Methodology

For reprint orders, please contact: info@becaris.com

Journal of **Comparative Effectiveness Research**

Bayesian hierarchical model-based network meta-analysis to overcome survival extrapolation challenges caused by data immaturity

Bart Heeg*[,1], Andre Verhoek[1], Gabriel Tremblay[2], Ofir Harari[2], Mohsen Soltanifar[2], Haitao Chu[3], Satrajit Roychoudhury[3,4] & Joseph C Cappelleri[3]

[1]Cytel RWAA, Weena 316, 3012 NJ, Rotterdam, The Netherlands

[2]Cytel RWAA, Canada

[3]Pfizer Inc, 445 Eastern Point Road, MS 8260-2502, Groton, CT 06340, USA

[4]Pfizer Inc., 235 E 42nd St, New York, NY 10017, USA.

*Author for correspondence: Tel.: +31 010 3330 877; bart.heeg@cytel.com

Aim: This research evaluated standard Weibull mixture cure (WMC) network meta-analysis (NMA) with Bayesian hierarchical (BH) WMC NMA to inform long-term survival of therapies. **Materials & methods:** Four trials in previously treated metastatic non-small-cell lung cancer with PD-L1 >1% were used comparing docetaxel with nivolumab, pembrolizumab and atezolizumab. Cure parameters related to a certain treatment class were assumed to share a common distribution. **Results:** Standard WMC NMA predicted cure rates were 0.03 (0.01; 0.07), 0.18 (0.12; 0.24), 0.07 (0.02; 0.15) and 0.03 (0.00; 0.09) for docetaxel, nivolumab, pembrolizumab and atezolizumab, respectively, with corresponding incremental life years (LY) of 3.11 (1.65; 4.66), 1.06 (0.41; 2.37) and 0.42 (-0.57; 1.68). The Bayesian hierarchical-WMC-NMA rates were 0.06 (0.03; 0.10), 0.17 (0.11; 0.23), 0.12 (0.05; 0.20) and 0.12 (0.03; 0.23), respectively, with incremental LY of 2.35 (1.04; 3.93), 1.67 (0.68; 2.96) and 1.36 (-0.05; 3.64). **Conclusion:** BH-WMC-NMA impacts incremental mean LYs and cost–effectiveness ratios, potentially affecting reimbursement decisions.

Tweetable abstract: Bayesian hierarchical model-based network meta-analysis for survival outcomes are shown to overcome data immaturity issues in the evidence network, impact incremental mean life years and cost–effectiveness ratios, affecting reimbursement decisions.

First draft submitted: 2 September 2022; Accepted for publication: 21 December 2022; Published online: 18 January 2023

Keywords: bayesian hierarchical modelling • immunotherapies • indirect treatment comparisons • mixture cure • network meta-analysis • oncology • survival extrapolation

The lifetime costs and consequences of interventions must be assessed for health economic modeling purposes. This evaluation usually requires extrapolation of the within-trial estimated survival curve, based on commonly used seven standard parametric distributions: exponential, generalized gamma, gamma, Gompertz, loglogistic, lognormal and Weibull [1].

Immune-oncology (IO) therapies and gene therapies potentially result in negligible disease related risk of events for a fraction of patients in the long term [2,3]. For these indications, standard parametric survival models may not provide clinically plausible extrapolations, as they do not capture flattening of the Kaplan–Meier curves well [4–7]. Mixture and non mixture cure, parametric mixture and spline models are suggested instead, for the flexibility to capture such shape [5–10]. The statistical cure rates based on mixture and non-mixture cure models are often estimated with large uncertainty, especially in scenarios when data are immature [11]. Guidance on extrapolating survival data indicates that external data from other treatments may be used to reduce this uncertainty.

Bayesian hierarchical modeling (BHM) frameworks assume that treatment effects are exchangeable across subgroups and trials, in other words, drawn from the same distribution of effects. BHMs have been applied in adaptive

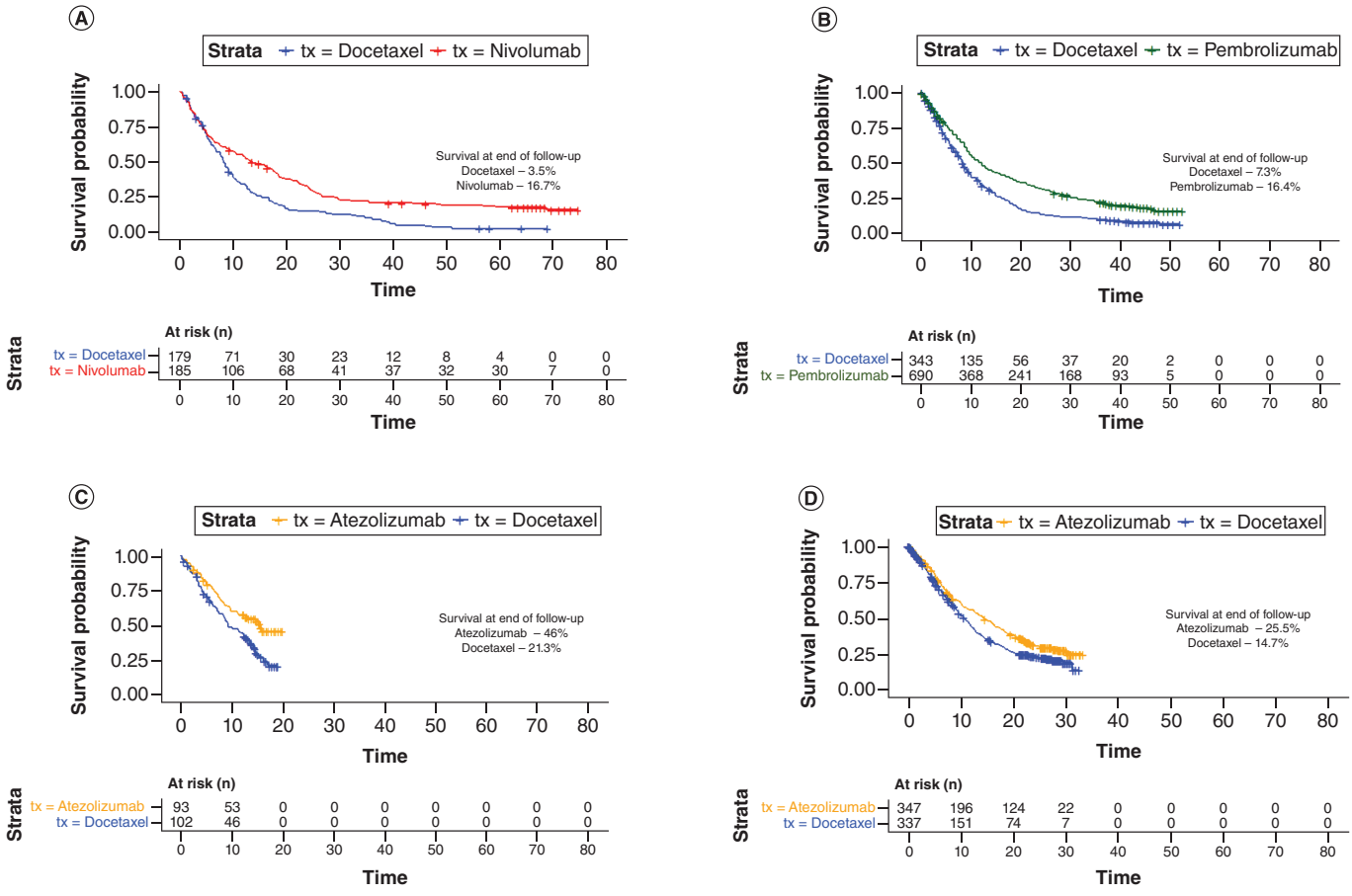

Figure 1. Reconstructed Kaplan–Meier curves in evidence network of overall survival. Kaplan–Meier curve in evidence network of OS for the trial comparing docetaxel with **(A)** nivolumab, **(B)** pembrolizumab and **(C & D)** atezolizumab.
OS: Overall survival

basket trial designs, borrowing external control data in randomized trials, extrapolation of adult data in pediatric population, tumor agnostic health technology assessment (HTA) submissions, in network meta-analyses (NMA) considering both real-world evidence and randomized clinical trials, and NMAs incorporating nonignorable missingness [12–20].

With immature data, mixture cure models may not be able to accurately estimate the cured fraction or may result in large uncertainty [11]. By specifying class effects on the cure rate using BHM framework, the therapies investigated in trials with relative short follow-up can borrow information from therapies of the same class with mature or longer follow-up.

The current research aimed to compare a standard Weibull mixture cure (WMC) NMA with a Bayesian hierarchical WMC NMA (BH WMCNMA) approach, and assessed the impact of the variance over the class effects on the cure rates in the BH WMC NMA approach.

Materials & methods

Data

For demonstrative purposes, data were selected from a published NMA in previously treated metastatic non-small-cell lung cancer programmed death-ligand 1 (PD-L1) >1% [21], comparing docetaxel with nivolumab (pooled analyses from CheckMate 017 and 057 [22]), pembrolizumab (from KEYNOTE-010 [23]) and atezolizumab (from POPLAR [24] and OAK [25]) (Figure 1). A publication with pooled analyses was used for CheckMate 017 and CheckMate 057 as the studies only considered squamous and non squamous patients, respectively. The same evidence network and data was previously used in Heeg *et al.* [11]. Only the overall survival data from these four treatments were considered to keep visual presentation and tabulation simple, and to better illustrate the

methodology. Pseudo individual patient-level data for each intervention was obtained by reconstructing time-to-event (TTE) data digitized from published Kaplan–Meier curves using Engauge Digitizer software [26], and from the algorithm published by Guyot *et al.* [27]. The corresponding evidence network is presented in the Supplementary Figure 1.

Standard WMC NMA

The WMC NMA model is defined using the following notations:

The cure rate π_{ij} for index trial i, $i = 1, \ldots, I$ and index treatment j, $j = 0, \ldots, J$, is defined as $\pi_{ij} = \frac{1}{1+\exp(\beta_{0_i}+\beta_{1_j})}$ in which $\beta_{0_i} + \beta_{1_j}$ is the log odds of cure for study i, $i = 1, \ldots, I$ and treatment j, $j = 0, \ldots, J$, with β_{0_i} the log odds for cure for study i, $i = 1, \ldots, I$, of the docetaxel ($j = 0$) and β_{1_j} the log odds ratio of cure of treatment j ($j = 1, 2, \ldots, J$) versus docetaxel, with $\beta_{1_o} = 0$. In order to have a positive scale and shape parameters for a Weibull distribution, the scale and shape for the study i, $i = 1, \ldots, I$ and treatment j, $j = 0, \ldots, J$, are defined by exp $(\lambda_{0_i} + \lambda_{1_j})$ and exp $(\gamma_{0_i} + \gamma_{1_j})$, respectively, where λ_{0_i} are γ_{0_i} the log scale and log shape of docetaxel for the study i, $i = 1, \ldots, I$ and λ_{1_j} and γ_{1_j} are the relative treatment effect of treatment j, $j = 0, \ldots, J$, versus docetaxel for the scale and shape, respectively, with $\lambda_{1_o} = \gamma_{1_o} = 0$.

The density function of the Weibull distribution is defined as:

$$f_{ij}^{ds}(t) = \frac{\exp\left(\gamma_{0_i} + \gamma_{1_j}\right)}{\exp\left(\lambda_{0_i} + \lambda_{1_j}\right)}\left(\frac{\tau}{\exp\left(\lambda_{0_i} + \lambda_{1_j}\right)}\right)^{\left(\exp\left(\gamma_{0_i}+\gamma_{1_j}\right)-1\right)} \exp\left(-\left(\frac{\tau}{\exp\left(\lambda_{0_i} + \lambda_{1_j}\right)}\right)^{\exp\left(\gamma_{0_i}+\gamma_{1_j}\right)}\right)$$

where t is the survival time, and the corresponding Weibull survival function is

$$Weib_{ij}^{ds}(t) = \exp\left(-\left(\frac{\tau}{\exp\left(\lambda_{0_i} + \lambda_{1_j}\right)}\right)^{\left(\exp\left(\gamma_{o_i}+\gamma_{1_j}\right)\right)}\right)$$

The disease-specific survival (mixture-cure model [MCM]) is written as $S_{ij}^{ds}(t) = \pi_{ij} + (1 - \pi_{ij}) Weib_{ij}^{ds}(t)$. The MCM in this paper is used to model relative survival–survival is modeled as the product of the MCM survival with general population survival [8].

Let k denote the subject in the ith trial, $i = 1, \ldots, I$ and jth treatment group $j = 0, \ldots, J$, ($k = 1,2, \ldots, n_{ij}$) and t_{ijk} denote the observed survival or censoring time, with δ_{ijk} (= 0, 1) with 0 indicating censoring and 1 indicating an event. Let Age_{ijk} be the mean age of the patient until event or censoring at t_{ijk}, $S^*(Age)_{ijk}$) denote the survival probability for general population at Age_{ijk} and $S_{ij}(Age_{ijk}, t_{ijk})$ be the all-cause survival for index trial i, $i = 1, \ldots, I$ and index treatment j, $j = 0, \ldots, J$, at time t_{ijk} at age Age_{ijk}, $k = 1,2, \ldots, n_{ij}$, corrected for the distribution of males and females in the reference trial, then $S_{ij}(Age_{ijk}, t_{ijk}) = (\pi_{ij} + (1 - \pi_{ij}) S_{ij}^{ds}(t_{ijk})) S \times (Age_{ijk})$ [8].

Thus, the log likelihood contribution for the kth subject, $k = 1,2, \ldots, n_{ij}$, in trial i, $i = 1, \ldots, I$ and treatment group j, $j = 0, \ldots, J$, is

$$ln\, L_{ijk} = ln\left[\pi_{ij} + \left(1 - \pi_{ij}\right) S_{ij}^{ds}\left(t_{ijk}\right)\right] + ln\left[S \times \left(Age_{ijk}\right)\right]$$
$$+ \delta_{ijk}\, ln\left(h \times \left(Age_{ijk}\right) + \frac{\left(1-\pi_{ij}\right) f_{ij}^{ds}\left(t_{ijk}\right)}{\pi_{ij}+\left(1-\pi_{ij}\right) S_{ij}^{ds}\left(t_{ijk}\right)}\right)$$

with $h^*(Age_{ijk})$ the general population hazard at Age_{ijk} corrected for gender difference in the reference trial.

Weak-informative a priori distributions are used:

$$\lambda_{1_j},\ \lambda_{0_i} \sim N(0, 5),\ \beta_{1_j},\ \beta_{0_i} \sim N(0, 1) \text{and} \gamma_{1_j},\ \gamma_{0_i} \sim N(0, 5)$$

where N(0, 5) is the normal distribution with a mean of zero and a standard deviation of 5.

BHM WMC NMA

Two types, or, classes of therapies-docetaxel and iIOs are distinguished in this evidence network. Only class effects on the cure parameter are applied; although, the model can be extended to have class effects for the shape and scale as well.

For the full BHM, the cure rate is written as

$$\pi_{iji} = \frac{1}{1 + \exp\left(\rho_i + \varepsilon_i + \delta_j\right)}$$

where

$$\rho_l = \begin{cases} \rho 1, & \text{with l = 0 for reference therapy(docetaxel)} \\ \rho 1 + \rho 2 & \text{with l = 1 for immunotherapies} \end{cases}$$

with $\rho 1$, $\rho 2 \sim N(0, 5)$ defining the overall cure rate $\rho 1$ for the reference therapy and overall treatment effect on cure $\rho 2$.

$\epsilon_i \sim N(0, \tau)$ being the parameter that represents trial effects on the cure rate $i = 1, \ldots, I$
$\delta_j \sim N(0, \sigma)$ represents treatment effects on the cure rate, $j = 0, \ldots, J$, with $\delta_j = 0$ for docetaxel.
$\tau \sim U(0, \tau')$
$\sigma \sim U(0, \sigma')$

with τ' and σ' predefined maximum values for uniform distribution (U).

The impact of the BHM will be assessed in several scenario analyses in which the allowed heterogeneity over the class effects is varied by different choices of τ' and σ':

- $\tau' = \sigma' = 0.01$
- $\tau' = \sigma' = 0.10$
- $\tau' = \sigma' = 1.00$
- $\tau' = \sigma' = 10.0$

In the current analyses, study 1 is CheckMate 017 and 057 (pooled), study 2 is KEYNOTE-010, study 3 is OAK and study 4 is POPLAR.

Treatment 1 is docetaxel, treatment 2 is nivolumab, treatment 3 is pembrolizumab and treatment 4 is atezolizumab.

The BHM is closely related to the random-effects model. A random-effects model is a model in which the treatment effect of a specific treatment is different per trial. In the current BHM, the treatment effect is fixed across trials by treatment. However, the treatment effect of each individual IO may differ, dependent upon the data and variance assumed over the IO class effect in that common distribution. If no variance is allowed, one assumes that the same treatment effect applies for all IOs across all trials. If you allow for variance, one assumes that the same treatment effect applies for all IOs across all trials, but there will be differences in the treatment effect between the individual IOs (depending on the level of variance allowed).Other statistical considerations.

Other statistical considerations

All conducted NMA approaches considered relative mortality. Specifically, they accounted for the UK general population mortality [28] in the likelihood function following the approach suggested by Andersson *et al.*, 2013 [29], based on cubic splines and applied by van Oostrum *et al.*, 2021 [30] on parametric distributions. UK life tables were used with a mean age of 62 years and a sex distribution of 55% males, based on CheckMate 017 and 057 [31].

The results were assessed twice-once with a mature reference trial (i.e., CheckMate) and once with an immature reference trial (i.e., OAK). The tested WMC NMA approaches were compared based on leave one out information criterion (LOOIC), mean and incremental mean survival and corresponding uncertainty, defined as the width of the 95% credible intervals of the mean and incremental mean survival.

Validations of the atezolizumab predicted cure rates were based on the final OAK and POPLAR intention-to-treat (ITT) data-cuts [32]. These ITT populations also included PD-L1 <1% patients and are for reference only.

The analyses were performed using the RStan package in R Statistical Software (version 1.2-0) [33] and were fitted with weakly informative priors [34]. The default Hamilton Monte Carlo was used as a model fitting algorithm within RStan [35]. The models were run with three chains of 28,600 iterations, of which 30% were burn-in iterations, to generate the posteriors for the defined parameters. We used set seed of 2, a thinner of 2, adapt-delta = 0.95 and max_treedepth of 10. Convergence of the three chains was tested using the Rhat test, as introduced by Gelman and Rubin in 1992 [36]. In addition, effective sample size, autocorrelation and trace plots were assessed.

Results

The predicted cure rates were analyzed individually (i.e., not in an NMA framework) according to the Bayesian Weibull MCM. Nivolumab and pembrolizumab showed differentiating cure rates compared with docetaxel, with upper limits of the docetaxel credible intervals not including the mean cure rates of the other two therapies. In CheckMate, the rates were 0.02 (0.00; 0.06) for docetaxel and 0.18 (0.11; 0.25) for nivolumab; in KEYNOTE 0.08 (0.05; 0.12) for docetaxel and 0.16 (0.12; 0.20) for pembrolizumab. The POPLAR study demonstrated cure rates of 0.03 (0.00; 0.19) for docetaxel and 0.10 (0.00; 0.49) atezolizumab. In OAK, however, the estimated cure rate of docetaxel (0.16 [0.06; 0.23]) was very high compared with the other three cure rates for docetaxel and did not differentiate from atezolizumab (0.16 [0.00; 0.28]), making the results for atezolizumab uncertain. The clinical reason for this higher cure rate was unknown, but it very likely influenced the NMA results. The follow-up for atezolizumab was likely insufficient for the Weibull MCM to capture cure rate in the tail (Figure 1A–D).

Table 1 provides an overview of the LOOIC, predicted cure rates, predicted means and incremental mean survival by model and treatment using OAK and the CheckMate trials as reference therapy for the base case and the BH WMC NMA scenarios. Figure 2A–E present the corresponding predicted survival over time for the base-case, non hierarchical WMC NMA and the four BH WMC NMA scenarios with varying τ' and σ' and with CheckMate as the reference trial. The visual fit of the standard WMC NMA (Supplementary Figure 2) and of the most stringent BH WMC NMA with $\tau' = \sigma' = 0.01$ (Supplementary Figure 3) compared with the KM of the individual trials are presented in the Supplementary Materials. Although, the BH WMC NMA was more restrictive, the visual fit was still good. Also, the LOOIC (Table 1) of the BHM models did not differ substantially from the standard WMC NMA. Supplementary Figure 1 presents the fit and comparators of the four trials for the standard WMC NMA and the BH WMC NMA for $\tau' = \sigma' = 0.01$.

With CheckMate as the reference trial, the standard WMC NMA predicted cure rates of 0.03 (0.01; 0.07), 0.18 (0.12; 0.24), 0.07 (0.02; 0.15) and 0.03 (0.00; 0.09) for docetaxel, nivolumab, pembrolizumab and atezolizumab, respectively. The corresponding mean life years were 1.62 (1.07; 2.44), 4.76 (3.50; 6.18), 2.70 (1.63; 4.53) and 2.01 (1.22; 3.60), respectively. The BH WMC NMA model with little variance on cure class effects ($\tau' = \sigma' = 0.01$) showed cure rates of 0.07 (0.05; 0.10), 0.17 (0.14; 0.20), 0.17 (0.14; 0.20) and 0.17 (0.14; 0.20), respectively. The corresponding predicted mean life years were 2.45 (1.87; 3.12), 4.60 (3.84; 5.36), 4.79 (4.01; 5.57) and 4.57 (3.79; 5.38) for docetaxel, nivolumab, pembrolizumab and atezolizumab, respectively. In the BH WMC NMA scenarios, the uncertainty of the atezolizumab cure rates was substantially reduced. However, the higher the variance (τ' and σ'), the higher the uncertainty around the atezolizumab cure rates with 0.17 (0.14; 0.20) for $\tau' = \sigma' = 0.01$ and 0.07 (0.00; 0.21) for $\tau' = \sigma' = 10$, with CheckMate as the reference trial.

The reference trial had limited impact on the BH WMC NMA models with smaller variance (τ' and $\sigma' = 0.01$), with very similar predicted incremental life expectancies for nivolumab, pembrolizumab and atezolizumab compared with docetaxel. However, the more variance that was allowed in terms of τ' and σ', the more the predicted cure rates and mean (incremental) survival differed by the reference trial. The same was true for the standard WMC NMA with OAK as the reference trial, which showed an incremental mean survival when compared with docetaxel of 1.54 (-0.06; 11.17), 0.76 (0.24; 3.90) and 0.52 (-1.33; 2.82) for nivolumab, pembrolizumab and atezolizumab, respectively. The corresponding numbers with CheckMate as reference trial were 3.11 (1.65; 4.66), 1.06 (0.41; 2.37) and 0.42 (-0.57; 1.68), respectively (Table 1).

The model parameters corresponding credible intervals, effective sample sizes and Rhat's for standard WMC and the BH WMC NMA with $\tau' = \sigma' = 0.01$, 0.1, 1 and 10 and the corresponding density plots are presented in the Supplementary Materials. The autocorrelation and trace plots for the cure parameters for placebo and cure treatment effect of the IOs are also presented in the supplementary materials (Supplementary Figure 5). There were some divergent cases in BH WMC NMA runs, but not in the standard WMC runs. The divergent cases (<1%) were examined, but were all individual cases (e.g., no 'blocks' of consecutive iterations). Furthermore, as

Table 1. Overview of leave one out information criterion by model, and overview of cure rates plus 95% credible intervals and mean predicted life expectancy plus 95% credible intervals by treatment and model and reference therapy.

Model		LOOIC	OAK reference trial				CHECKMATE reference trial			
			Docetaxel	Nivolumab	Pembrolizumab	Atezolizumab	Docetaxel	Nivolumab	Pembrolizumab	Atezolizumab
Cure rates										
Base-case NMAs		13462.0	0.04 [0.00; 0.20]	0.17 [0.00; 0.70]	0.07 [0.00; 0.36]	0.04 [0.00; 0.28]	0.03 [0.01; 0.07]	0.18 [0.12; 0.24]	0.07 [0.02; 0.15]	0.03 [0.00; 0.09]
BHM NMAs	$\tau' = \sigma' = 0.01$	13464.6	0.07 [0.05; 0.10]	0.17 [0.14; 0.20]	0.17 [0.14; 0.20]	0.17 [0.14; 0.20]	0.07 [0.05; 0.10]	0.17 [0.14; 0.20]	0.17 [0.14; 0.20]	0.17 [0.14; 0.20]
	$\tau' = \sigma' = 0.10$	13464.6	0.07 [0.05; 0.11]	0.17 [0.14; 0.21]	0.17 [0.14; 0.21]	0.17 [0.14; 0.22]	0.07 [0.05; 0.10]	0.17 [0.14; 0.20]	0.17 [0.14; 0.20]	0.17 [0.14; 0.20]
	$\tau' = \sigma' = 1.00$	13463.9	0.06 [0.03; 0.09]	0.17 [0.11; 0.23]	0.12 [0.05; 0.20]	0.12 [0.03; 0.20]	0.06 [0.03; 0.10]	0.17 [0.11; 0.23]	0.12 [0.05; 0.20]	0.12 [0.03; 0.23]
	$\tau' = \sigma' = 10.0$	13463.0					0.05 [0.02; 0.09]	0.17 [0.11; 0.24]	0.10 [0.03; 0.18]	0.07 [0.00; 0.21]
Mean survival										
Base-case NMAs		13462.0	1.32 [0.84; 5.04]	2.91 [0.92; 15.39]	2.12 [1.19; 8.54]	1.82 [1.26; 6.89]	1.62 [1.07; 2.44]	4.76 [3.50; 6.18]	2.70 [1.63; 4.53]	2.00 [1.22; 3.60]
BHM NMAs	$\tau' = \sigma' = 0.01$	13464.6	2.48 [1.86; 3.22]	4.63 [3.77; 5.57]	4.82 [4.00; 5.70]	4.61 [3.84; 5.41]	2.45 [1.87; 3.12]	4.60 [3.84; 5.36]	4.79 [4.01; 5.57]	4.57 [3.79; 5.38]
	$\tau' = \sigma' = 0.10$	13464.6	2.50 [1.85; 3.29]	4.67 [3.69; 5.79]	4.82 [3.91; 5.83]	4.65 [3.73; 5.68]	2.44 [1.86; 3.14]	4.58 [3.77; 5.49]	4.74 [3.92; 5.61]	4.57 [3.68; 5.55]
	$\tau' = \sigma' = 1.00$	13463.9	2.14 [1.44; 2.96]	4.52 [3.34; 5.88]	3.84 [2.37; 5.43]	3.52 [1.77; 5.94]	2.15 [1.45; 2.98]	4.53 [3.35; 5.88]	3.86 [2.36; 5.44]	3.57 [1.77; 5.95]
	$\tau' = \sigma' = 10.0$	13463.0					1.87 [1.19; 2.80]	4.65 [3.41; 6.04]	3.23 [1.87 5.20]	2.53 [1.27 5.59]
Incremental mean survival compared with docetaxel										
Base case		13462.0		1.54 [-0.06; 11.17]	0.76 [0.24; 3.90]	0.52 [-1.33; 2.82]		3.11 [1.65; 4.66]	1.06 [0.41; 2.37]	0.42 [-0.57; 1.68]
BHM	$\tau' = \sigma' = 0.01$	13464.6		2.14 [1.14; 3.15]	2.33 [1.39; 3.25]	2.12 [1.18; 3.02]		2.13 [1.15; 3.10]	2.33 [1.42; 3.21]	2.12 [1.21; 3.01]
	$\tau' = \sigma' = 0.10$	13464.6		2.16 [1.12; 3.27]	2.31 [1.35; 3.28]	2.14 [1.15; 3.17]		2.13 [1.11; 3.18]	2.28 [1.37; 3.19]	2.11 [1.15; 3.12]
	$\tau' = \sigma' = 1.00$	13463.9		2.36 [1.03; 3.92]	1.65 [0.67; 2.96]	1.33 [-0.05; 3.61]		2.35 [1.04; 3.93]	1.67 [0.68; 2.96]	1.36 [-0.05; 3.64]
	$\tau' = \sigma' = 10.0$	13463.0						2.75 [1.25; 4.37]	1.33 [0.50; 2.77]	0.66 [-0.45; 3.35]

BHM: Bayesian hierarchical modeling; LOOIC: Leave one out information criterion; NMA: Network meta-analysis.

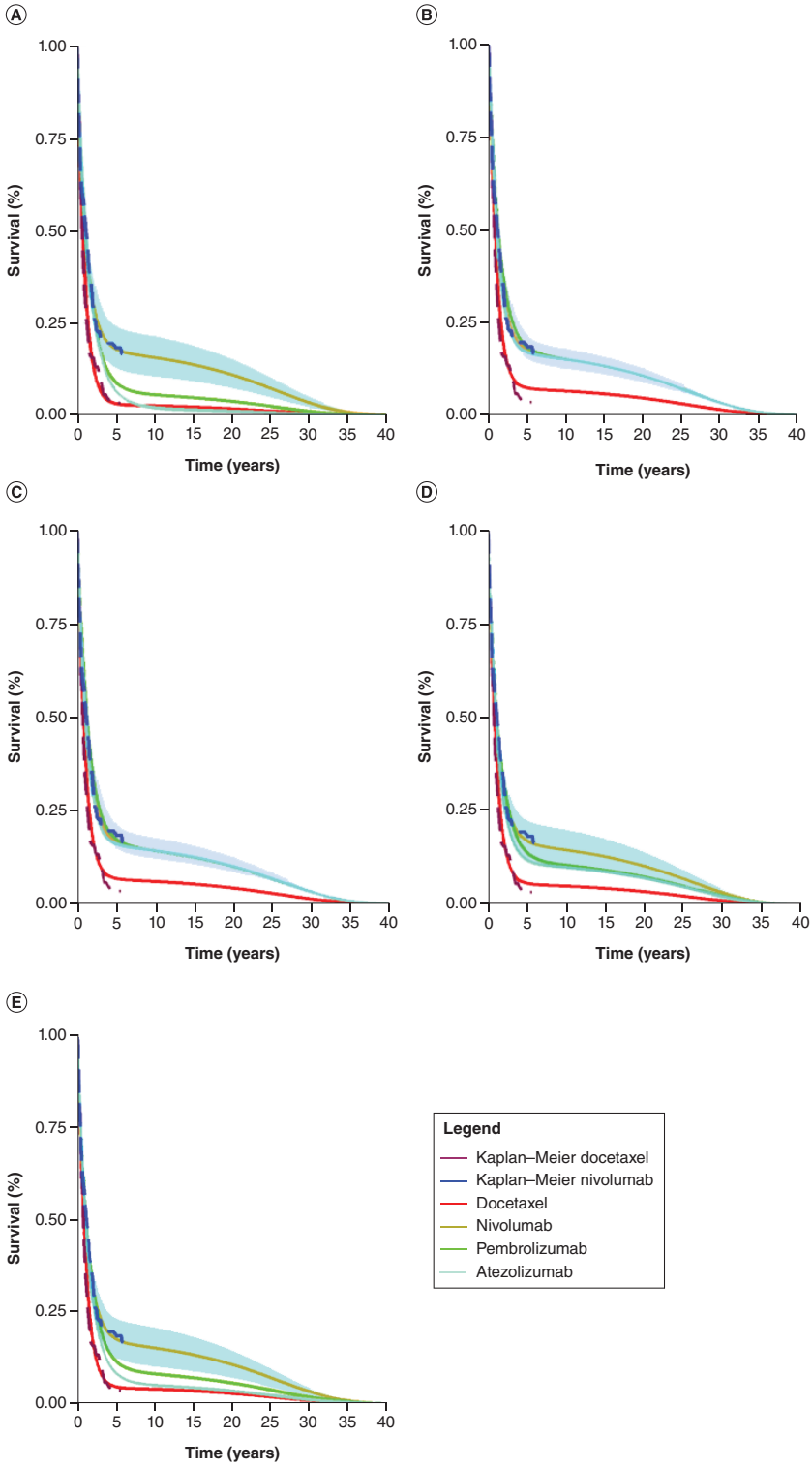

Figure 2. Predicted survival and corresponding 95% credible interval on therapy with longest predicted survival.
Weibull mixture cure predicted mean survival for docetaxel, nivolumab, pembrolizumab and atezolizumab with corresponding uncertainty for the therapy with the longest predicted survival for **(A)** base case Weibull mixture cure network meta-analysis model; **(B)** BHM mixture cure model with limited variance cure class effects with $\tau' = \sigma' = 0.01$; **(C)** BHM mixture cure model with moderate variance over cure class effect with $\tau' = \sigma' = 0.1$; **(D)** BHM mixture cure model with substantial variance over cure class effects with $\tau' = \sigma' = 1$ and **(E)** BHM mixture cure model with large variance over cure class effects with $\tau' = \sigma' = 10$.
BHM: Bayesian hierarchical modeling.

Rhat was 1 for all parameters, the effective sample sizes were high. Upon visual inspection the density, trace and autocorrelation plots seemed reasonable and the divergent cases were deemed to have no impact.

Discussion

A key issue in oncology HTA submissions is the immaturity of trial data, especially for OS. However, several novel therapies, like immune-oncology treatments, have mature datasets for long-term survival, or even cure for a fraction of patients. Standard and flexible extrapolation techniques cannot capture these survival tails based on immature data only, and are therefore likely to underestimate survival. HTA guidance suggests that external data may be leveraged from long-term survival extrapolations [11]. In that context, TTE BH WMC NMA models can leverage mature data of therapies in the evidence network to inform survival extrapolations of same-class therapies with immature survival data.

The current research assessed this approach in an evidence network of patients with previously treated metastatic non-small-cell lung cancer PD-L1 >1%, where IOs were compared with docetaxel. The Bayesian hierarchical approach was embedded in a WMC NMA. Class effects (common distributions) on the cure rates for docetaxel and on treatment effects on cure of IOs were specified in the BHM to leverage the mature data for pembrolizumab and nivolumab, in order to inform the extrapolations of immature atezolizumab data, in terms of predicted cure rates. These BH WMC NMA models were compared with a standard non hierarchical WMC NMA approach.

The standard WMC NMA could not capture differentiating cure rates for atezolizumab 0.03 (0.00; 0.09) compared with docetaxel 0.03 (0.01; 0.07), but was able to capture differentiating cure rates for nivolumab 0.18 (0.12; 0.24) and pembrolizumab 0.07 (0.02; 0.15). The BH WMC NMA scenario had moderate variance ($\tau' = \sigma' = 1$) over the class effects, with predicted cure rates of 0.06 (0.03; 0.10), 0.17 (0.11; 0.23), 0.12 (0.05; 0.20) and 0.12 (0.03; 0.23) for docetaxel, nivolumab, pembrolizumab and atezolizumab, respectively.

Regarding validation of the predicted cure rates of atezolizumab, the reported 4-year survival percentages were 15 and 8% in OAK, and 16 and 9% in POPLAR for atezolizumab and docetaxel, respectively. Both studies included ITT populations (PD-L1 <1%), which typically have poorer survival outcomes than PD-L1 >1%. This was illustrated by Mazieres *et al.*, 2021, in which most events in the tail of the ITT population were explained by the PD-L1 <1% population [32]. Therefore, these long-term survival percentages might be higher for the PD-L1 atezolizumab in the PD-L1 >1% population, as the evidence network for this paper only considered PD-L1 >1%. These ITT 4-year survival percentages aligned with the atezolizumab cure rates predicted by most BHM models (Table 1). The base case WMC predicted lower cure rates for docetaxel (3%) and atezolizumab (3%).

These findings are key in the global reimbursement process, which relies on long-term survival extrapolations in cost–effectiveness analyses. The base-case WMC NMA showed the incremental mean survival predictions compared with docetaxel (with CheckMate as the reference trial) of 3.11 (1.65; 4.66), 1.06 (0.41; 2.37) and 0.42 (-0.57; 1.68) life years for nivolumab, pembrolizumab and atezolizumab, respectively. For instance, the BH WMC NMA model with $\tau' = \sigma' = 0.10$ estimated 2.13 (1.11; 3.18), 2.28 (1.37; 3.19) and 2.11 (1.15; 3.12) years, respectively. The base-case WMC NMA and the BH WMC NMA model would clearly result in contrasting incremental cost–effectiveness ratios for nivolumab, pembrolizumab and atezolizumab and would therefore result in different reimbursement decisions.

The BH WMC NMA approach presented here can easily be extended to many parametric TTE NMA approaches (e.g., standard parametric, fractional polynomials, spline, mixture, mixture and non mixture cure and piecewise NMAs) [37–39]. This paper only assumed class effects on cure parameters, not on shape and scale parameters, since the focus was on long-term survival prediction. However, class effects can also be applied on the shape and scale parameters of the Weibull distribution describing survival of the uncured fraction. This would probably further reduce differences in immune-oncology therapies in the evidence network in terms of outcomes.

The BH WMC NMA approach is particularly advantageous when there is a difference in maturity across the trials in the evidence network of same-class therapies. The variances τ' and σ' over the class effects have a clear impact on the outcomes. The amount of variance that is assumed in the base case should depend on clinical plausibility of the extrapolations, as informed by clinical expert opinions. However, evidence from this type of TTE BH WMC NMA in HTA submissions should include full transparency, by showing standard TTE NMA extrapolations and BH WMC NMA extrapolations with different levels of variance; in addition, the base-case extrapolation method for modeling purposes should be validated by expert opinion.

Regarding class effects, it is unclear if there are differences in the efficacy and safety of atezolizumab, which is a programmed death-1 (PD-1), versus nivolumab and pembrolizumab which are PD-L1s [40]. For demonstrating the

use of BH WMC NMA with selective immature data in the evidence network, PD-1s and PD-L1s were assumed to be of a similar class. Variance, however, was allowed over the treatment effects of the PD-1 and PD-L1s in the evidence network. Therefore, variations in treatment effects across the therapies in the network were allowed, depending on the amount of variance.

The BH WMC NMA TTE NMA approach does not work in cases where there is no evidence network. In this case then, the approach by Soikkeli *et al.* [41], which defined priors based on historical data to inform trial-based immature survival extrapolations based on standard parametric curves, could be extended by using external data to define priors for the cure rate of the standard-of-care arm to inform the MCM predictions. This mature external data could be based on previous trials or real-world evidence, if the homogeneity assumption is not violated. Alternatively, a frequentist approach as suggested by Pennington *et al.* could be considered [42].

In the fixed effects model, the BHM naturally extends to three or more arm studies. In random effects, both BHM and standard time to event NMAs are more complex due to the need to avoid double counting. Consistency needs to be assessed in any NMA [43,44].

This study was for demonstration purposes only. Although both MCM and non-MCMs are able to statistically estimate cure fractions, one should always consider the clinical plausibility of establishing cure, given a treatment mechanism in a particular disease. In case cure is clinically plausible but mortality is a rare event, one may consider earlier end points, such as progression free survival or relapse free survival to estimate a cure fraction. These earlier end points include more events, and therefore the evidence is more likely to be mature.

Conclusion

This research concluded that the BH WMC NMA approach can overcome selected immature trial data in an evidence network. This has important consequences for predicted mean life years which, in turn, affects incremental cost–effectiveness ratios and reimbursement decisions.

Summary points

- For health technology assessment, lifetime survival extrapolations are needed for health economic modelling.
- For a fraction of patients in some oncology indications, immune-oncology (IO) and gene therapies result in negligible disease related risk of events in the long term.
- For these lifetime survival extrapolations, flexible parametric approaches (e.g., mixture and non mixture cure) are suggested in order to capture the shape of the hazard over time.
- In health technology assessment, overall survival data from the pivotal trial is often immature.
- The statistical cure rates when data is immature, based on mixture and non mixture cure models, are either not estimable or often estimated with large uncertainty.
- Guidance on extrapolating survival data indicates that external data from other treatments may be used to reduce this uncertainty.
- Bayesian hierarchical modeling frameworks may assume that treatment effects are exchangeable, in other words, drawn from the same distribution of effects.
- We assumed a class effect on the cure treatment effects of IO therapies in a Bayesian hierarchical mixture cure network meta-analyses, based on a network of three IOs and four trials, of which two were relatively immature.
- This research concludes that when applied in a Bayesian hierarchical NMA framework, mixture cure models can overcome selected immature trial data in an evidence network.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/cer-2022-0159

Financial & competing interests disclosure

B Heeg, A Verhoek, G Tremblay, O Harari, M Soltanifar are employees of Cytel which has received consulting fees from various manufacturers unrelated to this research. JC Cappelleri, H Chu and S Roychoudhury are employees of Pfizer. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

References

1. Latimer NR. NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials – extrapolation with patient-level data. (2011). http://www.nicedsu.org.uk
2. Soerensen AV, Ellebaek E, Bastholt L, Schmidt H, Donia M, Svane IM. Improved progression-free long-term survival of a nation-wide patient population with metastatic melanoma. *Cancers (Basel)* 12(9), 2591 (2020).
3. Wolchok JD, Chiarion-Sileni V, Gonzalez R *et al.* Long-term outcomes with nivolumab plus ipilimumab or nivolumab alone versus ipilimumab in patients with advanced melanoma. *J. Clin. Oncol.* 40(2), 127–137 (2022).
4. Rutherford MJ, Lambert PC, Sweeting MJ *et al.* NICE DSU Technical Support Document 21. In: *Flexible Methods for Survival Analysis.* National Institute for Health and Care Excellence (NICE), UK (2020). www.nicedsu.org.uk
5. 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* 37(9), 1129–1138 (2019).
6. Bullement A, Latimer NR, Bell Gorrod H. Survival extrapolation in cancer immunotherapy: a validation-based case study. *Value Health* 22(3), 276–283 (2019).
7. Gibson E, Koblbauer I, Begum N *et al.* Modelling the survival outcomes of immuno-oncology drugs in economic evaluations: a systematic approach to data analysis and extrapolation. *Pharmacoeconomics* 35(12), 1257–1270 (2017).
8. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* 8(3), 576–594 (2006).
9. Lambert PC, Thompson JR, Weston CL, Dickman PW. Estimating and modeling the cure fraction in population-based cancer survival analysis. *Biostatistics* 8(3), 576–594 (2007).
10. Cislo PR, Emir B, Cabrera J, Li B, Alemayehu D. Finite mixture models, a flexible alternative to standard modeling techniques for extrapolated mean survival times needed for cost-effectiveness analyses. *Value Health* 24(11), 1643–1650 (2021).
11. Heeg B, Garcia A, Beekhuizen SV *et al.* Novel and existing flexible survival methods for network meta-analyses. *J. Comp. Eff. Res.* doi:10.2217/cer-2022-0044 (2022) (Epub ahead of print).
12. Efthimiou O, Debray TP, van Valkenhoef G *et al.* GetReal in network meta-analysis: a review of the methodology. *Res. Synth. Methods* 7(3), 236–263 (2016).
13. Thall PF, Simon R. Practical Bayesian guidelines for phase IIB clinical trials. *Biometrics* 50(2), 337–349 (1994).
14. Thall PF, Wathen JK, Bekele BN, Champlin RE, Baker LH, Benjamin RS. Hierarchical Bayesian approaches to phase II trials in diseases with multiple subtypes. *Stat. Med.* 22(5), 763–780 (2003).
15. Murphy P, Glynn D, Dias S *et al.* Modelling approaches for histology-independent cancer drugs to inform NICE appraisals: a systematic review and decision-framework. *Health Technol. Assess.* 25(76), 1–228 (2021).
16. Herson J. Predictive probability early termination plans for phase II clinical trials. *Biometrics* 35(4), 775–783 (1979).
17. Tan SB, Machin D. Bayesian two-stage designs for phase II clinical trials. *Stat. Med.* 21(14), 1991–2012 (2002).
18. Heitjan DF. Bayesian interim analysis of phase II cancer clinical trials. *Stat. Med.* 16(16), 1791–1802 (1997).
19. Zhang J, Chu H, Hong H, Virnig BA, Carlin BP. Bayesian hierarchical models for network meta-analysis incorporating nonignorable missing ness. *Stat. Methods Med. Res.* 26(5), 2227–2243 (2017).
20. Zhang J, Ko CW, Nie L, Chen Y, Tiwari R. Bayesian hierarchical methods for meta-analysis combining randomized-controlled and single-arm studies. *Stat. Methods Med. Res.* 28(5), 1293–1310 (2019).
21. Kim J, Cho J, Lee MH, Lim JH. Relative efficacy of checkpoint inhibitors for advanced NSCLC according to programmed death-ligand-1 expression: a systematic review and network meta-analysis. *Sci. Rep.* 8(1), 11738–11738 (2018).
22. Borghaei H, Gettinger S, Vokes EE *et al.* Five-year outcomes from the randomized, phase III trials CheckMate 017 and 057: nivolumab versus docetaxel in previously treated non-small-cell lung cancer. *J. Clin. Oncol.* 39(7), 723–733 (2021).
23. Herbst RS, Garon EB, Kim DW *et al.* Long-term outcomes and retreatment among patients with previously treated, programmed death-ligand 1-positive, advanced non-small-cell lung cancer in the KEYNOTE-010 Study. *J. Clin. Oncol.* 38(14), 1580–1590 (2020).
24. Fehrenbacher L, Spira A, Ballinger M *et al.* Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase II randomised controlled trial. *Lancet* 387(10030), 1837–1846 (2016).
25. Fehrenbacher L, von Pawel J, Park K *et al.* Updated efficacy analysis including secondary population results for OAK: a randomized phase III study of atezolizumab versus docetaxel in patients with previously treated advanced non-small cell lung cancer. *J. Thorac. Oncol.* 13(8), 1156–1170 (2018).
26. Mitchell M, Muftakhidinov B, Winchen T. Engauge digitizer software. (2020). https://markummitchell.github.io/engauge-digitizer/

27. Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Med. Res. Methodol.* 12(1), 9 (2012).
28. Office for National Statistics. National life tables, UK: 2016 to 2018. (2019). www.ons.gov.uk/releases/nationallifetablesuk2016to2018
29. Andersson TM, Dickman PW, Eloranta S, Lambe M, Lambert PC. Estimating the loss in expectation of life due to cancer using flexible parametric survival models. *Stat. Med.* 32(30), 5286–5300 (2013).
30. van Oostrum I, Ouwens M, Remiro-Azócar A *et al.* Comparison of parametric survival extrapolation approaches incorporating general population mortality for adequate health technology assessment of new oncology drugs. *Value Health* 24(9), 1294–1301 (2021).
31. Office for National Statistics. National life tables, UK: 2016 to 2018. (2019). www.ons.gov.uk/releases/nationallifetablesuk2016to2018
32. Mazieres J, Rittmeyer A, Gadgeel S *et al.* Atezolizumab versus docetaxel in pretreated patients with NSCLC: final results from the randomized phase II POPLAR and phase III OAK clinical trials. *J. Thorac. Oncol.* 16(1), 140–150 (2021).
33. Mazieres J, Rittmeyer A, Gadgeel S *et al.* Package 'netmeta'. Network meta-analysis using frequentist methods (Version 0.7-0). (2015). https://link.springer.com/book/10.1007/978-3-319-21416-0
34. Stan Development Team. RStan: the R interface to Stan. R package version, 1.2-0. https://mc-stan.org/rstan/
35. Hoffman MD, Gelman A. The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.* 15(1), 1593–1623 (2014).
36. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Stat. Sci.* 7(4), 457–472 (1992).
37. Ouwens MJ, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. *Res. Synthesis Meth.* 1(3-4), 258–271 (2010).
38. Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Med. Res. Methodol.* 11(1), 1–14 (2011).
39. Freeman SC, Carpenter JR. Bayesian one-step IPD network meta-analysis of time-to-event data using Royston–Parmar models. *Res. Synthesis Meth.* 8(4), 451–464 (2017).
40. Banna GL, Cantale O, Bersanelli M *et al.* Are anti-PD1 and anti-PD-L1 alike? The non-small-cell lung cancer paradigm. *Onc. Rev.* 14(2), 490 (2020).
41. Soikkeli F, Hashim M, Ouwens M, Postma M, Heeg B. Extrapolating survival data using historical trial-based a priori distributions. *Value Health* 22(9), 1012–1017 (2019).
42. Pennington M, Grieve R, der Meulen JV, Hawkins N. Value of external data in the extrapolation of survival data: a study using the NJR data set. *Value Health* 21(7), 822–829 (2018).
43. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J. Am. Stat. Assoc.* 101(474), 447–459 (2006).
44. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials. *Med. Decis. Making* 33(5), 641–656 (2013).