PharmacoEconomics
https://doi.org/10.1007/s40273-024-01425-4

PRACTICAL APPLICATION

MPES-R: Multi-Parameter Evidence Synthesis in R for Survival Extrapolation—A Tutorial

Ash Bullement[1,2] · Mark Edmondson-Jones[2,3] · Patricia Guyot[4] · Nicky J. Welton[5] · Gianluca Baio[6] · Matthew Stevenson[1] · Nicholas R. Latimer[1,2]

Accepted: 11 August 2024

Abstract

Survival extrapolation often plays an important role in health technology assessment (HTA), and there are a range of different approaches available. Approaches that can leverage external evidence (i.e. data or information collected outside the main data source of interest) may be helpful, given the extent of uncertainty often present when determining a suitable survival extrapolation. One of these methods is the multi-parameter evidence synthesis (MPES) approach, first proposed for use in HTA by Guyot et al., and more recently by Jackson. While MPES has potential benefits over conventional extrapolation approaches (such as simple or flexible parametric models), it is more computationally complex and requires use of specialist software. This tutorial presents an introduction to MPES for HTA, alongside a user-friendly, publicly available operationalisation of Guyot's original MPES that can be executed using the statistical software package R. Through two case studies, both Guyot's and Jackson's MPES approaches are explored, along with sensitivity analyses relevant to HTA. Finally, the discussion section of the tutorial details important considerations for analysts considering use of an MPES approach, along with potential further developments. MPES has not been used often in HTA, and so there are limited examples of how it has been used and perceived. However, this tutorial may aid future research efforts exploring the use of MPES further.

Key Points for Decision Makers

Multi-parameter evidence synthesis (MPES) allows for external evidence to be combined with trial data to inform survival extrapolations.

This tutorial presents an introduction to MPES for health technology assessment (HTA), via two different specifications: Guyot's MPES and Jackson's MPES.

Despite their value in terms of methodological progress and the ability to leverage multiple evidence sources in one overall model, MPES approaches do not guarantee accurate extrapolations, as different applications will give different extrapolations, and so describing and justifying assumptions is crucial.

1 Introduction

Survival extrapolation often plays an important role in health technology assessment (HTA), and there are a range of different approaches available that can be used to produce estimated survival curves [1, 2]. Approaches that can leverage external evidence (that is, data or information collected outside the main data source of interest) may be helpful, given the extent of uncertainty often present when determining a suitable survival extrapolation. The use of external evidence to support extrapolation of survival outcomes for HTA is recognised as an area requiring further research but is not a new one [2, 3].

A recent systematic review conducted by some of the authors identified methods leveraging external evidence for survival extrapolation from as early as 2005 [4, 5]. One of the methods identified in this systematic review was the

✉ Ash Bullement
abullement@deltahat.com

1 School of Medicine and Population Health, Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, UK

2 Delta Hat, Nottingham, UK

3 Population Health Sciences, University of Leicester, Leicester, UK

4 Sanofi, Gentilly, France

5 Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

6 Department of Statistical Science, University College London, London, UK

Published online: 29 August 2024

approach of Guyot et al. [6], which used a multi-parameter evidence synthesis (MPES) approach to extrapolate survival over a lifetime horizon. This MPES approach brought in evidence from multiple sources, and the authors used a case study in squamous cell carcinoma of the head and neck (SCCHN). More recently, Jackson presented an alternative MPES approach [7]. Given that HTA involves bringing together different sources of evidence to inform decision making, an MPES approach would seem particularly relevant, given that the method itself directly combines multiple sources of evidence, which was not a feature of the other methods identified in the systematic review.

A small number of other studies have applied an MPES approach to address different research questions, using Guyot's MPES specifically (since Jackson's MPES was only recently published). Vickers (2019) explored a variety of extrapolation techniques (including MPES) using a published database study that provided data for a cohort of patients with non-small-cell lung cancer (NSCLC) [8]. Chaudhary et al. [9] also explored different extrapolation methods for an NSCLC population, but made use of individual patient data (IPD) from two clinical trials of nivolumab: CheckMate-017 (NCT01642004) and CheckMate-057 (NCT01673867). Despite the publication of these more recent studies, the only statistical analysis code available in the public domain from which to execute Guyot's MPES is provided in the original publication appendix, developed using WinBUGS [10]. As the model fitting process may be considered complex for non-programmers, this may have contributed (at least in part) to the limited use of MPES for HTA.

In this tutorial, we introduce the MPES approach in general, and explain the key features of Guyot's and Jackson's MPES approaches. We then present a user-friendly, publicly available operationalisation of Guyot's MPES using the R interface for Stan: *'rstan'* [11, 12]. Following this, we compare Guyot's MPES to Jackson's MPES using two case studies, with the intention of providing researchers with further details of an MPES approach when applied in different contexts (including where the approaches can produce unexpected results). The main aim of this tutorial is to introduce the MPES approach in general and provide a means of running both Guyot's and Jackson's MPES approaches in the same, readily accessible software so that it can be more easily used for HTA purposes and further research.

2 What is MPES?

The term 'MPES' (multi-parameter evidence synthesis) can be used to describe various methods to combine evidence from multiple sources that may inform estimation of different model parameters that are linked in some way [13]. An MPES approach offers a means of combining multiple sources of relevant evidence to inform estimates of survival outcomes. These sources of evidence could include, for example, data collected from a clinical trial, a disease registry, and/or clinical expert opinion. In principle, producing a survival model that is informed by multiple sources of evidence is likely to be valuable to decision-makers faced with uncertainty in determining a plausible survival extrapolation.

An MPES approach can be considered as a comprehensive method that combines all inputs to produce a final survival model (with all key assumptions related to external evidence 'baked into' the model itself, reflecting additional 'ingredients')—see Fig. 1. However, as with baking, it is important that the ingredients are thoughtfully combined and are not 'lumped in' without due care and attention. Each source of evidence is expected to tell us something specific to that source, and therefore there needs to be careful consideration for how it could be used in the model.

3 Guyot's MPES

In their article, Guyot et al. present an illustration of how an MPES approach could be undertaken, noting several possible sensitivity analyses and some important assumptions [6]. Guyot's MPES uses four evidence sources: patient-level survival data (e.g. from a clinical trial), conditional survival data (e.g. from a disease registry), background mortality (BGM) data (e.g. from population statistics), and expectation concerning treatment effect over time (e.g. based on expert opinion). Full details of Guyot's MPES can be found in the original study [6], but a brief overview is provided in this tutorial. More information about the specific SCCHN case study that motivated Guyot's MPES, including the sources of each of the inputs, is provided in theElectronic Supplementary Material (ESM).

Guyot's MPES is estimated as a common model (i.e. one overarching model that is informed by all relevant inputs simultaneously). The starting point for the common model is a cubic spline model fitted to the ('true' or 'recreated') patient-level survival data with two components. One of these components represents the log cumulative hazard of death for the control group, and the other captures the effect of the active intervention. The external sources introduced into the model are data from a disease registry (for conditional survival) and for BGM—both influence the estimation of survival for the control group, i.e. the first component. Information is also added for the treatment effect, expressed based on a hazard ratio (HR), i.e. the second component. Knot locations for the cubic spline model are specified by the analyst, and do not need to be strictly located within the follow-up period of the patient-level survival data.

To 'add in' these additional pieces of information, Guyot et al., constructed likelihood functions for the external evidence sources, which were expressed in terms of the

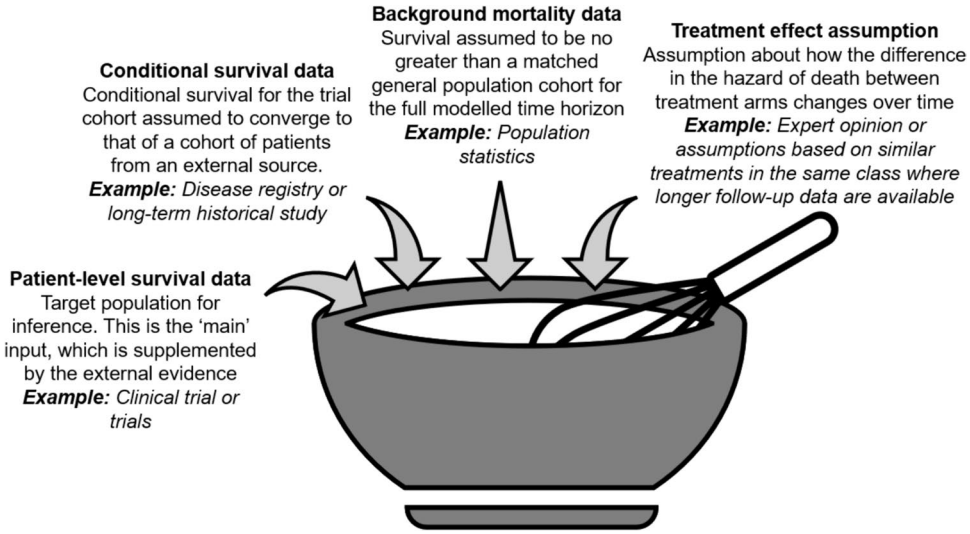

Fig. 1 Illustration of the MPES concept. *MPES* multi-parameter evidence synthesis

parameters of the model, and also determined the likelihood function for the patient-level data. This means that the authors established a common model that can be estimated using all the included evidence sources at the same time. While the original study makes reference to the data being added incrementally, it should be noted that the final model specification is for a common model, so there is no particular 'order' in which the evidence sources are added.

4 Jackson's MPES

Like Guyot's MPES, Jackson's MPES involves fitting a common model accounting for all input data and assumptions. However, Jackson's MPES uses an M-spline function, rather than a cubic spline function, as used by Guyot et al. A detailed explanation of Jackson's MPES is provided in the original publication, which includes an appendix that details the differences between M-splines and natural cubic splines [7]. The full publication highlights the availability of the '*survextrap*' R package, which can be used to execute Jackson's MPES, and examples can be accessed via this link: https://chjackson.github.io/survextrap/articles/examples.html.

Jackson's MPES assumes that the following evidence sources are available:

- Either 'true' or 'recreated' patient-level survival data—for example, from a clinical trial. The number of knots and knot locations can be specified by the analyst, but by default these are automatically selected by the '*survextrap*' function within the '*survextrap'* R package, based on quantiles of the event times in the data.
- At least one aggregate-level external dataset, providing counts of survivors over arbitrary time periods.

Jackson's MPES includes three 'special mechanisms' that may be useful for analysts to consider; these are as follows: relative survival, mixture-cure modelling, and incorporation of treatment effect waning. Importantly, these special mechanisms can be considered together – in fact, the relative survival option should be included if a cure model is fitted (so that the end model does not reflect cured people as immortal).

- The relative survival mechanism allows the user to focus the model fitting process on *excess* mortality (that is, mortality specifically related to the disease of interest, rather than *all-cause* mortality). This means that BGM data can be included in Jackson's MPES, but this is not mandatory.
- Mixture-cure modelling may be useful for estimating survival for potentially curative (or 'functionally curative') treatments associated with a survival plateau.
- Treatment effect waning may be helpful to explore the relationship between modelled survival estimates for multiple treatment groups, though as discussed later in this tutorial, Jackson's MPES considers treatment effect waning as a post hoc adjustment, rather than an input to the MPES model itself.

5 How do Guyot's and Jackson's MPES approaches compare?

Table 1 summarises the main similarities and differences between Guyot's MPES and Jackson's MPES. For more details, please refer to Jackson's article (see 'Additional File 1').

6 Implementation of Guyot's and Jackson's MPES

6.1 Preparing the Programming Code for Guyot's MPES

The statistical analysis software package R is recognised as one of the most popular software packages for the purpose of producing survival extrapolations for HTA (alongside others, such as Stata and SAS [14–16]). There is increased interest in the use of R for a variety of analyses associated with HTA, advocated by the 'R for HTA' academic consortium, given that R is freely available and open source [17]. Unlike Guyot's MPES, the more recent MPES approach of Jackson has been developed with a corresponding R package: '*survextrap'* [7]. Without an R interface, many HTA analysts may struggle to execute Guyot's MPES, and understand how it compares to Jackson's MPES.

Using the original code and input files, Guyot's MPES was re-programmed into R using the *'rstan'* package [11]. The re-programmed code was used to run the original model specification and the outputs were verified, both via digitisation (comparing the plotted survival curve outputs from the original study versus the re-programmed code)

Table 1 Comparison of Guyot's and Jackson's MPES

Guyot	Jackson
A restricted cubic spline model is used to represent the log-cumulative hazard over time. This can, theoretically, produce a negative hazard	An M-spline model is used to represent the hazard over time. This cannot produce a negative hazard
Knot locations are specified by the analyst. The original case study describes the approach taken to select knot locations, though no formal process is provided or included as part of a statistical software package	Knot locations are determined by default in the statistical software package, though these can be overridden by user inputs
Both methods rely on there being one source of patient-level data, usually from a clinical trial, which could be 'true' or 'recreated' data. While it may be possible to leverage meta-analysed results to inform some other aspects of the model (such as combining external data sources in Jackson's MPES, or combining expert opinion sources for Guyot's MPES), neither method is explicitly designed to evaluate a context where there are multiple clinical trials of interest	
The external data set is inputted as conditional survival count data, denoting start and stop times, and associated sample sizes	
Guyot's MPES approach assumes there is only one external data set providing conditional survival data in the population of relevant (i.e. outside of BGM)	Jackson's MPES approach allows for one or more external data sets for conditional survival
BGM is introduced based on a specific time point, at which time conditional survival for the control arm is expected to be the same as the age- and sex-adjusted general population	BGM is introduced by fitting a relative survival model (i.e. including data for the background expected hazard for each observed survival time based on patient characteristics, via an 'additive hazards' model)
An informative prior is included as part of the model fitting to describe the HR (i.e. this is considered probabilistically). This can be used, for example, to impose the expectation that the HR could be close to 1 at a time point after the end of follow-up for the clinical trial. The HR setting in Guyot's MPES is crucial to link the conditional survival and general population constraints applied to the 'control' arm to the 'active' arm	While Jackson's MPES can be specified as a proportional hazards model, there is no specific input that would allow for quantifying uncertainty about how the HR changes after the clinical trial. Instead, a post hoc adjustment can be applied to adjust for possible treatment effect waning (i.e. this is considered deterministically)
There is no formal constraint included within the model fitting process related to the risk of overfitting for many knots (though no more than three knots have been explored in this tutorial)	A prior smoothness constraint is included as part of the model fitting process, which aims to avoid overfitting in the presence of many knots
Guyot's MPES does not include any formal or specific methodology related to the possibility of a cured proportion of patients. This could, however, be implied based on the external conditional survival data used to inform the common model	Jackson's MPES includes the ability to fit a mixture-cure model, to reflect the possibility of a proportion of long-term survivors

BGM background mortality, *HR* hazard ratio, *MPES* multi-parameter evidence synthesis

and through verification with the original study authors. However, it should be noted that our re-programming of Guyot's MPES does not provide the full suite of functionality afforded by the *survextrap* package for Jackson's MPES (e.g. it does not immediately easily extract hazard estimates alongside credible intervals, which may be a helpful tool to guide model specification as shown by Jackson; see Figure 2 of Jackson [2023]) [7].

The original SCCHN case study has been documented previously both by Guyot et al. and Jackson. Therefore, we considered two other case studies to demonstrate how an MPES approach can be executed from first principles. The first of these considers a population with advanced melanoma, which was the subject of an HTA by the National Institute for Health and Care Excellence (NICE) (TA319) [18]. The second case study considers a population with NSCLC, which is the same cancer type used in the previous applications of an MPES approach by Chaudhury et al. and Vickers, and also formed the main clinical evidence base for NICE TA531 [8, 9, 19].

The case studies explored within this tutorial are intended to be illustrative of the methodology, rather than examples of where an MPES should have been used. We encourage readers to revisit the original case study by Guyot et al., in the context of SCCHN for a full description of how they identified relevant evidence, sought clinical expert opinion, and determined the MPES model specification [6].

6.2 Modifications of Guyot's MPES and Sensitivity Analyses

MPES allows for a flexible approach to model specification, as different settings and assumptions will be appropriate for different applications. The original WinBUGS version of Guyot's MPES is somewhat limited with respect to varying model settings, restricting the extent to which sensitivity analyses can be conducted around model assumptions. In our R version, we allowed more settings to be varied by the user, similar to how Jackson's MPES has been developed allowing different settings to be easily explored. For example, important sensitivity analyses could involve varying the number of knots, knot locations, HR adjustment in the long term, and BGM adjustment time points. Further details around sensitivity analyses using model settings are provided in the ESM.

7 Melanoma Case Study

For this case study, (re-created) patient-level data from the CA184-024 study were used, along with published long-term survival data from the American Joint Committee on Cancer (AJCC) [20, 21]. The CA184-024 study has three published Kaplan-Meier (KM) estimates of survival, reflecting 3, 4, and 5 years of minimum follow-up [20, 22, 23]. The AJCC data provide a KM estimate of survival for $n = 1158$ people with stage IV melanoma, which can be used to extract estimates of survival at 12-month intervals.

We used the earliest (3-year) data-cut from CA184-024 as our primary 'IPD' data source, and obtained survival estimates from the AJCC data between 4 and 15 years to represent our external data source. For the AJCC data, we derived numbers at risk over time based on the starting sample size, assuming no censoring (to provide the necessary input data for conditional survival). The AJCC data were used to inform the survival model only for the control arm. We initially did not include any BGM adjustment or make any assumption regarding the treatment effect. For model specification, we placed two internal knots at 3 years and 7 years (broadly in keeping with the rationale used by Guyot et al., in their original case study), with a boundary knot at 20 years (a time point after the end of the AJCC data) for Guyot's MPES and allowed Jackson's MPES to automate knot selection (i.e. allowing the *survextrap* function to determine where there appeared to be changes in the hazard function over time that would warrant additional knots to improve model fit).

The resultant models, using both MPES approaches, are shown in Fig. 2, panel A. Based on these results, there is a clear issue with Guyot's MPES that we had not anticipated—if the model for the active arm is not linked to the control arm in any way, the resultant extrapolations could be completely implausible. We added in an informative prior that the HR would be approximately 1 from 5 to 20 years to determine if this resolved the issue for Guyot's MPES, and left Jackson's MPES unchanged, which produced the models shown in Fig. 2, panel B and resolved the increasing survival over time issue. We then sought to demonstrate the impact BGM may have on results. For Jackson's MPES, this was facilitated by enabling the relative survival option (which means that the model itself represents *excess* hazard, instead of *overall* hazard); whereas for Guyot's MPES, we chose a time point of 30 years where conditional survival between the control arm and the general population were expected to be similar (though in keeping with Guyot's original case study, no equivalent constraint was imposed on the intervention arm). These changes resulted in the models shown in Fig. 2, panel C (which were similar to those in panel B without BGM, though Guyot's MPES was affected to a greater extent, which may be due to the specific incorporation of BGM at 30 years).

8 NSCLC Case Study

Results from the KeyNote-024 study of pembrolizumab for patients with previously *untreated*, programmed death-ligand 1 (PD-L1)-positive, advanced NSCLC (NCT02 142738) were used to generate re-created patient-level data

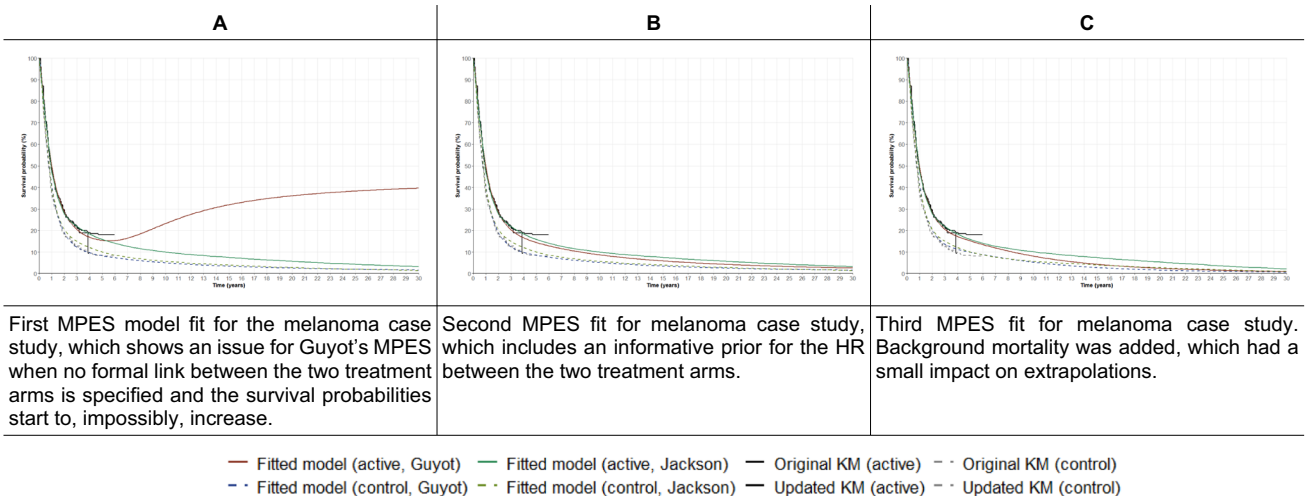

Fig. 2 Melanoma case study. 'Original' refers to the earliest published data-cut from the pivotal study. 'Updated' refers to the latest published data-cut from the pivotal study. *HR* hazard ratio, *KM* Kaplan-Meier, *MPES* multi-parameter evidence synthesis

[24]. Around the same time, pembrolizumab was being studied in a separate study in a previously *treated* population (KeyNote-010, NCT01905657) [25]. For this tutorial, a later data-cut from KeyNote-010 was used as the external evidence source, despite there being an important difference in the patient populations based on treatment history [26]. While this means the case study represents an artificial situation where the external data (KeyNote-010) were published later than the initial data-cut from KeyNote-024, it otherwise represents a possible use case for MPES. In addition, for the purposes of this tutorial, our objective is to demonstrate how these kinds of evidence sources could be used within an MPES.

We used the earliest data-cut from KeyNote-024 (median follow-up of 11.2 months) as our trial population of interest [24]. We extracted reported numbers at risk in 6-month intervals from KeyNote-010, for the pembrolizumab arm, between 2 and 7 years, based on the granularity of reporting in the study publication [27], and applied these data per the melanoma case study. For Guyot's MPES specifically, we assumed (arbitrarily) that the HR would be approximately 1 after 5 years of follow-up and included the same BGM assumptions as per the final melanoma models (relative survival for Jackson's MPES, BGM applied at 30 years as conditional survival for Guyot's MPES). For model specification, we placed two internal knots at 3 years and 7 years, with a boundary knot at 20 years for Guyot's MPES, and allowed Jackson's MPES to automate knot selection.

The resultant models, using both MPES approaches, are shown in Fig. 3, panel A. While Guyot's MPES appeared to work well, Jackson's MPES did not. We believe that this was due to the HR link between the arms in Guyot's MPES, which does not apply to Jackson's MPES, which may dampen the effect of the potentially optimistic external conditional survival data derived from KeyNote-010 (i.e. by establishing a formal link between the treatment arms wherein the HR is expected to be approximately 1 after 5 years, the intervention arm is 'forced' to not project too optimistic estimates of survival in the long term). To explore this further, we repeated the analysis by disabling the HR adjustment in Guyot's MPES and applied the BGM constraint to both arms (i.e. not just the control arm), while leaving Jackson's MPES unaffected—see Fig. 3, panel B. This confirmed our expectation that the KeyNote-010 data may not be an appropriate source of conditional survival estimate, as when the models are not constrained via the HR, we obtain extrapolations that are unrealistically optimistic (assessed through visual inspection of the overall survival data from the updated data-cut from KeyNote-024 versus the KeyNote-010 conditional survival data). Given that we identified that survival in KeyNote-010 appeared to be too optimistic compared to the KeyNote-024 population, we re-ran the analyses with conditional survival probabilities reduced by 10% for both MPES approaches (while leaving the HR adjustment in Guyot's MPES disabled), and obtained the results shown in Fig. 3, panel C.

9 Discussion

Using the software developed for this tutorial, and Jackson's '*survextrap*' R package, MPES can be applied in other contexts with relative ease. Through presenting two case studies, we show how different inputs can be prepared, how these could theoretically be altered, and consequently how they can change the survival extrapolations, demonstrating the

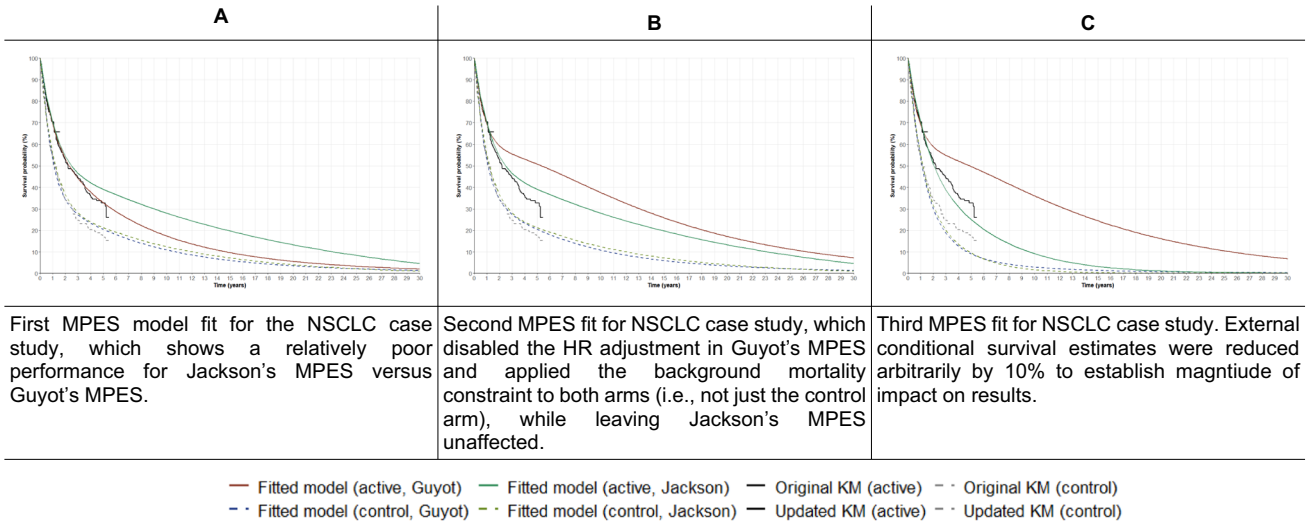

Fig. 3 NSCLC case study. 'Original' refers to the earliest published data-cut from the pivotal study. 'Updated' refers to the latest published data-cut from the pivotal study. *HR* hazard ratio, *KM* Kaplan-Meier, *MPES* multi-parameter evidence synthesis, *NSCLC* non-small-cell lung cancer

importance of justifying assumptions and sensitivity analysis. We also demonstrate how results can differ when using different MPES approaches, and compare the outcomes of these two different specifications (i.e. Guyot's and Jackson's MPES approaches).

From a practical perspective, we show that it is possible to run both MPES approaches in the same software package (R), which should make using both methods more straightforward to an audience less familiar with programming. While the purpose of this tutorial was not to establish which approach is better than the other, we have presented results using both MPES approaches to demonstrate where they appear to produce similar and different results.

9.1 When Should MPES be Used?

Before using MPES, it is first important to assess why it is being considered. In general, the decision to use MPES should be based on an expectation that the added complexity of this type of method (versus other, 'standard' approaches) is warranted given the expected survival profile for a given intervention within a specific disease area. This same logic applies to any complex method. It is therefore necessary to first substantiate (either via empirical evidence or clearly reasoned argument) that standard approaches do not, or will not, yield appropriate survival estimates for decision making. That being said, if there is useful external evidence, it would generally be sensible for this to be used in some way for survival analysis, and MPES provides one possible method to do this (though other methods are available—see Bullement et al. [4]). General guidance for MPES is provided in Fig. 4.

9.2 Advantages of MPES Approaches

To the best of the authors' knowledge, MPES is one of the few methods available to simultaneously leverage evidence from multiple sources to inform estimates of survival for HTA. Given that evidence synthesis plays a critical role in HTA, a clear advantage over 'standard' extrapolation techniques is the ability to make use of multiple relevant sources of evidence in the survival estimation process. Nevertheless, determining how to inform an MPES model is challenging, since there will doubtlessly be an element of heterogeneity between the data source representing the population of interest and supporting external evidence. In the NSCLC example, we assumed that the *conditional* survival for a previously *treated* cohort may be similar in the long term to the survival for a previously *untreated* cohort. This may be inappropriate if, for example, subsequent therapies influence long-term survival markedly or indeed if a large difference in patient characteristics is expected in the longer term.

9.3 Pitfalls Associated with MPES Approaches

Compared with standard frequentist approaches (i.e. fitting a parametric model to only the trial data, with informal comparisons with longer-term evidence), the model fitting process can take some time, which introduces challenges when considering the need to implement the outputs within a cost-effectiveness analysis. For example, on the lead author's computer, each of the case studies took just over 5 min to run for Guyot's MPES. Jackson's MPES is notably faster, taking less than 1 min for each case study. When considering the need to perform similar analyses across different endpoints,

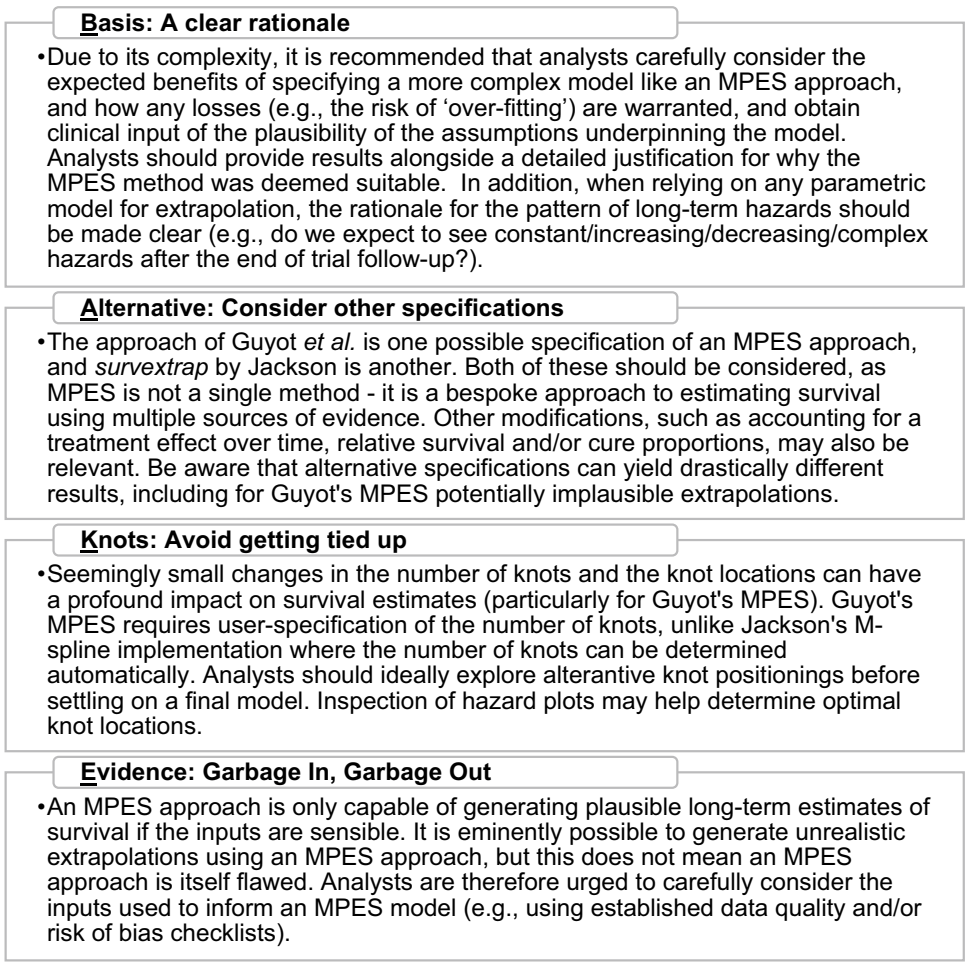

Fig. 4 B-A-K-E: Recommendations for analysts. *MPES* multi-parameter evidence synthesis

subgroups, and to inform sensitivity analyses (within the confines of HTA process timelines), run time may impose challenges when attempting to use an MPES method for HTA.

A further potential issue with using MPES approaches for HTA is how the outputs of this analysis should be combined with all other inputs in a cost-effectiveness analysis. For example, the MPES approach may need to be used as a baseline survival estimate to then combine with some form of indirect treatment comparison (ITC), such as a network meta-analysis. There are some time-varying ITC methods, such as those that use fractional polynomials, that are incompatible with the current forms of both MPES approaches proposed by Guyot et al. and Jackson. It may also be necessary to consider how estimates of survival interact with other model features, such as treatment discontinuation over time, and how to appropriately integrate survival extrapolations within probabilistic sensitivity analysis (taking into consideration the relationship with other cost-effectiveness model inputs).

9.4 Choosing Between MPES Approaches

As noted previously, Jackson's MPES was developed more recently than Guyot's MPES, and is a different specification of an MPES. Jackson's MPES uses an M-spline, whereas Guyot's MPES uses a restricted cubic spline. Perhaps one of the most notable differences between these methods, as Jackson explains, is that M-splines have more favourable properties when considering that hazards should always take a positive value (see the Appendix of Jackson's paper for details) [7]. Jackson's MPES also includes 'special mechanisms' that may be useful for HTA (e.g. relative survival). However, it should also be noted that Jackson's MPES typically specifies a relatively large number of knots, versus Guyot's MPES. For the base-case analyses presented in this tutorial, the default settings of the '*survextrap'* package fitted models with nine knots (for all three case studies).

There are also some notable differences between Guyot's and Jackson's MPES approaches with respect to capturing treatment effect. Guyot's MPES allows the user to specify an

informative prior on the long-term treatment effect as part of the model fitting process, whereas Jackson's MPES does not. Treatment effect waning is a topic frequently raised as part of HTA decision making [28]. Hypothetically, both MPES approaches could be repeated using different long-term effect assumptions, to understand how much this influences survival extrapolations, and by extension, cost-effectiveness results.

9.5 Limitations of Our Study and Avenues for Future Research

In this tutorial, we made some modifications to Guyot's MPES to account for a range of contexts where MPES may be considered. If only single-arm study data are available, a single-arm version of Guyot's MPES could be used, though suitable long-term data for the relevant treatment may be difficult to identify. We also explored models without specifying an informative prior for the HR and/or including a BGM adjustment. Across the scenarios explored, results can be greatly affected by these settings (and may be clinically implausible). Alternative specifications of both MPES approaches (via bespoke programming for Guyot's MPES, or using the pre-built functionality of Jackson's MPES) would also be helpful avenues of further research to provide further information on inputs that have a substantial impact on the predictions made by this alternative application of MPES.

As part of this tutorial, we developed the B-A-K-E recommendations to help guide analysts looking to use MPES for survival extrapolation in HTA (Fig. 4). However, it should be acknowledged that we were not able to follow all of the guidance set out in these recommendations using the two example case studies to demonstrate the two MPES methods. For practical reasons, we did not seek clinical input to determine the model specifications, but accepted those proposed in original publications. Despite this, we believe the recommendations are important to follow if an MPES approach is considered to inform HTA decision-making, particularly given the lack of examples where it has been used in practice.

There are several potential further modifications and refinements to 'off-the-shelf' MPES programs that may be avenues for further research. Firstly, Guyot's MPES includes a somewhat crude approach to accounting for BGM to address implausible long-term estimates of survival. Re-specifying Guyot's MPES as a relative survival model may get around this issue (and is a feature of Jackson's MPES). In addition, the ability to combine multiple sources of evidence within the model fitting process may be relevant under some circumstances (e.g. meta-analysed data from multiple clinical trials), though this is not currently possible in the default specifications of either MPES approach (without, for example, introducing a number of customised informative priors). Similarly, there may be cases where an adjustment to external evidence could be warranted (e.g. adjusting the external study data in our NSCLC case study to account for important differences in patient characteristics). Finally, both the original specification and our re-specification of Guyot's MPES are not currently capable of evaluating comparisons of more than two treatment groups simultaneously, which may be relevant under some circumstances, particularly when considering overlap with methods for deriving ITCs, which would require additional coding.

In HTA, it is commonplace for companies to assume that the trial population is generalisable to that of the 'real-world' population for whom reimbursement is sought. Simultaneously, it is also generally recognised that clinical trial populations tend to be relatively fitter compared to real-world populations, due to eligibility criteria often excluding people with comorbidities, receiving concomitant medications, or otherwise having vulnerabilities due to age [33]. Consequently, it is important to reflect upon the relevance of each data source to the decision problem population for any survival analysis. In the original SCCHN example, Guyot et al. assumed that the Bonner et al. [34] population was directly relevant to the decision problem population, and adjusted the external Surveillance, Epidemiology, and End Results (SEER) Program data to 'match' this population [6].

The current implementation of both MPES approaches does not allow for the analyst to directly control the extent to which external evidence influences survival extrapolations, beyond an arbitrary scaling of the sample size inputs for the external evidence. For example, while the sample size of an external data source may be related to how uncertain estimates of survival derived from these data may be, this is distinct from the concept of how similar the survival experience of the external data is to that of the decision problem population. Ideally, an MPES method would allow for a user-specified control for how much the model is influenced by external evidence (akin to a Bayesian power prior approach [35]), but this is both conceptually and practically difficult, and so remains an area for further research. There is also a range of emergent literature concerning examples of eliciting and incorporating expert opinion for survival, which would also be relevant to consider for MPES [29–32].

10 Conclusion

Several survival analysis guidance documents comment that incorporating external evidence into extrapolations is likely to be helpful and important, but methods for doing so have been lacking. In this tutorial, we demonstrate the use of two specifications of an MPES approach that allow this to be done. However, use of these approaches does not guarantee

accurate extrapolations, different applications of MPES will give different extrapolations, and therefore describing and justifying assumptions is crucial. And investigating how successful these methods are is an important area for further research. Nevertheless, these two MPES approaches represent valuable progress in methods for survival modelling that incorporate multiple evidence sources. This tutorial facilitates further exploration of these methods, aligned with the sentiment raised by Jackson in the presentation of the M-spline MPES: "to improve confidence in [flexible Bayesian evidence synthesis methods], more work to demonstrate their use in a wide range of applications would be helpful" (Jackson, 2023, p.13) [7].

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40273-024-01425-4.

Declarations

Funding No specific funding was provided for this study. Latimer is supported by Yorkshire Cancer Research (Award reference number S406NL). Welton is supported by the NIHR Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol.

Conflict of interest Guyot and Welton are co-authors of the original Guyot et al. (2017) MPES publication referenced throughout this article.

Authors' contributions The idea for the tutorial was conceived by Bullement, Baio, Stevenson, and Latimer. The R code required to generate the results was developed by Bullement and Edmondson-Jones. Guyot and Welton provided further insight into the development of the original MPES, which guided the development of the R code. All co-authors contributed to the overall flow of the tutorial, including inputting to the set of sensitivity analyses, presentation of recommendations, and key discussion points.

Data and code availability The statistical analysis code and input data used to generate all results presented in this tutorial, for both MPES approaches, is provided in the appendix (ESM).

References

1. Latimer N. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data. 2011. http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf. Accessed 31 Dec 2020.
2. Rutherford MJ, Lambert PC, Sweeting MJ, Pennington R, Crowther MJ, Abrams KR, et al. NICE DSU Technical Support Document 21. Flexible methods for survival analysis. 2020. http://nicedsu.org.uk/wp-content/uploads/2020/11/NICE-DSU-Flex-Surv-TSD-21_Final_alt_text.pdf. Accessed 31 Dec 2020.
3. Jackson C, Stevens J, Ren S, Latimer N, Bojke L, Manca A, et al. Extrapolating survival from randomized trials using external data: a review of methods. Med Decis Making. 2017;37(4):377–90.
4. Bullement A, Stevenson MD, Baio G, Shields GE, Latimer NR. A systematic review of methods to incorporate external evidence into trial-based survival extrapolations for health technology assessment. Med Decis Making. 2023;2023:2729898.
5. Lambert PC, Smith LK, Jones DR, Botha JL. Additive and multiplicative covariate regression models for relative survival incorporating fractional polynomials for time-dependent effects. Stat Med. 2005;24(24):3871–85.
6. Guyot P, Ades AE, Beasley M, Lueza B, Pignon JP, Welton NJ. Extrapolation of survival curves from cancer trials using external information. Med Decis Making. 2017;37(4):353–66.
7. Jackson-Christopher H. survextrap: a package for flexible and transparent survival extrapolation. BMC Med Res Methodol. 2023;23(1):282.
8. Vickers A. An Evaluation of survival curve extrapolation techniques using long-term observational cancer data. Med Decis Making. 2019;39(8):926–38.
9. Chaudhary MA, Edmondson-Jones M, Baio G, Mackay E, Penrod JR, Sharpe DJ, et al. Use of advanced flexible modeling approaches for survival extrapolation from early follow-up data in two nivolumab trials in advanced NSCLC with extended follow-up. Med Decis Making. 2023;43(1):91–109.
10. Lunn DJ, Thomas A, Best N, Spiegelhalter DJ. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Stat Comput. 2000;10:325–37.
11. Stan Development Team. RStan: the R interface to Stan. R package version 2.32.6. 2024. https://mc-stan.org/.
12. Stan Development Team. Stan reference manual. Version 2.35. 2024. https://mc-stan.org/docs/reference-manual/
13. Ades AE, Welton NJ, Caldwell D, Price M, Goubar A, Lu G. Multiparameter evidence synthesis in epidemiology and medical decision-making. J Health Serv Res Policy. 2008;13(Suppl 3):12–22.
14. R Core Team. R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. 1995. https://www.R-project.org/. Accessed 28 Oct 2022.
15. StataCorp. Stata statistical software. College Station: StataCorp LP. 1985.
16. SAS Institute Inc. SAS: SAS Institute Inc. 1972.
17. R for HTA. R for Health Technology Assessment (HTA). 2023. https://r-hta.org/. Accessed 27 Oct 2023.
18. National Institute for Health and Care Excellence (NICE). NICE TA319: Ipilimumab for previously untreated advanced (unresectable or metastatic) melanoma. 2014. https://www.nice.org.uk/guidance/ta319/. Accessed 31 Dec 2020.
19. National Institute for Health and Care Excellence (NICE). NICE TA531: Pembrolizumab for untreated PD-L1-positive metastatic non-small-cell lung cancer. 2018. https://www.nice.org.uk/guidance/ta531/. Accessed 27 Oct 2023.
20. Robert C, Thomas L, Bondarenko I, O'Day S, Weber J, Garbe C, et al. Ipilimumab plus dacarbazine for previously untreated metastatic melanoma. N Engl J Med. 2011;364(26):2517–26.
21. Balch CM, Buzaid AC, Soong SJ, Atkins MB, Cascinelli N, Coit DG, et al. Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. J Clin Oncol. 2001;19(16):3635–48.
22. Maio M, Bondarenko I, Robert C, Thomas L, Garbe C, Testori A, et al. 1127P—four-year survival update for metastatic melanoma (MM) patients (PTS) treated with ipilimumab (IPI) + dacarbazine (DTIC) on phase 3 study CA184-024. Ann Oncol. 2012;23:367.
23. Maio M, Grob JJ, Aamdal S, Bondarenko I, Robert C, Thomas L, et al. Five-year survival rates for treatment-naive patients with advanced melanoma who received ipilimumab plus dacarbazine in a phase III trial. J Clin Oncol. 2015;33(10):1191–6.
24. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl J Med. 2016;375(19):1823–33.

25. Herbst RS, Baas P, Kim DW, Felip E, Pérez-Gracia JL, Han JY, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet. 2016;387(10027):1540–50.
26. Herbst RS, Garon EB, Kim DW, Cho BC, Gervais R, Perez-Gracia JL, et al. Five year survival update from KEYNOTE-010: pembrolizumab versus docetaxel for previously treated, programmed death-ligand 1-positive advanced NSCLC. J Thorac Oncol. 2021;16(10):1718–32.
27. Reck M, Rodríguez-Abreu D, Robinson AG, Hui R, Csőszi T, Fülöp A, et al. Five-year outcomes with pembrolizumab versus chemotherapy for metastatic non-small-cell lung cancer with PD-L1 tumor proportion score ≥ 50. J Clin Oncol. 2021;39(21):2339–49.
28. Armoiry X, Wang-Steverding X, Connock M, Grove A, Clarke A, Arun T, et al. Is the assumption of waning of treatment effect applied consistently across NICE technology appraisals? A case-study focusing on disease-modifying therapies for treatment of multiple sclerosis. Int J Technol Assess Health Care. 2022;38(1):e83.
29. Ayers D, Cope S, Towle K, Mojebi A, Marshall T, Dhanda D. Structured expert elicitation to inform long-term survival extrapolations using alternative parametric distributions: a case study of CAR T therapy for relapsed/refractory multiple myeloma. BMC Med Res Methodol. 2022;22(1):272.
30. 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):182.
31. Cooney P, White A. Direct incorporation of expert opinion into parametric survival models to inform survival extrapolation. Med Decis Making. 2023;43(3):325–36.
32. Willigers BJA, Ouwens M, Briggs A, Heerspink HJL, Pollock C, Pecoits-Filho R, et al. The role of expert opinion in projecting long-term survival outcomes beyond the horizon of a clinical trial. Adv Ther. 2023;40(6):2741–51.
33. Tan YY, Papez V, Chang WH, Mueller SH, Denaxas S, Lai AG. Comparing clinical trial population representativeness to real-world populations: an external validity analysis encompassing 43 895 trials and 5 685 738 individuals across 989 unique drugs and 286 conditions in England. lancet Healthy longevity. 2022;3(10):e674–89.
34. Bonner JA, Harari PM, Giralt J, Azarnia N, Shin DM, Cohen RB, et al. Radiotherapy plus cetuximab for squamous-cell carcinoma of the head and neck. N Engl J Med. 2006;354(6):567–78.
35. Ibrahim JG, Chen MH, Gwon Y, Chen F. The power prior: theory and applications. Stat Med. 2015;34(28):3724–49.