ScienceDirect

Contents lists available at **sciencedirect.com**
Journal homepage: **www.elsevier.com/locate/jval**

Methodology

Predicting Survival for Chimeric Antigen Receptor T-Cell Therapy: A Validation of Survival Models Using Follow-Up Data From ZUMA-1

Sachin Vadgama, MSc, Jess Mann, MSc, Zahid Bashir, MBBS, Clare Spooner, MBBS, Graham P. Collins, DPhil, Ash Bullement, MSc

ABSTRACT

Objectives: Survival extrapolation for chimeric antigen receptor T-cell therapies is challenging, owing to their unique mechanistic properties that translate to complex hazard functions. Axicabtagene ciloleucel is indicated for the treatment of relapse or refractory diffuse large B-cell lymphoma after 2 or more lines of therapy based on the ZUMA-1 trial. Four data snapshots are available, with minimum follow-up of 12, 24, 36, and 48 months. This analysis explores how survival extrapolations for axicabtagene ciloleucel using ZUMA-1 data can be validated and compared.

Methods: Three different parametric modeling approaches were applied: standard parametric, spline-based, and cure-based models. Models were compared using a range of metrics, across the 4 data snapshot, including visual fit, plausibility of long-term estimates, statistical goodness of fit, inspection of hazard plots, point-estimate accuracy, and conditional survival estimates.

Results: Standard and spline-based parametric extrapolations were generally incapable of fitting the ZUMA-1 data well. Cure-based models provided the best fit based on the earliest data snapshot, with extrapolations remaining consistent as data matured. At 48 months, the maximum survival overestimate was 8.3% (Gompertz mixture-cure model) versus the maximum underestimate of 33.5% (Weibull standard parametric model).

Conclusions: Where a plateau in the survival curve is clinically plausible, cure-based models may be helpful in making accurate predictions based on immature data. The ability to reliably extrapolate from maturing data may reduce delays in patient access to potentially lifesaving treatments. Additional research is required to understand how models compare in broader contexts, including different treatments and therapeutic areas.

Keywords: chimeric antigen receptor T-cell, mixture-cure model, non-Hodgkin lymphoma, survival extrapolation.

VALUE HEALTH. 2021; ■(■):■–■

Introduction

Axicabtagene ciloleucel (axi-cel) is an autologous anti–CD-19 chimeric antigen receptor T-cell (CAR T-cell) therapy, indicated for the treatment of adult patients with relapsed or refractory (R/R) non-Hodgkin lymphoma (including diffuse large B-cell lymphoma [DLBCL] and primary mediastinal large B-cell lymphoma), after ≥2 lines of systemic therapy.[1] CAR T-cell therapies have been described as advanced cancer treatments belonging to a new generation of cancer immunotherapies, which involve collecting and genetically modifying patients' immune cells to treat their cancer.[2,3]

Axi-cel has been studied in the ZUMA-1 clinical trial: a single-arm, multicenter, phase I/II study of adults with R/R aggressive non-Hodgkin lymphoma (NCT02348216). Initial findings from ZUMA-1 were published in 2017, with a minimum follow-up of 12 months and median follow-up of 15.4 months for 101 treated patients.[4] Three subsequent database locks (data snapshots) were later published, with median follow-up of 27.1, 39.1, and 51.1 months.[5–7] With the latest data snapshot, 43.6% of the treated patients (n = 44/101) were still alive at 4 years and the median overall survival (OS) was (25.8 months).[7]

Survival estimates from clinical trials are usually evaluated using the Kaplan-Meier (KM) method, which estimates the proportion of patients still alive over time while considering some subjects are censored before the event of interest occurred. Where data are incomplete, estimates of survival toward the end of follow-up are typically based on a smaller "at risk" population and are subject to greater uncertainty. This is especially challenging for populations that may experience long-term survival benefits that extend over many years, as is expected for a proportion of patients receiving axi-cel.

The potential for the efficacy of cellular immunotherapy, such as axi-cel, in R/R DLBCL is illustrated by the observation of a graft-versus-lymphoma effect in a study by Bishop et al.[8] In this study, patients with R/R DLBCL were treated with donor lymphocyte infusions after relapse after an allogeneic stem-cell transplant (SCT), and long-term remissions were documented. Nevertheless,

other factors may be important in contributing to long-term survival, including cellular factors (eg, cell dose, early expansion after infusion) and patient factors (eg, extranodal sites of disease and disease burden at infusion).[9] In some cases and similar to an allogeneic SCT, CAR T-cell therapy may lead to persistence of antitumor T-cells (in this case autologous anti–CD-19 directed), which may be important in contributing to durable progression-free survival.[10]

The accurate estimation of lifetime survival outcomes for patients treated with a new health technology is especially important for cost-effectiveness analysis as part of health technology assessment (HTA). Nevertheless, trial-based estimates of costs and benefits are only available for the relatively short period of follow-up. Evaluating the technology in terms of its costs and benefits only within the trial period would lead to biased cost-effectiveness estimates. Hence, decision makers will often require information on the intervention's lifetime cost and benefits, versus the established standard of care, to determine whether the technology represents value for money if adopted.

Lifetime survival estimates are factored into the determination of the incremental cost-effectiveness ratio—a measure of cost-effectiveness commonly used in HTA decision making. The selection of survival extrapolation techniques can have a profound impact on the estimate of the incremental cost-effectiveness ratio[11-13] and indeed was the case in the National Institute for Health and Care Excellence (NICE) assessment of axi-cel (TA559).[14] The choice of one extrapolation method over another may mean the difference between a positive or negative reimbursement decision. Therefore, selecting the most plausible and appropriate model in the context of the appraised intervention to extrapolate estimates of survival is of paramount importance and is therefore the focus of this study, with axi-cel being a case study.

A range of studies have previously explored possible means of validating model estimates. Cope et al[15] combined trial-based estimates of survival with clinical expert opinion and referred to the accuracy of projections based on annualized point estimates. Ouwens et al[16] focused instead on the "area-under-the-curve" to compare different models, equivalent to estimating restricted mean survival time, and also presented plots of the estimated hazard function to inform model selection. Klijn et al[17] compared estimates of the conditional survival between models as an alternative means of judging model fit. Nevertheless, there is no established consensus on how to choose the best-fitting models, in spite of there being a large number of studies concerning how to undertake survival analysis for HTA.[11,13,18-21]

This study builds on the growing research for survival extrapolation in cancer immunotherapy, focusing on the ZUMA-1 trial data. A variety of models to extrapolate survival estimates over a lifetime horizon were considered and compared using a range of possible metrics. We then aimed to identify which methods may be considered the most appropriate to inform lifetime survival estimates for the purposes of cost-effectiveness analysis and HTA of CAR T-cell therapy.

Methods

Inspection of Available Data

The ZUMA-1 trial comprises 111 patients enrolled in phase II, of which 101 received axi-cel infusion, recruited from May 19, 2015, to September 15, 2016, across 22 sites in the United States and Israel.[4] The primary endpoint for phase II of ZUMA-1 was overall response rate, with secondary outcomes including safety, duration of response, progression-free survival, and OS, the latter of which is the focus of our study.

To date, 4 data snapshots have been reported with 12-, 24-, 36-, and 48-month minimum follow-up times. Figure 1 compares the corresponding KM estimates for OS across each of the 4 data snapshots. The KM estimates for each data snapshot are overlayed to allow for an inspection of how the KM estimate changed as further follow-up data were made available from ZUMA-1.

Median OS should be interpreted with care, because it may not serve as an accurate reflection of the average outcome of treatment. This is demonstrated in Figure 1 where median OS is reached at 25.8 months, before which almost half of the trial participants had an event and after which approximately the other half was censored, leading to a plateau in the KM estimate. It is also important to acknowledge the number of patients still at risk after 48 months, after which there are limited numbers of patients to inform the KM estimate.

Survival Models Considered

Guidance for the conduct and selection of survival extrapolation methods for cost-effectiveness analysis has been previously published, although the survival for patients treated with CAR T-cell therapies warrants consideration of flexible methods given the unique mechanism of action of CAR T-cell therapies.[13,18] More recently, the NICE Decision Support Unit (DSU) published a technical support document (TSD) 21: Flexible Methods for Survival Analysis; which describes in more detail other options that may also be important to consider in the presence of complex hazard functions.[19] Nevertheless, NICE DSU TSD 21 highlights that the methods presented ought not to be considered as "an extended list of survival methods to 'try out' on data."[19] Instead, the authors encourage the consideration of the likely hazard/survival functions and to then choose appropriate methods accordingly. Therefore, we aimed to explore some of the methods noted in previously published studies and guidelines, based on a range of approaches that may be deemed suitable in the context of the ZUMA-1 trial.

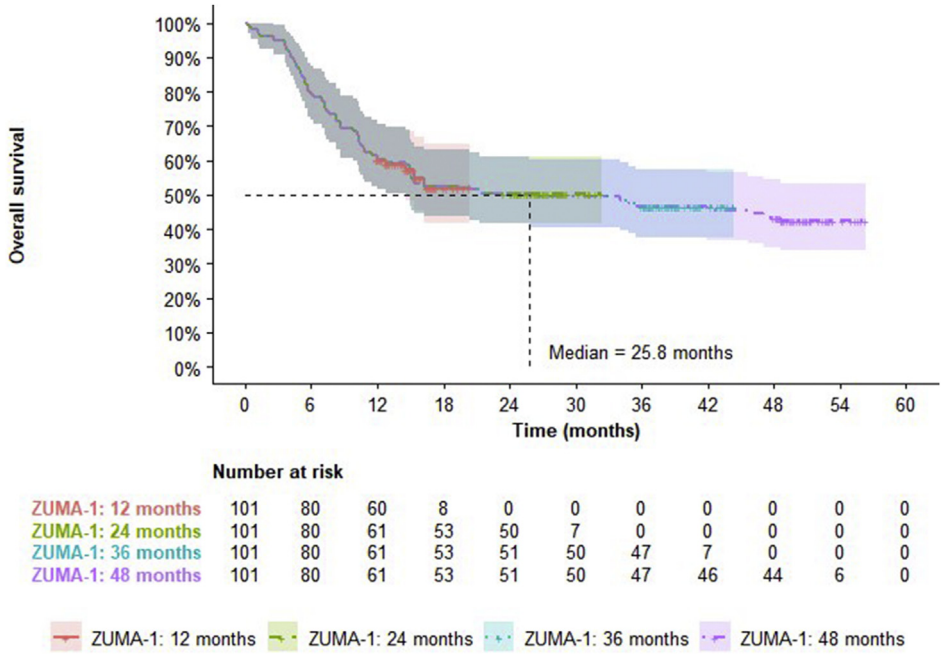
In addition to the selection of methods, it is noted that NICE DSU TSD 21 recommends that background mortality be incorporated into survival models to avoid "extremely implausible" projections and that incorporation of background mortality for cure-based models is "essential."[19] Therefore, to account for background population mortality consistently across the models, all models in the present study were fitted within a relative survival framework, which considers the age- and sex-matched hazard rate of the general population when extrapolating survival. Relative survival is used extensively in the context of population-based cancer registry analysis, where the outcome of interest is often the difference in observed survival for the population under consideration versus the expected survival of the equivalent age- and sex-adjusted general population.[22]

Three broad categories of parametric models were considered within the analysis: "standard" models, spline-based models, and cure-based models. Further details of the models fitted, along with the statistical software packages used and source of information for background mortality, are provided within the Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015.

Evaluation of Survival Models

There is no universally recognized "gold standard" approach to definitively establish the most suitable survival extrapolation, and so judgment of the most appropriate model has historically involved consideration of several factors. With this in mind, we considered a range of different approaches to compare the different extrapolation methods. Furthermore, we elicited the

Figure 1. Kaplan-Meier estimates of overall survival from ZUMA-1 across 4 data snapshots. Shaded area represents the 95% confidence interval around the Kaplan-Meier estimator, and tick marks represent censored observations.



opinion of an expert physician to determine which models were most realistic.

The following approaches to compare models were considered:

- Visual fit of the model to the KM estimator
- Plausibility of long-term estimates based on input from a clinical expert
- Akaike's and Bayesian information criteria
- Inspection of hazard plots
- Point-estimate accuracy versus the KM estimate at specific time points
- Conditional survival estimates (ie, the probability of patients surviving until $t = x + \delta$, given that they have survived up until $t = x$)

We also developed a pragmatic approach to ascertain which methods seemed to provide a good fit to the ZUMA-1 data in general, while accounting for the number of data cuts available. In this approach, we calculated the sum of absolute errors of a given parametric model fitted to each of the 4 data snapshots at a specific time point (in the base-case analysis, we considered a time point of 48 months); we term this the Vadgama Deviance (VD).

$$VD = \sum_{i}^{N} \left| S(t = LOE)_{i=N}^{KM} - S(t = LOE)_{i}^{Model} \right|$$

where i is the data cut being used to inform the curve (ie, 12, 24, 36, or 48 months), N is the total number of data cuts (4 in this case), LOE is the last observable exit time for the latest data cut N, $S(t)^{KM}$ is the survival function evaluated at time t as determined by the KM estimator, and $S(t)^{Model}$ is the survival function evaluated at time t as estimated by the model. VD is given in terms of percentage. Nevertheless, it should be noted that this is intended to merely serve as a summary of how the different models fared in terms of their fit to the 4 data snapshots, as opposed to a fully robust means of formally establishing goodness of fit.

Results

Description of Model Fits for Each Data Snapshot

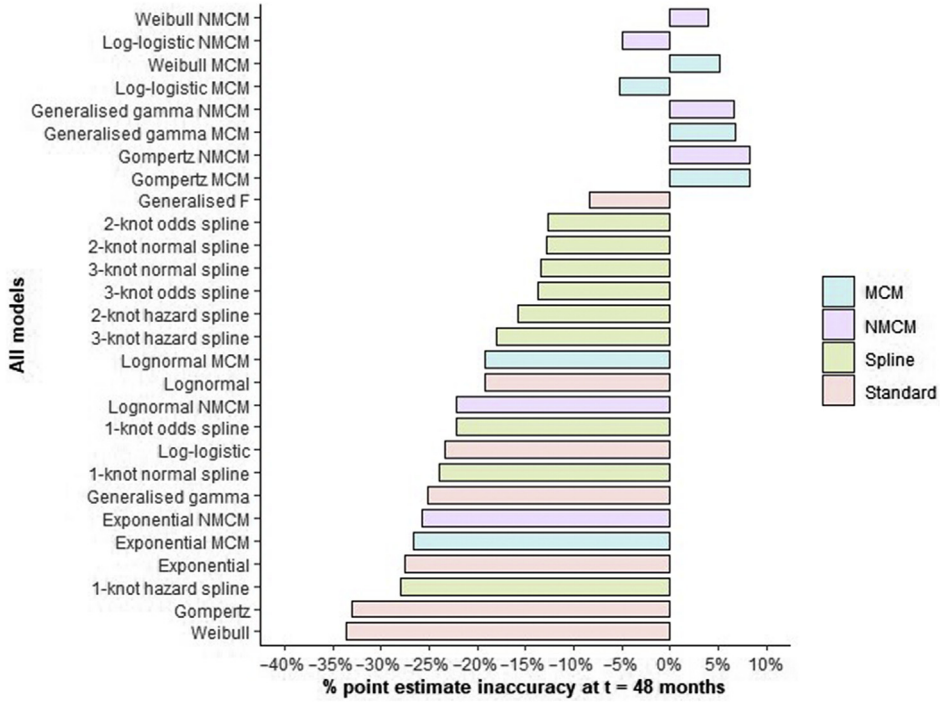
Full model results are provided within the Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015; nevertheless, for brevity, top-level findings for each model type are discussed below.

The difference in projected survival at the latest, 48-month data snapshot for each model fitted to the 12-month data snapshot is presented in Figure 2. Figure 2 shows that the mixture-cure model (MCM) and non-mixture-cure models (NMCMs) better predicted survival at 48 months versus the standard and spline-based models. Some of the MCMs and NMCMs overestimate the KM estimator slightly at 48 months, but all the standard and spline-based models underestimate survival (maximum overestimate of 8.3% for the Gompertz MCM vs maximum underestimate of 33.5% for the standard Weibull model).

The standard parametric models generally did not fit the ZUMA-1 data well, across each data snapshot. These models were shown to consistently underestimate OS even with more mature data from ZUMA-1 (Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015). This finding suggests that these models are not capable of appropriately reflecting the complex pattern of hazards associated with axi-cel, with maybe the exception of the Gompertz model, which was able to reflect a "plateau" in the OS curve in more mature data snapshots.

The spline-based models provided a better visual fit to the KM estimates versus the standard parametric models. This is likely due to the ability for the spline-based models to better consider a

Figure 2. Point-estimate inaccuracy of models fitted to 12M data, versus the 48M Kaplan-Meier estimator evaluated at t = 48M. Models ranked in order of smallest to highest deviance to last observed KM OS survival probability.



KM indicates Kaplan-Meier; M, month; MCM, mixture-cure model; NMCM, non-mixture-cure model; OS, overall survival; t, time.

more complex hazard function with turning points (see Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015 for hazard plots). Still, the ability for these models to accurately project survival based on later data snapshots was limited, because the models did not seem capable of reflecting the expected plateau, which was not fully established in the earlier data snapshots of ZUMA-1.

Some of the MCMs reflected a plateau in the OS curve based on the 12-month data (generalized gamma, Gompertz, and Weibull), whereas others (exponential, log-logistic, and lognormal) did not exhibit a clear plateau in the OS curve. For the later data snapshots, all the MCMs converged to produce similar projections to those based on the generalized gamma, Gompertz, and Weibull MCMs fitted to the 12-month data.

As part of the MCM model fitting process, a "cure fraction" (π) is produced, which is a model parameter, as opposed to an expectation of the proportion of patients that may be "cured" of their disease. The MCMs fitted to the 12-month data snapshot that provided the poorest fit to the 48-month KM had the lowest cure fraction π of the MCMs considered. This implies that some choices of parametric model for the "uncured fraction" (ie, $1 - \pi$) may result in a predominantly "uncured" group when fitted to less mature data, leading to unrealistically pessimistic extrapolations. For the later data snapshots, estimates of the π were stable regardless of the parametric form, implying that more mature data allow for a clearer distinction between the "uncured" and "cured" fractions regardless of the distributional choice.

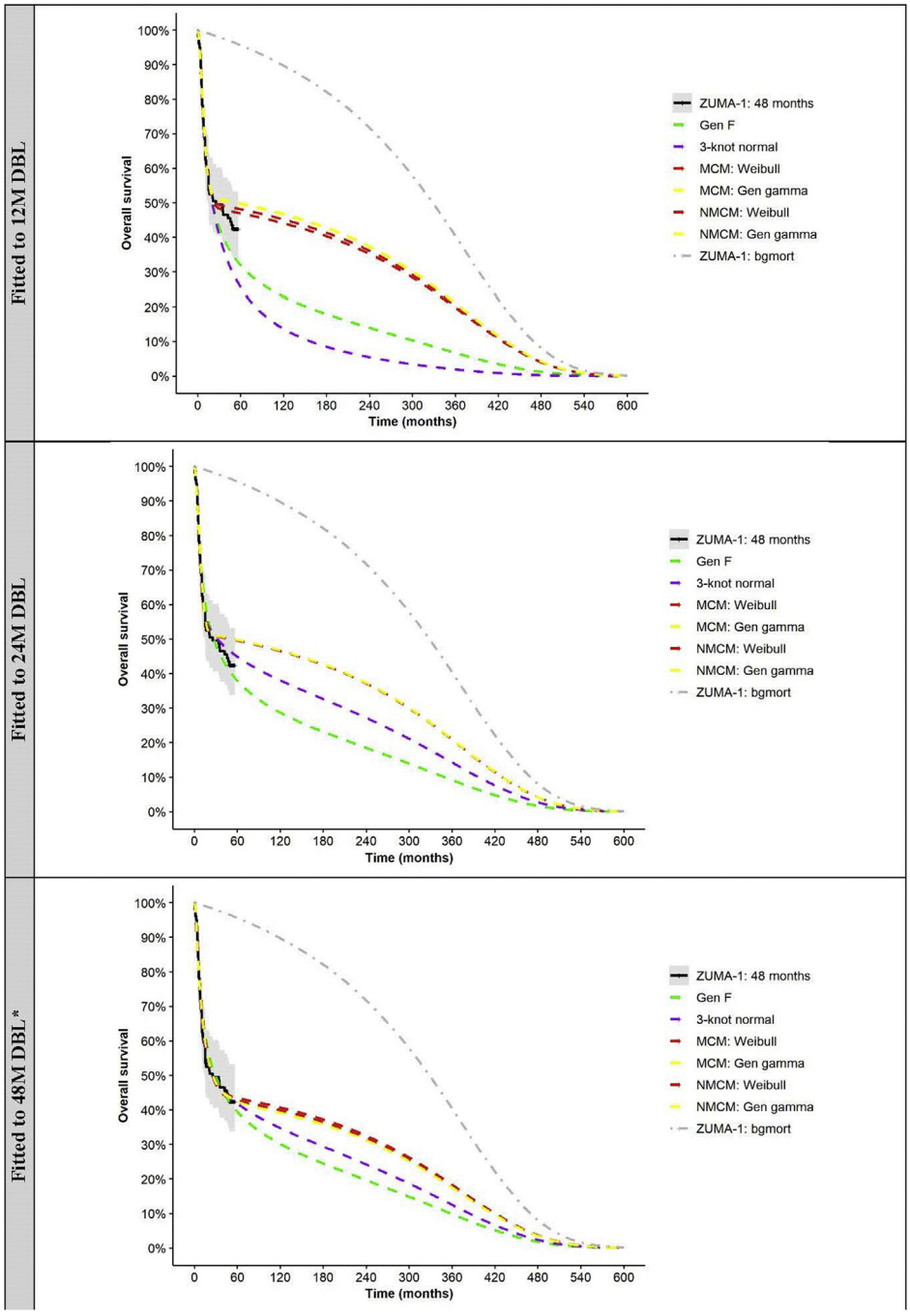
The NMCMs provided generally similar fits to the MCMs; nevertheless, there was slightly more variation between the different NMCMs and the MCMs (though not to a large extent). Importantly, the NMCM and MCM variations based on the same distribution choice (eg, Weibull) exhibited similar fits, implying that the choice of MCM or NMCM is unlikely to result in substantially dissimilar extrapolations—instead, the choice of distribution is of greater influence on the model fit. This is aligned with the findings of previous studies that have explored NMCMs and MCMs.[23,24]

Outside the 48-month time point, all the standard parametric model estimates were within the 95% confidence interval (CI) of the KM estimator up until 18 months, but after 52 months none of the estimates fell within the 95% CI (see Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015). This suggests a systematic underestimation of survival for all the standard parametric models, which is aligned with the results shown in Figure 2. The MCM estimates fell within the 95% CI for all time points for all but the exponential and lognormal MCMs fitted to the 12-month data snapshot, where the estimates fell outside the 95% CI after 24- and 27-month time points, respectively. Nevertheless, as the data matured, these models converged to that of the other distributional forms.

Presentation of "Most Plausible" Models

For simplicity, we selected the "most plausible" models fitted to the 12-month data, from each of the different classes of models: standard, spline-based, and both types of cure-based models. Models were selected according to visual and statistical goodness of fit and clinical plausibility of projections. This was to avoid comparing a large quantity of models, some of which produce near-identical extrapolations, or extrapolations that are clearly implausible. The resultant extrapolations for these models are presented in Figure 3, with the reasons for selecting these models provided in the Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015.

Figure 3. Comparison of model fits to each data snapshot. The 36M plot is provided in the Supplemental Appendix for completeness, but is not presented here in the interest of conciseness as it similar to the 24M plot.



Bgmort indicates background mortality; DBL, database lock; Gen, generalized; M, month; MCM, mixture-cure model; NMCM, non-mixture-cure model.

Comparison of "Most Plausible" Models

A summary of how the different models performed is provided in Table 1. Based on the models fitted to the 12-month data snapshot, the cure-based models provided the closest fit to the KM estimate based on the 48-month data snapshot. Statistical goodness-of-fit scores for all models are provided in the Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015.

DLBCL is a highly aggressive disease with patients relapsing quickly if they do not respond to CAR T-cell therapy. Clinically, it is believed that patients who survive past 12 months are substantially more likely to benefit from long-term survival; hence, below we calculate the proportion of patients projected to survive until 48 months (minimum follow-up in latest data snapshot) given that they survived until 12 months (minimum follow-up in earliest data snapshot). We find that cure-based models were able to closely approximate the ZUMA-1 data and the noncure-based

Table 1. Model performance for "most plausible" models.

Model	Error in point-estimate of survival at 48-mo when fitted to 12-mo data snapshot, %	% Alive at 48-mo given alive at 12-mo, %	VD for "most plausible" models fitted to each data snapshot evaluated at 48-mo, % (rank)
Kaplan-Meier	—	72.13	—
Gen F	−8.36	56.82	15.7 (5)
3-knot normal	−13.48	48.41	18.4 (6)
MCM: Weibull	5.09	78.52	13.0 (2)
MCM: Ggam	6.82	81.52	14.8 (4)
NMCM: Weibull	3.89	76.40	12.2 (1)
NMCM: Ggam	6.60	81.07	14.6 (3)

Gen indicates generalized; Ggam, generalized gamma; MCM, mixture-cure model; NMCM, non-mixture-cure model; VD, Vadgama Deviance.

models fitted to the more mature data seemed to underestimate conditional survival to a greater extent than any of the cure models (eg, the standard generalized F model fitted to the 48-month data snapshot underestimated conditional survival by approximately 6%).

In addition to the numerical results considered, hazard-based plots were also produced (see Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015). Inspection of the hazard plots demonstrated that the cure-based models were capable of better reflecting the shape of the hazard function (ie, an initial peak followed by a fall, then an increase due to all-cause mortality).

The findings from the pragmatic approach taken to provide an overall measure of the goodness of fit for each model across each data snapshot is provided in Table 1 (full results provided in Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015). These results further corroborate that the cure-based models were overall more accurate than the standard and spline-based models—in terms of the VD.

Discussion

In this study, we set out to compare the accuracy of different survival extrapolation methods based on data from the ZUMA-1 clinical trial. As a result of sequential data snapshots of ZUMA-1 being made available for analysis, we were able to explore how extrapolations differed with increasingly maturing data. This type of analysis has not been considered within the context of CAR T-cell therapy previously and provides evidence of an initial emergent (yet uncertain) plateau in the survival curve, which is later shown to be robust with further follow-up data. It can also be seen from our findings that some of the less flexible approaches produce extrapolations that at face value seem similar to the cure-based models; for example, the Gompertz model fitted to the later data snapshots (see Supplemental Appendix in Supplemental Materials found at https://doi.org/10.1016/j.jval.2021.10.015). Nevertheless, inspection of hazard plots show that the underlying hazard function from ZUMA-1 is expected to be nonmonotonic, and so the Gompertz model is unlikely to provide both (1) a good fit to the KM estimate and (2) a plausible long-term extrapolation.

Our analysis demonstrates that the cure-based models (MCMs and NMCMs) provided the most accurate estimates of survival compared with the other parametric approaches (standard and spline-based models). The cure-based models produced the least amount of variation of the models considered in our study, even when using 12 months follow-up data. Previous research in the topic of survival extrapolation for novel cancer therapies led to similar conclusions.[16,25] Nevertheless, relatively little research has been conducted within the context of evaluating survival extrapolations for CAR T-cell therapy specifically, and so the generalizability of findings between different types of cancer immunotherapy is yet to be established.

Although cure-based models may be criticized for appearing overly optimistic based on immature data, the clinical rationale and the shape of the hazard function provide further support for the specification of these models in the case of axi-cel. Therefore, cure-based models may provide a useful survival extrapolation technique to account for the heterogeneity in outcomes seen in CAR T-cell therapy clinical trials. Moreover, as the specification of a cure-model based on the earlier data snapshot was shown to project reasonable and accurate survival outcomes in the longer term, these models should be seriously considered to guide earlier decision making for patient access to these potentially lifesaving treatments.

Although the additional follow-up data from ZUMA-1 reduces the uncertainty in OS estimation, data are not available over a lifetime horizon, and so the choice of the most appropriate extrapolation is still subject to uncertainty. A recent study by Grant et al[24] showed that cure-based models fitted to complete data are unlikely to provide a good fit. It may therefore be the case that the cure models provide a reasonable fit to the currently available data from ZUMA-1, but were complete data available, the same models may not fit the data well. In our research, we observed a good fit to the data and stabilization of cure models after 24 months of follow-up data. Relatedly, cure models require the strong assumption of a "statistically cured" population within the cohort. Nevertheless, to date there are no long-term follow-up studies for patients with DLBCL treated with CAR T-cell therapy.

A particular area of future research should ascertain the longevity of excess mortality of death experienced by patients treated in later lines with CAR T-cell therapies. These patients typically endure multiple prior lines of therapy, for example the population in ZUMA-1 had a median of 3 prior lines of therapy and 21% of patients had relapsed post SCT. It is unclear whether these patients would experience the same excess mortality if treated in earlier lines. Nevertheless, in the context of HTA, this uncertainty is usually adequately accounted for in the economic models. Treatments given after axi-cel may also influence long-term survival; nevertheless, in the context of ZUMA-1, only 3 patients went to receive an SCT post axi-cel.[26] It was not possible to robustly undertake meaningful survival analyses for this very small subgroup of patients, and it would be unlikely to have a large effect on

long-term survival estimates. Similarly, in the spirit of survival analysis methods commonly used in HTA, we did not perform statistical analysis to predict long-term survival at the individual patient level because it requires granular data on prognostic factors and further reduces our sample size; nevertheless, this is an area of future research. As we have showed, thus far, cure models seem to provide a reasonable fit for the purposes of decision making in the short term.

Further research is required to more conclusively ascertain how to best extrapolate survival outcomes for patients treated with CAR T-cell therapy. Our study focuses on methods to directly estimate OS, through specifying a statistical model. Nevertheless, other approaches may also be helpful to consider within the context of informing HTA. For example, Batteson et al[27] considered a range of approaches to both directly and indirectly estimate OS for patients with melanoma treated with adjuvant nivolumab. Repeating the analyses described in our study and those conducted by others for different patient populations would further increase the understanding of the most appropriate methods to produce suitable survival estimates for HTA. Further research is also needed to establish the "gold standard" metrics required to demonstrate the accuracy of extrapolated models to the KM estimator at increasingly mature data cuts.

Conclusions

In this study, we set out to establish which methods may be appropriate to estimate the lifetime survival for patients with R/R DLBCL treated with axi-cel, a novel CAR T-cell therapy, and which methods may be deemed inappropriate. Paraphrasing George Box,[28] all of the models considered within our analysis are inherently "wrong" (in the sense that they are a simplification of reality), but the estimation of an accurate survival model is clearly useful for HTA decision making.

Through our approach, we have shown that accurate modeling is possible even when based on limited OS data in aggressive lymphoma. This was made possible through the implementation of a more flexible survival modeling approach, namely, cure models that acknowledge the heterogeneity of patients and outcomes through the specification of a "cure fraction." Although cure models were moderately optimistic, they provided the most accurate estimation of long-term survival on several metrics. This framework can provide HTA decision makers with a more realistic tool to estimate the long-term benefit of innovative and potentially transformative therapies such as axi-cel in 3-level R/R DLBCL and therefore better inform cost-effective resource allocation decision making in the presence of "immature" trial data.

Although our findings are consistent with similar studies performed for different types of immunotherapy in other cancers, further research is still required to understand how these different modeling methods compare in other indications, with even more mature data, across different types of cancer immunotherapy, and whether a cure assumption still holds in the long-term for these patients and standardization of metrics for reporting model accuracy compared with the KM estimator. Although out of scope for this piece of research, we acknowledge that cure models can be used in a variety of disease; nevertheless, the pathophysiology of the disease and mechanism of action of subsequent treatment must be carefully considered before they are implemented.

Supplemental Material

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2021.10.015.

Article and Author Information

Accepted for Publication: October 31, 2021

Published Online: xxxx

doi: https://doi.org/10.1016/j.jval.2021.10.015

Author Affiliations: Kite, a Gilead Company, Stockley Park, Uxbridge, England, UK (Vadgama, Bashir, Spooner); Department of Medicine, University College London, England, UK (Vadgama); Delta Hat Ltd, Nottingham, England, UK (Mann, Bullement); Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, England, UK (Collins); School of Health and Related Research, University of Sheffield, Sheffield, England, UK (Bullement).

Correspondence: Sachin Vadgama, MSc, Gilead Science Europe, 2 Roundwood Avenue, Uxbridge, UB11 1AF. Email: sachin.vadgama@gilead.com

Author Contributions: *Concept and Design*: Vadgama, Mann, Bashir, Bullement
Acquisition of data: Vadgama, Bashir
Analysis and interpretation of data: Vadgama, Mann, Bashir, Spooner, Collins, Bullement
Drafting of manuscript: Vadgama, Mann, Bashir, Spooner, Collins, Bullement
Critical revision of paper for important intellectual content: Vadgama, Mann, Bashir, Collins, Bullement
Statistical analysis: Vadgama, Mann, Bullement
Provision of study materials or patients: Vadgama
Obtaining funds: Vadgama
Administrative, technical, or logistical support: Vadgama, Spooner
Supervision: Vadgama

Conflict of Interest Disclosures: Messrs Vadgama and Bashir and Ms Spooner reported receiving funding and grants from Kite, a Gilead Company. Messrs Vadgama, Mann, Bashir, and Bullement and Ms Spooner reported receiving support for conception, analysis, and manuscript writing from Delta Hat Limited. Mr Vadgama reported receiving grants from Kite to fund his position within the company and reported receiving an honorary research contract with the Department of Medicine, University College London. Messrs Mann and Bullement are employees of Delta Hat Limited, a privately held consultancy that received funding from Kite, a Gilead Company, and reported over the past 36 months working on projects with several pharmaceutical companies. None of these projects are directly related to the work reported within the manuscript. Ms Spooner is an employee of Kite, a Gilead Company, and reported stock ownership in Gilead Sciences. Dr Collins reported receiving funding from Kite, a Gilead Company; reported receiving consulting fees from Gilead, Novartis, Takeda, Roche, Bristol Myers Squibb, Merck & Co, Astra Zeneca, Daiichi Sankyo, Incyte, Celleron, and ADC Therapeutics; reported receiving payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events from Gilead, Novartis, Takeda, Roche, Bristol Myers Squibb, Merck & Co, Astra Zeneca, Daiichi Sankyo, Incyte, Celleron, and ADC Therapeutics; reported receiving support for attending meetings and/or travel from Takeda and Roche; and reported participation on a Data Safety Monitoring Boards or Advisory Boards for Gilead, Novartis, Takeda, Roche, Bristol Myers Squibb, Merck & Co, Astra Zeneca, Daiichi Sankyo, Incyte, Celleron, and ADC Therapeutics. No other disclosures were reported.

Funding/Support: This research received funding from Kite, A Gilead Company.

Acknowledgment: The authors thank the patients, families, investigators, coinvestigators, and study team members for the ZUMA-1 trial, without whom this study would not have been possible. The authors also thank Julia Snider and Gab Castaigne, who are employees of Kite, a Gilead Company, for their review of the manuscript.

REFERENCES

1. Yescarta: summary of product characteristics (SmPC). European Medicines Agency (EMA). https://www.ema.europa.eu/en/documents/product-

information/yescarta-epar-product-information_en.pdf. Accessed November 18, 2020.
2. Definition of CAR T-cell therapy - NCI dictionary of cancer terms. National Cancer Institute (NCI). https://www.cancer.gov/publications/dictionaries/cancer-terms/def/car-t-cell-therapy. Accessed November 18, 2020.
3. Chaplin S. CAR-T cell therapy: personalised immunotherapy for cancer. *Prescriber*. 2018;29(12):26–29.
4. Neelapu SS, Locke FL, Bartlett NL, et al. Axicabtagene Ciloleucel CAR T-cell therapy in refractory large B-cell lymphoma. *N Engl J Med*. 2017;377(26):2531–2544.
5. Locke FL, Ghobadi A, Jacobson CA, et al. Long-term safety and activity of axicabtagene ciloleucel in refractory large B-cell lymphoma (ZUMA-1): a single-arm, multicentre, phase 1-2 trial. *Lancet Oncol*. 2019;20(1):31–42.
6. Neelapu SS, Rossi JM, Jacobson CA, et al. CD19-loss with preservation of other B-cell lineage features in patients with large B cell lymphoma who relapsed post-axi-cel. *Blood*. 2019;134(suppl 1):203.
7. Jacobson C, Locke FL, Ghobadi A, et al. Long-term survival and gradual recovery of B cells in patients with refractory large B cell lymphoma treated with axicabtagene ciloleucel (axi-cel). *Blood*. 2020;136(suppl 1):40–42.
8. Bishop MR, Dean RM, Steinberg SM, et al. Clinical evidence of a graft-versus-lymphoma effect against relapsed diffuse large B-cell lymphoma after allogeneic hematopoietic stem-cell transplantation. *Ann Oncol*. 2008;19(11):1935–1940.
9. Vercellino L, Di Blasi R, Kanoun S, et al. Predictive factors of early progression after CAR T-cell therapy in relapsed/refractory diffuse large B-cell lymphoma. *Blood Adv*. 2020;4(22):5607–5615.
10. Awasthi R, Pacaud L, Waldron E, et al. Tisagenlecleucel cellular kinetics, dose, and immunogenicity in relation to clinical factors in relapsed/refractory DLBCL. *Blood Adv*. 2020;4(3):560–572.
11. Bell Gorrod H, Kearns B, Stevens J, et al. A review of survival analysis methods used in NICE technology appraisals of cancer treatments: consistency, limitations, and areas for improvement. *Med Decis Making*. 2019;39(8):899–909.
12. 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.
13. Latimer NR. NICE DSU technical support: document 14:. Survival analysis for economic evaluations alongside clinical trials - extrapolation with patient-level data. National Institute for Health and Care Excellence (NICE). http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf. Accessed January 18, 2019.
14. TA559: axicabtagene ciloleucel for treating diffuse large B-cell lymphoma and primary mediastinal large B-cell lymphoma after 2 or more systemic therapies. National Institute for Health and Care Excellence (NICE). https://www.nice.org.uk/guidance/ta559. Accessed November 18, 2020.
15. 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.
16. Ouwens MJNM, Mukhopadhyay P, Zhang Y, Huang M, Latimer N, Briggs A. Estimating lifetime benefits associated with immuno-oncology therapies: challenges and approaches for overall survival extrapolations. *Pharmacoeconomics*. 2019;37(9):1129–1138.
17. Klijn S, Hofstra M, Malcolm B, Johannesen KM. CN1 - validating survival extrapolations in first line treatment of renal cell carcinoma using literature-based conditional survival probabilities. *Value in Health*. 2018;21(suppl 3):S3.
18. Bagust A, Beale S. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. *Med Decis Making*. 2014;34(3):343–351.
19. Rutherford MJ, Lambert PC, Sweeting MJ, et al. NICE DSU technical support: document 21:. Flexible methods for survival analysis. National Institute for Health and Care Excellence (NICE). http://nicedsu.org.uk/wp-content/uploads/2020/11/NICE-DSU-Flex-Surv-TSD-21_Final_alt_text.pdf. Accessed December 13, 2020.
20. Grieve R, Hawkins N, Pennington M. Extrapolation of survival data in cost-effectiveness analyses: improving the current state of play. *Med Decis Making*. 2013;33(6):740–742.
21. Collett D. *Modelling Survival Data in Medical Research*. 3rd ed. New York, NY: Chapman and Hall/CRC; 2014.
22. Dickman PW, Coviello E. Estimating and modeling relative survival. *STATA J*. 2015;15(1):186–215.
23. Martinez EZ, Achcar JA, Jácome AA, Santos JS. Mixture and non-mixture cure fraction models based on the generalized modified Weibull distribution with an application to gastric cancer data. *Comput Methods Programs Biomed*. 2013;112(3):343–355.
24. Grant TS, Burns D, Kiff C, Lee D. A case study examining the usefulness of cure modelling for the prediction of survival based on data maturity. *Pharmacoeconomics*. 2020;38(4):385–395.
25. Bullement A, Latimer NR, Bell Gorrod H. Survival extrapolation in cancer immunotherapy: a validation-based case study. *Value in Health*. 2019;22(3):276–283.
26. ZUMA-1 clinical study report (data on file). Gilead Sciences (Kite, A Gilead Company). Accessed March 16, 2021.
27. Batteson R, Hart R, Hemstock M, et al. Modelling survival of patients treated with adjuvant nivolumab who have melanoma with lymph node involvement or metastatic disease after complete resection. *Pharmacoecon Open*. 2020;4(2):343–351.
28. Box GEP. Science and statistics. *J Am Stat Assoc*. 1976;71(356):791–799.