

This is a repository copy of *An Efficient Method for Computing Expected Value of Sample Information for Survival Data from an Ongoing Trial*.

White Rose Research Online URL for this paper:  
<https://eprints.whiterose.ac.uk/182027/>

Version: Published Version

---

**Article:**

Vervaart, Mathyn, Strong, Mark, Claxton, Karl P [orcid.org/0000-0003-2002-4694](https://orcid.org/0000-0003-2002-4694) et al. (3 more authors) (2021) An Efficient Method for Computing Expected Value of Sample Information for Survival Data from an Ongoing Trial. *Medical Decision Making*. 272989X211068019. ISSN 1552-681X

<https://doi.org/10.1177/0272989X211068019>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:  
<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# An Efficient Method for Computing Expected Value of Sample Information for Survival Data from an Ongoing Trial

Medical Decision Making

1–14

© The Author(s) 2021



Article reuse guidelines:

[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)

DOI: 10.1177/0272989X211068019

[journals.sagepub.com/home/mdm](http://journals.sagepub.com/home/mdm)

Mathyn Vervaart , Mark Strong , Karl P. Claxton, Nicky J. Welton, Torbjørn Wisløff , and Eline Aas

**Background.** Decisions about new health technologies are increasingly being made while trials are still in an early stage, which may result in substantial uncertainty around key decision drivers such as estimates of life expectancy and time to disease progression. Additional data collection can reduce uncertainty, and its value can be quantified by computing the expected value of sample information (EVS), which has typically been described in the context of designing a future trial. In this article, we develop new methods for computing the EVS of extending an existing trial's follow-up, first for an assumed survival model and then extending to capture uncertainty about the true survival model. **Methods.** We developed a nested Markov Chain Monte Carlo procedure and a nonparametric regression-based method. We compared the methods by computing single-model and model-averaged EVS for collecting additional follow-up data in 2 synthetic case studies. **Results.** There was good agreement between the 2 methods. The regression-based method was fast and straightforward to implement, and scales easily included any number of candidate survival models in the model uncertainty case. The nested Monte Carlo procedure, on the other hand, was extremely computationally demanding when we included model uncertainty. **Conclusions.** We present a straightforward regression-based method for computing the EVS of extending an existing trial's follow-up, both where a single known survival model is assumed and where we are uncertain about the true survival model. EVS for ongoing trials can help decision makers determine whether early patient access to a new technology can be justified on the basis of the current evidence or whether more mature evidence is needed.

## Highlights

- Decisions about new health technologies are increasingly being made while trials are still in an early stage, which may result in substantial uncertainty around key decision drivers such as estimates of life-expectancy and time to disease progression. Additional data collection can reduce uncertainty, and its value can be quantified by computing the expected value of sample information (EVS), which has typically been described in the context of designing a future trial.
- In this article, we have developed new methods for computing the EVS of extending a trial's follow-up, both where a single known survival model is assumed and where we are uncertain about the true survival model. We extend a previously described nonparametric regression-based method for computing EVS, which we demonstrate in synthetic case studies is fast, straightforward to implement, and scales easily to include any number of candidate survival models in the EVS calculations.
- The EVS methods that we present in this article can quantify the need for collecting additional follow-up data before making an adoption decision given any decision-making context.

## Corresponding Author

Mathyn Vervaart, Department of Health Management and Health Economics, University of Oslo, Forskningsveien 3A, Harald Schjelderups hus, Oslo, 0373, Norway; ([mathyn.vervaart@medisin.uio.no](mailto:mathyn.vervaart@medisin.uio.no)).

## Keywords

bayesian decision theory, computational methods, economic evaluation model, expected value of sample information, generalized additive model, model averaging, Monte Carlo methods, nonparametric regression, survival data

Date received: April 09, 2021; accepted: November 30, 2021

## Introduction

The expected value of sample information (EVSI) quantifies the expected value to the decision maker of reducing uncertainty through the collection of additional data,<sup>1,2</sup> for example, a future randomized controlled trial. Although a few studies have considered the use of EVSI methods at interim analyses of adaptive trials,<sup>3</sup> overall little research has been done on EVSI for trials that are ongoing at the point of decision making.

In the past decade, the European Medicines Agency has introduced regulatory mechanisms that are aimed at accelerating the licensing of new pharmaceuticals, such as adaptive pathways<sup>4</sup> and conditional marketing authorizations.<sup>5</sup> When evidence is obtained from a trial at an early stage, the events of interest, such as disease progression or death, may have only been observed in a small proportion of patients. Health care authorities therefore have to issue guidance on new pharmaceuticals based on less mature evidence than previously, resulting in greater uncertainty about clinical and cost-effectiveness. With this comes an increased risk of recommending a technology that reduces net health benefit.<sup>6</sup>

Additional evidence can be valuable as it can lead to better decisions that improve health and/or reduce resource use.<sup>6</sup> Positive adoption decisions can be costly or difficult to reverse and may remove the incentives for manufacturers to provide additional data. When a trial

is ongoing at the point of decision making, for example, when follow-up is continued for regulatory purposes, there may therefore be value in delaying the adoption decision until additional data have been collected in the ongoing trial and uncertainty has reduced.<sup>7</sup> In this context, there will be a tradeoff between granting early access to a new technology that may turn out to reduce health benefits and waiting for uncertainty to be reduced through ongoing data collection with a potential loss of health benefits while waiting. When the manufacturer is already committed to continuing the ongoing trial, the option to delay a decision is relevant even in a policy context in which the decision maker does not have the formal authority to commission research. The value of delaying the decision could be quantified, at least in theory, by computing the EVSI for the additional follow-up data.

Estimates of life expectancy and time to disease progression are often key drivers of cost-effectiveness, particularly in oncology. However, immature data means that there may be substantial uncertainty around these estimates, and they rely on extrapolation beyond the trial follow-up period.<sup>8</sup> The choice of the survival distribution for extrapolation can have major implications for cost-effectiveness, and uncertainty surrounding this choice can be accounted for by model averaging, which may improve the quality of the extrapolations compared with selecting a single model.<sup>9</sup> A potential benefit of continuing an ongoing trial is to reduce the structural uncertainty as to the most appropriate survival distribution. However, to the best of the authors' knowledge, there is no guidance on how to compute EVSI for survival data from a trial that is ongoing at the point of decision making nor on how to account for structural uncertainty about the choice of survival model in the EVSI calculations.

In this article, we present algorithms for computing the EVSI of extending a trial's follow-up with and without accounting for structural uncertainty. The algorithms are based on nested Markov Chain Monte Carlo (MCMC) methods and a fast nonparametric regression-based method.<sup>10</sup> The nonparametric regression-based method<sup>10</sup> is generally more practical than other EVSI approximation methods, as it neither requires nested Monte Carlo computations nor importance sampling.<sup>11</sup>

---

Department of Health Management and Health Economics, University of Oslo, Oslo, Norway (MV, EA); Norwegian Medicines Agency, Oslo, Norway (MV); School of Health and Related Research, University of Sheffield, Sheffield, UK (MS); Centre for Health Economics, University of York, York, UK (KPC); Department of Economics and Related Studies, University of York, York, UK (KPC); Population Health Sciences, University of Bristol, Bristol, UK (NJW); Department of Community Medicine, UiT The Arctic University of Norway, Oslo, Norway (TTW); Norwegian Institute of Public Health, Oslo, Norway (TTW). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided entirely by a grant from the Norwegian Research Council through NordForsk (298854). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

The article is structured as follows. In the second section, we describe single-model and model-averaged EVSI algorithms for survival data from an ongoing trial. In the third section, we compare the EVSI algorithms in 2 illustrative case studies, and in a final section, we conclude with a brief discussion.

## Method

### *EVSI for an Ongoing Study Collecting Time-to-Event Data*

*Decision problem and model definition.* We assume a decision problem with  $d = 1, \dots, D$  decision options. The net benefit of option  $d$  is  $\text{NB}(d, \boldsymbol{\theta})$ , and we have a cost-effectiveness model that predicts this quantity, given a vector of  $p$  possibly correlated model input parameters,  $\boldsymbol{\theta} = \{\theta_1, \dots, \theta_p\}$ . Our current judgments about the vector  $\boldsymbol{\theta}$  is represented by the joint probability distribution  $p(\boldsymbol{\theta})$ . Our goal is to choose the decision option with the greatest net benefit.

*EVSI for further follow-up in an ongoing study.* The EVSI for a new study that will provide (as yet uncollected) data,  $\mathbf{x}$ , is defined as:

$$\text{EVSI}(\text{new study}) = \mathbb{E}_{\mathbf{x}}[\max_d \mathbb{E}_{\boldsymbol{\theta}|\mathbf{x}}\{\text{NB}(d, \boldsymbol{\theta})\}] - \max_d \mathbb{E}_{\boldsymbol{\theta}}\{\text{NB}(d, \boldsymbol{\theta})\}, \quad (1)$$

where the first term is the expected value of a decision based on our beliefs about  $\boldsymbol{\theta}$  given the new data,  $p(\boldsymbol{\theta}|\mathbf{x})$ , and the second term is the expected value of a decision based on our beliefs about  $\boldsymbol{\theta}$  given current information alone,  $p(\boldsymbol{\theta})$ .<sup>12</sup> We now imagine that data  $\mathbf{x}$  have been collected during a given follow-up period for this study, which we denote as time  $t_1$ . This could be an interim analysis or the end of the study follow-up period.

The value of extending the follow-up from current time  $t_1$  to some future point  $t_2$  is given by

$$\text{EVSI}(\text{ongoing study}) = \mathbb{E}_{\tilde{\mathbf{x}}|\mathbf{x}}[\max_d \mathbb{E}_{\boldsymbol{\theta}|\mathbf{x}, \tilde{\mathbf{x}}}\{\text{NB}(d, \boldsymbol{\theta})\}] - \max_d \mathbb{E}_{\boldsymbol{\theta}|\mathbf{x}}\{\text{NB}(d, \boldsymbol{\theta})\}, \quad (2)$$

where the first term is the expected value of a decision based on our beliefs about  $\boldsymbol{\theta}$  given both new data,  $\tilde{\mathbf{x}}$ , collected between  $t_1$  and  $t_2$ , and data,  $\mathbf{x}$ , collected between time 0 and  $t_1$ . The second term is the expected value of a decision based on our beliefs about  $\boldsymbol{\theta}$  given only the information collected up until  $t_1$ . See Appendix A for a fuller explanation.

*Specifying current beliefs about model parameters for an ongoing study.* The distribution for the cost-effectiveness model parameters given knowledge at  $t_1$   $p(\boldsymbol{\theta}|\mathbf{x})$  can be defined either in a fully Bayesian manner, by updating (possibly vague) prior information about  $\boldsymbol{\theta}$  with data  $\mathbf{x}$ , or by fitting a standard frequentist statistical model to  $\mathbf{x}$  and obtaining the maximum likelihood estimate for  $\boldsymbol{\theta}$  along with some expression of uncertainty and treating this as a Bayesian posterior. In the absence of strong prior information about  $\boldsymbol{\theta}$ , the 2 methods will produce very similar distributions for  $p(\boldsymbol{\theta}|\mathbf{x})$ , even with relatively little data.<sup>13</sup>

*Specifying the likelihood for ongoing time-to-event data and left truncation.* To compute EVSI, we must define the data-generating distribution for the follow-up data between  $t_1$  and  $t_2$ ,  $p(\tilde{\mathbf{x}}|\boldsymbol{\theta})$ . We first consider the structure of the data we will observe. We assume our study has 2 arms, new treatment and standard care, and that  $N$  participants are recruited into each arm. Data,  $\mathbf{x}$ , collected from time 0 to  $t_1$ , take the form of a vector of times to death, time to end of follow-up, or time to loss to follow-up, whichever is soonest. Survival times for those alive at  $t_1$  are censored. If we continue to collect data  $\tilde{\mathbf{x}}$  from  $t_1$  to  $t_2$ , we may observe times to death for the participants whose observations were censored at  $t_1$ . Survival times for those alive at  $t_2$  or lost to follow-up are now the only observations censored. Table 1 illustrates the structure of the data for 1 arm of a study with follow-up at 12 and 24 mo.

Survival times are usually assumed to arise from a data-generating process that can be described using a parametric model, the form of which must be chosen by the analyst.<sup>14</sup> Censoring is common when collecting time-to-event data, as the follow-up time may not be long enough to observe the endpoint of interest for all individuals in the trial, and some individuals may be lost to follow-up.<sup>15</sup> The likelihood function for survival data,  $\mathbf{x}$ , obtained up until  $t_1$  for a model with hazard function  $h(\cdot)$  and survivor function  $S(\cdot)$ , is

$$\text{Likelihood } p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n_1} h(x_i, \boldsymbol{\theta})^{\delta_i} S(x_i, \boldsymbol{\theta}), \quad (3)$$

where  $i$  indexes the  $n_1 = N$  study participants at risk at time 0, where the censoring indicator  $\delta_i = 1$  when  $x_i$  is an observed event,  $\delta_i = 0$  when  $x_i$  is a censored observation, and where  $\boldsymbol{\theta}$  are the parameters of the survival distribution. The observed data set at time point  $t_1$  consists of the  $n_1$  survival times and censoring indicators,  $\mathbf{x} = \{x_1, \dots, x_{n_1}, \delta_1, \dots, \delta_{n_1}\}$ .

The data collected between time points  $t_1$  and  $t_2$  is denoted  $\tilde{\mathbf{x}} = \{\tilde{x}_1, \dots, \tilde{x}_{n_2}, \tilde{\delta}_1, \dots, \tilde{\delta}_{n_2}\}$ , where  $n_2$  is the

**Table 1** Structure of Data for 1 Arm of a Study with Follow-up at 12 and 24 mo<sup>a</sup>

ID	Follow-up at $t_1 = 12$ Mo			Follow-up at $t_2 = 24$ Mo			Outcome
	Survival Time	Censoring Indicator, $\delta$	At Risk at $t_2$	Survival Time	Censoring Indicator, $\delta$		
1	9.3	1	No	—	—		Died at 9.3 mo
2	12.0 <sup>b</sup>	0	Yes	13.4	1		Died at 13.4 mo
3	12.0 <sup>b</sup>	0	Yes	24.0 <sup>b</sup>	0		Alive at 24.0 mo
4	6.7 <sup>b</sup>	0	No	—	—		LFU at 6.7 mo
5	12.0 <sup>b</sup>	0	Yes	15.9 <sup>b</sup>	0		LFU at 15.9 mo

LFU, lost to follow-up.

<sup>a</sup>Five participants are shown. Data are denoted  $\mathbf{x} = \{(9.3, 12, 12, 6.7, 12), (1, 0, 0, 0, 0)\}$  for observations up until  $t_1 = 12$  mo and  $\tilde{\mathbf{x}} = \{(13.4, 24, 15.9), (1, 0, 0)\}$  for observations between  $t_1$  and  $t_2 = 24$  mo.

<sup>b</sup>Observation censored ( $\delta = 0$ ).

number of study participants at risk at  $t_1$ . The likelihood function for  $\tilde{\mathbf{x}}$  is left truncated at  $t_1$  to reflect that events beyond  $t_1$  are conditional on not having occurred prior to  $t_1$ .<sup>16</sup> Unlike censoring, which contributes to the likelihood by plugging in a survival factor for censored observations as well as observed survival times, truncation does not add any data points to the likelihood. This distinction is important, because we want to avoid double counting the observed data  $\mathbf{x}$  when we compute the likelihood for the ongoing study data  $\tilde{\mathbf{x}}$ . The left-truncated likelihood has an additional term in the denominator that renormalizes the truncated distribution so that it integrates to 1, that is,

$$\text{Left-truncated likelihood } p_{LT}(\tilde{\mathbf{x}}|\boldsymbol{\theta}) = \prod_{i=1}^{n_2} \frac{h(\tilde{x}_i, \boldsymbol{\theta})^{\delta_i} S(\tilde{x}_i, \boldsymbol{\theta})}{S(t_1, \boldsymbol{\theta})}. \quad (4)$$

Once we have derived the posterior distribution for the model parameters given data at  $t_1$ ,  $p(\boldsymbol{\theta}|\mathbf{x})$ , and the likelihood for the ongoing follow-up data,  $p_{LT}(\tilde{\mathbf{x}}|\boldsymbol{\theta})$ , we require a method for actually computing expression (2). In almost all realistic applications, this will require numerical methods. Nested Monte Carlo can be used, but this is computationally expensive. A regression-based approach is much quicker,<sup>10</sup> and this is described along with the Monte Carlo approach in Appendix B.

We are now in a position to describe methods for computing EVSI that account for uncertainty about the choice of survival model.

#### Model-Averaged EVSI for an Ongoing Study Accounting for Survival Model Uncertainty

*Survival model uncertainty and model averaging.* In this section, “model” refers to the survival model for the

time-to-event data  $p(\mathbf{x}|\boldsymbol{\theta})$ , not the cost-effectiveness model,  $\text{NB}(d, \boldsymbol{\theta})$ . In many real applications, we will be uncertain about which survival model is most appropriate and should be used to extrapolate the data beyond the observed follow-up period  $t_1$ , although we may be comfortable with proposing a candidate set of models,  $\mathcal{M} = M_r$ ,  $r = 1, \dots, R$ , that covers plausible approximations of the data-generating process, that is, the set is  $\mathcal{M}$  – *open* in the terminology used by Bernardo and Smith.<sup>17</sup> In these circumstances, we may account for model uncertainty using predictive model averaging and average over model predictions using model weights based on each model’s predictive ability.<sup>18,19</sup> After observing data  $\mathbf{x}$  at time  $t_1$ , we place probability weight  $P(M_r|\mathbf{x})$  on the  $r^{\text{th}}$  model producing the best predictions, with  $\sum_{r=1}^R P(M_r|\mathbf{x}) = 1$ .

The net benefit function for decision option  $d$  given model  $M_r$  and parameters  $\boldsymbol{\theta}_r$  is denoted  $\text{NB}(d, \boldsymbol{\theta}_r, M_r)$ . Taking the expectation over both parameters and models after observing data  $\mathbf{x}$  up to time point  $t_1$  gives us

$$\begin{aligned} \text{Model-averaged } \text{NB}_d|\mathbf{x} &= \sum_{r=1}^R \{ \mathbb{E}_{\boldsymbol{\theta}_r|\mathbf{x}, M_r} \text{NB}(d, \boldsymbol{\theta}_r, M_r) P(M_r|\mathbf{x}) \} \\ &= \mathbb{E}_{\mathcal{M}|\mathbf{x}} [ \mathbb{E}_{\boldsymbol{\theta}_r|\mathbf{x}, M_r} \{ \text{NB}(d, \boldsymbol{\theta}_r, M_r) \} ] \\ &= \mathbb{E}_{\boldsymbol{\theta}_r, \mathcal{M}|\mathbf{x}} \{ \text{NB}(d, \boldsymbol{\theta}_r, M_r) \}, \end{aligned} \quad (5)$$

and the optimal choice at time point  $t_1$  is the decision  $d$  that maximizes this expectation.

*EVSI for an ongoing study accounting for model uncertainty.* Additional follow-up data  $\tilde{\mathbf{x}}$  will not only update our judgments about parameters,  $p(\boldsymbol{\theta}_r|\mathbf{x}, \tilde{\mathbf{x}}, M_r)$ , but will also update our judgments about the relative plausibility of each model,  $P(M_r|\mathbf{x}, \tilde{\mathbf{x}})$ , for each model  $r = 1, \dots, R$ .

The EVSI for an ongoing study, where we average over models, is given by

$$\begin{aligned} \text{Model-averaged EVSI} &= \mathbb{E}_{\tilde{\mathbf{x}}|\mathbf{x}}[\max_d \mathbb{E}_{\boldsymbol{\theta}_r, \mathcal{M}|\mathbf{x}, \tilde{\mathbf{x}}} \{ \text{NB}(d, \boldsymbol{\theta}_r, M_r) \}] \\ &\quad - \max_d \mathbb{E}_{\boldsymbol{\theta}_r, \mathcal{M}|\mathbf{x}} \{ \text{NB}(d, \boldsymbol{\theta}_r, M_r) \}, \end{aligned} \quad (6)$$

which is identical to equation (2), except that expectations are now taken over models as well as parameters (see Appendix C for a derivation).

To compute equation (6) we will need a method for generating plausible data sets  $\tilde{\mathbf{x}}$  from  $p(\tilde{\mathbf{x}}|\mathbf{x})$ , the distribution of the follow-up data given the observed data, which takes account of the fact that we now consider plausible a number of different data-generating models. We will also need to define model probabilities given observed data,  $P(M_r|\mathbf{x})$ , and then find a method for computing posterior model probabilities  $P(M_r|\mathbf{x}, \tilde{\mathbf{x}})$ , given each sampled future plausible data set  $\tilde{\mathbf{x}}$ . We address the issue of defining model probabilities given observed data first.

*Deriving model probabilities given observed data up until  $t_1$ .* We assume that before we see the observed data  $\mathbf{x}$ , that we are indifferent about the ‘‘correct’’ model, so  $P(M_r) = 1/R$  for all  $r$ . After we observe data  $\mathbf{x}$ , we use the Akaike’s Information Criterion (AIC)<sup>20</sup> to derive posterior model probabilities giving greater weight to models with better predictive ability (according to Kullback-Leibler divergence), as described by Jackson et al.<sup>18</sup> We set

$$P(M_r|\mathbf{x}) = \frac{\exp\{-0.5 \text{AIC}_r(\mathbf{x})\}}{\sum_{r=1}^R \exp\{-0.5 \text{AIC}_r(\mathbf{x})\}}, \quad (7)$$

where

$$\text{AIC}_r(\mathbf{x}) = -2 \log\{p(\mathbf{x}|\hat{\boldsymbol{\theta}}_r)\} + 2u_r.$$

The term  $\hat{\boldsymbol{\theta}}_r$  is the maximum likelihood estimate for the parameters of model  $M_r$ , and  $u_r$  is the number of parameters in model  $M_r$ .

*Generating plausible ongoing follow-up data sets,  $\tilde{\mathbf{x}}$ , that we may observe between  $t_1$  and  $t_2$ .* Plausible data sets from the distribution  $p(\tilde{\mathbf{x}}|\mathbf{x})$  are generated as follows. First, we sample a model  $M_r^{(k)}$  with probability  $P(M_r|\mathbf{x})$  given by equation (7). Next, we draw a sample  $\boldsymbol{\theta}_r^{(k)}$  from the distribution of the parameters of our chosen model  $p(\boldsymbol{\theta}_r|\mathbf{x}, M_r^{(k)})$ . Finally, we generate a data set  $\tilde{\mathbf{x}}^{(k)}$  from the distribution of the data  $p(\tilde{\mathbf{x}}|\boldsymbol{\theta}_r^{(k)}, M_r^{(k)})$  given the sampled parameter values  $\boldsymbol{\theta}_r^{(k)}$  and model  $M_r^{(k)}$ . We can repeat this

process  $k = 1, \dots, K$  times to generate an arbitrary number of data sets.

*Updating model probabilities given ongoing follow-up data from  $t_1$  to  $t_2$ .* We can derive our posterior model probabilities at time point  $t_2$ , for data set  $\tilde{\mathbf{x}}^{(k)}$ , via Bayes theorem:

$$P(M_r|\mathbf{x}, \tilde{\mathbf{x}}^{(k)}) = \frac{p(\tilde{\mathbf{x}}^{(k)}|M_r, \mathbf{x})P(M_r|\mathbf{x})}{\sum_{r=1}^R p(\tilde{\mathbf{x}}^{(k)}|M_r, \mathbf{x})P(M_r|\mathbf{x})}, \quad (8)$$

where  $p(\tilde{\mathbf{x}}^{(k)}|M_r, \mathbf{x})$  is the marginal likelihood (‘‘marginal’’ because we have integrated out the model parameters):

$$p(\tilde{\mathbf{x}}^{(k)}|M_r, \mathbf{x}) = \int_{\Theta} p(\tilde{\mathbf{x}}^{(k)}|M_r, \boldsymbol{\theta}_r) p(\boldsymbol{\theta}_r|M_r, \mathbf{x}) d\boldsymbol{\theta}_r.$$

We use bridge sampling to approximate the marginal likelihood, which is a form of importance sampling that has been shown to give good approximations in a wide range of settings.<sup>21–24</sup> The key notion behind bridge sampling is that the marginal likelihood can be written as the ratio of 2 expectations, each of which can be estimated via importance sampling. The name ‘‘bridge’’ reflects the incorporation in the estimator of a density function that ‘‘bridges’’ (i.e., has good overlap with) the 2 densities from which samples are drawn. A detailed tutorial on the bridge sampling method is given in the article by Gronau et al.,<sup>23</sup> and the method is straightforward to implement in the R package `bridgesampling`.<sup>25</sup> Given the bridge sampling estimates of  $p(\tilde{\mathbf{x}}^{(k)}|M_r, \mathbf{x})$  for each model, posterior model probabilities are trivial to compute via expression (8).

As with single-model EVSI, computing model-averaged EVSI (expression [6]) will require numerical methods. Nested Monte Carlo and a regression-based approach are described in Appendix D. In the next section, we will apply these methods in a synthetic case study.

*Synthetic case study.* We will model survival with and without accounting for survival model uncertainty.

*Decision problem and model definition.* Our decision problem is to determine which of 2 treatment options has the longest mean survival: a new treatment ( $d = 1$ ) or standard care ( $d = 2$ ).

In the single-model case, survival is assumed to follow a Weibull distribution, and the net benefit of each treatment option is assumed to equal the restricted mean

**Table 2.** Weibull and Gamma distribution parameters for the synthetic case study datasets

	Increasing Hazard Case Study		Decreasing Hazard Case Study	
	New Treatment	Standard Care	New Treatment	Standard Care
Weibull shape, $k$	1.10	1.10	0.60	0.60
Weibull scale, $\lambda$	70.00	50.00	80.00	57.00
Gamma shape, $\alpha$	1.80	1.80	0.80	0.80
Gamma rate, $\beta$	0.04	0.04	0.01	0.01

survival time, given an overall time horizon of  $t_h = 240$  mo (i.e., the area under the survival curve from 0 to 240 mo). So the net benefit function is:

$$\text{NB}(d, \boldsymbol{\theta}_d) = \int_0^{t_h} \exp\left\{-\left(\frac{t}{e^{\theta_{\lambda d}}}\right)^{e^{\theta_{kd}}}\right\} dt, \quad (9)$$

where the model parameters are the log-transformed Weibull shape and scale parameters,  $\boldsymbol{\theta}_d = (\theta_{kd}, \theta_{\lambda d})$ . Computing the restricted mean survival for distributions other than the exponential requires numerical integration, but easy-to-use functions are available in the R package `flexsurv`.<sup>26</sup>

In the model-averaged case, the decision problem is as above, but we assume we are uncertain about the choice of survival model,  $M_r$ , to extrapolate the observed data beyond the current follow-up period  $t_1$ . We assume that our set of plausible models  $\mathcal{M}$  contains the following 4 parametric distributions: Weibull ( $r = 1$ ), Gamma ( $r = 2$ ), log-normal ( $r = 3$ ), and log-logistic ( $r = 4$ ).

*Generating synthetic case study data sets,  $\mathbf{x}$ , collected up to  $t_1 = 12$  months.* We generated 2 synthetic case study data sets: one in which the hazard of death is monotonically increasing and the other in which it is monotonically decreasing. For each case study, we generated a data set with 200 participants per trial arm with a maximum follow-up of  $t_1 = 12$  mo. We denote the data sets  $\mathbf{x}_1$  for new treatment and  $\mathbf{x}_2$  for standard care.

To explore the performance of the method when the survival model was misspecified, we generated survival times evenly spaced from either a Weibull or a Gamma distribution, using the 0.005<sup>th</sup>, 0.015<sup>th</sup>, ..., 0.985<sup>th</sup>, 0.995<sup>th</sup> quantiles from each distribution (i.e., 100 evenly spaced quantiles that avoid 0 and 1). We could have randomly generated survival times, but this would have just added additional Monte Carlo error when assessing the methods for computing EVSI. The parameters of the Weibull and Gamma distributions that we used to generate the synthetic case study data sets are shown in Table 2.

We enrolled all patients in the trial at  $t_0 = 0$  and right censored the data sets at  $t_1 = 12$  mo. We assumed no loss to follow-up and did not apply any other censoring. Figure 1 shows the Kaplan-Meier plots for the 2 synthetic case study data sets.

#### *Initial Trial Analysis at $t_1 = 12$ Months*

For each synthetic case study, we analyzed the 2 trial arms separately. We fitted all 4 models to the data from each arm and estimated the model parameters using maximum likelihood (as implemented in the `flexsurv` function).<sup>26</sup> We assumed that our judgments about the log-transformed parameters for each survival model conditional on the observed data up to  $t_1$ ,  $p(\boldsymbol{\theta}_r|\mathbf{x})$ , are represented by a bivariate normal distribution with the mean vector and covariance matrix derived from the maximum likelihood estimation. We computed the AIC for each model fit and derived model probability weights via equation (7).

Net benefits, AICs, and model probabilities are shown in Table 3, and means and covariances for each model are reported in Appendix G.

The expected net benefits (mean survival times) assuming a single Weibull model computed via Equation (9) are 50.96 versus 44.01 mo (incremental = 6.95 mo) for the increasing hazard data set and 84.81 versus 77.85 mo (incremental = 6.97 mo) for the decreasing hazard data set. The expected value of perfect information (EVPI) values, computed via Monte Carlo simulation with a sample size of  $10^5$ , are 4.93 and 6.33 mo for the increasing and decreasing hazard data set, respectively.

The model-averaged net benefits, weighted by model probabilities, were 72.93 versus 62.36 mo (incremental = 10.57 mo) for the increasing hazard data set and 93.31 versus 85.85 mo (incremental = 7.46 mo) for the decreasing hazard data set. The model-averaged EVPI values are 10.32 and 9.97 mo for the respective data sets.

*Generating plausible ongoing follow-up data sets,  $\tilde{\mathbf{x}}$ , for the EVSI computation.* Both the nested Monte Carlo and

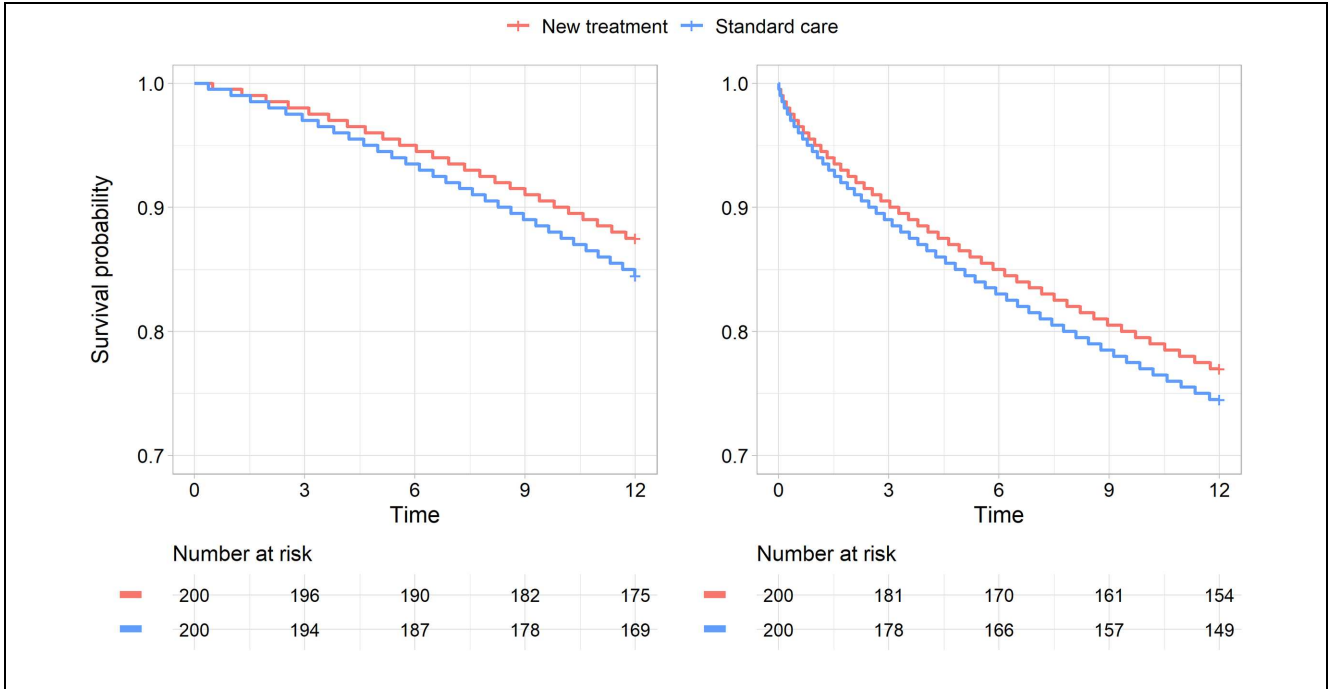


Figure 1 Kaplan-Meier plots for the increasing hazard data set (left) and decreasing hazard data set (right).

Table 3 Mean Survival, Akaike’s Information Criterion (AIC), and Prior Model Probabilities  $P(M_r|x)$  for the 2 Hypothetical Data Sets

	Increasing Hazard Data Set			Decreasing Hazard Data Set		
	Net Benefit (Mean Survival)	AIC(x)	$P(M_r x)$	Net Benefit (Mean Survival)	AIC(x)	$P(M_r x)$
New treatment						
Weibull	50.96	277.58	0.26	84.81	437.06	0.29
Gamma	57.71	277.57	0.26	74.41	437.08	0.29
Log-normal	110.43	277.97	0.22	123.49	438.45	0.14
Log-logistic	79.28	277.58	0.26	105.98	437.14	0.28
Weighted average	72.93			93.31		
Standard care						
Weibull	44.01	329.26	0.28	77.85	470.37	0.30
Gamma	49.42	329.29	0.28	66.99	470.40	0.29
Log-normal	98.43	330.18	0.18	116.25	472.01	0.13
Log-logistic	71.00	329.33	0.27	99.70	470.47	0.28
Weighted average	62.36			85.85		
Incremental values						
Weighted average	10.57			7.46		

regression-based EVSI methods require a set of sampled ongoing follow-up data sets for each trial arm, denoted  $\tilde{x}_1$  and  $\tilde{x}_2$ . We generated  $k = 1, \dots, K$  data sets with  $K = 6000$  for each trial arm, where the  $k^{th}$  data set was generated as follows.

In the single-model case, we first sampled log-shape and log-scale values  $(\theta_1^{(k)})$  for new treatment and  $(\theta_2^{(k)})$  for standard care) from the bivariate normal distributions in Appendix G. We computed the net benefit for each decision option, given the sampled parameters,  $NB(d, \theta_d^{(k)})$



and stored this (these values are required for the regression-based approximation). For each arm, we then sampled  $n$  survival times from a truncated Weibull distribution (see Appendix E) with the sampled shape and scale values, where  $n$  was the number of patients who were still alive in the trial arm at  $t_1 = 12$  mo. Finally, survival times were censored at the proposed endpoint for the ongoing data collection,  $t_2$ .

In the model-averaged case, we first chose a model  $M_r^{(k)}$  with probability  $P(M_r|\mathbf{x})$ , before sampling  $\theta_r^{(k)}$  from the bivariate normal distribution  $p(\theta_r|\mathbf{x})$  for the chosen model  $M_r^{(k)}$  and generating the  $n$  survival times for each arm. The remainder of the data-generation step is as above.

*Computing EVSI for ongoing follow-up via nested Monte Carlo.* To sample from the posterior distributions,  $p(\theta_d|\mathbf{x}_d, \tilde{\mathbf{x}}_d^{(k)})$ , we used Hamiltonian Monte Carlo (HMC) as implemented in the package rstan.<sup>27</sup> HMC is a Metropolis-Hastings MCMC algorithm with a particularly efficient sampling scheme that reduces Monte Carlo sampling error, therefore requiring fewer posterior samples for any inference. The package rstan is an R interface to the Stan language.<sup>28</sup> An alternative option would have been to use OpenBUGS.<sup>29</sup>

In the single-model case, for each outer loop sampled data set,  $k = 1, \dots, 6000$ , we averaged the net benefit functions over  $J = 2000$  inner loop posterior samples of the model parameters and stored the maximum net benefit of the 2 treatment options. We then averaged these maximized net benefits and subtracted the expected value of a decision based on current information to obtain the EVSI following expression (5) in Appendix B.

In the model-averaged case, for each outer loop data set, we generated the  $J$  posterior samples of the model parameters for each of the  $r = 1, \dots, 4$  models (we needed to identify the truncated likelihood function for each model as we did for the Weibull example above, but this is straightforward (see Appendix E)). We weighted the parameter averaged net benefits  $\text{NB}_r^k(d)$  by the posterior model probabilities  $P(M_r|\tilde{\mathbf{x}}^{(k)})$  to give the posterior model-averaged expected net benefit and identified the treatment  $d$  that maximized this for iteration  $k = 1, \dots, 6000$ . We then subtracted the expected value of a decision based on current information to obtain the EVSI following expression (14) in Appendix D.

*Computing EVSI for ongoing follow-up via regression.* The generalized additive model (GAM) approach to computing EVSI for extending the follow-up until time  $t_2$  for the hypothetical example is as follows.

For each trial arm, we computed a low-dimensional summary statistic for each data set. A convenient choice here is the number of observed events  $e_d^{(k)}$  and the total time at risk  $y_d^{(k)}$  for each data set  $\tilde{\mathbf{x}}_d^{(k)}$ , that is,  $T(\tilde{\mathbf{x}}_d^{(k)}) = \{e_d^{(k)}, y_d^{(k)}\}$  for  $d = 1, 2$ .

Then, for each of the 2 decision options, we fitted a GAM regression model with the stored net benefits  $\text{NB}(d, \theta_d^{(k)})$  as the dependent variable and the two summary statistics,  $e_d^{(k)}$  and  $y_d^{(k)}$ , as independent variables. We allowed a smooth, arbitrary, nonlinear relationship between the independent and dependent variables, plus an arbitrary interaction between the independent variables, by specifying a tensor product cubic regression spline basis for the independent variables. This has the simple syntax `gam(nb_d ~ te(e_d, y_d))` in the `mgcv`<sup>30</sup> package in R. We extracted the GAM model fitted values  $\hat{g}_d^{(k)}$  from each regression model fit and estimated the EVSI using equation (9) in Appendix B.

The GAM-based approximation method for model-averaged EVSI is identical to that used in the single-model case.

## Results

### *EVSI Values for the Weibull Ongoing Data*

The nested Monte Carlo- and GAM-based EVSI estimates for additional follow-up times of 12, 24, 36, and 48 mo (i.e.,  $t_2 = 24, 36, 48, 60$  mo) are shown in Table 4. The methods used to estimate the standard errors of the nested Monte Carlo and GAM estimators are described in an appendix of the article by Strong et al.<sup>31</sup>

As expected, the EVSI reflects the diminishing marginal returns for increasing the follow-up duration and converges toward the EVPI. The EVSI varies depending on the underlying hazard pattern, even when point estimates of mean incremental survival benefit are similar (6.95 mo for the increasing hazard data set and 6.97 mo for the decreasing hazard data set). The increasing hazard data set has lower numbers of prior observed events and higher expected numbers of future events for the additional follow-up time than the decreasing hazard data set does, which—all else equal—is expected to result in greater EVSI values. This upward effect on EVSI is, however, canceled out by the downward effect of lower estimates of mean survival, resulting in greater EVSI values for the decreasing hazard data set than for the increasing hazard data set.

The GAM method agrees well with the MCMC method, with the benefit of a greatly reduced computational cost. The MCMC inner loop for the Monte Carlo

**Table 4** EVSI (SE) Values for Additional Follow-up Time for the 2 Hypothetical Data Sets Given a Weibull Distribution for the Survival Times

Additional Follow-up (mo)	Increasing Hazard Data Set		Decreasing Hazard Data Set	
	Nested Monte Carlo	GAM	Nested Monte Carlo	GAM
12	4.25 (0.09)	4.28 (0.08)	4.41 (0.10)	4.46 (0.10)
24	4.58 (0.09)	4.62 (0.06)	5.20 (0.11)	5.27 (0.09)
36	4.68 (0.09)	4.71 (0.05)	5.45 (0.11)	5.54 (0.08)
48	4.74 (0.09)	4.77 (0.04)	5.55 (0.11)	5.65 (0.07)

EVSI, expected value of sample information; GAM, generalized additive model.

<sup>a</sup>EVPI values are 4.93 and 6.33, respectively. Total computation times for the analyses in the table are 24,808 s (nested Monte Carlo) and 36 s (GAM).

**Table 5** EVSI (SE) Values for Additional Follow-up Time for the Two Hypothetical Data Sets Given a Weibull Distribution for the Survival Times

Additional Follow-up (mo)	Increasing Hazard Data Set		Decreasing Hazard Data Set	
	Nested Monte Carlo	GAM	Nested Monte Carlo	GAM
12	7.50 (0.18)	7.52 (0.14)	6.69 (0.15)	6.70 (0.13)
24	8.75 (0.20)	8.82 (0.10)	8.09 (0.18)	8.16 (0.11)
36	9.43 (0.21)	9.44 (0.08)	8.71 (0.19)	8.76 (0.09)
48	9.77 (0.22)	9.74 (0.07)	8.96 (0.19)	9.01 (0.08)

EVSI, expected value of sample information; GAM, generalized additive model.

<sup>a</sup>EVPI values are 10.32 and 9.97, respectively. Total computation times for the analyses in the table are 289,211 s (nested Monte Carlo) and 37 s (GAM).

method used parallel processing, but even with this additional efficiency, the regression method was approximately 700 times faster than the nested Monte Carlo method was. We used a machine running Windows 10 with an Intel Core i9 CPU with 15 threads running on 8 cores at 2.40 GHz and with 32 GB RAM.

Of note is that the standard errors for the nested Monte Carlo estimator slightly increase with increasing follow-up duration, while the opposite is true for the GAM estimator. This is due to different mechanisms through which the effective sample size of the generated data  $\tilde{x}$  affects the standard errors of the nested Monte Carlo and GAM estimators, which is further explained in Appendix F.

#### Model-Averaged EVSI Values

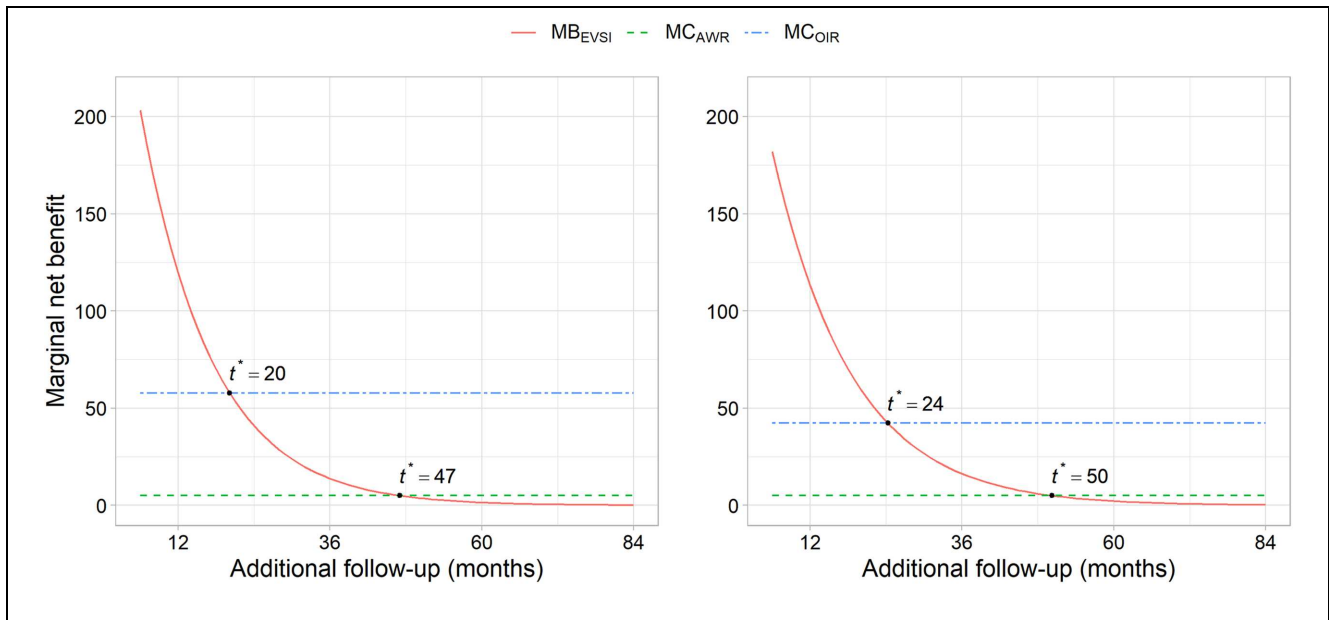
The nested Monte Carlo- and GAM-based model-averaged EVSI estimates for additional follow-up times of 12, 24, 36, and 48 mo (i.e.,  $t_2 = 24, 36, 48, 60$  mo) are shown in Table 5.

As expected, the EVSI converges toward the EVPI as follow-up time increases, and there is good agreement

between the 2 methods. The model-averaged EVSI values for additional follow-up are greater than the Weibull model EVSI (Table 4), which reflects the additional value in reducing model as well as parameter uncertainty. The GAM method is approximately 8000 times faster than the nested Monte Carlo method.

*Expected net benefit of sampling.* The net value of additional data collection can be quantified by computing the expected net benefit of sampling (ENBS).<sup>32</sup> In the context of an ongoing study, the ENBS is the difference between the EVSI for collecting additional data between  $t_1$  and  $t_2$  and the expected cost of continuing the study and potential health benefits foregone if approval is withheld. When the ENBS is positive, it is worthwhile to continue the study and collect more data before making an adoption decision.

If the adoption decision is reversible, then there are 2 decision options given that the new technology is expected to improve net health benefits: “approval with research” (AWR), which refers to approval while additional data are being collected, or “only in research” (OIR), which means a decision to approve or reject is



**Figure 2** Marginal benefit ( $MB_{EVSI}$ ), marginal cost of “approval with research” ( $MC_{AWR}$ ) and, marginal cost of “only in research” ( $MC_{OIR}$ ) given different durations of additional follow-up. Estimates are based on the model-averaged EVSI analyses for the increasing hazard data set (left) and decreasing hazard data set (right), trial costs of 5 life-months per month, 5 new patients receiving treatment each month, and a decision time horizon of 10 y.

withheld until additional data have been collected.<sup>6</sup> An adoption decision may also be reversible with a cost, in which case the EVSI for AWR will be lower than for OIR.<sup>7</sup> For example, some irrecoverable costs, such as high initial treatment costs that are offset only by later health benefits, may be avoided if treatment initiation could be delayed until additional data have been collected.<sup>33</sup> If these avoidable costs are large, OIR may potentially be more appropriate than AWR, even if the decision is reversible. OIR may also be recommended if the new technology is not expected to improve net health benefits, but there is value in collecting additional data. If the adoption decision is irreversible, or approval would mean that further research could not be conducted, then AWR is not available and OIR may be the only option. In these circumstances, opportunity costs, in terms of potential net health benefits foregone, will be incurred while the research is being conducted if the new technology is expected to improve net health benefits.

Establishing population ENBS requires an assessment of the number of current and future patients who may benefit from additional data collection over the decision relevance time horizon.<sup>34</sup> The cost of continuing an ongoing study will primarily consist of variable (per patient) costs, including marginal incremental treatment

costs and marginal reporting costs. Fixed study costs that have already been incurred will not affect the decision to continue the ongoing study or not.

Figure 2 illustrates that if AWR is recommended, the marginal benefit in terms of population model-averaged EVSI equals the marginal cost of continuing the trial at 47 and 50 mo of additional follow-up for the increasing and decreasing hazard data sets, respectively. These are the time points at which the ENBS is at a maximum. If OIR is recommended, the ENBS is at a maximum when the marginal benefit of delaying the decision until more data have been collected equals the marginal cost of continuing the trial and withholding approval, which is at 20 and 24 mo of additional follow-up for the increasing and decreasing hazard data sets, respectively.

## Discussion

EVSI is useful not only for informing the design of a future trial but also for deciding whether an ongoing study should continue in order to collect additional data before making an adoption decision. This article is the first to set out generic EVSI algorithms for survival data from an ongoing trial with or without accounting for survival model uncertainty. The EVSI algorithms generalize

to any decision context in which structural uncertainty is present, provided that the analyst is able to derive probability weights for the competing scenarios.

### *Strengths and Limitations*

The nonparametric regression-based method is fast and straightforward to implement, even when we include consideration of model uncertainty. In fact, extending the method to include model uncertainty does not increase the complexity or computation time. The nested Monte Carlo procedure, on the other hand, is extremely computationally demanding when we include model uncertainty.

Although we considered only 2 treatment options in the synthetic case study, the EVSI methods described in this article extend to any number of treatment options that are being compared. This requires the generation of data for each treatment arm for which additional data will be collected conditional on samples from the distribution of the model parameters, which could be drawn from independently fitted survival models or from a joint model based on proportional hazards or an accelerated failure time assumption. The net benefit for each treatment arm can then be regressed on the treatment arm-specific summary statistics of the generated data, and EVSI can be computed the usual way following the algorithms in this article.

When a large part of the relevant time horizon is unobserved, the clinical plausibility of the survival extrapolations is often of greater importance than the mathematical fit to the observed data.<sup>14</sup> Deriving prior model probabilities from purely statistical measures such as AIC may therefore not always be appropriate when data are immature, since these measures do not reflect the plausibility of the extrapolations.<sup>8</sup> This became evident in the synthetic case studies, as the AIC-based prior model probabilities of the log-normal and log-logistic models were similar to those of the Weibull and Gamma models for the increasing hazard data set, despite the fact that the former 2 models do not allow for monotonically increasing hazards and therefore cannot capture the true underlying hazard pattern.

Similar to previous work,<sup>9,35</sup> we have viewed model uncertainty in terms of a discrete model space, which can be addressed by model averaging. An alternative view on model uncertainty could involve indexing candidate models within a continuous model space, using a single very flexible model that includes all the models the analyst believes plausible. For example, the generalized F distribution includes most commonly used parametric survival distributions as special cases.<sup>36</sup> In this case, the EVSI algorithms in this article would reduce to the

single-model case. This approach, however, requires the specification of a prior that appropriately reflects uncertainty in choosing between alternative functional forms within the flexible model, which may not be straightforward. Flexible models such as the generalized F distribution, Royston-Parmar spline models, or fractional polynomials are also prone to overfitting and may not always provide reliable predictions of mean survival, particularly when data are immature.<sup>9</sup>

For the purposes of describing the new method, we assumed that the survival distribution in the future unobserved time period is the same as in the observed period. This is a simplifying assumption that may not hold in real-life settings. Most importantly, we may have additional uncertainty about the postobservation period that is not captured by the uncertainty encoded in the survival model probability distribution. For example, the duration of the treatment effect is conditional on multiple factors such as the biological effect mechanism, treatment-stopping rules, compliance, and side effects.<sup>37</sup> The extrapolation of trial data may therefore have to be supplemented with external evidence<sup>38</sup> and assumptions about disease progression and mechanisms of action of the treatments that reflects additional knowledge and uncertainty. This typically involves eliciting expert opinion.<sup>39</sup> We also did not consider flexible parametric models such as Royston-Parmar spline-based models<sup>40</sup> or mixture cure models<sup>41</sup> in the synthetic case studies. Although the EVSI methods described in this article apply equally to any survival distribution and underlying assumptions (including those regarding the duration of the treatment effect), they require the generation of plausible data sets that obey all the model rules, which may not be straightforward for complex study designs. This is a common limitation of existing EVSI methods, and more research in this area may be needed.

In the synthetic case studies, we assumed all patients had the same follow-up at  $t_1$ . In clinical trials, patients are usually recruited over a period of time, which means the individual follow-up times will vary at  $t_1$ . In these circumstances, additional follow-up will provide more information not only about the tail of the survival curve (from patients who were enrolled early) but also about the central part (from patients who were enrolled later).

We did not consider sequential trial designs,<sup>42</sup> which require EVSI to be recalculated after each observation and to account for all the possible ways in which future patients may be allocated to the trial arms or when to stop the trial.<sup>34</sup> This can give rise to a large number of subproblems that may have to be solved using dynamic programming methods, which can be computationally very demanding.

### Policy Implications

Immature evidence leads to a high level of decision uncertainty, which may result in the uptake of technologies that reduce net health benefit. The decision-making context in which trials are ongoing and evidence is immature is particularly pronounced for new oncology drugs. The purpose of the Cancer Drug Fund (CDF) in the United Kingdom, for example, is to enable early patient access to promising new cancer drugs while allowing evidential uncertainty to be reduced through ongoing data collection. In the period between 2017 and July 2018, the National Institute for Health and Care Excellence (NICE) recommended more than half of the appraised cancer drugs through the CDF, typically because of concerns about immature survival data.<sup>43</sup>


EVSI will depend on both the study design and the decision context<sup>44,45</sup> but also on whether the trial results generalize to multiple jurisdictions<sup>46,47</sup> and whether the adoption decision can be fully implemented.<sup>48,49</sup> The benefit of additional data collection can be realized only when trial results are reported.<sup>50</sup> An assessment is therefore required of when the ongoing trial might report and at which point the adoption decision can be revisited.<sup>51</sup> Risk-sharing agreements between a manufacturer and payer may potentially modify the value of collecting additional data as well as the expected net benefit of access to a new technology.<sup>6,47,52</sup> The option to enroll more patients into an ongoing trial should also be considered if it has a positive net value.


The EVSI algorithms in this article can help decision makers determine whether early patient access to a new technology can be justified on the basis of the current evidence or whether more mature evidence is needed. Unlike most of the existing work on EVSI that primarily targets commissioners and funders of research, EVSI for ongoing trials also addresses the policy context of decision makers who do not have the remit to commission additional research.


### Acknowledgments

We thank researchers at the Centre for Health Economics in York for their comments on this work during a seminar in April 2020. We also thank the editor and 3 anonymous reviewers for their comments.

### ORCID iDs

Mathyn Vervaart  <https://orcid.org/0000-0003-2678-0132>

Mark Strong  <https://orcid.org/0000-0003-1486-8233>

Torbjørn Wisløff  <https://orcid.org/0000-0002-7539-082X>

### Data Availability

The analysis code used in this study is available from Github at <https://github.com/matverv/evsi-survival-ongoing-trial>.

### Supplemental Material

Supplementary material for this article is available on the *Medical Decision Making* website at <http://journals.sagepub.com/home/mdm>.

### References

- Schlaifer R. *Probability and Statistics for Business Decisions*. 1st ed. New York: McGraw-Hill; 1959.
- Raiffa H, Schlaifer R. *Applied Statistical Decision Theory*. Boston: Division of Research, Graduate School of Business Administration, Harvard University; 1961.
- Flight L, Arshad F, Barnsley R, et al. A review of clinical trials with an adaptive design and health economic analysis. *Value Health*. 2019;22(4):391–8.
- European Medicines Agency. Adaptive pathways. 2018. Available from: <https://www.ema.europa.eu/en/human-regulatory/research-development/adaptive-pathways>. Accessed October 19, 2020.
- European Medicines Agency. Conditional marketing authorisation. 2018. Available from: <https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/conditional-marketing-authorisation>. Accessed October 19, 2020.
- Claxton K, Palmer S, Longworth L, et al. A comprehensive algorithm for approval of health technologies with, without, or only in research: the key principles for informing coverage decisions. *Value Health*. 2016;19(6):885–91.
- Eckermann S, Willan AR. The option value of delay in health technology assessment. *Med Decis Making*. 2008;28(3):300–5.
- Gallacher D, Kimani P, Stallard N. Extrapolating parametric survival models in health technology assessment: a simulation study. *Med Decis Making*. 2021;41(1):37–50.
- Gallacher D, Kimani P, Stallard N. Extrapolating parametric survival models in health technology assessment using model averaging: a simulation study. *Med Decis Making*. 2021;41(4):476–84.
- Strong M, Oakley JE, Brennan A, Breeze P. Estimating the expected value of sample information using the probabilistic sensitivity analysis sample: a fast, nonparametric regression-based method. *Med Decis Making*. 2015;35(5):570–83.
- Heath A, Kunst N, Jackson C, et al. Calculating the expected value of sample information in practice: considerations from 3 case studies. *Med Decis Making*. 2020;40(3):314–26.
- Ades AE, Lu G, Claxton K. Expected value of sample information calculations in medical decision modeling. *Med Decis Making*. 2004;24(2):207–27.

13. Albert J. *Bayesian Computation with R*. 2. Dordrecht (the Netherlands): Springer; 2009.
14. Latimer NR. Survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making*. 2013;33(6):743–54.
15. Collett D. *Modelling Survival Data in Medical Research*. 3rd ed. London: Chapman and Hall/CRC; 2015.
16. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer Science & Business Media; 2013.
17. Bernardo JM, Smith AFM. *Bayesian Theory*. Chichester (UK): Wiley; 1994.
18. Jackson CH, Thompson SG, Sharples LD. Accounting for uncertainty in health economic decision models by using model averaging. *J R Stat Soc Ser A Stat Soc*. 2009;172(2):383–404.
19. Jackson CH, Sharples LD, Thompson SG. Structural and parameter uncertainty in Bayesian cost-effectiveness models. *J R Stat Soc Ser C Appl Stat*. 2010;59(2):233–53.
20. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Csaki F, eds. *Proceedings of the 2nd International Symposium on Information Theory*. Budapest: Akademiai Kiado; 1973. p 267–81.
21. Meng X-L, Wong WH. Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. *Stat Sin*. 1996;6(4):831–60.
22. Frühwirth-Schnatter S. Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *Econ J*. 2004;7(1):143–67.
23. Gronau QF, Sarafoglou A, Matzke D, et al. A tutorial on bridge sampling. *J Math Psychol*. 2017;81:80–97.
24. Wong JST, Forster JJ, Smith PWF. Properties of the bridge sampler with a focus on splitting the MCMC sample. *Stat Comput*. 2020;30(4):799–816.
25. Gronau QF, Singmann H, Wagenmakers E-J. Bridgesampling: an R package for estimating normalizing constants. *J Stat Softw*. 2020;92(1):1–29.
26. Jackson C. Flexsurv: a platform for parametric survival modeling in R. *J Stat Softw*. 2016;70(1):1–33.
27. Guo J, Gabry J, Goodrich B, Weber S. Rstan: R interface to stan. 2020. Available from: <https://CRAN.R-project.org/package=rstan>
28. Gelman A, Lee D, Guo J. Stan: a probabilistic programming language for bayesian inference and optimization. *J Educ Behav Stat*. 2015;40(5):530–43.
29. Lunn D, Spiegelhalter D, Thomas A, Best N. The BUGS project: evolution, critique and future directions. *Stat Med*. 2009;28(25):3049–67.
30. Wood S. Mgcov: mixed GAM computation vehicle with automatic smoothness estimation. 2020. Available from: <https://CRAN.R-project.org/package=mgcv>
31. Strong M, Oakley JE, Brennan A. Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample: a nonparametric regression approach. *Med Decis Making*. 2014;34(3):311–26.
32. Claxton K, Posnett J. An economic approach to clinical trial design and research priority-setting. *Health Econ*. 1996;5(6):513–24.
33. McKenna C, Soares M, Claxton K, et al. Unifying research and reimbursement decisions: case studies demonstrating the sequence of assessment and judgments required. *Value Health*. 2015;18(6):865–75.
34. Briggs A, Claxton K, Sculpher M. *Decision Modelling for Health Economic Evaluation*. 1st ed. Oxford (UK): Oxford University Press; 2006.
35. Jackson CH, Bojke L, Thompson SG, Claxton K, Sharples LD. A framework for addressing structural uncertainty in decision models. *Med Decis Making*. 2011;31(4):662–74.
36. Cox C. The generalized F distribution: an umbrella for parametric survival analysis. *Stat Med*. 2008;27(21):4301–12.
37. Eckermann S. *Health Economics from Theory to Practice*. Cham (UK): Springer International Publishing; 2017.
38. Jackson C, Stevens J, Ren S, et al. Extrapolating survival from randomized trials using external data: a review of methods. *Med Decis Making*. 2017;37(4):377–90.
39. O’Hagan A. Expert knowledge elicitation: subjective but scientific. *Am Stat*. 2019;73(suppl 1):69–81.
40. Royston P, Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*. 2002;21(15):2175–97.
41. Angelis RD, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. *Stat Med*. 1999;18(4):441–54.
42. Armitage P. The search for optimality in clinical trials. *Int Stat Rev*. 1985;53(1):15–24.
43. Walton MJ, O’Connor J, Carroll C, Claxton L, Hodgson R. A review of issues affecting the efficiency of decision making in the NICE single technology appraisal process. *Pharmacoecon Open*. 2019;3(3):403–10.
44. Eckermann S, Karnon J, Willan AR. The value of value of information: best informing research design and prioritization using current methods. *Pharmacoeconomics*. 2010;28(9):699–709.
45. Willan AR, Eckermann S. Value of information and pricing new healthcare interventions. *Pharmacoeconomics*. 2012;30(6):447–59.
46. Eckermann S, Willan AR. Globally optimal trial design for local decision making. *Health Econ*. 2009;18(2):203–16.
47. Eckermann S, Willan AR. Optimal global value of information trials: better aligning manufacturer and decision maker interests and enabling feasible risk sharing. *Pharmacoeconomics*. 2013;31(5):393–401.
48. Willan AR, Eckermann S. Optimal clinical trial design using value of information methods with imperfect implementation. *Health Econ*. 2010;19(5):549–61.

49. Grimm SE, Dixon S, Stevens JW. Assessing the expected value of research studies in reducing uncertainty and improving implementation dynamics. *Med Decis Making*. 2017;37(5):523–33.
50. Rothery C, Strong M, Koffijberg HE, et al. Value of information analytical methods: report 2 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force. *Value Health*. 2020;23(3):277–86.
51. Griffin SC, Claxton KP, Palmer SJ, Sculpher MJ. Dangerous omissions: the consequences of ignoring decision uncertainty. *Health Econ*. 2011;20(2):212–24.
52. Grimm SE, Strong M, Brennan A, Wailoo AJ. The HTA risk analysis chart: visualising the need for and potential value of managed entry agreements in health technology assessment. *Pharmacoeconomics*. 2017;35(12):1287–96.