



Daly, C. H., Maconachie, R., Ades, A. E., & Welton, N. J. (2021). A non-parametric approach for jointly combining evidence on progression free and overall survival time in network meta-analysis. *Research Synthesis Methods*. <https://doi.org/10.1002/jrsm.1539>

Publisher's PDF, also known as Version of record

License (if available):  
CC BY

Link to published version (if available):  
[10.1002/jrsm.1539](https://doi.org/10.1002/jrsm.1539)

[Link to publication record in Explore Bristol Research](#)  
PDF-document




This is the final published version of the article (version of record). It first appeared online via Wiley at [10.1002/jrsm.1539](https://doi.org/10.1002/jrsm.1539). Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available: <http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# A non-parametric approach for jointly combining evidence on progression free and overall survival time in network meta-analysis

Caitlin H. Daly<sup>1</sup>  | Ross Maconachie<sup>2</sup> | A. E. Ades<sup>1</sup>  | Nicky J. Welton<sup>1</sup> 

<sup>1</sup>Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

<sup>2</sup>Value, Access & Devolved Nations, Merck Sharp and Dohme (UK) Ltd, London, UK

## Correspondence

Nicky J. Welton, Population Health Sciences, Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol BS8 2PS, UK.  
Email: nicky.welton@bristol.ac.uk

## Funding information

National Institute of Health and Care Excellence, Grant/Award Number: N/A; Medical Research Council, Grant/Award Number: MR/R025223/1

## Abstract

Randomised controlled trials of cancer treatments typically report progression free survival (PFS) and overall survival (OS) outcomes. Existing methods to synthesise evidence on PFS and OS either rely on the proportional hazards assumption or make parametric assumptions which may not capture the diverse survival curve shapes across studies and treatments. Furthermore, PFS and OS are not independent; OS is the sum of PFS and post-progression survival (PPS). Our aim was to develop a non-parametric approach for jointly synthesising evidence from published Kaplan–Meier survival curves of PFS and OS without assuming proportional hazards. Restricted mean survival times (RMST) are estimated by the area under the survival curves (AUCs) up to a restricted follow-up time. The correlation between AUCs due to the constraint that  $OS > PFS$  is estimated using bootstrap re-sampling. Network meta-analysis models are given for RMST for PFS and PPS and ensure that  $OS = PFS + PPS$ . Both additive and multiplicative network meta-analysis models are presented to obtain relative treatment effects as either differences or ratios of RMST. The methods are illustrated with a network meta-analysis of treatments for stage IIIA–N2 non-small cell lung cancer. The approach has implications for health economic models of cancer treatments, which require estimates of the mean time spent in the PFS and PPS health-states. The methods can be applied to a single time-to-event outcome, and so have wide applicability in any field where time-to-event outcomes are reported, the proportional hazards assumption is in doubt, and survival curve shapes differ across studies and interventions.

## KEYWORDS

network meta-analysis, oncology, restricted mean survival time, survival analysis, time-to-event outcomes

## Highlights

### What is already known

Network meta-analyses of cancer therapies typically pool hazard ratios or parameters of parametric survival distributions for progression-free survival

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *Research Synthesis Methods* published by John Wiley & Sons Ltd.

(PFS) and overall survival (OS) separately. These methods either assume proportional hazards or that survival curves come from the same family across studies and treatments, and do not account for the structural relationship between PFS and OS.

#### **What is new**

We have presented a network meta-analysis model to estimate differences or ratios of the restricted mean survival time for PFS and post progression survival (PPS) based on joint estimates of areas under the curve for PFS and OS. The approach does not assume proportional hazards or any parametric form for the survival curves.

#### **Potential impact for research synthesis methods readers outside authors field**

The methods can be applied to a single time-to-event outcome, so have wide applicability in any field where time-to-event outcomes are reported, the proportional hazards assumption is violated or questionable, and survival curve shapes differ across studies or treatments.

## **1 | INTRODUCTION**

Evidence synthesis within a decision-making framework aims to pool and compare all relevant evidence on the efficacy of treatments within a decision set. If there are more than two treatments under consideration, then network meta-analysis (NMA) can be used to synthesise results whilst preserving the randomisation in the included randomised controlled trials (RCTs).<sup>1,2</sup> In the context of decision-making for cancer therapies, the objective is to identify well-tolerated treatments that extend the time until disease progresses (worsens) and to extend life. RCTs of cancer therapies therefore typically report results for progression free survival (PFS) and overall survival (OS), available as published Kaplan–Meier survival curves and/or hazard ratios (HRs). However, decision models comparing cost-effectiveness of different treatments rely on estimates of the difference in expected (or mean) survival times<sup>3</sup> often obtained by extrapolating parametric survival curves.<sup>4</sup>

The most common approach to the synthesis of survival outcomes is to pool HRs,<sup>5,6</sup> and to analyse PFS and OS separately. Whilst pooling HRs obtained from proportional hazards models has the advantage that it does not make any assumptions about the parametric form of the survival curves in the studies,<sup>7</sup> it does make the strong assumption that the HR is constant over time within each study and treatment comparison (the proportional hazards assumption). If proportional hazards do not hold across all studies and treatment comparisons, then pooling HRs can lead to misleading results, and can be particularly problematic when the results of the synthesis are used in extrapolations to predict long-term survival.<sup>4,8–10</sup> Furthermore, if

HRs are not constant over time, then they will be confounded with study follow-up time, which will introduce heterogeneity in pairwise meta-analysis and may introduce inconsistency in network meta-analysis.

The proportional hazards assumption can be assessed using log-cumulative hazard plots,<sup>4,7,11</sup> visual inspection of the survival curves,<sup>4,11</sup> or by comparing model fit with a model that does not assume proportional hazards.<sup>7</sup> If proportional hazards are not supported for some studies or treatment comparisons, then alternative methods are required. If a particular parametric survival model is appropriate across all studies and treatments, then one or more of the parameters from that model can be pooled. This approach includes accelerated failure time models where it is assumed that the effect of treatment is to accelerate or decelerate time to event by a proportional constant (the acceleration factor [AF]),<sup>7</sup> and synthesis of AFs have been proposed for pairwise meta-analysis.<sup>12,13</sup> Further flexibility can be captured by modelling more than one parameter of a survival distribution. For example, a bivariate meta-analysis model can be put on the shape and scale parameters of a Weibull model, where shape and scale may depend on treatment and study.<sup>14,15</sup> This approach can be generalised to other parametric distributions<sup>16</sup> and piecewise exponential models, where exponential distributions with different rates are assumed for different segments of the curves.<sup>17</sup> More flexible parametric models can also be used such as fractional polynomial models,<sup>8</sup> which include many of the standard parametric models as a special case, and restricted cubic spline models.<sup>18</sup>

A drawback of fitting parametric curves is that it constrains the survival curves to have the same shape across studies and arms, characterised by a single parametric

model. Although it is considered good practise to assume the same parametric curve across arms within a single study,<sup>4</sup> finding a single parametric model that fits all the data well becomes less likely as more treatments (that may have differing mechanisms of action) and studies are included in the NMA. Whilst parametric models could potentially incorporate different shapes across studies, and more flexible models can be fitted (e.g. fractional polynomial and spline models), these models are complex to fit and in practise there is usually insufficient data to fit them. Furthermore, model selection criteria (such as the AIC and BIC) give more weight to the fit of the early part of the curves, and less weight to the tails of the curves.<sup>11</sup> This means that selected models may not characterise well the shape of the later part of the curves, which can have a substantial impact on predicted long-term survival.<sup>10</sup> Increasingly oncology products are targeted at subgroups of patients that may have a long-lasting response to treatment, meaning that these treatments may have a very different shaped survival curve tails than previous treatment options.<sup>19</sup>

To overcome these limitations, the restricted mean survival time (RMST), calculated as the area under the Kaplan–Meier curve up to a specified follow-up time period, has been proposed as a non-parametric summary from RCTs reporting survival outcomes,<sup>20,21</sup> and Wei et al. have proposed its use for pairwise meta-analysis with a single survival outcome.<sup>22</sup> The RMST is an attractive summary for the synthesis of survival outcomes because it provides a single intuitive summary measure. Differences in life expectancy over a period of time are easily interpreted and meaningful for patients, and directly correspond to the parameters used in health economic models to assess cost-effectiveness of cancer therapies.<sup>23</sup> The clinical validity of treatment effects acting on differences in mean time to event over a fixed period is a priori no less reasonable that assuming treatments act on the shape and scale parameters of parametric distributions and may be easier for clinical experts to interpret and assess.

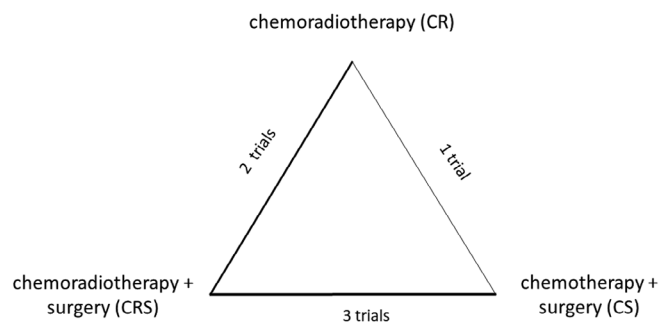
Analysing OS and PFS separately is problematic because these outcomes are not independent; OS is the sum of PFS and post progression survival (PPS),<sup>23</sup> and at the very least must conform to the natural constraint that OS is greater than PFS. It should also be noted that making separate parametric assumptions for PFS and OS implies a parametric assumption for PPS due to the constraint that  $OS = PFS + PPS$ , and the implied parametric form for PPS may not be clinically plausible if the OS and PFS models are fitted independently. Bivariate methods have been proposed to jointly synthesise PFS and OS,<sup>24</sup> however, these do not take into consideration the structural relationship between the outcomes, and rely on either individual patient data being available or for correlations between the summary measures of interest

(e.g. HRs or parameters of parametric models) to be available, which is unusual in practice. We propose instead putting the synthesis model on PFS and PPS which are not structurally related (although may still be correlated).

Our objective is to extend the RMST approach of Wei et al.<sup>22</sup> to jointly synthesise relative treatment effects from PFS and OS Kaplan–Meier curves in a network meta-analysis, providing a method that does not assume proportional hazards nor make parametric assumptions, and conforms to the constraint that  $OS = PFS + PPS$ . We motivate and illustrate the approach using a NMA from the National Institute for Health and Care Excellence (NICE) guideline on non-small cell lung cancer (NSCLC).<sup>25</sup> We begin by describing the NSCLC example and show how to obtain the area under the curve (AUC) summaries from published Kaplan–Meier curves, accounting for constraints between the OS and PFS outcomes. We then describe the evidence synthesis model to obtain estimates of treatment differences or ratios of RMST for PFS and PPS. We present the results from the NSCLC example, and end with a discussion of further adaptations and evidence sources that can be used to extrapolate RMST beyond the restricted follow-up time.

## 2 | MOTIVATING EXAMPLE: NON-SMALL CELL LUNG CANCER

The methods proposed in this paper were motivated by a network of six trials evaluating three treatments for stage IIIA-N2 non-small cell lung cancer (Figure 1).<sup>25</sup> All trials provided Kaplan–Meier curves for PFS and overall survival (OS). Visual inspection of the Kaplan–Meier curves revealed that the proportional hazards assumption did not appear to hold, as the survival curves crossed at least once in every study (Appendix A.1). Consequently, traditional pooling of hazards ratios was not considered appropriate. In addition, the shapes of the survival curves were different across studies, suggesting that it was not



**FIGURE 1** Network diagram of comparisons for which direct evidence on differences in restricted mean survival time up to 5 years is available. Lines are proportional to the number of studies that compare the two connected treatments

appropriate to synthesise the evidence under an assumption of a single parametric model. A non-parametric approach to evidence synthesis was therefore required.

## 2.1 | Data extraction

Data were extracted from the Kaplan–Meier curves using a validated algorithm that makes use of the digitised curves as well as data on the numbers at risk and total number of events.<sup>26</sup> For each treatment group within each study, this produces a set of pseudo individual patient data (survival times and censor times) that produce Kaplan–Meier curves similar to those published. This was done for both the PFS and OS curves for each study. Generally, the reconstructed data provided similar hazard ratios and median survival times to those reported in each study (Appendix A.2) although the hazard ratio differs a little for PFS in three studies.<sup>27–29</sup>

## 2.2 | Calculating area under the Kaplan–Meier curves

Kaplan–Meier curves were fitted to the extracted data using the `survfit` function from the survival package in R (v. 3.4.2).<sup>30,31</sup> This package calculates the area under the Kaplan–Meier curves from randomisation  $t_0 = 0$  to a specified truncated follow up time  $t_T$  as a Reimann sum:

$$\text{AUC}_{\text{KM}}(t_T) = \sum_{i=1}^N (t_i - t_{i-1}) \widehat{S}_{\text{KM}}(t_{i-1}),$$

where

$$N = \begin{cases} (\text{number of distinct event times between } t_0 \text{ and } t_T) & \text{if an event occurs at } t_T \\ (\text{number of distinct event times between } t_0 \text{ and } t_T) + 1 & \text{otherwise,} \end{cases}$$

$t_i$  are the ordered event times, and  $\widehat{S}_{\text{KM}}(t_{i-1})$  is the Kaplan–Meier estimate of the probability of survival at time  $t_{i-1}$ . The variance of the AUC is estimated as,<sup>32</sup>

$$\widehat{V}(\text{AUC}_{\text{KM}}(t_T)) = \sum_{i=1}^{N-1} \frac{d_{(i)}}{n_{(i)}(n_{(i)} - d_{(i)})} \left( \sum_{j=i}^{N-1} (t_{j+1} - t_j) \widehat{S}_{\text{KM}}(t_{j+1}) \right)^2,$$

where  $d_{(i)}$  is the number of patients who experienced an event at time  $t_i$  and  $n_{(i)}$  is the number of people at risk at time  $t_i$ .

For the lung-cancer example, we considered a truncated follow-up time of  $t_T = 5$  years as being sufficiently long enough to capture the important differences between treatments on PFS and OS, while allowing the inclusion of most of the studies (five of the six included trials reported Kaplan–Meier curves up until at least this period). This also coincided with an assumption made by the NICE guideline committee, that patients who are alive and progression-free at 5 years will remain progression-free until death.<sup>25</sup> The extracted areas under the Kaplan–Meier curves up to 5 years for each of the five trials with 5-year follow-up are presented in Table 1. We outline how the methods could be extended to include the study with less than 5 year follow-up in the discussion section.

## 2.3 | Calculating the structurally-induced correlations between the area under the survival curves of progression-free survival and overall survival

The area under the survival curves (AUCs) for PFS and OS must meet the constraint that OS is greater than PFS at each time-point, which induces correlations between the AUC for PFS and the AUC for OS. We estimated this correlation using non-parametric bootstrapping, where 5000 pairs of PFS and OS datasets were resampled with replacement from the reconstructed data<sup>35</sup> (R code given in Appendix A3). If there were any pairs of PFS and OS datasets where the OS Kaplan–Meier curve fell below the PFS Kaplan–Meier curve, then the paired sample was discarded, and the bootstrapping process continued until 5000 bootstrap samples were obtained. For each sampled PFS and OS dataset, the AUCs were computed and the correlation between the bootstrapped PFS and OS AUCs are provided in Table 1.

## 3 | METHODS

### 3.1 | Joint network meta-analysis model for restricted mean survival time for progression-free survival and overall survival

Let  $y_{i,k}^{\text{PFS}}$  and  $y_{i,k}^{\text{OS}}$  be the estimated AUCs up to  $t_T$  years for study  $i$ , arm  $k$  for PFS and OS respectively, with covariance matrix  $V_{i,k}$  for the PFS and OS AUC outcomes at  $t_T$ . We assume the AUCs follows a bivariate normal likelihood:

TABLE 1 Extracted data from the reconstructed Kaplan–Meier PFS and OS curves from each trial reporting 5-year follow-up data

Study	Treatment	PFS		OS		AUC <sub>KM(5)</sub> correlation
		AUC <sub>KM(5)</sub>	SE	AUC <sub>KM(5)</sub>	SE	
Albain 2009 <sup>27</sup>	CR	1.55	0.11	2.33	0.15	0.09
	CRS	1.95	0.13	2.42	0.15	0.27
Eberhardt 2015 <sup>33</sup>	CR	2.41	0.23	3.09	0.21	0.25
	CRS	2.49	0.22	3.30	0.21	0.22
Katakami 2011 <sup>28</sup>	CS	1.60	0.28	2.88	0.30	0.17
	CRS	2.15	0.35	3.19	0.30	0.31
Pless 2015 <sup>29</sup>	CS	1.86	0.18	2.90	0.19	0.03
	CRS	2.13	0.19	2.94	0.18	0.13
van Meerbeeck 2007 <sup>34</sup>	CR	1.52	0.12	2.11	0.12	0.23
	CS	1.48	0.12	1.96	0.13	0.26

Note: AUC restricted to 5-year follow-up for PFS and OS are reported, with standard errors and correlations between PFS and OS AUCs.

Abbreviations: AUC, area under the (Kaplan–Meier) curve; CR, chemoradiotherapy; CRS, chemoradiotherapy + surgery; CS, chemotherapy + surgery; OS, overall survival; PFS, progression-free survival; SE, standard error.

$$\begin{pmatrix} y_{i,k}^{\text{PFS}} \\ y_{i,k}^{\text{OS}} \end{pmatrix} \sim \text{Normal} \left( \begin{pmatrix} \theta_{i,k}^{\text{PFS}} \\ \theta_{i,k}^{\text{OS}} \end{pmatrix}, V_{i,k} \right),$$

where  $\theta_{i,k}^{\text{PFS}}$  and  $\theta_{i,k}^{\text{OS}}$  are the RMST up to  $t_T$  years for study  $i$ , arm  $k$ , for PFS and OS respectively. AUCs are a form of average survival, and by the central limit theorem are expected to be asymptotically normally distributed, hence the choice of likelihood.

For OS, the RMST is defined as the sum of the RMST for PFS and post-progression survival (PPS):

$$\theta_{i,k}^{\text{OS}} = \theta_{i,k}^{\text{PFS}} + \theta_{i,k}^{\text{PPS}}. \quad (1)$$

Treatment effects are given for RMST for PFS and PPS and may be considered as additive or multiplicative. For the additive treatment effect model, the NMA model<sup>6</sup> for PFS is:

$$\theta_{i,k}^{\text{PFS}} = \mu_i^{\text{PFS}} + \delta_{i,k}^{\text{PFS}}, \quad (2)$$

where  $\mu_i^{\text{PFS}}$  is the baseline RMST for PFS in study  $i$ , and  $\delta_{i,k}^{\text{PFS}}$  is the difference in RMST for treatment in arm  $k$  relative to the treatment in arm 1 in study  $i$ , which may be modelled as either a fixed or random effect:

$$\begin{aligned} \delta_{i,k}^{\text{PFS}} &= d_{t_{i,k}}^{\text{PFS}} - d_{t_{i,1}}^{\text{PFS}} && \text{fixed effect model} \\ \delta_{i,k}^{\text{PFS}} &\sim N(d_{t_{i,k}}^{\text{PFS}} - d_{t_{i,1}}^{\text{PFS}}, \sigma_{\text{PFS}}^2) && \text{random effects model,} \end{aligned} \quad (3)$$

where  $d_k^{\text{PFS}}$  is the difference in RMST for treatment  $k$  relative to treatment 1 ( $d_1^{\text{PFS}} = 0$ ) and  $\sigma_{\text{PFS}}$  is the

between-study standard deviation in treatment differences in RMST.

Similarly, the additive NMA model for PPS is given by:

$$\theta_{i,k}^{\text{PPS}} = \mu_i^{\text{PPS}} + \delta_{i,k}^{\text{PPS}}, \quad (4)$$

and

$$\begin{aligned} \delta_{i,k}^{\text{PPS}} &= d_{t_{i,k}}^{\text{PPS}} - d_{t_{i,1}}^{\text{PPS}} && \text{fixed effect model} \\ \delta_{i,k}^{\text{PPS}} &\sim N(d_{t_{i,k}}^{\text{PPS}} - d_{t_{i,1}}^{\text{PPS}}, \sigma_{\text{PPS}}^2) && \text{random effects model.} \end{aligned} \quad (5)$$

For the multiplicative model, the NMA models are specified on the log-scale for PFS and PPS, so Equations (2) and (4) are replaced by:

$$\log(\theta_{i,k}^{\text{PFS}}) = \mu_i^{\text{PFS}} + \delta_{i,k}^{\text{PFS}}, \quad (6)$$

and

$$\log(\theta_{i,k}^{\text{PPS}}) = \mu_i^{\text{PPS}} + \delta_{i,k}^{\text{PPS}}, \quad (7)$$

where  $\mu_i^{\text{PFS}}$  and  $\mu_i^{\text{PPS}}$  are the baseline log-RMSTs for study  $i$ , and  $\delta_{i,k}^{\text{PFS}}$  and  $\delta_{i,k}^{\text{PPS}}$  are the log-ratio of RMSTs for treatment in arm  $k$  relative to the treatment in arm 1 in study  $i$ , for PFS and PPS respectively. Equations (1), (3), and (5) are unchanged, but the interpretation changes so that  $d_k^{\text{PFS}}$  is the log-ratio of RMST for treatment  $k$  relative to treatment 1 ( $d_1^{\text{PFS}} = 0$ ), and  $\sigma_{\text{PFS}}$  is the between-study

standard deviation in log-ratio of RMST, and similarly for PPS.

## 4 | MODEL IMPLEMENTATION AND CRITIQUE

### 4.1 | Model implementation and priors

We took a Bayesian approach to model estimation. All analyses were run in WinBUGS version 1.4.3.<sup>36</sup> Normal  $(0,100^2)$  prior distributions are given to the trial-specific baselines  $\mu_i^{\text{PFS}}, \mu_i^{\text{PPS}}$  and for the treatment effects on AUC,  $d_k^{\text{PFS}}, d_k^{\text{PPS}}$ . These priors were chosen to be very uninformative, so that the posteriors are driven by the data rather than the priors. Note that for the additive model the trial-specific baselines must be positive (which was the case in our results), however an alternative prior specification to ensure this is to use a flat truncated normal prior.

In the case of random effects models, the between study standard deviations  $\sigma_{\text{PFS}}, \sigma_{\text{PPS}}$  for the treatment effects on AUC for PFS and PPS, respectively, were assigned Uniform(0,5) priors. In all the models run, convergence was assessed using the Brooks-Gelman-Rubin diagnostic and was satisfactory by 20,000 simulations for all outcomes. All results presented are based on a further sample of 40,000 iterations on two chains. All WinBUGS code and data is available in Appendices A4–A7.

### 4.2 | Assessing model fit

The posterior mean of the residual deviance, which measures the magnitude of the differences between the observed data and the model predictions of the data, was used to assess the goodness of fit of each model.<sup>37</sup> Smaller values are preferred, and in a well-fitting model the posterior mean residual deviance should be close to the number of data points in the network (each study arm contributes one data point).<sup>37</sup> The deviance information criterion (DIC) was used to compare models. The DIC is equal to the sum of the posterior mean deviance and the effective number of parameters,  $p_D$ , and thus penalises model fit with model complexity.<sup>37</sup> Lower values are preferred and differences of at least five points were considered meaningful.<sup>37</sup>

### 4.3 | Assessing heterogeneity and inconsistency

Heterogeneity concerns the differences in treatment effects between trials within each treatment contrast,

while consistency concerns the differences between the direct and indirect evidence informing the treatment contrasts.<sup>38,39</sup> Heterogeneity was assessed by comparing the fit of fixed and random effects NMA models. The estimated between-study standard deviation in treatment effects was also inspected to assess heterogeneity. Inconsistency was assessed by comparing the fit of the chosen consistency model (fixed or random effects) to an “inconsistency”, or unrelated mean effects, model.<sup>38,39</sup> The latter is equivalent to having separate, unrelated, meta-analyses for every pairwise contrast, with a common between-study variance parameter assumed in the case of random effects models. Scatter plots of the deviance values for each data point predicted by the consistency vs. inconsistency models were also inspected, where points notably below the line of equality may suggest evidence of inconsistency.<sup>2</sup>

## 5 | RESULTS FOR THE NON-SMALL CELL LUNG CANCER SYNTHESIS

For both the additive and multiplicative effect models, there were no meaningful differences between the fixed and random effects models in terms of the posterior mean residual deviance and DIC (Table 2). This is perhaps not surprising as there were at most two studies per comparison with which to estimate the between studies standard deviation in the random effects models. Results are therefore presented for the fixed effect models. To test whether there was any evidence of inconsistency we compared the fit of the NMA (consistency) model with the fit of the unrelated mean effects (inconsistency) model. For both the additive and multiplicative models there was no improvement in the posterior mean deviance when the consistency assumption was relaxed, and the NMA (consistency) model was preferred based on the DIC (Table 2). Figure 2 shows the contribution of each study arm to the posterior mean residual deviance under the inconsistency model vs the consistency (NMA) model for (a) the additive model and (b) the multiplicative model. There are no points below the line of equality, indicating no evidence of inconsistency for either the additive or multiplicative model. There is however one data-point (van Meerbeeck 2007) which has a very high deviance under the inconsistency model for the multiplicative model. This is the only study directly comparing CS versus CR, which gives a highly uncertain estimate of PPS (effectively equal to the prior), which leads to the high deviance for this study. In the NMA (consistency) model the CS versus CR estimate is identifiable due to borrowing strength from the evidence network under the

TABLE 2 Model fit statistics based on 5-year follow-up data

Model		Posterior median between-study SD (95% CrI)	Posterior mean residual deviance <sup>a</sup>	Effective number of data-points, pD	DIC
Additive	Fixed effect, consistency	–	15.85	14.02	29.87
	Random effects, consistency	PFS: 0.16 (0.01, 1.20) PPS: 0.20 (0.01, 1.51)	17.24	16.35	33.59
	Fixed effect, inconsistency	–	17.63	16.07	33.70
Multiplicative	Fixed effect, consistency	–	17.35	14.55	31.90
	Random effects, consistency	PFS: 0.09 (0.00, 0.71) PPS: 0.33 (0.01, 2.84)	20.53	18.47	39.00
	Fixed effect, inconsistency	–	25.06	19.41	44.47

<sup>a</sup>Compared to 20 total data points.

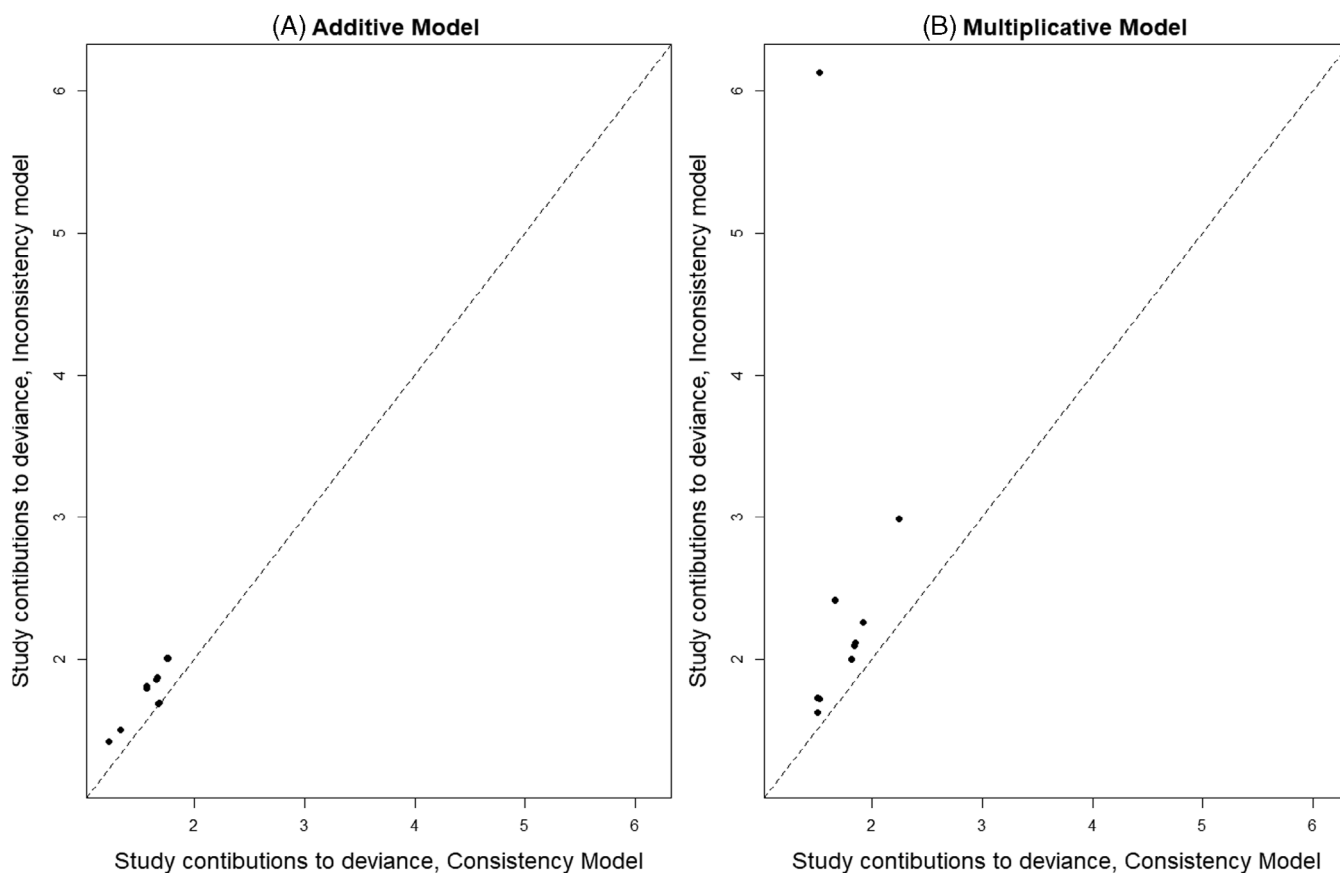
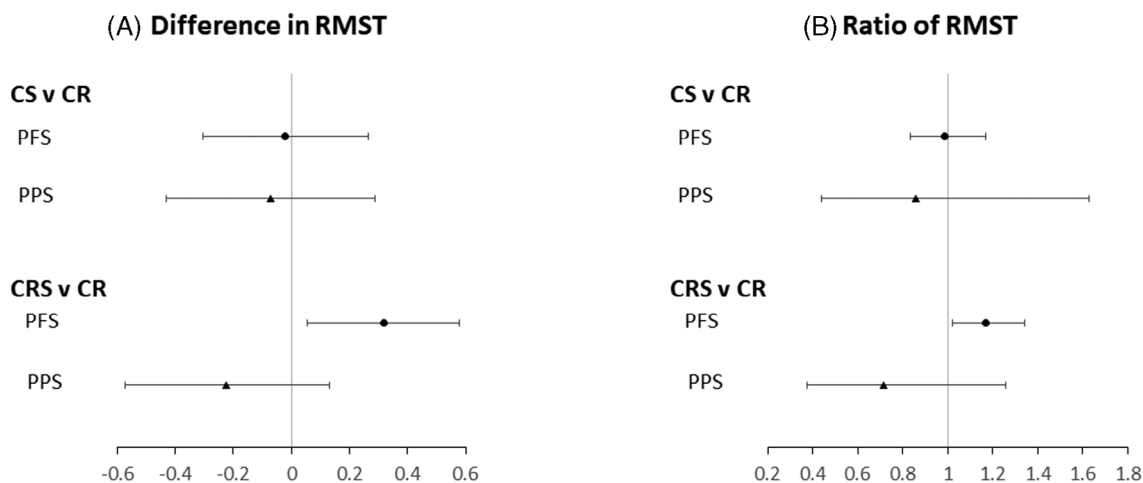


FIGURE 2 Deviance contributions from the fixed effect consistency and inconsistency models for area under the Kaplan–Meier curves pooled by (a) the additive model and (b) the multiplicative model. Points are plotted for each study and treatment arm. Points where the deviance contribution is much lower under the inconsistency model than the consistency model (i.e. points below the line of equality) indicate potential inconsistency

consistency assumption. The additive model gives a slightly better fit (posterior mean deviance 15.85, DIC 29.87) than the multiplicative model (posterior mean deviance 17.35, DIC 31.90), but the difference is not meaningful (Table 2) and we present results from both models.

Figure 3a shows the estimated treatment differences in RMST from the additive model. Differences in RMST can be interpreted as the additional expected time (in years) spent in a survival state (PFS or PPS) over a 5-year period. There are no differences in either PFS or PPS for CS relative to CR. There is evidence of an





**FIGURE 3** Relative treatment effects (a) RMST differences from the additive model and (b) RMST ratios from the multiplicative model. CR, chemoradiotherapy; CRS, chemoradiotherapy + surgery; CS, chemosurgery; PFS, progression free survival; PPS, post progression survival; RMST, restricted mean survival time

Median rank (95% CrI)	Additive model		Multiplicative model	
	PFS	PPS	PFS	PPS
CR	2 (2,3)	1 (1,3)	2 (2,3)	1 (1,3)
CS	3 (2,3)	2 (1,3)	3 (2,3)	2 (1,3)
CRS	1 (1,1)	3 (1,3)	1 (1,2)	3 (1,3)

**TABLE 3** Median rank and 95% credible interval for rank for each treatment and outcome under the additive and multiplicative models. Rank 1 has the greatest survival and rank 3 the least

increase in PFS for CSR relative to CR of 0.32 years with 95% CrI (0.05, 0.58). This may to some extent associated with a shorter time spent in PPS of  $-0.22$  years with 95% CrI ( $-0.57, 0.12$ ), although note that this interval crosses 0 (no effect). Quality of life is typically higher prior to disease progression in oncology, so CRS is likely to improve quality of life compared with CR or CS. Figure 3b shows the estimated ratios of RMST from the multiplicative model. Ratios of RMST can be interpreted as the proportional increase in time spent in a survival state (PFS or PPS) over a 5-year period. As for the additive model, the multiplicative model finds that there is evidence of an improvement in PFS for CRS relative to CR with an estimated ratio of RMST of 1.17 times longer spent in PFS over a 5-year period, with 95% CrI (1.02, 1.34). The 95% credible intervals (CrIs) corresponding to the estimated ratios of PPS are notably wider than those for PFS (Figure 3b).

NMA estimates can be summarised in terms of treatment rankings. This is achieved in a Bayesian analysis by ranking the treatment estimates at each iteration of the Markov chain Monte Carlo simulation, and then forming the posterior summaries for the treatment ranks. Table 3 shows the posterior median rank and 95% credible intervals for the rank for each treatment, outcome (PFS and PPS) for the additive and multiplicative models. Under

both models CRS is ranked first for PFS (median rank 1, 95% CrI [1,1]), and there is a high degree of uncertainty around the rankings for PPS.

## 6 | DISCUSSION

We have presented a non-parametric approach for jointly pooling individual patient data reconstructed from Kaplan–Meier curves for progression-free and overall survival. This simple approach allows one to estimate treatment effects on the restricted mean progression-free and post-progression survival times without assuming proportional hazards, allows for different shapes of survival curves, and respects the natural constraint that OS should be equal to or greater than PFS.

The appropriateness of assuming that the treatment effects are additive or multiplicative depends on both statistical and clinical judgement. In our example we found no meaningful differences in model fit, and qualitative conclusions were similar, supporting the use of the more easily interpreted additive model in the NICE guideline.<sup>25</sup> Modelling treatment effects as ratios may be warranted if the data is skewed,<sup>40</sup> which is generally the case for survival outcomes.<sup>41</sup> We observed this was the case for some of the distributions of AUC for PFS and OS created by

non-parametric bootstrapping ( $-0.08 < \text{skewness} < 0.15$ ), and so multiplicative treatment effects may be appropriate, although in this example results were robust to choice of model. Modelling multiplicative treatment effects also has the advantage that when used to generate absolute RMST the predictions are always positive. However, the additive model has the advantage that the difference in RMST for OS between two treatments may be computed by simply adding the difference in RMST for PFS and the difference in RMST for PPS; whereas in the multiplicative model, the ratio of RMST for OS between two treatments cannot be similarly derived.

We chose to estimate rRMST as the area under the Kaplan–Meier curves estimated by a Reimann sum. There are other estimators for RMST, and in a meta-analysis context, the Reimann sum has been compared with a pseudo-values estimator based on non-parametric jack-knife samples, as well as the area under a flexible parametric survival curve, through a simulation study.<sup>22</sup> Wei et al.<sup>22</sup> observed similar bias and mean squared error (MSE) for the Reimann sum and pseudo-values estimators, while lower MSE was observed for the flexible parametric survival method when the survival time followed a Weibull distribution. Given our proposed approach is motivated by situations where parametric models are not suitable, and the Reimann sum is readily available in standard software (e.g. survival package in R), we suggest using this estimator.

Jointly modelling the AUCs for PFS and OS requires an estimate of the correlation between the two. We chose to estimate this correlation through non-parametric bootstrapping, but note this only captures the correlation that is induced through the requirement that the AUC for OS is greater than the AUC for PFS. Any further correlation resulting from PPS depending on time to progression cannot be captured without the joint IPD for both outcomes which is rarely available. It may be possible to estimate this correlation from external information. For example, there is considerable literature in the oncology field investigating the association between PFS and OS,<sup>42</sup> although we note that correlation estimates might not be available for the restricted time considered in the NMA. Sensitivity analyses may be conducted to assess the impact of assumed correlation estimates. We previously ran the analysis using stronger, positive correlation estimates, and found the overall results were robust to this. In general, we observed that as correlation decreased, but remained positive, the uncertainty in the estimated relative effects increased.

A limitation of our method is that only studies reporting results at the restricted follow-up time (5 years in the NSCLC example) can be included. If there are studies with shorter follow-up there are several options. Studies

with shorter follow-up could simply be excluded, which may be a reasonable approach if follow-up is so short that those studies contain very little information, and most studies have sufficient follow-up to be included. An alternative approach is to choose the restricted follow-up time to be short enough for all studies to be included. This has the advantage that all studies can be included but loses information in the later part of the survival curves from the studies with longer follow-up. Another approach is to regress AUC at the restricted follow-up time against AUC at earlier follow-up times, based on studies that allow both to be derived, and use the regression equation to predict AUC at the restricted follow-up time in those studies where it is missing. Finally parametric models could be fitted to studies with shorter follow-up so that the AUC up to the restricted follow-up time can be obtained by extrapolation. This does however lose the non-parametric advantages of the method for those studies where extrapolation is necessary. For the NSCLC example we ran a sensitivity analysis using a 4-year restricted follow-up time so that the Girard study<sup>43</sup> could be included. We found that the conclusions were very similar, but the Girard study<sup>43</sup> (which was based on small patient numbers) was identified as a potential outlier. For this reason, and because we prefer longer follow-up if possible, the 5-year restricted follow-up time omitting the Girard study<sup>43</sup> is presented here.

In our example, all trials reported Kaplan–Meier curves for both PFS and OS. Whilst it is standard to report both in oncology trials, it may be the case that some included trials may only report the Kaplan–Meier curve for either PFS or OS, creating a missing data problem. Modelling strategies to jointly synthesis PFS and OS with missing outcomes have been proposed assuming proportional hazards,<sup>14</sup> and extending those methods to the non-parametric approach described here is worth future investigation.

It is challenging to model the data from the NSCLC example with standard methods due to violation of proportional hazards and different shaped survival curves (Appendix A1). Appendix A8 shows that the conclusions drawn from fitting a proportional hazards (PH) model are similar to those from the AUC method, although the conclusion that CRS is more effective than CR is stronger from the AUC model (Figure A8.1). The relative effects from fitting an accelerated failure time (AFT) model however give different conclusions from the AUC model, showing some evidence that CR is more effective than CS for both PFS and OS, although the credible intervals do cross the line of no effect (Figure A8.1). Only the AUC approach gives estimates for PPS. To obtain estimates of RMST from standard methods assumptions regarding the survival curve for the reference treatment are required, which is not necessary for the AUC method. Fitting

parametric curves to the CR arms of the studies which included CR indicates that the generalised gamma distribution is an appropriate fit based on AIC for most studies and outcomes (Table A8.1). The generalised gamma distribution has the AFT property for modelling treatment effects, and so relative effects (log time-ratios) from the AFT model can be applied to a reference generalised gamma distribution, and the RMST difference then calculated. It is not straightforward to apply a log-hazard ratio to a generalised gamma distribution. Appendix A8 shows that the RMST differences from an AFT model AFT approach gives different estimates of RMST difference than those obtained from the non-parametric AUC approach, which makes fewer assumptions. The results from the AFT approach shows CS is less effective than CR for both PFS and OS, whereas the AUC approach shows no effect (Figure A8.2). In addition, the AUC approach shows a benefit of CRS compared with CS, whereas results from the AFT approach shows no effect (Figure A8.2). We consider the results from the AUC approach to be more robust than the AFT results since it makes less assumptions.

To further illustrate the difficulties in applying standard methods Appendix A9 provides a synthesis of simulated data for three studies with two treatments and a single survival outcome where proportional hazards is violated and survival curve shapes differ between studies and treatments (Figure A9.1). The treatment effect estimates clearly show a benefit whether based on RMST difference, log hazard ratio, or log time-ratio (Figure A9.2), however very different estimates are obtained for RMST difference using the different methods (Figure A9.3). The PH method gives different results depending on the study used to estimate survival on treatment 1, and the AFT method overestimates the RMST difference compared with the AUC method which makes less assumptions and gives the closest estimate to the true-pooled RMST difference from which the study data were simulated.

To fully explore the performance of the different methods for the synthesis of survival outcomes would require a comprehensive simulation study following the ADEMP procedure.<sup>44</sup> This would be a substantial piece of work which explores the performance of a range of parametric, non-parametric, and flexible models for the joint synthesis of PFS and OS outcomes in a network meta-analysis to estimate restricted mean differences. Simulation scenarios should include factors such as the degree to which proportional hazards (or proportional time ratios) do or do not hold, varying survival curve shapes across studies and treatments in the network, the restriction time, follow-up times in the studies, degree of censoring in the studies, choice of distribution and data-source for the survival curve for the reference treatment used to obtain RMST difference, and so on.

Reimbursement decisions for new cancer therapies consider both cost-effectiveness and effectiveness. Cost-effectiveness is assessed using economic models that typically model movements between three states (progression free, post-progression and death), usually taking a partitioned survival analysis approach.<sup>23</sup> These models require estimates of the mean time spent in the PFS state and the mean time spent in the PPS state over a time-horizon chosen to capture the period where there may be differential effects, often over a patient's lifetime. To obtain absolute RMST estimates for each treatment, the relative treatment effects (differences or ratios of RMSTs) may be applied to a baseline RMST on the reference treatment. The evidence source used for the baseline RMST should be considered representative of the population of interest, which could be the most recent relative RCT in the absence of other information. Note the restricted follow-up time for the RMSTs used by our method is likely to be less than the time horizon required in an economic model because it is limited by the follow-up periods of the included RCTs. Predicting lifetime mean survival time beyond the restricted follow-up time,  $t_T$ , requires information on long-term survival conditional on survival status at  $t_T$  using data from an external source.<sup>10</sup> For treatments that have been in use for some time, this information may be available from registries, cohorts, and other routine data sources, and we can combine the RMST estimates with an estimate of life-expectancy conditional on surviving the first 5 years from registry data (e.g. Surveillance, Epidemiology, and End Results (SEER) Program<sup>45</sup>) in a population matched to the characteristics of the RCT populations. However, registry data rarely collects information on disease progression, and will not have long-term evidence on newer treatments. It is likely therefore that assumptions will need to be made to extrapolate beyond the restricted follow-up time,  $t_T$ . For the NSCLC guideline it was assumed that patients who are alive and progression-free at 5 years would remain progression-free until death,<sup>25</sup> and survival conditional on being alive at 5 years was independent of initial treatment and taken from SEER.<sup>45</sup>

The methods presented can also be applied to the case where there is a single time-to-event outcome. A univariate Normal likelihood would be given for the AUC estimate

$$y_{i,k} \sim N(\theta_{i,k}, \text{se}_{i,k}^2),$$

and the NMA model put on the RMST parameters  $\theta_{i,k}$  as in Equations (2) and (3) for the additive model and Equations (3) and (6) for the multiplicative model. The approach therefore has applicability outside of oncology,

in situations where there is a time-to-event outcome and the proportional hazards assumption is in question.

In summary, we have presented a non-parametric approach to jointly synthesise relative treatment effects from PFS and OS Kaplan–Meier curves, which does not assume proportional hazards, conforms to the constraints on PFS and OS, and provides inputs required for a partitioned survival cost-effectiveness model. The estimates can be combined with external sources on long-term survival to obtain estimates of mean time progression-free and post-progression.

## ACKNOWLEDGEMENTS

CHD, AEA, and NJW were supported by the National Institute of Health and Care Excellence (NICE) Guidelines Technical Support Unit at the University of Bristol, with funding from the NICE Centre for Guidelines, and RM was employed by NICE for the period when this work was conducted. NJW and AEA were also partly funded by the UK Medical Research Council grant MR/R025223/1.

## CONFLICT OF INTEREST

The views expressed in this publication are those of the authors and not necessarily those of NICE. The authors declare no conflict of interest.

## AUTHOR CONTRIBUTION

CHD conducted statistical analyses, drafted and reviewed the manuscript. RM provided the data, context, and interpretation of the results, and reviewed and commented on the manuscript. AEA conceived the methodology, advised on the statistical analyses, and reviewed and commented on the manuscript. NJW conducted and advised on the statistical analyses, and drafted, revised, and reviewed the manuscript.

## DATA AVAILABILITY STATEMENT

All data is available in the supplementary materials (the summary measures for each study) or by request from the authors (the reconstructed individual patient data for each study).

## ORCID

Caitlin H. Daly  <https://orcid.org/0000-0002-8272-797X>

A. E. Ades  <https://orcid.org/0000-0001-7822-3552>

Nicky J. Welton  <https://orcid.org/0000-0003-2198-3205>

## REFERENCES

- Higgins JPT, Whitehead A. Borrowing strength from external trials in a meta-analysis. *Stat Med*. 1996;15:2733-2749.
- Dias S, Ades AE, Welton NJ, Jansen JP, Sutton AJ. *Network Meta-Analysis for Decision Making*. Wiley; 2018:488.
- Briggs A, Sculpher M, Claxton K. *Decision Modelling for Health Economic Evaluation*. Oxford University Press; 2006.
- Latimer, N., NICE DSU Technical Support Document 14: Undertaking survival analysis for economic evaluations alongside clinical trials—extrapolation with patient-level data. 2011, NICE Decision Support Unit, <http://www.nicesdu.org.uk>.
- Guyot P, Welton NJ, Ouwens MJNM, Ades AE. Survival time outcomes in randomised controlled trials and meta-analyses: the parallel universes of efficacy and cost-effectiveness. *Value Health*. 2011;14:640-646.
- Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Mak*. 2013;33:607-617.
- Collett D, ed. *Modelling Survival Data in Medical Research*. 2nd ed. Chapman & Hall/CRC; 2003.
- Jansen JP. Network meta-analysis of survival data with fractional polynomials. *BMC Med Res Methodol*. 2011;11(1):1-14.
- Jackson C, Stevens J, Ren S, et al. Extrapolating survival from randomized trials using external data: a review of methods. *Med Decis Mak*. 2017;37:377-390.
- Guyot P, Ades AE, Beasley M, Lueza B, Pignon JP, Welton NJ. Extrapolation of survival curves from cancer trials using external information. *Med Decis Mak*. 2016;37:353-366.
- Bagust A, Beale S. Survival analysis and extrapolation modeling of time-to-event clinical trial data for economic evaluation: an alternative approach. *Med Decis Mak*. 2014;34:343-351.
- Barrett J, Farewell VT, Siannis F, Tierney J, Higgins JP. Two-stage meta-analysis of survival data from individual participants using percentile ratios. *Stat Med*. 2012;31:4296-4308.
- Siannis F, Barrett JK, Farewell VT, Tierney JF. One-stage parametric meta-analysis of time-to-event outcomes. *Stat Med*. 2010;29:3030-3045.
- Welton NJ, Willis SR, Ades AE. Synthesis of survival and disease progression outcomes for health technology assessment of cancer therapies. *Res Synth Methods*. 2010;1:239-257.
- Ouwens MJNM, Philips Z, Jansen JP. Network meta-analysis of parametric survival curves. *Res Synth Methods*. 2010;1:258-271.
- Cope S, Chan K, Jansen J. Multivariate network meta-analysis of survival function parameters. *Value Health*. 2017;20:A401-A402.
- Lu G, Ades AE, Sutton AJ, Cooper NJ, Briggs AH, Caldwell DM. Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med*. 2007;26(20):3681-3699.
- Freeman S, Carpenter J. Bayesian one-step IPD network meta-analysis of time-to-event data using Royston-Parmar models. *Res Synth Methods*. 2017;8:451-464.
- Harris S, Brown J, Lopez J, Yap TA. Immuno-oncology combinations: raising the tail of the survival curve. *Cancer Biol Med*. 2016;13:171-193.
- Royston P, Parmar M. The use of restricted mean survival time to estimate the treatment effect in randomized clinical trial when the proportional hazards assumption is in doubt. *Stat Med*. 2011;30:2409-2421.
- Royston P, Parmar MKB. 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;13:152-152.
- Wei Y, Royston P, Tierney JF, Parmar MKB. Meta-analysis of time-to-event outcomes from randomized trials using restricted

- mean survival time: application to individual participant data. *Stat Med.* 2015;34:2881-2898.
23. Woods B, Sideris E, Palmer S, Latimer N, Soares M. *NICE DSU Technical Support Document 19: Partitioned Survival Analysis for Decision Modelling in Health Care: a Critical Review.* University of Sheffield; 2017.
  24. Riley RD, Jackson D, Salanti G, et al. Multivariate and network meta-analysis of multiple outcomes and multiple treatments: rationale, concepts, and examples. *BMJ.* 2017;358:j3932.
  25. National Institute for Health and Care Excellence. *Lung Cancer: Diagnosis and Management.* National Institute for Health and Care Excellence; 2019.
  26. Guyot P, Ades AE, Ouwens MJNM, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Med Res Methodol.* 2012;12:9.
  27. Albain K, Swann RS, Rusch VW, et al. Radiotherapy plus chemotherapy with or without surgical resection for stage III non-small-cell lung cancer: a phase III randomised controlled trial. *Lancet.* 2009;374:379-386.
  28. Katakami N, Tada H, Mitsudomi T, et al. A phase 3 study of induction treatment with concurrent chemoradiotherapy versus chemotherapy before surgery in patients with pathologically confirmed N2 stage IIIA non-small cell lung cancer (WJTOG9903). *Cancer.* 2012;118:6126-6135.
  29. Pless M, Stupp R, Ris HB, et al. Induction chemoradiation in stage IIIA/N2 non-small-cell lung cancer: a phase 3 randomised trial. *Lancet.* 2015;386:1049-1056.
  30. R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing; 2018.
  31. Therneau, T., A Package for Survival Analysis in R. R package version 3.2-13. 2015.
  32. Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data.* 2nd ed. Springer-Verlag; 2003.
  33. Eberhardt W, Pöttgen C, Gauler TC, et al. Phase III study of surgery versus definitive concurrent Chemoradiotherapy boost in patients with Resectable stage IIIA(N2) and selected IIIB non-small-cell lung cancer after induction chemotherapy and concurrent Chemoradiotherapy (ESPA-TUE). *J Clin Oncol.* 2015;33:4194-4201.
  34. van Meerbeeck J, Kramer GW, van Schil P, et al. Randomized controlled trial of resection versus radiotherapy after induction chemotherapy in stage IIIA-N2 non-small-cell lung cancer. *J Natl Cancer Inst.* 2007;99:442-450.
  35. Efron B, Tibshiranie RJ. *An Introduction to the Bootstrap.* Chapman & Hall; 1993.
  36. Spiegelhalter, D., A. Thomas, N. Bes, D. Lunn, WinBUGS User Manual Version 1.4 January 2003. Upgraded to Version 1.4.3. 2007.
  37. Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc B.* 2002; 64(4):583-616.
  38. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE. *NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials.* University of Bristol; 2011.
  39. 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 Mak.* 2013;33:641-656.
  40. Spiegelman D, VanderWeele TJ. Evaluating public health interventions: 6. Modeling ratios or differences? Let the data tell us. *Am J Public Health.* 2017;107(7):1087-1091.
  41. Saad ED, Zalberg JR, Péron J, Coart E, Burzykowski T, Buyse M. Understanding and communicating measures of treatment effect on survival: can we do better? *J Natl Cancer Inst.* 2017;110(3):232-240.
  42. Davis S, Tappenden P, Cantrell A. *A Review of Studies Examining the Relationship between Progression-Free Survival and Overall Survival in Advanced or Metastatic Cancer.* Decision Support Unit; 2012.
  43. Girard N, Mornex F, Douillard JY, et al. Is neoadjuvant chemoradiotherapy a feasible strategy for stage IIIA-N2 non-small cell lung cancer? Mature results of the randomized IFCT-0101 phase II trial. *Lung Cancer.* 2010;69:86-93.
  44. Morris T, White I, Crowther M. Using simulation studies to evaluate statistical methods. *Stat Med.* 2019;38:2074-2102.
  45. Fuller CD, Wang SJ, Thomas CR. Conditional survival in head and neck squamous cell carcinoma: results from the SEER dataset 1973–1998. *Cancer.* 2007;109:1331-1343.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Daly CH, Maconachie R, Ades AE, Welton NJ. A non-parametric approach for jointly combining evidence on progression free and overall survival time in network meta-analysis. *Res Syn Meth.* 2022;1-12. doi:10.1002/jrsm.1539