INSPIRE HEALTH

Inspire Health Journal
ISSN 3110-9438, Vol. 1, No. 2, 2026
doi.org/10.65718/inspireHealth.2026.2008

InspireQuill.org
Inspire Academic Press (Canada)

Research Paper OPEN ACCESS

Bayesian Hierarchical Mixture Cure Modeling for Survival Analysis in Oncology Trials

Prachi Patel [1, *]

[1] Harrisburg University of Science and Technology, PA, United States
[*] Corresponding Author: prachi25898@gmail.com

Abstract

Survival analysis is an important part of clinical trials in oncology, especially when it comes to figuring out the long-term benefits of treatment and informing health technology assessment (HTA). This paper presents a Bayesian hierarchical framework for mixture cure models, addressing the complexities in estimating survival outcomes in oncology trials. Traditional survival models often fall short when an intervention is effective only for a subset of the population, leading to substantial right censoring and potentially misleading conclusions. By leveraging the correlation between multiple event types, such as overall survival (OS) and progression-free survival (PFS), the proposed model allows for the borrowing of information across event types, resulting in more accurate extrapolations beyond observed data. The model's effectiveness is demonstrated using the Checkmate 067 trial data, showing improved stability in cure fraction estimates, even with limited follow-up. The approach offers significant advantages for health economic evaluations, particularly in estimating long-term survival and cost-effectiveness metrics like quality-adjusted life years (QALYs). The hierarchical model's ability to incorporate prior knowledge and its robustness against small sample sizes underscore its value in clinical decision-making and policy formulation.

Keywords: Oncology trials; Mixture cure model; Bayesian; Survival analysis; Oncology; Health economic evaluation; Long-term extrapolation.

1. Introduction

Health Technology Assessment (HTA) organizations, such as the National Institute for Health and Care Excellence (NICE) in England, place a lot of emphasis on interventions that affect the time until a significant event, like disease progression or mortality in cancer trials [1]. A new treatment might, for instance, try to prolong the period of time before a patient dies or their illness worsens. HTA modeling thus relies heavily on survival data [2-3].
Survival data from clinical trials are often reported both during the trial and at its conclusion. This enables ongoing evaluations and can guide decisions such as halting a trial early due to lack of efficacy or safety concerns. However, immediate utilization of collected data is usually impractical due to the necessity of cleaning and addressing data quality issues, including missing information. Consequently, snapshots of trial data, referred to as data-cuts, are taken at predefined intervals based on study protocols. These intervals may be defined by specific dates, follow-up durations, or the number of enrolled patients.

Academic Editor: Jaafar Alghazo

Received: 17/09/2025
Revised: 23/11/2025
Accepted: 18/02/2025
Published: 11/03/2026

Citation
Patel, P. (2026). Bayesian Hierarchical Mixture Cure Modeling for Survival Analysis in Oncology Trials. *Inspire Health Journal, 1*(2), 77-94.

The *mean survival time*, which represents long-term treatment effects, must be estimated in order to properly evaluate the health and economic impact of a new intervention using survival data. Mean survival time necessitates extrapolation beyond the observed data period, in contrast to commonly reported survival statistics like median survival, which are usually employed in biostatistical analysis. Given that available data (e.g., from randomized trials) often cover a limited timeframe and exhibit substantial censoring, it becomes necessary to *extrapolate* survival curves to estimate outcomes over a patient's lifetime. Consequently, HTA modeling generally employs fully parametric survival analysis rather than semiparametric approaches like Cox regression, as recommended in NICE's Decision Support Unit (DSU) guidance [1].

One of the striking advances in the field of cancer research in recent years has been the introduction of immuno-oncology treatment, which aids the immune system in identifying and destroying cancerous cells [4]. Immune checkpoint inhibitors, interleukin-2 (IL-2), and oncolytic virus therapy have all been used for melanoma management. Of these, immune checkpoint blockade combination therapies have proved significantly successful in triggering immune responses against cancer, causing tumor regression in many patients [5]. The development has led to better survival outcomes and higher rates of long-term survival (LTS) among cancer patients. But complete responders (CRs) continue to have poorer survival than the overall population.

In clinical trials evaluating melanoma therapies such as ipilimumab and nivolumab, survival curves often exhibit a plateau effect, where a substantial proportion of LTS patients maintain stable survival probabilities over time [6], [7]. This suggests that a fraction of patients achieve durable remission, although uncertainty remains due to small sample sizes and potential data limitations. To account for this phenomenon, researchers commonly use *mixture cure models* (MCMs), which classify patients into cured and non-cured groups. In many MCM applications, the survival of the non-cured fraction is modeled using Cox proportional hazard models.

However, survival plateaus indicating LTS proportions may not be evident at the time of a data-cut, including the final planned data-cut of a trial. Therefore, it is necessary to extrapolate survival curves to account for potential plateaus in long-term projections. Ideally, these extrapolations should remain consistent across multiple data-cuts [8]. This study explores principled methods for achieving such consistency when trials include multiple treatment arms and endpoints. By considering multiple endpoints simultaneously, information about the plateau from one endpoint may help refine estimates for another, ensuring internal coherence and maximizing the utility of available data.

Within the frequentist paradigm, prior work includes multilevel modeling approaches in MCMs that employ random effects to account for hierarchical structures in hazard functions and cure probability estimates [9]. Some studies have explored correlation structures between survival outcomes and cure fractions using bivariate normal distributions [10, 11]. Instead of this approach, the study imposes dependencies through hyperparameters. To the knowledge, no existing research has developed a Bayesian fully parametric MCM with a multi-level structure for analyzing multiple endpoints within a trial.

This paper is organized as follows. Section 2 describes the motivating dataset. Section 3 introduces the standard mixture cure model, followed by its extension into a Bayesian hierarchical framework. Section 4 applies the novel modeling approach to the dataset and compares it with independent models. Finally, Section 5 presents key findings and outlines future research directions.

2. Background

This study is driven by a comprehensive evaluation of melanoma treatment strategies, with a particular focus on the extitCheckMate 067 trial [6-7, 12]. This phase 3, randomized, double-blind clinical trial enrolled individuals aged 18 years or older diagnosed with previously untreated, unresectable stage III or IV melanoma. The primary objective was to compare the therapeutic effectiveness of nivolumab alone, nivolumab combined with ipilimumab, and ipilimumab alone in patients with metastatic melanoma. Participants were assigned to one of the three

treatment groups in a 1:1:1 ratio, with stratification based on PD-L1 expression levels, BRAF mutation status, and the American Joint Committee on Cancer (AJCC) metastasis stage.

The trial assessed the impact of nivolumab, an inhibitor targeting the programmed death-1 (PD-1) receptor, and ipilimumab, a monoclonal antibody directed against cytotoxic T-lymphocyte–associated antigen 4 (CTLA-4). A total of 945 participants were allocated to the following three treatment regimens:

1) Ipilimumab monotherapy (3 mg/kg IV infusion every 3 weeks for four cycles)
2) Nivolumab monotherapy (3 mg/kg IV infusion every 2 weeks)
3) Combination therapy consisting of nivolumab (1 mg/kg IV) and ipilimumab (3 mg/kg IV) every 3 weeks for four cycles, followed by nivolumab (3 mg/kg IV) every 2 weeks

The study employed two primary endpoints: *progression-free survival* (PFS), defined as the duration from randomization until disease advancement or death, and *overall survival* (OS), measured as the time from randomization to death from any cause. The analysis incorporated patient-level data on PFS and OS, along with relevant covariates such as sex, age at trial enrollment, and trial site location.

To evaluate survival extrapolation, I considered three data-cuts: the final study follow-up at 60 months, an intermediate cut at 30 months, and an early cut at 12 months. The 30month cut aligns with the study's planned 28-month data-cut, representing a stage where meaningful inferences about OS can be drawn. Meanwhile, the 12-month cut was chosen based on the median PFS durations across treatment groups, allowing for an assessment of early-stage survival trends. Earlier data-cuts would likely yield insufficient information for reliable extrapolation. Patients who had not yet experienced an event at the time of a data-cut were censored accordingly.

The reader is encouraged to refer to previous publications for a comprehensive analysis of the CheckMate 067 trial outcomes, particularly the 5-year database lock (DBL) assessment [7]. In brief, the median (95% CI) progression-free survival (PFS) durations for the different treatment arms were 3.12 (2.85-3.78) months for ipilimumab, 7.45 (5.60-11.30) months for nivolumab, and 12.30 (9.80-21.50) months for the combination therapy involving nivolumab and ipilimumab. Likewise, the median (95% CI) overall survival (OS) durations were 21.5 (18.10-27.20) months for ipilimumab, 39.2 (32.80-63.50) months for nivolumab, while for the combination therapy, the median was not attained.

Figure 1 presents the Kaplan-Meier survival curves for both progression-free survival (PFS) and overall survival (OS) across all therapeutic groups in the CheckMate 067 phase 3 study. The observed trends in PFS and OS suggest a prolonged survival plateau, with the PFS curves displaying this phenomenon more distinctly. Updated results from a 7.5year follow-up further confirm the persistence of treatment benefits, particularly in the cohort receiving the combination of nivolumab and ipilimumab, thereby strengthening the evidence for sustained long-term survival [13].

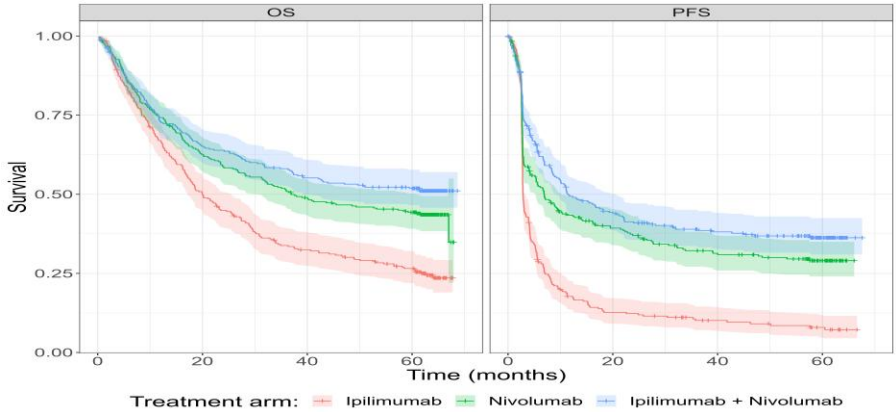

Figure 1. Kaplan-Meier survival curves for LFI and TS, including 95% confidence intervals, for patients treated with αblockers, β-blockers, or their combination.

A standard frequentist approach has been previously applied to these data, employing separate mixture cure models for each treatment regimen and outcome (PFS or OS) [14]. The estimated proportion of long-term survivors (LTSs) for OS ranged from 16%–26% for ipilimumab, 38%–46% for nivolumab, and 49%–54% for the combination therapy. Similarly, for PFS, the estimated LTS proportions were 9%–13% for ipilimumab, 29%–33% for nivolumab, and 38%–40% for the combination.

3. Modeling Framework

Consider a study where N individuals are observed, with recorded (non-negative) times to either disease progression (T_j^{prog}) or death (T_j^{death}) for each subject $j = 1,...,N$. Let T_j denote the observed event time for subject j. In the motivating example, separate sets of T_j values exist for both PFS and OS. Specifically, PFS times are defined as in Equation 1.

$$T_j = \min(T_j^{prog}, T_j^{death}) \quad (1)$$

whereas OS times are given by $T_j = T_j^{death}$. Since some observations may be censored, I introduce a censoring time C_j and an event indicator $\delta_j = \mathrm{I}(T_j < C_j)$, where $\delta_j = 1$ signifies an observed event, and $\delta_j = 0$ indicates censoring. The actual recorded time is therefore $t_j = \min(T_j, C_j)$.

Studies often collect covariate information on each subject, represented as a vector $z_j = (z_{1j},...,z_{Pj})^\top$. These covariates may include demographic factors, comorbidities, and treatment assignment. The complete dataset for each individual can be summarized as $D_j = (t_j, \delta_j, z_j)$. For modeling, I assume that the observed times follow a parametric probability distribution: $t_j \sim f(t_j \mid \psi)$, where the parameter set $\psi = (\lambda(z), \eta(z))$ comprises a location parameter $\eta(z)$ (indicating central tendency or scale) and a shape parameter $\lambda(z)$ (describing dispersion). Typically, only $\eta(z)$ directly depends on covariates, simplifying the model. To ensure positivity, a generalized linear formulation is employed as shown in Equation 2.

$$h(\eta) = \gamma_0^\eta + {}^{\mathrm{X}}\gamma_p^\eta z_{pj} [+...] \ , \ p=1 \quad (2)$$

where γ_0^η is an intercept, and γ_p^η are the regression coefficients. The function $h(\cdot)$ is typically the logarithm.

3.1. Standard Mixture Cure Model

If evidence suggests a sustained survival plateau, a mixture cure model can be employed. The key parameter in such models is the cure fraction, denoted by π. In this context, cures do not imply disease eradication but rather a durable response leading to long-term survival. Because mixture cure models explicitly classify subjects into cured and noncured groups, the model parameters are split accordingly $\phi = (\psi^b, \psi^u)$, where b refers to the background risk and u to the excess risk due to disease. The survival function for individual j in a mixture cure model is given in Equation 3.

$$S(t_j \mid \phi, z_j) = S_b(t_j \mid \psi^b, z_j)[\pi + (1 - \pi)S_u(t_j \mid \psi^u, z_j)] \quad (3)$$

where $S(t) = 1 - \int_0^t f(s \mid \psi)\,\mathrm{d}s$ denotes the survival probability at time t. Here, $S_b(t \mid \psi^b, \boldsymbol{z})$ represents background mortality, while $S_u(t \mid \psi^u, \boldsymbol{z})$ models excess mortality due to disease. This formulation consists of two sub-models: (i) an incidence model for the cure fraction π, and (ii) a latency model for the survival distribution among non-cured individuals. These components are interlinked, as the uncured group cannot be directly observed. For further discussion on cure models in oncology, see [15].

The model can be extended to accommodate multiple treatments, denoted as T, where in the given scenario, $T = 3$. The cure fraction for treatment $t = 1,...,T$ can be represented using the fixed-effect model as in Equation 4.

$$\mathrm{logit}(\theta_t) = \gamma_t^\theta \ [+ \dots] \quad (4)$$

where γ_t^{θ} are the regression parameters capturing the influence of treatment t. The term [+...] can incorporate a frailty component in the model. The treatment index for each individual is included in the covariate set **z**, so that θ is replaced by $\theta(\mathbf{z})$ in equation (2).

It is crucial to note that, while incorporating all treatment arms into a single framework, I do not enforce any predefined association among them. Particularly, I do not impose proportional hazards, thus providing enhanced adaptability when the proportional hazards assumption does not hold.

Several challenges arise when applying traditional cure modeling techniques in a health technology assessment (HTA) context. Firstly, when endpoints are correlated, such as the commonly analyzed overall survival (OS) and progression-free survival (PFS), such interdependence should be considered. Information from one set of event times can potentially enhance inference for another. Conducting independent analyses may yield counterintuitive results. In cases involving multiple endpoint types within the same trial, the cure fraction is clinically interpreted as the proportion of long-term survivors (LTS). Consequently, one might expect a single underlying cure fraction. A discrepancy in observed survival plateaus between OS and PFS could present an unintuitive dichotomy in LTS proportions.

Secondly, right-censoring is a common issue in time-toevent data, arising from administrative censoring or loss to follow-up. If censoring occurs too early, accurately modeling survival for HTA becomes challenging. Specifically, OS times are inherently bounded below by PFS times, leading to increased censoring at a fixed study cut-off point. Leveraging multiple endpoints collectively can mitigate data scarcity, improving model fit and extrapolation for long-term survival estimations. This aspect is particularly critical in health economics to accurately assess the overall cost-effectiveness of an intervention.

Furthermore, in scenarios with heavy censoring across all endpoints—such as early interim analyses, the observed data may not adequately reflect the true underlying survival curves, including cure plateaus. Here, integrating supplementary sources of information becomes essential. Bayesian methodologies provide a natural framework for incorporating external evidence, such as historical data or expert opinions. In the absence of strong prior information, one may reasonably assume that cure fractions for different endpoints are drawn from similar distributions. Practically, this can be implemented by controlling the global variance parameter governing the cure fractions. In the limiting case where $\sigma_t = 0$, complete data pooling occurs, effectively reducing endpoint-specific variability [16].

3.2. Hierarchical Mixture Cure Model

In this section, I introduce an extended hierarchical mixture cure model to jointly analyze multiple endpoints and address the issues highlighted in Section III-A1. I introduce an additional index to denote the endpoint type, where $e = 1,...,E$ (typically, $E = 2$ for PFS and OS). A separate mixture cure model, analogous to equation (2), is specified for each endpoint e.

I explore three alternative approaches for estimating the cure fraction for treatment t and endpoint e, denoted as θ_{te}.

Firstly, the standard approach involves modeling each cure fraction independently without pooling which assumes complete independence across endpoints.

$$\theta_{te} \perp\!\!\!\perp \theta_{te'}, \qquad e,e' = 1,...,E. \tag{5}$$

Secondly, I consider a fully pooled model where a common cure fraction is assumed across all endpoints:

$$\theta_{te} = \theta_{te'}, \qquad e,e' = 1,...,E. \tag{6}$$

This approach leverages the assumption that, over sufficiently long follow-up periods, the cure fraction remains consistent across endpoints. In particular, PFS can serve as a proxy for OS when OS data are heavily censored. However, this method may overlook genuine variability in cure fractions across different endpoints.

Lastly, I propose a Bayesian hierarchical framework that balances these two extremes by introducing partial pooling, allowing information borrowing between endpoints while preserving endpoint-specific heterogeneity. This hierarchical mixture cure model optimally integrates data across endpoints to improve inference reliability.
Lastly, the methodology introduced in this work presents a middle ground between the two extremes, referred to as partial pooling. A hierarchical model is formulated for the remission fraction, assuming exchangeability across all response variables *m* within each therapeutic group *n*. It is postulated that a singular, overarching ”global” remission fraction exists, from which individual response-specific remission fractions are derived, consistent with equation (??). However, under this framework, the parameters $\rho_n, \gamma_n{}^{\rho}$ are no longer tied to distinct responses but instead represent globally shared values. By leveraging this global remission fraction, the second-tier response-specific remission fractions are expressed as in Equation 7.

$$\text{logit}(\rho_{nm}) \sim \text{Normal}(\mu_n, \tau_n^2),\ m = 1, \ldots, M. \quad (7)$$

where $\mu_n = \text{logit}(\rho_n)$ and $\tau_n{}^2$ signifies the variance across different responses. In this scenario, the response-specific remission fractions are denoted as ρ_{RFS}, ρ_{EFS}. Figure 2 illustrates a directed acyclic graphical (DAG) representation of the Bayesian hierarchical mixture remission model for two responses. Corresponding DAGs for the complete pooling and no pooling approaches are provided in the Appendix.
It is crucial to highlight that pooling is strictly confined to the remission fractions. The survival durations remain unpooled; parameters linked to the non-remission survival models (ψ, γ^{λ} in the latency model) are separately estimated using survival data for each response. For example, under the complete pooling assumption, RFS and EFS are not constrained to follow identical survival distributions. However, in both the complete and partial pooling frameworks, an indirect pooling mechanism influences the latency models via the incidence model. Consequently, the resemblance in remission fractions shapes the overall structure of the survival distributions.

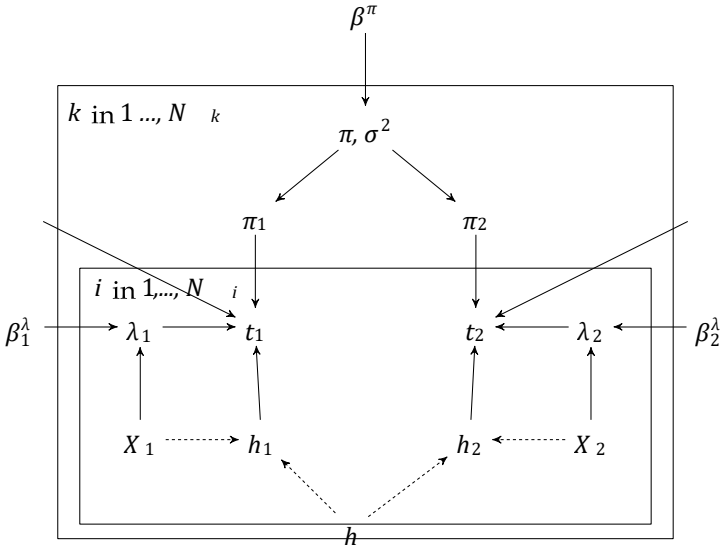

Figure 2. A hierarchical mixture cure model DAG for a trial with two end-points. Solid lines represent stochastic and dashed lines deterministic relationships.

4. Results

This section applies the hierarchical models to the CheckMate 067 dataset. The modeling choices specific to this analysis are first detailed, followed by a presentation of the results. The computational implementation was performed using the Stan inference engine [17], executed from R v.4.3.1 [18] via the rstan package on a Windows 10 system. Each model was run with 4 Markov chains of 1000 iterations each, incorporating a burn-in period of

100 iterations. Convergence diagnostics included checking effective sample sizes, evaluating Markov chain standard error, and conducting posterior predictive checks. Additional details concerning the algorithm and model validation procedures are included in the Appendix.

The developed code has been structured into an R package, facilitating generalizable analyses applicable to an arbitrary number of endpoints with new datasets.

4.1. Baseline Survival

In numerous cure fraction models, it is customary to assume $S_b(t,x) = 1$ for all relevant time points, which holds under short-term considerations. However, since the focus extends across an individual's entire lifespan within a Health Technology Assessment (HTA) framework, I incorporate general population all-cause mortality as a competing risk. Consequently, the uncured cohort experiences background mortality risk alongside the primary event of interest.

To incorporate this, I utilized the World Health Organization (WHO) life tables, selecting the latest available data from 2016 [19] to inform the background mortality component within the mixture cure model. The baseline hazards correspond to the expected mortality rates for individuals at the age of event occurrence. The dataset is adjusted based on country, age, and gender, ensuring a refined characterization of patient profiles within the trial. WHO life tables provide conditional death probabilities in five-year intervals up to age 85, followed by a constant annual mortality rate for individuals above 85, assuming no survival beyond 100 years. Further elaboration is available in the supplementary material.

Within a Bayesian framework, multiple approaches exist for modeling background mortality. In this study, WHO hazard point estimates were treated as fixed values. Given the sample size, I assumed that these estimates provide sufficiently accurate mortality rates, rendering uncertainty incorporation less critical. This also ensures consistency across different model fits. The fixed WHO estimates for individual i in equation (2) are denoted as $S_b(t_i \mid \theta_b, x_i) = S^*_i$. It is plausible that trial participants exhibit worse background survival compared to general population estimates, leading to potential overestimation of survival when relying solely on WHO life tables. Several strategies can mitigate this issue. One approach involves leveraging complete responders, clinically confirmed as cured, to estimate a posterior hazard ratio relative to WHO baselines, subsequently used as a prior in a two-step modeling approach. Alternatively, expert knowledge may inform a prior distribution, either directly on the hazard ratio or a more interpretable natural scale, such as mean lifetime, subsequently transformed.

4.2. Prior Specification

To facilitate computation in the latency model, covariates were centered, and vague priors were assigned on a logarithmic scale for overall survival (OS) and progression-free survival (PFS) rate coefficients: $\log(\lambda_{OS})$ and $\log(\lambda_{PFS})$. The baseline scenario assumes:

$$\beta_{PFS,}{}^{\lambda}{}_{0} \sim \text{Normal}(-3, 0.5), \qquad \beta_{OS,}{}^{\lambda}{}_{0} \sim \text{Normal}(-3, 0.5),$$
$$\beta_{PFS,age}{}^{\lambda} \sim \text{Normal}(0, 0.01), \qquad \beta_{OS,age}{}^{\lambda} \sim \text{Normal}(0, 0.01).$$

The prior distributions for auxiliary parameters were contingent on the survival distribution. Specifically, for log-logistic and Weibull distributions, $\psi \sim \text{Gamma}(1,1)$; for Gompertz, $\psi \sim \text{Gamma}(1,1000)$; and for log-normal, the standard deviation prior was $\psi \sim \text{Gamma}(1,2)$. For the incidence model, the global cure fraction was modeled using a fixed-effect logistic regression. Independent priors were assigned to treatment coefficients $\beta_k^{\pi} \sim \text{Normal}(-0.1, 0.2)$.

This implies a prior mean cure fraction of approximately 0.5, with around a 10% chance of exceeding 0.6 and a 1% chance of surpassing 0.7. An alternative approach is defining the cure fraction directly on the $[0,1]$ scale via a Beta prior, $\pi \sim \text{Beta}(a,b)$, simplifying interpretation but making parameter selection less intuitive. The Beta

distribution parameters can be inferred through mean and variance transformation. Prior predictive plots and further discussion are provided in the supplementary material. For complete data analysis, random effect variance on the global cure fraction was modeled using a minimally informative halfNormal prior $\sigma_k \sim \text{Normal}(0,2.5^2)\text{I}(0,\infty)$ [20], where $\text{I}(0,\infty)$ denotes truncation at zero. This choice was optimal for this study. As a general guideline, when the number of groups exceeds five, an uninformative uniform prior on standard deviation suffices, whereas a Gamma(2,0.1) prior may be preferable when hierarchical variance is small [21].

For early-phase trials, where cure fraction estimation is unreliable due to the absence of observable plateaus, Bayesian hierarchical modeling is advantageous. Under this setting, a shared distribution for endpoint cure fractions is reasonable, effectively imposing a narrow prior on between-group standard deviation at the global cure fraction level.

To implement this, I employed a *Penalized Complexity (PC)* prior [22], favoring model simplicity while penalizing deviations. Specifically, the global cure fraction variance was modeled using an exponential prior $\sigma_k \sim \text{Exp}(\rho)$, where ρ is derived by setting a threshold value σ_0 and probability $p(\sigma > \sigma_0) = \alpha$. This leads to the expression $\rho = -\log(\alpha)/\sigma_0$. For instance, for $p(\pi_k > 0.55) = 0.005$, $\sigma_0 = 0.18$ implies $\rho = 25$, leading to $\sigma \sim \text{Exp}(25)$. Similarly, for $p(\pi_k > 0.525) = 0.005$, $\sigma_0 = 0.08$ implies $\rho = 55$, yielding $\sigma \sim \text{Exp}(55)$. The latter prior was applied at the 12-month data-cut. Prior uncertainty implications on cure fraction scale are depicted in Figure 3, with $\rho = 55$ producing a sharper peak favoring the base model.

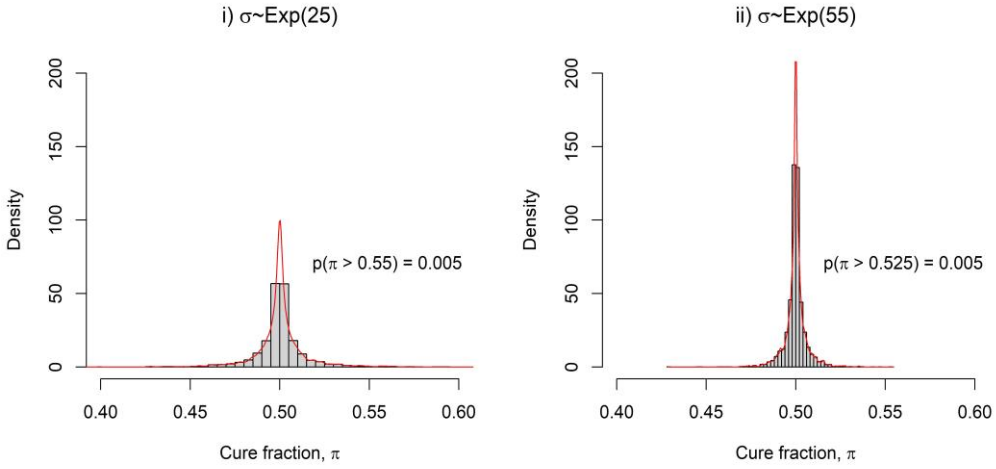

Figure 3. Penalized Complexity Prior Distributions on the Remission Fraction Scale Corresponding to Different Intergroup Variances at the Global Remission Fraction Level of the Bayesian Hierarchical Mixture Remission Model.

4.3. Parametric Models for Sampling Variability

To characterize the sampling variability in the observed event times for OS and PFS, I employed various parametric distributions, including exponential, Weibull, Gompertz, lognormal, and log-logistic. These distributions align with the recommendations set forth by NICE guidelines for Health Technology Assessments (HTA) [1]. It is a common interpretation in the HTA literature that all these distributions should be fitted to a dataset. However, this rigid approach does not necessarily reflect prior knowledge of the problem. In practice, only a subset of these models may be appropriate for a given dataset, making it crucial to evaluate and select the most plausible ones.

4.4. Model Evaluation

To evaluate model adequacy, I employed the widely used Watanabe-Akaike Information Criterion (WAIC), supplemented by leave-one-out cross-validation (LOO), as described in the Appendix [23]. These evaluation measures offer distinct advantages over traditional methods such as AIC and DIC. The WAIC is calculated as the difference between the log pointwise predictive density and the estimated number of effective parameters, where a lower WAIC value signifies a superior model fit. As a fully Bayesian metric, WAIC remains invariant to reparameterization and can be conveniently derived from posterior distributions generated through Stan output.

The hierarchical model parameters and their dimensions are α^{π} (3), τ (3), $\alpha^{\lambda}_{\mathbf{PFS}}$ (2), $\alpha^{\lambda}_{\mathbf{OS}}$ (2), $\psi_{\mathbf{OS}}$ (1), $\psi_{\mathbf{PFS}}$ (1), $\gamma_{\mathbf{OS}}$ (3), $\gamma_{\mathbf{PFS}}$ (3). Limited research has been conducted on model assessment methods specifically tailored to mixture cure models. Traditional methods, such as Schoenfeld residuals for latency sub-models [24] and concordance measures like AUC [25], are applicable primarily when individual cure statuses are known. Since my model assumes a single population cure fraction, adaptations of these techniques are necessary for proper evaluation.

4.5. Analysis Using 60-Month Data

This section details the mixture cure model findings utilizing the complete 60-month dataset from the CheckMate 067 trial. An example showcasing the Bayesian hierarchical model's posterior survival curves is depicted in Figure 4, while corresponding plots for alternative distributions and a combined analysis are available in the Appendix.

The baseline survival trend follows a steady linear decline over time, remaining uniform across the visual representations. This consistency is anticipated due to the application of WHO life tables for the initial estimates. A visual evaluation confirms an adequate model fit, particularly for the PFS Kaplan-Meier curves, which illustrate a pronounced early decline before leveling off. The exponential model aligns well with the tail distribution, which is crucial in survival analysis.

Median Survival Durations: The model fits demonstrate a high degree of similarity. For the exponential model, the projected PFS median survival durations are 6, 10, and 13 months for ipilimumab, nivolumab, and their combination, respectively. These values surpass the Kaplan-Meier estimates by 80%, 35%, and 7%, highlighting eventual convergence. Likewise, the OS median survival durations are estimated at 24, 40, and 58 months. Among uncured individuals, the median survival times are 6 months for PFS and 16 months for OS under the exponential model.

Restricted Mean Survival Time (RMST): RMST provides an intuitive measure, summarizing the average survival duration over a fixed period. Over a 60-month observation window, the RMSTs (95% credible intervals, CrI) for the uncured subgroup under the exponential model are 21.2 (18.9, 23.7) months for OS and 7.02 (6.41, 7.79) months for PFS. RMSTs for the OS endpoint are 30.1 (27.6, 32.5), 37.9 (35.4, 40.3), and 41.3 (38.6, 43.8) months for ipilimumab, nivolumab, and the combination therapy, respectively. For PFS, the RMSTs are estimated at 12.5 (10.5, 14.8), 25.1 (22.0, 28.5), and 28.9 (25.8, 32.1) months. The RMST for cured patients corresponds with the general population estimate of 58.2 months.

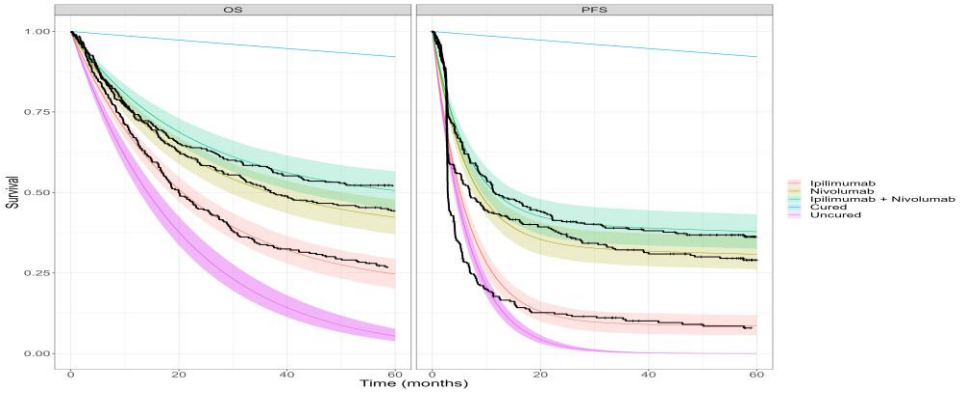

Figure 4. Bayesian hierarchical mixture cure model posterior mean survival curves for uncured fraction assumed exponential for OS and PFS, and for ipilimumab, nivolumab, and combination therapies. The black lines are Kaplan-Meier curves from the overall CheckMate 067 trial data.

Cure Fraction Analysis: Figures 5 and 6 display posterior distributions of cure fractions for the hierarchical and separate mixture cure models, respectively. The global cure fraction posterior distributions are relatively broad, spanning 0.1-0.55 for ipilimumab, 0.2-0.6 for nivolumab, and 0.2-0.65 for the combination treatment. Despite this breadth, the data influence endpoint cure fractions, pulling estimates downward for ipilimumab and upward for the combination treatment. For ipilimumab, the posterior distribution of the PFS cure fraction is close to zero within its credible interval. Meanwhile, for nivolumab and combination treatments, the global cure fraction posterior mean is positioned between the OS and PFS means. The hierarchical model introduces slight shrinkage, adjusting PFS cure fractions upward and OS cure fractions downward compared to the separate model.

Survival Plateau: Survival curves plateau when the uncured survival probability reaches zero. Since background mortality is informed by external WHO life tables, insights beyond this point remain limited. These plateaus occur earlier for PFS and for the log-normal distribution.

Comparison of Hierarchical vs. Separate Models: Differences between the hierarchical and separate models are minimal. Given only two exchangeable endpoints and a relatively weak prior, the effect of hierarchical modeling is constrained. However, slight shrinkage effects are observed, with hierarchical estimates pulling OS and PFS cure fractions toward the global mean.

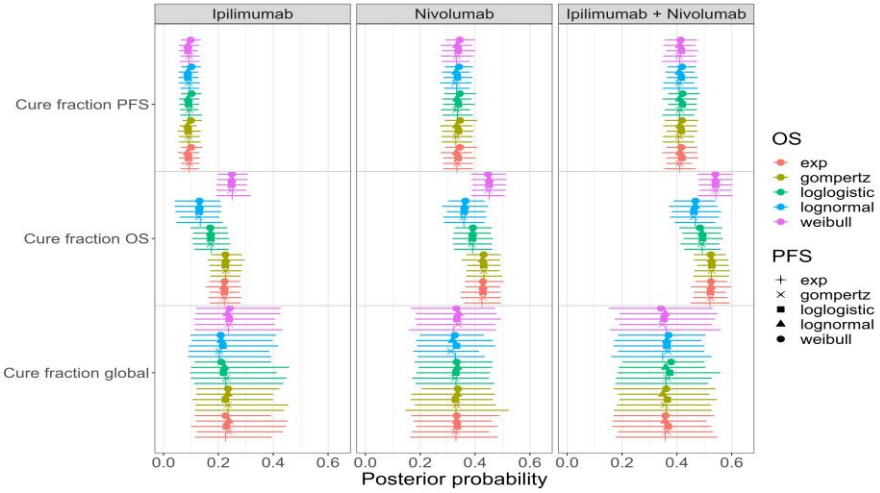

Figure 5. Hierarchical mixture cure model posterior cure fraction forest plots with 95% credible intervals for (i) ipilimumab only (ii) nivolumab only and (iii) combination treatment.

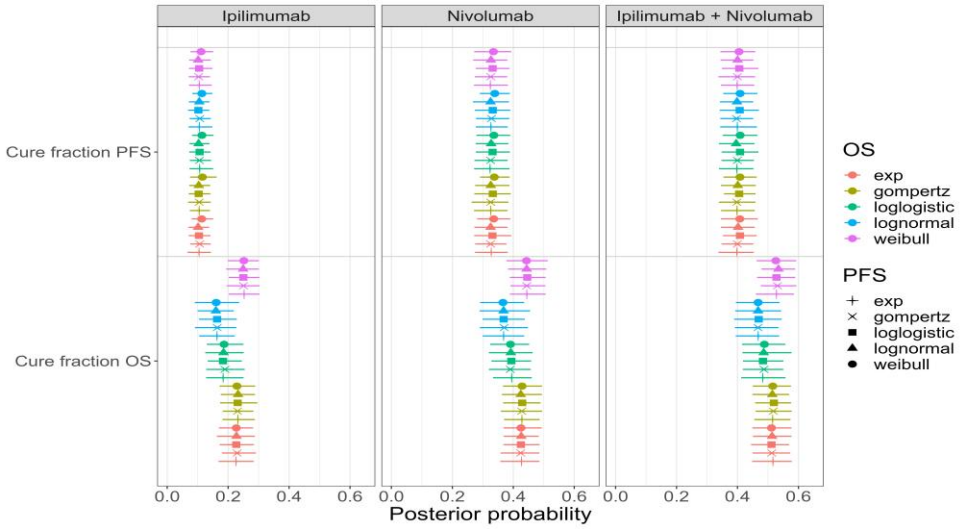

Figure 6. Separate mixture cure models posterior cure fraction forest plots with 95% credible intervals for (i) ipilimumab only (ii) nivolumab only and (iii) combination treatment.

4.6. Analysis of Data-Cut Results

In this section, we analyze the effects of employing mixture cure models on prior data snapshots from the CheckMate 067 trial dataset. When examining the complete dataset covering a span of up to 60 months, the contrast between independent and nested models indicates comparable parameter estimations. This consistency is anticipated due to the dataset's maturity, the constrained number of overarching data points, and the application of a weakly informative prior for the global variance hyperparameter.

For illustrative purposes, I present findings using the penalized complexity prior, incorporating an exponential distribution with a rate parameter of $\lambda = 55$ for the 12-month data-cut scenario while maintaining all previously established priors, including weak priors for the mean π_k. As more data become available over time, the reliance on this prior diminishes, aligning with the principles of adaptive informationsharing methods like power priors. These approaches allow for flexible borrowing, where a power parameter dictates the extent of information transfer. In a simple case involving normally distributed historical data under a flat prior, the power prior follows a normal distribution with a variance scaled in proportion to the sample size and the power parameter set by the researcher.

Figures 7 and 8 display forest plots of cure fraction estimates for 12, 30, and 60 months for the separate and hierarchical models, respectively. These results correspond to both the OS exponential with PFS exponential model and the OS log-normal with PFS log-normal model.

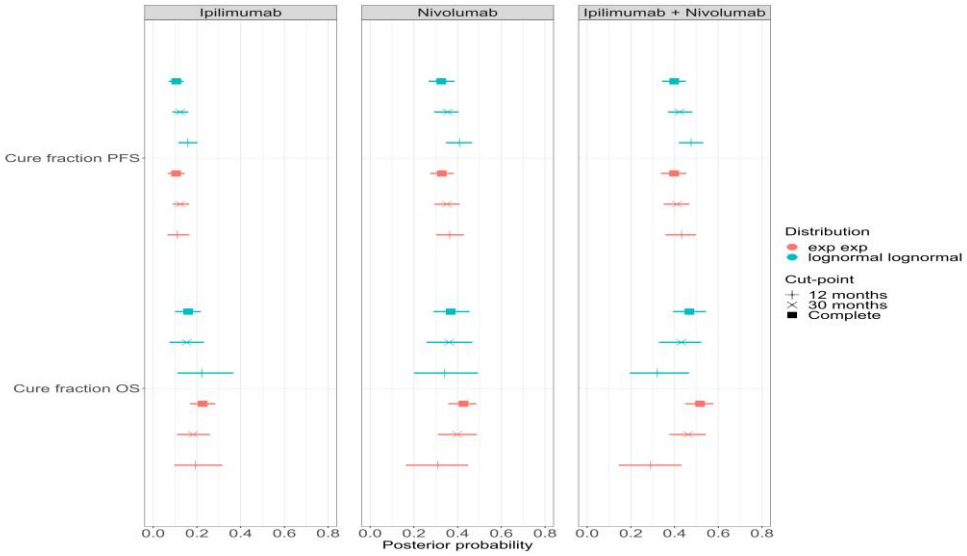

Figure 7. Posterior cure fraction forest plot with 95% credible intervals for study data with cutpoint censoring at 12 months, 30 months, and the complete dataset using the separate mixture cure models.

Examining Figure 7, I observe greater uncertainty in the OS survival estimates from the separate exponential model compared to its hierarchical counterpart shown in Figure 8. Furthermore, for the log-normal model, both separate and hierarchical OS survival curves tend to overestimate the cure fraction at earlier data-cuts across all treatment groups. However, as data matures, the cure fraction estimates converge downward toward the complete dataset values. Conversely, within the separate models across all treatment arms, the exponential model at 12 months underestimates the OS cure fraction, progressively approaching the complete dataset estimate from below. This behavior is less distinct than that of the log-normal model, likely due to greater estimation uncertainty in the exponential model. These differing patterns suggest contrasting model behaviors, but determining the correct trajectory remains uncertain until further data accrual. Notably, in the hierarchical model, the OS cure fraction estimates exhibit increased stability and certainty, closely aligning with the final dataset estimates even at the 12-month mark.

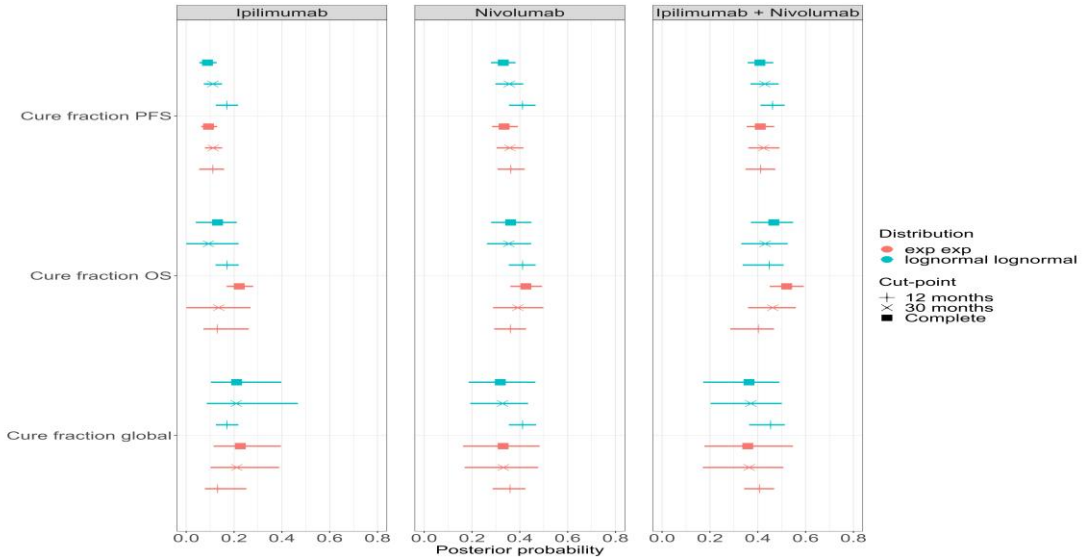

Figure 8. Posterior cure fraction forest plot with 95% credible intervals for study data with cut-point censoring at 12 months, 30 months, and the complete dataset using the hierarchical mixture cure model.

Figure 9 illustrates the survival curves corresponding to the 12-month data-cut for both separate and hierarchical models under the exponential distribution assumption. Equivalent plots for the 30-month data-cut are provided in the Appendix.

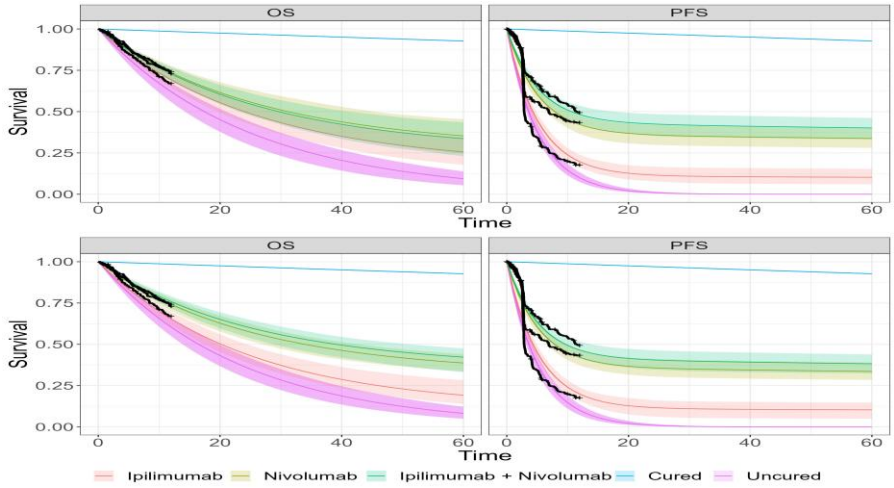

Figure 9. Mixture cure models' posterior survival curves with 95% credible intervals using 12-month cut-point data for separate (top) and hierarchical (bottom) models. Both PFS and OS employ exponential uncured survival curves. The black lines represent observed Kaplan-Meier survival curves.

Table 1 summarizes the WAIC point estimates (and standard errors) for separate models and all distributions. Analyzing the uncured Restricted Mean Survival Times (RMSTs) across models, the hierarchical exponential model yields an OS RMST of 20.6 (18.4, 23.1) months and a PFS RMST of 6.72 (6.15, 7.34) months. In contrast, the separate model estimates OS at 22.3 (18.5, 26.9) months and PFS at 5.68 (4.9, 6.45) months. For the log-normal hierarchical model, the OS RMST is 6.6 (6.17, 7.13) months, while PFS is 3.98 (3.84, 4.12) months. The

corresponding values from the separate model are 6.59 (6.18, 7.07) months for OS and 3.98 (3.85, 4.1) months for PFS. WAIC point estimates (and standard errors) for separate models and all distributions.

Table 1. WAIC point estimates (and standard errors) for separate models and all distributions.

OS Distn.	PFS Distn.	ELPD†	pD‡	WAIC
exp	exp	-5066.38 (81.4)	10.46 (0.43)	10132.76 (162.81)
	Gompertz	-5066.68 (81.46)	10.38 (0.46)	10133.37 (162.92)
	loglogistic	-4971.73 (81.67)	10.36 (0.28)	9943.46 (163.34)
	lognormal	-4732.41 (93.18)	18.56 (0.99)	9464.83 (186.36)
	Weibull	-5064.87 (82.86)	14.39 (0.86)	10129.74 (165.72)
Gompertz	exp	-5066.72 (81.44)	10.51 (0.45)	10133.43 (162.89)
	Gompertz	-5066.76 (81.49)	10.19 (0.44)	10133.53 (162.97)
	loglogistic	-4972.24 (81.7)	10.7 (0.28)	9944.47 (163.4)
	lognormal	-4732.23 (93.3)	18.01 (0.95)	9464.46 (186.61)
	Weibull	-5064.46 (82.92)	13.5 (0.78)	10128.91 (165.83)
loglogistic	exp	-5062.51 (81.89)	10.79 (0.4)	10125.02 (163.78)
	Gompertz	-5062.64 (82.01)	10.26 (0.41)	10125.27 (164.02)
	loglogistic	-4968.45 (82.36)	11.63 (0.28)	9936.89 (164.72)
	lognormal	-4728.25 (94.2)	18.18 (0.92)	9456.51 (188.4)
	Weibull	-5060.18 (83.31)	13.79 (0.69)	10120.37 (166.62)
lognormal	exp	-4973.43 (85.26)	14.63 (0.55)	9946.86 (170.53)
	Gompertz	-4973.48 (85.22)	14.31 (0.51)	9946.96 (170.44)
	loglogistic	-4879.07 (86.55)	15.07 (0.44)	9758.14 (173.09)
	lognormal	-4638.43 (100.4)	21.3 (0.96)	9276.85 (200.8)
	Weibull	-4972.48 (87.07)	19.54 (0.91)	9944.95 (174.13)
Weibull	Exp	-5064.74 (81.52) -	11.11 (0.45)	10129.48 (163.04)
	Gompertz	5065.16 (81.59) -4970.64	11.16 (0.44)	10130.32 (163.18)
	Loglogistic	(81.87)	11.68 (0.35)	9941.27 (163.73)
	Lognormal	-4730.93 (93.78)	19.04 (0.96)	9461.85 (187.56)
	Weibull	-5063.2 (83.09)	14.87 (0.83)	10126.41 (166.17)

†ELPD: Expected log pointwise predictive density; ‡pD : Effective number of parameters; WAIC: widely applicable information criterion.

Table 2 presents the extrapolated survival estimates at 60 months based on 12- and 30-month data-cuts. When the survival curve reaches a stable plateau, it effectively represents the cure fraction estimate.

Table 2. Estimated survival probabilities for distinct and hierarchical models at 60 months using data cut points of 12, 30, and 60 months, models employ either exponential or log-normal distributions for both DFS and RFS.

Endpoint	Model	Distribution	Cut-point (months)	Drug A	Drug B	Combination
DFS	hier	exp	12	0.22 [0.16, 0.31]	0.37 [0.31, 0.42]	0.45 [0.36, 0.51]
DFS	hier	exp	30	0.21 [0.13, 0.29]	0.36 [0.31, 0.41]	0.50 [0.43, 0.58]
DFS	hier	exp	60	0.26 [0.22, 0.31]	0.34 [0.29, 0.39]	0.54 [0.49, 0.59]
DFS	hier	lognormal	12	0.30 [0.25, 0.36]	0.40 [0.35, 0.45]	0.52 [0.46, 0.57]
DFS	hier	lognormal	30	0.24 [0.18, 0.32]	0.35 [0.30, 0.40]	0.52 [0.46, 0.59]
DFS	hier	lognormal	60	0.27 [0.23, 0.32]	0.33 [0.28, 0.38]	0.54 [0.49, 0.59]
DFS	separate	exp	12	0.28 [0.20, 0.36]	0.37 [0.31, 0.43]	0.38 [0.26, 0.48]
DFS	separate	exp	30	0.24 [0.19, 0.30]	0.34 [0.29, 0.40]	0.50 [0.43, 0.55]
DFS	separate	exp	60	0.27 [0.23, 0.32]	0.32 [0.27, 0.37]	0.53 [0.48, 0.58]
DFS	separate	lognormal	12	0.36 [0.28, 0.44]	0.40 [0.34, 0.45]	0.44 [0.35, 0.54]
DFS	separate	lognormal	30	0.29 [0.24, 0.34]	0.35 [0.29, 0.39]	0.51 [0.45, 0.57]
DFS	separate	lognormal	60	0.29 [0.24, 0.34]	0.32 [0.27, 0.37]	0.54 [0.48, 0.59]
RFS	hier	exp	12	0.12 [0.07, 0.17]	0.42 [0.37, 0.47]	0.41 [0.35, 0.47]
RFS	hier	exp	30	0.12 [0.09, 0.16]	0.44 [0.37, 0.52]	0.42 [0.36, 0.48]
RFS	hier	exp	60	0.10 [0.07, 0.14]	0.45 [0.40, 0.51]	0.41 [0.35, 0.47]
RFS	hier	lognormal	12	0.18 [0.13, 0.22]	0.50 [0.44, 0.54]	0.46 [0.41, 0.51]
RFS	hier	lognormal	30	0.12 [0.09, 0.16]	0.46 [0.41, 0.52]	0.43 [0.37, 0.48]
RFS	hier	lognormal	60	0.10 [0.07, 0.14]	0.46 [0.41, 0.51]	0.41 [0.35, 0.46]
RFS	separate	exp	12	0.12 [0.08, 0.17]	0.38 [0.28, 0.48]	0.43 [0.36, 0.49]
RFS	separate	exp	30	0.13 [0.09, 0.17]	0.44 [0.37, 0.51]	0.41 [0.35, 0.46]
RFS	separate	exp	60	0.12 [0.09, 0.16]	0.45 [0.40, 0.51]	0.40 [0.34, 0.45]
RFS	separate	lognormal	12	0.17 [0.13, 0.21]	0.46 [0.38, 0.55]	0.47 [0.42, 0.52]
RFS	separate	lognormal	30	0.14 [0.09, 0.17]	0.46 [0.39, 0.53]	0.42 [0.37, 0.48]
RFS	separate	lognormal	60	0.12 [0.09, 0.15]	0.46 [0.41, 0.51]	0.40 [0.34, 0.45]

5. Discussion

This paper introduces a Bayesian hierarchical mixture cure model to generate comprehensive survival curves exhibiting a plateau effect. Such survival curves can be effectively utilized in Health Technology Assessment (HTA) to inform patient lifetime estimations. The approach was applied to the CheckMate 067 trial dataset for overall survival (OS) and progression-free survival (PFS) endpoints at varying artificial data-cut points. By comparing the hierarchical model with its independent model counterpart, it was demonstrated that the former provides superior performance, making it both a theoretically justified and practically applicable choice in clinical trials involving multiple treatment arms and endpoints.

One of the key advantages of employing a Bayesian framework instead of a frequentist approach is its ability to flexibly model complex hierarchical structures, particularly in cases of multilevel or nested data. Frequentist models, on the other hand, typically require rigid assumptions regarding covariance structures. Furthermore, the hierarchical Bayesian model is more robust in settings with limited sample sizes due to its capacity to incorporate prior information and facilitate information sharing across endpoints.

A crucial aspect of this methodology is the specification of informative priors, particularly for cure fractions, which play a pivotal role in early-stage trials where long-term survival data are scarce. This prior information helps stabilize cure fraction estimates. However, excessive borrowing of information across endpoints may not always be appropriate when the endpoints exhibit stark differences. In such cases, a partial exchangeability framework can be employed instead [26]. Unlike some previous frequentist methods, this Bayesian framework allows the incorporation of contextual knowledge regarding model parameters through prior distributions. Expert opinion or external datasets can be leveraged to enhance inference stability, particularly when data sparsity is an issue. For

this study, various priors were explored. Weakly informative priors ensure that the posterior distribution is predominantly influenced by observed data, even with small sample sizes. However, incorporating additional clinical knowledge can refine prior selection, enhancing interpretability and inference quality.

For instance, suppose clinical expertise suggests that a particular drug treatment results in a cure fraction exceeding 30%. Additionally, if a cancer patient enrolled in the study is aged 60, it is reasonable to assume with high confidence that they will not survive beyond 40 years. When faced with extensive right censoring or missing data, regularizing inference through prior knowledge can significantly impact model stability. A straightforward approach to constraining survival estimates involves blending survival curves [27]. However, establishing practical methodologies for eliciting such prior information remains a challenge, necessitating formalized protocols [28].

Regulatory bodies like the European Medicines Agency (EMA) and the U.S. Food and Drug Administration (FDA) may appreciate the hierarchical model's ability to incorporate prior skepticism. This prevents naive overestimation of cure rates, particularly when data availability is limited. The hierarchical model has demonstrated an ability to yield more precise cure rate estimates, particularly in early data-cut scenarios (especially under the exponential model). This highlights its potential for producing reliable estimates at earlier trial stages, which could support regulatory and reimbursement decision making.

For this study, censoring due to artificial data-cuts was simulated by truncating patient data at a fixed time point. While real-world scenarios might involve censoring at each patient's last visit date, this distinction was not essential for the purposes of this analysis. The chosen data-cut points were primarily intended to illustrate the theoretical advantages of the proposed approach rather than strictly follow specific trial protocols.

Future research should validate the hierarchical mixture cure model across diverse datasets and implement data-cuts aligned with actual clinical trial protocols. The framework is extendable beyond two endpoints to multiple endpoints. Previous frequentist studies have applied similar hierarchical approaches to hospital clusters and recurrent event data [9-10] or dose-response interspecies extrapolation [29]. Future studies may explore whether additional endpoint data can enhance model performance, even if the endpoints do not exhibit plateauing behavior. Moreover, refining background survival curves to better reflect cohort-specific characteristics (e.g., long-term survival among complete responders) may bridge the gap between statistical and clinical cure definitions. Short-term clinical responder data could be integrated with external sources, such as WHO life tables, as demonstrated in [30].

Another perspective on data-cut selection is determining the optimal follow-up time that maximizes expected overall utility. The stopping utility should account for survival extrapolation robustness, financial and logistical constraints, accelerated drug market entry, and minimized patient exposure to adverse effects. Moreover, the stopping criteria must ensure reliable parameter estimation, keeping inferred values within an acceptable deviation from the true values. This framework could be extended to support decision-making in drug selection based on order statistics.

This perspective aligns with value-of-information (VoI) analysis [31]. For example, one could assess the additional utility of gathering further data at 12 and 30 months. In practice, a discrete set of pre-determined cut-points may be more effective than attempting to define a general continuous rule. Beyond real-time trial analysis, the hierarchical mixture cure model could be leveraged in trial design to determine optimal follow-up durations and data-cut selection. Given the hierarchical structure's capacity to extract additional information, trials could potentially achieve equivalent statistical power with reduced sample sizes and shorter durations.

The adoption of hierarchical mixture cure models has broad implications for HTA. Cost-effectiveness evaluations hinge on stable estimates of treatment benefits, as fluctuating estimates over time could lead to inconsistent cost-effectiveness ratios and differing optimal decisions. Lee (2019) [32] examined the cost-effectiveness of nivolumab plus ipilimumab versus ipilimumab alone using CheckMate 067 data, employing both a partitioned survival model and a Markov state-transition model. The study compared cost-effectiveness estimates derived from 18-month (OS

data unavailable) and 36-month (OS data available) data-cuts. Notably, the inclusion of OS data yielded an increase of over one quality-adjusted life-year (QALY) across both treatment arms.

6. Conclusion

This study proposes a Bayesian hierarchical mixture cure model to create comprehensive survival curves that reflect the long-term plateau effects frequently observed in immunotherapy trials. Health Technology Assessment (HTA) needs these kinds of curves because they need accurate lifetime survival estimates to make decisions. The suggested framework was used on the CheckMate 067 dataset to look at overall survival (OS) and progression-free survival (PFS) at different artificial data-cut points. Results show that the hierarchical formulation works better than independent models because it lets endpoints share information while still being flexible. The Bayesian method lets you use prior knowledge and hierarchical structures, which makes the parameters more stable and the estimates more accurate, especially when the sample sizes are small or the follow-up data are limited. Informative priors for cure fractions help keep inference stable during the early stages of a trial, and partial exchangeability can help avoid too much borrowing when endpoints are very different. The hierarchical model yielded more accurate cure rate estimates in early data-cut scenarios, underscoring its utility for interim clinical analyses and regulatory assessment. The framework provides a theoretically robust and practically viable methodology for survival extrapolation, facilitating more consistent health technology assessment outcomes and enhancing evidence generation in multi-endpoint clinical trials.

Data Availability Statement
Not applicable.

Funding
This work was supported without any funding.

Conflicts of Interest
The author declares no conflicts of interest.

Ethical Approval and Consent to Participate
Not applicable.

References

[1]. Latimer, N. (2013). *NICE DSU technical support document no. 14: Survival analysis for economic evaluations alongside clinical trials—Extrapolation with patient-level data*. National Institute for Health and Care Excellence.
[2]. Demiris, N., & Sharples, L. D. (2006). Bayesian evidence synthesis to extrapolate survival estimates in cost-effectiveness studies. *Statistics in Medicine, 25*(11), 1960–1975.
[3]. Jackson, C. H., Sharples, L. D., & Thompson, S. G. (2010). Survival models in health economic evaluations: Balancing fit and parsimony to improve prediction. *The International Journal of Biostatistics, 6*(1), Article 34.
[4]. Ouwens, M. J., Mukhopadhyay, P., Zhang, Y., Huang, M., Latimer, N., & Briggs, A. H. (2019). Estimating lifetime benefits associated with immuno-oncology therapies: Challenges and approaches for overall survival extrapolations. *Pharmacoeconomics, 37*(9), 1129–1138.
[5]. Khair, D. O., Bax, H. J., Mele, S., Crescioli, S., Pellizzari, G., Khiabany, A., Nakamura, M., Harris, R. J., French, E., Hoffmann, R. M., Williams, I. P., Cheung, A., Thair, B., Beales, C. T., Touizer, E., Signell, A. W., Tasnova, N. L., Spicer, J. F., Josephs, D. H., … Karagiannis, S. N. (2019). Combining immune checkpoint inhibitors: Established and emerging targets and strategies to improve outcomes in melanoma. *Frontiers in Immunology, 10*, 453.
[6]. Wolchok, J. D., Chiarion-Sileni, V., Gonzalez, R., Rutkowski, P., Grob, J.-J., Cowey, C. L., Lao, C. D., Wagstaff, J., Schadendorf, D., Ferrucci, P. F., Smylie, M., Dummer, R., Hill, A., Hogg, D., Haanen, J., Carlino, M. S., Bechter, O., Maio, M., Marquez-Rodas, I., … Larkin, J. (2017). Overall survival with combined nivolumab and ipilimumab in advanced melanoma. *New England Journal of Medicine, 377*, 1345–1356.
[7]. Larkin, J., Chiarion-Sileni, V., Gonzalez, R., Grob, J.-J., Rutkowski, P., Lao, C. D., Cowey, C. L., Schadendorf, D., Wagstaff, J., Dummer, R., Ferrucci, P. F., Smylie, M., Hogg, D., Hill, A., Marquez-Rodas, I., Haanen, J., Guidoboni, M.,

Maio, M., Schoffski, P., … Wolchok, J. D. (2019). Five-year survival with combined nivolumab and ipilimumab in advanced melanoma. *New England Journal of Medicine, 381*(16), 1535–1546.
[8]. Bullement, A., Willis, A., Amin, A., Schlichting, M., Hatswell, A. J., & Bharmal, M. (2020). Evaluation of survival extrapolation in immuno-oncology using multiple pre-planned data cuts: Learnings to aid in model selection. *BMC Medical Research Methodology, 20*(1), 1–14.
[9]. Lai, X., & Yau, K. K. W. (2009). Multilevel mixture cure models with random effects. *Biometrical Journal, 51*, 456–466.
[10]. Lai, X., & Yau, K. K. W. (2008). Long-term survivor model with bivariate random effects: Applications to bone marrow transplant and carcinoma study data. *Statistics in Medicine, 27*, 5692–5708.
[11]. Tan, S. H., Abrams, K. R., & Bujkiewicz, S. (2018). Bayesian multiparameter evidence synthesis to inform decision making: A case study in metastatic hormone-refractory prostate cancer. *Medical Decision Making, 38*(7), 834–848.
[12]. Hodi, F. S., Chiarion-Sileni, V., Gonzalez, R., Grob, J. J., Rutkowski, P., Cowey, C. L., Lao, C. D., Schadendorf, D., Wagstaff, J., Dummer, R., Ferrucci, P. F., Smylie, M., Hill, A., Hogg, D., Marquez-Rodas, I., Jiang, J., Rizzo, J., Larkin, J., & Wolchok, J. D. (2018). Nivolumab plus ipilimumab or nivolumab alone versus ipilimumab alone in advanced melanoma (CheckMate 067): 4-year outcomes of a multicentre, randomised, phase 3 trial. *The Lancet Oncology, 19*(11), 1480–1492.
[13]. Hodi, F., Chiarion-Sileni, V., Lewis, K., Grob, J.-J., Rutkowski, P., Lao, C., Cowey, C., Schadendorf, D., Wagstaff, J., Dummer, R., Queirolo, P., Smylie, M., Butler, M., Hill, A., Marquez-Rodas, I., Haanen, J., Durani, P., Wolchok, J., & Larkin, J. (2022). Long-term survival in advanced melanoma for patients treated with nivolumab plus ipilimumab in CheckMate 067. *Journal of Clinical Oncology, 40*, 9522–9522.
[14]. Mohr, P., Larkin, J., Paly, V., Azocar, A. R., Baio, G., Kurt, M., Amadi, A., Rizzo, J., Johnson, H., Moshyk, A., Kotapati, S., & Middleton, M. (2020). Estimating long-term survivorship in patients with advanced melanoma treated with immune-checkpoint inhibitors: Analyses from the phase III CheckMate 067 trial. *Annals of Oncology, 31*, S747.
[15]. Yu, X., De Angelis, R., Andersson, T., Lambert, P., O'Connell, D., & Dickman, P. (2013). Estimating the proportion cured of cancer: Some practical advice for users. *Cancer Epidemiology, 37*, 836–842.
[16]. Nikolaidis, G. F., Woods, B., Palmer, S., & Soares, M. O. (2021). Classifying information-sharing methods.
[17]. Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1).
[18]. R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
[19]. World Health Organization. (2020). *Life tables by country*. http://apps.who.int/gho/data/
[20]. Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis, 1*(3), 515–533.
[21]. Chung, Y., Rabe-Hesketh, S., Dorie, V., Gelman, A., & Liu, J. (2013). A nondegenerate penalized likelihood estimator for variance parameters in multilevel models. *Psychometrika*, 685–709.
[22]. Simpson, D., Rue, H., Riebler, A., Martins, T. G., & Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science, 32*(1), 1–28.
[23]. Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413–1432.
[24]. Wileyto, E. P., Li, Y., Chen, J., & Heitjan, D. F. (2013). Assessing the fit of parametric cure models. *Biostatistics, 14*, 340–350.
[25]. Peng, Y., & Yu, B. (2021). *Cure models*. Chapman and Hall/CRC.
[26]. Neuenschwander, B., Wandel, S., Roychoudhury, S., & Bailey, S. (2016). Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics, 15*(2), 123–134.
[27]. Che, Z., Green, N., & Baio, G. (2022). Blended survival curves: A new approach to extrapolation for time-to-event outcomes from clinical trials in health technology assessment. *Medical Decision Making*.
[28]. O'Hagan, A. (2019). Expert knowledge elicitation: Subjective but scientific. *The American Statistician, 73*(sup1), 69–81.
[29]. DuMouchel, W. H., & Harris, J. E. (1983). Bayes methods for combining the results of cancer studies in humans and other species. *Journal of the American Statistical Association, 78*(382), 293–308.
[30]. Jackson, C., Stevens, J., Ren, S., Latimer, N., Bojke, L., Manca, A., & Sharples, L. D. (2017). Extrapolating survival from randomized trials using external data: A review of methods. *Medical Decision Making, 37*(4), 377–390.
[31]. Heath, A., Manolopoulou, I., & Baio, G. (2017). A review of methods for analysis of the expected value of information. *Medical Decision Making, 37*(7), 747–758.
[32]. Lee, D., Amadi, A., Sabater, J., Ellis, J., Johnson, H., Kotapati, S., McNamara, S., Walker, A., Cooper, M., Patterson, K., Roskell, N., & Meng, Y. (2019). Can we accurately predict cost effectiveness without access to overall survival data? The case study of nivolumab in combination with ipilimumab for the treatment of patients with advanced melanoma in England. *PharmacoEconomics – Open, 3*(1), 43–54.