Research Synthesis Methods (2025), **16**: 650–671
doi:10.1017/rsm.2025.21

Research Synthesis Methods

RESEARCH ARTICLE

One-step parametric network meta-analysis models using the exact likelihood that allow for time-varying treatment effects

Harlan Campbell[1], Dylan Maciel[1], Keith Chan[1], Jeroen P. Jansen[1], Sven Klijn[2], Kevin Towle[1], Bill Malcolm[3] and Shannon Cope[1]

[1]Health Economics and Outcomes Research, Precision AQ, Vancouver, BC, Canada
[2]Bristol Myers Squibb, Princeton, NJ, USA
[3]Bristol Myers Squibb, Uxbridge, UK

Corresponding author: Shannon Cope; Email: shannon.cope@precisionvh.com

Received: 31 January 2024; **Revised:** 10 March 2025; **Accepted:** 11 March 2025; published online 15 May 2025

Keywords: comparative effectiveness research; evidence synthesis; network meta-analysis; parametric models; time-to-event outcomes

Abstract

The importance of network meta-analysis (NMA) methods for time-to-event (TTE) that do not rely on the proportional hazard (PH) assumption is increasingly recognized in oncology, where clinical trials evaluating new interventions versus standard comparators often violate this assumption. However, existing NMA methods that allow for time-varying treatment effects do not directly leverage individual events and censor times that can be reconstructed from Kaplan–Meier curves, which may be more accurate than discrete hazards. They are also challenging to implement given reparameterizations that rely on discrete hazards. Additionally, two-step methods require assumptions regarding within-study normality and variance. We propose a one-step fully Bayesian parametric individual patient data (IPD)-NMA model that fits TTE data with the exact likelihood and allows for time-varying treatment effects. We define fixed or random effects with the following distributions: Weibull, Gompertz, log-normal, log-logistic, gamma, or generalized gamma distributions. We apply the one-step model to a network of randomized controlled trials (RCTs) evaluating multiple interventions for advanced melanoma and compare results with those obtained with the two-step approach. Additionally, a simulation study was performed to compare the proposed one-step method to the two-step method. The one-step method allows for straightforward model selection among the "standard" distributions, now including gamma and generalized gamma, with treatment effects on either the scale alone or with multivariate treatment effects. Generalized gamma offers flexibility to model U-shaped hazards within a network of RCTs, with accessible interpretation of parameters that simplifies to exponential, Weibull, log-normal, or gamma in special cases.

Highlights

What is already known

Network meta-analysis (NMA) methods that do not rely on the proportional hazard assumption (for time-to-event outcomes) are often required in oncology. Individual event and censor times from clinical trials can typically be reconstructed from published Kaplan–Meier curves, but existing NMA methods used to synthesize these trials do not directly leverage this individual-level data.

This article was awarded Open Data and Open Materials badges for transparent practices. See the Data availability statement for details.

© Bristol-Myers Squibb Company, 2025. Published by Cambridge University Press on behalf of The Society for Research Synthesis Methodology. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

What is new

We propose using a one-step fully Bayesian parametric NMA model that leverages individual event and censor times with the exact likelihood and allows for time-varying treatment effects. By using the exact likelihood specification, this model offers a flexible fit while avoiding certain assumptions/approximations required with existing approaches. The Bayesian framework ensures that model selection and interpretation is straightforward and efficient computation for fitting the model can be done using Stan software.

Potential impact for RSM readers outside the authors' field

Flexible NMA methods are particularly important when there are differences in the survival distributions for the different treatments being compared, and relative treatment effects need to be extrapolated beyond the available trial data for a cost-effectiveness analysis. In this context, the one-step parametric NMA model offers a valuable tool that leverages individual-level data.

1. Introduction

Network meta-analysis (NMA) is a widely used statistical technique that expands upon standard pairwise meta-analysis, enabling simultaneous direct and indirect comparisons of multiple interventions within a unified statistical model.[1–3] Given that randomized controlled trials (RCTs) in oncology rarely compare the new intervention to all relevant comparators, NMAs are often needed to synthesize time-to-event outcomes (TTE), such as progression-free survival (PFS) and overall survival (OS).[4]

The synthesis of TTE outcomes is typically based on hazard ratios (HRs)[5] derived from the Cox proportional hazards (PH) model.[6,7] However, for many reasons, the PH assumption may be implausible,[8–10] for example, due to treatment effects that vary over time, treatment crossover, delayed treatment effects, competing risks, time-varying covariates, and differential dropout.[11–13] When the PH assumption does not hold, NMAs that rely on trial-specific HRs may yield biased estimates that result in poor predictions,[14] which are important for cost-effectiveness modeling.

Several methods have been proposed as alternatives to NMAs based on reported HRs, which have been reviewed by Cope et al.[14] Crowther et al.[15] proposed a model for individual patient data (IPD) meta-analysis using Poisson regression models that allows for non-proportional hazards by prespecifying specific time-points at which the HR between two treatments is allowed to change. Freeman and Carpenter[16] proposed NMA models using cubic splines to model the log-cumulative hazard functions. While flexible, these cubic spline models are restricted to HRs that specifically vary with log-time. Another possibility, proposed by Petit et al.[17] is to consider the difference in restricted mean survival time at a prespecified time horizon as an alternative to the HR and to fit a standard univariate NMA. While practical, prespecifying the time horizon of interest in accordance with clinical interest may be challenging.

Ouwens et al.[18] and Jansen[19] proposed using multivariate parametric NMA models that allow for the HR to change over time. A single parametric survival distribution is used to model interventions in each trial with multivariate relative treatment effects, which are pooled and indirectly compared across trials. This means that interventions can impact both the shape and scale of a selected distribution, resulting in a time-varying treatment effect, which improves model fit and predictions. However, these models were proposed prior to the development of an algorithm to reconstruct event and censor times from the published Kaplan–Meier (KM) curves.[20] Consequently, they derived discrete hazards from "binned" events over time assuming a binomial likelihood per interval, which may be less accurate than assuming the actual event and censor times follow the likelihood for a particular distribution. Moreover, parameters for these models were transformed to a linear scale, and as a consequence, implementation of these models can be challenging given that the parameterization does not align with standard software (i.e., flexsurv in R).[21]

To address these limitations, Cope et al.[22] proposed a two-step NMA approach, first fitting parametric survival distributions using frequentist maximum likelihood estimation to each arm of

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

each trial (using the flexsurv R package[23]). In the second step, the arm-specific parameter estimates, their standard errors, and the correlation between the parameters were synthesized using a Bayesian multivariate Normal NMA.[24]

Burke et al.[25] and Debray et al.[26] discuss the pros and cons of two-step approaches for IPD meta-analysis relative to one-step approaches. Specifically, the two-step approach for NMA by Cope et al.[22] offers practical benefits in terms of using established statistical software packages and possibly improving computational burden.[14,27] However, it was only applied to a subset of relevant standard distributions (Weibull, Gompertz, log-normal, and log-logistic), with parameters transformed to be on a linear scale as in Ouwens et al.[18] It also relies on bootstrapping to account for parameter correlation[26,28] and requires assumptions regarding within-study normality and variance.[25,29] Given that these assumptions can result in certain unexpected biases, Jackson and White[29] recommend that, when possible, to use one-step methods that make fewer normality assumptions. Therefore, there is a need for one-step NMA models with time-varying treatment effects that can also be applied to distributions such as gamma and generalized gamma that relate the available IPD (whether available or reconstructed) directly to the likelihoods of interest without complex reparameterizations.

In this article, a one-step parametric IPD-NMA framework for TTE outcomes is proposed, which allows for time-varying (i.e., multivariate) treatment effects with the exact likelihood. In Section 2, we extend existing one-step PH models (exponential, Weibull, and Gompertz) or accelerated failure time models (Weibull, log-logistic, gamma, log-normal, and generalized gamma)[27,30,31] to have time-varying treatment effects. In Section 3, we evaluate the proposed method in a simulation study, and in Section 4, we apply the model to a network of RCTs evaluating multiple interventions for advanced melanoma regarding OS. We conclude in Section 5 with a general discussion, where we explain that proposed models provide the foundation for extensions to the multilevel network meta-regression (ML-NMR) framework to synthesize IPD and aggregate data (AD), while adjusting for potential prognostic factors and effect modifiers.[31]

2. A parametric IPD-NMA model for TTE outcomes

We follow the parametrization used by Phillippo[31] and adopt the reference treatment parameterization defining t_{ijk} and y_{ijk} as the event time and censoring indicator, respectively, of individual i, in study j, with treatment k, for i in 1,. . .,N, for j in 1,. . .,J, and for k in 1,. . .,K. Importantly, $k = 1$ is considered the "reference treatment." We outline the logic and formulas for the Weibull, which can be extended to other standard distributions.

Phillippo[31] defines a PH Weibull likelihood for IPD (without covariates) as

$$L_{i,j,k}\left(t, y_{i,j,k}\right) = S_{j,k}(t)\lambda_{j,k}(t)^{y_{i,j,k}}, \tag{1}$$

where the hazard function is

$$\lambda_{j,k}(t) = \upsilon_j t^{(\upsilon_j - 1)}\alpha_j \exp\left(\gamma_k\right) \tag{2}$$

and the survivor function is

$$S_{j,k}(t) = \exp\left(-\alpha_j \exp\left(\gamma_k\right) t^{\upsilon_j}\right), \tag{3}$$

where $\gamma_1 = 0$, as treatment 1 is the "reference treatment." As such, each study has a unique baseline hazard function ($\lambda_{j,1}(t) = \upsilon_j t^{\upsilon_j - 1}\alpha_j$) defined by the shape parameter, υ_j, and scale parameter, α_j, for study j, in 1,. . .,J. This specification restricts the scale and shape parameters to $\alpha_j > 0$ and $\upsilon_j > 0$, for j in 1,. . .,J, and for k in 1,. . .,K. It is noteworthy that when the shape parameter, υ_j, is <1, the hazard rate decreases over time, whereas when $\upsilon_j$is >1, the hazard rate increases over time. When $\upsilon_j = 1$, the hazard rate is constant over time, and the Weibull reduces to the exponential distribution.

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

The treatment effect is represented by γ_k, and Phillippo[31] proposes that priors for the model parameters can be specified as

$$\log\left(\alpha_j\right) \sim \text{Normal}\left(0, 100^2\right), \text{for } j \text{ in } 1, \ldots, J; \tag{4}$$

$$\upsilon_j \sim \text{Uniform}\left(0, +\infty\right), \text{for } j \text{ in } 1, \ldots, J; \text{and} \tag{5}$$

$$\gamma_k \sim \text{Normal}\left(0, 100^2\right), \text{for } k \text{ in } 2, \ldots, K. \tag{6}$$

We extend the PH Weibull model proposed by Phillippo[31] by defining the hazard function as:

$$\lambda_{j,k}(t) = \upsilon_{j,k} t^{\left(\upsilon_{j,k}-1\right)} \alpha_{j,k} \tag{7}$$

and the survivor function as

$$S_{j,k}(t) = \exp\left(-\alpha_{j,k} t^{\upsilon_{j,k}}\right), \tag{8}$$

where $\alpha_{j,k}$ is the scale parameter and $\upsilon_{j,k}$ is the shape parameter for the k-th treatment in the j-th study. An NMA model with an arm-based likelihood[31] is then specified as

$$\begin{pmatrix} \log\left(\alpha_{j,k}\right) \\ \log\left(\upsilon_{j,k}\right) \end{pmatrix} = \begin{pmatrix} \mu_{1,j} \\ \mu_{2,j} \end{pmatrix} + \begin{pmatrix} \delta_{1,j,k} \\ \delta_{2,j,k} \end{pmatrix}, \tag{9}$$

where $\delta_{1,j,1} = \delta_{2,j,1} = 0$, for j in 1,…,J. It is noteworthy that, in this notation, the first subscript used for μ and δ differentiates between the shape and scale parameters. Also note that the PH Weibull model from Phillippo[31] can be recovered by setting: $\delta_{1,j,k} = \gamma_k$, $\mu_{1,j} = \alpha_j$, $\exp\left(\mu_{2,j}\right) = \upsilon_j$, and $\exp\left(\delta_{2,j,k}\right) = 1$.

A fixed-effect (FE) NMA model sets $\delta_{1,j,k} = d_{1,k}$ and $\delta_{2,j,k} = d_{2,k}$, for k in 1,…,K, so that treatment effects are not study-specific. Alternatively, a random-effects (RE) NMA model is specified with study-specific treatment effects such that, for j in 1,…,J, we define the following multivariate normal distribution:

$$\begin{pmatrix} \delta_{1,j,2} \\ \delta_{2,j,2} \\ \vdots \\ \delta_{1,j,K} \\ \delta_{2,j,K} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} d_{1,2} \\ d_{2,2} \\ \vdots \\ d_{1,K} \\ d_{2,K} \end{pmatrix}, \Sigma_\delta\right), \tag{10}$$

where Σ_δ, the $2(K-1)$ by $2(K-1)$ covariance matrix (following Cope et al.[22]), is defined as

$$\Sigma_\delta = \begin{pmatrix} \Sigma & 1/2\,\Sigma & \cdots & 1/2\,\Sigma \\ & \Sigma & \vdots & 1/2\,\Sigma \\ & & \ddots & \vdots \\ & & & \Sigma \end{pmatrix}, \text{with}$$

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \tag{11}$$

For both the FE and RE models, the $d_{1,k}$ and $d_{2,k}$ parameters represent the relative effect of treatment k versus treatment 1 within each of the J study populations (see White et al.[32]).

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

All Bayesian models require defining priors for all model parameters. For j in 1,. . .,J, and for k in 2,. . .,K, wide Normal priors with a mean of 0 and a variance of 1,000 can be specified for $\mu_{1,j}$, $\mu_{2,j}$ and for $d_{1,k}$ and $d_{2,k}$ (also following Cope et al.[22]):

$$\begin{aligned} \mu_{1,j} &\sim \text{Normal}\,(0, 1,000), \\ \mu_{2,j} &\sim \text{Normal}\,(0, 1,000), \\ d_{1,k} &\sim \text{Normal}\,(0, 1,000), \text{and} \\ d_{2,k} &\sim \text{Normal}\,(0, 1,000). \end{aligned}$$

In the RE NMA, a prior is also needed for Σ (or alternatively for the individual σ_1, σ_2, and ρ parameters). Following Cope et al.[22] and Jansen [19] we can define the following Inverse-Wishart prior:

$$\Sigma \sim \text{IW}\,(P, 2)\,,$$

where P is a given 2 by 2 scale matrix.

It is noteworthy that, alternatively, a somewhat less complex RE model could be defined whereby RE are specified for only one of the two treatment effect parameters. This simpler model could be useful when there are reasons to believe that one parameter varies between studies while the rest remain relatively constant. For example, a model could define study-specific treatment effects with respect to the scale parameter, while the shape parameter would remain fixed across studies. For instance, one could define, for k in 1, . . .,K, $\delta_{2,j,k}$=$d_{2,k}$, and for j in 1,. . ., J:

$$\begin{pmatrix} \delta_{1,j,2} \\ \delta_{1,j,3} \\ \vdots \\ \delta_{1,j,K} \end{pmatrix} \sim \text{MVN}\left(\begin{pmatrix} d_{1,2} \\ d_{1,3} \\ \vdots \\ d_{1,K} \end{pmatrix}, \begin{pmatrix} \sigma^2 & \sigma^2/2 & \cdots & \sigma^2/2 \\ & \sigma^2 & \vdots & \sigma^2/2 \\ & & \ddots & \vdots \\ & & & \sigma^2 \end{pmatrix} \right). \tag{12}$$

We also note that alternative prior specifications are also possible, and various parameterizations of the variance–covariance matrix specified in equations (11) and (12) could be considered (see Wei and Higgins[33]). Our multivariate Weibull NMA model can be easily generalized to other parametric distributions. We outline standard parametric models in Table 1, including the Weibull, Gompertz, log-normal, log-logistic, gamma, and generalized gamma distributions. The generalized gamma distribution is characterized by three parameters and in our NMA model; only two of these three parameters are study- and treatment-specific. The third parameter, Q, is assumed to have the same value for all individuals across all studies and treatments within the network.

3. Simulation study

In this section, we consider a simple simulation study to investigate the validity of the proposed one-step model and compare it to the two-step model. In this simulation study, both the proposed one-step model and two-step model from Cope et al.[22] are used to analyze 5,000 simulated datasets, assuming FE with the Weibull distribution. Jackson and White[29] suggest that a one-step approach may be preferable due to the "hidden normality" assumptions required with a two-step approach, which can, at least in theory, "have serious implications for the accuracy of the resulting statistical inference." The simulation study is based on the work in Section 7.3 of Phillippo,[31] where a single simulated dataset is considered consisting of two RCTs: one trial comparing treatments A versus B, and the other trial comparing treatments A versus C. Here, we extend the network to include a third trial comparing treatments C versus D. In the Supplementary Material, additional analyses are presented to demonstrate how the proposed one-step model fits when different distributions are assumed.

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

Table 1. *Survival and hazard functions for standard parametric survival models.*

Survival distribution	Parameters: Flexsurv	Parameters: Model	Survival and hazard functions
Weibull (PH parameterization from Flexsurv)	a	$\upsilon_{j,k}$	$S_{j,k}(t) = \exp\left(-\alpha_{j,k} t^{\upsilon_{j,k}}\right)$
	m	$\alpha_{j,k}$	$\lambda_{j,k}(t) = \upsilon_{j,k} t^{\left(\upsilon_{j,k}-1\right)} \alpha_{j,k}$
Gompertz	a	$\log\left(\upsilon_{j,k}\right)$	$S_{j,k}(t) = \exp\left(-\frac{\alpha_{j,k}}{\log\left(\upsilon_{j,k}\right)}\left(\exp\left(t\log\left(\upsilon_{j,k}\right)\right)-1\right)\right)$
	b	$\alpha_{j,k}$	$\lambda_{j,k}(t) = \alpha_{j,k}\exp\left(t\log\left(\upsilon_{j,k}\right)\right)$
Log-normal	μ	$\log\left(\alpha_{j,k}\right)$	$S_{j,k}(t) = 1-\Phi\left(\rho_{j,k}(t)\right)$
	σ	$\upsilon_{j,k}$	$\lambda_{j,k}(t) = \frac{\phi\left(\rho_{j,k}(t)\right)}{t\upsilon_{j,k}\Phi\left(\rho_{j,k}(t)\right)}$
Log-logistic	b	$\alpha_{j,k}$	$S_{j,k}(t) = \frac{1}{1+\left(\frac{t}{\alpha_{j,k}}\right)^{\upsilon_{j,k}}}$
	a	$\upsilon_{j,k}$	$\lambda_{j,k}(t) = \frac{\left(\frac{\upsilon_{j,k}}{\alpha_{j,k}}\right)\left(\frac{t}{\alpha_{j,k}}\right)^{\upsilon_{j,k}-1}}{1+\left(\frac{t}{\alpha_{j,k}}\right)^{\upsilon_{j,k}}}$
Gamma	α	$\upsilon_{j,k}$	$S_{j,k}(t) = 1-\frac{1}{\Gamma\left(\upsilon_{j,k}\right)}\gamma\left(\upsilon_{j,k}, t\alpha_{j,k}\right)$
	σ	$\alpha_{j,k}$	$\lambda_{j,k}(t) = \frac{\alpha_{j,k}{}^{\upsilon_{j,k}}\, t^{\upsilon_{j,k}-1}\exp\left(-\alpha_{j,k}t\right)}{\Gamma\left(\upsilon_{j,k}\right)S_{j,k}(t)}$
Generalized gamma	μ	$\log\left(\alpha_{j,k}\right)$	$S_{j,k}(t) = 1-\frac{\gamma\left(Q^{-2}, Q^{-2}\exp\left(Q\rho_{j,k}(t)\right)\right)}{\Gamma\left(Q^{-2}\right)}$, if $Q \neq 0$ $S_{j,k}(t) = 1-\Phi\left(\rho_{j,k}(t)\right)$, if $Q = 0$
	σ	$\upsilon_{j,k}$	$\lambda_{j,k}(t) = \frac{\lvert Q\rvert\left(Q^{-2}\right)^{Q^{-2}}}{\upsilon_{j,k}t\Gamma\left(Q^{-2}\right)S_{j,k}(t)} \times$
	Q	Q	$\exp\left\{Q^{-2}\left[Q\rho_{j,k}(t)-\exp\left(Q\rho_{j,k}(t)\right)\right]\right\}$

Note: For reference, the parameterization used ("standard parameterization") corresponds to that used in the flexsurv R package[23] following the equivalences listed in the "Parameters" columns. Recall that our model specification (equation (9)) restricts the scale and shape parameter to be strictly positive (i.e., $\alpha_{j,k} > 0$ and $\upsilon_{j,k} > 0$, for j in 1,. . .,*J*, and for k in 1,. . .,*K*). Note that $\phi(\bullet)$ and $\Phi(\bullet)$ denote the probability density function and the cumulative density function of the standard normal distribution, respectively; $\Gamma(\bullet)$ and $\gamma(\bullet,\bullet)$ denote the gamma function and the incomplete gamma function, respectively; and $\rho_{j,k}(t) = \left[\log(t) - \log\left(\alpha_{j,k}\right)\right]/\upsilon_{j,k}$.

3.1. Data-generating mechanism

As illustrated in Figure 1, each artificial dataset includes survival data for three two-arm RCTs: Study $j = 1$ compares treatment $k = 2$ to $k = 1$ (henceforth, "Study AB" comparing intervention B with A); Study $j = 2$ compares treatment $k = 3$ to $k = 1$ ("Study AC"); and Study $j = 3$ compares treatment $k = 4$ to $k = 3$ ("Study CD"). Each study consists of N_{study} individuals randomly assigned (with a 1:1 randomization ratio) to one of the two treatment arms. We considered two scenarios: (1) with $N_{study} = 36$ and (2) with $N_{study} = 100$.

We simulated survival times to obtain time-varying treatment effects. Specifically, survival times were simulated from Weibull distributions corresponding to the FE Weibull model outlined in equations (7)–(9) using the cumulative distribution function inversion method (as implemented in the R package simsurv[34]) with the following parameters: $\mu_{1,1} = \mu_{1,2} = \mu_{1,3} = 1.5$, $\mu_{2,1} = \mu_{2,2} = \mu_{2,3} = 0.5$, $d_{1,2} = 1.2$, $d_{2,2} = 0.6$, $d_{1,3} = 0.5$, $d_{2,3} = -0.3$, $d_{1,4} = -0.5$, and $d_{2,4} = -0.7$.

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

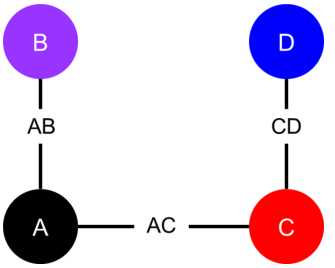


Figure 1. *Network diagram of artificial randomized controlled trials.*

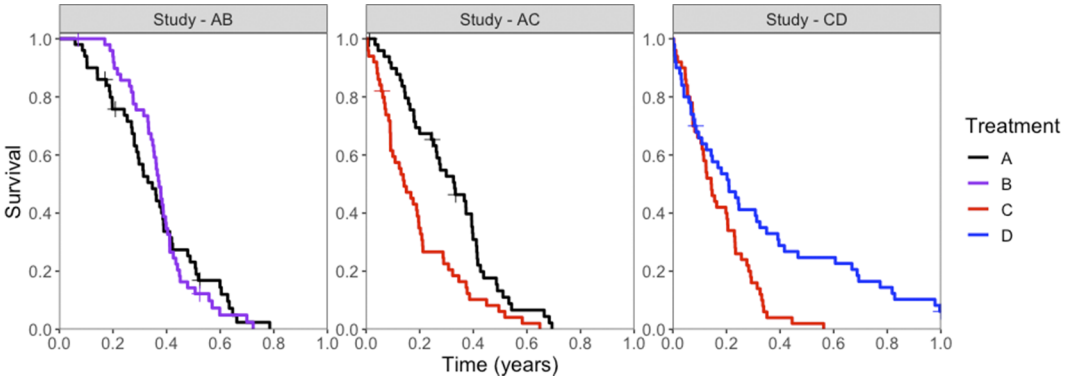


Figure 2. *Kaplan–Meier survival curves for simulated event times, for each treatment (colors) in each study (panels). Censored events are marked with a cross ("+").*

Potential censoring times were simulated independently of event times from a Uniform(0,1) distribution, and a random 10% of individuals were selected for potential censoring. An individual was ultimately censored if their survival time was greater than their potential censoring time. All individuals who had not experienced an event by 1 year were censored at 1 year. KM plots illustrate the event and censor times for one of the simulated datasets in Figure 2.

3.2. Methods to estimate treatment effects based on the simulated data

The relative treatment effects for the competing interventions based on the artificial RCTs in the network were assessed using a one-step FE IPD-NMA model assuming a Weibull distribution for event times (as outlined in equations (7)–(9)). For comparison, we also fit the data using the two-step approach of Cope et al.,[22] also assuming an FE model and a Weibull distribution. Both the one-step and two-step models used the parameterization specified in Table 1. The parameters for these models were estimated using the Markov Chain Monte Carlo (MCMC) method using R (packages: rstan, loo, and flexsurv[23]) and Stan,[35–37] where an initial series of 2,000 iterations from the sampler was discarded (i.e., burn-in) and inferences were based on subsequent 2,000 iterations using four chains.

Results for each dataset were summarized in terms of the posterior medians and 95% credible intervals (CrIs). For each of the two N_{study} scenarios, across all 5,000 datasets, the mean bias, mean credible interval coverage, and mean credible interval width were calculated.

3.3. Results

Results are summarized in Table 2 and suggest that the NMA estimates obtained from both the one-step and two-step models have little to no bias and that both methods provide 95% CrIs with approximately

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

Table 2. *Results from the simulation study: for each parameter, the average estimate (averaged over the 2,000 simulated datasets); the bias (average estimate—truth); 95% CrI coverage (proportion of simulation for which the 95% CrI contained the true parameter value); and average 95% CrI width (averaged over the 5,000 simulated datasets).*

			Estimate		Bias		95% CrI coverage		95% CrI width	
N_{study}	Parameter	Truth	Two-step	One-step	Two-step	One-step	Two-step	One-step	Two-step	One-step
36	$d_{1,2}$	1.20	1.316	1.299	0.116	***0.099***	0.941	0.939	2.524	2.534
	$d_{1,3}$	0.50	0.555	0.550	0.055	***0.050***	0.943	0.941	2.131	2.138
	$d_{1,4}$	−0.50	−0.544	−0.541	−0.044	***−0.041***	0.932	0.933	2.938	2.949
	$d_{2,2}$	0.60	0.604	0.604	0.004	0.004	0.933	0.942	1.053	1.081
	$d_{2,3}$	−0.30	−0.289	−0.29	0.011	***0.010***	0.935	0.940	1.053	1.078
	$d_{2,4}$	−0.70	−0.703	−0.705	***−0.003***	−0.005	0.935	0.940	1.510	1.543
	$\mu_{1,1}$	1.50	1.630	1.584	0.130	***0.084***	0.944	0.944	1.348	1.357
	$\mu_{1,2}$	1.50	1.634	1.588	0.134	***0.088***	0.940	0.942	1.353	1.362
	$\mu_{3,3}$	1.50	1.624	1.575	0.124	***0.075***	0.935	0.937	2.692	2.700
	$\mu_{2,1}$	0.50	0.559	0.541	0.059	***0.041***	0.927	0.944	0.749	0.766
	$\mu_{2,2}$	0.50	0.557	0.540	0.057	***0.040***	0.927	0.941	0.750	0.767
	$\mu_{3,3}$	0.50	0.551	0.529	0.051	***0.029***	0.928	0.940	1.286	1.314
100	$d_{1,2}$	1.20	1.244	1.238	0.044	***0.038***	0.948	0.947	1.421	1.425
	$d_{1,3}$	0.50	0.514	0.512	0.014	***0.012***	0.949	0.948	1.205	1.205
	$d_{1,4}$	−0.50	−0.522	−0.521	−0.022	***−0.021***	0.943	0.943	1.664	1.665
	$d_{2,2}$	0.60	0.603	0.603	0.003	0.003	0.947	0.946	0.627	0.633
	$d_{2,3}$	−0.30	−0.301	−0.301	−0.001	−0.001	0.950	0.950	0.627	0.632
	$d_{2,4}$	−0.70	−0.705	−0.706	***−0.005***	−0.006	0.950	0.950	0.900	0.906
	$\mu_{1,1}$	1.50	1.541	1.525	0.041	***0.025***	0.944	0.941	0.772	0.774
	$\mu_{1,2}$	1.50	1.544	1.528	0.044	***0.028***	0.944	0.944	0.775	0.776
	$\mu_{3,3}$	1.50	1.546	1.529	0.046	***0.029***	0.945	0.945	1.517	1.518
	$\mu_{2,1}$	0.50	0.521	0.515	0.021	***0.015***	0.935	0.938	0.446	0.450
	$\mu_{2,2}$	0.50	0.522	0.516	0.022	***0.016***	0.938	0.948	0.448	0.451
	$\mu_{3,3}$	0.50	0.521	0.513	0.021	***0.013***	0.947	0.948	0.766	0.771

Note: Survival times were simulated from Weibull distributions corresponding to the FE Weibull model with the following parameters: $\mu_{1,1} = \mu_{1,2} = \mu_{1,3} = 1.5$, $\mu_{2,1} = \mu_{2,2} = \mu_{2,3} = 0.5$, $d_{1,2} = 1.2$, $d_{2,2} = 0.6$, $d_{1,3} = 0.5$, $d_{2,3} = -0.3$, $d_{1,4} = -0.5$, and $d_{2,4} = -0.7$. Bolded values indicate smaller bias value (comparing two-step vs. one-step).

correct coverage (i.e., the 95% CrIs obtained contain the true parameter value ~95% of the time). For almost all parameter estimates, the one-step method appears to be less biased than the two-step method. The average widths of the 95% CrIs are very similar for the one-step and two-step models, suggesting that both approaches are equally efficient. Comparing results from the $N_{study} = 36$ scenario with those from the $N_{study} = 100$ scenario, we see that with smaller sample sizes (i.e., with $N_{study} = 36$), the estimates have larger biases, and the 95% CrIs are perhaps slightly too narrow. Finally, with respect to the computational resources required, the one-step model took about twice as long to fit compared to the two-step model when $N_{study} = 36$ and about three times as long to fit when $N_{study} = 100$.

4. Illustrative examples with existing empirical data

The illustrative example was based on a network of studies concerning the treatment of advanced (Stage IIIc or IV) melanoma. Details of the network have been presented previously in Jansen et al. [38]

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

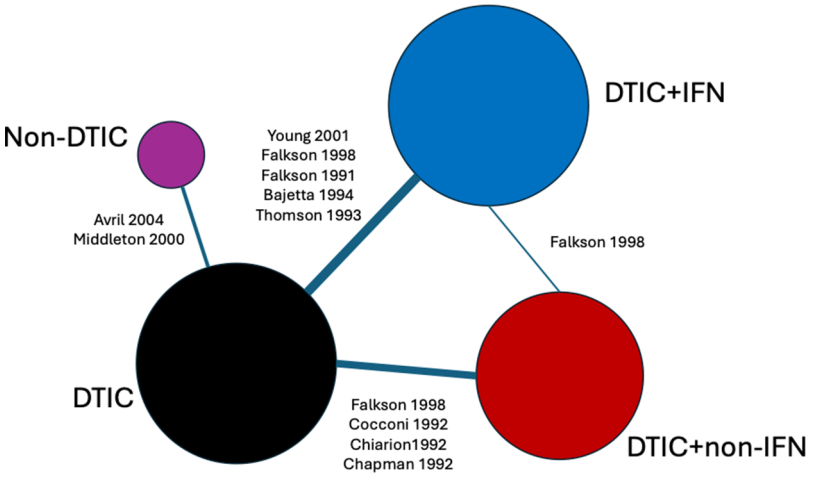


Figure 3. *Network of evidence for melanoma randomized controlled trials. Node size and line thickness correspond to the number of studies, including the treatment and the treatment comparison. Abbreviations: DTIC, dacarbazine; IFN, interferon (IFN).*

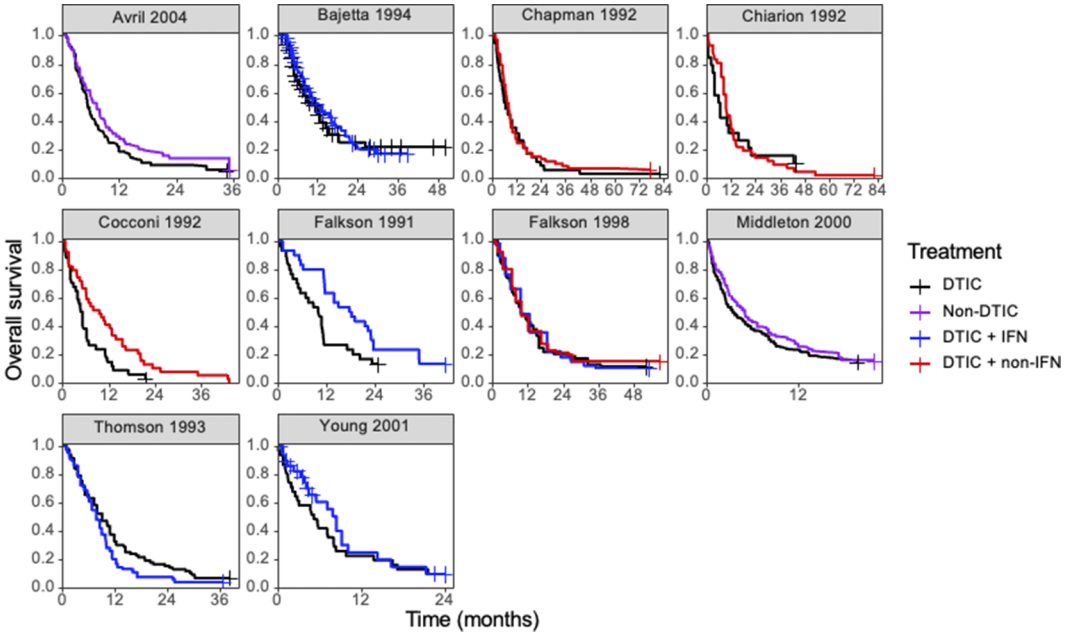


Figure 4. *Kaplan–Meier plots of the reconstructed individual event and censoring times obtained for each randomized controlled trial. Abbreviations: DTIC, dacarbazine; IFN, interferon (IFN).*

and Cope et al.[22] Notably, these studies are somewhat dated, having been published between 1991 and 2004, and we refer readers to Boutros et al.[39] for a review of the current treatment landscape.

Ten RCTs identified in a systematic literature review formed a connected network, illustrated in Figure 3, evaluating four treatments: dacarbazine (DTIC) monotherapy, DTIC + Interferon (IFN), DTIC + Non-IFN, and Non-DTIC. For each treatment arm in each RCT, the reported KM curves were digitized (DigitizeIt; http://www.digitizeit.de/), and the reconstructed individual event and censoring times were obtained using the Guyot et al.[20] algorithm (which was recently recommended by Saluja et al.[40] following an assessment of its reliability, accuracy, and precision) (see Figure 4).

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

Table 3. *For each of the 10 studies in the melanoma network: the number of patients per arm, the number of overall survival events per arm, the median survival time per arm, and the p-value obtained from applying the Grambsch and Therneau test for proportional hazard (PH) assumption.*

Study	Arm	*N* at risk	Total number of events	Median survival time (months)	*p*-value obtained from the Grambsch and Therneau test for PH
Avril et al.[41]	DTIC	117	111	5.44	0.534
	Non-DTIC	112	105	7.36	
Bajetta et al.[42]	DTIC	82	49	11.80	0.167
	DTIC + IFN	76	48	12.84	
Chapman et al.[43]	DTIC	121	117	6.65	0.359
	DTIC + non-IFN	119	112	8.10	
Chiarion Sileni et al.[44]	DTIC	19	17	6.84	0.021
	DTIC + non-IFN	41	40	9.61	
Cocconi et al.[45]	DTIC	34	33	5.06	0.805
	DTIC + non-IFN	39	39	9.67	
Falkson et al.[46]	DTIC	30	26	10.12	0.063
	DTIC + IFN	30	26	17.85	
Falkson et al.[47]	DTIC	69	61	9.73	0.912
	DTIC + IFN	68	61	9.69	
	DTIC + non-IFN	66	56	9.97	
Middleton et al.[48]	DTIC	149	127	3.43	0.198
	Non-DTIC	156	132	5.04	
Thomson et al.[49]	DTIC	82	76	9.20	0.689
	DTIC + IFN	87	83	7.83	
Young et al.[50]	DTIC	31	28	5.20	0.169
	DTIC + IFN	30	20	8.43	

For each of the 10 studies in the melanoma network, Table 3 lists the number of patients per arm, the number of OS events per arm, the median survival time per arm, and the *p*-value obtained from applying the Grambsch and Therneau test for PH.[51] Notably, there is evidence of non-PH in the Chiarion 1992 study, which compared DTIC and DTIC + non-IFN (*p*-value = 0.021), and therefore, according to Cope et al.[14] and recent guidance (e.g., "if the PH assumption is deemed to be implausible for one or more comparisons in the network, then (network) meta-analysis of HRs should not be carried out"[52]), a method that does *not* require the PH assumption should be used for analysis.

4.1. Methods

We fit the one-step and two-step multivariate NMA models with both the fixed and the REs using MCMC with R and Stan[35–37] as described in Section 3. We defined REs on both the shape and scale parameters (following equation (10)) and specified an Inverse-Wishart prior for Σ (following equation (11)) with:

$$P = \begin{pmatrix} 1/10 & 0 \\ 0 & 1/10 \end{pmatrix}.$$

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

Consistent with Cope et al.,[22] we considered four different distributions for the likelihood: the Weibull, the Gompertz, the log-normal, and log-logistic. It is noteworthy that, for the two-step approach, the parameterization originally used by Cope et al.[22] was based on models originally proposed by Ouwens et al.[18] In contrast, we fit all models with the more familiar parameterization detailed in Table 1.

The approximate leave-one-out (LOO) information criterion (LOOIC) was used to select the most appropriate model. The LOOIC is similar to the DIC in that smaller values signal better model fit, but, unlike the DIC, it is invariant to parametrization and corresponds to a model's predictive performance (integrating over the posterior distribution of the parameters) (see Vehtari et al.[53]). Convergence for each parameter was assessed using the Gelman-Rubin statistic.[54,55] Checks for divergent transitions were also performed (see Betancourt et al.[56]).

4.2. Results

All models converged and appeared to fit the data reasonably well (see predicted survival curves in Figures 5 and 6, and MCMC trace plots for the log-logistic models in Figures 7 and 8). However, for all models to successfully converge, the RStan default settings for the "adapt_delta" and "max_treedepth" control variables needed to be changed following current recommendations (see details in code provided in the Supplemental Materials).[37] On a laptop computer (Macbook Pro, with M1 chip and 16 GB memory), the two-step models all took less than a minute to fit, whereas the one-step models took between 1 and 188 minutes (with RE models taking about two and a half times longer to fit than FE models; see Table 4).

The LOOIC suggests that the log-logistic distribution was most appropriate (Table 4), which is consistent with findings by Cope et al.[22] based on the sum of treatment-arm specific AIC values per distribution in the first step of their analysis. The parameter estimates and 95% CrIs for the one-step log-logistic random-effects (and fixed-effect) model are very similar to those obtained using the two-step NMA method (Table 5). The small discrepancies between the estimates may be explained by the assumptions required in the two-step NMA regarding within-study estimates in terms of normality and standard errors and/or by the different priors required for the one-step and two-step models. These findings were consistent across the alternative distributions for the likelihood evaluated (Table 6), which are summarized in terms of the predicted survivals in Figures 5 and 6.

Figure 9 plots the estimated survival curves from the FE and RE one-step models fit to a population with the same baseline risk as the Avril 2004 study population. Within the first 6 months, survival appears highest for DTIC + non-IFN. However, the DTIC + non-IFN and non-DTIC survival curves cross shortly after 6 months (illustrating the effect of time-varying treatment effects), and among all four treatments, long-term survival appears highest for non-DTIC.

The treatment effect estimates can be summarized in many ways, and Cope and Jansen[57] consider how various quantitative summaries compare using the same melanoma network as an illustrative example.[57] Suppose one was specifically interested in OS at 1 year and at 2 years, and was interested in comparing DTIC + IFN and non-DTIC, a treatment comparison for which there is no direct evidence. Table 7 lists the relevant estimates obtained from the one-step and two-step log-logistic models fit to a population with the same baseline risk as the Avril 2004 study population. Briefly, with the one-step FE log-logistic model, the OS at 1 year with DTIC + IFN is estimated to be 20% (95% CrI: [12%, 30%]), slightly lower than with non-DTIC: 30% (95% CrI: [24%, 37%]). The estimated difference in OS at 1 year is −10% (95% CrI: [−20%, 1%]), and at 2 years is −6% (95% CrI: [−11%, −1%]). These findings were consistent for one-step and two-step models. Credible intervals are notably wider with RE models relative to FE models.

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

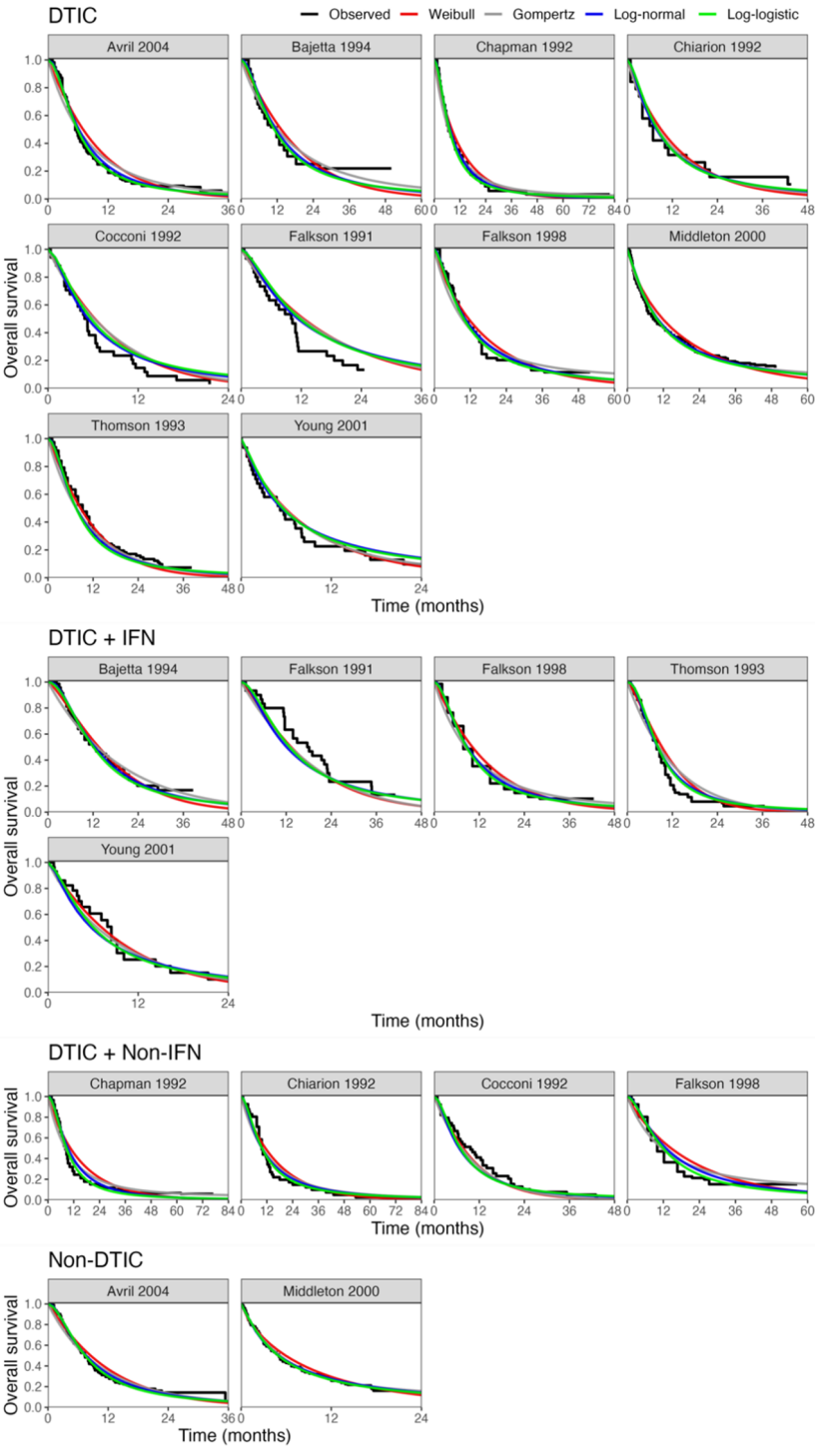


Figure 5. *Fitted survival functions for all distributions from FE one-step model (Weibull, Gompertz, log-normal, and log-logistic) and Kaplan–Meier curves by treatment arm. Abbreviations: DTIC, dacarbazine; IFN, interferon (IFN).*

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

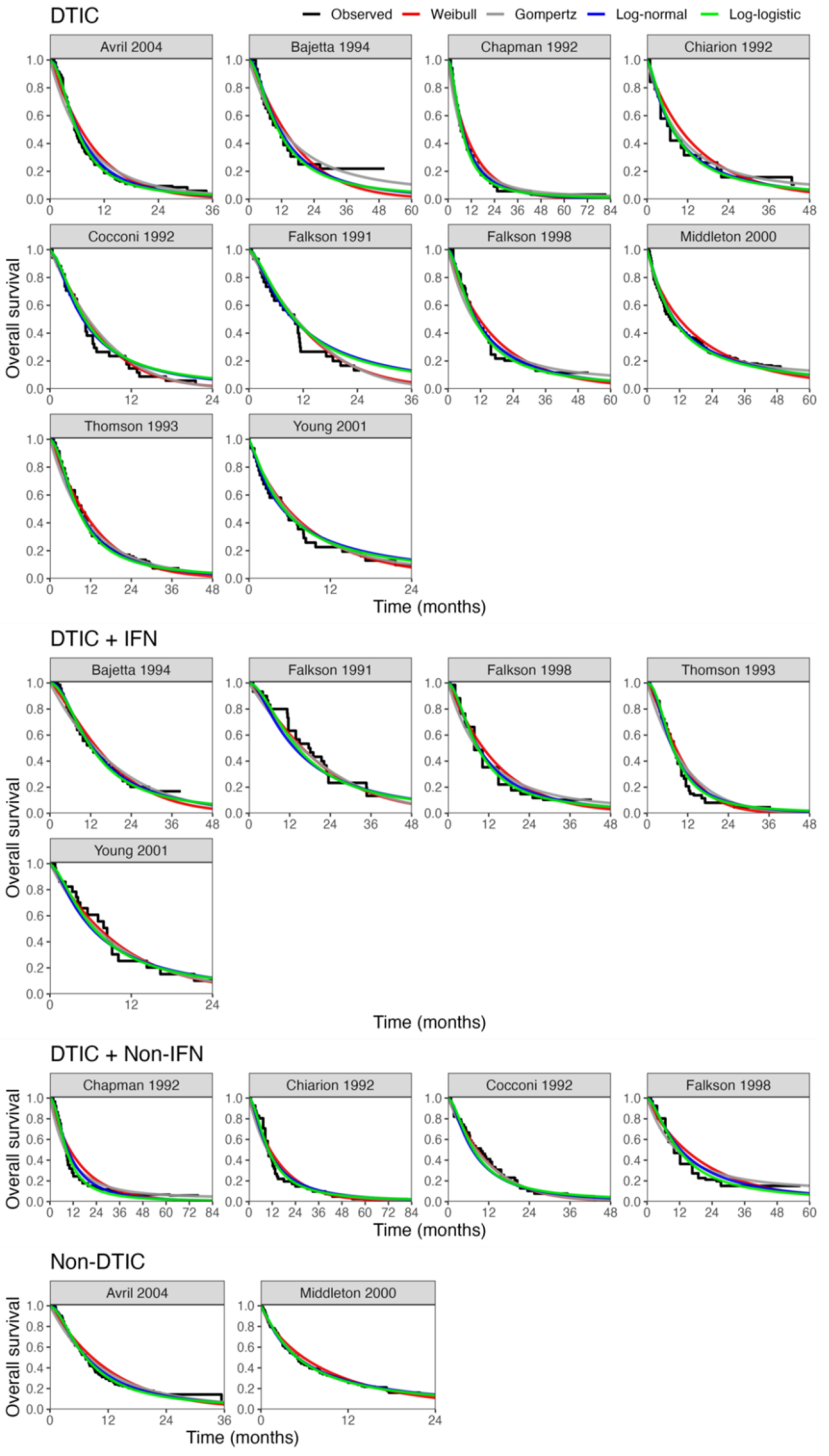


Figure 6. *Fitted study-specific survival functions (using study-specific baseline risk [$\mu_{1,j}$ and $\mu_{1,j}$] and relative treatment effects [$\delta_{1,j,k}$ and $\delta_{2,j,k}$]) for all distributions from the RE one-step model (Weibull, Gompertz, log-normal, and log-logistic) and Kaplan–Meier curves by treatment arm. Abbreviations: DTIC, dacarbazine; IFN, interferon (IFN).*

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

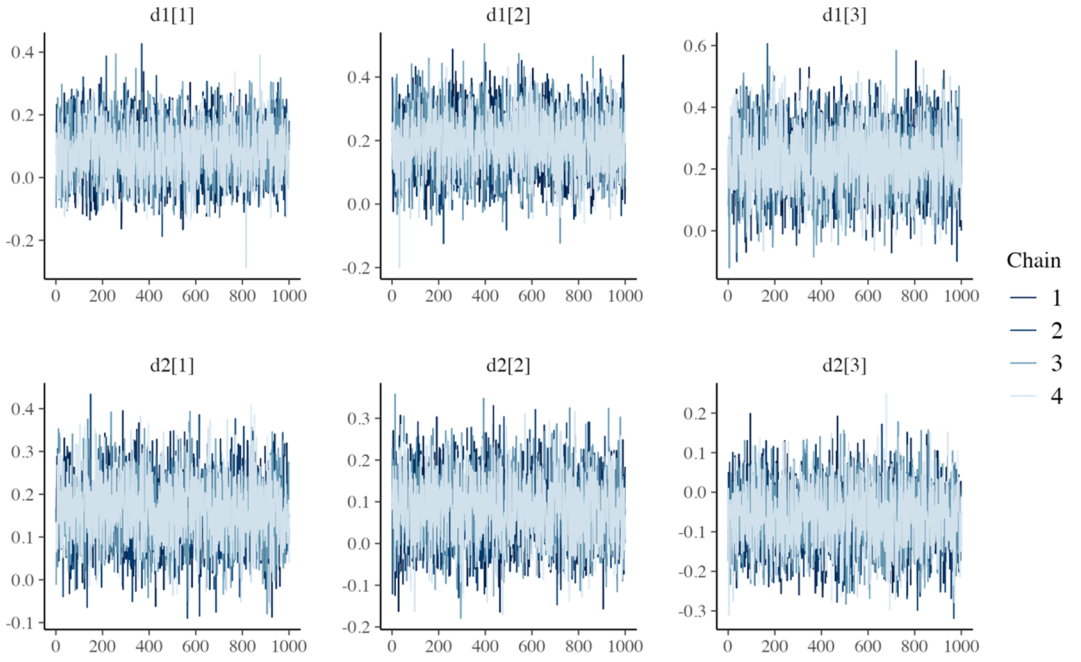


Figure 7. *MCMC trace plots for relative treatment effect parameters of the one-step FE log-logistic NMA model.*

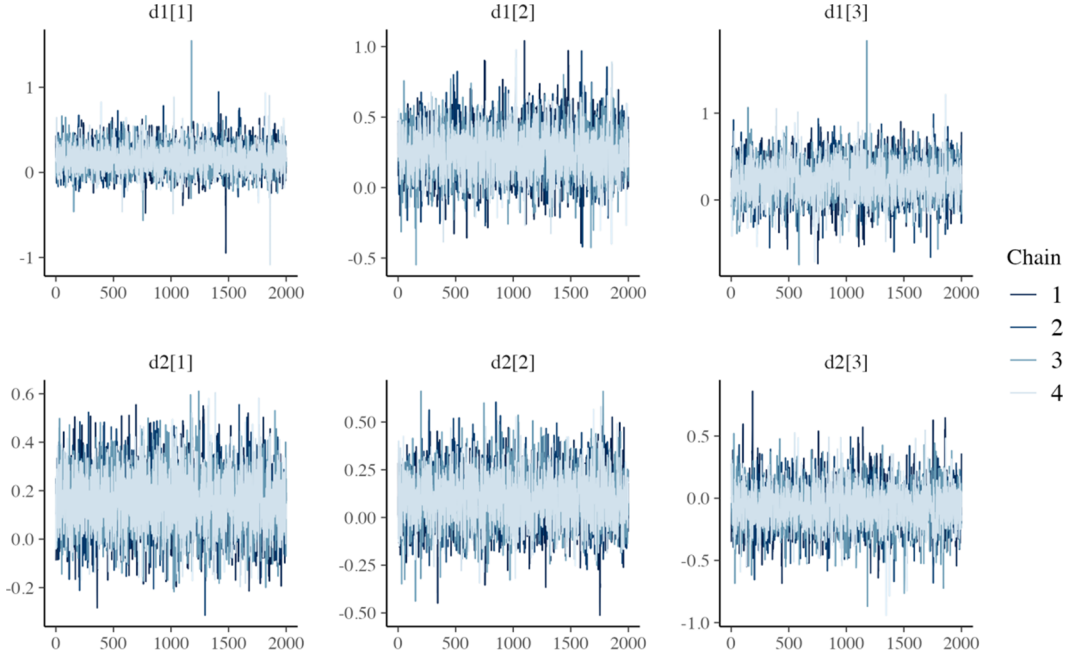


Figure 8. *MCMC trace plots for relative treatment effect parameters of the one-step RE log-logistic NMA model.*

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

Table 4. *Computational sampling time required for each model and leave-one-out information criterion (LOOIC).*

	Computational sampling time (minutes)	LOOIC	SE of LOOIC
RE model			
Log-logistic	3	9357.51	77.32
Log-normal	2	9365.48	76.28
Gompertz	188	9560.78	75.69
Weibull	136	9628.08	70.45
FE model			
Log-logistic	1	9353.32	77.55
Log-normal	1	9358.66	76.44
Gompertz	67	9549.57	75.69
Weibull	53	9623.45	70.46

Note: The LOOIC can be used to select the most appropriate model. The distribution with the lowest LOOIC values, the log-logistic in this case (for both random effects [REs] and fixed effect [FE]), is considered the "best" model. Abbreviation: SE, standard error.

Table 5. *Parameter estimates (posterior medians and 95% CrIs) obtained with random-effects (REs) and fixed-effect (FE) log-logistic NMA models: (1) two-step multivariate network meta-analysis model (Cope et al.[22]) versus (2) the proposed one-step IPD NMA.*

Parameter	2-step log-logistic	1-step log-logistic
RE model		
DTIC + IFN ($d_{1,2}$)	0.171 (−0.093, 0.477)	0.158 (−0.113, 0.459)
DTIC + non-IFN ($d_{1,3}$)	0.251 (−0.048, 0.581)	0.243 (−0.060, 0.567)
Non-DTIC ($d_{1,4}$)	0.239 (−0.185, 0.639)	0.233 (−0.177, 0.627)
DTIC + IFN ($d_{2,2}$)	0.159 (−0.062, 0.377)	0.160 (−0.071, 0.385)
DTIC + non-IFN ($d_{2,3}$)	0.092 (−0.146, 0.342)	0.095 (−0.145, 0.350)
Non-DTIC ($d_{2,4}$)	−0.067 (−0.379, 0.246)	−0.066 (−0.373, 0.253)
FE model		
DTIC + IFN ($d_{1,2}$)	0.096 (−0.065, 0.252)	0.089 (−0.075, 0.248)
DTIC + non-IFN ($d_{1,3}$)	0.197 (0.024, 0.371)	0.192 (0.021, 0.364)
Non-DTIC ($d_{1,4}$)	0.231 (0.034, 0.424)	0.232 (0.036, 0.432)
DTIC + IFN ($d_{2,2}$)	0.157 (0.014, 0.299)	0.161 (0.018, 0.307)
DTIC + non-IFN ($d_{2,3}$)	0.081 (−0.067, 0.230)	0.086 (−0.063, 0.232)
Non-DTIC ($d_{2,4}$)	−0.061 (−0.208, 0.086)	−0.062 (−0.214, 0.087)

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

Table 6. *Parameter estimates (posterior medians and 95% CrIs) obtained with random-effects (REs) and fixed-effect (FE) NMA models: (1) two-step multivariate network meta-analysis model (Cope et al.[22]) versus (2) the proposed one-step IPD NMA.*

Parameter	Weibull 2-step	Weibull 1-step	Gompertz 2-step	Gompertz 1-step	Log-normal 2-step	Log-normal 1-step
RE model						
DTIC + IFN ($d_{1,2}$)	−0.475 (−1.006, 0.083)	−0.472 (−1.018, 0.054)	−0.240 (−0.597, 0.102)	−0.235 (−0.593, 0.108)	0.139 (−0.114, 0.436)	0.135 (−0.138, 0.440)
DTIC + non-IFN ($d_{1,3}$)	−0.264 (−0.809, 0.240)	−0.270 (−0.788, 0.231)	−0.198 (−0.575, 0.153)	−0.192 (−0.588, 0.165)	0.269 (−0.036, 0.605)	0.264 (−0.039, 0.584)
Non-DTIC ($d_{1,4}$)	−0.185 (−0.716, 0.359)	−0.188 (−0.705, 0.345)	−0.287 (−0.702, 0.154)	−0.282 (−0.732, 0.148)	0.202 (−0.217, 0.617)	0.217 (−0.192, 0.627)
DTIC + IFN ($d_{2,2}$)	0.123 (−0.078, 0.324)	0.124 (−0.078, 0.327)	0.016 (−0.092, 0.127)	0.017 (−0.098, 0.128)	−0.140 (−0.349, 0.065)	−0.141 (−0.349, 0.068)
DTIC + non-IFN ($d_{2,3}$)	0.004 (−0.201, 0.218)	0.009 (−0.199, 0.232)	−0.003 (−0.124, 0.119)	−0.002 (−0.128, 0.128)	−0.064 (−0.296, 0.152)	−0.063 (−0.292, 0.166)
Non-DTIC ($d_{2,4}$)	−0.008 (−0.270, 0.258)	−0.004 (−0.275, 0.267)	0.017 (−0.164, 0.196)	0.016 (−0.161, 0.201)	0.075 (−0.206, 0.356)	0.070 (−0.215, 0.364)
FE model						
DTIC + IFN ($d_{1,2}$)	−0.412 (−0.888, 0.062)	−0.469 (−0.933, −0.004)	−0.171 (−0.421, 0.084)	−0.175 (−0.429, 0.076)	0.073 (−0.091, 0.230)	0.067 (−0.099, 0.235)
DTIC + non-IFN ($d_{1,3}$)	−0.235 (−0.659, 0.177)	−0.257 (−0.684, 0.172)	−0.172 (−0.409, 0.069)	−0.173 (−0.406, 0.063)	0.224 (0.055, 0.401)	0.223 (0.046, 0.398)
Non-DTIC ($d_{1,4}$)	−0.163 (−0.519, 0.198)	−0.158 (−0.525, 0.202)	−0.240 (−0.499, 0.013)	−0.244 (−0.500, 0.004)	0.212 (0.020, 0.403)	0.215 (0.015, 0.411)
DTIC + IFN ($d_{2,2}$)	0.129 (−0.003, 0.262)	0.146 (0.007, 0.281)	0.015 (−0.004, 0.034)	0.016 (−0.004, 0.035)	−0.134 (−0.265, −0.009)	−0.137 (−0.265, −0.007)
DTIC + non-IFN ($d_{2,3}$)	0.002 (−0.130, 0.133)	0.013 (−0.122, 0.144)	−0.001 (−0.017, 0.015)	−0.000 (−0.016, 0.015)	−0.049 (−0.174, 0.078)	−0.052 (−0.176, 0.077)
Non-DTIC ($d_{2,4}$)	−0.022 (−0.153, 0.111)	−0.021 (−0.156, 0.113)	0.009 (−0.017, 0.035)	0.010 (−0.017, 0.037)	0.071 (−0.060, 0.203)	0.070 (−0.058, 0.201)

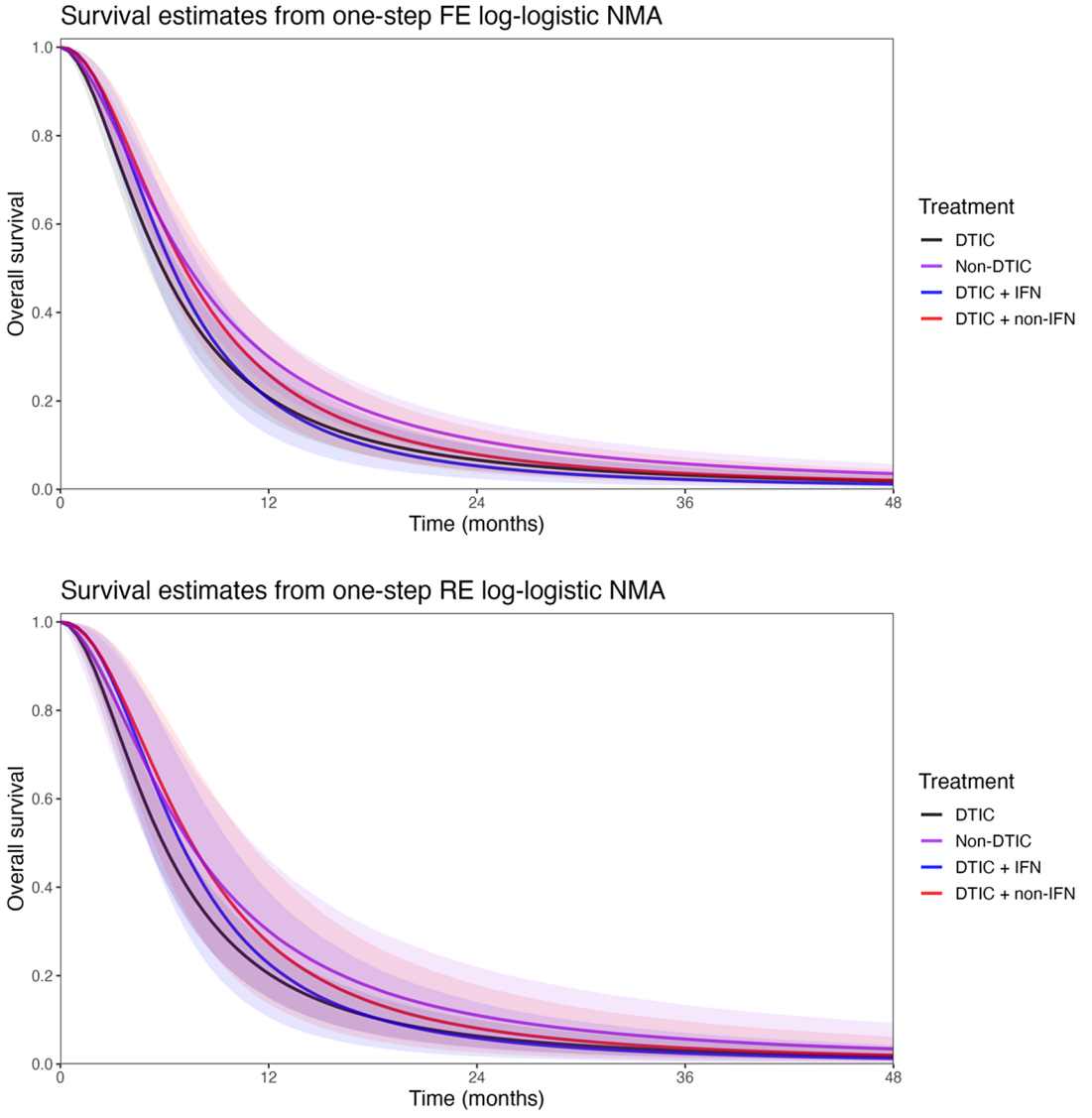


Figure 9. *Posterior estimates of overall survival (with shaded 95% CrIs) comparing DTIC, DTIC + IFN, DTIC + non-IFN, and Non-DTIC in the Avril 2004 population from the one-step NMA FE (top panel) and RE (bottom panel) models. Abbreviations: DTIC, dacarbazine; IFN, interferon (IFN).*

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

Table 7. *Estimates obtained from the one-step and two-step log-logistic models regarding overall survival at 1 year and at 2 years comparing DTIC + IFN and non-DTIC fit for a population with baseline risk of the Avril 2004 study population.*

	DTIC + IFN ($k = 2$)		Non-DTIC ($k = 4$)		DTIC + IFN versus non-DTIC	
	Two-step	One-step	Two-step	One-step	Two-step	One-step
FE log-logistic with Avril 2004 population						
Overall survival at 1 year (%)	20.49 [12.38, 30.23]	20.29 [12.16, 30.28]	29.96 [23.37, 36.61]	30.05 [23.82, 36.94]	−9.36 [−19.06, 1.45]	−9.65 [−19.79, 0.99]
Overall survival at 2 years (%)	5.29 [2.42, 10.05]	5.28 [2.36, 10.04]	11.15 [7.42, 15.54]	11.27 [7.73, 15.79]	−5.74 [−10.71, −0.51]	−5.90 [−10.87, −0.68]
RE log-logistic with Avril 2004 population						
Overall survival at 1 year (%)	23.19 [11.26, 39.65]	22.73 [10.61, 38.99]	30.10 [14.76, 46.91]	30.12 [15.17, 46.38]	−6.96 [−26.72, 14.22]	−7.16 [−26.56, 14.95]
Overall survival at 2 years (%)	5.93 [1.89, 13.93]	5.86 [1.77, 13.98]	11.02 [3.68, 21.91]	11.04 [3.59, 21.97]	−4.88 [−16.39, 5.25]	−4.97 [−16.28, 5.27]

5. Discussion

The proposed multivariate NMA models for TTE outcomes provide a one-step framework that allows for time-varying treatment effects, which leverage exact (or reconstructed) event and censor times for the studies in a connected network of evidence. By using the exact likelihood specification, we avoid the assumptions regarding within-study normality and variance required in the two-step method (Cope et al.[22]). This allows the entire model to be fit within a Bayesian framework, facilitating straightforward model selection and interpretation. Potential concerns regarding computational burden are mitigated by using Stan software, which provides a more efficient MCMC sampling than previous software (i.e., WinBUGS or JAGS) through the implementation of the Hamiltonian Monte Carlo algorithm (see Monnahan et al.[58]).

While we discussed using the LOOIC for model selection, often, valuable insight regarding the plausibility of different models may also be obtained from clinical experts[59–61] and observational studies.[62] Moreover, when there is no reason to believe that treatment effects vary over time, simpler models with fewer parameters may be more appropriate. Cope et al.[14] previously proposed a stepwise process exploring standard parametric models with treatment effect on scale alone, followed by models with multivariate treatment effects (scale and shape) for each trial and network (or other more flexible models), which builds upon the process outlined by Latimer et al.[59] for survival analysis of a single RCT in context of cost-effectiveness analysis. The proposed one-step models allow this process to be applied to all the "standard" distributions, now including gamma and generalized gamma. Further, the generalized gamma distribution may be particularly useful to improve fit over other standard distributions to a network of RCTs given the flexibility to model U-shaped hazards with this distribution[63] as well as the nature of its nested distributions (i.e., exponential, Weibull, log-normal, and gamma).

The simulation study in Section 3 found that the proposed one-step method and the previously proposed two-step method provide broadly similar results. When study sample sizes are especially small, the one-step method may be preferred, in line with previous recommendations.[29] However, due to finite computational resources, the simulation study was limited in that we were unable to consider fitting RE models and did not consider a wide range of sample sizes and prior specifications. More extensive simulation studies may be helpful to better understand the impact of model misspecification, the differences between fitting FE and RE models, and the differences between using one-step and two-step methods.[64]

The illustrative melanoma example highlighted an application to a network of multiple trials, where we demonstrated the feasibility of accounting for between-study heterogeneity in terms of both the shape and the scale parameters. However, simpler random-effect models specified for only one of the two treatment effect parameters may often be sufficient. We advise consulting clinicians to consider whether between-study heterogeneity is most likely to affect the scale versus the shape of the distributions. Future research should consider the merits of using different priors and parameterizations for the variance–covariance matrix,[65,66] potentially based on incorporating external evidence on the between-trial heterogeneity.[67]

It may be of interest to extend the proposed methods to nonparametric IPD-NMA models. While potentially flexible, nonparametric models can be considerably more complex and may not be among the first NMA models to consider when following recommended model selection procedures in the broader cost-effectiveness framework.[59] Nonetheless, there may be utility in extending the current framework to include additional nonparametric methods.

During the development of this publication, Phillippo et al. proposed the extension the ML-NMR framework for TTE data, considering networks in which IPD-level covariate data is only available for a subset of trials within the network (with aggregate-level covariate data available for the remaining studies).[68] This allows for the adjustment of prognostic factors and effect modifiers when information on these is available. Maciel et al.[69] demonstrated the feasibility of applying ML-NMR for TTE outcomes with a univariate treatment effect. The main advantage of NMA models with a multivariate

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

treatment effect as proposed is that they do not rely on a PH assumption across studies and treatment comparisons. This flexibility may be important when there are differences in the survival distributions for the treatments compared, and relative treatment effects need to be extrapolated beyond the available trial data for a cost-effectiveness analysis.

Author contributions. Concept and design: H.C., D.M., K.C., J.J., S.K., K.T., B.M., and S.C.; Statistical analysis: D.M. and H.C.; Analysis and interpretation of results: H.C., D.M., K.C., J.J., S.K., K.T., B.M., and S.C.; Drafting of the manuscript: H.C. and S.C.; Critical revision of the paper for important intellectual content: H.C., D.M., K.C., J.J., S.K., K.T., B.M., and S.C.; Obtaining funding: S.K. and S.C.; Administrative, technical, or logistic support: K.T.; Supervision: S.C.

Competing interest statement. S.K. and B.M. are employees of Bristol Myers Squibb. H.C., D.M., K.C., J.J., K.T., and S.C. are employees of Precision AQ, a consulting firm that received funding from Bristol Myers Squibb for this study.

Data availability statement. We provide Stan code for both the two-step and one-step FE and RE models in the Supplementary Material. Also in the Supplementary Material, R code and data are provided to fit the one-step NMA models to the illustrative melanoma example from Section 4.

Funding statement. Funding for this study was provided by Bristol Myers Squibb.

Supplementary material. To view supplementary material for this article, please visit http://doi.org/10.1017/rsm.2025.21.

References

[1] Ades AE. A chain of evidence with mixed comparisons: models for multi-parameter synthesis and consistency of evidence. *Stat Med.* 2003;22(19): 2995–3016. https://doi.org/10.1002/sim.1566
[2] Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. 2005;331(7521): 897–900. https://doi.org/10.1136/bmj.331.7521.897
[3] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20): 3105–3124. https://doi.org/10.1002/sim.1875
[4] Ades AE, Welton NJ, Dias S, Phillippo DM, Caldwell DM. Twenty years of network meta-analysis: continuing controversies and recent developments. *Res Synth Methods*. 2024;15(5): 702–727.
[5] Fay MP, Li F. Causal interpretation of the hazard ratio in randomized clinical trials. *Clin Trials*. 2024;21(5): 623–635.
[6] Williamson PR, Smith CT, Hutton JL, Marson AG. Aggregate data meta-analysis with time-to-event outcomes. *Stat Med*. 2002;21(22): 3337–3351. https://doi.org/10.1002/sim.1303
[7] Batson S, Greenall G, Hudson P. Review of the reporting of survival analyses within randomised controlled trials and the implications for meta-analysis. *PloS One*. 2016;11(5): e0154870. https://doi.org/10.1371/journal.pone.0154870
[8] Li H, Han D, Hou Y, Chen H, Chen Z. Statistical inference methods for two crossing survival curves: a comparison of methods. *PLOS ONE.* 2015;10(1): e0116774. https://doi.org/10.1371/journal.pone.0116774
[9] Schadendorf D, Hodi FS, Robert C, et al. Pooled analysis of long-term survival data from Phase II and Phase III trials of ipilimumab in unresectable or metastatic melanoma. *J Clin Oncol*. 2015;33(17): 1889–1894. https://doi.org/10.1200/jco.2014.56.2736
[10] Royston P, Parmar M. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013;13: 152. https://doi.org/10.1186/1471-2288-13-152
[11] Magirr D. Non-proportional hazards in immuno-oncology: is an old perspective needed? *Pharm Stat*. 2021;20(3): 512–527.
[12] Ristl R, Ballarini NM, Götte H, Schüler A, Posch M, König F. Delayed treatment effects, treatment switching and heterogeneous patient populations: how to design and analyze RCTs in oncology. *Pharm Stat*. 2021;20(1): 129–145.
[13] Rahman R, Fell G, Ventz S, et al. Deviation from the proportional hazards assumption in randomized phase 3 clinical trials in oncology: prevalence, associated factors, and implications. *Clin Cancer Res*. 2019;25(21): 6339–6345.
[14] Cope S, Chan K, Campbell H, et al. A comparison of alternative network meta-analysis methods in the presence of nonproportional hazards: a case study in first-line advanced or metastatic renal cell carcinoma. *Value Health*. 2022;26(4): 465–476. https://doi.org/10.1016/j.jval.2022.11.017
[15] Crowther MJ, Riley RD, Staessen JA, Wang J, Gueyffier F, Lambert PC. Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Med Res Methodol*. 2012;12: 34. https://doi.org/10.1186/1471-2288-12-34
[16] Freeman S, Carpenter J. Bayesian one-step IPD network meta-analysis of time-to-event data using royston-parmar models. *Res Synth Methods*. 2017;8(4): 451–464. https://doi.org/10.1002/jrsm.1253

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

[17] Petit C, Blanchard P, Pignon J, Lueza B. Individual patient data network meta-analysis using either restricted mean survival time difference or hazard ratios: is there a difference? a case study on locoregionally advanced nasopharyngeal carcinomas. *Syst Rev*. 2019;8(1): 96. https://doi.org/10.1186/s13643-019-0984-x
[18] Ouwens MJ, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. *Res Synth Methods*. 2010;1(3–4): 258–271.
[19] Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Med Res Methodol*. 2011;11: 61. https://doi.org/10.1186/1471-2288-11-61
[20] 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*. 2012;12: 9. https://doi.org/10.1186/1471-2288-12-9
[21] Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Res Synth Methods*. 2012;3(2): 80–97.
[22] Cope S, Chan K, Jansen J. Multivariate network meta-analysis of survival function parameters. *Res Synth Methods*. 2020;11(3): 443–456. https://doi.org/10.1002/jrsm.1405
[23] Jackson CH. flexsurv: a platform for parametric survival modeling in R. *J Stat Softw*. 2016; 70: i80
[24] Achana FA, Cooper NJ, Bujkiewicz S, et al. Network meta-analysis of multiple outcome measures accounting for borrowing of information across outcomes. *BMC Med Res Methodol*. 2014;14: 92. https://doi.org/10.1186/1471-2288-14-92
[25] Burke DL, Ensor J, Riley RD. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat Med*. 2017;36(5): 855–875.
[26] Debray TPA, Moons KGM, Abo-Zaid GMA, Koffijberg H, Riley RD. Individual participant data meta-analysis for a binary outcome: one-stage or two-stage? *PloS One*. 2013;8(4): e60650.
[27] Freeman SC, Cooper NJ, Sutton AJ, Crowther MJ, Carpenter JR, Hawkins N. Challenges of modelling approaches for network meta-analysis of time-to-event outcomes in the presence of non-proportional hazards to aid decision making: application to a melanoma network. *Stat Methods Med Res*. 2022;31(5): 839–861. https://doi.org/10.1177/09622802211070253
[28] Kanters S, Karim ME, Thorlund K, Anis A, Bansback N. When does the use of individual patient data in network meta-analysis make a difference? A simulation study. *BMC Med Res Methodol*. 2021;21(1): 21. https://doi.org/10.1186/s12874-020-01198-2
[29] Jackson D, White IR. When should meta-analysis avoid making hidden normality assumptions? *Biom J*. 2018;60(6): 1040–1058.
[30] Saramago P, Chuang LH, Soares MO. Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data. *BMC Med Res Methodol*. 2014;14: 105. https://doi.org/10.1186/1471-2288-14-105
[31] Phillippo DM. *Calibration of Treatment Effects in Network Meta-Analysis Using Individual Patient Data*. University of Bristol; 2019. https://research-information.bris.ac.uk/en/studentTheses/calibration-of-treatment-effects-in-network-meta-analysis-using-i
[32] White IR, Turner RM, Karahalios A, Salanti G. A comparison of arm-based and contrast-based models for network meta-analysis. *Stat Med*. 2019;38(27): 5197–5213.
[33] Wei Y, Higgins JPT. Estimating within-study covariances in multivariate meta-analysis with multiple outcomes. *Stat Med*. 2013;32(7): 1191–1205.
[34] Brilleman SL, Wolfe R, Moreno-Betancur M, Crowther MJ. Simulating survival data using the simsurv R package. *J Stat Softw*. 2021;97: 1–27.
[35] Phillippo DM. *multinma: Bayesian Network Meta-Analysis of Individual and Aggregate Data.* R package version 0.8.0. 2025; https://doi.org/10.5281/zenodo.3904454.
[36] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2018.
[37] Stan Development Team. Stan Modeling Language Users Guide and Reference Manual Version 2.27. 2019; https://mc-stan.org
[38] Jansen JP, Vieira MC, Cope S. Network meta-analysis of longitudinal data using fractional polynomials. *Stat Med*. 2015;34(15): 2294–2311. https://doi.org/10.1002/sim.6492
[39] Boutros A, Croce E, Ferrari M, et al. The treatment of advanced melanoma: current approaches and new challenges. *Crit Rev Oncol Hematol*. 2024: 104276.
[40] Saluja R, Cheng S, delos Santos KA, Chan KKW. Estimating hazard ratios from published Kaplan-Meier survival curves: a methods validation study. *Res Synth Methods*. 2019;10(3): 465–475.
[41] Avril MF, Aamdal S, Grob JJ, et al. Fotemustine compared with dacarbazine in patients with disseminated malignant melanoma: a phase III study. *J Clin Oncol*. 2004;22(6): 1118–1125.
[42] Bajetta E, Di Leo A, Zampino MG, et al. Multicenter randomized trial of dacarbazine alone or in combination with two different doses and schedules of interferon alfa-2a in the treatment of advanced melanoma. *J Clin Oncol*. 1994;12(4): 806–811.
[43] Chapman PB, Einhorn LH, Meyers ML, et al. Phase III multicenter randomized trial of the Dartmouth regimen versus dacarbazine in patients with metastatic melanoma. *J Clin Oncol*. 1999;17(9): 2745–2751.
[44] Chiarion Sileni V, Nortilli R, Aversa SM, et al. Phase II randomized study of dacarbazine, carmustine, cisplatin and tamoxifen versus dacarbazine alone in advanced melanoma patients. *Melanoma Res*. 2001;11(2): 189–196.
[45] Cocconi G, Bella M, Calabresi F, et al. Treatment of metastatic malignant melanoma with dacarbazine plus tamoxifen. *N Engl J Med*. 1992;327(8): 516–523.

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press

[46] Falkson CI, Falkson G, Falkson HC. Improved results with the addition of interferon alfa-2b to dacarbazine in the treatment of patients with metastatic malignant melanoma. *J Clin Oncol*. 1991;9(8): 1403–1408.
[47] Falkson CI, Ibrahim J, Kirkwood JM, Coates AS, Atkins MB, Blum RH. Phase III trial of dacarbazine versus dacarbazine with interferon alpha-2b versus dacarbazine with tamoxifen versus dacarbazine with interferon alpha-2b and tamoxifen in patients with metastatic malignant melanoma: an eastern cooperative oncology group study. *J Clin Oncol*. 1998;16(5): 1743–1751.
[48] Middleton MR, Grob JJ, Aaronson N, et al. Randomized phase III study of temozolomide versus dacarbazine in the treatment of patients with advanced metastatic malignant melanoma. *J Clin Oncol*. 2000;18(1): 158–166.
[49] Thomson DB, Adena M, McLeod GR, et al. Interferon-alpha 2a does not improve response or survival when combined with dacarbazine in metastatic malignant melanoma: results of a multi-institutional Australian randomized trial. *Melanoma Res*. 1993;3(2): 133–138.
[50] Young AM, Marsden J, Goodman A, Burton A, Dunn JA. Prospective randomized comparison of dacarbazine (DTIC) versus DTIC plus interferon-alpha (IFN-alpha) in metastatic melanoma. *Clin Oncol*. 2001;13(6): 458–465.
[51] Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*. 1994;81(3): 515–526.
[52] Assessment HC-Mscgoht. *Practical Guideline for Quantitative Evidence Synthesis: Direct and Indirect Comparisons*. 2024. https://health.ec.europa.eu/document/download/1f6b8a70-5ce0-404e-9066-120dc9a8df75_en?filename=hta_practical-guideline_direct-and-indirect-comparisons_en.pdf
[53] Vehtari A, Gelman A, Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat Comput*. 2017;27: 1413–1432.
[54] Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. *Statist Sci*. 1992;7: 457–511.
[55] Gelman A. *Bayesian Data Analysis*, 3rd ed. Boca Raton : Texts in Statistical Science; 2013.
[56] Betancourt M, Girolami M. Hamiltonian Monte Carlo for hierarchical models. *Curr Trends Bayesian Methodol Appl*. 2015;79(30): 2–4.
[57] Cope S, Jansen JP. Quantitative summaries of treatment effect estimates obtained with network meta-analysis of survival curves to inform decision-making. *BMC Med Res Methodol*. 2013;13: 147. https://doi.org/10.1186/1471-2288-13-147
[58] Monnahan CC, Thorson JT, Branch TA. Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo. *Methods Ecol Evol*. 2017;8(3): 339–348.
[59] Latimer NR. Survival analysis for economic evaluations alongside clinical trials–extrapolation with patient-level data: inconsistencies, limitations, and a practical guide. *Med Decis Making*. 2013;33(6): 743–754. https://doi.org/10.1177/0272989x12472398
[60] Rutherford M, Lambert P, Sweeting M, Pennington R, Crowther M, Abrams K. *NICE DSU Technical Support Document 21: Flexible Methods for Survival Analysis*. 2020. http://nicedsu.org.uk/
[61] Latimer N. *NICE DSU Technical Support Document 14: Survival analysis for Economic Evaluations Alongside Clinical Trials-Extrapolation with Patient-Level Data*. 2011. http://nicedsu.org.uk/wp-content/uploads/2016/03/NICE-DSU-TSD-Survival-analysis.updated-March-2013.v2.pdf
[62] Ruiz-Morales JM, Swierkowski M, Wells JC, et al. First-line sunitinib versus pazopanib in metastatic renal cell carcinoma: results from the international metastatic renal cell carcinoma database consortium. *Eur J Cancer*. 2016;65: 102– 108. https://doi.org/10.1016/j.ejca.2016.06.016
[63] Cox C, Chu H, Schneider MF, Munoz A. Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Stat Med*. 2007;26(23): 4352–4374.
[64] Heinze G, Boulesteix AL, Kammer M, Morris TP, White IR. Simulation Panel of the SI. Phases of methodological research in biostatistics—building the evidence base for new methods. *Biom J*. 2024;66(1): e2200222.
[65] Wang Z, Lin L, Hodges JS, Chu H. The impact of covariance priors on arm-based Bayesian network meta-analyses with binary outcomes. *Stat Med*. 2020;39(22): 2883–2900.
[66] Alvarez I, Niemi J, Simpson M. Bayesian inference for a covariance matrix. *arXiv preprint arXiv:14084050*. 2014.
[67] Turner RM, Domínguez-Islas CP, Jackson D, Rhodes KM, White IR. Incorporating external evidence on between-trial heterogeneity in network meta-analysis. *Stat Med*. 2019;38(8): 1321–1335.
[68] Phillippo DM, Dias M, Ades AE, Welton NJ. Multilevel network meta-regression for general likelihoods: synthesis of individual and aggregate data with applications to survival analysis. *arXiv preprint arXiv:2401.12640*. 2024.
[69] Maciel D, Cope S, Towle K, et al. MSR31 applying multi-level network meta-regression (ML-NMR) to a case study in triple-class exposed (TCE) relapsed/refractory multiple myeloma (RRMM). *Value Health*. 2023;26(6): S283.

Cite this article: Campbell H, Maciel D, Chan K, Jansen JP, Klijn S, Towle K, Malcolm B, Cope S. One-step parametric network meta-analysis models using the exact likelihood that allow for time-varying treatment effects. *Research Synthesis Methods*. 2025;16: 650–671. https://doi.org/10.1017/rsm.2025.21

https://doi.org/10.1017/rsm.2025.21 Published online by Cambridge University Press