain the lost weight back quickly, once they switch to insulin treatment.

Table 13: results of the manual sensitivity analysis. Incremental incidence is the number of events in the dapagliflozin group minus the number of events in the SU group

Dapagliflozin weight change	**SU weight change**	**Incremental QALYs**	**Incremental incidence of heart failures (events per cohort of 1000 patients)**
-3.3	1.36	0.61	-8.1
0	1.36	0.50	-7.3

If accurate, these findings would contradict earlier research claiming that the weight loss associated with dapagliflozin is a major driver of its effect on patients' quality of life [30, 31]. Unfortunately, there is no way to validate the findings, since there is of course no clinical evidence for a situation where dapagliflozin does not lead to weight loss, but has otherwise identical effects to those that are

currently known. This uncovers a potential flaw in the methodology of this study: when the original model is passed input parameters which are far from the available clinical evidence, it might produce unrealistic estimates, which the metamodel will reproduce. Usually there is no way to determine how valid the results are, apart from expert judgement.

The problem is not specific to metamodeling per se: the metamodel will still accurately estimate the behavior of the original model, and as such serve its intended purpose. Rather, the issue is whether applying inputs without clinical evidence to the original model is a sensible thing to do.

6.2.2. Time investment

I previously emphasized the massive decrease in computation time needed to simulate results from the metamodel compared to a PSA in a conventional health economic model (section 5.4). While the creation of the metamodel itself took many hours, the vast majority of this time was spend on research and development of a suitable metamodeling method. The coding and execution of the metamodel in R or another programming language is fairly simple and straightforward. A reasonably experienced R programmer with an established methodology to build on could probably program a functional model within a day. For reference, the entire R code used in this thesis is around 700 lines long (including the Stan code, which is integrated into the R code as a character string).

Far more problematic is the generation of representative training data from the original model, which will most often be coded in spreadsheet software like Microsoft Excel. The drawbacks of using Excel for health technology assessment have been discussed elsewhere [52], and I will focus only on one particular aspect: spreadsheet models make it difficult to automatically run the model with varying input values and store the results. In the case of this study, running 40 experiments and documenting the results took roughly four hours and involved lots of manual, error prone work. This effectively limits the number of inputs that can be incorporated into the metamodel, since each additional parameter significantly increases the amount of manual work involved.

If the model were coded in a proper programming language, this would be far less of a problem. In that case, it is easy to write a loop which repeatedly runs the model with randomized inputs and stores the results. There is then no real limit to the number of inputs used: the computation time will increase with each input (in the case of this study, the time which the Stan model takes to produce an adequate number of parameter samples); however, this is less of a problem, since this process requires no supervision by the user.

Unfortunately, the use of modern programming languages for health economic modelling is not widespread yet. The problem of having to work with spreadsheet models will persist for the time being. I see this as the major limitation of my methodology: with the time-consuming and error-prone work involved in producing adequate training data, the creation of a metamodel may not be worth the effort in many otherwise promising use-cases.

6.2.3. Timeline

The timeline analysis in section 5.3.2 models the incremental costs of dapagliflozin treatment compared to SU treatment at different points time. Specifically, it models the cost-effectiveness evaluation at each point in time (i.e. what the Cardiff model would put out, given that exact scenario) (Figure 16) and the probability that the treatment will be cost-saving or not (Figure 17). This is *not* the same as the lifetime costs of a patient starting at a particular time, since the simulation assumes that once a patient started treatment, the price of dapagliflozin stays constant. For instance, if a patient starts treatment in 2027, the price remains fixed at € 517.44, even though it actually falls

shortly after. This means that the lifetime costs for that patient would actually be slightly lower. The modelled estimates are not “wrong”, but they answer a different question than one might expect.

The actual lifetime costs of a patient are arguably more relevant in the context of value-based payments. However, the current version of the Cardiff model does not allow to calculate this. It would require to pass an entire timeline of prices to the model (instead of just a single price) and then have patients start treatment at different points on that timeline. Implementing this requires modifying the VBA and C++ code of the model, which I was unable to do due to my limited proficiency in these languages. The calculated estimates would also be fundamentally different from and incomparable to ordinary cost-effectiveness evaluations. Nonetheless, such an analysis might produce very useful results especially in the context of value-based payments, and should be explored in future research.

6.2.4. (Un-)normality of the original model’s distribution

The approximation via a multivariate normal distribution assumes that the distribution of cost-effectiveness pairs simulated by the original model is approximately Gaussian, or normally distributed. A strong correlation between costs and effects is not problematic. However, if the distribution is multimodal[6] or otherwise irregularly shaped, the approximation will be inaccurate and produce potentially biased and misleading results.

Figure 18 provides a visual illustration: examples 1 and 2 are approximately Gaussian, therefore they can be approximated well through a multivariate normal distribution. Example 3 has an irregular and asymmetric shape; it is not well approximated with a normal distribution. Example 4 has two separate modes around which the cost-effectiveness pairs cluster. It is not well approximated either.

Note that these are extreme examples - in many cases a normal approximation will still provide useful results, even when the target distribution has an irregular shape. As discussed earlier, the goal of metamodeling is not to reproduce the original model, but to approximate it well enough to accurately answer all questions of interest. Proper model validation is therefore crucial.

[6] A multimodal distribution has more than one maximum.

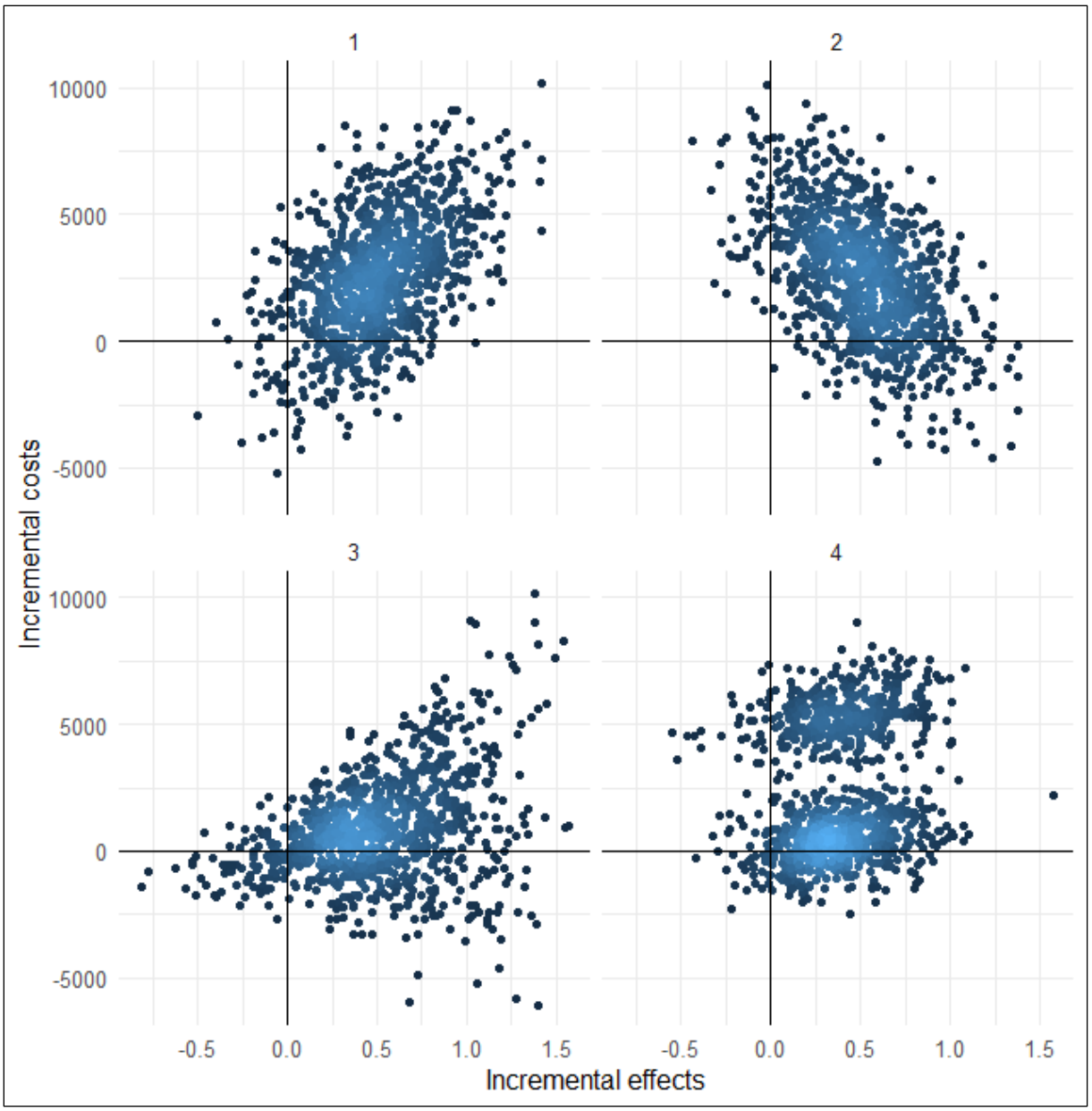


Figure 18: different shapes that the distribution of simulated cost-effectiveness pairs could take. Simulated data, not connected to the case study.

6.3. Conclusion and recommendations

I hope that my research report demonstrated the functionality and validity of applying the metamodel technique to health economic simulation models. The implementation is relatively straightforward, and the results add a valuable perspective to the cost-effectiveness of dapagliflozin treatment in the Netherlands. I can give an unconditional recommendation to use metamodeling for estimating the consequences of price changes, as long as careful model validation is included in the analysis. For changes in clinical effectiveness, metamodeling can in principle also be used, but the results have to be critically examined on a case-by-case basis. Even if the metamodel accurately approximates the estimates of the original model, the estimates should be evaluated in the light of available clinical evidence and not be taken at face-value.

The main practical challenge is the generation of suitable training data for the metamodel. As indicated earlier, this is far less of a problem when the original model is coded in a proper

programming language instead of a spreadsheet. I thus support calls for the widespread adoption of modern programming languages in health technology assessment [52]. The process could be sped up even further if there was some kind of standardized framework regarding the construction of health economic models in R or Python. For ideas of what such a framework could look like, I refer the reader to a series of publications by Alarid-Escudero and Krijkamp et al, providing general coding guidelines [53] as well as specific tutorials for cohort-state transition models [54] and patient microsimulation models [55].

For the case study at hand, the metamodel demonstrated a good predictive accuracy despite the assumption that the predictors have linear effect on the outcomes. This will not always be the case. Gaussian process regression is a candidate for a more general metamodeling method which can approximate arbitrary nonlinear effects. Many authors (e.g. Gramacy [56]) have written about the use of Gaussian processes for metamodeling of simulation experiments, and I see no reason why such methods could not be adapted to health economic models. It has to be noted, however, that Gaussians processes are more difficult than ordinary regression methods. Learning to use them requires dedicated research and practice. An accessible introduction to Gaussian processes with applications in health economics would thus be a valuable addition to the literature.

7. Appendix

7.1.1. Cardiff model inputs

7.1.1.1. Health economic assumptions

- Model: Cardiff model using UKPDS 68 risk equations and DECLARE data.
- Cohort size for simulations: 250 for training data, 10000 for test data, 1000 for manual sensitivity analysis (Table 13)
- Time horizon: lifetime (implemented as 40 years).
- Cycle length: 6 months.
- Perspective: societal
- Discount rates: 4% for costs, 1.5% for effects.

7.1.1.2. Patient population

- T2DM patients with insufficient glycemic control on metformin alone. The model cohort was considered representative of Dutch patients who would be eligible to receive dapagliflozin added to metformin. *Table* 14 shows the precise inputs.
- In case SE was not reported and could not be calculated, it was assumed to be 1% of the mean.

Table 14: Cardiff model inputs, patient characteristics and clinical risk factors

Demographics			**Source**
Age	58.40	0.32	Nauck et al. 2014[57, 58]
Proportion female (%)	0.448	0.02	Nauck et al. 2014[57, 58]
Diabetes duration (Years)	6.32	0.19	Nauck et al. 2014[57, 58]
Height (m)	1.67	0.0167	Nauck et al. 2014[57, 58], SE assumed
Proportion Afro-Caribbean (%)	0.062	0.00062	Nauck et al. 2014[57, 58], SE assumed
Proportion Indian (%)	0.076	0.00076	Nauck et al. 2014[57, 58], SE assumed
Proportion smokers (%)	0.176	0.00176	Nauck et al. 2014[57, 58], SE assumed
Clinical risk factors			
HbA1c (%)	7.72	0.025	Nauck et al. 2014[57, 58]
Total-Cholesterol (mg/dL)	182.91	1.829	Nauck et al. 2011[58], SE assumed
HDL-Cholesterol (mg/dL)	46.02	0.46	Nauck et al. 2011[58], SE assumed
LDL-Cholesterol (mg/dL)	103.25	1.033	Nauck et al. 2011[58], SE assumed
SBP (mmHg)	133.30	1.333	Nauck et al. 2011[58], SE assumed
DBP (mmHg)	80.6	0.806	Nauck et al. 2011[58], SE assumed
Weight (Kg)	88.02	0.88	Nauck et al. 2014[57, 58], SE assumed
eGFR (ml/min/1.73^2)	89.95	0.9	Nauck et al. 2014[57, 58], SE assumed
Haemoglobin (g/dl)	13.71	0.137	Hayes et al 2013[59], SE assumed
Albuminuria (mg/l)	47.00	0.47	Hayes et al 2013[59], SE assumed
White blood cell count (10^6)	6.80	0.068	Hayes et al 2013[59], SE assumed
Heart rate (bpm)	72.00	0.72	Hayes et al 2013[59]

Glucose variability	37.26	0.19	DECLARE baseline[38, 60]
Haematocrit (%)	41.12	0.17	Nauck et al. 2011[58]
Ischemic heart disease	0.09375	0.009	Nauck et al. 2011[58] based on % with prior history of CVD, distributed equally over ischemic heart disease and myocardial infarction (assumption)
Myocardial infarction	0.09375	0.009	Nauck et al. 2011[58] based on % with prior history of CVD, distributed equally over ischemic heart disease and myocardial infarction (assumption)
Retinopathy	0.058	0.0058	Nauck et al. 2011[58]
All other history	0	0	Assumption
All medication use	0	0	Assumption

Abbreviations: HDL, high-density lipoprotein; kg, kilogram; LDL, low-density lipoprotein; m, meter; mmHg, millimeter of mercury; mmol/L, millimole/liter; SBP, systolic blood pressure.

7.1.1.3. Treatment profiles

- HbA1c threshold for therapy change: 8% (assumed Dutch clinical practice).
- In case SE was not reported and could not be calculated, it was assumed to be 10% of the mean.

Table 15: Cardiff model inputs, treatment profiles

	1st line				**2nd line**		**3rd line**	
	Comparator: MET + SU Mean (SE)	**Source**	**Intervention: MET + dapagliflozin Mean (SE)**	**Source**	**Intervention 1: MET + insulin Mean (SE)**	**Source**	**MET + Insulin + Bolus Mean (SE)**	**Source**
ΔHbA1c in year 1 (%)	-0.67 (0.0638)	Barnett et al. 2016[61]	-0.67 (0.1582)	Barnett et al. 2016[61]	-0.13 (0.5230)	Lozano-Ortega et al. 2016[62]	-1.68 (0.168)	Hollander et al. (2008)[63]
HbA1c months benefit in year 1	12.00 (0.00)	Assumption	12.00 (0.00)	Assumption	12.00 (0.00)	Assumption	12.00 (0.00)	Assumption
ΔSBP (mmHg)	0.90 (0.09)	Nauck et al. 2014[57]	-3.80 (0.38)	Nauck et al. 2014[57]	1.31 (2.1990)	Assumed equal to MET + SU + insulin, Lozano-Ortega et al. 2016[62]	0.00	Hollander et al. (2008)[63]
ΔDBP (mmHg)	-0.40 (0.04)	Nauck et al. 2011[58]	-1.60 (0.02)	Nauck et al. 2011[58]	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption
ΔTotal Cholesterol (mg/dL)	-0.50 (0.05)	Nauck et al. 2011[58]	1.27 (0.13)	Nauck et al. 2011[58]	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption
ΔHDL Cholesterol (mg/dL)	-0.034 (0.003)	Nauck et al. 2011[58]	1.26 (0.13)	Nauck et al. 2011[58]	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption

ΔWeight (kg)	1.36 (0.09)	Barnett et al. 2016[61]	-3.30 (0.90)	Barnett et al. 2016[61]	3.20 (1.8546)	Lozano-Ortega et al. 2016[62]	4.20 (0.42)	Hollander et al. (2008)[63]
Years of maintained weight loss	2.00 (0.00)	Nauck et al. 2014[57], Charokopou et al. 2015[64]	2.00 (0.00)	Nauck et al. 2014[57], Charokopou et al. 2015[64]	1.00 (0.00)	Assumption	1.00 (0.00)	Assumption
Natural annual weight gain (kg)	0.00 (0.00)	Nauck et al. 2014[57]	0.00 (0.00)	Nauck et al. 2014[57]	0.1 (0.01)	Assumption	0.1 (0.01)	Assumption
Years to loss of weight effect	0.00 (0.00)	Assumption	1.00 (0.00)	Assumption	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption
Glucose variability change	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption
Haematocrit change	0.390 (0.130)	Nauck et al. 2011[58]	2.860 (0.140)	Nauck et al. 2011[58]	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption
Annual number of symptomatic hypo	0.408 (0.082)	Charokopou et al. 2015 [30]	0.035 (0.006)	Charokopou et al. 2015 [30]	0.011 (0.0022)	Charokopou et al. 2015 [30]	0.616 (0.123)	Charokopou et al. 2015 [30]
Annual number of severe hypo	0.007 (0.001)	Charokopou et al. 2015 [30]	0 (0)	Charokopou et al. 2015 [30]	0.037 (0.0074)	Charokopou et al. 2015 [30]	0.022 (0.004)	Charokopou et al. 2015 [30]
Annual probability of UTI	0.064 (0.0064)	Nauck et al. 2014[57]	0.108 (0.0108)	Nauck et al. 2014[57]	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption
Annual probability of GI	0.027 (0.0027)	Nauck et al. 2014[57]	0.123 (0.0123)	Nauck et al. 2014[57]	0.00 (0.00)	Assumption	0.00 (0.00)	Assumption
Annual probability of discontinuation	0.0515 (0.00515)	Nauck et al. 2014[57]	0.00246 (0.000246)	Nauck et al. 2014[57]	0.042 (0.0042)	Assumed equal to MET+SU+insulin. Holman et al. 2007[65]	0.00 (0.00)	Assumption

Abbreviations: BMI, body mass index; Dapa, dapagliflozin; DPP-4i, dipeptidyl peptidase-4 inhibitors; GI, genital infection; HbA1c, glycated hemoglobin; HDL, high-density lipoprotein; hypo, hypoglycemia; kg, kilogram; MET, metformin; mmHg, millimeter of mercury; mmol/L, millimole/liter; SBP, systolic blood pressure; SU, sulfonylurea; UTI, urinary tract infection; Y1, year 1; Y1>1, subsequent years; Y1≥1, year 1 and subsequent years.

7.1.1.4. Costs

Table 16: calculation of baseline SU price. The weighted mean is calculated by multiplying each price with the respective market share and then summing everything up.

	Costs per day	**Market share**	**Source**
Glibenclamide	€ 0.10	0.01	CIBG [36]
Tolbutamide	€ 0.11	0.16	CIBG [36]
Gliclazide (Diamicron)	€ 0.11	0.63	CIBG [36]
Glimepiride (Amaryl)	€ 0.02	0.20	CIBG [36]
Mean (weighted) SU	€ 0.09	1.00	

Table 17: Cardiff model inputs, non-treatment costs

Direct costs of disease and treatment-related events, 2019 price level		
Hospitalisation for angina	€2,700 in Y1, €666 in Y>1	Soekhlal et al. 2013[66]
Myocardial infarction (incl. Y1 cost of subsequent events)	€20,389 in Y1, €1,239 in Y>1	Greving et al. (2011)[67]
Myocardial infarction (fatal)	€18,146 in Y1	Greving et al. (2011)[67]
Congestive heart failure	€11,280 in Y≥1	Postmus et al. (2011)[68]
Congestive heart failure (fatal)‡	€10,039 in Y1	Assumption
Stroke (incl. Y1 cost of subsequent events)	€39,349 in Y1, €4,947 in Y>1	Baeten et al. (2010)[69]
Stroke (fatal)	€22,429 in Y1	Baeten et al. (2010)[69]
Amputation (incl. Y1 cost of subsequent events)	€17,261 in Y1, €647 in Y>1	Niessen et al. (2003)[70]
Amputation (fatal)§	€9,839 in Y1	Assumption
Blindness	€2,737 in Y≥1	Niessen et al. (2003)[70]
End-stage renal disease	€91,074 in Y≥1	De Vries et al. (2016)[71]
Ulcer	€2,383 in Y1	Redekop et al. 2003[72]
PCI	€18,583 in Y1, €1,519 in Y>1	Osnabrugge et al. 2015[73]
CABG	€22,166 in Y1, €1,278 in Y>1	Osnabrugge et al. 2015[73]
Non-coronary revascularisation	€2,561 in Y1	Spronk et al. 2008[74]
Severe hypo event	€522	De Groot et al. 2018[75]
Symptomatic hypo event	€3	De Groot et al. 2018[75]
UTI and GI	€46.50 from a health care perspective, €88.18 from a societal perspective	NHG (2013)[76] and Hakkaart-van Roijen et al. (2016)[77]
Discontinuation	€35.25 from a health care perspective, €76.92 from a societal perspective	Hakkaart-van Roijen et al. (2016)[77]
Annual indirect costs of disease and treatment-related events, 2019 price level¶		
Ischaemic heart disease	€0 in base case, €1,153 in scenario	Clarke et al. (2008)[78]
Myocardial infarction	€0 in base case, €8,699 in scenario	Isaaz et al. (2010)[79]
Congestive heart failure	€0 in base case, €8,699 in scenario	Ericson et al. (2011)[80]
Stroke	€0 in base case, €8,699 in scenario	Lindgren et al. (2008)[81]
Amputation	€0 in base case, €6,423 in scenario	Fisher et al. (2003)[82]
Blindness	€0 in base case, €8,699 in scenario	Frick et al. (2003)[83]
End-stage renal disease	€0 in base case, €8,699 in scenario	Naim et al. (2010)[84]

‡ Costs for fatal congestive heart failure are unknown. The ratio of costs for fatal compared with non-fatal congestive heart failure was assumed to be equal to the ratio of costs for fatal compared with non-fatal myocardial infarction.
§ Costs for fatal amputation are unknown. The ratio of costs for fatal compared with non-fatal amputation was assumed to be equal to the ratio of costs for fatal compared with non-fatal stroke.
¶ Productivity losses are applied below the age of 66.

7.1.1.5. Utilities

Table 18: Cardiff model inputs, utilities

Disutilities		
Diabetes-related event disutilities		
Ischaemic heart disease	0.042 (0.008)	Sullivan et al. (2016)[85]
Myocardial infarction	0.047 (0.005)	Sullivan et al. (2016)[85]
Congestive heart failure	0.050 (0.007)	Sullivan et al. (2016)[85]
Stroke	0.060 (0.007)	Sullivan et al. (2016)[85]
Amputation	0.095 (0.040)	Sullivan et al. (2016)[85]

Blindness	0.045 (0.009)	Sullivan et al. (2016)[85]
End-stage renal disease	0.038 (0.011)	Sullivan et al. (2016)[85]
Ulcer	0.042 (0.007)	Sullivan et al. (2016)[85]
PCI	0.042 (0.008)	Sullivan et al. (2016)[85], assumed equal to ischemic heart disease
CABG	0.042 (0.008)	Sullivan et al. (2016)[85], assumed equal to ischemic heart disease
Non-coronary revascularisation	0.045 (0.013)	Sullivan et al. (2016)[85]
Treatment-related event disutilities		
Symptomatic hypo	0.014 (0.0014)	Currie et al. (2006)[86]
Nocturnal hypo	0	Assumed
Severe hypo	0.047 (0.0047)	Currie et al. (2006)[86]
UTI	0.025 (0.0025)	Sullivan et al. (2016)[85]
GI	0.038 (0.0038)	Sullivan et al. (2016)[85]
BMI disutilities		
BMI (per unit increase)	0.017 (0.005)	Grandy et al. (2014)[87]
BMI (per unit decrease)	-0.047 (0.005)	Grandy et al. (2014)[87]

Abbreviations: BMI, body mass index; Dapa, dapagliflozin; DPP-4i, dipeptidyl peptidase-4 inhibitors; GI, genital infection; HbA1c, glycated hemoglobin; HDL, high-density lipoprotein; hypo, hypoglycemia; kg, kilogram; MET, metformin; mmHg, millimeter of mercury; mmol/L, millimole/liter; SBP, systolic blood pressure; SU, sulfonylurea; UTI, urinary tract infection; Y1, year 1; Y1>1, subsequent years; Y1≥1, year 1 and subsequent years.

7.1.2. Training data

Table 19: input parameters obtained through LHS sampling

Index	weighteffect	met	dapa	SU	met_dapa	met_SU	met_ins	met_ins_bolus
1	-5.48	24.69	412.82	25.2	493.51	105.89	365.59	1220.27
2	-3.91	28.12	599.36	7.19	683.48	91.31	369.02	1223.7
3	-1.76	13.39	155.41	12.4	224.8	81.79	354.29	1208.97
4	-3.79	11.27	763.65	24.45	830.92	91.72	352.17	1206.85
5	-4.28	6.49	491.94	3.25	554.43	65.74	347.39	1202.07
6	-0.6	20.75	751.89	11.77	828.64	88.52	361.65	1216.33
7	-4.06	30.64	374.06	22.65	460.7	109.29	371.54	1226.22
8	-1.32	3.94	844.87	4.89	904.81	64.83	344.84	1199.52
9	-2.01	23.77	668.57	21.62	748.34	101.39	364.67	1219.35
10	-5.38	16.49	630.04	13.47	702.53	85.96	357.39	1212.07
11	-2.24	34.42	177.96	26.85	268.38	117.27	375.32	1230
12	-6.04	31.81	316.66	15.95	404.47	103.76	372.71	1227.39
13	-2.72	21.11	290.53	7.74	367.64	84.85	362.01	1216.69
14	-1.57	12.41	700.81	31.65	769.22	100.06	353.31	1207.99
15	-4.35	4.88	221.97	26.45	282.85	87.33	345.78	1200.46
16	-1.43	35.81	822.72	8.76	914.53	100.57	376.71	1231.39
17	-2.51	5.76	334.43	36.38	396.19	98.14	346.66	1201.34
18	-3.6	29.34	437.06	32.57	522.4	117.91	370.24	1224.92
19	-2.33	10.51	76.69	19.39	143.2	85.9	351.41	1206.09
20	-4.5	9.86	557.48	30.19	623.34	96.05	350.76	1205.44
21	-5.75	19.66	187.61	35.1	263.27	110.76	360.56	1215.24
22	-0.09	18.12	45.7	28.69	119.82	102.81	359.02	1213.7
23	-2.86	3.24	352.26	6.12	411.5	65.36	344.14	1198.82
24	-1.85	16.03	548.81	20.47	620.84	92.5	356.93	1211.61

25	-4.86	18.56	125.1	18.39	199.66	92.95	359.46	1214.14
26	-0.37	35.36	525.85	16.47	617.21	107.83	376.26	1230.94
27	-5.25	23.02	652.96	28.34	731.98	107.36	363.92	1218.6
28	-3.38	26.38	588.41	9.96	670.79	92.34	367.28	1221.96
29	-6.26	0.4	461.27	17.54	517.67	73.94	341.3	1195.98
30	-0.69	15.07	245.47	0.11	316.54	71.18	355.97	1210.65
31	-3.14	14.32	57.16	37.42	127.48	107.74	355.22	1209.9
32	-4.72	8.87	873.8	11.23	938.67	76.1	349.77	1204.45
33	-1.14	26.52	792.12	30.73	874.64	113.25	367.42	1222.1
34	-6.37	33.1	253.66	33.44	342.76	122.54	374	1228.68
35	-5.88	7.75	98.13	4.09	161.88	67.84	348.65	1203.33
36	-6.49	28.55	735.07	1.39	819.62	85.94	369.45	1224.13
37	-0.32	2.69	451.66	22.97	510.35	81.66	343.59	1198.27
38	-0.93	22.01	905.15	2.66	983.16	80.67	362.91	1217.59
39	-4.99	1.17	888.5	35.47	945.67	92.64	342.07	1196.75
40	-3.03	32.52	11.89	14.77	100.41	103.29	373.42	1228.1

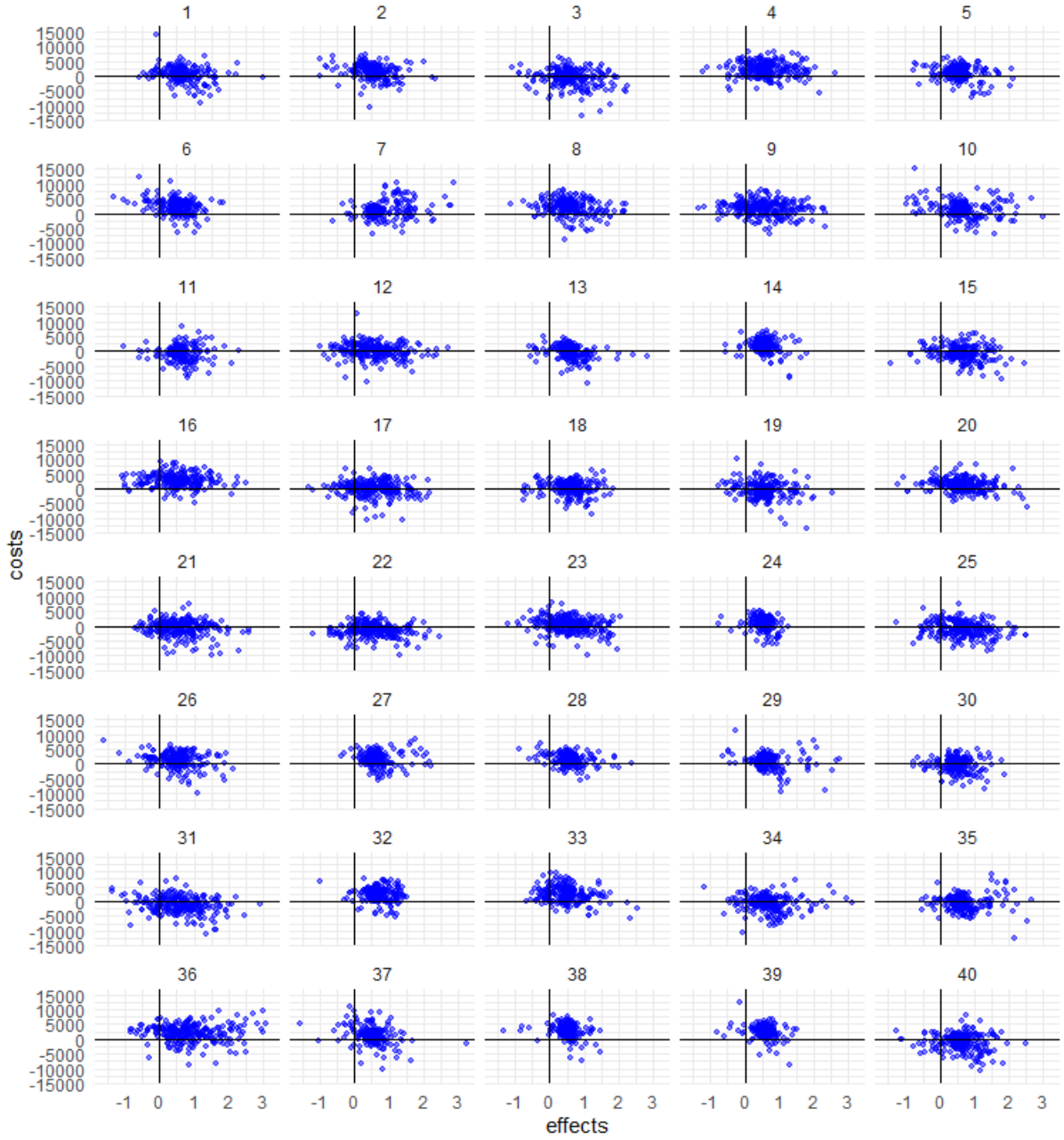


Figure 19: visualization of training data obtained from the Cardiff model. Each of the 40 sets was obtained by putting the respective inputs from Table 19 into the Cardiff model and drawing 250 samples from the Cardiff model

7.1.3. References

1. Chatterjee, A., et al. *Innovative pharma contracts: When do value-based arrangements work?* 2017; Available from: www.mckinsey.com/industries/pharmaceuticals-and-medical-products/our-insights/innovative-pharma-contracts-when-do-value-based-arrangements-work.
2. Neumann, P.J., et al., *Future directions for cost-effectiveness analyses in health and medicine.* 2018. **38**(7): p. 767-777.
3. Betancourt, M. *Probabilistic Modeling and Statistical Inference*. 2019; Available from: betanalpha.github.io/assets/case_studies/modeling_and_inference.html.
4. Betancourt, M. *Markov chain Monte Carlo*. 2020; Available from: github.com/betanalpha/knitr_case_studies/tree/master/markov_chain_monte_carlo.

5. Betancourt, M. *An Introduction to Stan*. 2020; Available from: betanalpha.github.io/assets/case_studies/stan_intro.
6. Hoffman, M.D. and A.J.J.M.L.R. Gelman, *The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo.* 2014. **15**(1): p. 1593-1623.
7. McElreath, R., *Statistical rethinking: A Bayesian course with examples in R and Stan*. 2020: CRC press.
8. Greenland, S., et al., *Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations.* 2016. **31**(4): p. 337-350.
9. Baio, G., *Bayesian methods in health economics*. 2012: CRC Press.
10. Briggs, A., M. Sculpher, and K. Claxton, *Decision modelling for health economic evaluation*. 2006: Oup Oxford.
11. Ge, H., K. Xu, and Z. Ghahramani, *Turing: A Language for Flexible Probabilistic Inference*, in *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, S. Amos and P.-C. Fernando, Editors. 2018, PMLR: Proceedings of Machine Learning Research. p. 1682--1690.
12. Degeling, K., et al., *A scoping review of metamodeling applications and opportunities for advanced health economic analyses.* 2019. **19**(2): p. 181-187.
13. Jalal, H., et al., *Linear regression metamodeling as a tool to summarize and present simulation model results.* 2013. **33**(7): p. 880-890.
14. Degeling, K., et al., *Introduction to Metamodeling for Reducing Computational Burden of Advanced Analyses with Health Economic Models: A Structured Overview of Metamodeling Methods in a 6-Step Application Process.* 2020. **40**(3): p. 348-363.
15. Rojnik, K. and K.J.V.i.h. Naveršnik, *Gaussian process metamodeling in Bayesian value of information analysis: a case of the complex health economic model for breast cancer screening.* 2008. **11**(2): p. 240-250.
16. Jalal, H. and F.J.M.D.M. Alarid-Escudero, *A Gaussian approximation approach for value of information analysis.* 2018. **38**(2): p. 174-188.
17. Alam, M.F. and A. Briggs, *Artificial neural network metamodel for sensitivity analysis in a total hip replacement health economic model.* 2020. **20**(6): p. 629-640.
18. Jalal, H. and F. Alarid-Escudero, *BayCANN: Streamlining Bayesian Calibration with Artificial Neural Network Metamodeling.* 2020.
19. Lipton, Z.C.J.Q., *The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery.* 2018. **16**(3): p. 31-57.
20. Stevenson, M., J. Oakley, and J.J.M.D.M. Chilcott, *Gaussian process modeling in conjunction with individual patient simulation modeling: a case study describing the calculation of cost-effectiveness ratios for the treatment of established osteoporosis.* 2004. **24**(1): p. 89-100.
21. Volksgezondheidenzorg. *Diabetes mellitus → Cijfers & Context → Huidige situatie*. 2020 [cited 2021 15-04]; Available from: https://www.volksgezondheidenzorg.info/onderwerp/diabetes-mellitus/cijfers-context/huidige-situatie.
22. Volksgezondheidenzorg. *Diabetes mellitus → Kosten → Zorguitgaven*. 2020 [cited 2021 15-04]; Available from: https://www.volksgezondheidenzorg.info/onderwerp/diabetes-mellitus/cijfers-context/huidige-situatie.
23. Guariguata, L., et al., *Global estimates of diabetes prevalence for 2013 and projections for 2035.* 2014. **103**(2): p. 137-149.
24. Eurpean Medicines Agency. *Forxiga*. 2020; Available from: www.ema.europa.eu/en/medicines/human/EPAR/forxiga.
25. Zorginstituut Nederland. *Regeling zorgverzekering*. 2021; Available from: https://www.farmacotherapeutischkompas.nl/algemeen/regeling-zorgverzekering.
26. Nauck, M.A., et al., *Dapagliflozin versus glipizide as add-on therapy in patients with type 2 diabetes who have inadequate glycemic control with metformin: a randomized, 52-week, double-blind, active-controlled noninferiority trial.* 2011. **34**(9): p. 2015-2022.

27. Rijksoverheid. *Forse besparing door lagere maximumprijzen geneesmiddelen*. 2019; Available from: www.rijksoverheid.nl/actueel/nieuws/2019/12/19/forse-besparing-door-lagere-maximumprijzen-geneesmiddelen.
28. Register, E.P. *EP2069374 - CRYSTALLINE SOLVATES OF (1S)-1,5-ANHYDRO-1-C-(3-((PHENYL) METHYL) PHENYL)-D-GLUCITOL DERIVATIVES WITH ALCOHOLS AS SGLT2 INHIBITORS FOR THE TREATMENT OF DIABETES* 2020; Available from: https://register.epo.org/application?number=EP07784499&tab=main.
29. *European Patent Convention Art. 63 (1)*. 2020.
30. Charokopou, M., et al., *Cost-effectiveness of dapagliflozin versus DPP-4 inhibitors as an add-on to metformin in the treatment of type 2 diabetes mellitus from a UK healthcare system perspective.* 2015. **15**(1): p. 1-11.
31. Sabale, U., et al., *Cost-effectiveness of dapagliflozin (Forxiga®) added to metformin compared with sulfonylurea added to metformin in type 2 diabetes in the Nordic countries.* 2015. **9**(1): p. 39-47.
32. Clarke, P., et al., *The impact of diabetes-related complications on healthcare costs: results from the United Kingdom Prospective Diabetes Study (UKPDS Study No. 65).* 2003. **20**(6): p. 442-450.
33. Wiviott, S.D., et al., *Dapagliflozin and cardiovascular outcomes in type 2 diabetes.* 2019. **380**(4): p. 347-357.
34. McEwan, P., et al., *The cost-effectiveness of dapagliflozin in treating high risk patients with type 2 diabetes mellitus: an economic evaluation using data from the DECLARE-TIMI 58 trial.* 2020.
35. McEwan, P., et al., *Evaluation of the costs and outcomes from changes in risk factors in type 2 diabetes using the Cardiff stochastic simulation cost-utility model (DiabForecaster).* 2006. **22**(1): p. 121-129.
36. CIBG. *Wet geneesmiddelenprijzen*. 2020 [cited 2021 25-01]; Available from: www.farmatec.nl/prijsvorming/wet-geneesmiddelenprijzen.
37. Nederland, Z. *Wat is het uitgiftetarief?* 2021.
38. HEOR, *Cardiff T2DM Model with integrated DECLARE risk equations - Technical Report.* 2019.
39. Paisley, A.J., et al., *Dapagliflozin: a review on efficacy, clinical effectiveness and safety.* 2013. **22**(1): p. 131-140.
40. Barnett, A., et al., *Systematic review and network meta-analysis to compare dapagliflozin with other diabetes medications in combination with metformin for adults with type 2 diabetes.* 2014. **6**: p. 006.
41. Team, S.D. *RStan: the R interface to Stan*. 2020 2020.
42. guide, S.U.s. *22.12 Standardizing predictors and outputs*. 2021 [cited 2021 19-04]; Available from: mc-stan.org/docs/2_26/stan-users-guide/standardizing-predictors-and-outputs.html.
43. Carnell, R. *lhs: Latin Hypercube Samples*. 2020; Available from: https://bertcarnell.github.io/lhs/.
44. Lewandowski, D., D. Kurowicka, and H.J.J.o.m.a. Joe, *Generating random correlation matrices based on vines and extended onion method.* 2009. **100**(9): p. 1989-2001.
45. Agency, T.N.M. *Price and reimbursement list*. 2021; Available from: https://legemiddelverket.no/english/public-funding-and-pricing/maximum-price#list-of-products-with-maximum-prices.
46. Service, N.H. *Dictionary of medicines and devices (dm+d)*. 2021; Available from: https://www.nhsbsa.nhs.uk/pharmacies-gp-practices-and-appliance-contractors/dictionary-medicines-and-devices-dmd.
47. invaliditeitsverzekering, R.v.z.-e. *Farmaceutische specialiteiten - Lijsten van prijzen en vergoedingsbasis*. 2021; Available from: https://www.inami.fgov.be/nl/themas/kost-terugbetaling/door-ziekenfonds/geneesmiddel-gezondheidsproduct/terugbetalen/specialiteiten/Paginas/farma-specialiteiten-lijst-prijs-basis.aspx.

48. Vondeling, G.T., et al., *The impact of patent expiry on drug prices: a systematic literature review.* 2018. **16**(5): p. 653-660.
49. Fischer, J., et al., *Analogue-based drug discovery*. 2010: Wiley-VCH Mörlenbach, Germany.
50. Hering, S., et al., *Can lifecycle management safeguard innovation in the pharmaceutical industry?* 2018. **23**(12): p. 1962-1973.
51. NICE. *Guide to the Methods of Technology Appraisal 2013* 2013; Available from: https://www.nice.org.uk/process/pmg9.
52. Incerti, D., et al., *R you still using Excel? the advantages of modern software tools for health technology assessment.* 2019. **22**(5): p. 575-579.
53. Alarid-Escudero, F., et al., *A need for change! A coding framework for improving transparency in decision modeling.* 2019. **37**(11): p. 1329-1339.
54. Alarid-Escudero, F., et al., *Cohort State-Transition Models in R: A Tutorial.* 2020.
55. Krijkamp, E.M., et al., *Microsimulation modeling for health decision sciences using R: a tutorial.* 2018. **38**(3): p. 400-422.
56. Gramacy, R.B., *Surrogates: Gaussian process modeling, design, and optimization for the applied sciences*. 2020: Chapman and Hall/CRC.
57. Nauck, M., et al., *Durability of glycaemic efficacy over 2 years with dapagliflozin versus glipizide as add-on therapies in patients whose type 2 diabetes mellitus is inadequately controlled with metformin.* Diabetes, Obesity and Metabolism, 2014. **16**(11): p. 1111-1120.
58. Nauck, M.A., et al., *Dapagliflozin versus glipizide as add-on therapy in patients with type 2 diabetes who have inadequate glycemic control with metformin: a randomized, 52-week, double-blind, active-controlled noninferiority trial.* Diabetes care, 2011. **34**(9): p. 2015-2022.
59. Hayes, A., et al., *UKPDS outcomes model 2: a new version of a model to simulate lifetime health outcomes of patients with type 2 diabetes mellitus using data from the 30 year United Kingdom Prospective Diabetes Study: UKPDS 82.* 2013. **56**(9): p. 1925-1933.
60. Wiviott, S.D., et al., *Dapagliflozin and cardiovascular outcomes in type 2 diabetes.* New England Journal of Medicine, 2019. **380**(4): p. 347-357.
61. Barnett, A., et al., *Systematic review and network meta-analysis to compare dapagliflozin with other diabetes medications in combination with metformin for adults with type 2 diabetes.* Intern Med. S, 2014. **6**: p. 006.
62. Lozano-Ortega, G., et al., *Network meta-analysis of treatments for type 2 diabetes mellitus following failure with metformin plus sulfonylurea.* Current medical research and opinion, 2016. **32**(5): p. 807-816.
63. Hollander, P., et al., *A 52-week, multinational, open-label, parallel-group, noninferiority, treat-to-target trial comparing insulin detemir with insulin glargine in a basal-bolus regimen with mealtime insulin aspart in patients with type 2 diabetes.* Clinical therapeutics, 2008. **30**(11): p. 1976-1987.
64. Charokopou, M., et al., *Cost-effectiveness of dapagliflozin versus DPP-4 inhibitors as an add-on to Metformin in the Treatment of Type 2 Diabetes Mellitus from a UK Healthcare System Perspective.* BMC health services research, 2015. **15**(1): p. 496.
65. Holman, R.R., et al., *Addition of biphasic, prandial, or basal insulin to oral therapy in type 2 diabetes.* New England Journal of Medicine, 2007. **357**(17): p. 1716-1730.
66. Soekhlal, R., et al., *Treatment costs of acute myocardial infarction in the Netherlands.* Netherlands Heart Journal, 2013. **21**(5): p. 230-235.
67. Greving, J., et al., *Statin treatment for primary prevention of vascular disease: whom to treat? Cost-effectiveness analysis.* Bmj, 2011. **342**: p. d1672.
68. Postmus, D., et al., *A trial-based economic evaluation of 2 nurse-led disease management programs in heart failure.* American heart journal, 2011. **162**(6): p. 1096-1104.
69. Baeten, S.A., et al., *Lifetime health effects and medical costs of integrated stroke services-a non-randomized controlled cluster-trial based life table approach.* Cost Effectiveness and Resource Allocation, 2010. **8**(1): p. 21.

70. Niessen, L.W., et al., *Lifetime health effects and costs of diabetes treatment.* The Netherlands journal of medicine, 2003.
71. de Vries, E.F., T.J. Rabelink, and W.B. van den Hout, *Modelling the cost-effectiveness of delaying end-stage renal disease.* Nephron, 2016. **133**(2): p. 89-97.
72. Redekop, W.K., et al., *The cost effectiveness of Apligraf® treatment of diabetic foot ulcers.* Pharmacoeconomics, 2003. **21**(16): p. 1171-1183.
73. Osnabrugge, R.L., et al., *Cost-effectiveness of percutaneous coronary intervention versus bypass surgery from a Dutch perspective.* Heart, 2015. **101**(24): p. 1980-1988.
74. Spronk, S., et al., *Cost-effectiveness of endovascular revascularization compared to supervised hospital-based exercise training in patients with intermittent claudication: a randomized controlled trial.* Journal of vascular surgery, 2008. **48**(6): p. 1472-1480.
75. de Groot, S., et al., *A cost of illness study of hypoglycaemic events in insulin-treated diabetes in the Netherlands.* BMJ open, 2018. **8**(3): p. e019864.
76. NHG. *Urineweginfecties*. 2013 September 2018]; Available from: https://www.nhg.org/standaarden/samenvatting/urineweginfecties.
77. Hakkaart-van Roijen L, v.d.L.N., Bouwmans C, Kanters T, Swan Tan S, *Kostenhandleiding: Methodologie van kostenonderzoek en referentieprijzen voor economische evaluaties in de gezondheidszorg*. 2016.
78. Clarke, P., et al., *Estimating the cost of complications of diabetes in Australia using administrative health-care data.* Value in health, 2008. **11**(2): p. 199-206.
79. Isaaz, K., et al., *Return to work after acute ST-segment elevation myocardial infarction in the modern era of reperfusion by direct percutaneous coronary intervention.* Archives of cardiovascular diseases, 2010. **103**(5): p. 310-316.
80. Ericson, L., L. Bergfeldt, and I. Björholt, *Atrial fibrillation: the cost of illness in Sweden.* The European Journal of Health Economics, 2011. **12**(5): p. 479-487.
81. Lindgren, P., E.-L. Glader, and B. Jönsson, *Utility loss and indirect costs after stroke in Sweden.* European Journal of Cardiovascular Prevention & Rehabilitation, 2008. **15**(2): p. 230-233.
82. Fisher, K., R. Hanspal, and L. Marks, *Return to work after lower limb amputation.* International Journal of Rehabilitation Research, 2003. **26**(1): p. 51-56.
83. Frick, K.D. and A. Foster, *The magnitude and cost of global blindness: an increasing problem that can be alleviated.* American journal of ophthalmology, 2003. **135**(4): p. 471-476.
84. Naim A, D.M., Wagner S, Piech C. *Assessing Work Productivity Loss and Disability Among Chronic Kidney Disease Sufferers in the United States*. 2010 September 2018]; Available from: http://www.kantarhealth.com/docs/publications-citations/aohc_10_kidney_disease-work.
85. Sullivan, P.W. and V.H. Ghushchyan, *EQ-5D scores for diabetes-related comorbidities.* Value in Health, 2016. **19**(8): p. 1002-1008.
86. Currie, C.J., et al., *Multivariate models of health-related utility and the fear of hypoglycaemia in people with diabetes.* Current medical research and opinion, 2006. **22**(8): p. 1523-1534.
87. Grandy, S., et al., *Changes in weight loss-related quality of life among type 2 diabetes mellitus patients treated with dapagliflozin.* 2014. **16**(7): p. 645-650.

7.1.4. Code

```
#### Packages ####
library(rstan)
library(rethinking)
library(tidyverse)
library(bayesplot)
bayesplot_theme_set(theme_minimal())
library(MASS)
library(readxl)
library(ggpointdensity)
library(LHS)

#### Section 4.1: A short introduction to Bayesian statistics ####

# small sample
y <- 15
x <- 20

n_samples <- 101000
p <- rep( NA , n_samples )
p[1] <- 0.5

for ( i in 2:n_samples ) {
  p_new <- rnorm( 1 , p[i-1] , 0.1 )
  if ( p_new < 0 ) p_new <- abs( p_new )
  if ( p_new > 1 ) p_new <- 2 - p_new
  q0 <- dbinom( y , x , p[i-1] ) * dbeta(p[i-1], 8, 8)
  q1 <- dbinom( y , x , p_new ) * dbeta(p_new, 8, 8)
  p[i] <- ifelse( runif(1) < q1/q0 , p_new , p[i-1] )
}

p_small <- data.frame(p = p[1001:n_samples], source = "small sample")

# Table 1
precis(p, prob = 0.9)

# large sample
y <- 75
x <- 100

n_samples <- 101000
p <- rep( NA , n_samples )
p[1] <- 0.5

for ( i in 2:n_samples ) {
  p_new <- rnorm( 1 , p[i-1] , 0.1 )
  if ( p_new < 0 ) p_new <- abs( p_new )
  if ( p_new > 1 ) p_new <- 2 - p_new
  q0 <- dbinom( y , x , p[i-1] ) * dbeta(p[i-1], 8, 8)
  q1 <- dbinom( y , x , p_new ) * dbeta(p_new, 8, 8)
  p[i] <- ifelse( runif(1) < q1/q0 , p_new , p[i-1] )
```

```
}

p_large <- p
p_large <- data.frame(p = p[1001:n_samples], source = "large sample")

# Figure 1
p <- rbind(p_small, p_large)

ggplot(p) +
  geom_function(fun = function(x) dbeta(x,8,8), aes(colour = "Prior"), size = 1) +
  geom_density(aes(p, colour = "Posterior"), size = 1) +
  scale_x_continuous(limits =  c(0,1)) +
  scale_color_manual(name = "", values = c("Prior" = "blue", "Posterior" = "red")) +
  xlab("proportion male patients") +
  ylab("probability density") +
  geom_vline(xintercept = mean(p), linetype = 2, size = 1, color = "grey") +
  geom_vline(xintercept = 0.5, linetype = 2, size = 1, color = "grey") +
  facet_wrap(. ~ source) +
  theme_minimal() +
  theme( text = element_text(size = 12))

# Section 4.5: simulating data sets ####

price_met <- 18.26
price_dapa <- 459.85
price_SU <-33.46

dispensing_fees <- 4*7
consumables_ins <- 284.9
consumables_ins_bolus <- 1139.58
weighteffect <- -6.6

price_met + price_dapa + 2*dispensing_fees
price_met + price_SU + 2*dispensing_fees
price_met + 2*dispensing_fees + consumables_ins
price_met + 2*dispensing_fees + consumables_ins_bolus

# Table 17
N <- 40
sim_factors <- maximinLHS(N, 4)

sim_params <- data.frame(index = 1:N)
sim_params$met <- round( sim_factors[,1] * price_met * 2 , 2)
sim_params$dapa <- round( sim_factors[,2] * price_dapa * 2 , 2)
sim_params$SU <- round( sim_factors[,3] * price_SU * 2 , 2)
sim_params$met_dapa <- sim_params$met + sim_params$dapa + 2*dispensing_fees
sim_params$met_SU <- sim_params$met + sim_params$SU + 2*dispensing_fees
sim_params$met_ins <- sim_params$met + 2*dispensing_fees + consumables_ins
sim_params$met_ins_bolus <- sim_params$met + 2*dispensing_fees + consumables_ins_bolus
sim_params$weighteffect <- round( sim_factors[,4] * weighteffect , 2)
```

```
sim_random20 <- data.frame(X1 = runif(20, 0, 1),
                           X2 = runif(20, 0, 1),
                           source = "20 simulations")

sim_random2000 <- data.frame(X1 = runif(2000, 0, 1),
                             X2 = runif(2000, 0, 1),
                             source = "2000 simulations")

sim_random <- rbind(sim_random20, sim_random2000)

# Figure 5
ggplot(sim_random, aes(X1, X2)) +
  geom_point(size = 2) +
  facet_wrap(source ~ .) +
  theme_grey() +
  theme( text = element_text(size = 14))

N <- 20
sim_factors <- maximinLHS(N, 2)

sim_factors <- data.frame(sim_factors)
sim_factors$source <- "LHS"

# Figure 6
ggplot(sim_factors, aes(X1, X2)) +
  geom_point(size = 2, fill = "black") +
  theme_grey() +
  theme( text = element_text(size = 14))

#### Load training data ####

df <- read_excel("[…]", sheet = 1) %>% # insert path to training data here
  rename("effects" = 1,
         "costs" = 2)

dlist <- list(effects = df$effects, costs = df$costs, weight = df$weighteffect,
              pDapa = df$dapa, pMet = df$met, pSU = df$SU, N = nrow(df))

#### Section 7.2: training data ####

# Figure 19
ggplot(df) +
  geom_point(aes(effects, costs), size = 1, color = "blue", alpha = 0.5) +
  facet_wrap( ~ index, ncol = 5) +
  geom_vline(xintercept = 0) +
  geom_hline(yintercept = 0) +
  theme_minimal()



#### Section 4.6: Priors ####
```

```
## Prior model, unconditioned on the data ##
prior_model <- stan(model_code =
                      "data {
   int N;
   vector[N] costs;
   vector[N] effects;
   vector[N] weight;
   vector[N] pDapa;
   vector[N] pMet;
   vector[N] pSU;
                      }

  generated quantities {
   real alphaC;
   real betaC_dapa;
   real betaC_SU;
   real betaC_met;
   real alphaE;
   real betaE_weight;
   real betaC_weight;
   vector<lower=0>[2] sigma;

   real muE;
   real muC;
   vector[2] MU;
   vector[2] y_sim;

   real alphaE_std = normal_rng(0, 0.5);
   real betaE_weight_std = normal_rng(0, 0.5);
   real betaC_weight_std = normal_rng(0, 0.5);
   real alphaC_std = normal_rng(0, 0.5);
   real betaC_dapa_std = normal_rng(0, 0.5);
   real betaC_SU_std = normal_rng(0, 0.5);
   real betaC_met_std = normal_rng(0, 0.5);
   matrix[2,2] Rho = lkj_corr_rng( 2, 2 );
   vector[2] sigma_std;
   sigma_std[1] = exponential_rng(1);
   sigma_std[2] = exponential_rng(1);


   alphaC = sd(costs) * (alphaC_std - betaC_weight_std * mean(weight) / sd(weight) -
betaC_dapa_std * mean(pDapa) / sd(pDapa)
         - betaC_SU_std * mean(pSU) / sd(pSU) - betaC_met_std * mean(pMet) / sd(pMet))
         + mean(costs);
   betaC_dapa = betaC_dapa_std * sd(costs) / sd(pDapa);
   betaC_SU = betaC_SU_std * sd(costs) / sd(pSU);
   betaC_met = betaC_met_std * sd(costs) / sd(pMet);

   alphaE = sd(effects) * (alphaE_std - betaE_weight_std * mean(weight) / sd(weight)) +
mean(effects);
   betaE_weight = betaE_weight_std * sd(effects) / sd(weight);
   betaC_weight = betaC_weight_std * sd(costs) / sd(weight);
```

```
   sigma[1] = sd(effects) * sigma_std[1];
   sigma[2] = sd(costs) * sigma_std[2];

    muE = alphaE + betaE_weight * ( - 3.3) ;
    muC = alphaC + betaC_weight * ( - 3.3) +
              betaC_dapa * 459.85 + betaC_SU * 33.46 + betaC_met * 18.26 ;
    MU = [muE, muC]';


   y_sim = multi_normal_rng( MU , quad_form_diag(Rho , sigma) );

}", data = dlist , chains = 4, cores = 4, warmup = 500, iter = 3000, algorithm = "Fixed_param")

prior <- rstan::extract(prior_model)
prior.matrix <- as.matrix(prior_model)

# Figure 7
mcmc_areas(prior.matrix,
        pars = c("betaC_dapa_std", "betaC_SU_std", "betaC_met_std",
              "betaE_weight_std", "betaC_weight_std"),
        prob = 0.9) +
 scale_x_continuous(limits = c(-2.5, 2.5)) +
 theme( text = element_text(size = 14))

mcmc_areas(prior.matrix,
        pars = c("Rho[1,2]"),
        prob = 0.9)

sim_df <- data.frame(effects = prior$y_sim[,1],
                costs = prior$y_sim[,2])

# Figure 8
ggplot(sim_df) +
 geom_pointdensity(aes(effects, costs), size = 1.5, alpha = 0.75, show.legend = FALSE) +
 geom_hline(yintercept = 0) +
 geom_vline(xintercept = 0) +
 theme_minimal() +
 theme( text = element_text(size = 14)) +
 xlab( "Incremental effects" ) +
 ylab( "Incremental costs" )

#### Section 5.1: Parameter estimates ####

## Actual model, conditioned on the data ##

multinormal_model <- stan(model_code =
                  "data {
   int N;
   vector[N] costs;
   vector[N] effects;
   vector[N] weight;
   vector[N] pDapa;
```

```
  vector[N] pMet;
  vector[N] pSU;
                    }
 transformed data {
  vector[N] costs_std;
  vector[N] effects_std;
  vector[N] pDapa_std;
  vector[N] pMet_std;
  vector[N] pSU_std;
  vector[N] weight_std;

  costs_std = (costs - mean(costs)) / sd(costs);
  effects_std = (effects - mean(effects)) / sd(effects);
  pDapa_std = (pDapa - mean(pDapa)) / sd(pDapa);
  pMet_std = (pMet - mean(pMet)) / sd(pMet);
  pSU_std = (pSU - mean(pSU)) / sd(pSU);
  weight_std = (weight - mean(weight)) / sd(weight);
 }
 parameters {
  real alphaC_std;
  real betaC_dapa_std;
  real betaC_SU_std;
  real betaC_met_std;
  real alphaE_std;
  real betaE_weight_std;
  real betaC_weight_std;
  corr_matrix[2] Rho;
  vector<lower=0>[2] sigma_std;
 }
 model {
  vector[N] muE;
  vector[N] muC;
  // priors
  alphaE_std ~ normal(0, 0.5);
  betaE_weight_std ~ normal(0, 0.5);
  betaC_weight_std ~ normal(0, 0.5);
  alphaC_std ~ normal(0, 0.5);
  betaC_dapa_std ~ normal(0, 0.5);
  betaC_SU_std ~ normal(0, 0.5);
  betaC_met_std ~ normal(0, 0.5);
  Rho ~ lkj_corr( 2 );
  sigma_std ~ exponential(1);

  // sampling of incremental effects
  for (i in 1:N) {
    muE[i] = alphaE_std + betaE_weight_std * weight_std[i] ;
    muC[i] = alphaC_std + betaC_weight_std * weight_std[i] +
              betaC_dapa_std * pDapa_std[i] + betaC_SU_std * pSU_std[i] + betaC_met_std *
pMet_std[i] ;
  }

  {
```

```
   vector[2] YY[N];
   vector[2] MU[N];
   for ( j in 1:N ) MU[j] = [ muE[j] , muC[j] ]';
   for ( j in 1:N ) YY[j] = [ effects_std[j] , costs_std[j] ]';
   YY ~ multi_normal( MU , quad_form_diag(Rho , sigma_std) );
   }
}
 generated quantities {
   // unstandardized parameters
   real alphaC;
   real betaC_dapa;
   real betaC_SU;
   real betaC_met;
   real alphaE;
   real betaE_weight;
   real betaC_weight;
   vector<lower=0>[2] sigma;

   alphaC = sd(costs) * (alphaC_std - betaC_weight_std * mean(weight) / sd(weight) -
betaC_dapa_std * mean(pDapa) / sd(pDapa)
         - betaC_SU_std * mean(pSU) / sd(pSU) - betaC_met_std * mean(pMet) / sd(pMet))
         + mean(costs);
   betaC_dapa = betaC_dapa_std * sd(costs) / sd(pDapa);
   betaC_SU = betaC_SU_std * sd(costs) / sd(pSU);
   betaC_met = betaC_met_std * sd(costs) / sd(pMet);

   alphaE = sd(effects) * (alphaE_std - betaE_weight_std * mean(weight) / sd(weight)) +
mean(effects);
   betaE_weight = betaE_weight_std * sd(effects) / sd(weight);
   betaC_weight = betaC_weight_std * sd(costs) / sd(weight);
   sigma[1] = sd(effects) * sigma_std[1];
   sigma[2] = sd(costs) * sigma_std[2];
}", data = dlist , chains = 4, cores = 4, warmup = 500, iter = 3000)

# Table 7 and Table 8
precis_model <- as.data.frame( precis(multinormal_model, 3, prob = 0.9, digits = 4) )

post <- rstan::extract(multinormal_model)
post.matrix <- as.matrix(multinormal_model)

# Figure 9
mcmc_areas(post.matrix,
        pars = c("betaC_dapa_std", "betaC_SU_std", "betaC_met_std",
              "betaE_weight_std", "betaC_weight_std"),
        prob = 0.9) +
  legend_text(size = 14)


#### Section 5.2: posterior predictive simulation and model validation ####
N <- 10000
price_met <- 18.26
price_dapa <- 459.85
```

```
price_SU <- 33.46
weight <- -3.3

sim <- matrix(NA, nrow = 2, ncol = N)
for (i in 1:N) {
  mu <- c(
    (post$alphaE[i] + post$betaE_weight[i] * weight),
    (post$alphaC[i] + post$betaC_weight[i] * weight +
       post$betaC_dapa[i] * price_dapa + post$betaC_SU[i] * price_SU + post$betaC_met[i] *
price_met))
  Sigma <- matrix( c( post$sigma[i,1]^2, post$Rho[i,1,2]*post$sigma[i,1]*post$sigma[i,2],
post$Rho[i,1,2]*post$sigma[i,1]*post$sigma[i,2], post$sigma[i,2]^2 ), 2, 2)
  sim[,i] <- mvrnorm(1, mu, Sigma)
}

sim_df <- data.frame(effects = sim[1,],
                     costs = sim[2,],
                     source = "Metamodel")

Cardiff_model_df <- read.csv2("[…]") %>% # insert path of validation data here
  rename("effects" = 1,
         "costs" = 2) %>%
  dplyr::select(1:2) %>%
  mutate(source = "Cardiff model")

sim_df <- rbind(sim_df, Cardiff_model_df)

# Table 10
sim_df_summary <- sim_df %>% group_by(source) %>%
  summarise(costs_mean = mean(costs), effects_mean = mean(effects), cost_sd = sd(costs),
effects_sd = sd(effects))

CEAC <- data.frame( threshold = seq(0, 50000, length.out = 1000) )
CEAC$P_Metamodel <- rep(0, 1000)
CEAC$P_Cardiff_model <- rep(0, 1000)
for (i in 1:1000) {
  CEAC$P_Metamodel[i] <- sum (sim_df[sim_df$source == "Metamodel",2] / sim_df[sim_df$source
== "Metamodel",1] < CEAC$threshold[i] ) / N
  CEAC$P_Cardiff_model[i] <- sum (sim_df[sim_df$source == "Cardiff model",2] /
sim_df[sim_df$source == "Cardiff model",1] < CEAC$threshold[i] ) / N
}

# Figure 10
ggplot(sim_df) +
  geom_pointdensity(aes(effects, costs), size = 1.5, alpha = 0.75, show.legend = FALSE) +
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0) +
  facet_wrap(. ~ source) +
  xlab( "Incremental QALYs" ) +
  ylab( "Incremental costs" ) +
  theme_minimal() +
  theme( text = element_text(size = 14))
```

```
# Figure 11
ggplot(CEAC) +
  geom_line(aes(threshold, P_Metamodel, color = "Metamodel"), size = 1) +
  geom_line(aes(threshold, P_Cardiff_model, color = "Cardiff model"), size = 1) +
  geom_vline(xintercept = 20000, linetype = 2, size = 1, color = "grey") +
  ylab( "Probability cost-effectiveness") +
  xlab( "Cost-effectiveness threshold") +
  theme_minimal() +
  scale_color_manual(name = "", values = c("Metamodel" = "blue", "Cardiff model" = "black")) +
  scale_y_continuous(limits = c(0, 1)) +
  theme( text = element_text(size = 14)) +
  theme(legend.position="bottom")

## Price dapa variation ##
M <- 1000
N <- 10000
price_met <- 18.26
price_dapa <- seq(0,1000, length.out = M)
price_SU <- 33.46
weight <- -3.3


sim_eff <- matrix(NA, nrow = N, ncol = M)
sim_costs <- matrix(NA, nrow = N, ncol = M)
ICER <- matrix(NA, nrow = N, ncol = M)
for (j in 1:M) {
  for (i in 1:N) {
    mu <- c(
      (post$alphaE[i] + post$betaE_weight[i] * weight),
      (post$alphaC[i] + post$betaC_weight[i] * weight +
        post$betaC_dapa[i] * price_dapa[j] + post$betaC_SU[i] * price_SU + post$betaC_met[i] *
price_met))
    Sigma <- matrix( c( post$sigma[i,1]^2, post$Rho[i,1,2]*post$sigma[i,1]*post$sigma[i,2],
post$Rho[i,1,2]*post$sigma[i,1]*post$sigma[i,2], post$sigma[i,2]^2 ), 2, 2)
    sim <- mvrnorm(1, mu, Sigma)
    sim_eff[i,j] <- sim[1]
    sim_costs[i,j] <- sim[2]
    ICER[i,j] <- sim[2] / sim[1]
  }
}

mean_costs <- apply(sim_costs, 2, mean)


ci50 <- apply(sim_costs, 2, PI, prob = 0.5)
ci90 <- apply(sim_costs, 2, PI, prob = 0.90)

pred <- data.frame(price_dapa = price_dapa,
                   mean_costs = mean_costs,
                   ci50_lower = ci50[1,],
                   ci50_upper = ci50[2,],
```

```
            ci90_lower = ci90[1,],
            ci90_upper = ci90[2,])

# Figure 12
ggplot(pred) +
  geom_line(aes(price_dapa, mean_costs), size = 1) +
  geom_ribbon(aes(x = price_dapa, ymin = ci50_lower, ymax = ci50_upper, alpha = "50% credibility
interval"), fill = "blue") +
  geom_ribbon(aes(x = price_dapa, ymin = ci90_lower, ymax = ci90_upper, alpha = "90% credibility
interval"), fill = "blue") +
  scale_alpha_manual(name = "", values = c("50% credibility interval" = 0.4, "90% credibility interval"
= 0.2) ) +
  ylab("Incremental costs") +
  xlab( "Price dapagliflozin" ) +
  geom_vline(xintercept = 459.85, linetype = 2, size = 1, color = "grey") +
  geom_hline(yintercept = 0, linetype = 2, size = 1, color = "grey") +
  theme_minimal() +
  theme(legend.position="bottom") +
  theme( text = element_text(size = 14))


P_costsaving <- rep(NA, M)
P_costeffective <- rep(NA, M)
for (i in 1:M) {
  P_costsaving[i] <- sum( ICER[,i] <= 0) / N
  P_costeffective[i] <- sum( ICER[,i] <= 20000) / N
}

CEAC <- data.frame(price_dapa = price_dapa,
            P = c( P_costsaving, P_costeffective ),
            source = c( rep("cost-saving", M), rep("cost-effective", M) ) )

# Figure 13
ggplot(CEAC) +
  geom_line(aes(price_dapa, P), size = 1) +
  ylab( "Probability") +
  xlab( "Price dapagliflozin" ) +
  facet_wrap(. ~ source, dir = "v") +
  theme_minimal() +
  geom_vline(xintercept = 459.85, linetype = 2, size = 1, color = "grey") +
  geom_hline(yintercept = 0.5, linetype = 2, size = 1, color = "grey") +
  scale_y_continuous(limits = c(0, 1)) +
  theme( text = element_text(size = 14))

## weight parameter variation ##
M <- 1000
N <- 10000
price_met <- 18.26
price_dapa <- 459.85
price_SU <- 33.46
weight <- seq( -6.6, 0, length.out = M)
```

```
sim_eff <- matrix(NA, nrow = N, ncol = M)
sim_costs <- matrix(NA, nrow = N, ncol = M)
ICER <- matrix(NA, nrow = N, ncol = M)
for (j in 1:M) {
 for (i in 1:N) {
   mu <- c(
    (post$alphaE[i] + post$betaE_weight[i] * weight[j]),
    (post$alphaC[i] + post$betaC_weight[i] * weight[j] +
      post$betaC_dapa[i] * price_dapa + post$betaC_SU[i] * price_SU + post$betaC_met[i] *
price_met))
   Sigma <- matrix( c( post$sigma[i,1]^2, post$Rho[i,1,2]*post$sigma[i,1]*post$sigma[i,2],
post$Rho[i,1,2]*post$sigma[i,1]*post$sigma[i,2], post$sigma[i,2]^2 ), 2, 2)
   sim <- mvrnorm(1, mu, Sigma)
   sim_eff[i,j] <- sim[1]
   sim_costs[i,j] <- sim[2]
   ICER[i,j] <- sim[2] / sim[1]
 }
}

# costs
mean_costs <- apply(sim_costs, 2, mean)
ci50 <- apply(sim_costs, 2, PI, prob = 0.5)
ci90 <- apply(sim_costs, 2, PI, prob = 0.90)

pred_costs <- data.frame(weight = weight,
                mean = mean_costs,
                source = "incremental costs",
                ci50_lower = ci50[1,],
                ci50_upper = ci50[2,],
                ci90_lower = ci90[1,],
                ci90_upper = ci90[2,])

# effects
mean_eff <- apply(sim_eff, 2, mean)
ci50 <- apply(sim_eff, 2, PI, prob = 0.5)
ci90 <- apply(sim_eff, 2, PI, prob = 0.90)

pred_eff <- data.frame(weight = weight,
                  mean = mean_eff,
                  source = "incremental effects",
                  ci50_lower = ci50[1,],
                  ci50_upper = ci50[2,],
                  ci90_lower = ci90[1,],
                  ci90_upper = ci90[2,])

pred <- rbind(pred_eff, pred_costs)

# Figure 14
ggplot(pred) +
 geom_line(aes(weight, mean), size = 1) +
```

```
  geom_ribbon(aes(x = weight, ymin = ci50_lower, ymax = ci50_upper, alpha = "50% credibility
interval"), fill = "blue") +
  geom_ribbon(aes(x = weight, ymin = ci90_lower, ymax = ci90_upper, alpha = "90% credibility
interval"), fill = "blue") +
  scale_alpha_manual(name = "", values = c("50% credibility interval" = 0.4, "90% credibility interval"
= 0.2) ) +
  ylab( "Incremental QALYs" ) +
  xlab( "Weight increment" ) +
  geom_vline(xintercept = -3.3, linetype = 2, size = 1, color = "grey") +
  facet_wrap(. ~ source, dir = "v", scales = "free") +
  theme_minimal() +
  theme(legend.position="bottom") +
  theme( text = element_text(size = 14))


# CEAC
P_costsaving <- rep(NA, M)
P_costeffective <- rep(NA, M)
for (i in 1:M) {
  P_costsaving[i] <- sum( ICER[,i] <= 0) / N
  P_costeffective[i] <- sum( ICER[,i] <= 20000) / N
}

CEAC <- data.frame(weight = weight,
                   P = c( P_costsaving, P_costeffective ),
                   source = c( rep("cost-saving", M), rep("cost-effective", M) ) )

# Figure 15
ggplot(CEAC) +
  geom_line(aes(weight, P), size = 1) +
  ylab( "Probability") +
  xlab( "Weight increment" ) +
  facet_wrap(. ~ source, dir = "v") +
  theme_minimal() +
  geom_vline(xintercept = -3.3, linetype = 2, size = 1, color = "grey") +
  geom_hline(yintercept = 0.5, linetype = 2, size = 1, color = "grey") +
  scale_y_continuous(limits = c(0, 1)) +
  theme( text = element_text(size = 14))

# timeline
price_forecast <- read_excel("[…]", sheet = 3) # insert path to price forecast here (table 11)
price_dapa <- price_forecast$price_dapa
price_met <- price_forecast$price_met
price_SU <- price_forecast$price_SU
weight <- -3.3
N <- 10000

sim_eff <- matrix(NA, nrow = N, ncol = nrow(price_forecast))
sim_costs <- matrix(NA, nrow = N, ncol = nrow(price_forecast))
ICER <- matrix(NA, nrow = N, ncol = nrow(price_forecast))
for (j in 1:nrow(price_forecast)) {
  for (i in 1:N) {
```

```
   mu <- c(
     (post$alphaE[i] + post$betaE_weight[i] * weight),
     (post$alphaC[i] + post$betaC_weight[i] * weight +
       post$betaC_dapa[i] * price_dapa[j] + post$betaC_SU[i] * price_SU[j] + post$betaC_met[i] *
price_met[j]))
   Sigma <- matrix( c( post$sigma[i,1]^2, post$Rho[i,1,2]*post$sigma[i,1]*post$sigma[i,2],
post$Rho[i,1,2]*post$sigma[i,1]*post$sigma[i,2], post$sigma[i,2]^2 ), 2, 2)
   sim <- mvrnorm(1, mu, Sigma)
   sim_eff[i,j] <- sim[1]
   sim_costs[i,j] <- sim[2]
   ICER[i,j] <- sim[2] / sim[1]
  }
}

mean_costs <- apply(sim_costs, 2, mean)
ci50 <- apply(sim_costs, 2, PI, prob = 0.5)
ci90 <- apply(sim_costs, 2, PI, prob = 0.90)

pred <- data.frame(date = price_forecast$date,
         mean_costs = mean_costs,
         ci50_lower = ci50[1,],
         ci50_upper = ci50[2,],
         ci90_lower = ci90[1,],
         ci90_upper = ci90[2,])

# Figure 16
ggplot(pred) +
  geom_line(aes(date, mean_costs)) +
  geom_point(aes(date, mean_costs)) +
  geom_ribbon(aes(x = date, ymin = ci50_lower, ymax = ci50_upper, alpha = "50% credibility
interval"), fill = "blue") +
  geom_ribbon(aes(x = date, ymin = ci90_lower, ymax = ci90_upper, alpha = "90% credibility
interval"), fill = "blue") +
  scale_alpha_manual(name = "", values = c("50% credibility interval" = 0.4, "90% credibility interval"
= 0.2) ) +
  ylab("Incremental costs") +
  xlab( "Date") +
  geom_hline(yintercept = 0, linetype = 2, size = 1, color = "grey") +
  theme_minimal() +
  theme(legend.position="bottom") +
  theme( text = element_text(size = 14))

P_costsaving <- rep(NA, nrow(price_forecast))
P_costeffective <- rep(NA, nrow(price_forecast))
for (i in 1:nrow(price_forecast)) {
  P_costsaving[i] <- sum( ICER[,i] <= 0) / N
  P_costeffective[i] <- sum( ICER[,i] <= 20000) / N
}

CEAC <- data.frame(date = price_forecast$date,
                   P = c( P_costsaving, P_costeffective ),
```

```
            source = c( rep("cost-saving", nrow(price_forecast)), rep("cost-effective",
nrow(price_forecast)) ) )

# Figure 17
ggplot(CEAC[CEAC$source == "cost-saving",]) +
  geom_line(aes(date, P), size = 1) +
  ylab( "Probability cost-saving") +
  theme_minimal() +
  geom_hline(yintercept = 0.5, linetype = 2, size = 1, color = "grey") +
  scale_y_continuous(limits = c(0, 1)) +
  theme( text = element_text(size = 14))

#### Section 6.2.4: 6.2.4.          (un-)normality of the original model's ditribution ####

# 1
rho <- 0.5
x <- mvrnorm(1000,
         mu = c(0.5, 2500),
         Sigma = matrix( c(0.3^2, 0.3 * 2500 * rho, 0.3 * 2500 * rho, 2500^2), ncol = 2
         ))

df1 <- data.frame(effects = x[,1],
              costs = x[,2],
              sim = 1)


# 2
rho <- -0.5
x <- mvrnorm(1000,
         mu = c(0.5, 2500),
         Sigma = matrix( c(0.3^2, 0.3 * 2500 * rho, 0.3 * 2500 * rho, 2500^2), ncol = 2
         ))

df2 <- data.frame(effects = x[,1],
              costs = x[,2],
              sim = 2)


# 3

x <- matrix(NA, nrow = 1000, ncol = 2)
x[,1] <- rnorm(1000, 0.5, 0.4)
x[,2] <- rnorm(1000, x[,1] * 2000, (exp(x[,1])) * 1000)

df3 <- data.frame(effects = x[,1],
              costs = x[,2],
              sim = 3)



# 4
```

```
random = runif(1000, 0, 1)
x <- matrix(NA, nrow = 1000, ncol = 2)
x[,1] <- rnorm(1000, 0.4, 0.3)
x[,2] <- rnorm(1000, x[,1] * 1000 + ifelse(random < 0.6, 0, 1) * 5000, 1000)

df4 <- data.frame(effects = x[,1],
                  costs = x[,2],
                  sim = 4)

df_total <- as.tibble(rbind(df1, df2, df3, df4))

# Figure 18
ggplot(df_total) +
  geom_pointdensity(aes(effects, costs), size = 1.5, alpha = 1, show.legend = FALSE) +
  geom_hline(yintercept = 0) +
  geom_vline(xintercept = 0) +
  facet_wrap(. ~ sim) +
  theme_minimal() +
  theme( text = element_text(size = 12)) +
  xlab( "Incremental effects" ) +
  ylab( "Incremental costs" )
```