Estimating the Long-Term Outcomes Associated With Immuno-Oncology Therapies: Challenges and Approaches for Overall Survival Extrapolations

Min Huang, PhD, AstraZeneca, Gaithersburg, MD, USA; Nicholas Latimer, PhD, University of Sheffield, Sheffield, UK; Yiduo Zhang, PhD, Pralay Mukhopadhyay, PhD, and Mario Ouwens, PhD, AstraZeneca, Gaithersburg, MD, USA; and Andrew Briggs, DPhil, University of Glasgow, Glasgow, UK

**KEY POINTS**

Traditional parametric survival models are commonly used to estimate long-term survival in oncology health technology assessments, however, they cannot adequately represent complex hazard functions and may not be appropriate for modelling the underlying mechanism of action associated with immuno-oncology treatments.

Novel techniques for survival extrapolation, such as flexible parametric models, parametric mixture models and mixture cure models, and landmark-based response models can characterize complex hazard functions with turning points and changing slopes.

Cure, parametric mixture and landmark models may provide more insight into the potential mechanism of action of immuno-oncology therapies, compared with the more mechanistic traditional parametric survival models, or spline models.

**TRADITIONAL SURVIVAL MODELING: LIMITATIONS FOR IMMUNO-ONCOLOGY**

Cost-effectiveness analyses are key factors in reimbursement decisions made on new healthcare interventions around the world. Cancer treatments usually affect survival, and therefore cost-effectiveness analyses must estimate costs and benefits associated with competing treatment options over a lifetime period [1-4]. Trials have limited follow-up and in almost all instances, it is necessary to extrapolate beyond the trial data to estimate lifetime survival. Traditional parametric survival models are usually used for this task, whereby it is assumed that survival follows a particular underlying distribution. Different distributions can have a large impact on long-term survival estimates and although model choice is not straightforward, some guidance exists to help with this problem [5,6]. However, the extrapolation problem appears to be becoming more difficult—and even more important—with the development of immuno-oncology treatments, which have a number of unique characteristics, such as delayed effects, potential long-term survivors, and less mature survival data [7].

When treatments have delayed effects and long-term survivors, the implication is that the hazard function follows a more complex pattern than is modeled by the traditional parametric survival distributions. In the context of clinical trials of IO treatments in the metastatic cancer setting, the risk (or hazard) of death may be relatively low initially, due to trial inclusion criteria resulting in trial populations that are relatively fit compared to the more general disease population. However, given that trial participants have a severe disease, the hazard of death is likely to rise in the short-term. If treatment has a delayed effect, or only works for a proportion of subjects, the hazard may begin to fall, or at least its gradient may become less steep. In fact, in the longer term the situation may be even more complicated, as hazards may change again, increasing due to age-related mortality risks.

Traditional parametric models cannot adequately represent complex hazard functions with turning points. They also may be inadequate for representing underlying biomedical processes such as durable responses that have been observed for some IO treatments [8,9]. The models most commonly used in health technology assessment (HTA) are Weibull and exponential models [5,6]. Exponential models assume that the hazard remains constant across time, whilst Weibull models can represent hazards that either monotonically increase or monotonically decrease and nest the exponential as a special case. Other traditional models, such as Gompertz, log-logistic, log normal, and generalized Gamma models are either similarly restricted, or allow only a small amount of increased flexibility. Like Weibull models, Gompertz models can only represent hazards that increase or decrease monotonically but the rate of change has to be exponential. Log-logistic, log normal, and generalized Gamma models are able to represent hazards that initially increase and then decrease, but cannot characterize a second turning point or additional important changes in the slope of the hazard.

Therefore, if we believe that the hazard function associated with a new IO treatment (or any other treatment) is likely to have a turning point and/or important changes in slope over time, it is necessary to look beyond the standard parametric distributions when attempting to model long-term survival. Flexible parametric models [10,11], parametric mixture models and mixture cure models [12,13], and landmark-based response models [14] each provide a modelling approach that can characterize complex hazard functions with turning points and changing slopes. Each approach has strengths and weaknesses. There is little information in the literature regarding direct comparisons of these approaches; they have not been used largely in HTA decision making.
NOVEL TECHNIQUES FOR SURVIVAL MODELING: A CASE STUDY

In this case study, we introduced a comprehensive list of relevant methods for extrapolating overall survival (OS) data and illustrated their application to a clinical trial data set, the ATLANTIC study. We also compared and contrasted the model outcomes and provided insights on the tradeoffs in model selection in the context of the probable underlying biomedical processes for IO treatments. ATLANTIC is a phase II, open-label, single-arm trial of durvalumab in patients with stage IIIB–IV non-small cell lung cancer with World Health Organization performance status 0 or 1, who received at least 2 prior systemic treatment regimens, including 1 platinum-based. The trial results were presented at a plenary session entitled Immune Checkpoint Inhibitors in Advanced NSCLC at the World Conference on Lung Cancer [15].

We first fitted the OS data with the standard parametric models. Akaike information criterion and Bayesian information criterion suggested the log-normal as the best-fitting standard model to the trial data. Therefore, we selected the log-normal model as the benchmark against which to compare and evaluate different, more complex survival modeling approaches. We recognize that providing the best fit to the observed data does not mean that the log-normal model represents the best option for extrapolating beyond the trial period, but we decided to use it as a standard parametric model benchmark in order to avoid an unwieldy number of comparisons.

FLEXIBLE PARAMETRIC MODELS

Spline-based models [10,11] are flexible parametric models defined by piecewise polynomials. The point at which polynomials “join” are called knots. The modelled hazards are smoothed at the “knots” where the distributions change.

We considered that there were 2 key turning points in the observed hazard function of the data, and therefore 1-knot and 2-knot spline models were used to model the ATLANTIC OS data.

Although allowing more knots increases the flexibility of curve fitting, which often provides a better fit to observed data, it is important to select an approach that balances the flexibility in capturing the observed hazards and the risk of overfitting the data (eg, segmenting the data too thinly resulting in extrapolations based on a small amount of data). Indeed, fit to the observed data is often of secondary importance to the credibility of the extrapolated portion of the curve.

MIXTURE CURE MODELS

Mixture cure models were introduced more than 50 years ago [16]. They have been proposed recently to model survival of emerging cancer therapies, (eg, IO therapies) [13,17], as evidence has shown that these treatments may offer long-term survival (“cure”) to certain patients in some indications. Mixture cure models can address the heterogeneity induced by the fraction of “cured” patients whose OS prognosis is assumed to be similar to that of the general population (depending upon how the “cure” is defined) instead of their counterparts who were not “cured.”

The key assumption in this approach is the plausibility of a cure. Based on the assumption, the cure models estimate the percentage of patients who are cured and estimate the survival function for patients who are not cured. The risk of death of a cured patient is based on the background mortality of the general population.

We fitted the ATLANTIC OS data with a cure model and modeled this population as a mixture of cured and uncured patients, assuming a Weibull distribution for the survival of the uncured population. For background mortality, we used age- and gender-matched UK life tables based on data for the years 2012 to 2014.

Compared with the standard log-normal function, the mixture cure model predicts a much larger long-term survival rate and patient life expectancy.

In the absence of long-term follow-up data, a scientifically grounded and consistent approach in survival extrapolations is desirable, and should serve as the foundation to demonstrate the potential value of immuno-oncology treatments.

PARAMETRIC MIXTURE MODELS

Parametric mixture models are a more general approach to address population heterogeneity. They can be used to model 2 (or more) distinct groups, without assuming a “cure.” We used a mixture of 2 Weibull distributions, and the model estimated a probability for each patient of belonging to each group. On average, the probability of belonging to group one (the first mixture) was 68%, and the probability of belonging to group 2 (the second mixture) was 32%, with group one being represented by a superior survival distribution.

RESPONSE-BASED LANDMARK MODELS

Based on the strong correlation between response and survival for IO therapies, response-based landmark models were considered as another approach to model the heterogeneity in overall survival in the ATLANTIC study.

These models distinguish “responders” from “non-responders” and model survival for each group separately. Response is declared, and subsequent survival is modeled, from a landmark point to avoid the bias that responders by definition have to survive to the point at which response is assessed. For the ATLANTIC study, we defined the landmark at 2 months, as the first tumor response assessment occurred at 8 weeks after treatment initiation in this trial. Response categories were defined as follows:

- Responder (R): Patients who remain progression free 2 months or more from start of treatment
- Non-responder (NR): Patients who progress or are censored prior to 2 months

Fifty-four percent of patients were categorized as responders and 46% as non-responders. The OS data after the landmark point show clear differentiation between the 2 response groups. Standard parametric models were fitted to each of the 2 groups, and the exponential distribution was selected for both groups as the best fit based on AIC and BIC, suggesting a constant hazard over time for each of the response categories. Visual inspection of both the survival function and the hazard function suggest that the landmark model provided a close fit to the observed data.

Value & Outcomes Spotlight JANUARY/FEBRUARY 2018 I 29
KEY LEARNING AND IMPLICATIONS

In this study, we explored a comprehensive list of alternative survival models using a data set from a recent IO clinical trial. All models provided a close fit to the observed OS data; however, their tails—the projections beyond the trial period—are very different [18]. Consequently, they estimate very different outcomes for mean OS, which is a measure of great importance used in cost-effectiveness models and health economic evaluation.

Therefore, validation of long-term survival projections is critical for choosing the “right” survival model. Relevant internal and external benchmarks should be cross-examined carefully and long-term clinical and real-world data should be identified and used to validate the survival extrapolations. At the same time, survival modeling should account for clinical rationale and validity. Among the models that were examined in this study, only the mixture model (mixture cure model as a special case, if a “cure” can be supported) and the landmark model offer potential insight into the biomedical mechanism of action. Nevertheless, in this application, these 2 approaches paint quite different pictures.

Although it is appealing to use a cure modeling approach to account for long-term survivors in cancer treatments, the validity of the assumed “cure” remains as the main challenge. The assumption of cure can only be verified with long-term follow-up data. However, for studies with limited follow-up, the key assumption cannot be validated. In addition, the estimate of the cure fraction could be sensitive to the choice of parametric distribution. Therefore, cure models could generate misleading results.

Parametric mixture models can be used to model distinct survival distributions within data when a cure is not supported. These provide great flexibility in the shape of hazard and survival functions, although the flexibility of the modeling approach may lead to a high level of uncertainty in parameter estimates.

Response-based landmark models acknowledge the distinction between responders and non-responders and have a strong clinical rationale, provided the response measure is prognostic. It is therefore important to demonstrate that the chosen response measure is a reliable surrogate for survival. In addition, the model can be sensitive to the choice of the landmark time-point. The landmark should be selected such that it minimizes the loss of patients who die prior to the landmark, and accounts for vast majority of responders. In our case study, a key finding was that the OS projection for responders was the key driver of uncertainty, because the patients who were still alive at the end of trial follow-up period were mostly responders.

In general terms, we prefer the cure, parametric mixture, and landmark models in that they provide more insight into the potential mechanism of action of IO therapies than the more mechanistic traditional parametric survival models, or spline models. Nevertheless, due to the limited follow-up in the current ATLANTIC trial, more data are required to provide a definitive choice as to which model offers the best fit and the most appropriate option for extrapolation.

We are still at an early stage in the development of IO therapies. For many indications and many therapies, the mechanism of action is yet to be fully understood and the long-term survival trend of patients is yet to be observed. This adds to the challenge of modeling long-term survival in IO. It also highlights the importance of advancing the basic scientific understanding on the mechanism of action and collecting long-term outcome data from clinical trial studies and real-world evidence studies. In the absence of long-term follow-up data, a scientifically grounded and consistent approach in survival extrapolations is desirable, and should serve as the foundation to demonstrate the potential value of IO treatments. 

REFERENCES


Additional information: 
The preceding article was based on a workshop presentation presented at the ISPOR 20th Annual European Congress. To view this presentation, go to https://www.ispor.org/Event/ReleasedPresentations/2017Glasgow#workshoppresentations