Using evidence from randomised controlled trials in economic models: What information is relevant and is there a minimum amount of sample data required to make decisions?

John W. Stevens, School of Health and Related Research, University of Sheffield

Abstract

Evidence from randomised controlled trials (RCTs) is used to support regulatory approval and reimbursement decisions. I discuss how these decisions are typically made and argue that the amount of sample data and regulatory authorities concerns over multiplicity are irrelevant when making reimbursement decisions. Decision analytic models (DAMs) are usually necessary to meet the requirements of an economic evaluation. DAMs involve inputs relating to health benefits and resource use that represent unknown true population parameters. Evidence about parameters may come from a variety of sources including RCTs, and uncertainty about parameters is represented by their joint posterior distribution. Any impact of multiplicity is mitigated through the prior distribution. I illustrate our perspective with three examples: the estimation of a treatment effect on a rare event; the number of RCTs available in a meta-analysis; and the estimation of population mean overall survival. I conclude by recommending that reimbursement decisions should be followed by an assessment of the value of sample information and the DAM revised as necessary to include any new sample data that may be generated.

Key points:

Evidence about the true values of uncertain parameters in decision analytic models can be from sources external to a clinical trial, including expert opinion.

The amount of sample data that is available and regulatory concerns over multiplicity is irrelevant when making reimbursement decisions.

A reimbursement decision should be followed by an assessment of the value of sample information.

1. Introduction

An economic evaluation involves the systematic evaluation of the relative impact of technologies on clinical outcomes, resource use, and other aspects relating to the health of patients. The aim of an economic evaluation is to make decisions regarding the most likely cost-effective technology to adopt amongst two or more competing technologies and to facilitate the development of guidance on its use in clinical practice. Health technologies include new medicinal products, medical devices, diagnostic tests, surgical procedures and health promotion activities. An economic evaluation of a medicinal product can be performed using a study-based cost-effectiveness analysis or by constructing a decision analytic model (DMA). For the purpose of reimbursement decision making, study-based cost-effectiveness analyses are of limited value because they rarely, if ever, meet the requirements of an economic evaluation [1].

Evidence about the clinical effectiveness of medicinal products will typically be available from one or more randomised controlled trials (RCTs) as well as from other sources. The primary purpose of a clinical trial is to assess the effect of a new treatment relative to a control treatment. Regulatory authorities have developed guidance to avoid conscious or unconscious biases by requiring that all

important aspects regarding the design and conduct of a clinical trial, and the main statistical analyses are specified in a protocol before the clinical trial begins [2]. A particular concern is multiplicity, which can arise as a consequence of, for example, multiple primary variables, multiple comparisons between treatments, and sequential clinical trials involving repeated analyses of accumulating data. Issues regarding multiplicity should also be of concern in the context of DAMs, although the way the multiplicity is accounted for depends on the analytical framework. In addition, sample sizes for secondary endpoints and the number of RCTs available for a meta-analysis may be relatively small, which raises the question whether there is sufficient information to estimate parameters in DAMs in order to make reimbursement decisions. In this paper, I discuss issues regarding multiplicity and sample size, and their impact on drug regulatory approval and reimbursement decisions.

In Section 2, I outline the regulatory authorities concern regarding multiplicity and the way risk is controlled using the frequentist Type I error probability. In Section 3, I present the incremental cost-effectiveness ratio and net (monetary) benefit as criteria for reimbursement decisions, and discuss how uncertain parameters in DAMs should be estimated. In Section 4, I conclude with a discussion and recommendation.

2. Drug Registration

A key feature of the design of a clinical trial is the specification of an appropriate sample size, and the standard frequentist approach to sample size determination is based on Neyman-Pearson theory. Let Θ represent the set of all possible values of an unknown population parameter θ of the distribution of a random variable X then the Neyman-Pearson approach to statistics separates the parameter space according to null (H_0) and alternative (H_1) hypotheses, which can be written in general form as:

$$H_0: \theta \in \Theta_0$$

$$H_1: \theta \in \Theta_0^c$$

where Θ_0 is a subset of the parameter space and Θ_0^c is its compliment. Thus, two types of error are possible: 1) the statistical test leads to the rejection of the null hypothesis even though it is true (Type I error), and 2) the statistical test leads to the acceptance of the null hypothesis even though it is false (Type II error) [3].

At the design stage, the conventional approach to fixed sample size determination involves a primary variable, a test statistic, $T = t(X)$, null and alternative hypotheses, the Type I error probability and the power of the test (1 minus the Type II error probability). Power is often set at 0.80 or 0.90, and it is recommended that this is made as large as possible, particularly when clinical trials are difficult or impossible to repeat. The Type I error probability is the main concern for the regulator and is usually set at 0.05 or less [2]. Multiplicity can inflate the Type I error probability. In the simple case of a clinical trial comparing two treatments, one primary variable and one hypothesis to be tested, there is no impact on the Type I error probability. In other situations, various approaches can be taken to control the Type I error probability when it might be inflated.

At the analysis stage, tests of statistical significance are usually presented using a Fisherian approach. The Fisherian approach to significance testing requires a test statistic, and the probability

distribution for the test statistic under the null hypothesis. Then, given the observed value of the test statistic, $T = t(x)$, and the null hypothesis, the probability of observing a value as extreme or more extreme than that observed is calculated i.e. $P_0\big(t(X) \geq t(x)\big)$. The resulting probability is the P-value with small values indicating that either a rare event has occurred or the null hypothesis is false [4]. When there is multiplicity, point and interval estimates of treatment effects depend on the procedure used to control the Type I error probability (the probability of rejecting at least one null hypothesis), although methods for generating point and interval estimates that are consistent with the testing procedure in that they produce unbiased estimates with the required coverage are not always available [5]. In addition, there is ambiguity over the appropriate way to interpret estimates of treatment effect associated with outcomes that were not specified as part of the primary objectives and such analyses are generally regarded as exploratory. Nevertheless, such outcomes may be important when making reimbursement decisions.

3. Reimbursement

3.1. A framework for decision making

A DAM provides a mathematical representation of the way in which changes in health over time and the resources used affect quality-adjusted life years and total costs relevant to a target patient population. A DAM can be thought of as a function where the inputs represent unknown true population parameters. Acknowledging uncertainty about the unknown true population parameters is achieved by treating them as random variables with a specified (joint) probability distribution. Generating the distribution of the outputs of the model (i.e. population mean costs and benefits) is known as uncertainty analysis in the statistical literature, although it is referred to as probabilistic sensitivity analysis in the health economic literature [6]; sensitivity analysis strictly involves assessing how uncertainty in the inputs affects uncertainty in the output [7]. An important justification for the use of probabilistic sensitivity analysis in order to correctly generate population mean costs and benefits is in the context of non-linear models.

Decisions are often made with respect to the incremental cost-effectiveness ratio, (ρ), defined as

$$\rho = (\gamma_2 - \gamma_1)/(\mu_2 - \mu_1),$$

where γ_i is the population mean cost for treatment i and μ_i is the population mean benefit for treatment i. In practice, the appropriate basis for the decision is whether net (monetary) benefit,

$$\beta(K) = K(\mu_2 - \mu_1) - (\gamma_2 - \gamma_1),$$

is positive for some given value of the acceptable cost per unit health benefit, K [8]. Also of interest, at least in terms of quantifying decision uncertainty, is the probability that $\beta(K)$ is positive such that

$$Q(K) = P\{\beta(K) > 0|\boldsymbol{x}\},$$

where $\boldsymbol{x}$ represents the sample data and $Q(K)$ is a function known as the cost-effectiveness acceptability curve. The act of treating unknown true values as random variables and making probabilistic statements about population parameters have meaning only from a Bayesian perspective [9].

3.2. Parameter Estimation

The purpose of a clinical trial and a meta-analysis of a medicinal product are to estimate relative treatments effects; baseline parameters such as those representing the response to the control treatment (e.g. log odds) are treated as nuisance parameters and are relevant only in that they are used in the estimation of study-specific treatment effects. Estimation is normally performed on an additive scale such as the log odds ratio and log hazard ratio, which allows both pooling of relative treatment effects across clinical trials and, with suitable information on the background risk in a target population, estimation of absolute effects [10] as required in an economic evaluation. Statistical analysis proceeds by specifying a model, known as a likelihood function, which links sample data to the parameters, θ, in the model; a linear predictor, $\eta = \boldsymbol{X\beta}$, where $\boldsymbol{X}$ is a design matrix of independent explanatory variables representing treatment and other potentially relevant covariates, and $\boldsymbol{\beta}$ is a vector of unknown parameters; and a link function, $g(\theta) = \eta$, which maps values of θ to values between $-\infty$ and ∞ [11]. In a generalised linear model framework the unknown parameters, $\boldsymbol{\beta}$, can be estimated from a frequentist perspective using maximum likelihood or using a Bayesian approach. However, as discussion in Section 3.1, an economic evaluation strictly requires a Bayesian perspective and the generation of a (joint) posterior distribution [12].

A Bayesian analysis synthesises two sources of information about the unknown parameters: the sample data, which expresses the relative plausibility of different values of the parameters, expressed formerly by the likelihood function, $f(x|\theta)$, and a prior density, $f(\theta)$, which represents other information that is available in addition to the sample data [13]. Bayes' theorem synthesises the two sources of information through the equation:

$$f(\theta|x) \propto f(x|\theta)f(\theta)$$

where '$\propto$' is the proportionality symbol and $f(\theta|x)$ is the posterior distribution. The proportionality symbol expresses the fact that the product of the likelihood and prior density must be scaled to integrate to one over the sample space of the parameter for it to be a probability distribution. In principle, the prior density should represent genuine prior beliefs in any Bayesian analysis and should incorporate correlation between multiple parameters. In addition, concerns over multiplicity are irrelevant from a Bayesian perspective because adjustments for multiplicity are made explicitly through the prior distribution. In practice, when there is a large amount of sample data, the data will provide the majority of the evidence about parameters and will dominate any reasonable prior beliefs so that the posterior distribution will be similar to the likelihood function. In these circumstances, the cost and effort involved in formulating a prior distribution may not be justified. Indeed, it is common for many Bayesian analyses to make use of so-called non-informative prior distributions representing a state of prior ignorance [14], sometimes without any consideration to the impact on posterior distributions. However, in the absence of sufficient sample data, posterior distributions will not represent reasonable posterior beliefs unless prior distributions represent reasonable prior beliefs. Thus, it is precisely in situations when there is limited sample data with which to estimate parameters that the Bayesian approach and the incorporation of external information allows the estimation of parameters. Although the generation of a (joint) posterior distribution is required to propagate uncertainty through an economic model, a Bayesian estimate of location will not generally be unbiased, and the extent of the bias will depend on the prior distribution. However, the posterior mean has zero expected bias and is consistent in the sense that as the sample size tends to ∞, the posterior probability that the true value is the Bayesian estimate

tends to one [15]. In what follows, I discuss the estimation of parameters using evidence from RCTs with reference to three specific examples: the estimation of a treatment effect on a rare event, the number of clinical trials available in a meta-analysis, and the estimation of population mean overall survival.

Whether there is sufficient sample data with which to estimate parameters relates not simply to the underlying sample size but also the number of events. For example, consider an RCT that provides evidence about the probabilities of a rare event in a control and experimental treatment group. Assume that events in each treatment group are possible so that the odds of an event in each treatment group is estimable, that this is the only available RCT and the objective is to characterise the uncertainty associated with the odds ratio. In the presence of zero events in one of the treatment groups, a frequentist estimate of the odds ratio and 95% confidence interval is sometimes generated by adding 0.5 to each cell, although other correction factors have been proposed that outperform adding a constant 0.5 in the context of a frequentist meta-analysis [15]. Alternatively, given that a Binomial likelihood supports values of the observed number of events $r \in \{0,1,2,\dots,n\}$, where n is the sample size, the odds ratio could be estimated in a generalised linear model framework using a Binomial likelihood with a logit link function. However, in some circumstances, frequentist estimates of log odds ratios using this approach may give rise to excessively large standard errors so that coverage tends to one rather than the nominal level. A Bayesian analysis with conventional non-informative prior distributions for the log odds associated with the control group, $N(0, 1000)$, and log odds ratio between the experimental and control treatment groups, $N(0, 1000)$, may also give rise to implausibly large values for the odds ratio. In this situation effort should be given to specifying prior distributions that reflect reasonable prior beliefs so as to eliminate implausibly extreme parameter values. Furthermore, clinical trials with zero events in each treatment group do not provide sample information about the treatment effect [11] so that all information about it would need to come from external sources.

Evidence about parameters from multiple clinical trials can be combined using meta-analysis models [12]. A fixed effect model assumes that all clinical trials provide an estimate of the same underlying treatment effect and that any differences in sample estimates are the result of sampling variation alone, and also allows for a conditional inference given the available clinical trials [16]. However, heterogeneity in treatment effects is expected between clinical trials because they will not usually follow the same protocol. A random effects meta-analysis assumes that the true underlying treatment effects in each study follow some distribution. In a Bayesian analysis it is assumed that the true study-specific treatment effects, θ_i, from study i are related but not identical (i.e. that they are exchangeable) such that:

$$\theta_i|\mu,\tau \sim N(\mu,\tau^2)$$

where μ represents the underlying population mean and τ represents the between-study standard deviation [17]. The same model is used from a frequentist perspective and there are several estimators available with which to estimate the between-study standard deviation, although there is limited information about which performs best when the number of clinical trials is small (< 5) [18] and estimates tend to be imprecise [17]. It has become common to specify uniform prior distributions for between-study standard deviations such as a $U(0,2)$ for the between-study standard deviation on the logit scale. However, in the absence of a sufficient number of clinical trials

with which to estimate the between-study standard deviation from the sample data alone, the prior distribution will not be non-informative [19]. When so-called non-informative prior distributions are used in meta-analyses with few clinical trials, they are unlikely to reflect reasonable prior beliefs with the consequence that posterior distributions will not reflect reasonable posterior beliefs and will include unreasonably large parameter values for treatment effects. Faced with such results, some analysts may decide to adopt a fixed effect model, thereby ignoring what is believed about heterogeneity *a priori* [17]. In addition, this approach asserts with probability one that $\tau = 0$, and implies that it is not possible to learn about the true value of τ with additional sample data even though there is uncertainty about its true value. A Bayesian solution would rule out implausible values using an informative prior distribution, which could be done using a weakly informative prior distribution based on a half-t distribution [19], the predictive distribution for the between-study standard deviation in a future meta-analysis [20] or by formal elicitation of experts' beliefs [21].

Overall survival often provides an important contribution to a DAM in disease areas such as cancer where interest is in estimating the population mean quality adjusted life years. However, clinical trials may be designed to test hypotheses about intermediate outcomes such as progression-free survival, follow-up may be short relative to what would be necessary to test hypotheses about overall survival and there may be few deaths. In addition, there may be other issues that affect the estimate of the relative effect of treatments on overall survival such as switching from the control treatment to the experimental treatment group, although I will not discuss this issue further. In the absence of sufficient sample data there is uncertainty regarding the true model from which the data arose, and the decision regarding relative cost-effectiveness is often sensitive to the model choice. There is an extensive literature on modelling time-to-event data, including standard parametric models that are members of the generalised F distribution [22], flexible parametric models [23], piecewise models, and mixture models which include the standard cure rate model as a special case [24], and a growing body of work in the health economic literature on fitting parametric survival functions to time-to-event data [25-29], combining evidence on time-to-event outcomes across multiple clinical trials [30-32], and on incorporating external information in addition to sample data [33, 34]. A limitation with standard parametric models is that they only capture certain shapes for the hazard function; for example, the hazard function of a generalised F distribution can be only decreasing, decreasing but not necessarily monotone and arc shaped, while the hazard function of the generalised gamma subfamily can be only monotonically increasing and decreasing, bathtub and arc-shaped. The appropriateness of each of these models depends on true model from which the data arose which requires input from clinical experts about the underlying disease process, the mechanism of action of different treatments and the expected risk of an event over time as well as from the data analyst when using criteria such as measures of relative goodness-of-fit between models.

It is not sufficient to attempt to fit complex parametric survival models such as a four-parameter generalised F distribution unless there is reasonable clinical justification for it, and attempting to do so based on the sample data alone with limited follow-up and events may simply result in a model with parameters that fail to converge whether or not it is the true model. When complex parametric survival models such as spline models, fractional polynomials and cure rate models are considered, there may not be sufficient sample data alone with which to estimate parameters, although this would not negate their relevance and it is important to incorporate any relevant external evidence about parameters including registry data [34] alone or in combination with elicitation of experts'

beliefs [35, 36]. Finally, there are several ways in which uncertainty about a model might be assessed: 1) A single encompassing model can be constructed based on the set of possible alternative models. Prior probabilities would then be assigned to each of the alternative models to give a prior distribution over the encompassing models. 2) For a set of possible models, the robustness of decisions can be assessed against different model choices. 3) A single model can be specified and used to consider the extent to which it is adequate for decision making [37]. In the health economic literature the focus has tended to be on the second approach.

4. Discussion

The frequentist approach to statistics involves testing a limited number of hypotheses and using procedures that control the Type I error probability. Using this approach to decide on the marketing authorisation of new medicinal products is reasonable by regulatory authorities which are risk averse. However, it is unhelpful in the context of an economic evaluation in which a decision has to be made whether to reimburse a new product; for a decision-maker whose utility function is net benefit, the decision should be based on whether there is positive expected net benefit [8].

Probabilistic sensitivity analysis of DAMs is inherently Bayesian and the Bayesian approach to statistics and its applicability in the context of making reimbursement decisions has long since been established. Sample data with which to estimate treatment effects from one or more clinical trials may have various limitations with respect to the target parameter(s): sample sizes may be small; outcomes may have been secondary in the clinical trial protocol; the duration of follow-up may be relatively short; there may be few or no events associated with each treatment group; there may be concerns over the conduct of the study; there may only be a few clinical trials available with which to conduct a meta-analysis. Nevertheless, where evidence is directly or indirectly relevant to a parameter of interest, the task of the analyst is to specify a model that can be used to express the current state of uncertainty about the parameters supported by both the sample data and external evidence. Once that has been done then it is irrelevant how much sample data is available or what status an outcome had in a clinical trial protocol; all information about the parameters of interest is contained in the posterior distribution.

Some analysts are reluctant to incorporate prior information, preferring instead to estimate parameters using maximum likelihood. In those situations when this is possible, point estimates will generally correspond to the Bayesian posterior mode and will ultimately require an assumption of asymptotic multivariate normality when approximating joint posterior distributions with sampling distributions used in probabilistic sensitivity analyses. Although such approximations may be reasonable in some situations, it is precisely in situations when there is limited sample data with which to estimate parameters that the Bayesian approach and the incorporation of external information is important to represent genuine uncertainty.

The decision whether or not to reimburse a new medicinal product involves uncertainty whether it is the correct decision. Although decision uncertainty can be quantified using a cost-effectiveness acceptability curve, this does not tell us which parameters might affect the decision and whether learning about them would be beneficial. Sculpher et al [38] discussed a framework for decision making under uncertainty, including when it can be concluded that there is sufficient evidence to make a decision about the cost-effectiveness of a new technology. A decision should be made whether to acquire more information to reduce the uncertainty associated with influential

parameters using expected value of sample information [8, 39]. It should also be recognised that the decision whether to reimburse a new medicinal product and to collect additional sample information are conditional on the model structure. If additional sample information is collected then the model structure should be reviewed and amended as necessary before repeating the process.

In principle, there is no requirement to have any sample data and we could simply make use of information in the form of experts' beliefs. Indeed, there may be value to a company which is developing a new medicinal product in doing just that before conducting an expected value of sample information analysis to identify those parts of a DAM that are most influential prior to conducting a Phase 3 clinical trial and generating any sample data.

In conclusion, all evidence from a randomised clinical trial is relevant when characterising uncertainty about parameters in an economic model irrespective of the amount of sample data that is available or the status an outcome had in a clinical trial protocol. However, it is necessary to account properly for uncertainty, including prior information to account for multiplicity, when generating a joint posterior distribution.

Compliance with Ethical Standards

No funding was provided in support of this manuscript.

There are no conflicts of interest.

The opinions presented are the opinions of the author.

References

1. Sculpher M, Claxton K, Drummond M, McCabe C: **Whither trial-based economic evaluation for health care decision making?** *Health Economics* 2006, **15:**677-687.
2. **ICH harmonised tripartite guideline. Statistical principles for clinical trials 1998. ICH.E9** [http://www.ifpma.org/ich1.html]
3. Senn S: *Statistical Issues in Drug Development.* Wiley; 2008.
4. Senn S: **Two cheers for P-Values?** *Journal of Epidemiology and Biostatistics* 2001, **6:**193-204.
5. **Points to consider on multiplicty issues in clinical trials** [http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf]
6. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, Brazier J, O'Hagan A: **Probabilistic sensitivity analysis for NICE technology assessment: not an optuional extra.** *Health Economics* 2005, **14:**339-347.
7. Oakley JE, O'Hagan A: **Probabilistic sensitivity analysis of complex models: a Bayesian approach.** *Journal of the Royal Statistical Society B* 2004, **66:**751-769.
8. Claxton K: **The irrelevance of inference: a decision-making approach to the stochastic evaluation of health care technologies.** *Journal of Health Economics* 1999, **18:**341-364.
9. O'Hagan A, Stevens JW: **Inference for the cost-effectiveness acceptability curve and cost-effectiveness ratio.** *Pharmacoeconomics* 2000, **17:**339-349.
10. Senn S: **Contoversies concerning randomisation and additivity in clinical trials.** *Statistics in Medicine* 2004, **23:**3729-3753.
11. Dias S, Sutton AJ, Ades AE, Welton NJ: **Evidence synthesis for decision making 2: A generlaised linera modelling framework for pairwise and network meta-analysis of randomised controlled trials.** *Medical Decision Making* 2013, **33:**607-617.
12. Welton NJ, Sutton AJ, Cooper NJ, Abrams KR, Ades AE: *Evidence synthesis for decision making in healthcare.* Wiley; 2012.
13. Stevens JW, O'Hagan A: **Incorporation of genuine prior information in cost-effectiveness analysis of clinical trial data.** *International Journal of Technology Assessment in Health Care* 2002, **18:**782-790.
14. Kass RE, Wasserman L: **The selection of prior distributions by formal rules.** *JAMA* 1996, **91:**1343-1370.
15. Sweeting MJ, Sutton AJ, Lambert PC: **What to addto nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data.** *Statistics in Medicine* 2004, **23:**1351-1375.
16. Hedges LV, Vevea JL: **Fixed- and random-effects models in meta-analysis.** *Psychological Methods* 1998, **3:**486-504.
17. Higgins JPT, Thompson SG, Spiegelhalter DJ: **A re-evaluation of random effects meta-analysis.** *Journal of the Royal Statistical Society A* 2009, **172:**137-159.
18. Veroniki AA, Jacjson D, Viechtbauer W, Bender R, Bowden J, Knapp G, Kuss O, Higgins JPT, Langan D, Salanti G: **Methods to estimate the between-study variance and its uncertainty in meta-analysis.** *Research Synthesis Methods* 2016, **7:**55-79.
19. Gelman A: **Prior distributions for variance parameters in hierarchical models.** *Bayesian Analysis* 2006, **1:**515-533.
20. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JPT: **Predicting the extent of heterogebneity in meta-analysis using empirical data from the Cochrane Database of Systematic Review.** *International Journal of Epidemiology* 2012, **41:**818-827.
21. Ren S, Oakley J, Stevens J: **Incorporating Genuine Prior Information about Between-Study Heterogeneity in Random Effects Pairwise and Network Meta-analyses.** *Medical Decision Making* 2018, **38:**531-542.
22. Cox C: **The generalised F distribution: An umbrella for parametric survival analysis.** *Statistics in Medicine* 2008, **27:**4301-4312.

23. Royston P, Parmar MKB: **Flexible parametric proprtional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects.** *Statistics in Medicine* 2002, **21:**2175-2197.
24. Lambert PC, Thompson JR: **Estimating and modelling the cure fraction in population-based cancer survival analysis.** *Biotatistics* 2007, **8:**576-594.
25. Latimer N: **Survival analysis for economic evaluatuions alongside clinical trials - extrapolation with patient-level data: inconsistences, limitations and practical guide.** *Medical Decision Making* 2013, **33:**743-754.
26. Ishak KJ, Kreif N, Benedict A, Muszbek N: **Overview of parametric survival analysis for health-economic applications.** *Pharmacoeconomics* 2013, **31:**663-675.
27. Bagust A, Beale S: **Survival analysis and extrapolation modelling of time-to-event clinicakl trial data for economic evaluation: An alternative approach.** *Medical Decision Making* 2014, **34:**343-351.
28. Tremblay G, Haines P, Briggs A: **A criterion-based approach for the systematic and transparent extrapolation of clinical trial survival data.** *Journal of Health Economics and Outcomes Research* 2015, **2:**147-160.
29. Gibson E, Koblbauer I, Begum N, Dranitsaris G, Liew D, McEwan P, Monfared AAT: **Modelling the survival outcomes of immuno-oncology drugs in economic evaluations: A systematic approach to data analysis and extrapolation.** *Pharmacoeconomics* 2017, **35:**1257-1270.
30. Ouwens MJNM, Philips Z, Jansen JP: **Network meta-analysis of parametric survival curves.** *Research Synthesis Methods* 2010, **1:**258-271.
31. Jansen JP: **Network meta-analysis of survival data with fractional polynomials.** *BMC Medical Research Methodology* 2011, **11**.
32. Jansen JP, Cope S: **Meta-regression models to address heterogeneity andf inconsistency in network meta-analysis of survival outcomes.** *BMC Medical Research Methodology* 2012, **12**.
33. Jackson C, Stevens J, Ren S, Latimer N, Bojke L, Manca A, Sharples L: **Extrapolating survival from randomised trials using external data: A review of methods.** *Medical Decision Making* 2017, **37:**377-390.
34. Guyot P, Ades AE, Beasley M, Lueza B, Pignon J-P, Welton NJ: **Extrapolation of survival curves from cancer trials using external information.** *Medical Decision Making* 2017, **37:**353-366.
35. O'Hagan A, Buck CE, Daneshkhah A, Eiser JR, Garthwaite PH, Jenkinson DJ, Oakley JE, Rakow T: *Uncertain judgements. Eliciting experts' beliefs.*: John Wiley & Sons, Ltd; 2006.
36. Bojke L, Grigore B, Jankovic D, Peters J, Soares M, Stein K: **Informing reimbursement decisions using cost-effectiveness modelling: A guide to the process of generating elicited prior to capture model uncertainties.** *Pharmacoeconomics* 2017, **35:**867-877.
37. O'Hagan A, Forster J: *Kendall's Advanced Theory of Statistics. Volume 2B. Bayesian Inference.* Arnold; 2004.
38. Sculpher M, Claxton K: **Establishing the cost-effectivenss of new pharmaceuticals under conditions of uncertainty - when is there sufficient evidence?** *Value in Health* 2005, **8:**433-446.
39. 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.** *Medical Decision Making* 2015, **35:**570-583.