CADTH Methods and Guidelines

Extrapolating Clinical Evidence Within Economic Evaluations

Methods Review

Authors: Doug Coyle, Alex Haines, Karen Lee

Authors and Contributors

Authors

Doug Coyle, PhD, Professor, School of Epidemiology and Public Health, University of Ottawa

Alex Haines, BSc, MSc, Manager, Health Economics, CADTH

Karen Lee, MA, Director, Health Economics, CADTH

Table of Contents

List of Tables

Key Messages

- Extrapolation is a fundamental part of many economic evaluations, though there are numerous methods of extrapolating the same dataset. The method that is most appropriate for the decision problem is contingent on several factors.

- From the literature, recommendations for best practices when it comes to the extrapolation of clinical data have been identified. These pertain to validation of output, use of external evidence, treatment waning, proportional hazards, characterization of uncertainty, reporting of results, and choice of decision model.

- Gaps remain in the literature that require further research regarding what is the minimum amount of data to extrapolate reliably and how to robustly extrapolate relative effects for multiple treatment comparisons.

Introduction

Guidelines for the economic evaluation of health care interventions recommend that the time horizon used should be long enough to capture all relevant differences in the future costs and outcomes associated with the interventions being compared. In most cases, this implies that a lifetime horizon should be adopted.[1,2] Analysts, however, frequently have to incorporate assumptions relating to the long-term clinical progression for a disease because of the limited length of follow-up in the available evidence base.[3,4] Such assumptions allow characterization of the natural history of the disease beyond the duration of evidence.[5]

To overcome the limited duration of clinical data, methods for extrapolation are used to allow the estimation of unobserved data based on establishing the relationship between observed data for a parameter (e.g., probability of survival or event) and time.[5-7] Extrapolation is typically required for both the long-term natural history of the disease and the effect of treatment on the disease course. As the price of new technologies is often either set or negotiated based on the margins of cost-effectiveness, minor changes with respect to extrapolation can have a major effect on the optimal decision.[8]

The current CADTH guidance document for economic evaluation specifies that appropriate methods for extrapolating estimated effectiveness parameters to longer-term effects should be adopted. The purpose of this report is to expand on the current guidance to provide more specific advice relating to the practice of extrapolation for clinical parameters such as transition probabilities and relative effects. However, there may be valid concerns relating to the extrapolation of cost and utility parameters outside of the original context that is not a focus of this report.[9]

This report will first provide a brief introduction of what is meant by extrapolation and its connected relationship with interpolation. Following this, the appropriate theoretical paradigm from which to base this discussion will be described and the current CADTH guidance relating to extrapolation based on this paradigm will be summarized. This foundation will allow a discussion of the current literature with respect to

extrapolation, which will provide the basis for updated guidance on extrapolation for economic evaluations within Canada.

What is Extrapolation?

Based on CADTH guidelines for economic evaluation, in most circumstances a lifetime horizon is required. As models need to reflect the change in health status of a cohort over time, consideration of how individual patients move from 1 health state to another within finite time periods are required. Models, therefore, require intervention-specific parameters, such as transition probabilities, with recognition that these values may not be constant with time.

Extrapolation refers to the need to adapt clinical data that relates to a specific limited duration to provide estimates for probabilities over the full length of the time horizon of the economic model. Extrapolation may include assumptions relating to how such probabilities vary with time or can assume that probabilities are constant with time. Within economic evaluations, parametric survival analysis is commonly adopted to estimate this relationship between time and the probability of an event.[6,7] Survival analysis is an analytic method used to estimate the expected time duration until a specific event (often mortality) occurs. The more generic term of time-to-event analysis reflects that the expected time duration may relate to any well-defined end point (e.g., disease progression, incidence of event).

The object of primary interest with survival analysis is the survival function (denoted as S_t with t representing time). The survival function is the probability that an individual will not have an event by the specific timepoint t. The survival function is nonincreasing with t bound by 0 and ∞ and S_t bound by 1 and 0. Thus, by definition, S_0 = 1 and S_∞ = 0. If, the shape of the function was known for all values of t up to ∞, then life expectancy could be estimated as simply the area under the survival function. For economic evaluation, where we wish to estimate the proportion of a cohort in specific states at different time points, we can use the survival function to estimate the probability (P) of the event of interest occurring within a discrete period or cycle. This is simply the complement of the ratio of the survival function at the end of the period and the survival function at the beginning of the period; for example, if the cycle length of the model is 1 month, the probability of an event occurring in the fifth cycle is:

$$1 - \frac{S_5}{S_4}$$

It is necessary to be clear that parametric survival analysis estimates the mathematical relationship between time and survival *only* for the period for which we have data and **not for the period beyond**. Therefore, when discussing extrapolation, it is necessary to define 3 related concepts: regression, interpolation, and extrapolation.

Consider a finite dataset of observations of 2 variables, X and Y. Regression involves estimating a mathematical relationship between 1 variable (X) and the value of the other variable (Y) by finding a shape that best fits the limited observations that you have. Interpolation is the process of using the results of the regression to estimate the value of Y from the value of X over the range of values of X you have already

observations, thus allowing estimates of Y for values of X with no observations. If the value of X is outside this range, then the process of estimating the associated value of Y is extrapolation.

The observations made within survival analyses are for specific individuals and are the observed period (t) for an individual (i) that relates to either the time until the last recorded observation for that individual or the time until an event (e.g., death). From the available observations, we can derive a plot of the Kaplan-Meier estimator (KM curve), which is a series of horizontal steps that decline over time: the X-axis will relate to time and the Y-axis to the survival probability. If we have a very large sample, the KM curve will closely approximate a "true" survival function. Given the typically limited sample size available, analysts can adopt parametric survival analysis as a regression-based approach to estimate the mathematical relationship between time and survival **for the period we have data** (from t_0 to the maximum value of t within our sample [t_{max}]). Interpolation is, in this context, the process of using the results of the parametric survival analysis to estimate the survival function at a specific time for the range of t_0 to t_{max}. The estimated survival function from the parametric survival analysis can differ from the actual KM estimator at this time point. Thus, the parametric survival function represents the underlying data-generating mechanism while the KM curve represents a single realization of this.

Survival functions can take different parametric forms. Analysis can adopt an exponential function where the rate of events is assumed constant or can adopt parametric forms of increasing complexity that allow for time-varying rates. The focus here is not to provide guidance on which parametric forms to adopt but to provide the basis by which the choice between parametric forms should be made.

Parametric survival analysis can then provide a means for extrapolation beyond the time horizon of the clinical trial evidence. By applying the results of the parametric survival analysis, it is possible to provide estimates of S_t for values of t greater than t_{max}.

Theoretical Paradigm

Before discussing best practices with respect to extrapolation, it is necessary to specify the theoretical paradigm that provides the basis for specifying such practices.

Given economic evaluation's use in facilitating health care decisions within a publicly funded health care system in Canada, current CADTH guidance is based on a social decision-making viewpoint.[1] This viewpoint is based on the premise that the decision-maker, acting on behalf of a socially legitimate higher authority, seeks to maximize the degree to which an explicit policy objective is achieved, subject to resource availability. For health-related decisions in the Canadian context, it is assumed the health care decision-maker wishes to maximize a health-related outcome, such as population health, within an exogenous budget constraint; although this may not be the sole goal of the decision-maker or the populations they serve. This is generally aligned with the stated mandates of health authorities and payers that comprise the Canadian health system. Given that decision-makers require the social legitimacy of the higher authority (e.g., government in power), decisions should be reflective of what the general population considers to be socially valuable. It is these social values that the researcher should endeavour to reflect in the evaluation.

Thus, further guidance related specifically to extrapolation must reflect the theoretical paradigm, with the aim of producing unbiased estimates of long-term costs and outcomes (e.g., life-years and quality-adjusted life-years [QALYs]) for alternative treatment options.

Current CADTH Guidance With Respect to Extrapolation

Current guidance with respect to extrapolation focuses heavily on survival (i.e., time-to-event) analysis.[1]

The guidance recognizes that economic evaluations frequently rely on parametric models to extrapolate from shorter-term parameter estimates to longer-term effects. In such situations, current guidance recommends that models follow the Survival Model Selection Process Algorithm developed by the Decision Support Unit commissioned by the National Institute for Health and Care Excellence (NICE).[10]

The guidance highlights 2 issues of central importance when considering the appropriateness of an extrapolation method: assumptions relating to the rate of events beyond the observed data for a base treatment and assumptions relating to the relative effects of other treatments beyond the treatment period.

Current guidance argues that researchers should report and justify the percentage of the estimated effect that occurs beyond the observed data. This can be measured as the ratio of the estimated incremental QALYs over the period for which clinical effectiveness data are available to the estimated incremental QALYs over the entire time horizon of the model. For example, consider the situation where a clinical trial had a duration of 3 years. The model predicts an incremental benefit in terms of QALYs of 0.2 over a 3-year period and 0.5 over the lifetime horizon. The percentage of the estimated effect that occurs during the observed data is 40% (0.2 out of 0.5). Therefore, the remaining 60% of the estimated effect occurs outside of the observed data.

Guidance also suggests that researchers should report and justify the percentage of the estimated incremental benefit that is accumulated after treatment is stopped. This can be measured as the ratio of incremental QALY gains during the period on treatment to the estimated incremental QALYs over the entire time horizon of the model. If in the previously provided example treatment is given for a maximum of 2 years and the estimated QALY gains at 2 years was 0.18, then only 36% of the incremental benefit is incurred while the patient is on treatment (0.18 out of 0.5). The remaining 64% of incremental benefit from the treatment is incurred after the patient comes off treatment.

Providing a justification regarding whether both values are clinically realistic will help researchers and decision-makers assess the suitability of the extrapolation methods. Expert judgment may be helpful in this regard.

The guidance also covers the choice of parametric model for extrapolation, the consideration of parameter uncertainty through probabilistic analysis, and assumptions relating to the duration and magnitude of the clinical effect beyond the study — often referred to as consideration of the waning of treatment effect. Although appropriate, the current guidance is limited in the extent to which it provides detailed guidance with respect to extrapolation. This report will provide enhanced guidance on these topics and incorporating the current literature on this topic within the theoretical paradigm.

Review of Current Literature

A targeted review of the literature relating to methodology for extrapolation was conducted.

Almost Complete Data

When data are complete, authors have argued there is little if any incentive to model the relationship between the survival function and time as an empirical distribution as the period of observation is available and the period for which data need to be extrapolated is limited.[11] This logic may similarly be applied where data are almost complete.

In this context, when a parametric survival analysis is conducted to allow extrapolation, the standard measures of fit typically advocated for such analyses (Akaike information criterion [AIC] and Bayesian information criterion [BIC]) may have value.[5] A more useful approach to adopt in situations where data are almost complete, however, is to compare the modelled estimates of life-years and QALYs for the specific intervention with estimates from the trial data (this is referred to as the restricted means approach).[12,13] The restricted means approach involves estimating survival time empirically based on the area under the survival curve from time 0 to a specified time point t^*. Ideally, t^* is prespecified, with further analysis assessing data maturity potentially leading to a revised value, t^*_{final}. Thus, a comparison with the modelled estimates of life-years and QALYs up until t^*_{final} provides a useful validation check.

In progressive diseases, the probabilities of disease progression and the probability of death need to be estimated. The time without progression or death is often referred to as progression free survival (PFS) and the time to death is often referred to as overall survival (OS). For advanced cancer models, PFS data may be mature enough to be considered complete, but this may not hold for OS. Thus, comparing the modelled estimates of both PFS and OS up to t^*_{final} with the restricted means estimate, as previously described, would be pertinent. However, methods will be required to obtain longer-term probabilities relating to mortality through extrapolation.[14]

Sufficient Data to Extrapolate

When data, with respect to the timing of events, are incomplete it is necessary to recognize that estimates of the survival function post the period for which data are unavailable are unreliable no matter what approach is adopted for modelling the relationship between time and the survival function.[15] Discussion relating to methods of extrapolation for short-term clinical data tend to focus on whether there are sufficient data from which to meaningfully extrapolate for long-term outcomes and, if so, the appropriate basis for such extrapolation.

With respect to the former issue, the need to consider whether there is sufficient data to extrapolate may be obvious. Authors state that for there to be a degree of confidence in the predicted shape of a survival function, there needs to be a sufficiently large proportion of the survival function that relates to the period of observation.[15,16] However, there is no suggested definition for how to determine what constitutes a sufficiently large proportion. Simulation studies exploring this issue may be highly beneficial in determining

when there is sufficient data to attempt extrapolation. It is likely that what amounts to sufficient data will be context specific, varying by factors such as disease and treatment effect size.

Choice of Survival Model

When data with respect to the timing of events is incomplete, but judged to be sufficiently mature to consider extrapolation, the key issue is how to choose between alternative survival functions. Standard guidance is that models for survival extrapolation should be parsimonious but not too parsimonious, however, this is not the basis for choosing between alternative functions.[11]

The different parametric methods that are commonly used to estimate survival functions (e.g., exponential, Weibull, Gompertz, Generalized gamma, log-logistic, and log-normal distributions) differ in terms of assumptions relating to the shape of the distribution of survival times. More complex but more flexible functional forms such as restricted cubic spline model and cure models have been suggested; these avoid some of the restrictive assumptions related to the more typically used models.[7,17-20] A purported advantage of such complex models is improved model fit.[21] While spline models may be attractive in recognizing the nonstandard shape of both hazards and hazard ratios for the period covered by the trial duration, there is the potential that the modelling of long-term outcomes is based on extrapolation from the tail of the KM curve and thus derived from a lesser volume of data.

NICE has developed a model selection process algorithm that was referenced in the CADTH economic evaluation guidelines.[1,5] The algorithm refers to exponential, Weibull, Gompertz, log-logistic, and log-normal as "standard" parametric models and suggests that they should be used unless they are found to be unsuitable. The suitability of models should be assessed through statistical fit and plausibility. The need to balance statistical fit and clinical plausibility is advocated by other authors.[19,22]

The algorithm is a useful basis to further explore the issue of model selection. Criticisms of the algorithm relate to the reliance on statistical fit as a criterion for model selection, too much focus on "standard" models, and lack of detail on what is meant by plausibility.[11,23]

For the purposes of extrapolation, comparisons of the statistical fit (e.g., Akaike information criterion [AIC] and Bayesian information criterion [BIC]) are of questionable value.[11,24] Given that the tails of a survival function are not routinely covered by clinical trials, models with equally good fit can give very different (especially extrapolated) survival estimates and thus impact estimated cost-effectiveness ratios.[24-27] The measures of fit merely provide estimates of how well the potential survival function fits the data for which we already have information (i.e., they relate to interpolation not extrapolation).

The algorithm states that standard survival models should be considered first, and only if they are unsuitable should other models be explored. This follows the general guidance that models should be parsimonious but not too parsimonious, but little detail on what determines suitability is provided.

The NICE algorithm suggests that the choice of function should be based partially on plausibility and partially on the coherence of the modelled survival function with external data.[5,10] However, limited guidance on what this involves is provided. Given the limited applicability of statistical fit measures in this context, external

evidence should provide the basis for either informing the choice of long-term survival function or informing the parameters within the function.[15,28] This approach is argued to allow consideration of the entirety of the available data.[29] External evidence can relate to the formal inclusion of expert opinion, adopting established elicitation techniques such as SHELF.[15,23,30] Inclusion of external evidence can also focus on external data; for instance, registry data relating to long-term survival for the specific disease of interest can guide the shape of the long-term survival curve.[31] Similarly, general population data can guide both the shape of the long-term curve and provide upper bounds to the survival function.[15,25] In circumstances where disease-specific mortality is low, extrapolation can be based on the shape of the curve of age-specific mortality. In addition to providing more valid estimates of long-term survival, studies have demonstrated that combining short-term trial data with external data better reflects uncertainty.[24] Furthermore, in situations where a large and continued treatment effect is assumed, the need to use external data becomes paramount.[31]

Uncertainty Around Extrapolation

Parametric survival analysis allows for estimation of both the survival function and the uncertainty around it. A further issue of importance is how to specify uncertainty around the extrapolation from short-term data. Within the period for which data are available, the uncertainty around the parameters of the parametric survival model is a function of sample size. The uncertainty estimates obtained, however, relate to **interpolation** and not **extrapolation**. Within the extrapolated period, uncertainty must also relate to the predictive accuracy of the survival model given that the purpose of extrapolation in this context is to provide unbiased estimates of incremental effects beyond the time horizon of the clinical evidence base. Thus, there is increasing uncertainty as the survival function is estimated for time points beyond the period for which data are available.[32] The typical approach adopted is to use the same expected values and uncertainty for the parameters of the survival function (which is used to derive hazards and transition probabilities) for all time points; implicitly assuming that the uncertainty for the extrapolated period can be derived from the uncertainty around the interpolation. This approach is inappropriate in that the uncertainty around unobserved data should be considered both separately from and greater than the uncertainty around observed data.

To be consistent with the nature of the data available, and to mirror related disciplines such as sociology, reliability, and environmental analysis, parameters related to the probability of events in the extrapolated period should be considered distinct from similar parameters for the interpolated period; furthermore, the uncertainty with respect to the probability of events occurring should be assumed to increase with time away from the observed period.[33-35]

Economic evaluations that incorporate value of information analyses frequently conclude that the greatest information value relates to estimates of transition probabilities and clinical effectiveness.[36,37] However, if such analyses do not distinguish between observed and unobserved periods, they may be reaching incorrect conclusions. When uncertainty regarding parameters based on unobserved versus observed data are considered distinct, a more accurate value of repeating short-term studies versus studies of longer duration will be obtained.

Proportional Hazards Assumption

Most survival models make the assumption of constant proportional hazards between treatment alternatives, although few if any test the validity of this assumption.[5,8,24] Adoption of the proportional hazards assumption is likely due to its current preponderance within the medical literature and its favour within such entities as the CONSORT statement.[8,38] This is despite consistent concerns over its adoption within the statistics literature.[39,40]

Within economic evaluation, however, we need to consider the relevance of constant proportional hazards over the lifetime horizon of the model and not the short term of a clinical trial.[11] The assumption of proportional hazards is unlikely to hold long term in most cases.[24] Caution has been expressed with assuming proportional hazards with new interventions and this assumption seems particularly unconvincing when comparing treatments with different mechanisms of action.[11,41] Furthermore, in situations where the study population is heterogeneous and the relative treatment effects vary by patient characteristics, the proportional hazards cannot hold because the characteristics of the alive population will change with time.[11]

In most circumstances, the relative hazard ratio will vary by time.[24,42] The relationship between the relative hazard ratio with time will likely take a U shape.[24] In the immediate posttreatment commencement stage, treatment choice will have little effect on hazards and in some cases possibly a negative effect (e.g., surgery). For the next period, the relative hazard ratio is likely to decline monotonically — this is referred to as the acceleration phase and is frequently the main period covered by randomized controlled trials. After this period, the relative hazard ratio is likely to trend toward 1, which represents the deceleration phase. Thus, in such circumstances, the estimated hazard ratio within a trial is a function of follow-up duration, not a measure of the true effect size. To address this issue, it may appear relevant to ask analysts to test whether the proportional hazards assumption holds in the short (trial) term with the focus of providing evidence that it holds — not the typical approach for this analysis. In a review of submissions to NICE, no studies were found to directly test for the proportional hazards assumption.[8] Even if conducted, testing for nonproportional hazards within a trial setting may be uninformative.[43] A test for nonproportionality may fail with short-term data but be significant if longer-term data were available.[17] Thus, testing for proportional hazards for the trial period is not relevant to the assumption relating to the exploration of the relative treatment effects long term, as it only confirms if the assumption cannot be rejected for the trial duration.

The hazard ratio within a clinical trial can be seen as the "average" of time-dependent hazard ratios within the trial.[13] One approach that has been previously adopted in the literature is to assume no relative effect during the initial period but then model a constant hazard ratio after a specific time point.[44] This will lead to an estimated hazard ratio applied in the long-term that will be larger than the "average" hazard ratio if estimated for the whole trial period. This is analogous to assuming the trial period is reflective of the previously discussed initial and acceleration periods but with no assumed deceleration period. This approach is likely to lead to greater bias than the standard approach of adopting trial-based hazard ratios.

Given this, proportional hazards can only be assumed if a credible argument based on the epidemiology of disease and the mechanism of action of the interventions is provided. If a proportional hazards assumption is made, then comparison of the differences in life-years and QALYs from the restricted means approach

(using the previously detailed approach) and the modelled differences up until t^*_{final} will provide a useful validation check; providing insight into by how much the approach to extrapolation may impact estimates of long-term outcomes. Furthermore, if proportional hazards are to be assumed, then the estimated hazard ratio must be consistent with the survival function (i.e., it must come from the parametric model and not from the clinical trial as these will not have the same numerical value).[8]

Alternatives to the Proportional Hazards Assumption

Given the previously discussed concerns related to the proportional hazards assumption, analysts may choose to estimate the survival functions for each treatment independently. In such cases, analyses should plot the relationship between relative hazards over time to check the plausibility of the implied relative effects. A U-shaped relationship would be expected and deviations from this would require a coherent argument. Furthermore, if a U-shaped relationship is found, then there must be face validity with respect to the duration of the acceleration and deceleration periods.

In this circumstance, comparison of the differences in life-years and QALYs from the restricted means approach using the previoulsy described approach and the modelled differences up until t^*_{final} will again provide a useful validation check.

Waning of Treatment Effects

Given the short-term nature of clinical trial data, in circumstances where either treatment will be delivered for a time horizon of longer than the trial duration, or that a treatment is assumed to have a continued effect on event rates postdelivery, assumptions relating to the continuation of long-term treatment effect must be considered carefully.[14,28]

It has been suggested that the change in proportional hazards over time within a trial could be modelled to determine the likely change over an extended time period.[24] However, such modelling is unlikely to be helpful as trials are unlikely to cover a sufficient period to model the time dependence of the estimate of proportional hazards from which to extrapolate long-term clinical effects.

Both CADTH and NICE have suggested 3 alternative scenarios to consider with respect to waning of treatment effect: no waning of effect, no effect beyond trial duration, and a decline of effect up to a specified time period — the time to cessation of effect ($T_{cessation}$).[1,45]

The first 2 options are unlikely to be the case as it is unlikely that there is the same continued effect long-term or that there is no effect beyond the trial duration. A further suggestion is to assume that $T_{cessation}$ is uncertain.[28,31] With appropriate bounds around this variable, this suggestion will mimic the other 3 approaches with $T_{cessation} = 0$ representing no effect beyond trial duration, $T_{cessation} = \infty$ representing no waning of treatment effect, and all other values of $T_{cessation}$ representing the third scenario.

Multiple Treatment Comparisons

There is little guidance in the literature with respect to extrapolation of treatment effects when considering multiple treatment options, and there appears to be limited alternatives to adoption of the proportional

hazards assumptions.[24] It is also unclear how such analyses can adequately consider waning of treatment effects. Thus, given the validity concerns relating to these assumptions, analyses with multiple treatment comparisons requiring extrapolation of treatment effects must be considered with caution. A scenario analysis comparing a new treatment explicitly to the treatment(s) that it is compared to in a randomized controlled trial that adheres to all the previously outlined recommendations would be highly informative.

Allowance for Causal Relationships

When extrapolating event rates beyond the duration of clinical evidence, economic models need to ensure that the statistical modelling techniques explicitly consider causal relationships between time, health status, treatment, and mortality.[29] Techniques such as Markov models can, but do not always, account for this explicitly; techniques such as partitioned survival models do not account for this explicitly.[29] Partitioned survival models are often argued to be advantageous as they can be created with less data than what is required for Markov models[46] and they are a more straightforward approach.[47,48] The choice of model, however, should be based on statistical appropriateness and not on data availability and simplicity of analysis.[49]

The need to consider causal relationships is especially pertinent to progressive diseases such as cancer. In such cases, if long-term data are unavailable, alternative scenarios should be considered. For example, if treatment is stopped postprogression, then event rates (such as mortality) postprogression should be considered equal across treatments unless a plausible and coherent argument for an alternative assumption is made. Thus, it should be expected that for treatments that are shown to delay progression, outcomes postprogression such as estimated QALYs should be equal to or worse than the treatment alternatives unless justification is provided.[14] It is recommended that estimated outcomes for both the preprogression and postprogression phases for each treatment alternative be reported and that any differences in the latter be justified.

Restricted Means Analysis

The restricted means approach involves estimating survival time empirically based on the area under the survival curve from time 0 to a specified time point t^*.[13,17] Ideally, t^* is prespecified, with further analysis assessing data maturity potentially leading to a revised value (t^*_{final}). Comparison of restricted means has been suggested as a more appropriate measure of treatment effects than assumptions of proportional hazards.[13,17] In progressive diseases, when time-to-event curves relating to both survival and PFS are available, the restricted means approach can be extended to allow measurement of QALYs during the period up to t^*_{final}.[50]

In the presence of near complete data, the restricted means approach can be used as a validation check if analysts choose to adopt a parametric model because the estimates of life expectancy and QALYs from both approaches should be near identical. However, the determination of what represents near complete data remains unclear. When data are incomplete, a comparison with the modelled estimates of life-years and QALYs up until t^*_{final} provides a useful validation check both for individual treatment estimates and estimated differences. Thus, in all circumstances, the restricted means approach provides a useful validation for

the interpolated period but not for the extrapolated period. One further advantage of the restricted means approach is that methods have been developed to allow the conduct of meta-analyses across clinical trials, with the potential that methods of evidence synthesis with multiple treatment comparators may be possible.[51]

Recommendations and CADTH Guidance

Based on the review of the literature, the following recommendations have been proposed as additions to the CADTH guidelines for economic evaluation. Alongside each recommendation CADTH has outlined guidance on how each issue should be addressed.

Table 1: Recommendations and CADTH Guidance

Recommendation	CADTH guidance
Recommendation 1: Estimates of modelled outcomes should be compared to empirical estimates using the restricted means approach. A comparison of the restricted means estimates with the modelled estimates of life-years and quality-adjusted life-years for the period where data are available provides a useful validation check both for individual treatment estimates and estimated differences in all circumstances. However, this is not a validation check relating to extrapolation.	Analysts should ensure that restricted mean estimates of modelled outcomes replicate the period where clinical data are available; this is a key aspect of model validation.
Recommendation 2: Further research should be conducted to provide guidance on the minimal data required for extrapolation. If there are incomplete data with respect to the timing of events, then estimates of the survival function will be unreliable. For there to be a degree of confidence in the predicted shape of a survival function, there needs to be a sufficiently large proportion of the survival function that relates to the period of observation. Simulation studies exploring this issue may be highly beneficial in determining when there is sufficient data to attempt extrapolation.	Extrapolation predictions have sometimes been poor when relying on clinical trial data. Further research is needed to note when clinical trial evidence is sufficient to provide accurate extrapolations of future benefit. CADTH notes this as an important research gap moving forward.
Recommendation 3: Consideration of external evidence should be the main criteria for choice of survival function (not statistical fit). Given the limited applicability of statistical fit measures with respect to extrapolation, external evidence should provide the basis for either informing the choice of long-term survival function or for informing the parameters within the function. External evidence can relate to the formal inclusion of expert opinion and to external data either specific to the disease of interest or relating to the general population.	Analysts should provide evidence outside of their clinical studies to justify the choice of survival function used (e.g., from long-term epidemiological studies or administrative databases). This should relate to what is known about the condition modelled, and be supplemented by evidence known about the technology(ies) under review (i.e., disease modifying, delaying disease, symptom improvement, curative). Uncertainties with associated external evidence should be thoroughly explored.

Recommendation	CADTH guidance
Recommendation 4: Uncertainty with respect to the hazards derived from the survival function should be assumed to increase with time away from the period for which data are available. Parameters related to the probability of events in the extrapolated period should be considered as distinct from similar parameters for the interpolated period and the uncertainty with respect to the probability of events occurring should be assumed to increase with time away from the observed period.	Analysts should account for the increasing levels of uncertainty of effect the further the model extends beyond the time frame for which data are available; that is, uncertainty in the interpolated (observed) period is not reflective of the uncertainty that exists when no evidence is available (the period for which data are extrapolated).
Recommendation 5: When modelling the hazards of 2 comparators, proportional hazards should only be assumed if a credible argument can be made that this will hold in the long term. In most cases, the assumption of proportional hazards is unlikely to hold in the long term. Thus, proportional hazards can only be assumed if a credible argument based on the epidemiology of the disease and the mechanism of action of the interventions is provided.	Analysts must assess whether there is a credible argument (based on the epidemiology of the disease or the mechanism of action of the interventions) to justify the assumption of proportional hazards beyond the period of time for which trial data exists. Alternative assumptions regarding the size of the HR should be considered within scenario analyses.
Recommendation 6: In instances where proportional hazards are not assumed and survival functions for treatment alternatives are estimated independently, analysts should report the relationship between relative hazards and time to check plausibility. If analysts choose to estimate the survival functions for each treatment independently, it is important to provide a plot of the relationship between relative hazards from the individual survival curves and time (**throughout the model's time horizon)**, to check the plausibility of the implied relative effects. A U-shaped relationship would be expected and deviations from this would require a coherent argument.	Where proportional hazards are not assumed and survival functions for treatment alternatives are estimated independently, it is important to provide a plot of the relationship between relative hazards from the individual survival curves and time (throughout the model time horizon), to assess the plausibility of the implied relative effects. A U-shaped relationship would be expected, consistent with waning of treatment effect. If deviations from the U-shaped relationship are used, details and analyses must be provided for justification. If the relative hazards suggest an increasing relative effect beyond the trial horizon, the model must allow for the examination of similar or attenuating effects in the extrapolation period.
Recommendation 7: Analysis should incorporate an explicit consideration of the waning of treatment effect with a plausible and coherent argument provided for the method chosen within the base case. Waning of treatment effect should be incorporated into the analysis and a plausible and coherent argument should be provided for the choice of approach. Three alternative scenarios with respect to waning of treatment effect are routinely advanced: no waning of effect, no effect beyond trial duration, and a decline of effect up to a specified time point. Given the improbability of the first 2 scenarios, the latter approach is likely the most valid. A revision to this third approach is to assume that the duration of effect is uncertain. Models should be flexible enough to allow analysts to consider all 4 approaches, with appropriate scenario analyses provided.	Analysts should ensure that models allow for the consideration of waning of treatment effect, with justification for assumptions made. At a minimum, analysts should ensure that 3 scenarios can be considered: no waning of effect (i.e., proportional hazards or increasing HR over time), no additional effect beyond trial duration (HR is 1 thereafter), and a decline of effect up to a specified time point (HR decreases either to 1 or below 1); this time point should be modifiable.

Recommendation	CADTH guidance
Recommendation 8: Given the concerns over extrapolation and the limited options with respect to extrapolation of relative treatment effects when considering multiple mutually exclusive treatment options, in such circumstances, the results of analysis should be treated with extreme caution. There is little guidance in the literature with respect to extrapolation of treatment effects when comparing multiple treatment options. Thus, analyses with multiple treatment comparisons requiring extrapolation of treatment effects must be considered with caution. A scenario analysis comparing a new treatment explicitly to the treatment(s) that it is compared to in a randomized controlled trial is useful.	When multiple relevant comparators exist, given the complexity of assessing relative treatment effects, analysts should ensure that a comparison that reflects the clinical data (e.g., trial) can be conducted to ensure validity and that it is reported as a scenario analysis. CADTH notes this is an area where further research is needed.
Recommendation 9: Analyses incorporating extrapolation must appropriately consider causal relationships between event rates and time-varying parameters. Thus, models that directly consider these relationships should be adopted. When extrapolating event rates beyond the duration of clinical evidence, economic models need to ensure that the statistical modelling techniques explicitly consider causal relationships between the probability of events and time, health status, and treatment. Models that cannot directly consider such causal relationships should be used only as scenario analyses.	Models should directly incorporate the probability of events and how they change over time and with changes in patients' statuses. A partition survival model structure is not recommended as it relies on an assumption of independence between curves. Given the inflexibility in their approach, this could preclude CADTH from conducting a robust assessment of the evidence.

HR = hazard ratio.

References

1. CADTH. *Guidelines for the Economic Evaluation of Health Technologies :* Canada *(*4th Edition*)*.; 2017.

2. National Institute for Health and Care Excellence. Guide to the methods of technology appraisal 2013. *Natl Inst Heal Care Excell*. Published online 2013. doi:10.2165/00019053-200826090-00002

3. Buxton MJ, Drummond MF, Van Hout BA, et al. Modelling in Ecomomic Evaluation: An Unavoidable Fact of Life. *Health Econ*. Published online 1997. doi:10.1002/(sici)1099-1050(199705)6:3<217::aid-hec267>3.3.co;2-n PubMed

4. Tappenden P, Chilcott J, Ward S, Eggington S, Hind D, Hummel S. Methodological issues in the economic analysis of cancer treatments. *Eur J Cancer*. Published online 2006. doi:10.1016/j.ejca.2006.08.010 PubMed

5. Latimer NR. Survival analysis for economic evaluations alongside clinical trials - Extrapolation with patient-level data: Inconsistencies, limitations, and a practical guide. *Med Decis Mak*. Published online 2013. doi:10.1177/0272989X12472398

6. Ishak KJ, Kreif N, Benedict A, Muszbek N. Overview of parametric survival analysis for health-economic applications. *Pharmacoeconomics*. Published online 2013. doi:10.1007/s40273-013-0064-3 PubMed

7. Crowther MJ, Lambert PC. A general framework for parametric survival analysis. *Stat Med*. Published online 2014. doi:10.1002/sim.6300 PubMed

8. Guyot P, Welton NJ, Ouwens MJNM, Ades AE. Survival time outcomes in randomized, controlled trials and meta-analyses: The parallel universes of efficacy and cost-effectiveness. *Value Heal*. 2011;14(5):640-646. doi:10.1016/j.jval.2011.01.008 PubMed

9. Bojke L, Manca A, Asaria M, Mahon R, Ren S, Palmer S. How to Appropriately Extrapolate Costs and Utilities in Cost-Effectiveness Analysis. *Pharmacoeconomics*. 2017;35(8):767-776. doi:10.1007/s40273-017-0512-6 PubMed

10. Latimer N. NICE DSU technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. *Sheff Rep by Decis Support Unit*. 2011;(0). http://www.nicedsu.org.uk/NICE DSU TSD Survival analysis.updated March 2013.v2.pdf.

11. Bagust A, Beale S. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: An alternative approach. *Med Decis Mak*. 2014;34(3):343-351. doi:10.1177/0272989X13497998 PubMed

12. Royston P, Parmar MKB. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Stat Med*. 2011;30(19):2409-2421. doi:10.1002/sim.4274 PubMed

13. Royston P, Parmar MKB. Restricted mean survival time: An alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013;13(1). doi:10.1186/1471-2288-13-152 PubMed

14. Beca J, Husereau D, Chan KKW, Hawkins N, Hoch JS. Oncology Modeling for Fun and Profit! Key Steps for Busy Analysts in Health Technology Assessment. *Pharmacoeconomics*. 2018;36(1):7-15. doi:10.1007/s40273-017-0583-4 PubMed

15. Jackson C, Stevens J, Ren S, et al. Extrapolating Survival from Randomized Trials Using External Data: A Review of Methods. *Med Decis Mak*. 2017;37(4):377-390. doi:10.1177/0272989X16639900 PubMed

16. Stevens JW. 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? *Pharmacoeconomics*. 2018;36(10):1135-1141. doi:10.1007/s40273-018-0681-y PubMed

17. Royston P, Lambert P. *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model*. Stata Press; 2011.

18. Hopwood P, Harvey A, Davies J, et al. Survey of the administration of quality of life (QL) questionnaires in three multicentre randomised trials in cancer. *Eur J Cancer*. 1998;34(1). doi:10.1016/S0959-8049(97)00347-X PubMed

19. Gibson E, Koblbauer I, Begum N, et al. Modelling the Survival Outcomes of Immuno-Oncology Drugs in Economic Evaluations: A Systematic Approach to Data Analysis and Extrapolation. *Pharmacoeconomics*. 2017;35(12):1257-1270. doi:10.1007/s40273-017-0558-5 PubMed

20. Whittington MD, McQueen RB, Ollendorf DA, et al. Long-term Survival and Cost-effectiveness Associated With Axicabtagene Ciloleucel vs Chemotherapy for Treatment of B-Cell Lymphoma. *JAMA Netw open*. 2019;2(2):e190035. doi:10.1001/jamanetworkopen.2019.0035 PubMed

21. Fagbamigbe AF, Karlsson K, Derks J, Petzold M. Performance evaluation of survival regression models in analysing Swedish dental implant complication data with frailty. *PLoS One*. 2021;16(1):e0245111. [doi:10.1371/journal.pone.0245111](#) [PubMed](#)

22. Jackson CH, Sharples LD, Thompson SG. Survival models in health economic evaluations: Balancing fit and parsimony to improve prediction. *Int J Biostat*. 2010;6(1). [doi:10.2202/1557-4679.1269](#) [PubMed](#)

23. Cope S, Ayers D, Zhang J, Batt K, Jansen JP. Integrating expert opinion with clinical trial data to extrapolate long-Term survival: A case study of CAR-T therapy for children and young adults with relapsed or refractory acute lymphoblastic leukemia. *BMC Med Res Methodol*. 2019;19(1):1-11. [doi:10.1186/s12874-019-0823-8](#) [PubMed](#)

24. Guyot P, Ades AE, Beasley M, Lueza B, Pignon JP, Welton NJ. Extrapolation of Survival Curves from Cancer Trials Using External Information. *Med Decis Mak*. 2017;37(4):353-366. [doi:10.1177/0272989X16670604](#) [PubMed](#)

25. Pennington M, Grieve R, der Meulen J Van, Hawkins N. Value of External Data in the Extrapolation of Survival Data: A Study Using the NJR Data Set. *Value Heal*. 2018;21(7):822-829. [doi:10.1016/j.jval.2017.12.023](#) [PubMed](#)

26. Connock M, Hyde C, Moore D. Cautions regarding the fitting and interpretation of survival curves: Examples from NICE single technology appraisals of drugs for cancer. *Pharmacoeconomics*. 2011;29(10):827-837. [doi:10.2165/11585940-000000000-00000](#) [PubMed](#)

27. Gerdtham UG, Zethraeus N. Predicting survival in cost-effectiveness analyses based on clinical trials. *Int J Technol Assess Health Care*. 2003;19(3):507-512. [doi:10.1017/S0266462303000436](#) [PubMed](#)

28. Grieve R, Hawkins N, Pennington M. Extrapolation of survival data in cost-effectiveness analyses: Improving the current state of play. *Med Decis Mak*. 2013;33(6):740-742. [doi:10.1177/0272989X13492018](#) [PubMed](#)

29. Hawkins N, Grieve R. Extrapolation of Survival Data in Cost-effectiveness Analyses: The Need for Causal Clarity. *Med Decis Mak*. 2017;37(4):337-339. [doi:10.1177/0272989X17697019](#) [PubMed](#)

30. O'Hagan A, Buck C, Daneshkhah A, et al. *Uncertain Judgements: Eliciting Experts' Probabilities.* Wiley; 2006.

31. Vickers A. An Evaluation of Survival Curve Extrapolation Techniques Using Long-Term Observational Cancer Data. *Med Decis Mak*. 2019;39(8):926-938. [doi:10.1177/0272989X19875950](#) [PubMed](#)

32. Kearns B, Stevens J, Ren S, Brennan A. How Uncertain is the Survival Extrapolation? A Study of the Impact of Different Parametric Survival Models on Extrapolated Uncertainty About Hazard Functions, Lifetime Mean Survival and Cost Effectiveness. *Pharmacoeconomics*. 2020;38(2):193-204. [doi:10.1007/s40273-019-00853-x](#) [PubMed](#)

33. Dietze MC. Prediction in ecology: A first-principles framework: A. *Ecol Appl*. 2017;27(7):2048-2060. [doi:10.1002/eap.1589](#) [PubMed](#)

34. Conn PB, Johnson DS, Boveng PL. On extrapolating past the range of observed data when making statistical predictions in ecology. *PLoS One*. 2015;10(10):1-16. [doi:10.1371/journal.pone.0141416](#) [PubMed](#)

35. Blackwell M, Honaker J, King G. A Unified Approach to Measurement Error and Missing Data: Overview and Applications. *Sociol Methods Res*. 2017;46(3):303-341. [doi:10.1177/0049124115585360](#)

36. Felli JC, Hazen GB. A Bayesian approach to sensitivity analysis. *Health Econ*. Published online 1999. [doi:10.1002/(SICI)1099-1050(199905)8:3<263::AID-HEC426>3.0.CO;2-S](#) [PubMed](#)

37. Willan AR. Clinical decision making and the expected value of information. *Clin Trials*. 2007;4(3):279-285. [doi:10.1177/1740774507079237](#) [PubMed](#)

38. Moher D, Hopewell S, Schulz KF, et al. ConSoRT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials David. *BMJ*. 2010;340:869. [doi:10.1136/bmj.c869](#) [PubMed](#)

39. Austin PC. Statistical power to detect violation of the proportional hazards assumption when using the Cox regression model. *J Stat Comput Simul*. 2018;88(3):533-552. [doi:10.1080/00949655.2017.1397151](#) [PubMed](#)

40. Aalen OO, Cook RJ, Røysland K. Does Cox analysis of a randomized survival study yield a causal treatment effect? *Lifetime Data Anal*. 2015;21(4). [doi:10.1007/s10985-015-9335-y](#) [PubMed](#)

41. Davies C, Briggs A, Lorgelly P, Garellick G, Malchau H. The "hazards" of extrapolating survival curves. *Med Decis Mak*. 2013;33(3):369-380. [doi:10.1177/0272989X12475091](#) [PubMed](#)

42. Alarid-Escudero F, Kuntz KM. Potential Bias Associated with Modeling the Effectiveness of Healthcare Interventions in Reducing Mortality Using an Overall Hazard Ratio. *Pharmacoeconomics*. 2020;38(3):285-296. doi:10.1007/s40273-019-00859-5 PubMed

43. Tremblay G, Haines P, Briggs A. A Criterion-based Approach for the Systematic and Transparent Extrapolation of Clinical Trial Survival Data. *J Heal Econ Outcomes Res*. 2015;2(2):147-169. doi:10.36469/9896

44. Davies A, Briggs A, Schneider J, et al. The ends justify the mean: Outcome measures for estimating the value of new cancer therapies. *Health Outcomes Res Med*. 2012;3(1):e25-e36. doi:10.1016/j.ehrm.2012.01.001

45. Latimer N. *NICE DSU Technical Support Document 14: Undertaking Survival Analysis for Economic Evaluations alongside Clinical Trials—Extrapolation with Patient-Level Data.*; 2011.

46. Cranmer H, Shields GE, Bullement A. A comparison of partitioned survival analysis and state transition multi-state modelling approaches using a case study in oncology. *J Med Econ*. 2020;23(10):1176-1185. doi:10.1080/13696998.2020.1796360 PubMed

47. Goeree R, Villeneuve J, Goeree J, Penrod JR, Orsini L, Monfared AAT. Economic evaluation of nivolumab for the treatment of second-line advanced squamous NSCLC in Canada: A comparison of modeling approaches to estimate and extrapolate survival outcomes. *J Med Econ*. 2016;19(6):630-644. doi:10.3111/13696998.2016.1151432 PubMed

48. Smare C, Lakhdari K, Doan J, Posnett J, Johal S. Evaluating Partitioned Survival and Markov Decision-Analytic Modeling Approaches for Use in Cost-Effectiveness Analysis: Estimating and Comparing Survival Outcomes. *Pharmacoeconomics*. 2020;38(1):97-108. doi:10.1007/s40273-019-00845-x PubMed

49. Brennan A, Chick SE, Davies R. A taxonomy of model structures for economic evaluation of health technologies. *Health Econ*. 2006;15(12):1295-1310. doi:10.1002/hec.1148 PubMed

50. Glasziou PP, Cole BF, Gelber RD, Hilden J, Simes RJ. Quality adjusted survival analysis with repeated quality of life measures. *Stat Med*. 1998;17(11):1215-1229. doi:10.1002/(SICI)1097-0258(19980615)17:11<1215::AID-SIM844>3.0.CO;2-Y PubMed

51. Wei Y, Royston P, Tierney JF, Parmar MKB. Meta-analysis of time-to-event outcomes from randomized trials using restricted mean survival time: Application to individual participant data. *Stat Med*. 2015;34(21):2881-2898. doi:10.1002/sim.6556 PubMed