National Institute for Health and Care Excellence

FINAL

Melanoma: assessment and management

[F] Evidence reviews for systemic and localised anticancer treatment for people with stage 4 (+ unresectable stage 3) melanoma

Network meta-analysis report

NICE Guideline NG14

Evidence reviews underpinning recommendations 1.7.1 to 1.7.2 and 1.8.6 to 1.8.17 and research recommendations in the NICE guideline

July 2022

Final

These evidence reviews were developed by the Guideline Updates Team

Disclaimer
Healthcare professionals are expected to take NICE clinical guidelines fully into account when exercising their clinical judgement. However, the guidance does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of each patient, in consultation with the patient and/or their guardian or carer.

ISBN: 978-1-4731-1322-0

Contents

Network meta-analysis

A.1 Background

Network meta-analysis (NMA) is a statistical technique that allows pooling of data for three or more interventions when the available evidence forms a connected network of intervention comparisons from trials, for example: evidence from trials comparing interventions A vs B, trials of B vs C and trials of C vs A. This enables both direct evidence (for example A vs B trials for the AvB comparison) and indirect evidence (for example A vs C and B vs C trials provide an indirect estimate of AvB) to be pooled. NMA combines all the available data into a single set of treatment effects that provide an ordering of intervention effectiveness, whilst respecting the randomisation in the included RCTs. The resulting estimates are easier to interpret than a series of pairwise comparisons, and because both direct and indirect evidence is pooled these are more precisely estimated. The estimates of treatment effect from the NMA provide a useful clinical summary of the results and facilitate the formation of recommendations based on the best available evidence. Having a single set of treatment effects that considers all the available evidence also facilitates cost effectiveness analysis.

The review for this guideline update comparing systemic chemotherapies in people with stage 4 and unresectable stage 3 melanoma formed a connected network of RCT evidence for the *BRAF* wild type and the mixed (*BRAF* wild type and mutant) melanoma populations and so an NMA was considered for these populations. For the *BRAF* mutant subgroup there were insufficient data to allow an NMA (see section A.4.2).

This topic was considered a high clinical priority for the guideline due to variations in practice and uncertainty about the most clinically and cost-effective strategy. It was also given the highest priority for new economic modelling. Given this, the committee agreed that network meta-analysis was warranted to facilitate cost effectiveness analysis and help decision making in this area.

The key outcomes of interest in this analysis were overall survival (OS) and progression-free survival (PFS) measured over time. These are typically reported as Kaplan-Meier (KM) curves that show the probability of death at respective time intervals and were identified through a systemic review of the literature. Visual inspection of the Kaplan Meier curves for each treatment revealed that the proportional hazards assumption did not appear to hold, and so traditional pooling of hazards ratios was not considered appropriate.

In this report, we describe the identification of evidence to inform the analysis and how it was prepared for analysis, which included fitting parametric models (and selecting most appropriate) and aggregating the data into time intervals for flexible models. We give an overview of the NMA methods considered and ultimately selected to estimate the treatment effects on the Kaplan Meier curves for OS and PFS. We performed NMA for each of these models and produced predicted survival curves from the models. We then describe how we selected models based on model fit and committees view on plausibility of the predicted survival curves and checked for inconsistency in the NMAs. We then present the results from the NMAs and the estimates to be inputted into the economic model.

A.2 Methods

A.2.1 Identification of evidence for the NMA

A.2.1.1 Population and comparators

The treatment pathway for stage 4 or unresectable stage 3 melanoma varies dependent on whether someone has *BRAF* wild type melanoma or if they have *BRAF* mutant melanoma. The systemic anticancer treatments available include:

- Immunotherapies: nivolumab with ipilimumab, pembrolizumab, nivolumab, and ipilimumab. These are available for people who have *BRAF* wild type melanoma and people with a *BRAF* mutation.
- *BRAF*/MEK inhibitors: dabrafenib with trametinib, encorafenib with binimetinib, vemurafenib, dabrafenib. These are available for people with a *BRAF* mutation.

Currently, *BRAF*/MEK inhibitors are given most frequently as dual therapies. While dabrafenib and vemurafenib are approved as monotherapies for those with *BRAF* mutations and ipilimumab is approved for all people with untreated, advanced melanoma, these are not commonly used in current clinical practice and are not considered within this analysis unless they provide evidence that is necessary to create a connected network (Section A.4).

A.2.1.2 Search

As part of this guideline update, we performed a systematic literature review on systemic treatments for advanced melanoma, which included the immunotherapy strategies and the *BRAF*/MEK therapies for people with *BRAF* mutant melanoma (Section A.2.1.1). Full details on this search can be found in Evidence Review F.

All studies that were included in the systematic review were assessed for inclusion in the NMA. Additionally, we reviewed NICE TAs for systemic and localised anticancer treatments for people with stage IV (or unresectable stage 3) melanoma. This review consisted of three parts.

1) NICE TAs for the population of interest, patients with stage 4 and unresectable stage 3 melanoma, were identified. All TAs for this population were included, regardless of their final appraisal determination. Thus, technologies that were recommended and technologies that were not recommended were included.
2) The TAs identified in Step 1 were read, and the clinical trials used in the companies' submissions were recorded.
3) The clinical trials identified in Step 2 were searched using clinicaltrials.gov and all papers indexed to said trials were requested and assessed to see if they contained KM curves suitable for inclusion in the NMA.

Finally, additional papers not identified through the previous two methods were identified by the committee and further searches to identify publications with longer follow-up for any the trials identified in the review.

Thus, the trials considered for inclusion in the NMA came from four sources: the systematic review undertaken as part of this guideline update, the additional review of NICE melanoma TAs, committee input and further supplementary searches. Trials that were not suitable for inclusion in the NMA were excluded, using the criteria for the systemic review of evidence (Appendix B:, Evidence Review F). If a trial was excluded at this stage, it was not considered further, and full texts of papers were not ordered for review. The full texts for all the papers identified were then ordered and reviewed to determine whether they reported data suitable for the NMA. The committee also advised whether certain trials were inappropriate for

inclusion in the network, and they were removed. A full list of all excluded trials and why they were excluded can be found in Section B.1.4.

A.2.1.3 KM curve selection

We then assessed all journal publications to select the most suitable paper for each trial in the network. Publications that did not have KM curves were not assessed further as they contained no data which could be used in the NMA. Several trials had multiple publications with distinctive KM curves, generally published over a number of years as longer follow-up data from the study becomes available for analysis.

In selecting curves from publications for trials included in the network, we followed four principles:

1. Publications needed to have relevant comparisons between at least two comparators of interest,
2. Publications with longer follow-up were prioritized,
3. Publications were prioritized if they had a first line (1L) treatment specific KM curves,
4. Publications that adjusted for treatment switching were considered for scenario analyses.

Publications needed to have relevant comparisons between at least two comparators of interest

For example, Robert et al, 2021 (1) presented a KM curve for PFS in people with or without emergent adverse events who received pembrolizumab. As this curve did not provide a comparison between two comparators, but rather presented analyses for a single comparator, this KM curve was not utilized in the network.

Publications with longer follow-up were prioritized

An illustrative example of this is with CheckMate 066, where several publications had KM curves. For this trial, Robert et al, 2020 (2) was preferred to Robert et al, 2015 (3) as it had 5 years of follow-up compared to 1 year of follow-up.

Publications were prioritized if they had a first line (1L) treatment specific KM curves

We prioritized first-line treatments for two reasons. First, our decision problem centred around the effectiveness of systemic anti-cancer treatments given as different lines of therapy. Thus, we need 1L treatment specific KM curves to assess the effectiveness of systemic anti-cancer treatments given as 1L treatments. As previously described, there was insufficient data publicly available to perform a network meta-analysis on systemic anti-cancer treatments given as further treatments (second or third line). Therefore, while we initially considered attempting a network meta-analysis for 2L or 3L treatments, it was not possible to perform given the absence of data. Second, treatment effects are known to vary based on which line of treatment they are given as. This difference may be explained biologically, as treatment naïve patients may have different immune responses, or they may be less well. Volume of disease is often a predictor of how someone will respond to treatment. Although typically unreported in trials, proxy measures such as M stage and LDH levels are usually reported. In the 1st and 2nd line phase III ipilimumab trials MDX010 and CA184, a higher proportion of people receiving 2nd line ipilimumab were M1C stage compared with those receiving 1st line ipilimumab. Another way this is demonstrated is that approximately 40% of patients in the real-world Sacco data went on to receive 2nd line treatment, the likely reason being they are less well. This difference may also be explained as self-sorting, if a person doesn't generate a strong response to a treatment, they are likely to progress and require further treatment. As an example, in KEYNOTE-006, KM curves for the entire study population as well as for those receiving first-line pembrolizumab or ipilimumab for advanced disease were presented. In these figures, there is an improved PFS

and OS response in those receiving treatment 1L. Therefore, KM curves for 1L treatment were prioritized to adequately reflect the treatment effects of 1L treatment.

Publications that adjusted for treatment switching were excluded

In several clinical trials, as per protocol, patients are allowed to switch treatments after disease progression or if it becomes clear there is a clinical benefit with one treatment. As such, if intention to treat analyses are performed as per most study protocols, these would underestimate the hazard ratio between the two treatments, as some of those who were randomized to a treatment determined to be less effective would also be receiving the clinical benefit of the better treatment. However, we ultimately decided against such analyses due to the lack of available data. Of the trials included in our network, only two had publications that adjusted for treatment switching, BREAK-3 (4) and BRF113220 (5). The use of these curves would therefore only provide us with data adjusted for treatment switching for two of the ten trials in our base case network. For such analyses to be useful, all treatments need to be adjusted for treatment switching. If only select treatments are adjusted for treatment switching, and the NMA is run, this will improve the treatment efficacy of the adjusted treatments and likely those connected to them in the network. For example, we have data that adjusts for treatment switching for BREAK-3, which would improve the treatment efficacy of dabrafenib monotherapy versus DTIC, and BRF113220 which would improve the treatment efficacy of dabrafenib + trametinib versus dabrafenib monotherapy. If the NMA is run with this updated data, dabrafenib monotherapy and dabrafenib + trametinib would appear better than they would if we used KM curves where treatment switching had not been adjusted for. Attempting such an analysis would unevenly impact the results based solely on which trials had treatment switching curves available. In the above example this would make dabrafenib monotherapy and dabrafenib + trametinib appear better, and the other targeted therapies which are connected to them in the network. Meanwhile, treatments that were not adjusted for treatment switching, such as the immunotherapies, would remain the same. As such we did not attempt these analyses because we felt it would introduce bias for which we could not account for into the NMA, namely an improvement in treatments based solely on whether analyses that adjust for treatment switching had been performed while all other treatments are left unadjusted. One could attempt to account for the lack of data, by determining a conversion factor from the trials where treatment switching was adjusted for and applying it to all other trials, but this requires assuming the adjustment observed in the analyses that account for treatment switching is the same across all trials and all treatments; we felt this was too large of an assumption to make. In an ideal situation, we would have had KM curves available that adjusted for treatment switching for all trials in the network and we accept this is a limitation of our analysis.

A.2.2 Networks

A summary of the networks is provided in Table 1.

Table 1: Summary of networks

Network	Population	Interventions	Length of follow-up
Network 1	*BRAF* mutant and *BRAF* wild type melanoma	Targeted therapy, immunotherapy	Any
Network 2	*BRAF* mutant melanoma	Targeted therapy, immunotherapy	Any
Network 3	*BRAF* wild type melanoma	Immunotherapy	Any[a]
Network 4	*BRAF* mutant and *BRAF* wild type	Immunotherapy	Any

Network	Population	Interventions	Length of follow-up
Network 5	*BRAF* mutant and *BRAF* wild type	Targeted therapy, immunotherapy	Long-term follow up only
Network 6	*BRAF* mutant and *BRAF* wild type	Immunotherapy	Long-term follow up only

(a) Although network 3 allowed for any amount of follow-up, it ended up having long-term follow up only. This is because its population was limited to those with BRAF wild type melanoma, which resulted in the exclusion of CheckMate 069 (as it only presented KM curves for a mixed population, not specifically those with BRAF wild type melanoma). CheckMate 069 was the only trial in this network with a short amount of follow-up, thereby leaving the remainder of the trials in this network with long-term follow up only.

A.2.2.1 Primary network

The population for the primary NMA (Network 1) is people with stage IV and unresectable stage III melanoma. This network consisted of a mixture of people with and without *BRAF* mutations.

All immunotherapy and *BRAF*/MEK inhibitor treatment strategies were included in the network: nivolumab, nivolumab with ipilimumab, ipilimumab, pembrolizumab, encorafenib with binimetinib, trametinib with dabrafenib, dabrafenib and vemurafenib.

A.2.2.2 Secondary networks

In addition to the broader advanced melanoma population, the committee wished to explore the possibility of making *BRAF* mutation-specific recommendations, given the difference in the treatment pathway between the wild type and the mutant subgroups. Therefore, additional secondary networks were also considered:

- Network 2: People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies.
- Network 3: People with *BRAF* wild type melanoma, with immunotherapy strategies only.

BRAF status is not generally considered to be an effect modifier for those receiving immunotherapies, suggesting that the treatment of immunotherapies is consistent between the *BRAF* mutant and *BRAF* wild type populations. However, the committee expressed some reservations regarding this assumption. As such, analyses of immunotherapy strategies only were conducted in the mixed (*BRAF* mutant and wild type) population for comparison with Network 3:

- Network 4: All people with melanoma (*BRAF* mutant and wild type), with immunotherapy strategies only.

A.2.2.3 Additional scenario analyses

As certain trials were limited in the amount of follow-up they had, we conducted a scenario analysis that only included trials with extended follow-up. Two trials had markedly shorter follow-up: BRIM-3, which despite having an OS curve with 60 months of follow-up, only had a PFS curve with 22 months of follow-up published, and CheckMate 069, which only had 24 months of follow-up for both PFS and OS. The follow-up for the trial with the next smallest amount is 45 months for PFS (BREAK-3) and 57 months for OS (COLUMBUS). Given the range of PFS and OS follow-up after excluding BRIM-3 and CheckMate 069 was 45-75 months, and 57-75 months respectively, no further consideration was given to excluding additional trials due to having limited follow-up. We conducted this analysis for the mixed population with all immunotherapy and *BRAF*/MEK inhibitor strategies (Network 5) and for the mixed population with immunotherapy strategies only (Network 6).

A.2.3 Extraction

Screengrabs of the KM curves for PFS and OS from studies designated for inclusion were saved as images.

A.2.4 Digitizing

The extracted KM curves were then digitized using either engauge digitizer, an open-source software that allows the digitization of KM curves (6), or R code (RStudio Version 1.4.1717), which does the same. We then used a validated algorithm on the digitized KM curves, as well as data on the numbers at risk and total number of events. This algorithm produces a set of individual patient data (survival times and censor times) for each treatment group for each study (7). This was done for both the PFS and OS curves.

A.2.5 Statistical methods

In preparation for this guideline update, we performed a preliminary review of methods for network meta-analysis with time to event outcomes, including methods that do not assume proportional hazards as it was considered very likely that the proportional hazards assumption would not be met. Though papers were available that discuss different methodology, there was an absence of information available to provide guidance on selecting one method over another. Freeman et al 2020 (8) summarises six main approaches to modelling time to event NMAs: 1) cox proportional hazards (PH) 2) restricted mean survival times 3) parametric models 4) piecewise exponential models 5) fractional polynomial models and 6) Royston-Parmar flexible parametric models.

In the absence of literature to inform the best method by which to perform this NMA, we considered fitting each of the methods identified in the literature, as described below.

A.2.5.1 Cox proportional hazards

We assessed the proportional hazards assumption in three ways: 1) statistical testing of the PH test based on weighted residuals developed by Grambsch and Therneau 2) graphical assessment of Schoenfeld residuals 3) graphical assessment of the log-log curves.

The Grambsch and Thernau test uses the Schoenfeld residuals to detect a linear trend with time. Where the lines overlap, particularly where this happens in the middle analysis time, it's visible in the smoothed line on the Schoenfeld residual plot as a bump (or dip), but where the smoother flattens out again, the test can't detect a slope, and so will return p > alpha. This is visible in CheckMate 069 (see Table 35) and CheckMate 067 (see Table 36). Where divergence from PH takes place at the beginning or end of analysis time, the test is much more sensitive (e.g., BRIM-3, Table 12). Therefore, while we present results for three tests of the PH assumption, we gave more weight to the log(-log) plots to assess the PH assumption.

Each network had trials where the proportional hazards assumption was not met (Appendix C:Cox proportional hazards), therefore it was not considered appropriate to fit a cox PH model for any network.

A.2.5.2 Restricted mean survival time

Restricted mean survival time (RMST) is the mean survival time accrued from randomisation up to T years. RMST can be estimated by the area under the survival curve up to time T, and the treatment effect estimated as the difference in AUCs between treatments. This measure does not assume proportional hazards and can be calculated regardless of the curve fitted to the data, including directly from the Kaplan-Meier curve, and so can allow for different survival distributions across studies. RMST has previously been used in NICE guidance where an NMA of TTE was performed (9,10)

Limitations of this approach are that 1) if studies differ in follow-up time, then either the RMST is restricted to the shortest follow-up or extrapolation methods are required and 2) external data is required to extrapolate beyond the restricted follow-up time. Because the follow-up varied between trials with the shortest follow-up at 22 months for PFS (BRIM-3) and 24 months for OS (CheckMate 069), the RMST approach was not considered suitable.

A.2.5.3 Generalized gamma

Whilst the Cox model is flexible because it makes no assumptions about the shape of the underlying baseline hazard, it makes a strong assumption about the treatment effect (proportional hazards). An alternative to the cox PH model, is to fit a parametric survival model which assumes a specific functional form for the baseline hazard, where treatment effects act on model parameters. This has the advantage of potentially relaxing the proportional hazards assumption, but the functional form can be too restrictive, resulting in curves that poorly fit the data.

We chose the generalized gamma distribution for this model because, unlike other parametric approaches which can be restrictive, the generalized gamma model is more flexible and also because other commonly used parametric models are special cases of the generalized gamma model (11).

We fit the generalized gamma model in a two-stage approach. Firstly, generalized gamma models were fit to each individual trial in RStudio Version 1.4.1717 to obtain relative treatment effects. We fit two generalized gamma models, one in which treatment effects depended only on the location parameter (an Accelerated Failure Time (AFT) model), and a more flexible model in which treatment effects depended on both the location and scale parameter. Secondly, we synthesized the treatment effect estimates from part 1 using a Bayesian framework within a standard fixed effect NMA model to estimate all parameters using Markov chain Monte Carlo simulation methods implemented in WinBUGS 14 for the model that was dependent on the location parameter only and in OpenBUGS v3.2.3 for the model that was dependent on both the location and scale parameters. Due to the limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

A.2.5.4 Piecewise exponential

The piecewise exponential assumes the hazard varies across time intervals, but is constant within time intervals, and the hazard ratio also varies across time-interval. Following the approach first detailed by Crowther et al, 2012 (12), we used a Poisson likelihood for number of events and time at risk on each piecewise interval to fit piecewise exponential models. These models were fit in a multi-stage approach: 1) We aggregated the reconstructed IPD, into number at risk at number of events in the time intervals for the piecewise models in RStudio Version 1.4.1717; 2) Using the aggregated data, we fit exponential models into each interval in WinBUGS 14. Due to the limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

We considered models with 1 cut point, 2 cut points and 3 cut points, and ultimately concluded that the model with 2 cut points at 12 and 18 months was the most appropriate for Network 1. The piecewise models that were run are listed below:

- 1 cut point
    - 6 months
    - 9 months
    - 12 months
    - 15 months
- 2 cut points
    - 6 and 12 months

    - 9 and 15 months
    - 9 and 18 months
    - 12 and 18 months
    - 12 and 20 months
    - 12 and 24 months
- 3 cut points
    - 6, 12 and 18 months
    - 12, 24 and 36 months

A.2.5.5 Fractional polynomials

Fractional polynomial models are very flexible models which allow flexibility in both the baseline hazard and the treatment effects and incorporate many parametric distributions as special cases. Following the approach first detailed by Jansen 2011 (13), we fit fractional polynomial models. These models were fit in a multi-stage approach: 1) We aggregated the reconstructed IPD, into number at risk at number of events in the time intervals for the fractional polynomial models in RStudio Version 1.4.1717; 2) Using the aggregated data we ran a fixed effect NMA using first-order fractional polynomials taking powers: -2, -1, -0.5, 0, 0.5, 1, 2, and 3 in WinBUGS 14. Given the difficulty in achieving model convergence for the first order fractional polynomials, despite a large burn-in and only running the models with two chains, we did not attempt to fit second order models as we found it incredibly unlikely that they would converge even with a substantially larger burn-in and again only using two chains.

The time intervals we used to obtain our aggregate data are as follows:

- Eight intervals: 0-6 months, 6-12 months, 12-18 months, 18-24 months, 24-30 months, 30-36 months, 36-42 months, and >42 months.

Due to the limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

A.2.5.6 Royston-Parmar flexible parametric model

The Royston-Parmar model is another flexible parametric model that estimates the baseline log-cumulative hazard and treatment effects using restricted cubic splines, and methods for its use in NMA have been developed by Freeman and Carpenter 2017 (14). Both the fractional polynomial and restricted cubic spline methods are very flexible models, and we would not expect there to be much difference in the results from these two approaches. Furthermore, given that we found that the results of the fractional polynomial models did not improve visual fit of predicted survival curves compared with the generalised and piecewise exponential models, we did not consider it likely that the restricted cubic spline models would improve visual fit either, and so did not consider these models further.

A.3 Implementation and model fit

A.3.1 Multi-arm trials

All BUGS code, available in A.8, can handle multi-arm studies (those with 3 or more arms). In our data-set there were three-arm studies, but no study had more than 3 arms. Because all models fitted were fixed effect models, it was not necessary to model the covariance structure in the random effects induced by multi-arm trials. However for the generalized gamma models, the likelihood was given to the estimated parameters from fitting a generalised gamma model to each study separately. These estimates are correlated due to (i) multiple comparisons from the same study for the 3-arm trials, and (ii) multiple parameters

estimated in the generalised gamma models with two treatment effect parameters. These correlations were incorporated using a multi-variate normal distribution for the likelihood.

A.3.2 Prior distributions

All models were fixed effect models, due to insufficient evidence to fit random effects models. Prior distributions were given to the treatment effects parameters. The generalized gamma with one treatment effect and piecewise exponential models were given non-informative normal prior distributions. The generalized gamma model with two treatment effects was given a non-informative bivariate normal prior distribution. The fractional polynomial models were given non-informative multivariate normal prior distributions.

A.3.3 Convergence

Convergence was assessed using the Brooks-Gelman-Rubin diagnostic plot.

For the generalised gamma convergence was satisfactory by 10,000 simulations for all outcomes. A further sample of 20,000 iterations per chain post-convergence was obtained on which all reported results were based. For the piecewise exponential convergence was satisfactory by 40,000 simulations for all outcomes. A further sample of 80,000 iterations per chain post-convergence was obtained on which all reported results were based. The generalized gamma dependent on location parameter alone and all piecewise exponential models were run with 3 chains, each with a different set of initial values, to check that the model had converged through the mixing of chain via history plots, and results were not influenced by the initial values. The generalized gamma model dependent on both location and scale parameters was run with 2 chains, again, each with a different set of initial values to check that the model had converged through the mixing of chain via history plots, and results were not influenced by initial values.

For the fractional polynomial models, a burn-in of 30,000 simulations was used. Not all first order fractional polynomials converged with a burn-in of this size. For those first order fractional polynomial models which did converge, a further sample of 30,000 iterations per chain post-convergence was obtained on which all reported results were based. The first order fractional polynomials were run with 2 chains, each with a different set of initial values, to check that the model had converged through the mixing of chain via history plots, and results were not influenced by the initial values. The reduction in the number of chains for the first order fractional polynomial models was necessitated by the fact that WinBUGS lacked sufficient memory space to store results for even one parameter when run with 3 chains with the required burn-in to achieve convergence and sample size. Even when clearing set nodes after burn-in, so that WinBUGS was only storing the post burn-in samples, still resulted in insufficient memory space to store results with 3 chains.

A.3.4 Model fit

For the generalised gamma models, we assessed the goodness of fit of the model by calculating the Akaike information criterion (AIC). This is equal to the sum of the deviance at the maximum likelihood estimate of parameters and twice the total number parameters and thus penalizes model fit with model complexity. We calculated the AIC for each trial in part one of the process. These values were then summed to obtain the cumulative AIC for each model, with a lower AIC value indicating a better fitting model.

For the piecewise exponential models, we assessed the goodness of fit of the model by calculating the deviance information criterion (DIC) (obtained from running the NMAs in WinBUGS). This is equal to the sum of the posterior mean deviance and the effective number of parameters and thus penalizes model fit with model complexity (15). We follow the advice of past research, where differences in DIC over 5 are considered important (16). It is

important to note in shifting the cut point for the piecewise models, one is then aggregating the IPD data differently. It is possible the DIC would no longer be informative because in aggregating the data with different cut points, the likelihoods would then be different and the DIC wouldn't be informative. However, we believe that while the data has been aggregated at different time points and using a different numbers of time points, it is ultimately still the same data, it has merely been divided differently. Therefore, while it has not been proven here, we believe the DIC values are still informative. Each event and time at risk are contributing to the deviance, so when summed over all those contributions you are in effect comparing like with like. The choice of cut point will determine which segment the observations will contribute to and that may give a better or worse fit. Furthermore, given the memoryless property of the Poisson likelihood, although we did not test this here, we believe that mathematically DIC would remain informative regardless of using different cut points (i.e., different aggregate data is used).

For the fractional polynomial models, we assessed the goodness of fit of the model by fitting several models in R calculating the AIC. Lower AIC values were presumed to indicate a better model fit. However, this was complicated in two ways. First, not all fractional polynomial models converged. Thus, occasionally the power that had the lowest AIC could not be used. In this instance, we selected the model with the lowest AIC value from the powers that converged. Additionally, we assessed goodness of fit of the model by calculating the DIC. In some instances, AIC and DIC were not in agreement, with AIC pointing to one power being the preferred model and DIC pointing to a different power being the preferred model. When this happened, we relied on visual inspection to select between the two powers.

A.3.5 Model validation

Model selection followed the guiding principle, 'the preferred method should balance goodness of fit and complexity of approach while retaining clinical plausibility.' Gibson et al, 2017 (17).

Upon fitting the models, it was necessary to select which model to use for both PFS and OS to incorporate into the economic model. To assess clinical plausibility, models were presented at multiple points to the full committee and at times a sub-committee composed of those specializing in oncology and pharmacy. These meetings entailed providing a visual inspection of the survival curves, and a discussion to assess the clinical plausibility of the predicted survival data over periods where IPD data exists, and extrapolations where survival is predicted beyond where we have data. We compared the treatment-specific KM curves from the trials versus the predicted treatment-specific survival curves, calculated by applying the treatment effect estimates from the NMA to the reference curve.

In an ideal scenario, statistical measures would allow for direct comparison of the models, using DIC values. However, this was not possible for this NMA as each of the models are using different data, therefore the likelihoods are not estimating the same thing and therefore cannot be used for decision making. While the IPD data is indeed the same, the generalized gamma model fits parametric models to the IPD data, and it is the treatment effects from this analysis that are then used in the NMA. For the piecewise models the IPD data is converted into aggregate data with different cut points. Finally, for the fractional polynomial models the IPD data is again converted into aggregate data, but these intervals are different than the one for the piecewise models, so it is not possible to compare the DIC values between these two models.

Figure 1: Decision bracket showing the process used to select the best fitting model

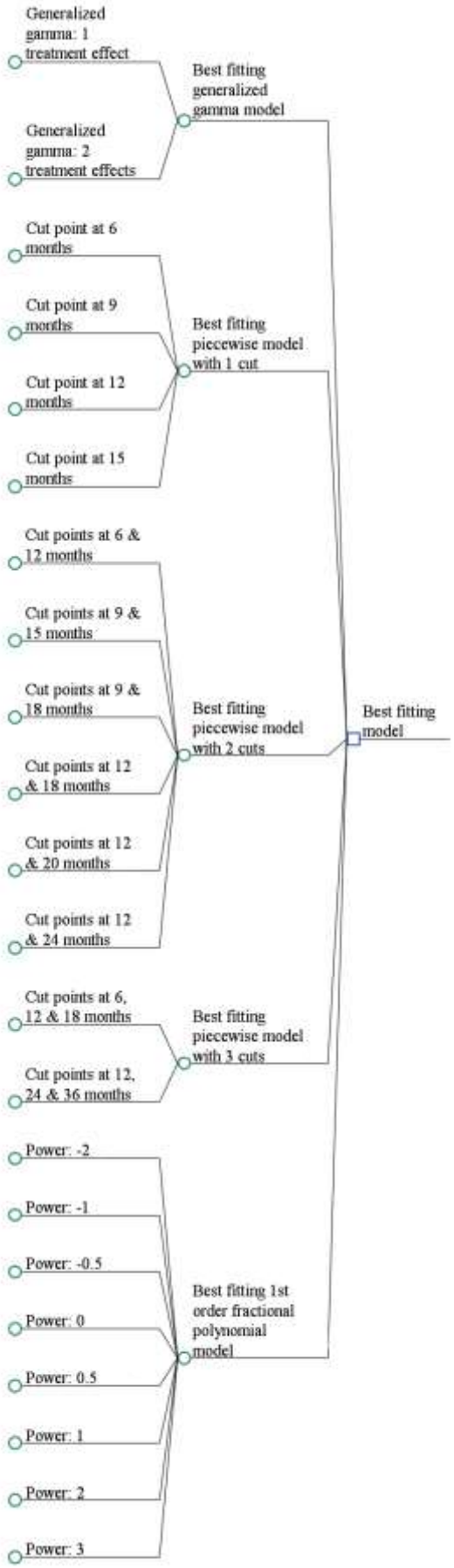


Figure 1 summarizes how we ultimately selected the best fitting model for PFS and OS. First, we followed the principals described in A.3.4 to select the best fitting model for each method (we did not select a best fitting piecewise model, but rather considered them all). From there, we considered the best fitting generalized gamma model, best fitting piecewise model with 1 cut, best fitting piecewise model with 2 cuts, best fitting piecewise model with 3 cuts and best fitting first order fractional polynomial model to select the best fitting model. Further detail on this process is given in the results for PFS A.5.1.1 and OS A.5.1.2.

A.3.6 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 (18,19).

Typically, heterogeneity is assessed by comparing the fit of fixed and random effects NMA models. However, as we did not fit random effects NMA models due to the limited number of studies for each comparison, we were unable to assess heterogeneity in this way. However, we still present the estimated between study standard deviation in treatment effects to assess heterogeneity. For PFS, this is done by presenting boxplots of the deviance for each data point. For OS, this is done by presenting boxplots of the residual deviance.

Inconsistency was assessed by comparing the fit of the fixed effects model to an "inconsistency", or unrelated mean effects, model (18,19). The latter is equivalent to having separate, unrelated, meta-analyses for every pairwise contrast. Note that inconsistency can only be assessed when there are closed loops of direct evidence on 3 treatments that are informed by at least 3 distinct trials (20).

A.3.7 Model outputs

The principal summary measure varied based upon the type of NMA methodology used.

For the generalized gamma model with one treatment effect, the principal outcome was the mu parameter, which can be described as a time-ratio.

For the generalized gamma model with one treatment effect, the principal outcomes were the mu and sigma parameter. The mu parameter can still be described as a time-ratio, however, sigma is interpreted as the shape parameter, as it effects the shape of the distribution.

For the piecewise exponential models, the principal outcomes were hazard ratios associated with the intervals derived from the selected cut points.

For the fractional polynomial models, the principal outcomes are β_0 and β_1, where β_0 is the treatment effect on the log-hazard at time 0, and the treatment effect on β_1 is the treatment effect on the (non-linear) relationship of the log-hazard over time.

Additional summary measures for each model include plots with survival predictions (curves) for each treatment and ranking plots for survival at 60 months, showing the likelihood of each treatment occupying each rank.

A.4 Evidence and networks

A.4.1 Summary of studies included in the effectiveness evidence

Table 2: Summary of clinical studies included in the network meta-analysis

Trial	Interventions (n)	Population	*BRAF* status	Previous treatment	Outcome	Follow-up time	Paper	Networks[a]
BREAK-3	Dabrafenib (n=187) Dacarbazine – hereafter referred to by its abbreviation - DTIC (n=63)	Patients with metastatic melanoma (Stage IV or unresectable stage III)	Patients had to have a *BRAF*^{V600E} mutation as per trial protocol	Patients with previously untreated *BRAF*^{V600E} mutant	PFS	45 months	Hauschild et al, 2020 (21)	Network 1 Network 2 Network 5
					OS	65 months	Hauschild et al, 2020 (21)	
BRF113220	Dabrafenib 150mg + Trametinib 2mg (n=54) Dabrafenib 150 mg + trametinib 1 mg (n=54)[b] Dabrafenib (n=54)	Patients with stage IIIC or IV melanoma	*BRAF*$^{V600E/K}$ mutant melanoma per trial protocol	BRAFi and MEKi naïve at initial study enrolment. Previous chemotherapy: n=7 (13%), n=15 (28%), n=12 (22%) by arm. Previous immunotherapy: n=13 (24%), n=16 (30%), n=8 (15%) by arm.	PFS	61 months	Long et al, 2018 (22)	Network 1 Network 2 Network 5
					OS	66 months	Long et al, 2018 (22)	
BRIM-3	Vemurafenib (n=337)	Patients with stage IIIC or	Patients with *BRAF*^{V600}	Patients who were treatment naïve	PFS	22 months	McArthur et al, 2014 (23)	Network 1 Network 2

[a] Network 1: mixed population and all comparators; Network 2: BRAF mutant and all comparators; Network 3: BRAF wild type and immunotherapies; Network 4: mixed population and immunotherapies; Network 5: mixed population and all comparators, data limited to long-term follow up

[b] Dabrafenib 150 mg + trametinib 1 mg arm in BRF113220 not included in the analysis, not representative of clinical practice

Trial	Interventions (n)	Population	*BRAF* status	Previous treatment	Outcome	Follow-up time	Paper	Networks[a]
	DTIC (n=338)	stage IV melanoma	mutations per trial protocol		OS	60 months	Chapman et al, 2017 (24)	
CheckMate 066	Nivolumab (n=210) DTIC (n=208)	Patients with stage III or IV melanoma	Patients without a *BRAF* mutation	Previously untreated per protocol. Prior adjuvant systemic therapy: n=32 (15.2%), n=36 (17.3%) by arm.	PFS	75 months	Robert et al, 2020 (2)	Network 1 Network 3 Network 4 Network 5 Network 6
					OS	75 months	Robert et al, 2020 (2)	
CheckMate 067	Nivolumab + ipilimumab (n=314) Nivolumab (n=316) Ipilimumab (n=315)	Patients with stage III (unresectable) or stage IV melanoma	*BRAF* mutation: n=101 (32.2%), n=100 (31.6%), 97 (30.8%). No mutation: 213 (67.8%), 216 (68.4%), 218 (69.2%) by arm.	No prior systemic treatment for unresectable or metastatic melanoma per trial protocol.	PFS	69 months	Larkin et al, 2019 (25)	Network 1 Network 3 Network 4 Network 5 Network 6
					OS	69 months	Larkin et al, 2019 (25)	
CheckMate 069	Nivolumab + ipilimumab (n=95)	Patients with unresectable stage III or IV melanoma	*BRAF* mutation-positive tumours:	Previously untreated per trial protocol	PFS	24 months	Hodi et al, 2016 (26)	Network 1 Network 4

Trial	Interventions (n)	Population	*BRAF* status	Previous treatment	Outcome	Follow-up time	Paper	Networks[a]
	Ipilimumab (n=47)		23 (24%), 10 (21%) BRAF wild-type tumours: 72 (76%), 37 (79%) by arm.		OS	24 months	Hodi et al, 2016 (26)	
COLUMBUS	Encorafenib + binimetinib (n=192) Vemurafenib (n=191) Encorafenib (n=194)[b]	Patients with unresectable stage IIIB/C or IV melanoma	Patients with $BRAF^{V600E/K}$ mutations per trial protocol	Patients were treatment-naïve or had progressed on or after previous first-line immunotherapy per trial protocol. Prior immunotherapy: 57 (30%), 57 (30%), 58 (30%) by arm.	PFS	54 months	Ascierto et al, 2020 (27)	Network 1 Network 2 Network 5
					OS	57 months	Ascierto et al, 2020 (27)	
COMBI-d	Dabrafenib + trametinib (n=211)	Patients with unresectable stage IIIC or stage IV melanoma	Patients with *BRAF* Val600Glu or Val600Ly	Previous immunotherapy: 57 (27%), 61 (29%) by arm.	PFS	71 months	Robert et al, 2019 (28) supplementary appendix	Network 1 Network 2 Network 5

[b] Although Columbus was a three-arm trial comparing encorafenib + binimetinib, vemurafenib monotherapy, and encorafenib monotherapy, this publication only presents KM curves for encorafenib + binimetinib and vemurafenib monotherapy. No KM curves for encorafenib monotherapy are presented. Additional publications do provide KM curves for all three treatments; however this comes at the cost of reduced follow-up. Therefore, we prioritized this publication as it gave longer follow-up. Furthermore, as encorafenib is not given as monotherapy in clinical practice, we did not feel it's absence from the NMA was significant – either mathematically as it would represent a single spur to the NMA, or clinically, as its results from the NMA wouldn't ultimately be used in the economic modelling.

Trial	Interventions (n)	Population	*BRAF* status	Previous treatment	Outcome	Follow-up time	Paper	Networks[a]
	Dabrafenib (n=212)		s mutations as determined by PCR		OS	77 months	Robert et al, 2019 (28) supplementary appendix	
COMBI-v	Dabrafenib + trametinib (n=352) Vemurafenib (n=352)	Patients with unresectable stage IIIC or IV melanoma	BRAF V600E or V600K mutations centrally determined with investigational use of THxID BRAF assay (bioMérieux)	Previously untreated per trial protocol	PFS	69 months	Robert et al, 2019 (28) supplementary appendix	Network 1 Network 2 Network 5
					OS	72 months	Robert et al, 2019 (28) supplementary appendix	
KEYNOTE-006	Pembrolizumab 10mg/kg every 2 weeks (n=279) Pembrolizumab 10mg/kg every 3 weeks (n=277) Ipilimumab (n=278)	Patients with unresectable stage III or IV melanoma	*BRAF*^{V600} mutation: 98 (35.1%), 97 (35.0%), 107 (38.5%).	Lines of previous systemic therapy: 0 – 183 (65.6%), 185 (66.8%), 181 (65.1%); 1 – 96 (34.4%), 91 (32.9%), 97 (34.9%).	PFS	50 months	Robert et al, 2019 (29)	Network 1 Network 3 (OS only) Network 4 Network 5 Network 6
					OS	64 months	Robert et al, 2019 (29)	

Table 3: Targeted therapy trial population characteristics

Trial	BREAK-3		BRF113220[a]		BRIM-3		COLUMBUS[b]		COMBI-d		COMBI-v	
Treatment (n)	DTIC (63)	Dabrafenib (187)	Dabrafenib (54)	Dabrafenib + trametinib 2mg (54)	DTIC (338)	Vemurafenib (337)	Vemurafenib (191)	Encorafenib + binimetinib (192)	Dabrafenib (212)	Dabrafenib + trametinib (211)	Vemurafenib (352)	Dabrafenib + trametinib (352)
Age, median (range)	50 (21-82)	53 (22-93)	50 (18-82)	58 (27-79)	52 (17-86)	56 (21-86)	56 (21-82)	57 (20-89)	56.5 (22-86)	55 (22-89)	54 (18-88)	55 (18-91)
Sex												
Female – no. (%)	26 (41%)	75 (40%)	25 (46%)	20 (37%)	157 (46%)	137 (41%)	80 (42%)	77 (40%)	98 (46%)	100 (47%)	172 (49%)	144 (41%)
Male – no. (%)	37 (59%)	112 (60%)	29 (54%)	34 (63%)	181 (54%)	200 (59%)	111 (58%)	115 (60%)	114 (54%)	111 (53%)	180 (51%)	208 (59%)
Eastern Cooperative Oncology Group (ECOG) performance status												
0 – no. (%)	44 (70%)	124 (66%)	34 (63%)	35 (65%)	230 (68%)	229 (68%)	140 (73%)	136 (71%)	150 (71%)	155 (74%)	248 (70%)	248 (71%)
1 – no. (%)	16 (25%)	62 (33%)	20 (37%)	19 (35%)	108 (32%)	108 (32%)	51 (27%)	56 (29%)	61 (29%)	55 (26%)	104 (30%)	102 (29%)
BRAF mutation												
V600E – no. (%)	(100%)[c]	(100%)[c]	48 (83%)	47 (87%)	(100%)[c]	(100%)[c]	168 (88%)	170 (89%)	181 (85%)	179 (85%)	317 (90%)	312 (90%)
V600K – no. (%)	-	-	9 (17%)	7 (13%)	-	-	23 (12%)	22 (11%)	30 (14%)	32 (15%)	34 (10%)	34 (10%)
Metastasis stage												
M0 – no. (%)	1 (2%)	6 (3%)	1 (2%)	0 (0%)	13 (4%)	20 (6%)	11 (6%)	9 (5%)	10 (5%)	5 (2%)	26 (7%)	14 (4%)
M1a – no. (%)	10 (16%)	23 (12%)	11 (20%)	6 (11%)	40 (12%)	34 (10%)	24 (13%)	26 (14%)	31 (15%)	19 (9%)	50 (14%)	55 (16%)
M1b – no. (%)	12 (19%)	34 (18%)	5 (9%)	10 (19%)	65 (19%)	62 (18%)	31 (16%)	34 (18%)	32 (15%)	45 (21%)	67 (19%)	61 (17%)

Trial	BREAK-3		BRF113220[a]		BRIM-3		COLUMBUS[b]		COMBI-d		COMBI-v	
M1c – no. (%)	40 (63%)	124 (66%)	37 (69%)	38 (70%)	220 (65%)	221 (66%)	125 (65%)	123 (64%)	138 (65%)	142 (67%)	208 (59%)	221 (63%)
Lactate dehydrogenase levels												
> ULN – no. (%)	19 (30%)	67 (36%)	27 (50%)	22 (41%)	196 (58%)	195 (58%)	52 (27%)	55 (29%)	71 (34%)	77 (37%)	114 (32%)	118 (34%)
≤ ULN – no. (%)	43 (68%)	119 (64%)	27 (50%)	32 (59%)	142 (42%)	142 (42%)	139 (73%)	137 (71%)	140 (66%)	133 (63%)	238 (68%)	233 (66%)

Abbreviations: number (no.); upper limit normal (ULN).
The above information was extracted from the papers hyperlinked in the table above. The Kaplan-Meier curves which were digitized and ultimately used in the NMA are listed in Table 2.
(a) BRF113220 had a third arm, Dabrafenib 150 mg + trametinib 1 mg arm, which is not included here as this was arm was not included in the NMA as this treatment dosing is not given in clinical practice.
(b) COLUMBUS had a third arm, encorafenib monotherapy, which is not included here as this arm was not included in the NMA as this treatment is not given in clinical practice.
(c) Trial protocol states participants must have a $BRAF^{V600E}$ mutation, which authors report was confirmed with genetic testing.

Table 4: Immunotherapy therapy trial population characteristics

Trial	CheckMate 066		CheckMate 067			CheckMate 069		KEYNOTE-006		
Treatment (n)	DTIC (n=208)	Nivolumab (n=210)	Ipilimumab (n=311)	Nivolumab (n=313)	Nivolumab + ipilimumab (n=313)	Ipilimumab (n=47)	Nivolumab + ipilimumab (n=95)	Ipilimumab (n=278)	Pembrolizumab 10mg/kg Q2W (n=279)	Pembrolizumab 10mg/kg Q3W (n=277)
Age, median (range)	66 (26-87)	64 (18-86)	60.8[a] (18-89)	58.7[a] (25-90)	59.3[a] (18-88)	67 (31-80)	64 (27-87)	62 (18-88)	61 (18-89)	63 (22-89)
Sex										
Female – no. (%)	83 (39.9%)	89 (42.4%)	113 (35.9%)	114 (36.1%)	108 (34.4%)	15 (31.9%)	32 (33.7%)	116 (41.7%)	118 (42.3%)	103 (37.2%)
Male – no. (%)	125 (60.1%)	121 (57.6%)	202 (64.1%)	202 (63.9%)	206 (65.6%)	32 (68.1%)	63 (66.3%)	162 (58.3%)	161 (57.7%)	174 (62.8%)
Eastern Cooperative Oncology Group (ECOG) performance status										
0 – no. (%)	121 (58.2%)	148 (70.5%)	224 (71.1%)	238 (75.3%)	230 (73.2%)	37 (78.7%)	79 (83.2%)	188 (67.6%)	196 (70.3%)	189 (68.2%)

Trial	CheckMate 066		CheckMate 067			CheckMate 069		KEYNOTE-006		
1 – no. (%)	84 (40.4%)	60 (28.6%)	91 (28.9%)	77 (24.4%)	83 (26.4%)	10 (21.3%)	14 (14.7%)	90 (32.4%)	83 (29.7%)	88 (31.8%)
BRAF status										
No Mutation – no. (%)	204 (98.1%)	202 (96.2%)	218 (69.2%)	216 (68.4%)	213 (67.8%)	37 (78.7%)	72 (75.8%)	171 (61.5%)	181 (64.9%)	180 (65%)
Mutation – no. (%)	-	-	97 (30.8%)	100 (31.6%)	101 (32.2%)	10 (21.3%)	23 (24.2%)	107 (38.5%)	98 (35.1%)	97 (35%)
Not reported – no. (%)	4 (1.9%)	8 (3.8%)	-	-	-	-	-	-	-	-
Metastasis stage										
M0 – no. (%)	81 (38.9%)	82 (39%)	132 (41.9%)	132 (41.8%)	133 (42.4%)	5 (10.6%)	8 (8.4%)	14 (5%)	9 (3.2%)	9 (3.2%)
M1a – no. (%)						8 (17%)	15 (15.8%)	30 (10.8%)	21 (7.5%)	34 (12.3%)
M1b – no. (%)						12 (25.5%)	27 (28.4%)	52 (18.7%)	64 (22.9%)	41 (14.8%)
M1c – no. (%)	127 (61.1%)	128 (61%)	183 (58.1%)	184 (58.2%)	181 (57.6%)	21 (44.7%)	44 (46.3%)	177 (63.7%)	179 (64.2%)	189 (68.2%)
Lactate dehydrogenase levels										
> ULN – no. (%)	74 (35.6%)	74 (35.6%)	115 (36.5%)	112 (35.4%)	114 (36.3%)	11 (23.4%)	24 (25.3%)	91 (32.7%)	81 (29%)	98 (35.4%)
≤ ULN – no. (%)	125 (60.1%)	120 (57.1%)	194 (61.6%)	196 (62%)	199 (63.4%)	36 (76.6%)	70 (73.7%)	187 (67.3%)	198 (71%)	179 (64.6%)

Abbreviations: number (no.); once every 2 weeks (Q2W); once every 3 weeks (Q3W); upper limit normal (ULN).
The above information was extracted from the papers hyperlinked in the table above. The Kaplan-Meier curves which were digitized and ultimately used in the NMA are listed in Table 2.
(a) CheckMate 067 reports a mean age. Thus, this number is the mean age, not the median as it is in the other columns.

A.4.2 Network diagrams

Six networks were constructed for each analysis, that were defined by the population's *BRAF* status (mutant, wild type, or mixed population), the included comparators and the length of follow-up time of included studies (see Table 1 for a summary of each network).

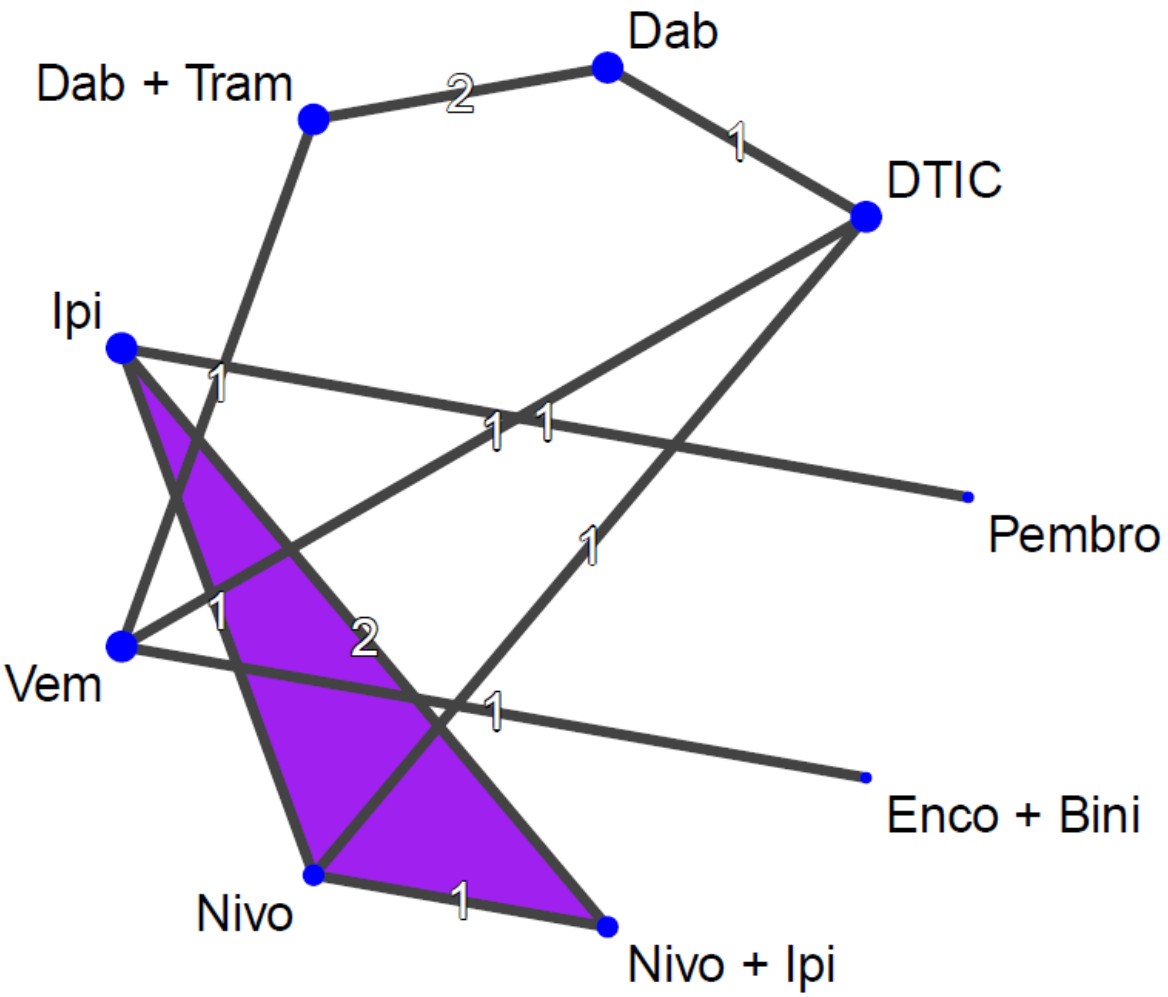
Figure 2 to Figure 7 present the network diagrams for each of our analyses (see Section A.2.2 for descriptions of each analysis). The size of the points corresponds to the number of participants across the trials for the comparator, the number of each of the lines refers to the number of trials that contain a comparison between the two comparators that the line connects, and the purple shading refers to a trial with three arms.

It is important to note that, while the networks for some of these analyses were the same (i.e. Figure 4 and Figure 7), the datasets that fed into the networks differed. In Figure 7, the analysis excluded CheckMate 069 on the basis of insufficient follow-up time and used KM curves from CheckMate 067 and KEYNOTE-006 that were not stratified by *BRAF* subtype, meaning that treatment estimates were calculated from a mixed population of patients with *BRAF* wild type and *BRAF* mutant subtypes. In Figure 4, the analysis excluded CheckMate 069 due to the absence of *BRAF*-specific KM curves, rather than insufficient follow-up and used KM curves from CheckMate 067 and KEYNOTE-006 that were specific to *BRAF* wild type patients only.

For PFS, Network 3 still has one trial (KEYNOTE-006) which has data from a population of all people with melanoma (*BRAF* mutant and wild type), as no PFS curve for only people with *BRAF* wild type melanoma was available.

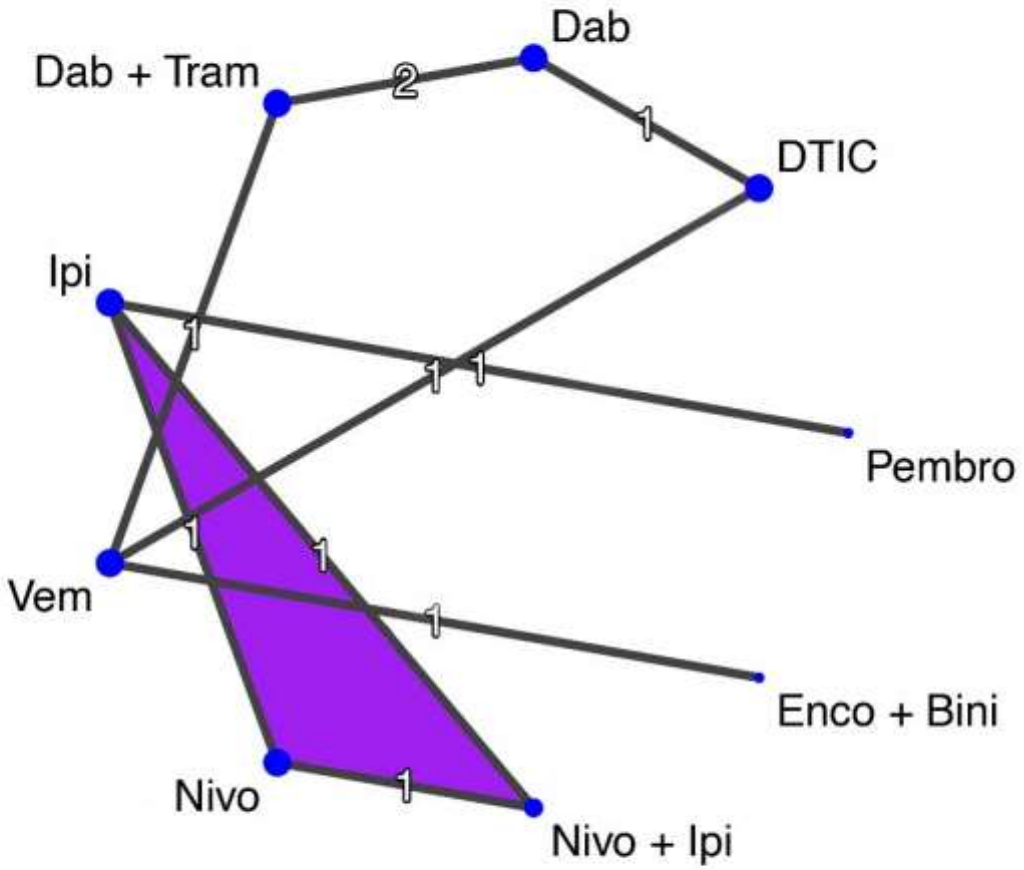
In Network 5 and Network 6, CheckMate 069, providing a comparison between ipilimumab and nivolumab & ipilimumab, is removed from the network due to its short follow-up period. BRIM-3 is also excluded from Network 5; therefore, we lose the connection between DTIC and vemurafenib in this network.

Figure 2: Network 1 – People with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies.



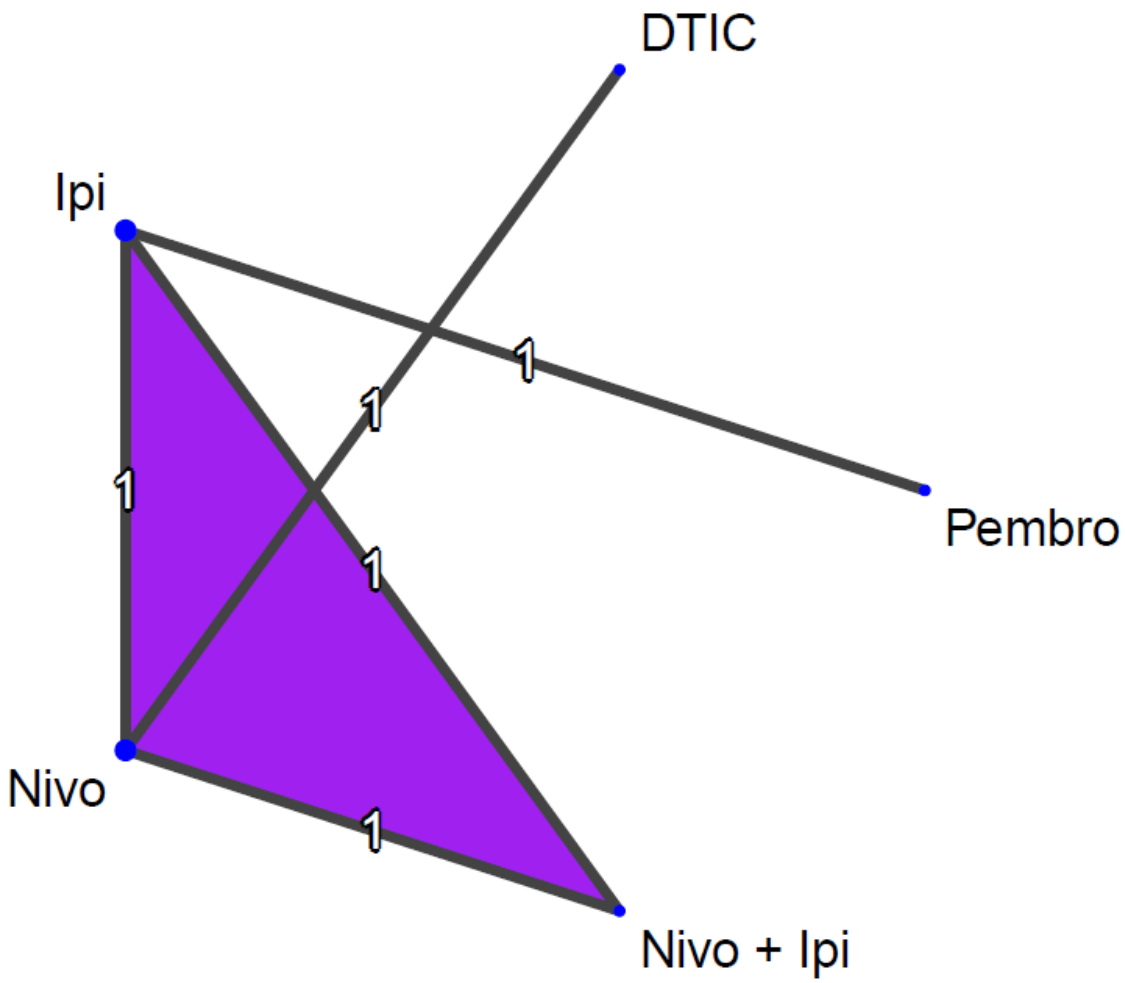
The blue circles are the network nodes, and their size is proportional to the number of studies including each treatment. The numbers on each line represent the number of studies involved in each direct comparison. The region shaded purple indicates a 3-arm trial. Bini = Binimetinib, Dab = Dabrafenib, DTIC = Dacarbazine, Enco = Encorafenib, Ipi = Ipilimumab, Nivo = Nivolumab, Pembro = Pembrolizumab, Tram = Trametinib, Vem = Vemurafenib.

Figure 3: Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies.



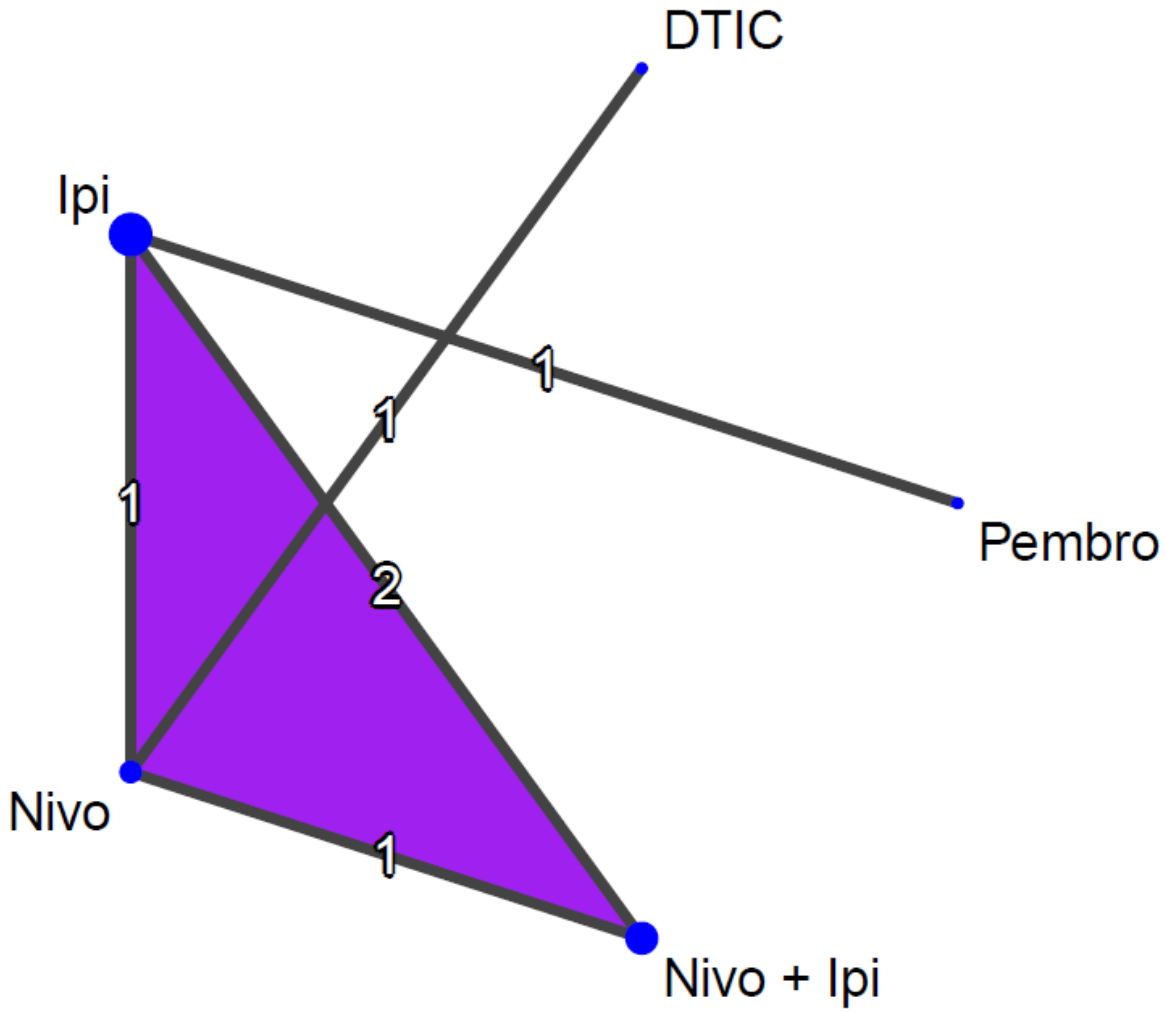
Network 2 differs from Network 1 in that it removes CheckMate 066, which compares DTIC with Nivolumab and CheckMate 069 which compares Ipilimumab with Nivolumab + Ipilimumab. These changes to the network can be seen in two ways. First, there is no longer a line connecting Nivolumab with DTIC. Second, the number on the line connecting Ipilimumab to Nivolumab + Ipilimumab changes from 2 to 1, as now there is only one trial making this comparison. The most significant omission is Checkmate 066 – it's removal from the network means this network is no longer fully connected, but rather consists of two sub-networks, one comprised of the immunotherapies, and one comprised of targeted therapies. It is the comparison between DTIC and Nivolumab in Checkmate 066 which connects the two subnetworks. The loss of this trials results in a broken network comprised of two subnetworks. The blue circles are the network nodes, and their size is proportional to the number of studies including each treatment. The numbers on each line represent the number of studies involved in each direct comparison. The region shaded purple indicates a 3-arm trial. Bini = Binimetinib, Dab = Dabrafenib, DTIC = Dacarbazine, Enco = Encorafenib, Ipi = Ipilimumab, Nivo = Nivolumab, Pembro = Pembrolizumab, Tram = Trametinib, Vem = Vemurafenib.

Figure 4: Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only.



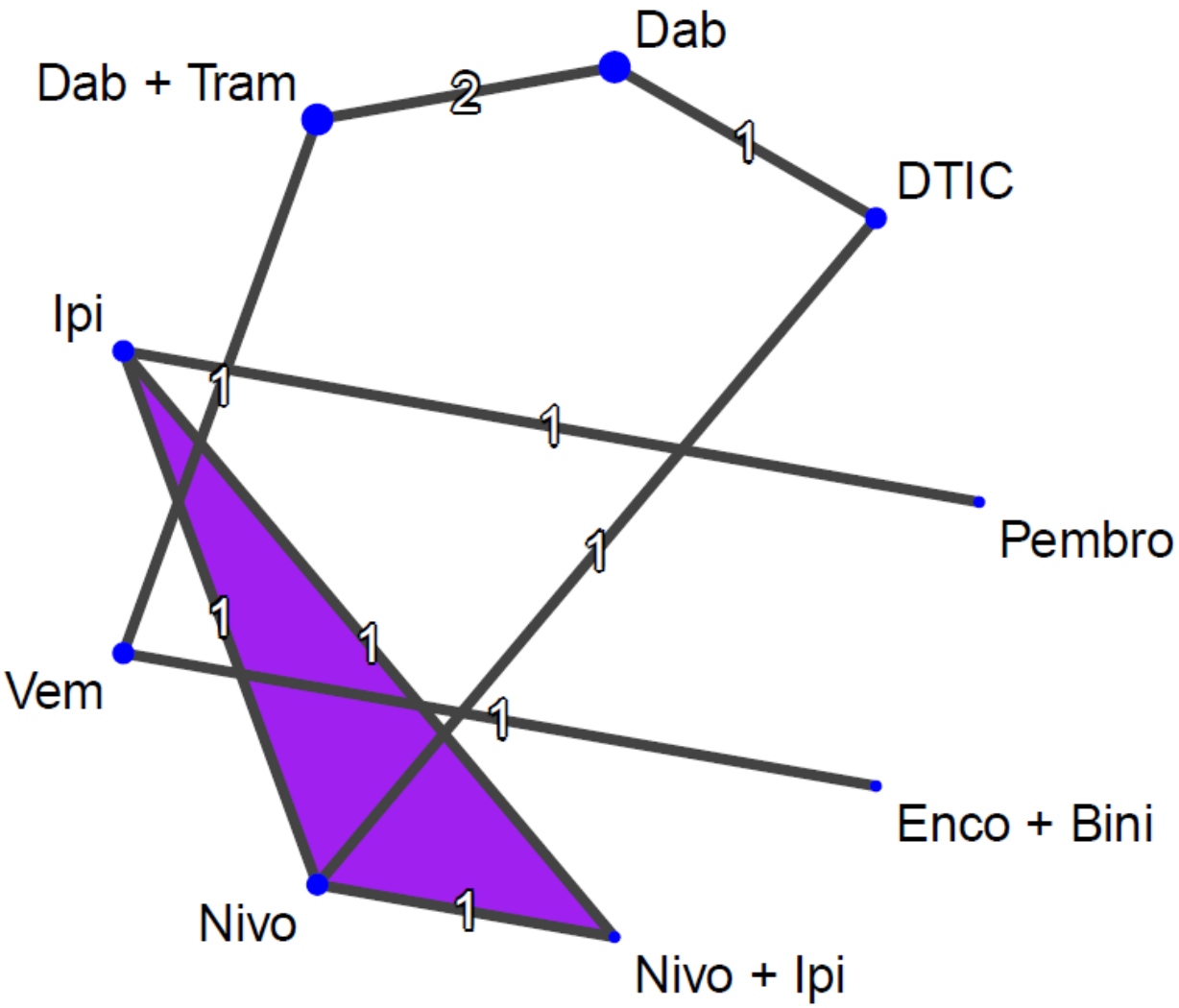
Network 3 consists of people with BRAF wild type melanoma only. However, KEYNOTE-006 only had a BRAF wildtype KM curve for OS. As such, the PFS analysis for network 3 includes one trial (KEYNOTE-006) that is based on a mixed population of people with BRAF wild type melanoma and people with BRAF mutant melanoma. The blue circles are the network nodes, and their size is proportional to the number of studies including each treatment. The numbers on each line represent the number of studies involved in each direct comparison. The region shaded purple indicates a 3-arm trial. DTIC = Dacarbazine, Ipi = Ipilimumab, Nivo = Nivolumab, Pembro = Pembrolizumab.

Figure 5: Network 4 - All people with melanoma (*BRAF* mutant and wild type), with immunotherapy strategies only.



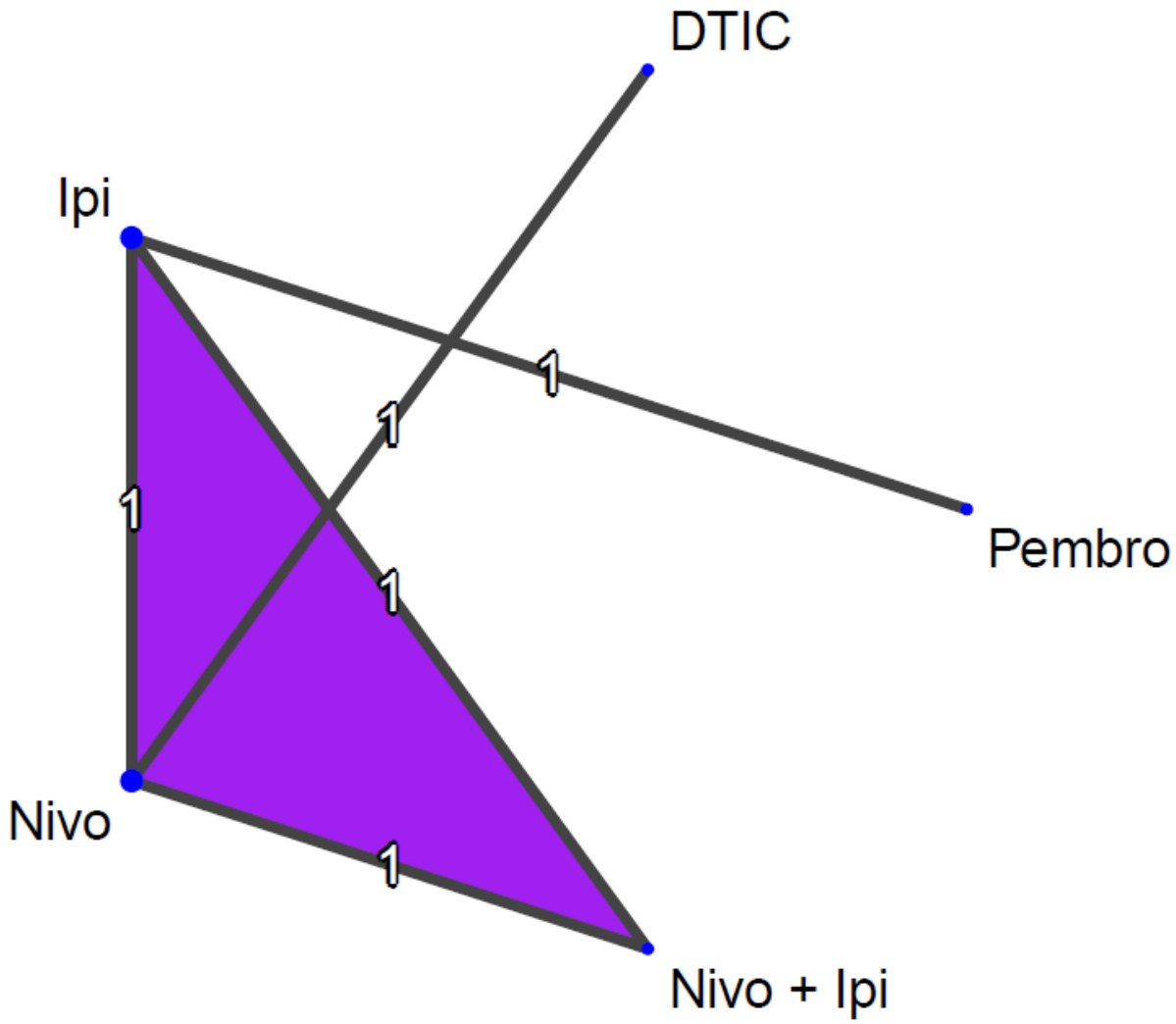
Network 4 differs from Network 3 in two ways. First, it includes CheckMate 069. CheckMate 069 was excluded from network 3 as its population is mixed, that is it includes both people with BRAF wild type melanoma and people with BRAF mutant melanoma. No KM curve was available for CheckMate 069 for people with BRAF wild type melanoma alone, only the curve for the mixed population was available. Second, the KM curves for CheckMate 067 and KEYNOTE-006 utilized in this network are of mixed populations, that is they include people with BRAF wild type and BRAF mutant melanoma. The blue circles are the network nodes, and their size is proportional to the number of studies including each treatment. The numbers on each line represent the number of studies involved in each direct comparison. The region shaded purple indicates a 3-arm trial. DTIC = Dacarbazine, Ipi = Ipilimumab, Nivo = Nivolumab, Pembro = Pembrolizumab.

Figure 6: People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up.



Network 5 differs from network 1 in that it drops BRIM-3, which compared Vemurafenib monotherapy against DTIC and CheckMate 069, which compared Ipilimumab monotherapy against Nivolumab + Ipilimumab from the analysis. Both trials had shorter follow-up. Namely, BRIM-3 was limited to 22 months of follow-up for PFS, and CheckMate 069 was limited to 24 months of follow-up for both PFS and OS. These changes to the network can be seen in two ways. First, there is no longer a line connecting Vemurafenib with DTIC. Second, the number on the line connecting Ipilimumab to Nivolumab + Ipilimumab changes from 2 to 1, as now there is only one trial making this comparison. The blue circles are the network nodes, and their size is proportional to the number of studies including each treatment. The numbers on each line represent the number of studies involved in each direct comparison. The region shaded purple indicates a 3-arm trial. Bini = Binimetinib, Dab = Dabrafenib, DTIC = Dacarbazine, Enco = Encorafenib, Ipi = Ipilimumab, Nivo = Nivolumab, Pembro = Pembrolizumab, Tram = Trametinib, Vem = Vemurafenib.

Figure 7: People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up.

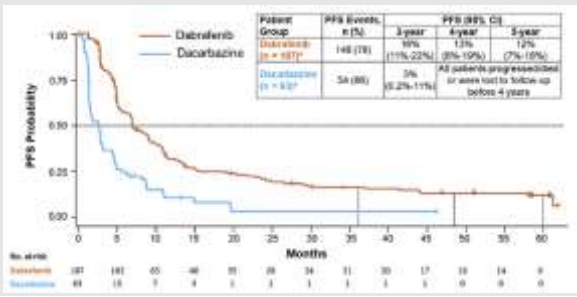
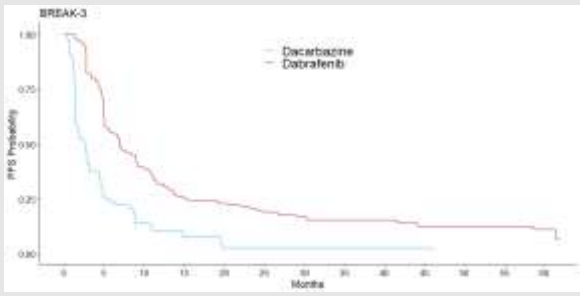
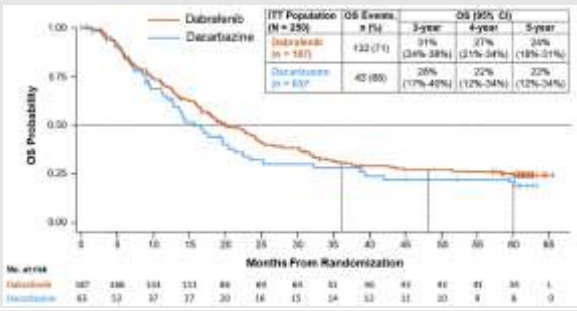
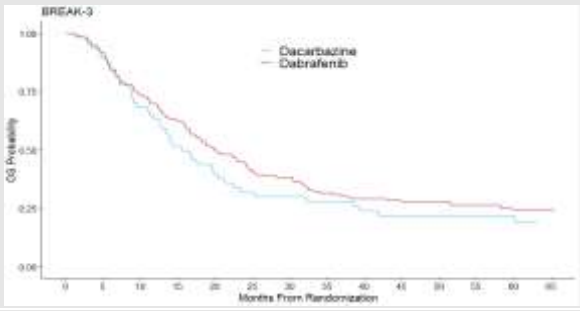
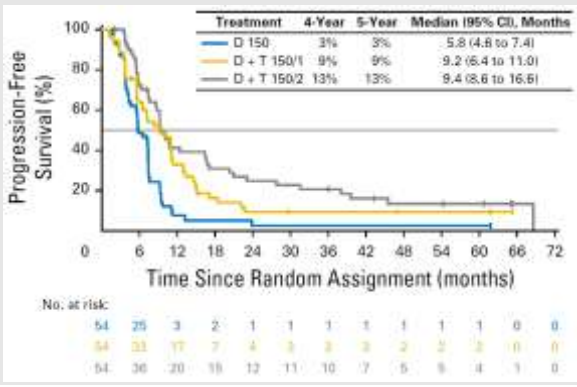
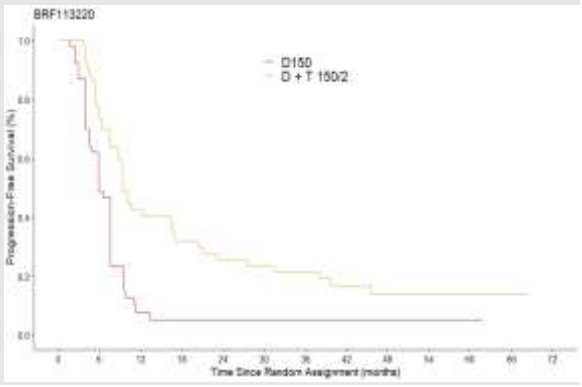
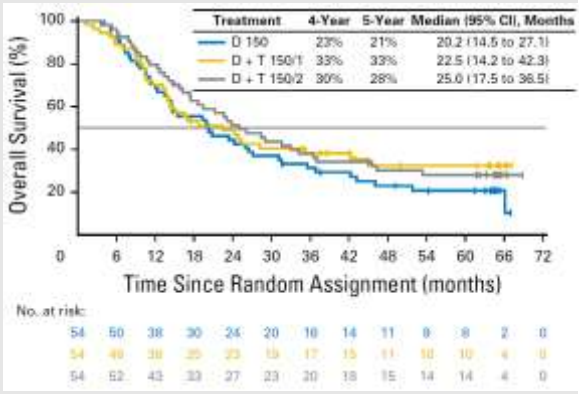
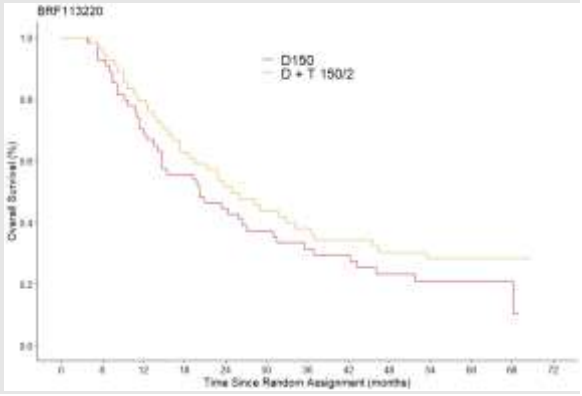


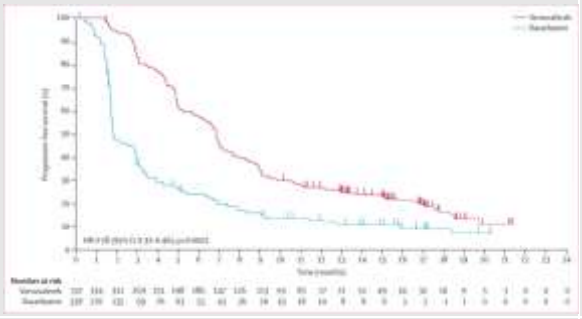
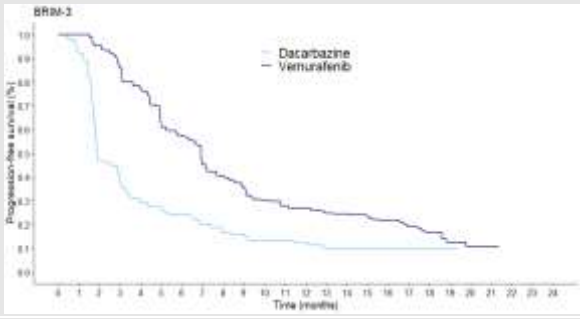
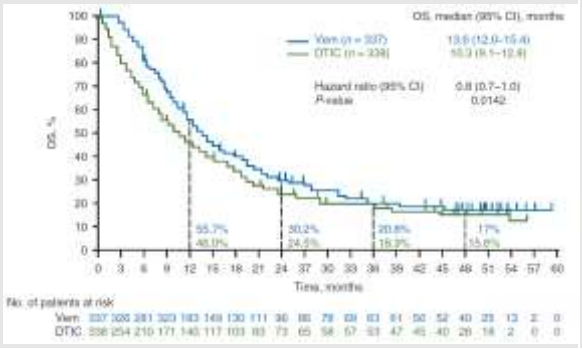
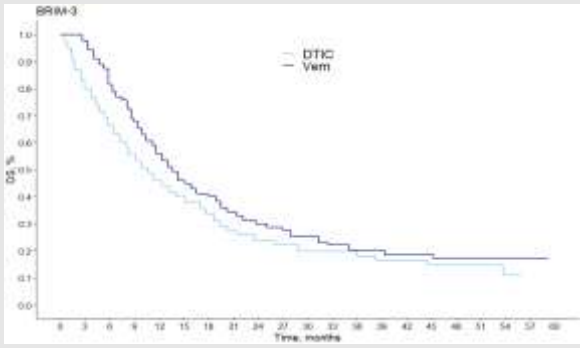
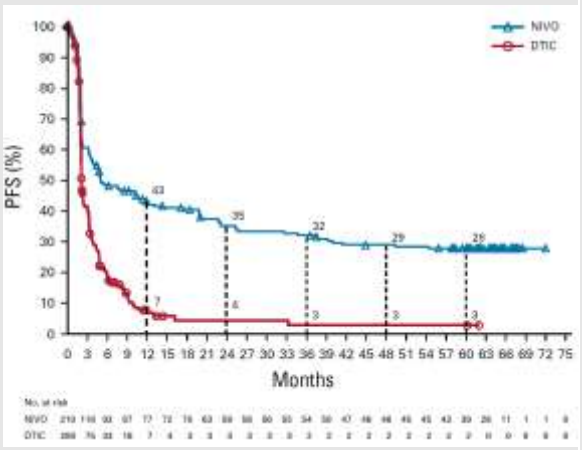
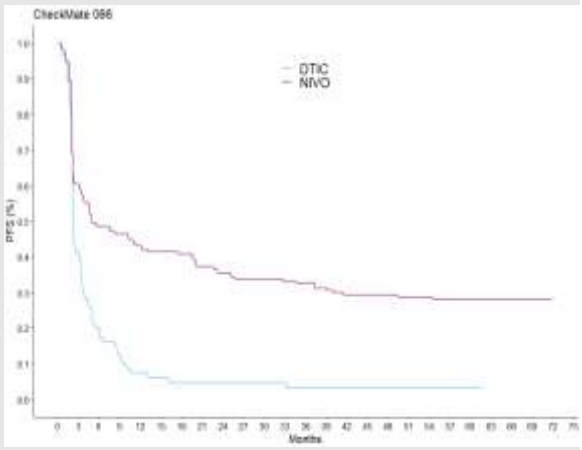
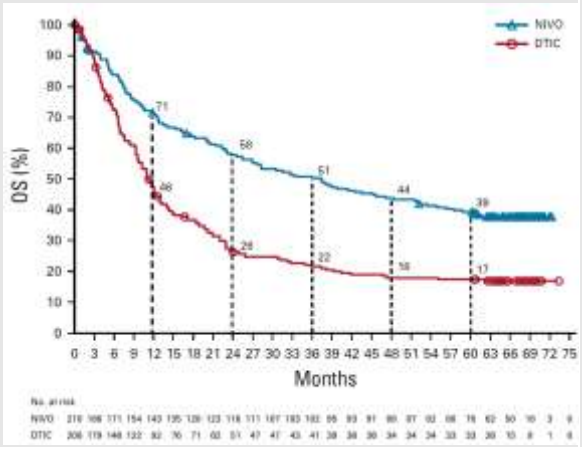
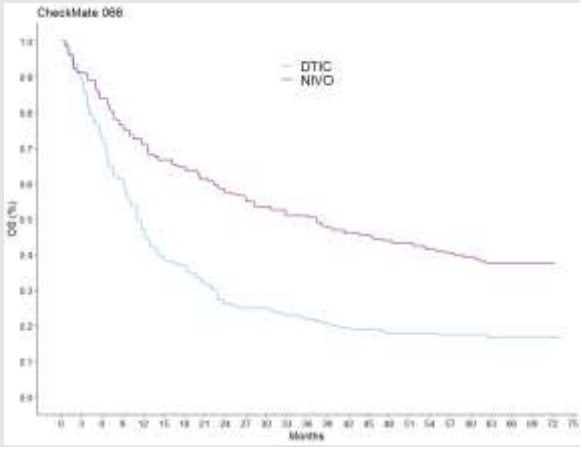
Though identical in shape and the trials included network 6 differs from network 3 in that it is a mixed population. Specifically, network 6 uses a KM curve for CheckMate 067 and KEYNOTE-006 that is comprised of a mixed population, people with BRAF wild type melanoma and people with BRAF mutant melanoma. Like network 3, CheckMate 069 is excluded from this network, however on different grounds. Where network 3 excludes CheckMate 069 because it only provides KM curves that are of a mixed population, network 6 excludes CheckMate 069 because of its limited follow-up. Blue circles are the network nodes, and their size is proportional to the number of studies including each treatment. The numbers on each line represent the number of studies involved in each direct comparison. The region shaded purple indicates a 3-arm trial. DTIC = Dacarbazine, Ipi = Ipilimumab, Nivo = Nivolumab, Pembro = Pembrolizumab.

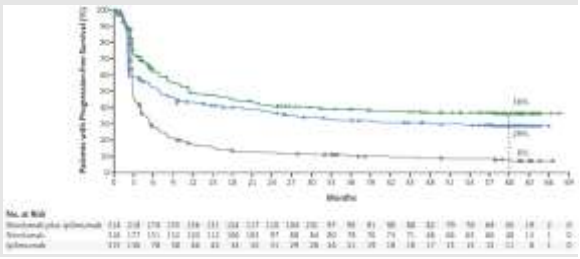
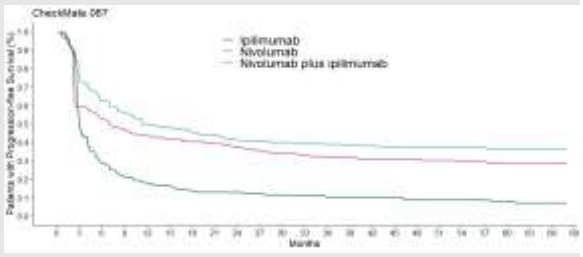
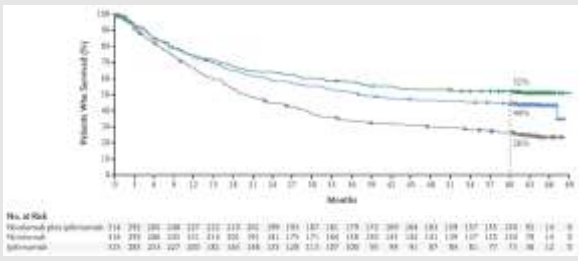
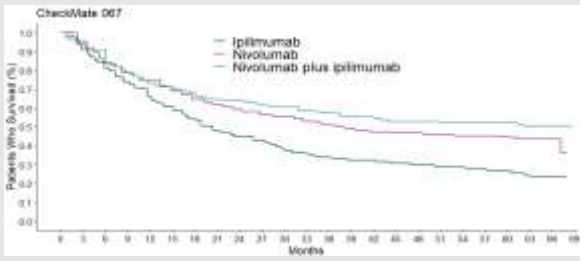
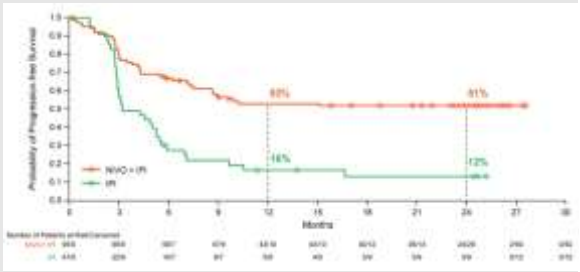
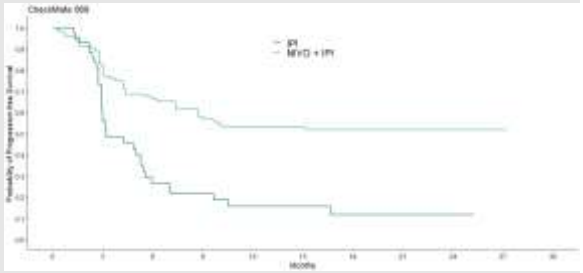
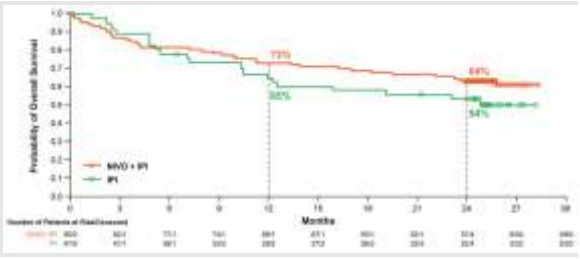
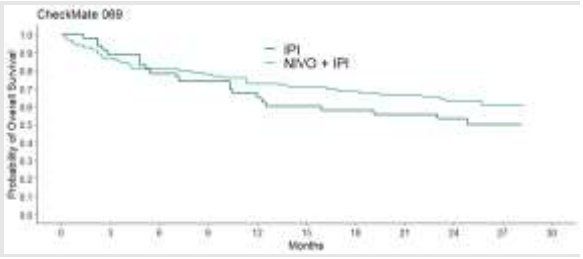
A.4.3 Extraction and digitizing

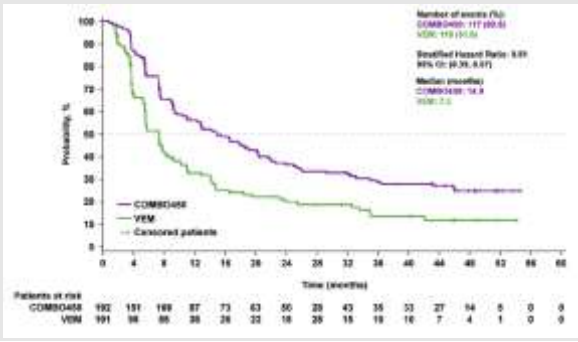
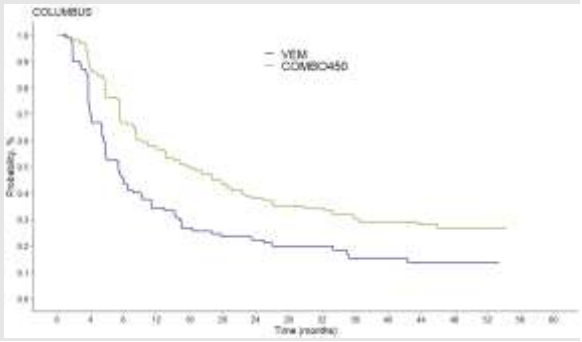
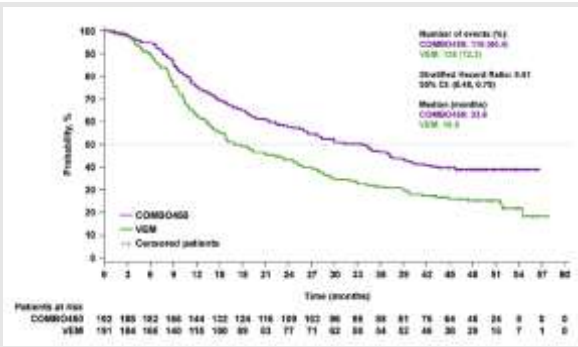
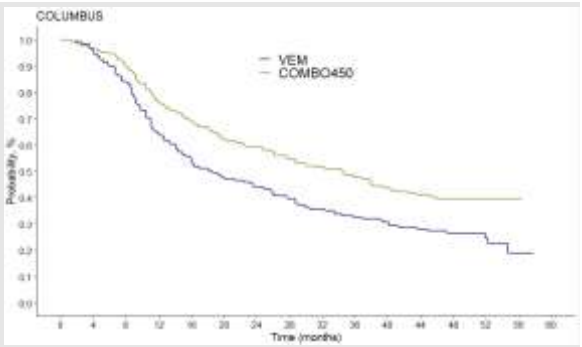
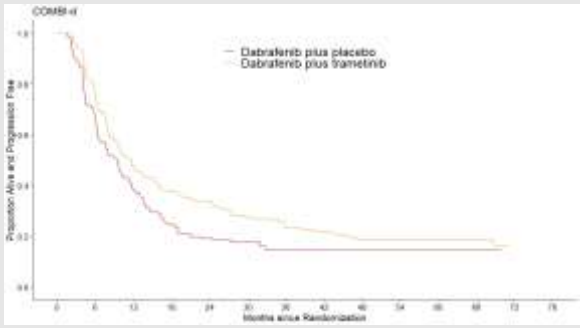
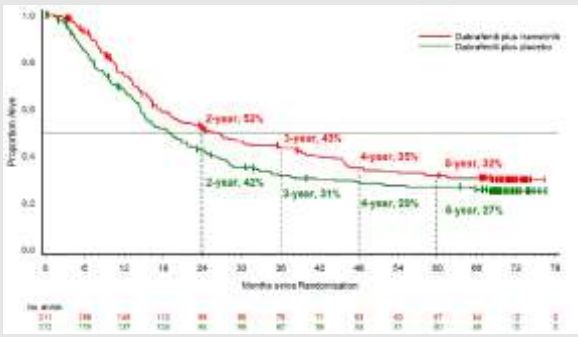
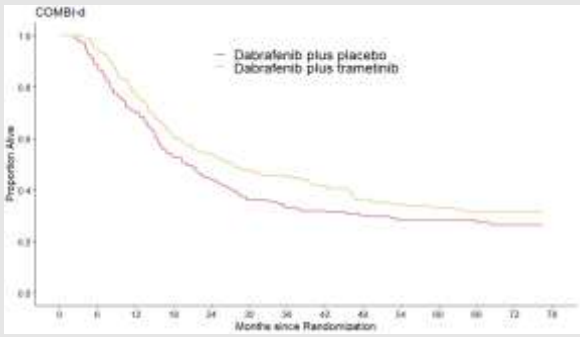
Presented below are the extracted KM curves from the included trials and the digitized versions of them.

Table 5: Extracted and digitized KM curves included in the networks

	Extracted	Digitized
BREAK-3		
PFS		
	Not published	HR: 0.3706 (0.2709-0.5071), p=9e-09
OS		
	Not published	HR: 0.8299 (0.5879-1.171), p=0.3
BRF113220		
PFS		
	HR: 0.44 (0.28-0.67), no p-value	HR: 0.4253 (0.2732-0.6622), p=1e-04
OS		
	HR: 0.76 (0.49-1.18), no p-value	HR: 0.7631 (0.4929-1.181), p=0.2
BRIM-3		

	Extracted	Digitized
PFS		
	HR: 0.38 (0.32-0.46), p<0.0001	HR: 0.4044 (0.3369-0.4854), p=<2e-16
OS		
	HR: 0.81 (0.67-0.98), p=0.03	HR: 0.7932 (0.6693-0.9402), p=0.008
CheckMate 066		
PFS		
	HR: 0.4 (0.33-0.54), p<0.0001	HR: 0.4433 (0.3486-0.5637), p=2e-11
OS		

	Extracted	Digitized
	HR: 0.5 (0.40-0.63), p<0.0001	HR: 0.5048 (0.3988-0.6388), p1e-08
CheckMate 067		
PFS		
	HR Nivo vs. Ipi: 0.53 (0.44-0.64), p<0.001 HR N+I vs. Ipi: 0.42 (0.35-0.51), p<0.001 HR N+I vs. Nivo: 0.79 (0.64-0.96) no p-value	HR Nivo vs. Ipi: 0.59 (0.49-0.71), p=3.28e-08 HR N+I vs. Ipi: 0.43 (0.36-0.53), p=<2e-16 HR N+I vs. Nivo: 0.73 (0.60-0.90) p=0.00227
OS		
	HR Nivo vs. Ipi: 0.63 (0.52-0.76), p<0.001 HR N+I vs. Ipi: 0.52 (0.42-0.64), p<0.001 HR N+I vs. Nivo: 0.83 (0.67-1.03) no p-value	HR Nivo vs. Ipi: 0.63 (0.52-0.77), p=5.99e-06 HR N+I vs. Ipi: 0.53 (0.43-0.65), p=2.06e-09 HR N+I vs. Nivo: 0.83 (0.67-1.04) p=0.107
CheckMate 069		
PFS		
	HR: 0.36 (0.22-0.56), p<0.0001	HR: 0.3602 (0.2263-0.5735), p=3e-05
OS		
	HR: 0.74 (0.43-1.26), p=0.26	HR: 0.7454 (0.4381-1.268), p=0.279
COLUMBUS		

	Extracted	Digitized
PFS		
	HR: 0.54 (0.41-0.71), no p-value	HR: 0.557 (0.427-0.7264), p=2e-05
OS		
	HR: 0.61 (0.48-0.79), no p-value	HR: 0.6534 (0.5077-0.8411), p=9e-04
COMBI-d		
PFS		
	Not published	HR: 0.7256 (0.5752-0.9152), p=0.007
OS		
	Not published	HR: 0.8007 (0.6301-1.017), p=0.0688
COMBI-v		

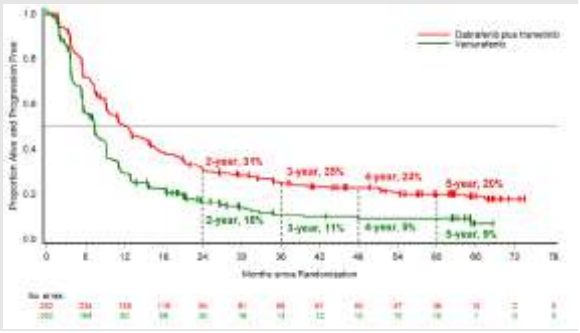
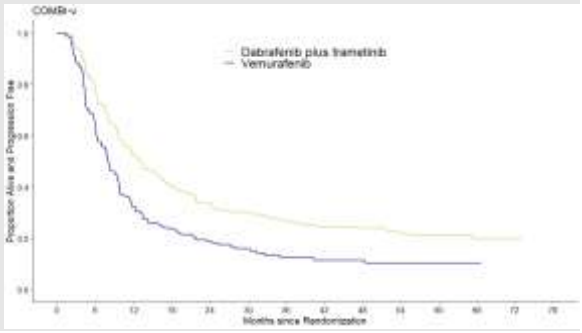
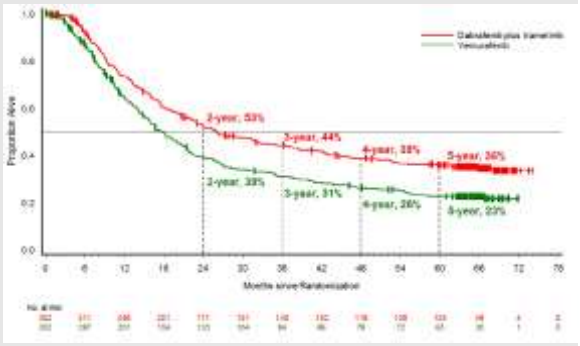
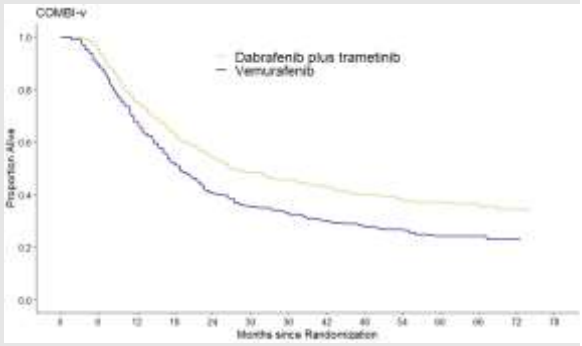
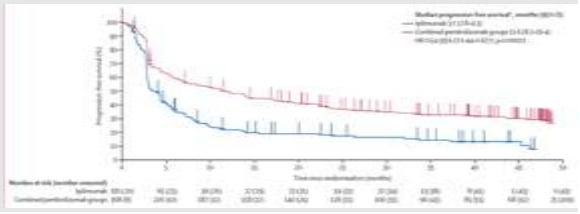
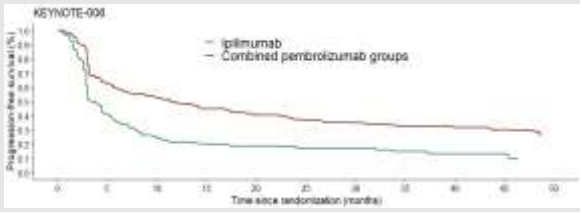
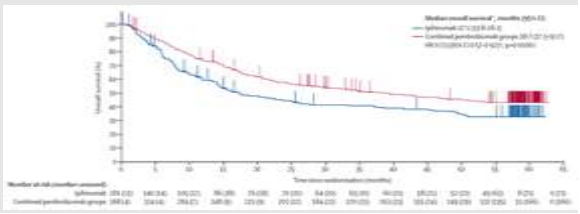
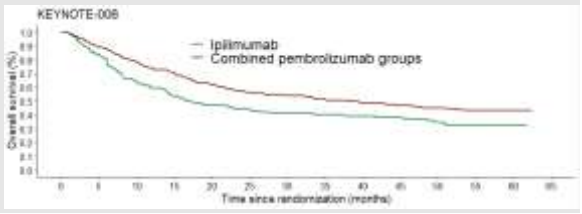
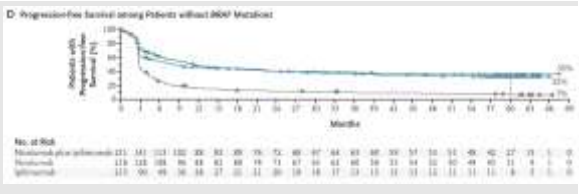
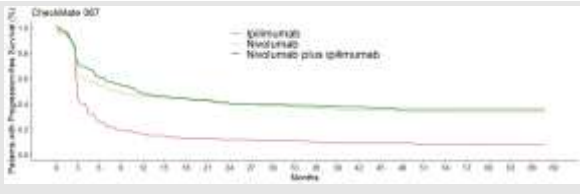
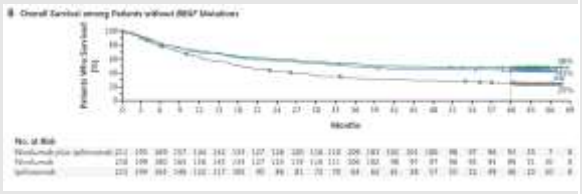
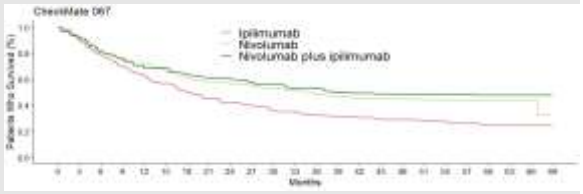
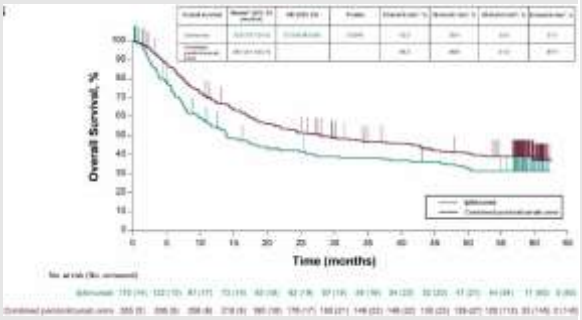
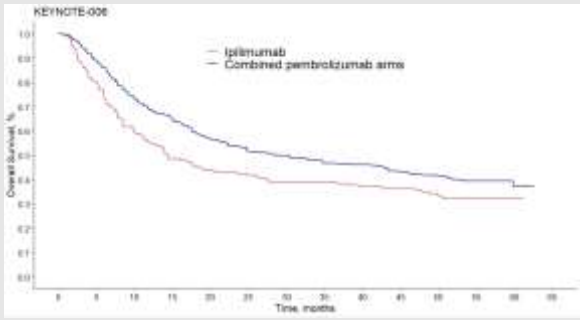
	Extracted	Digitized
PFS		
	Not published	HR: 0.6162 (0.5122-0.7412), p=3e-07
OS		
	Not published	HR: 0.7044 (0.5833-0.8506), p=3e-04
KEYNOTE-006		
PFS		
	HR: 0.54 (0.44-0.67), p=0.0001	HR: 0.5136 (0.4161-0.634), p=2e-09
OS		
	HR: 0.73 (0.57-0.92), p=0.0036	HR: 0.7022 (0.5563-0.8865), p=0.004

Table 6: Extracted and digitized KM curves included in the BRAF wildtype scenario analysis network

	Extracted	Digitized
CheckMate 067		
PFS		

	Extracted	Digitized
	Not published	HR Nivo vs. Ipi: 0.48 (0.38-0.60), p=1.46e-10 H N+I vs Ipi: 0.43 (0.34-0.54), p=1.07e-12 H N+I vs Nivo: 0.90 (0.71-1.15), p=0.415
OS		
	Not published	HR Nivo vs. Ipi: 0.66 (0.52-0.83), p=0.000531 H N+I vs Ipi: 0.59 (0.46-0.76), p=3.3e-05 H N+I vs Nivo: 0.90 (0.70-1.17), p=0.43
	KEYNOTE-006	
PFS	No BRAF wild type curve available	No digitized curve as there was no BRAF wild type curve available
OS		
	NP	HR: 0.745 (0.5884-0.9432), p=0.0145

A.5 Results

A.5.1 Model selection and validation

A summary of assessment of proportional hazards (PH) for each PFS and OS curve for each trial in the network is presented in Appendix C:. In most trials, the PH assumption was not met. As such, Cox PH models were not considered further. Additionally, due to the variation in follow-up between trials and the need for external data to generate extrapolations, the RMST model was ruled out a priori. Finally, the Royston-Parmar model was not run based on the results of the fractional polynomial model which did not result in an improved fit but resulted in increased model complexity.

We fit fractional polynomial, piecewise exponential, and generalised gamma models to PFS and OS data for each network. The process for selecting the best fitting PFS model is discussed in Section A.5.1.1 and the results of the model are presented in Section A.5.1.2. The process for selecting the best fitting OS model is discussed in Section A.5.2 and the results of the model are presented in Section A.5.3.

A.5.1.1 Progression-free survival

A summary of each model fit and why it was ultimately not selected as the best fitting model is presented in Table 7.

As seen in Appendix E.1 (Table 62, Table 63, Table 64, Table 65, Table 66), the best fitting PFS model (according to the smallest DIC value) for each network overall was the 2-cut point piecewise exponential model with cut points at 12 & 18 months (except Network 5, which had a better fit for a 2-cut point piecewise exponential model with cut points at 12 & 24 months – although this gives improbable results). This selection was based on discussions with the committee about their experience with the nature of the hazards associated with melanoma anticancer treatments over time, the availability of data in each trial for each time interval and the model fit. The fit statistics for each model in each network are presented in Appendix E.1.

We did not use a later cut-point in the piecewise model because for the long-term section of the curves (>20 months) there is insufficient data to estimate the relative effects, specifically the link connecting dabrafenib with DTIC. The DTIC arm of the BREAK-3 trial has no events occurring after 20 months, which in turn makes it impossible to run the NMA. To run a piecewise exponential NMA one needs to have events in all arms of the trials for each period of time over which data is aggregated. Furthermore, removing this trial is also not an option as the only trial connecting the targeted treatments to the immunotherapies in the network would be BRIM-3, which also is limited in its follow-up for PFS, having only 22 months of follow-up.

The fractional polynomial models were ruled out due to implausible PFS predictions. As shown in Figure 110-Figure 114 (Appendix F.1.1.8), these models frequently predicted that progression-free survival for the systemic cancer treatments was worse than DTIC, and at times was 0.

The generalized gamma model with two treatment effects (location and shape parameter) was deemed to be a poor fit to the data (Figure 8-Figure 12: most notable in Figure 9 encorafenib + binimetinib, Figure 10 nivolumab monotherapy, and Figure 11 nivolumab + ipilimumab where the observed KM data begins to plateau, however the generalized gamma predictions continue to plummet only plateauing much later than what is observed in the KM data). To validate this, the subcommittee was shown a selection of predictions for the generalized gamma model compared with piecewise models with one cut point and two cut points. The subcommittee agreed that the generalized gamma model was a poor fit to the observed KM trial data and had incredibly pessimistic extrapolations that they didn't believe

were realistic. As such, the subcommittee agreed with our assessment that the generalized gamma model was not the best fitting PFS model and could be discarded.

The committee therefore decided between the different piecewise exponential models that had been fit. The committee was presented with 3 models with 1 cut point (6 months, 9 months, 12 months, 15 months), and 6 models with 2 cut points (6 & 12 months, 9 & 15 months, 9 & 18 months, 12 & 18 months, 12 & 20 months, and 12 & 24 months).

While the committee noted the single cut point models was preferred for certain treatments (encorafenib + binimetinib), the two-cut point model was preferred for others (dabrafenib + trametinib, nivolumab + ipilimumab, nivolumab, pembrolizumab). However, the committee noted that the predictions from these models still lacked clinical plausibility. Based on DIC, the best fitting model with one cut point was cut point at 15 months, and the best fitting 2 cut point model was the model with cut points at 12 & 18 months.

The subcommittee was then presented with these additional analyses where they decided that the best fitting PFS model overall was the 2-cut point model with cut points at 12 & 18 months.

Table 7: Summary table of PFS models fit and rationale for non-selection

Model	Network					
	1	2	3	4	5	6
Generalized gamma						
Location parameter alone	> AIC compared with Location and Scale Parameter	NA	> AIC compared with Location and Scale Parameter	> AIC compared with Location and Scale Parameter	> AIC compared with Location and Scale Parameter	> AIC compared with Location and Scale Parameter
Location and scale parameter	Lacked clinical plausibility compared with the piecewise 2 cut point model	NA	Lacked clinical plausibility compared with the piecewise 2 cut point model	Lacked clinical plausibility compared with the piecewise 2 cut point model	Lacked clinical plausibility compared with the piecewise 2 cut point model	Lacked clinical plausibility compared with the piecewise 2 cut point model
Piecewise exponential						
1 cut – 6 months	> DIC compared with 1 cut – 15 months	NA	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months
1 cut – 9 months	> DIC compared with 1 cut – 15 months	NA	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months
1 cut – 12 months	> DIC compared with 1 cut – 15 months	NA	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 15 months
1 cut – 15 months	Although less complex, it lacked clinical plausibility compared	NA	Although less complex, it lacked clinical plausibility compared	Although less complex, it lacked clinical plausibility compared	Although less complex, it lacked clinical plausibility compared	Although less complex, it lacked clinical plausibility compared

Model	Network					
	1	2	3	4	5	6
	with the 2 cut – 12 & 18 month model		with the 2 cut – 12 & 18 month model	with the 2 cut – 12 & 18 month model	with the 2 cut – 12 & 24 month model	with the 2 cut – 12 & 18 month model
2 cuts – 6 & 12 months	> DIC compared with 2 cut – 12 & 18 months	NA	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 24 months	> DIC compared with 2 cut – 12 & 18 months
2 cuts – 9 & 15 months	> DIC compared with 2 cut – 12 & 18 months	NA	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 24 months	> DIC compared with 2 cut – 12 & 18 months
2 cuts – 9 & 18 months	> DIC compared with 2 cut – 12 & 18 months	NA	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 24 months	> DIC compared with 2 cut – 12 & 18 months
2 cuts – 12 & 18 months	Selected as best fitting model	NA	Selected as best fitting model	Selected as best fitting model	> DIC compared with 2 cut – 12 & 24 months	Selected as best fitting model
2 cuts – 12 & 20 months	NA	NA	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 24 months	> DIC compared with 2 cut – 12 & 18 months
2 cuts – 12 & 24 months	NA	NA	> DIC compared with 2 cut – 12 & 18 months	NA	Selected as best fitting model	> DIC compared with 2 cut – 12 & 18 months
3 cuts – 6, 12 & 18 months	Increased model complexity without a significantly better fit compared with 2 cuts – 12 & 18 months	NA	> DIC compared with 3 cut – 12, 24 & 36	Increased model complexity without a significantly better fit compared with 2 cuts – 12 & 18 months	> DIC compared with 3 cut – 12, 24 & 36	> DIC compared with 3 cut – 12, 24 & 36
3 cuts – 12, 24 & 36 months	NA	NA	Increased model complexity without a significantly better fit compared with 2 cuts – 12 & 18 months	NA	Increased model complexity without a significantly better fit compared with 2 cuts – 12 & 18 months	Increased model complexity without a significantly better fit compared with 2 cuts – 12 & 18 months
Fractional polynomial 8 interval model						

Model	Network					
	1	2	3	4	5	6
-2	> DIC compared with -0.5	NA	> DIC compared with 0	Poor fit and high model complexity compared with piecewise and gen gamma	> DIC compared with -0.5	> DIC compared with -0.5
-1	> DIC compared with -0.5	NA	> DIC compared with 0	Did not converge	Did not converge	> DIC compared with -0.5
-0.5	Poor fit and high model complexity compared with piecewise and gen gamma	NA	> DIC compared with 0	Did not converge	Poor fit and high model complexity compared with piecewise and gen gamma	Poor fit and high model complexity compared with piecewise and gen gamma
0	Did not converge	NA	Poor fit and high model complexity compared with piecewise and gen gamma	Did not converge	Did not converge	Did not converge
0.5	Did not converge	NA	> DIC compared with 0	Did not converge	Did not converge	> DIC compared with -0.5
1	Did not converge	NA	Did not converge	Did not converge	Did not converge	Did not converge
2	Did not converge	NA	Did not converge	Did not converge	Did not converge	Did not converge
3	> DIC compared with -0.5	NA	> DIC compared with 0	> DIC compared with -2	> DIC compared with -0.5	> DIC compared with -0.5

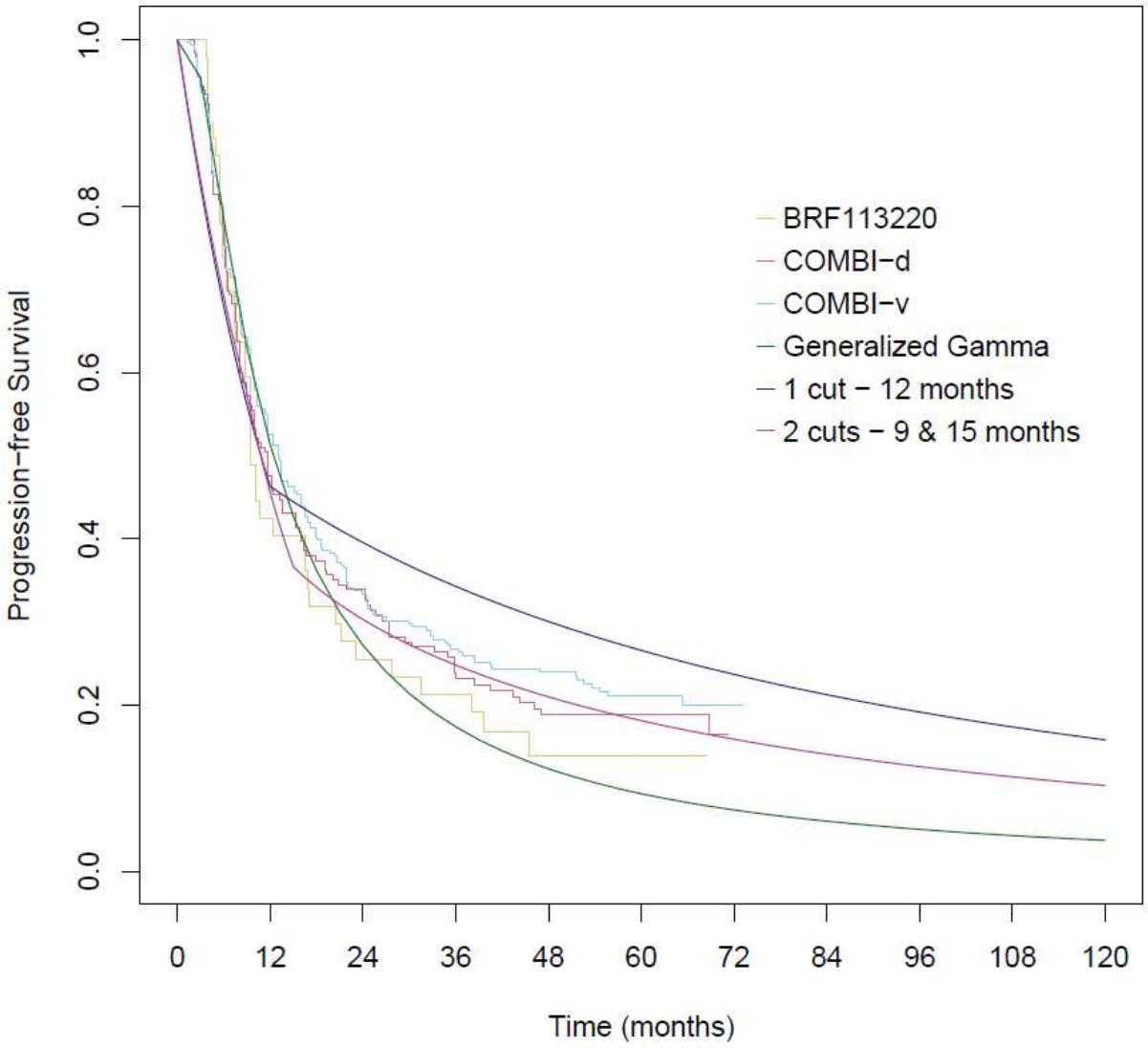


Figure 8: Predicted PFS for generalized gamma two treatment effect, piecewise exponential 1 cut at 12 months, piecewise exponential with 2 cuts 9 & 15 months, compared with observed PFS data for dabrafenib + trametinib

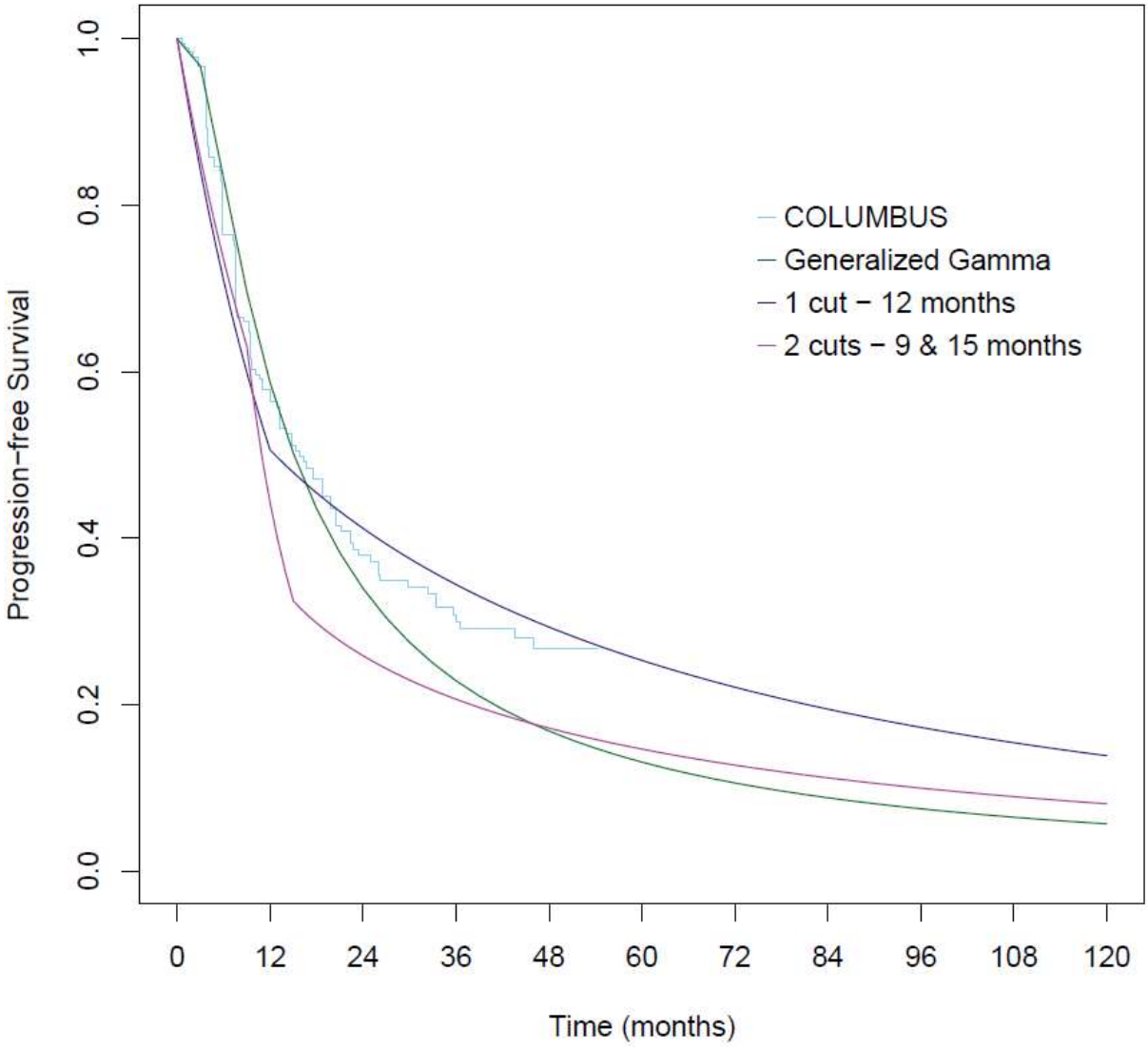


Figure 9: Predicted PFS for generalized gamma two treatment effect, piecewise exponential 1 cut at 12 months, piecewise exponential with 2 cuts 9 & 15 months, compared with observed PFS data for encorafenib + binimetinib

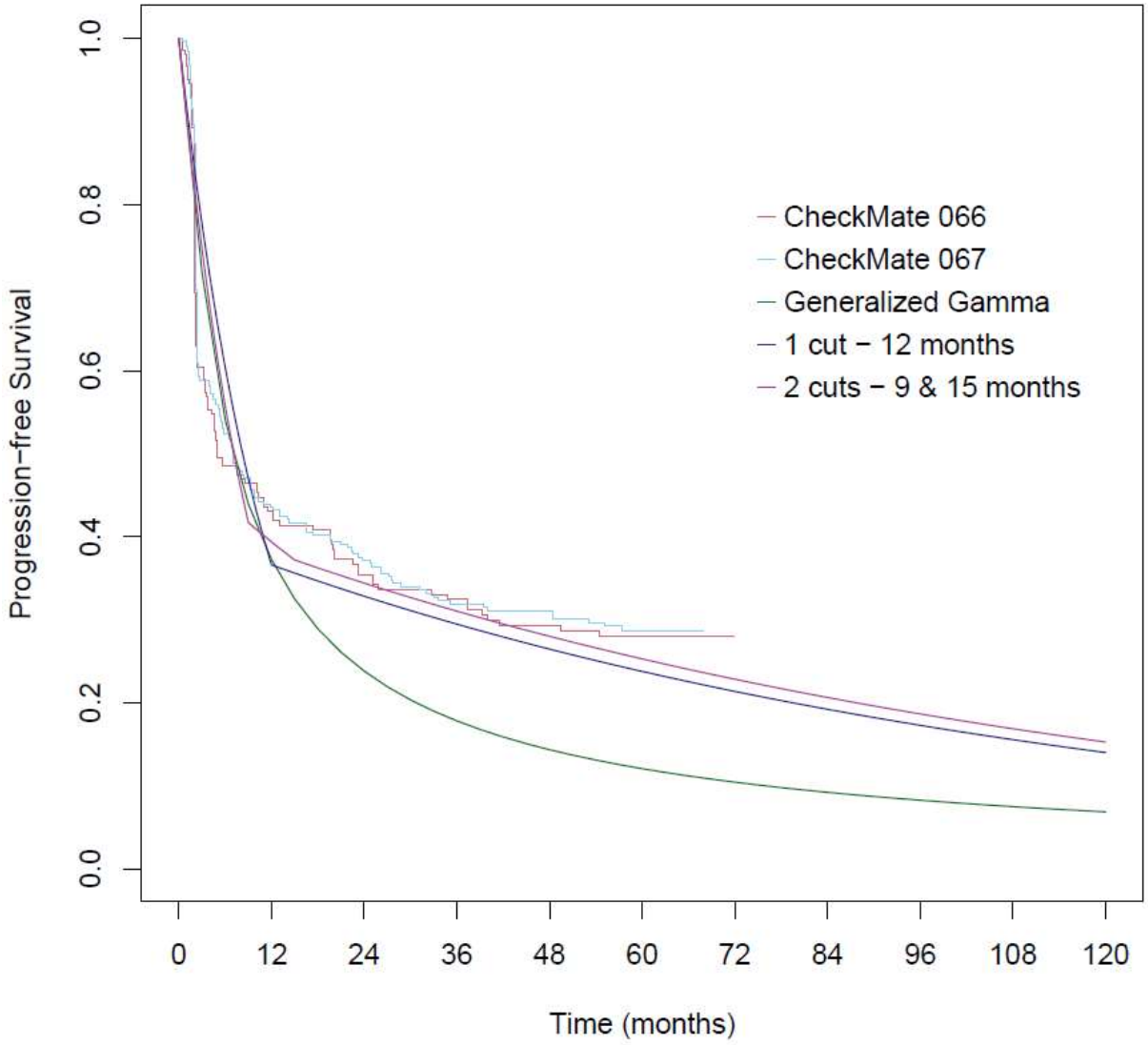


Figure 10: Predicted PFS for generalized gamma two treatment effect, piecewise exponential 1 cut at 12 months, piecewise exponential with 2 cuts 9 & 15 months, compared with observed PFS data for nivolumab

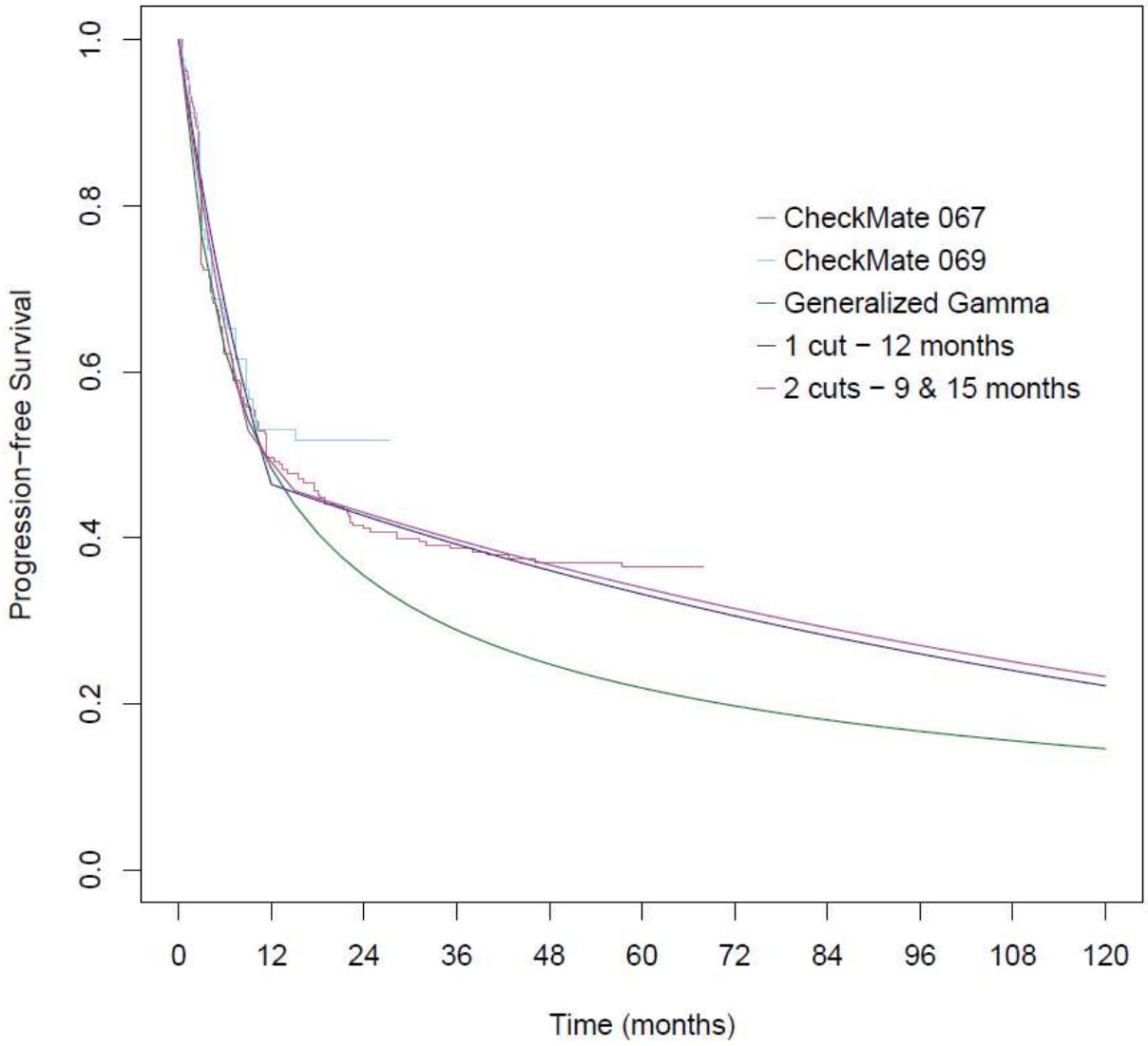


Figure 11: Predicted PFS for generalized gamma two treatment effect, piecewise exponential 1 cut at 12 months, piecewise exponential with 2 cuts 9 & 15 months, compared with observed PFS data for nivolumab + ipilimumab

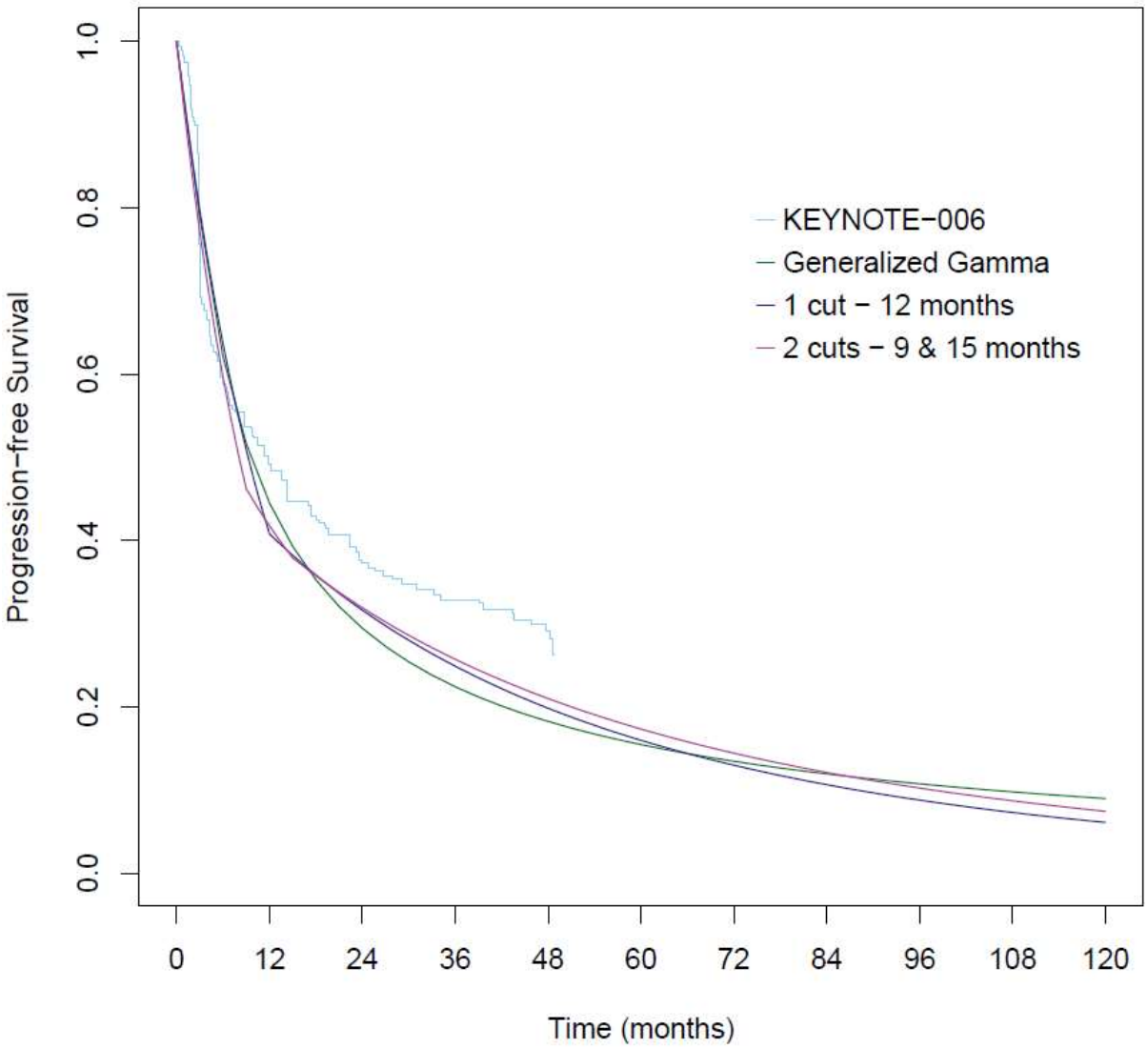


Figure 12: Predicted PFS for generalized gamma two treatment effect, piecewise exponential 1 cut at 12 months, piecewise exponential with 2 cuts 9 & 15 months, compared with observed PFS data for pembrolizumab

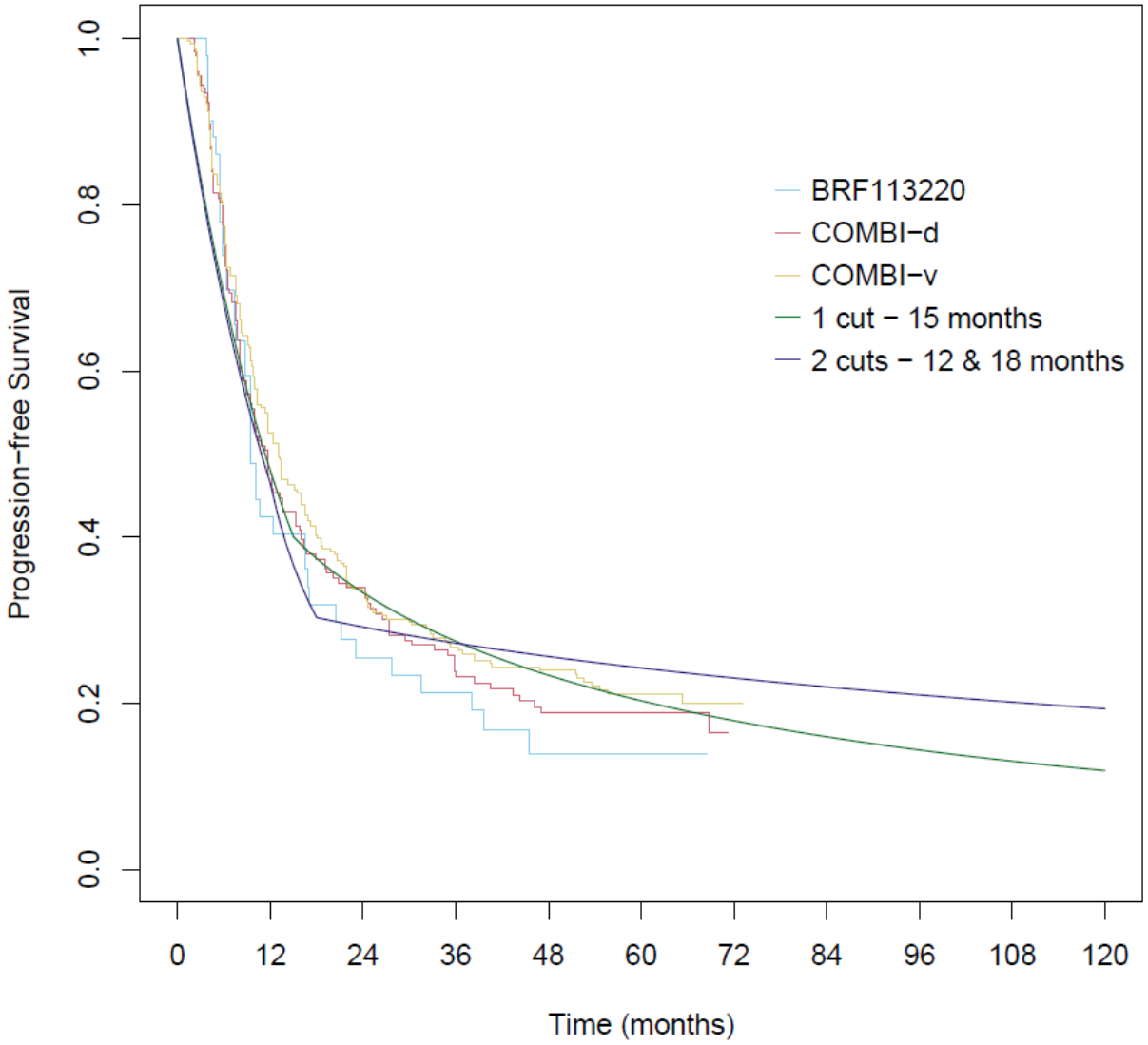


Figure 13: Predicted PFS for best fitting 1 and 2 cut points piecewise models compared with observed PFS data for dabrafenib + trametinib

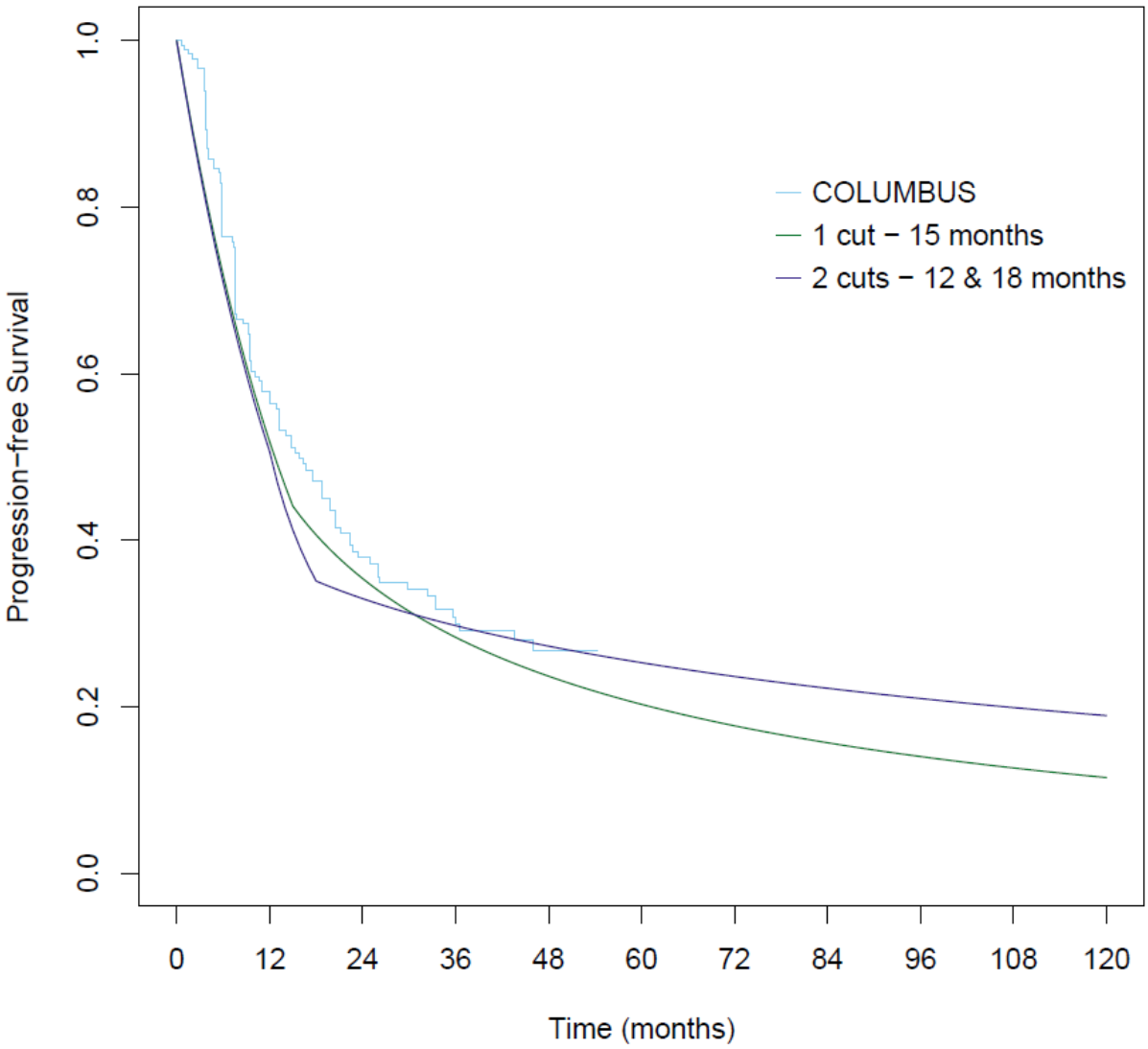


Figure 14: Predicted PFS for best fitting 1 and 2 cut points piecewise models compared with observed PFS data for encorafenib + binimetinib

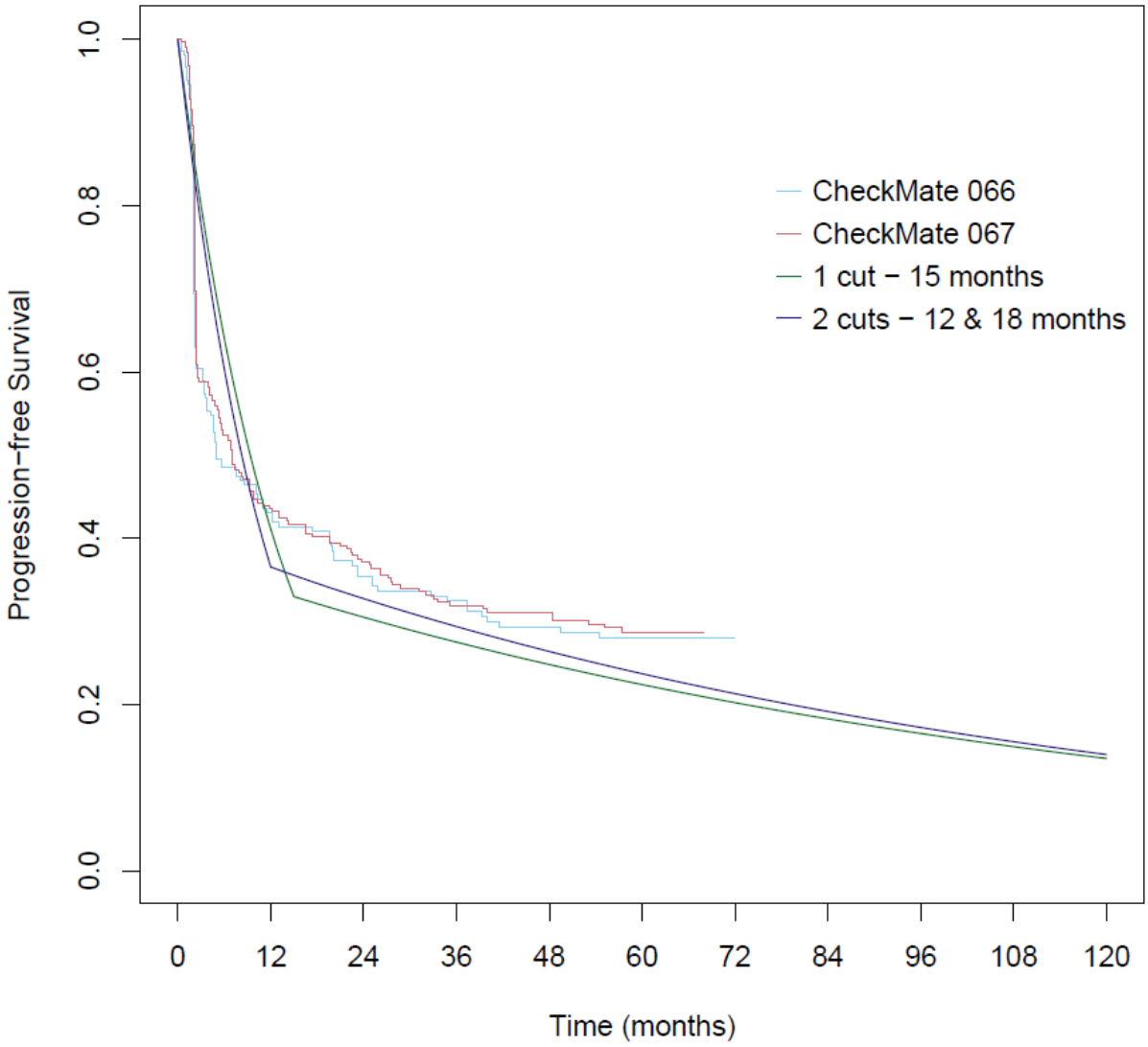


Figure 15: Predicted PFS for best fitting 1 and 2 cut points piecewise models compared with observed PFS data for nivolumab

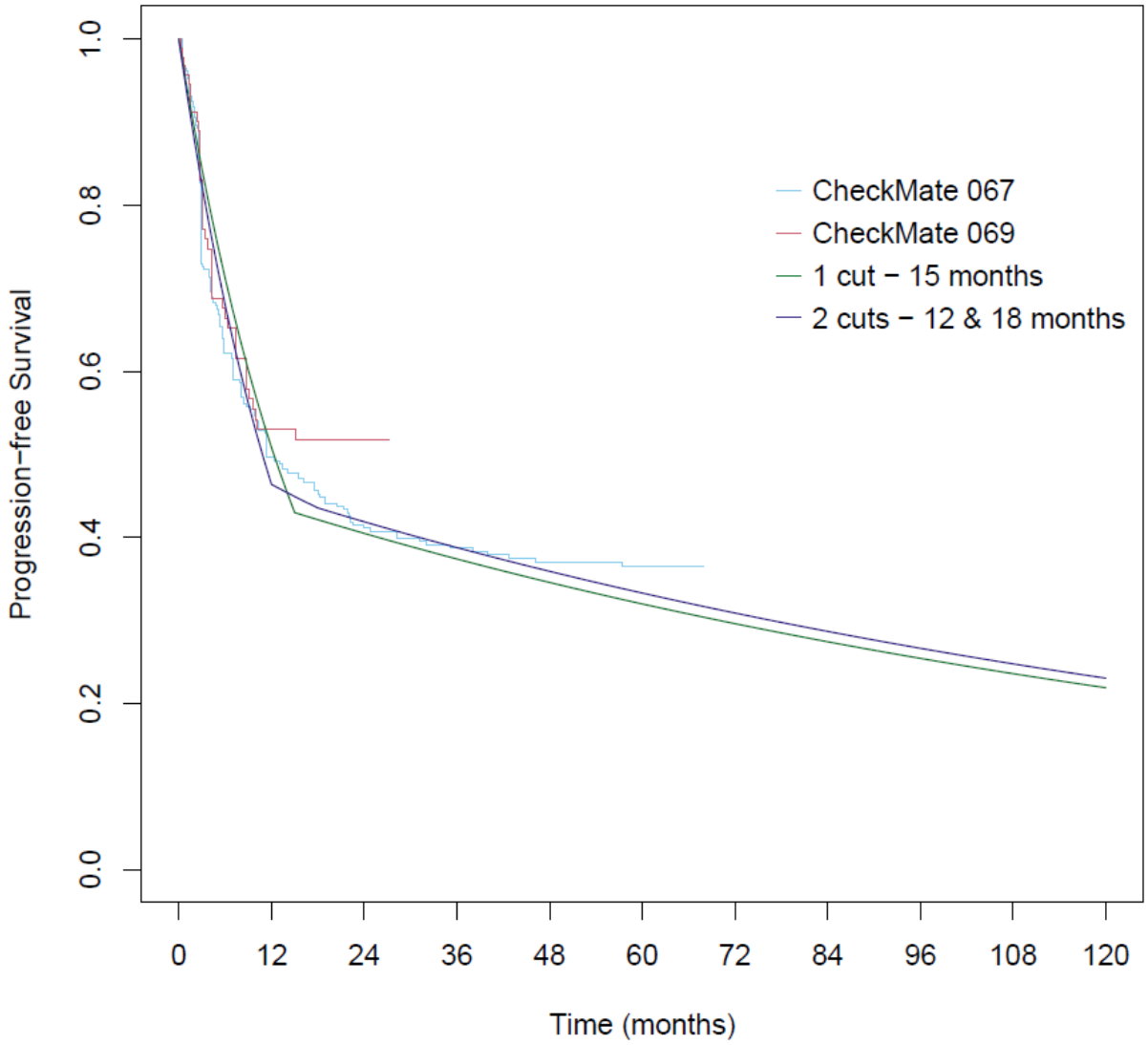


Figure 16: Predicted PFS for best fitting 1 and 2 cut points piecewise models compared with observed PFS data for nivolumab + ipilimumab

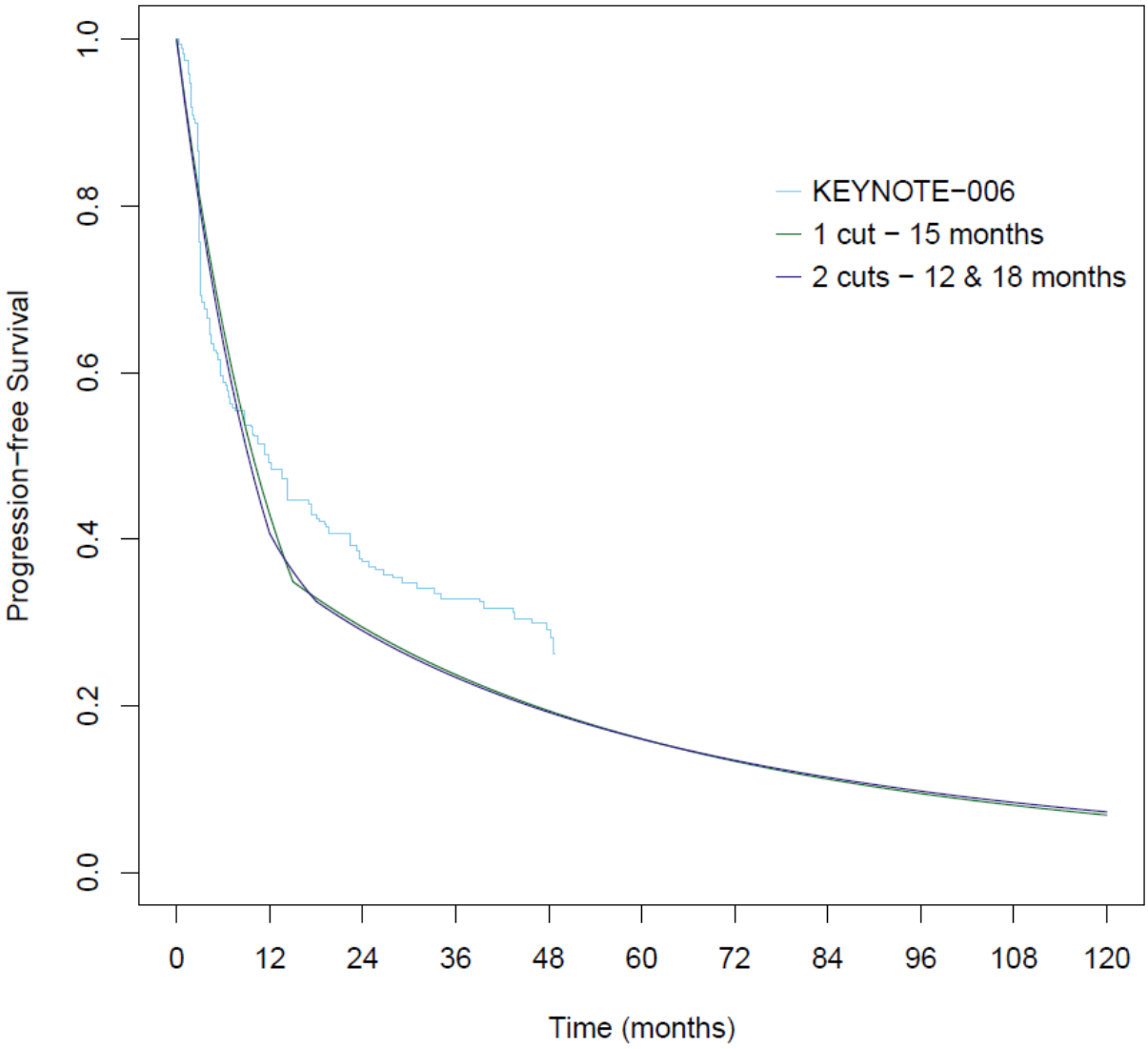


Figure 17: Predicted PFS for best fitting 1 and 2 cut points piecewise models compared with observed PFS data for pembrolizumab

A.5.1.2 Overall survival

A summary of each model fit and why it was ultimately not selected as the best fitting model is presented in Table 8.

The generalized gamma model with two treatment effects was selected as the most plausible model for OS. The outcomes of the model that was best fitting are presented in Section A.5.3.

The predicted OS outcomes for the fractional polynomial models by treatment were still lower than the observed OS from the trials. However, while these models may not match well to the observed KM data, their long-term extrapolations may be more plausible than other models due to their levelling off. Additionally, although a direct comparison using the DIC values is not possible, these models were incredibly complex requiring significant time to run. Thus, while it may not be possible to directly compare the complexity of the models using a measure such as DIC, it is reasonable to say the FP models are more complicated models. In the end, because the FP models were more complicated and did not provide a substantially improved fit, these models were immediately ruled out for OS and discarded.

In examining the piecewise models, both with a single cut point at 6 months, and two cut points at 12 & 18 months, it was apparent that while the two cut point model provided adequate fits to some treatments, for pembrolizumab, the piecewise model provided an incredibly pessimistic long-term extrapolation, shown in Figure 18, which was not clinically plausible. As such, any piecewise model used for OS, would require using the estimates of OS for pembrolizumab which were found to be poor fits. This inclusion alone was enough reason to strike the piecewise model, as model selection required assessing what model provided the best fit to the most data. The poor fit to pembrolizumab would have disadvantaged it in the economic model compared with other treatments.

The generalized gamma model with two treatment effects was a better fit for OS data when assessing goodness of fit based on AIC and based on visual inspection. The subcommittee was presented with survival predictions from the two-treatment effect generalized gamma in comparison with the observed KM data, and they agreed that on balance, the generalized gamma two treatment effect model represented the best fit to OS data.

Table 8: Summary table of OS models fit and rationale for non-selection

Model	Network					
	1	2	3	4	5	6
Generalized gamma						
Location parameter alone	> AIC compared with Location and Scale Parameter	NA	> AIC compared with Location and Scale Parameter	> AIC compared with Location and Scale Parameter	> AIC compared with Location and Scale Parameter	> AIC compared with Location and Scale Parameter
Location and scale parameter	Selected as best fitting model	NA	Selected as best fitting model	Selected as best fitting model	Selected as best fitting model	Selected as best fitting model
Piecewise exponential						
1 cut – 6 months	Although less complex, it lacked clinical plausibility compared with the 2 cut – 12 & 18 month model	NA	Although less complex, it lacked clinical plausibility compared with the 2 cut – 12 & 18 month model	> DIC compared with 1 cut – 15 months	Although less complex, it lacked clinical plausibility compared with the 2 cut – 6 & 12 month model	Although less complex, it lacked clinical plausibility compared with the 2 cut – 12 & 18 month model
1 cut – 9 months	> DIC compared with 1 cut – 6 months	NA	> DIC compared with 1 cut – 6 months	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 6 months	> DIC compared with 1 cut – 6 months
1 cut – 12 months	> DIC compared with 1 cut – 6 months	NA	> DIC compared with 1 cut – 6 months	> DIC compared with 1 cut – 15 months	> DIC compared with 1 cut – 6 months	> DIC compared with 1 cut – 6 months
1 cut – 15 months	> DIC compared with 1 cut – 6 months	NA	> DIC compared with 1 cut – 6 months	Although less complex, it lacked clinical plausibility compared	> DIC compared with 1 cut – 6 months	> DIC compared with 1 cut – 6 months

Model	Network					
	1	2	3	4	5	6
				with the 2 cut – 12 & 18 month model		
2 cuts – 6 & 12 months	> DIC compared with 2 cut – 12 & 18 months	NA	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 18 months	Lacked clinical plausibility compared with the gen gamma location and scale model	> DIC compared with 2 cut – 12 & 18 months
2 cuts – 9 & 15 months	> DIC compared with 2 cut – 12 & 18 months	NA	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 6 & 12 months	> DIC compared with 2 cut – 12 & 18 months
2 cuts – 9 & 18 months	> DIC compared with 2 cut – 12 & 18 months	NA	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 6 & 12 months	> DIC compared with 2 cut – 12 & 18 months
2 cuts – 12 & 18 months	Lacked clinical plausibility compared with the gen gamma location and scale model	NA	Selected as best fitting model	Selected as best fitting model	> DIC compared with 2 cut – 6 & 12 months	Selected as best fitting model
2 cuts – 12 & 20 months	> DIC compared with 2 cut – 12 & 18 months	NA	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 12 & 18 months	> DIC compared with 2 cut – 6 & 12 months	> DIC compared with 2 cut – 12 & 18 months
2 cuts – 12 & 24 months	NA	NA	> DIC compared with 2 cut – 12 & 18 months	NA	> DIC compared with 2 cut – 6 & 12 months	> DIC compared with 2 cut – 12 & 18 months
3 cuts – 6, 12 & 18 months	Increased model complexity without a significantly better fit compared with 2 cuts – 12 & 18 months	NA	> DIC compared with 3 cut – 12, 24 & 36	Increased model complexity without a significantly better fit compared with 2 cuts – 12 & 18 months	> DIC compared with 3 cut – 12, 24 & 36	> DIC compared with 3 cut – 12, 24 & 36
3 cuts – 12, 24 & 36 months	NA	NA	Increased model complexity without a significantly better fit	NA	Increased model complexity without a significantly better fit	Increased model complexity without a significantly better fit

Model	Network					
	1	2	3	4	5	6
			compared with 2 cuts – 12 & 18 months		compared with 2 cuts – 6 & 12 months	compared with 2 cuts – 12 & 18 months
Fractional polynomial 8 interval model						
-2	> DIC compared with 0	NA	> DIC compared with 0.5	> DIC compared with 0	> DIC compared with -0.5	> DIC compared with 0.5
-1	> DIC compared with 0	NA	> DIC compared with 0.5	> DIC compared with 0	> DIC compared with -0.5	> DIC compared with 0.5
-0.5	> DIC compared with 0	NA	> DIC compared with 0.5	> DIC compared with 0	Poor fit and high model complexity compared with piecewise and gen gamma	> DIC compared with 0.5
0	Poor fit and high model complexity compared with piecewise and gen gamma	NA	> DIC compared with 0.5	Poor fit and high model complexity compared with piecewise and gen gamma	Did not converge	> DIC compared with 0.5
0.5	Did not converge	NA	Poor fit and high model complexity compared with piecewise and gen gamma	Did not converge	Did not converge	Poor fit and high model complexity compared with piecewise and gen gamma
1	Did not converge	NA	Did not converge	Did not converge	Undefined real result	Did not converge
2	Did not converge	NA	Did not converge	Did not converge	Undefined real result	Did not converge
3	> DIC compared with 0	NA	> DIC compared with 0.5	> DIC compared with 0	> DIC compared with -0.5	> DIC compared with 0.5

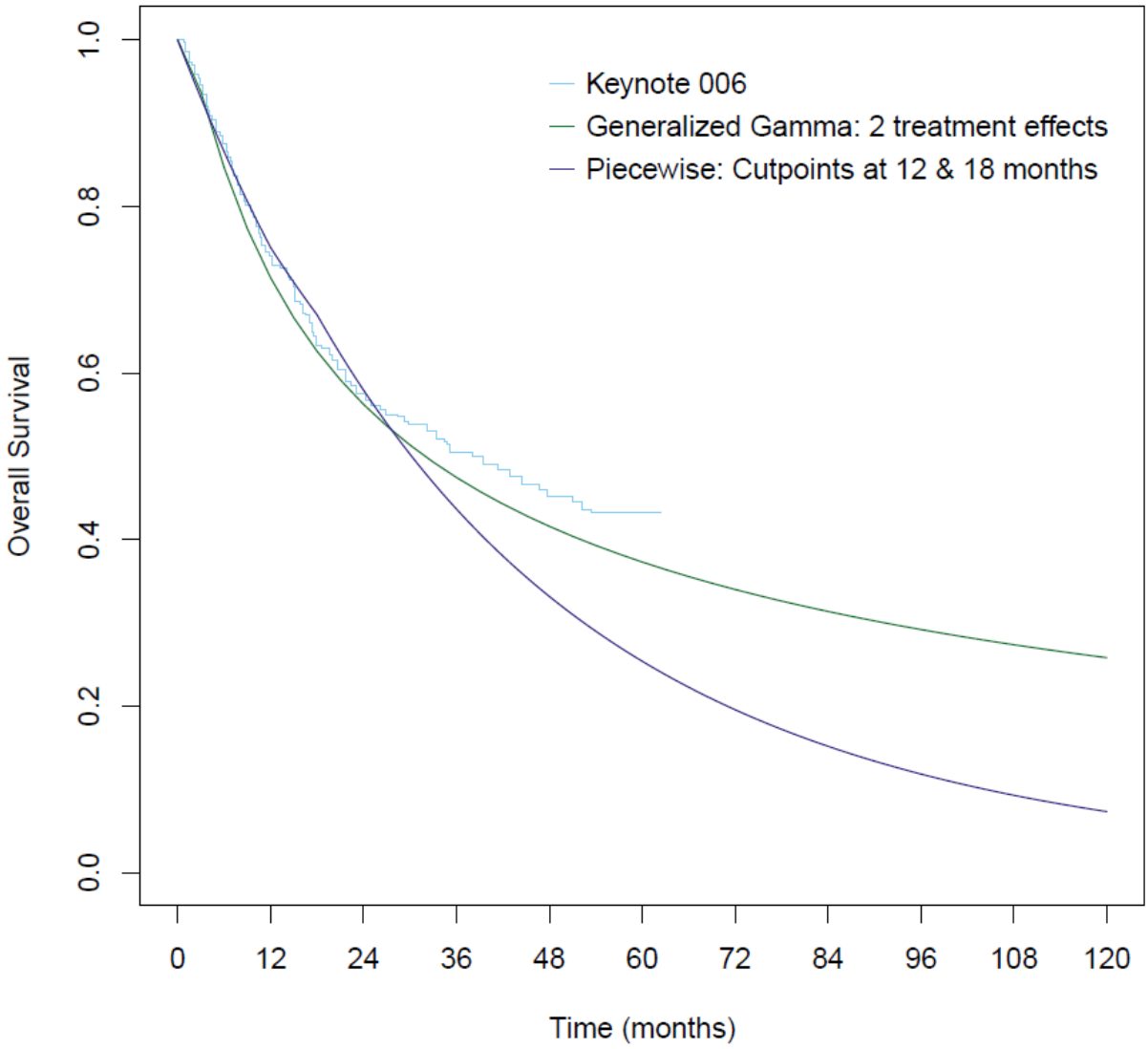


Figure 18: Predicted pembrolizumab overall survival data from generalized gamma two treatment effect models and piecewise model with cut points at 12 & 18 months compared with observed survival data

A.5.2 Selected PFS model for each network

In sections A.5.2.1-A.5.2.6 we present the results of the best fitting model for PFS for each network. However, we present the results of the models that were not selected in Appendix D, Appendix E:, and Appendix F:.

A.5.2.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

For Network 1, the piecewise exponential model with two cut points at 12 and 18 months was the best fitting model. Model fit statistics for network 1 are given in Table 9. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

Table 9: Model fit statistics

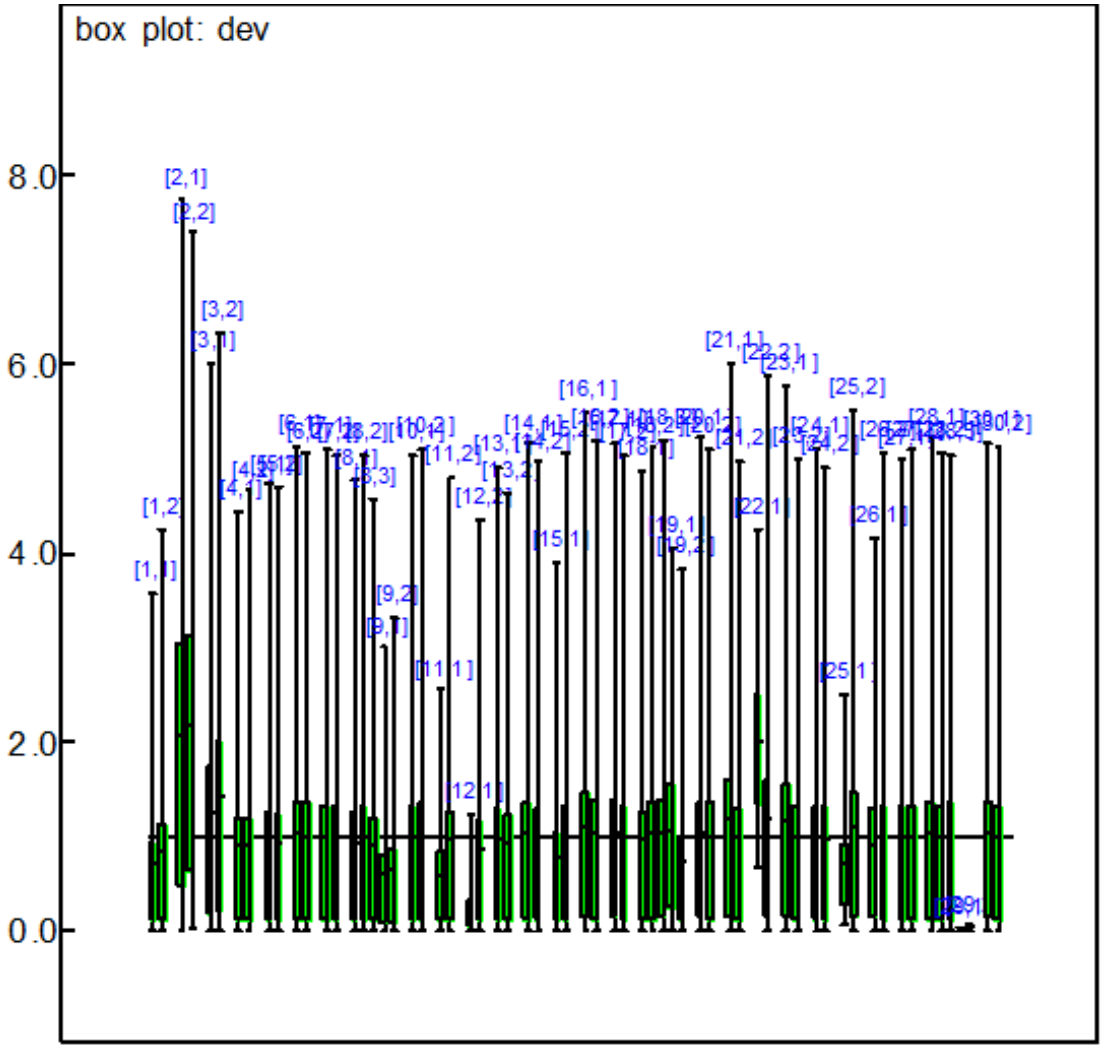
Model	Dbar	pD	DIC
Fixed effect – consistency	351.9	52.4	404.3
Fixed effect - inconsistency	352.9	53.8	406.7

In Figure 19, we can see that neither arm of the BRF113220 data is predicted well over the first interval (1-12 months). This poor fit suggests placing the first cut point at a different location may result in an improved fit for both arms of the BRF113220 data. We see exactly this in Figure 20 – where the cut points are placed at 6 and 12 months. Placing the cut points here results in an improved fit for both arms of the BRF113220 data in interval 1. However, this comes at the expense of a poor fit for both arms of the BRF113220 data in interval 2 (7-12 months). This highlights the balancing act required in determining the optimal cut point placements in a piecewise exponential model – while moving the location of the cut points may result in an improved fit for certain data points, it has the potential to provide an even worse fit for other data points. It is important to recognize in determining the ideal cut point placements one is trying to find the optimal placement for the entire network, and this may not fit all of the data well. However, finding the best cut point location on balance remains more important than resulting in an improved fit for individual trials. All of this is reflected in Table 62, where we see amongst the models with two cut points, placing the cut points at 12 and 18 months has the lowest DIC value, indicating a better fitting model.

We also see in Figure 19 that the first arm (dabrafenib monotherapy) of BRF113220 is not predicted well over the third interval (19-120 months).

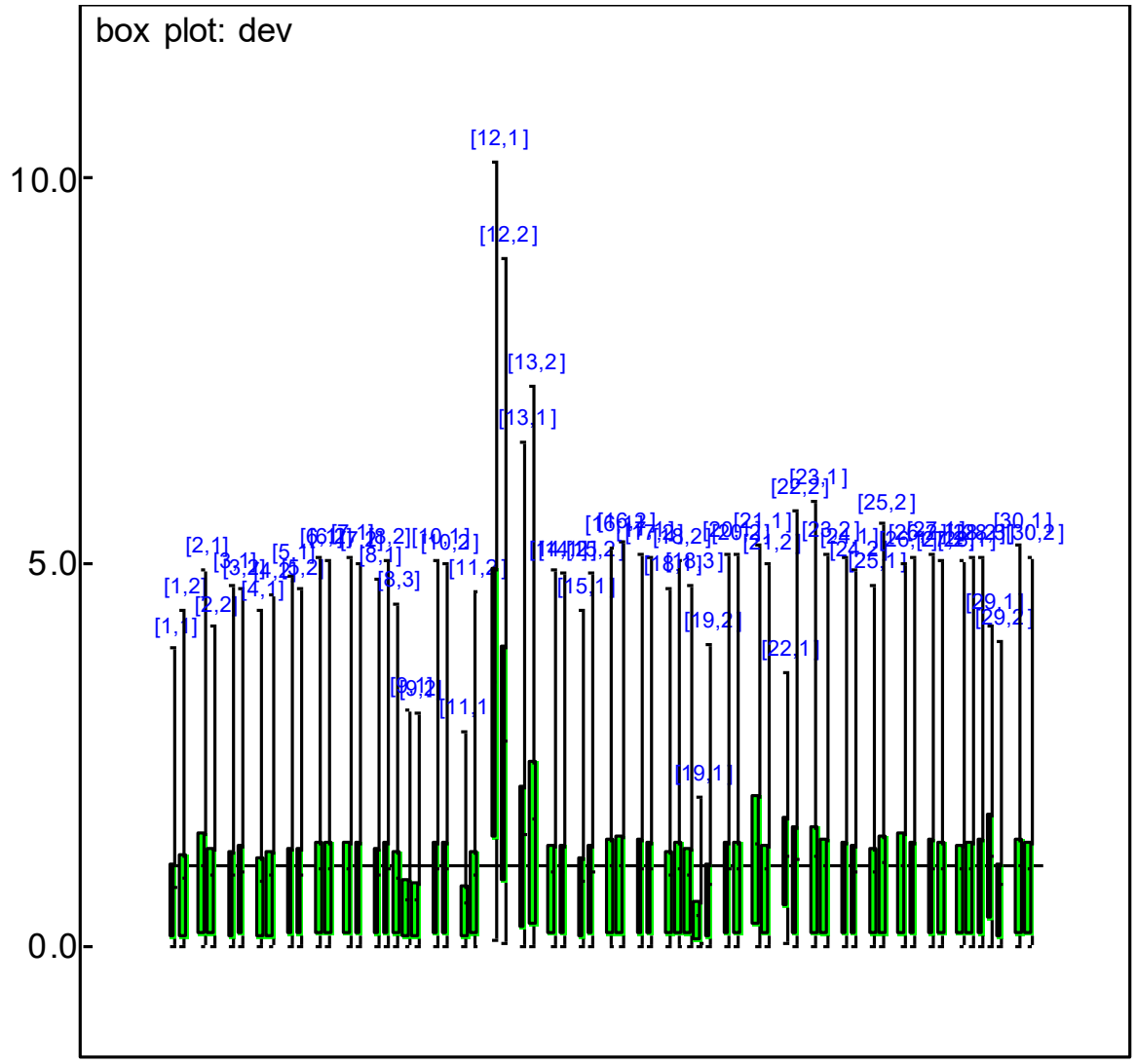
Notably, CheckMate 069, has a relatively small contribution to the residual deviance over its third interval (see [29,1] & [29,2]). This is likely due to CheckMate 069 having only 24 months of follow-up, meaning in the third interval for the aggregate data spans from just 18-24 months. Thus, there is relatively little data here, which in turn equates to a smaller relative contribution to the residual deviance. The remainder of the studies have normal (or as expected) contributions to the residual deviance.

Figure 19: box plot of the deviance contribution for each arm of aggregated data (2 cut points at 12 & 18 months) to the residual deviance – from WinBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. Numbers above the lines represent the aggregate data the plot corresponds to. There are two numbers in the brackets above each box plot. The first number corresponds to a line of aggregated data (note the line of aggregated data in the excel.csv does not always match to the study – for instance line 8 of the aggregated data refers to study 5, as such, for clarity we present the numbers and their corresponding study here: 1, 11, & 21 = BREAK-3; 2, 12 & 22 = BRF113220; 3, 13 & 23 = COMBI-d; 4, 14 & 24 = COMBI-v; 5, 15 & 25 = BRIM-3; 6, 16 & 26 = CheckMate 066; 7, 17 & 27 = COLUMBUS; 8, 18 & 28 = CheckMate 067; 9, 19 & 29 = CheckMate 069; 10, 20 & 30 = KEYNOTE-006) and the piecewise interval. Numbers 1-10 correspond to the first interval (1-12 months), numbers 11-20 correspond to the second interval (13-18 months) and numbers 21-30 correspond to the third interval (19-120 months). The second number corresponds to the study arm. For example, [2,1] corresponds to the aggregated data for BRF113220, in the first interval for arm 1; [28,3] corresponds to the aggregate data for CheckMate 067, in the third interval, for arm 3. The horizontal line indicates a contribution to residual deviance of 1, which is expected from each arm.

Figure 20: box plot of the deviance contribution for each arm of aggregated data (2 cut points at 6 & 12 months) to the residual deviance – from WinBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. Numbers above the lines represent the aggregate data the plot corresponds to. There are two numbers in the brackets above each box plot. The first number corresponds to a line of aggregated data (note the line of aggregated data in the excel.csv does not always match to the study – for instance line 8 of the aggregated data refers to study 5, as such, for clarity we present the numbers and their corresponding study here: 1, 11, & 21 = BREAK-3; 2, 12 & 22 = BRF113220; 3, 13 & 23 = COMBI-d; 4, 14 & 24 = COMBI-v; 5, 15 & 25 = BRIM-3; 6, 16 & 26 = CheckMate 066; 7, 17 & 27 = COLUMBUS; 8, 18 & 28 = CheckMate 067; 9, 19 & 29 = CheckMate 069; 10, 20 & 30 = KEYNOTE-006) and the piecewise interval. Numbers 1-10 correspond to the first interval (1-6 months), numbers 11-20 correspond to the second interval (7-12 months) and numbers 21-30 correspond to the third interval (13-120 months). The second number corresponds to the study arm. For example, [2,1] corresponds to the aggregated data for BRF113220, in the first interval for arm 1; [28,3] corresponds to the aggregate data for CheckMate 067, in the third interval, for arm 3. The horizontal line indicates a contribution to residual deviance of 1, which is expected from each arm.

No evidence of inconsistency was found, with model fit and DIC being marginally lower for the consistency model (Table 9). The area below the line of equality in Figure 21 highlights where the inconsistency model better predicted data points. Of note, there is one data point where the inconsistency model provides a better prediction, however, most other data points see no improvement in fit.

Figure 21: Deviance contributions from the fixed effect consistency and inconsistency models

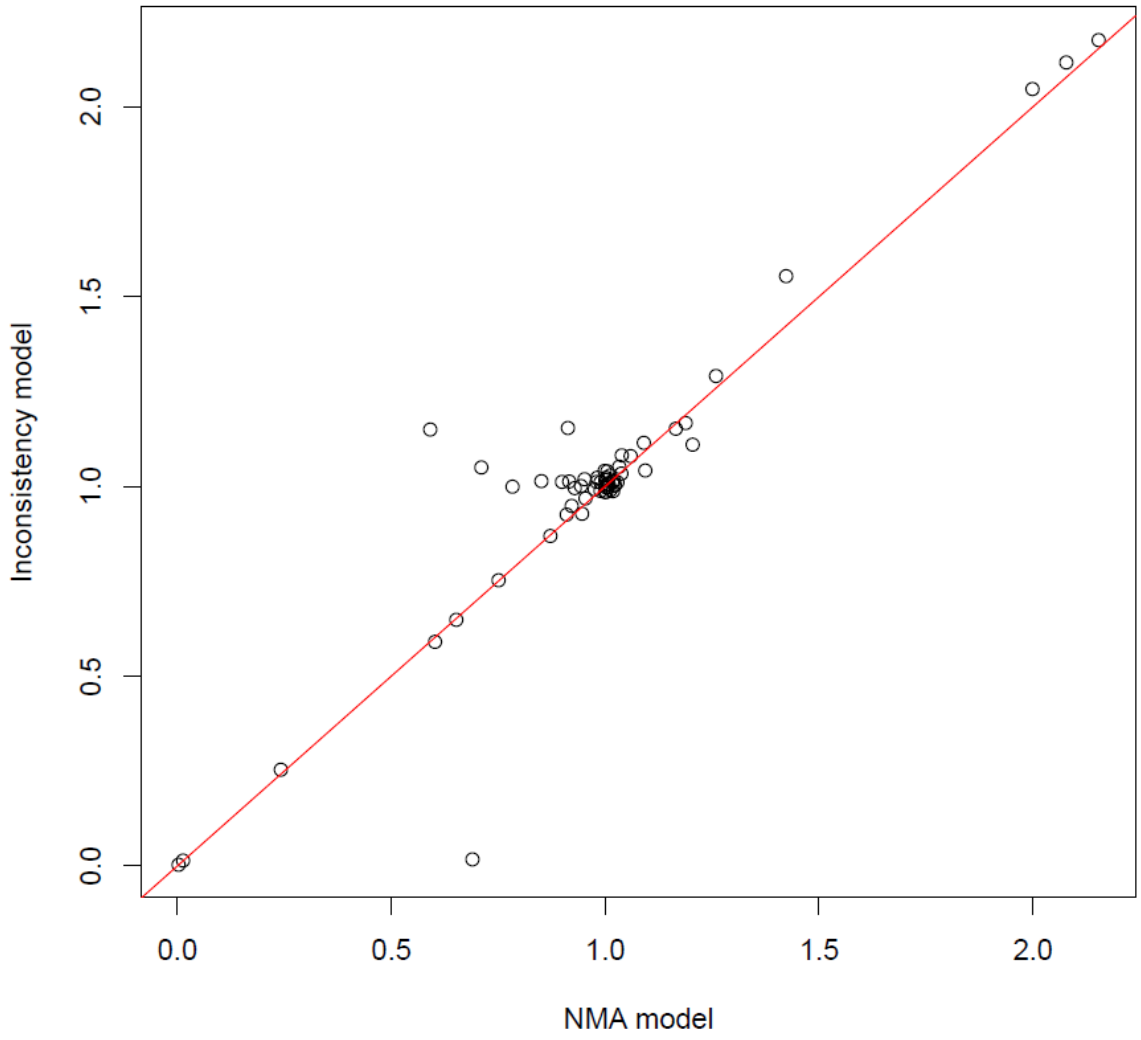


Table 10: Fixed effect PFS NMA results (piecewise exponential model with 2 cut points at 12 and 18 months)

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		
dabrafenib vs. DTIC	0.399	(0.313; 0.511)
dabrafenib + trametinib vs. DTIC	0.268	(0.212; 0.338)
ipilimumab vs. DTICC	0.692	(0.506; 0.94)
vemurafenib vs. DTIC	0.439	(0.37; 0.519)
nivolumab vs. DTIC	0.351	(0.275; 0.446)
nivolumab + ipilimumab vs. DTIC	0.267	(0.193; 0.367)
encorafenib + binimetinib vs. DTIC	0.236	(0.166; 0.336)
pembrolizumab vs. DTIC	0.313	(0.214; 0.456)
Interval 2: 13-18 months		
dabrafenib vs. DTIC	1.718	(0.474; 7.846)
dabrafenib + trametinib vs. DTIC	0.937	(0.276; 4.129)

ipilimumab vs. DTIC	0.451	(0.064; 4.433)
vemurafenib vs. DTIC	1.167	(0.366; 4.826)
nivolumab vs. DTIC	0.135	(0.023; 1.099)
nivolumab + ipilimumab vs. DTIC	0.138	(0.019; 1.353)
encorafenib + binimetinib vs. DTIC	0.753	(0.174; 3.955)
pembrolizumab vs. DTIC	0.461	(0.05; 5.54)
Interval 3: 19-120 months		
dabrafenib vs. DTIC	0.567	(0.141; 3.963)
dabrafenib + trametinib vs. DTIC	0.514	(0.115; 3.931)
ipilimumab vs. DTIC	2.356	(0.363; 26.233)
vemurafenib vs. DTIC	0.797	(0.17; 6.366)
nivolumab vs. DTIC	1.373	(0.232; 14.44)
nivolumab + ipilimumab vs. DTIC	0.941	(0.145; 10.444)
encorafenib + binimetinib vs. DTIC	0.789	(0.14; 6.987)
pembrolizumab vs. DTIC	2.509	(0.31; 32.819)

Table 10 presents the model predicted coefficients for each pairwise comparison and Figure 22 presents the predicted PFS survival curves.

- First interval: 0-12 months
    - The hazard ratio in the first interval was smallest for encorafenib + binimetinib, then dabrafenib + trametinib, then nivolumab + ipilimumab, then pembrolizumab monotherapy and finally nivolumab monotherapy. These hazard ratios were 0.237 (95% CI 0.166-0.334), 0.267 (95% CI 0.213-0.336), 0.268 (95% CI 0.195-0.334), 0.314 (95% CI 0.217-0.455), and 0.352 (95% CI 0.277-0.443) respectively. These point estimates can be interpreted as a 76.3%, 73.3%, 73.2%, 68.6% and 64.8% reduction in the chance of disease progression respectively for those treated with encorafenib + binimetinib, dabrafenib + trametinib, nivolumab + ipilimumab, pembrolizumab monotherapy and nivolumab monotherapy compared with those treated with DTIC.
- Second interval: 12-18 months
    - For the second interval, there was no evidence that any of these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the second interval was smallest for nivolumab monotherapy, then nivolumab +ipilimumab, then pembrolizumab monotherapy, then encorafenib + binimetinib, and finally dabrafenib + trametinib. These hazard ratios were 0.145 (95% CI 0.024-1.36), 0.148 (95% CI 0.019-1.637), 0.495 (95% CI 0.049-7.822), 0.776 (95% CI 0.185-4.039), and 0.963 (95% CI 0.29-3.939) respectively. These point estimates can be interpreted as an 85.5%, 85.2%, 50.5%, 22.4% and 3.7% reduction in the chance of disease progression respectively for those treated with nivolumab monotherapy, nivolumab +ipilimumab, pembrolizumab monotherapy, encorafenib + binimetinib, and dabrafenib + trametinib compared with those treated with DTIC.
- Third interval: >18 months
    - For the third interval, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the third interval was smallest for

dabrafenib + trametinib, then encorafenib + binimetinib, then nivolumab + ipilimumab, then nivolumab monotherapy, and finally pembrolizumab monotherapy. These hazard ratios were 0.512 (95% CI 0.109-2.815), 0.791 (95% CI 0.125-5.376), 0.977 (95% CI 0.166-8.298), 1.417 (95% CI 0.267-11.531), and 2.614 (95% CI 0.344-23.951) respectively. These point estimates can be interpreted as an 48.8%, 20.9%, 2.3%, reduction in the chance of disease progression respectively for those treated with dabrafenib + trametinib, encorafenib + binimetinib and nivolumab + ipilimumab respectively and a 41.7% and 161.4% increase in the chance of disease progression respectively for those treated with nivolumab monotherapy and pembrolizumab monotherapy, compared with those treated with DTIC.

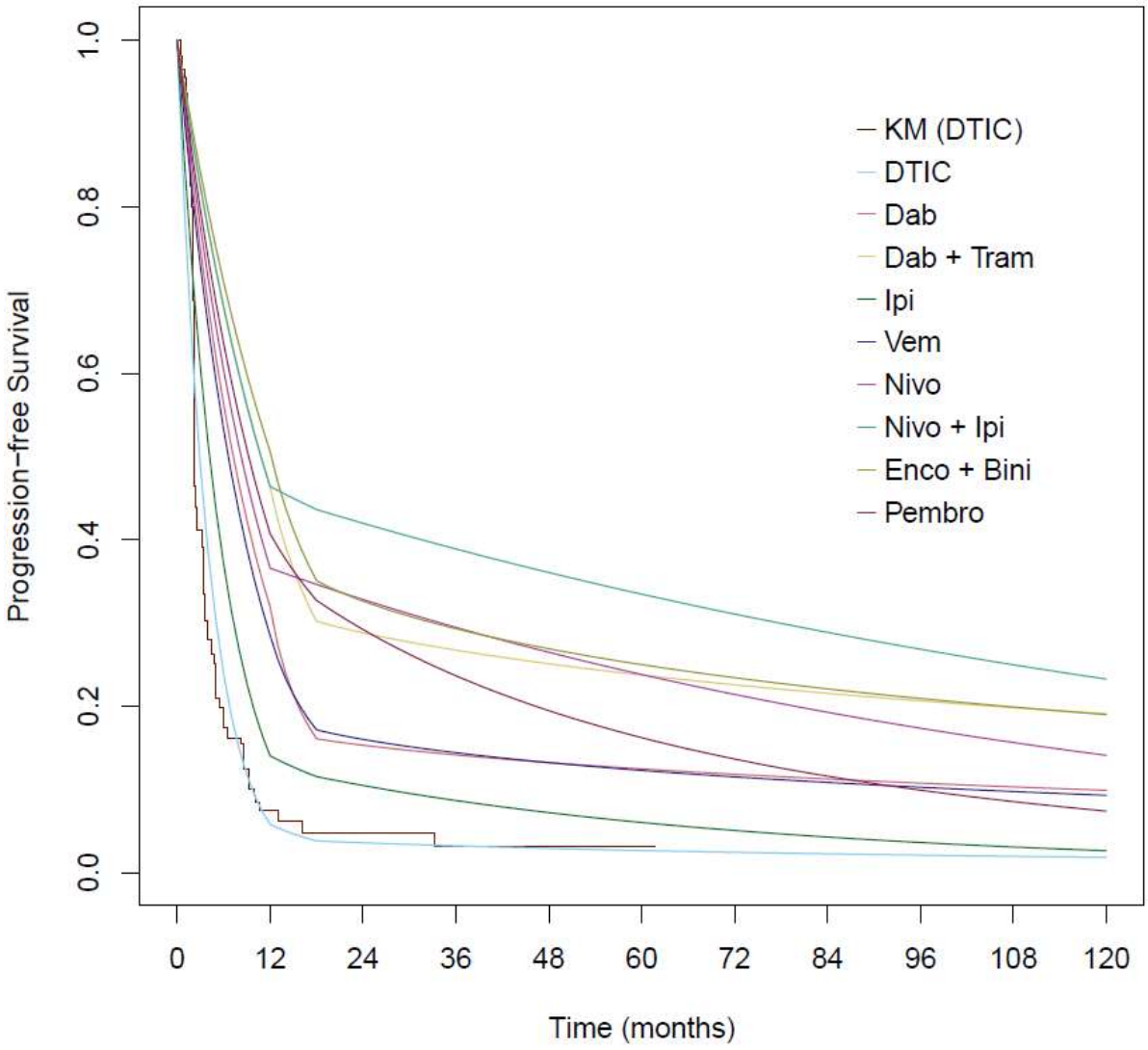


Figure 22: Survival curves for piecewise PFS model with two cut points at 12 and 18 months

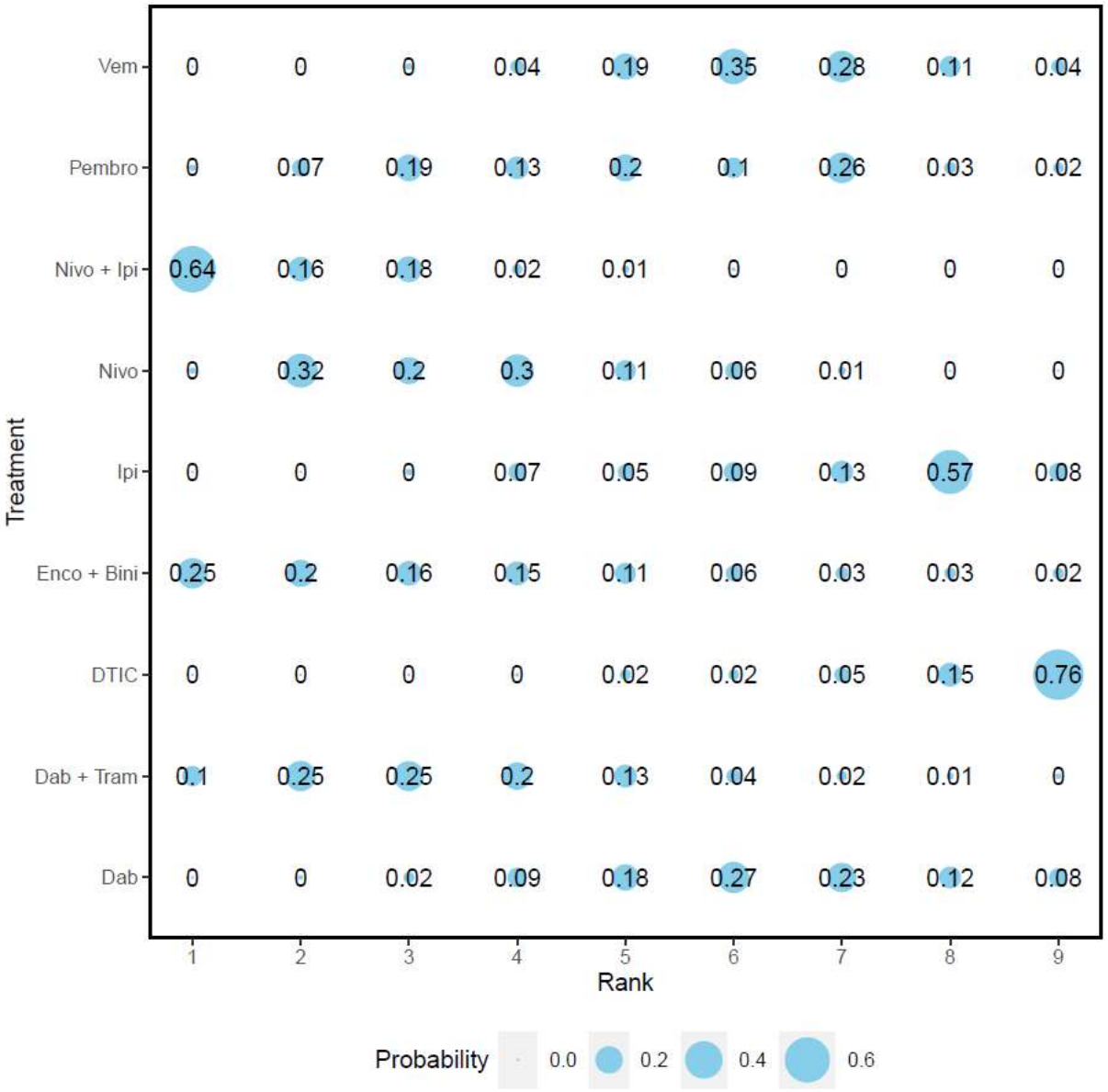


Figure 23: Ranking plot for survival at 60 months for piecewise PFS model with two cut points at 12 and 18 months

These results are reflected in the ranking plots of progression-free survival at 60 months Figure 23:

- Nivolumab + ipilimumab had an average ranking of 1.63 with a 64% probability of being the best treatment. This ranking is driven by nivolumab + ipilimumab having a low hazard ratio for both the first and second interval. This means by 18 months, nivolumab + ipilimumab has the greatest proportion of patients who are progression free, which is maintained until 120 months, even if the proportion of people progression free does decrease further from months 18-120 drawing it closer to both dabrafenib + trametinib and encorafenib + binimetinib.
- Nivolumab monotherapy had an average ranking of 3.42 with a 0% probability of being the best treatment. This ranking can be explained with two important pieces of information. Nivolumab has a worse hazard ratio in the first interval than either dabrafenib + trametinib or encorafenib + binimetinib. Although dabrafenib + trametinib has a worse hazard ratio in the second interval, it has a better hazard ratio in the final interval. This means that while dabrafenib + trametinib has a greater proportion of

progression events than nivolumab monotherapy from approximately month 15-48 months, dabrafenib + trametinib is better than nivolumab monotherapy at >48 months. This explains why nivolumab monotherapy appears worse than the top three treatments, nivolumab + ipilimumab, encorafenib + binimetinib and dabrafenib + trametinib. Nivolumab monotherapy does appear better than pembrolizumab monotherapy and this is because the hazard ratio for pembrolizumab monotherapy in the second and third interval is worse than nivolumab monotherapy. Therefore, while in the first interval pembrolizumab monotherapy had a smaller hazard ratio compared with nivolumab, due to the second and third interval hazard ratios nivolumab monotherapy overtakes pembrolizumab monotherapy and has a greater proportion of people progression-free.

- Pembrolizumab monotherapy had an average ranking of 5.07 with a 0% probability of being the best treatment. Pembrolizumab monotherapy appears the worst treatment for of the treatments of interest for progression-free survival. This is already covered above but is primarily driven by the hazard ratios for the second and third interval which push it far below the other treatments.

- Encorafenib + binimetinib had an average ranking of 3.09 with a 25% probability of being the best treatment. As already discussed briefly with the description of nivolumab + ipilimumab, encorafenib + binimetinib has the smallest hazard ratio for the first interval. However, it has worse hazard ratios for the second interval which pushes it below nivolumab + ipilimumab. Furthermore, the hazard ratio for encorafenib + binimetinib for the third interval is far worse than the hazard ratio for dabrafenib + trametinib which is the two treatments come together by 120 months.

- Dabrafenib + trametinib had an average ranking of 3.26 with a 10% probability of being the best treatment. As already discussed, the hazard ratio for dabrafenib + trametinib for intervals 1 and 2 is why it is pushed below both nivolumab + ipilimumab and encorafenib + binimetinib. However, dabrafenib + trametinib's hazard ratio for the interval is better than the hazard ratio for encorafenib + binimetinib which is why you see the two curves come together by month 120.

A.5.2.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

A.5.2.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

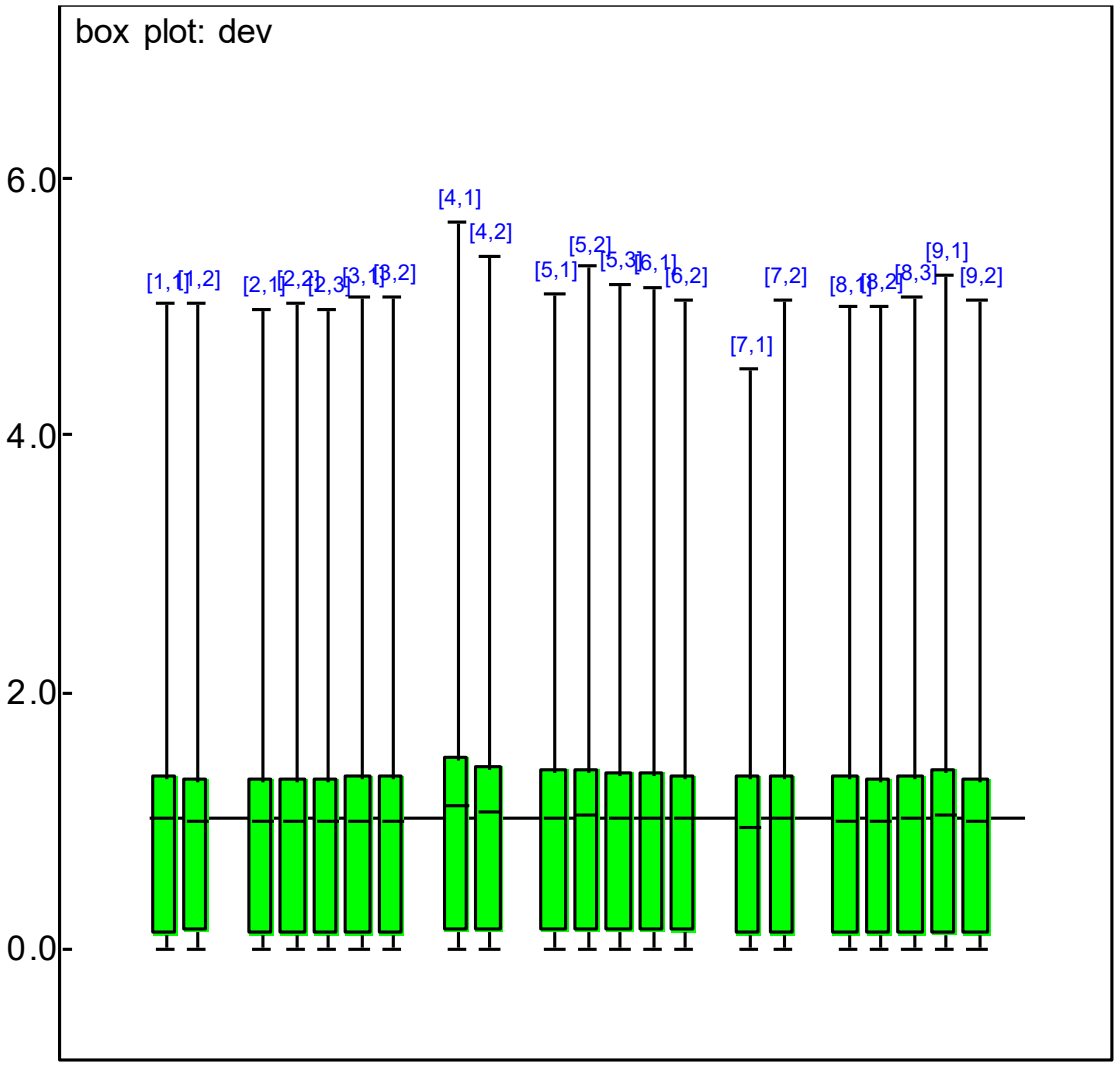
For Network 3, the piecewise exponential model with two cut points at 12 and 18 months was the best fitting model. Model fit statistics for network 3 are given in Table 11. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

Table 11: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect	122.882	20.626	143.508

In Figure 24, we can see that all studies have normal (or as expected) contributions to the residual deviance. Thus, it appears the predictions from the NMA are a good fit to the observed data points.

Figure 24: box plot of the deviance contribution for each arm of aggregated data to the residual deviance – from WinBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. There are two numbers in the brackets above each box plot. The first number corresponds to a line of aggregated data (for clarity we present the numbers and their corresponding study here: 1, 4, & 7 = CheckMate 066; 2, 5 & 8 = CheckMate 067; 3, 6 & 9 = KEYNOTE-006) and the piecewise interval. Numbers 1-3 correspond to the first interval (1-12 months), numbers 4-6 correspond to the second interval (13-18 months) and numbers 7-9 correspond to the third interval (19-120 months). The second number corresponds to the study arm. For example, [1,1] corresponds to aggregate data for CheckMate 066, in the first interval, for arm 1; [8,3] corresponds to aggregate data for CheckMate 067, in the third interval, for arm 3. The horizontal line indicates a contribution to residual deviance of 1, which is expected from each arm.

It was not possible to assess inconsistency in this network as it only had a single loop which was composed of a three-arm trial, and it is not possible for a three-arm trial to be inconsistent with itself.

Table 12: Fixed effect PFS NMA results (piecewise exponential model with 2 cut points at 12 and 18 months)

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		
ipilimumab vs. DTIC	0.827	(0.592; 1.15)
nivolumab vs. DTIC	0.351	(0.276; 0.444)
nivolumab + ipilimumab vs. DTIC	0.311	(0.216; 0.444)
pembrolizumab vs. DTIC	0.374	(0.25; 0.558)
Interval 2: 13-18 months		
ipilimumab vs. DTIC	0.721	(0.067; 9.365)
nivolumab vs. DTIC	0.137	(0.022; 1.1)
nivolumab + ipilimumab vs. DTIC	0.253	(0.026; 3.086)
pembrolizumab vs. DTIC	0.73	(0.057; 12.025)
Interval 3: 19-120 months		
ipilimumab vs. DTIC	2.191	(0.285; 32.492)
nivolumab vs. DTIC	1.335	(0.221; 17.167)
nivolumab + ipilimumab vs. DTIC	1.024	(0.146; 13.943)
pembrolizumab vs. DTIC	2.337	(0.254; 40.609)

Table 12 presents the model predicted coefficients for each pairwise comparison and Figure 25 presents the predicted PFS survival curves.

- First interval: 0-12 months
  - The hazard ratio in the first interval was smallest for nivolumab + ipilimumab, then nivolumab monotherapy and finally pembrolizumab monotherapy. These hazard ratios were 0.311 (95% CI 0.218-0.445), 0.351 (95% CI 0.278-0.443), 0.376 (95% CI 0.252-0.566) respectively. These point estimates can be interpreted as a 68.9%, 64.9% and 62.4% reduction in the chance of disease progression respectively for those treated with nivolumab + ipilimumab, nivolumab monotherapy and pembrolizumab monotherapy compared with those treated with DTIC.
- Second interval: 12-18 months
  - For the second interval, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the second interval was smallest for nivolumab monotherapy, then nivolumab + ipilimumab and finally pembrolizumab monotherapy. These hazard ratios were 0.141 (95% CI 0.024-1.211), 0.266 (95% CI 0.03-3.025), 0.796 (95% CI 0.071-10.979) respectively. These point estimates can be interpreted as an 85.9%, 73.4% and 20.4%, reduction in the chance of disease progression respectively for those treated with nivolumab monotherapy, nivolumab + ipilimumab and pembrolizumab monotherapy compared with those treated with DTIC.
- Third interval: >18 months
  - For the third interval, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates.

The point estimate of the hazard ratio in the second interval was smallest for nivolumab + ipilimumab, then nivolumab monotherapy and finally pembrolizumab monotherapy. These hazard ratios were 0.945 (95% CI 0.145-10.816), 1.219 (95% CI 0.216-12.453), 2.061 (95% CI 0.232-24.779) respectively. These point estimates can be interpreted as an 5.5% reduction in the chance of disease progression respectively for those treated with nivolumab + ipilimumab and a 21.9% and 106.1% increase in the chance of disease progression respectively for those treated with nivolumab monotherapy and pembrolizumab monotherapy compared with those treated with DTIC.

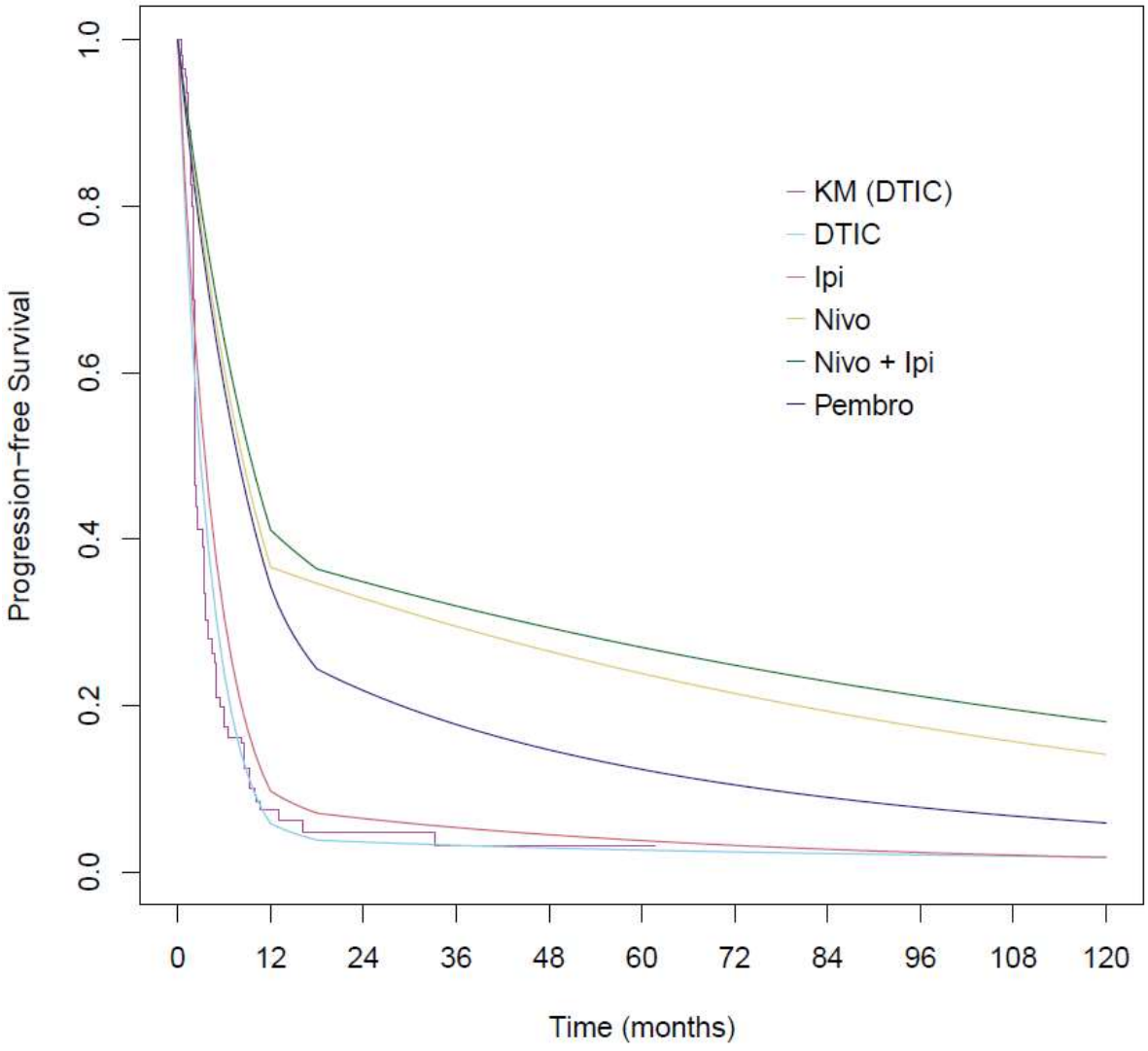


Figure 25: Survival curves for piecewise PFS model with two cut points at 12 and 18 months

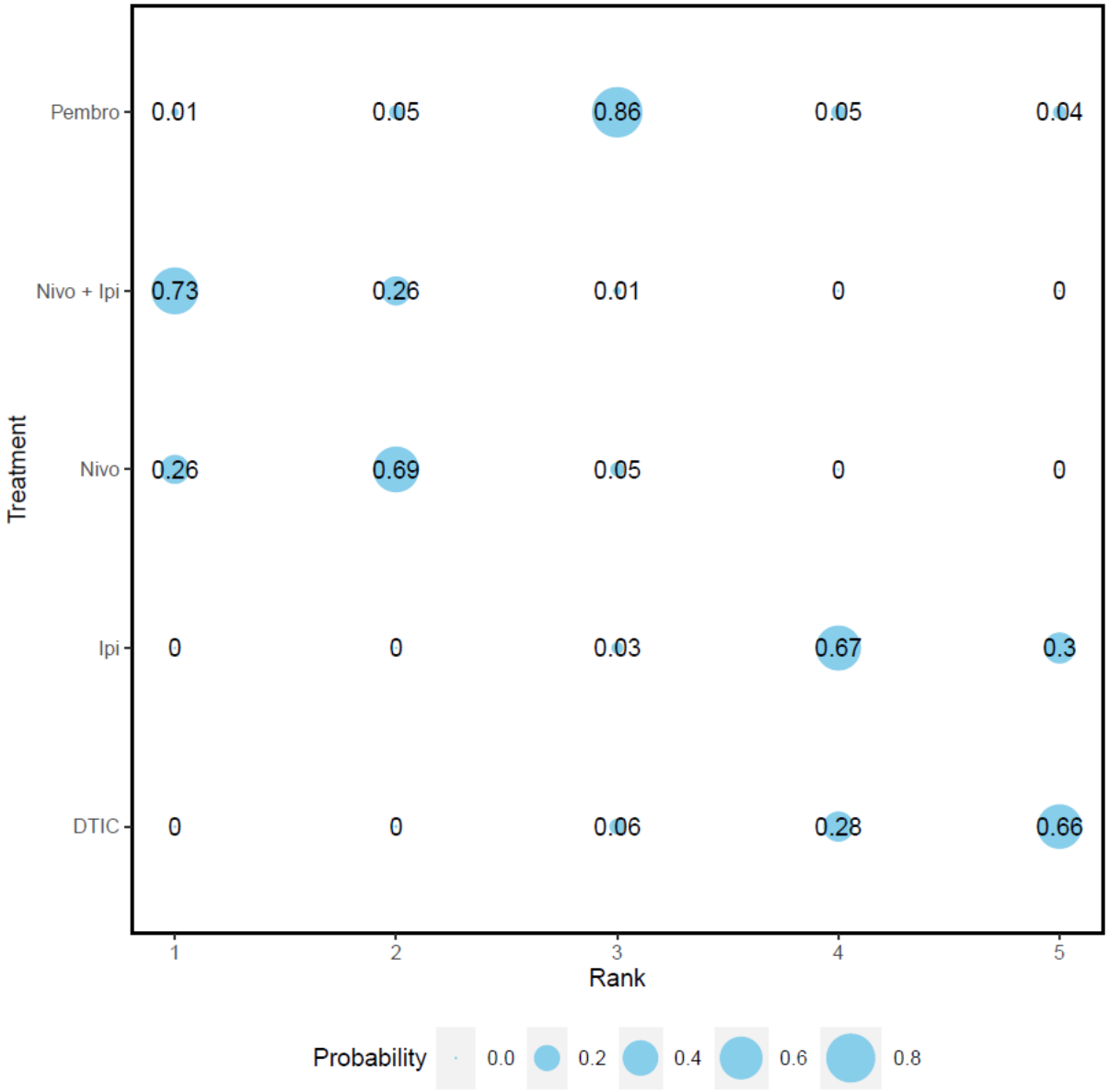


Figure 26: Ranking plot for survival at 60 months for piecewise PFS model with two cut points at 12 and 18 months

These results are reflected in the ranking plots of progression-free survival at 60 months Figure 26:

- Nivolumab + ipilimumab had an average ranking of 1.28 with a 73% probability of being the best treatment. This ranking is driven by nivolumab + ipilimumab having a low hazard ratio for all three intervals. Though in the second interval the point estimate of the hazard ratio for nivolumab + ipilimumab is worse than nivolumab monotherapy, it is not enough for nivolumab + ipilimumab to dip below nivolumab. In the final interval, the point estimate of the hazard ratio for nivolumab monotherapy is worse than nivolumab + ipilimumab so it is not possible for it to have a greater proportion of people who are progression free.
- Nivolumab monotherapy had an average ranking of 1.79 with a 26% probability of being the best treatment. This ranking is explained above as to why nivolumab monotherapy is worse than nivolumab + ipilimumab. Nivolumab monotherapy ends up with a better ranking than pembrolizumab monotherapy as nivolumab

monotherapy has smaller hazard ratios than pembrolizumab in all three intervals. This is why nivolumab comes out better in the treatment rankings.

- Pembrolizumab monotherapy had an average ranking of 3.09 with a 1% probability of being the best treatment. Pembrolizumab monotherapy appears the worst treatment for of the treatments of interest for progression-free survival. This is already covered above but is primarily driven by the hazard ratios for the second and third interval which push it far below the other treatments.

A.5.2.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

For Network 4, the piecewise exponential model with two cut points at 12 and 18 months was the best fitting model. Model fit statistics for network 1 are given in Table 13. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

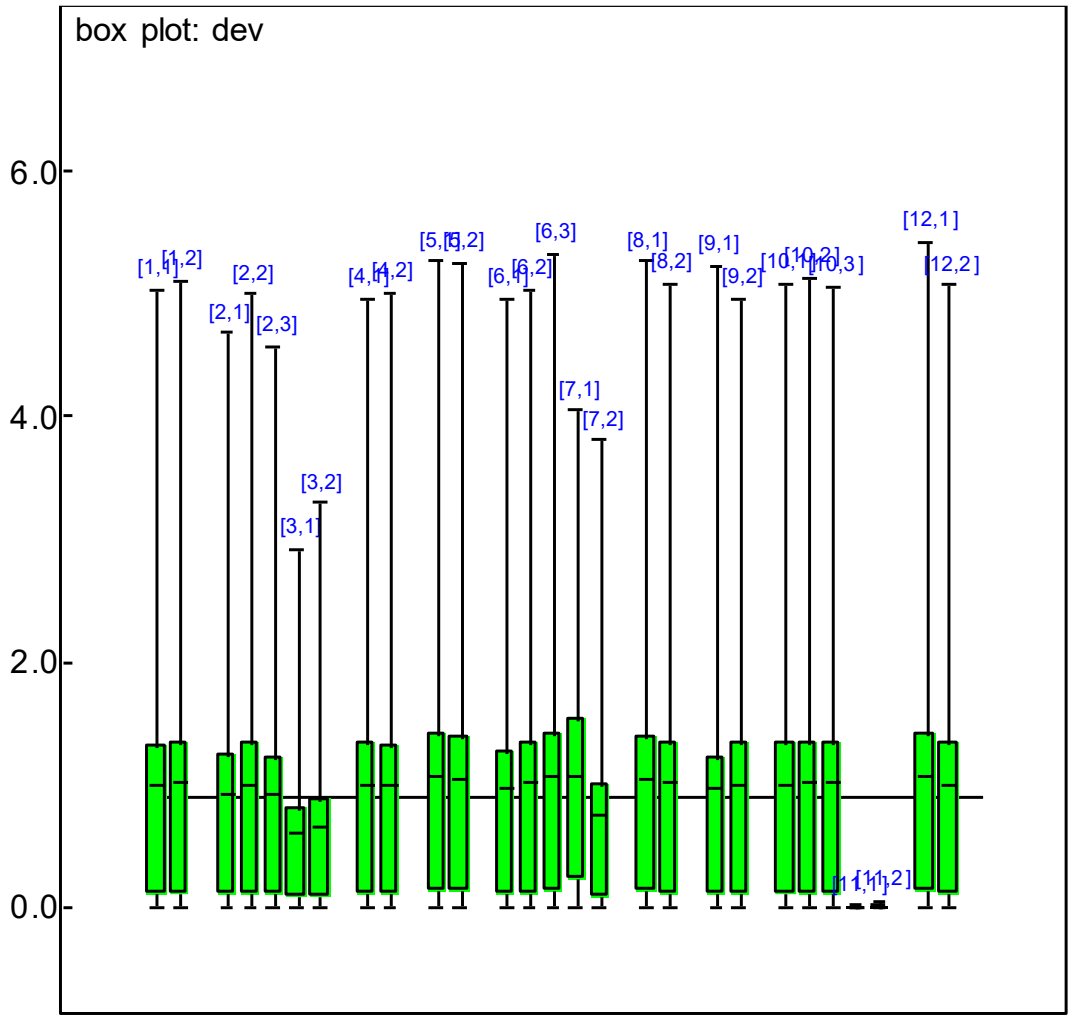
Table 13: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect	144.818	22.537	167.355

In Figure 27, we can see that all studies have normal (or as expected) contributions to the residual deviance. Notably, CheckMate 069, has a relatively small contribution to the residual deviance over its third interval (see [11,1] & [11,2]). This is likely due to CheckMate 069 having only 24 months of follow-up, meaning in the third interval for the aggregate data spans from just 18-24 months. Thus, there is relatively little data here, which in turn equates

to a smaller relative contribution to the residual deviance. Overall, it appears the predictions from the NMA are a good fit to the observed data points.

Figure 27: box plot of the deviance contribution for each arm of aggregated data to the residual deviance – from WinBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. There are two numbers in the brackets above each box plot. The first number corresponds to a line of aggregated data (for clarity we present the numbers and their corresponding study here: 1, 5, & 9 = CheckMate 066; 2, 6 & 10 = CheckMate 067; 3, 7 & 11 = CheckMate 069; 4, 8 & 12 = KEYNOTE-006) and the piecewise interval. Numbers 1-4 correspond to the first interval (1-12 months), numbers 5-8 correspond to the second interval (13-18 months) and numbers 9-12 correspond to the third interval (19-120 months). The second number corresponds to the study arm. For example, [1,1] corresponds to aggregate data for CheckMate 066, in the first interval, for arm 1; [10,3] corresponds to aggregate data for CheckMate 067, in the third interval, for arm 3. The horizontal line indicates a contribution to residual deviance of 1, which is expected from each arm.

It was not possible to assess inconsistency in this network as it only had a single loop, which due to the coding, disappeared under the treatment labelling we used.

Table 14: Fixed effect PFS NMA results (piecewise exponential model with 2 cut points at 12 and 18 months)

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		

ipilimumab vs. DTIC	0.691	(0.509; 0.939)
nivolumab vs. DTIC	0.351	(0.276; 0.444)
nivolumab + ipilimumab vs. DTIC	0.267	(0.193; 0.368)
pembrolizumab vs. DTIC	0.312	(0.214; 0.457)
Interval 2: 13-18 months		
ipilimumab vs. DTIC	0.435	(0.06; 3.732)
nivolumab vs. DTIC	0.131	(0.023; 0.968)
nivolumab + ipilimumab vs. DTIC	0.133	(0.018; 1.172)
pembrolizumab vs. DTIC	0.449	(0.049; 5.629)
Interval 3: 19-120 months		
ipilimumab vs. DTIC	2.537	(0.397; 51.111)
nivolumab vs. DTIC	1.46	(0.253; 28.531)
nivolumab + ipilimumab vs. DTIC	1.001	(0.159; 19.648)
pembrolizumab vs. DTIC	2.725	(0.35; 62.992)

Table 14 presents the model predicted coefficients for each pairwise comparison and Figure 28 presents the predicted PFS survival curves.

- First interval: 0-12 months
    - The hazard ratio in the first interval was smallest for nivolumab + ipilimumab, then pembrolizumab monotherapy and finally nivolumab monotherapy. These hazard ratios were 0.267 (95% CI 0.193-0.368), 0.312 (95% CI 0.214-0.457), 0.351 (95% CI 0.276-0.444) respectively. These point estimates can be interpreted as a 73.3%, 68.8% and 64.9% reduction in the chance of disease progression respectively for those treated with nivolumab + ipilimumab, pembrolizumab monotherapy and nivolumab monotherapy compared with those treated with DTIC.
- Second interval: 12-18 months
    - For the second interval, nivolumab monotherapy had the smallest hazard ratio at 0.131 (95% CI 0.023-0.968). For nivolumab + ipilimumab and pembrolizumab monotherapy, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the second interval was 0.133 (95% CI 0.018-1.172) and 0.449 (95% CI 0.049-5.629) for nivolumab + ipilimumab and pembrolizumab monotherapy respectively. These point estimates can be interpreted as an 86.9%, 86.7% and 55.1%, reduction in the chance of disease progression respectively for those treated with nivolumab monotherapy, nivolumab + ipilimumab and pembrolizumab monotherapy compared with those treated with DTIC.
- Third interval: >18 months
    - For the third interval, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the second interval was smallest for nivolumab + ipilimumab, then nivolumab monotherapy and finally pembrolizumab monotherapy. These hazard ratios were 1.001 (95% CI 0.159-19.648), 1.46 (95% CI 0.253-28.531), 2.725 (95% CI 0.35-62.992)

respectively. These point estimates can be interpreted as a 0.1%, 46% and 172.5% increase in the chance of disease progression respectively for those treated with nivolumab + ipilimumab, nivolumab monotherapy and pembrolizumab monotherapy compared with those treated with DTIC.

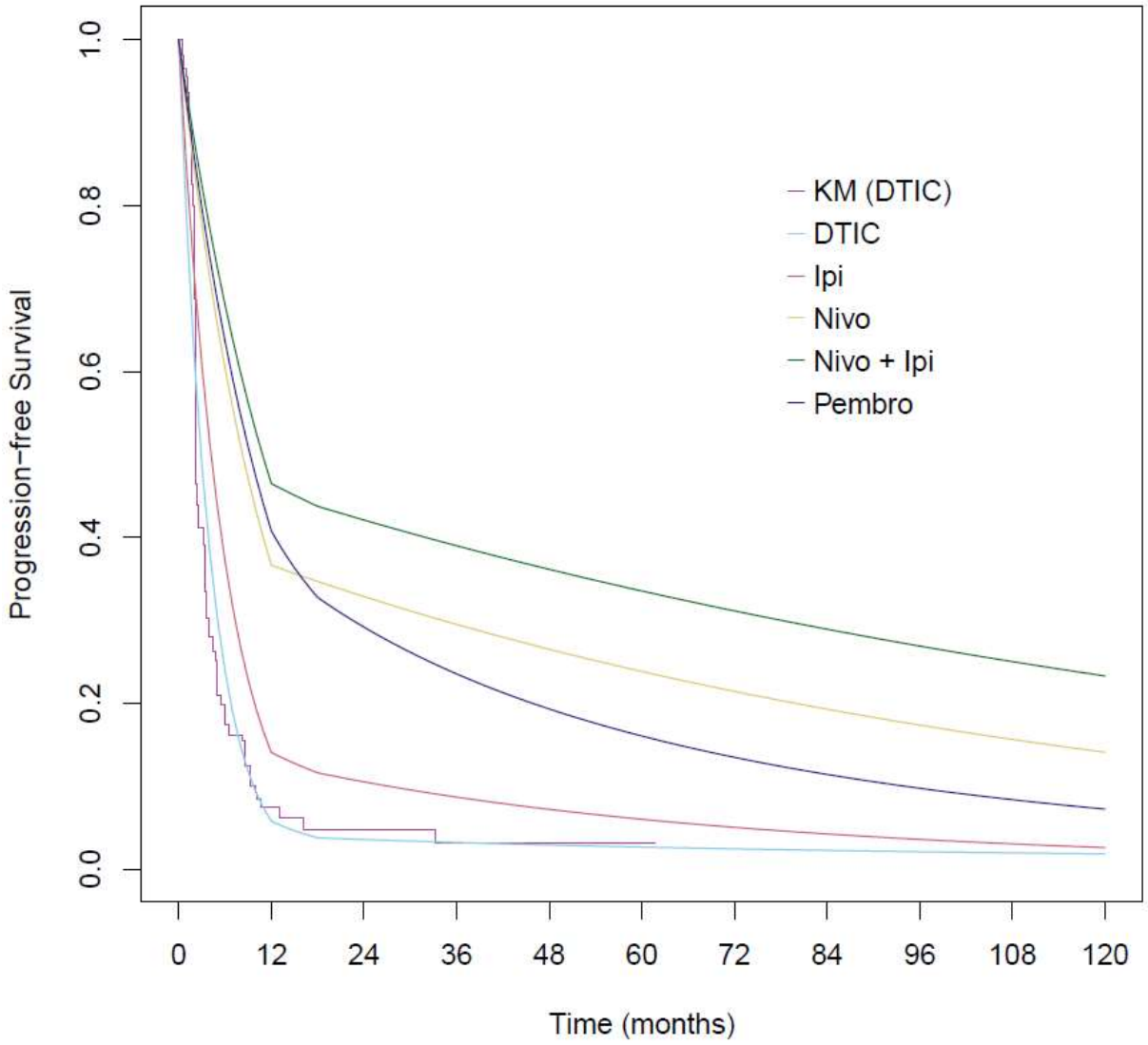


Figure 28: Survival curves for piecewise PFS model with two cut points at 12 and 18 months

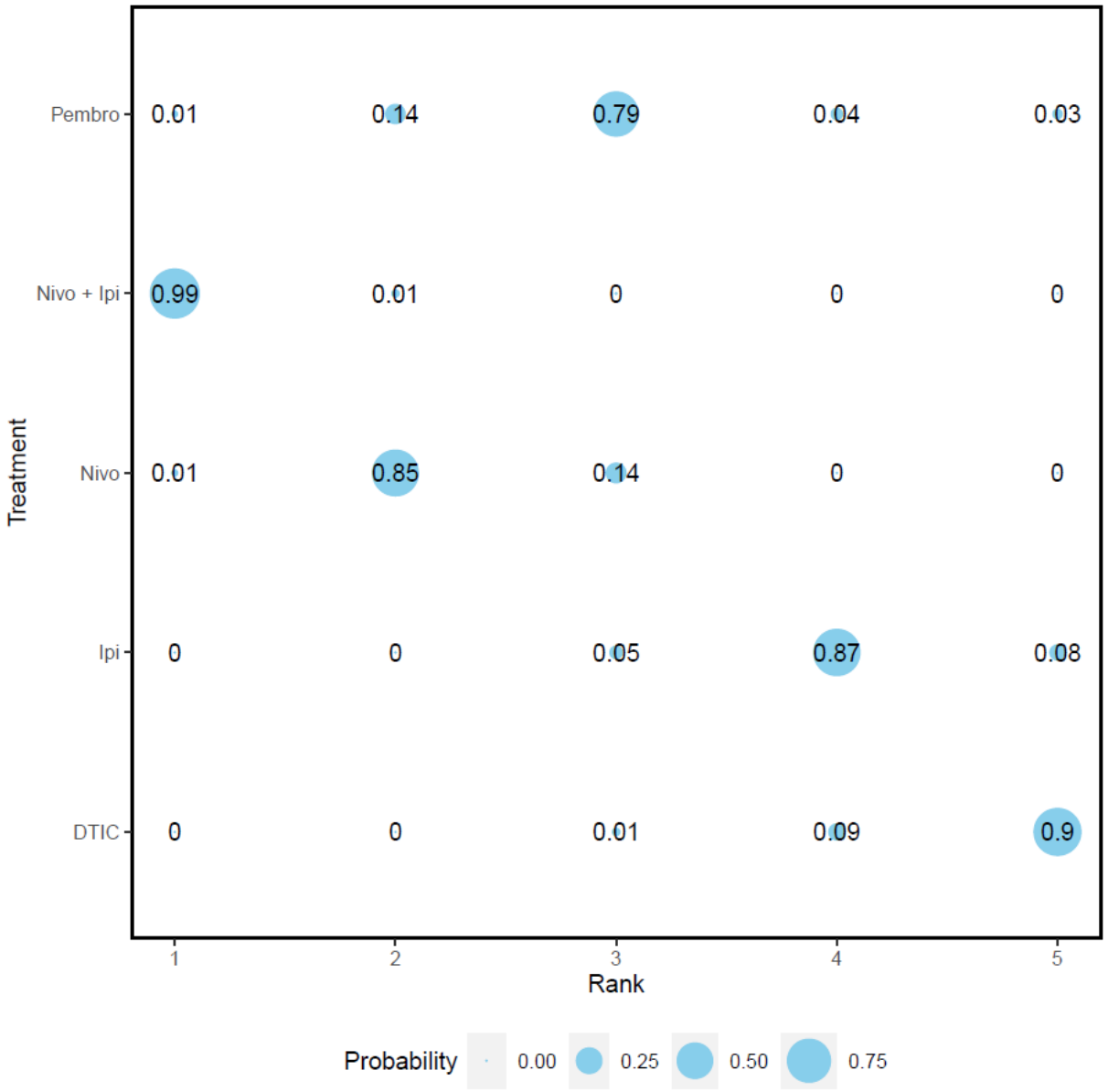


Figure 29: Ranking plot for survival at 60 months for piecewise PFS model with two cut points at 12 and 18 months

These results are reflected in the ranking plots of progression-free survival at 60 months Figure 29:

- Nivolumab + ipilimumab had an average ranking of 1.19 with a 99% probability of being the best treatment. This ranking is driven by nivolumab + ipilimumab having a low hazard ratio for all three intervals. Though in the second interval the point estimate of the hazard ratio for nivolumab + ipilimumab is worse than nivolumab monotherapy, it is not enough for nivolumab + ipilimumab to dip below nivolumab. In the final interval, the point estimate of the hazard ratio for nivolumab monotherapy is worse than nivolumab + ipilimumab so it is not possible for it to have a greater proportion of people who are progression free.
- Nivolumab monotherapy had an average ranking of 2.13 with a 1% probability of being the best treatment. This ranking is explained above as to why nivolumab monotherapy is worse than nivolumab + ipilimumab. Nivolumab monotherapy ends up with a better ranking than pembrolizumab monotherapy as nivolumab monotherapy has a hazard ratio only slightly larger than pembrolizumab monotherapy

in the first interval and has a smaller hazard ratio than pembrolizumab in the final 2 intervals. Therefore, nivolumab comes out better in the treatment rankings.

- Pembrolizumab monotherapy had an average ranking of 2.97 with a 1% probability of being the best treatment. Pembrolizumab monotherapy appears the worst treatment for of the treatments of interest for progression-free survival. This is already covered above but is primarily driven by the hazard ratios for the second and third interval which push it far below the other treatments.

A.5.2.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

For Network 5, the piecewise exponential model with two cut points at 12 and 18 months was the best fitting model. Model fit statistics for network 5 are given in Table 15. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

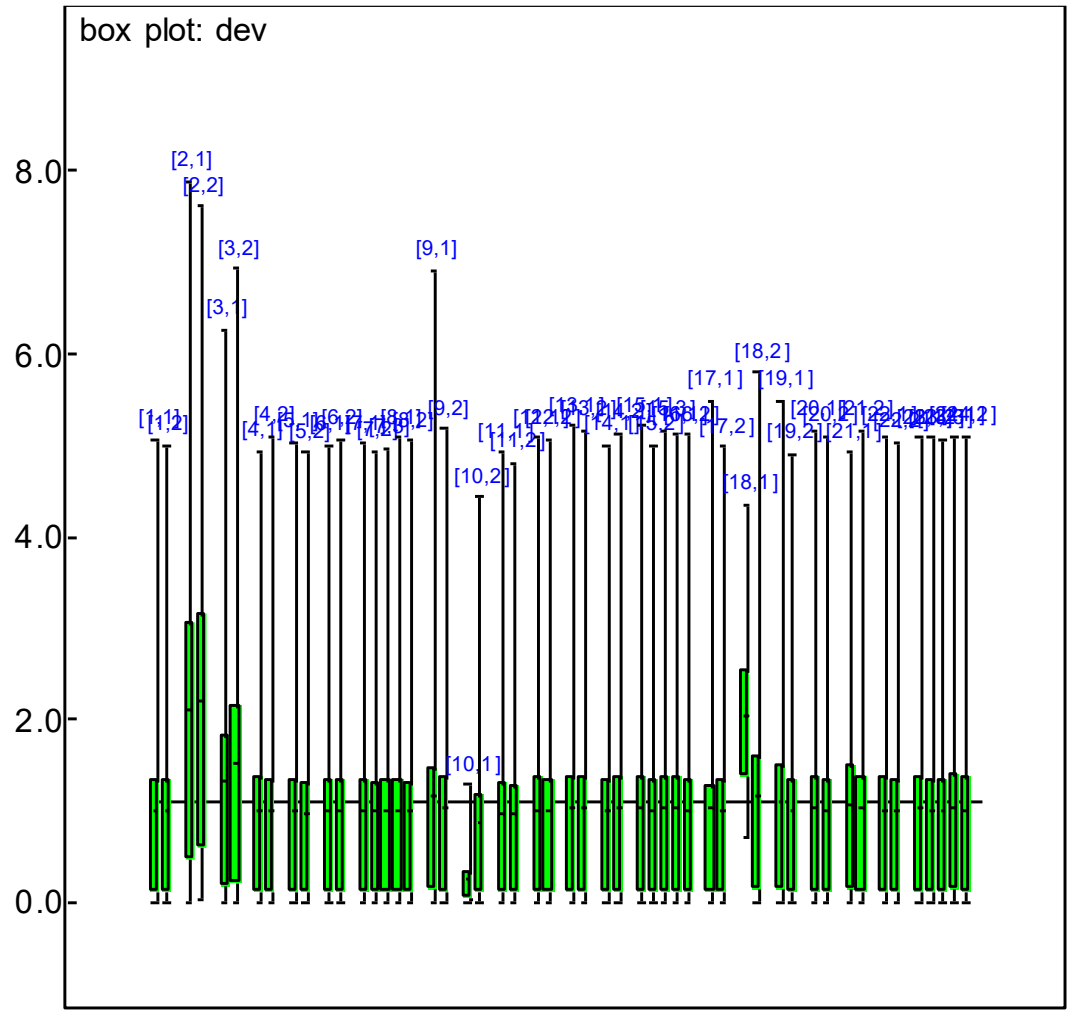
Table 15: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect	304.835	47.267	352.102

In Figure 30, we can see that neither arm of the BRF113220 data is predicted well over the first interval (1-12 months). Additionally, the first arm (dabrafenib monotherapy) of

BRF113220 is not predicted well over the third interval (19-120 months). The remainder of the studies have normal (or as expected) contributions to the residual deviance.

Figure 30: box plot of the deviance contribution for each arm of aggregated data to the residual deviance – from WinBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. There are two numbers in the brackets above each box plot. The first number corresponds to a line of aggregated data (note the line of aggregated data in the excel.csv does not always match to the study – for instance line 6 of the aggregated data refers to study 5, as such, for clarity we present the numbers and their corresponding study here: 1, 9, & 17 = BREAK-3; 2, 10 & 18 = BRF113220; 3, 11 & 19 = COMBI-d; 4, 12 & 20 = COMBI-v; 5, 13 & 21 = CheckMate 066; 6, 14 & 22 = COLUMBUS; 7, 15 & 23 = CheckMate 067; 8, 16 & 24 = KEYNOTE-006) and the piecewise interval. Numbers 1-8 correspond to the first interval (1-12 months), numbers 9-16 correspond to the second interval (13-18 months) and numbers 17-24 correspond to the third interval (19-120 months). The second number corresponds to the study arm. For example, [2,1] corresponds to aggregate data for BRF113220, in the first interval, for arm 1; [23,3] corresponds to aggregate data for CheckMate 067, in the third interval, for arm 3. The horizontal line indicates a contribution to residual deviance of 1, which is expected from each arm.

It was not possible to assess inconsistency in this network as it only had a single loop which was composed of a three-arm trial, and it is not possible for a three-arm trial to be inconsistent with itself. Effectively, this means the network is a long chain of evidence.

Table 16: Fixed effect PFS NMA results (piecewise exponential model with 2 cut points at 12 and 18 months

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		
dabrafenib vs. DTIC	0.4	(0.29; 0.559)
dabrafenib + trametinib vs. DTIC	0.269	(0.18; 0.404)
ipilimumab vs. DTIC	0.687	(0.507; 0.938)
vemurafenib vs. DTIC	0.44	(0.279; 0.702)
nivolumab vs. DTIC	0.351	(0.277; 0.445)
nivolumab + ipilimumab vs. DTIC	0.269	(0.196; 0.372)
encorafenib + binimetinib vs. DTIC	0.238	(0.138; 0.415)
pembrolizumab vs. DTIC	0.311	(0.213; 0.454)
Interval 2: 13-18 months		
dabrafenib vs. DTIC	1.631	(0.22; 67.627)
dabrafenib + trametinib vs. DTIC	0.862	(0.102; 39.134)
ipilimumab vs. DTIC	0.402	(0.056; 3.633)
vemurafenib vs. DTIC	1.046	(0.113; 46.805)
nivolumab vs. DTIC	0.131	(0.023; 0.956)
nivolumab + ipilimumab vs. DTIC	0.138	(0.02; 1.194)
encorafenib + binimetinib vs. DTIC	0.669	(0.06; 30.447)
pembrolizumab vs. DTIC	0.415	(0.045; 4.904)
Interval 3: 19-120 months		
dabrafenib vs. DTIC	0.455	(0.12; 3.083)
dabrafenib + trametinib vs. DTIC	0.402	(0.092; 2.869)
ipilimumab vs. DTIC	2.635	(0.337; 40.731)
vemurafenib vs. DTIC	0.613	(0.126; 4.716)
nivolumab vs. DTIC	1.536	(0.227; 22.511)
nivolumab + ipilimumab vs. DTIC	1.055	(0.141; 15.927)
encorafenib + binimetinib vs. DTIC	0.607	(0.103; 5.392)
pembrolizumab vs. DTIC	2.763	(0.288; 50.199)

It was not possible to assess inconsistency in this network as it only had a single loop which was composed of a three-arm trial, and it is not possible for a three-arm trial to be inconsistent with itself. Effectively, this means the network is a long chain of evidence.

Table 16 presents the model predicted coefficients for each pairwise comparison and Figure 31 presents the predicted PFS survival curves.

- First interval: 0-12 months
    - The hazard ratio in the first interval was smallest for encorafenib + binimetinib, then tied between dabrafenib + trametinib and nivolumab + ipilimumab, then pembrolizumab monotherapy and finally nivolumab monotherapy. These hazard ratios were 0.238 (95% CI 0.138-0.415), 0.269 (95% CI 0.18-0.404), 0.269 (95% CI 0.196-0.372), 0.311 (95% CI 0.213-0.454), and 0.351 (95% CI 0.277-0.445) respectively. These point estimates can be interpreted as a 76.2%, 73.1%, 73.1%, 68.9% and 64.9% reduction in the chance of disease

progression respectively for those treated with encorafenib + binimetinib, dabrafenib + trametinib, nivolumab + ipilimumab, pembrolizumab monotherapy and nivolumab monotherapy compared with those treated with DTIC.

- Second interval: 12-18 months
    - For the second interval, nivolumab monotherapy had the smallest hazard ratio at 0.131 (95% CI 0.023-0.956). For nivolumab + ipilimumab, pembrolizumab monotherapy, encorafenib + binimetinib and dabrafenib + trametinib there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the second interval was 0.138 (95% CI 0.023-0.956), 0.415 (95% CI 0.045-4.904), 0.669 (95% CI 0.06-30.447), 0.862 (95% CI 0.102-39.134) for nivolumab + ipilimumab, pembrolizumab monotherapy, encorafenib + binimetinib and dabrafenib + trametinib respectively. These point estimates can be interpreted as an 86.9%, 86.2%, 58.5%, 33.1% and 13.8% reduction in the chance of disease progression respectively for those treated with nivolumab monotherapy, nivolumab +ipilimumab, pembrolizumab monotherapy, encorafenib + binimetinib, and dabrafenib + trametinib compared with those treated with DTIC.
- Third interval: >18 months
    - For the third interval, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the third interval was smallest for dabrafenib + trametinib, then encorafenib + binimetinib, then nivolumab + ipilimumab, then nivolumab monotherapy, and finally pembrolizumab monotherapy. These hazard ratios were 0.402 (95% CI 0.092-2.869), 0.607 (95% CI 0.103-5.392), 1.055 (95% CI 0.141-15.927), 1.536 (95% CI 0.227-22.511), and 2.763 (95% CI 0.288-50.199) respectively. These point estimates can be interpreted as an 59.8% and 39.3% reduction in the chance of disease progression respectively for those treated with dabrafenib + trametinib and encorafenib + binimetinib respectively and a 5.5%, 163.5% and 176.3% increase in the chance of disease progression respectively for those treated with nivolumab + ipilimumab, nivolumab monotherapy and pembrolizumab monotherapy, compared with those treated with DTIC.

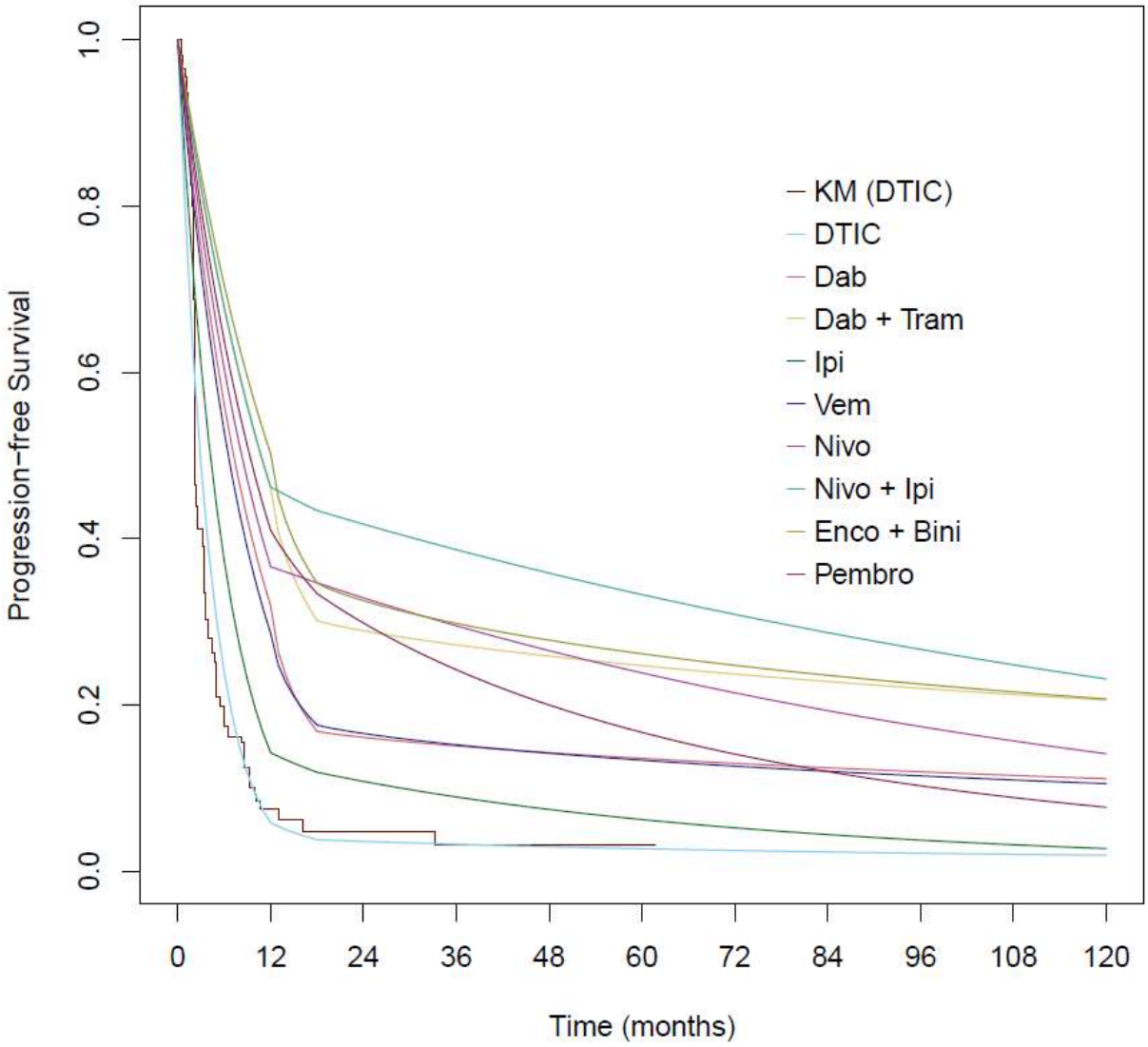


Figure 31: Survival curves for piecewise PFS model with two cut points at 12 and 18 months

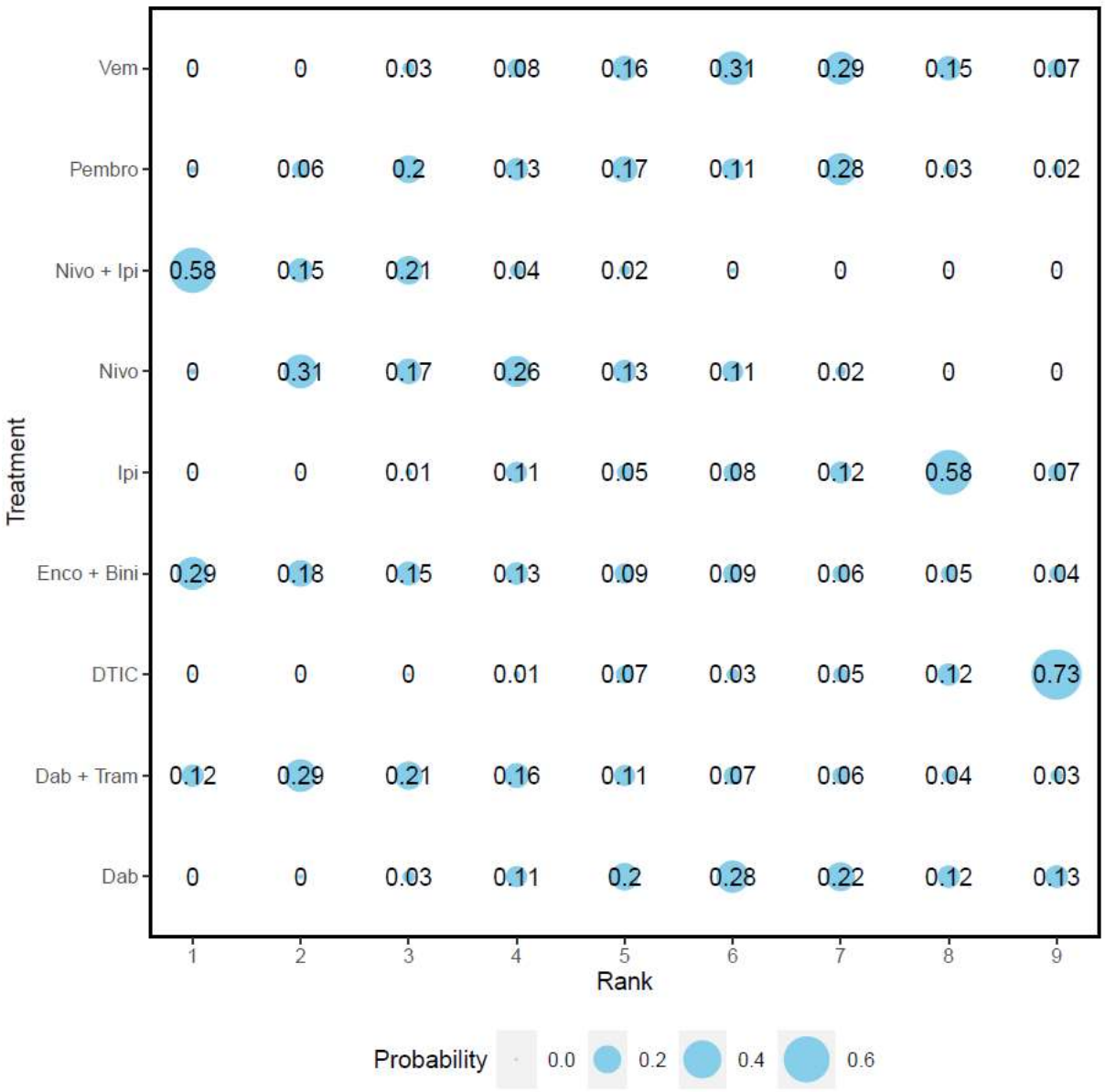


Figure 32: Ranking plot for survival at 60 months for piecewise PFS model with two cut points at 12 and 18 months

These results are reflected in the ranking plots of progression-free survival at 60 months Figure 32:

- Nivolumab + ipilimumab had an average ranking of 1.77 with a 58% probability of being the best treatment. This ranking is driven by nivolumab + ipilimumab having a low hazard ratio for both the first and second interval. This means by 18 months, nivolumab + ipilimumab has the greatest proportion of patients who are progression free, which is maintained until 120 months, even if the proportion of people progression free does decrease further from months 18-120 drawing it closer to both dabrafenib + trametinib and encorafenib + binimetinib.

- Nivolumab monotherapy had an average ranking of 3.62 with a 0% probability of being the best treatment. This ranking can be explained with two important pieces of information. Nivolumab has a worse hazard ratio in the first interval than either dabrafenib + trametinib or encorafenib + binimetinib. Although dabrafenib + trametinib has a worse hazard ratio in the second interval, it has a better hazard ratio in the final interval. This means that while dabrafenib + trametinib has a greater proportion of

progression events than nivolumab monotherapy from approximately month 15-48 months, dabrafenib + trametinib is better than nivolumab monotherapy at >48 months. This explains why nivolumab monotherapy appears worse than the top three treatments, nivolumab + ipilimumab, encorafenib + binimetinib and dabrafenib + trametinib. Nivolumab monotherapy does appear better than pembrolizumab monotherapy and this is because the hazard ratio for pembrolizumab monotherapy in the second and third interval is worse than nivolumab monotherapy. Therefore, while in the first interval pembrolizumab monotherapy had a smaller hazard ratio compared with nivolumab, due to the second and third interval hazard ratios nivolumab monotherapy overtakes pembrolizumab monotherapy and has a greater proportion of people progression-free.

- Pembrolizumab monotherapy had an average ranking of 6.21 with a 0% probability of being the best treatment. Pembrolizumab monotherapy appears the worst treatment of the treatments of interest for progression-free survival. This is already covered above but is primarily driven by the hazard ratios for the second and third interval which push it far below the other treatments.

- Encorafenib + binimetinib had an average ranking of 3.79 with a 29% probability of being the best treatment. As already discussed briefly with the description of nivolumab + ipilimumab, encorafenib + binimetinib has the smallest hazard ratio for the first interval. However, it has worse hazard ratios for the second interval which pushes it below nivolumab + ipilimumab. Furthermore, the hazard ratio for encorafenib + binimetinib for the third interval is far worse than the hazard ratio for dabrafenib + trametinib which is why the two treatments come together by 120 months.

- Dabrafenib + trametinib had an average ranking of 3.95 with a 12% probability of being the best treatment. As already discussed, the hazard ratio for dabrafenib + trametinib for intervals 1 and 2 is why it is pushed below both nivolumab + ipilimumab and encorafenib + binimetinib. However, dabrafenib + trametinib's hazard ratio for the third interval is better than the hazard ratio for encorafenib + binimetinib which is why you see the two curves come together by month 120.

A.5.2.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

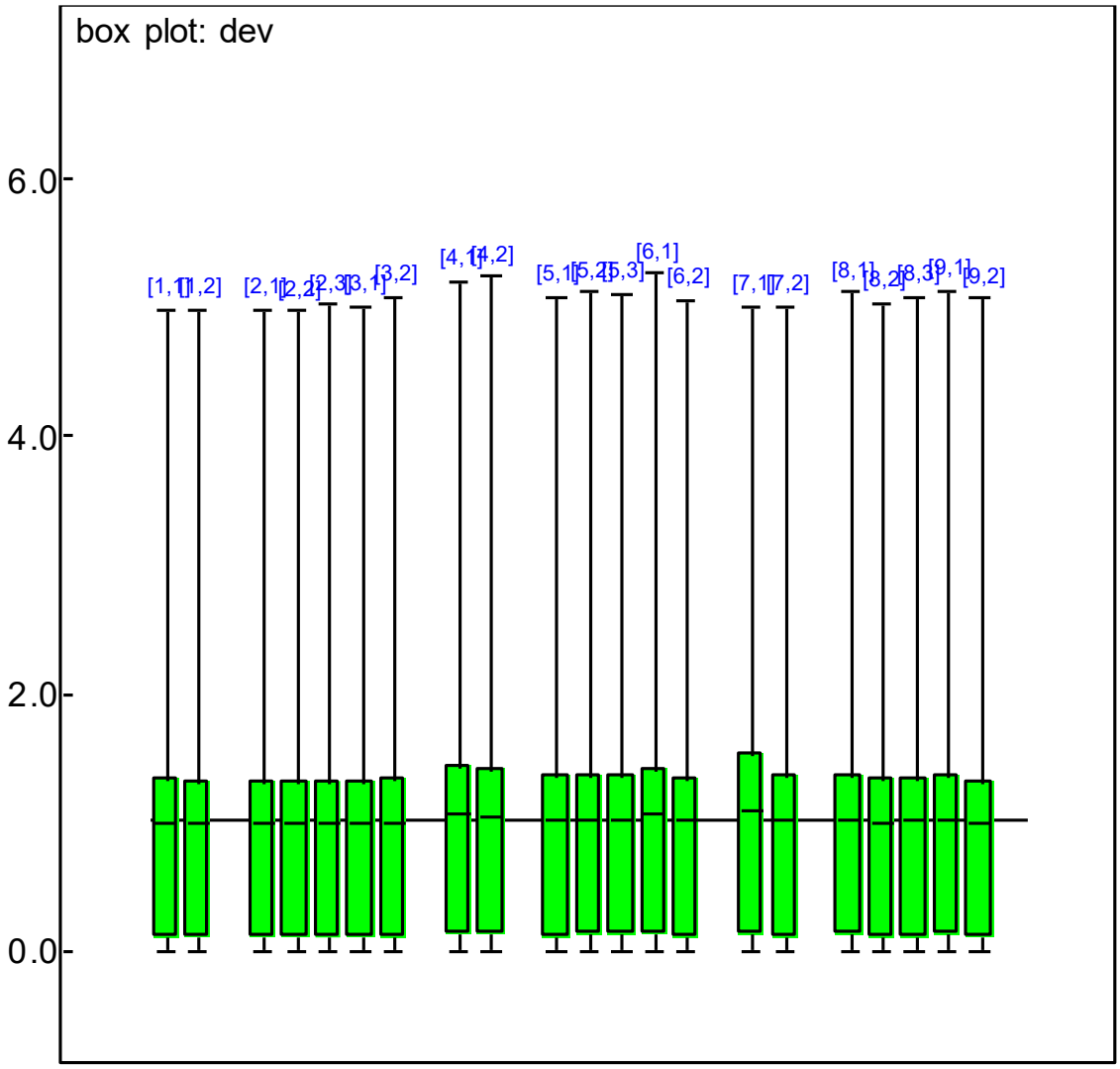
For Network 6, the piecewise exponential model with two cut points at 12 and 18 months was the best fitting model. Model fit statistics for network 6 are given in Table 17. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

Table 17: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect	127.132	20.697	147.829

In Figure 33, we can see that all studies have normal (or as expected) contributions to the residual deviance. Thus, it appears the predictions from the NMA are a good fit to the observed data points.

Figure 33: box plot of the deviance contribution for each arm of aggregated data to the residual deviance – from WinBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. There are two numbers in the brackets above each box plot. The first number corresponds to a line of aggregated data (for clarity we present the numbers and their corresponding study here: 1, 4, & 7 = CheckMate 066; 2, 5 & 8 = CheckMate 067; 3, 6 & 9 = KEYNOTE-006) and the piecewise interval. Numbers 1-3 correspond to the first interval (1-12 months), numbers 4-6 correspond to the second interval (13-18 months) and numbers 7-9 correspond to the third interval (19-120 months). The second number corresponds to the study arm. For example, [1,1] corresponds to aggregate data for CheckMate 066, in the first interval, for arm 1; [8,3] corresponds to aggregate data for CheckMate 067, in the third interval, for arm 3. The horizontal line indicates a contribution to residual deviance of 1, which is expected from each arm.

It was not possible to assess inconsistency in this network as it only had a single loop which was composed of a three-arm trial, and it is not possible for a three-arm trial to be inconsistent with itself.

Table 18: Fixed effect PFS NMA results (piecewise exponential model with 2 cut points at 12 and 18 months)

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		
ipilimumab vs. DTIC	0.686	(0.507; 0.928)
nivolumab vs. DTIC	0.351	(0.277; 0.443)
nivolumab + ipilimumab vs. DTIC	0.269	(0.194; 0.37)
pembrolizumab vs. DTIC	0.31	(0.213; 0.451)
Interval 2: 13-18 months		
ipilimumab vs. DTIC	0.427	(0.06; 3.747)
nivolumab vs. DTIC	0.136	(0.024; 0.986)
nivolumab + ipilimumab vs. DTIC	0.145	(0.02; 1.266)
pembrolizumab vs. DTIC	0.443	(0.047; 5.15)
Interval 3: 19-120 months		
ipilimumab vs. DTIC	2.385	(0.305; 44.523)
nivolumab vs. DTIC	1.4	(0.212; 24.361)
nivolumab + ipilimumab vs. DTIC	0.957	(0.131; 17.832)
pembrolizumab vs. DTIC	2.537	(0.256; 53.678)

Table 18 presents the model predicted coefficients for each pairwise comparison and Figure 34 presents the predicted PFS survival curves.

- First interval: 0-12 months
  - The hazard ratio in the first interval was smallest for nivolumab + ipilimumab, then pembrolizumab monotherapy and finally nivolumab monotherapy. These hazard ratios were 0.269 (95% CI 0.194-0.37), 0.31 (95% CI 0.213-0.451), 0.351 (95% CI 0.277-0.443) respectively. These point estimates can be interpreted as a 73.1%, 69% and 64.9% reduction in the chance of disease progression respectively for those treated with nivolumab + ipilimumab, pembrolizumab monotherapy and nivolumab monotherapy compared with those treated with DTIC.
- Second interval: 12-18 months
  - For the second interval, nivolumab monotherapy had the smallest hazard ratio at 0.136 (95% CI 0.024-0.986). For nivolumab + ipilimumab and pembrolizumab monotherapy, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the second interval was 0.145 (95% CI 0.02-1.266) and 0.443 (95% CI 0.047-5.15) for nivolumab + ipilimumab and pembrolizumab monotherapy respectively. These point estimates can be interpreted as an 86.4%, 85.5% and 55.7%, reduction in the chance of disease progression respectively for those treated with nivolumab monotherapy, nivolumab + ipilimumab and pembrolizumab monotherapy compared with those treated with DTIC.
- Third interval: >18 months
  - For the third interval, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one,

the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the second interval was smallest for nivolumab + ipilimumab, then nivolumab monotherapy and finally pembrolizumab monotherapy. These hazard ratios were 0.957 (95% CI 0.131-17.832), 1.4 (95% CI 0.212-24.361), 2.537 (95% CI 0.256-53.678) respectively. These point estimates can be interpreted as a 4.3% reduction in the chance of disease progression for those treated with nivolumab + ipilimumab and a 40% and 153.7% increase in the chance of disease progression respectively for those treated with nivolumab monotherapy and pembrolizumab monotherapy compared with those treated with DTIC.

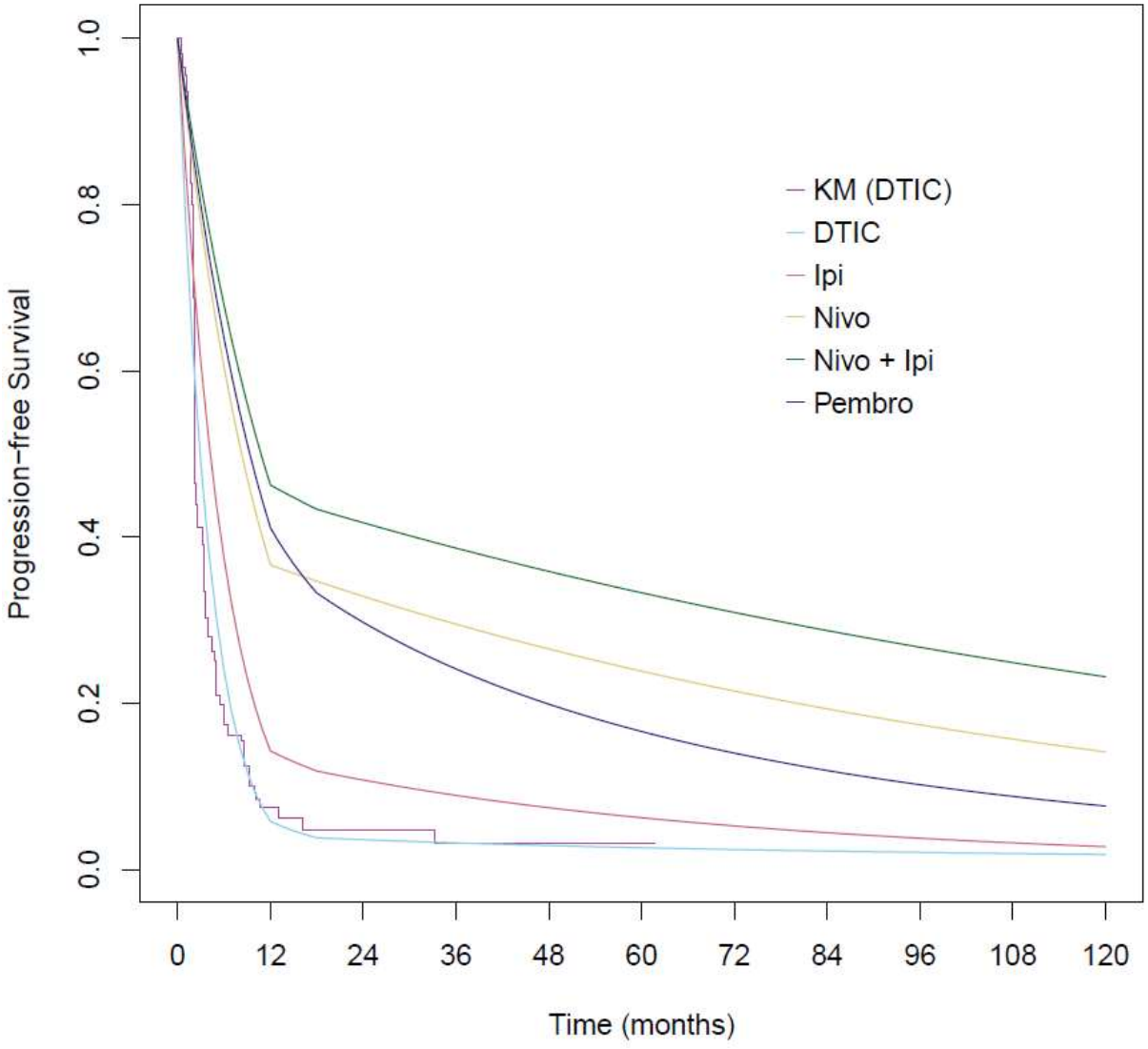


Figure 34: Survival curves for piecewise PFS model with two cut points at 12 and 18 months

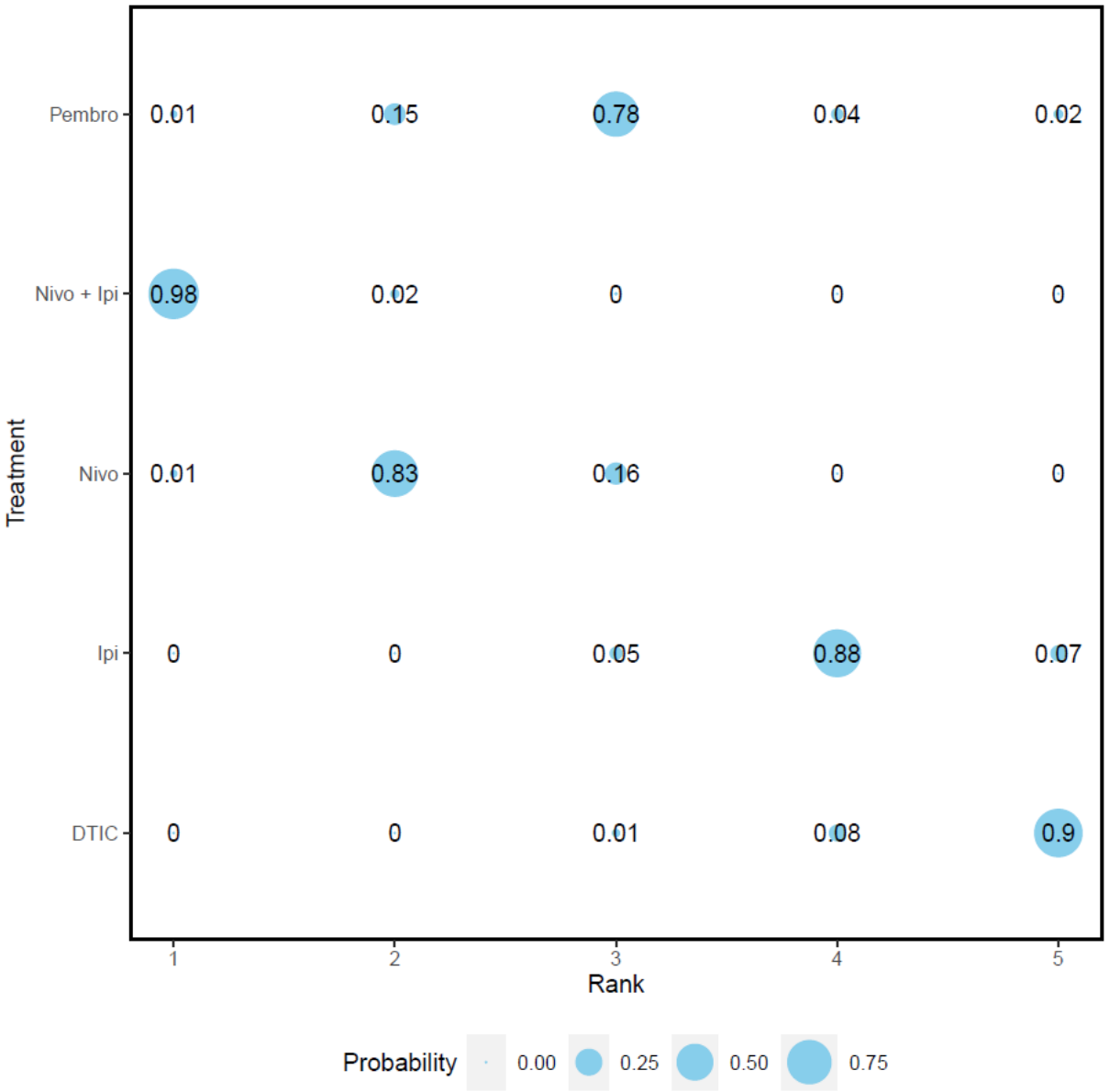


Figure 35: Ranking plot for survival at 60 months for piecewise PFS model with two cut points at 12 and 18 months

These results are reflected in the ranking plots of progression-free survival at 60 months Figure 35:

- Nivolumab + ipilimumab had an average ranking of 1.02 with a 98% probability of being the best treatment. This ranking is driven by nivolumab + ipilimumab having a low hazard ratio for all three intervals. Though in the second interval the point estimate of the hazard ratio for nivolumab + ipilimumab is worse than nivolumab monotherapy, it is not enough for nivolumab + ipilimumab to dip below nivolumab. In the final interval, the point estimate of the hazard ratio for nivolumab monotherapy is worse than nivolumab + ipilimumab so it is not possible for it to have a greater proportion of people who are progression free.
- Nivolumab monotherapy had an average ranking of 2.15 with a 1% probability of being the best treatment. This ranking is explained above as to why nivolumab monotherapy is worse than nivolumab + ipilimumab. Nivolumab monotherapy ends up with a better ranking than pembrolizumab monotherapy as nivolumab monotherapy has a hazard ratio only slightly larger than pembrolizumab monotherapy

in the first interval and has a smaller hazard ratio than pembrolizumab in the final 2 intervals. Therefore, nivolumab comes out better in the treatment rankings.

- Pembrolizumab monotherapy had an average ranking of 2.91 with a 1% probability of being the best treatment. Pembrolizumab monotherapy appears the worst treatment for of the treatments of interest for progression-free survival. This is already covered above but is primarily driven by the hazard ratios for the second and third interval which push it far below the other treatments.

A.5.3 Selected OS model for each Network

In sections A.5.3.1-A.5.3.6 we present the results of the best fitting model for OS for each network. However, we present the results of the models that were not selected in Appendix D, Appendix E, and Appendix F.

A.5.3.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

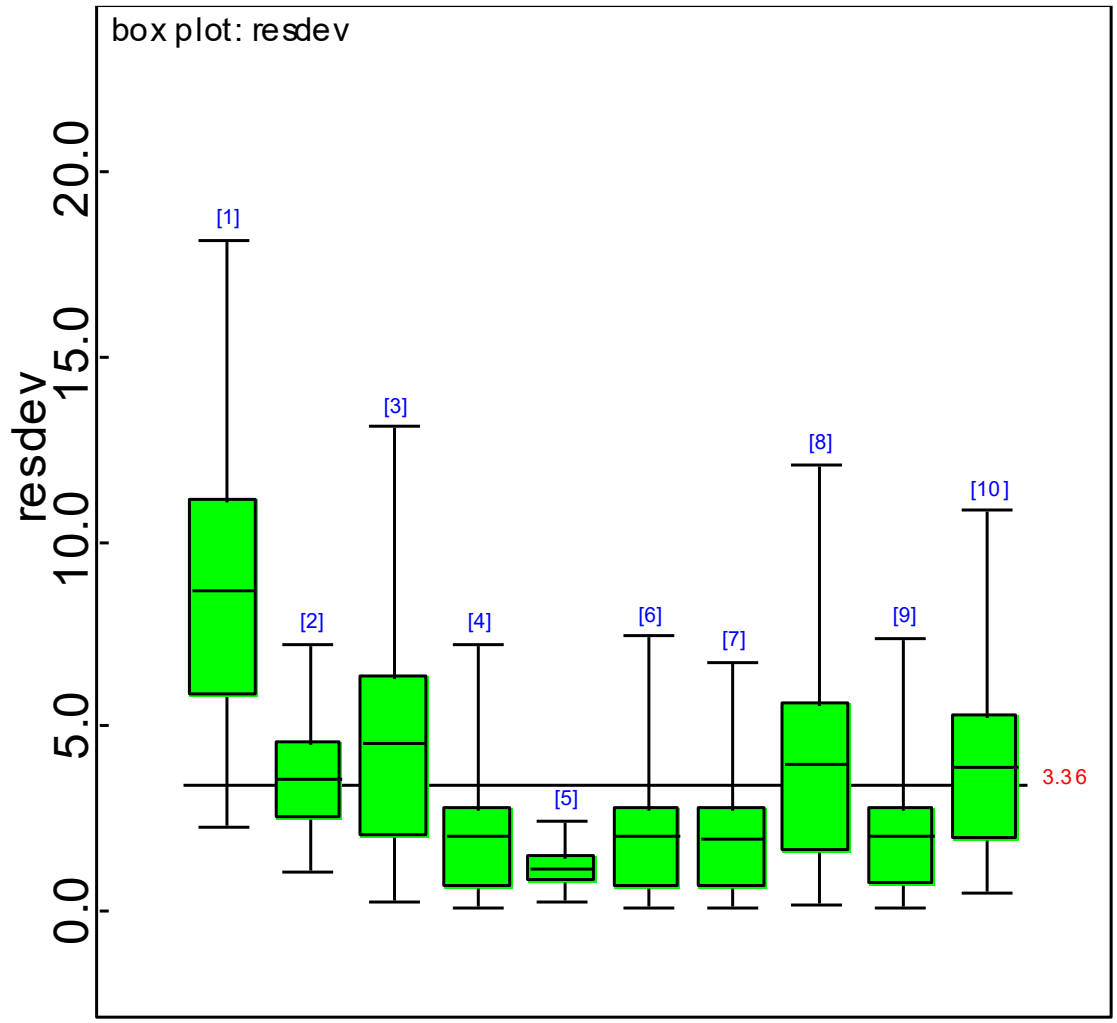
For Network 1, the generalized gamma model with two treatment effects was the best fitting model. Model fit statistics for network 1 are given in Table 19. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

Table 19: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect – consistency	-17.6	16	-1.6
Fixed effect - inconsistency	-30.8	18	-12.8

In Figure 36, we can see that BREAK-3 is not predicted well and has an especially large contribution to the residual deviance. BREAK-3 is a two-arm study comparing DTIC versus dabrafenib monotherapy. Additionally, we can see those studies 2, 3, and 8, or BRF113220, BRIM-3 and COMBI-v respectively have larger than expected contributions to the residual deviance (though smaller in comparison with BREAK-3). BRF113220 is a two-arm trial comparing dabrafenib monotherapy versus dabrafenib + trametinib, BRIM-3 is a two-arm trial comparing DTIC versus vemurafenib monotherapy and COMBI-v is a two-arm trial comparing dabrafenib + trametinib versus vemurafenib monotherapy. In Figure 48 which presents the box plots of the total deviance contribution for each study for network 5 – which drops BRIM-3 and CheckMate 069 from the network as they have shorter follow-up relative to the rest of the trials in the network, we observe normal contributions to the residual deviance from each trial. While it is not entirely clear what is causing this, in dropping BRIM-3 (which is linked to the other targeted therapy trials in network 1 where we see larger contributions to the residual deviance BREAK-3, BRF113220, and COMBI-v), we observe that the predictions from the NMA are a better fit to the observed data points.

Figure 36: box plot of the total deviance contribution for each study to the residual deviance – from OpenBUGS - consistency

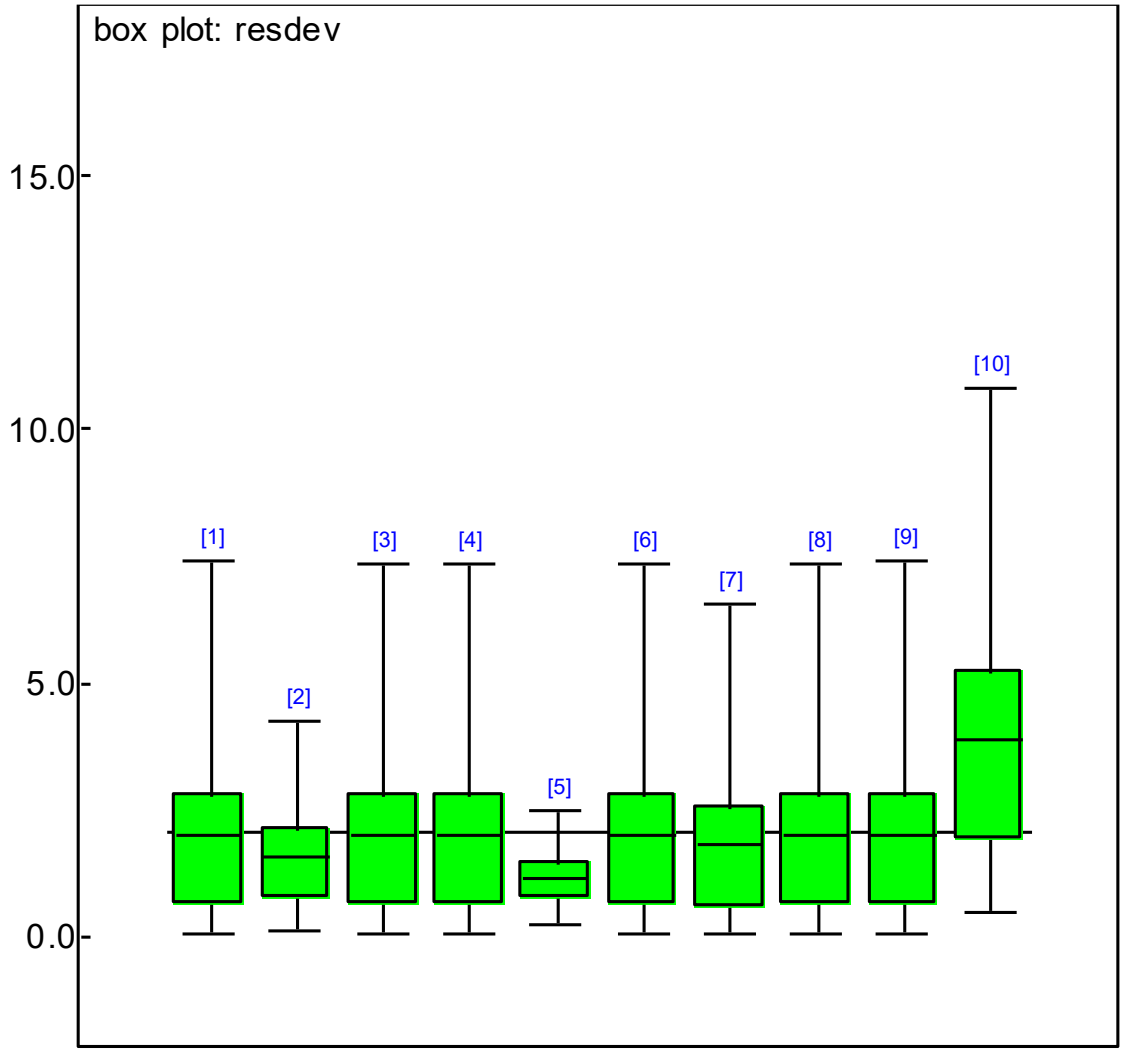


The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. Numbers above the lines represent the study the plot corresponds to (1 = BREAK-3; 2 = BRF113220; 3 = BRIM-3; 4 = CheckMate 066; 5 = CheckMate 069; 6 = COLUMBUS; 7 = COMBI-d; 8 = COMBI-v; 9 = KEYNOTE-006; 10 = CheckMate 067). The horizontal line the mean deviance contribution. For 2-arm trials, we would expect the deviance contribution to be 2 for a good fitting model, since the model sums over 2 outcomes. For 3 arm trials we would expect the deviance contribution to be 4 (2 outcomes x 2 relative effects = 4).

Evidence of inconsistency was found, with model fit and DIC being lower for the inconsistency model (Table 19).

In Figure 37, we can see that all studies have normal (or as expected) contributions to the residual deviance. Thus, it appears the predictions from the inconsistency model are a better fit to the observed data points compared with the consistency model.

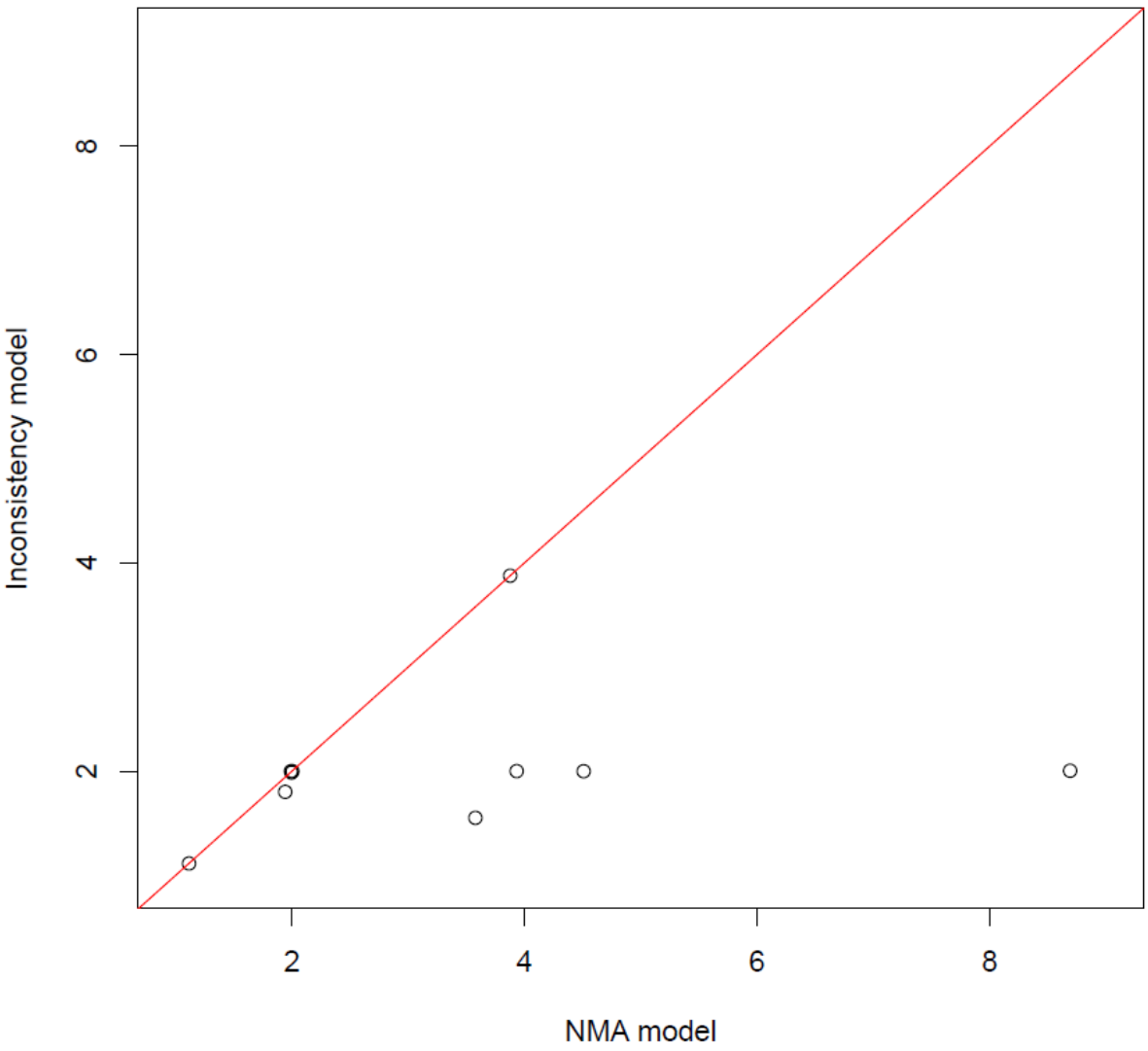
Figure 37: box plot of the total deviance contribution for each study to the residual deviance – from OpenBUGS - inconsistency



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. Numbers above the lines represent the study the plot corresponds to (1 = BREAK-3; 2 = BRF113220; 3 = BRIM-3; 4 = CheckMate 066; 5 = CheckMate 069; 6 = COLUMBUS; 7 = COMBI-d; 8 = COMBI-v; 9 = KEYNOTE-006; 10 = CheckMate 067). The horizontal line the mean deviance contribution. For 2-arm trials, we would expect the deviance contribution to be 2 for a good fitting model, since the model sums over 2 outcomes. For 3 arm trials we would expect the deviance contribution to be 4 (2 outcomes x 2 relative effects = 4).

The area below the line of equality in Figure 38 highlights where the inconsistency model better predicted data points. The loop between studies 1, 2, 3, and 5 gives inconsistent results, with all four of these points falling below the line of equality.

Figure 38: Deviance contributions from the fixed effect consistency and inconsistency models



In Table 20 we present only the results that were not in agreement between the fixed effect consistency and inconsistency models. Of the 8 comparisons where we can obtain estimates from the inconsistency model (dabrafenib versus DTIC, vemurafenib versus DTIC, nivolumab versus DTIC, vemurafenib versus dabrafenib + trametinib, nivolumab versus ipilimumab, nivolumab + ipilimumab versus ipilimumab, encorafenib + binimetinib versus vemurafenib and pembrolizumab versus ipilimumab) only the results for dabrafenib versus DTIC and vemurafenib versus DTIC weren't in agreement. While the point estimates of the mean of the posterior distribution vary between the two models, highlighting the inconsistency in the network, there is considerable overlap between the credible intervals, which can be seen in Figure 39. Thus, the impact of the observed inconsistency is limited.

Table 20: Discordant results from the fixed effect consistency and inconsistency models

Comparison	Fixed effect – consistency		Fixed effect – inconsistency	
	Mean of posterior distribution	95% Credible Interval	Mean of posterior distribution	95% Credible Interval
Time-Ratio				
dabrafenib vs. DTIC	1.513	(1.181; 1.935)	1.191	(0.832; 1.706)
vemurafenib vs. DTIC	1.54	(1.282; 1.85)	1.716	(1.4; 2.102)
Shape parameter				
dabrafenib vs. DTIC	0.86	(0.717; 1.033)	1.088	(0.846; 1.4)
vemurafenib vs. DTIC	0.707	(0.628; 0.798)	0.657	(0.576; 0.749)

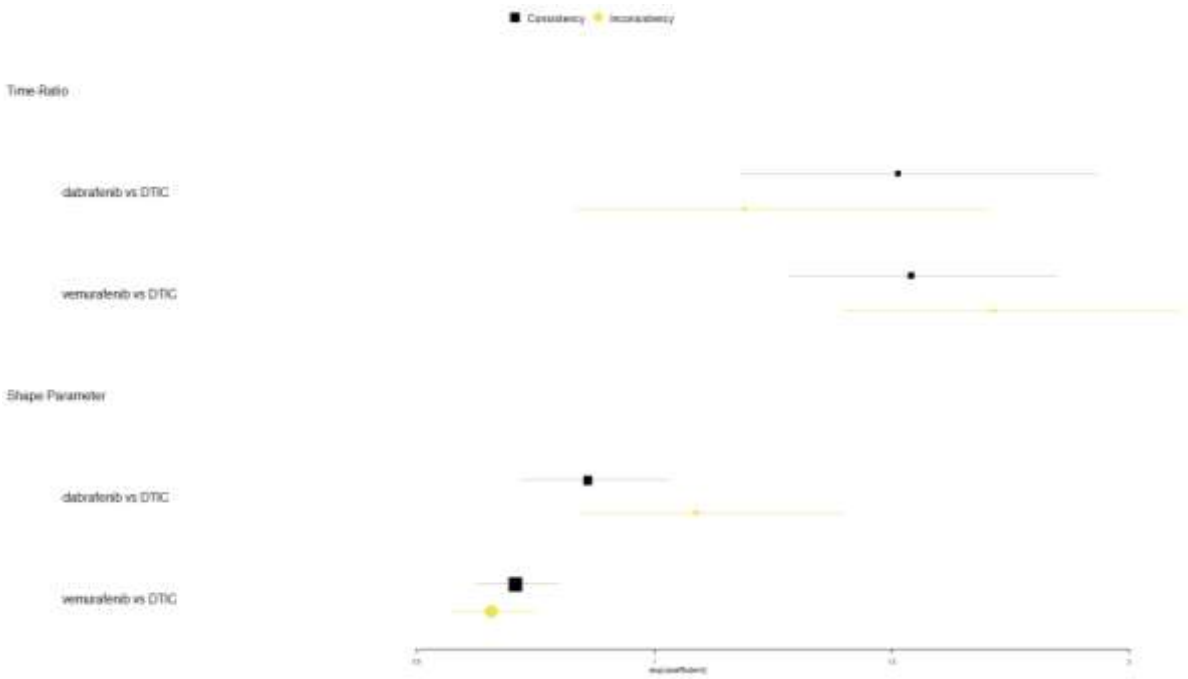


Figure 39: Forrest plots of discordant NMA results from the fixed effect consistency and inconsistency models - highlighting the degree to which the credible intervals overlap

Table 21: Fixed effect OS NMA results for generalized gamma with location and scale parameters

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
dabrafenib vs. DTIC	1.522	(1.181; 1.935)
dabrafenib + trametinib vs. DTIC	2.142	(1.692; 2.68)
ipilimumab vs. DTIC	1.424	(0.8762; 2.188)

Comparison	exp(coefficient)	95% Credible Interval
vemurafenib vs. DTIC	1.545	(1.282; 1.85)
nivolumab vs. DTIC	2.533	(1.739; 3.593)
nivolumab + ipilimumab vs. DTIC	3.343	(1.922; 5.426)
encorafenib + binimetinib vs. DTIC	2.431	(1.73; 3.32)
pembrolizumab vs. DTIC	2.502	(1.362; 4.213)
Shape parameter		
dabrafenib vs. DTIC	0.864	(0.7174; 1.033)
dabrafenib + trametinib vs. DTIC	0.774	(0.6531; 0.9107)
ipilimumab vs. DTIC	1.272	(1.005; 1.593)
vemurafenib vs. DTIC	0.709	(0.6275; 0.7982)
nivolumab vs. DTIC	1.534	(1.282; 1.826)
nivolumab + ipilimumab vs. DTIC	2.067	(1.614; 2.618)
encorafenib + binimetinib vs. DTIC	0.887	(0.7023; 1.106)
pembrolizumab vs. DTIC	1.516	(1.118; 2.013)

Table 21 presents the model predicted coefficients for each pairwise comparison and Figure 40 presents the predicted OS survival curves.

Nivolumab + ipilimumab had the highest point estimate for the scale parameter which can be interpreted as a time ratio 3.343 (95% CI 1.922-5.426). The time ratio can be interpreted as the likelihood of dying comes 3.343 times slower for nivolumab + ipilimumab compared to DTIC (the reference treatment). This in turn means people given nivolumab + ipilimumab live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 2.067 (1.614-2.618) which changes the shape of the survival curve.

Nivolumab monotherapy had the second-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.533 (95% CI 1.739-3.593). The time ratio can be interpreted as the likelihood of dying comes 2.533 times slower for nivolumab monotherapy compared to DTIC (the reference treatment). This in turn means people given nivolumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.534 (1.282-1.826).

Pembrolizumab monotherapy had the third-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.502 (95% CI 1.362-4.213). The time ratio can be interpreted as the likelihood of dying comes 2.502 times slower for pembrolizumab monotherapy compared to DTIC (the reference treatment). This in turn means people given pembrolizumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.516 (1.118-2.013).

Encorafenib + Binimetinib had the fourth-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.431 (95% CI 1.73-3.32). The time ratio can be interpreted as the likelihood of dying comes 2.431 times slower for encorafenib + binimetinib compared to DTIC (the reference treatment). This in turn means people given encorafenib + binimetinib live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 0.887 (0.7023-1.106) respectively.

Dabrafenib + Trametinib had the fifth-best point estimate for the scale parameter which can be interpreted as a time ratio 2.142 (95% CI 1.692-2.68). The time ratio can be interpreted as the likelihood of dying comes 2.142 times slower for dabrafenib + trametinib compared to

DTIC (the reference treatment). This in turn means people given dabrafenib + trametinib live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 0.774 (0.6531-0.9107) respectively.

The other treatments in the network, dabrafenib monotherapy, ipilimumab monotherapy and vemurafenib monotherapy in general had worse point estimates for both the scale and shape parameters compared with dabrafenib + trametinib. This is reflected in the survival curves in Figure 40 where they are some of the lowest curves depicted. Additionally, the committee advised that dabrafenib and vemurafenib are not given as monotherapy, and ipilimumab is not frequently given due to its toxicity. Thus, in addition to performing poorly in the network, they are not clinically prevalent.

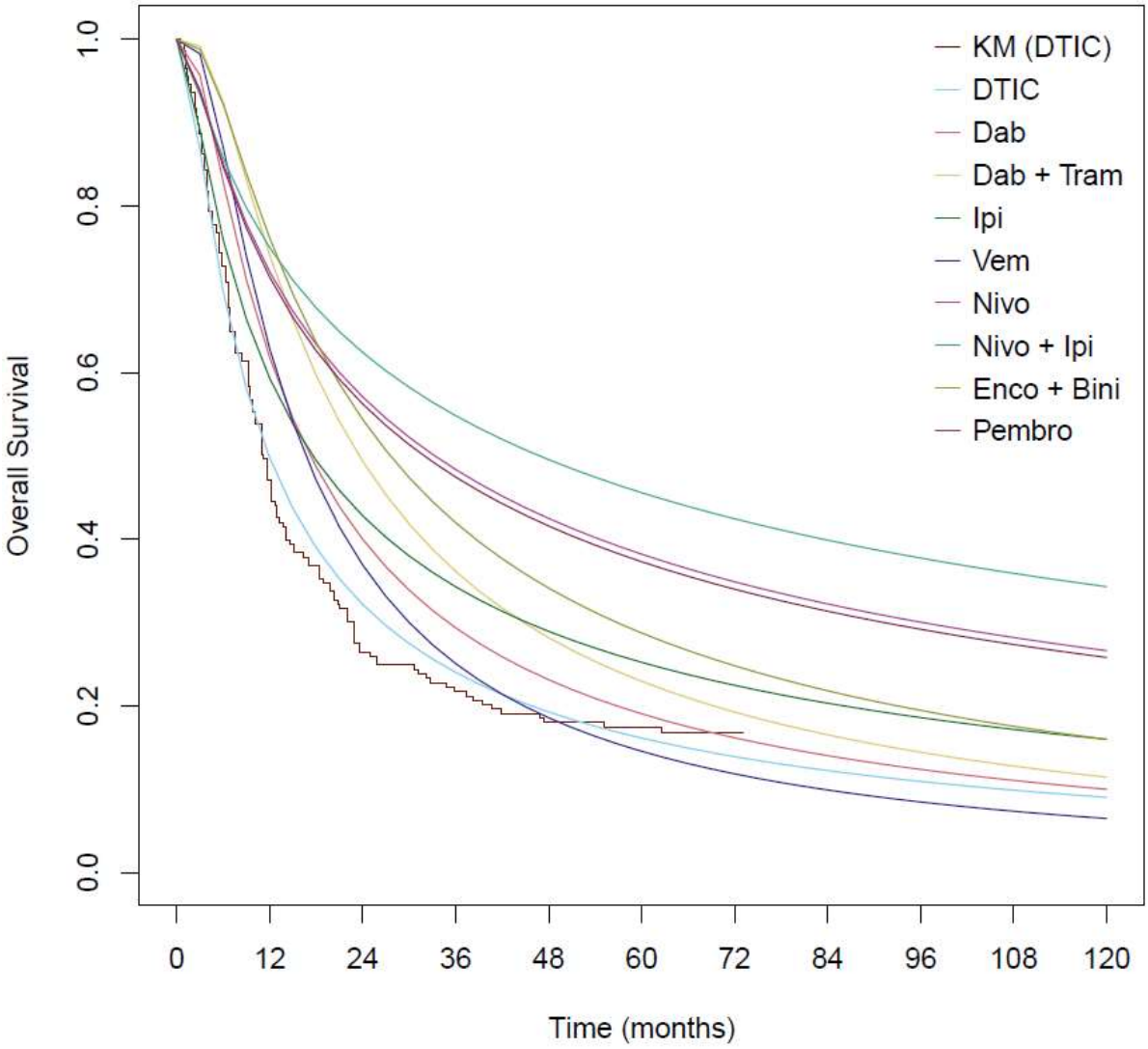


Figure 40: Survival curves for generalized gamma OS model on location and scale parameters

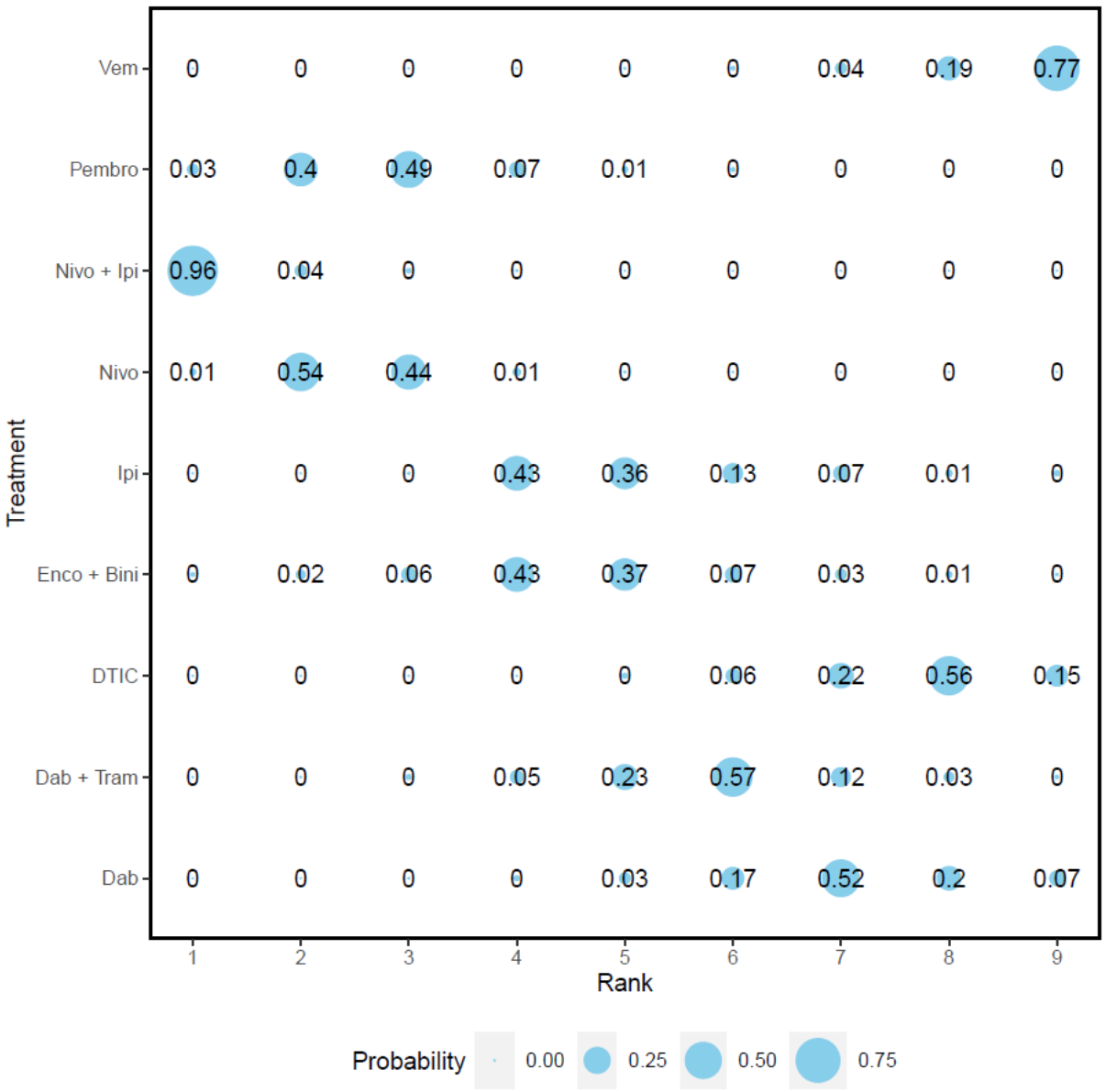


Figure 41: Ranking plot for survival at 60 months for generalized gamma OS model on location and scale parameters

These results are reflected in the ranking plots of overall survival at 60 months Figure 41:

- Nivolumab + ipilimumab had an average ranking of 1.04 with a 96% probability of being the best treatment
- Nivolumab monotherapy had a median ranking of 2.45 with a 1% probability of being the best treatment
- Pembrolizumab monotherapy had a median ranking of 2.63 with a 3% probability of being the best treatment
- Encorafenib + binimetinib had a median ranking of 4.5 with a 0% probability of being the best treatment
- Dabrafenib + trametinib had a median ranking of 5.85 with a 0% probability of being the best treatment

A.5.3.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

A.5.3.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

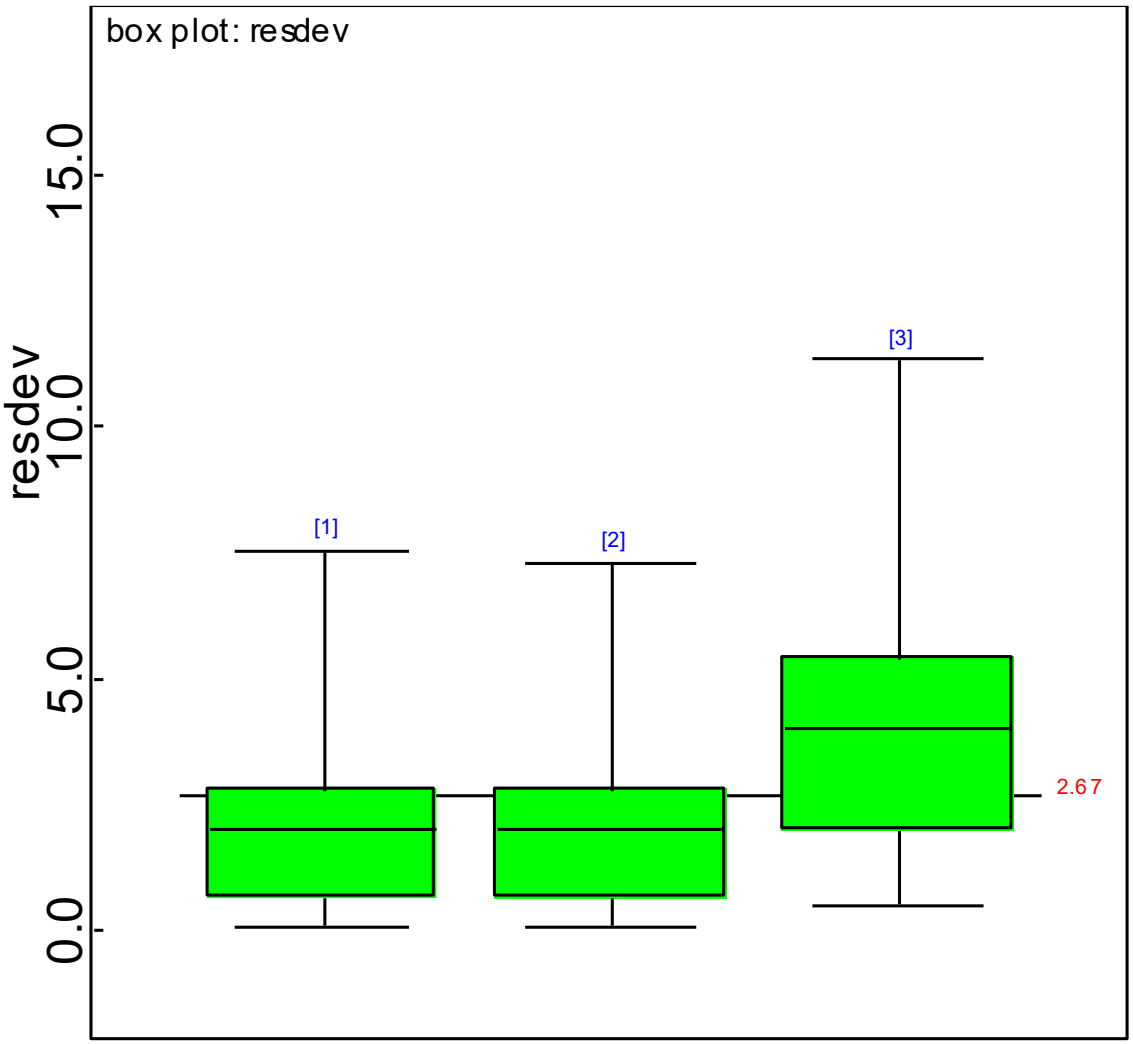
For Network 3, the generalized gamma model with two treatment effects was the best fitting model. Model fit statistics for network 3 are given in Table 22. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

Table 22: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect	-10.51	8.013	-2.497

In Figure 42, we can see that all studies have normal (or as expected) contributions to the residual deviance. Thus, it appears the predictions from the NMA are a good fit to the observed data points.

Figure 42: box plot of the total deviance contribution for each study to the residual deviance – from OpenBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. Numbers above the lines represent the study the plot corresponds to (1 = CheckMate 066; 2 = KEYNOTE-006; 3 = CheckMate 067). The horizontal line the mean deviance contribution. For 2-arm trials, we would expect the deviance contribution to be 2 for a good fitting model, since the model sums over 2 outcomes. For 3 arm trials we would expect the deviance contribution to be 4 (2 outcomes x 2 relative effects = 4).

It was not possible to assess inconsistency in this network as it only had a single loop which was composed of a three-arm trial, and it is not possible for a three-arm trial to be inconsistent with itself.

Table 23: Fixed effect OS NMA results for generalized gamma with location and scale parameters

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
ipilimumab vs. DTIC	1.427	(0.8165; 2.348)

Comparison	exp(coefficient)	95% Credible Interval
nivolumab vs. DTIC	2.532	(1.742; 3.557)
nivolumab + ipilimumab vs. DTIC	3.094	(1.624; 5.421)
pembrolizumab vs. DTIC	2.433	(1.256; 4.315)
Shape parameter		
ipilimumab vs. DTIC	1.092	(0.8433; 1.385)
nivolumab vs. DTIC	1.533	(1.28; 1.822)
nivolumab + ipilimumab vs. DTIC	1.703	(1.301; 2.198)
pembrolizumab vs. DTIC	1.16	(0.8405; 1.563)

Table 23 presents the model predicted coefficients for each pairwise comparison and Figure 43 presents the predicted OS survival curves.

Nivolumab + ipilimumab had the highest point estimate for the scale parameter which can be interpreted as a time ratio 3.094 (95% CI 1.624-5.421). The time ratio can be interpreted as the likelihood of dying comes 3.094 times slower for nivolumab + ipilimumab compared to DTIC (the reference treatment). This in turn means people given nivolumab + ipilimumab live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.703 (95% CI 1.301-2.198) which changes the shape of the survival curve.

Nivolumab monotherapy had the second-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.532 (95% CI 1.742-3.557). The time ratio can be interpreted as the likelihood of dying comes 2.532 times slower for nivolumab monotherapy compared to DTIC (the reference treatment). This in turn means people given nivolumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.533 (1.28-1.822).

Pembrolizumab monotherapy had the third-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.433 (95% CI 1.256-4.315). The time ratio can be interpreted as the likelihood of dying comes 2.433 times slower for pembrolizumab monotherapy compared to DTIC (the reference treatment). This in turn means people given pembrolizumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.16 (0.8405-1.563).

Ipilimumab monotherapy, the other treatment in the network had the worst point estimates for both the scale and shape parameters. This is reflected in Figure 43 where ipilimumab has a worse survival curve than nivolumab + ipilimumab, nivolumab monotherapy or pembrolizumab monotherapy. Additionally, as stated above ipilimumab is not frequently given due to its toxicity. Thus, in addition to performing poorly in the network, ipilimumab is not clinically prevalent.

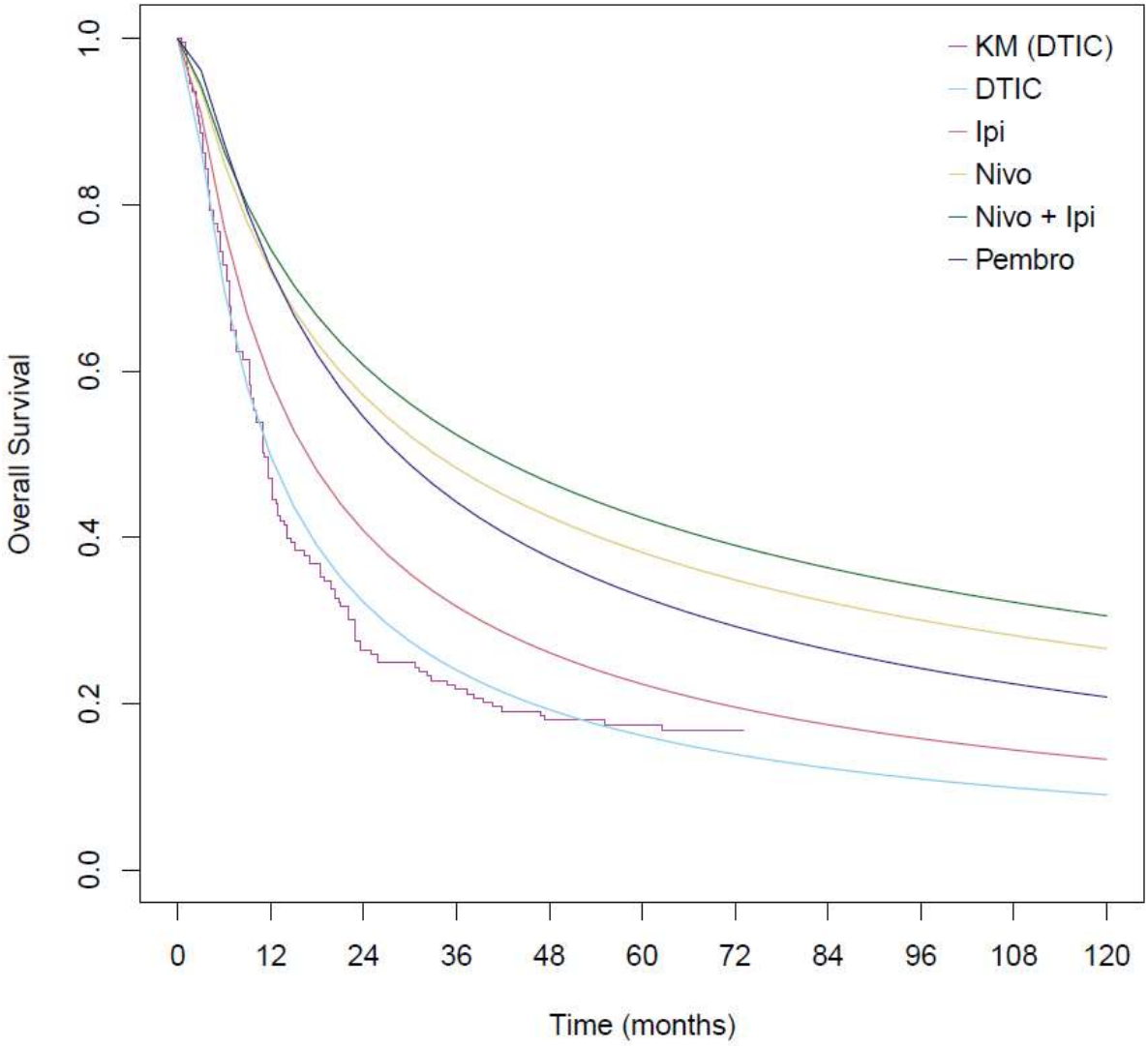


Figure 43: Survival curves for generalized gamma OS model on location and scale parameters

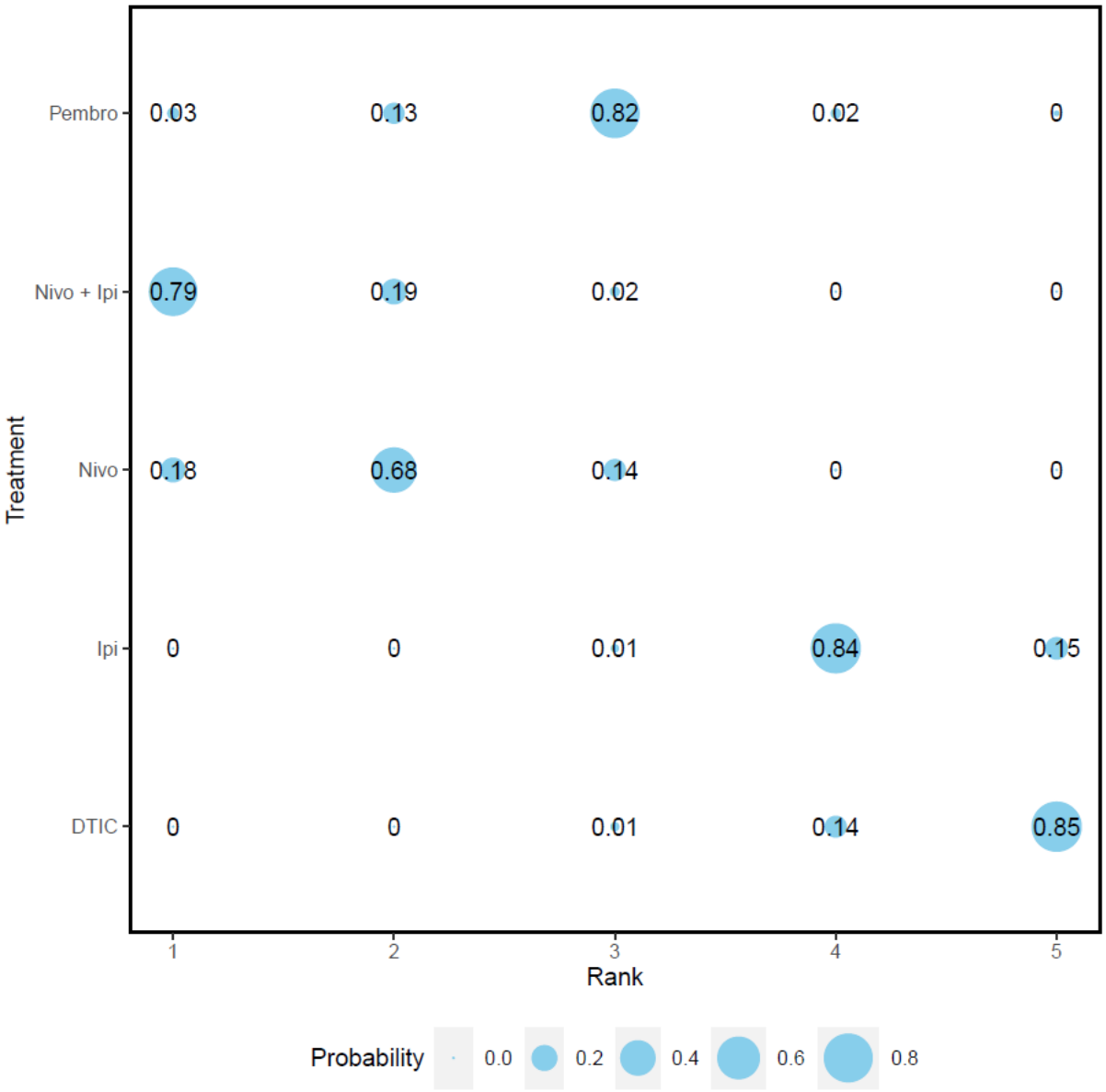


Figure 44: Ranking plot for survival at 60 months for generalized gamma OS model on location and scale parameters

These results are reflected in the ranking plots of overall survival at 60 months Figure 44:

- Nivolumab + ipilimumab had an average ranking of 1.23 with a 79% probability of being the best treatment
- Nivolumab monotherapy had an average ranking of 1.96 with a 18% probability of being the best treatment
- Pembrolizumab monotherapy had an average ranking of 2.83 with a 3% probability of being the best treatment

A.5.3.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

For Network 4, the generalized gamma model with two treatment effects was the best fitting model. Model fit statistics for network 4 are given in Table 24. Due to a limited number of

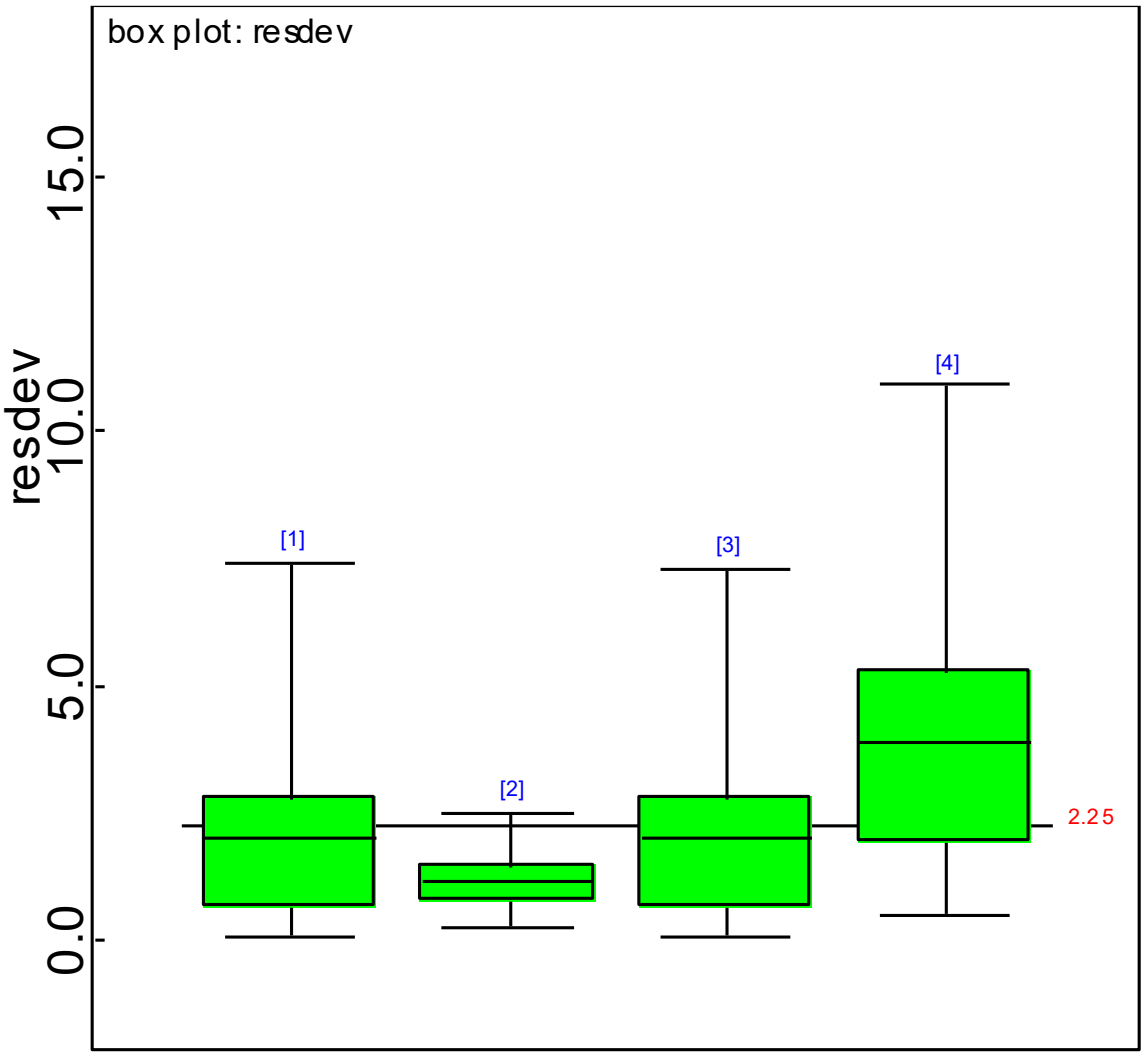
studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

Table 24: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect	-11.793	7.946	-3.847

In Figure 45, we can see that all studies have normal (or as expected) contributions to the residual deviance. Notably, study 2, which corresponds to CheckMate 069, has a relatively small contribution to the residual deviance compared to other studies. This is likely due to CheckMate 069 having only 24 months of follow-up, whereas the other studies had a range in follow-up from 64-75 months. The shorter amount of follow-up equates to less data which in turn equates to a relatively smaller contribution to the residual deviance. Overall, it appears the predictions from the NMA are a good fit to the observed data points.

Figure 45: box plot of the total deviance contribution for each study to the residual deviance – from OpenBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. Numbers above the lines represent the study the plot corresponds to (1 = CheckMate 066; 2 = CheckMate 069; 3 = KEYNOTE-006; 4 = CheckMate 067). The horizontal line the mean deviance contribution. For 2-arm trials, we would expect the deviance contribution to be 2 for a good fitting model, since the model sums over 2 outcomes. For 3 arm trials we would expect the deviance contribution to be 4 (2 outcomes x 2 relative effects = 4).

It was not possible to assess inconsistency in this network as it only had a single loop, which due to the coding, disappeared under the treatment labelling we used.

Table 25: Fixed effect OS NMA results for generalized gamma with location and scale parameters

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
ipilimumab vs. DTIC	1.421	(0.8792; 2.179)
nivolumab vs. DTIC	2.531	(1.745; 3.566)

Comparison	exp(coefficient)	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	3.338	(1.93; 5.397)
pembrolizumab vs. DTIC	2.501	(1.367; 4.186)
Shape parameter		
ipilimumab vs. DTIC	1.27	(1.004; 1.587)
nivolumab vs. DTIC	1.533	(1.28; 1.823)
nivolumab + ipilimumab vs. DTIC	2.064	(1.61; 2.607)
pembrolizumab vs. DTIC	1.515	(1.118; 2.005)

Table 25 presents the model predicted coefficients for each pairwise comparison and Figure 46 presents the predicted OS survival curves.

Nivolumab + ipilimumab had the highest point estimate for the scale parameter which can be interpreted as a time ratio 3.338 (95% CI 1.93-5.397). The time ratio can be interpreted as the likelihood of dying comes 3.338 times slower for nivolumab + ipilimumab compared to DTIC (the reference treatment). This in turn means people given nivolumab + ipilimumab live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 2.064 (95% CI 1.61-2.607) which changes the shape of the survival curve.

Nivolumab monotherapy had the second-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.531 (95% CI 1.745-3.566). The time ratio can be interpreted as the likelihood of dying comes 2.531 times slower for nivolumab monotherapy compared to DTIC (the reference treatment). This in turn means people given nivolumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.533 (1.28-1.823).

Pembrolizumab monotherapy had the third-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.501 (95% CI 1.367-4.186). The time ratio can be interpreted as the likelihood of dying comes 2.501 times slower for pembrolizumab monotherapy compared to DTIC (the reference treatment). This in turn means people given pembrolizumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.515 (1.118-2.005).

Ipilimumab monotherapy, the other treatment in the network had the worst point estimates for both the scale and shape parameters. This is reflected in Figure 46 where ipilimumab has a worse survival curve than nivolumab + ipilimumab, nivolumab monotherapy or pembrolizumab monotherapy. Additionally, as stated above ipilimumab is not frequently given due to its toxicity. Thus, in addition to performing poorly in the network, ipilimumab is not clinically prevalent.

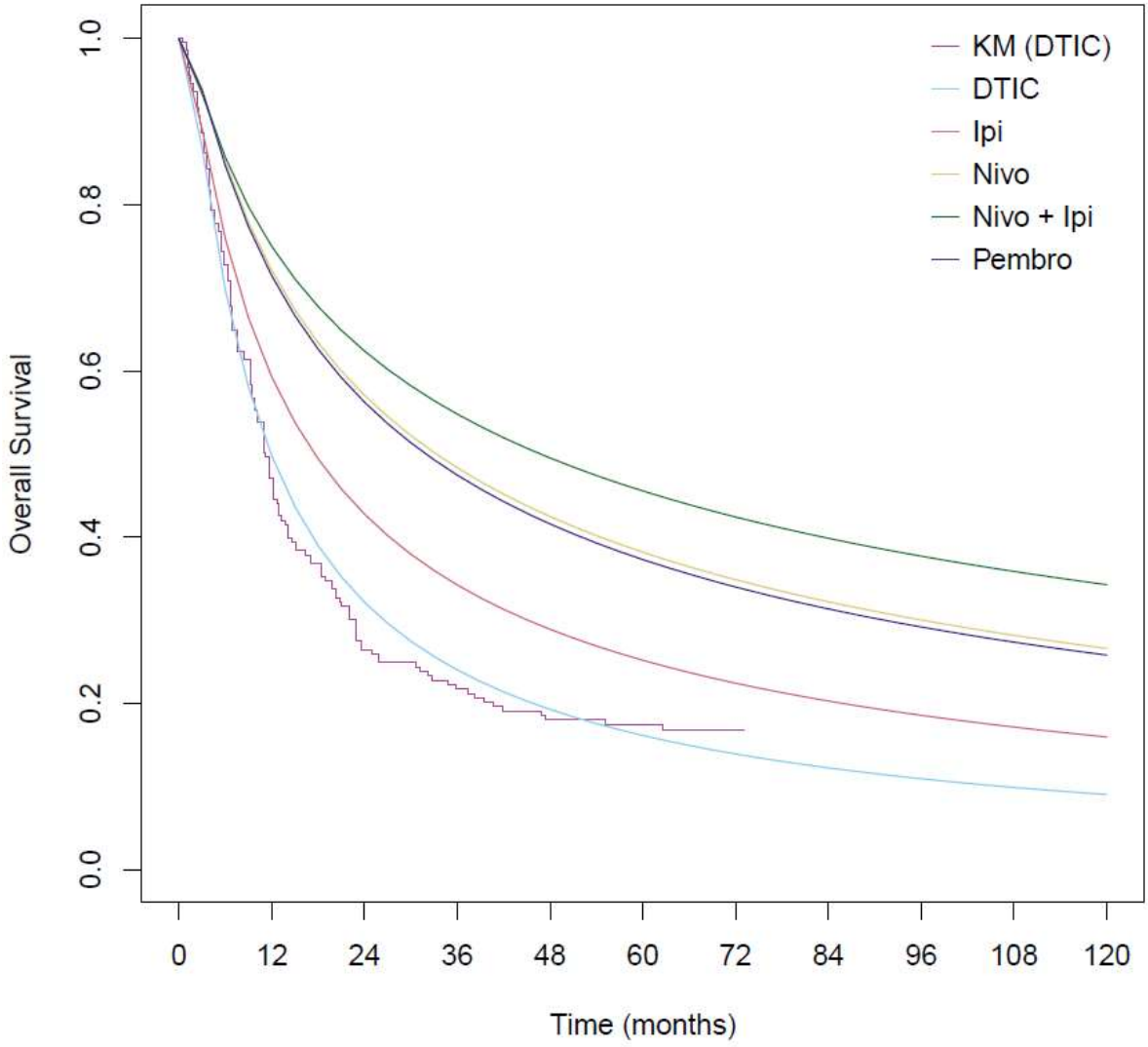


Figure 46: Survival curves for generalized gamma OS model on location and scale parameters

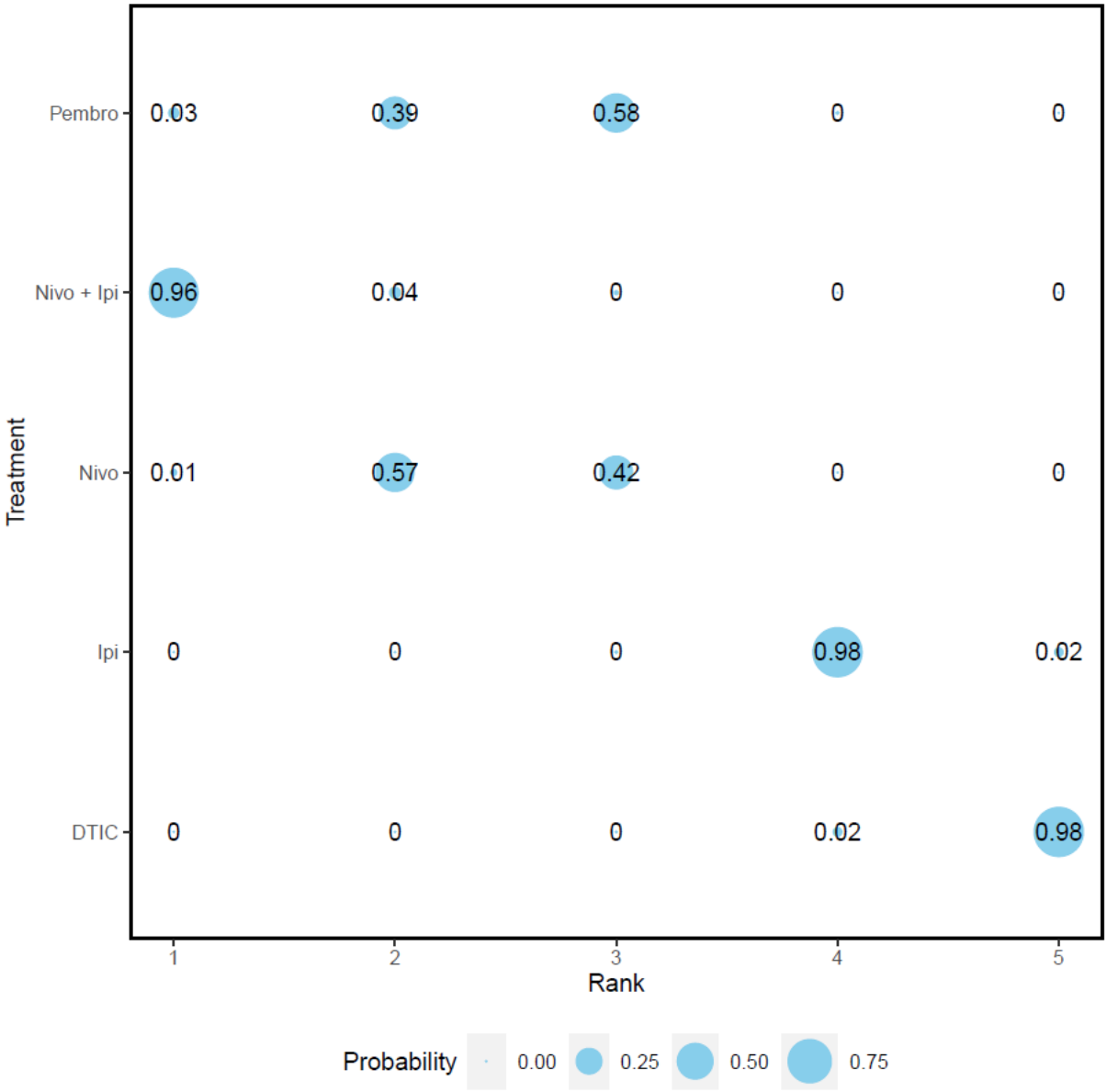


Figure 47: Ranking plot for survival at 60 months for generalized gamma OS model on location and scale parameters

These results are reflected in the ranking plots of progression-free survival at 60 months:

- Nivolumab + ipilimumab had an average ranking of 1.04 with a 96% probability of being the best treatment
- Nivolumab monotherapy had an average ranking of 2.41 with a 1% probability of being the best treatment
- Pembrolizumab monotherapy had an average ranking of 2.55 with a 3% probability of being the best treatment

A.5.3.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

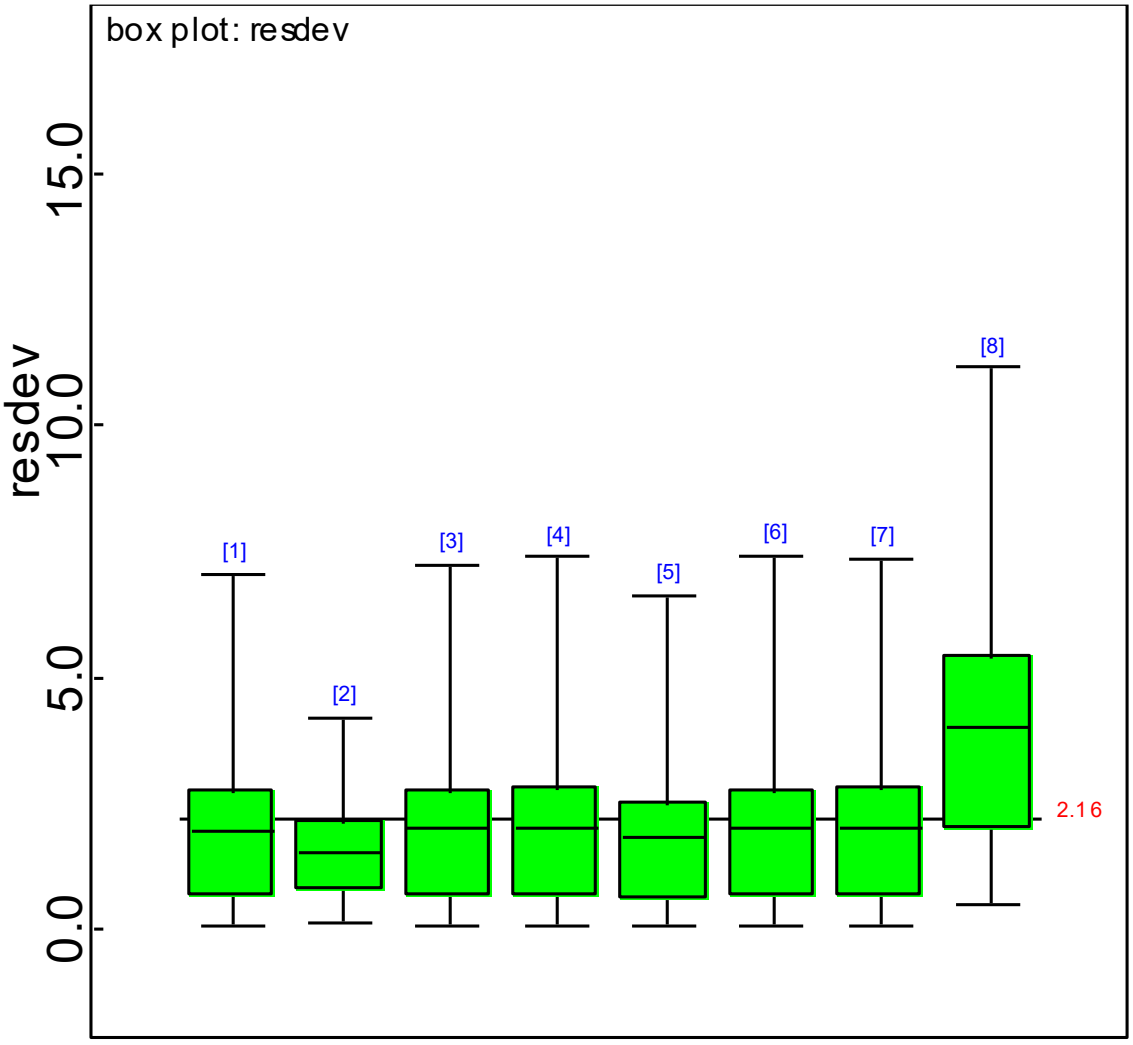
For Network 5, the generalized gamma model with two treatment effects was the best fitting model. Model fit statistics for network 5 are given in Table 26. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

Table 26: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect	-12.253	7.947	-4.305

In Figure 48, we can see that all studies have normal (or as expected) contributions to the residual deviance. Thus, it appears the predictions from the NMA are a good fit to the observed data points.

Figure 48: box plot of the total deviance contribution for each study to the residual deviance – from OpenBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. Numbers above the lines represent the study the plot corresponds to (1 = BREAK-3; 2 = BRF113220; 3 = CheckMate 066; 4 = COLUMBUS; 5 = COMBI-d; 6 = COMBI-v; 7 = KEYNOTE-006; 8 = CheckMate 067). The horizontal line the mean deviance contribution. For 2-arm trials, we would expect the deviance contribution to be 2 for a good fitting model, since the model sums over 2 outcomes. For 3 arm trials we would expect the deviance contribution to be 4 (2 outcomes x 2 relative effects = 4).

It was not possible to assess inconsistency in this network as it only had a single loop which was composed of a three-arm trial, and it is not possible for a three-arm trial to be inconsistent with itself. Effectively, this means the network is a long chain of evidence.

Table 27: Fixed effect OS NMA results for generalized gamma with location and scale parameters

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
dabrafenib vs. DTIC	1.202	(0.8336; 1.69)
dabrafenib + trametinib vs. DTIC	1.629	(1.071; 2.39)
ipilimumab vs. DTIC	1.414	(0.8696; 2.172)
vemurafenib vs. DTIC	1.121	(0.7039; 1.707)
nivolumab vs. DTIC	2.533	(1.737; 3.593)
nivolumab + ipilimumab vs. DTIC	3.405	(1.943; 5.555)
encorafenib + binimetinib vs. DTIC	1.764	(1.016; 2.864)
pembrolizumab vs. DTIC	2.486	(1.354; 4.185)
Shape parameter		
dabrafenib vs. DTIC	1.095	(0.8444; 1.402)
dabrafenib + trametinib vs. DTIC	1.087	(0.7918; 1.462)
ipilimumab vs. DTIC	1.279	(1.01; 1.603)
vemurafenib vs. DTIC	1.081	(0.7593; 1.498)
nivolumab vs. DTIC	1.534	(1.283; 1.826)
nivolumab + ipilimumab vs. DTIC	2.051	(1.596; 2.602)
encorafenib + binimetinib vs. DTIC	1.352	(0.902; 1.959)
pembrolizumab vs. DTIC	1.525	(1.123; 2.026)

Table 27 presents the model predicted coefficients for each pairwise comparison and Figure 49 the predicted OS survival curves.

Nivolumab + ipilimumab had the highest point estimate for the scale parameter which can be interpreted as a time ratio 3.405 (95% CI 1.943-5.555). The time ratio can be interpreted as the likelihood of dying comes 3.405 times slower for nivolumab + ipilimumab compared to DTIC (the reference treatment). This in turn means people given nivolumab + ipilimumab live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 2.051 (1.596-2.602) which changes the shape of the survival curve.

Nivolumab monotherapy had the second-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.533 (95% CI 1.737-3.593). The time ratio can be interpreted as the likelihood of dying comes 2.533 times slower for nivolumab monotherapy compared to DTIC (the reference treatment). This in turn means people given nivolumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.534 (1.283-1.826).

Pembrolizumab monotherapy had the third-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.486 (95% CI 1.354-4.185). The time ratio can be interpreted as the likelihood of dying comes 2.486 times slower for pembrolizumab monotherapy compared to DTIC (the reference treatment). This in turn means people given pembrolizumab monotherapy live longer (improved overall survival

compared with DTIC). The treatment effect on the shape parameter was 1.525 (1.123-2.026).

Encorafenib + Binimetinib had the fourth-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 1.764 (95% CI 1.016-2.864). The time ratio can be interpreted as the likelihood of dying comes 1.764 times slower for encorafenib + binimetinib compared to DTIC (the reference treatment). This in turn means people given encorafenib + binimetinib live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.352 (0.902-1.959) respectively.

Dabrafenib + Trametinib had the fifth-best point estimate for the scale parameter which can be interpreted as a time ratio 1.629 (95% CI 1.071-2.39). The time ratio can be interpreted as the likelihood of dying comes 1.629 times slower for dabrafenib + trametinib compared to DTIC (the reference treatment). This in turn means people given dabrafenib + trametinib live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.087 (0.7918-1.462) respectively.

The other treatments in the network, dabrafenib monotherapy, ipilimumab monotherapy and vemurafenib monotherapy in general had worse point estimates for both the scale and shape parameters compared with dabrafenib + trametinib. This is reflected in the survival curves in Figure 49 where they are some of the lowest curves depicted. Additionally, the committee advised that dabrafenib and vemurafenib are not given as monotherapy, and ipilimumab is not frequently given due to its toxicity. Thus, in addition to performing poorly in the network, they are not clinically prevalent.

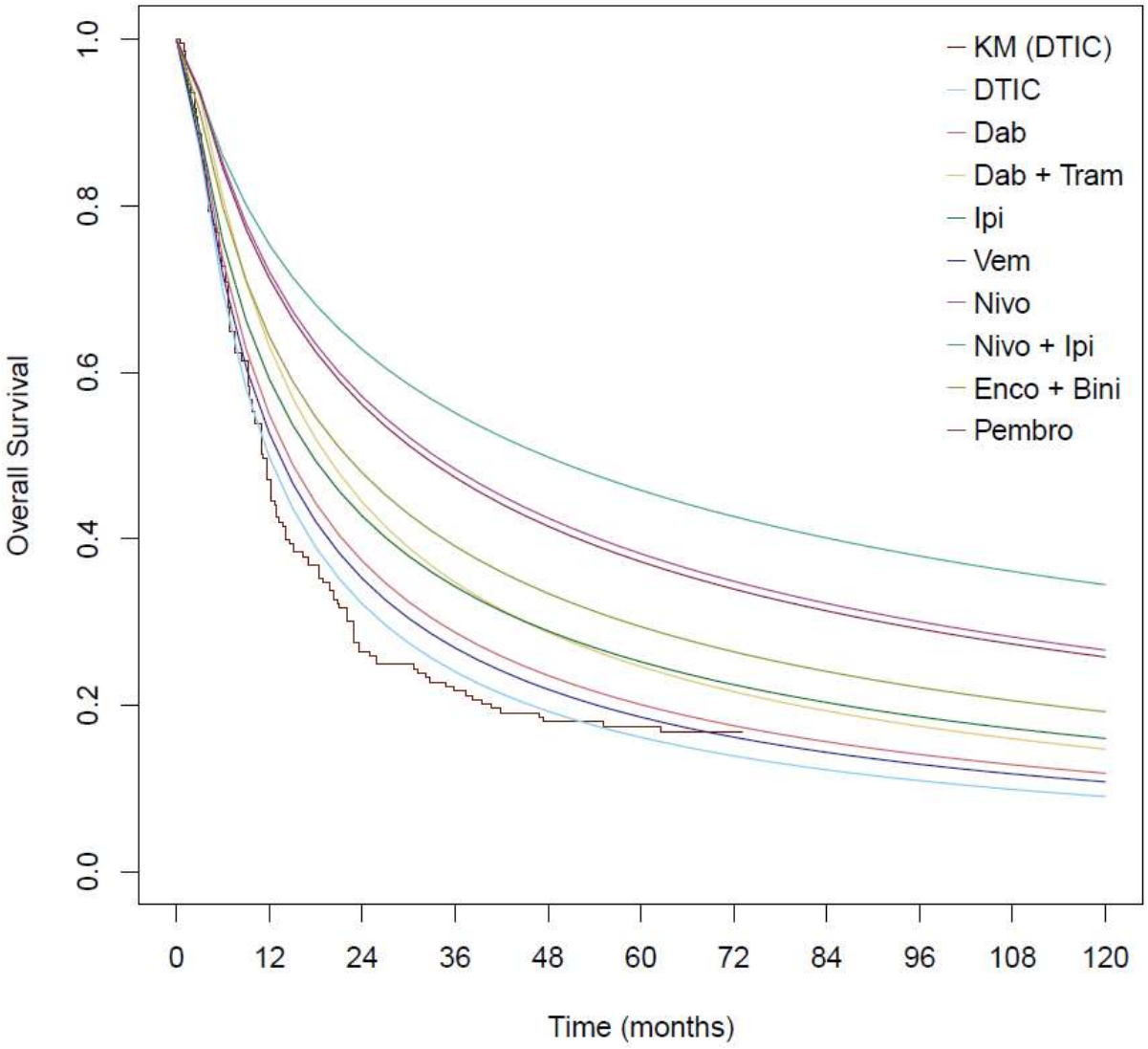


Figure 49: Survival curves for generalized gamma OS model on location and scale parameters

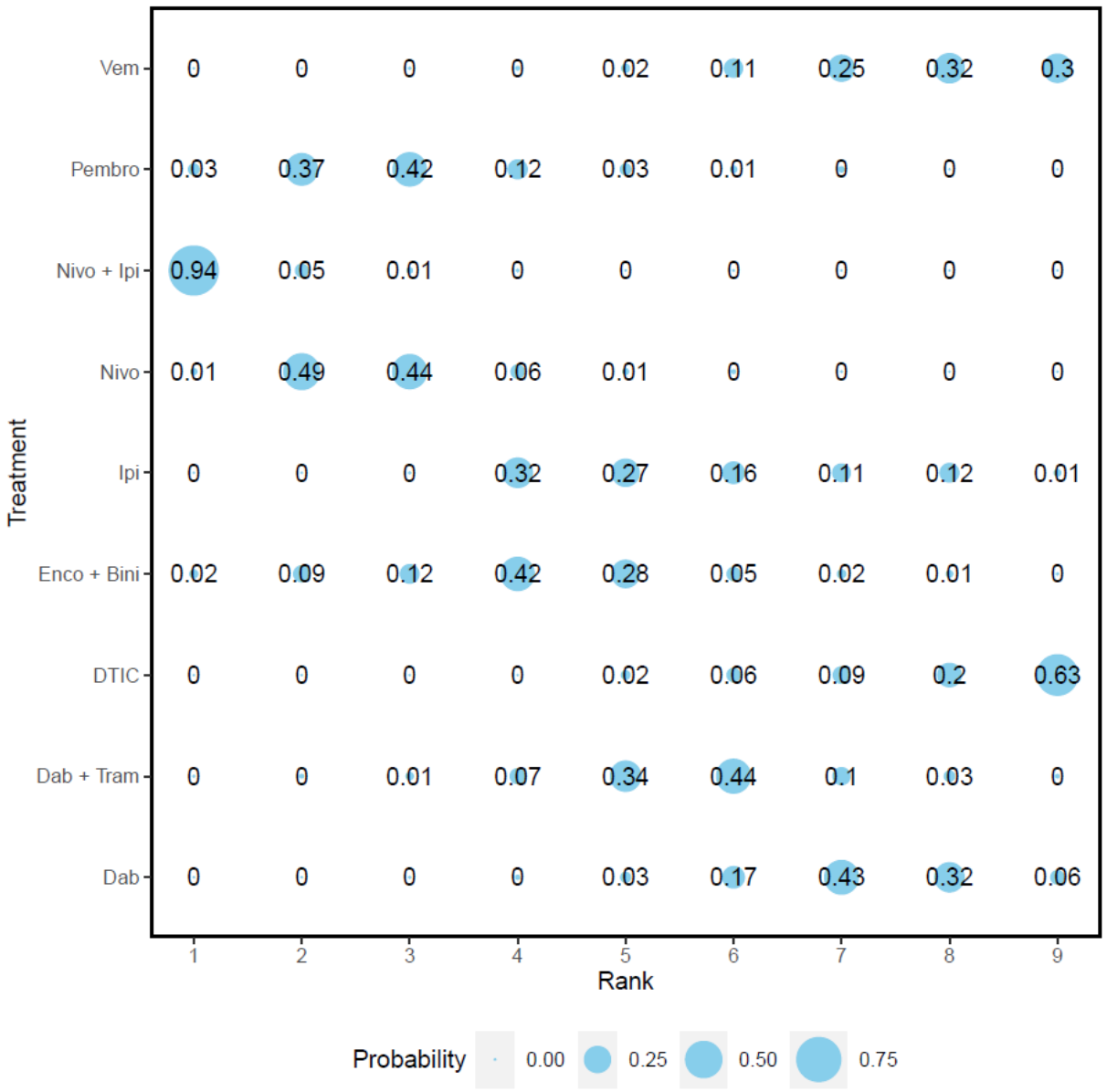


Figure 50: Ranking plot for survival at 60 months for generalized gamma OS model on location and scale parameters

These results are reflected in the ranking plots of overall survival at 60 months Figure 50:

- Nivolumab + ipilimumab had an average ranking of 1.07 with a 94% probability of being the best treatment
- Nivolumab monotherapy had an average ranking of 2.6 with a 1% probability of being the best treatment
- Pembrolizumab monotherapy had an average ranking of 2.72 with a 3% probability of being the best treatment
- Encorafenib + binimetinib had an average ranking of 4.16 with a 2% probability of being the best treatment
- Dabrafenib + trametinib had an average ranking of 5.59 with a 0% probability of being the best treatment

A.5.3.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

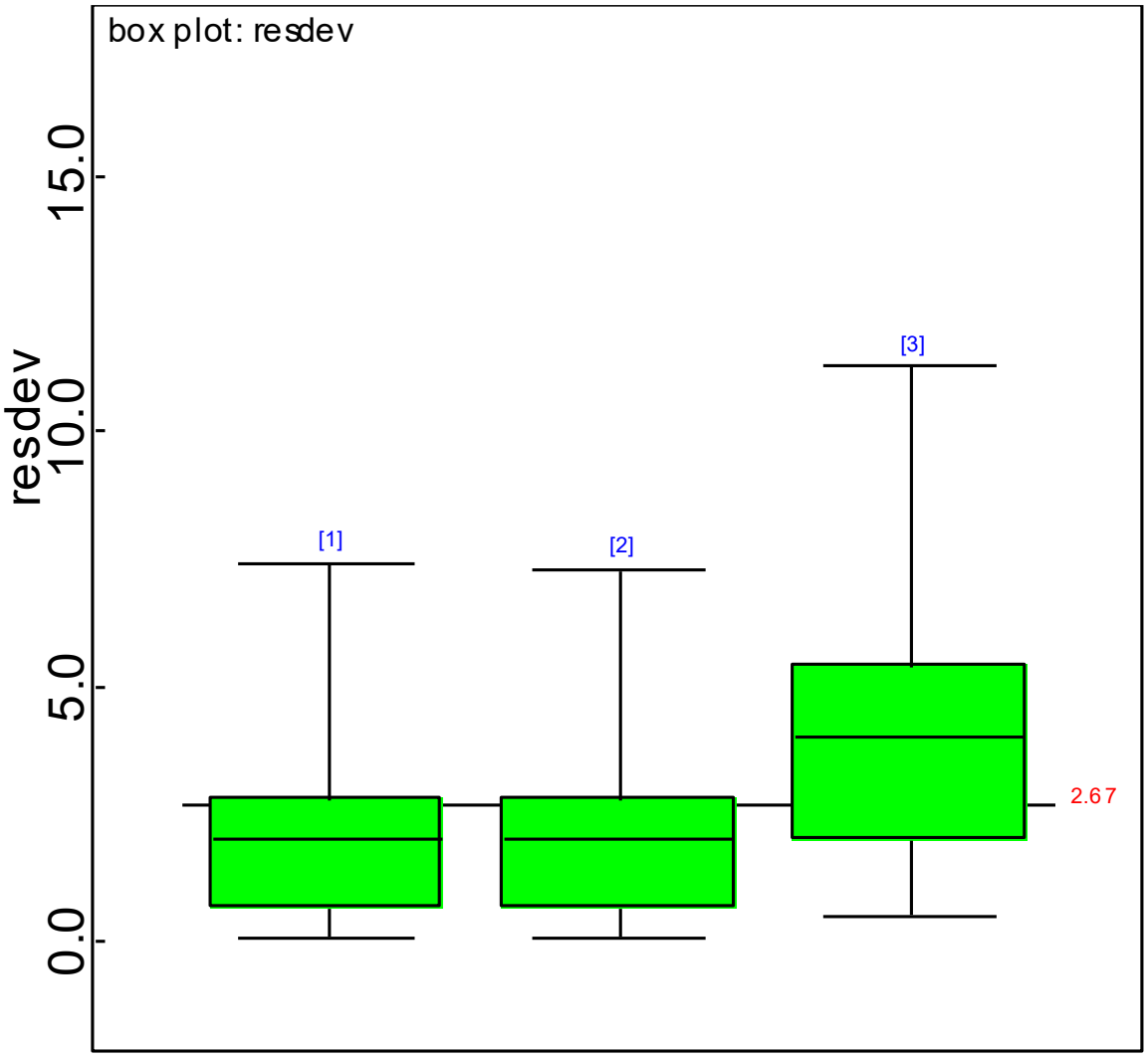
For Network 6, the generalized gamma model with two treatment effects was the best fitting model. Model fit statistics for network 6 are given in Table 28. Due to a limited number of studies for each comparison, we were limited to a standard fixed effect NMA and were unable to run a random effects NMA.

Table 28: Model fit statistics

Model	Dbar	pD	DIC
Fixed effect	-27.033	15.991	-11.042

In Figure 51, we can see that all studies have normal (or as expected) contributions to the residual deviance. Thus, it appears the predictions from the NMA are a good fit to the observed data points.

Figure 51: box plot of the total deviance contribution for each study to the residual deviance – from OpenBUGS



The boxes represent the interquartile range, the line across the box represents the median, and the whiskers represents the 95% credible intervals. Numbers above the lines represent the study the plot corresponds to (1 = CheckMate 066; 2 = KEYNOTE-006; 3 = CheckMate 067). The horizontal line the mean deviance contribution. For 2-arm trials, we would expect the deviance contribution to be 2 for a good fitting model, since the model sums over 2 outcomes. For 3 arm trials we would expect the deviance contribution to be 4 (2 outcomes x 2 relative effects = 4).

It was not possible to assess inconsistency in this network as it only had a single loop which was composed of a three-arm trial, and it is not possible for a three-arm trial to be inconsistent with itself.

Table 29: Fixed effect OS NMA results for generalized gamma with location and scale parameters

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
ipilimumab vs. DTIC	1.412	(0.8718; 2.167)
nivolumab vs. DTIC	2.531	(1.744; 3.567)
nivolumab + ipilimumab vs. DTIC	3.401	(1.948; 5.536)
pembrolizumab vs. DTIC	2.485	(1.357; 4.166)
Shape parameter		
ipilimumab vs. DTIC	1.277	(1.009; 1.596)
nivolumab vs. DTIC	1.533	(1.279; 1.822)
nivolumab + ipilimumab vs. DTIC	2.048	(1.592; 2.591)
pembrolizumab vs. DTIC	1.524	(1.124; 2.015)

Table 29 presents the model predicted coefficients for each pairwise comparison and Figure 52 presents the predicted OS survival curves.

Nivolumab + ipilimumab had the highest point estimate for the scale parameter which can be interpreted as a time ratio 3.401 (95% CI 1.948-5.536). The time ratio can be interpreted as the likelihood of dying comes 3.401 times slower for nivolumab + ipilimumab compared to DTIC (the reference treatment). This in turn means people given nivolumab + ipilimumab live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 2.048 (95% CI 1.592-2.591) which changes the shape of the survival curve.

Nivolumab monotherapy had the second-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.531 (95% CI 1.744-3.567). The time ratio can be interpreted as the likelihood of dying comes 2.531 times slower for nivolumab monotherapy compared to DTIC (the reference treatment). This in turn means people given nivolumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.533 (1.279-1.822).

Pembrolizumab monotherapy had the third-best point estimate point estimate for the scale parameter which can be interpreted as a time ratio 2.485 (95% CI 1.357-4.166). The time ratio can be interpreted as the likelihood of dying comes 2.485 times slower for pembrolizumab monotherapy compared to DTIC (the reference treatment). This in turn means people given pembrolizumab monotherapy live longer (improved overall survival compared with DTIC). The treatment effect on the shape parameter was 1.524 (1.124-2.015).

Ipilimumab monotherapy, the other treatment in the network had the worst point estimates for both the scale and shape parameters. This is reflected in Figure 52 where ipilimumab has a worse survival curve than nivolumab + ipilimumab, nivolumab monotherapy or pembrolizumab monotherapy. Additionally, as stated above ipilimumab is not frequently given due to its toxicity. Thus, in addition to performing poorly in the network, ipilimumab is not clinically prevalent.

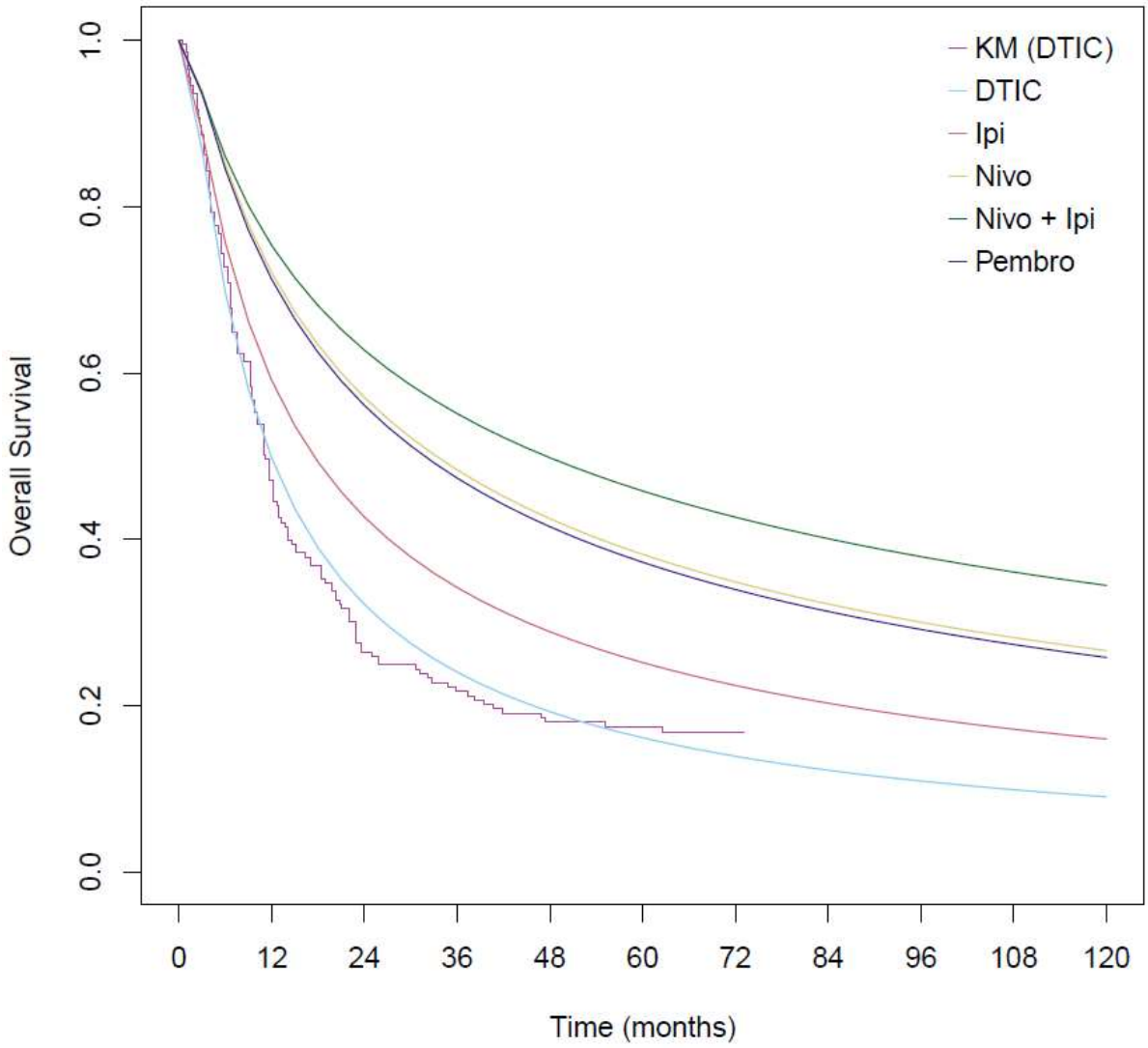


Figure 52: Survival curves for generalized gamma OS model on location and scale parameters

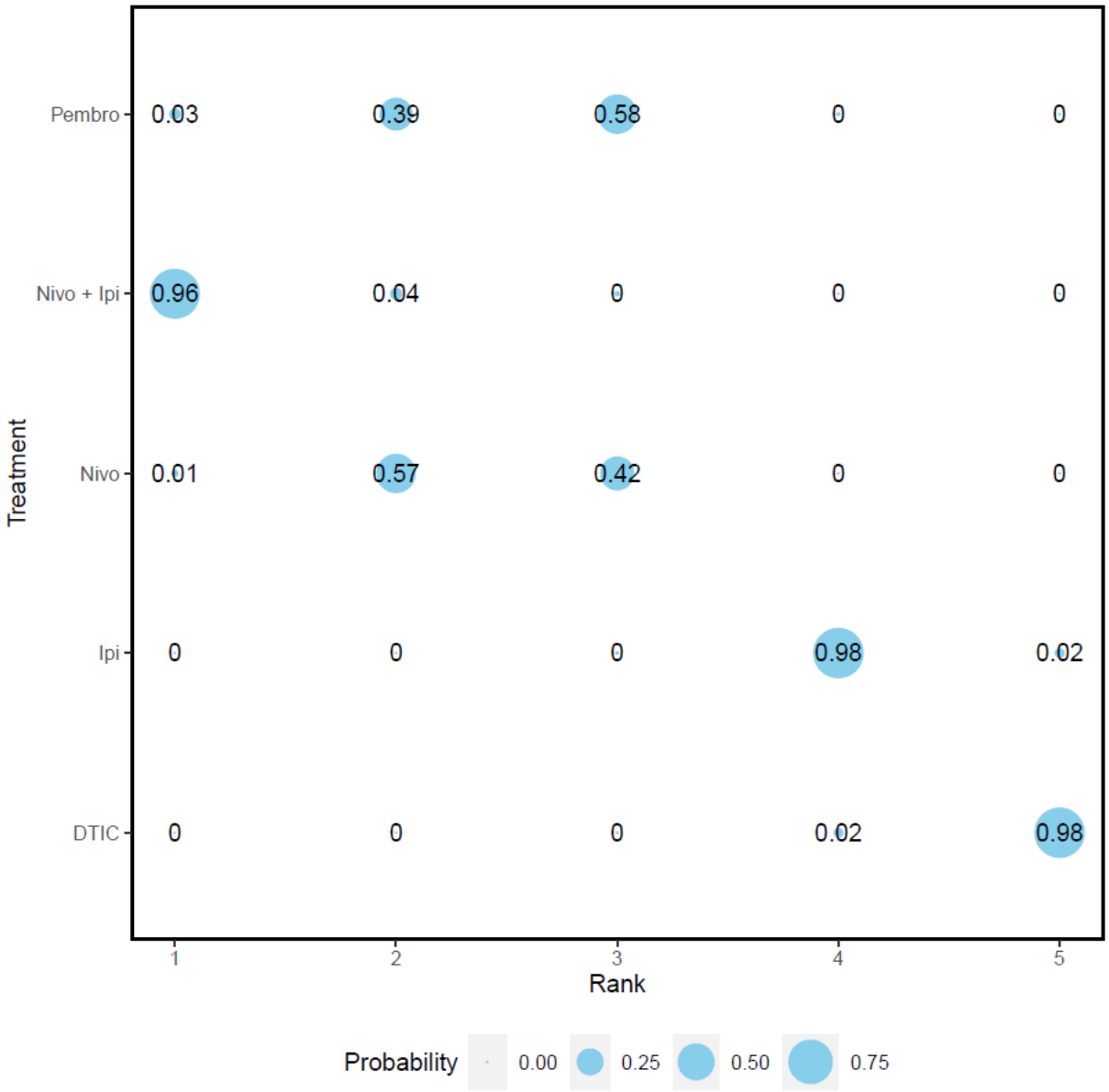


Figure 53: Ranking plot for survival at 60 months for generalized gamma OS model on scale and shape parameters

These results are reflected in the ranking plots of progression-free survival at 60 months Figure 53:

- Nivolumab + ipilimumab had an average ranking of 1.04 with a 96% probability of being the best treatment
- Nivolumab monotherapy had an average ranking of 2.41 with a 1% probability of being the best treatment
- Pembrolizumab monotherapy had an average ranking of 2.55 with a 3% probability of being the best treatment

A.6 Discussion

A.6.1 Progression-free survival

For progression-free survival, we fitted piecewise models where we estimated 3 hazard ratios relative to DTIC, one for each time interval of the model.

Network 1: People with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

The relative rankings of each treatment in each interval differed, which is in part a reflection of the mechanism of action for each of the treatments. In the first interval (0 to 12 months), encorafenib + binimetinib followed by dabrafenib + trametinib had the greatest effect on PFS relative to DTIC. These were then followed by nivolumab + ipilimumab, pembrolizumab monotherapy and nivolumab monotherapy. This is not unexpected, given that targeted therapies are considered to provide a quicker response than immunotherapies. In the second interval (12 to 18 months) and the third interval (greater than 18 months), there was no evidence that any of these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. The point estimate of the hazard ratio in the second interval was smallest for nivolumab monotherapy, then nivolumab +ipilimumab, then pembrolizumab monotherapy, then encorafenib + binimetinib, and finally dabrafenib + trametinib. The point estimate of the hazard ratio in the third interval was smallest for dabrafenib + trametinib, then encorafenib + binimetinib, then nivolumab + ipilimumab, then nivolumab monotherapy, and finally pembrolizumab monotherapy.

These results are reflected in the ranking plots of progression-free survival at 60 months. Nivolumab + ipilimumab had an average ranking of 1.63 with a 64% probability of being the best treatment. Dabrafenib + trametinib had an average ranking of 3.26, and encorafenib + binimetinib had an average ranking of 3.09. Nivolumab monotherapy and pembrolizumab monotherapy had average rankings of 3.42 and 5.07 respectively.

Based on these results, it can be stated with reasonable confidence that nivolumab + ipilimumab is the best treatment for achieving the greatest progression-free survival benefit. The NMA then suggests that the next best treatments with respect to this outcome are encorafenib + binimetinib, then dabrafenib + trametinib, then nivolumab monotherapy then pembrolizumab monotherapy.

When these survival projections were validated by clinical experts, it was noted that the long-term projections for pembrolizumab were implausibly low, and from clinical experience it would not be the treatment with the poorest outcome. A comparison of the KM plots from CM-006 and the different survival curves generated by each NMA survival model also suggests that the outcome from the NMA for pembrolizumab is too pessimistic. However, the results of the alternative NMA piecewise models do not offer any more valid long-term extrapolations, suggesting that the issue lies with the data that was inputted into the model rather than the choice of model itself. The outcome for pembrolizumab appears to be primarily driven by the hazard ratios for the second and third interval which push it far below the other treatments. In clinical practice, pembrolizumab and nivolumab are generally considered to be relatively equivalent given their similar modes of action.

In Figure 54, we present the PFS KM data for nivolumab and pembrolizumab together. We advise against any direct comparisons of the two treatments based on this diagram, as overlaying KM curves is not mathematically validated way to indirectly compare treatments that have not been compared to head-to-head in a clinical trial, only an NMA can do this. However, overlaying the KM curves here does allow us to examine the data that is being used in the NMA, and attempt to understand why the NMA is producing the results that it is.

Figure 54: Digitized PFS KM data for nivolumab (CheckMate 066 & CheckMate 067) and pembrolizumab (KEYNOTE-006)

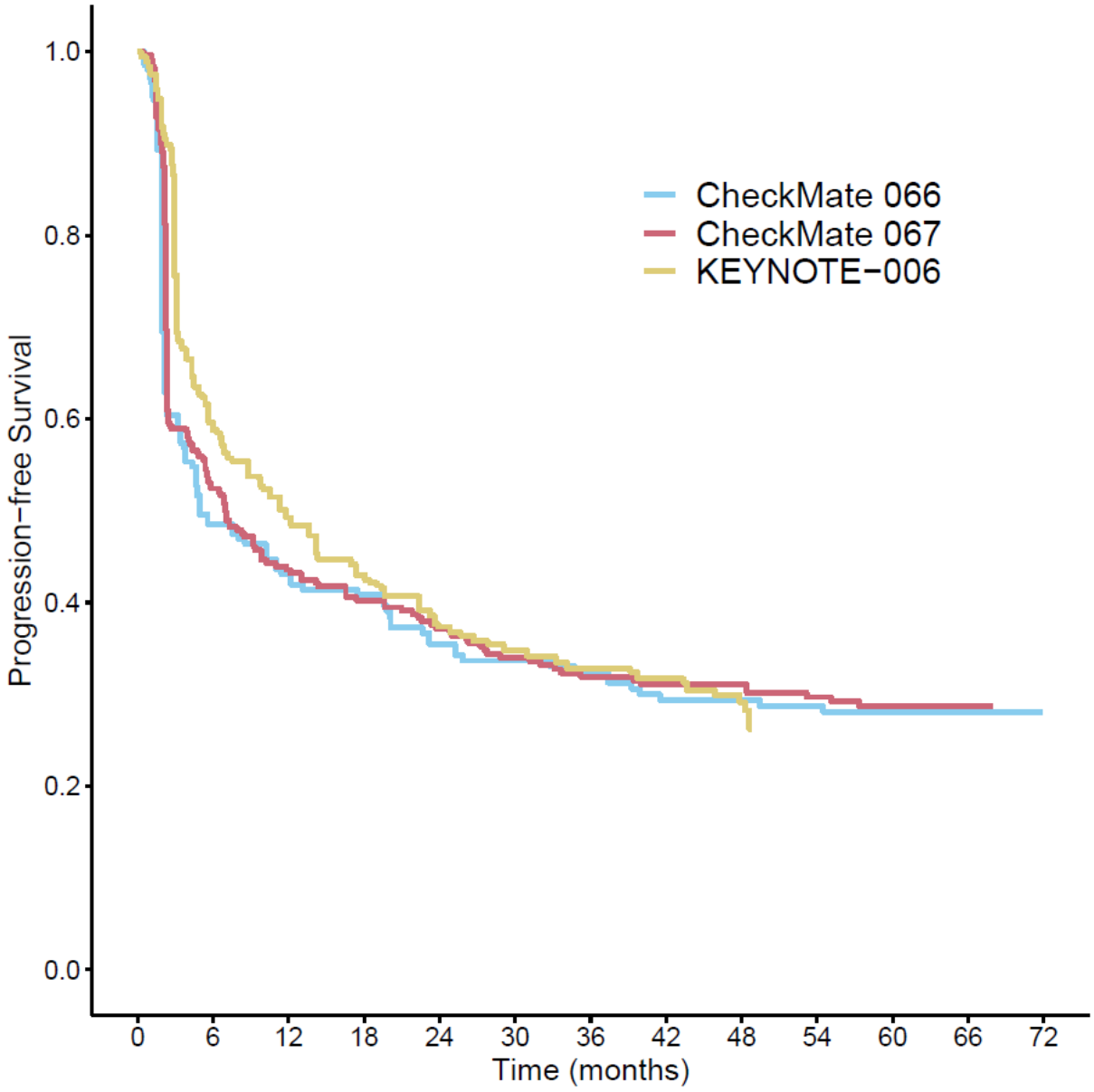


Figure 54 shows that there are similarities between the two treatments. But KEYNOTE-006 has a shorter follow-up period than both CheckMate 066 and CheckMate 067 (48 months versus approximately 66 months and 72 months respectively). Additionally, where nivolumab has plateaued by approximately 42 months in both CheckMate 066 and CheckMate 067, pembrolizumab still appears to be dropping in KEYNOTE-006, with a sharp drop at 48 months. It is likely a combination of both factors that contribute to the PFS NMA results where pembrolizumab appears worse than nivolumab. When thinking about the aggregate data the NMA uses, with a cut point at 18 months, pembrolizumab has data from 18-48 months. Over this time period, the data suggests that pembrolizumab is still seeing the PFS curve drop. As such, you get a more pessimistic hazard ratio for pembrolizumab in the final interval, 18-120 months, thereby making pembrolizumab appear worse. Whereas with nivolumab, with a cut point at 18 months, there is data up until 66 months and 72 months for CheckMate 066 and CheckMate 067 respectively. Over this time period, the data suggests that nivolumab is no longer dropping but has plateaued. Thus, you get a more optimistic hazard ratio for nivolumab for the final interval, 18-120 months. Of course, as already noted, both hazard ratios for nivolumab and pembrolizumab over the final interval show no evidence that these treatments reduce the hazard ratio compared with DTIC (as each credible interval

includes one, the line of no effect). But, as we also noted, there is a high degree of uncertainty in these estimates, which is likely a reflection of the smaller amount of data from which these estimates are being generated (i.e., most people have been censored by this point, thus there is a smaller population to get estimates from). It would have been preferrable in this situation if data had been available for KEYNOTE-006 extending up until 66 months, as it does for nivolumab in both CheckMate 066 and CheckMate 067. Then, we would be able to access in the NMA if pembrolizumab is worse than nivolumab for PFS. However, in the absence of such data, we are left with the results of the NMA and our best attempt at understanding why we get such results.

Network 2: People with a BRAF mutation for both targeted therapies and immunotherapies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

Network 5: People with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies using only trials that had long-term follow-up

The results of this analysis, that excluded two trials with only 20 months of follow up from the network, can be considered analogous to Network 1 (with no such restrictions on trial follow up time). The exclusion of two trials in the network resulted in some changes to the point estimates. In interval 1, the point estimate of the hazard ratio for dabrafenib + trametinib, nivolumab + ipilimumab and encorafenib + binimetinib increased, the point estimate of the hazard ratio for pembrolizumab monotherapy decreased, and the point estimate of the hazard ratio for nivolumab monotherapy didn't change. In interval 2: the point estimate of the hazard ratio for nivolumab + ipilimumab increased, the point estimate of the hazard ratios for dabrafenib + trametinib, nivolumab monotherapy, encorafenib + binimetinib and pembrolizumab monotherapy decreased. In interval 3, the point estimate of the hazard ratios for nivolumab monotherapy, nivolumab + ipilimumab, and pembrolizumab monotherapy increased, the point estimate of the hazard ratios for dabrafenib + trametinib and encorafenib + binimetinib decreased).

Similar to Network 1, nivolumab + ipilimumab had the best average ranking (1.77) and greatest probability of being the best treatment (58%). The results of this network show our base case network results are robust to different assumptions, in this case the exclusion of trials with shorter follow-up.

Network 3: People with BRAF wild type melanoma, with immunotherapy strategies only

The hazard ratio in the first interval was smallest for nivolumab + ipilimumab, then nivolumab monotherapy and finally pembrolizumab monotherapy. For the second and third intervals, there was no evidence that these treatments reduced the hazard ratio compared with DTIC (as each credible interval includes one, the line of no effect), but there is a high degree of uncertainty in the estimates. In the second interval, the point estimate of the hazard ratio in the second interval was smallest for nivolumab monotherapy, then nivolumab + ipilimumab and finally pembrolizumab monotherapy, and in the third interval, the point estimate of the hazard ratio in the second interval was smallest for nivolumab + ipilimumab, then nivolumab monotherapy and finally pembrolizumab monotherapy.

Nivolumab + ipilimumab had the best average ranking (1.28) and greatest probability of being the best treatment (73%). Based on these results, it can be stated with reasonable confidence that nivolumab + ipilimumab is the best treatment for achieving the greatest progression-free survival benefit, followed by nivolumab monotherapy than pembrolizumab monotherapy.

Network 4: People with BRAF mutant and wild type melanoma, with immunotherapy strategies only

In this network, nivolumab + ipilimumab had a more favourable average ranking and higher probability of being the best treatment than in Network 3, that only included trials where people with *BRAF* wild type melanoma were included. Both pembrolizumab and nivolumab had similarly low probabilities of being the most effective treatment, although pembrolizumab had a slightly lower average ranking than nivolumab.

There were some changes to the point estimates for this network relative to network 3 (Interval 1: the hazard ratio for nivolumab + ipilimumab and pembrolizumab monotherapy decreased, and the hazard ratio for nivolumab monotherapy didn't change; Interval 2: the hazard ratio for nivolumab increased, the hazard ratios for nivolumab + ipilimumab and pembrolizumab monotherapy decreased; Interval 3: the hazard ratios for nivolumab monotherapy and nivolumab + ipilimumab increased, the hazard ratio for pembrolizumab monotherapy decreased).

Network 6: People with BRAF mutant and wildtype melanoma, receiving immunotherapies only, limited to trials with long follow-up

The results of this analysis were similar to that for Network 5, with nivolumab + ipilimumab having the most favourable average ranking and highest probability of being the best treatment. There were some changes to the point estimates for this network relative to network 3 (Interval 1: the hazard ratio for nivolumab + ipilimumab and pembrolizumab monotherapy decreased, and the hazard ratio for nivolumab monotherapy didn't change; Interval 2: the hazard ratios for nivolumab + ipilimumab, nivolumab monotherapy and pembrolizumab monotherapy decreased; Interval 3: the hazard ratios for nivolumab monotherapy and pembrolizumab monotherapy increased, the hazard ratio for nivolumab + ipilimumab monotherapy decreased).

Based on these results, it can be stated with reasonable confidence that nivolumab + ipilimumab is the best treatment for achieving the greatest progression-free survival benefit, followed by nivolumab monotherapy than pembrolizumab monotherapy. The results of this network show our base case network results are robust to different assumptions, in this case the mixed population of people with and without BRAF mutations compared with a population of people that are only without a BRAF mutation. The results of Network 4 and 6 are notable in a few ways, with an even greater surety that nivolumab + ipilimumab is the best treatment. And while pembrolizumab performs better than nivolumab in the first interval which was not seen in Network 3, this only has a slight effect in increasing the probability that pembrolizumab monotherapy is the second-best treatment from 5% in Network 3 to 12% in Network 4.

A.6.2 Overall survival

Network 1: People with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

As previously noted, there was evidence of inconsistency, with model fit and DIC being lower for the inconsistency model (Table 19). However, as we also have already noted, of the 8 comparisons where we are able to obtain estimates from the inconsistency model, only 2 (dabrafenib versus DTIC, and vemurafenib versus DTIC) disagree with consistency model. Furthermore, the confidence intervals for the consistency and inconsistency models overlap for these comparisons. Thus, while the point estimates may differ, due to the considerable overlap in the confidence intervals we are less concerned with any observed inconsistency.

Furthermore, as noted by NICE's Technical Support Unit, 'inconsistency in one part of the network does not necessarily imply that the entire body of evidence is to be considered suspect' (19). In this instance, the two estimates where issues of inconsistency arise, are not given in clinical practice. Thus, while it may be possible the NMA is producing 'deviant' estimates for dabrafenib and vemurafenib, it is of little clinical significance since neither of these treatments are given as monotherapies anymore. Although we are unable to isolate

what is causing the inconsistency, we do propose a number of things at the trial level which may be a factor including but not limited to the shapes of the curves are different, there are differences in the trial populations, or the trials could have had differences in treatment switching (either allowing versus not allowing for it, or different proportions of people switching, or different benefits achieved through switching). Furthermore, despite the inconsistency observed in these two estimates, a recent publication provides support for our overall finding that initial treatment with nivolumab + ipilimumab is more effective than either targeted therapy regimen (30).

Based on these results, it can be stated with reasonable confidence that nivolumab + ipilimumab is the best treatment for achieving the greatest overall survival benefit, followed by nivolumab monotherapy, pembrolizumab monotherapy, encorafenib + binimetinib and finally dabrafenib + trametinib. However, both nivolumab and pembrolizumab monotherapy have similar effects to each other.

Network 2: People with a BRAF mutation for both targeted therapies and immunotherapies

As already noted, Network 2 was not a fully connected network. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

Network 5: People with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies using only trials that had long-term follow-up

Network 5 resulted in some changes to the point estimates (scale parameter: nivolumab monotherapy remained the same, nivolumab + ipilimumab increased, and pembrolizumab monotherapy, encorafenib + binimetinib, and dabrafenib + trametinib all decreased; shape parameter: nivolumab monotherapy remained the same, nivolumab + ipilimumab decreased, and pembrolizumab monotherapy, encorafenib + binimetinib, and dabrafenib + trametinib all increased), however the overall rankings remained the same, nivolumab + ipilimumab has the best point estimate for both the scale and shape parameter time ratios, nivolumab monotherapy has the second-best estimate, pembrolizumab monotherapy the third-best, encorafenib + binimetinib the fourth-best, and finally dabrafenib + trametinib the fifth-best.

Based on these results, it can be stated with reasonable confidence that nivolumab + ipilimumab is the best treatment for achieving the greatest overall survival benefit, followed by nivolumab monotherapy, pembrolizumab monotherapy, encorafenib + binimetinib and finally dabrafenib + trametinib. However, both nivolumab and pembrolizumab monotherapy have similar effects to each other. The results of this network show our base case network results are robust to different assumptions, in this case the exclusion of trials with shorter follow-up.

Network 3: People with BRAF wild type melanoma, with immunotherapy strategies only

Based on these results, it can be stated with reasonable confidence that nivolumab + ipilimumab is the best treatment for achieving the greatest overall survival benefit in people with BRAF wildtype melanoma, followed by nivolumab monotherapy and finally pembrolizumab monotherapy. However, both nivolumab and pembrolizumab monotherapy have similar effects to each other.

Network 4: People with BRAF mutant and wild type melanoma, with immunotherapy strategies only

Network 4 resulted in some changes to the point estimates (scale parameter: nivolumab monotherapy decreased, nivolumab + ipilimumab and pembrolizumab monotherapy both increased; shape parameter: nivolumab monotherapy remained the same, nivolumab + ipilimumab and pembrolizumab monotherapy both increased), however the overall rankings remained the same, nivolumab + ipilimumab has the best point estimate for both the scale

and shape parameter time ratios, nivolumab monotherapy has the second-best estimate and pembrolizumab monotherapy the third-best.

Network 6: People with BRAF mutant and wildtype melanoma, receiving immunotherapies only, limited to trials with long follow-up

Network 6 resulted in some changes to the point estimates (scale parameter: nivolumab monotherapy decreased, nivolumab + ipilimumab and pembrolizumab monotherapy both increased; shape parameter: nivolumab monotherapy remained the same, nivolumab + ipilimumab and pembrolizumab monotherapy both increased). However, the overall rankings remained the same, nivolumab + ipilimumab has the best point estimate for both the scale and shape parameter time ratios, nivolumab monotherapy has the second-best estimate and pembrolizumab monotherapy the third-best.

Based on these results, it can be stated with reasonable confidence that nivolumab + ipilimumab is the best treatment for achieving the greatest overall survival benefit, followed by nivolumab monotherapy and finally pembrolizumab monotherapy. However, both nivolumab and pembrolizumab monotherapy have similar effects to each other. The results of this network show our base case network results are robust to different assumptions, in this case using a mixed population. An important point to consider is the improvement of nivolumab + ipilimumab when using a mixed population. This appears to be driven in large part by the KM curves from CheckMate067 (Larkin 2019). For the mixed BRAF population, there is clear separation between nivolumab + ipilimumab and nivolumab, with nivolumab having a higher proportion of patients who are alive (i.e., nivolumab + ipilimumab is the optimal treatment for overall survival – Fig 1A). However, for the KM curve for people without BRAF mutations, there is less separation apparent between the two curves, with the two curves even coming together at certain points. By the end of the diagram, nivolumab + ipilimumab does have a higher proportion of patients alive, the difference between the two treatments is less noticeable in this population of exclusively patients without BRAF mutations. This data seemingly explains why nivolumab + ipilimumab performs better in a mixed BRAF population than it does in an exclusively BRAF wildtype population.

A.6.3 Strengths

This analysis explored 5 different methodologies for the synthesis of time to event outcomes in a network meta-analysis: the Cox PH, RMST, generalized gamma parametric models, piecewise models, and fractional polynomial models. In a recent review of NICE guidance, it was found that the most frequently reported outcome when an NMA of TTE data is performed, is the hazard ratio (31). However, Freeman et al 2020 note that if the proportional hazards assumption is not met, a hazard ratio is not an appropriate outcome measure.

We found that the proportional hazards assumption was violated for each of our networks, which led us to consider models that do not rely on the proportional hazards assumption, ultimately fitting three such models. We consider it a strength of our analysis that we did not fit PH model when the PH assumption clearly wasn't met. As previously noted, there is a gap in the literature with regards to what NMA methodology for TTE data one should use if PH is not met. Thus, we consider it a strength of our analysis that we considered and fitted a breadth of models using different methodologies. Although, it is clear from this exercise that while several methods exist that don't rely on proportional hazards, there is a gap in the literature as to determining which of these methods is best. Further research is needed in this area. Fitting these models allowed us to pick the best fitting model for both PFS and OS. Given PFS and OS are frequently model drivers in partitioned survival models, obtaining the best fitting model was a significant goal for us.

Additionally, providing the coding for a range of different models with the same network will hopefully be of use to others in the future who are fitting NMAs for TTE outcomes.

A.6.4 Limitations

There are several limitations to our NMA that can broadly be classed in three ways. Broader methodological limitations, data limitations, and model specific limitations.

We note several methodological limitations. It has been proposed that when performing NMAs of TTE data, PFS and OS should be jointly modelled, as was done in the NICE Lung Cancer N2 model (9,10,32). This is because PFS and OS are correlated, should be treated as such with joint modelling. Although this has been done with hazard ratios (which isn't appropriate here due to lack of PH) and RMST, to our knowledge, joint modelling of PFS and OS has not been attempted before with parametric models, piecewise models or FP models (33). Further work is warranted, not only to assess the impact of joint modelling of PFS and OS in comparison with modelling them individually, but also to develop code that is widely available so that people can model these jointly.

A further limitation is that because we summarised the data in different formats to input to the different models, the likelihoods not comparable and so meant that we could not use statistical measure such as DIC or AIC to compare between different survival models (although these could be used to compare models of the same type). Considering this, visual inspection became the primary tool by which we were able to assess the best fitting model. Further research into this area is needed, ideally in a way that allows all models to have the same likelihood, which would allow someone to use DIC to determine the best fitting model.

We also note several limitations with regards to the data used in our model. First, as already noted, there was no KM curve for people without a BRAF mutation for Keynote 006. This means that for people that are without a BRAF mutation, the PFS network still has one trial using a mixed population.

Additionally, a few trials do not have data published with longer follow-up. This can be seen with BRIM-3 and CheckMate069, which both have only 20 months of follow-up for PFS, and Checkmate069 also only has only 24 months of follow-up for OS. This in turn forces one to either use methodology where reduced follow-up does not have a significant impact (i.e., means a RMST model cannot be used), or may force a consideration as to whether it is possible to remove the trial from the network.

Finally, the data available as already noted has not been adjusted for treatment switching. Although this impacts all treatments, it is unclear the magnitude by which it affects all treatments. It would be useful in the decision-making context to have data on treatment switching. Additionally, finding KM curves for 1L treatment and by BRAF mutation status at times proved challenging. Reporting on results from clinical trials could alleviate this difficulty in the future by making such KM curves and analyses easily available (assuming of course such analyses were done).

We also note several model specific limitations. Although we ran two generalized gamma models, we did not run a generalized gamma model with three treatment effects. A model with three treatment effects would have the benefit of increased flexibility, however this would come at increased model complexity. Additionally, although we ran a number of piecewise models that were informed by clinical judgement as to where cut points should be placed, this was an iterative process. Further research is needed on a set of principles one can use to determine where best to place cut points for piecewise exponential models for cancer treatments. While we understand there is likely variation between treatments based on disease severity and other modifiers, if such research was done, it would provide a useful starting point for those seeking to run piecewise exponential NMA TTE models. For the fractional polynomial models, we aggregated the data into 8 intervals. Increasing the number of aggregated data intervals may have resulted in improved estimates and improved convergence for some of the fractional polynomial models. However, the fractional polynomials took a significant amount of time to run and those fractional polynomial models that we had fitted hadn't resulted in a better fit but had resulted in dramatically increased

model complexity. We therefore do not think that refinements to the fractional polynomials is likely to change our model selection. However, we do believe that further research is needed into the number of intervals required for fractional polynomial models, as increasing the number of intervals will lead to a longer run-time and greater model complexity but will also likely achieve a better fit. Examining the trade-offs associated with differing numbers of intervals would be of use to anyone fitting a fractional polynomial model. Finally, we did not run the Royston-Parmar model. While we believe this is justifiable as it is also a flexible parametric model like the fractional polynomials, and as such is likely to dramatically increase complexity without a guarantee of being a better fit, without having run the model we cannot say this for certain.

A.7 Conclusions

Nivolumab + ipilimumab was the best treatment consistently within each network that we explored, for improving both progression-free survival and overall survival in people with advanced melanoma. Notably, this result held for people with *BRAF* mutant as well as *BRAF* wild type melanoma. The other two immunotherapies in the analysis, pembrolizumab and nivolumab, showed promise as a means of improving survival outcomes, being ranked just below nivolumab + ipilimumab for most networks.

Targeted treatment dual strategies, dabrafenib + trametinib and encorafenib + binimetinib, were less effective than nivolumab + ipilimumab and the immunotherapies for progression-free survival and overall survival. However, they had more favourable outcomes than chemotherapy and single agent targeted therapy.

The majority of the trials providing data informing the network were considered to be of low risk of bias, and none were considered to be at high risk of bias. All trials were considered directly relevant to the decision problem, with only very minor concerns regarding the study design (see Evidence Review F for a full discussion). However, evaluating the changing event hazard over time of treatment strategies with different modes of action and corresponding response patterns led to challenges in selecting a single model that was a good fit to every single treatment in the network. This was further compounded by small numbers of patients and events in the latter period of some of the trials, which had implications for the extrapolation of survival. These limitations should be borne in mind when interpreting results, and the plausibility of long-term survival projections may be enhanced through adjustments to the curves based on clinical experience and external sources of evidence, such as considering general population mortality or assuming equivalence between two treatment strategies after a certain time point (see HE report for details of how the NMA models were implemented in the economic model).

A.8 Code

Portions of this code have been adapted after they were graciously provided by Suzanne Freeman (34).

A.8.1 Digitizing R code (Guyot algorithm)

```
# Digitizing code


#------------------------------------------------------------------------
# Instructions
```

```
#------------------------------------------------------------------------

# The following code has two parts. The first allows you to digitize KM curves
# by clicking on an points on an image you upload. The second part uses the
# validated Guyot code to create IPD data from the digitized curves. If you
# have already digitized curves in an external program (such as Enguage digitizer)
# you can upload the relevant CSV files and proceed directly to part two to
# create IPD data.

#------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------

library(digitize)
library("data.table")
library("ggplot2")
library("qpcR")
library("dplyr")
library(survival)
library(survminer)

# Start with clean environment
rm(list=ls())

# Set working directory
setwd()

#------------------------------------------------------------------------
# Load and format data
#------------------------------------------------------------------------

# The image of the KM curves is preferably a png but other types should be usable
```

```
Inputfile <- "KMcurves/BREAK-3_OS_KM.png"


# Number at risk (NaR) should be a .csv file and should be in format:
# col1=time, col2=Line1NaR, col3=Line2NaR
NaRData <- read.csv("BREAK-3_OS_NaR.csv")
TotalEventsLine1 <- NA # if reported then how many events
TotalEventsLine2 <- NA # if reported then how many events


# TO UPDATE BASED ON WHICH ARMS ARE INCLUDED
Line1TreatmentName <- "Dabrafenib"
Line2TreatmentName <- "Dacarbazine"
StudyName <- "BREAK-3_OS_"


#-----------------------------------------------------------------------
# Step 1: Digitizing curves (clicking on them)
#-----------------------------------------------------------------------


# Parts 1 and 2 (manually converting graph into x,y co-ordinates) must be run
# line by line the rest of the code can be run at once


#-----------------------------------------------------------------------
# Part 1: Digitise first line
#-----------------------------------------------------------------------


###!~!~!~!~!~! RUN THIS LINE BY ITSELF (Do NOT run entire program at once)
Line1 <- digitize(Inputfile) # follow instructions (4 clicks + 4 numbers for axes, then click
data points)


#save the data for line1
write.table(Line1, file = paste(StudyName,Line1TreatmentName,"_data",".csv",sep=""),
sep = "\t", row.names = F)


#-----------------------------------------------------------------------
```

```
# Part 2: Digitise second line

#----------------------------------------------------------------------------


###!~!~!~!~!~! RUN THIS LINE BY ITSELF (Do NOT run entire program at once)

Line2 <- digitize(Inputfile) # follow instructions (4 clicks + 4 numbers for axes, then click data points)


#save the data for line2

write.table(Line2, file = paste(StudyName,Line2TreatmentName,"_data",".csv",sep=""), sep = "\t", row.names = F)


#----------------------------------------------------------------------------

# Step 2: Generate IPD data using Guyot method

#----------------------------------------------------------------------------


#### OPTIONAL: If you have already digitised and have the KM time and event data ####

#### in two column "time" then "survival" .csv format you can skip all previous ####

#### steps and use the lines below to read the data in ####

# Line1<-read.csv("BREAK-3_OS_dab_data.csv")

# Line2<-read.csv("BREAK-3_OS_daca_data.csv")


#----------------------------------------------------------------------------

# Part 1: Format data

#----------------------------------------------------------------------------


# Format data for algorithm

names(NaRData) <- c("t", "Line1.NaR","Line2.NaR")

names(Line1) <- c("Line1.t", "Line1.S")

names(Line2) <- c("Line2.t", "Line2.S")

TotalEvents <- c(TotalEventsLine1,TotalEventsLine2)


# Use this if you have cumulative incidence and need to change to survival curves

#Line1$Line1.S <- 1-Line1$Line1.S
```

```
#Line2$Line2.S <- 1-Line2$Line2.S


Line1 <- Line1 %>%
  na.omit() %>%
  slice(1) %>%
  mutate(Line1.t = 0, Line1.S = 1) %>%
  bind_rows(Line1 %>%
              slice(2:n())) %>%
  mutate(Line1.t = ifelse(row_number()==1 | Line1.t>=lag(Line1.t), Line1.t, lag(Line1.t)),
         Line1.S = ifelse(row_number()==1 | Line1.S<=lag(Line1.S), Line1.S, lag(Line1.S)))
Line2 <- Line2 %>%
  na.omit() %>%
  slice(1) %>%
  mutate(Line2.t = 0, Line2.S = 1) %>%
  bind_rows(Line2 %>%
              slice(2:n())) %>%
  mutate(Line2.t = ifelse(row_number()==1 | Line2.t>=lag(Line2.t), Line2.t, lag(Line2.t)),
         Line2.S = ifelse(row_number()==1 | Line2.S<=lag(Line2.S), Line2.S, lag(Line2.S)))


# Need to add NAs to the bottom of the shorter vector to prevent recycling of the shorter
table in R
if(nrow(Line1)>nrow(Line2))
{Padding<-data.frame(Line2.t=rep(NA,nrow(Line1)-
nrow(Line2)),Line2.S=rep(NA,nrow(Line1)-nrow(Line2)))
Line2<-rbind(Line2,Padding)}
if(nrow(Line2)>nrow(Line1))
{Padding<-data.frame(Line1.t=rep(NA,nrow(Line2)-
nrow(Line1)),Line1.S=rep(NA,nrow(Line2)-nrow(Line1)))
Line1<-rbind(Line1,Padding)}


InputLines <- cbind(Line1,Line2)


#------------------------------------------------------------------------
# Part 2: Guyot Code to generate IPD data
```

```
#------------------------------------------------------------------------

# Data shouldn't have any points with t<0 and the probabilities should all be 0<=p<=1
# Anything outside this range will break the algorithm

for (arm.id in 0:1)
{
  tot.events<-TotalEvents[arm.id+1]



  #Read in survival times
  datSurv <- InputLines
  tblSurv <- datSurv %>% dplyr::select(t.S = ifelse(arm.id==0, "Line1.t", "Line2.t"),
                                       S = ifelse(arm.id==0, "Line1.S", "Line2.S")) %>%
    filter(!is.na(t.S))
  maxDat <- max(tblSurv %>% dplyr::select(t.S))


  datNaR <-NaRData
  t.risk<-unlist(datNaR %>% mutate(t=ifelse(t<maxDat, t, maxDat)) %>% dplyr::select("t"))
  n.risk<-unlist(datNaR %>% mutate(t=ifelse(t<maxDat, t, maxDat)) %>%
dplyr::select(ifelse(arm.id==0, "Line1.NaR", "Line2.NaR")))
  n.int<-length(n.risk)


  t.S <- rep(0, n.int-1)
  S <- rep(0, n.int-1)


  for (x in 2:n.int){
    t.S[x-1] <- as.numeric(t.risk[x])
    S[x-1] <- tblSurv %>%
      filter(t.S <= as.numeric(t.risk[x])) %>%
      slice(which.max(t.S)) %>%
      pull(S)
  }
```

```
tbl <- tibble(t.S, S)
tbl2 <- tblSurv %>%
  bind_rows(tbl) %>%
  arrange(t.S) %>%
  distinct()
t.S <- tbl2 %>% pull(t.S)
S <- tbl2 %>% pull(S)
lower <- rep(0, n.int)
upper <- rep(0, n.int)
for (x in 1:n.int){
  lower[x] <- tbl2%>%
    filter(t.S <= as.numeric(t.risk[x])) %>%
    summarise(n = n()) %>%
    pull(n)
  # upper[x]<-max(lower[x], which(t.S < as.numeric(t.risk[x+1])))
  upper[x] <- tbl2%>%
    filter(t.S <= as.numeric(t.risk[x+1])) %>%
    summarise(n = n()) %>%
    pull(n)
}
upper[n.int] <- tbl2%>%
  summarise(n = n()) %>%
  pull(n)
n.t<- upper[n.int]

#Initialise vectors
arm<-rep(arm.id,n.risk[1])
n.censor<- rep(0,(n.int-1))
n.hat<-rep(n.risk[1]+1,n.t)
cen<-rep(0,n.t)
d<-rep(0,n.t)
KM.hat<-rep(1,n.t)
```

```
  last.i<-rep(1,n.int)
  sumdL<-0
  if (n.int > 1){
    #Time intervals 1,...,(n.int-1)
    for (i in 1:(n.int-1)){
      #First approximation of no. censored on interval i
      n.censor[i]<- round(n.risk[i]*S[lower[i+1]]/S[lower[i]]- n.risk[i+1])
      #Adjust tot. no. censored until n.hat = n.risk at start of interval (i+1)
      while((n.hat[lower[i+1]]>n.risk[i+1])||((n.hat[lower[i+1]]<n.risk[i+1])&&(n.censor[i]>0))){
        if (n.censor[i]<=0){
          cen[lower[i]:upper[i]]<-0
          n.censor[i]<-0
        }
        if (n.censor[i]>0){
          cen.t<-rep(0,n.censor[i])
          for (j in 1:n.censor[i]){
            cen.t[j]<- t.S[lower[i]] +
              j*(t.S[lower[(i+1)]]-t.S[lower[i]])/(n.censor[i]+1)
          }
          #Distribute censored observations evenly over time. Find no. censored on each time
interval.
          cen[lower[i]:upper[i]]<-hist(cen.t,breaks=t.S[lower[i]:lower[(i+1)]],
                                       plot=F)$counts
        }
        #Find no. events and no. at risk on each interval to agree with K-M estimates read
from curves
        n.hat[lower[i]]<-n.risk[i]
        last<-last.i[i]
        for (k in lower[i]:upper[i]){
          if (i==1 & k==lower[i]){
            d[k]<-0
            KM.hat[k]<-1
          }
```

```
        else {
          d[k]<-round(n.hat[k]*(1-(S[k]/KM.hat[last])))
          KM.hat[k]<-KM.hat[last]*(1-(d[k]/n.hat[k]))
        }
        n.hat[k+1]<-n.hat[k]-d[k]-cen[k]
        if (d[k] != 0) last<-k
      }
      n.censor[i]<- n.censor[i]+(n.hat[lower[i+1]]-n.risk[i+1])
    }
    if (n.hat[lower[i+1]]<n.risk[i+1]) n.risk[i+1]<-n.hat[lower[i+1]]
    last.i[(i+1)]<-last
  }
 }
 #Time interval n.int.
 if (n.int>1){
  #Assume same censor rate as average over previous time intervals.
  n.censor[n.int]<- min(round(sum(n.censor[1:(n.int-1)])*(t.S[upper[n.int]]-
                                                t.S[lower[n.int]])/(t.S[upper[(n.int-1)]]-t.S[lower[1]])),
n.risk[n.int])
 }
 if (n.int==1){n.censor[n.int]<-0}
 if (n.censor[n.int] <= 0){
  cen[lower[n.int]:(upper[n.int]-1)]<-0
  n.censor[n.int]<-0
 }
 if (n.censor[n.int]>0){
  cen.t<-rep(0,n.censor[n.int])
  for (j in 1:n.censor[n.int]){
   cen.t[j]<- t.S[lower[n.int]] +
    j*(t.S[upper[n.int]]-t.S[lower[n.int]])/(n.censor[n.int]+1)
  }
  cen[lower[n.int]:(upper[n.int]-1)]<-hist(cen.t,breaks=t.S[lower[n.int]:upper[n.int]],
                                plot=F)$counts
```

```
  }
  #Find no. events and no. at risk on each interval to agree with K-M estimates read from
curves
  n.hat[lower[n.int]]<-n.risk[n.int]
  last<-last.i[n.int]
  for (k in lower[n.int]:upper[n.int]){
    if(KM.hat[last] !=0){
      d[k]<-round(n.hat[k]*(1-(S[k]/KM.hat[last])))} else {d[k]<-0}
    KM.hat[k]<-KM.hat[last]*(1-(d[k]/n.hat[k]))
    n.hat[k+1]<-n.hat[k]-d[k]-cen[k]
    #No. at risk cannot be negative
    if (n.hat[k+1] < 0) {
      n.hat[k+1]<-0
      cen[k]<-n.hat[k] - d[k]
    }
    if (d[k] != 0) last<-k
  }
  #If total no. of events reported, adjust no. censored so that total no. of events agrees.
  if (!is.na(tot.events)){
    if (n.int>1){
      sumdL<-sum(d[1:upper[(n.int-1)]])
      #If total no. events already too big, then set events and censoring = 0 on all further
time intervals
      if (sumdL >= tot.events){
        d[lower[n.int]:upper[n.int]]<- rep(0,(upper[n.int]-lower[n.int]+1))
        cen[lower[n.int]:(upper[n.int]-1)]<- rep(0,(upper[n.int]-lower[n.int]))
        n.hat[(lower[n.int]+1):(upper[n.int]+1)]<- rep(n.risk[n.int],(upper[n.int]+1-lower[n.int]))
      }
    }
    #Otherwise adjust no. censored to give correct total no. events
    if ((sumdL < tot.events)|| (n.int==1)){
      sumd<-sum(d[1:upper[n.int]])
      while ((sumd > tot.events)||((sumd< tot.events)&&(n.censor[n.int]>0))){
```

```
    n.censor[n.int]<- n.censor[n.int] + (sumd - tot.events)
    if (n.censor[n.int]<=0){
      cen[lower[n.int]:(upper[n.int]-1)]<-0
      n.censor[n.int]<-0
    }
    if (n.censor[n.int]>0){
      cen.t<-rep(0,n.censor[n.int])
      for (j in 1:n.censor[n.int]){
        cen.t[j]<- t.S[lower[n.int]] +
          j*(t.S[upper[n.int]]-t.S[lower[n.int]])/(n.censor[n.int]+1)
      }
      cen[lower[n.int]:(upper[n.int]-1)]<-hist(cen.t,breaks=t.S[lower[n.int]:upper[n.int]],
                                               plot=F)$counts
    }
    n.hat[lower[n.int]]<-n.risk[n.int]
    last<-last.i[n.int]
    for (k in lower[n.int]:upper[n.int]){
      d[k]<-round(n.hat[k]*(1-(S[k]/KM.hat[last])))
      KM.hat[k]<-KM.hat[last]*(1-(d[k]/n.hat[k]))
      if (k != upper[n.int]){
        n.hat[k+1]<-n.hat[k]-d[k]-cen[k]
        #No. at risk cannot be negative
        if (n.hat[k+1] < 0) {
          n.hat[k+1]<-0
          cen[k]<-n.hat[k] - d[k]
        }
      }
      if (d[k] != 0) last<-k
    }
    sumd<- sum(d[1:upper[n.int]])
  }
}
```

```
  }
 #
write.table(matrix(c(t.S,n.hat[1:n.t],d,cen),ncol=4,byrow=F),paste(strPath,strFileOutputKM,
sep=""),sep="/t")


assign(paste0("SummaryData",arm.id+1),data.table(matrix(c(t.S,n.hat[1:n.t],d,cen),ncol=4,
byrow=F)))
  outputtable<-data.frame(matrix(c(t.S,n.hat[1:n.t],d,cen),ncol=4,byrow=F))
  ### Now form IPD ###
  #Initialise vectors
  t.IPD<-rep(t.S[n.t],n.risk[1])
  event.IPD<-rep(0,n.risk[1])
  #Write event time and event indicator (=1) for each event, as separate row in t.IPD and
event.IPD
  k=1
  for (j in 1:n.t){
    if(d[j]!=0){
      t.IPD[k:(k+d[j]-1)]<- rep(t.S[j],d[j])
      event.IPD[k:(k+d[j]-1)]<- rep(1,d[j])
      k<-k+d[j]
    }
  }
  #Write censor time and event indicator (=0) for each censor, as separate row in t.IPD and
event.IPD
  for (j in 1:(n.t-1)){
    if(cen[j]!=0){
      t.IPD[k:(k+cen[j]-1)]<- rep(((t.S[j]+t.S[j+1])/2),cen[j])
      event.IPD[k:(k+cen[j]-1)]<- rep(0,cen[j])
      k<-k+cen[j]
    }
  }
  #Output IPD


assign(paste0("PatientData",arm.id+1),data.frame(matrix(c(t.IPD,event.IPD,arm),ncol=3,by
row=F)))
  #write.table(IPD,paste(strPath,strFileOutputIPD,sep=""),sep="/t",col.names=TRUE)
```

```
}


## Store results in SummaryData1 & SummaryData2, and PatientData1 & PatientData2

PatientData1$Treatment <- Line1TreatmentName

PatientData2$Treatment <- Line2TreatmentName

SummaryData1$Treatment <- Line1TreatmentName

SummaryData2$Treatment <- Line2TreatmentName


# Combine PatientData1 and PatientData2

Patientdata <- rbind(PatientData1,PatientData2)


# Combine SummaryData1 and SummaryData2

Summarydata <- rbind(SummaryData1, SummaryData2)


# Change working directory for exporting results

setwd()


# Save patient level data

write.csv(Patientdata,
paste0("SummaryLevel",StudyName,gsub("[^[:alnum:]]","",Sys.time()),".csv")) # outputs file
with system date and time


# Save summary level data

write.csv(Summarydata,
paste0("SummaryLevel",StudyName,gsub("[^[:alnum:]]","",Sys.time()),".csv")) # outputs file
with system date and time


#---------------------------------------------------------------------------

# Part 3: Check if code is returning what we wanted or not by fititng KM by

# treatment

#---------------------------------------------------------------------------


# Prepare data to fit Kaplan Meier

km <- Surv(time = CombinedData[['X1']], event = CombinedData[['X2']])
```

```
#Fit KM, stratifying by treatment
km_treatment<-survfit(km~Treatment,data=CombinedData,type='kaplan-
meier',conf.type='log')

# Make KM plot
ggsurvplot(km_treatment)

# Make KM plot with confidence intervals
ggsurvplot(km_treatment,conf.int = 'True')

# Fit cox PH
cox <- coxph(km~Treatment, data=CombinedData)

# Summary of cox PH
summary(cox)
```

A.8.2 KM curve plot R code

```
# Plot KM curves for each trial
#------------------------------------------------------------------------
# Instructions
#------------------------------------------------------------------------

# The following code creates KM plots for all trials included in the network.

#------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------

# Load libraries
library(survival)
```

```
library(survminer)
library(dplyr)
library(RColorBrewer)

# Start with empty environment
rm(list=ls())

# Set the working directory
setwd()

#-------------------------------------------------------------------------
# Import data and make plots
#-------------------------------------------------------------------------

# Import data
data <- read.csv("../Data/melanoma_os_ipd_nc.csv")
df1 <- data[data$studyCode==1,]
df2 <- data[data$studyCode==2,]
df3 <- data[data$studyCode==3,]
df4 <- data[data$studyCode==4,]
df5 <- data[data$studyCode==5,]
df6 <- data[data$studyCode==6,]
df7 <- data[data$studyCode==7,]
df8 <- data[data$studyCode==8,]
df9 <- data[data$studyCode==9,]
df10 <- data[data$studyCode==10,]

# Get median survival by printing fit
fit1 <- survfit(Surv(time, event) ~ arm, data = df1)
fit2 <- survfit(Surv(time, event) ~ arm, data = df2)
fit3 <- survfit(Surv(time, event) ~ arm, data = df3)
fit4 <- survfit(Surv(time, event) ~ arm, data = df4)
```

```
fit5 <- survfit(Surv(time, event) ~ arm, data = df5)
fit6 <- survfit(Surv(time, event) ~ arm, data = df6)
fit7 <- survfit(Surv(time, event) ~ arm, data = df7)
fit8 <- survfit(Surv(time, event) ~ arm, data = df8)
fit9 <- survfit(Surv(time, event) ~ arm, data = df9)
fit10 <- survfit(Surv(time, event) ~ arm, data = df10)

# Create lists for fit and df
fit <- list(fit1, fit2, fit3, fit4, fit5, fit6, fit7, fit8, fit9, fit10)
df_list <- list(df1, df2, df3, df4, df5, df6, df7, df8, df9, df10)

# Create plots for each trial using lists
ggsurvplot_list(fit=fit, data=df_list, censor=F)

# Using colours from the Safe palette from rcartocolor
colors=c("#88CCEE", "#CC6677", "#DDCC77", "#117733", "#332288", "#AA4499",
        "#44AA99", "#999933", "#882255", "#661100", "#6699CC", "#888888")

# Make plot 1
p1 <- ggsurvplot(fit=fit1, data = df1, censor=F, palette=c(colors[1], colors[2]),
          xlab="Months From Randomization", ylab="OS Probability",
          xlim=c(0, 65), ylim=c(0, 1), title="BREAK-3",
          legend=c(0.5,0.9), legend.title="", legend.labs=c("Dacarbazine", "Dabrafenib"),
          font.legend=18)
p1$plot + scale_x_continuous(breaks=sort(c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55,
60, 65))) + scale_y_continuous(breaks=sort(c(0, 0.25, 0.50, 0.75, 1.00)))

# Save plot
dev.copy(pdf, "Break3_os_nc.pdf")
dev.off()

# Make plot 2
p2 <- ggsurvplot(fit=fit2, data = df2, censor=F, palette=c(colors[2], colors[3]),
```

```
        xlab="Time Since Random Assignment (months)", ylab="Overall Survival (%)",
        xlim=c(0, 72), ylim=c(0, 1), title="BRF113220",
        legend=c(0.5,0.9), legend.title="", legend.labs=c("D150", "D + T 150/2"),
        font.legend=18)
p2$plot + scale_x_continuous(breaks=sort(c(0, 6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66,
72))) + scale_y_continuous(breaks=sort(c(0, 0.2, 0.4, 0.6, 0.8, 1)))


# Save plot
dev.copy(pdf, "BRF113220_os_nc.pdf")
dev.off()


# Make plot 3
p3 <- ggsurvplot(fit=fit3, data = df3, censor=F, palette=c(colors[1], colors[5]),
        xlab="Time, months", ylab="OS, %",
        xlim=c(0, 60), ylim=c(0, 1), title="BRIM-3",
        legend=c(0.5,0.9), legend.title="", legend.labs=c("DTIC", "Vem"),
        font.legend=18)
p3$plot + scale_x_continuous(breaks=sort(c(0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36,
39, 42, 45, 48, 51, 54, 57, 60))) + scale_y_continuous(breaks=sort(c(0, 0.1, 0.2, 0.3, 0.4,
0.5, 0.6, 0.7, 0.8, 0.9, 1)))


# Save plot
dev.copy(pdf, "BRIM-3_os_nc.pdf")
dev.off()


# Make plot 4
p4 <- ggsurvplot(fit=fit4, data = df4, censor=F, palette=c(colors[1], colors[6]),
        xlab="Months", ylab="OS (%)",
        xlim=c(0, 72), ylim=c(0, 1), title="CheckMate 066",
        legend=c(0.5,0.9), legend.title="", legend.labs=c("DTIC", "NIVO"),
        font.legend=18)
p4$plot + scale_x_continuous(breaks=sort(c(0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36,
39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69, 72, 75))) + scale_y_continuous(breaks=sort(c(0,
0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)))
```

```
# Save plot

dev.copy(pdf, "CheckMate_066_os_nc.pdf")

dev.off()


# Make plot 5

p5 <- ggsurvplot(fit=fit5, data = df5, censor=F, palette=c(colors[4], colors[6], colors[7]),

                 xlab="Months", ylab="Patients Who Survived (%)",

                 xlim=c(0, 66), ylim=c(0, 1), title="CheckMate 067",

                 legend=c(0.5,0.9), legend.title="", legend.labs=c("Ipilimumab", "Nivolumab",
"Nivolumab plus ipilimumab"),

                 font.legend=18)

p5$plot + scale_x_continuous(breaks=sort(c(0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, 36,
39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 69))) + scale_y_continuous(breaks=sort(c(0, 0.1, 0.2,
0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)))


# Save plot

dev.copy(pdf, "CheckMate_067_os_nc.pdf")

dev.off()


# Make plot 6

p6 <- ggsurvplot(fit=fit6, data = df6, censor=F, palette=c(colors[4], colors[7]),

                 xlab="Months", ylab="Probability of Overall Survival",

                 xlim=c(0, 30), ylim=c(0, 1), title="CheckMate 069",

                 legend=c(0.5,0.9), legend.title="", legend.labs=c("IPI", "NIVO + IPI"),

                 font.legend=18)

p6$plot + scale_x_continuous(breaks=sort(c(0, 3, 6, 9, 12, 15, 18, 21, 24, 27, 30))) +
scale_y_continuous(breaks=sort(c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)))


# Save plot

dev.copy(pdf, "CheckMate_069_os_nc.pdf")

dev.off()


# Make plot 7

p7 <- ggsurvplot(fit=fit7, data = df7, censor=F, palette=c(colors[5], colors[8]),
```

```
        xlab="Time (months)", ylab="Probability, %",
        xlim=c(0, 60), ylim=c(0, 1), title="COLUMBUS",
        legend=c(0.5,0.9), legend.title="", legend.labs=c("VEM", "COMBO450"),
        font.legend=18)
p7$plot + scale_x_continuous(breaks=sort(c(0, 4, 8, 12, 16, 20, 24, 28, 32, 36, 40, 44, 48,
52, 56, 60))) + scale_y_continuous(breaks=sort(c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8,
0.9, 1)))


# Save plot
dev.copy(pdf, "COLUMBUS_os_nc.pdf")
dev.off()


# Make plot 8
p8 <- ggsurvplot(fit=fit8, data = df8, censor=F, palette=c(colors[2], colors[3]),
        xlab="Months since Randomization", ylab="Proportion Alive",
        xlim=c(0, 78), ylim=c(0, 1), title="COMBI-d",
        legend=c(0.5,0.9), legend.title="", legend.labs=c("Dabrafenib plus placebo",
"Dabrafenib plus trametinib"),
        font.legend=18)
p8$plot + scale_x_continuous(breaks=sort(c(0, 6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66,
72, 78))) + scale_y_continuous(breaks=sort(c(0, 0.2, 0.4, 0.6, 0.8, 1)))


# Save plot
dev.copy(pdf, "COMBI-d_os_nc.pdf")
dev.off()


# Make plot 9
p9 <- ggsurvplot(fit=fit9, data = df9, censor=F, palette=c(colors[3], colors[5]),
        xlab="Months since Randomization", ylab="Proportion Alive",
        xlim=c(0, 78), ylim=c(0, 1), title="COMBI-v",
        legend=c(0.5,0.9), legend.title="", legend.labs=c("Dabrafenib plus trametinib",
"Vemurafenib"),
        font.legend=18)
p9$plot + scale_x_continuous(breaks=sort(c(0, 6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66,
72, 78))) + scale_y_continuous(breaks=sort(c(0, 0.2, 0.4, 0.6, 0.8, 1)))
```

```
# Save plot
dev.copy(pdf, "COMBI-v_os_nc.pdf")
dev.off()


# Make plot 10
p10 <- ggsurvplot(fit=fit10, data = df10, censor=F, palette=c(colors[4], colors[10]),
                  xlab="Time since randomization (months)", ylab="Overall survival (%)",
                  xlim=c(0, 65), ylim=c(0, 1), title="KEYNOTE-006",
                  legend=c(0.5,0.9), legend.title="", legend.labs=c("Ipilimumab", "Combined pembrolizumab groups"),
                  font.legend=18)
p10$plot + scale_x_continuous(breaks=sort(c(0, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65))) + scale_y_continuous(breaks=sort(c(0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1)))


# Save plot
dev.copy(pdf, "Keynote_006_os_nc.pdf")
dev.off()
```

A.8.3 Network diagram R code

```
# Create a network plot
#-------------------------------------------------------------------------
# Instructions
#-------------------------------------------------------------------------

# The following code tests generates a network plot. To do so you need a csv
# file with the following columns, Study name, t1 (treatment 1 - numeric code),
# t2 (treatment 2 - numeric code), lhr (log hazard ratio), se (standard error).
# While not required to make the netmeta object, additional columns in the csv file
# for the studyCode, treatments 1 and 2 (full names of treatments), and labels
# 1 and 2 (abbreviated treatment names) may be helpful to the user.
```

```
#------------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------------

# Load library
library(netmeta)

# Start with empty environment
rm(list=ls())

# Set the working directory
setwd()

#------------------------------------------------------------------------------
# Load data
#------------------------------------------------------------------------------

# Load the data
data <- read.csv("netgraph_data_os_nc.csv")

#------------------------------------------------------------------------------
# Make network plot
#------------------------------------------------------------------------------

# Create a netmeta object
a <- netmeta(data$lhr, data$se, treat1=data$t1, treat2=data$t2, studlab=data$Study,
reference=1)
netgraph(a)

# Treatment labels
lab <- c("DTIC", "Dab", "Dab + Tram", "Ipi", "Vem", "Nivo", "Nivo + Ipi", "Enco + Bini",
        "Pembro")
```

```
# Plot network
netgraph(a, labels=lab, offset=0.02, plastic=F, col="#434343", multiarm=T,
col.multiarm="purple", points=T,
        col.points="blue", number.of.studies = T, cex=1.5,
        cex.points=c(3,3,3,3,3,2,2,1,1))

# Save network plot as a PDF file
dev.copy(pdf, "netgraph_os_nc.pdf")
dev.off()
```

A.8.4 Proportional hazards tests R code

```
# Tests for proportional hazards (PH) assumptions

#------------------------------------------------------------------------
# Instructions
#------------------------------------------------------------------------

# The following code tests the proportional hazards assumption for IPD data.
# There are three tests performed, the Grambsch and Thernau statistical test,
# the Schoenfeld residual plot, and the log-log curve plot.

#------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------

library(survival)
library(broom)
library(metafor)
library(survminer)
```

```
# Start with an empty environment
rm(list=ls())

# Set the working directory
setwd()

#-------------------------------------------------------------------------
# Import and format data
#-------------------------------------------------------------------------

# Import the data and split into data frames for each trial
data <- read.csv("../Data/melanoma_os_ipd_nc.csv")
df1 <- data[data$studyCode==1,]
df1$arm <- as.factor(df1$arm)
df2 <- data[data$studyCode==2,]
df2$arm <- as.factor(df2$arm)
df3 <- data[data$studyCode==3,]
df3$arm <- as.factor(df3$arm)
df4 <- data[data$studyCode==4,]
df4$arm <- as.factor(df4$arm)
df5 <- data[data$studyCode==5,]
df5$arm <- as.factor(df5$arm)
df6 <- data[data$studyCode==6,]
df6$arm <- as.factor(df6$arm)
df7 <- data[data$studyCode==7,]
df7$arm <- as.factor(df7$arm)
df8 <- data[data$studyCode==8,]
df8$arm <- as.factor(df8$arm)
df9 <- data[data$studyCode==9,]
df9$arm <- as.factor(df9$arm)
df10 <- data[data$studyCode==10,]
df10$arm <- as.factor(df10$arm)
```

```
# Fit the Cox PH model
cox1 <- coxph(formula = Surv(time, event) ~ arm, data=df1)
coxph(formula = Surv(time, event) ~ arm, data=df1)
cox2 <- coxph(formula = Surv(time, event) ~ arm, data=df2)
coxph(formula = Surv(time, event) ~ arm, data=df2)
cox3 <- coxph(formula = Surv(time, event) ~ arm, data=df3)
coxph(formula = Surv(time, event) ~ arm, data=df3)
cox4 <- coxph(formula = Surv(time, event) ~ arm, data=df4)
coxph(formula = Surv(time, event) ~ arm, data=df4)
cox5 <- coxph(formula = Surv(time, event) ~ arm, data=df5)
coxph(formula = Surv(time, event) ~ arm, data=df5)
cox6 <- coxph(formula = Surv(time, event) ~ arm, data=df6)
coxph(formula = Surv(time, event) ~ arm, data=df6)
cox7 <- coxph(formula = Surv(time, event) ~ arm, data=df7)
coxph(formula = Surv(time, event) ~ arm, data=df7)
cox8 <- coxph(formula = Surv(time, event) ~ arm, data=df8)
coxph(formula = Surv(time, event) ~ arm, data=df8)
cox9 <- coxph(formula = Surv(time, event) ~ arm, data=df9)
coxph(formula = Surv(time, event) ~ arm, data=df9)
cox10 <- coxph(formula = Surv(time, event) ~ arm, data=df10)
coxph(formula = Surv(time, event) ~ arm, data=df10)

#-------------------------------------------------------------------------
# Test 1: Grambsch and Thernau statistical test
#-------------------------------------------------------------------------

# Test the PH assumption
test1 <- cox.zph(cox1)
test2 <- cox.zph(cox2)
test3 <- cox.zph(cox3)
test4 <- cox.zph(cox4)
```

```
test5 <- cox.zph(cox5)
test6 <- cox.zph(cox6)
test7 <- cox.zph(cox7)
test8 <- cox.zph(cox8)
test9 <- cox.zph(cox9)
test10 <- cox.zph(cox10)

#-------------------------------------------------------------------------
# Test 2: Schoenfeld residual plots
#-------------------------------------------------------------------------

# Plot Schoenfeld residuals
ggcoxzph(test1)
ggcoxzph(test2)
ggcoxzph(test3)
ggcoxzph(test4)
ggcoxzph(test5)
ggcoxzph(test6)
ggcoxzph(test7)
ggcoxzph(test8)
ggcoxzph(test9)
ggcoxzph(test10)

#-------------------------------------------------------------------------
# Test 3: log-log plots
#-------------------------------------------------------------------------

# Save data needed for log-log plots
s1 <- Surv(df1$time, df1$event)
s2 <- Surv(df2$time, df2$event)
s3 <- Surv(df3$time, df3$event)
s4 <- Surv(df4$time, df4$event)
```

```
s5 <- Surv(df5$time, df5$event)
s6 <- Surv(df6$time, df6$event)
s7 <- Surv(df7$time, df7$event)
s8 <- Surv(df8$time, df8$event)
s9 <- Surv(df9$time, df9$event)
s10 <- Surv(df10$time, df10$event)

# Plot log-log
plot(survfit(s1 ~ df1$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s2 ~ df2$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s3 ~ df3$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s4 ~ df4$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s5 ~ df5$txCode), col=c("blue", "red", "green"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s6 ~ df6$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s7 ~ df7$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s8 ~ df8$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s9 ~ df9$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
plot(survfit(s10 ~ df10$txCode), col=c("blue", "red"), fun="cloglog",
    xlab="Log(time)", ylab="log(-log(survival probability))")
```

A.8.5 Generalized gamma

A.8.5.1 One treatment effect

A.8.5.1.1 R code

```
# Generalized Gamma NMA for Time to Event Outcomes - 1 treatment effect

#------------------------------------------------------------------------
# Instructions
#------------------------------------------------------------------------
# The following code runs a generalized gamma NMA for time to event outcome IPD data
# using WinBUGS. The code has 2 parts. The first part runs the NMA.
# The second part uses the results from part 1 to make plots and estimate
# the area under the curve for different treatments. Each part and each step has
# more detail of what the code does at that point.

#------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------

# Load required packages
library(survival)
library(flexsurv)
library(R2WinBUGS)
library(pracma)
library(reshape)
library(reshape2)
library(ggplot2)
library(ggpubr)

# Start with clean environment
rm(list=ls())
```

```
# Set working directory
setwd()

#------------------------------------------------------------------------
# Part 1: Run NMA
#------------------------------------------------------------------------

# Part 1 of the analysis runs the NMA and produces results

# Step 1 fits AFT models with one treatment effect in R to each study in the network

# Step 2 is only required if there are 3 arm trials in the network. This step
# reformats the data from step 1. This is necessary as if you have a 3 arm network,
# Step 1 will return two lines of data for it. One comparing Treatment A vs B, and
# the second comparing A v C. This step will combine these rows so you have one row
# with all three treatments in it, rather than two rows each only having two
# treatments in them.

# Step 3 formats the data as the 'bugs_data' frame to be read into WinBUGS
# later using R2WinBUGS

# Step 4 sets the initial values for the NMA.

# Step 5 runs the NMA in WinBUGS using R2WinBUGS. Note, to run the NMA you must
have
# the WinBUGS model saved as a text file. In this case it is 'FE_model.txt'
# The code will then produce and saves results, and allows you to check the
# density and trace plots.

#------------------------------------------------------------------------
# Step 1: Run AFT analysis to obtain data to feed into WinBUGS
#------------------------------------------------------------------------
```

```
# Import data

data <- read.csv("../../Data/melanoma_os_ipd_nc.csv")

d1 <- data[data$studyCode==1,]

d2 <- data[data$studyCode==2,]

d3 <- data[data$studyCode==3,]

d4 <- data[data$studyCode==4,]

d5 <- data[data$studyCode==5,]

d6 <- data[data$studyCode==6,]

d7 <- data[data$studyCode==7,]

d8 <- data[data$studyCode==8,]

d9 <- data[data$studyCode==9,]

d10 <- data[data$studyCode==10,]


# Fit gengamma parametric model to 2 arm trials

gg1 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d1, dist="gengamma",
method="BFGS")

gg2 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d2, dist="gengamma",
method="BFGS")

gg3 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d3, dist="gengamma",
method="BFGS")

gg4 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d4, dist="gengamma",
method="BFGS")

gg6 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d6, dist="gengamma",
method="BFGS")

gg7 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d7, dist="gengamma",
method="BFGS")

gg8 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d8, dist="gengamma",
method="BFGS")

gg9 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d9, dist="gengamma",
method="BFGS")

gg10 <- flexsurvreg(formula=Surv(time,event) ~ arm, data=d10, dist="gengamma",
method="BFGS")


# Fit gengamma parametric model to3 arm trials - need to convert arm to a factor variable

d5$treat <- factor(d5$arm, labels=c("A", "B", "C"))

gg5 <- flexsurvreg(formula=Surv(time,event) ~ treat, data=d5, dist="gengamma",
method="BFGS")
```

```
# Vector of beta coefficients
beta <- c(gg1$coefficients[["arm"]], gg2$coefficients[["arm"]], gg3$coefficients[["arm"]],
          gg4$coefficients[["arm"]], gg6$coefficients[["arm"]], gg7$coefficients[["arm"]],
          gg8$coefficients[["arm"]], gg9$coefficients[["arm"]], gg10$coefficients[["arm"]],
          gg5$coefficients[["treatB"]], gg5$coefficients[["treatC"]])

# Vector of beta SE
se <- c(sqrt(diag(gg1$cov))["arm"], sqrt(diag(gg2$cov))["arm"], sqrt(diag(gg3$cov))["arm"],
        sqrt(diag(gg4$cov))["arm"], sqrt(diag(gg6$cov))["arm"], sqrt(diag(gg7$cov))["arm"],
        sqrt(diag(gg8$cov))["arm"], sqrt(diag(gg9$cov))["arm"], sqrt(diag(gg10$cov))["arm"],
        sqrt(diag(gg5$cov))["treatB"], sqrt(diag(gg5$cov))["treatC"])

# Vector of LCI
lci <- c(gg1$res[4,2], gg2$res[4,2], gg3$res[4,2], gg4$res[4,2],
         gg6$res[4,2], gg7$res[4,2], gg8$res[4,2], gg9$res[4,2], gg10$res[4,2],
         gg5$res[4,2], gg5$res[5,2])

# Vector of UCI
uci <- c(gg1$res[4,3], gg2$res[4,3], gg3$res[4,3], gg4$res[4,3],
         gg6$res[4,3], gg7$res[4,3], gg8$res[4,3], gg9$res[4,3], gg10$res[4,3],
         gg5$res[4,3], gg5$res[5,3])

# Calculate Exponentiated beta
expbeta <- exp(beta)

# Create a data frame
aft_data <- data.frame(STUDY=c("BREAK-3", "BRF113220", "BRIM-3", "CheckMate 066",
"CheckMate 069", "COLUMBUS",
                               "COMBI-d", "COMBI-v", "KEYNOTE-006", "CheckMate 067",
"CheckMate 067"),
                       MEAN=beta, MEANSE=se, lci=lci, uci=uci, expbeta=expbeta, COV=NA,
                       t1=c(1, 2, 1, 1, 4, 5, 2, 3, 4, 4, 4),
```

```
                    t2=c(2, 3, 5, 6, 7, 8, 3, 5, 9, 6, 7))


# Calculate variance of baseline treatment - Checkmate 67

gg5[["cov"]]

V67 <- gg5$cov[5,4]


# Add variance to 3 arm trial in aft_data frame

aft_data$COV[10:11] <- V67


# Calculate AIC

AIC_os_nc_scale<- c(gg1$AIC, gg2$AIC, gg3$AIC, gg4$AIC, gg5$AIC, gg6$AIC,
gg7$AIC, gg8$AIC, gg9$AIC, gg10$AIC)


#### OPTIONAL: Can save aft_data as a CSV file ####

# write.csv(aft_data,file="Gamma results/aft_data_os_nc.csv")


#------------------------------------------------------------------------

# Step 2: Reshape aft_data for WinBUGS analysis (only needed if there are 3 arm trials)

#------------------------------------------------------------------------


#### OPTIONAL: If you already have aft_data (obtained from step 1) ####

#### You can skip step 1 and load aft_data here ####

# data <- read.csv("Gamma results/aft_data_os_nc.csv")


# This step is required as three arm trials will occupy two rows of data in the csv file

# created in part one. So this step reformats the data so that three arm trials are in

# one row rather than 2 separate ones.


# Reformat data

df<-subset(aft_data, select=-c(uci, lci, expbeta))

df$arm<- df$na <- rep(NA,nrow(df))

for (i in 1:nrow(df)){

  df_sub<-subset(df,STUDY==STUDY[i])
```

```
  df$arm[i]<-rank(df_sub$t2)[df_sub$t2==df$t2[i]]+1
  df$na[i]<- nrow(df_sub)+1
}

df2<-reshape(df,timevar="arm", idvar=c("STUDY","na"), direction="wide")
df2<- df2[order(df2$na),]
df2$hash<- rep("#",nrow(df2))
df3<-
df2[,c("MEAN.2","MEANSE.2","MEAN.3","MEANSE.3","COV.2","t1.2","t2.2","t2.3","na","ha
sh","STUDY")]

reshaped.data<-rename(df3,c(MEAN.2="y[,2]", MEANSE.2="se[,2]", MEAN.3="y[,3]",
MEANSE.3="se[,3]", COV.2="V[]",
                t1.2="t1[]", t2.2="t2[]",       t2.3="t3[]", na=          "na", hash=     "#",
STUDY="study"))

reshaped.data$y1 <- 0
reshaped.data$se1 <- 0

#---------------------------------------------------------------------------
# Step 3: Prep data to load in to WinBUGS as bugs_data
#---------------------------------------------------------------------------

# Set the location for WinBUGS
bugs.directory <- "C:/Program Files (x86)/WinBUGS14"

# WinBUGS burn-in & simulation size
num.sims <- 20000
burn.in <- 10000

# Number of studies
ns <- nrow(reshaped.data)

# Number of treatments
```

```
nt <- max(reshaped.data$t2)


# Number of arms in each trial

reshaped.data$na <- 2

reshaped.data$na[reshaped.data$study=="CheckMate 067"] <- 3


# Create arrays to load into bugs_data frame

y <- array(c(reshaped.data$y1, reshaped.data$`y[,2]`, reshaped.data$`y[,3]`), dim=c(ns,3))

se <- array(c(reshaped.data$se1, reshaped.data$`se[,2]`, reshaped.data$`se[,3]`),
dim=c(ns,3))

t <- array(c(reshaped.data$t1, reshaped.data$t2, reshaped.data$t3), dim=c(ns,3))


# Create bugs_data to load into WinBUGS ns2 corresponds to the number of two arm
trials

# and ns3 the number of three arm trials.

bugs_data <- list(ns2=9, ns3=1, nt=nt, t=t, y=y, se=se, na=reshaped.data$na,
V=reshaped.data$V)


#------------------------------------------------------------------------

# Step 4: Set initial values

#------------------------------------------------------------------------


# Set initial values chain one

d1 <- c(NA, rep(0, nt-1))


# Set initial values chain two

d2 <- c(NA, rep(0.1, nt-1))


# Set initial values chain three

d3 <- c(NA, rep(-0.1, nt-1))


# Make a list of initial values for all chains

fe_inits <- list(list(d=d1),
                 list(d=d2),
```

```
            list(d=d3))


#---------------------------------------------------------------------------
# Step 5: Fit FE model in WinBUGS
#---------------------------------------------------------------------------

# Run NMA in WinBUGS
bugs.fe <- bugs(data=bugs_data, inits=fe_inits,
            parameters.to.save=c("d", "TR", "best", "prob", "rk", "totresdev"),
            model.file="FE_model.txt", clearWD=F,
            summary.only=FALSE, n.iter=(num.sims+burn.in),
            n.sims=num.sims, n.burnin=burn.in, n.chains=3,
            bugs.seed=212034, bugs.directory=bugs.directory,
            debug=TRUE, DIC=TRUE)

# Save results in a data frame
fe_results <- as.data.frame(bugs.fe$summary)

# Save results in csv file
write.csv(fe_results,file="NMA results/fe_results_os_nc.csv")

# Check results
results2 <- bugs.fe$sims.matrix[,grep("d",rownames(bugs.fe$summary))]
results2 <- cbind(rep(0,dim(results2)[1]),results2)
summary(results2)

results_mcmc<-mcmc(results2)
par(mfrow=c(3,2))

# Check autocorrelation
autocorr.plot(results_mcmc[,2:9])
```

```
# Check trace for convergence
traceplot(results_mcmc[,2])
traceplot(results_mcmc[,3])
traceplot(results_mcmc[,4])
traceplot(results_mcmc[,5])
traceplot(results_mcmc[,6])
traceplot(results_mcmc[,7])
traceplot(results_mcmc[,8])
traceplot(results_mcmc[,9])

# Histograms of posterior distributions
densplot(results_mcmc[,2])
densplot(results_mcmc[,3])
densplot(results_mcmc[,4])
densplot(results_mcmc[,5])
densplot(results_mcmc[,6])
densplot(results_mcmc[,7])
densplot(results_mcmc[,8])
densplot(results_mcmc[,9])

#------------------------------------------------------------------------
# Part 2: Results - Survival plot, Area under curve, Rank plots
#------------------------------------------------------------------------

# Part two of the analysis is only dealing with the results obtained from part 1

# Step 6 generates a plot with survival curves for each treatment

# Step 7 calculate the area under the curves for each treatment

# Step 8 generates a ranking plot, showing the liklihood of each treatment occupying each
rank
```

```
#### OPTIONAL: If you already have result data, obtained from Part 1 ####
#### You can skip all previous steps and load result data here ####
fe_results <- read.csv("NMA results/fe_results_os_nc.csv")

#------------------------------------------------------------------------
# Step 6: Plot survival curves for each treatment
#------------------------------------------------------------------------

# Fit generalised gamma model just for the DTIC arm of the CheckMate 066 trial
sc <- data[data$studyCode==4 & data$txCode==1,]

ma <- flexsurvreg(formula=Surv(time,event)~1,
                  data=sc, dist="gengamma", method="BFGS")

# Identify coefficicents needed for predicting survival
mu <- ma$coefficients["mu"]
sigma <- exp(ma$coefficients["sigma"])
q <- ma$coefficients["Q"]

# Store treatment effects from NMA for each treatment
trt2 <- fe_results$mean[1]
trt3 <- fe_results$mean[2]
trt4 <- fe_results$mean[3]
trt5 <- fe_results$mean[4]
trt6 <- fe_results$mean[5]
trt7 <- fe_results$mean[6]
trt8 <- fe_results$mean[7]
trt9 <- fe_results$mean[8]

#Calculate survival across 120 months
x <- seq(0,120,3)
S.trt1 <- 1-pgengamma(x, mu = mu, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)
```

```
S.trt2 <- 1-pgengamma(x, mu = mu+trt2, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)

S.trt3 <- 1-pgengamma(x, mu = mu+trt3, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)

S.trt4 <- 1-pgengamma(x, mu = mu+trt4, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)

S.trt5 <- 1-pgengamma(x, mu = mu+trt5, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)

S.trt6 <- 1-pgengamma(x, mu = mu+trt6, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)

S.trt7 <- 1-pgengamma(x, mu = mu+trt7, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)

S.trt8 <- 1-pgengamma(x, mu = mu+trt8, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)

S.trt9 <- 1-pgengamma(x, mu = mu+trt9, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)


graph_data <- data.frame(time=x, trt1=S.trt1, trt2=S.trt2, trt3=S.trt3, trt4=S.trt4, trt5=S.trt5,
trt6=S.trt6,
                         trt7=S.trt7, trt8=S.trt8, trt9=S.trt9)


# Save graph data

write.csv(graph_data, file="Plots & AUC/graph_data_os_nc.csv")


# Calculate KM estimate for DTIC from CheckMate 066

ipd_data <- sc[sc$txCode==1,]

KM.est<-survfit(Surv(time,event)~1, data=ipd_data, type="kaplan-meier", conf.int=FALSE)


# Using colours from the Safe palette from rcartocolor

colors=c("#88CCEE", "#CC6677", "#DDCC77", "#117733", "#332288", "#AA4499",
         "#44AA99", "#999933", "#882255", "#661100", "#6699CC", "#888888")


# Start by plotting the Kaplan-Meier DTIC curve for the CheckMate 066 trial

plot(KM.est,xlab="Time (months)",ylab="Overall Survival",xaxt="n",yaxt="n",main="
",xlim=c(0,120),ylim=c(0,1),

     mark.time=FALSE, col=color[10], conf.int=F)
```

```
#Add y axis (2 specifies that axis goes on the left of the plot)
axis(2, at=c(0, 0.2, 0.4, 0.6, 0.8, 1))

#Add x axis (1 specified that axis goes at the bottom of the plot)
axis(1, at=c(0, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120))

# Add survival curves for each treatment
lines(x,S.trt1, col=color[1])
lines(x,S.trt2, col=color[2])
lines(x,S.trt3, col=color[3])
lines(x,S.trt4, col=color[4])
lines(x,S.trt5, col=color[5])
lines(x,S.trt6, col=color[6])
lines(x,S.trt7, col=color[7])
lines(x,S.trt8, col=color[8])
lines(x,S.trt9, col=color[9])

# Add legend
legend("topright",
       c("KM", "DTIC", "Dab", "Dab + Tram", "Ipi",
         "Vem", "Nivo", "Nivo + Ipi", "Enco + Bini", "Pembro"),
       col=c(color[10], color[1], color[2], color[3], color[4], color[5],
             color[6], color[7], color[8], color[9]),
       lty=c(1,1,1,1), ncol=3, text.width=20, box.lty=0, y.intersp = 2)

# save plot
dev.copy(pdf, "Plots & AUC/survival_plot_os_nc.pdf")
dev.off()

#-------------------------------------------------------------------------
# Step 7: Calculate area under curve for different treatments
#-------------------------------------------------------------------------
```

```
# Create data frame to store AUC results
auc <- data.frame(trt=c(1:9), auc=NA)

# Calculate AUC
auc$auc[1] <- trapz(graph_data$time, graph_data$trt1)
auc$auc[2] <- trapz(graph_data$time, graph_data$trt2)
auc$auc[3] <- trapz(graph_data$time, graph_data$trt3)
auc$auc[4] <- trapz(graph_data$time, graph_data$trt4)
auc$auc[5] <- trapz(graph_data$time, graph_data$trt5)
auc$auc[6] <- trapz(graph_data$time, graph_data$trt6)
auc$auc[7] <- trapz(graph_data$time, graph_data$trt7)
auc$auc[8] <- trapz(graph_data$time, graph_data$trt8)
auc$auc[9] <- trapz(graph_data$time, graph_data$trt9)

# Save area under curves results as CSV file
write.csv(auc, file="Plots & AUC/auc_os_nc.csv")

#------------------------------------------------------------------------
# Step 8: Rank plot
#------------------------------------------------------------------------

# Keep rows for ranking only
rankdata <- fe_results[26:106,]

# Variable for rank
rankdata$rank_code <- rep(9:1, 9)

# Restrict probability to 2 decimal places
rankdata$prob <- round(rankdata$mean, 2)

# Add a treatment label
```

```
rankdata$Treatment <- c(rep("DTIC", 9), rep("Dab", 9), rep("Dab + Tram", 9), rep("Ipi", 9), rep("Vem", 9),
                        rep("Nivo", 9), rep("Nivo + Ipi", 9), rep("Enco + Bini", 9), rep("Pembro", 9))


# Rename mean column
names(rankdata)[names(rankdata)=="mean"] <- "Probability"


# Plot with text
q <- ggplot(rankdata, aes(x=rank_code, y=Treatment)) +
  geom_point(aes(size=Probability), shape=21, colour="skyblue", fill="skyblue") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black", fill=NA, size=1),
        legend.position="bottom") +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_discrete(name="Treatment") +
  geom_text(aes(label=prob))
q


# save plot
dev.copy(pdf, "Plots & AUC/rank_plot_os_nc.pdf")
dev.off()
```

A.8.5.1.2 WinBUGS code

```
# Normal likelihood, identity link
# Trial-level data given as treatment differences
# Fixed effects model for multi-arm trials


model{                          # *** PROGRAM STARTS


for(i in 1:ns2) {                  # LOOP THROUGH 2-ARM STUDIES
   y[i,2] ~ dnorm(delta[i,2],prec[i,2]) # normal likelihood for 2-arm trials
```

```
#Deviance contribution for trial i
   resdev[i] <- (y[i,2]-delta[i,2])*(y[i,2]-delta[i,2])*prec[i,2]
 }

for(i in (ns2+1):(ns2+ns3)) {        # LOOP THROUGH THREE-ARM STUDIES
   for (k in 1:(na[i]-1)) {    # set variance-covariance matrix
      for (j in 1:(na[i]-1)) {
         Sigma[i,j,k] <- V[i]*(1-equals(j,k)) + var[i,k+1]*equals(j,k)
       }
    }
   Omega[i,1:(na[i]-1),1:(na[i]-1)] <- inverse(Sigma[i,,])  #Precision matrix
# multivariate normal likelihood for 3-arm trials
   y[i,2:na[i]] ~ dmnorm(delta[i,2:na[i]],Omega[i,1:(na[i]-1),1:(na[i]-1)])
#Deviance contribution for trial i
   for (k in 1:(na[i]-1)){  # multiply vector & matrix
      ydiff[i,k]<- y[i,(k+1)] - delta[i,(k+1)]
      z[i,k]<- inprod2(Omega[i,k,1:(na[i]-1)], ydiff[i,1:(na[i]-1)])
    }
   resdev[i]<- inprod2(ydiff[i,1:(na[i]-1)], z[i,1:(na[i]-1)])
 }

for(i in 1:(ns2+ns3)){               #   LOOP THROUGH ALL STUDIES
    for (k in 2:na[i]) {            #  LOOP THROUGH ARMS
      var[i,k] <- pow(se[i,k],2)   # calculate variances
      prec[i,k] <- 1/var[i,k]      # set precisions
      delta[i,k] <-  d[t[i,k]] - d[t[i,1]]
    }
 }

totresdev <- sum(resdev[])           #Total Residual Deviance

d[1]<-0       # treatment effect is zero for reference treatment
```

```
# vague priors for treatment effects
for (k in 2:nt){
        d[k] ~ dnorm(0,.0001)
}

# convert to time ratio
for (k in 2:nt){
        TR[k] <- exp(d[k])
}


# ranking on relative scale
    for (k in 1:nt) {
        rk[k] <- rank(d[],k) # assumes events are “bad”
        best[k] <- equals(rk[k],1) #calculate probability that treat k is best
    for (h in 1:nt){
        prob[h,k] <- equals(rk[k],h) } # calculates probability that treat k is h-th best
    }
}                           # *** PROGRAM ENDS
```

A.8.5.2 Two treatment effects

A.8.5.2.1 R code

```
# AFT NMA for Time to Event Outcomes - 2 treatment effects
#-------------------------------------------------------------------------
# Instructions
#-------------------------------------------------------------------------
# The following code enables users to run a generalized gamma NMA with
# 2 treatment effects in OpenBUGS. The code has 2 parts. The first part fits
# parametric curves to the data to get the outputs required to run the NMA
# in OpenBUGS. The user can then copy the data into OpenBUGS and run the
```

```
# analysis directly in OpenBUGS. The second part uses results from the NMA
# in OpenBUGS to make plots and estimate the area under the curve for
# different treatments. Each part and each step has more detail of what the
# code does at that point.
#------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------
# Load required packages
library(survival)
library(flexsurv)
library(pracma)
library(reshape)
library(reshape2)
library(ggplot2)
library(ggpubr)

# Start with clean environment
rm(list=ls())

# Set working directory
setwd()

#------------------------------------------------------------------------
# Part 1: Run NMA
#------------------------------------------------------------------------

# Part 1 of the analysis runs the NMA and produces results

# Step 1 fits AFT models with two treatment effects in R to each study in the network

# Step 2 formats the data as the 'bugs_data' frame to be read into winbugs
# later using R2winbugs
```

```
# Step 3 sets the initial values for the NMA.


# Step 4 runs the NMA in OpenBUGS using R2Winbugs. Note, to run the NMA you must have
# the winbugs model saved as a text file. In this case it is 'FE_model.txt'
# The code will then produce and saves results, and allows you to check the
# density and trace plots.
#--------------------------------------------------------------------------
# Step 1: Run AFT analysis to obtain data to feed into OpenBUGS
#--------------------------------------------------------------------------


# Import data
data <- read.csv("../../Data/melanoma_os_ipd_nc.csv")
d1 <- data[data$studyCode==1,]
d2 <- data[data$studyCode==2,]
d3 <- data[data$studyCode==3,]
d4 <- data[data$studyCode==4,]
d5 <- data[data$studyCode==5,]
d6 <- data[data$studyCode==6,]
d7 <- data[data$studyCode==7,]
d8 <- data[data$studyCode==8,]
d9 <- data[data$studyCode==9,]
d10 <- data[data$studyCode==10,]


# Fit gengamma parametric model to 2 arm trials
gg1 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d1,
dist="gengamma", method="BFGS")
gg2 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d2,
dist="gengamma", method="BFGS")
gg3 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d3,
dist="gengamma", method="BFGS")
gg4 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d4,
dist="gengamma", method="BFGS")
```

```
gg6 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d6,
dist="gengamma", method="BFGS")

gg7 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d7,
dist="gengamma", method="BFGS")

gg8 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d8,
dist="gengamma", method="BFGS")

gg9 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d9,
dist="gengamma", method="BFGS")

gg10 <- flexsurvreg(formula=Surv(time,event) ~ arm + sigma(arm), data=d10,
dist="gengamma", method="BFGS")


# Fit gengamma parametric model to3 arm trials - need to convert arm to a factor variable

d5$treat <- factor(d5$arm, labels=c("A", "B", "C"))

gg5 <- flexsurvreg(formula=Surv(time,event) ~ treat + sigma(treat), data=d5,
dist="gengamma", method="BFGS")


# Vector of treatment effect on mu, arm2 vs arm1

dmu2 <- c(gg1$coefficients[["arm"]], gg2$coefficients[["arm"]], gg3$coefficients[["arm"]],

        gg4$coefficients[["arm"]], gg6$coefficients[["arm"]], gg7$coefficients[["arm"]],

        gg8$coefficients[["arm"]], gg9$coefficients[["arm"]], gg10$coefficients[["arm"]],

        gg5$coefficients[["treatB"]])


# Vector of treatment effect on mu, arm3 vs arm1

dmu3 <- c(rep(NA, 9), gg5$coefficients[["treatC"]])


# Vector of treatment effects on sigma, arm2 vs arm1

dsigma2 <- c(gg1$coefficients[["sigma(arm)"]], gg2$coefficients[["sigma(arm)"]],
gg3$coefficients[["sigma(arm)"]],

        gg4$coefficients[["sigma(arm)"]], gg6$coefficients[["sigma(arm)"]],
gg7$coefficients[["sigma(arm)"]],

        gg8$coefficients[["sigma(arm)"]], gg9$coefficients[["sigma(arm)"]],
gg10$coefficients[["sigma(arm)"]],

        gg5$coefficients[["sigma(treatB)"]])


# Vector of treatment effect on sigma, arm3 vs arm1

dsigma3 <- c(rep(NA, 9), gg5$coefficients[["sigma(treatC)"]])
```

```
# Precision for each trial
prec1<-solve(gg1[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])
prec2<-solve(gg2[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])
prec3<-solve(gg3[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])
prec4<-solve(gg4[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])
prec5<-solve(gg5[["cov"]][c("treatB","sigma(treatB)","treatC",
"sigma(treatC)"),c("treatB","sigma(treatB)","treatC", "sigma(treatC)")])
prec6<-solve(gg6[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])
prec7<-solve(gg7[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])
prec8<-solve(gg8[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])
prec9<-solve(gg9[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])
prec10<-solve(gg10[["cov"]][c("arm","sigma(arm)"),c("arm","sigma(arm)")])

# Vector of elements in precision matrices: dmu2
P_11 <- c(prec1["arm","arm"], prec2["arm","arm"], prec3["arm","arm"], prec4["arm","arm"],
        prec6["arm","arm"], prec7["arm","arm"], prec8["arm","arm"], prec9["arm","arm"],
        prec10["arm","arm"], prec5["treatB","treatB"])

# Vector of elements in precision matrices: dsigma2
P_22 <- c(prec1["sigma(arm)","sigma(arm)"], prec2["sigma(arm)","sigma(arm)"],
        prec3["sigma(arm)","sigma(arm)"], prec4["sigma(arm)","sigma(arm)"],
        prec6["sigma(arm)","sigma(arm)"], prec7["sigma(arm)","sigma(arm)"],
        prec8["sigma(arm)","sigma(arm)"], prec9["sigma(arm)","sigma(arm)"],
        prec10["sigma(arm)","sigma(arm)"], prec5["sigma(treatB)","sigma(treatB)"])

#Vector of elements in precision matrices:dmu3
P_33 <- c(rep(NA, 9), prec5["treatC","treatC"])

#Vector of elements in precision matrices:dsigma3
P_44 <- c(rep(NA, 9), prec5["sigma(treatC)","sigma(treatC)"])

#Vector of precisions between dmu2 and dsigma2
```

```
P_12 <- c(prec1["arm","sigma(arm)"], prec2["arm","sigma(arm)"],
prec3["arm","sigma(arm)"],
        prec4["arm","sigma(arm)"], prec6["arm","sigma(arm)"], prec7["arm","sigma(arm)"],
        prec8["arm","sigma(arm)"], prec9["arm","sigma(arm)"], prec10["arm","sigma(arm)"],
        prec5["treatB","sigma(treatB)"])


#Vector of precisions between dmu2 and dmu3
P_13 <- c(rep(NA, 9), prec5["treatB","treatC"])


#Vector of precisions between dmu2 and dsigma3
P_14 <- c(rep(NA, 9), prec5["treatB","sigma(treatC)"])


#Vector of precisions between dsigma2 and dmu3
P_23 <- c(rep(NA, 9), prec5["sigma(treatB)","treatC"])


#Vector of precisions between dsigma2 and dsigma3
P_24 <- c(rep(NA, 9), prec5["sigma(treatB)","sigma(treatC)"])


#Vector of precisions between dmu3 and dsigma3
P_34 <- c(rep(NA, 9), prec5["treatC","sigma(treatC)"])


na<- c(rep(2,9), rep(3,1))


t1<- c(1, 2, 1, 1, 4, 5, 2, 3, 4, 4)
t2<- c(2, 3, 5, 6, 7, 8, 3, 5, 9, 6)
t3<- c(rep(NA,9),7)


STUDY<- c("BREAK-3", "BRF113220", "BRIM-3", "CheckMate 066", "CheckMate 069",
        "COLUMBUS", "COMBI-d", "COMBI-v", "KEYNOTE-006", "CheckMate 067")
hash<- rep("#",10)


# Create a data frame
df <- data.frame(dmu2, dmu3, dsigma2, dsigma3,
```

```
            P_11, P_22, P_33, P_44, P_12, P_13, P_14, P_23, P_24, P_34,
            t1, t2, t3, na, hash, STUDY)


# Create new data frame that renames columns for OpenBUGS analysis
aft_data<-rename(df,c(dmu2="y[,1]", dsigma2="y[,2]", dmu3="y[,3]", dsigma3="y[,4]",
                      P_11="P[,1,1]", P_22="P[,2,2]",  P_33="P[,3,3]", P_44="P[,4,4]",
                      P_12="P[,1,2]", P_13="P[,1,3]", P_14="P[,1,4]",
                      P_23="P[,2,3]", P_24="P[,2,4]", P_34="P[,3,4]",
                      t1="t[,1]", t2="t[,2]",    t3="t[,3]", na=  "na[]", hash=   "#", STUDY="STUDY"))


# Calculate AIC
AIC_os_nc_scale<- c(gg1$AIC, gg2$AIC, gg3$AIC, gg4$AIC, gg5$AIC, gg6$AIC,
gg7$AIC, gg8$AIC, gg9$AIC, gg10$AIC)


#### OPTIONAL: Can save aft_data as a CSV file ####
# aft_data file is ready to load into OpenBUGS to run NMA
# write.csv(aft_data,file="Gamma results/aft_data_os_nc.csv")


#------------------------------------------------------------------------
# Part 2: Results - Survival plot, Area under curve, Rank plots
#------------------------------------------------------------------------


# Part two of the analysis is only dealing with the results obtained from part 1


# Step 2 generates a plot with survival curves for each treatment


# Step 3 calculate the area under the curves for each treatment


# Step 4 generates a ranking plot, showing the liklihood of each treatment occupying each
rank


#### OPTIONAL: If you already have result data, obtained from Part 1 ####
#### You can skip all previous steps and load result data here ####
```

```
# fe_results <- read.csv("NMA results/fe_results_os_nc.csv")


#------------------------------------------------------------------------
# Step 2: Plot survival curves for each treatment
#------------------------------------------------------------------------


# Fit generalised gamma model just for the DTIC arm of the CheckMate 066 trial
sc <- data[data$studyCode==4 & data$txCode==1,]


ma <- flexsurvreg(formula=Surv(time,event)~1,
                  data=sc, dist="gengamma", method="BFGS")


# Identify coefficicents needed for predicting survival
mu <- ma$coefficients["mu"]
sigma <- exp(ma$coefficients["sigma"])
q <- ma$coefficients["Q"]


# Store treatment effects for each treatment
trt<-matrix(rep(NA,18),9,2)
for(i in 2:9){
  for (j in 1:2){
    nodename<-paste("d[",i,",",j,"]",sep="")
    trt[i,j]<- fe_results$mean[fe_results$node==nodename]
  }
}


#Calculate survival across 120 months
x <- seq(0,120,3)
S.trt1 <- 1-pgengamma(x, mu = mu, sigma = sigma, Q=q, lower.tail = TRUE, log.p =
FALSE)
S.trt2 <- 1-pgengamma(x, mu = mu+trt[2,1], sigma = sigma+trt[2,2], Q=q, lower.tail =
TRUE, log.p = FALSE)
S.trt3 <- 1-pgengamma(x, mu = mu+trt[3,1], sigma = sigma+trt[3,2], Q=q, lower.tail =
TRUE, log.p = FALSE)
```

```
S.trt4 <- 1-pgengamma(x, mu = mu+trt[4,1], sigma = sigma+trt[4,2], Q=q, lower.tail =
TRUE, log.p = FALSE)

S.trt5 <- 1-pgengamma(x, mu = mu+trt[5,1], sigma = sigma+trt[5,2], Q=q, lower.tail =
TRUE, log.p = FALSE)

S.trt6 <- 1-pgengamma(x, mu = mu+trt[6,1], sigma = sigma+trt[6,2], Q=q, lower.tail =
TRUE, log.p = FALSE)

S.trt7 <- 1-pgengamma(x, mu = mu+trt[7,1], sigma = sigma+trt[7,2], Q=q, lower.tail =
TRUE, log.p = FALSE)

S.trt8 <- 1-pgengamma(x, mu = mu+trt[8,1], sigma = sigma+trt[8,2], Q=q, lower.tail =
TRUE, log.p = FALSE)

S.trt9 <- 1-pgengamma(x, mu = mu+trt[9,1], sigma = sigma+trt[9,2], Q=q, lower.tail =
TRUE, log.p = FALSE)


graph_data <- data.frame(time=x, trt1=S.trt1, trt2=S.trt2, trt3=S.trt3, trt4=S.trt4, trt5=S.trt5,
trt6=S.trt6,
                         trt7=S.trt7, trt8=S.trt8, trt9=S.trt9)


# Save graph data

write.csv(graph_data, file="Plots & AUC/graph_data_os_nc.csv")


# Calculate KM estimate for DTIC from CheckMate 066

ipd_data <- sc[sc$txCode==1,]

KM.est<-survfit(Surv(time,event)~1, data=ipd_data, type="kaplan-meier", conf.int=FALSE)


# Using colours from the Safe palette from rcartocolor

colors=c("#88CCEE", "#CC6677", "#DDCC77", "#117733", "#332288", "#AA4499",

        "#44AA99", "#999933", "#882255", "#661100", "#6699CC", "#888888")


# Start by plotting the Kaplan-Meier DTIC curve for the CheckMate 066 trial

plot(KM.est,xlab="Time (months)",ylab="Overall Survival",xaxt="n",yaxt="n",main="
",xlim=c(0,120),ylim=c(0,1),

    mark.time=FALSE, col=color[10], conf.int=F)


#Add y axis (2 specifies that axis goes on the left of the plot)

axis(2, at=c(0, 0.2, 0.4, 0.6, 0.8, 1))
```

```
#Add x axis (1 specified that axis goes at the bottom of the plot)
axis(1, at=c(0, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120))

# Add survival curves for each treatment
lines(x,S.trt1, col=color[1])
lines(x,S.trt2, col=color[2])
lines(x,S.trt3, col=color[3])
lines(x,S.trt4, col=color[4])
lines(x,S.trt5, col=color[5])
lines(x,S.trt6, col=color[6])
lines(x,S.trt7, col=color[7])
lines(x,S.trt8, col=color[8])
lines(x,S.trt9, col=color[9])

# Add legend
legend("topright",
       c("KM", "DTIC", "Dab", "Dab + Tram", "Ipi",
         "Vem", "Nivo", "Nivo + Ipi", "Enco + Bini", "Pembro"),
       col=c(color[10], color[1], color[2], color[3], color[4], color[5],
             color[6], color[7], color[8], color[9]),
       lty=c(1,1,1,1), ncol=3, text.width=20, box.lty=0, y.intersp = 2)

# save plot
dev.copy(pdf, file="Plots & AUC/survival_plot_os_nc.pdf")
dev.off()

#-------------------------------------------------------------------------
# Step 3: Calculate area under curve for different treatments
#-------------------------------------------------------------------------

# Create data frame to store AUC results
auc <- data.frame(trt=c(1:9), auc=NA)
```

```
# Calculate AUC
auc$auc[1] <- trapz(graph_data$time, graph_data$trt1)
auc$auc[2] <- trapz(graph_data$time, graph_data$trt2)
auc$auc[3] <- trapz(graph_data$time, graph_data$trt3)
auc$auc[4] <- trapz(graph_data$time, graph_data$trt4)
auc$auc[5] <- trapz(graph_data$time, graph_data$trt5)
auc$auc[6] <- trapz(graph_data$time, graph_data$trt6)
auc$auc[7] <- trapz(graph_data$time, graph_data$trt7)
auc$auc[8] <- trapz(graph_data$time, graph_data$trt8)
auc$auc[9] <- trapz(graph_data$time, graph_data$trt9)

# Save area under curves results as CSV file
write.csv(auc, file="Plots & AUC/auc_os_nc.csv")

#------------------------------------------------------------------------
# Step 4: Rank plots
#------------------------------------------------------------------------

# Keep rows for ranking only
rankdata <- fe_results[51:212,]

# Variable for rank
rankdata$rank_code <- rep(9:1, each = 2, len = 18)

# Restrict probability to 2 decimal places
rankdata$prob <- round(rankdata$mean, 2)

# Add a treatment label
rankdata$Treatment <- c(rep("DTIC", 18), rep("Dab", 18), rep("Dab + Tram", 18), rep("Ipi",
18), rep("Vem", 18),
                        rep("Nivo", 18), rep("Nivo + Ipi", 18), rep("Enco + Bini", 18), rep("Pembro",
18))
```

```
# Rename mean column
names(rankdata)[names(rankdata)=="mean"] <- "Probability"

# Create dummy indicator to show even/odd
row_odd <- seq_len(nrow(rankdata)) %% 2

# Create variable for parameter
rankdata$parameter <- rep(c("mu", "sigma"), length.out=nrow(rankdata))

# Store rankings for treatments as two data frame, one for scale, one for shape
p1 <- rankdata[row_odd == 1,]
p2 <- rankdata[row_odd == 0,]

# Create data frame for mu (location parameter)
rp1.1 <- p1[1:9,]
rp2.1 <- p1[10:18,]
rp3.1 <- p1[19:27,]
rp4.1 <- p1[28:36,]
rp5.1 <- p1[37:45,]
rp6.1 <- p1[46:54,]
rp7.1 <- p1[55:63,]
rp8.1 <- p1[64:72,]
rp9.1 <- p1[73:81,]

# Create data frame for sigma (scale parameter)
rp1.2 <- p2[1:9,]
rp2.2 <- p2[10:18,]
rp3.2 <- p2[19:27,]
rp4.2 <- p2[28:36,]
rp5.2 <- p2[37:45,]
rp6.2 <- p2[46:54,]
```

```
rp7.2 <- p2[55:63,]
rp8.2 <- p2[64:72,]
rp9.2 <- p2[73:81,]

# Combine mu and sigma (location and scale parameters)
rp1 <- rbind(rp1.1, rp1.2)
rp2 <- rbind(rp2.1, rp2.2)
rp3 <- rbind(rp3.1, rp3.2)
rp4 <- rbind(rp4.1, rp4.2)
rp5 <- rbind(rp5.1, rp5.2)
rp6 <- rbind(rp6.1, rp6.2)
rp7 <- rbind(rp7.1, rp7.2)
rp8 <- rbind(rp8.1, rp8.2)
rp9 <- rbind(rp9.1, rp9.2)

# DTIC plot with text
a <- ggplot(rp1, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="DTIC treatment effects rankings") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
a

# save plot
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_DTIC.pdf")
dev.off()
```

```
# Dabrafenib plot with text
b <- ggplot(rp2, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="Dabrafenib treatment effects rankings") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
b


# save plot
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_Dab.pdf")
dev.off()


# Dabrafenib + Trametinib plot with text
c <- ggplot(rp3, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="Dabrafenib + Trametinib treatment effects rankings") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
c


# save plot
```

```
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_D+T.pdf")
dev.off()


# Ipilimumab plot with text
d <- ggplot(rp4, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="Ipilimumab scale and shape parameter rankings") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
d


# save plot
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_Ipi.pdf")
dev.off()


# Vemurafenib plot with text
e <- ggplot(rp5, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="Vemurafenib treatment effects rankings") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
```

```
e


# save plot
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_Vem.pdf")
dev.off()


# Nivolumab plot with text
f <- ggplot(rp6, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="Nivolumab treatment effects rankings") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
f


# save plot
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_Nivo.pdf")
dev.off()


# Nivolumab + Ipilimumab plot with text
g <- ggplot(rp7, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="Nivolumab + Ipilimumab treatment effects rankings") +
```

```
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
g


# save plot
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_N+I.pdf")
dev.off()


# Encorafenib + Binimetinib plot with text
h <- ggplot(rp8, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="Encorafenib + Binimetinib treatment effects rankings") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
h


# save plot
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_E+B.pdf")
dev.off()


# Pembrolizumab plot with text
i <- ggplot(rp9, aes(x=rank_code, y=prob, group=parameter)) +
  geom_line(aes(linetype=parameter, color=parameter)) +
  geom_point(aes(shape=parameter, color=parameter)) +
  scale_size_area(max_size=10) +
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
```

```
  scale_y_continuous(name="Probability", limits=c(0, 1), breaks=seq(.1,10,.1)) +
  geom_text(aes(label=prob, hjust=1.25)) +
  labs(title="Pembrolizumab treatment effects rankings") +
  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),
        legend.position="right", plot.title=element_text(hjust = 0.5))
i


# save plot
dev.copy(pdf, file="Plots & AUC/Rank plots/os_nc/rank_plot_os_nc_Pembro.pdf")
dev.off()
```

A.8.5.2.2 OpenBugs code

```
# Normal likelihood, identity link
# Trial-level data given as treatment differences
# Fixed effects model for multi-arm trials
#Multivariate model for 2 outcomes, capturing both between outcome and between arm
correlations


model{                          # *** PROGRAM STARTS


for(i in 1:ns2) {                 # LOOP THROUGH 2-ARM STUDIES
          P2[i,1,1]<-P[i,1,1]
          P2[i,2,2]<-P[i,2,2]
          P2[i,1,2]<- P[i,1,2]
          P2[i,2,1]<- P[i,1,2]
# Bivariate normal likelihood for treatment effects on mu and sigma for 2-arm trials.
#delta indexed delta[i,k,j] for study i, arm k, outcome j
    y[i,1:2] ~ dmnorm(delta[i,2,1:2],P2[i,1:2,1:2])


#Deviance contribution for trial i
   for (j in 1:2){  # LOOP over outcomes to perform matrix multiplication for deviance
```

```
        ydiff2[i,j]<- y[i,j] - delta[i,2,j]
        z2[i,j]<- inprod(P2[i,j,1:2], ydiff2[i,1:2])
      }
    resdev[i]<- inprod(ydiff2[i,1:2], z2[i,1:2])
 }

for(i in (ns2+1):(ns2+ns3)) {        # LOOP THROUGH THREE-ARM STUDIES
          for (j in 2:4){ # Complete the lower triangle of the Covariance matrix
                    for (k in 1:(j-1)){
                    P[i,j,k]<-P[i,k,j]
                    }
          }

# multivariate normal likelihood for 3-arm trials  with 2 outcomes
    y[i,1:4] ~ dmnorm(delta3[i,1:4],P[i,1:4,1:4])

#Deviance contribution for trial i
    for (j in 1:4){  # LOOP over outcomes/arms to perform matrix multiplication for deviance
        ydiff3[i,j]<- y[i,j] - delta3[i,j]
        z3[i,j]<- inprod(P[i,j,1:4], ydiff3[i,1:4])
      }
    resdev[i]<- inprod(ydiff3[i,1:4], z3[i,1:4])

delta3[i,1] <- delta[i,2,1]
delta3[i,2]<- delta[i,2,2]
delta3[i,3]<- delta[i,3,1]
delta3[i,4]<- delta[i,3,2]
  }

for(i in 1:(ns2+ns3)){               #   LOOP THROUGH ALL STUDIES
     for (k in 2:na[i]) {            #  LOOP THROUGH ARMS
        delta[i,k,1] <-  d[t[i,k],1] - d[t[i,1],1]        #NMA model for treatment effect on location
(mu)
```

```
      delta[i,k,2] <-  d[t[i,k],2] - d[t[i,1],2]         #NMA model for treatment effect on scale
(sigma)
    }
 }

totresdev <- sum(resdev[])      #Total Residual Deviance

for (j in 1:2){
          d[1,j]<-0       # treatment effect is zero for reference treatment
          # vague priors for treatment effects
          for (k in 2:nt){
                    d[k,j] ~ dnorm(0,.0001)
                    TR[k,j] <- exp(d[k,j]) # convert to time ratio scale
          }
}

# ranking on relative scale
for (j in 1:2){
     for (k in 1:nt) {
          rk[k,j] <- rank(d[,j],k) # assumes events are “bad”
          best[k,j] <- equals(rk[k,j],1) #calculate probability that treat k is best

                    for (h in 1:nt){
           prob[h,k,j] <- equals(rk[k,j],h) # calculates probability that treat k is h-th best
          }
          }
}

}                              # *** PROGRAM ENDS

list(ns2=9, ns3=1, nt=9)
```

```
y[,1]  y[,3]  y[,2]  y[,4]  P[,1,1]  P[,2,2]  P[,3,3]  P[,4,4]  P[,1,2]  P[,1,3]  P[,1,4]  P[,2,3]
       P[,2,4]  P[,3,4]  t[,1]  t[,2]  t[,3]  na[]  #  STUDY

0.174211599  NA  0.083688862  NA  34.44512428  69.18889783  NA  NA  -
17.13441166  NA  NA  NA  NA  NA  1  2  NA  2  #
       BREAK-3

0.237010567  NA  0.118751012  NA  33.91145757  39.15659597  NA  NA  -
15.46521305  NA  NA  NA  NA  NA  2  3  NA  2  #
       BRF113220

0.539495054  NA  -0.42058256  NA  96.00179654  233.5234029  NA  NA  -
26.71583306  NA  NA  NA  NA  NA  1  5  NA  2  #
       BRIM-3

0.91358103  NA  0.423157457  NA  31.85713769  133.415734  NA  NA  -
18.34036005  NA  NA  NA  NA  NA  1  6  NA  2  #
       CheckMate 066

0.530626253  NA  0.624837131  NA  3.395119688  20.91591087  NA  NA  -
1.904401561  NA  NA  NA  NA  NA  4  7  NA  2  #
       CheckMate 069

0.444280299  NA  0.219398872  NA  59.92072053  120.5827157  NA  NA  -
29.46186397  NA  NA  NA  NA  NA  5  8  NA  2  #
       COLUMBUS

0.325487325  NA  -0.045416189  NA  94.38127217  136.6262169  NA  NA  -
54.84588435  NA  NA  NA  NA  NA  2  3  NA  2  #
       COMBI-d

-0.377802956  NA  -0.007406443  NA  125.8292377  211.1987756  NA  NA  -
71.12682112  NA  NA  NA  NA  NA  3  5  NA  2  #
       COMBI-v

0.552114023  NA  0.172231184  NA  43.37935746  140.2109692  NA  NA  -
33.60837725  NA  NA  NA  NA  NA  4  9  NA  2  #
       KEYNOTE-006

0.592372992  0.869332508  0.185073734  0.471000563  65.33935238  245.6840198
       40.04653699  230.3821351  -53.79968047  -17.37649906  22.35620934
       17.3378917  -93.91963822  -42.21596963  4  6  7  3  #
       CheckMate 067


END

list(d=structure(.Data=c(NA,NA,0,0,0,  0,0,0,0,0,  0,0,0,0,0,  0,0,0),
.Dim=c(9,2)))


list(d=structure(.Data=c(NA,NA,-1,2,-2,  0.5,0.5,-0.5,-1,-1.5,  0.5,0.5,-0.5,-1,-1.5,
       0.5,0.5,-0.5), .Dim=c(9,2)))
```

A.8.5.2.3 R code to rank overall survival at 60 months

```
# Generalized Gamma - Ranking Survival at 60 months

#------------------------------------------------------------------------
# Instructions
#------------------------------------------------------------------------

# The following code uses coda from OpenBUGS to create a data frame
# that ranks all treatments on overall survival at 60 months. This
# data frame can than be used to produce a ranking plot.

#------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------

# Load required packages
library(survival)
library(flexsurv)
library(MASS)
library(coda)

# Start with clean environment
rm(list=ls())

# Set the working directory
setwd()

#------------------------------------------------------------------------
# Import and prep data
#------------------------------------------------------------------------

# Import IPD Data
```

```
data <- read.csv("../../../../Data/melanoma_os_ipd_nc.csv")


#Read in coda samples for treatment effects d on mu and log(sigma)

dsims <- read.coda("d_coda_nc.out", "d_coda_nc.ind")


#No treatments

nt <- 9


#No. simulations

nsims <- 10000


# Add treatment labels

trtnames <- c("DTIC", "Dab", "Dab + Tram", "Ipi", "Vem", "Nivo", "Nivo + Ipi", "Enco + Bini",
"Pembro")


#------------------------------------------------------------------------

# Generalised Gamma Fit for reference curve CheckMate 066 DTIC arm (trt 1)

#------------------------------------------------------------------------


# Fit generalised gamma model just for the DTIC arm of the CheckMate 066 trial

df <- data[data$studyCode==4 & data$txCode==1,]


gengamt1 <- flexsurvreg(formula=Surv(time,event)~1,

                data=df, dist="gengamma", method="BFGS")


gengamt1$coefficients

gengamt1$cov


#Generate 1000 random samples for the parameters of the Generalised Gamma

rcoefs<-mvrnorm(nsims,gengamt1$coefficients,gengamt1$cov)


mu<-matrix(nrow=nsims,ncol=nt)

sigma<- matrix(nrow=nsims,ncol=nt)
```

```
Q<-rcoefs[,"Q"]
mu[,1]<-rcoefs[,"mu"]
sigma[,1]<-exp(rcoefs[,"sigma"])

# Store treatment effects for each treatment
nodename<-matrix(nrow=nt, ncol=2)
for(t in 2:nt){
  for (j in 1:2){
    nodename[t,j]<-paste("d[",t,",",j,"]",sep="")
  }
mu[,t]<- mu[,1]+dsims[,nodename[t,1]]
sigma[,t]<-exp(rcoefs[,"sigma"]+dsims[,nodename[t,2]])
  }

Surv60<-matrix(nrow=nsims,ncol=nt)
for (t in 1:nt){
  Surv60[,t]<-1-pgengamma(60,mu[,t],sigma[,t],Q)
}

#Rankings based on 60month Survival. Rank 1 == highest 60m survival
rk<-nt+1- t(apply(Surv60,1,rank))
colnames(rk)<- trtnames
summary(rk)

count<-matrix(nrow=nt, ncol=nt)
for (t in 1:nt){
    count[t,]<- tabulate(rk[,t], nbins=9)
  }
prob <- as.data.frame(count/nsims)

# Write as CSV
write.csv(prob, file="../rankings_os_nc.csv")
```

A.8.6 Piecewise

A.8.6.1 R code

A.8.6.1.1 NMA and results code

```
# Piecewise NMA for Time to Event Outcomes - 2 cut points

#------------------------------------------------------------------------
# Instructions
#------------------------------------------------------------------------
# The following code runs a Piecewise NMA for time to event outcome IPD data
# using WinBUGS. The code has 2 parts. The first part runs the NMA.
# The second part uses the results from part 1 to make plots and estimate
# the area under the curve for different treatments. Each part and each step has
# more detail of what the code does at that point.

#------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------

# Load libraries
library(survival)
library(doBy)
library(R2WinBUGS)
library(reshape)
library(reshape2)
library(ggplot2)
library(coda)
library(pracma)

# Start with clean environment
rm(list=ls())
```

```
# Set working directory
setwd()

#------------------------------------------------------------------------
# Part 1: Run NMA
#------------------------------------------------------------------------

# Part 1 of the analysis runs the NMA and produces results

# Step 1 runs the anova parameterization to obtain the piecewise data
# required to run the model.

# Step 2 formats the data as the 'bugs_data' frame to be read into WinBUGS
# later using R2WinBUGS

# Step 3 sets the initial values for the NMA.

# Step 4 runs the NMA in WinBUGS using R2Winbugs. Note, to run the NMA you must have
# the WinBUGS model saved as a text file. In this case it is 'FE_model.txt'
# The code will then produce and saves results, and allows you to check the
# density and trace plots.

#------------------------------------------------------------------------
# Step 1: Run Piecewise analysis to obtain data to feed into WinBUGS
#------------------------------------------------------------------------

# Start with IPD data
data <- read.csv("../../../Data/melanoma_os_ipd_nc.csv")

# Add treatment as a factor variable
data$treatment[data$txCode==1] <- "DTIC"
```

```
data$treatment[data$txCode==2] <- "Dab"
data$treatment[data$txCode==3] <- "Dab + Tram"
data$treatment[data$txCode==4] <- "Ipi"
data$treatment[data$txCode==5] <- "Vem"
data$treatment[data$txCode==6] <- "Nivo"
data$treatment[data$txCode==7] <- "Nivo + Ipi"
data$treatment[data$txCode==8] <- "Enco + Bini"
data$treatment[data$txCode==9] <- "Pembro"
data$treatment <- as.factor(data$treatment)

# Import function for aggregating data (this calls on the anova
# function to aggregate data, this will be shared as 'Anova function')
source("../../anova_data.R")

# Select time points for aggregating data (not this is the location in the r code you
# can select different time points if you would like, but if you change the time points
# here, you will need to change the WinBUGS code as well)
timepoints=c(12, 18, 120)

# Time points including zero
timepoints2=c(0, 12, 18, 120)

# Empty data frame for aggregated data
anova <- data.frame(spgrp=NA, treatment=NA, trialid=NA, y=NA, nevents=NA,
                    natrisk=NA, y.max=NA, start=NA, time=NA, trt=NA, treatnumf=NA,
                    studynumf=NA)

# Apply function
anova <- anova_data(timepoints=timepoints, timepoints2=timepoints2, ref.study=3,
                    df=data)

# Converting aggregate melanoma csv files to wide format
```

```
anova$t[anova$treatment=="DTIC"] <- 1
anova$t[anova$treatment=="Dab"] <- 2
anova$t[anova$treatment=="Dab + Tram"] <- 3
anova$t[anova$treatment=="Ipi"] <- 4
anova$t[anova$treatment=="Vem"] <- 5
anova$t[anova$treatment=="Nivo"] <- 6
anova$t[anova$treatment=="Nivo + Ipi"] <- 7
anova$t[anova$treatment=="Enco + Bini"] <- 8
anova$t[anova$treatment=="Pembro"] <- 9

anova$E<-anova$y
df<-subset(anova, select=-c(treatment,y,natrisk,y.max,start,time))
df$arm<- df$na <- rep(NA,nrow(df))
for (i in 1:nrow(df)){
  df_sub<-subset(df,trialid==trialid[i]&spgrp==spgrp[i])
  df$arm[i]<-rank(df_sub$t)[df_sub$t==df$t[i]]
  df$na[i]<- nrow(df_sub)
}

piecewise<-reshape(df,timevar="arm", idvar=c("spgrp","trialid","na"), direction="wide" )

#### OPTIONAL: Can save piecewise as a CSV file ####
# write.csv(piecewise, "Aggregate data/piecewise_os_nc.csv")

#------------------------------------------------------------------------
# Step 2: Prep data to load in to WinBUGS as bugs_data
#------------------------------------------------------------------------

#### OPTIONAL: If you already have piecewise data (obtained from step 1) ####
#### You can skip step 1 and load piecewise here ####
# piecewise <- read.csv("Aggregate data/piecewise_os_nc.csv")
```

```
# Set the location for WinBUGS
bugs.directory <- "C:/Program Files (x86)/WinBUGS14"

# WinBUGS burn-in & simulation size
num.sims <- 20000
burn.in <- 10000

# Number of studies
ns <- max(piecewise$trialid)

# Number of intervals
nint <- max(piecewise$spgrp)

# Number of treatments
ntrt <- max(piecewise$t.2)

# N - Number of rows of data (number of studies * number of intervals)
N <- ns*nint

# Set the reference curve (in this case CM66 - note if you want to change the reference curve,
# you must do so here and in the WinBUGS code)
CM66 <- 4

# Create arrays to load into bugs_data frame
nevents <- array(c(piecewise$nevents.1, piecewise$nevents.2, piecewise$nevents.3), dim=c(N,3))
E <- array(c(piecewise$E.1, piecewise$E.2, piecewise$E.3), dim=c(N,3))
t <- array(c(piecewise$t.1, piecewise$t.2, piecewise$t.3), dim=c(N,3))

# Create bugs_data to load into WinBUGS
bugs_data <- list(nint=nint, ntrt=ntrt, N=N, CM66=CM66, nevents=nevents, E=E, t=t,
                  spgrp=piecewise$spgrp, trialid=piecewise$trialid, na=piecewise$na)
```

```
#---------------------------------------------------------------------------
# Step 3: Set Initial values
#---------------------------------------------------------------------------

d1 <- array(c(NA, rep(0.1, ntrt-1), NA, rep(0.2, ntrt-1)), dim=c(ntrt, nint))
d2 <- array(c(NA, rep(0.2, ntrt-1), NA, rep(-0.1, ntrt-1)), dim=c(ntrt, nint))
d3 <- array(c(NA, rep(-0.1, ntrt-1), NA, rep(0.1, ntrt-1)), dim=c(ntrt, nint))

mu1 <- array(rep(c(0.4, 0.5, 0.6, 0.2, 0.3, 0.1, 0.1),N), dim=c(ns, nint))
mu2 <- array(rep(c(0.3, 0.4, 0.5, 0.6, 0.7, -0.1, -0.2),N), dim=c(ns, nint))
mu3 <- array(rep(c(0.5, 0.6, 0.7, 0.1, -0.1,  0.2, 0.2),N), dim=c(ns, nint))

fe_inits <- list(list(d=d1, mu=mu1),
                 list(d=d2, mu=mu2),
                 list(d=d3, mu=mu3))

#---------------------------------------------------------------------------
# Step 4: Fit FE model in WinBUGS
#---------------------------------------------------------------------------

# Run NMA in WinBUGS
bugs.fe <- bugs(data=bugs_data, inits=fe_inits,
                parameters.to.save=c("d", "mu", "S", "Cum_H", "rk60", "totresdev"),
                model.file="FE_model.txt", clearWD=F,
                summary.only=FALSE, n.iter=(num.sims+burn.in),
                n.sims=num.sims, n.burnin=burn.in, n.chains=3,
                bugs.seed=212034, bugs.directory=bugs.directory,
                debug=TRUE, DIC=TRUE)

# Save results in a data frame
fe_results <- as.data.frame(bugs.fe$summary)
```

```
# Save results in csv file
write.csv(fe_results, "NMA results/fe_results_os_nc.csv")

# Check results
results2 <- bugs.fe$sims.matrix[,grep("d",rownames(bugs.fe$summary))]
results2 <- cbind(rep(0,dim(results2)[1]),results2)
summary(results2)

results_mcmc<-mcmc(results2)
par(mfrow=c(3,2))

# Check autocorrelation
autocorr.plot(results_mcmc[,2:9])

# Check trace for convergence
traceplot(results_mcmc[,2])
traceplot(results_mcmc[,3])
traceplot(results_mcmc[,4])
traceplot(results_mcmc[,5])
traceplot(results_mcmc[,6])
traceplot(results_mcmc[,7])
traceplot(results_mcmc[,8])
traceplot(results_mcmc[,9])

# Histograms of posterior distributions
densplot(results_mcmc[,2])
densplot(results_mcmc[,3])
densplot(results_mcmc[,4])
densplot(results_mcmc[,5])
densplot(results_mcmc[,6])
densplot(results_mcmc[,7])
```

```
densplot(results_mcmc[,8])
densplot(results_mcmc[,9])


#------------------------------------------------------------------------
# Part 2: Results - Survival plot, Area under curve, Rank plots
#------------------------------------------------------------------------

# Part two of the analysis is only dealing with the results obtained from part 1

# Step 5 generates a plot with survival curves for each treatment

# Step 6 calculate the area under the curves for each treatment

# Step 7 generates a ranking plot, showing the liklihood of each treatment occupying each rank

#### OPTIONAL: If you already have result data, obtained from Part 1 ####
#### You can skip all previous steps and load result data here ####
# fe_results <- read.csv("NMA results/fe_results_os_nc.csv")

#------------------------------------------------------------------------
# Step 5: Plot survival curves for each treatment
#------------------------------------------------------------------------

# Create graph data frame, this has the predicted survival by treatment from
# the NMA
graph_data <- data.frame(time=c(0:120),
                         trt1=c(1, fe_results$mean[55:174]),
                         trt2=c(1, fe_results$mean[175:294]),
                         trt3=c(1, fe_results$mean[295:414]),
                         trt4=c(1, fe_results$mean[415:534]),
                         trt5=c(1, fe_results$mean[535:654]),
                         trt6=c(1, fe_results$mean[655:774]),
```

```
                trt7=c(1, fe_results$mean[775:894]),
                trt8=c(1, fe_results$mean[895:1014]),
                trt9=c(1, fe_results$mean[1015:1134]))

# Using colours from the Safe palette from rcartocolor
colors=c("#88CCEE", "#CC6677", "#DDCC77", "#117733", "#332288", "#AA4499",
        "#44AA99", "#999933", "#882255", "#661100", "#6699CC", "#888888")

# Kaplan-Meier data for DTIC arm from CheckMate 066
ipd_data <- data[data$studyCode==4 & data$txCode==1,]
KM.est<-survfit(Surv(time,event)~1, data=ipd_data, type="kaplan-meier", conf.int=FALSE)

# Start by plotting the Kaplan_Meier DTIC curve
plot(KM.est,xlab="Time (months)",ylab="Overall Survival",xaxt="n",yaxt="n",main="
",xlim=c(0,120),ylim=c(0,1),
    mark.time=FALSE, col=color[10], conf.int=F)
#Add y axis (2 specifies that axis goes on the left of the plot)
axis(2, at=c(0, 0.2, 0.4, 0.6, 0.8, 1))
#Add x axis (1 specified that axis goes at the bottom of the plot)
axis(1, at=c(0, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120))

# Add prediction lines
lines(graph_data$time, graph_data$trt1, col=color[1])
lines(graph_data$time, graph_data$trt2, col=color[2])
lines(graph_data$time, graph_data$trt3, col=color[3])
lines(graph_data$time, graph_data$trt4, col=color[4])
lines(graph_data$time, graph_data$trt5, col=color[5])
lines(graph_data$time, graph_data$trt6, col=color[6])
lines(graph_data$time, graph_data$trt7, col=color[7])
lines(graph_data$time, graph_data$trt8, col=color[8])
lines(graph_data$time, graph_data$trt9, col=color[9])

# Add legend
```

```
legend("topright",
       c("KM", "DTIC", "Dab", "Dab + Tram", "Ipi", "Vem",
         "Nivo", "Nivo + Ipi", "Enco + Bini", "Pembro"),
       col=c(color[10], color[1], color[2], color[3], color[4],
             color[5], color[6], color[7], color[8], color[9]),
       lty=c(1,1,1,1), ncol=3, text.width=20, box.lty=0, y.intersp = 2)

# save plot
dev.copy(pdf, "Plots & AUC/survival_plot_os_nc.pdf")
dev.off()

#-------------------------------------------------------------------------
# Step 6: Calculate area under curve for different treatments
#-------------------------------------------------------------------------

# Create data frame to store AUC results
auc <- data.frame(trt=c(1:9), auc=NA)

# Calculate AUC
auc$auc[1] <- trapz(graph_data$time, graph_data$trt1)
auc$auc[2] <- trapz(graph_data$time, graph_data$trt2)
auc$auc[3] <- trapz(graph_data$time, graph_data$trt3)
auc$auc[4] <- trapz(graph_data$time, graph_data$trt4)
auc$auc[5] <- trapz(graph_data$time, graph_data$trt5)
auc$auc[6] <- trapz(graph_data$time, graph_data$trt6)
auc$auc[7] <- trapz(graph_data$time, graph_data$trt7)
auc$auc[8] <- trapz(graph_data$time, graph_data$trt8)
auc$auc[9] <- trapz(graph_data$time, graph_data$trt9)

# Save area under curves results as CSV file
write.csv(auc, file="Plots & AUC/auc_os_nc.csv")
```

```
#-------------------------------------------------------------------------

# Step 7: Rank plot

#-------------------------------------------------------------------------


# Keep rows for ranking only

rankdata <- fe_results[2215:2295,]


# Variable for rank

rankdata$rank_code <- rep(9:1, 9)


# Restrict probability to 2 decimal places

rankdata$prob <- round(rankdata$mean, 2)


# Add a treatment label

rankdata$Treatment <- c(rep("DTIC", 9), rep("Dab", 9), rep("Dab + Tram", 9), rep("Ipi", 9),
rep("Vem", 9),

                        rep("Nivo", 9), rep("Nivo + Ipi", 9), rep("Enco + Bini", 9), rep("Pembro", 9))


# Rename mean column

names(rankdata)[names(rankdata)=="mean"] <- "Probability"


# Plot with text

q <- ggplot(rankdata, aes(x=rank_code, y=Treatment)) +

  geom_point(aes(size=Probability), shape=21, colour="skyblue", fill="skyblue") +

  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),

        legend.position="bottom") +

  scale_size_area(max_size=10) +

  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +

  scale_y_discrete(name="Treatment") +

  geom_text(aes(label=prob))

q
```

```
# save plot
dev.copy(pdf, "Plots & AUC/rank_plot_os_nc.pdf")
dev.off()
```

A.8.6.1.2 Anova Function

```
# Create function that takes a generated dataset and formats the data ready to apply the anova
# parameterisation
anova_data <- function(timepoints, timepoints2, ref.study=1, df){

  # Split the data at timepoints
  df2 <- survSplit(Surv(time, event) ~., data=df,
                   cut=timepoints, episode ="timegroup")

  # Calculate offset
  df2$y <- df2$time - df2$tstart

  # Add a variable that equals one for all patients - this is so the number at risk
  # can be calculated when we collapse the data
  df2$n <- 1

  # Collapse data
  df3 <- summaryBy(y + event + n ~ timegroup + treatment + studyCode, FUN=c(sum, max), data=df2)
  df3 <- subset(df3, select=-c(event.max, n.max))
  names(df3) <- c("spgrp", "treatment", "trialid", "y", "nevents", "natrisk", "y.max")

  # Add in a start time variable
  df3$start <- NA
  for(i in unique(df3$spgrp)){
    df3$start[df3$spgrp==i] <- timepoints2[i]
  }
```

```
  # Add in a time variable (i.e. how long since time 0 to max value of y for each row)
  df3$time <- df3$start + df3$y.max

  # Return the formatted dataset
  return(df3)

}
```

A.8.6.2 WinBUGS code

```
# Fixed treatment effect model: Piecewise Constant Hazards Model

model{

  for (i in 1:N){ #Loop over studies and time-periods
          mu[trialid[i],spgrp[i]]~dnorm(0,.0001) #Priors for log-hazard on control arm indexed by study and time-period
          for (k in 1:na[i]){                                    #Loop over study arms
                  nevents[i,k]~dpois(theta[i,k])  #Poisson likelihood for number of events
                  theta[i,k]<- lambda[i,k]*E[i,k]          #Event rate is the mean event rate multiplied by total exposure time at risk, E
                  log(lambda[i,k])<-  mu[trialid[i],spgrp[i]] + d[t[i,k],spgrp[i]] - d[t[i,1],spgrp[i]] # NMA model for log-hazards for time period.
        dev[i,k] <- 2*((theta[i,k]-nevents[i,k]) + nevents[i,k]*log(nevents[i,k]/theta[i,k])) #Deviance contribution
          }
#  summed residual deviance contribution for this trial
    resdev[i] <- sum(dev[i,1:na[i]])
  }
totresdev <- sum(resdev[])          #Total Residual Deviance

# PRIORS
 for (s in 1:nint){
```

```
        d[1,s]<- 0
        for (k in 2:ntrt){
                d[k,s]~dnorm(0,.0001)
        }
}

# Calculate survival using CheckMate 066 as baseline
for (s in 1:nint){
log(hazard[1,s])<- mu[CM66,s]                    #DTIC taken from CheckMate 066 study, trialid CM66
        for (k in 2:ntrt){
                log(hazard[k,s])<- mu[CM66,s] + d[k,s]
        }
}

# Months 1-12
for(k in 1:ntrt){
 for(m in 1:12) {
     Cum_H[k,m] <- m*hazard[k, 1] # Cumulative hazard over time by treatment
     TT[k,m] <- 1 - exp(-Cum_H[k,m]) # mortality over time by treatment
     S[k,m] <- 1 - TT[k,m] # Survival over time by treatment
  }
}

# Months 13-18
for(k in 1:ntrt){
 for(m in 13:18) {
     Cum_H[k,m] <- Cum_H[k,12] + ((m-12)*hazard[k,2]) # Cumulative hazard over time by treatment
     TT[k,m] <- 1 - exp(-Cum_H[k,m]) # mortality over time by treatment
     S[k,m] <- 1 - TT[k,m] # Survival over time by treatment
  }
}
```

```
# Months 19-120
for(k in 1:ntrt){
        for(m in 19:120) {
    Cum_H[k,m] <- Cum_H[k,18] + ((m-18)*hazard[k,3]) # Cumulative hazard over time by treatment
    TT[k,m] <- 1 - exp(-Cum_H[k,m]) # mortality over time by treatment
    S[k,m] <- 1 - TT[k,m] # Survival over time by treatment
  }
}

# Rank treatments at 60 months
for (k in 1:ntrt) {
   for (r in 1:ntrt) {
     rk60[k,r] <- equals(ranked(S[,60],r),S[k,60])
   }
  }

}                              # *** PROGRAM ENDS
```

A.8.7 Fractional polynomial code

A.8.7.1 R code

A.8.7.1.1 Anova parameterization and Roche code to fit FP models to compare AIC values

```
# Fractional Polynomial model - 1st Order fixed effect
# Use Roche code to run fractional polynomial NMA for melanoma network

#-------------------------------------------------------------------------
# Instructions
#-------------------------------------------------------------------------
# The following code uses IPD data to create aggregate data to run a FP NMA in
# Winbugs. This code has two parts. The first part runs the anova
```

```
# parameterization to obtain the aggregate data required to run the NMA.
# The second part uses Roche code to fit multiple FP models to compare AIC values.

#------------------------------------------------------------------------
# Load required packages and clean environment
#------------------------------------------------------------------------

# Start with a clean environment
rm(list = ls())

# Load libraries
library(survival)
library(doBy)

# Set the working directory
setwd()

#------------------------------------------------------------------------
# Part 1: Run ANOVA on IPD data to obtain aggregate data to feed into WinBUGS
#------------------------------------------------------------------------

# Load IPD data
data <- read.csv("../Data/melanoma_os_ipd_nc.csv")

# Add treatment as a factor variable
data$treatment[data$txCode==1] <- "DTIC"
data$treatment[data$txCode==2] <- "Dab"
data$treatment[data$txCode==3] <- "Dab + Tram"
data$treatment[data$txCode==4] <- "Ipi"
data$treatment[data$txCode==5] <- "Vem"
data$treatment[data$txCode==6] <- "Nivo"
data$treatment[data$txCode==7] <- "Nivo + Ipi"
```

```
data$treatment[data$txCode==8] <- "Enco + Bini"
data$treatment[data$txCode==9] <- "Pembro"
data$treatment <- as.factor(data$treatment)

# Select source for anova function
source("anova_data.R")

# Select time points for aggregating data
timepoints=c(6, 12, 18, 24, 30, 36, 42, 120)

# Time points including zero
timepoints2=c(0, 6, 12, 18, 24, 30, 36, 42, 120)

# Apply function
anova <- data.frame(spgrp=NA, treatment=NA, trialid=NA, y=NA, nevents=NA,
                    natrisk=NA, y.max=NA, start=NA, time=NA, trt=NA, treatnumf=NA,
                    studynumf=NA)

anova <- anova_data(timepoints=timepoints, timepoints2=timepoints2, ref.study=4,
                    df=data)

# Add trt column
anova$trt <- anova$treatment

# Add treatment number as a variable
anova$treatnumf[anova$treatment=="DTIC"] <- 1
anova$treatnumf[anova$treatment=="Dab"] <- 2
anova$treatnumf[anova$treatment=="Dab + Tram"] <- 3
anova$treatnumf[anova$treatment=="Ipi"] <- 4
anova$treatnumf[anova$treatment=="Vem"] <- 5
anova$treatnumf[anova$treatment=="Nivo"] <- 6
anova$treatnumf[anova$treatment=="Nivo + Ipi"] <- 7
```

```
anova$treatnumf[anova$treatment=="Enco + Bini"] <- 8
anova$treatnumf[anova$treatment=="Pembro"] <- 9
anova$treatnumf <- as.integer(anova$treatnumf)

# Add study number as a variable
anova$studynumf[anova$trialid==1] <- 1
anova$studynumf[anova$trialid==2] <- 2
anova$studynumf[anova$trialid==3] <- 3
anova$studynumf[anova$trialid==4] <- 4
anova$studynumf[anova$trialid==5] <- 5
anova$studynumf[anova$trialid==6] <- 6
anova$studynumf[anova$trialid==7] <- 7
anova$studynumf[anova$trialid==8] <- 8
anova$studynumf[anova$trialid==9] <- 9
anova$studynumf[anova$trialid==10] <- 10
anova$studynumf <- as.integer(anova$studynumf)

#### OPTIONAL: Can save piecewise as a CSV file ####
write.csv(anova, "FP 8/aggregate_data_os_nc.csv")

#------------------------------------------------------------------------
# Step 2: Roche code to fit several FP models and to compare the AIC values
#------------------------------------------------------------------------

# Save data as km for use in Roche code below
km <- anova

#list of models to be fitted - 1st order FP models
models <- list(
  "Exponential" = list(b1=function(x){0},b2=function(x){0}),
  "Weibull,p1=0" = list(b1=function(x){log(x)},b2=function(x){0}),
  "p1=0.5" = list(b1=function(x){x^0.5},b2=function(x){0}),
```

```
"Gompertz,p1=1" = list(b1=function(x){x},b2=function(x){0}),
"p1=2" = list(b1=function(x){x^2},b2=function(x){0}),
"p1=3" = list(b1=function(x){x^3},b2=function(x){0}),
"p1=-0.5" = list(b1=function(x){x^-0.5},b2=function(x){0}),
"p1=-1" = list(b1=function(x){x^-1},b2=function(x){0}),
"p1=-2" = list(b1=function(x){x^-2},b2=function(x){0}),
"Second order, p1=3, p2=3" = list(b1=function(x){x^3},b2=function(x){x^3*log(x)}),
"Second order, p1=3, p2=2" = list(b1=function(x){x^3},b2=function(x){x^2}),
"Second order, p1=3, p2=1" = list(b1=function(x){x^3},b2=function(x){x}),
"Second order, p1=3, p2=0.5" = list(b1=function(x){x^3},b2=function(x){x^-0.5}),
"Second order, p1=3, p2=0" = list(b1=function(x){x^3},b2=function(x){log(x)}),
"Second order, p1=3, p2=-0.5" = list(b1=function(x){x^3},b2=function(x){x^-0.5}),
"Second order, p1=3, p2=-1" = list(b1=function(x){x^3},b2=function(x){x^-1}),
"Second order, p1=3, p2=-2" = list(b1=function(x){x^3},b2=function(x){x^-2}),
"Second order, p1=2, p2=2" = list(b1=function(x){x^2},b2=function(x){x^2*log(x)}),
"Second order, p1=2, p2=1" = list(b1=function(x){x^2},b2=function(x){x}),
"Second order, p1=2, p2=0.5" = list(b1=function(x){x^2},b2=function(x){x^0.5}),
"Second order, p1=2, p2=0" = list(b1=function(x){x^2},b2=function(x){log(x)}),
"Second order, p1=2, p2=-0.5" = list(b1=function(x){x^2},b2=function(x){x^-0.5}),
"Second order, p1=2, p2=-1" = list(b1=function(x){x^2},b2=function(x){x^-1}),
"Second order, p1=2, p2=-2" = list(b1=function(x){x^2},b2=function(x){x^-2}),
"Second order, p1=1, p2=1" = list(b1=function(x){x},b2=function(x){x*log(x)}),
"Second order, p1=1, p2=0.5" = list(b1=function(x){x},b2=function(x){x^0.5}),
"Second order, p1=1, p2=0" = list(b1=function(x){x},b2=function(x){log(x)}),
"Second order, p1=1, p2=-0.5" = list(b1=function(x){x},b2=function(x){x^-0.5}),
"Second order, p1=1, p2=-1" = list(b1=function(x){x},b2=function(x){x^-1}),
"Second order, p1=1, p2=-2" = list(b1=function(x){x},b2=function(x){x^-2}),
"Second order, p1=0.5, p2=0.5" = list(b1=function(x){x^0.5},b2=function(x){x^0.5*log(x)}),
"Second order, p1=0.5, p2=0" = list(b1=function(x){x^0.5},b2=function(x){log(x)}),
"Second order, p1=0.5, p2=-0.5" = list(b1=function(x){x^0.5},b2=function(x){x^-0.5}),
"Second order, p1=0.5, p2=-1" = list(b1=function(x){x^0.5},b2=function(x){x^-1}),
"Second order, p1=0.5, p2=-2" = list(b1=function(x){x^0.5},b2=function(x){x^-2}),
```

```
  "Second order, p1=0, p2=0" = list(b1=function(x){log(x)},b2=function(x){log(x)*log(x)}),
  "Second order, p1=0, p2=-0.5" = list(b1=function(x){log(x)},b2=function(x){x^-0.5}),
  "Second order, p1=0, p2=-1" = list(b1=function(x){log(x)},b2=function(x){x^-1}),
  "Second order, p1=0, p2=-2" = list(b1=function(x){log(x)},b2=function(x){x^-2}),
  "Second order, p1=-0.5, p2=-0.5" = list(b1=function(x){x^-0.5},b2=function(x){x^-
0.5*log(x)}),
  "Second order, p1=-0.5, p2=-1" = list(b1=function(x){x^-0.5},b2=function(x){x^-1}),
  "Second order, p1=-0.5, p2=-2" = list(b1=function(x){x^-0.5},b2=function(x){x^-2}),
  "Second order, p1=-1, p2=-1" = list(b1=function(x){x^-1},b2=function(x){x^-1*log(x)}),
  "Second order, p1=-1, p2=-2" = list(b1=function(x){x^-1},b2=function(x){x^-2}),
  "Second order, p1=-2, p2=-2" = list(b1=function(x){x^-2},b2=function(x){x^-2*log(x)})
)

#Fit all models
fit.KM.NMA<-function(bf){
  km.new=km
  km.new$beta1=bf[[1]](km.new$time)
  km.new$beta2=bf[[2]](km.new$time)
  #model formula
  f=cbind(nevents,natrisk-
nevents)~treatnumf+studynumf+treatnumf*beta1+treatnumf*beta2+studynumf*beta1+stud
ynumf*beta2
  glm(f,family=binomial(link=cloglog),data=km.new)
}
fits=lapply(models,fit.KM.NMA)

#Get AIC from each model
aics=lapply(fits,AIC)

#Print the AICs
data.frame(AIC=round(unlist(aics),2))

# Sort AIC into ascending order
a <- data.frame(AIC=round(unlist(aics),2))
```

```
sort(a[,1])
```

A.8.7.1.2 Anova function

```
# Create function that takes a generated dataset and formats the data ready to apply the anova

# parameterisation


anova_data <- function(timepoints, timepoints2, ref.study=1, df){

# Split the data at timepoints

  df2 <- survSplit(Surv(time, event) ~., data=df,

                   cut=timepoints, episode ="timegroup")


  # Calculate offset

  df2$y <- df2$time - df2$tstart


  # Add a variable that equals one for all patients - this is so the number at risk

  # can be calculated when we collapse the data

  df2$n <- 1


  # Collapse data

  df3 <- summaryBy(y + event + n ~ timegroup + treatment + studyCode, FUN=c(sum, max), data=df2)

  df3 <- subset(df3, select=-c(event.max, n.max))

  names(df3) <- c("spgrp", "treatment", "trialid", "y", "nevents", "natrisk", "y.max")


  # Add in a start time variable

  df3$start <- NA

  for(i in unique(df3$spgrp)){

    df3$start[df3$spgrp==i] <- timepoints2[i]

  }


  # Add in a time variable (i.e. how long since time 0 to max value of y for each row)
```

```
  df3$time <- df3$start + df3$y.max

  # Return the formatted dataset
  return(df3)

}
```

A.8.7.1.3 Fit NMA and use results to make plots

```
# Fractional Polynomial model - 1st Order fixed effect

#---------------------------------------------------------------------------
# Instructions
#---------------------------------------------------------------------------

# The following code runs a fractional polynomial network meta-analysis (NMA)
# for time to event outcome IPD data using winbugs. The code has 2 parts.
# The first part runs the NMA. The second part uses the results from part 1
# to make plots and estimate the area under the curve for different treatments.
# Each part and each step has more detail of what the code does at that point.

#---------------------------------------------------------------------------
# Load required packages and clean environment
#---------------------------------------------------------------------------

# Start with a clean environment
rm(list = ls())

# Load libraries
library(coda)
library(survival)
library(R2WinBUGS)
```

```
library(pracma)
library(reshape2)
library(ggplot2)

# Set the working directory
setwd()

#--------------------------------------------------------------------------
# Part 1: Run NMA
#--------------------------------------------------------------------------

# Part 1 of the analysis runs the NMA and produces results

# Step 1 Formats aggregate data previously prepared as the 'bugs_data'
# frame to be read into WinBUGS

# Step 2 sets the initial values for the NMA.

# Step 3 runs the NMA in WinBUGS using R2Winbugs. Note, to run the NMA you must have
# the winbugs model saved as a text file. In this case it is 'FE_1st_order_model_nc.txt'
# The code will then produce and saves results, and allows you to check the
# density and trace plots.

#--------------------------------------------------------------------------
# Step 1: Prep data to load in to WINBUGS as bugs_data
#--------------------------------------------------------------------------

anova <- read.csv("FP 8/aggregate_data_os_nc.csv")

# Create a treatment code variable as an integer
anova$txCode[anova$treatment=="DTIC"] <- 1
anova$txCode[anova$treatment=="Dab"] <- 2
```

```
anova$txCode[anova$treatment=="Dab + Tram"] <- 3
anova$txCode[anova$treatment=="Ipi"] <- 4
anova$txCode[anova$treatment=="Vem"] <- 5
anova$txCode[anova$treatment=="Nivo"] <- 6
anova$txCode[anova$treatment=="Nivo + Ipi"] <- 7
anova$txCode[anova$treatment=="Enco + Bini"] <- 9
anova$txCode[anova$treatment=="Pembro"] <- 10

# Order data
anova <- anova[order(anova$trialid, anova$txCode, anova$spgrp),]

# Need to number the treatment arms within each trial
anova$arm[anova$treatment=="DTIC" | anova$treatment=="Ipi"] <- 1
anova$arm[anova$treatment=="Nivo" | anova$treatment=="Pembro"] <- 2
anova$arm[anova$treatment=="Enco + Bini"] <- 2
anova$arm[anova$trialid==5 & anova$treatment=="Nivo + Ipi"] <- 3
anova$arm[anova$trialid==2 & anova$treatment=="Dab"] <- 1
anova$arm[anova$trialid==8 & anova$treatment=="Dab"] <- 1
anova$arm[anova$trialid==1 & anova$treatment=="Dab"] <- 2
anova$arm[anova$trialid==2 & anova$treatment=="Dab + Tram"] <- 2
anova$arm[anova$trialid==8 & anova$treatment=="Dab + Tram"] <- 2
anova$arm[anova$trialid==9 & anova$treatment=="Dab + Tram"] <- 1
anova$arm[anova$trialid==3 & anova$treatment=="Vem"] <- 2
anova$arm[anova$trialid==9 & anova$treatment=="Vem"] <- 2
anova$arm[anova$trialid==7 & anova$treatment=="Vem"] <- 1
anova$arm[anova$trialid==6 & anova$treatment=="Nivo + Ipi"] <- 2

# Check all arms coded
anova$arm==1 | anova$arm==2 | anova$arm==3

# Set the location for WinBUGS
bugs.directory <- "C:/Program Files (x86)/WinBUGS14"
```

```
# WinBUGS burn-in & simulation size
num.sims <- 30000
burn.in <- 30000

# Fractional polynomial powers (not this is where you can change
# which power you are using in the model)
P1 <- -2

# Length of time intervals
anova$length <- anova$time-anova$start

# Number of treatments
nt <- length(unique(anova$treatment))

# Number of studies
ns <- length(unique(anova$trialid))

# Number of rows in dataset
N <- nrow(anova)

# Maximum time
maxt <- 120

# Mean & precision
mean <- c(0,0)
prec <- array(c(0.0001, 0, 0, 0.0001), dim=c(2,2))

# Number of treatment arms for each trial
na <- c(2, 2, 2, 2, 3, 2, 2, 2, 2, 2)

# Treatment in each trial arm - This fills the down the columns first
```

```
t <- array(data=c(1, 2, 1, 1, 4, 4, 5, 2, 3, 4,
                  2, 3, 5, 6, 6, 7, 8, 3, 5, 9,
                  NA, NA, NA, NA, 7, NA, NA, NA, NA, NA), dim=c(10, 3))


# Create bugs_data to load into WinBUGS
bugs_data <- list(s=anova$trialid, r=anova$nevents, z=anova$natrisk, a=anova$arm,
time=anova$time,
                  dt=anova$length, P1=P1, N=N, nt=nt, ns=ns, maxt=maxt, mean=mean,
prec=prec,
                  t=t,  na=na)


#---------------------------------------------------------------------------
# Step 2: Set Initial values
#---------------------------------------------------------------------------


# Create initial values for model
d1 <- array(c(NA, rep(0.1, nt-1), NA, rep(0.2, nt-1)), dim=c(nt,2))
d2 <- array(c(NA, rep(0.2, nt-1), NA, rep(-0.1, nt-1)), dim=c(nt,2))


mu1 <- array(rep(c(0.4, 0.5, 0.6, 0.2, 0.3, 0.1, 0.1),4), dim=c(ns,2))
mu2 <- array(rep(c(0.3, 0.4, 0.5, 0.6, 0.7, -0.1, -0.2),4), dim=c(ns,2))


inits <- list(list(d=d1, mu=mu1),
              list(d=d2, mu=mu2))


# Optional: This saves the bugs_data and inits as text files you can then
# copy and paste into WinBUGS. With some of the larger datasets for
# the fractional polynomial models, this was done as running R2WinBUGS
# would crash due to the large amount of data.
# bugs.data(bugs_data)
# bugs.data(list(d=d1, mu=mu1), data.file="inits1.txt")
# bugs.data(list(d=d2, mu=mu2), data.file="inits2.txt")
```

```
#------------------------------------------------------------------------
# Step 3: Fit FE model in WinBUGS
#------------------------------------------------------------------------

# Run NMA in WinBUGS
bugs.object <- bugs(data=bugs_data, inits=inits,
                    parameters.to.save=c("d", "S", "rk60"),
                    model.file="FE_1st_order_model_nc.txt", clearWD=F,
                    summary.only=FALSE, n.iter=(num.sims+burn.in),
                    n.sims=num.sims, n.burnin=burn.in, n.chains=2,
                    bugs.seed=212034, bugs.directory=bugs.directory,
                    debug=F, DIC=TRUE)

# Save results in a data frame
results <- as.data.frame(bugs.object$summary)

# Save results in csv file
write.csv(results,file="FP 8/negtwo/NMA results/results_os_nc.csv")

# Check results
results2 <- bugs.object$sims.matrix[,grep("d",rownames(bugs.object$summary))]
results2 <- cbind(rep(0,dim(results2)[1]),results2)
summary(results2)

results_mcmc<-mcmc(results2)
par(mfrow=c(3,3))

#Check autocorrelation
autocorr.plot(results_mcmc[,2:18])

# Check trace for convergence
traceplot(results_mcmc[,2])
```

```
traceplot(results_mcmc[,3])
traceplot(results_mcmc[,4])
traceplot(results_mcmc[,5])
traceplot(results_mcmc[,6])
traceplot(results_mcmc[,7])
traceplot(results_mcmc[,8])
traceplot(results_mcmc[,9])
traceplot(results_mcmc[,10])
traceplot(results_mcmc[,11])
traceplot(results_mcmc[,12])
traceplot(results_mcmc[,13])
traceplot(results_mcmc[,14])
traceplot(results_mcmc[,15])
traceplot(results_mcmc[,16])
traceplot(results_mcmc[,17])
traceplot(results_mcmc[,18])

# Histograms of posterior distributions
densplot(results_mcmc[,2])
densplot(results_mcmc[,3])
densplot(results_mcmc[,4])
densplot(results_mcmc[,5])
densplot(results_mcmc[,6])
densplot(results_mcmc[,7])
densplot(results_mcmc[,8])
densplot(results_mcmc[,9])
densplot(results_mcmc[,10])
densplot(results_mcmc[,11])
densplot(results_mcmc[,12])
densplot(results_mcmc[,13])
densplot(results_mcmc[,14])
densplot(results_mcmc[,15])
```

```
densplot(results_mcmc[,16])
densplot(results_mcmc[,17])
densplot(results_mcmc[,18])

#------------------------------------------------------------------------
# Part 2: Results - Survival plot, Area under curve, Rank plots
#------------------------------------------------------------------------

# Part two of the analysis is only dealing with the results obtained from part 1

# Step 4 generates a plot with survival curves for each treatment

# Step 5 calculate the area under the curves for each treatment

# Step 6 generates a ranking plot, showing the liklihood of each treatment occupying each
rank

#### OPTIONAL: If you already have result data, obtained from Part 1 ####
#### You can skip all previous steps and load result data here ####
# results <- read.csv("FP 8/negtwo/NMA results/results_os_nc.csv")

#------------------------------------------------------------------------
# Step 4: Plot survival curves for each treatment
#------------------------------------------------------------------------

# Create graph data frame, this has the predicted survival by treatment from
# the NMA
graph_data <- data.frame(time=c(0:120),
                         trt1=c(1, results$mean[17:136]),
                         trt2=c(1, results$mean[137:256]),
                         trt3=c(1, results$mean[257:376]),
                         trt4=c(1, results$mean[377:496]),
                         trt5=c(1, results$mean[497:616]),
```

```
                trt6=c(1, results$mean[617:736]),
                trt7=c(1, results$mean[737:856]),
                trt8=c(1, results$mean[857:976]),
                trt9=c(1, results$mean[977:1096]))

# Using colours from the Safe palette from rcartocolor
colors=c("#88CCEE", "#CC6677", "#DDCC77", "#117733", "#332288", "#AA4499",
        "#44AA99", "#999933", "#882255", "#661100", "#6699CC", "#888888")

# Kaplan-Meier data for DTIC arm from CheckMate 066
data <- read.csv("../Data/melanoma_os_ipd_nc.csv")
ipd_data <- data[data$studyCode==4 & data$txCode==1,]
KM.est<-survfit(Surv(time,event)~1, data=ipd_data, type="kaplan-meier", conf.int=FALSE)

# Start by plotting the Kaplan_Meier DTIC curve
plot(KM.est,xlab="Time (months)",ylab="Overall Survival",xaxt="n",yaxt="n",main="
",xlim=c(0,120),ylim=c(0,1),
    mark.time=FALSE, col=color[10], conf.int=F)
#Add y axis (2 specifies that axis goes on the left of the plot)
axis(2, at=c(0, 0.2, 0.4, 0.6, 0.8, 1))
#Add x axis (1 specified that axis goes at the bottom of the plot)
axis(1, at=c(0, 12, 24, 36, 48, 60, 72, 84, 96, 108, 120))

# Add prediction lines
lines(graph_data$time, graph_data$trt1, col=color[1])
lines(graph_data$time, graph_data$trt2, col=color[2])
lines(graph_data$time, graph_data$trt3, col=color[3])
lines(graph_data$time, graph_data$trt4, col=color[4])
lines(graph_data$time, graph_data$trt5, col=color[5])
lines(graph_data$time, graph_data$trt6, col=color[6])
lines(graph_data$time, graph_data$trt7, col=color[7])
lines(graph_data$time, graph_data$trt8, col=color[8])
lines(graph_data$time, graph_data$trt9, col=color[9])
```

```
# Add legend
legend("topright",
       c("KM", "DTIC", "Dab", "Dab + Tram", "Ipi", "Vem", "Nivo", "Nivo + Ipi", "Enco + Bini",
"Pembro"),
       col=c(color[10], color[1], color[2], color[3], color[4], color[5], color[6], color[7], color[8],
color[9],
             color[9]),
       lty=c(1,1,1,1), ncol=3, text.width=20, box.lty=0)

# save plot
dev.copy(pdf, "FP 8/negtwo/Plots & AUC/survival_plot_os_nc.pdf")
dev.off()

#------------------------------------------------------------------------
# Step 5: Calculate area under curve for different treatments
#------------------------------------------------------------------------

# Calculate AUC at 60 months
graph60 <- graph_data[1:61, ]

# Create data frame to store AUC results
auc60 <- data.frame(trt=c(1:9), auc=NA)

# Calculate AUC
auc60$auc[1] <- trapz(graph60$time, graph60$trt1)
auc60$auc[2] <- trapz(graph60$time, graph60$trt2)
auc60$auc[3] <- trapz(graph60$time, graph60$trt3)
auc60$auc[4] <- trapz(graph60$time, graph60$trt4)
auc60$auc[5] <- trapz(graph60$time, graph60$trt5)
auc60$auc[6] <- trapz(graph60$time, graph60$trt6)
auc60$auc[7] <- trapz(graph60$time, graph60$trt7)
auc60$auc[8] <- trapz(graph60$time, graph60$trt8)
```

```
auc60$auc[9] <- trapz(graph60$time, graph60$trt9)


# Save area under curves results as CSV file

write.csv(auc60, file="FP 8/negtwo/Plots & AUC/auc60_os_nc.csv")


#------------------------------------------------------------------------

# Step 6: Rank plot

#------------------------------------------------------------------------


# Keep rows for ranking only

rankdata <- results[1097:1177,]


# Variable for rank

rankdata$rank_code <- rep(9:1, 9)


# Restrict probability to 2 decimal places

rankdata$prob <- round(rankdata$mean, 2)


# Add a treatment label

rankdata$Treatment <- c(rep("DTIC", 9), rep("Dab", 9), rep("Dab + Tram", 9), rep("Ipi", 9),
rep("Vem", 9),

                        rep("Nivo", 9), rep("Nivo + Ipi", 9), rep("Enco + Bini", 9), rep("Pembro", 9))


# Rename mean column

names(rankdata)[names(rankdata)=="mean"] <- "Probability"


# Plot with text

q <- ggplot(rankdata, aes(x=rank_code, y=Treatment)) +

  geom_point(aes(size=Probability), shape=21, colour="skyblue", fill="skyblue") +

  theme(panel.background=element_blank(), panel.border=element_rect(colour="black",
fill=NA, size=1),

        legend.position="bottom") +

  scale_size_area(max_size=10) +
```

```
  scale_x_continuous(name="Rank", limits=c(1, 9), breaks=seq(1,9,1)) +
  scale_y_discrete(name="Treatment") +
  geom_text(aes(label=prob))
q


# save plot
dev.copy(pdf, "FP 8/negtwo/Plots & AUC/rank_plot_os_nc.pdf")
dev.off()
```

A.8.7.2 WinBUGS code

```
#Fixed effects 1st order fractional polynomial model (e.g. Weibull (P1=0) and Gompertz (P1=1))
model{                          # *** PROGRAM STARTS
for (j in 1:N){                                                 # LOOP THROUGH EVENTS


# time in months transformed according to power P1
  timen[j]<-(time[j])
  timen1[j]<-(equals(P1,0)*log(timen[j])+(1-equals(P1,0))*pow(timen[j],P1) )


  r[j]~dbin(p[j], z[j])              # likelihood according to eq.
  p[j]<-1-exp(-h[j]*dt[j])                  # hazard rate in each interval standardized by unit of time


#Fixed effects model
# hazard over time according to FP
  log(h[j])<-Alpha[s[j],a[j],1]+Alpha[s[j],a[j],2]*timen1[j]


# Deviance contribution
      rhat[j]<- p[j] * z[j] # expected value of the numerators
      dev[j] <- 2 * (r[j] * (log(r[j])-log(rhat[j]))  +  (z[j]-r[j]) * (log(z[j]-r[j]) - log(z[j]-rhat[j])))
}
```

```
totresdev<- sum(dev[])


for (i in 1:ns){                # LOOP THROUGH STUDIES
  for (k in 1:na[i]){            # LOOP THROUGH ARMS
    Alpha[i,k,1]<-mu[i,1]+d[t[i,k],1]-d[t[i,1],1] # model for linear predictor of alpha_0
    Alpha[i,k,2]<-mu[i,2]+d[t[i,k],2]-d[t[i,1],2] # model for linear predictor of alpha_1
   }
 }


#priors
for (i in 1:ns){                          # LOOP THROUGH STUDIES
  mu[i,1:2] ~ dmnorm(mean[1:2],prec[,])   # vague priors for all trial baselines
 }
d[1,1]<-0                                    # alpha_0 treatment effect is zero for reference
treatment
d[1,2]<-0                                          # alpha_1 treatment effect is zero for
reference treatment


for (k in 2:nt){                              # LOOP THROUGH TREATMENTS
  d[k,1:2] ~ dmnorm(mean[1:2],prec[,])  # vague priors for treatment effects
 }


#Output
for (m in 1:maxt){                              # create time points for output
  time1[m]<-(equals(P1,0)*log(m) + (1-equals(P1,0))*pow(m,P1)    )
 }


#Hazard ratios over time for all possible contrasts
for (c in 1:(nt-1)){
  for (k in (c+1):nt){
    for (m in 1:maxt){
      log(HR[c,k,m])<-(d[k,1]-d[c,1])+(d[k,2]-d[c,2])*time1[m]
```

```
    }
  }
 }

# Provide estimates of survival probabilities over time by treatment
for (k in 1:nt){
  alpha0[k]<-mu[4,1]+d[k,1]       # alpha_0 by treatment using baseline from study 4
  alpha1[k]<-mu[4,2]+d[k,2]       # alpha_1 by treatment using baseline from study 4

  for (m in 1:maxt){
    log(HAZARD[k,m])<-alpha0[k]+alpha1[k]*time1[m]  #hazard over time by treatment
            CUM_H[k,m]<-sum(HAZARD[k,1:m]) # cumulative hazard over time by treatment
    T[k,m]<-1-exp(-CUM_H[k,m])          # mortality over time by treatment
            S[k,m]<-1-T[k,m]             # survival over time by treatment
   }
 }

# Rank treatments at 60 months
for (l in 1:nt) {
    for (m in 1:nt) {
      rk60[l,m] <- equals(ranked(S[,60],m),S[l,60])

    }

  }
}                          # *** PROGRAM ENDS
```

References

1. Robert C, Hwu WJ, Hamid O, Ribas A, Weber JS, Daud AI, et al. Long-term safety of pembrolizumab monotherapy and relationship with clinical outcome: A landmark analysis in patients with advanced melanoma. Eur J Cancer [Internet]. 2021 Feb 1 [cited 2021 Dec 30];144:182–91. Available from: http://www.ejcancer.com/article/S0959804920313332/fulltext

2. Robert C, Long G V., Brady B, Dutriaux C, Di Giacomo AM, Mortier L, et al. Five-year outcomes with nivolumab in patients with wild-type BRAF advanced melanoma. J Clin Oncol. 2020 Nov 20;38(33):3937–46.

3. Robert C, Schachter J, Long G V., Arance A, Grob JJ, Mortier L, et al. Pembrolizumab versus Ipilimumab in Advanced Melanoma. N Engl J Med [Internet]. 2015 Jun 25 [cited 2021 Dec 30];372(26):2521–32. Available from: https://www.nejm.org/doi/full/10.1056/nejmoa1503093

4. Latimer NR, Abrams KR, Amonkar MM, Stapelkamp C, Swann RS. Adjusting for the Confounding Effects of Treatment Switching—The BREAK-3 Trial: Dabrafenib Versus Dacarbazine. Oncologist [Internet]. 2015 Jul [cited 2021 Dec 18];20(7):798–805. Available from: /pmc/articles/PMC4492231/

5. Latimer NR, Amonkar MM, Stapelkamp C, Sun P. Adjusting for confounding effects of treatment switching in a randomized phase II study of dabrafenib plus trametinib in BRAF V600+ metastatic melanoma. Melanoma Res [Internet]. 2015 [cited 2021 Dec 18];25(6):528–36. Available from: https://journals.lww.com/melanomaresearch/Fulltext/2015/12000/Adjusting_for_confounding_effects_of_treatment.9.aspx

6. Mitchell M, Muftakhidinov B, Winchen T, et al. Engauge Digitizer Software [Internet]. [cited 2022 Jan 13]. Available from: http://markummitchell.github.io/engauge-digitizer

7. 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 [Internet]. 2012 Feb 1 [cited 2021 Dec 17];12(1):1–13. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-9

8. 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 [Internet]. 2022 Jan 19 [cited 2022 Jan 22];096228022110702. Available from: https://journals.sagepub.com/doi/10.1177/09622802211070253?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed

9. 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 Synth Methods [Internet]. 2021 Dec 12 [cited 2021 Dec 23]; Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jrsm.1539

10. National Institute for Health and Care Excellence. Evidence reviews for the clinical and cost effectiveness of treatment regimen for the treatment of Stage IIIA-N2 NSCLC [Internet]. NICE. 2018. Available from: https://www.nice.org.uk/guidance/ng122/documents/evidence-review-3

11. Collett D. Modelling Survival Data in Medical Research. 3rd ed. New York: CRC press; 2015.

12. 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 [Internet]. 2012 Mar 23 [cited 2021 Dec 17];12(1):1–14. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-34

13. Jansen JP. Network meta-analysis of survival data with fractional polynomials. BMC Med Res Methodol [Internet]. 2011 May 6 [cited 2021 Dec 17];11(1):1–14. Available from: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-11-61

14. Freeman SC, Carpenter JR. Bayesian one-step IPD network meta-analysis of time-to-event data using Royston-Parmar models. Res Synth Methods [Internet]. 2017 Dec 1 [cited 2021 Dec 17];8(4):451–64. Available from: https://pubmed.ncbi.nlm.nih.gov/28742955/

15. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. J R Stat Soc Ser B (Statistical Methodol [Internet]. 2002 Oct 1 [cited 2021 Dec 30];64(4):583–639. Available from: https://onlinelibrary.wiley.com/doi/full/10.1111/1467-9868.00353

16. MRC Biostatistics Unit. DIC: deviance information criterion [Online] [Internet]. 2015. Available from: https://www.mrc-bsu.cam.ac.uk/software/bugs/the-bugs-project-dic/

17. Gibson E, Koblbauer I, Begum N, Dranitsaris G, Liew D, McEwan P, et al. Modelling the Survival Outcomes of Immuno-Oncology Drugs in Economic Evaluations: A Systematic Approach to Data Analysis and Extrapolation. Pharmacoeconomics [Internet]. 2017 Dec 1 [cited 2021 Dec 30];35(12):1257–70. Available from: https://link.springer.com/article/10.1007/s40273-017-0558-5

18. 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 [Internet]. 2013 Jul 26 [cited 2021 Dec 30];33(5):641–56. Available from: https://journals.sagepub.com/doi/10.1177/0272989X12455847?url_ver=Z39.88-2003&rfr_id=ori%3Arid%3Acrossref.org&rfr_dat=cr_pub++0pubmed

19. 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 [Internet]. 2011. Available from: http://www.nicedsu.org.uk/

20. van Valkenhoef G, Dias S, Ades AE, Welton NJ. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. Res Synth Methods [Internet]. 2016 Mar 1 [cited 2021 Dec 30];7(1):80–93. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/jrsm.1167

21. Hauschild A, Ascierto PA, Schadendorf D, Grob JJ, Ribas A, Kiecker F, et al. Long-term outcomes in patients with BRAF V600-mutant metastatic melanoma receiving dabrafenib monotherapy: Analysis from phase 2 and 3 clinical trials. Eur J Cancer [Internet]. 2020 Jan 1 [cited 2021 Dec 30];125:114–20. Available from: http://www.ejcancer.com/article/S0959804919308159/fulltext

22. Long G V., Eroglu Z, Infante J, Patel S, Daud A, Johnson DB, et al. Long-term outcomes in patients with BRAF V600-mutant metastatic melanoma who received dabrafenib combined with trametinib. J Clin Oncol. 2018 Mar 1;36(7):667–73.

23. McArthur GA, Chapman PB, Robert C, Larkin J, Haanen JB, Dummer R, et al. Safety and efficacy of vemurafenib in BRAFV600E and BRAFV600K mutation-positive melanoma (BRIM-3): extended follow-up of a phase 3, randomised, open-label study.

Lancet Oncol [Internet]. 2014 Mar 1 [cited 2021 Dec 30];15(3):323–32. Available from: http://www.thelancet.com/article/S1470204514700129/fulltext

24. Chapman PB, Robert C, Larkin J, Haanen JB, Ribas A, Hogg D, et al. Vemurafenib in patients with BRAFV600 mutation-positive metastatic melanoma: final overall survival results of the randomized BRIM-3 study. Ann Oncol [Internet]. 2017 Oct 1 [cited 2021 Dec 30];28(10):2581–7. Available from: http://www.annalsofoncology.org/article/S0923753419349439/fulltext

25. Larkin J, Chiarion-Sileni V, Gonzalez R, Grob J-J, Rutkowski P, Lao CD, et al. Five-Year Survival with Combined Nivolumab and Ipilimumab in Advanced Melanoma. N Engl J Med [Internet]. 2019 Sep 27;381(16):1535–46. Available from: https://doi.org/10.1056/NEJMoa1910836

26. Hodi FS, Chesney J, Pavlick AC, Robert C, Grossmann KF, McDermott DF, et al. Combined nivolumab and ipilimumab versus ipilimumab alone in patients with advanced melanoma: 2-year overall survival outcomes in a multicentre, randomised, controlled, phase 2 trial. Lancet Oncol [Internet]. 2016 Nov 1 [cited 2021 Dec 17];17(11):1558–68. Available from: /pmc/articles/PMC5630525/

27. Ascierto PA, Dummer R, Gogas HJ, Flaherty KT, Arance A, Mandala M, et al. Update on tolerability and overall survival in COLUMBUS: landmark analysis of a randomised phase 3 trial of encorafenib plus binimetinib vs vemurafenib or encorafenib in patients with BRAF V600–mutant melanoma. Eur J Cancer [Internet]. 2020 Feb 1 [cited 2021 Dec 17];126:33–44. Available from: https://pubmed.ncbi.nlm.nih.gov/31901705/

28. Robert C, Grob JJ, Stroyakovskiy D, Karaszewska B, Hauschild A, Levchenko E, et al. Five-Year Outcomes with Dabrafenib plus Trametinib in Metastatic Melanoma. N Engl J Med [Internet]. 2019 Jun 4;381(7):626–36. Available from: https://doi.org/10.1056/NEJMoa1904059

29. Robert C, Ribas A, Schachter J, Arance A, Grob J-J, Mortier L, et al. Pembrolizumab versus ipilimumab in advanced melanoma (KEYNOTE-006): post-hoc 5-year results from an open-label, multicentre, randomised, controlled, phase 3 study. Lancet Oncol [Internet]. 2019 Sep 1;20(9):1239–51. Available from: https://doi.org/10.1016/S1470-2045(19)30388-2

30. Atkins MB, Lee SJ, Chmielowski B, Ribas A, Tarhini AA, Truong T-G, et al. DREAMseq (Doublet, Randomized Evaluation in Advanced Melanoma Sequencing): A phase III trial—ECOG-ACRIN EA6134. J Clin Oncol. 2021 Dec 20;39(36_suppl):356154–356154.

31. Freeman SC, Sutton AJ, Cooper NJ. Uptake of methodological advances for synthesis of continuous and time-to-event outcomes would maximize use of the evidence base [Internet]. Vol. 124, Journal of Clinical Epidemiology. J Clin Epidemiol; 2020 [cited 2021 Dec 17]. p. 94–105. Available from: https://pubmed.ncbi.nlm.nih.gov/32407766/

32. Dias S, Ades AE, Welton NJ, Jansen JP SA. Network Meta-Analysis of Survival Outcomes. In: Network Meta-Analysis for Decision-Making. Wiley; 2018. p. 293–322.

33. Bujkiewicz S, Achana F, Papanikos T, Riley RD, Abrams KR. NICE DSU Technical Support Document 20: Multivariate meta-analysis of summary data for combining treatment effects on correlated outcomes and evaluating surrogate endpoints. 2019 [cited 2021 Dec 30]; Available from: www.nicedsu.org.uk

34. Freeman S. Melanoma NMA [Internet]. 2021. Available from: github.com/SCFreeman/Melanoma_NMA

35. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, et al. The

PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations. Ann Intern Med [Internet]. 2015 Jun 2;162(11):777–84. Available from: https://doi.org/10.7326/M14-2385

Appendices

Appendix B: Summary of studies

B.1.1 Summary of studies from the systematic review included as clinical evidence

Table 30: Summary of trials and papers from the systematic review included as clinical evidence

Trial	Interventions	Reference
Immunotherapy studies		
ABC	Nivolumab 1 mg/kg combined with ipilimumab 3 mg/kg every 3 weeks for four doses, then nivolumab 3 mg/kg every 2 weeks Nivolumab 3 mg/kg every 2 weeks	Long 2018
CheckMate 037	Nivolumab 3 mg/kg every 2 weeks DTIC 1000mg/m2, Powder for IV solution, IV, every 3 weeks or carboplatin Area under the concentration-time curve (AUC) 6, solution for injection, IV, every 3 weeks	Larkin 2018
CheckMate 064	Nivolumab 3 mg/kg every 2 weeks up to 6 doses in induction period than 3 mg/kg every 2 weeks, followed by ipilimumab 3 mg/kg solution every 3 weeks up to 4 doses in induction period Ipilimumab 3 mg/kg every 3 weeks up to 4 doses in induction period, followed by nivolumab 3 mg/kg every 2 weeks up to 6 doses in induction period and 3 mg/kg every 2 weeks	Weber 2016
CheckMate 066	Nivolumab 3 mg/kg every 2 weeks DTIC 1000mg/m2 every 3 weeks	Robert 2020
CheckMate 067	Nivolumab 1 mg/kg plus ipilimumab 3 mg/kg every 3 weeks for 4 doses followed by nivolumab 3 mg/kg every 2 weeks Nivolumab 3 mg/kg every 2 weeks plus ipilimumab-matched placebo Ipilimumab 3 mg/kg every 3 weeks for 4 doses plus nivolumab-matched placebo	Larkin 2019
CheckMate 069	Nivolumab 1 mg/kg + ipilimumab 3 mg/kg every 3 weeks for 4 doses, then Nivolumab 3 mg/kg every 2 weeks Ipilimumab 3 mg/kg every 3 weeks for 4 doses plus Nivolumab-matched placebo	Hodi 2016
KEYNOTE-002	Pembrolizumab 2mg/kg Pembrolizumab 10mg/kg Chemotherapy	Hamid 2017
KEYNOTE-006	Pembrolizumab 10mg/kg every 2 weeks Pembrolizumab 10mg/kg every 3 weeks Ipilimumab 3mg/kg every 3 weeks	Robert 2019
Targeted therapy studies		
BREAK-3	Dabrafenib 150 mg b.i.d. DTIC 1000 mg/m2 every 3 weeks	Grob 2014

Trial	Interventions	Reference
Immunotherapy studies		
BRIM-3	Vemurafenib 960mg b.i.d. DTIC 1000mg/m2 every 3 weeks	Chapman 2017
BRF1132220	Dabrafenib 150mg plus trametinib 2mg Dabrafenib 150mg plus trametinib 1mg Dabrafenib 150mg	Long 2018
COLUMBUS	Encorafenib 450 mg q.d. plus binimetinib 45 mg b.i.d. Encorafenib 300mg q.d. Vemurafenib 960mg b.i.d.	Ascierto 2020
COMBI-d	Dabrafenib 150 mg b.i.d. + trametinib 2 mg q.d. Dabrafenib 150 mg b.i.d.	Long 2015
COMBI-v	Dabrafenib 150 mg b.i.d. + trametinib 2 mg q.d. Vemurafenib 960 b.i.d.	Robert 2019

B.1.2 Summary of studies included as part of the review of NICE melanoma TAs

As part of this guideline update, we reviewed NICE TAs for systemic and localised anticancer treatments for people with stage IV (or unresectable stage 3) melanoma. This review identified 12 relevant melanoma TAs. The TAs, as well as the clinical trials cited in their appraisal submissions to NICE, and all papers indexed to those trials on clinicaltrials.gov are listed in Table 31 below.

Table 31: Summary of trials and papers included as part of the review of NICE melanoma TAs

TA #	TA Title	Clinical trials referenced	Papers indexed on clinicaltrials.gov
TA562	Encorafenib with binimetinib for unresectable or metastatic BRAF V600 mutation-positive melanoma	COLUMBUS	Gogas 2021, Gogas 2019, Dummer 2018, Dummer 2018
TA414	Cobimetinib in combination with vemurafenib for treating unresectable or metastatic BRAF V600 mutation-positive melanoma	coBRIM	Ascierto 2020, de la Cruz-Merino 2017, Dréno 2017, Ascierto 2016, Larkin 2014
TA410	Talimogence laherparepvec for treating unresctable metastatic melanoma	OPTiM	Andtbacka 2019, Kaufman 2017, Andtbacka 2015
TA400	Nivolumab in combination with ipilimumab for treating advanced melanoma	CheckMate 067	Larkin 2019, Hodi 2018, Wolchok 2018, Larkin 2018, Long 2017, Schadendorf 2017
		CheckMate 069	Postow 2018, Hodi 2016
TA396	Trametinib in combination with dabrafenib for treating unresectable or metastatic melanoma	COMBI-d	Schadendorf 2021, Syeda 2021, Robert 2019, Long 2016, Long 2015, Schadendorf 2015, Menzies 2014, Long 2014
		COMBI-v	Schadendorf 2021, Robert 2019, Long 2016, Grob 2014, Robert 2014

TA #	TA Title	Clinical trials referenced	Papers indexed on clinicaltrials.gov
		BRF113220	Long 2017, Long 2016, Long 2016, Corcoran 2015, Latimer 2015, Johnson 2014, Frederick 2014, Carlino 2013, Flaherty 2012
TA384	Nivolumab for treating advanced (unresectable or metastatic) melanoma	CheckMate 037	Larkin 2017, Weber 2015
		CheckMate 066	Robert 2020, Ascierto 2019, Long 2018, Long 2017, Robert 2014
		CheckMate 067	Previously covered – see TA400
TA366	Pembrolizumab for advanced melanoma not previously treated with ipilimumab	KEYNOTE-001	Hamid 2021, Robert 2020, Lala 2020, van Vugt 2019, Garon 2019, Leighl 2019, Wang 2019, Hamid 2019, Hamid 2018, Joseph 2018, Brogden 2018, Robert 2017, Shaverdian 2017, Hui 2017, Daud 2016, Ribas 2016, Hodi 2016, Garon 2015, Hamid 2013
		KEYNOTE-006	Hamid 2021, Robert 2021, Lala 2021, van Vugt 2019, Robert 2019, Wang 2019, Hamid 2018, Carlino 2018, Petrella 2017, Schachter 2017, Robert 2015
TA357	Pembrolizumab for treating advanced melanoma after disease progression with ipilimumab	KEYNOTE-001	Previously covered – see TA366
		KEYNOTE-002	Robert 2020, Lala 2020, van Vugt 2019, Wang 2019, Hamid 2018, Hamid 2017, Schadendorf 2016, Ribas 2015
TA321	Dabrafenib for treating unresectable or metastatic BRAF V600 mutation-positive melanoma	BREAK-3	Hauschild 2020, Santiago-Walker 2015, Latimer 2015, Grob 2014, Ouellet 2014, Hauschild 2012
TA319	Ipilimumab for previously untreated advanced (unresectable or metastatic) melanoma	CA184-024	Maio 2015, Schadendorf 2015, Robert 2011
		MDX010-08	Hersh 2010
		BREAK-3	Previously covered – see TA321
		BRIM-3	Ascierto 2020, Chapman 2017, Yamazaki 2015, Frederick 2014, McArthur 2014, Lacouture 2013, Su 2012, Chapman 2011
TA269	Vemurafenib for treating locally advanced or metastatic BRAF V600 mutation-positive malignant melanoma	BRIM-3	Previously covered – see TA319
TA268	Ipilimumab for previously treated advanced (unresectable or metastatic) melanoma	MDX010-20	Larkin 2015, Koguchi 2015, Schadendorf 2015, Johnson 2015, Hatswell 2014, McDermott 2013, Robert 2013, Weber 2013, Revicki 2012, Hodi 2010
		CA184-022	Schadendorf 2015, Wolchok 2009
		CA184-007	Schadendorf 2015

B.1.3 Summary of papers put forward by committee and supplementary search results

The committee put forward two pieces of information they thought may be useful:

- The supplementary appendix to Robert C, Grob JJ, Stroyakovskiy D, et al. Five-year outcomes with dabrafenib plus trametinib in metastatic melanoma. N Engl J Med 2019;381:626-36. DOI: 10.1056/NEJMoa1904059
- Dummer. ASCO 2021. Abstr 9507

Only one additional paper was found during the supplementary searches, which is listed below:

- Ascierto PA, Dummer R, Gogas HJ, Flaherty KT, Arance A, Mandala M, Liszkay G, Garbe C, Schadendorf D, Krajsova I, Gutzmer R. Update on tolerability and overall survival in COLUMBUS: landmark analysis of a randomised phase 3 trial of encorafenib plus binimetinib vs vemurafenib or encorafenib in patients with BRAF V600–mutant melanoma. European Journal of Cancer. 2020 Feb 1;126:33-44.

B.1.4 Excluded studies

Trials and publications that were not used in the NMA and the reasons for exclusion are detailed in Table 32 and Table 33.

Table 32: Studies excluded from NMA and exclusion reasons

Trial	Reason
ABC	
Long, Georgina V; Atkinson, Victoria; Lo, Serigne; Sandhu, Shahneen; Guminski, Alexander D; Brown, Michael P; Wilmott, James S; Edwards, Jarem; Gonzalez, Maria; Scolyer, Richard A; Menzies, Alexander M; McArthur, Grant A; Combination nivolumab and ipilimumab or nivolumab alone in melanoma brain metastases: a multicentre randomised phase 2 study.; The Lancet. Oncology; 2018; vol. 19 (no. 5); 672-681	Study population was patients with diagnosed brain metastases. While other studies included patients with brain metastases, their populations were not exclusively patients with brain metastases. As such it was felt this population was different compared to other trials in the network and it was inappropriate to include. Additionally, this trial was investigating Nivolumab combined with ipilimumab, both of which are already included in the network through CheckMate 067. Finally, this trial had a relatively smaller number of participants (n=79) and a short duration of follow-up (24 months). While neither of these were the primary reasons for exclusion, they were additional factors supporting exclusion.
CheckMate 037	
Larkin, James; Minor, David; D'Angelo, Sandra; Neyns, Bart; Smylie, Michael; Miller, Wilson H Jr; Gutzmer, Ralf; Linette, Gerald; Chmielowski, Bartosz; Lao, Christopher D; Lorigan, Paul; Grossmann, Kenneth; Hassel, Jessica C; Sznol, Mario; Daud, Adil; Sosman, Jeffrey; Khushalani, Nikhil; Schadendorf, Dirk; Hoeller, Christoph; Walker, Dana; Kong, George; Horak, Christine; Weber, Jeffrey; Overall Survival in Patients With Advanced Melanoma Who Received Nivolumab Versus Investigator's Choice Chemotherapy in CheckMate 037: A Randomized, Controlled, Open-Label Phase III Trial.; Journal of clinical oncology : official journal of the American	Patients in CheckMate 037 were randomized to receive either Nivolumab or Investigator's choice chemotherapy (ICC), which could be either DTIC or Carboplatin and Paclitaxel. Inclusion of CheckMate 037 would require either classing ICC as DTIC, or introducing a new node in the network, ICC. However, the network already has CheckMate 066 in which Nivolumab was compared against DTIC. Thus, if we were to consider classing ICC as DTIC, we would not be introducing a new comparison to the model but would be increasing the sample size for an existing comparison. Furthermore, the published results of CheckMate 037 noted a difference in

Trial	Reason
Society of Clinical Oncology; 2018; vol. 36 (no. 4); 383-390	responses between DTIC compared with paclitaxel or carboplatin. Given this information, we felt it was best not to consider ICC as equivalent to DTIC. After this decision not to consider ICC equivalent to DTIC was made, we decided to exclude from the network as this comparison would have been between an existing treatment and a treatment not already in the network. Therefore, its inclusion would have limited use for improving our estimates of treatments effects. As such this trial was excluded.
CheckMate 064	
Weber, Jeffrey S; Gibney, Geoff; Sullivan, Ryan J; Sosman, Jeffrey A; Slingluff, Craig L Jr; Lawrence, Donald P; Logan, Theodore F; Schuchter, Lynn M; Nair, Suresh; Fecher, Leslie; Buchbinder, Elizabeth I; Berghorn, Elmer; Ruisi, Mary; Kong, George; Jiang, Joel; Horak, Christine; Hodi, F Stephen; Sequential administration of nivolumab and ipilimumab with a planned switch in patients with advanced melanoma (CheckMate 064): an open-label, randomised, phase 2 trial.; The Lancet. Oncology; 2016; vol. 17 (no. 7); 943-955	Patients in CheckMate 064 were randomized to receive either nivolumab followed by ipilimumab or ipilimumab followed by nivolumab. As such, this comparison was not between two different treatments but rather a sequence. As such this trial was excluded.
KEYNOTE-002	
Hamid, Omid; Puzanov, Igor; Dummer, Reinhard; Schachter, Jacob; Daud, Adil; Schadendorf, Dirk; Blank, Christian; Cranmer, Lee D; Robert, Caroline; Pavlick, Anna C; Gonzalez, Rene; Hodi, F Stephen; Ascierto, Paolo A; Salama, April K S; Margolin, Kim A; Gangadhar, Tara C; Wei, Ziwen; Ebbinghaus, Scot; Ibrahim, Nageatte; Ribas, Antoni; Final analysis of a randomised trial comparing pembrolizumab versus investigator-choice chemotherapy for ipilimumab-refractory advanced melanoma.; European journal of cancer (Oxford, England : 1990); 2017; vol. 86; 37-45	Patients in KEYNOTE-002 were randomized to receive pembrolizumab 2mg/kg, pembrolizumab 10mg/kg or chemotherapy. As chemotherapy was not part of the network, this would introduce a comparison between a treatment already in the network (pembrolizumab 2mg/kg) with a treatment not in the network (chemotherapy). Therefore, its inclusion would have limited use for improving our estimates of treatments effects. As such this trial was excluded.
CheckMate 511	
Lebbe, Celeste, Meyer, Nicolas, Mortier, Laurent et al. (2019) Evaluation of Two Dosing Regimens for Nivolumab in Combination With Ipilimumab in Patients With Advanced Melanoma: Results From the Phase IIIb/IV CheckMate 511 Trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 37(11): 867-875	Patients in CheckMate 511 were randomized to receive either Nivolumab 3mg/kg plus Ipilimumab 1mg/kg or Nivolumab 1mg/kg plus Ipilimumab 3mg/kg. As such, this comparison was not between two different treatments but rather the same treatment (Nivolumab+Ipilimumab) at different doses. Therefore, this trial was excluded.
METRIC	
Robert C, Flaherty K, Nathan P, Hersey P, Garbe C, Milhem M, Demidov L, Mohr P, Hassel JC, Rutkowski P, Dummer R. Five-year outcomes from a phase 3 METRIC study in patients with BRAF V600 E/K–mutant advanced or metastatic melanoma. European Journal of Cancer. 2019 Mar 1;109:61-9.	Patients in METRIC were randomized to receive either Trametinib monotherapy or DTIC. We initially included this trial in the network. However, the results of this trial proved to greatly affect the results of the network. This is because the OS HR for METRIC crosses the line of no effect (0.7034-1.209). This results in DTIC appearing better, as there is another

Trial	Reason
	treatment that performs indistinguishable from it. On removing METRIC from the network, more credible estimates of PFS and OS were obtained as verified by the committee. A further justification for its removal is the fact that Trametinib is not given in current clinical practice as a monotherapy. Both due to its effect on the network's results and it not reflecting clinically meaningful treatments (Trametinib monotherapy), this trial was excluded.
NCT01515189	
Ascierto PA, Del Vecchio M, Robert C, Mackiewicz A, Chiarion-Sileni V, Arance A, Lebbé C, Bastholt L, Hamid O, Rutkowski P, McNeil C. Ipilimumab 10 mg/kg versus ipilimumab 3 mg/kg in patients with unresectable or metastatic melanoma: a randomised, double-blind, multicentre, phase 3 trial. The Lancet Oncology. 2017 May 1;18(5):611-22.	Patients in NCT01515189 were randomized to receive either 10mg/kg ipilimumab or 3mg/kg ipilimumab. As such, this comparison was not between two different treatments but rather the same treatment (Ipilimumab) at different doses. Therefore, this trial was excluded.
S1320	
Algazi, A.P., Othus, M., Daud, A.I. et al. (2020) Continuous versus intermittent BRAF and MEK inhibition in patients with BRAF-mutated melanoma: a randomized phase 2 trial. Nature Medicine 26(10): 1564-1568	Patients in S1320 were randomized to receive either continuous or intermittent dosing of dabrafenib + trametinib. As such, this comparison was not between two different treatments but rather the same treatment (dabrafenib + trametinib) at different dosing schedules. Therefore, this trial was excluded.
coBRIM	
Excluded prior to full text screening	Patients in coBRIM were randomized to receive either vemurafenib + cobimetinib or vemurafenib. Vemurafenib + cobimetinib was not recommended for routine commissioning in NICE TA414. As such, this trial was excluded as it was a comparison between a treatment not recommended by NICE and a treatment already in the network.
OPTiM	
Excluded prior to full text screening	Patients in OPTiM received Talimogence laherparepvec. This is a localised treatment rather than a systemic anticancer treatment. Therefore, this trial was excluded.
KEYNOTE-001	
Excluded prior to full text screening	Patients in KEYNOTE-001 were randomized to receive different doses of pembrolizumab. As such, this comparison was not between two different treatments but rather the same treatment (Pembrolizumab) at different doses. Therefore, this trial was excluded.
CA184-024	
Excluded prior to full text screening	Patients in CA184-024 were randomized to receive Ipilimumab + DTIC or DTIC. As Ipilimumab + DTIC was not part of the network, this would introduce a comparison between a treatment already in the network (Ipilimumab) with a treatment not in the network (Ipilimumab + DTIC). Therefore, its inclusion would have

Trial	Reason
	limited use for improving our estimates of treatments effects. Additionally, giving Ipilimumab + DTIC is not a treatment combination reflected in clinical practice. For both reasons, this trial was excluded.
MDX010-08	
Excluded prior to full text screening	Patients in MDX010-08 were randomized to receive Ipilimumab + DTIC or Ipilimumab. As Ipilimumab + DTIC was not part of the network, this would introduce a comparison between a treatment already in the network (Ipilimumab) with a treatment not in the network (Ipilimumab + DTIC). Therefore, its inclusion would have limited use for improving our estimates of treatments effects. Additionally, giving Ipilimumab + DTIC is not a treatment combination reflected in clinical practice. For both reasons, this trial was excluded.
MDX010-20	
Excluded prior to full text screening	Patients in MDX010-20 were randomized to receive Ipilimumab + a peptide vaccine, Ipilimumab, or a peptide vaccine. As Ipilimumab + a peptide vaccine was not part of the network, this would introduce a comparison between a treatment already in the network (Ipilimumab) with a treatment not in the network (Ipilimumab + a peptide vaccine). Therefore, its inclusion would have limited use for improving our estimates of treatments effects. Additionally, giving Ipilimumab + a peptide vaccine is not reflective of current clinical practice. For both reasons, this trial was excluded.
CA184-022	
Excluded prior to full text screening	Patients in KEYNOTE-001 were randomized to receive different doses of ipilimumab. As such, this comparison was not between two different treatments but rather the same treatment (Ipilimumab) at different doses. Therefore, this trial was excluded.
CA184-007	
Excluded prior to full text screening	Patients in CA184-007 were randomized to receive Ipilimumab or Ipilimumab + Budesonide. As Ipilimumab + Budesonide was not part of the network, this would introduce a comparison between a treatment already in the network (Ipilimumab) with a treatment not in the network (Ipilimumab + Budesonide). Therefore, its inclusion would have limited use for improving our estimates of treatments effects. Additionally, Ipilimumab + Budesonide is not reflective of current clinical practice. For both reasons, this trial was excluded.

Table 33: Publications that were not used in the NMA, but were assessed for suitable KM curves

Paper	Reason
BRIM-3	
Ascierto PA, Ribas A, Larkin J, McArthur GA, Lewis KD, Hauschild A, Flaherty KT, McKenna E, Zhu Q, Mun Y, Dréno B. Impact of initial treatment and prognostic factors on postprogression survival in BRAF-mutated metastatic melanoma treated with dacarbazine or vemurafenib±cobimetinib: a pooled analysis of four clinical trials. Journal of translational medicine. 2020 Dec;18(1):1-2.	KM curves present for post-progression overall survival, however, none of these curves represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Yamazaki N, Kiyohara Y, Sugaya N, Uhara H. Phase I/II study of vemurafenib in patients with unresectable or recurrent melanoma with BRAFV 600 mutations. The Journal of dermatology. 2015 Jul;42(7):661-6.	No KM curves
Frederick DT, Salas Fragomeni RA, Schalck A, Ferreiro-Neira I, Hoff T, Cooper ZA, Haq R, Panka DJ, Kwong LN, Davies MA, Cusack JC. Clinical profiling of BCL-2 family members in the setting of BRAF inhibition offers a rationale for targeting de novo resistance using BH3 mimetics. PloS one. 2014 Jul 1;9(7):e101286.	No KM curves
Lacouture ME, Duvic M, Hauschild A, Prieto VG, Robert C, Schadendorf D, Kim CC, McCormack CJ, Myskowski PL, Spleiss O, Trunzer K. Analysis of dermatologic events in vemurafenib-treated patients with melanoma. The oncologist. 2013 Mar;18(3):314.	No KM curves
Su F, Viros A, Milagre C, Trunzer K, Bollag G, Spleiss O, Reis-Filho JS, Kong X, Koya RC, Flaherty KT, Chapman PB. RAS mutations in cutaneous squamous-cell carcinomas in patients treated with BRAF inhibitors. New England Journal of Medicine. 2012 Jan 19;366(3):207-15.	No KM curves
Chapman PB, Hauschild A, Robert C, Haanen JB, Ascierto P, Larkin J, Dummer R, Garbe C, Testori A, Maio M, Hogg D. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. New England Journal of Medicine. 2011 Jun 30;364(26):2507-16.	KM curves present, but analyses with longer follow-up found in a different paper
CheckMate 066	
Ascierto PA, Long GV, Robert C, Brady B, Dutriaux C, Di Giacomo AM, Mortier L, Hassel JC, Rutkowski P, McNeil C, Kalinka-Warzocha E. Survival outcomes in patients with previously untreated BRAF wild-type advanced melanoma treated with nivolumab therapy: three-year follow-up of a randomized phase 3 trial. JAMA oncology. 2019 Feb 1;5(2):187-94.	KM curves present, but analyses with longer follow-up found in a different paper
Long GV, Tykodi SS, Schneider JG, Garbe C, Gravis G, Rashford M, Agrawal S, Grigoryeva E, Bello A, Roy A, Rollin L. Assessment of nivolumab exposure and clinical safety of 480 mg every 4 weeks flat-dosing schedule in	No KM curves

Paper	Reason
patients with cancer. Annals of Oncology. 2018 Nov 1;29(11):2208-13.	
Long GV, Weber JS, Larkin J, Atkinson V, Grob JJ, Schadendorf D, Dummer R, Robert C, Márquez-Rodas I, McNeil C, Schmidt H. Nivolumab for patients with advanced melanoma treated beyond progression: analysis of 2 phase 3 clinical trials. JAMA oncology. 2017 Nov 1;3(11):1511-9.	No KM curves
Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, Hassel JC, Rutkowski P, McNeil C, Kalinka-Warzocha E, Savage KJ. Nivolumab in previously untreated melanoma without BRAF mutation. New England journal of medicine. 2015 Jan 22;372(4):320-30.	KM curves present, but analyses with longer follow-up found in a different paper
CheckMate 067	
Hodi FS, Chiarion-Sileni V, Gonzalez R, Grob JJ, Rutkowski P, Cowey CL, Lao CD, Schadendorf D, Wagstaff J, Dummer R, Ferrucci PF. Nivolumab plus ipilimumab or nivolumab alone versus ipilimumab alone in advanced melanoma (CheckMate 067): 4-year outcomes of a multicentre, randomised, phase 3 trial. The Lancet Oncology. 2018 Nov 1;19(11):1480-92.	KM curves present, but analyses with longer follow-up found in a different paper
Wolchok JD, Chiarion-Sileni V, Gonzalez R, Rutkowski P, Grob JJ, Cowey CL, Lao CD, Wagstaff J, Schadendorf D, Ferrucci PF, Smylie M. Overall survival with combined nivolumab and ipilimumab in advanced melanoma. New England Journal of Medicine. 2017 Oct 5;377(14):1345-56.	KM curves present, but analyses with longer follow-up found in a different paper
Long GV, Weber JS, Larkin J, Atkinson V, Grob JJ, Schadendorf D, Dummer R, Robert C, Márquez-Rodas I, McNeil C, Schmidt H. Nivolumab for patients with advanced melanoma treated beyond progression: analysis of 2 phase 3 clinical trials. JAMA oncology. 2017 Nov 1;3(11):1511-9.	No KM curves
Schadendorf D, Larkin J, Wolchok J, Hodi FS, Chiarion-Sileni V, Gonzalez R, Rutkowski P, Grob JJ, Cowey CL, Lao C, Wagstaff J. Health-related quality of life results from the phase III CheckMate 067 study. European Journal of Cancer. 2017 Sep 1;82:80-91.	No KM curves
Larkin J, Chiarion-Sileni V, Gonzalez R, Grob JJ, Cowey CL, Lao CD, Schadendorf D, Dummer R, Smylie M, Rutkowski P, Ferrucci PF. Combined nivolumab and ipilimumab or monotherapy in untreated melanoma. New England journal of medicine. 2015 Jul 2;373(1):23-34.	PFS curve present, but analyses with longer follow-up found in a different paper
CheckMate 069	
Postow MA, Chesney J, Pavlick AC, Robert C, Grossmann K, McDermott D, Linette GP, Meyer N, Giguere JK, Agarwala SS, Shaheen M. Nivolumab and ipilimumab versus ipilimumab in	PFS curve present, but analyses with longer follow-up found in a different paper

Paper	Reason
untreated melanoma. New England Journal of Medicine. 2015 May 21;372(21):2006-17.	
COLUMBUS	
Gogas H, Dummer R, Ascierto PA, Arance A, Mandalà M, Liszkay G, Garbe C, Schadendorf D, Krajsová I, Gutzmer R, Sileni VC. Quality of life in patients with BRAF-mutant melanoma receiving the combination encorafenib plus binimetinib: Results from a multicentre, open-label, randomised, phase III study (COLUMBUS). European Journal of Cancer. 2021 Jul 1;152:116-28.	KM curves present by time to definitive 10% deterioration. However, these curves do not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Gogas HJ, Flaherty KT, Dummer R, Ascierto PA, Arance A, Mandala M, Liszkay G, Garbe C, Schadendorf D, Krajsova I, Gutzmer R. Adverse events associated with encorafenib plus binimetinib in the COLUMBUS study: incidence, course and management. European Journal of Cancer. 2019 Sep 1;119:97-106.	No KM curves
Dummer R, Ascierto PA, Gogas HJ, Arance A, Mandala M, Liszkay G, Garbe C, Schadendorf D, Krajsova I, Gutzmer R, Sileni VC. Overall survival in patients with BRAF-mutant melanoma receiving encorafenib plus binimetinib versus vemurafenib or encorafenib (COLUMBUS): a multicentre, open-label, randomised, phase 3 trial. The Lancet Oncology. 2018 Oct 1;19(10):1315-27.	OS curve present, but analyses with longer follow-up found in a different paper
Dummer R, Ascierto PA, Gogas HJ, Arance A, Mandala M, Liszkay G, Garbe C, Schadendorf D, Krajsova I, Gutzmer R, Chiarion-Sileni V. Encorafenib plus binimetinib versus vemurafenib or encorafenib in patients with BRAF-mutant melanoma (COLUMBUS): a multicentre, open-label, randomised phase 3 trial. The Lancet Oncology. 2018 May 1;19(5):603-15.	PFS curve present, but analyses with longer follow-up found in a different paper
Dummer. ASCO 2021. Abstr 9507	This is a meeting abstract. As such it was excluded, as we did not include abstracts. However, we acknowledge this abstract presents more data than the publication we used in the NMA, as it includes the KM curve for encorafenib monotherapy, which ours did not.
COMBI-d	
Schadendorf D, Robert C, Dummer R, Flaherty KT, Tawbi HA, Menzies AM, Banerjee H, Lau M, Long GV. Pyrexia in patients treated with dabrafenib plus trametinib across clinical trials in BRAF-mutant cancers. European Journal of Cancer. 2021 Aug 1;153:234-41.	No KM curves
Syeda MM, Wiggins JM, Corless BC, Long GV, Flaherty KT, Schadendorf D, Nathan PD, Robert C, Ribas A, Davies MA, Grob JJ. Circulating tumour DNA in patients with advanced melanoma treated with dabrafenib or dabrafenib plus trametinib: a clinical validation study. The Lancet Oncology. 2021 Mar 1;22(3):370-80.	KM curves present by ctDNA status. However, these curves do not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).

Paper	Reason
Robert C, Grob JJ, Stroyakovskiy D, Karaszewska B, Hauschild A, Levchenko E, Chiarion Sileni V, Schachter J, Garbe C, Bondarenko I, Gogas H. Five-year outcomes with dabrafenib plus trametinib in metastatic melanoma. New England Journal of Medicine. 2019 Aug 15;381(7):626-36.	KM curves present for single arm of the trial (Dabrafenib + trametinib), and factors identified as predictive. However, curves that only present a single arm are not suitable for inclusion, and curves that are by predictive factors do not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Long GV, Grob JJ, Nathan P, Ribas A, Robert C, Schadendorf D, Lane SR, Mak C, Legenne P, Flaherty KT, Davies MA. Factors predictive of response, disease progression, and overall survival after dabrafenib and trametinib combination treatment: a pooled analysis of individual patient data from randomised trials. The lancet oncology. 2016 Dec 1;17(12):1743-54.	KM curves present for pooled analyses of Dabrafenib + Trametinib from three trials, a comparison of Dabrafenib + Trametinib arms in different trials, survival after progression by sites of progression and factors identified as predictive. However, none of these curves represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Long GV, Stroyakovskiy D, Gogas H, Levchenko E, De Braud F, Larkin J, Garbe C, Jouary T, Hauschild A, Grob JJ, Chiarion-Sileni V. Dabrafenib and trametinib versus dabrafenib and placebo for Val600 BRAF-mutant melanoma: a multicentre, double-blind, phase 3 randomised controlled trial. The Lancet. 2015 Aug 1;386(9992):444-51.	KM curves present, but analyses with longer follow-up found in a different paper
Schadendorf D, Amonkar MM, Stroyakovskiy D, Levchenko E, Gogas H, De Braud F, Grob JJ, Bondarenko I, Garbe C, Lebbe C, Larkin J. Health-related quality of life impact in a randomised phase III study of the combination of dabrafenib and trametinib versus dabrafenib monotherapy in patients with BRAF V600 metastatic melanoma. European Journal of Cancer. 2015 May 1;51(7):833-40.	No KM curves
Menzies AM, Ashworth MT, Swann S, Kefford RF, Flaherty K, Weber J, Infante JR, Kim KB, Gonzalez R, Hamid O, Schuchter L. Characteristics of pyrexia in BRAFV600E/K metastatic melanoma patients treated with combined dabrafenib and trametinib in a phase I/II clinical trial. Annals of Oncology. 2015 Feb 1;26(2):415-21.	PFS curve present by Pyrexia status and by grade. However, these curves do not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Long GV, Stroyakovskiy D, Gogas H, Levchenko E, de Braud F, Larkin J, Garbe C, Jouary T, Hauschild A, Grob JJ, Chiarion Sileni V. Combined BRAF and MEK inhibition versus BRAF inhibition alone in melanoma. New England Journal of Medicine. 2014 Nov 13;371(20):1877-88.	KM curves present, but analyses with longer follow-up found in a different paper
COMBI-v	
Schadendorf D, Robert C, Dummer R, Flaherty KT, Tawbi HA, Menzies AM, Banerjee H, Lau M, Long GV. Pyrexia in patients treated with dabrafenib plus trametinib across clinical trials in BRAF-mutant cancers. European Journal of Cancer. 2021 Aug 1;153:234-41.	No KM curves

Paper	Reason
Robert C, Grob JJ, Stroyakovskiy D, Karaszewska B, Hauschild A, Levchenko E, Chiarion Sileni V, Schachter J, Garbe C, Bondarenko I, Gogas H. Five-year outcomes with dabrafenib plus trametinib in metastatic melanoma. New England Journal of Medicine. 2019 Aug 15;381(7):626-36.	KM curves present for single arm of the trial (Dabrafenib + trametinib), and factors identified as predictive. However, curves that only present a single arm are not suitable for inclusion, and curves that are by predictive factors do not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Long GV, Grob JJ, Nathan P, Ribas A, Robert C, Schadendorf D, Lane SR, Mak C, Legenne P, Flaherty KT, Davies MA. Factors predictive of response, disease progression, and overall survival after dabrafenib and trametinib combination treatment: a pooled analysis of individual patient data from randomised trials. The lancet oncology. 2016 Dec 1;17(12):1743-54.	KM curves present for pooled analyses of Dabrafenib + Trametinib from three trials, a comparison of Dabrafenib + Trametinib arms in different trials, survival after progression by sites of progression and factors identified as predictive. However, none of these curves represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Grob JJ, Amonkar MM, Karaszewska B, Schachter J, Dummer R, Mackiewicz A, Stroyakovskiy D, Drucis K, Grange F, Chiarion-Sileni V, Rutkowski P. Comparison of dabrafenib and trametinib combination therapy with vemurafenib monotherapy on health-related quality of life in patients with unresectable or metastatic cutaneous BRAF Val600-mutation-positive melanoma (COMBI-v): results of a phase 3, open-label, randomised trial. The Lancet Oncology. 2015 Oct 1;16(13):1389-98.	No KM curves
Robert C, Karaszewska B, Schachter J, Rutkowski P, Mackiewicz A, Stroiakovski D, Lichinitser M, Dummer R, Grange F, Mortier L, Chiarion-Sileni V. Improved overall survival in melanoma with combined dabrafenib and trametinib. New England Journal of Medicine. 2015 Jan 1;372(1):30-9.	KM curves present, but analyses with longer follow-up found in a different paper
KEYNOTE-006	
Hamid O, Robert C, Daud A, Carlino MS, Mitchell TC, Hersey P, Schachter J, Long GV, Hodi FS, Wolchok JD, Arance A. Long-term outcomes in patients with advanced melanoma who had initial stable disease with pembrolizumab in KEYNOTE-001 and KEYNOTE-006. European Journal of Cancer. 2021 Nov 1;157:391-402.	OS curves present by subsequent response (complete, partial, or stable disease). However, these curves do not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Robert C, Hwu WJ, Hamid O, Ribas A, Weber JS, Daud AI, Hodi FS, Wolchok JD, Mitchell TC, Hersey P, Dronca R. Long-term safety of pembrolizumab monotherapy and relationship with clinical outcome: A landmark analysis in patients with advanced melanoma. European Journal of Cancer. 2021 Feb 1;144:182-91.	PFS curve present by immune-mediated adverse events. However, this curve does not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Lala M, Li TR, de Alwis DP, Sinha V, Mayawala K, Yamamoto N, Siu LL, Chartash E, Aboshady H, Jain L. A six-weekly dosing schedule for pembrolizumab in patients with cancer based on evaluation using modelling and simulation. European Journal of Cancer. 2020 May 1;131:68-75.	No KM curves

Paper	Reason
Van Vugt MJ, Stone JA, Snyder ES, Lipka L, Turner DC, Chain A, Lala M, Li M, Robey SH, Kondic AG, De Alwis D. Immunogenicity of pembrolizumab in patients with advanced tumors. Journal for immunotherapy of cancer. 2019 Dec;7(1):1-8.	No KM curves
Wang M, Chen C, Jemielita T, Anderson J, Li XN, Hu C, Kang SP, Ibrahim N, Ebbinghaus S. Are tumor size changes predictive of survival for checkpoint blockade based immunotherapy in metastatic melanoma?. Journal for immunotherapy of cancer. 2019 Dec;7(1):1-0.	OS curve present by early tumour size changes. However, this curve does not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Hamid O, Robert C, Ribas A, Hodi FS, Walpole E, Daud A, Arance AS, Brown E, Hoeller C, Mortier L, Schachter J. Antitumour activity of pembrolizumab in advanced mucosal melanoma: a post-hoc analysis of KEYNOTE-001, 002, 006. British journal of cancer. 2018 Sep;119(6):670-4.	KM curves present for mucosal and nonmucosal melanoma by past treatment status with ipilimumab. Not only does this type of melanoma represent a subgroup of melanoma and is therefore more specific than our population of interest, but curves by past treatment status with ipilimumab do not represent the types of curves needed for this NMA (PFS and OS for different systemic anticancer treatments).
Carlino MS, Long GV, Schadendorf D, Robert C, Ribas A, Richtig E, Nyakas M, Caglevic C, Tarhini A, Blank C, Hoeller C. Outcomes by line of therapy and programmed death ligand 1 expression in patients with advanced melanoma treated with pembrolizumab or ipilimumab in KEYNOTE-006: a randomised clinical trial. European Journal of Cancer. 2018 Sep 1;101:236-43.	KM curves present, but analyses with longer follow-up found in a different paper
Petrella TM, Robert C, Richtig E, Miller Jr WH, Masucci GV, Walpole E, Lebbe C, Steven N, Middleton MR, Hille D, Zhou W. Patient-reported outcomes in KEYNOTE-006, a randomised study of pembrolizumab versus ipilimumab in patients with advanced melanoma. European journal of cancer. 2017 Nov 1;86:115-24.	No KM curves
Robert C, Schachter J, Long GV, Arance A, Grob JJ, Mortier L, Daud A, Carlino MS, McNeil C, Lotem M, Larkin J. Pembrolizumab versus ipilimumab in advanced melanoma. New England Journal of Medicine. 2015 Jun 25;372(26):2521-32.	KM curves present, but analyses with longer follow-up found in a different paper

Appendix C: Cox proportional hazards

Cox proportional hazards

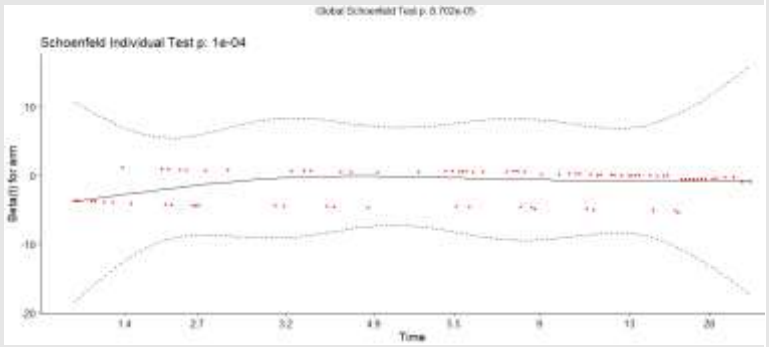
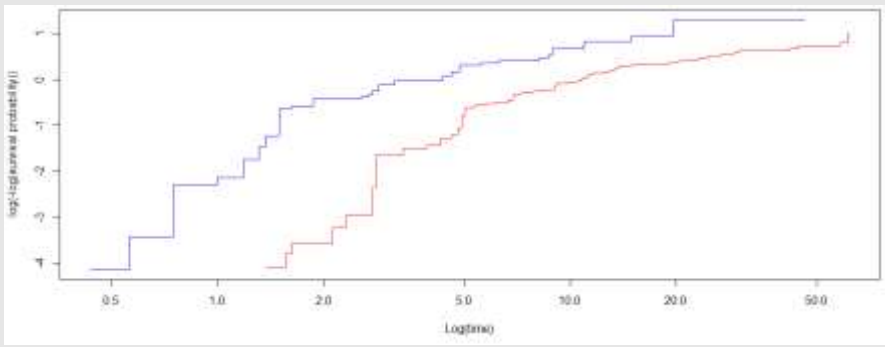
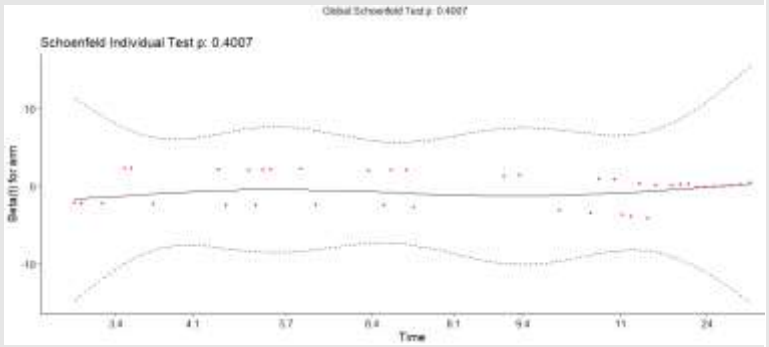
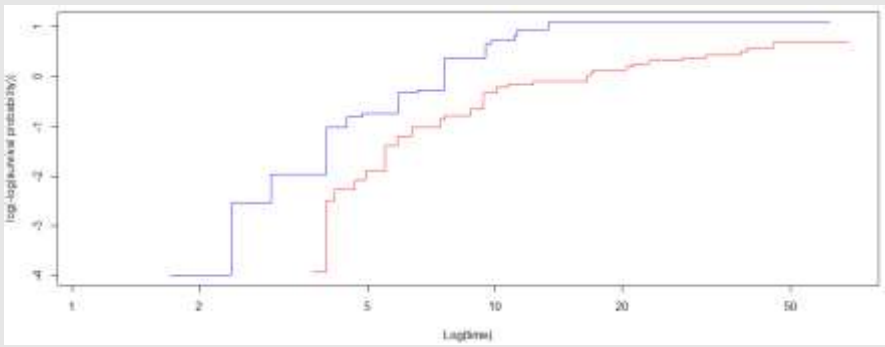
A summary of assessment of proportional hazards (PH) for each PFS and OS curve for each trial in the network is presented in Table 34. In most trials, the PH assumption was not met. As such, Cox PH models were not considered further.

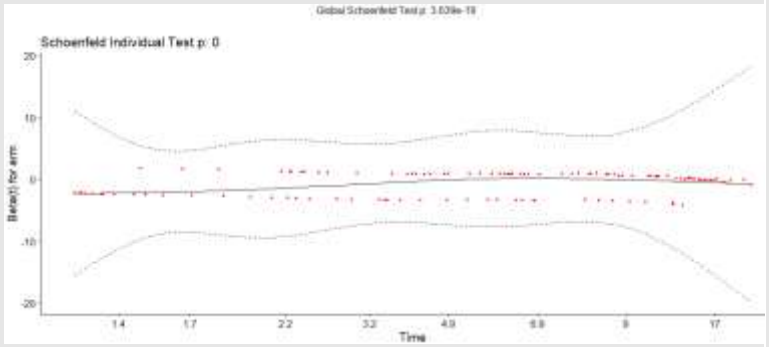
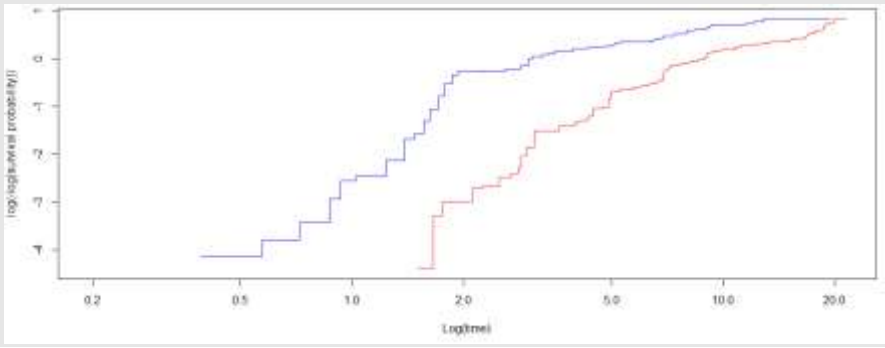
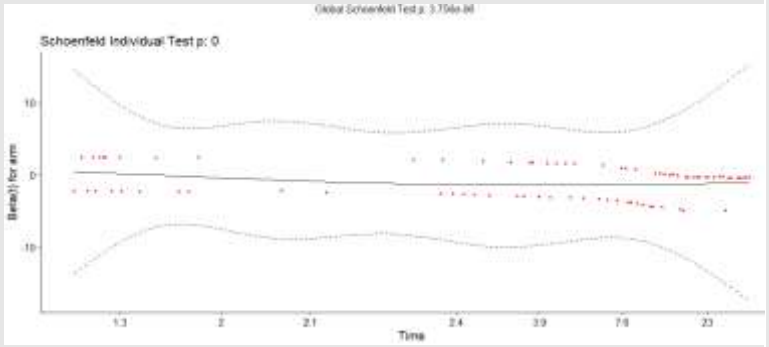
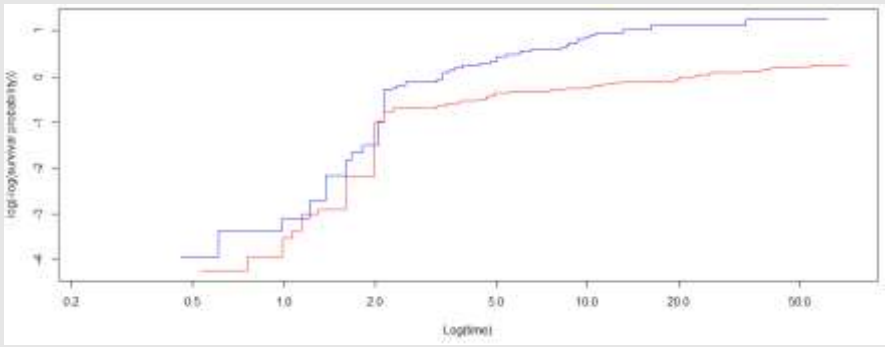
Table 34: Summary of proportional hazards assessments for each PFS and OS curve for every network

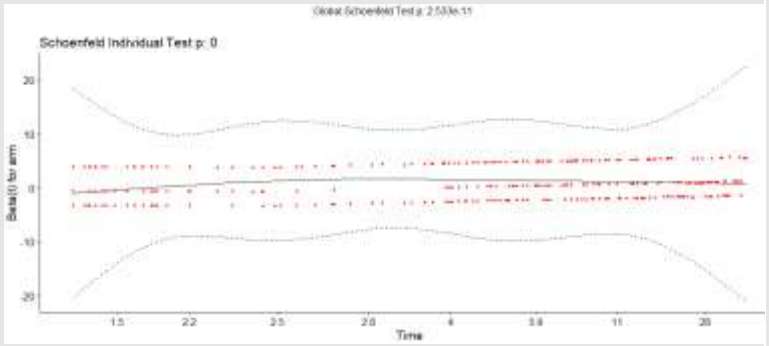
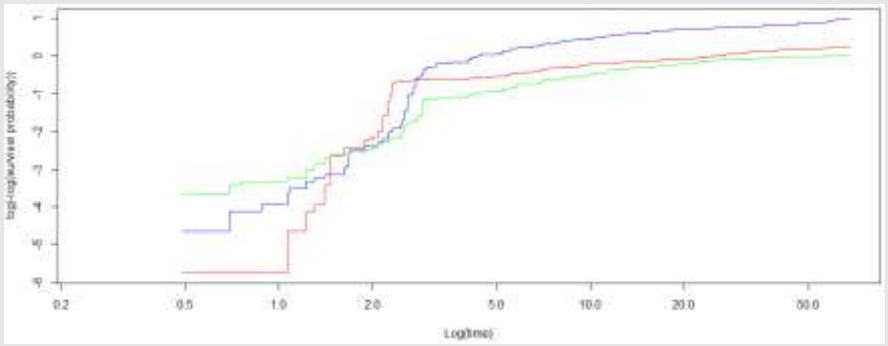
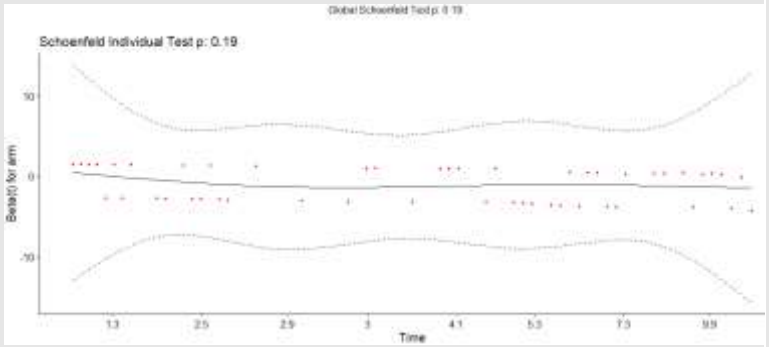
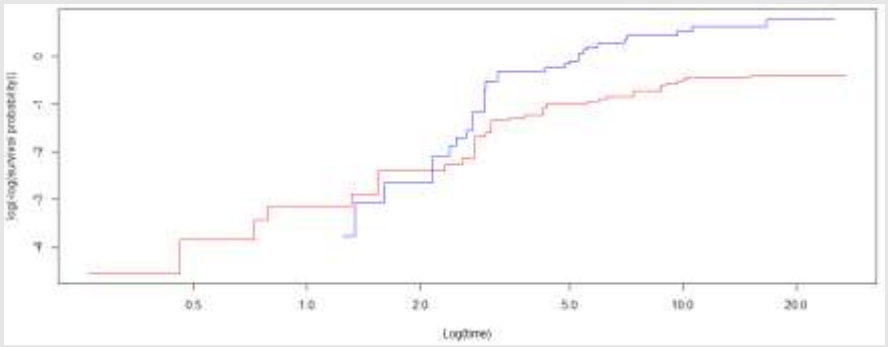
Trial	Network 1		Network 3		Network 4		Network 5		Network 6	
	PFS	OS	PFS	OS	PFS	OS	PFS	OS	PFS	OS
BREAK-3			-	-	-	-			-	-
BRF113220	✓	✓	-	-	-	-	✓	✓	-	-
BRIM-3			-	-	-	-	-	-	-	-
CheckMate 066										
CheckMate 067										
CheckMate 069			-	-			-	-	-	-
COLUMBUS			-	-	-	-			-	-
COMBI-d		✓	-	-	-	-		✓	-	-
COMBI-v	✓	✓	-	-	-	-	✓	✓	-	-
KEYNOTE-006	✓		✓		✓		✓		✓	
Trials where PH is met	3/10	3/10	1/3	0/3	1/4	0/4	3/8	3/8	1/3	0/3

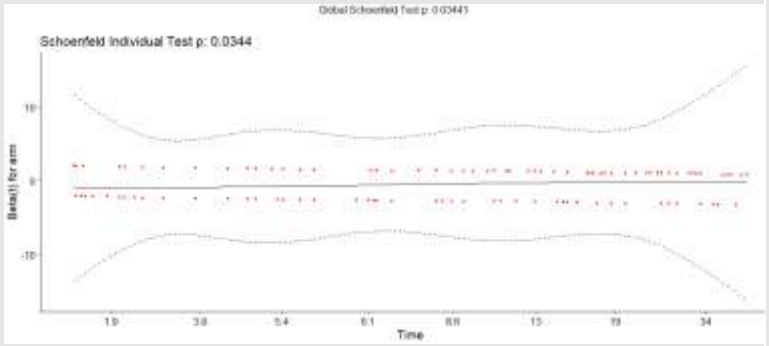
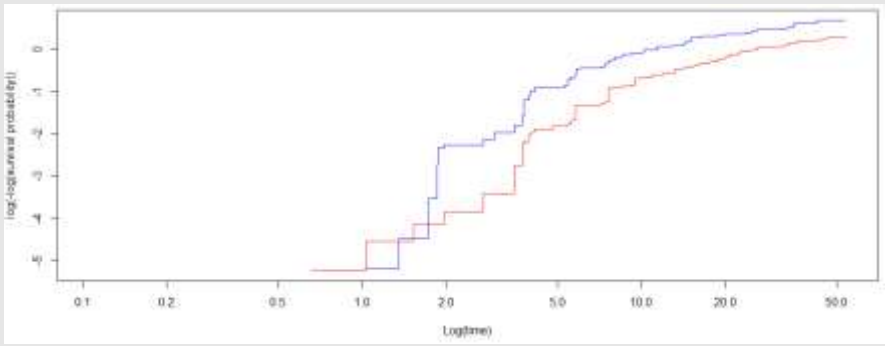
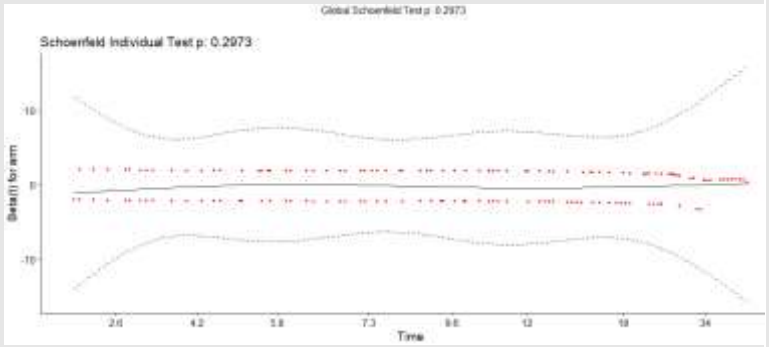
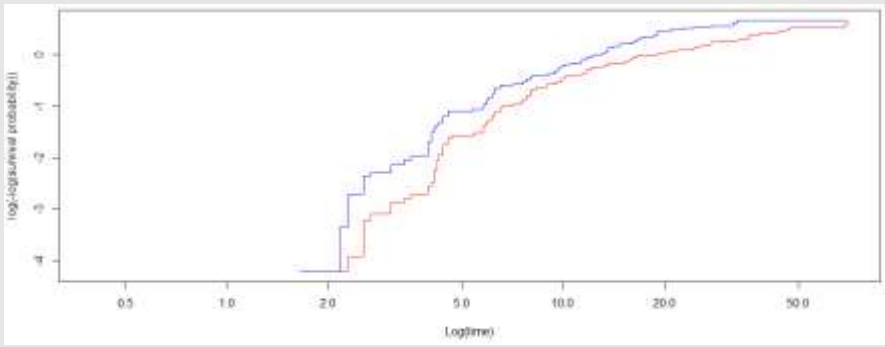
Cells shaded black indicate the proportional hazards assumption was not met. Cells shaded yellow with a check mark (✓) indicate the proportional hazards assumption was met. Cells with a dash (-) mean this trial was not included in that network.

Table 35: Tests of proportional hazards for progression-free survival in Network 1 (People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies)

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
BREAK-3		
8.7E-05		
BRF113220		
0.4		
BRIM-3		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
<2E-16		
CheckMate 066		
3.8E-06		
CheckMate 067		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
2.5E-11		
CheckMate 069		
0.19		
COLUMBUS		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
0.034		
COMBI-d		
0.3		
COMBI-v		

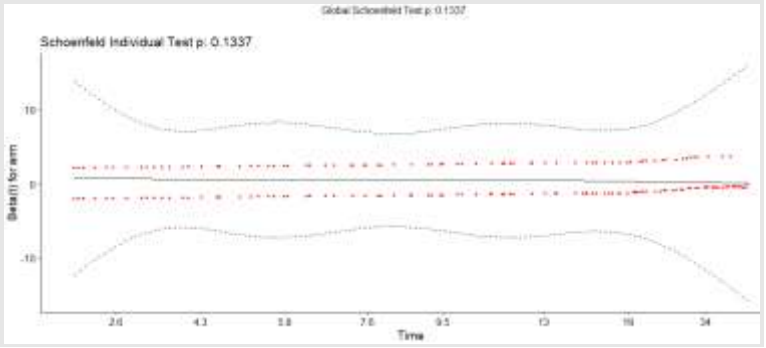
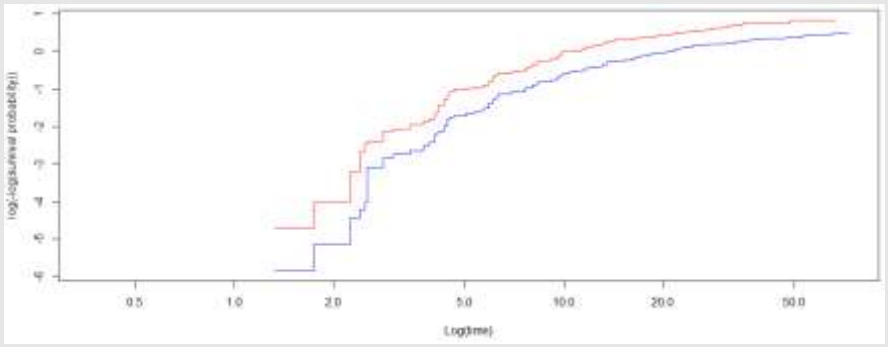
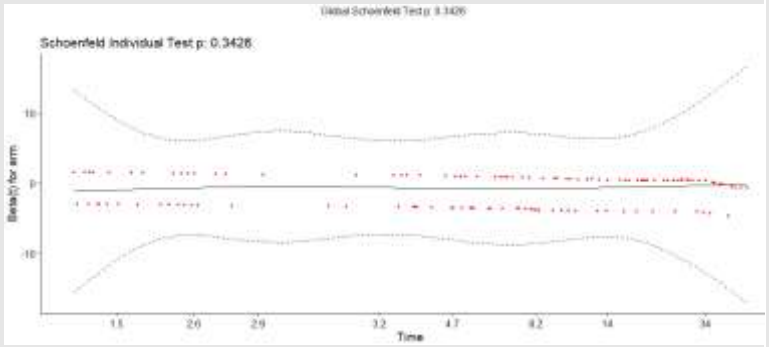
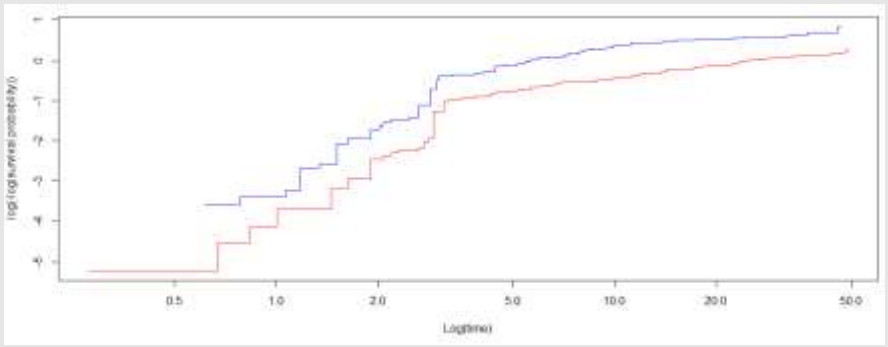
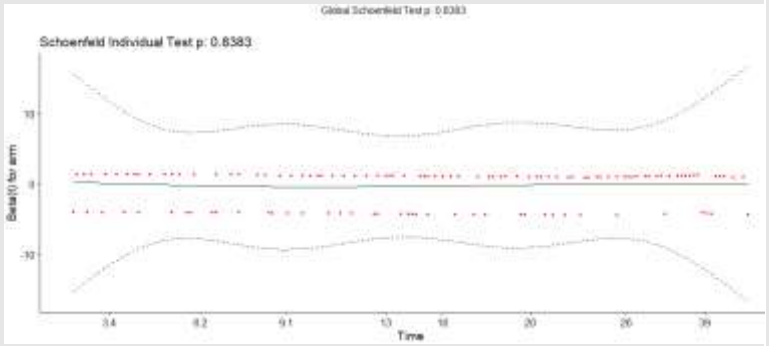
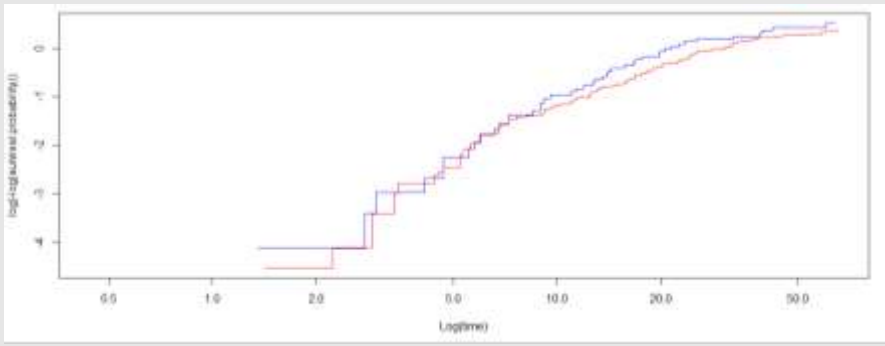
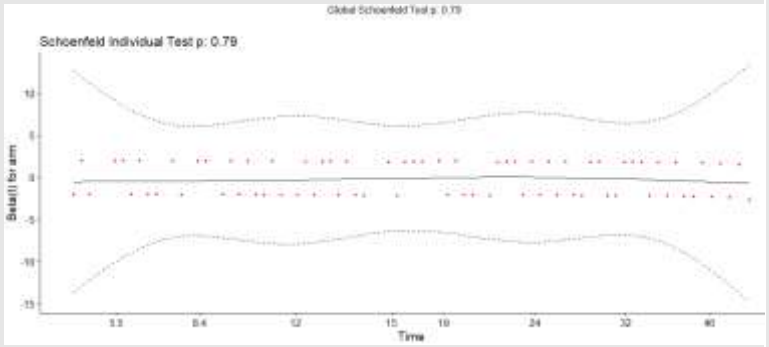
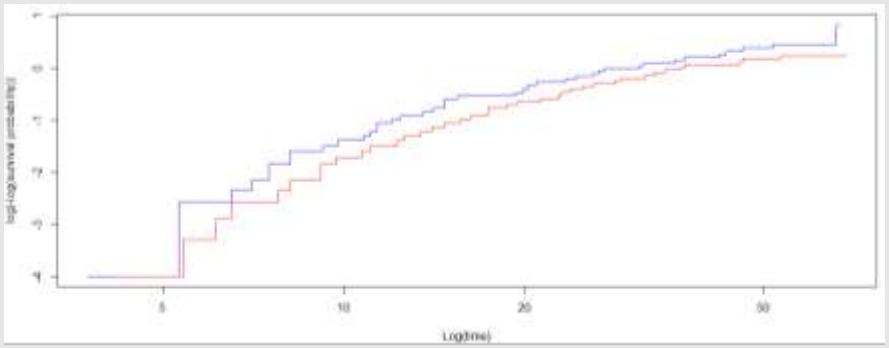
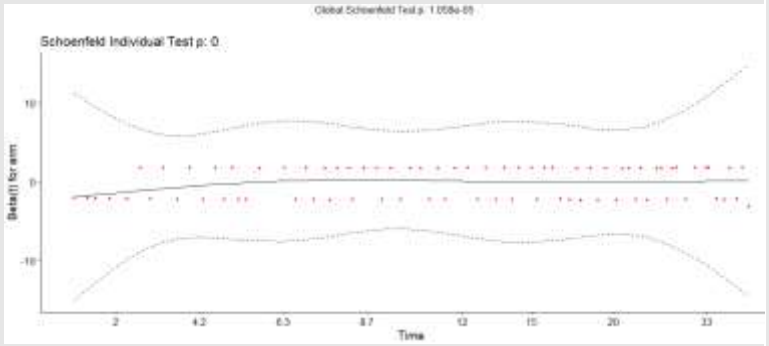
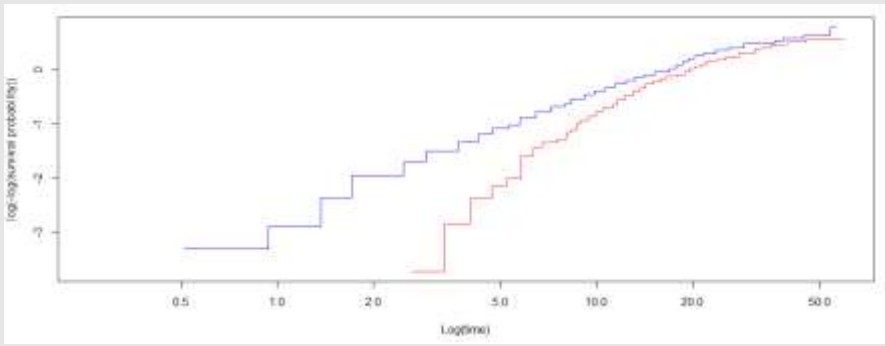
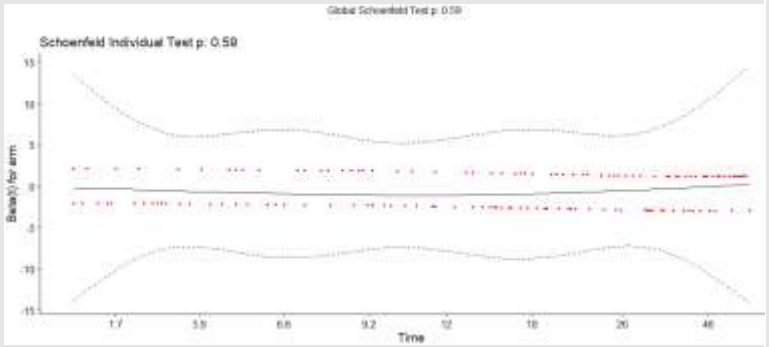
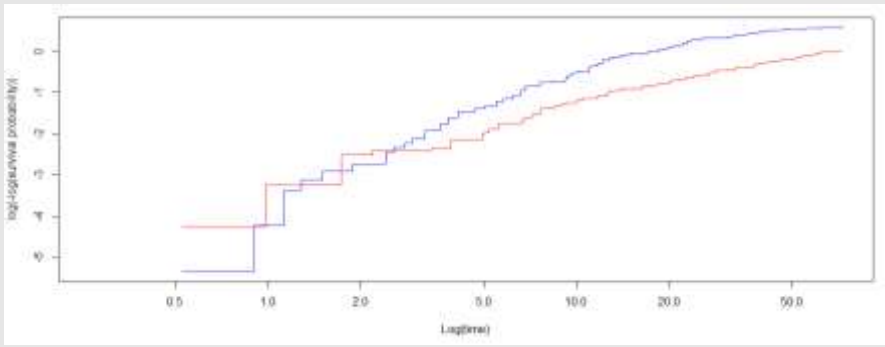
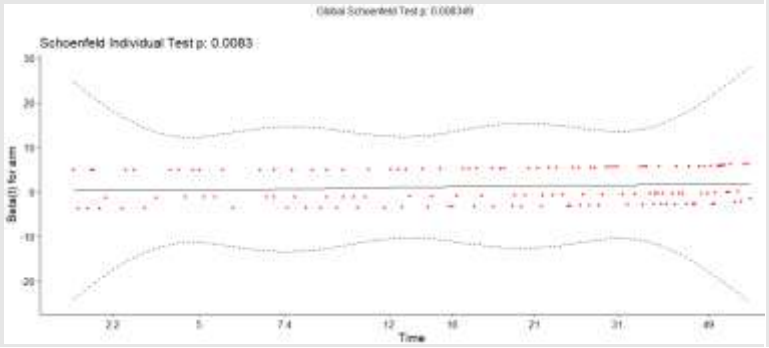
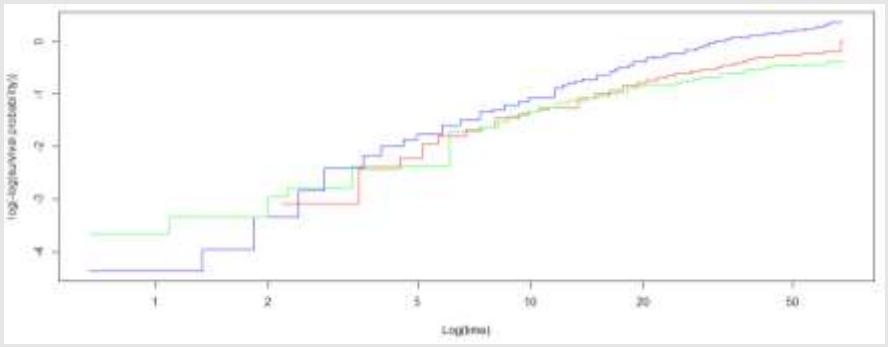
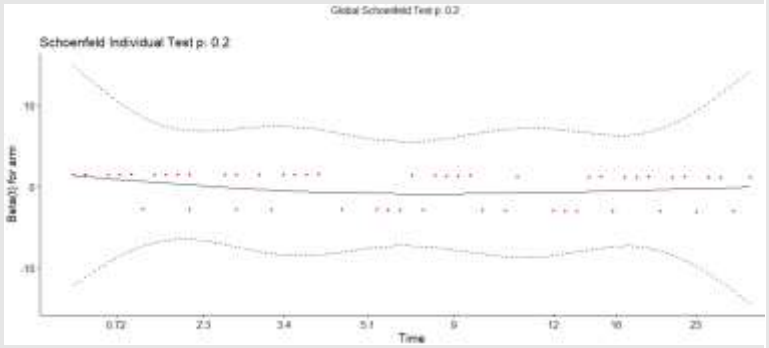
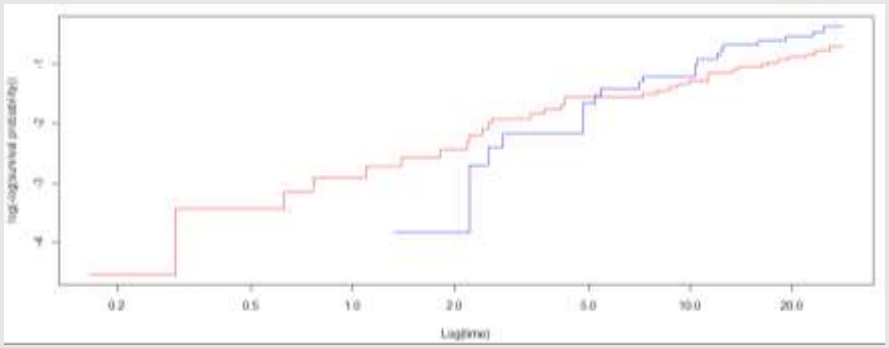
Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
0.13		
KEYNOTE-006		
0.34		

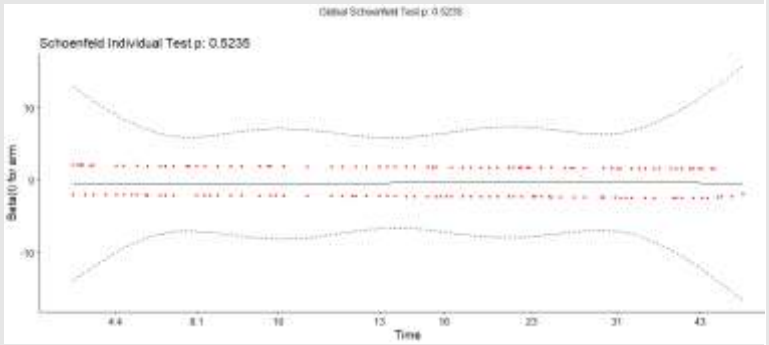
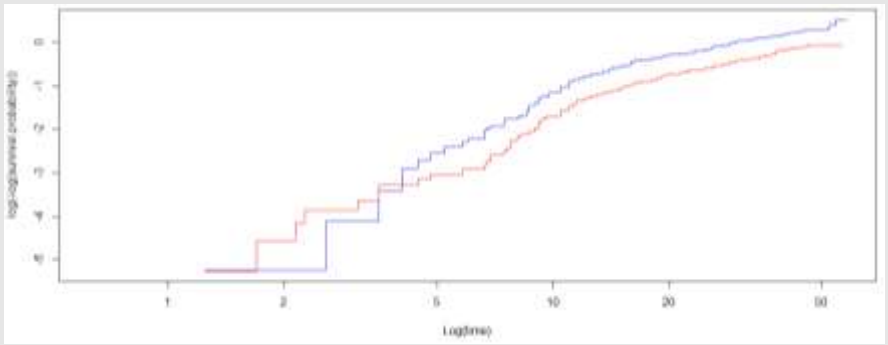
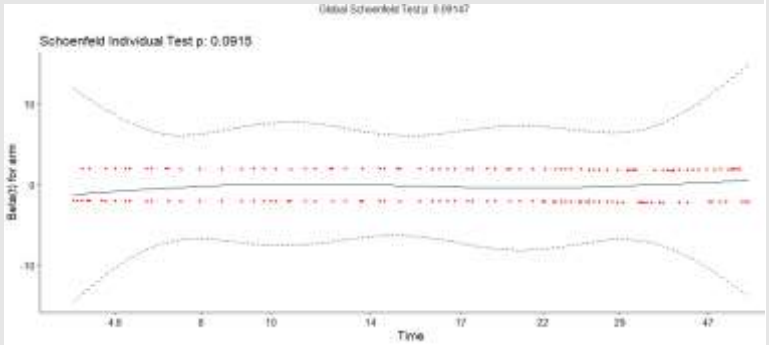
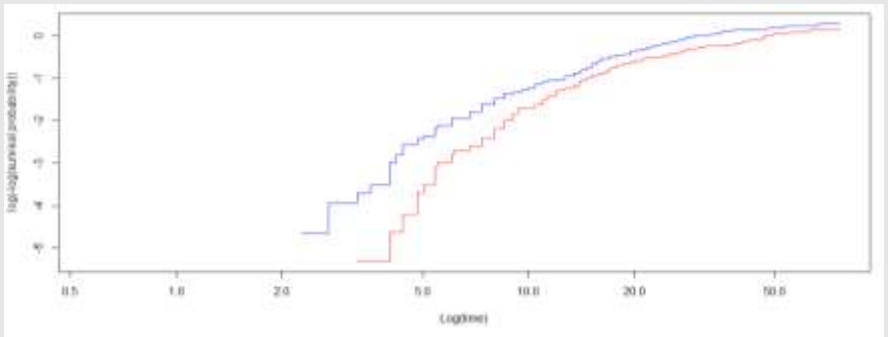
Table 36: Tests of proportional hazards for overall survival in Network 1 (People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies)

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
BREAK-3		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
0.84		
BRF113220		
0.79		
BRIM-3		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
1.1E-05		
CheckMate 066		
0.59		
CheckMate 067		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
0.0083		
CheckMate 069		
0.2		
COLUMBUS		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
0.52		
COMBI-d		
0.091		
COMBI-v		

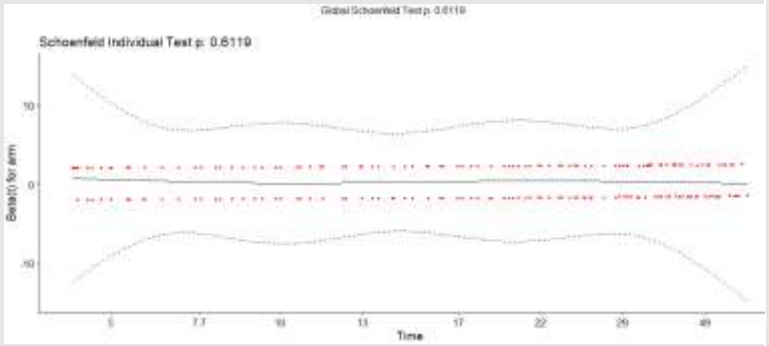
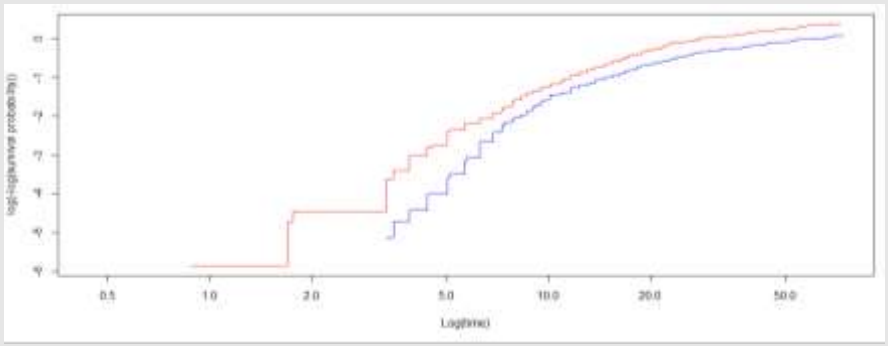
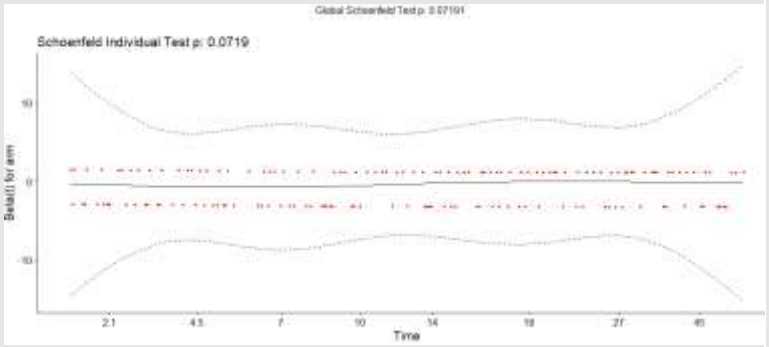
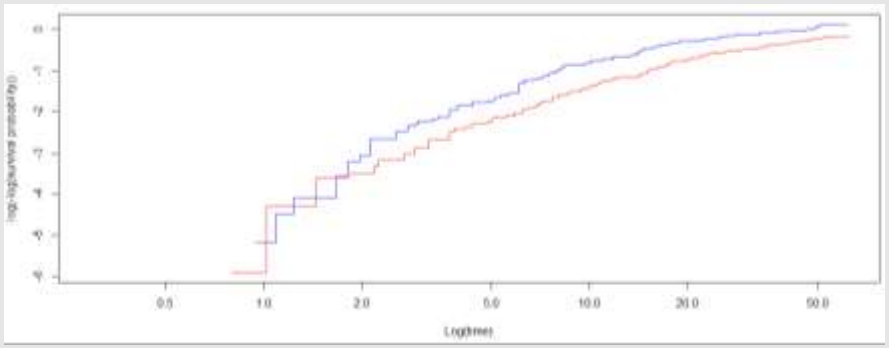
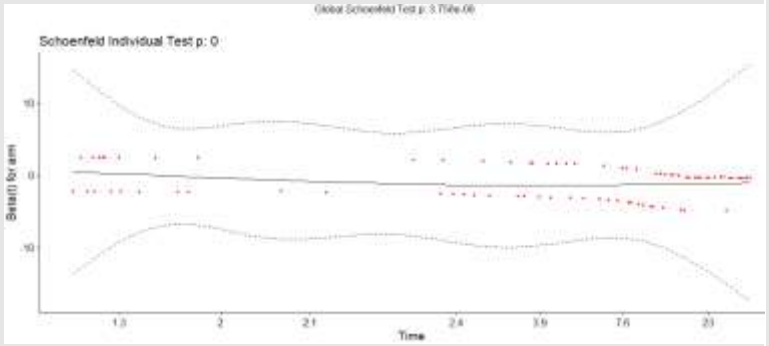
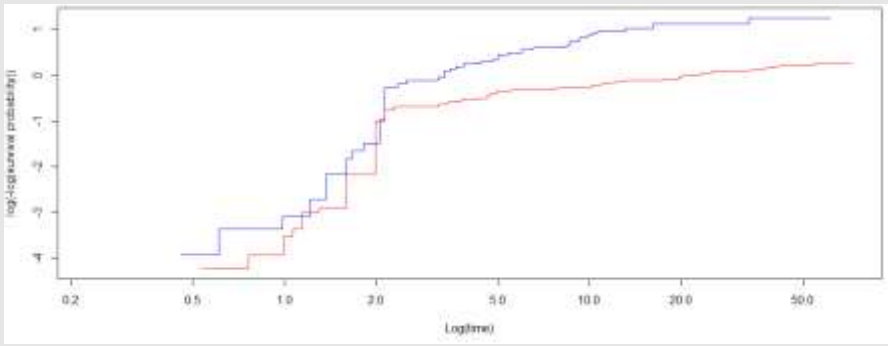
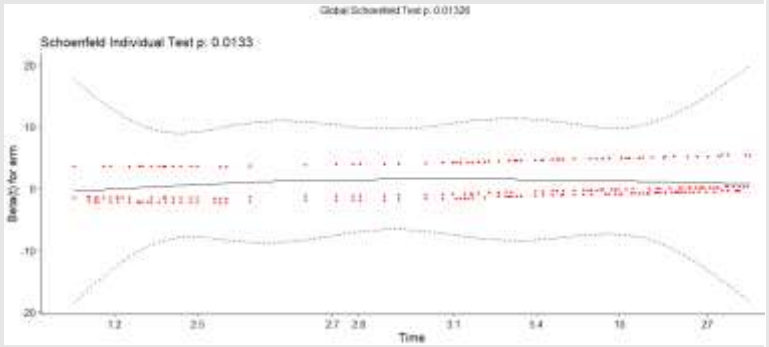
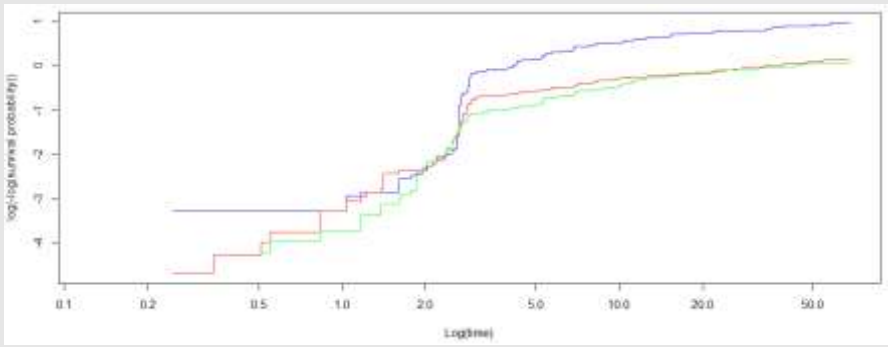
Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
0.61		
KEYNOTE-006		
0.072		

Table 37: Tests of proportional hazards for progression-free survival in Network 3 (*BRAF* wild type)

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
CheckMate 066		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
3.8E-06		
CheckMate 067		
0.013		
KEYNOTE-006		

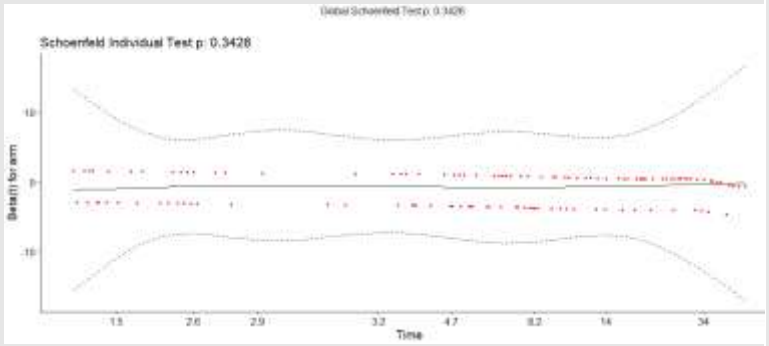
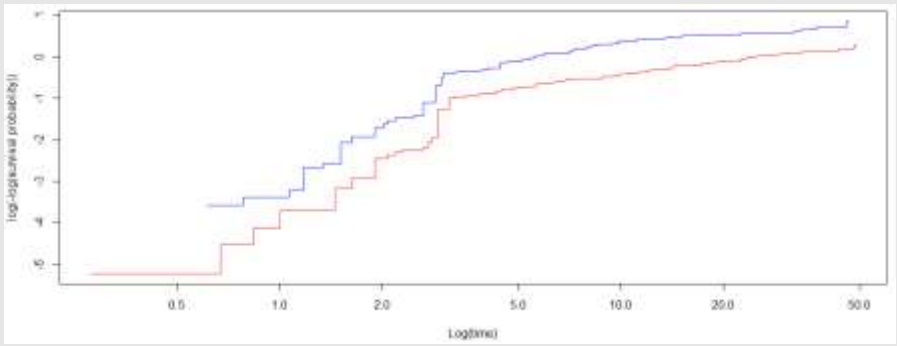
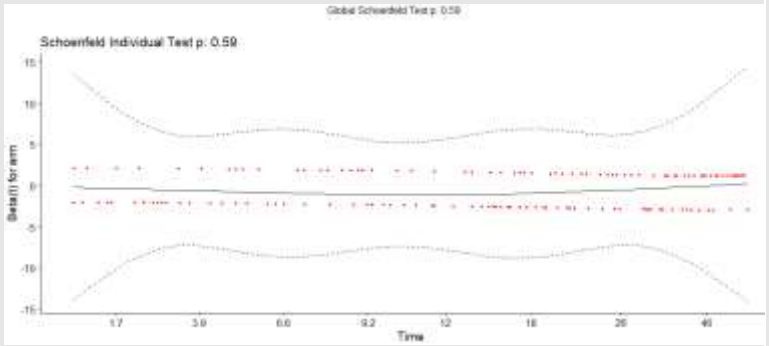
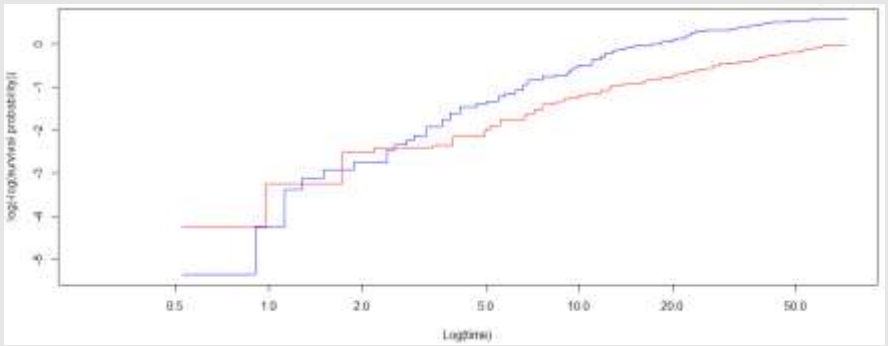
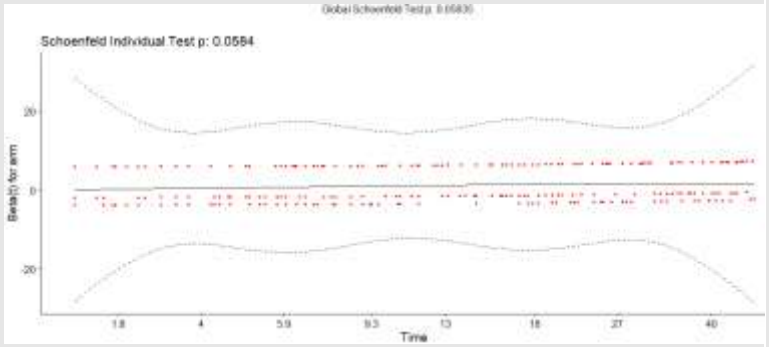
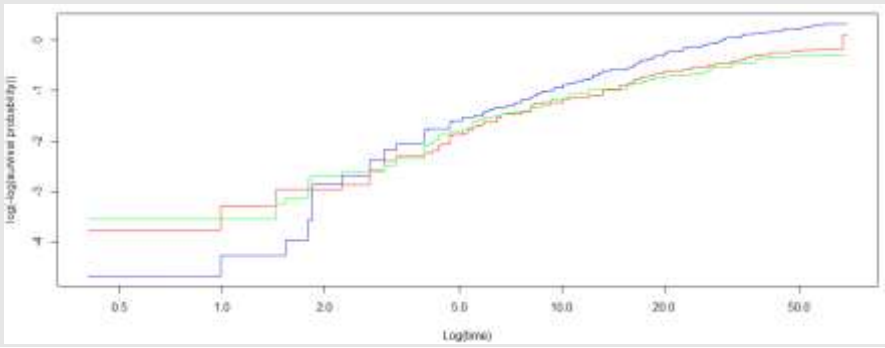
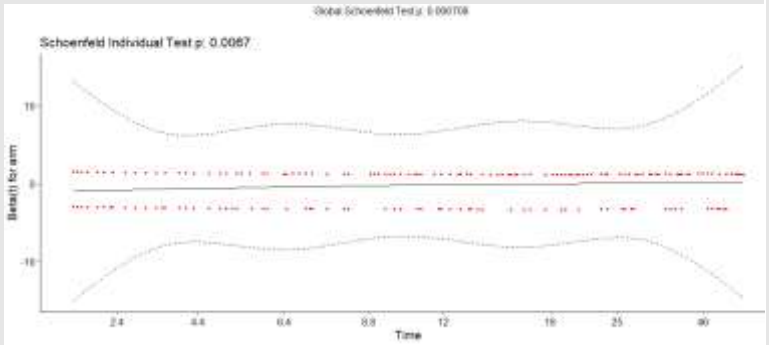
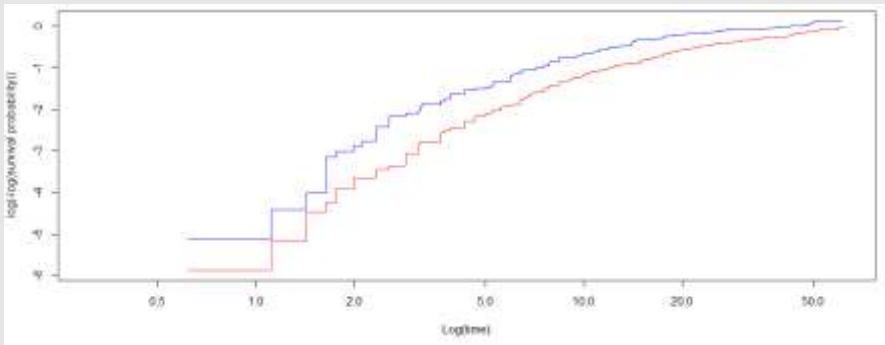
Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
0.34		

Table 38: Tests of proportional hazards for overall survival in Network 3 (*BRAF* wild type)

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
CheckMate 066		
0.59		
CheckMate 067		

Grambsch and Thernau (test p-value)	Schoenfeld residuals	log-log curves
0.058		
KEYNOTE-006		
0.0067		

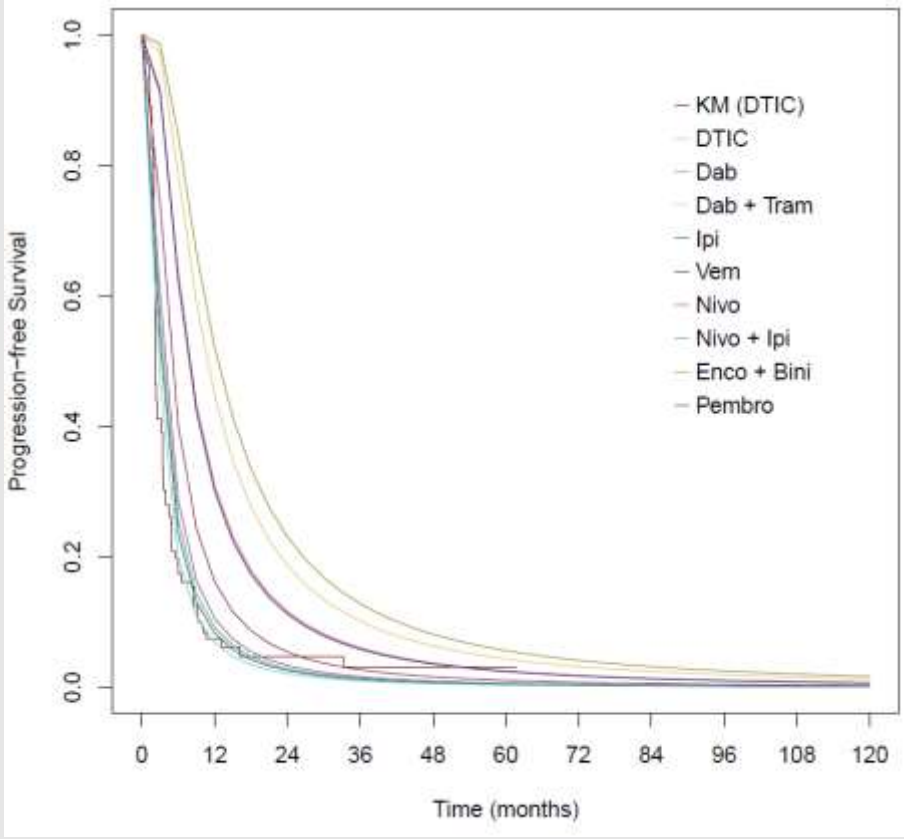
Appendix D: Generalized gamma

Comparing one treatment effect with two treatment effects

Although we successfully fit both the generalized gamma model with one treatment effect (location parameter alone) and two treatment effects (location and scale parameters), the two-treatment effect model was a better fit for all network for both PFS and OS. Goodness of fit was assessed both by examining the Akaike information criterion (AIC), where a smaller AIC value is indicative of a better fitting model, as well as visual inspection. Tables comparing the AIC values for the one treatment effect and two treatment effect models are in Appendix: D.1.1.1. Furthermore, side by side comparisons of the visual fit for the two models for network 1 are presented in Figure 55 and Figure 56. It is clear from these figures that the generalized gamma model with one treatment effect presents overly pessimistic projections for both PFS and OS that are not clinically plausible.

Figure 55: A comparison of survival curves for generalized gamma PFS models with one and two treatment effects for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Generalized gamma one treatment effect



Generalized gamma two treatment effects

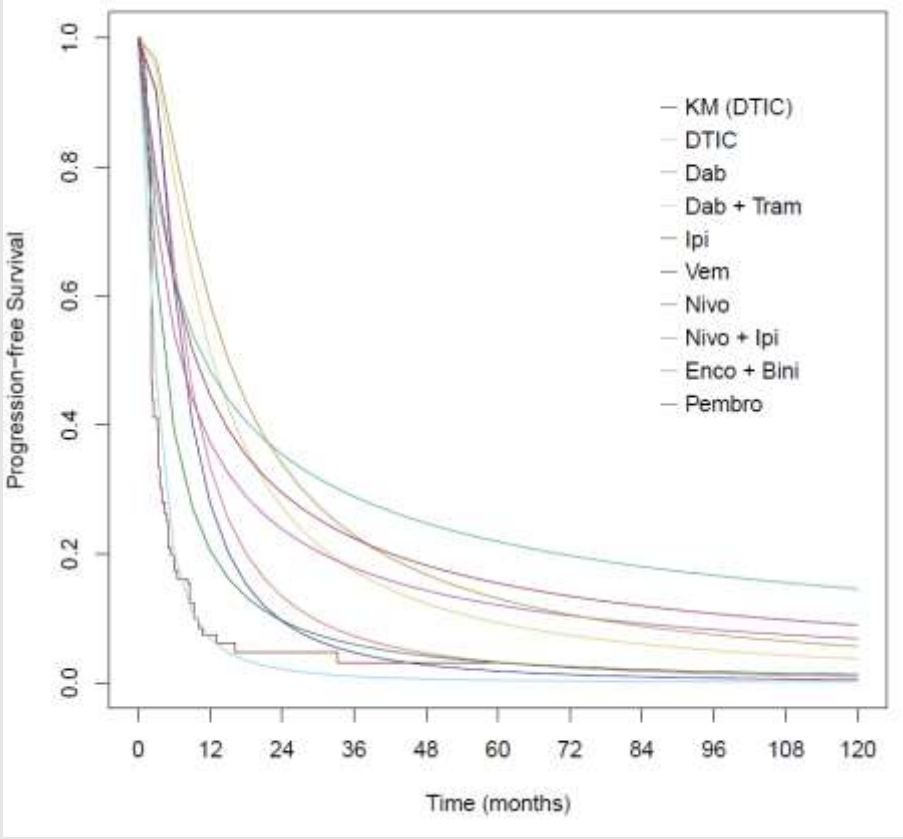
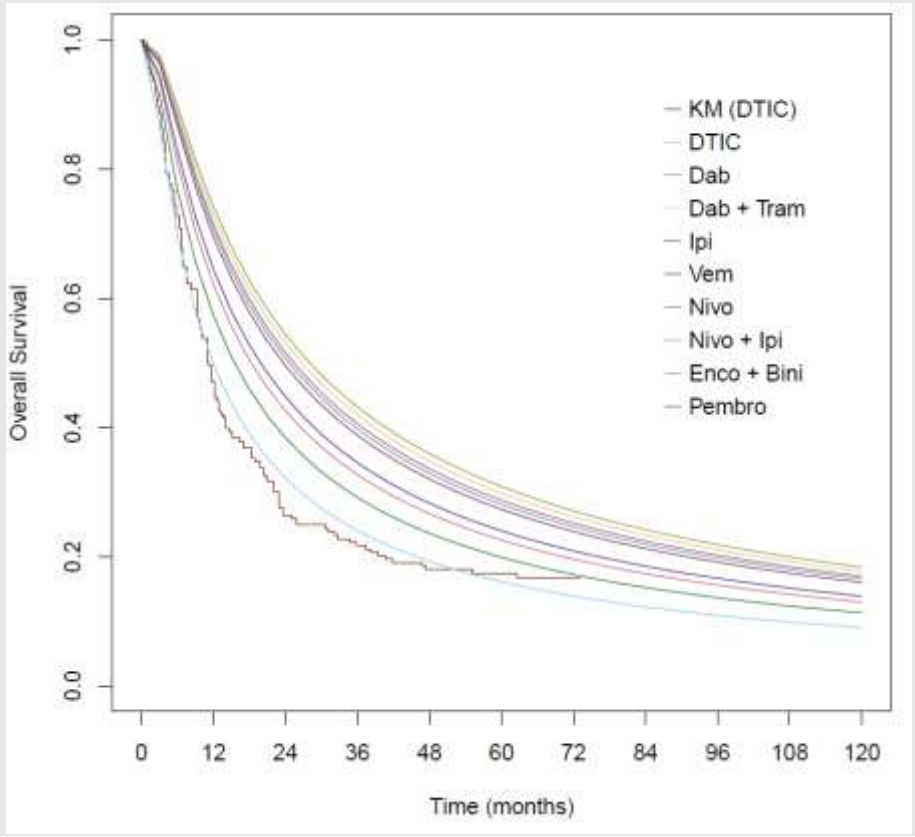
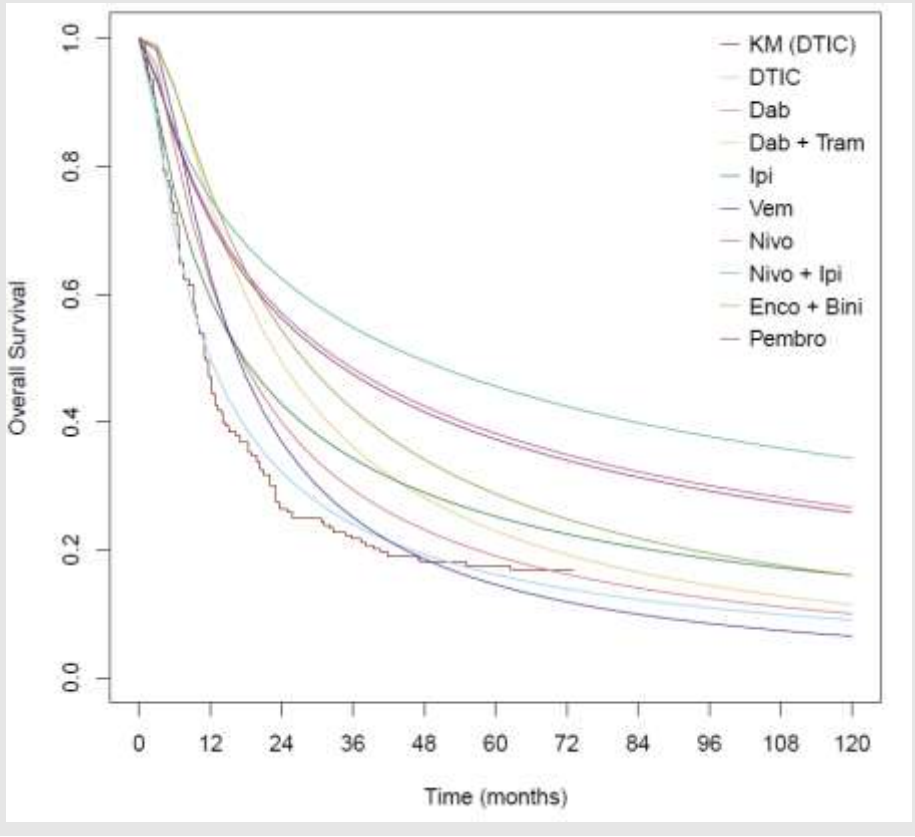


Figure 56: A comparison of survival curves for generalized gamma OS models with one and two treatments effect for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Generalized gamma one treatment effect



Generalized gamma two treatment effects



D.1.1.1 Comparisons of AIC values for the generalized gamma one and two treatment effect models

As seen in Table 39, Table 40, Table 41, Table 42 & Table 43, the generalized gamma model with two treatment effects had a lower AIC value for both PFS and OS for all networks versus the generalized gamma model with one treatment effects. Thus, based on AIC, the generalized gamma model with two treatment effects was deemed to be a better fitting model than the generalized gamma model with one treatment effect.

D.1.1.1.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Table 39: AIC values for the generalized gamma one and two treatment effect models for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Trial	PFS		OS	
	AIC for location parameter	AIC for location & scale parameters	AIC for location parameter	AIC for location & scale parameters
BREAK-3	**1296.2**	1298.1	**1558.5**	1560.1
BRF113220	582.6	**582.2**	**734.5**	736
BRIM-3	**2647.2**	2648.9	4267.6	**4230.1**
CheckMate 066	1855.6	**1809.8**	2600.3	**2580.7**
CheckMate 067	4604.6	**4507.8**	5371.5	**5338.6**
CheckMate 069	556.5	**521**	554.6	**549.9**
COLUMBUS	1736.7	**1730.9**	2262.5	**2259.4**
COMBI-d	**2199.7**	2200.5	**2496.7**	2498.5
COMBI-v	3541.8	**3533.5**	**4038**	4040
KEYNOTE-006	2873.9	**2848**	2988.3	**2987**
Total	21894.8	**21680.7**	26872.5	**26780.3**

*Smaller AIC values in **bold**.*

D.1.1.1.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

D.1.1.1.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 40: AIC values for the generalized gamma one and two treatment effect models for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Trial	PFS		OS	
	AIC for location parameter	AIC for location & scale parameters	AIC for location parameter	AIC for location & scale parameters
CheckMate 066	1855.6	**1809.8**	2600.3	**2580.7**
CheckMate 067	3182.2	**3153.5**	3732.9	**3712.2**

Trial	PFS		OS	
	AIC for location parameter	AIC for location & scale parameters	AIC for location parameter	AIC for location & scale parameters
KEYNOTE-006	2873.9	**2848**	**2895.3**	2897
Total	7911.6	**7811.4**	9228.5	**9189.8**

*Smaller AIC values in **bold**.*

D.1.1.1.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Table 41: AIC values for the generalized gamma one and two treatment effect models for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Trial	PFS		OS	
	AIC for location parameter	AIC for location & scale parameters	AIC for location parameter	AIC for location & scale parameters
CheckMate 066	1855.6	**1809.8**	2600.3	**2580.7**
CheckMate 067	4604.6	**4507.8**	5371.5	**5338.6**
CheckMate 069	556.5	**521**	554.6	**549.9**
KEYNOTE-006	2873.9	**2848**	2988.3	**2987**
Total	9890.5	**9686.6**	11514.6	**11456.2**

*Smaller AIC values in **bold**.*

D.1.1.1.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Table 42: AIC values for the generalized gamma one and two treatment effect models for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Trial	PFS		OS	
	AIC for location parameter	AIC for location & scale parameters	AIC for location parameter	AIC for location & scale parameters
BREAK-3	1296.2	1298.1	**1558.5**	1560.1
BRF113220	582.6	**582.2**	**734.5**	736
CheckMate 066	1855.6	**1809.8**	2600.3	**2580.7**
CheckMate 067	4604.6	**4507.8**	5371.5	**5338.6**
COLUMBUS	1736.7	**1730.9**	2262.5	**2259.4**
COMBI-d	2199.7	2200.5	**2496.7**	2498.5
COMBI-v	3541.8	**3533.5**	**4038**	4040
KEYNOTE-006	2873.9	**2848**	2988.3	**2987**
Total	18691.1	**18510.9**	22050.2	**22000.3**

*Smaller AIC values in **bold**.*

D.1.1.1.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Table 43: AIC values for the generalized gamma one and two treatment effect models for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Trial	PFS		OS	
	AIC for location parameter	AIC for location & scale parameters	AIC for location parameter	AIC for location & scale parameters
CheckMate 066	1855.6	**1809.8**	2600.3	**2580.7**
CheckMate 067	4604.6	**4507.8**	5371.5	**5338.6**
KEYNOTE-006	2873.9	**2848**	2988.3	**2987**
Total	9334	**9165.6**	10960	**10906.2**

*Smaller AIC values in **bold**.*

D.1.1.2 One treatment effect (location parameter alone) – NMA results

D.1.1.2.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Table 44: Fixed effect PFS & OS NMA results for generalized gamma with location parameter alone for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Comparison	PFS		OS	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
dabrafenib vs. DTIC	2.793	(2.359; 3.283)	1.547	(1.21; 1.946)
dabrafenib + trametinib vs. DTIC	3.837	(3.274; 4.468)	2.26	(1.79; 2.814)
ipilimumab vs. DTIC	1.138	(0.869; 1.467)	1.339	(0.83; 2.051)
vemurafenib vs. DTIC	2.729	(2.428; 3.055)	1.674	(1.37; 2.019)
nivolumab vs. DTIC	1.316	(1.084; 1.581)	2.166	(1.497; 3.026)
nivolumab + ipilimumab vs. DTIC	1.211	(0.911; 1.575)	2.125	(1.296; 3.285)
encorafenib + binimetinib vs. DTIC	4.492	(3.38; 5.852)	2.386	(1.704; 3.23)
pembrolizumab vs. DTIC	1.764	(1.209; 2.494)	2.067	(1.142; 3.461)

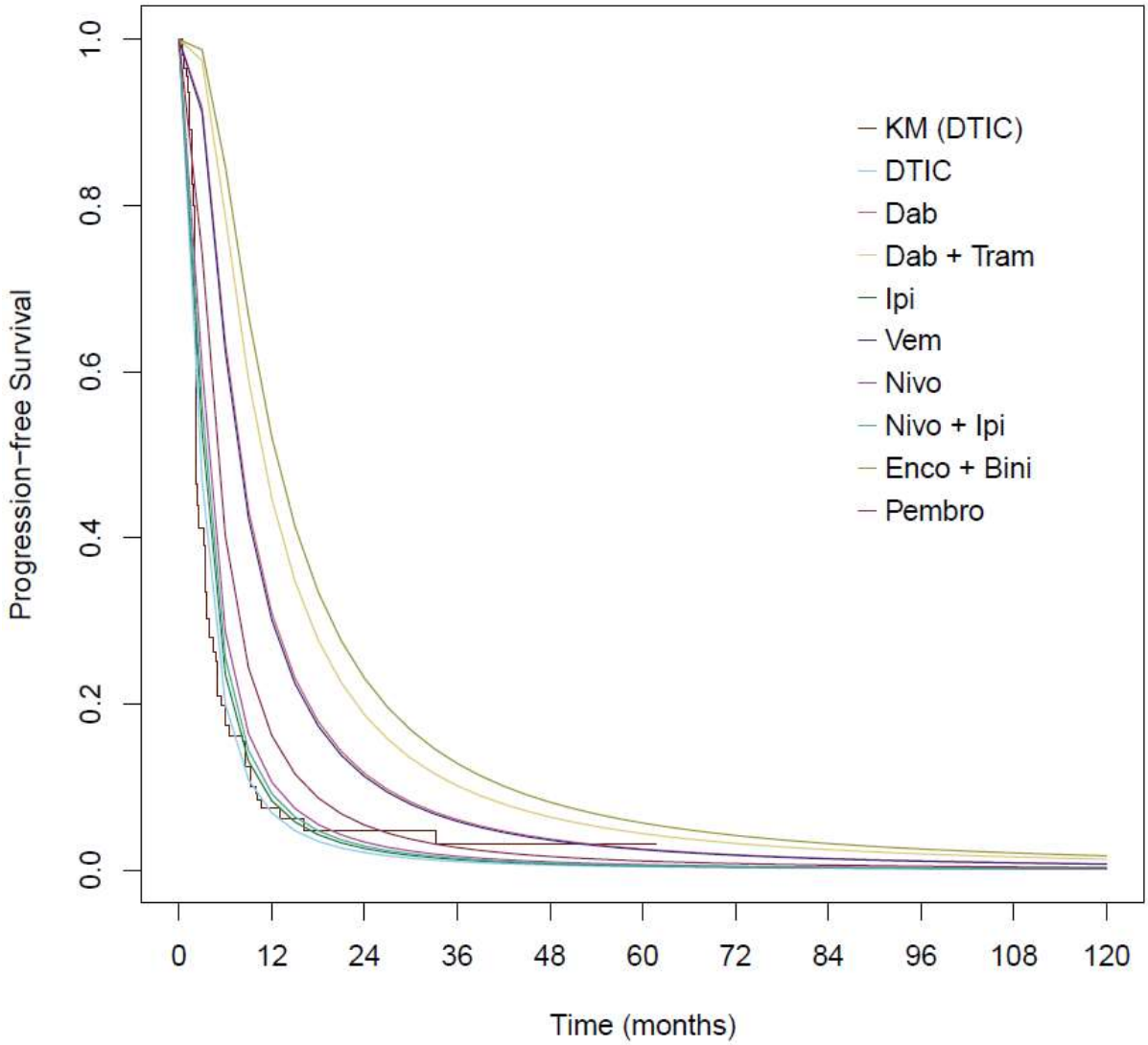


Figure 57: Survival curves for generalized gamma PFS model with one treatment effect for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

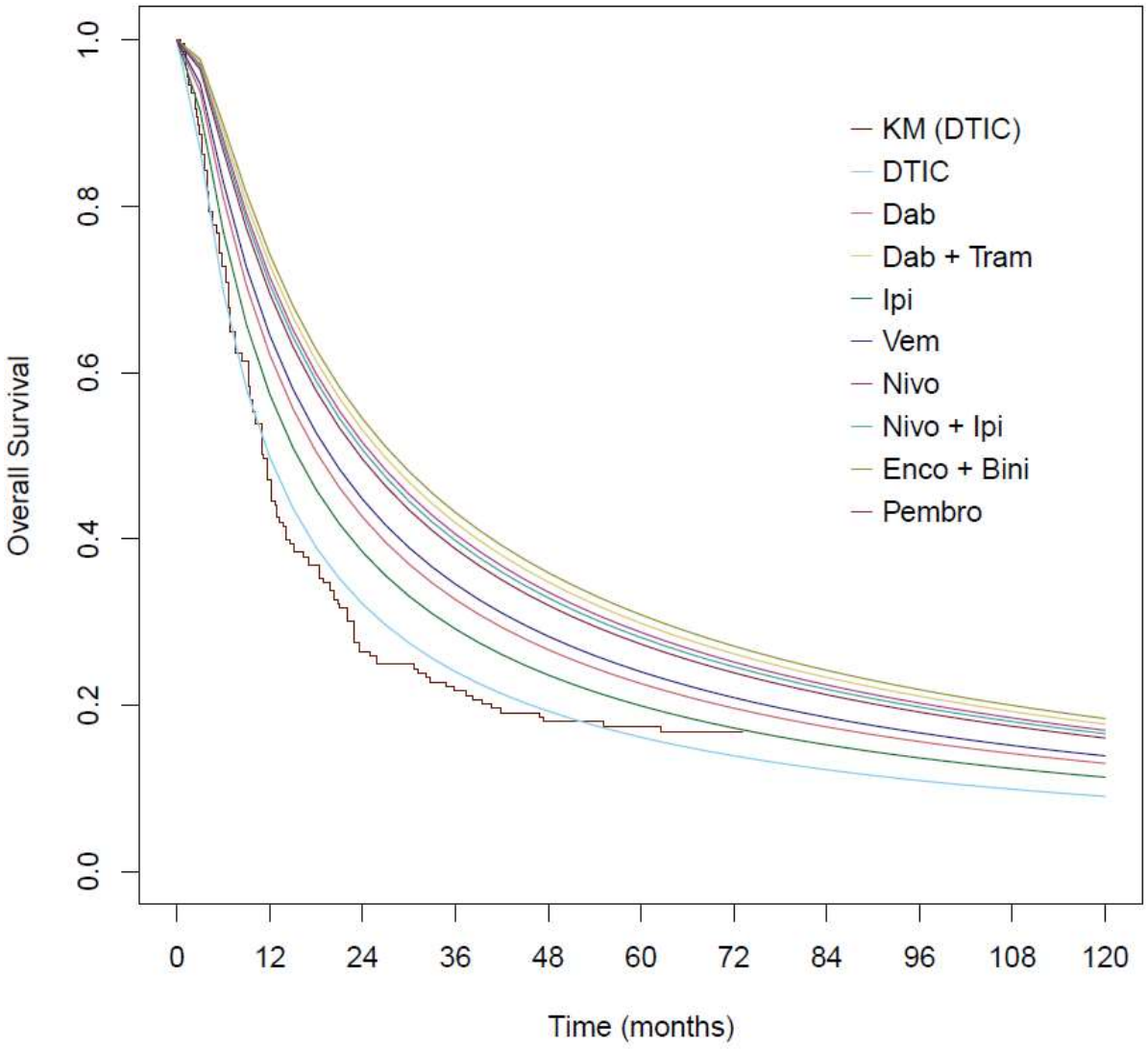


Figure 58: Survival curves for generalized gamma OS model with one treatment effect for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

D.1.1.2.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

D.1.1.2.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 45: Fixed effect PFS & OS NMA results for generalized gamma with location parameters for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Comparison	PFS		OS	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
ipilimumab vs. DTIC	0.836	(0.582; 1.163)	1.478	(0.835; 2.419)
nivolumab vs. DTIC	1.319	(1.083; 1.588)	2.179	(1.498; 3.063)
nivolumab + ipilimumab vs. DTIC	1.727	(1.207; 2.399)	2.361	(1.35; 3.85)
pembrolizumab vs. DTIC	1.298	(0.824; 1.933)	2.406	(1.24; 4.176)

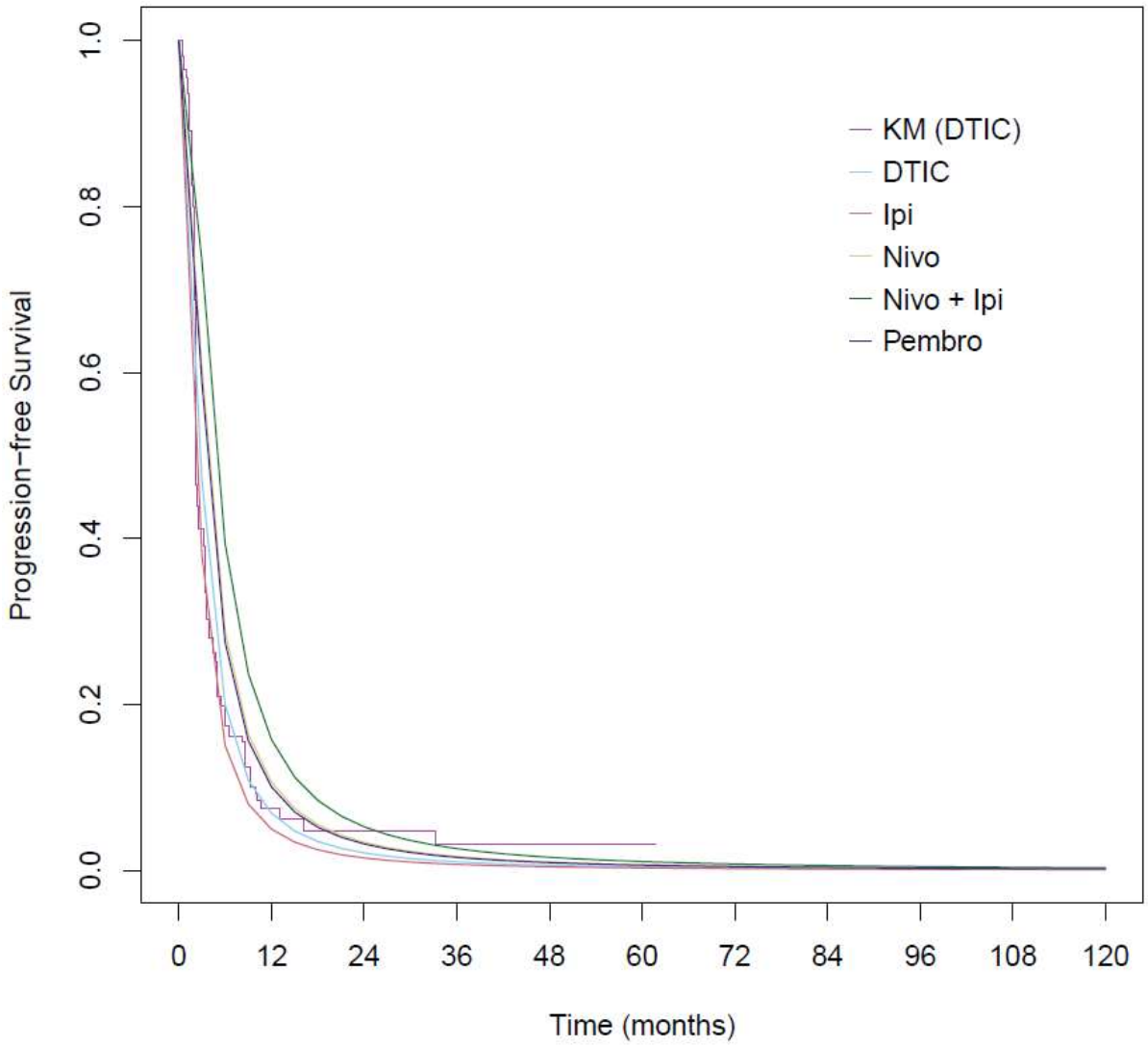


Figure 59: Survival curves for generalized gamma PFS model with one treatment effect for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

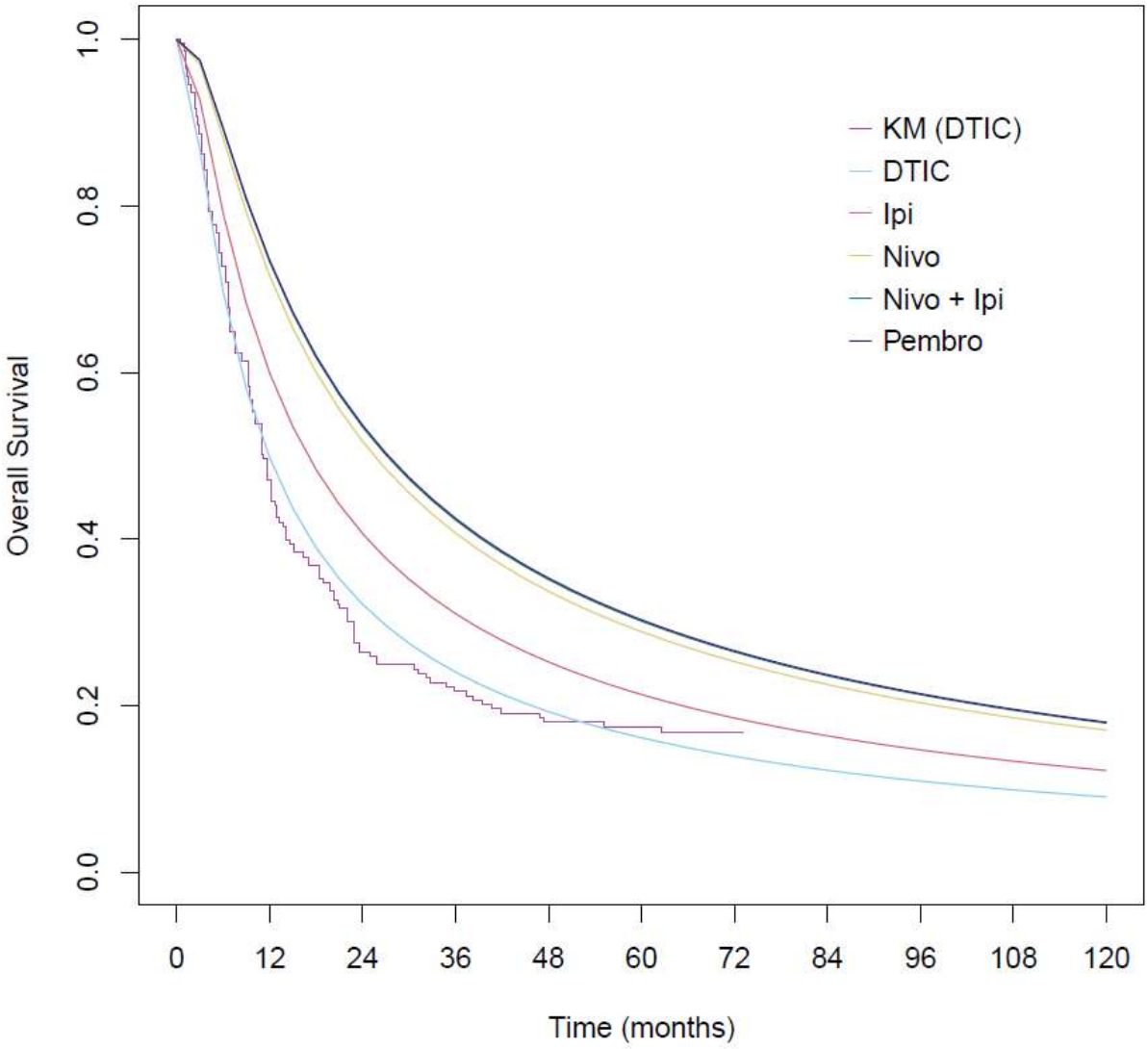


Figure 60: Survival curves for generalized gamma OS model with one treatment effect for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

D.1.1.2.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Table 46: Fixed effect PFS & OS NMA results for generalized gamma with location parameters for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Comparison	PFS		OS	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
ipilimumab vs. DTIC	1.142	(0.869; 1.474)	1.351	(0.828; 2.078)
nivolumab vs. DTIC	1.319	(1.082; 1.587)	2.179	(1.495; 3.049)

Comparison	PFS		OS	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	1.216	(0.912; 1.587)	2.142	(1.302; 3.326)
pembrolizumab vs. DTIC	1.773	(1.207; 2.498)	2.089	(1.138; 3.483)

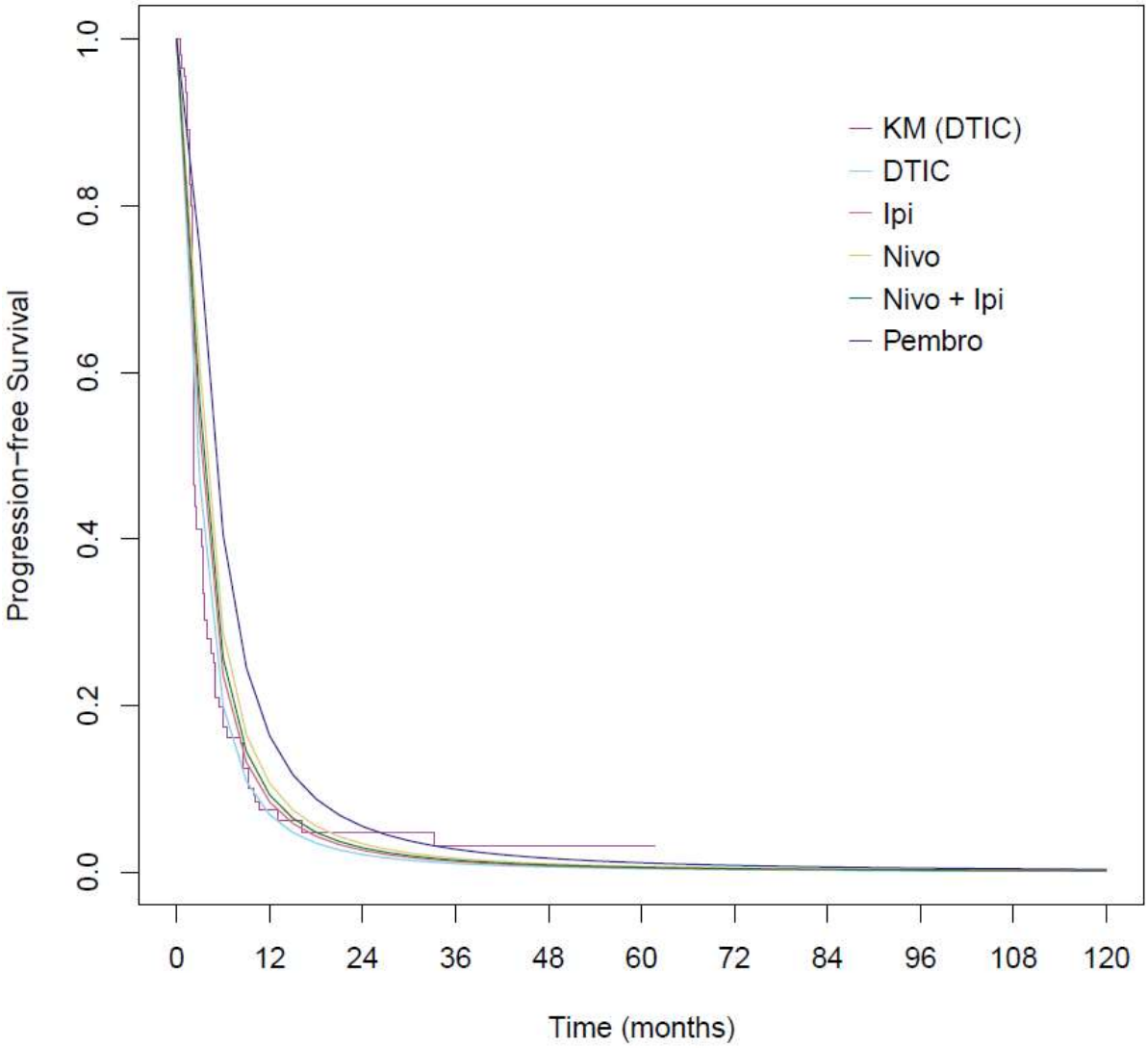


Figure 61: Survival curves for generalized gamma PFS model with one treatment effect for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

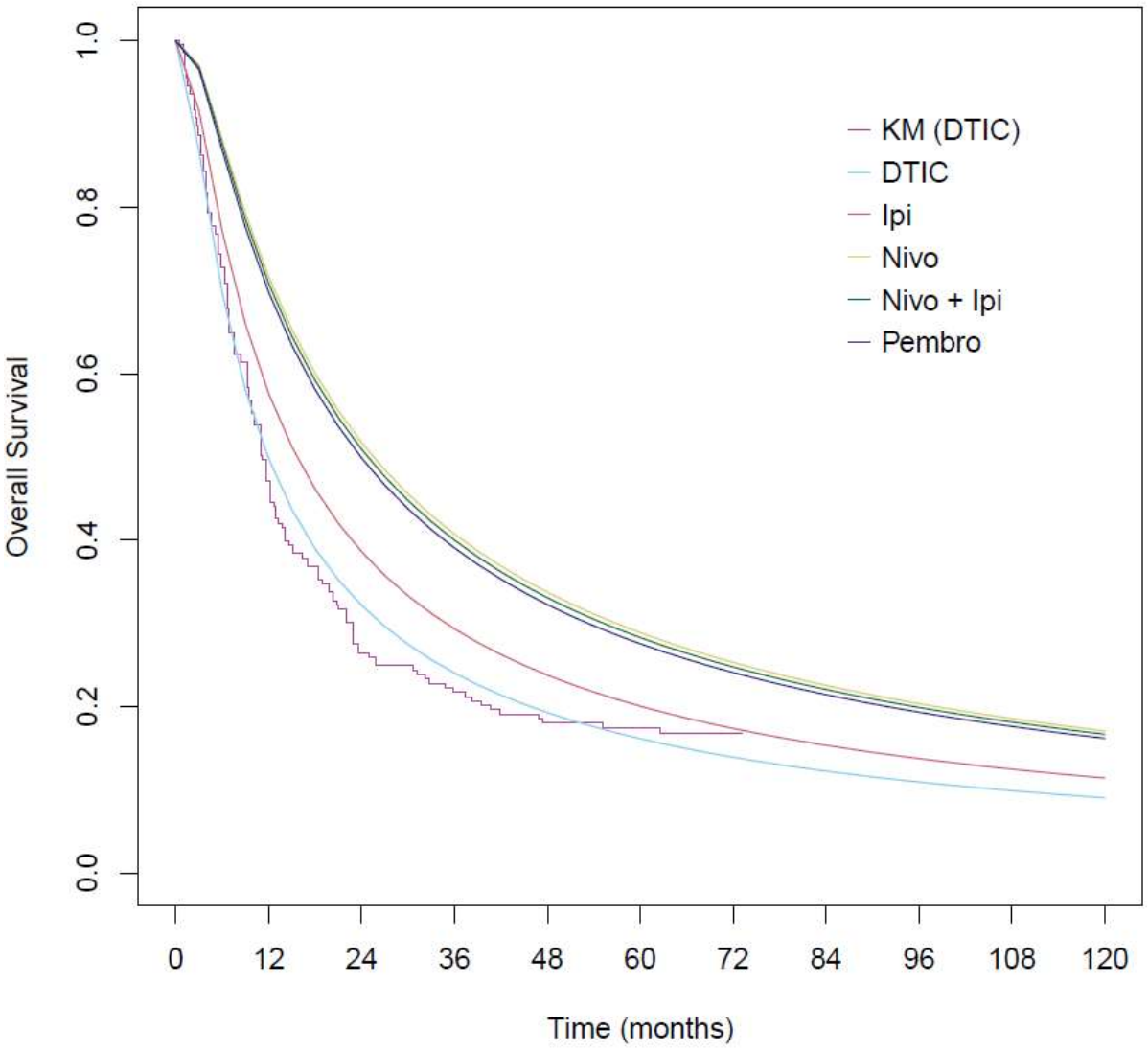


Figure 62: Survival curves for generalized gamma OS model with one treatment effect for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

D.1.1.2.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Table 47: Fixed effect PFS & OS NMA results for generalized gamma with location parameters for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Comparison	PFS		OS	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
dabrafenib vs. DTIC	3.055	(2.393; 3.862)	1.161	(0.807; 1.631)

Comparison	PFS		OS	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
dabrafenib + trametinib vs. DTIC	4.325	(3.257; 5.648)	1.584	(1.054; 2.306)
ipilimumab vs. DTIC	1.149	(0.878; 1.486)	1.324	(0.817; 2.04)
vemurafenib vs. DTIC	3.186	(2.321; 4.296)	1.096	(0.703; 1.649)
nivolumab vs. DTIC	1.316	(1.082; 1.585)	2.168	(1.495; 3.051)
nivolumab + ipilimumab vs. DTIC	1.196	(0.899; 1.564)	2.165	(1.318; 3.378)
encorafenib + binimetinib vs. DTIC	5.246	(3.462; 7.654)	1.564	(0.927; 2.507)
pembrolizumab vs. DTIC	1.781	(1.224; 2.509)	2.042	(1.13; 3.423)

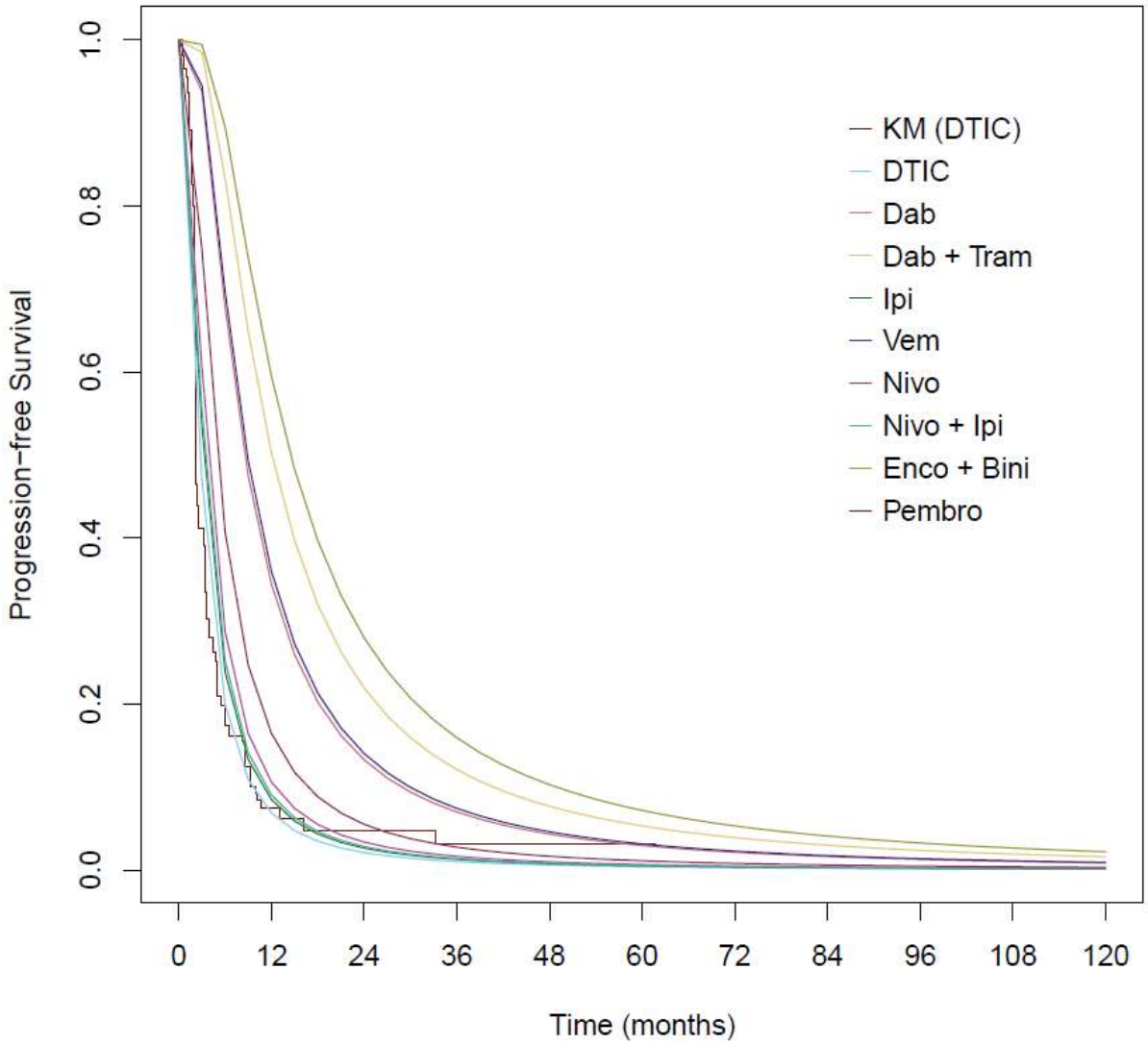


Figure 63: Survival curves for generalized gamma PFS model with one treatment effect for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

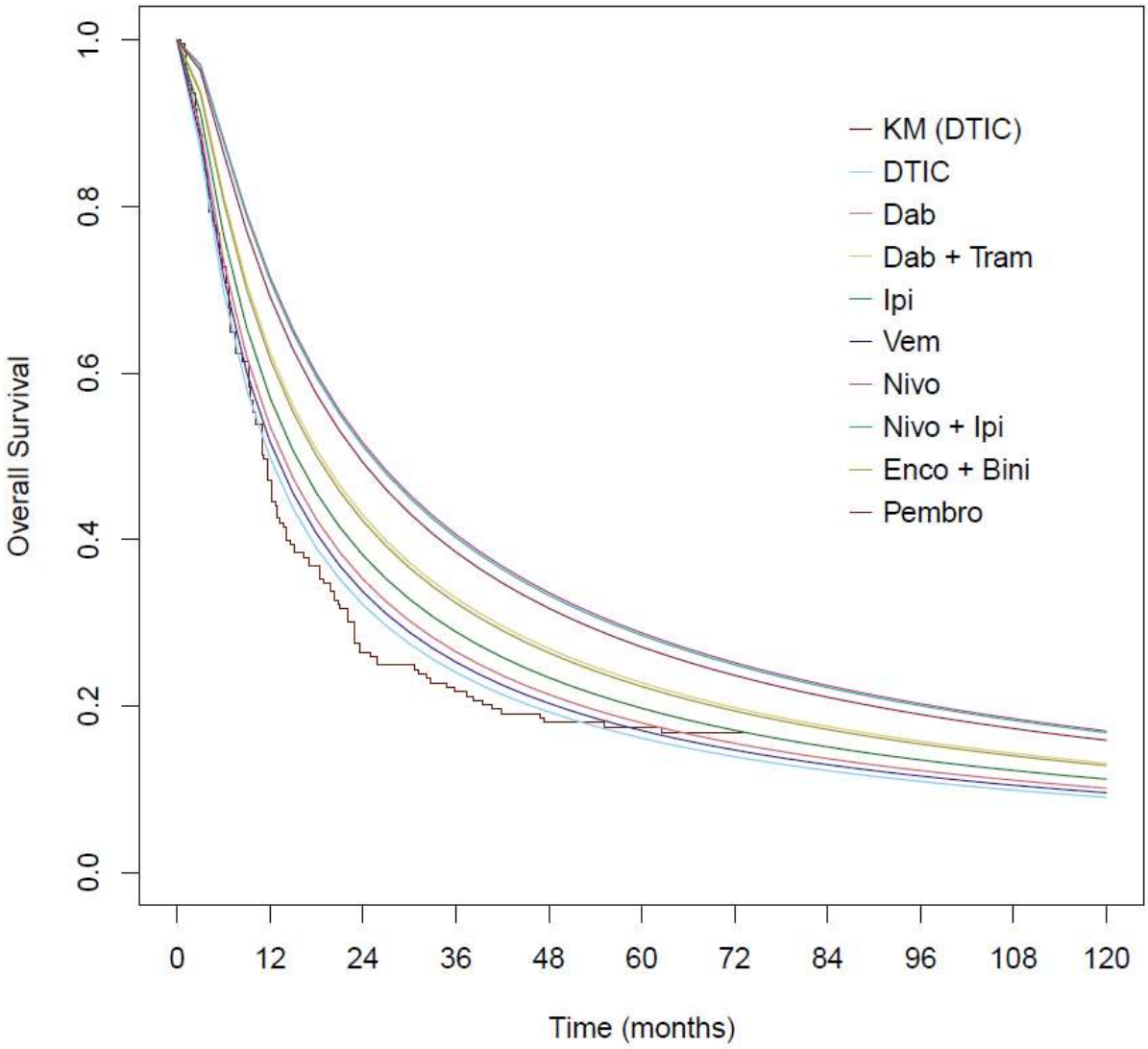


Figure 64: Survival curves for generalized gamma OS model with one treatment effect for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

D.1.1.2.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Table 48: Fixed effect PFS & OS NMA results for generalized gamma with location parameters for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Comparison	PFS		OS	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
ipilimumab vs. DTIC	1.153	(0.876; 1.49)	1.333	(0.815; 2.059)
nivolumab vs. DTIC	1.319	(1.081; 1.588)	2.179	(1.495; 3.052)

Comparison	PFS		OS	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	1.2	(0.899; 1.568)	2.179	(1.32; 3.394)
pembrolizumab vs. DTIC	1.79	(1.217; 2.523)	2.061	(1.121; 3.439)

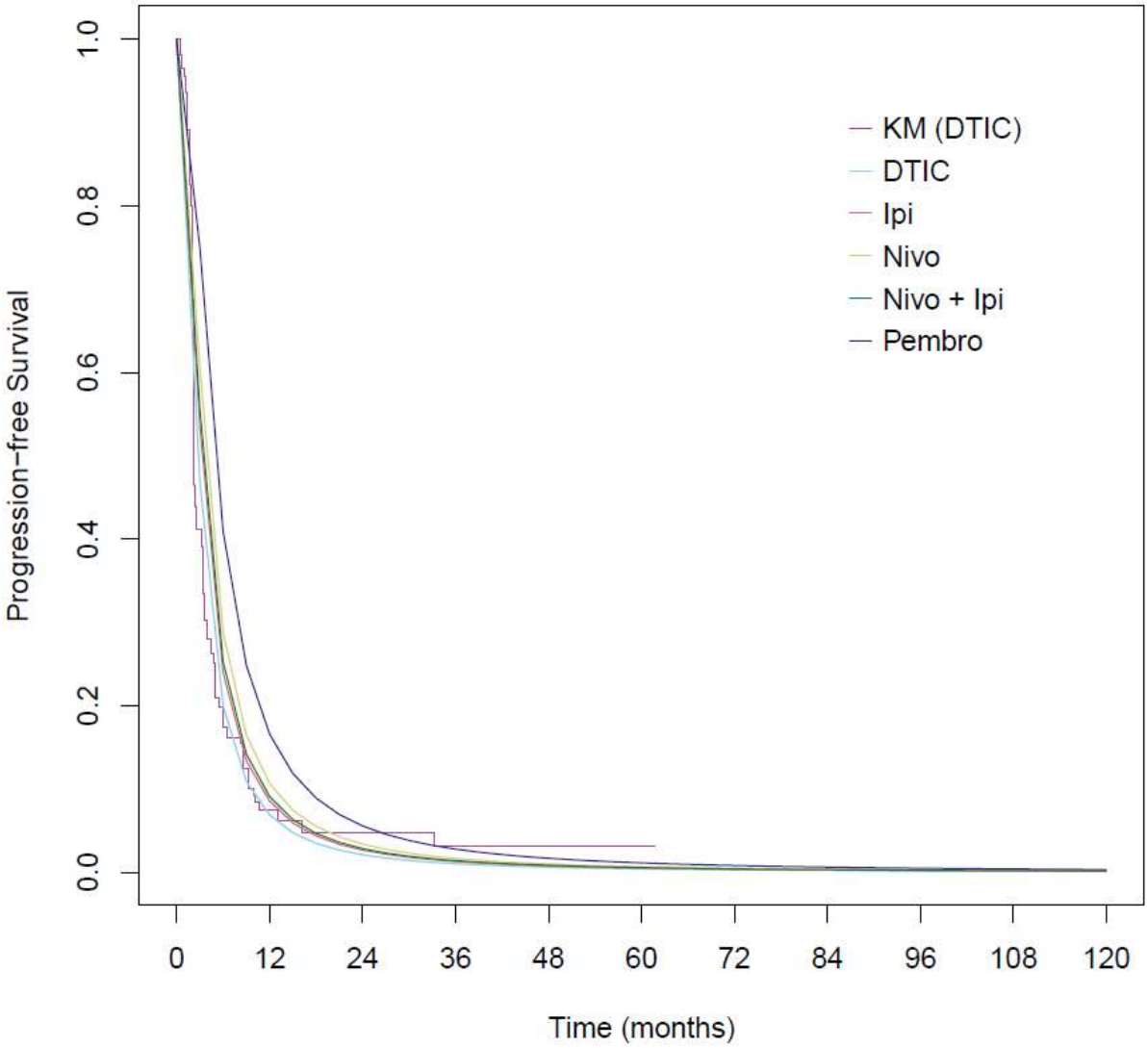


Figure 65: Survival curves for generalized gamma PFS model with one treatment effect for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

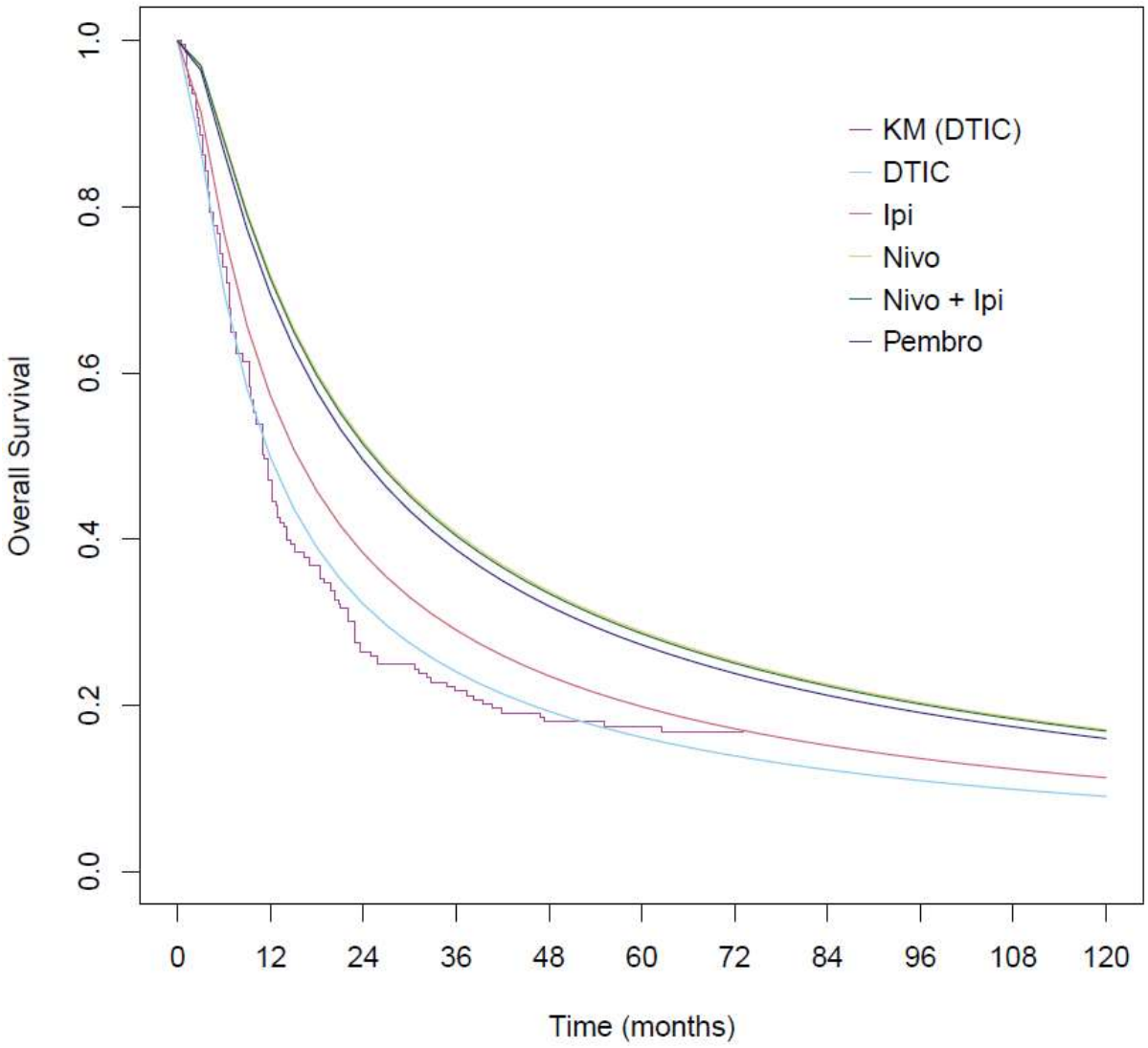


Figure 66: Survival curves for generalized gamma OS model with one treatment effect for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

D.1.1.2.7 Impact of different network assumptions on NMA results for generalized gamma models with one treatment effect

Table 49: A comparison of PFS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
dabrafenib vs. DTIC	2.793	(2.359; 3.283)	3.055	(2.393; 3.862)
dabrafenib + trametinib vs. DTIC	3.837	(3.274; 4.468)	4.325	(3.257; 5.648)
ipilimumab vs. DTIC	1.138	(0.869; 1.467)	1.149	(0.878; 1.486)

Comparison	Network 1		Network 5	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
vemurafenib vs. DTIC	2.729	(2.428; 3.055)	3.186	(2.321; 4.296)
nivolumab vs. DTIC	1.316	(1.084; 1.581)	1.316	(1.082; 1.585)
nivolumab + ipilimumab vs. DTIC	1.211	(0.911; 1.575)	1.196	(0.899; 1.564)
encorafenib + binimetinib vs. DTIC	4.492	(3.38; 5.852)	5.246	(3.462; 7.654)
pembrolizumab vs. DTIC	1.764	(1.209; 2.494)	1.781	(1.224; 2.509)

Table 50: A comparison of PFS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up), and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio						
dabrafenib vs. DTIC	0.836	(0.582; 1.163)	1.142	(0.869; 1.474)	1.153	(0.876; 1.49)
dabrafenib + trametinib vs. DTIC	1.319	(1.083; 1.588)	1.319	(1.082; 1.587)	1.319	(1.081; 1.588)
ipilimumab vs. DTIC	1.727	(1.207; 2.399)	1.216	(0.912; 1.587)	1.2	(0.899; 1.568)
vemurafenib vs. DTIC	1.298	(0.824; 1.933)	1.773	(1.207; 2.498)	1.79	(1.217; 2.523)

Table 51: A comparison of OS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
dabrafenib vs. DTIC	1.547	(1.21; 1.946)	1.161	(0.807; 1.631)
dabrafenib + trametinib vs. DTIC	2.26	(1.79; 2.814)	1.584	(1.054; 2.306)
ipilimumab vs. DTIC	1.339	(0.83; 2.051)	1.324	(0.817; 2.04)
vemurafenib vs. DTIC	1.674	(1.37; 2.019)	1.096	(0.703; 1.649)
nivolumab vs. DTIC	2.166	(1.497; 3.026)	2.168	(1.495; 3.051)

Comparison	Network 1		Network 5	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	2.125	(1.296; 3.285)	2.165	(1.318; 3.378)
encorafenib + binimetinib vs. DTIC	2.386	(1.704; 3.23)	1.564	(0.927; 2.507)
pembrolizumab vs. DTIC	2.067	(1.142; 3.461)	2.042	(1.13; 3.423)

Table 52: A comparison of OS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up), and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio						
dabrafenib vs. DTIC	1.478	(0.835; 2.419)	1.351	(0.828; 2.078)	1.333	(0.815; 2.059)
dabrafenib + trametinib vs. DTIC	2.179	(1.498; 3.063)	2.179	(1.495; 3.049)	2.179	(1.495; 3.052)
ipilimumab vs. DTIC	2.361	(1.35; 3.85)	2.142	(1.302; 3.326)	2.179	(1.32; 3.394)
vemurafenib vs. DTIC	2.406	(1.24; 4.176)	2.089	(1.138; 3.483)	2.061	(1.121; 3.439)

D.1.1.3 Two treatment effects (location and scale parameter) – NMA results

D.1.1.3.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Table 53: Fixed effect PFS NMA results for generalized gamma PFS model with two treatment effects for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
dabrafenib vs. DTIC	2.889	(2.4; 3.451)
dabrafenib + trametinib vs. DTIC	4.219	(3.542; 4.993)
ipilimumab vs. DTIC	1.472	(1.031; 2.046)
vemurafenib vs. DTIC	2.673	(2.36; 3.016)
nivolumab vs. DTIC	2.126	(1.583; 2.801)

Comparison	exp(coefficient)	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	3.07	(2.041; 4.443)
encorafenib + binimetinib vs. DTIC	5.1	(3.77; 6.745)
pembrolizumab vs. DTIC	2.929	(1.889; 4.336)
Shape parameter		
dabrafenib vs. DTIC	1.041	(0.8692; 1.238)
dabrafenib + trametinib vs. DTIC	1.213	(1.029; 1.422)
ipilimumab vs. DTIC	1.338	(1.023; 1.73)
vemurafenib vs. DTIC	0.958	(0.8479; 1.081)
nivolumab vs. DTIC	2.065	(1.676; 2.52)
nivolumab + ipilimumab vs. DTIC	3.028	(2.306; 3.921)
encorafenib + binimetinib vs. DTIC	1.287	(1.007; 1.62)
pembrolizumab vs. DTIC	2.097	(1.529; 2.82)

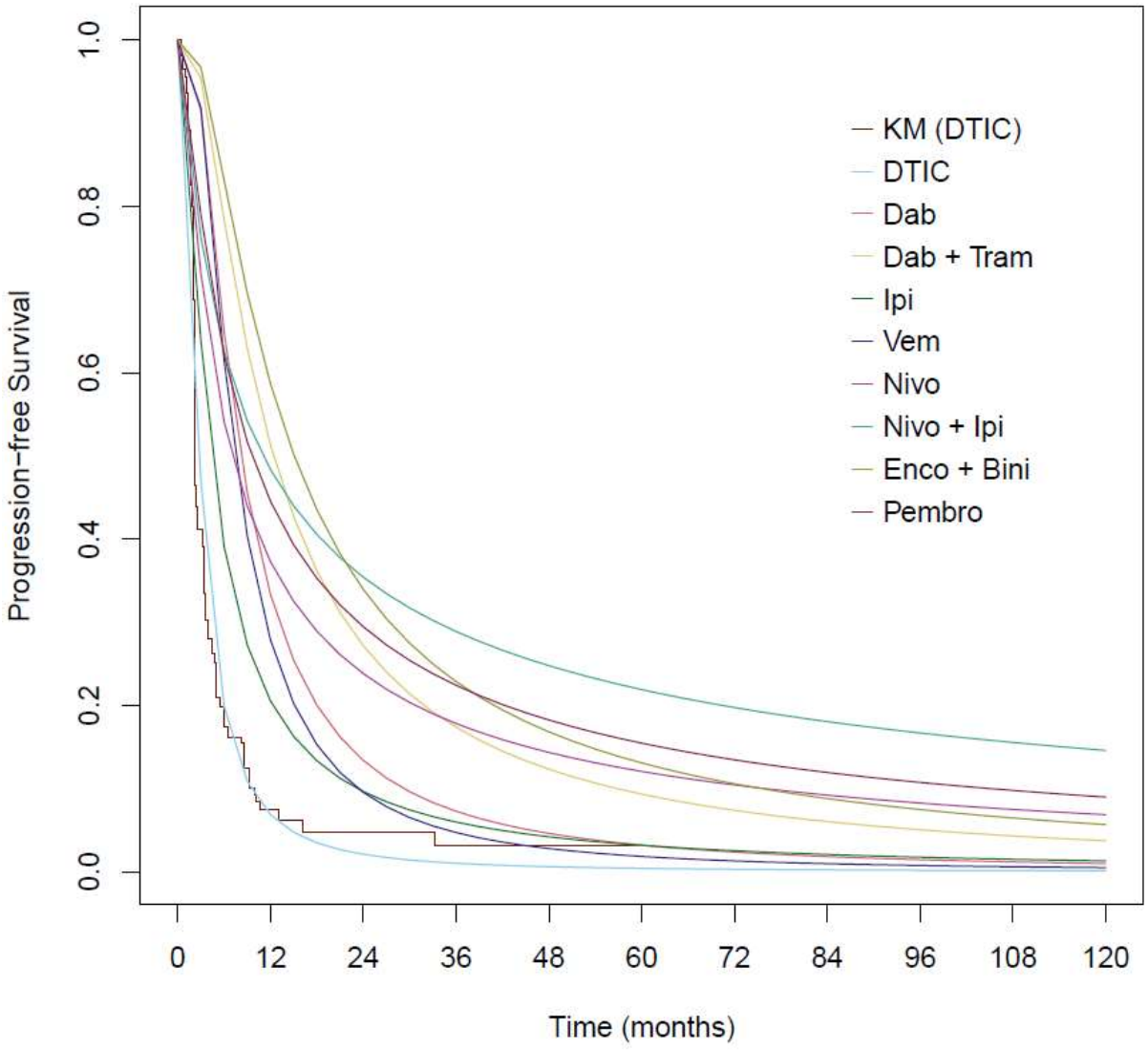


Figure 67: Survival curves for generalized gamma PFS model with two treatment effects for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

D.1.1.3.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

D.1.1.3.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 54: Fixed effect PFS NMA results for generalized gamma PFS model with two treatment effects for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		

Comparison	exp(coefficient)	95% Credible Interval
ipilimumab vs. DTIC	0.904	(0.5747; 1.37)
nivolumab vs. DTIC	2.124	(1.59; 2.781)
nivolumab + ipilimumab vs. DTIC	2.678	(1.625; 4.196)
pembrolizumab vs. DTIC	1.801	(1.071; 2.874)
Shape parameter		
ipilimumab vs. DTIC	1.296	(0.9795; 1.678)
nivolumab vs. DTIC	2.062	(1.673; 2.514)
nivolumab + ipilimumab vs. DTIC	2.016	(1.507; 2.646)
pembrolizumab vs. DTIC	2.031	(1.468; 2.736)

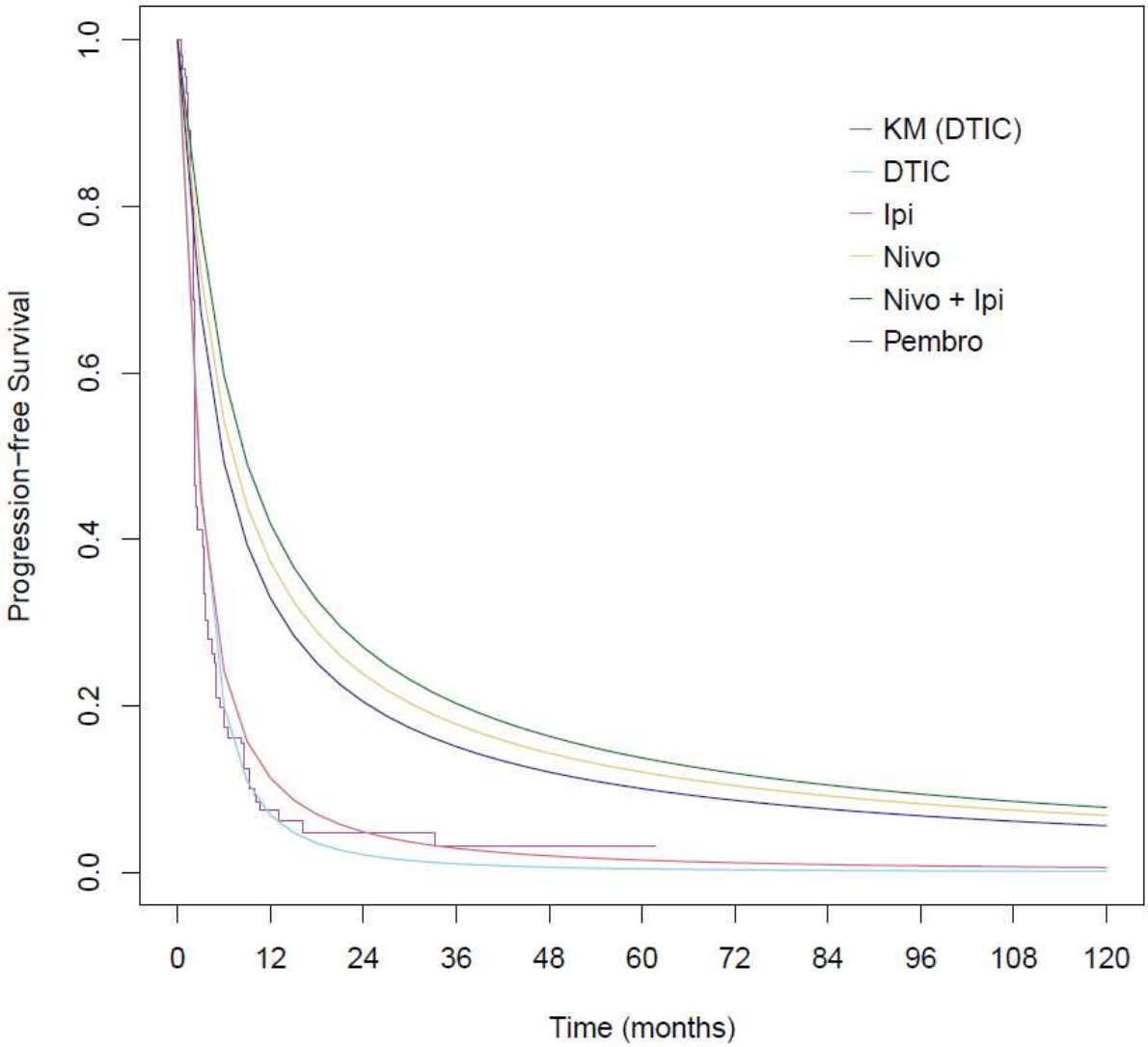


Figure 68: Survival curves for generalized gamma PFS model with two treatment effects for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

D.1.1.3.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Table 55: Fixed effect PFS NMA results for generalized gamma PFS model with two treatment effects for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
ipilimumab vs. DTIC	1.469	(1.035; 2.032)
nivolumab vs. DTIC	2.124	(1.588; 2.784)
nivolumab + ipilimumab vs. DTIC	3.064	(2.053; 4.403)
pembrolizumab vs. DTIC	2.926	(1.894; 4.308)

Comparison	exp(coefficient)	95% Credible Interval
Shape parameter		
ipilimumab vs. DTIC	1.335	(1.022; 1.719)
nivolumab vs. DTIC	2.062	(1.674; 2.52)
nivolumab + ipilimumab vs. DTIC	3.022	(2.306; 3.911)
pembrolizumab vs. DTIC	2.094	(1.528; 2.802)

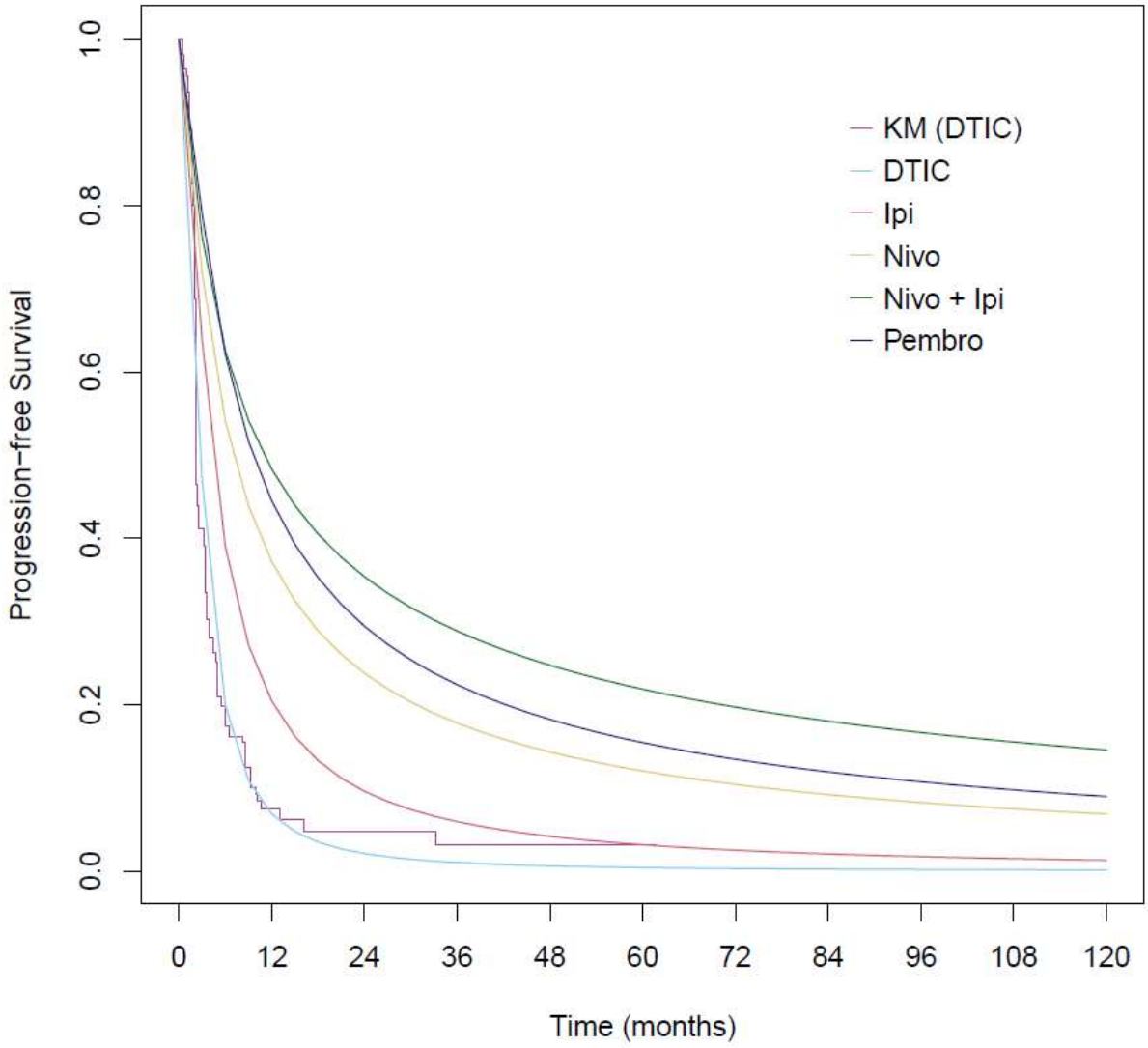


Figure 69: Survival curves for generalized gamma PFS model with two treatment effects for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

D.1.1.3.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Table 56: Fixed effect PFS NMA results for generalized gamma PFS model with two treatment effects for network 5 - people with *BRAF* wild type and

mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
dabrafenib vs. DTIC	3.086	(2.37; 3.957)
dabrafenib + trametinib vs. DTIC	4.619	(3.398; 6.153)
ipilimumab vs. DTIC	1.491	(1.044; 2.073)
vemurafenib vs. DTIC	3.012	(2.128; 4.16)
nivolumab vs. DTIC	2.126	(1.582; 2.801)
nivolumab + ipilimumab vs. DTIC	2.988	(1.973; 4.357)
encorafenib + binimetinib vs. DTIC	5.747	(3.68; 8.616)
pembrolizumab vs. DTIC	2.968	(1.912; 4.4)
Shape parameter		
dabrafenib vs. DTIC	1.038	(0.808; 1.319)
dabrafenib + trametinib vs. DTIC	1.212	(0.8916; 1.616)
ipilimumab vs. DTIC	1.378	(1.053; 1.783)
vemurafenib vs. DTIC	0.961	(0.6841; 1.318)
nivolumab vs. DTIC	2.065	(1.675; 2.52)
nivolumab + ipilimumab vs. DTIC	2.917	(2.219; 3.782)
encorafenib + binimetinib vs. DTIC	1.291	(0.8646; 1.865)
pembrolizumab vs. DTIC	2.16	(1.572; 2.909)

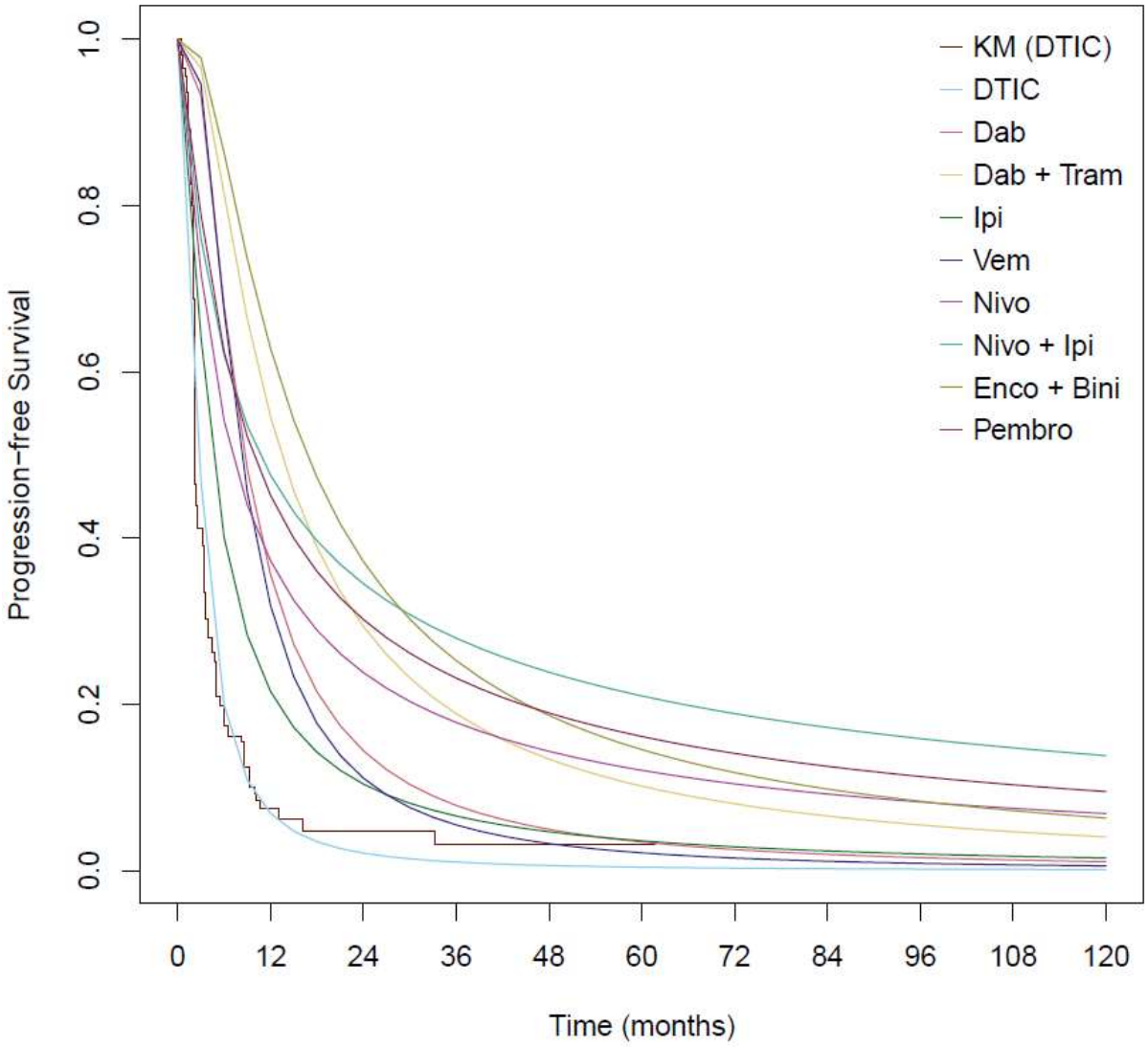


Figure 70: Survival curves for generalized gamma PFS model with two treatment effects for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

D.1.1.3.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Table 57: Fixed effect PFS NMA results for generalized gamma PFS model with two treatment effects for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Comparison	exp(coefficient)	95% Credible Interval
Time-Ratio		
ipilimumab vs. DTIC	1.488	(1.048; 2.06)
nivolumab vs. DTIC	2.124	(1.588; 2.784)
nivolumab + ipilimumab vs. DTIC	2.983	(1.981; 4.328)

Comparison	exp(coefficient)	95% Credible Interval
pembrolizumab vs. DTIC	2.965	(1.918; 4.372)
Shape parameter		
ipilimumab vs. DTIC	1.376	(1.052; 1.772)
nivolumab vs. DTIC	2.062	(1.675; 2.519)
nivolumab + ipilimumab vs. DTIC	2.911	(2.215; 3.767)
pembrolizumab vs. DTIC	2.157	(1.571; 2.888)

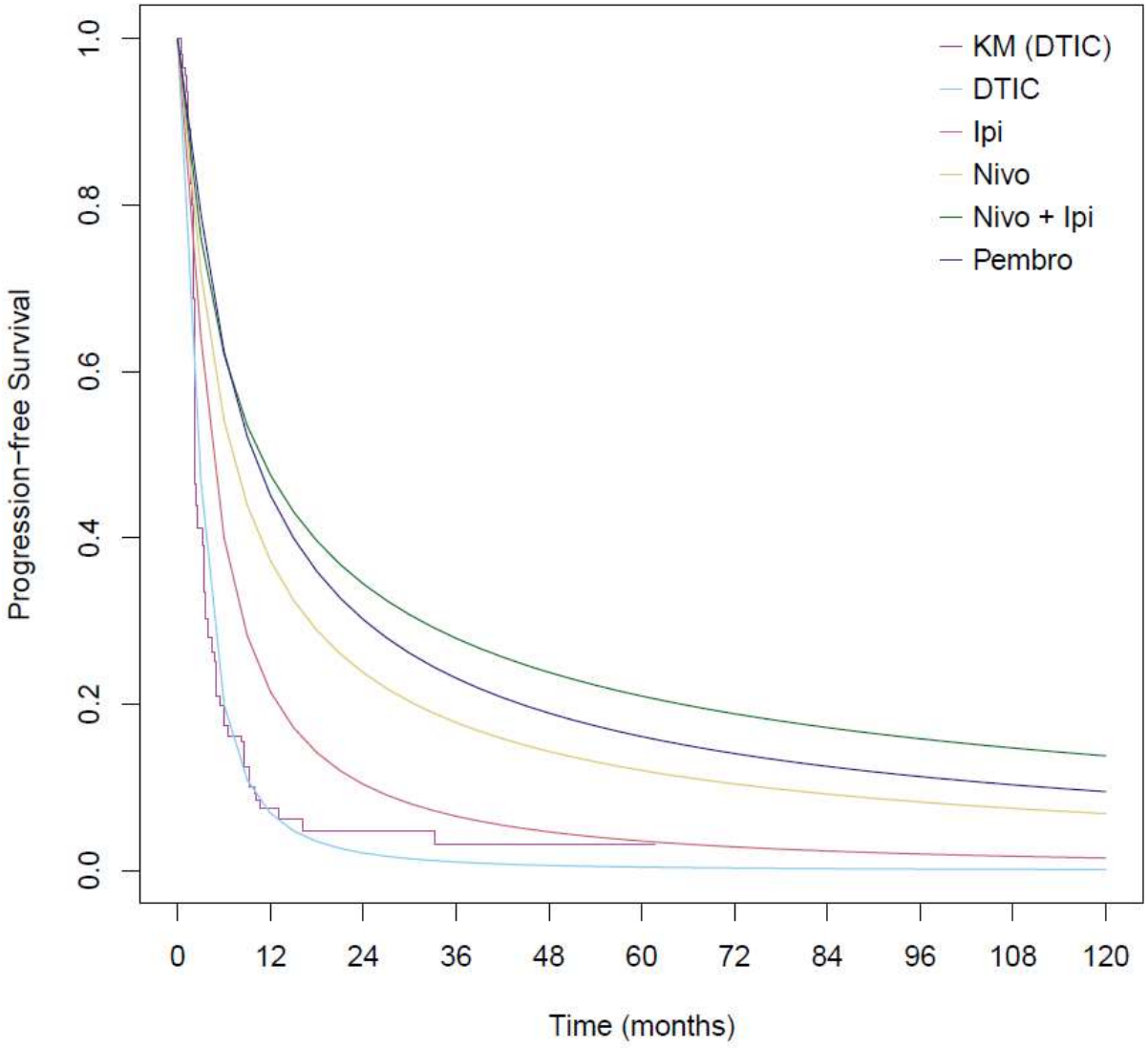


Figure 71: Survival curves for generalized gamma PFS model with two treatment effects for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

D.1.1.3.7 Impact of different network assumptions on NMA results for generalized gamma models with two treatment effects

Table 58: A comparison of PFS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
dabrafenib vs. DTIC	2.889	(2.4; 3.451)	3.086	(2.37; 3.957)

Comparison	Network 1		Network 5	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
dabrafenib + trametinib vs. DTIC	4.219	(3.542; 4.993)	4.619	(3.398; 6.153)
ipilimumab vs. DTIC	1.472	(1.031; 2.046)	1.491	(1.044; 2.073)
vemurafenib vs. DTIC	2.673	(2.36; 3.016)	3.012	(2.128; 4.16)
nivolumab vs. DTIC	2.126	(1.583; 2.801)	2.126	(1.582; 2.801)
nivolumab + ipilimumab vs. DTIC	3.07	(2.041; 4.443)	2.988	(1.973; 4.357)
encorafenib + binimetinib vs. DTIC	5.1	(3.77; 6.745)	5.747	(3.68; 8.616)
pembrolizumab vs. DTIC	2.929	(1.889; 4.336)	2.968	(1.912; 4.4)
Shape parameter				
dabrafenib vs. DTIC	1.041	(0.8692; 1.238)	1.038	(0.808; 1.319)
dabrafenib + trametinib vs. DTIC	1.213	(1.029; 1.422)	1.212	(0.8916; 1.616)
ipilimumab vs. DTIC	1.338	(1.023; 1.73)	1.378	(1.053; 1.783)
vemurafenib vs. DTIC	0.958	(0.8479; 1.081)	0.961	(0.6841; 1.318)
nivolumab vs. DTIC	2.065	(1.676; 2.52)	2.065	(1.675; 2.52)
nivolumab + ipilimumab vs. DTIC	3.028	(2.306; 3.921)	2.917	(2.219; 3.782)
encorafenib + binimetinib vs. DTIC	1.287	(1.007; 1.62)	1.291	(0.8646; 1.865)
pembrolizumab vs. DTIC	2.097	(1.529-2.82)	2.16	(1.572; 2.909)

Table 59: A comparison of PFS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up), and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio						
ipilimumab vs. DTIC	0.904	(0.5747; 1.37)	1.469	(1.035; 2.032)	1.488	(1.048; 2.06)

Comparison	Network 3		Network 4		Network 6	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
nivolumab vs. DTIC	2.124	(1.59; 2.781)	2.124	(1.588; 2.784)	2.124	(1.588; 2.784)
nivolumab + ipilimumab vs. DTIC	2.678	(1.625; 4.196)	3.064	(2.053; 4.403)	2.983	(1.981; 4.328)
pembrolizumab vs. DTIC	1.801	(1.071; 2.874)	2.926	(1.894; 4.308)	2.965	(1.918; 4.372)
Shape parameter						
ipilimumab vs. DTIC	1.296	(0.9795; 1.678)	1.335	(1.022; 1.719)	1.376	(1.052; 1.772)
nivolumab vs. DTIC	2.062	(1.673; 2.514)	2.062	(1.674; 2.52)	2.062	(1.675; 2.519)
nivolumab + ipilimumab vs. DTIC	2.016	(1.507; 2.646)	3.022	(2.306; 3.911)	2.911	(2.215; 3.767)
pembrolizumab vs. DTIC	2.031	(1.468; 2.736)	2.094	(1.528; 2.802)	2.157	(1.571; 2.888)

Table 60: A comparison of OS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio				
dabrafenib vs. DTIC	1.522	(1.181; 1.935)	1.202	(0.8336; 1.69)
dabrafenib + trametinib vs. DTIC	2.142	(1.692; 2.68)	1.629	(1.071; 2.39)
ipilimumab vs. DTIC	1.424	(0.8762; 2.188)	1.414	(0.8696; 2.172)
vemurafenib vs. DTIC	1.545	(1.282; 1.85)	1.121	(0.7039; 1.707)
nivolumab vs. DTIC	2.533	(1.739; 3.593)	2.533	(1.737; 3.593)
nivolumab + ipilimumab vs. DTIC	3.343	(1.922; 5.426)	3.405	(1.943; 5.555)
encorafenib + binimetinib vs. DTIC	2.431	(1.73; 3.32)	1.764	(1.016; 2.864)

Comparison	Network 1		Network 5	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
pembrolizumab vs. DTIC	2.502	(1.362; 4.213)	2.486	(1.354; 4.185)
Shape parameter				
dabrafenib vs. DTIC	0.864	(0.7174; 1.033)	1.095	(0.8444; 1.402)
dabrafenib + trametinib vs. DTIC	0.774	(0.6531; 0.9107)	1.087	(0.7918; 1.462)
ipilimumab vs. DTIC	1.272	(1.005; 1.593)	1.279	(1.01; 1.603)
vemurafenib vs. DTIC	0.709	(0.6275; 0.7982)	1.081	(0.7593; 1.498)
nivolumab vs. DTIC	1.534	(1.282; 1.826)	1.534	(1.283; 1.826)
nivolumab + ipilimumab vs. DTIC	2.067	(1.614; 2.618)	2.051	(1.596; 2.602)
encorafenib + binimetinib vs. DTIC	0.887	(0.7023; 1.106)	1.352	(0.902; 1.959)
pembrolizumab vs. DTIC	1.516	(1.118; 2.013)	1.525	(1.123; 2.026)

Table 61: A comparison of OS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up), and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
Time-Ratio						
ipilimumab vs. DTIC	1.427	(0.8165; 2.348)	1.421	(0.8792; 2.179)	1.412	(0.8718; 2.167)
nivolumab vs. DTIC	2.532	(1.742; 3.557)	2.531	(1.745; 3.566)	2.531	(1.744; 3.567)
nivolumab + ipilimumab vs. DTIC	3.094	(1.624; 5.421)	3.338	(1.93; 5.397)	3.401	(1.948; 5.536)
pembrolizumab vs. DTIC	2.433	(1.256; 4.315)	2.501	(1.367; 4.186)	2.485	(1.357; 4.166)
Shape parameter						
ipilimumab vs. DTIC	1.092	(0.8433; 1.385)	1.27	(1.004; 1.587)	1.277	(1.009; 1.596)

Comparison	Network 3		Network 4		Network 6	
	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval	exp(coefficient)	95% Credible Interval
nivolumab vs. DTIC	1.533	(1.28; 1.822)	1.533	(1.28; 1.823)	1.533	(1.279; 1.822)
nivolumab + ipilimumab vs. DTIC	1.703	(1.301; 2.198)	2.064	(1.61; 2.607)	2.048	(1.592; 2.591)
pembrolizumab vs. DTIC	1.16	(0.8405; 1.563)	1.515	(1.118; 2.005)	1.524	(1.124; 2.015)

Appendix E: Piecewise exponential

E.1 DIC comparisons for different piecewise models

E.1.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Table 62: DIC values for all piecewise exponential models for PFS and OS for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Cut point placement	DIC (smallest DIC value in **bold**)	
	PFS	OS
One cut point		
6 months	327.9	**312**
9 months	320.1	318.2
12 months	310.2	320
15 months	**306.6**	320
Two cut points		
6 and 12 months	448.9	451.8
9 and 15 months	435.9	452.9
9 and 18 months	423.8	453.2
12 and 18 months	**404.3**	**448.3**
12 and 20 months	[b]	449.6
12 and 24 months	[a]	[a]
Three cut points		
6, 12 and 18 months	**543.1**	**580.5**
12, 24 and 36 months	[a]	[a]

(a) The shortest amount of follow-up time in these networks is 22 months (BRIM-3) and 24 months (CheckMate 069) for PFS and OS respectively. As such, it is not possible to generate aggregate data beyond these time points. Thus, for this network, it's impossible to run piecewise models with cut points beyond 22 or 24 months for PFS and OS respectively.

(b) Although below the 22 months, using 20 months as the second cut placement resulted in the final interval, >20 months, of the BRIM-3 aggregate data having one arm with no data. Thus, it's impossible to run the PFS piecewise model with a second interval at 20 months.

E.1.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

E.1.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 63: DIC values for all piecewise exponential models for PFS and OS for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Cut point	**DIC** (smallest DIC value in **bold**)	
	PFS	**OS**
One cut point		
6 months	112.6	**111.8**
9 months	110.5	112.8
12 months	108.1	112.7
15 months	**106.5**	112.2
Two cut points		
6 and 12 months	154.1	159.9
9 and 15 months	149.7	158.8
9 and 18 months	149.1	159.9
12 and 18 months	**143.5**	**157.6**
12 and 20 months	143.9	158.3
12 and 24 months	144.6	158.3
Three cut points		
6, 12 and 18 months	**189.9**	204.9
12, 24 and 36 months	Did not converge	**198.7**

E.1.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Table 64: DIC values for all piecewise exponential models for PFS and OS for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Cut point placement	**DIC** (smallest DIC value in **bold**)	
	PFS	**OS**
One cut point		
6 months	137.4	137.6
9 months	135.4	138.1
12 months	130.6	137.9
15 months	**129.4**	**137.3**
Two cut points		

Cut point placement	DIC (smallest DIC value in **bold**)	
	PFS	OS
6 and 12 months	187.2	197.8
9 and 15 months	182.1	194.1
9 and 18 months	175.8	194.4
12 and 18 months	**167.4**	**191.3**
12 and 20 months	168.5	191.6
12 and 24 months	[a]	[a]
Three cut points		
6, 12 and 18 months	Did not converge	**250.8**
12, 24 and 36 months	[a]	[a]

(a) The shortest amount of follow-up time in these networks is 24 months (CheckMate 069) for both PFS and OS. As such, it is not possible to generate aggregate data beyond these time points. Thus, for this network, it's impossible to run piecewise models with cut points beyond 24 months for PFS or OS.

E.1.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Table 65: DIC values for all piecewise exponential models for PFS and OS for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Cut point	DIC (smallest DIC value in **bold**)	
	PFS	OS
One cut point		
6 months	273.7	**255.7**
9 months	268.8	263.7
12 months	262.9	265.9
15 months	**260.2**	266.1
Two cut points		
6 and 12 months	379.3	**371.2**
9 and 15 months	370	376.2
9 and 18 months	367.3	377.7
12 and 18 months	352.1	374.8
12 and 20 months	348[a]	376.4
12 and 24 months	**347.5[a]**	377.1
Three cut points		
6, 12 and 18 months	**469.3**	480.7
12, 24 and 36 months	Did not converge	**476.2**

(a) Although the piecewise models with cuts at 12 & 20 months, and 12 & 24 months have the smallest and second smallest DIC value respectively for network 5, neither of these return plausible estimates of PFS beyond 20 months for any of the targeted therapies – as shown in Figure 72. Specifically, the DTIC arm of BREAK-3 has no event after 20 months, so it is not possible to obtain credible estimates. This in turn affects the estimates for dabrafenib+trametinib, vemurafenib monotherapy, encorafenib+binimetinib as these treatments are connected in the network and therefore the network uses the unreliable estimates between dabrafenib and DTIC to inform them. As such, we considered the model with two cuts at 12 & 18 months to be the best fitting PFS model for network 5.

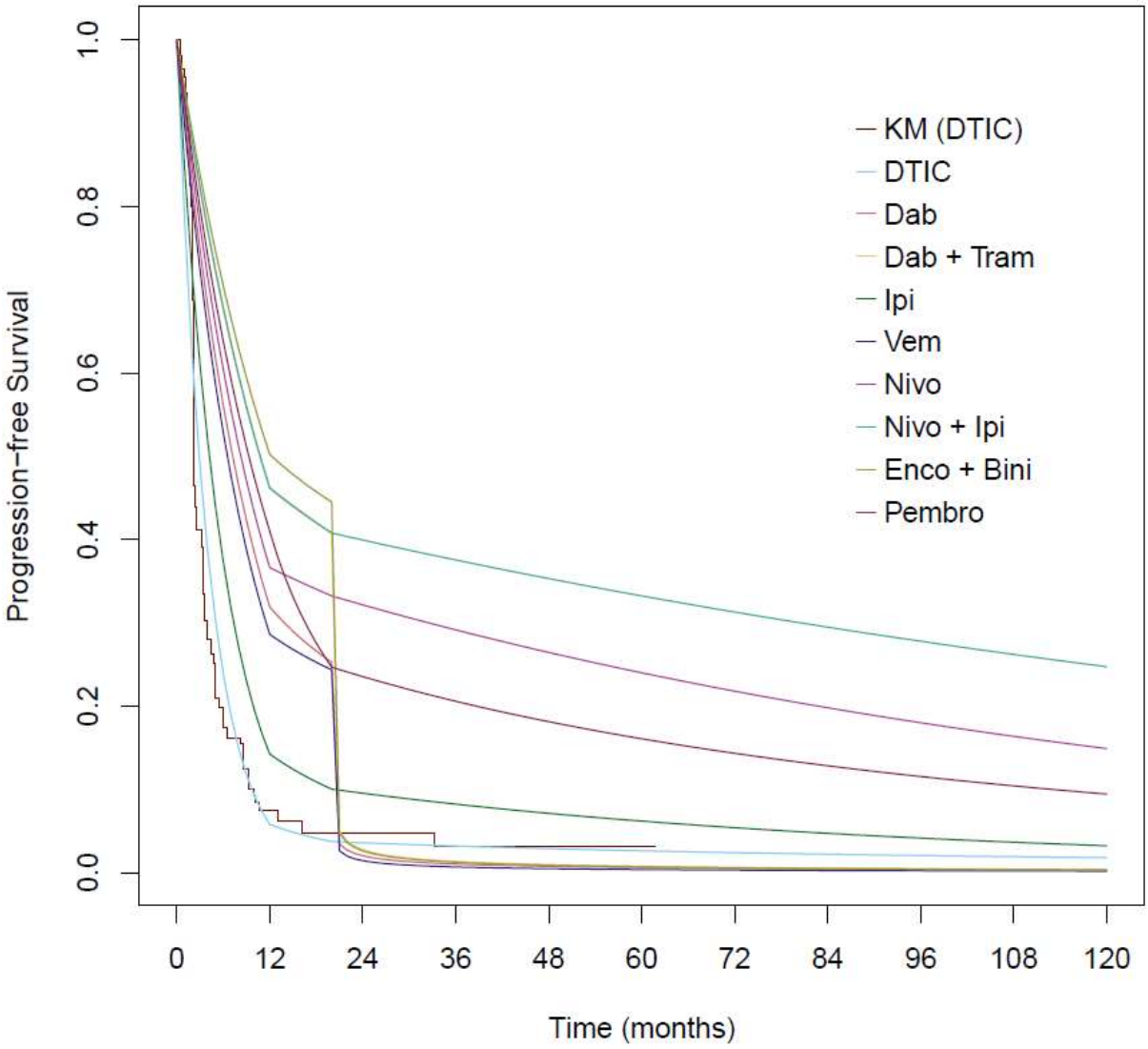


Figure 72: Survival curves for piecewise PFS model with two cut points at 12 and 20 months for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

E.1.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Table 66: DIC values for all piecewise exponential models for PFS and OS for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Cut point	**DIC** (smallest DIC value in **bold**)	
	PFS	**OS**
One cut point		
6 months	115.1	**113**
9 months	113.1	114.7
12 months	110.5	114.8
15 months	**109.3**	114.4
Two cut points		

Cut point	DIC (smallest DIC value in **bold**)	
	PFS	OS
6 and 12 months	158	162.9
9 and 15 months	154	162.3
9 and 18 months	153.6	163.4
12 and 18 months	**147.8**	**161.1**
12 and 20 months	148.2	161.8
12 and 24 months	148.7	162.1
Three cut points		
6, 12 and 18 months	**195.3**	208.8
12, 24 and 36 months	Did not converge	**204.2**

E.1.7 One cut point - NMA results

E.1.7.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Table 67: Fixed effect PFS & OS NMA results for the piecewise exponential model with the lowest DIC value with 1 cut point for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1	**1-15 months**		**1-6 months**	
dabrafenib vs. DTIC	0.411	(0.324; 0.522)	0.687	(0.385; 1.256)
dabrafenib + trametinib vs. DTIC	0.258	(0.207; 0.325)	0.237	(0.135; 0.411)
ipilimumab vs. DTIC	0.656	(0.485; 0.891)	0.666	(0.372; 1.186)
vemurafenib vs. DTIC	0.428	(0.361; 0.506)	0.491	(0.363; 0.661)
nivolumab vs. DTIC	0.314	(0.247; 0.398)	0.563	(0.361; 0.867)
nivolumab + ipilimumab vs. DTIC	0.238	(0.173; 0.327)	0.377	(0.203; 0.691)
encorafenib + binimetinib vs. DTIC	0.232	(0.165; 0.326)	0.258	(0.109; 0.591)
pembrolizumab vs. DTIC	0.296	(0.203; 0.432)	0.418	(0.201; 0.872)
Interval 2	**16-120 months**		**7-120 months**	
dabrafenib vs. DTIC	1.266	(0.313; 7.714)	0.842	(0.644; 1.106)
dabrafenib + trametinib vs. DTIC	1.097	(0.274; 6.449)	0.706	(0.555; 0.9)

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
ipilimumab vs. DTIC	1.489	(0.314; 11.393)	0.777	(0.539; 1.12)
vemurafenib vs. DTIC	1.55	(0.393; 9.235)	1.013	(0.838; 1.229)
nivolumab vs. DTIC	0.671	(0.164; 4.845)	0.417	(0.315; 0.553)
nivolumab + ipilimumab vs. DTIC	0.501	(0.111; 3.777)	0.377	(0.258; 0.549)
encorafenib + binimetinib vs. DTIC	1.31	(0.284; 8.619)	0.652	(0.471; 0.906)
pembrolizumab vs. DTIC	1.355	(0.24; 11.314)	0.562	(0.358; 0.889)

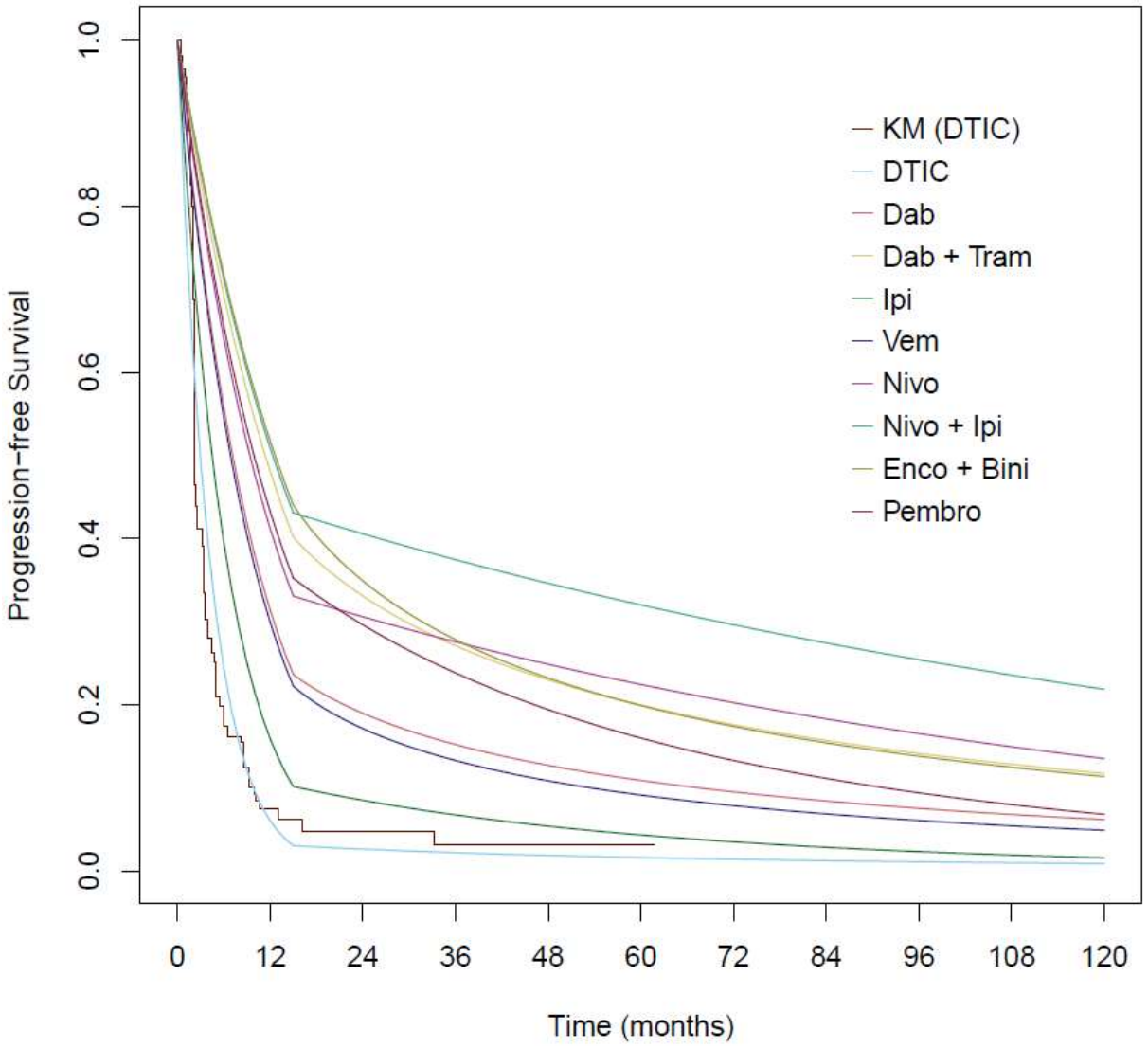


Figure 73: Survival curves for piecewise PFS model with one cut point at 15 months for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

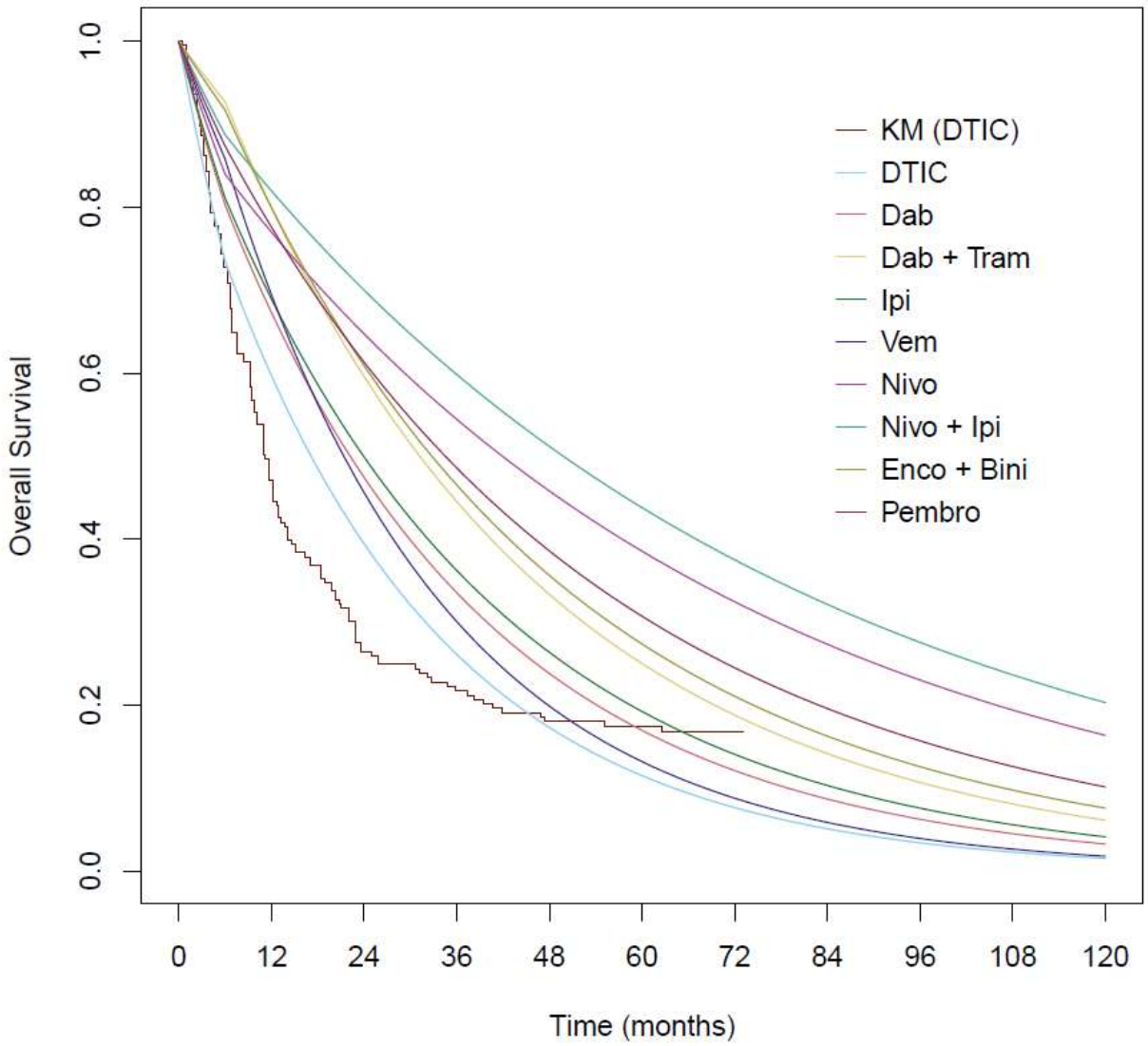


Figure 74: Survival curves for piecewise OS model with one cut point at 6 months for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

E.1.7.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

E.1.7.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 68: Fixed effect PFS & OS NMA results for the piecewise exponential model with the lowest DIC value with 1 cut point for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1	1-15 months		1-6 months	

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
ipilimumab vs. DTIC	0.789	(0.567; 1.099)	0.691	(0.375; 1.273)
nivolumab vs. DTIC	0.314	(0.248; 0.397)	0.565	(0.363; 0.868)
nivolumab + ipilimumab vs. DTIC	0.285	(0.2; 0.406)	0.621	(0.334; 1.148)
pembrolizumab vs. DTIC	0.356	(0.24; 0.531)	0.342	(0.165; 0.714)
Interval 2	**16-120 months**		**7-120 months**	
ipilimumab vs. DTIC	1.41	(0.304; 10.299)	0.778	(0.519; 1.157)
nivolumab vs. DTIC	0.597	(0.156; 3.717)	0.417	(0.314; 0.549)
nivolumab + ipilimumab vs. DTIC	0.496	(0.108; 3.48)	0.331	(0.215; 0.505)
pembrolizumab vs. DTIC	1.28	(0.229; 10.783)	0.717	(0.435; 1.18)

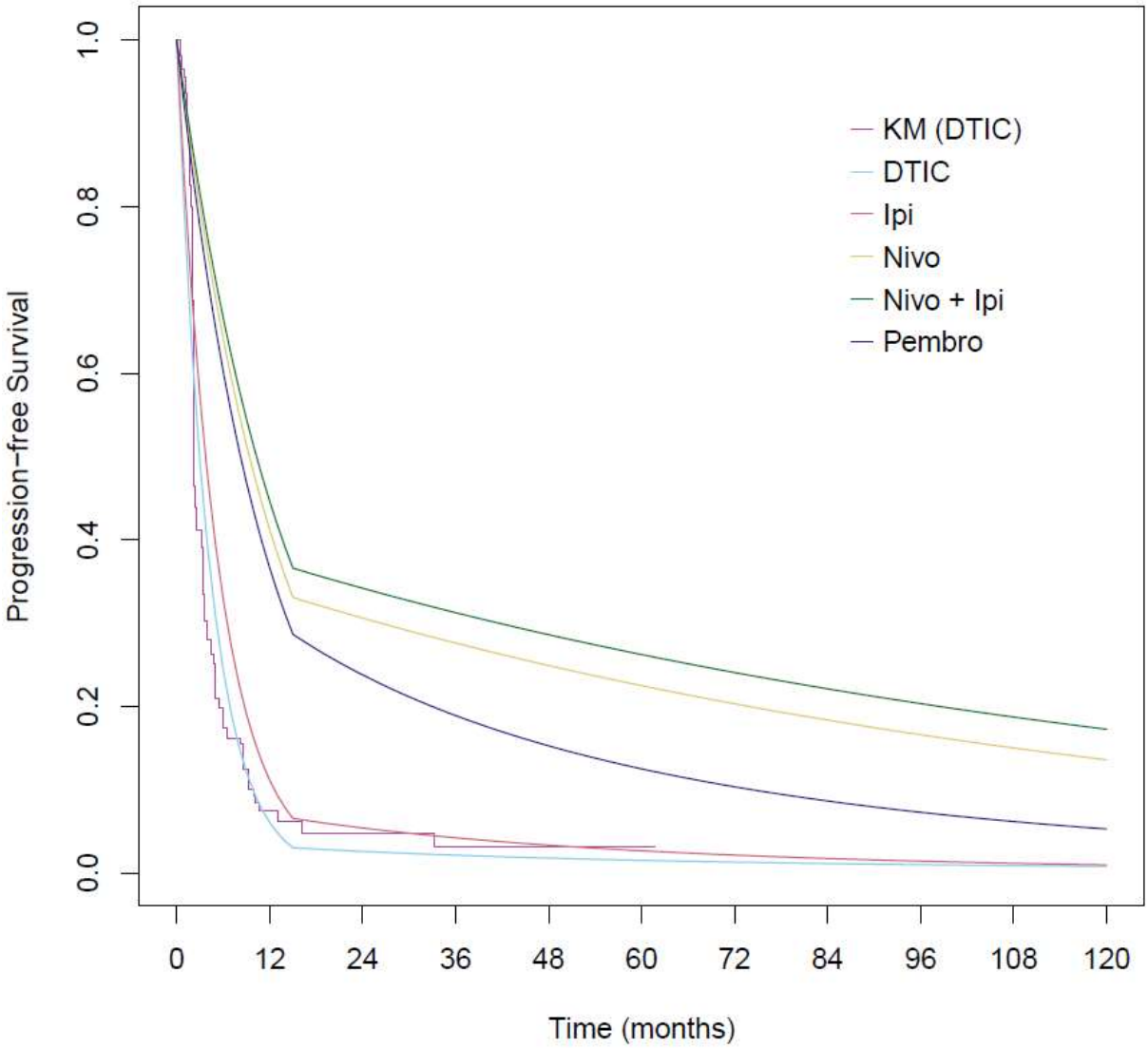


Figure 75: Survival curves for piecewise PFS model with one cut point at 15 months for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

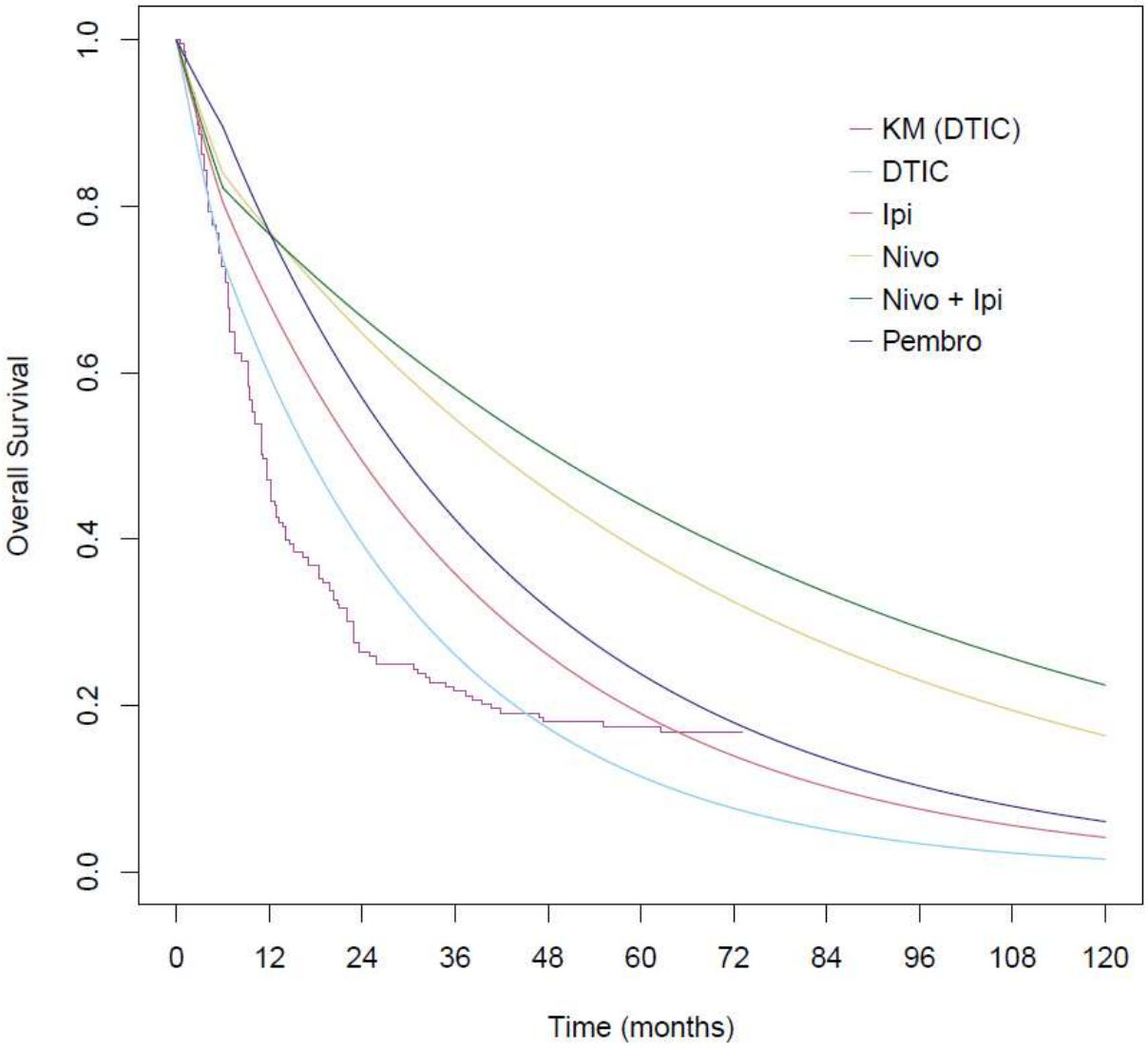


Figure 76: Survival curves for piecewise OS model with one cut point at 6 months for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

E.1.7.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Table 69: Fixed effect PFS & OS NMA results for the piecewise exponential model with the lowest DIC value with 1 cut point for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-15 months[a]				
ipilimumab vs. DTIC	0.655	(0.486; 0.886)	0.609	(0.411; 0.9)
nivolumab vs. DTIC	0.313	(0.248; 0.396)	0.457	(0.34; 0.612)

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	0.238	(0.174; 0.326)	0.429	(0.286; 0.642)
pembrolizumab vs. DTIC	0.296	(0.204; 0.428)	0.345	(0.212; 0.563)
Interval 2: 16-120 months[a]				
ipilimumab vs. DTIC	1.413	(0.301; 11.811)	1.184	(0.727; 1.929)
nivolumab vs. DTIC	0.64	(0.155; 4.993)	0.58	(0.388; 0.861)
nivolumab + ipilimumab vs. DTIC	0.476	(0.103; 3.92)	0.426	(0.254; 0.714)
pembrolizumab vs. DTIC	1.279	(0.225; 11.577)	1.229	(0.648; 2.312)

(a) Both PFS and OS models have the same cut points

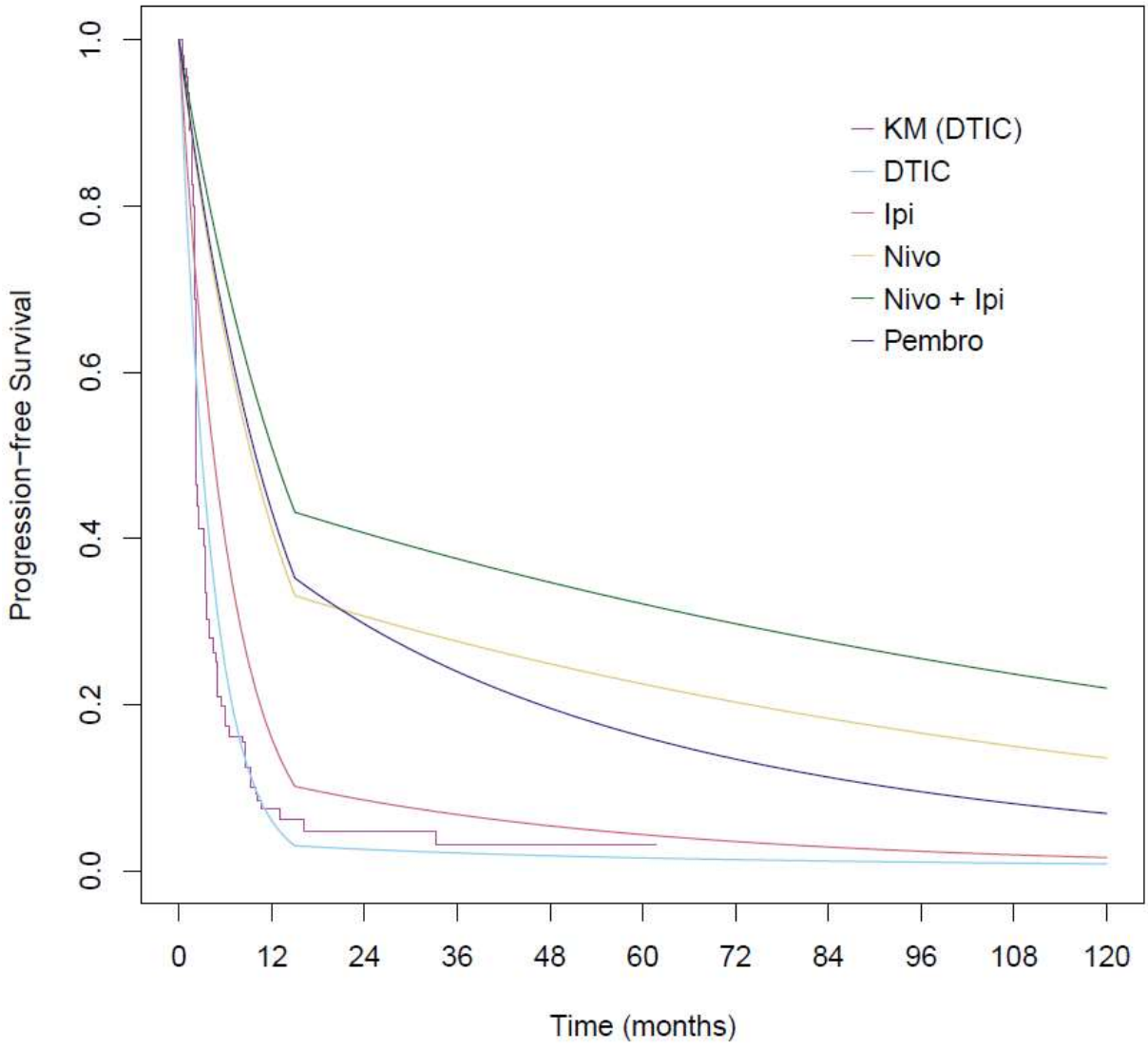


Figure 77: Survival curves for piecewise PFS model with one cut point at 15 months for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

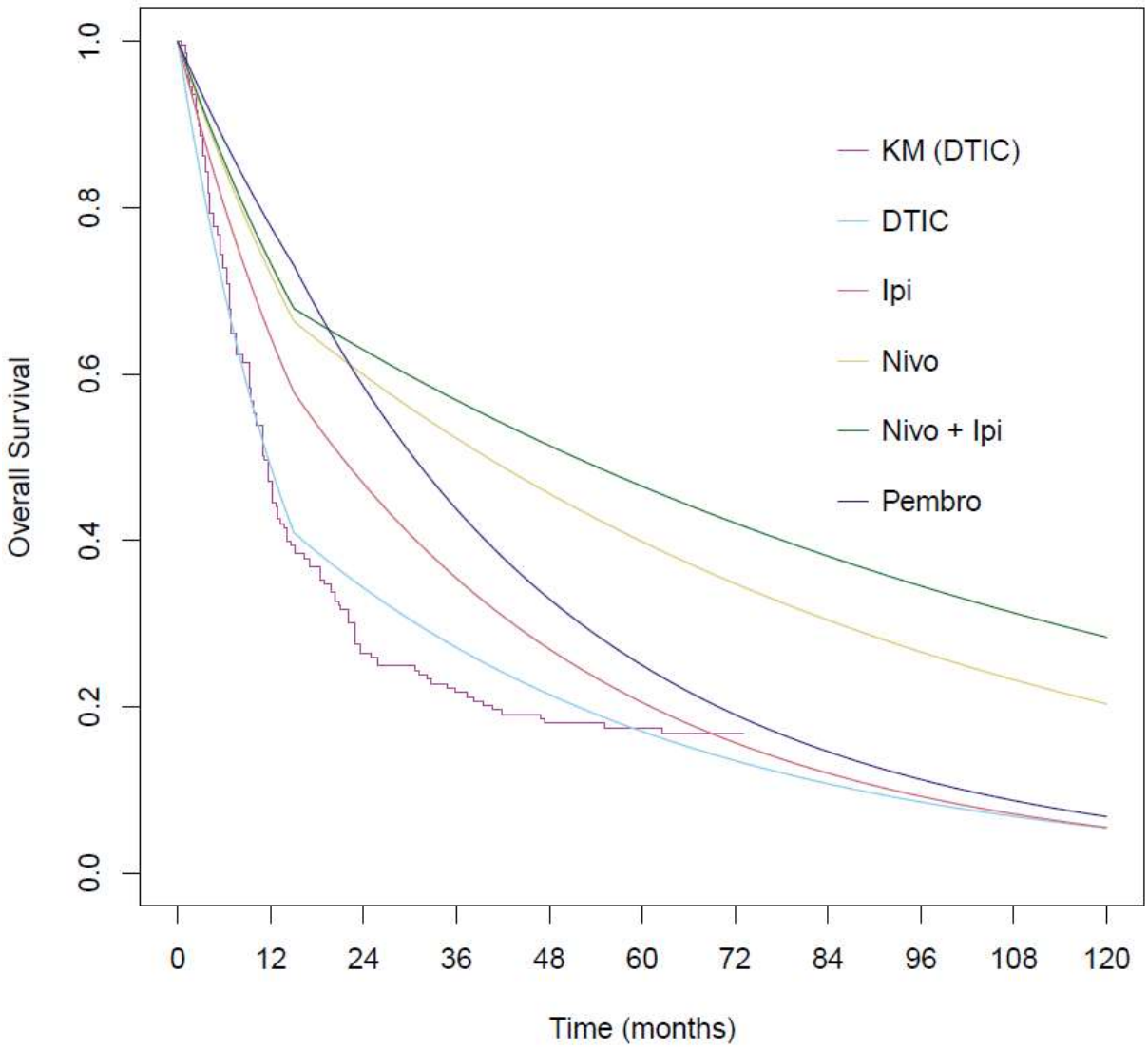


Figure 78: Survival curves for piecewise OS model with one cut point at 15 months for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

E.1.7.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Table 70: Fixed effect PFS & OS NMA results for the piecewise exponential model with the lowest DIC value with 1 cut point for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1	**1-15 months**		**1-6 months**	
dabrafenib vs. DTIC	0.406	(0.296; 0.565)	0.897	(0.423; 2.051)
dabrafenib + trametinib vs. DTIC	0.254	(0.171; 0.382)	0.378	(0.133; 1.111)

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
ipilimumab vs. DTIC	0.656	(0.48; 0.886)	0.705	(0.392; 1.264)
vemurafenib vs. DTIC	0.419	(0.267; 0.664)	0.927	(0.276; 3.274)
nivolumab vs. DTIC	0.315	(0.249; 0.398)	0.565	(0.365; 0.867)
nivolumab + ipilimumab vs. DTIC	0.243	(0.176; 0.333)	0.338	(0.177; 0.635)
encorafenib + binimetinib vs. DTIC	0.227	(0.132; 0.39)	0.487	(0.113; 2.116)
pembrolizumab vs. DTIC	0.296	(0.201; 0.43)	0.443	(0.213; 0.933)
Interval 2	**16-120 months**		**7-120 months**	
dabrafenib vs. DTIC	0.551	(0.133; 3.881)	0.806	(0.559; 1.197)
dabrafenib + trametinib vs. DTIC	0.44	(0.092; 3.304)	0.665	(0.431; 1.047)
ipilimumab vs. DTIC	1.444	(0.33; 10.176)	0.781	(0.543; 1.119)
vemurafenib vs. DTIC	0.588	(0.112; 4.655)	0.944	(0.584; 1.539)
nivolumab vs. DTIC	0.672	(0.173; 4.609)	0.416	(0.315; 0.549)
nivolumab + ipilimumab vs. DTIC	0.512	(0.118; 3.758)	0.371	(0.254; 0.54)
encorafenib + binimetinib vs. DTIC	0.496	(0.082; 4.25)	0.608	(0.347; 1.06)
pembrolizumab vs. DTIC	1.303	(0.247; 10.697)	0.565	(0.358; 0.886)

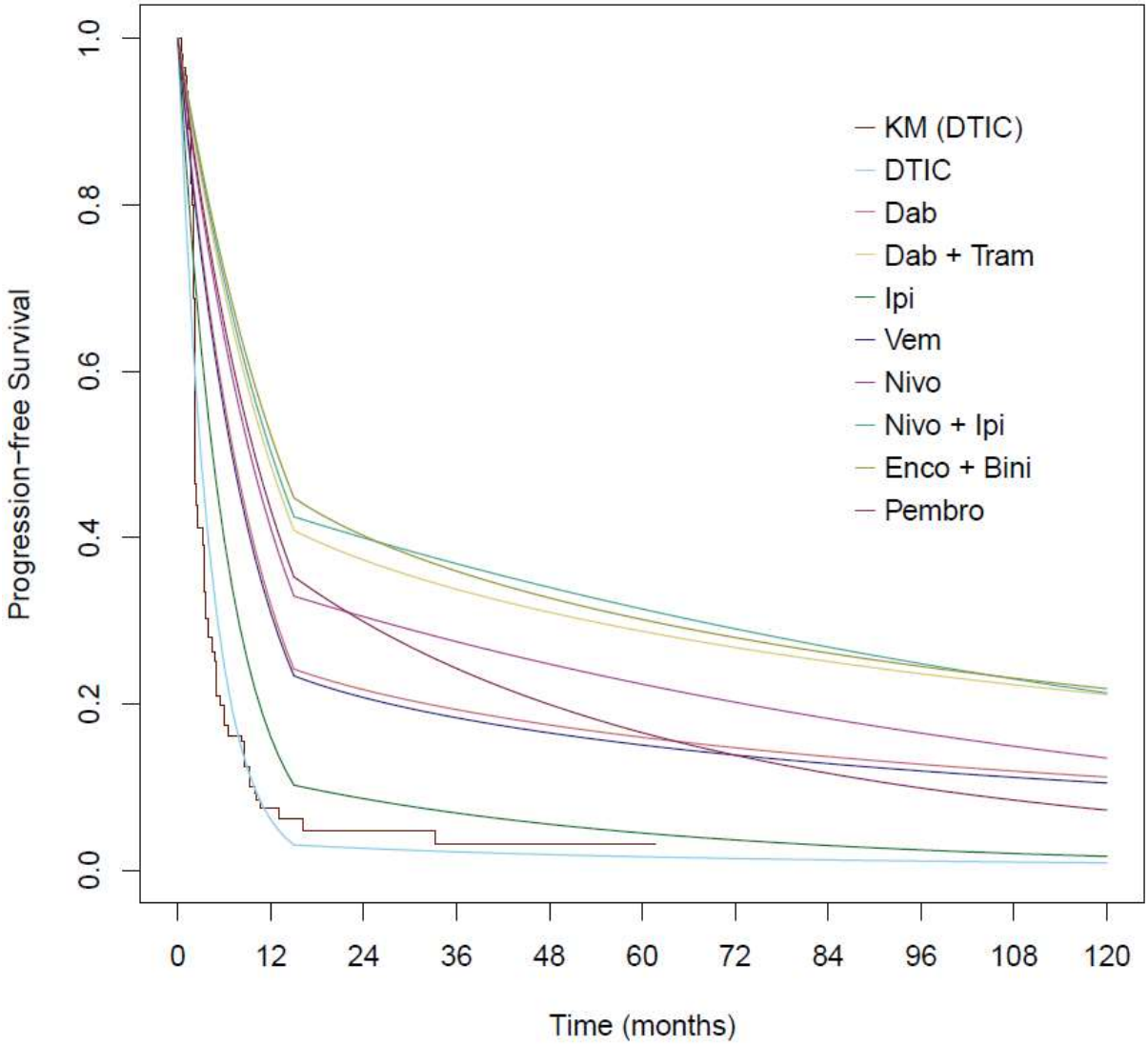


Figure 79: Survival curves for piecewise PFS model with one cut point at 15 months for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

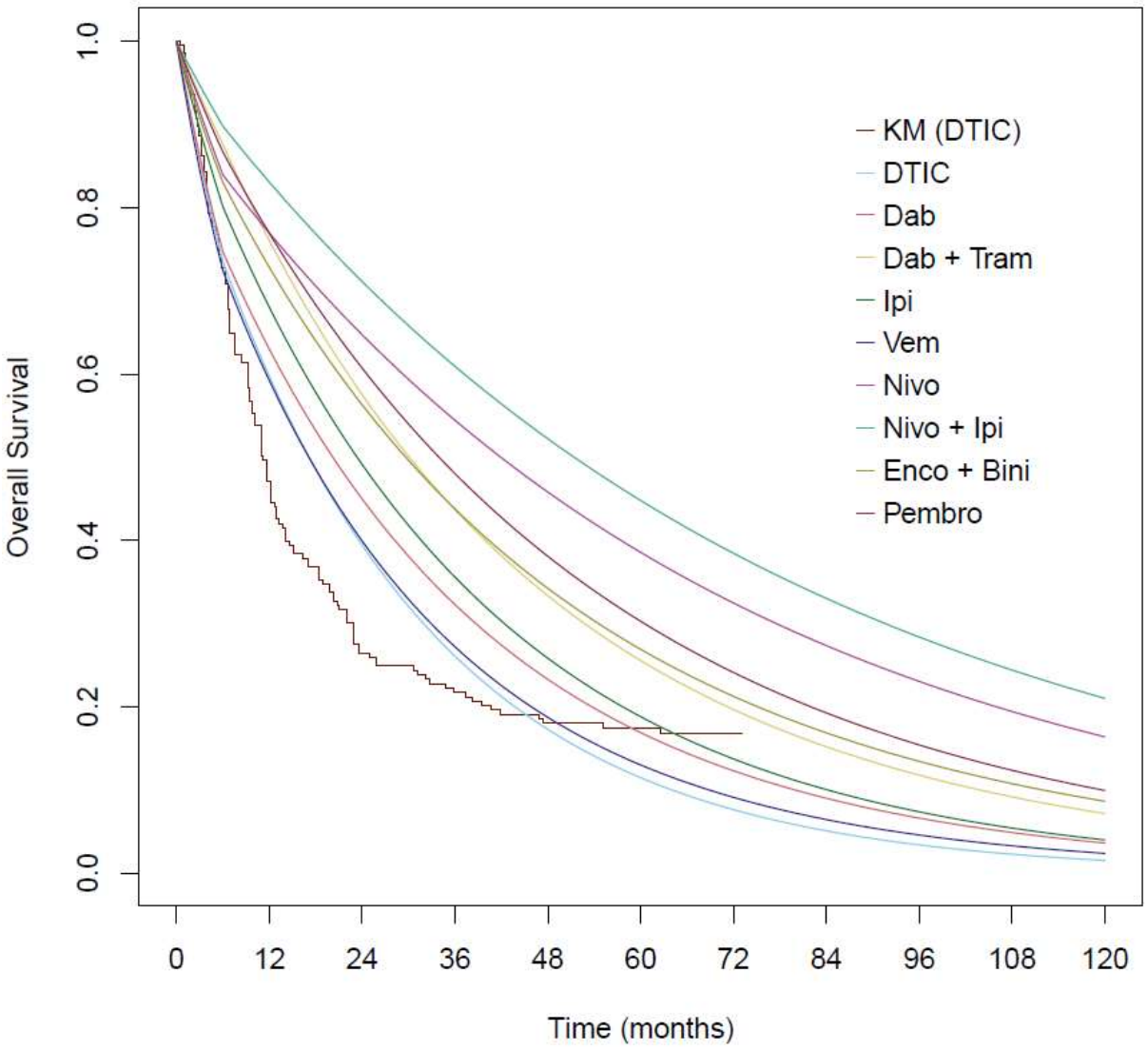


Figure 80: Survival curves for piecewise OS model with one cut point at 6 months for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

E.1.7.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Table 71: Fixed effect PFS & OS NMA results for the piecewise exponential model with the lowest DIC value with 1 cut point for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1	**1-15 months**		**1-6 months**	
ipilimumab vs. DTIC	0.654	(0.482; 0.888)	0.705	(0.401; 1.226)
nivolumab vs. DTIC	0.315	(0.248; 0.398)	0.562	(0.364; 0.853)

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	0.243	(0.176; 0.334)	0.334	(0.177; 0.616)
pembrolizumab vs. DTIC	0.295	(0.203; 0.429)	0.443	(0.215; 0.891)
Interval 2	**16-120 months**		**7-120 months**	
ipilimumab vs. DTIC	1.356	(0.288; 10.528)	0.781	(0.55; 1.098)
nivolumab vs. DTIC	0.637	(0.155; 4.702)	0.417	(0.317; 0.546)
nivolumab + ipilimumab vs. DTIC	0.485	(0.106; 3.743)	0.371	(0.256; 0.531)
pembrolizumab vs. DTIC	1.232	(0.215; 10.794)	0.565	(0.366; 0.877)

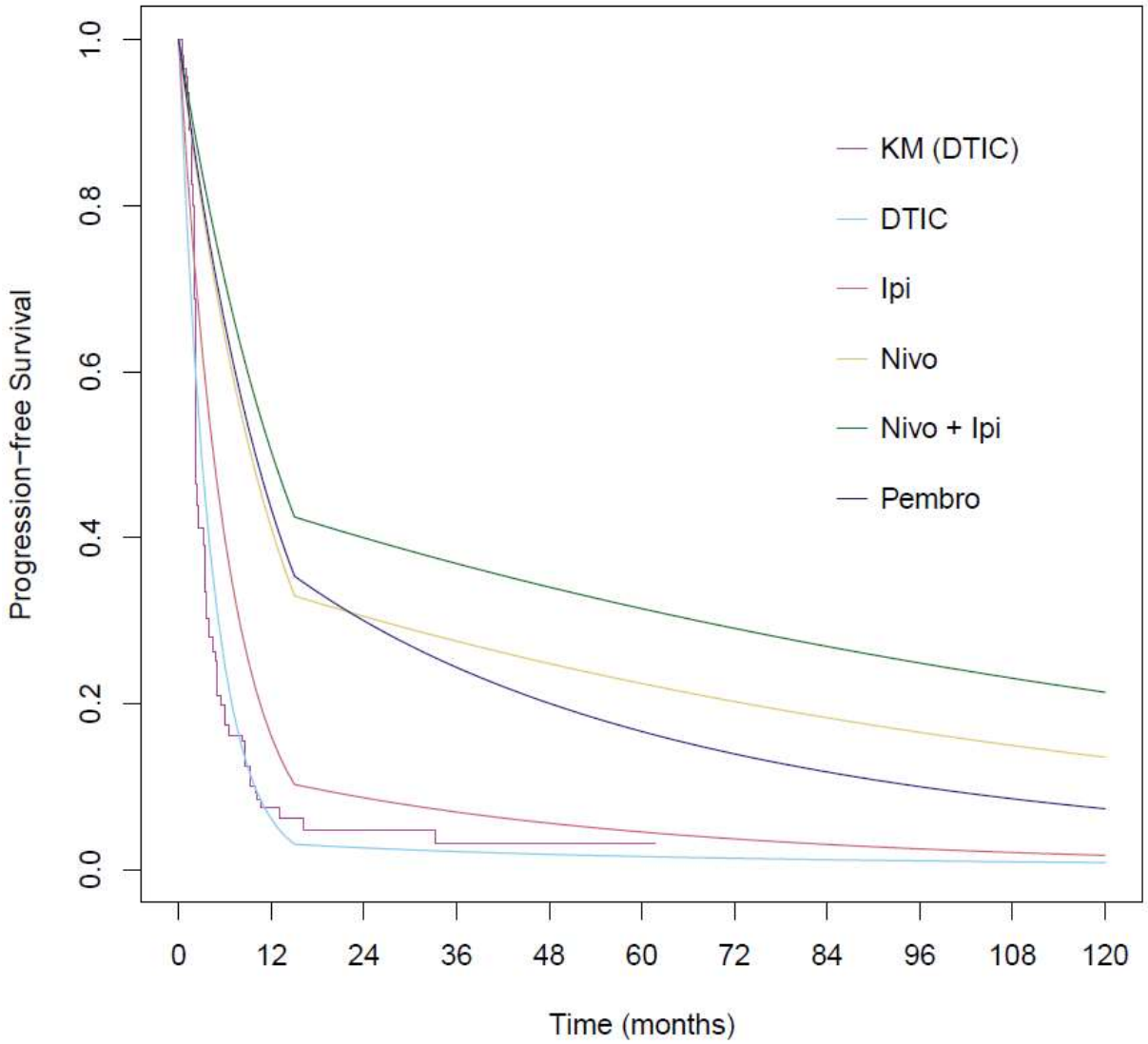


Figure 81: Survival curves for piecewise PFS model with one cut point at 15 months for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

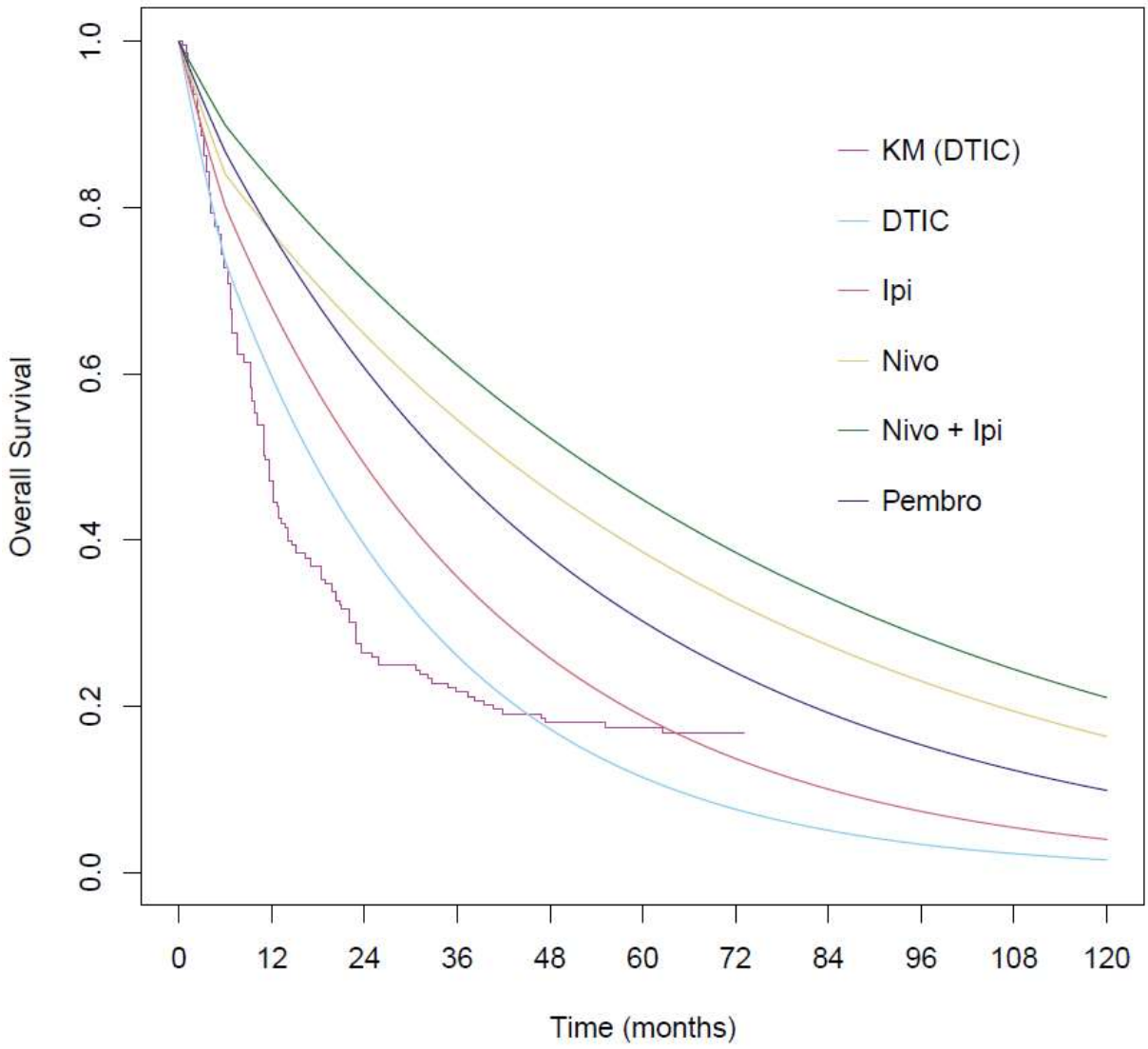


Figure 82: Survival curves for piecewise OS model with one cut point at 6 months for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

E.1.7.7 Impact of different network assumptions on NMA results for piecewise exponential models with one cut point

Table 72: DIC values for all piecewise exponential models with one cut point for PFS for each network

Cut point placement	Targeted therapy + Immunotherapy		Immunotherapy only		
	Network 1	Network 5	Network 3	Network 4	Network 6
One cut point					
6 months	327.9	273.7	112.6	137.4	115.1
9 months	320.1	268.8	110.5	135.4	113.1
12 months	310.2	262.9	108.1	130.6	110.5
15 months	**306.6**	**260.2**	**106.5**	**129.4**	**109.3**

*Smallest DIC value in **bold**.*

Table 73: A comparison of PFS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-15 months				
dabrafenib vs. DTIC	0.411	(0.324; 0.522)	0.406	(0.296; 0.565)
dabrafenib + trametinib vs. DTIC	0.258	(0.207; 0.325)	0.254	(0.171; 0.382)
ipilimumab vs. DTIC	0.656	(0.485; 0.891)	0.656	(0.48; 0.886)
vemurafenib vs. DTIC	0.428	(0.361; 0.506)	0.419	(0.267; 0.664)
nivolumab vs. DTIC	0.314	(0.247; 0.398)	0.315	(0.249; 0.398)
nivolumab + ipilimumab vs. DTIC	0.238	(0.173; 0.327)	0.243	(0.176; 0.333)
encorafenib + binimetinib vs. DTIC	0.232	(0.165; 0.326)	0.227	(0.132; 0.39)
pembrolizumab vs. DTIC	0.296	(0.203; 0.432)	0.296	(0.201; 0.43)
Interval 2: 16-120 months				
dabrafenib vs. DTIC	1.266	(0.313; 7.714)	0.551	(0.133; 3.881)
dabrafenib + trametinib vs. DTIC	1.097	(0.274; 6.449)	0.44	(0.092; 3.304)
ipilimumab vs. DTIC	1.489	(0.314; 11.393)	1.444	(0.33; 10.176)
vemurafenib vs. DTIC	1.55	(0.393; 9.235)	0.588	(0.112; 4.655)
nivolumab vs. DTIC	0.671	(0.164; 4.845)	0.672	(0.173; 4.609)
nivolumab + ipilimumab vs. DTIC	0.501	(0.111; 3.777)	0.512	(0.118; 3.758)
encorafenib + binimetinib vs. DTIC	1.31	(0.284; 8.619)	0.496	(0.082; 4.25)
pembrolizumab vs. DTIC	1.355	(0.24; 11.314)	1.303	(0.247; 10.697)

Table 74: A comparison of PFS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up),

and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-15 months						
ipilimumab vs. DTIC	0.789	(0.567; 1.099)	0.655	(0.486; 0.886)	0.654	(0.482; 0.888)
nivolumab vs. DTIC	0.314	(0.248; 0.397)	0.313	(0.248; 0.396)	0.315	(0.248; 0.398)
nivolumab + ipilimumab vs. DTIC	0.285	(0.2; 0.406)	0.238	(0.174; 0.326)	0.243	(0.176; 0.334)
pembrolizumab vs. DTIC	0.356	(0.24; 0.531)	0.296	(0.204; 0.428)	0.295	(0.203; 0.429)
Interval 2: 16-120 months						
ipilimumab vs. DTIC	1.41	(0.304; 10.299)	1.413	(0.301; 11.811)	1.356	(0.288; 10.528)
nivolumab vs. DTIC	0.597	(0.156; 3.717)	0.64	(0.155; 4.993)	0.637	(0.155; 4.702)
nivolumab + ipilimumab vs. DTIC	0.496	(0.108; 3.48)	0.476	(0.103; 3.92)	0.485	(0.106; 3.743)
pembrolizumab vs. DTIC	1.28	(0.229; 10.783)	1.279	(0.225; 11.577)	1.232	(0.215; 10.794)

Table 75: DIC values for all piecewise exponential models with one cut point for OS for each network

Cut point placement	Targeted therapy + Immunotherapy		Immunotherapy only		
	Network 1	Network 5	Network 3	Network 4	Network 6
One cut point					
6 months	**312**	**255.7**	**111.8**	137.6	**113**
9 months	318.2	263.7	112.8	138.1	114.7
12 months	320	265.9	112.7	137.9	114.8
15 months	320	266.1	112.2	**137.3**	114.4

*Smallest DIC value in **bold**.*

Table 76: A comparison of OS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months				

Comparison	Network 1		Network 5	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
dabrafenib vs. DTIC	0.687	(0.385; 1.256)	0.897	(0.423; 2.051)
dabrafenib + trametinib vs. DTIC	0.237	(0.135; 0.411)	0.378	(0.133; 1.111)
ipilimumab vs. DTIC	0.666	(0.372; 1.186)	0.705	(0.392; 1.264)
vemurafenib vs. DTIC	0.491	(0.363; 0.661)	0.927	(0.276; 3.274)
nivolumab vs. DTIC	0.563	(0.361; 0.867)	0.565	(0.365; 0.867)
nivolumab + ipilimumab vs. DTIC	0.377	(0.203; 0.691)	0.338	(0.177; 0.635)
encorafenib + binimetinib vs. DTIC	0.258	(0.109; 0.591)	0.487	(0.113; 2.116)
pembrolizumab vs. DTIC	0.418	(0.201; 0.872)	0.443	(0.213; 0.933)
Interval 2: 7-120 months				
dabrafenib vs. DTIC	0.842	(0.644; 1.106)	0.806	(0.559; 1.197)
dabrafenib + trametinib vs. DTIC	0.706	(0.555; 0.9)	0.665	(0.431; 1.047)
ipilimumab vs. DTIC	0.777	(0.539; 1.12)	0.781	(0.543; 1.119)
vemurafenib vs. DTIC	1.013	(0.838; 1.229)	0.944	(0.584; 1.539)
nivolumab vs. DTIC	0.417	(0.315; 0.553)	0.416	(0.315; 0.549)
nivolumab + ipilimumab vs. DTIC	0.377	(0.258; 0.549)	0.371	(0.254; 0.54)
encorafenib + binimetinib vs. DTIC	0.652	(0.471; 0.906)	0.608	(0.347; 1.06)
pembrolizumab vs. DTIC	0.562	(0.358; 0.889)	0.565	(0.358; 0.886)

Table 77: A comparison of OS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up),

and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months						
ipilimumab vs. DTIC	0.691	(0.375; 1.273)	0.663	(0.366; 1.184)	0.705	(0.401; 1.226)
nivolumab vs. DTIC	0.565	(0.363; 0.868)	0.562	(0.361; 0.866)	0.562	(0.364; 0.853)
nivolumab + ipilimumab vs. DTIC	0.621	(0.334; 1.148)	0.376	(0.201; 0.689)	0.334	(0.177; 0.616)
pembrolizumab vs. DTIC	0.342	(0.165; 0.714)	0.415	(0.197; 0.867)	0.443	(0.215; 0.891)
Interval 2: 7-120 months						
ipilimumab vs. DTIC	0.778	(0.519; 1.157)	0.778	(0.545; 1.111)	0.781	(0.55; 1.098)
nivolumab vs. DTIC	0.417	(0.314; 0.549)	0.417	(0.316; 0.55)	0.417	(0.317; 0.546)
nivolumab + ipilimumab vs. DTIC	0.331	(0.215; 0.505)	0.378	(0.261; 0.549)	0.371	(0.256; 0.531)
pembrolizumab vs. DTIC	0.717	(0.435; 1.18)	0.563	(0.359; 0.886)	0.565	(0.366; 0.877)

E.1.8 Two cut points – NMA results

E.1.8.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Table 78: Fixed effect OS NMA results for the piecewise exponential model with the lowest DIC value with 2 cut points for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		
dabrafenib vs. DTIC	0.762	(0.537; 1.091)
dabrafenib + trametinib vs. DTIC	0.523	(0.381; 0.714)
ipilimumab vs. DTIC	0.668	(0.43; 1.035)
vemurafenib vs. DTIC	0.695	(0.563; 0.854)
nivolumab vs. DTIC	0.478	(0.344; 0.659)
nivolumab + ipilimumab vs. DTIC	0.512	(0.326; 0.799)
encorafenib + binimetinib vs. DTIC	0.435	(0.279; 0.669)
pembrolizumab vs. DTIC	0.391	(0.229; 0.671)

Interval 2: 13-18 months		
dabrafenib vs. DTIC	0.725	(0.433; 1.246)
dabrafenib + trametinib vs. DTIC	0.65	(0.398; 1.064)
ipilimumab vs. DTIC	0.519	(0.219; 1.223)
vemurafenib vs. DTIC	1.042	(0.708; 1.539)
nivolumab vs. DTIC	0.366	(0.179; 0.74)
nivolumab + ipilimumab vs. DTIC	0.179	(0.07; 0.45)
encorafenib + binimetinib vs. DTIC	0.622	(0.299; 1.272)
pembrolizumab vs. DTIC	0.394	(0.14; 1.112)
Interval 3: 19-120 months		
dabrafenib vs. DTIC	0.931	(0.618; 1.435)
dabrafenib + trametinib vs. DTIC	0.75	(0.51; 1.107)
ipilimumab vs. DTIC	1.205	(0.716; 2.032)
vemurafenib vs. DTIC	1.027	(0.742; 1.415)
nivolumab vs. DTIC	0.597	(0.395; 0.91)
nivolumab + ipilimumab vs. DTIC	0.449	(0.26; 0.775)
encorafenib + binimetinib vs. DTIC	0.757	(0.45; 1.275)
pembrolizumab vs. DTIC	1.224	(0.612; 2.465)

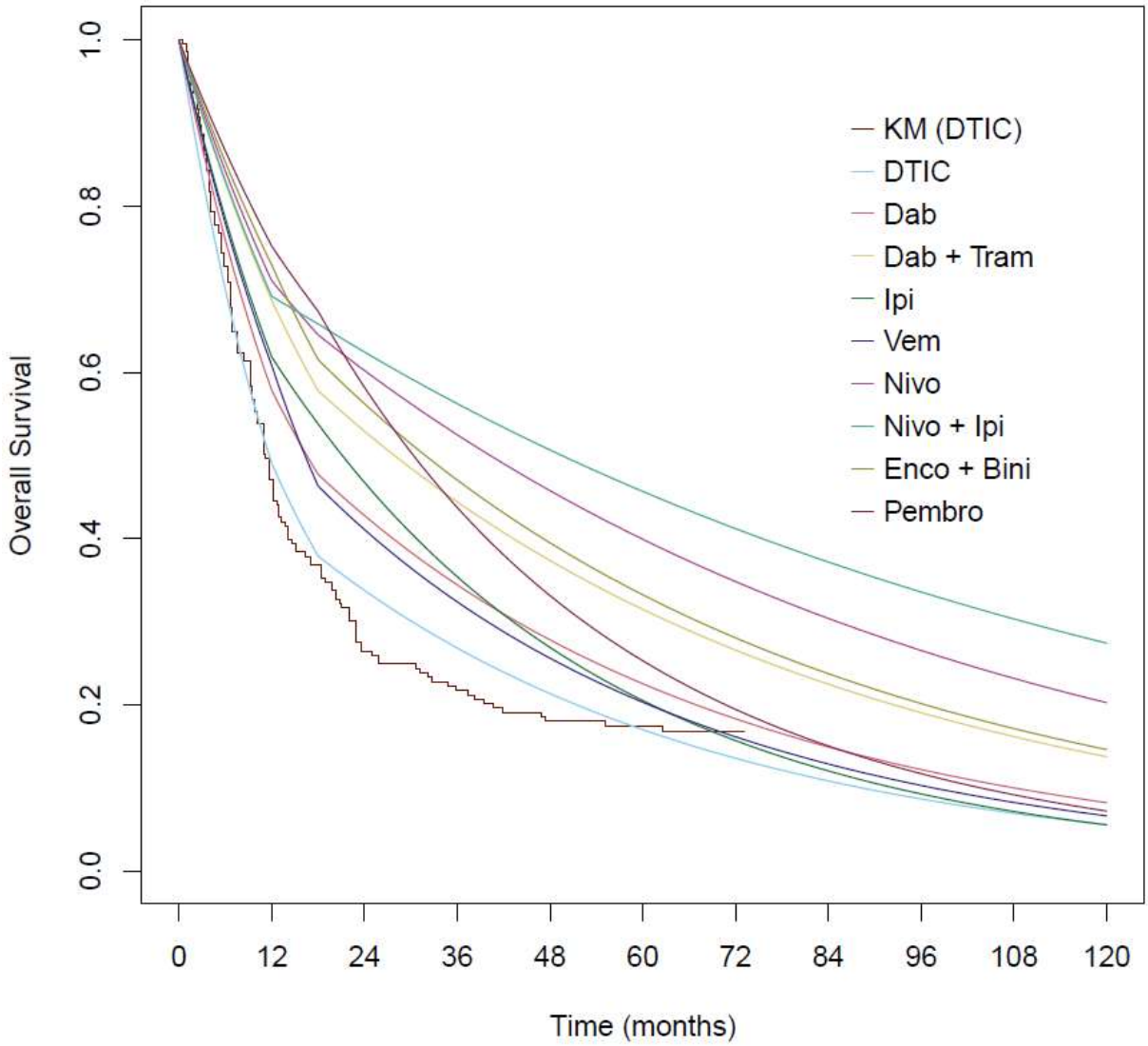


Figure 83: Survival curves for piecewise OS model with two cut points at 12 and 18 months for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

E.1.8.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

E.1.8.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 79: Fixed effect OS NMA results for the piecewise exponential model with the lowest DIC value with 2 cut points for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		
ipilimumab vs. DTIC	0.657	(0.415; 1.036)

nivolumab vs. DTIC	0.477	(0.345; 0.651)
nivolumab + ipilimumab vs. DTIC	0.509	(0.314; 0.815)
pembrolizumab vs. DTIC	0.406	(0.236; 0.701)
Interval 2: 13-18 months		
ipilimumab vs. DTIC	0.536	(0.221; 1.312)
nivolumab vs. DTIC	0.368	(0.181; 0.744)
nivolumab + ipilimumab vs. DTIC	0.24	(0.092; 0.623)
pembrolizumab vs. DTIC	0.401	(0.139; 1.191)
Interval 3: 19-120 months		
ipilimumab vs. DTIC	1.185	(0.653; 2.18)
nivolumab vs. DTIC	0.593	(0.392; 0.903)
nivolumab + ipilimumab vs. DTIC	0.467	(0.25; 0.878)
pembrolizumab vs. DTIC	1.531	(0.709; 3.391)

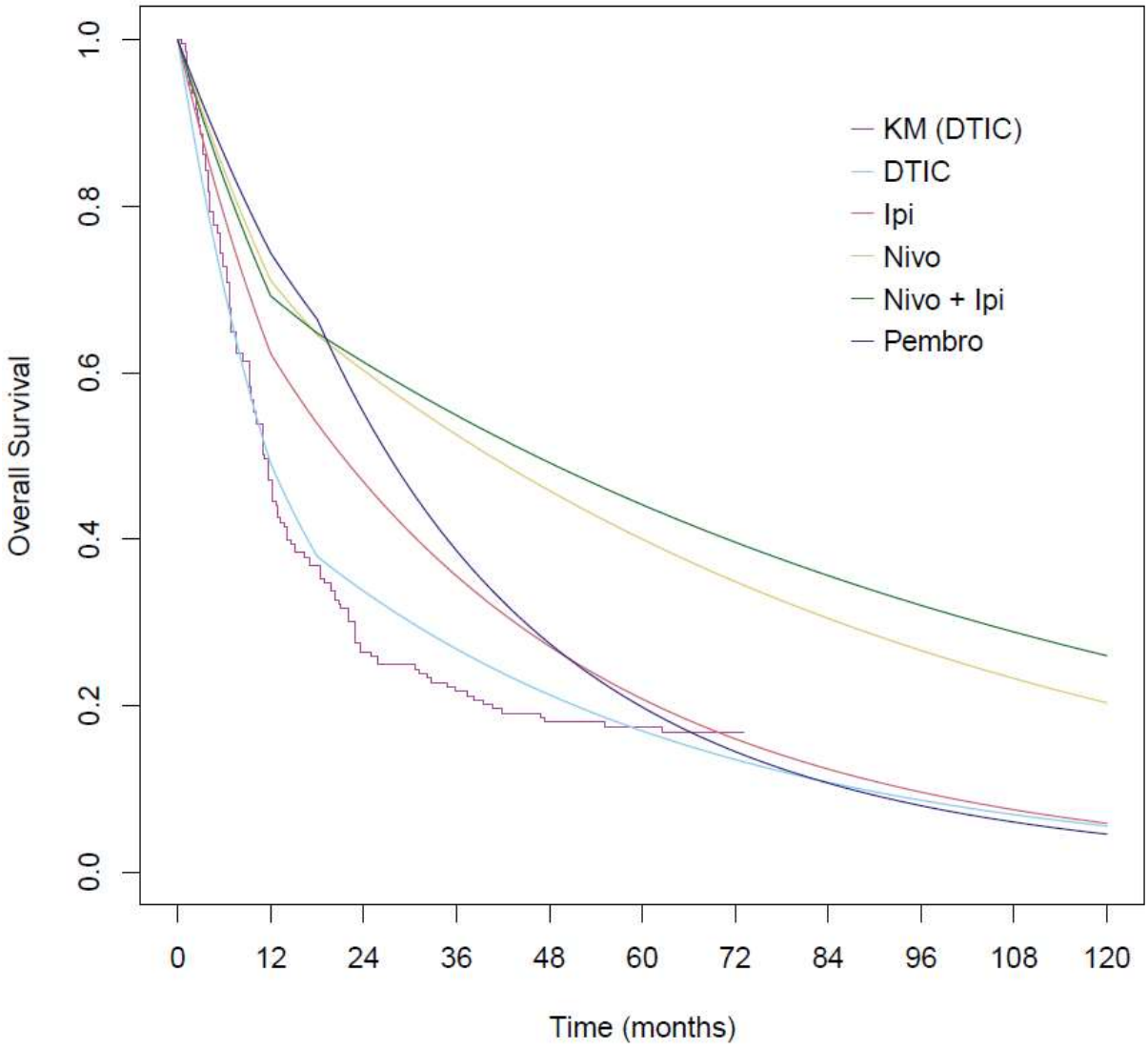


Figure 84: Survival curves for piecewise OS model with two cut points at 12 and 18 months for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

E.1.8.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Table 80: Fixed effect OS NMA results for the piecewise exponential model with the lowest DIC value with 2 cut point for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		
ipilimumab vs. DTIC	0.664	(0.43; 1.022)
nivolumab vs. DTIC	0.476	(0.344; 0.653)
nivolumab + ipilimumab vs. DTIC	0.509	(0.327; 0.793)
pembrolizumab vs. DTIC	0.388	(0.229; 0.663)
Interval 2: 13-18 months		
ipilimumab vs. DTIC	0.525	(0.217; 1.233)

nivolumab vs. DTIC	0.37	(0.176; 0.753)
nivolumab + ipilimumab vs. DTIC	0.18	(0.069; 0.459)
pembrolizumab vs. DTIC	0.401	(0.141; 1.146)
Interval 3: 19-120 months		
ipilimumab vs. DTIC	1.204	(0.704; 2.053)
nivolumab vs. DTIC	0.596	(0.393; 0.91)
nivolumab + ipilimumab vs. DTIC	0.448	(0.257; 0.778)
pembrolizumab vs. DTIC	1.225	(0.606; 2.535)

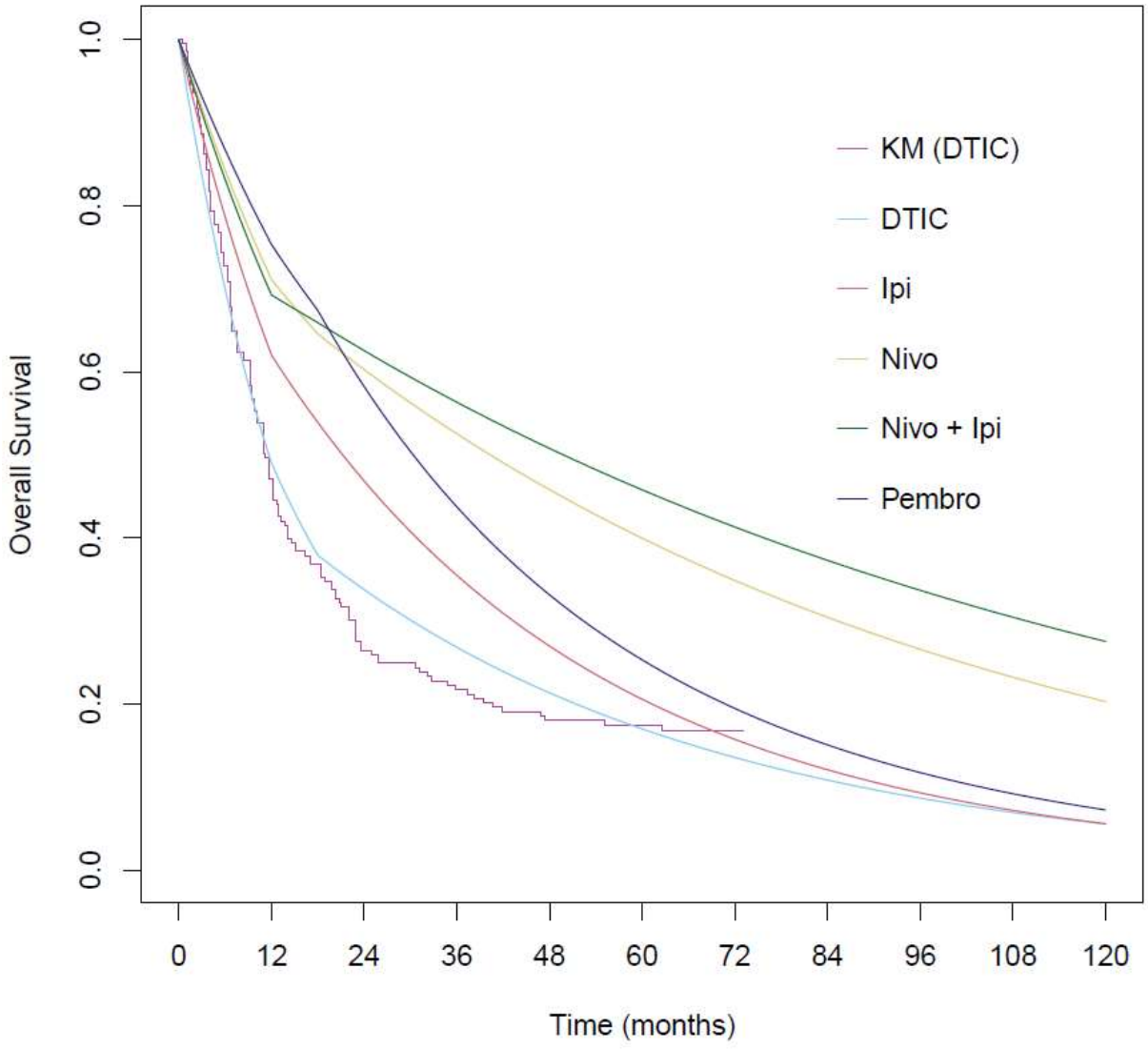


Figure 85: Survival curves for piecewise OS model with two cut points at 12 and 18 months for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

E.1.8.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Table 81: Fixed effect OS NMA results for the piecewise exponential model with the lowest DIC value with 2 cut points for network 5 - people with *BRAF*

wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months		
dabrafenib vs. DTIC	0.902	(0.426; 2.052)
dabrafenib + trametinib vs. DTIC	0.386	(0.138; 1.092)
ipilimumab vs. DTIC	0.699	(0.395; 1.245)
vemurafenib vs. DTIC	0.948	(0.279; 3.228)
nivolumab vs. DTIC	0.562	(0.361; 0.868)
nivolumab + ipilimumab vs. DTIC	0.336	(0.176; 0.628)
encorafenib + binimetinib vs. DTIC	0.501	(0.117; 2.126)
pembrolizumab vs. DTIC	0.438	(0.213; 0.917)
Interval 2: 7-12 months		
dabrafenib vs. DTIC	0.788	(0.408; 1.574)
dabrafenib + trametinib vs. DTIC	0.667	(0.304; 1.486)
ipilimumab vs. DTIC	0.628	(0.319; 1.206)
vemurafenib vs. DTIC	0.78	(0.328; 1.855)
nivolumab vs. DTIC	0.384	(0.237; 0.614)
nivolumab + ipilimumab vs. DTIC	0.704	(0.362; 1.358)
encorafenib + binimetinib vs. DTIC	0.491	(0.187; 1.305)
pembrolizumab vs. DTIC	0.335	(0.152; 0.737)
Interval 3: 13-120 months		
dabrafenib vs. DTIC	0.856	(0.539; 1.386)
dabrafenib + trametinib vs. DTIC	0.708	(0.416; 1.224)
ipilimumab vs. DTIC	0.972	(0.627; 1.523)
vemurafenib vs. DTIC	1.055	(0.584; 1.928)
nivolumab vs. DTIC	0.511	(0.362; 0.731)
nivolumab + ipilimumab vs. DTIC	0.327	(0.204; 0.524)
encorafenib + binimetinib vs. DTIC	0.72	(0.362; 1.425)
pembrolizumab vs. DTIC	0.874	(0.5; 1.572)

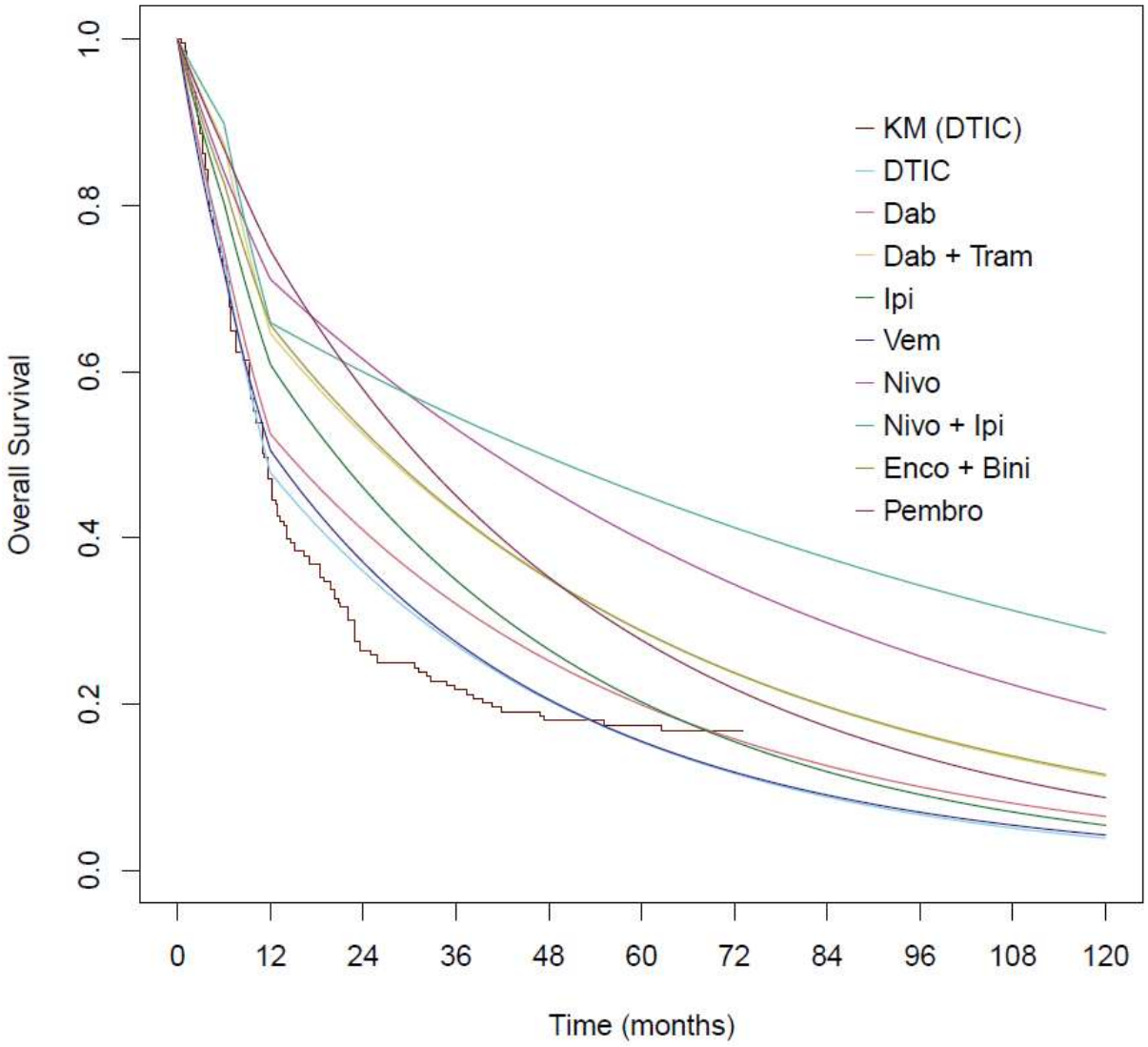


Figure 86: Survival curves for piecewise OS model with two cut points at 6 and 12 months for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

E.1.8.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Table 82: Fixed effect OS NMA results for the piecewise exponential model with the lowest DIC value with 2 cut points for network 6 – people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months		
ipilimumab vs. DTIC	0.667	(0.432; 1.019)
nivolumab vs. DTIC	0.477	(0.345; 0.652)
nivolumab + ipilimumab vs. DTIC	0.51	(0.324; 0.791)
pembrolizumab vs. DTIC	0.39	(0.23; 0.664)
Interval 2: 13-18 months		

ipilimumab vs. DTIC	0.518	(0.219; 1.232)
nivolumab vs. DTIC	0.362	(0.176; 0.735)
nivolumab + ipilimumab vs. DTIC	0.166	(0.063; 0.433)
pembrolizumab vs. DTIC	0.392	(0.141; 1.123)
Interval 3: 19-120 months		
ipilimumab vs. DTIC	1.215	(0.713; 2.069)
nivolumab vs. DTIC	0.593	(0.39; 0.901)
nivolumab + ipilimumab vs. DTIC	0.431	(0.245; 0.749)
pembrolizumab vs. DTIC	1.231	(0.61; 2.529)

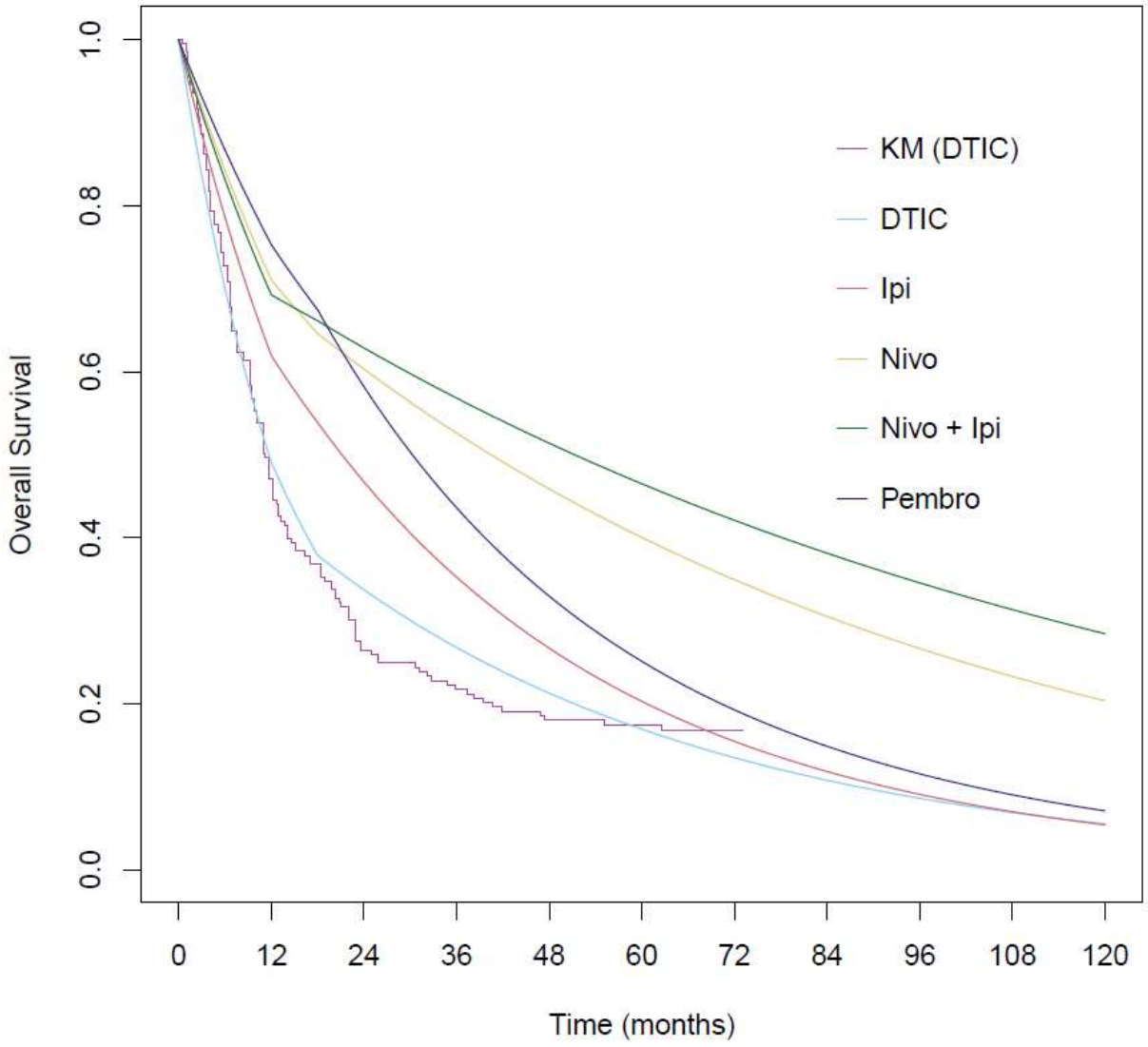


Figure 87: Survival curves for piecewise OS model with two cut points at 12 and 18 months for network 6 – people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

E.1.8.7 Impact of different network assumptions on NMA results for piecewise exponential models with two cut points

Table 83: DIC values for all piecewise exponential models with two cut points for PFS for each network

Cut point placement	Targeted therapy + Immunotherapy		Immunotherapy only		
	Network 1	Network 5	Network 3	Network 4	Network 6
Two cut points					
6 and 12 months	448.9	379.3	154.1	187.2	158
9 and 15 months	435.9	370	149.7	182.1	154
9 and 18 months	423.8	367.3	149.1	175.8	153.6

Cut point placement	Targeted therapy + Immunotherapy		Immunotherapy only		
	Network 1	Network 5	Network 3	Network 4	Network 6
12 and 18 months	**404.3**	352.1	**143.5**	**167.4**	**147.8**
12 and 20 months	[b]	348[c]	143.9	168.5	148.2
12 and 24 months	[a]	**347.5**[c]	144.6	[a]	148.7

(b) The shortest amount of follow-up time in these networks is 22 months (BRIM-3) and 24 months (CheckMate 069) for PFS and OS respectively. As such, it is not possible to generate aggregate data beyond these time points. Thus, for this network, it's impossible to run piecewise models with cut points beyond 22 or 24 months for PFS and OS respectively.

(c) Although below the 22 months, using 20 months as the second cut placement resulted in the final interval, >20 months, of the BRIM-3 aggregate data having one arm with no data. Thus, it's impossible to run the PFS piecewise model with a second interval at 20 months

(d) Although the piecewise models with cuts at 12 & 20 months, and 12 & 24 months have the smallest and second smallest DIC value respectively for network 5, neither of these return plausible estimates of PFS beyond 20 months for any of the targeted therapies. Specifically, the DTIC arm of BREAK-3 has no event after 20 months, so it is not possible to obtain credible estimates. This in turn affects the estimates for dabrafenib+trametinib, vemurafenib monotherapy, encorafenib+binimetinib as these treatments are connected in the network and therefore the network uses the unreliable estimates between dabrafenib and DTIC to inform them. As such, we considered the model with two cuts at 12 & 18 months to be the best fitting PFS model for network 5.

Table 84: A comparison of PFS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months				
dabrafenib vs. DTIC	0.399	(0.313; 0.511)	0.4	(0.29; 0.559)
dabrafenib + trametinib vs. DTIC	0.268	(0.212; 0.338)	0.269	(0.18; 0.404)
ipilimumab vs. DTIC	0.692	(0.506; 0.94)	0.687	(0.507; 0.938)
vemurafenib vs. DTIC	0.439	(0.37; 0.519)	0.44	(0.279; 0.702)
nivolumab vs. DTIC	0.351	(0.275; 0.446)	0.351	(0.277; 0.445)
nivolumab + ipilimumab vs. DTIC	0.267	(0.193; 0.367)	0.269	(0.196; 0.372)
encorafenib + binimetinib vs. DTIC	0.236	(0.166; 0.336)	0.238	(0.138; 0.415)
pembrolizumab vs. DTIC	0.313	(0.214; 0.456)	0.311	(0.213; 0.454)
Interval 2: 13-18 months				

Comparison	Network 1		Network 5	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
dabrafenib vs. DTIC	1.718	(0.474; 7.846)	1.631	(0.22; 67.627)
dabrafenib + trametinib vs. DTIC	0.937	(0.276; 4.129)	0.862	(0.102; 39.134)
ipilimumab vs. DTIC	0.451	(0.064; 4.433)	0.402	(0.056; 3.633)
vemurafenib vs. DTIC	1.167	(0.366; 4.826)	1.046	(0.113; 46.805)
nivolumab vs. DTIC	0.135	(0.023; 1.099)	0.131	(0.023; 0.956)
nivolumab + ipilimumab vs. DTIC	0.138	(0.019; 1.353)	0.138	(0.02; 1.194)
encorafenib + binimetinib vs. DTIC	0.753	(0.174; 3.955)	0.669	(0.06; 30.447)
pembrolizumab vs. DTIC	0.461	(0.05; 5.54)	0.415	(0.045; 4.904)
Interval 3: 19-120 months				
dabrafenib vs. DTIC	0.567	(0.141; 3.963)	0.455	(0.12; 3.083)
dabrafenib + trametinib vs. DTIC	0.514	(0.115; 3.931)	0.402	(0.092; 2.869)
ipilimumab vs. DTIC	2.356	(0.363; 26.233)	2.635	(0.337; 40.731)
vemurafenib vs. DTIC	0.797	(0.17; 6.366)	0.613	(0.126; 4.716)
nivolumab vs. DTIC	1.373	(0.232; 14.44)	1.536	(0.227; 22.511)
nivolumab + ipilimumab vs. DTIC	0.941	(0.145; 10.444)	1.055	(0.141; 15.927)
encorafenib + binimetinib vs. DTIC	0.789	(0.14; 6.987)	0.607	(0.103; 5.392)
pembrolizumab vs. DTIC	2.509	(0.31; 32.819)	2.763	(0.288; 50.199)

Table 85: A comparison of PFS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up),

and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months						
ipilimumab vs. DTIC	0.827	(0.592; 1.15)	0.691	(0.509; 0.939)	0.686	(0.507; 0.928)
nivolumab vs. DTIC	0.351	(0.276; 0.444)	0.351	(0.276; 0.444)	0.351	(0.277; 0.443)
nivolumab + ipilimumab vs. DTIC	0.31	(0.216; 0.444)	0.267	(0.193; 0.368)	0.269	(0.194; 0.37)
pembrolizumab vs. DTIC	0.374	(0.25; 0.558)	0.312	(0.214; 0.457)	0.31	(0.213; 0.451)
Interval 2: 13-18 months						
ipilimumab vs. DTIC	0.721	(0.067; 9.365)	0.435	(0.06; 3.732)	0.427	(0.06; 3.747)
nivolumab vs. DTIC	0.137	(0.022; 1.1)	0.131	(0.023; 0.968)	0.136	(0.024; 0.986)
nivolumab + ipilimumab vs. DTIC	0.253	(0.026; 3.086)	0.133	(0.018; 1.172)	0.145	(0.02; 1.266)
pembrolizumab vs. DTIC	0.73	(0.057; 12.025)	0.449	(0.049; 5.629)	0.443	(0.047; 5.15)
Interval 3: 19-120 months						
ipilimumab vs. DTIC	2.191	(0.285; 32.492)	2.537	(0.397; 51.111)	2.385	(0.305; 44.523)
nivolumab vs. DTIC	1.335	(0.221; 17.167)	1.46	(0.253; 28.531)	1.4	(0.212; 24.361)
nivolumab + ipilimumab vs. DTIC	1.024	(0.146; 13.943)	1.001	(0.159; 19.648)	0.957	(0.131; 17.832)
pembrolizumab vs. DTIC	2.337	(0.254; 40.609)	2.725	(0.35; 62.992)	2.537	(0.256; 53.678)

Table 86: DIC values for all piecewise exponential models with two cut points for OS for each network

Cut point placement	Targeted therapy + Immunotherapy		Immunotherapy only		
	Network 1	Network 5	Network 3	Network 4	Network 6
Two cut points					
6 and 12 months	451.8	**371.2**	159.9	197.8	162.9
9 and 15 months	452.9	376.2	158.8	194.1	162.3

Cut point placement	Targeted therapy + Immunotherapy		Immunotherapy only		
	Network 1	Network 5	Network 3	Network 4	Network 6
9 and 18 months	453.2	377.7	159.9	194.4	163.4
12 and 18 months	**448.3**	374.8	**157.6**	**191.3**	**161.1**
12 and 20 months	449.6	376.4	158.3	191.6	161.8
12 and 24 months	[a]	377.1	158.3	[a]	162.1

(a) The shortest amount of follow-up time in these networks is 22 months (BRIM-3) and 24 months (CheckMate 069) for PFS and OS respectively. As such, it is not possible to generate aggregate data beyond these time points. Thus, for this network, it's impossible to run piecewise models with cut points beyond 22 or 24 months for PFS and OS respectively.

Table 87: A comparison of OS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5[a]	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months				
dabrafenib vs. DTIC	0.762	(0.537; 1.091)	0.821	(0.502; 1.374)
dabrafenib + trametinib vs. DTIC	0.523	(0.381; 0.714)	0.584	(0.319; 1.082)
ipilimumab vs. DTIC	0.668	(0.43; 1.035)	0.666	(0.429; 1.032)
vemurafenib vs. DTIC	0.695	(0.563; 0.854)	0.798	(0.406; 1.584)
nivolumab vs. DTIC	0.478	(0.344; 0.659)	0.477	(0.344; 0.655)
nivolumab + ipilimumab vs. DTIC	0.512	(0.326; 0.799)	0.509	(0.324; 0.794)
encorafenib + binimetinib vs. DTIC	0.435	(0.279; 0.669)	0.5	(0.231; 1.085)
pembrolizumab vs. DTIC	0.391	(0.229; 0.671)	0.39	(0.227; 0.668)
Interval 2: 13-18 months				
dabrafenib vs. DTIC	0.725	(0.433; 1.246)	0.674	(0.332; 1.478)
dabrafenib + trametinib vs. DTIC	0.65	(0.398; 1.064)	0.588	(0.252; 1.437)

Comparison	Network 1		Network 5[a]	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
ipilimumab vs. DTIC	0.519	(0.219; 1.223)	0.527	(0.222; 1.245)
vemurafenib vs. DTIC	1.042	(0.708; 1.539)	0.916	(0.349; 2.515)
nivolumab vs. DTIC	0.366	(0.179; 0.74)	0.368	(0.179; 0.745)
nivolumab + ipilimumab vs. DTIC	0.179	(0.07; 0.45)	0.169	(0.064; 0.439)
encorafenib + binimetinib vs. DTIC	0.622	(0.299; 1.272)	0.546	(0.173; 1.768)
pembrolizumab vs. DTIC	0.394	(0.14; 1.112)	0.4	(0.144; 1.104)
Interval 3: 19-120 months				
dabrafenib vs. DTIC	0.931	(0.618; 1.435)	1.042	(0.561; 2.062)
dabrafenib + trametinib vs. DTIC	0.75	(0.51; 1.107)	0.864	(0.423; 1.874)
ipilimumab vs. DTIC	1.205	(0.716; 2.032)	1.227	(0.731; 2.074)
vemurafenib vs. DTIC	1.027	(0.742; 1.415)	1.213	(0.556; 2.782)
nivolumab vs. DTIC	0.597	(0.395; 0.91)	0.597	(0.397; 0.903)
nivolumab + ipilimumab vs. DTIC	0.449	(0.26; 0.775)	0.434	(0.251; 0.753)
encorafenib + binimetinib vs. DTIC	0.757	(0.45; 1.275)	0.895	(0.367; 2.239)
pembrolizumab vs. DTIC	1.224	(0.612; 2.465)	1.247	(0.623; 2.575)

(a) Although the two-cut point model with the smallest DIC value for network 5 is the one with cuts at 6 and 12 months, here we present the results for the model with cuts at 12 and 18 months. This is to allow for a meaningful comparison between the results of this model and the results of network 1, so one can see what impact using trials with long-term follow up only has on the NMA results.

Table 88: A comparison of OS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up),

and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-12 months						
ipilimumab vs. DTIC	0.657	(0.415; 1.036)	0.664	(0.43; 1.022)	0.667	(0.432; 1.019)
nivolumab vs. DTIC	0.477	(0.345; 0.651)	0.476	(0.344; 0.653)	0.477	(0.345; 0.652)
nivolumab + ipilimumab vs. DTIC	0.509	(0.314; 0.815)	0.509	(0.327; 0.793)	0.51	(0.324; 0.791)
pembrolizumab vs. DTIC	0.406	(0.236; 0.701)	0.388	(0.229; 0.663)	0.39	(0.23; 0.664)
Interval 2: 13-18 months						
ipilimumab vs. DTIC	0.536	(0.221; 1.312)	0.525	(0.217; 1.233)	0.518	(0.219; 1.232)
nivolumab vs. DTIC	0.368	(0.181; 0.744)	0.37	(0.176; 0.753)	0.362	(0.176; 0.735)
nivolumab + ipilimumab vs. DTIC	0.24	(0.092; 0.623)	0.18	(0.069; 0.459)	0.166	(0.063; 0.433)
pembrolizumab vs. DTIC	0.401	(0.139; 1.191)	0.401	(0.141; 1.146)	0.392	(0.141; 1.123)
Interval 3: 19-120 months						
ipilimumab vs. DTIC	1.185	(0.653; 2.18)	1.204	(0.704; 2.053)	1.215	(0.713; 2.069)
nivolumab vs. DTIC	0.593	(0.392; 0.903)	0.596	(0.393; 0.91)	0.593	(0.39; 0.901)
nivolumab + ipilimumab vs. DTIC	0.467	(0.25; 0.878)	0.448	(0.257; 0.778)	0.431	(0.245; 0.749)
pembrolizumab vs. DTIC	1.531	(0.709; 3.391)	1.225	(0.606; 2.535)	1.231	(0.61; 2.529)

E.1.9 Three cut points – NMA results

E.1.9.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Table 89: Fixed effect PFS & OS NMA results for the best fitting (lowest DIC) piecewise exponential model with 3 cut points for network 1 - people with BRAF

wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months				
dabrafenib vs. DTIC	0.346	(0.259; 0.464)	0.69	(0.386; 1.247)
dabrafenib + trametinib vs. DTIC	0.204	(0.152; 0.274)	0.236	(0.135; 0.409)
ipilimumab vs. DTIC	0.827	(0.598; 1.148)	0.672	(0.38; 1.181)
vemurafenib vs. DTIC	0.342	(0.279; 0.419)	0.491	(0.364; 0.664)
nivolumab vs. DTIC	0.496	(0.386; 0.635)	0.567	(0.367; 0.867)
nivolumab + ipilimumab vs. DTIC	0.346	(0.245; 0.487)	0.38	(0.206; 0.695)
encorafenib + binimetinib vs. DTIC	0.162	(0.104; 0.251)	0.259	(0.109; 0.595)
pembrolizumab vs. DTIC	0.439	(0.292; 0.66)	0.423	(0.206; 0.873)
Interval 2: 7-12 months				
dabrafenib vs. DTIC	0.799	(0.478; 1.352)	0.907	(0.57; 1.453)
dabrafenib + trametinib vs. DTIC	0.569	(0.358; 0.921)	0.813	(0.541; 1.225)
ipilimumab vs. DTIC	0.29	(0.112; 0.738)	0.651	(0.34; 1.248)
vemurafenib vs. DTIC	1.013	(0.682; 1.555)	0.992	(0.729; 1.351)
nivolumab vs. DTIC	0.113	(0.05; 0.242)	0.382	(0.236; 0.61)
nivolumab + ipilimumab vs. DTIC	0.134	(0.052; 0.342)	0.678	(0.356; 1.306)
encorafenib + binimetinib vs. DTIC	0.703	(0.36; 1.397)	0.625	(0.362; 1.068)
pembrolizumab vs. DTIC	0.108	(0.036; 0.325)	0.348	(0.159; 0.767)
Interval 3: 13-18 months				

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
dabrafenib vs. DTIC	1.774	(0.511; 8.045)	0.726	(0.431; 1.235)
dabrafenib + trametinib vs. DTIC	0.963	(0.296; 4.092)	0.65	(0.399; 1.063)
ipilimumab vs. DTIC	0.418	(0.059; 3.593)	0.521	(0.217; 1.217)
vemurafenib vs. DTIC	1.193	(0.385; 4.821)	1.043	(0.708; 1.539)
nivolumab vs. DTIC	0.128	(0.022; 0.912)	0.368	(0.178; 0.745)
nivolumab + ipilimumab vs. DTIC	0.129	(0.018; 1.102)	0.179	(0.07; 0.451)
encorafenib + binimetinib vs. DTIC	0.768	(0.185; 3.947)	0.622	(0.303; 1.291)
pembrolizumab vs. DTIC	0.429	(0.046; 4.938)	0.397	(0.141; 1.114)
Interval 4: 19-120 months				
dabrafenib vs. DTIC	0.653	(0.149; 5.613)	0.939	(0.618; 1.44)
dabrafenib + trametinib vs. DTIC	0.599	(0.122; 5.479)	0.757	(0.515; 1.117)
ipilimumab vs. DTIC	2.631	(0.323; 55.813)	1.198	(0.703; 2.069)
vemurafenib vs. DTIC	0.934	(0.178; 8.846)	1.037	(0.756; 1.434)
nivolumab vs. DTIC	1.523	(0.213; 31.312)	0.593	(0.39; 0.914)
nivolumab + ipilimumab vs. DTIC	1.042	(0.132; 22.488)	0.445	(0.254; 0.789)
encorafenib + binimetinib vs. DTIC	0.932	(0.145; 10.612)	0.765	(0.457; 1.278)
pembrolizumab vs. DTIC	2.768	(0.286; 70.035)	1.214	(0.606; 2.478)

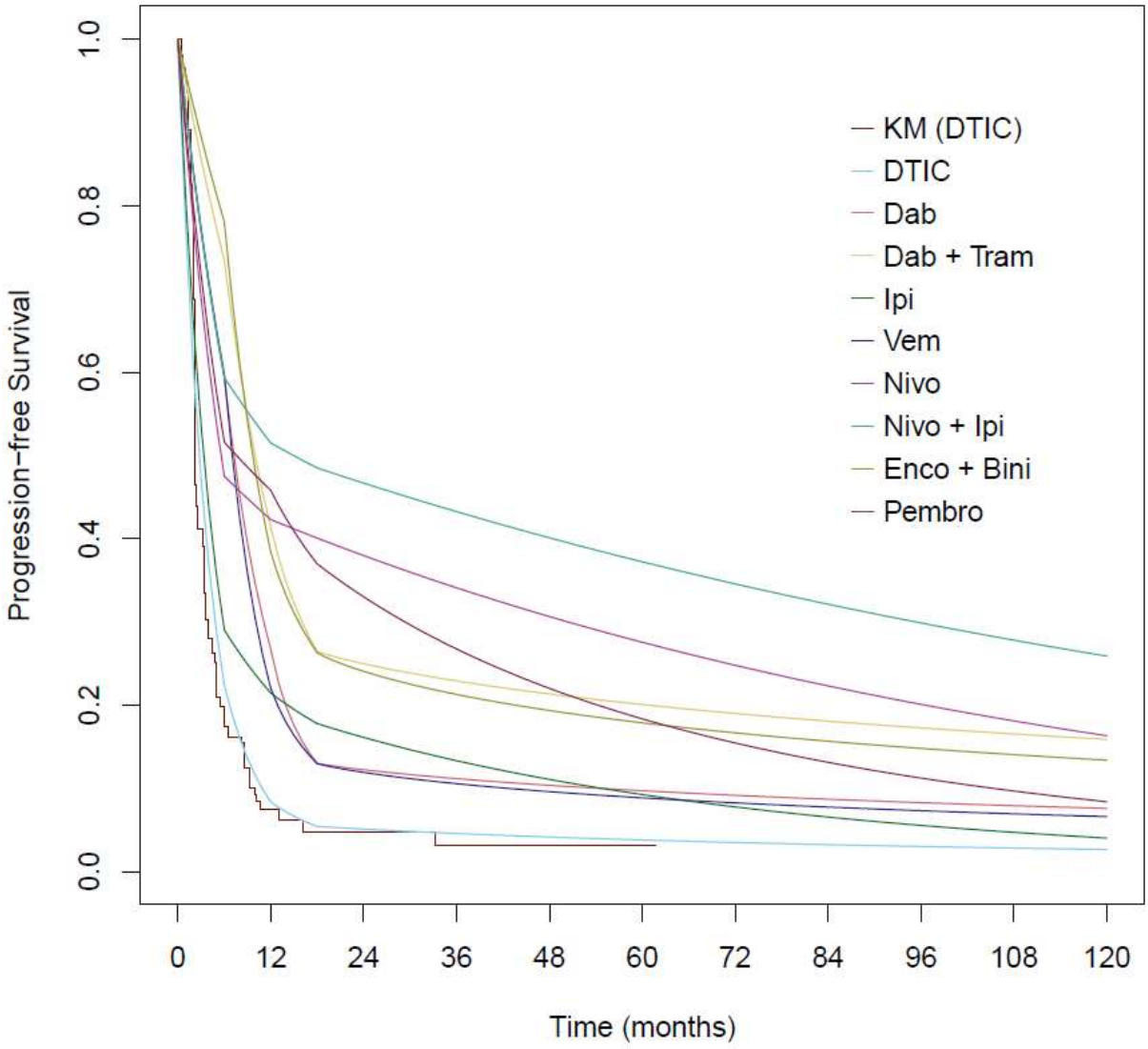


Figure 88: Survival curves for piecewise PFS model with three cut points at 6, 12 and 18 months for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

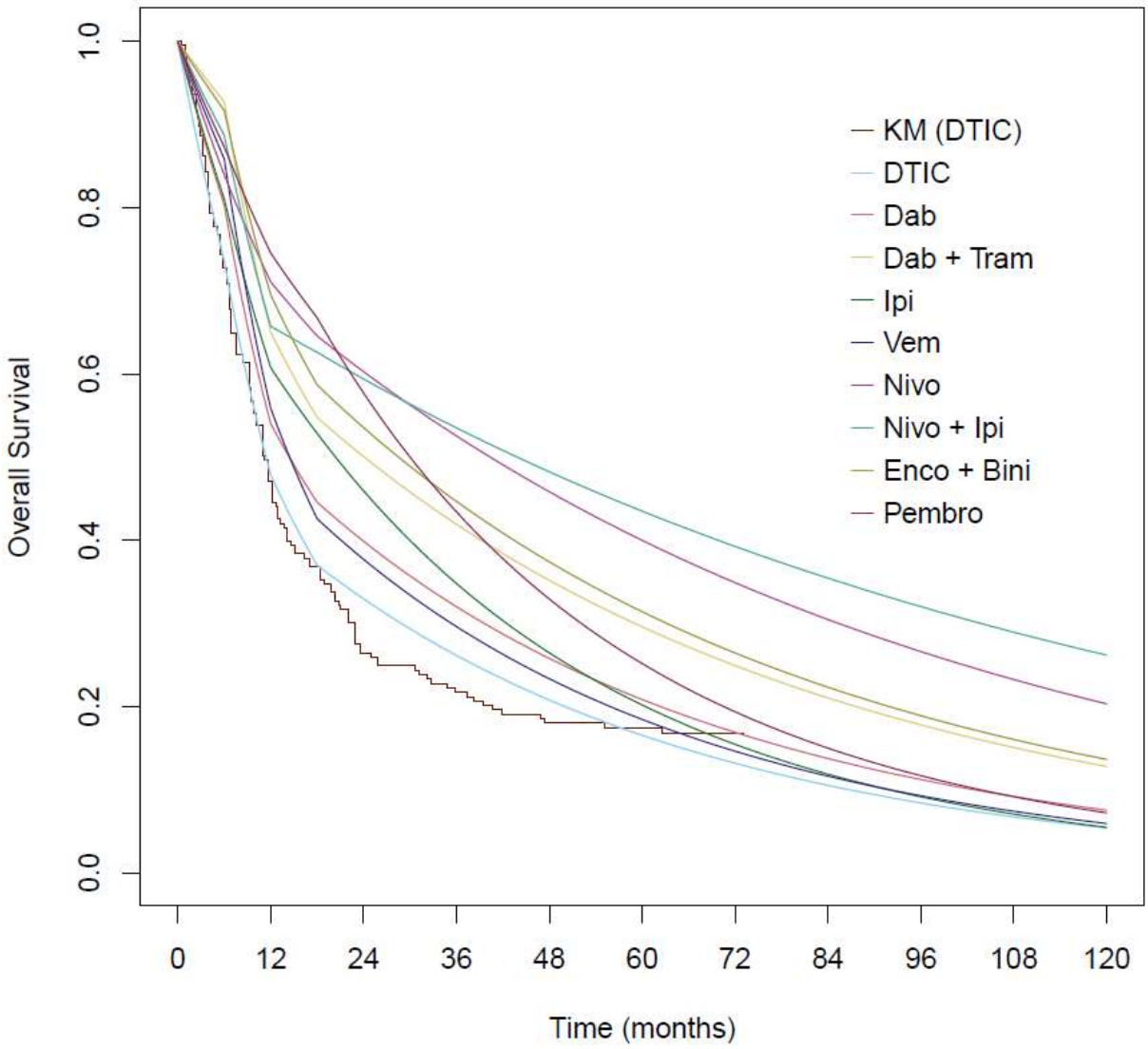


Figure 89: Survival curves for piecewise OS model with three cut points at 6, 12 and 18 months for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

E.1.9.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

E.1.9.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 90: Fixed effect PFS & OS NMA results for the best fitting (lowest DIC) piecewise exponential model with 3 cut points for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1	1-6 months		1-12 months	

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
ipilimumab vs. DTIC	0.968	(0.68; 1.388)	0.66	(0.417; 1.05)
nivolumab vs. DTIC	0.494	(0.385; 0.635)	0.478	(0.346; 0.658)
nivolumab + ipilimumab vs. DTIC	0.398	(0.27; 0.588)	0.51	(0.316; 0.822)
pembrolizumab vs. DTIC	0.512	(0.335; 0.795)	0.408	(0.239; 0.71)
Interval 2	**7-12 months**		**13-24 months**	
ipilimumab vs. DTIC	0.323	(0.108; 0.887)	0.658	(0.343; 1.282)
nivolumab vs. DTIC	0.111	(0.048; 0.239)	0.363	(0.221; 0.589)
nivolumab + ipilimumab vs. DTIC	0.152	(0.053; 0.416)	0.243	(0.116; 0.507)
pembrolizumab vs. DTIC	0.12	(0.035; 0.379)	0.639	(0.281; 1.509)
Interval 3	**13-18 months**		**25-36 months**	
ipilimumab vs. DTIC	0.69	(0.075; 7.576)	0.971	(0.332; 2.956)
nivolumab vs. DTIC	0.132	(0.023; 0.988)	0.633	(0.27; 1.566)
nivolumab + ipilimumab vs. DTIC	0.243	(0.028; 2.569)	0.535	(0.181; 1.633)
pembrolizumab vs. DTIC	0.711	(0.058; 9.718)	1.591	(0.376; 7.561)
Interval 4	**19-120 months**		**37-120 months**	
ipilimumab vs. DTIC	2.977	(0.338; 65.759)	2.268	(0.779; 7.029)
nivolumab vs. DTIC	1.806	(0.252; 38.784)	1.212	(0.584; 2.798)
nivolumab + ipilimumab vs. DTIC	1.398	(0.171; 31.312)	0.901	(0.294; 2.91)
pembrolizumab vs. DTIC	3.146	(0.292; 72.24)	2.146	(0.583; 8.44)

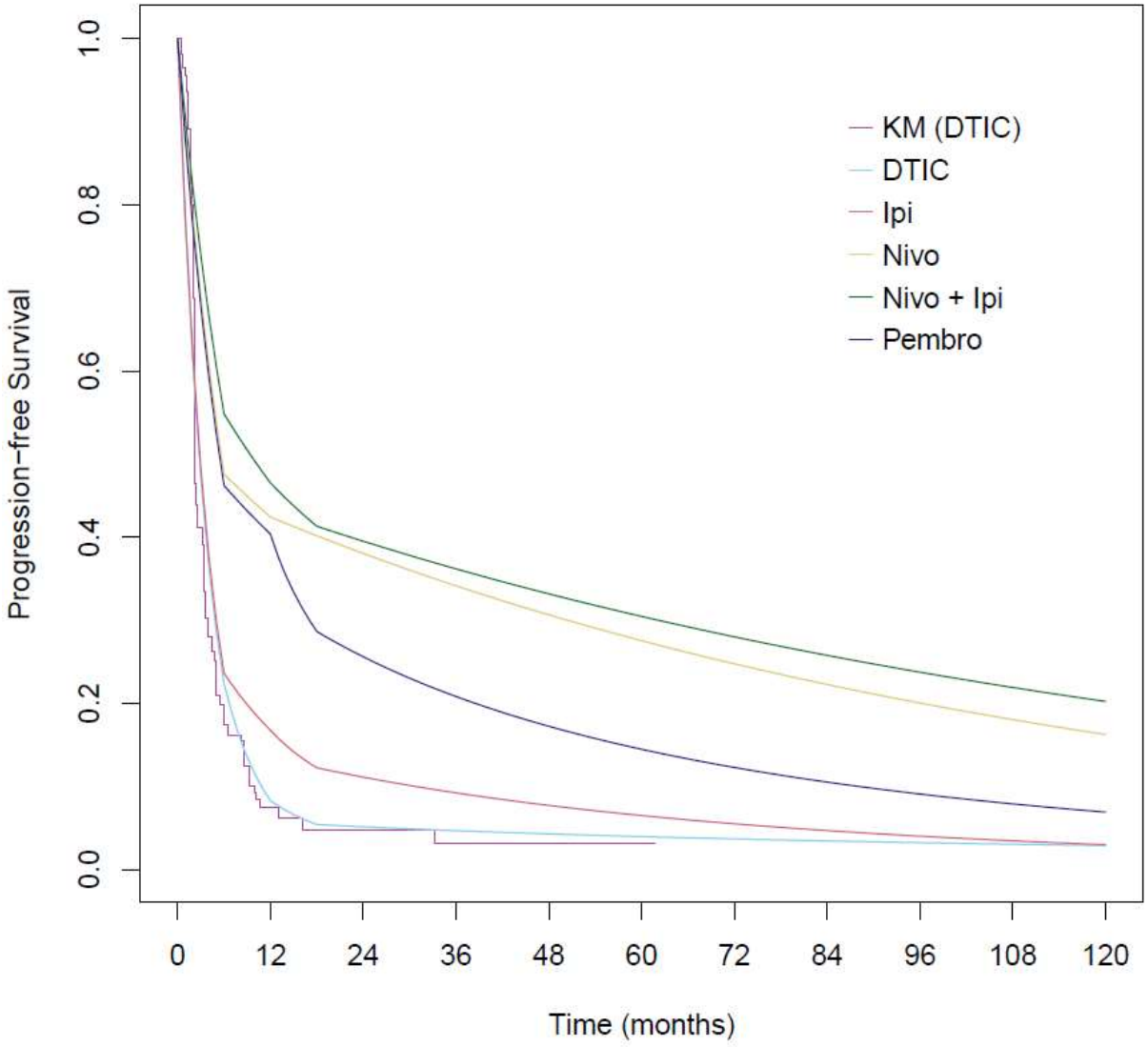


Figure 90: Survival curves for piecewise PFS model with three cut points at 6, 12 and 18 months for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

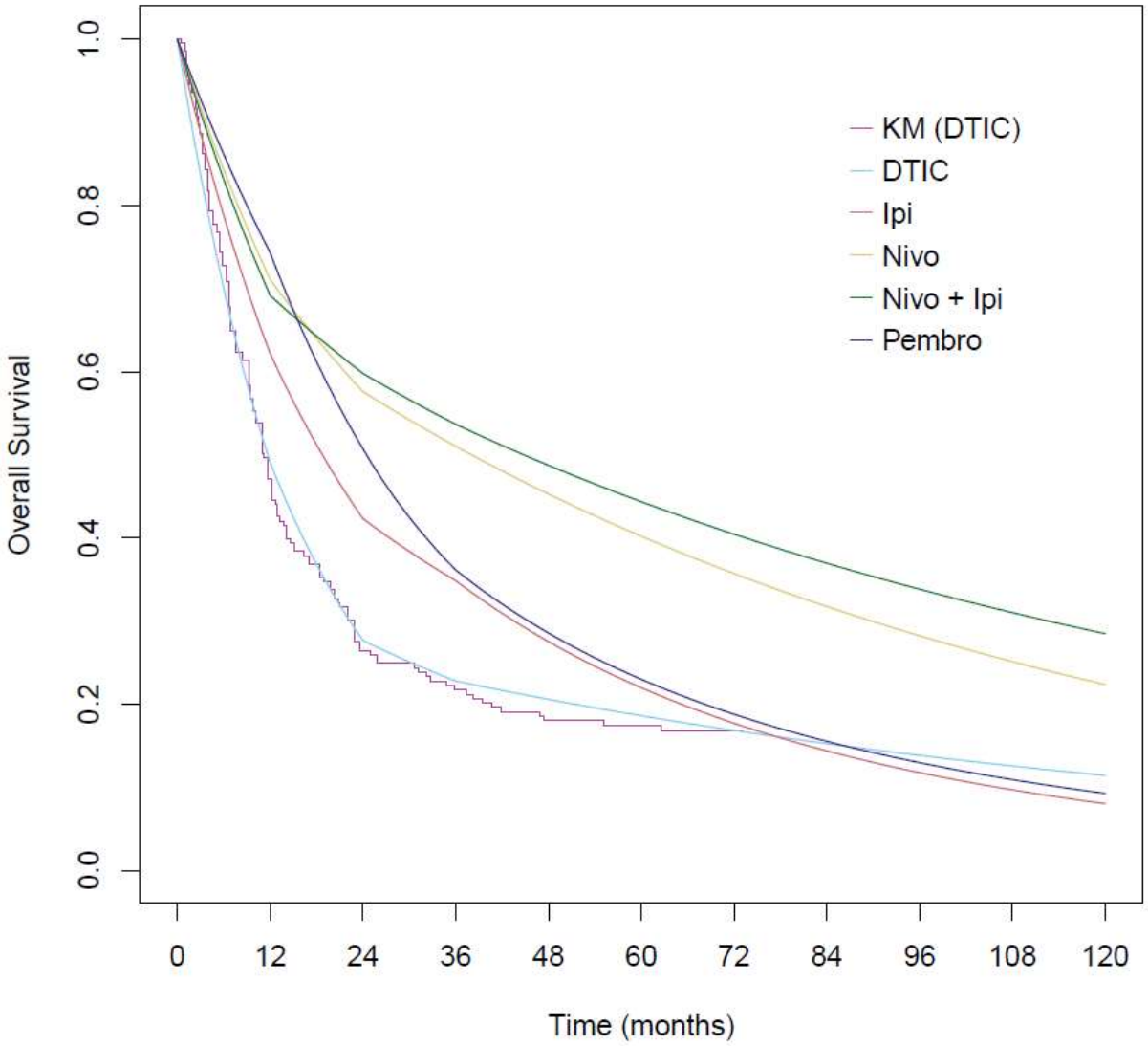


Figure 91: Survival curves for piecewise OS model with three cut points at 12, 24 and 36 months for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

E.1.9.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

As shown in Table 64, it was not possible to run a piecewise model with 3 cuts at 12, 24 and 36 months for PFS, and the model with cuts at 6, 12, and 18 months did not converge. Therefore, there are no results to present for a piecewise model with 3 cuts for PFS.

Table 91: Fixed effect OS NMA results for the best fitting (lowest DIC) piecewise exponential model with 3 cut point for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Comparison	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months		
ipilimumab vs. DTIC	0.665	(0.373; 1.175)
nivolumab vs. DTIC	0.563	(0.364; 0.864)
nivolumab + ipilimumab vs. DTIC	0.377	(0.205; 0.689)

pembrolizumab vs. DTIC	0.417	(0.201; 0.859)
Interval 2: 7-12 months		
ipilimumab vs. DTIC	0.647	(0.336; 1.238)
nivolumab vs. DTIC	0.382	(0.235; 0.606)
nivolumab + ipilimumab vs. DTIC	0.675	(0.35; 1.293)
pembrolizumab vs. DTIC	0.346	(0.158; 0.764)
Interval 3: 13-18 months		
ipilimumab vs. DTIC	0.518	(0.217; 1.184)
nivolumab vs. DTIC	0.368	(0.178; 0.732)
nivolumab + ipilimumab vs. DTIC	0.179	(0.069; 0.443)
pembrolizumab vs. DTIC	0.394	(0.143; 1.065)
Interval 4: 19-120 months		
ipilimumab vs. DTIC	1.209	(0.709; 2.026)
nivolumab vs. DTIC	0.598	(0.391; 0.906)
nivolumab + ipilimumab vs. DTIC	0.45	(0.259; 0.779)
pembrolizumab vs. DTIC	1.227	(0.602; 2.486)

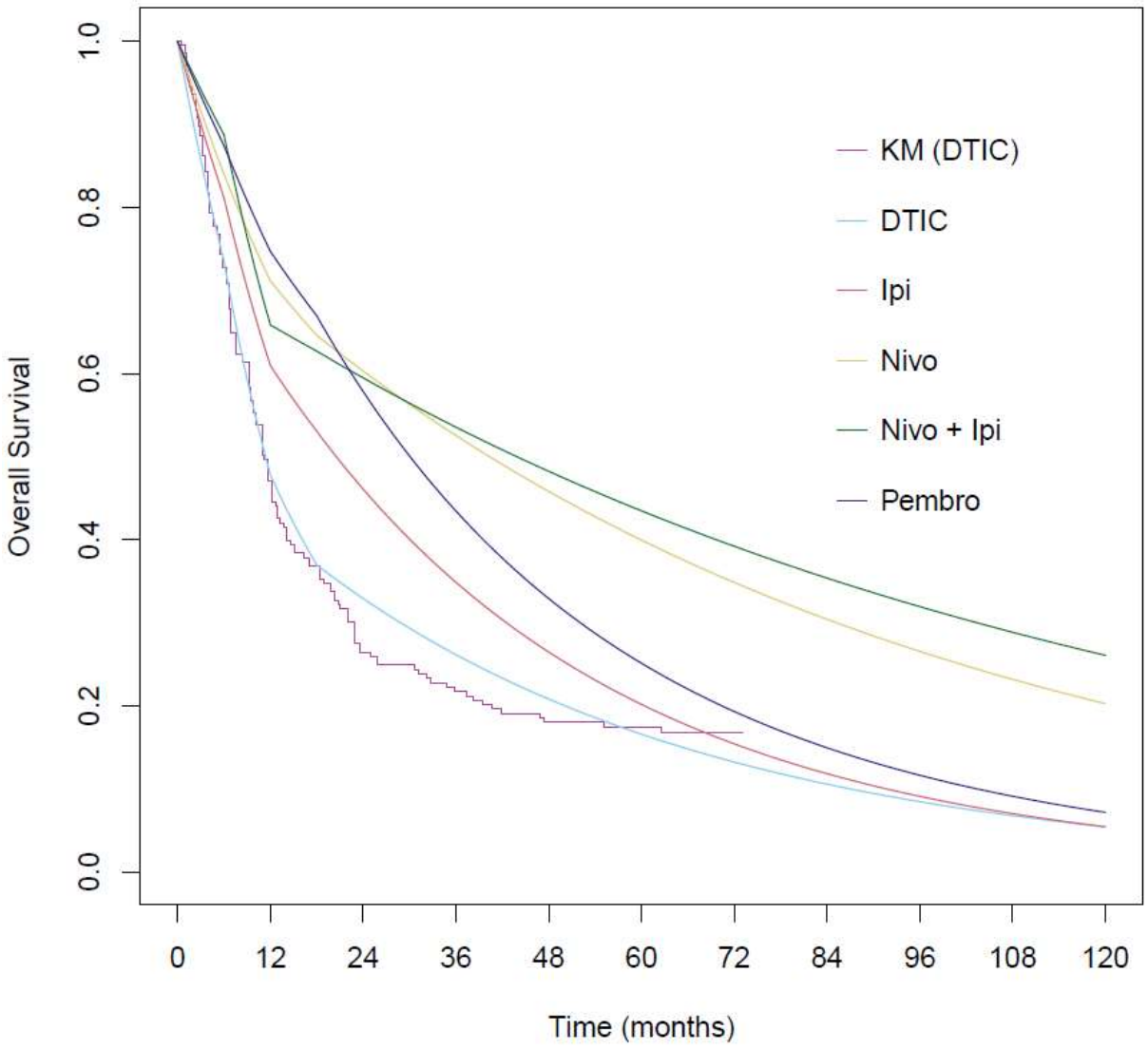


Figure 92: Survival curves for piecewise OS model with three cut points at 6, 12 and 18 months for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

E.1.9.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Table 92: Fixed effect PFS & OS NMA results for the best fitting (lowest DIC) piecewise exponential model with 3 cut points for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1	**1-6 months**		**1-12 months**	
dabrafenib vs. DTIC	0.351	(0.242; 0.509)	0.833	(0.507; 1.406)
dabrafenib + trametinib vs. DTIC	0.209	(0.126; 0.343)	0.594	(0.321; 1.105)

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
ipilimumab vs. DTIC	0.816	(0.587; 1.129)	0.668	(0.431; 1.035)
vemurafenib vs. DTIC	0.353	(0.196; 0.635)	0.813	(0.411; 1.628)
nivolumab vs. DTIC	0.493	(0.382; 0.633)	0.477	(0.345; 0.656)
nivolumab + ipilimumab vs. DTIC	0.346	(0.243; 0.489)	0.51	(0.324; 0.804)
encorafenib + binimetinib vs. DTIC	0.167	(0.082; 0.336)	0.509	(0.231; 1.13)
pembrolizumab vs. DTIC	0.433	(0.287; 0.646)	0.391	(0.227; 0.669)
Interval 2	**7-12 months**		**13-24 months**	
dabrafenib vs. DTIC	0.739	(0.35; 1.747)	0.674	(0.382; 1.228)
dabrafenib + trametinib vs. DTIC	0.515	(0.226; 1.295)	0.542	(0.276; 1.087)
ipilimumab vs. DTIC	0.277	(0.103; 0.71)	0.633	(0.34; 1.197)
vemurafenib vs. DTIC	0.904	(0.374; 2.386)	0.84	(0.395; 1.809)
nivolumab vs. DTIC	0.109	(0.048; 0.236)	0.365	(0.221; 0.596)
nivolumab + ipilimumab vs. DTIC	0.13	(0.05; 0.329)	0.205	(0.102; 0.408)
encorafenib + binimetinib vs. DTIC	0.626	(0.222; 1.892)	0.569	(0.236; 1.401)
pembrolizumab vs. DTIC	0.102	(0.033; 0.307)	0.545	(0.248; 1.213)
Interval 3	**13-18 months**		**25-36 months**	
dabrafenib vs. DTIC	1.683	(0.235; 34.09)	3.359	(0.906; 19.298)
dabrafenib + trametinib vs. DTIC	0.89	(0.113; 18.672)	2.158	(0.515; 13.158)
ipilimumab vs. DTIC	0.418	(0.056; 4.011)	1.108	(0.403; 3.062)
vemurafenib vs. DTIC	1.081	(0.127; 23.359)	2.516	(0.533; 16.445)

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
nivolumab vs. DTIC	0.133	(0.022; 1.068)	0.616	(0.259; 1.485)
nivolumab + ipilimumab vs. DTIC	0.142	(0.019; 1.377)	0.425	(0.149; 1.216)
encorafenib + binimetinib vs. DTIC	0.697	(0.067; 16.151)	1.731	(0.324; 12.182)
pembrolizumab vs. DTIC	0.425	(0.042; 5.49)	1.392	(0.375; 5.436)
Interval 2	**7-12 months**		**13-24 months**	
dabrafenib vs. DTIC	0.45	(0.109; 3.557)	0.656	(0.204; 2.361)
dabrafenib + trametinib vs. DTIC	0.394	(0.081; 3.418)	0.901	(0.242; 3.747)
ipilimumab vs. DTIC	2.894	(0.357; 110.609)	2.347	(0.922; 6.184)
vemurafenib vs. DTIC	0.598	(0.109; 5.436)	1.244	(0.297; 5.585)
nivolumab vs. DTIC	1.67	(0.232; 59.561)	1.198	(0.568; 2.688)
nivolumab + ipilimumab vs. DTIC	1.148	(0.143; 42.309)	0.933	(0.358; 2.534)
encorafenib + binimetinib vs. DTIC	0.596	(0.09; 5.972)	0.923	(0.181; 5.043)
pembrolizumab vs. DTIC	3.051	(0.304; 121.268)	1.873	(0.577; 6.334)

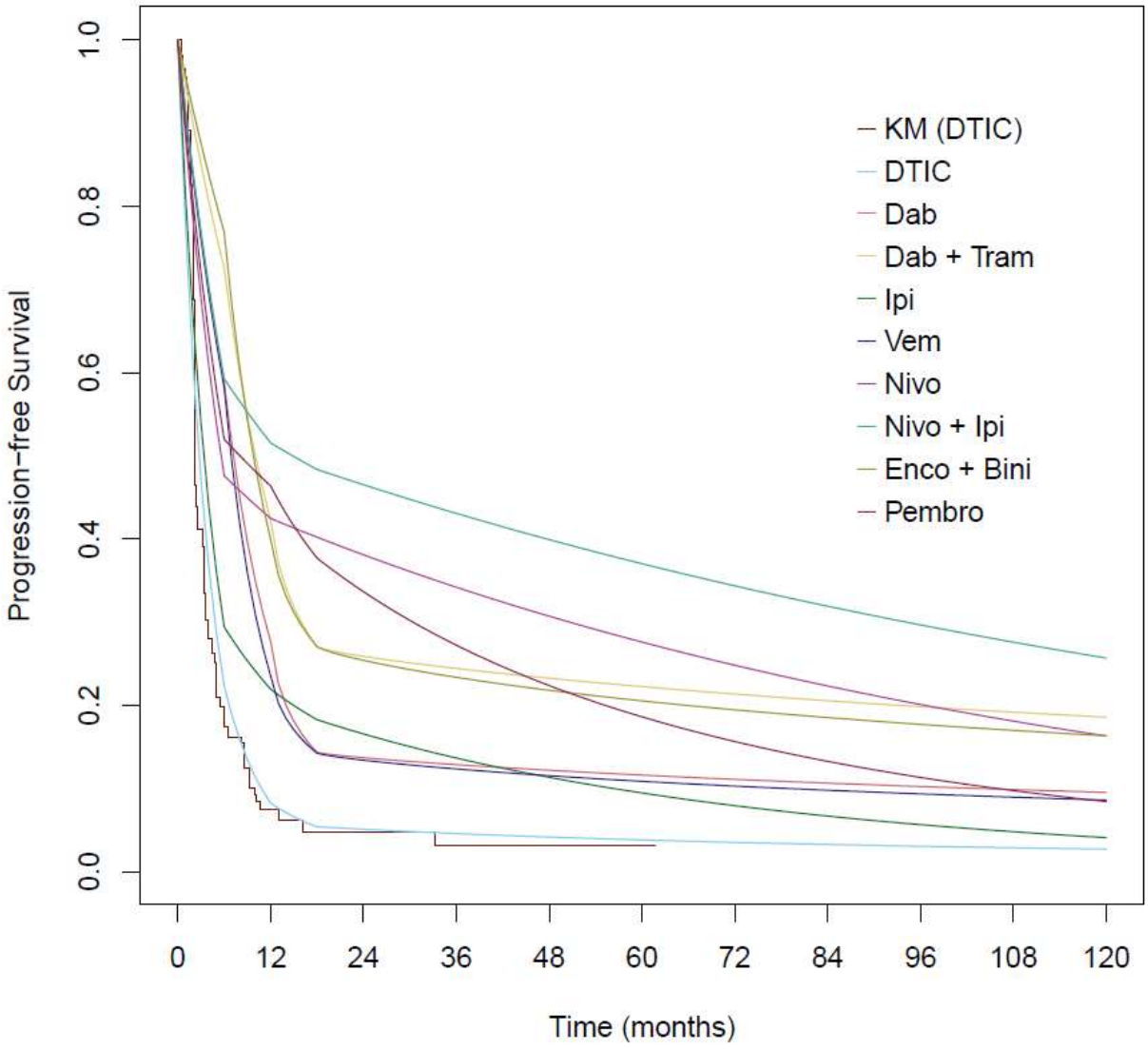


Figure 93: Survival curves for piecewise PFS model with three cut points at 6, 12 and 18 months for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

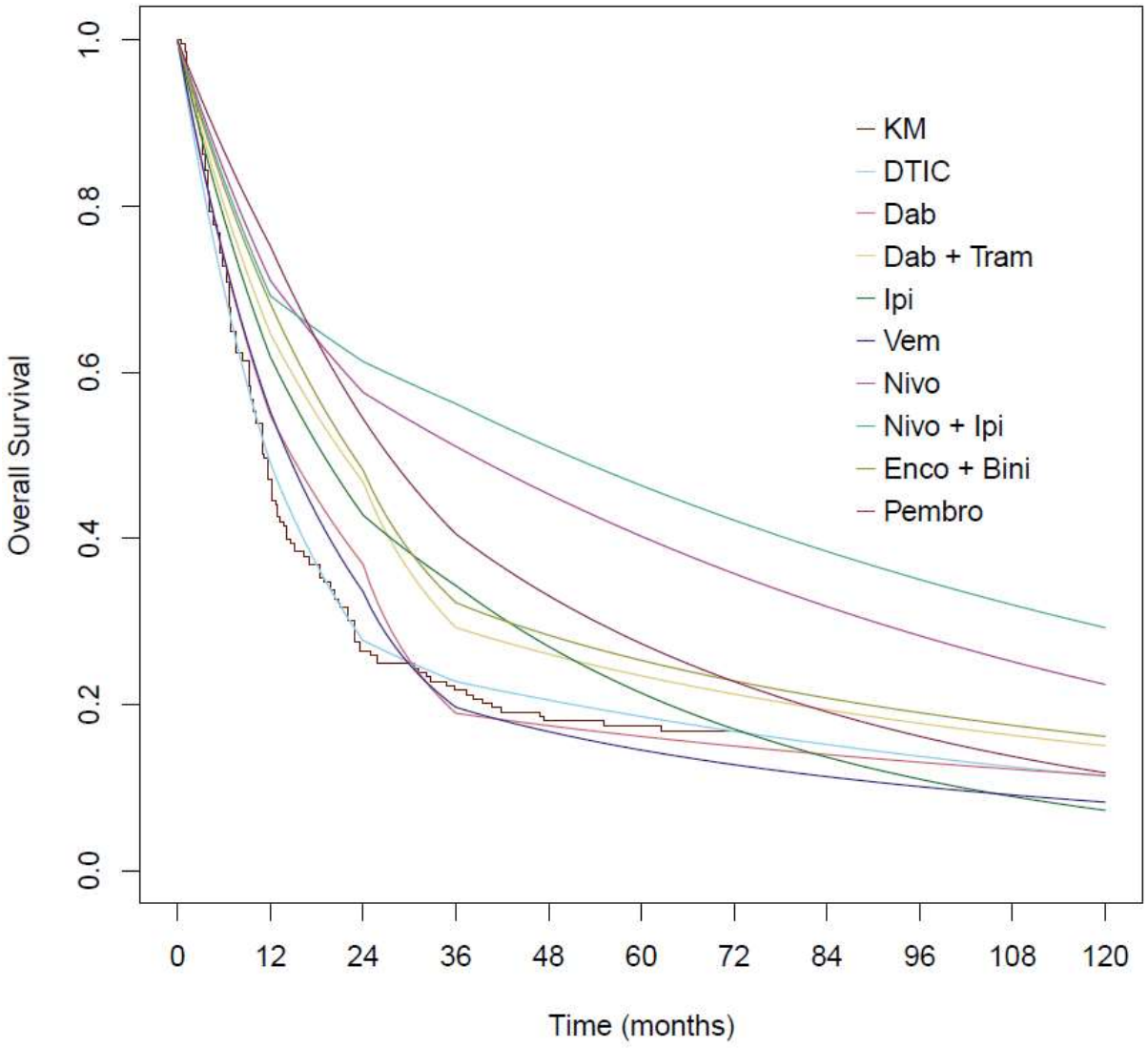


Figure 94: Survival curves for piecewise OS model with three cut points at 12, 24 and 36 months for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

E.1.9.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Table 93: Fixed effect PFS & OS NMA results for the best fitting (lowest DIC) piecewise exponential model with 3 cut points for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1	**1-6 months**		**1-12 months**	
ipilimumab vs. DTIC	0.82	(0.587; 1.14)	0.668	(0.431; 1.026)
nivolumab vs. DTIC	0.495	(0.383; 0.636)	0.478	(0.346; 0.654)

Comparison	PFS		OS	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	0.347	(0.243; 0.494)	0.51	(0.325; 0.795)
pembrolizumab vs. DTIC	0.434	(0.286; 0.654)	0.391	(0.231; 0.666)
Interval 2	**7-12 months**		**13-24 months**	
ipilimumab vs. DTIC	0.281	(0.106; 0.719)	0.63	(0.338; 1.17)
nivolumab vs. DTIC	0.11	(0.049; 0.24)	0.363	(0.219; 0.591)
nivolumab + ipilimumab vs. DTIC	0.131	(0.05; 0.332)	0.203	(0.101; 0.405)
pembrolizumab vs. DTIC	0.104	(0.034; 0.312)	0.542	(0.246; 1.192)
Interval 3	**13-18 months**		**25-36 months**	
ipilimumab vs. DTIC	0.401	(0.056; 3.662)	1.135	(0.412; 3.121)
nivolumab vs. DTIC	0.128	(0.024; 0.929)	0.629	(0.262; 1.535)
nivolumab + ipilimumab vs. DTIC	0.137	(0.02; 1.199)	0.436	(0.151; 1.268)
pembrolizumab vs. DTIC	0.412	(0.042; 4.816)	1.429	(0.383; 5.463)
Interval 4	**19-120 months**		**37-120 months**	
ipilimumab vs. DTIC	3.029	(0.345; 56.43)	2.38	(0.938; 6.495)
nivolumab vs. DTIC	1.764	(0.236; 32.59)	1.218	(0.581; 2.784)
nivolumab + ipilimumab vs. DTIC	1.212	(0.148; 23.22)	0.952	(0.366; 2.618)
pembrolizumab vs. DTIC	3.208	(0.307; 67.424)	1.915	(0.588; 6.619)

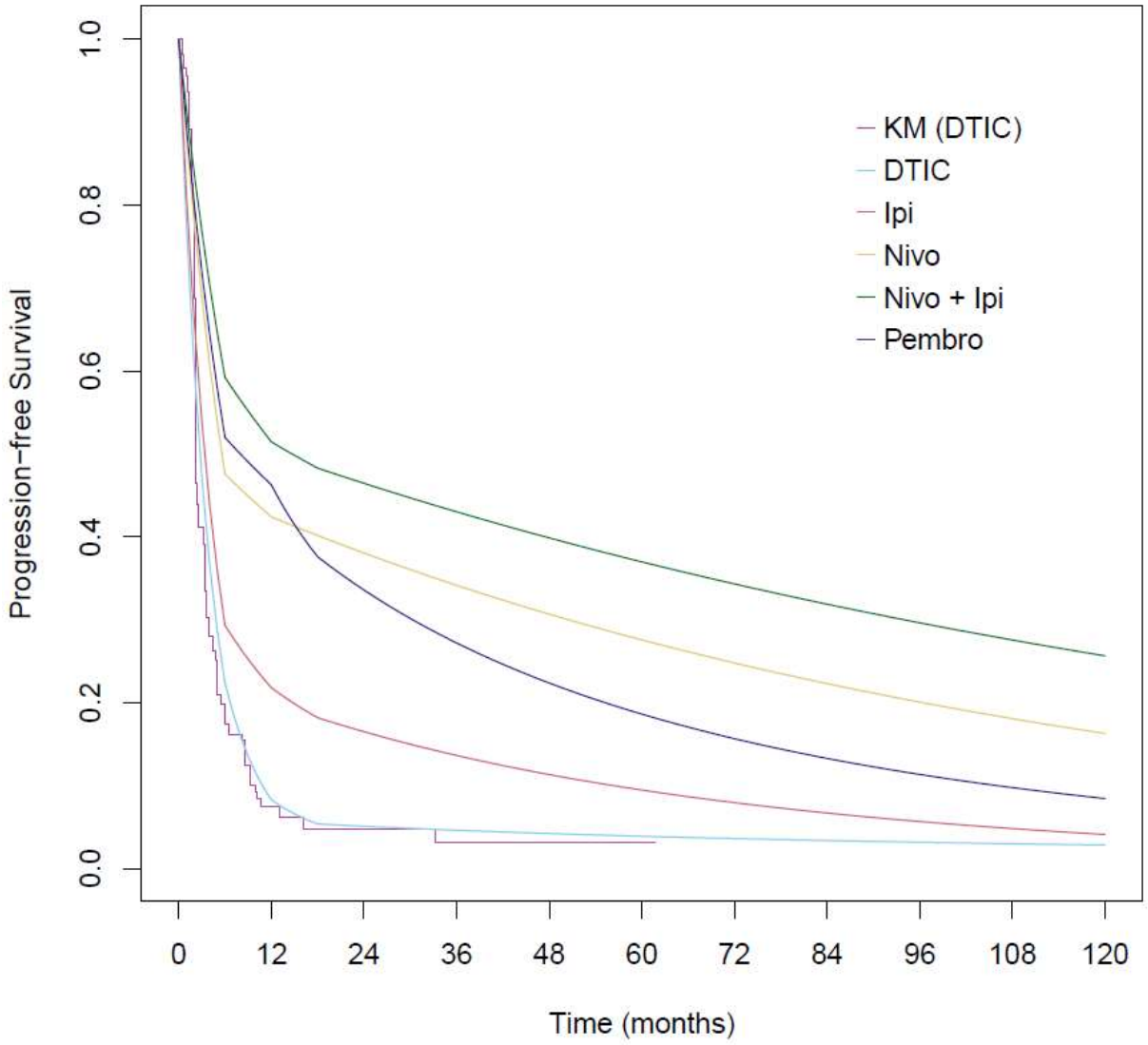


Figure 95: Survival curves for piecewise PFS model with three cut points at 6, 12 and 18 months for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

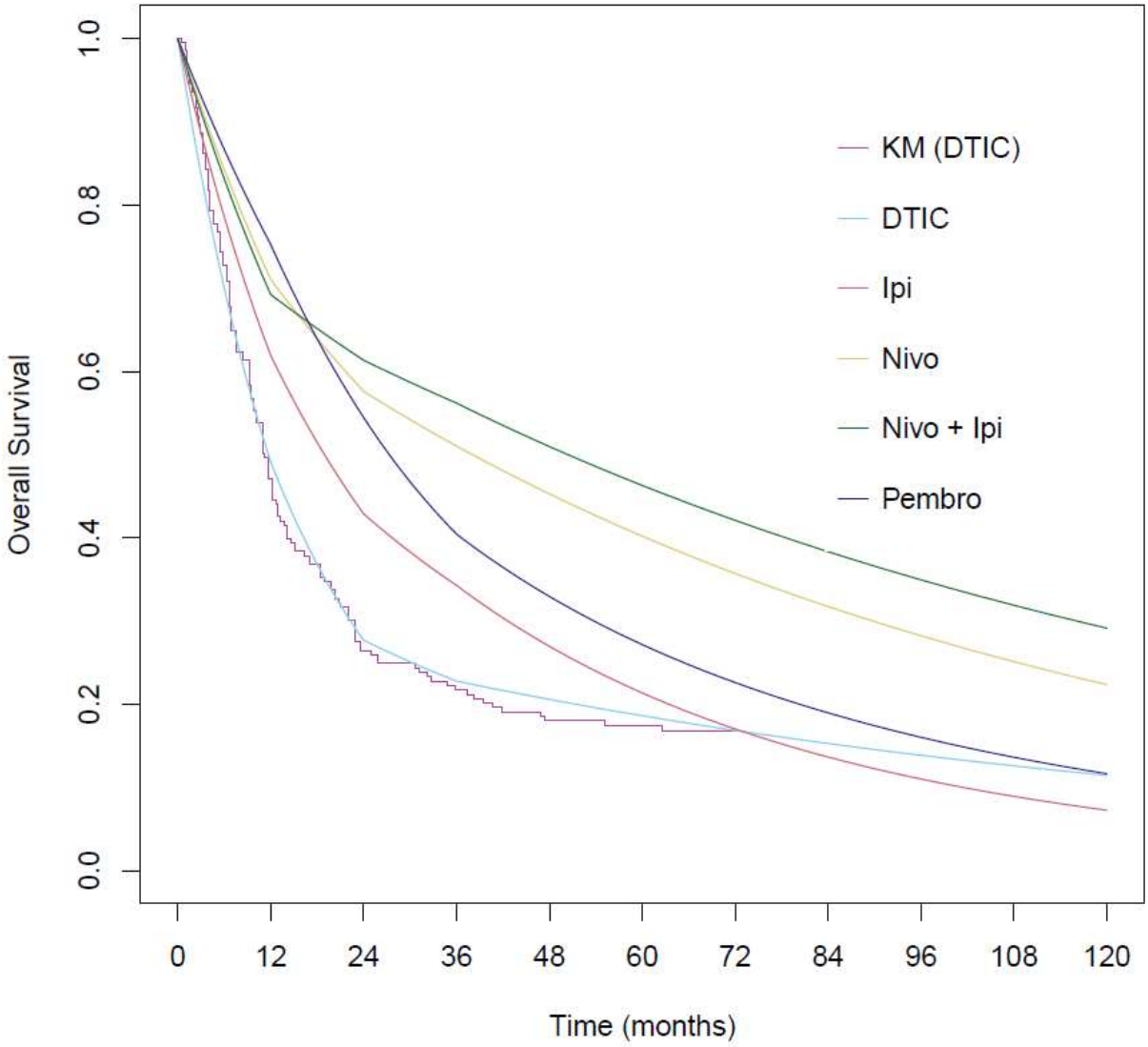


Figure 96: Survival curves for piecewise OS model with three cut points at 12, 24 and 36 months for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

E.1.9.7 Impact of different network assumptions on NMA results for piecewise exponential models with three cut points

Table 94: DIC values for all piecewise exponential models with three cut points for PFS for each network

Cut point placement	Targeted therapy + Immunotherapy		Immunotherapy only		
	Network 1	Network 5	Network 3	Network 4	Network 6
Three cut points					
6, 12 and 18 months	**543.1**	**469.3**	**189.9**	Did not converge	**195.3**
12, 24 and 36 months	[a]	Did not converge	Did not converge	[a]	Did not converge

(a) The shortest amount of follow-up time in these networks is 22 months (BRIM-3) and 24 months (CheckMate 069) for PFS and OS respectively. As such, it is not possible to generate aggregate data beyond these time points. Thus, for this network, it's impossible to run piecewise models with cut points beyond 22 or 24 months for PFS and OS respectively.

Table 95: A comparison of PFS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months				
dabrafenib vs. DTIC	0.346	(0.259; 0.464)	0.351	(0.242; 0.509)
dabrafenib + trametinib vs. DTIC	0.204	(0.152; 0.274)	0.209	(0.126; 0.343)
ipilimumab vs. DTIC	0.827	(0.598; 1.148)	0.816	(0.587; 1.129)
vemurafenib vs. DTIC	0.342	(0.279; 0.419)	0.353	(0.196; 0.635)
nivolumab vs. DTIC	0.496	(0.386; 0.635)	0.493	(0.382; 0.633)
nivolumab + ipilimumab vs. DTIC	0.346	(0.245; 0.487)	0.346	(0.243; 0.489)
encorafenib + binimetinib vs. DTIC	0.162	(0.104; 0.251)	0.167	(0.082; 0.336)
pembrolizumab vs. DTIC	0.439	(0.292; 0.66)	0.433	(0.287; 0.646)
Interval 2: 7-12 months				
dabrafenib vs. DTIC	0.799	(0.478; 1.352)	0.739	(0.35; 1.747)
dabrafenib + trametinib vs. DTIC	0.569	(0.358; 0.921)	0.515	(0.226; 1.295)
ipilimumab vs. DTIC	0.29	(0.112; 0.738)	0.277	(0.103; 0.71)
vemurafenib vs. DTIC	1.013	(0.682; 1.555)	0.904	(0.374; 2.386)
nivolumab vs. DTIC	0.113	(0.05; 0.242)	0.109	(0.048; 0.236)
nivolumab + ipilimumab vs. DTIC	0.134	(0.052; 0.342)	0.13	(0.05; 0.329)
encorafenib + binimetinib vs. DTIC	0.703	(0.36; 1.397)	0.626	(0.222; 1.892)
pembrolizumab vs. DTIC	0.108	(0.036; 0.325)	0.102	(0.033; 0.307)

Comparison	Network 1		Network 5	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 3: 13-18 months				
dabrafenib vs. DTIC	1.774	(0.511; 8.045)	1.683	(0.235; 34.09)
dabrafenib + trametinib vs. DTIC	0.963	(0.296; 4.092)	0.89	(0.113; 18.672)
ipilimumab vs. DTIC	0.418	(0.059; 3.593)	0.418	(0.056; 4.011)
vemurafenib vs. DTIC	1.193	(0.385; 4.821)	1.081	(0.127; 23.359)
nivolumab vs. DTIC	0.128	(0.022; 0.912)	0.133	(0.022; 1.068)
nivolumab + ipilimumab vs. DTIC	0.129	(0.018; 1.102)	0.142	(0.019; 1.377)
encorafenib + binimetinib vs. DTIC	0.768	(0.185; 3.947)	0.697	(0.067; 16.151)
pembrolizumab vs. DTIC	0.429	(0.046; 4.938)	0.425	(0.042; 5.49)
Interval 4: 19-120 months				
dabrafenib vs. DTIC	0.653	(0.149; 5.613)	0.45	(0.109; 3.557)
dabrafenib + trametinib vs. DTIC	0.599	(0.122; 5.479)	0.394	(0.081; 3.418)
ipilimumab vs. DTIC	2.631	(0.323; 55.813)	2.894	(0.357; 110.609)
vemurafenib vs. DTIC	0.934	(0.178; 8.846)	0.598	(0.109; 5.436)
nivolumab vs. DTIC	1.523	(0.213; 31.312)	1.67	(0.232; 59.561)
nivolumab + ipilimumab vs. DTIC	1.042	(0.132; 22.488)	1.148	(0.143; 42.309)
encorafenib + binimetinib vs. DTIC	0.932	(0.145; 10.612)	0.596	(0.09; 5.972)
pembrolizumab vs. DTIC	2.768	(0.286; 70.035)	3.051	(0.304; 121.268)

Table 96: A comparison of PFS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up),

and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3[a]		Network 6[a]	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months				
ipilimumab vs. DTIC	0.968	(0.68; 1.388)	0.82	(0.587; 1.14)
nivolumab vs. DTIC	0.494	(0.385; 0.635)	0.495	(0.383; 0.636)
nivolumab + ipilimumab vs. DTIC	0.398	(0.27; 0.588)	0.347	(0.243; 0.494)
pembrolizumab vs. DTIC	0.512	(0.335; 0.795)	0.434	(0.286; 0.654)
Interval 2: 7-12 months				
ipilimumab vs. DTIC	0.323	(0.108; 0.887)	0.281	(0.106; 0.719)
nivolumab vs. DTIC	0.111	(0.048; 0.239)	0.11	(0.049; 0.24)
nivolumab + ipilimumab vs. DTIC	0.152	(0.053; 0.416)	0.131	(0.05; 0.332)
pembrolizumab vs. DTIC	0.12	(0.035; 0.379)	0.104	(0.034; 0.312)
Interval 3: 13-18 months				
ipilimumab vs. DTIC	0.69	(0.075; 7.576)	0.401	(0.056; 3.662)
nivolumab vs. DTIC	0.132	(0.023; 0.988)	0.128	(0.024; 0.929)
nivolumab + ipilimumab vs. DTIC	0.243	(0.028; 2.569)	0.137	(0.02; 1.199)
pembrolizumab vs. DTIC	0.711	(0.058; 9.718)	0.412	(0.042; 4.816)
Interval 4: 19-120 months				
ipilimumab vs. DTIC	2.977	(0.338; 65.759)	3.029	(0.345; 56.43)
nivolumab vs. DTIC	1.806	(0.252; 38.784)	1.764	(0.236; 32.59)
nivolumab + ipilimumab vs. DTIC	1.398	(0.171; 31.312)	1.212	(0.148; 23.22)
pembrolizumab vs. DTIC	3.146	(0.292; 72.24)	3.208	(0.307; 67.424)

(a) As seen in Table 94, it was not possible to run a piecewise model with three cuts at 12, 24, and 36 months for network 4. Additionally, the piecewise model with 3 cuts at 6, 12, and 18 months did not converge for network 4. As such, we are unable to present any results in this table for network 4.

Table 97: DIC values for all piecewise exponential models with three cut points for OS for each network

Cut point placement	Targeted therapy + Immunotherapy		Immunotherapy only		
	Network 1	Network 5	Network 3	Network 4	Network 6
Three cut points					
6, 12 and 18 months	**580.5**	480.7	204.9	**250.8**	208.8
12, 24 and 36 months	[a]	**476.2**	**198.7**	[a]	**204.2**

(a) The shortest amount of follow-up time in these networks is 22 months (BRIM-3) and 24 months (CheckMate 069) for PFS and OS respectively. As such, it is not possible to generate aggregate data beyond these time points. Thus, for this network, it's impossible to run piecewise models with cut points beyond 22 or 24 months for PFS and OS respectively.

Table 98: A comparison of OS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5[a]	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months				
dabrafenib vs. DTIC	0.69	(0.386; 1.247)	0.908	(0.431; 2.031)
dabrafenib + trametinib vs. DTIC	0.236	(0.135; 0.409)	0.388	(0.139; 1.096)
ipilimumab vs. DTIC	0.672	(0.38; 1.181)	0.705	(0.389; 1.272)
vemurafenib vs. DTIC	0.491	(0.364; 0.664)	0.952	(0.292; 3.184)
nivolumab vs. DTIC	0.567	(0.367; 0.867)	0.564	(0.362; 0.868)
nivolumab + ipilimumab vs. DTIC	0.38	(0.206; 0.695)	0.338	(0.176; 0.639)
encorafenib + binimetinib vs. DTIC	0.259	(0.109; 0.595)	0.501	(0.119; 2.096)
pembrolizumab vs. DTIC	0.423	(0.206; 0.873)	0.442	(0.21; 0.927)
Interval 2: 7-12 months				
dabrafenib vs. DTIC	0.907	(0.57; 1.453)	0.778	(0.399; 1.611)

Comparison	Network 1		Network 5[a]	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
dabrafenib + trametinib vs. DTIC	0.813	(0.541; 1.225)	0.656	(0.296; 1.502)
ipilimumab vs. DTIC	0.651	(0.34; 1.248)	0.627	(0.321; 1.234)
vemurafenib vs. DTIC	0.992	(0.729; 1.351)	0.767	(0.32; 1.858)
nivolumab vs. DTIC	0.382	(0.236; 0.61)	0.383	(0.235; 0.615)
nivolumab + ipilimumab vs. DTIC	0.678	(0.356; 1.306)	0.703	(0.362; 1.37)
encorafenib + binimetinib vs. DTIC	0.625	(0.362; 1.068)	0.483	(0.183; 1.288)
pembrolizumab vs. DTIC	0.348	(0.159; 0.767)	0.335	(0.151; 0.75)
Interval 3: 13-18 months				
dabrafenib vs. DTIC	0.726	(0.431; 1.235)	0.665	(0.322; 1.443)
dabrafenib + trametinib vs. DTIC	0.65	(0.399; 1.063)	0.574	(0.243; 1.413)
ipilimumab vs. DTIC	0.521	(0.217; 1.217)	0.535	(0.222; 1.275)
vemurafenib vs. DTIC	1.043	(0.708; 1.539)	0.894	(0.335; 2.439)
nivolumab vs. DTIC	0.368	(0.178; 0.745)	0.369	(0.178; 0.759)
nivolumab + ipilimumab vs. DTIC	0.179	(0.07; 0.451)	0.171	(0.064; 0.447)
encorafenib + binimetinib vs. DTIC	0.622	(0.303; 1.291)	0.534	(0.169; 1.72)
pembrolizumab vs. DTIC	0.397	(0.141; 1.114)	0.408	(0.143; 1.162)
Interval 4: 19-120 months				
dabrafenib vs. DTIC	0.939	(0.618; 1.44)	1.042	(0.561; 2.009)
dabrafenib + trametinib vs. DTIC	0.757	(0.515; 1.117)	0.866	(0.424; 1.803)

Comparison	Network 1		Network 5[a]	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
ipilimumab vs. DTIC	1.198	(0.703; 2.069)	1.229	(0.723; 2.09)
vemurafenib vs. DTIC	1.037	(0.756; 1.434)	1.218	(0.557; 2.676)
nivolumab vs. DTIC	0.593	(0.39; 0.914)	0.596	(0.393; 0.908)
nivolumab + ipilimumab vs. DTIC	0.445	(0.254; 0.789)	0.435	(0.249; 0.762)
encorafenib + binimetinib vs. DTIC	0.765	(0.457; 1.278)	0.901	(0.374; 2.216)
pembrolizumab vs. DTIC	1.214	(0.606; 2.478)	1.24	(0.618; 2.494)

(a) Although the piecewise model with 3 cut points at 12, 24 and 36 months has a smaller DIC value (indicating a better fitting model) than the model with cuts at 6, 12, and 18 months for network 5, we present results for network 5 in this table from the model with cuts at 6, 12, and 18 months. This is to allow one to assess what impact different network assumptions has on the NMA results and is only possible if we are looking at results from the same model.

Table 99: A comparison of OS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up), and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
Interval 1: 1-6 months						
ipilimumab vs. DTIC	0.691	(0.371; 1.282)	0.665	(0.373; 1.175)	0.705	(0.393; 1.264)
nivolumab vs. DTIC	0.565	(0.363; 0.869)	0.563	(0.364; 0.864)	0.564	(0.36; 0.875)
nivolumab + ipilimumab vs. DTIC	0.62	(0.333; 1.157)	0.377	(0.205; 0.689)	0.337	(0.176; 0.637)
pembrolizumab vs. DTIC	0.342	(0.163; 0.725)	0.417	(0.201; 0.859)	0.443	(0.211; 0.926)
Interval 2: 7-12 months						
ipilimumab vs. DTIC	0.633	(0.3; 1.326)	0.647	(0.336; 1.238)	0.621	(0.323; 1.198)
nivolumab vs. DTIC	0.381	(0.233; 0.614)	0.382	(0.235; 0.606)	0.38	(0.236; 0.604)

Comparison	Network 3		Network 4		Network 6	
	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval	Hazard ratio	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	0.382	(0.173; 0.838)	0.675	(0.35; 1.293)	0.695	(0.362; 1.331)
pembrolizumab vs. DTIC	0.493	(0.206; 1.172)	0.346	(0.158; 0.764)	0.331	(0.15; 0.719)
Interval 3: 13-18 months						
ipilimumab vs. DTIC	0.528	(0.212; 1.285)	0.518	(0.217; 1.184)	0.531	(0.215; 1.242)
nivolumab vs. DTIC	0.365	(0.176; 0.73)	0.368	(0.178; 0.732)	0.368	(0.174; 0.741)
nivolumab + ipilimumab vs. DTIC	0.237	(0.088; 0.61)	0.179	(0.069; 0.443)	0.17	(0.063; 0.436)
pembrolizumab vs. DTIC	0.395	(0.131; 1.17)	0.394	(0.143; 1.065)	0.404	(0.14; 1.123)
Interval 4: 19-120 months						
ipilimumab vs. DTIC	1.206	(0.682; 2.177)	1.209	(0.709; 2.026)	1.23	(0.735; 2.096)
nivolumab vs. DTIC	0.6	(0.398; 0.917)	0.598	(0.391; 0.906)	0.599	(0.396; 0.911)
nivolumab + ipilimumab vs. DTIC	0.475	(0.257; 0.885)	0.45	(0.259; 0.779)	0.437	(0.252; 0.761)
pembrolizumab vs. DTIC	1.565	(0.725; 3.497)	1.227	(0.602; 2.486)	1.25	(0.621; 2.558)

(a) Although the piecewise model with 3 cut points at 12, 24 and 36 months has a smaller DIC value (indicating a better fitting model) than the model with cuts at 6, 12, and 18 months for network 5, we present results for network 5 in this table from the model with cuts at 6, 12, and 18 months. This is to allow one to assess what impact different network assumptions has on the NMA results and is only possible if we are looking at results from the same model.

Appendix F: Fractional Polynomials

F.1.1.1 Network 1 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Table 100: Convergence table for PFS and OS for 1st order fractional polynomial models for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
-2	Yes	1423.5	1130	Yes	1428.8	1693.6
-1	Yes	1317.2	1017.2	Yes	1417.8	1615.2
-0.5	Yes	1257[b]	944.6[c]	Yes	1403.8	1525.6

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
0	No	1211.6	-	Yes	1386.2[b]	1404.4[c]
0.5	No	1206.9[a]	-	No	1373.1	-
1	No	1263.7	-	No	1372.4[a]	-
2	No	1538.6	-	No	1400.1	-
3	Yes	1840.5	8.2E+13	Yes	1424.4	1.3E+14

(a) Lowest AIC of any power – model did not converge
(b) Lowest AIC of models that did converge
(c) Lowest DIC

As shown in Table 100, of the fractions that converged, -0.5 had both the lowest AIC and DIC values for PFS, and 0 had both the lowest AIC and DIC values for OS. The AIC values calculated in R prior to running the NMA were smallest with the fractions 0.5 and 1 for PFS and OS respectively. However, neither of these models converged.

Table 101: Fixed effect PFS and OS NMA results for best fitting fractional polynomial model for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

Comparison	PFS		OS	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	-0.5		0	
Pooled estimate for difference β_0				
dabrafenib vs. DTIC	0.826	(-0.28; 2.183)	-0.427	(-1.299; 0.311)
dabrafenib + trametinib vs. DTIC	0.169	(-0.972; 1.552)	-1.082	(-1.875; -0.401)
ipilimumab vs. DTIC	-1.094	(-2.622; 0.146)	-1.144	(-2.063; -0.153)
vemurafenib vs. DTIC	1.147	(-0.004; 2.466)	-0.821	(-1.355; -0.312)
nivolumab vs. DTIC	-2.29	(-3.519; -0.807)	-0.931	(-1.63; -0.235)
nivolumab + ipilimumab vs. DTIC	-2.36	(-3.785; -0.847)	-1.017	(-1.947; -0.008)
encorafenib + binimetinib vs. DTIC	1.08	(-0.437; 3.018)	-1.336	(-2.483; -0.283)
pembrolizumab vs. DTIC	-1.091	(-3.086; 0.355)	-1.768	(-2.816; -0.431)
Pooled estimate for difference β_1				
dabrafenib vs. DTIC	-3.777	(-7.347; -0.795)	0.088	(-0.162; 0.386)
dabrafenib + trametinib vs. DTIC	-2.929	(-6.486; 0.136)	0.225	(-0.011; 0.503)
ipilimumab vs. DTIC	2.363	(-0.885; 6.366)	0.335	(-0.013; 0.641)
vemurafenib vs. DTIC	-4.328	(-7.844; -1.262)	0.255	(0.067; 0.455)
nivolumab vs. DTIC	3.885	(0.276; 7.1)	0.092	(-0.152; 0.338)
nivolumab + ipilimumab vs. DTIC	3.417	(-0.392; 7.258)	0.056	(-0.283; 0.375)

Comparison	PFS		OS	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
encorafenib + binimetinib vs. DTIC	-5.409	(-10.63; -1.244)	0.287	(-0.076; 0.68)
pembrolizumab vs. DTIC	0.821	(-3.032; 6.16)	0.454	(-0.004; 0.828)

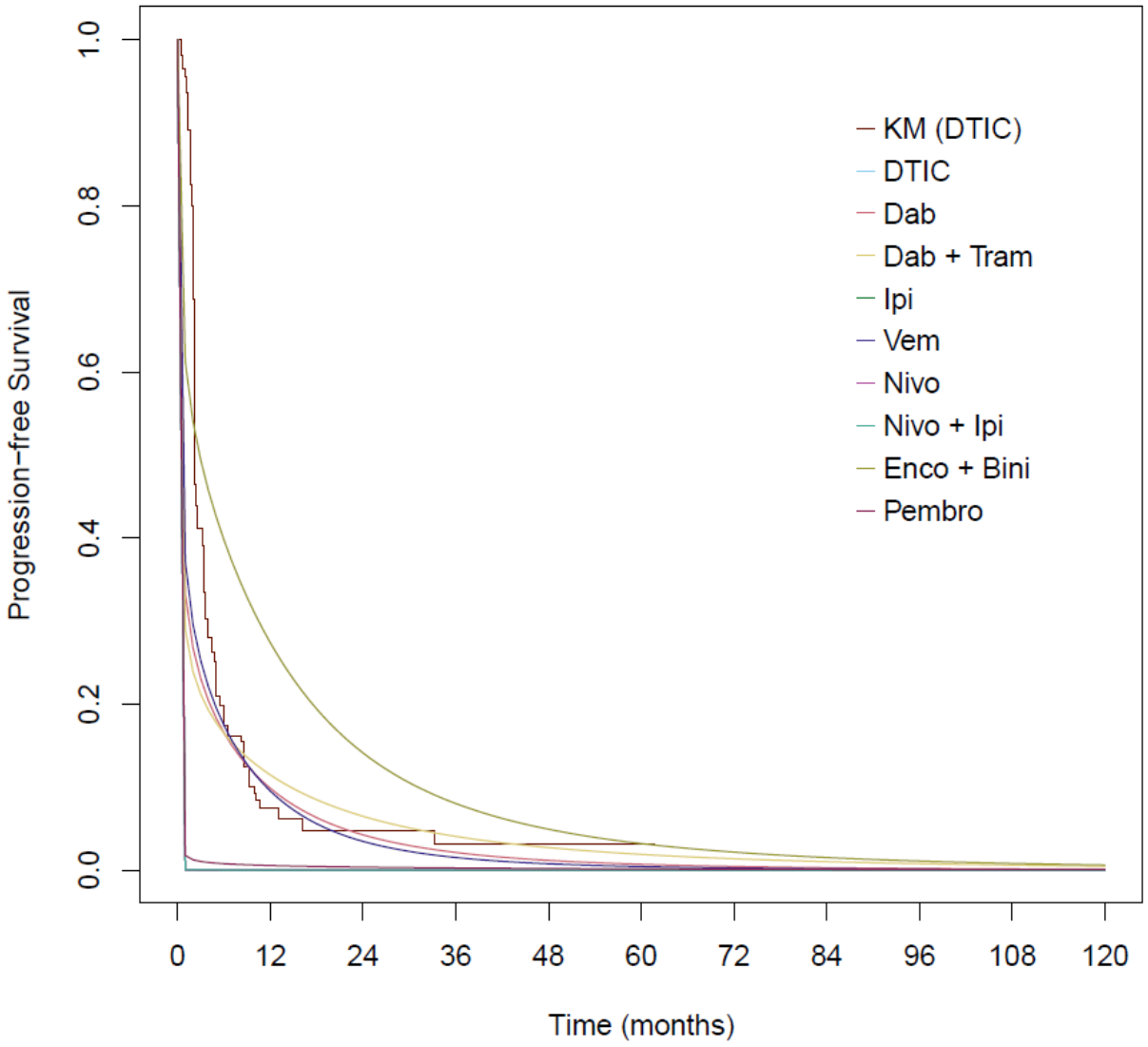


Figure 97: Survival curves for fractional polynomial PFS model with a power of -0.5 for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

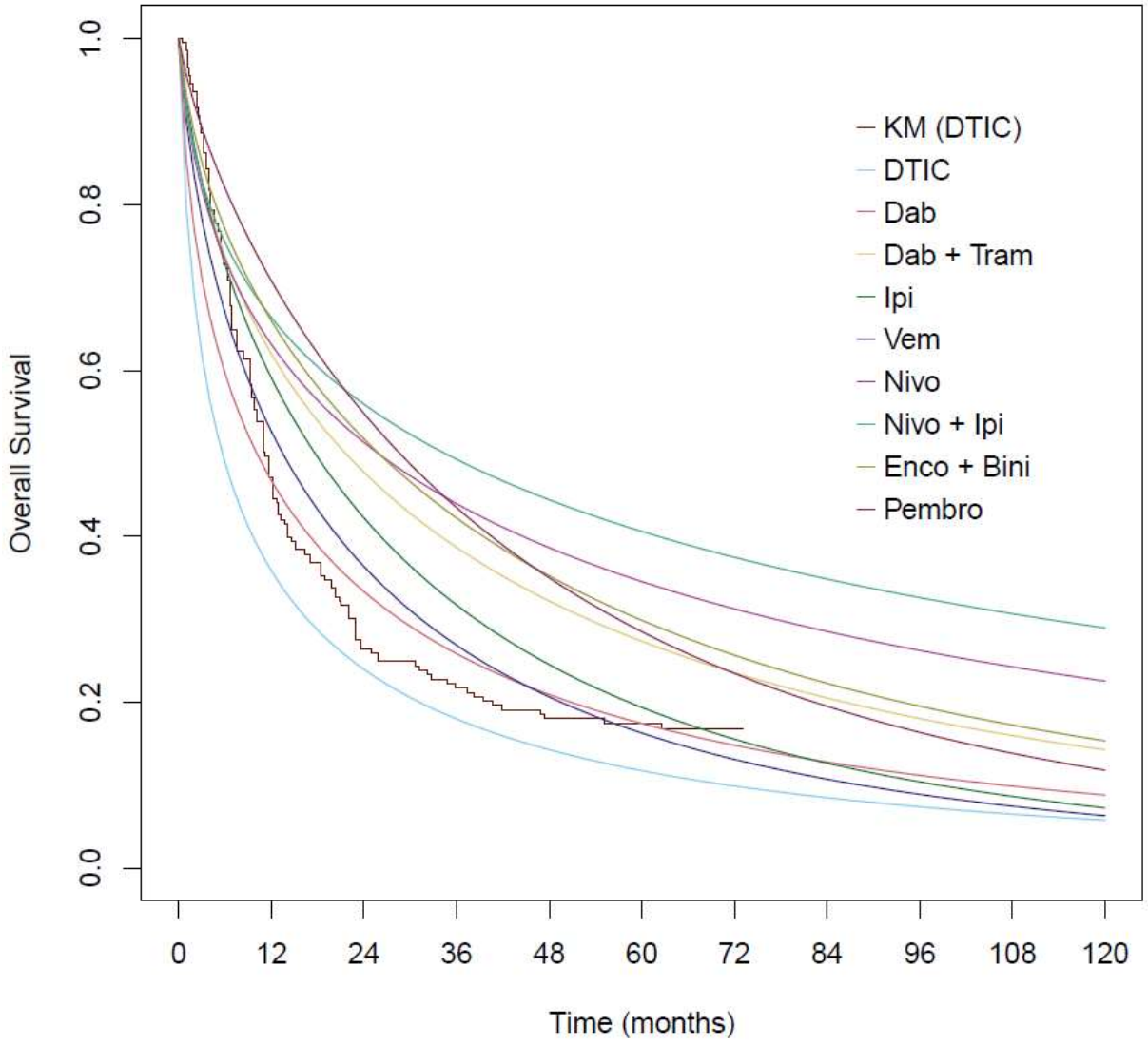


Figure 98: Survival curves for fractional polynomial OS model with a power of 0 for network 1 - people with BRAF wild type and mutant melanoma, all immunotherapy and targeted therapy strategies

F.1.1.2 Network 2 - People with *BRAF* mutant melanoma, with all immunotherapy and *BRAF*/MEK inhibitor strategies

As previously described, it was not possible to create a fully connected network for this analysis. Thus, while you can get treatment effects between treatments in the subnetworks, you cannot get overall treatment effects.

F.1.1.3 Network 3 - People with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 102: Convergence table for PFS and OS for 1st order fractional polynomial models for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
-2	Yes	318.2	329.3	Yes	434.2	564.7
-1	Yes	290.3	282.8	Yes	413.2	498
-0.5	Yes	283.4[a]	261.6	Yes	401.5	447.4
0	Yes	293.2	258.7	Yes	393.9[a]	393.3
0.5	Yes	330.8	285.4[c]	Yes	395.9	352.4[b]
1	No	401.4	-	No	408.9	-
2	No	609	-	No	448.3	-
3	Yes	807.8	2.5E+13	Yes	475	4.1E+13

(a) Lowest AIC of any power – model did converge
(b) Lowest DIC

As shown in Table 102, for PFS, -0.5 had the lowest AIC value, and 0 had the lowest DIC value. For OS, 0 had the lowest AIC value, and 0.5 had the lowest DIC value.

Table 103: Fixed effect PFS NMA results for best fitting fractional polynomial model assessed by AIC and DIC for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Comparison	PFS – Smallest AIC value		PFS – Smallest DIC value	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	-0.5		0	
Pooled estimate for difference β_0				
ipilimumab vs. DTIC	-0.791	(-2.167; 0.451)	0.776	(-0.676; 1.801)
nivolumab vs. DTIC	-1.985	(-3.047; -0.753)	-0.103	(-1.181; 0.872)
nivolumab + ipilimumab vs. DTIC	-1.682	(-2.912; -0.387)	-0.295	(-1.649; 0.915)
pembrolizumab vs. DTIC	-0.769	(-2.22; 0.548)	-0.36	(-2.007; 1.124)
Pooled estimate for difference β_1				
ipilimumab vs. DTIC	2.108	(-1.159; 5.72)	-0.376	(-0.894; 0.349)
nivolumab vs. DTIC	3.087	(-0.072; 5.779)	-0.358	(-0.837; 0.188)
nivolumab + ipilimumab vs. DTIC	1.974	(-1.514; 5.127)	-0.309	(-0.879; 0.342)
pembrolizumab vs. DTIC	0.516	(-2.977; 4.342)	-0.099	(-0.818; 0.695)

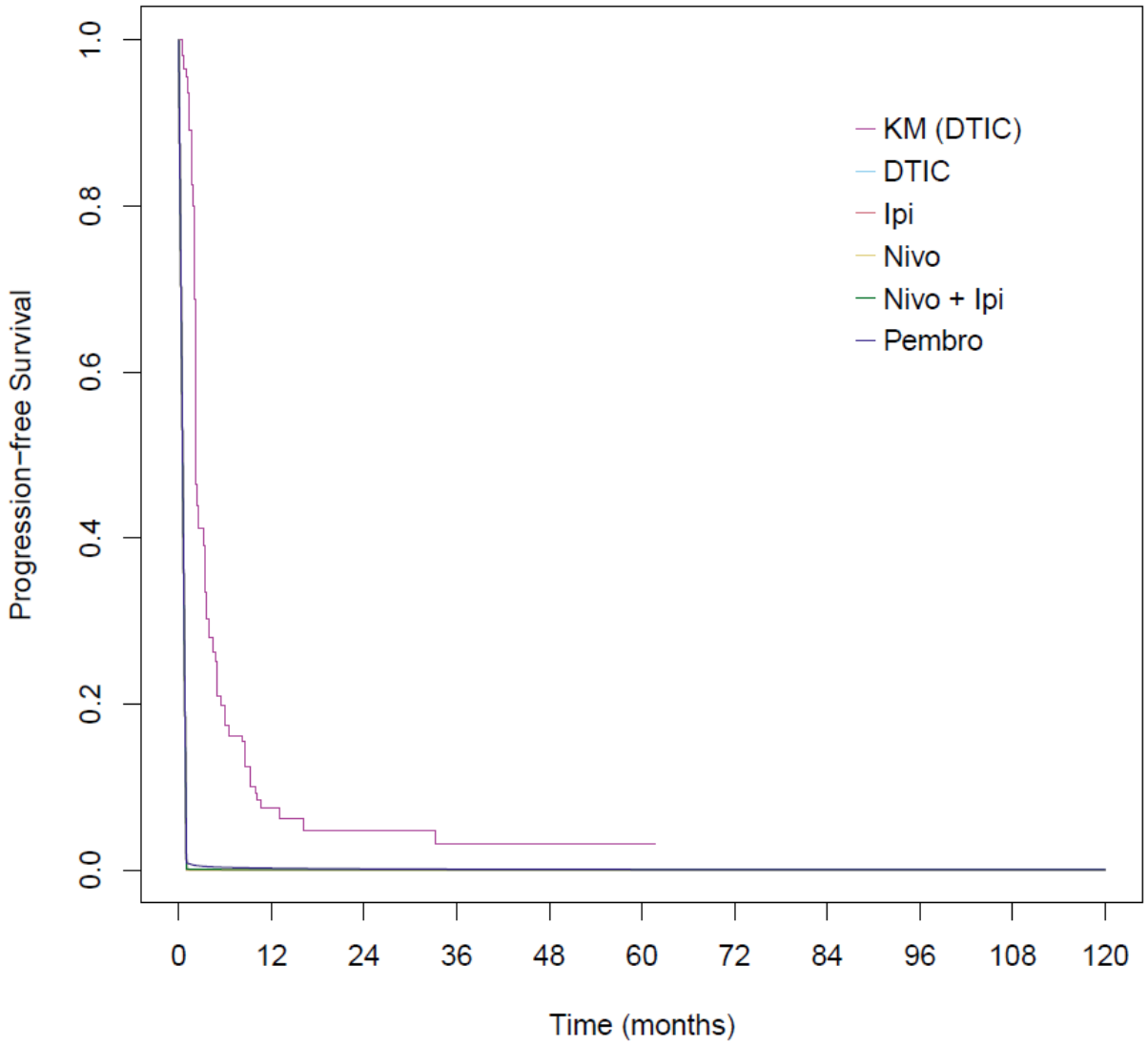


Figure 99: Survival curves for fractional polynomial PFS model with a power of -0.5 (smallest AIC value) for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

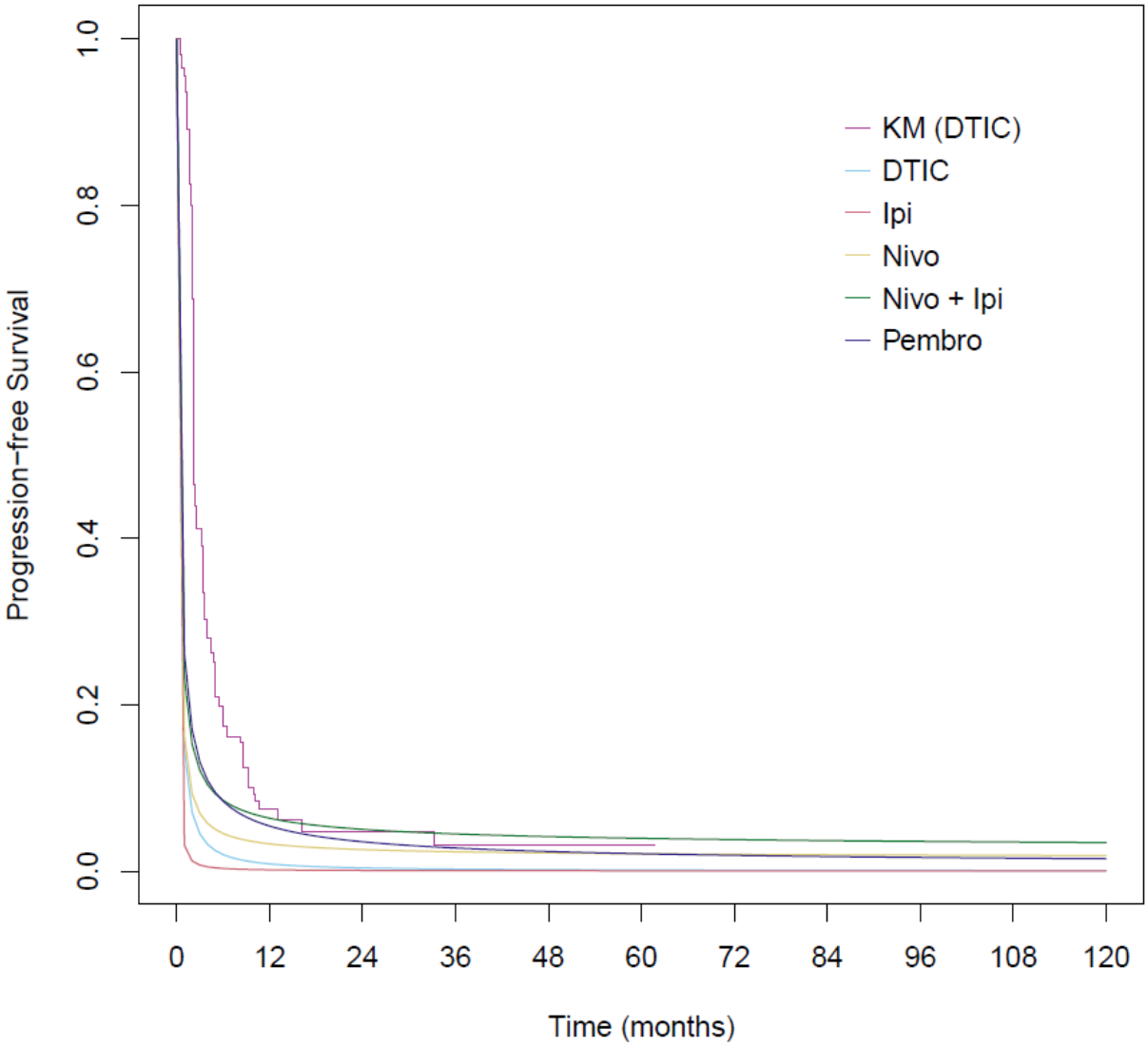


Figure 100: Survival curves for fractional polynomial PFS model with a power of 0 (smallest DIC value) for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Table 104: Fixed effect OS NMA results for best fitting fractional polynomial model assessed by AIC and DIC for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

Comparison	OS - Smallest AIC value		OS - Smallest DIC value	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	0		0.5	
Pooled estimate for difference β_0				
ipilimumab vs. DTIC	-0.934	(-2.07; 0.239)	-0.866	(-1.72; -0.093)
nivolumab vs. DTIC	-0.868	(-1.614; -0.018)	-0.974	(-1.549; -0.404)
nivolumab + ipilimumab vs. DTIC	-0.418	(-1.594; 0.705)	-0.667	(-1.518; 0.139)

Comparison	OS - Smallest AIC value		OS - Smallest DIC value	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
pembrolizumab vs. DTIC	-1.96	(-3.304; -0.275)	-1.72	(-2.685; -0.748)
Pooled estimate for difference β_1				
ipilimumab vs. DTIC	0.259	(-0.157; 0.654)	0.153	(-0.014; 0.354)
nivolumab vs. DTIC	0.071	(-0.228; 0.331)	0.073	(-0.053; 0.203)
nivolumab + ipilimumab vs. DTIC	-0.138	(-0.533; 0.276)	-0.028	(-0.208; 0.168)
pembrolizumab vs. DTIC	0.556	(-0.042; 1.05)	0.306	(0.083; 0.533)

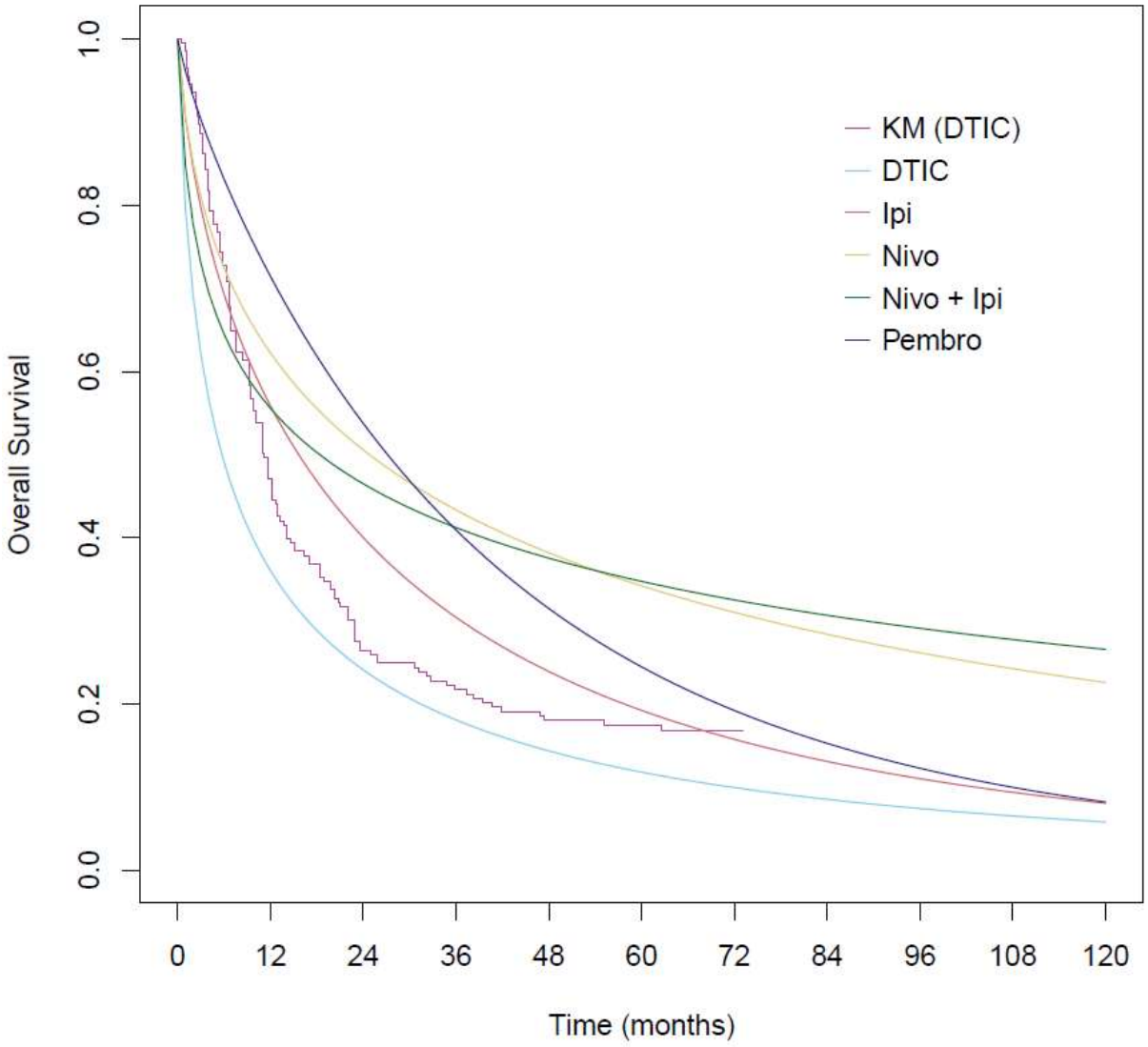


Figure 101: Survival curves for fractional polynomial OS model with a power of 0 (smallest AIC value) for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

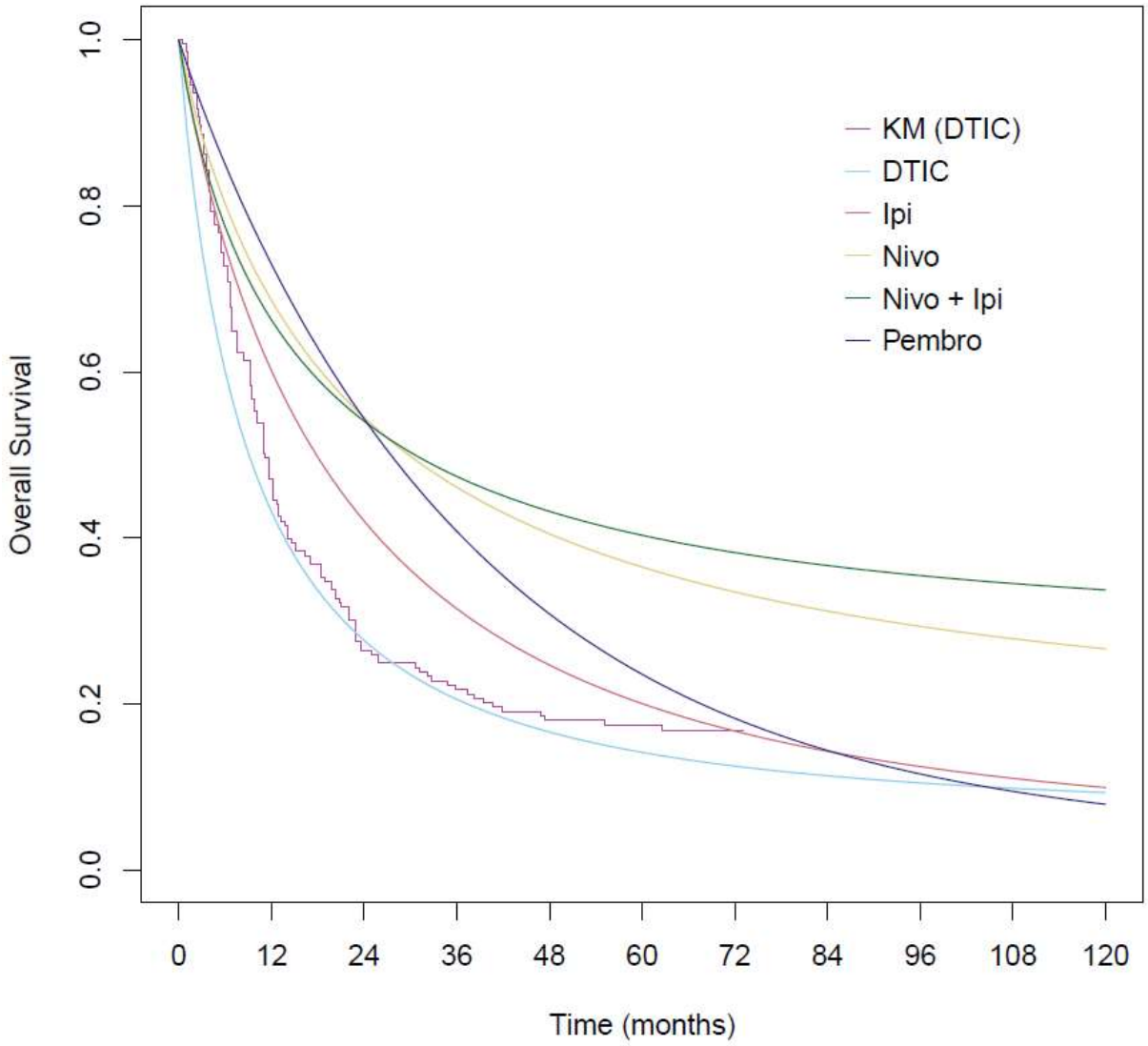


Figure 102: Survival curves for fractional polynomial OS model with a power of 0.5 (smallest DIC value) for network 3 - people with *BRAF* wild type melanoma, with immunotherapy strategies only

F.1.1.4 Network 4 - People with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Table 105: Convergence table for PFS and OS for 1st order fractional polynomial models for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
-2	Yes	385.1[c]	414.4[d]	Yes	513.7	657.7
-1	No	348.1	-	Yes	495.6	587.9
-0.5	No	338.1[a]	-	Yes	485.5	531.9
0	No	348.7	-	Yes	479.3[b]	471[d]
0.5	No	394.7	-	No	482	-

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
1	No	483.3	-	No	494.3	-
2	No	748.2	-	No	526.4	-
3	Yes	994.7	3.4E+13	Yes	544.0	5.5E+13

(a) Lowest AIC of any power – model did not converge
(b) Lowest AIC of any power – model did converge
(c) Lowest AIC of models that did converge
(d) Lowest DIC

As shown in Table 105, of the fractions that converged, -2 had both the lowest AIC and DIC values for PFS. For OS, 0 had both the lowest AIC and DIC values for OS. The AIC values calculated in R prior to running the NMA was smallest for the -0.5 fractions for PFS. However, this model did not converge.

Table 106: Fixed effect PFS and OS NMA results for best fitting fractional polynomial model for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

Comparison	PFS		OS	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	-2		0	
Pooled estimate for difference β_0				
ipilimumab vs. DTIC	-0.743	(-1.438; -0.043)	-0.934	(-1.801; -0.001)
nivolumab vs. DTIC	-1.822	(-2.405; -1.2)	-0.8	(-1.534; -0.084)
nivolumab + ipilimumab vs. DTIC	-1.888	(-2.566; -1.184)	-0.821	(-1.763; 0.064)
pembrolizumab vs. DTIC	-1.206	(-1.947; -0.468)	-1.562	(-2.752; -0.463)
Pooled estimate for difference β_1				
ipilimumab vs. DTIC	21.552	(-4.928; 49.46)	0.261	(-0.084; 0.568)
nivolumab vs. DTIC	40.525	(17.15; 63.481)	0.046	(-0.211; 0.306)
nivolumab + ipilimumab vs. DTIC	31.85	(4.898; 58.56)	-0.013	(-0.329; 0.324)
pembrolizumab vs. DTIC	16.626	(-12.31; 47.2)	0.38	(-0.005; 0.806)

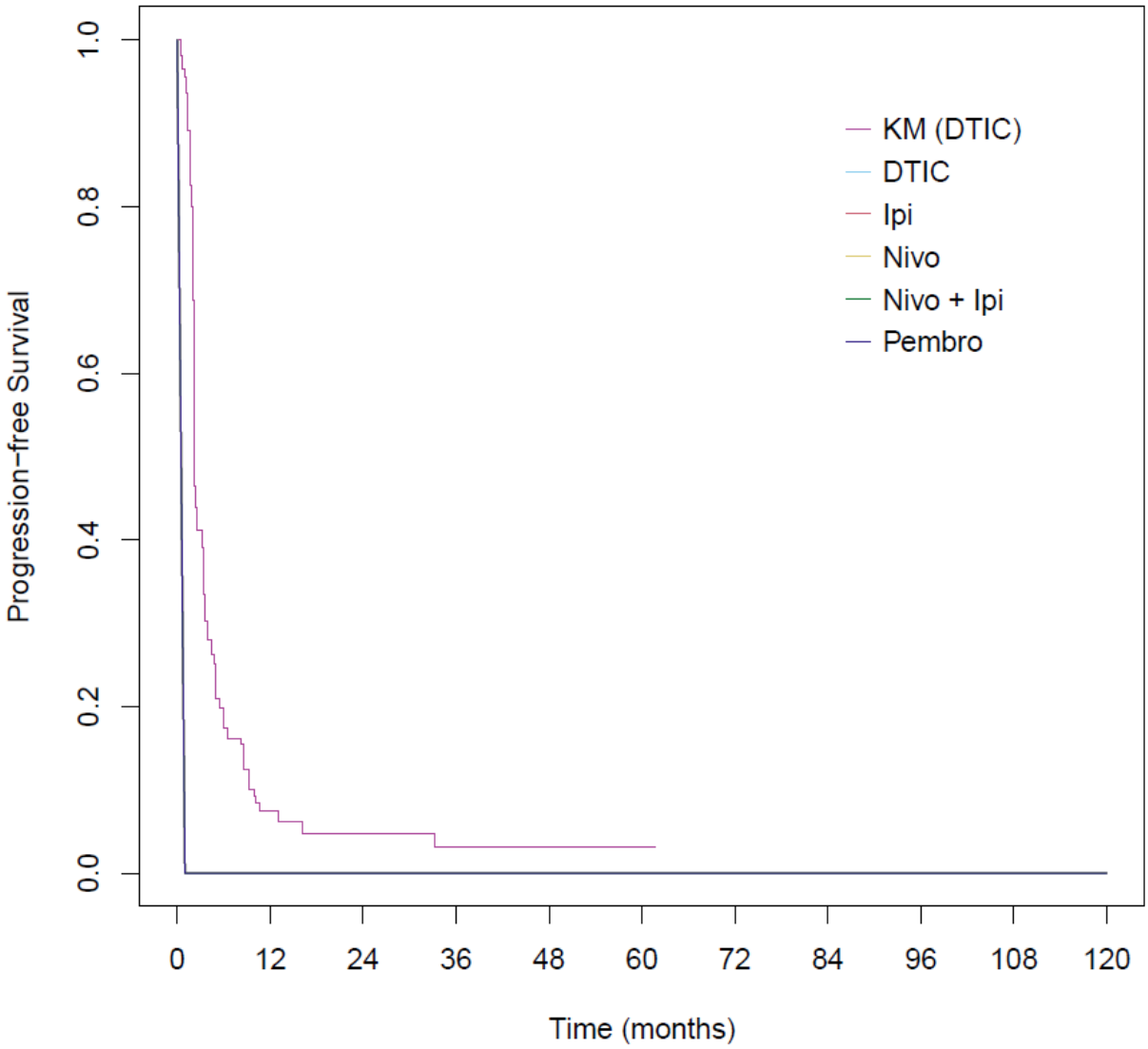


Figure 103: Survival curves for fractional polynomial PFS model with a power of -2 for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

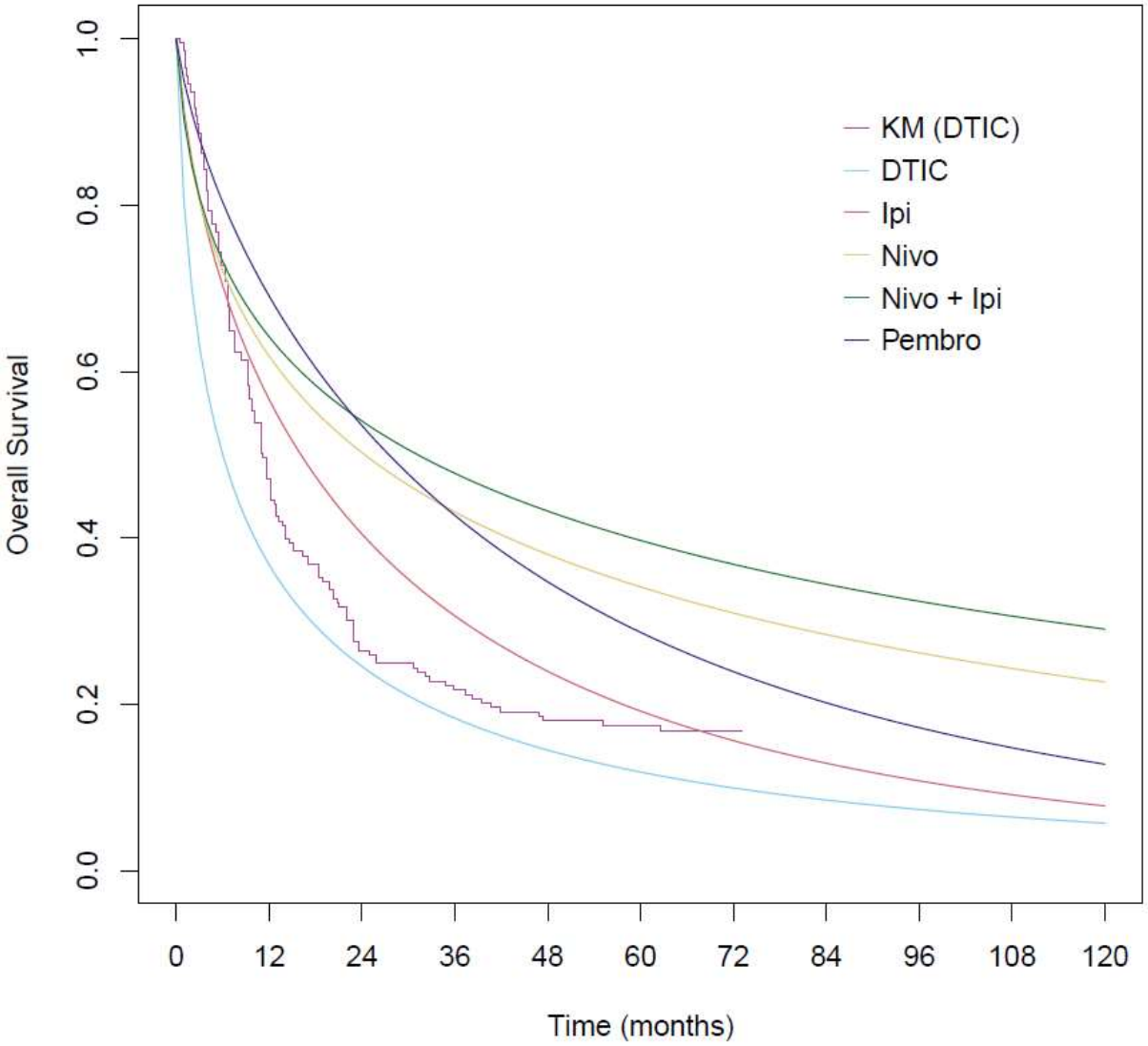


Figure 104: Survival curves for fractional polynomial OS model with a power of 0 for network 4 - people with *BRAF* wild type and mutant melanoma, with immunotherapy strategies only

F.1.1.5 Network 5 - People with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Table 107: Convergence table for PFS and OS for 1st order fractional polynomial models for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
-2	Yes	1213.4	1030.2	Yes	1129	1439.6
-1	No	1129.7	-	Yes	1129.4	1373.4
-0.5	Yes	1085.1[b]	850.1[c]	Yes	1123.9[b]	1295.9[c]
0	No	1056.9[a]	-	No	1115.3	-

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
0.5	No	1067	-	No	1109.1[a]	-
1	No	1131.7	-	Undefined real result	1110.4	-
2	No	1390.7	-	Undefined real result	1128.2	-
3	Yes	1655.4	7E+13	Yes	1139.9	1.1E+14

(a) Lowest AIC of any power – model did not converge
(b) Lowest AIC of models that did converge
(c) Lowest DIC

As shown in Table 107, of the fractions that converged, -0.5 had both the lowest AIC and DIC values for PFS and OS. The AIC values calculated in R prior to running the NMA were smallest with the fractions 0 and 0.5 for PFS and OS respectively. However, neither of these models converged.

Table 108: Fixed effect PFS and OS NMA results for best fitting fractional polynomial model for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

Comparison	PFS		OS	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	-0.5		-0.5	
Pooled estimate for difference β_0				
dabrafenib vs. DTIC	0.481	(-0.928; 1.71)	-0.219	(-1.168; 0.667)
dabrafenib + trametinib vs. DTIC	-0.204	(-1.61; 1.074)	-0.233	(-1.207; 0.722)
ipilimumab vs. DTIC	-1.077	(-3.088; 0.441)	0.345	(-0.396; 1.197)
vemurafenib vs. DTIC	0.779	(-0.714; 2.201)	0.174	(-0.865; 1.198)
nivolumab vs. DTIC	-2.292	(-3.957; -1.042)	-0.645	(-1.228; 0)
nivolumab + ipilimumab vs. DTIC	-2.298	(-4.093; -0.836)	-0.841	(-1.681; 0.045)
encorafenib + binimetinib vs. DTIC	0.754	(-1.133; 2.566)	-0.215	(-1.507; 0.972)
pembrolizumab vs. DTIC	-0.992	(-2.944; 0.528)	0.274	(-0.722; 1.291)
Pooled estimate for difference β_1				
dabrafenib vs. DTIC	-3.385	(-6.589; 0.399)	0.206	(-2.926; 3.538)
dabrafenib + trametinib vs. DTIC	-2.693	(-6.098; 1.188)	-0.752	(-4.21; 2.792)
ipilimumab vs. DTIC	2.279	(-1.663; 7.458)	-2.059	(-4.919; 0.582)
vemurafenib vs. DTIC	-4.306	(-8.198; -0.268)	-0.866	(-4.582; 2.906)
nivolumab vs. DTIC	3.877	(0.521; 8.147)	-0.184	(-2.292; 1.814)

Comparison	PFS		OS	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	3.261	(-0.66; 7.81)	-0.253	(-3.334; 2.735)
encorafenib + binimetinib vs. DTIC	-5.515	(-10.49; -0.212)	-1.045	(-5.405; 3.776)
pembrolizumab vs. DTIC	0.529	(-3.5; 5.64)	-2.93	(-6.427; 0.372)

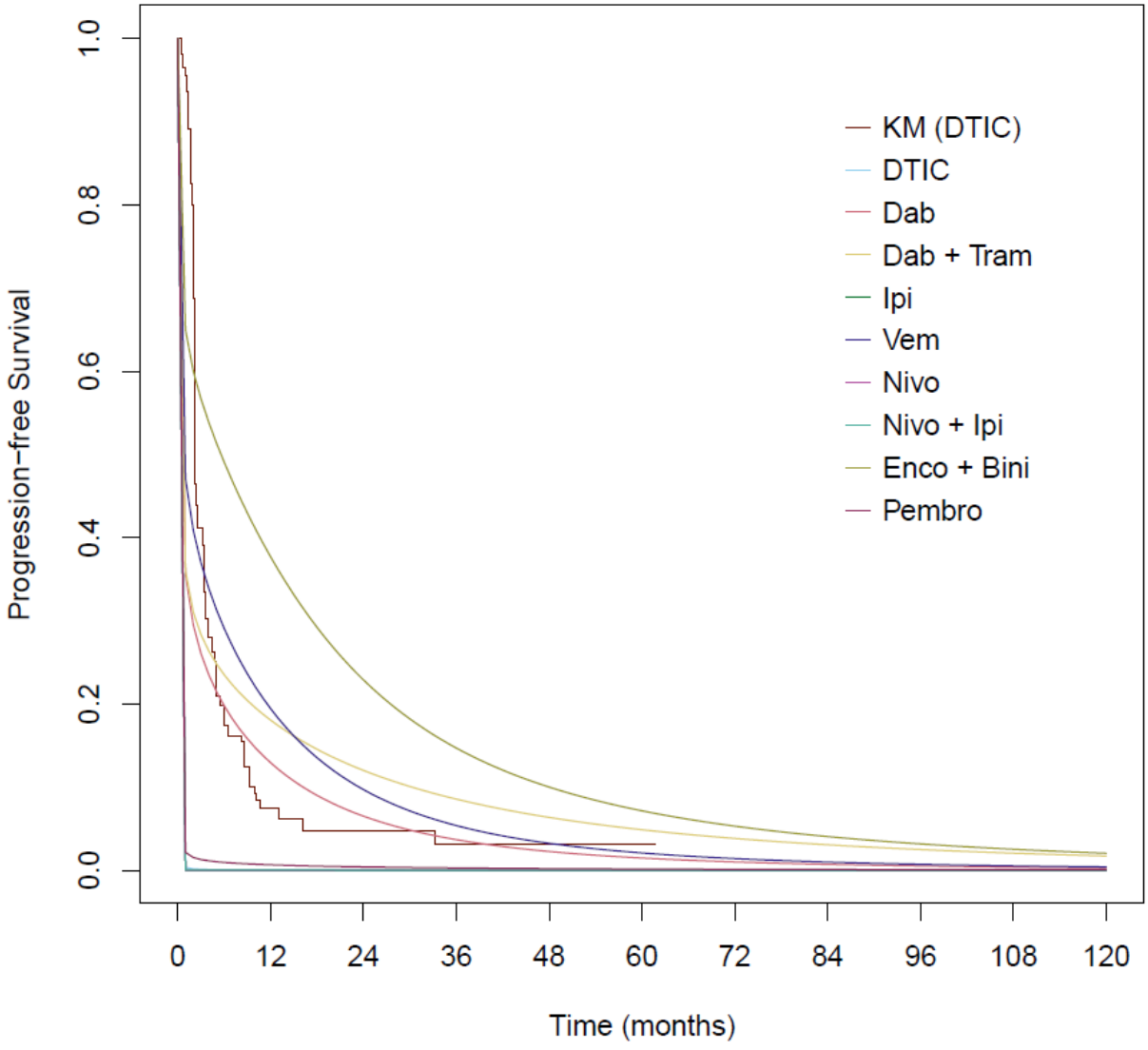


Figure 105: Survival curves for fractional polynomial PFS model with a power of -0.5 for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

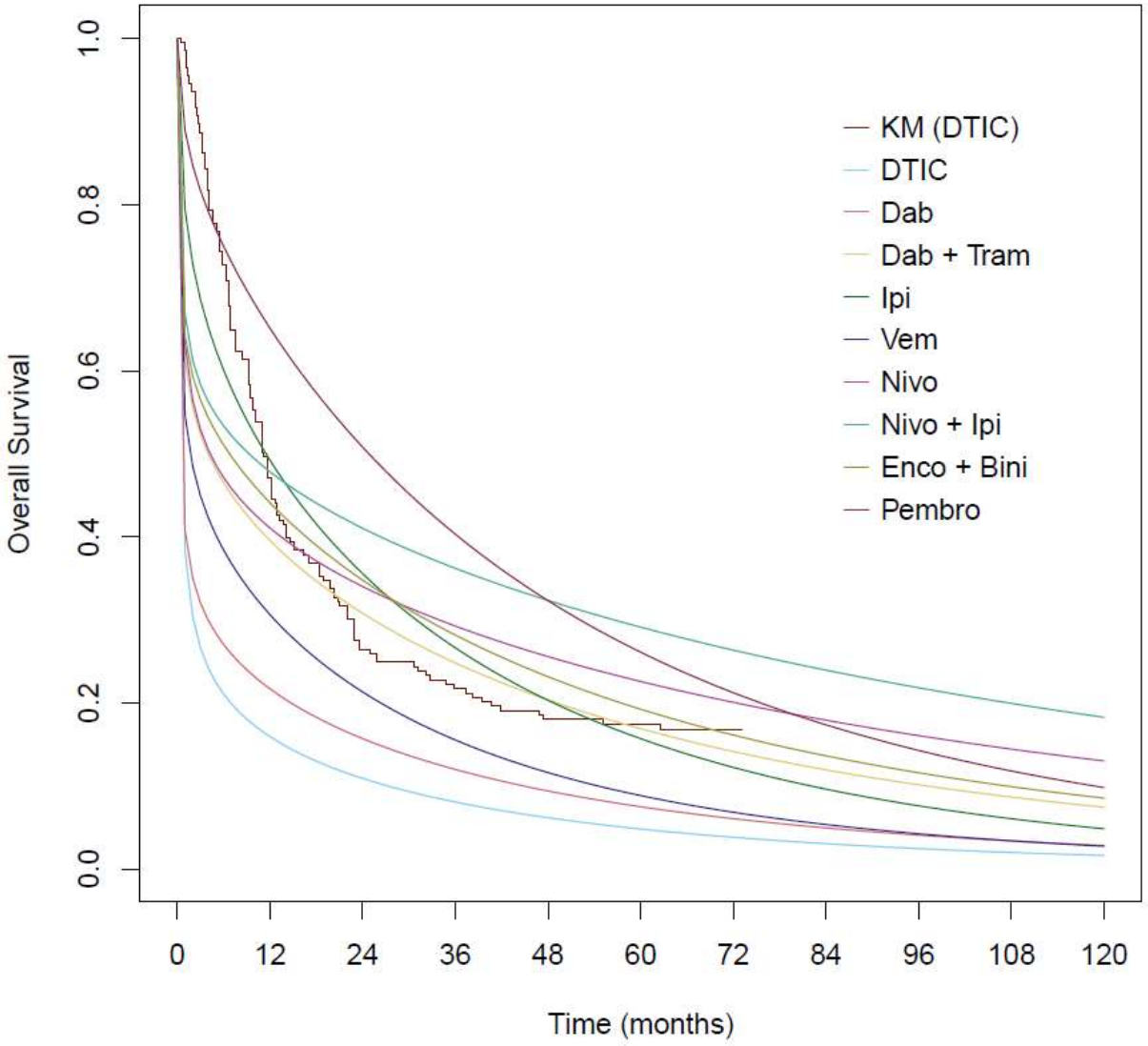


Figure 106: Survival curves for fractional polynomial OS model with a power of -0.5 for network 5 - people with *BRAF* wild type and mutant melanoma, all immunotherapy and targeted therapy strategies, studies with long-term follow-up

F.1.1.6 Network 6 - People with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Table 109: Convergence table for PFS and OS for 1st order fractional polynomial models for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
-2	Yes	340.9	373.5	Yes	470.7	617.2
-1	Yes	306.1	312.2	Yes	453.8	548.7
-0.5	Yes	296.8[a]	281.7[b]	Yes	443.8	493.2
0	No	307.5	-	Yes	437.2[a]	433.1

Power	PFS			OS		
	Convergence	AIC	DIC	Convergence	AIC	DIC
0.5	Yes	352.2	300.6	Yes	438.5	384.2[b]
1	No	437.7	-	No	448.9	-
2	No	691.1	-	No	477.4	-
3	Yes	927.9	3.1E+13	Yes	493.6	5.1E+13

(a) Lowest AIC of any power – model did converge
(b) Lowest DIC

As shown in Table 109, -0.5 had both the lowest AIC and DIC values for PFS. For OS, 0 had the lowest AIC value, and 0.5 had the lowest DIC value.

Table 110: Fixed effect PFS NMA results for best fitting fractional polynomial model for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Comparison	PFS	
	Median of posterior distribution	95% Credible Interval
p1 – power 1	-0.5	
Pooled estimate for difference β_0		
ipilimumab vs. DTIC	-0.896	(-2.065; 0.185)
nivolumab vs. DTIC	-2.115	(-3.29; -1.087)
nivolumab + ipilimumab vs. DTIC	-2.082	(-3.342; -0.981)
pembrolizumab vs. DTIC	-0.933	(-2.316; 0.426)
Pooled estimate for difference β_1		
ipilimumab vs. DTIC	1.806	(-1; 4.938)
nivolumab vs. DTIC	3.421	(0.713; 6.481)
nivolumab + ipilimumab vs. DTIC	2.694	(-0.293; 6.053)
pembrolizumab vs. DTIC	0.375	(-3.17; 4.022)

Table 111: Fixed effect OS NMA results for best fitting fractional polynomial model assessed by AIC and DIC for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

Comparison	OS - AIC		OS - DIC	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	0		0.5	
Pooled estimate for difference β_0				
ipilimumab vs. DTIC	-1.109	(-1.974; -0.036)	-0.98	(-1.716; -0.248)
nivolumab vs. DTIC	-0.951	(-1.665; -0.103)	-1.015	(-1.611; -0.437)

Comparison	OS - AIC		OS - DIC	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
nivolumab + ipilimumab vs. DTIC	-1.048	(-2.076; 0.127)	-1.115	(-1.887; -0.336)
pembrolizumab vs. DTIC	-1.794	(-2.851; -0.587)	-1.581	(-2.494; -0.638)
Pooled estimate for difference β_1				
ipilimumab vs. DTIC	0.326	(-0.033; 0.623)	0.183	(0.027; 0.346)
nivolumab vs. DTIC	0.099	(-0.192; 0.353)	0.083	(-0.046; 0.22)
nivolumab + ipilimumab vs. DTIC	0.059	(-0.335; 0.404)	0.059	(-0.107; 0.232)
pembrolizumab vs. DTIC	0.467	(0.051; 0.838)	0.255	(0.054; 0.463)

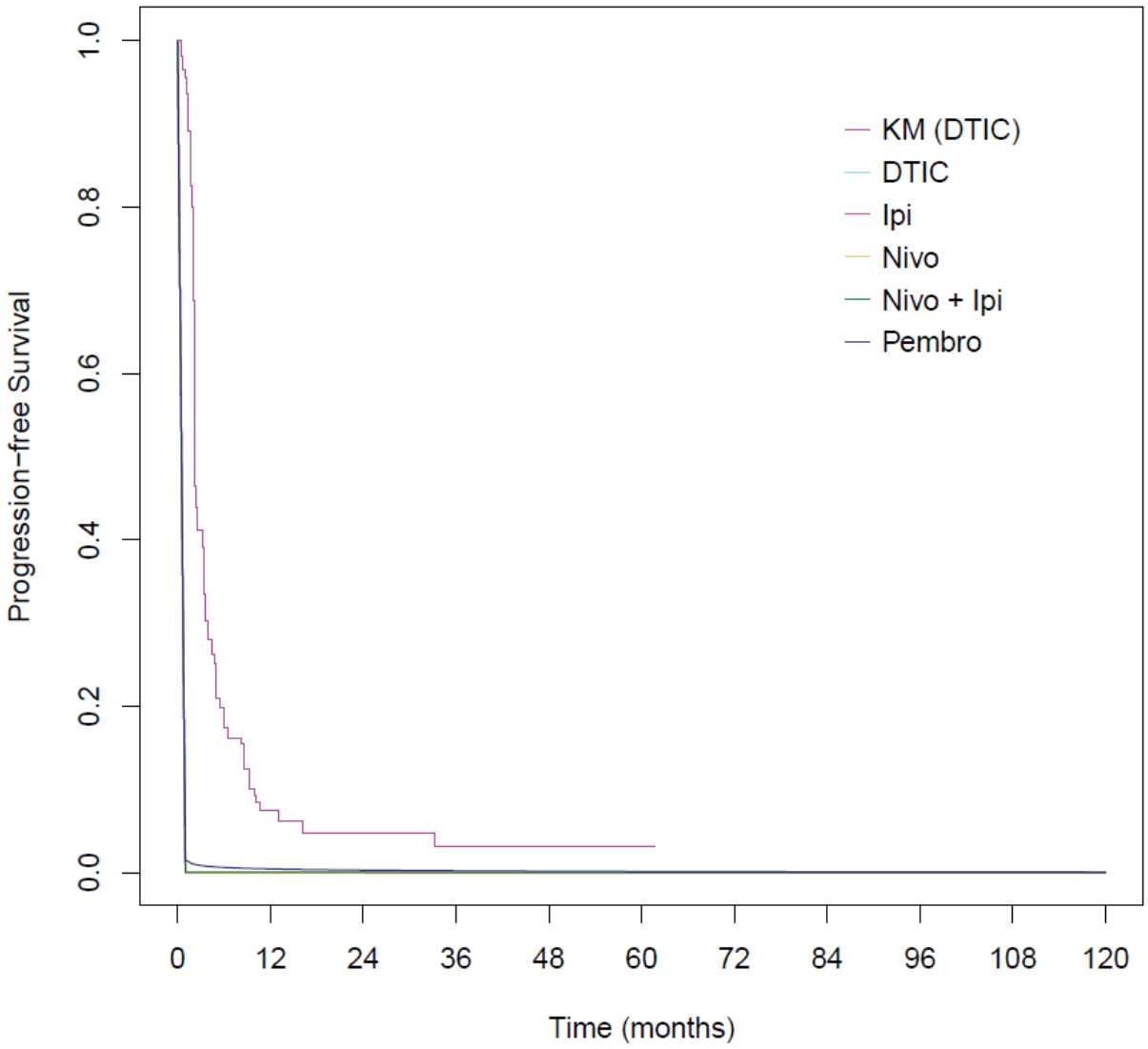


Figure 107: Survival curves for fractional polynomial PFS model with a power of -0.5 for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

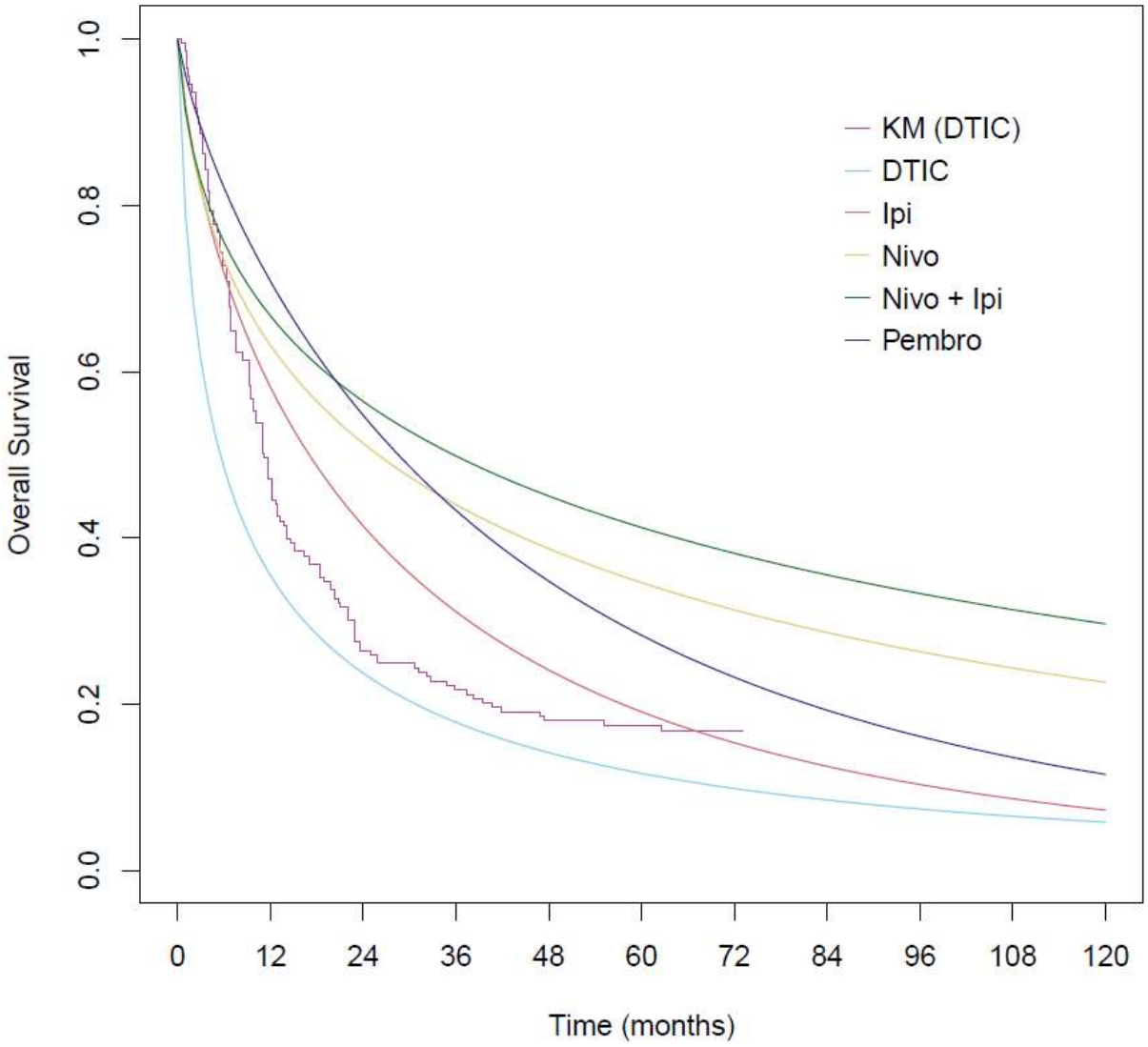


Figure 108: Survival curves for fractional polynomial OS model with a power of 0 (smallest AIC value) for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

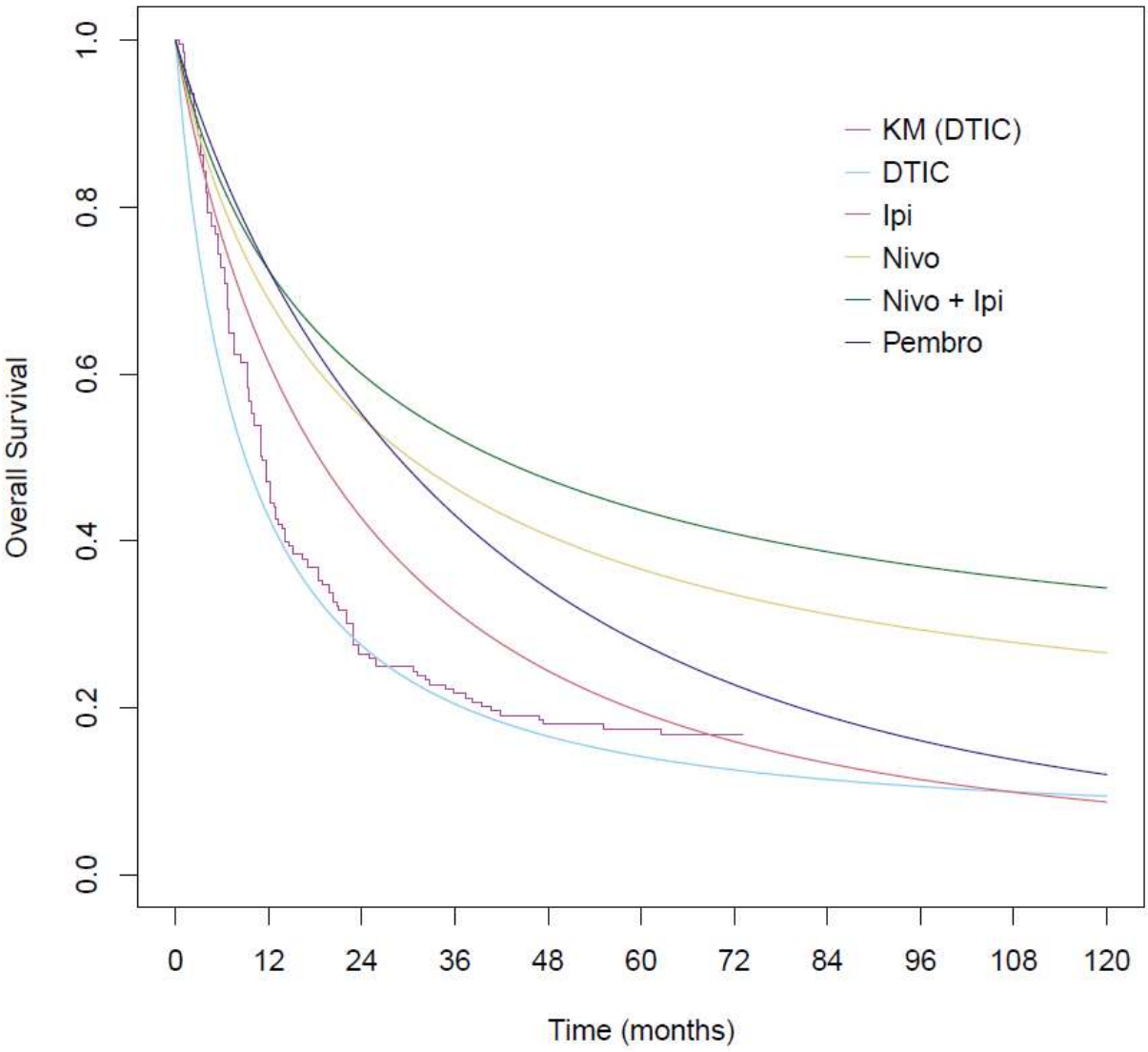


Figure 109: Survival curves for fractional polynomial OS model with a power of 0.5 (smallest DIC value) for network 6 - people with *BRAF* wild type and mutant melanoma, immunotherapy strategies, studies with long-term follow-up

F.1.1.7 Impact of different network assumptions on NMA results for fractional polynomial models

Table 112: A comparison of PFS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	-0.5		-0.5	
Pooled estimate for difference β_0				

Comparison	Network 1		Network 5	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
dabrafenib vs. DTIC	0.826	(-0.28; 2.183)	0.481	(-0.928; 1.71)
dabrafenib + trametinib vs. DTIC	0.169	(-0.972; 1.552)	-0.204	(-1.61; 1.074)
ipilimumab vs. DTIC	-1.094	(-2.622; 0.146)	-1.077	(-3.088; 0.441)
vemurafenib vs. DTIC	1.147	(-0.004; 2.466)	0.779	(-0.714; 2.201)
nivolumab vs. DTIC	-2.29	(-3.519; -0.807)	-2.292	(-3.957; -1.042)
nivolumab + ipilimumab vs. DTIC	-2.36	(-3.785; -0.847)	-2.298	(-4.093; -0.836)
encorafenib + binimetinib vs. DTIC	1.08	(-0.437; 3.018)	0.754	(-1.133; 2.566)
pembrolizumab vs. DTIC	-1.091	(-3.086; 0.355)	-0.992	(-2.944; 0.528)
Pooled estimate for difference β_1				
dabrafenib vs. DTIC	-3.777	(-7.347; -0.795)	-3.385	(-6.589; 0.399)
dabrafenib + trametinib vs. DTIC	-2.929	(-6.486; 0.136)	-2.693	(-6.098; 1.188)
ipilimumab vs. DTIC	2.363	(-0.885; 6.366)	2.279	(-1.663; 7.458)
vemurafenib vs. DTIC	-4.328	(-7.844; -1.262)	-4.306	(-8.198; -0.268)
nivolumab vs. DTIC	3.885	(0.276; 7.1)	3.877	(0.521; 8.147)
nivolumab + ipilimumab vs. DTIC	3.417	(-0.392; 7.258)	3.261	(-0.66; 7.81)
encorafenib + binimetinib vs. DTIC	-5.409	(-10.63; -1.244)	-5.515	(-10.49; -0.212)
pembrolizumab vs. DTIC	0.821	(-3.032; 6.16)	0.529	(-3.5; 5.64)

Table 113: A comparison of PFS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up) and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3[a]		Network 6[a]	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	-0.5		-0.5	
Pooled estimate for difference β_0				
ipilimumab vs. DTIC	-0.791	(-2.167; 0.451)	-0.896	(-2.065; 0.185)
nivolumab vs. DTIC	-1.985	(-3.047; -0.753)	-2.115	(-3.29; -1.087)
nivolumab + ipilimumab vs. DTIC	-1.682	(-2.912; -0.387)	-2.082	(-3.342; -0.981)
pembrolizumab vs. DTIC	-0.769	(-2.22; 0.548)	-0.933	(-2.316; 0.426)

Comparison	Network 3[a]		Network 6[a]	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
Pooled estimate for difference β_1				
ipilimumab vs. DTIC	2.108	(-1.159; 5.72)	1.806	(-1; 4.938)
nivolumab vs. DTIC	3.087	(-0.072; 5.779)	3.421	(0.713; 6.481)
nivolumab + ipilimumab vs. DTIC	1.974	(-1.514; 5.127)	2.694	(-0.293; 6.053)
pembrolizumab vs. DTIC	0.516	(-2.977; 4.342)	0.375	(-3.17; 4.022)

(b) As seen in Table 105, the fractional polynomial model with a power of -0.5 did not converge for network 4. As such, we are unable to present any results in this table for network 4.

Table 114: A comparison of OS estimates for network 1 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, any amount of follow up) and network 5 (BRAF mutant and BRAF wild type population, targeted therapy and immunotherapy, long-term follow up only)

Comparison	Network 1		Network 5	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	-0.5		-0.5	
Pooled estimate for difference β_0				
dabrafenib vs. DTIC	-0.082	(-0.684; 0.563)	-0.219	(-1.168; 0.667)
dabrafenib + trametinib vs. DTIC	-0.081	(-0.686; 0.545)	-0.233	(-1.207; 0.722)
ipilimumab vs. DTIC	0.215	(-0.604; 1.066)	0.345	(-0.396; 1.197)
vemurafenib vs. DTIC	0.368	(-0.12; 0.841)	0.174	(-0.865; 1.198)
nivolumab vs. DTIC	-0.724	(-1.363; -0.094)	-0.645	(-1.228; 0)
nivolumab + ipilimumab vs. DTIC	-0.953	(-1.829; -0.127)	-0.841	(-1.681; 0.045)
encorafenib + binimetinib vs. DTIC	-0.066	(-0.841; 0.719)	-0.215	(-1.507; 0.972)
pembrolizumab vs. DTIC	0.153	(-0.769; 1.157)	0.274	(-0.722; 1.291)
Pooled estimate for difference β_1				
dabrafenib vs. DTIC	-0.439	(-2.707; 1.662)	0.206	(-2.926; 3.538)
dabrafenib + trametinib vs. DTIC	-1.524	(-3.754; 0.534)	-0.752	(-4.21; 2.792)
ipilimumab vs. DTIC	-1.678	(-4.383; 1.061)	-2.059	(-4.919; 0.582)
vemurafenib vs. DTIC	-1.853	(-3.375; -0.251)	-0.866	(-4.582; 2.906)
nivolumab vs. DTIC	0.079	(-1.963; 2.15)	-0.184	(-2.292; 1.814)
nivolumab + ipilimumab vs. DTIC	0.222	(-2.419; 3.069)	-0.253	(-3.334; 2.735)

Comparison	Network 1		Network 5	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
encorafenib + binimetinib vs. DTIC	-1.852	(-4.693; 1.011)	-1.045	(-5.405; 3.776)
pembrolizumab vs. DTIC	-2.577	(-5.811; 0.475)	-2.93	(-6.427; 0.372)

Table 115: A comparison of OS estimates for network 3 (BRAF wild type population, immunotherapy, any amount of follow up), network 4 (BRAF mutant and BRAF wild type population, immunotherapy, any amount of follow up), and network 6 (BRAF mutant and BRAF wild type population, immunotherapy, long-term follow up only)

Comparison	Network 3		Network 4		Network 6	
	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval	Median of posterior distribution	95% Credible Interval
p1 – power 1	0		0		0	
Pooled estimate for difference β_0						
ipilimumab vs. DTIC	-0.934	(-2.07; 0.239)	-0.934	(-1.801; -0.001)	-1.109	(-1.974; -0.036)
nivolumab vs. DTIC	-0.868	(-1.614; -0.018)	-0.8	(-1.534; -0.084)	-0.951	(-1.665; -0.103)
nivolumab + ipilimumab vs. DTIC	-0.418	(-1.594; 0.705)	-0.821	(-1.763; 0.064)	-1.048	(-2.076; 0.127)
pembrolizumab vs. DTIC	-1.96	(-3.304; -0.275)	-1.562	(-2.752; -0.463)	-1.794	(-2.851; -0.587)
Pooled estimate for difference β_1						
ipilimumab vs. DTIC	0.259	(-0.157; 0.654)	0.261	(-0.084; 0.568)	0.326	(-0.033; 0.623)
nivolumab vs. DTIC	0.071	(-0.228; 0.331)	0.046	(-0.211; 0.306)	0.099	(-0.192; 0.353)
nivolumab + ipilimumab vs. DTIC	-0.138	(-0.533; 0.276)	-0.013	(-0.329; 0.324)	0.059	(-0.335; 0.404)
pembrolizumab vs. DTIC	0.556	(-0.042; 1.05)	0.38	(-0.005; 0.806)	0.467	(0.051; 0.838)

F.1.1.8 Model validation – progression-free survival

The fractional polynomial models were ruled out due to implausible PFS predictions. As shown in Figure 110-Figure 114, these models frequently predicted progression-free survival for the systemic cancer treatments was worse than DTIC, and at times was 0.

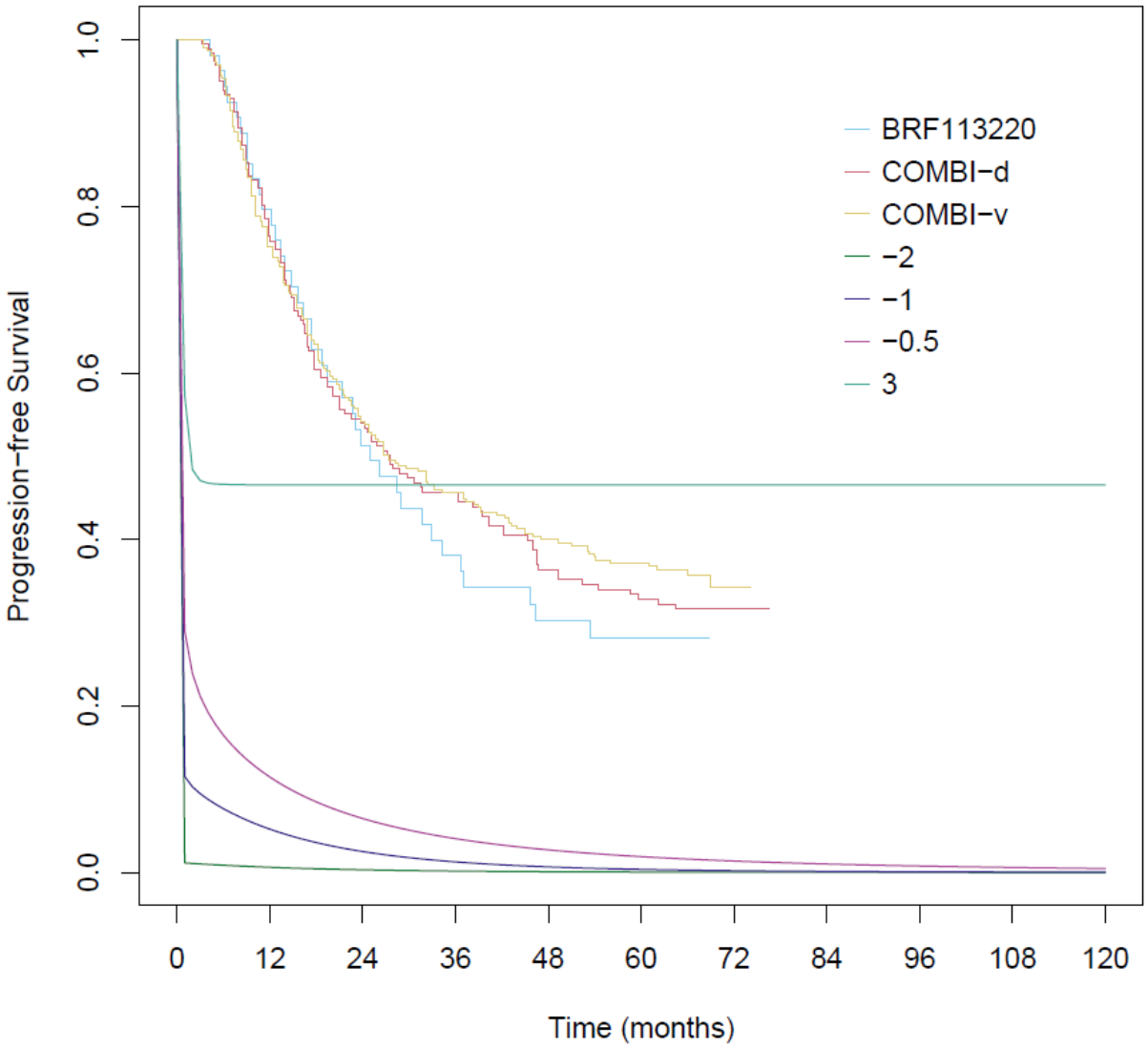


Figure 110: Predicted PFS from FP models that converged compared with observed KM data for Dabrafenib + Trametinib

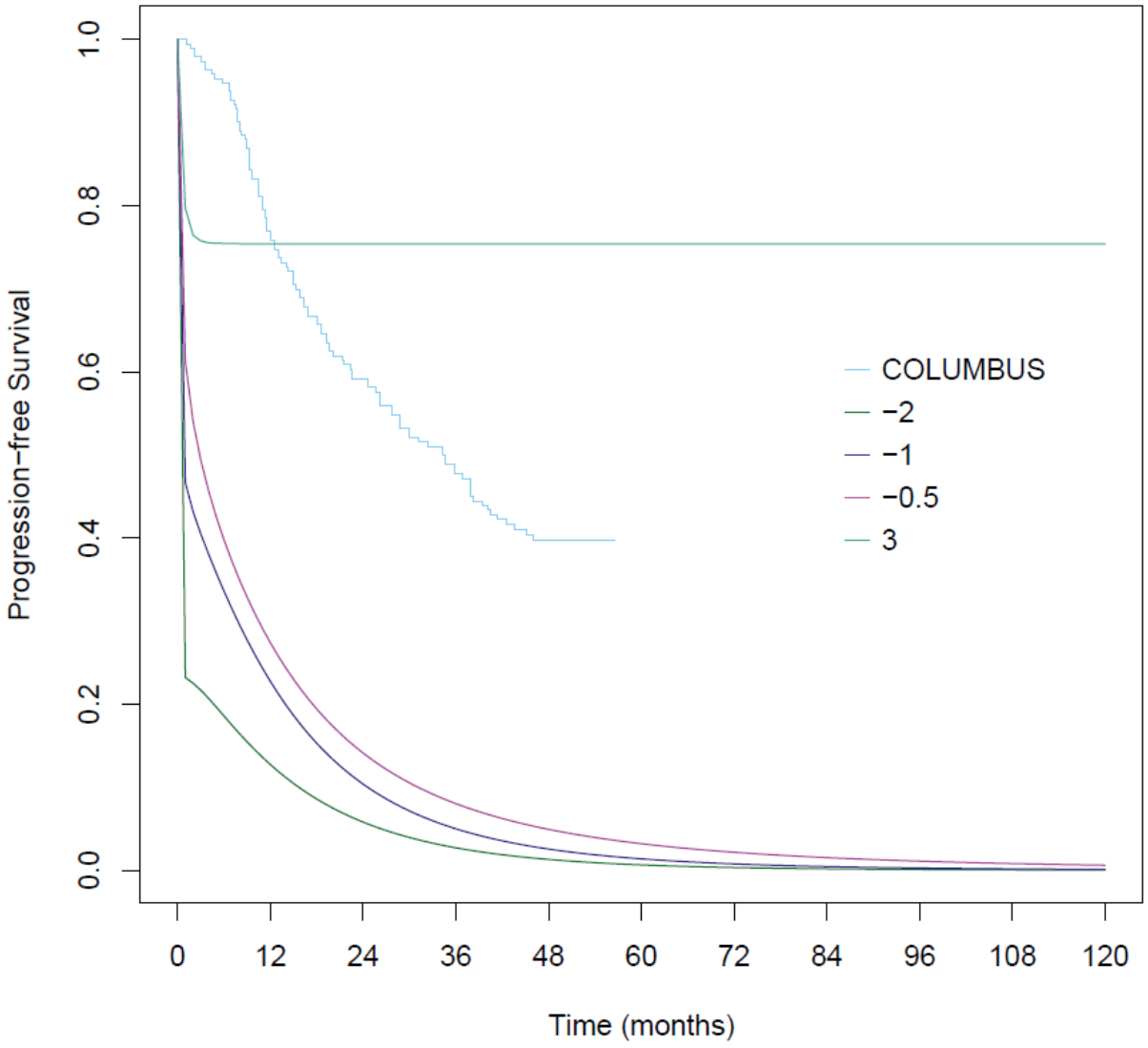


Figure 111: Predicted PFS from FP models that converged compared with observed KM data for Encorafenib + binimetinib

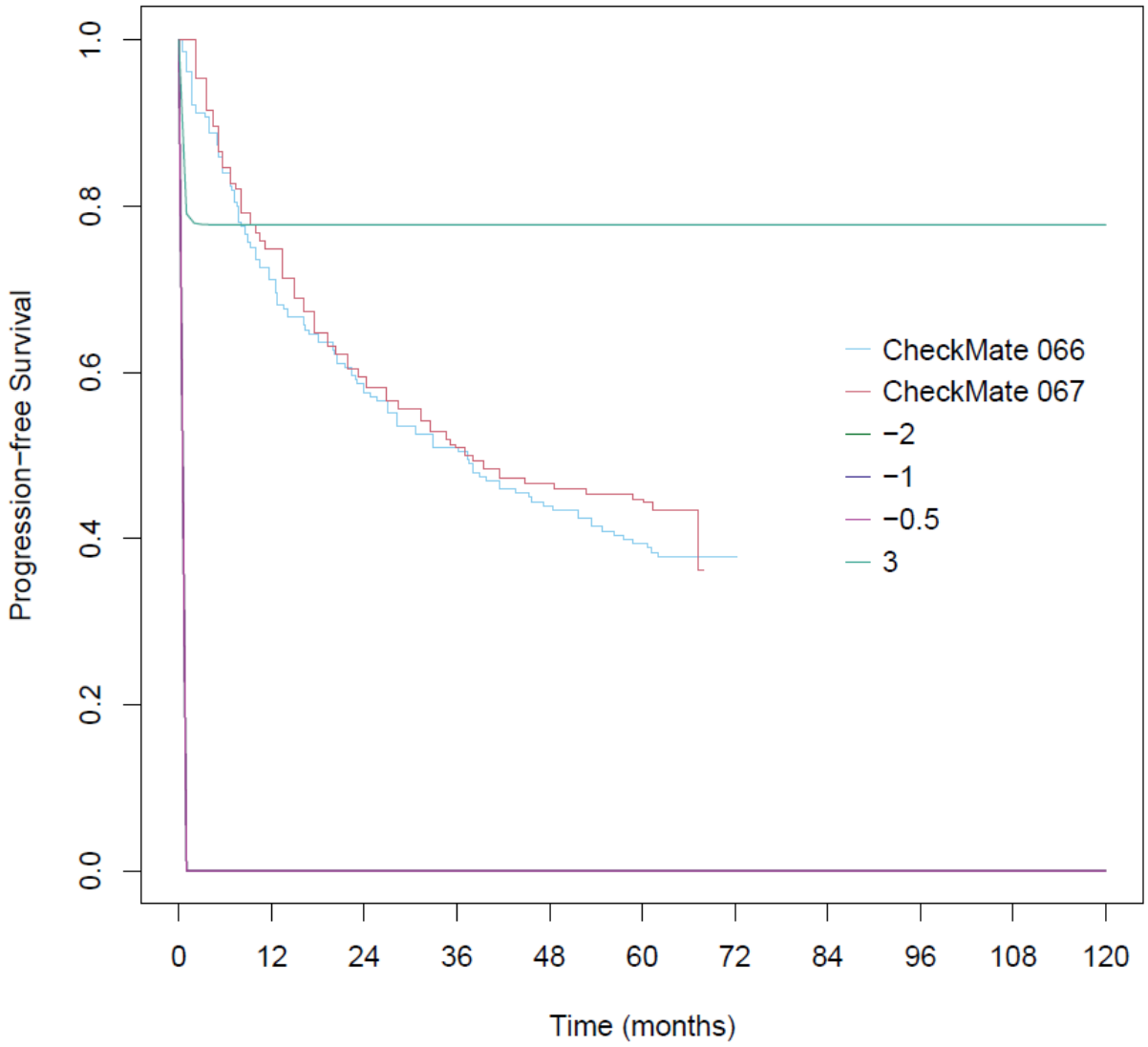


Figure 112: Predicted PFS from FP models that converged compared with observed KM data for nivolumab

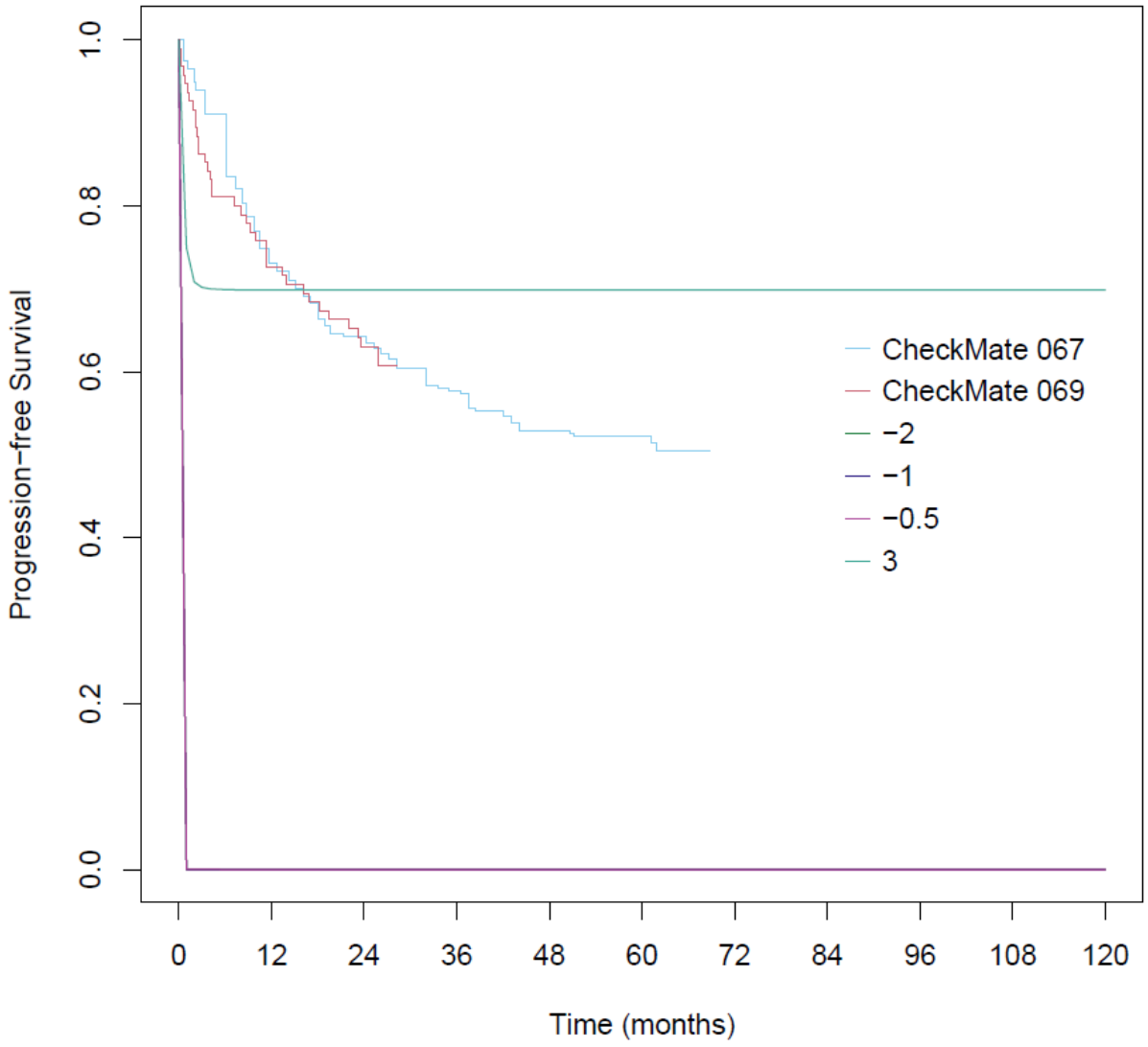


Figure 113: Predicted PFS from FP models that converged compared with observed KM data for nivolumab + ipilimumab

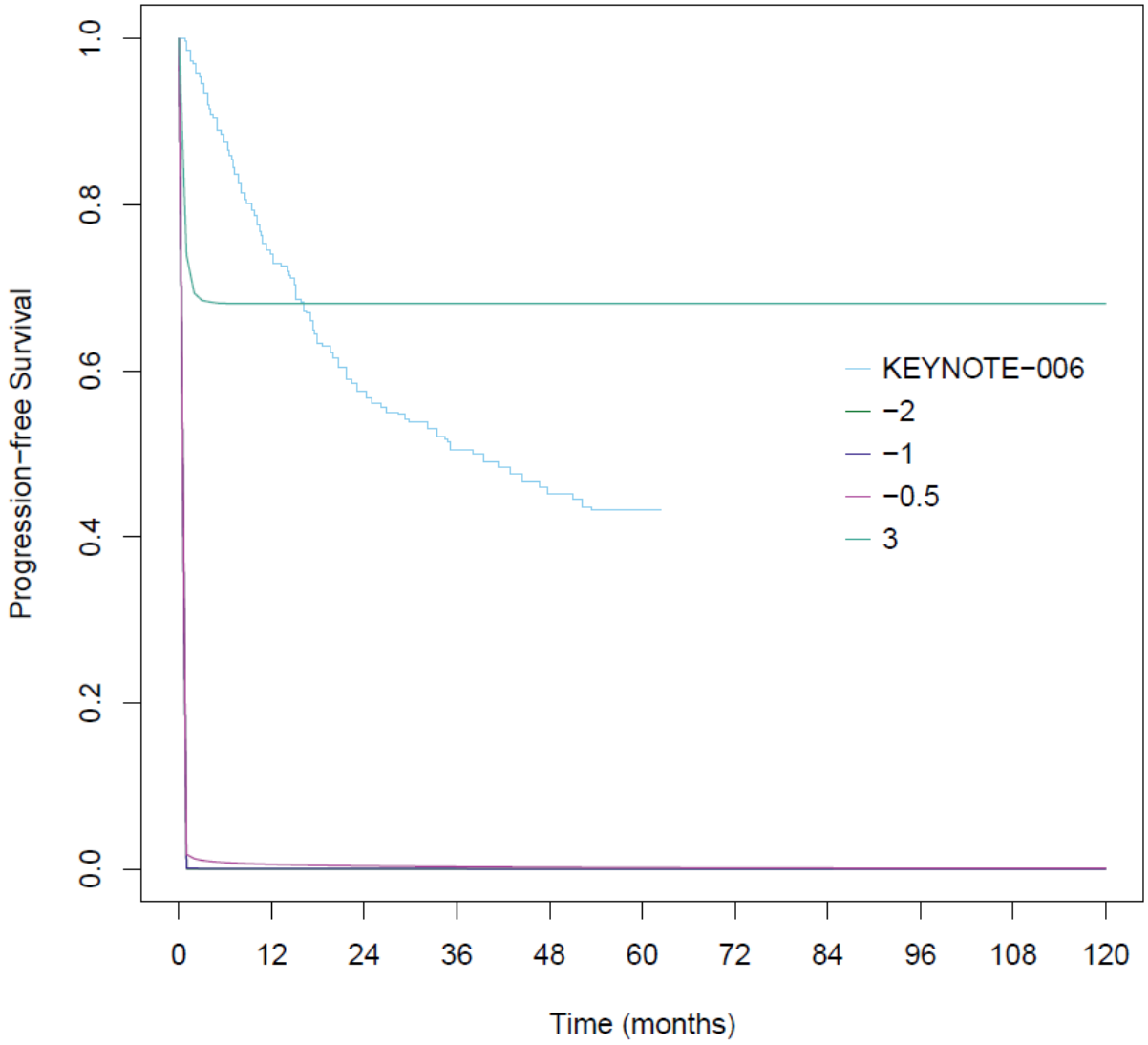


Figure 114: Predicted PFS from FP models that converged compared with observed KM data for pembrolizumab

F.1.1.9 Model validation – overall survival

The predicted OS outcomes for the fractional polynomial models by treatment were still lower than the observed OS from the trials. However, while these models may not match well to the observed KM data, their long-term extrapolations may be more plausible than other models due to their levelling off. Additionally, although a direct comparison using the DIC values is not possible, these models were incredibly complex requiring significant time to run. Thus, while it may not be possible to directly compare the complexity of the models using a measure such as DIC, it is reasonable to say the FP models are more complicated models. In the end, because the FP models were more complicated and did not provide a substantially improved fit, these models were immediately ruled out for OS and discarded.

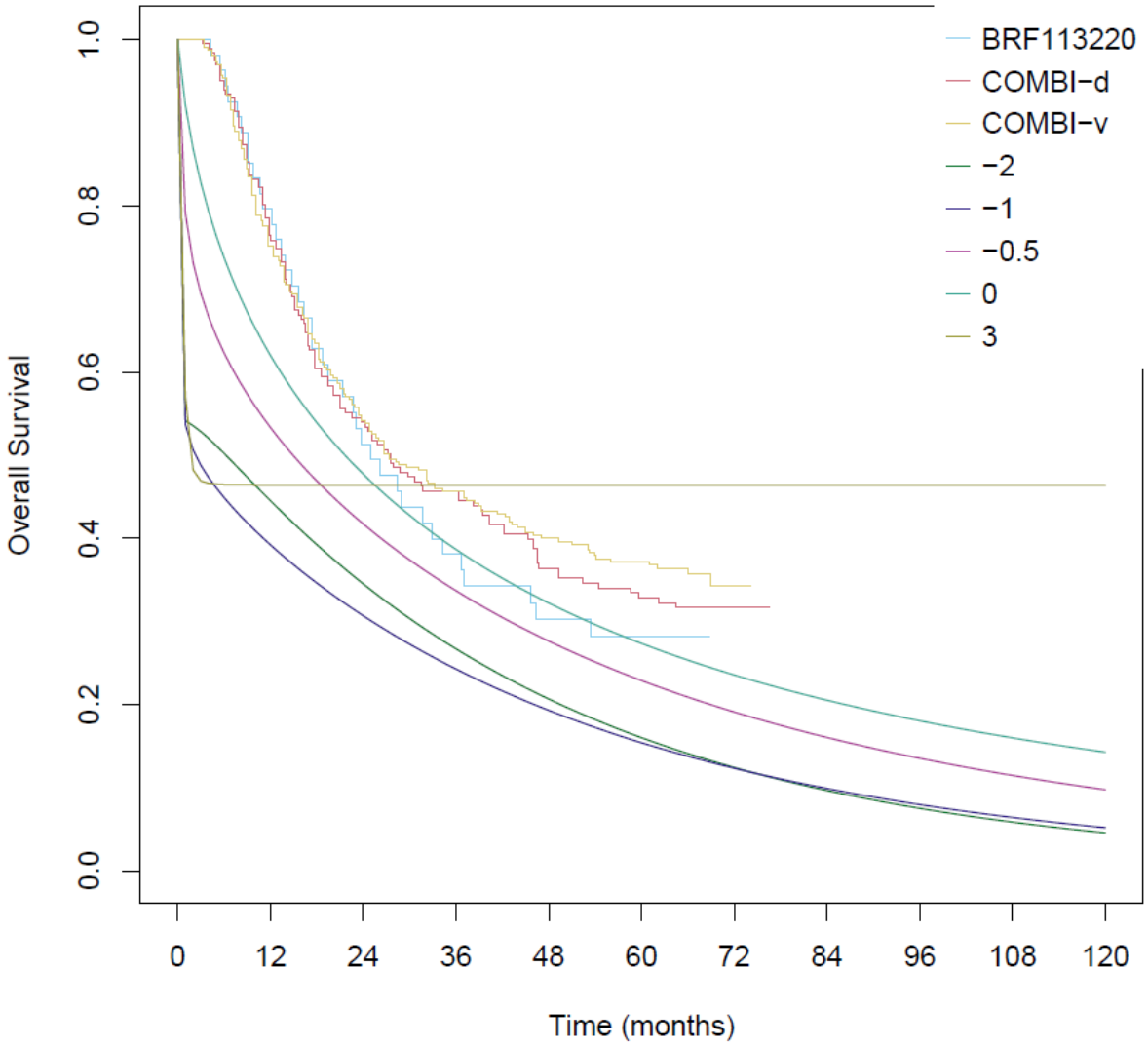


Figure 115: Predicted OS from FP models that converged compared with observed KM data for Dabrafenib + Trametinib

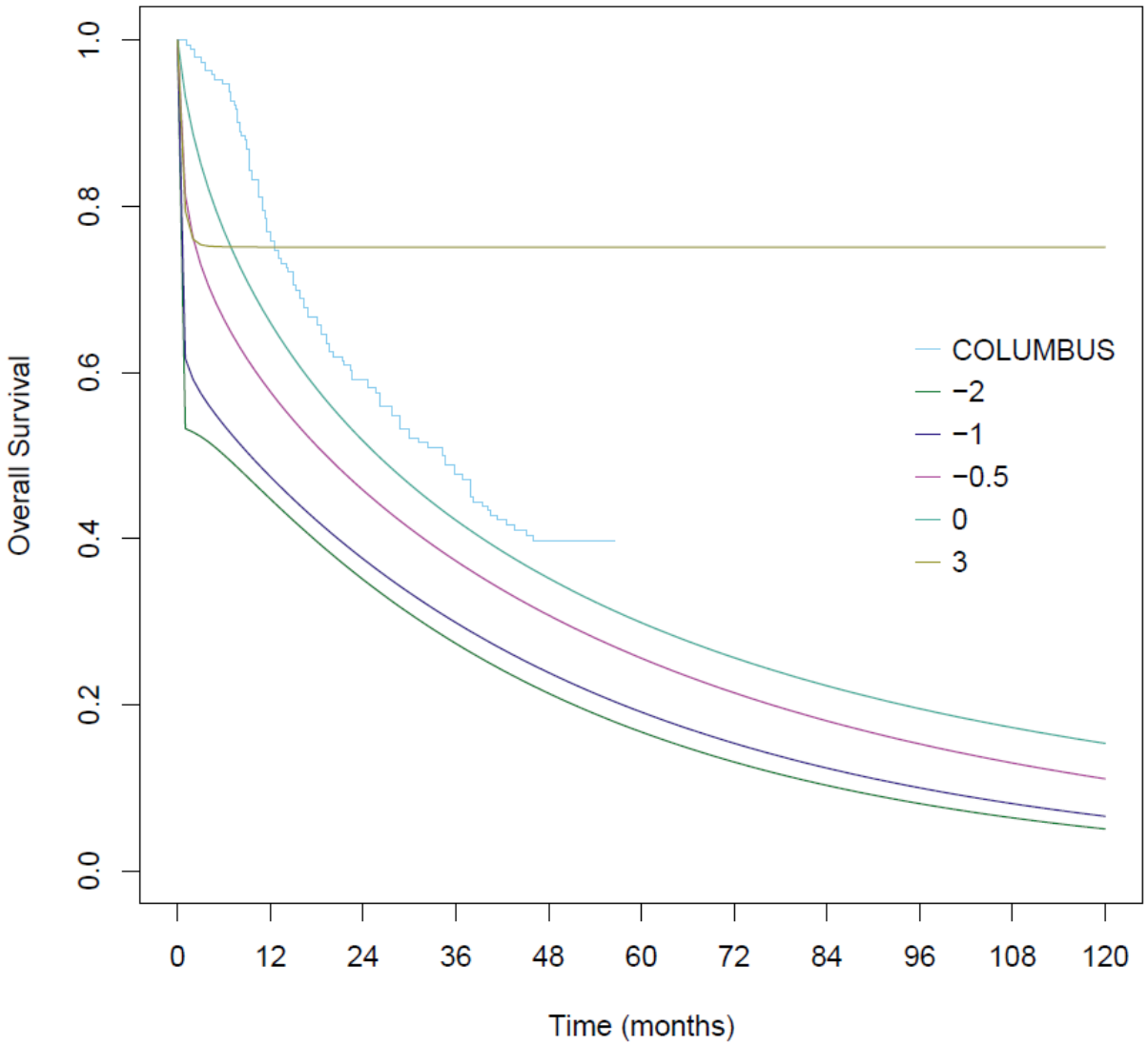


Figure 116: Predicted OS from FP models that converged compared with observed KM data for encorafenib + binimetinib

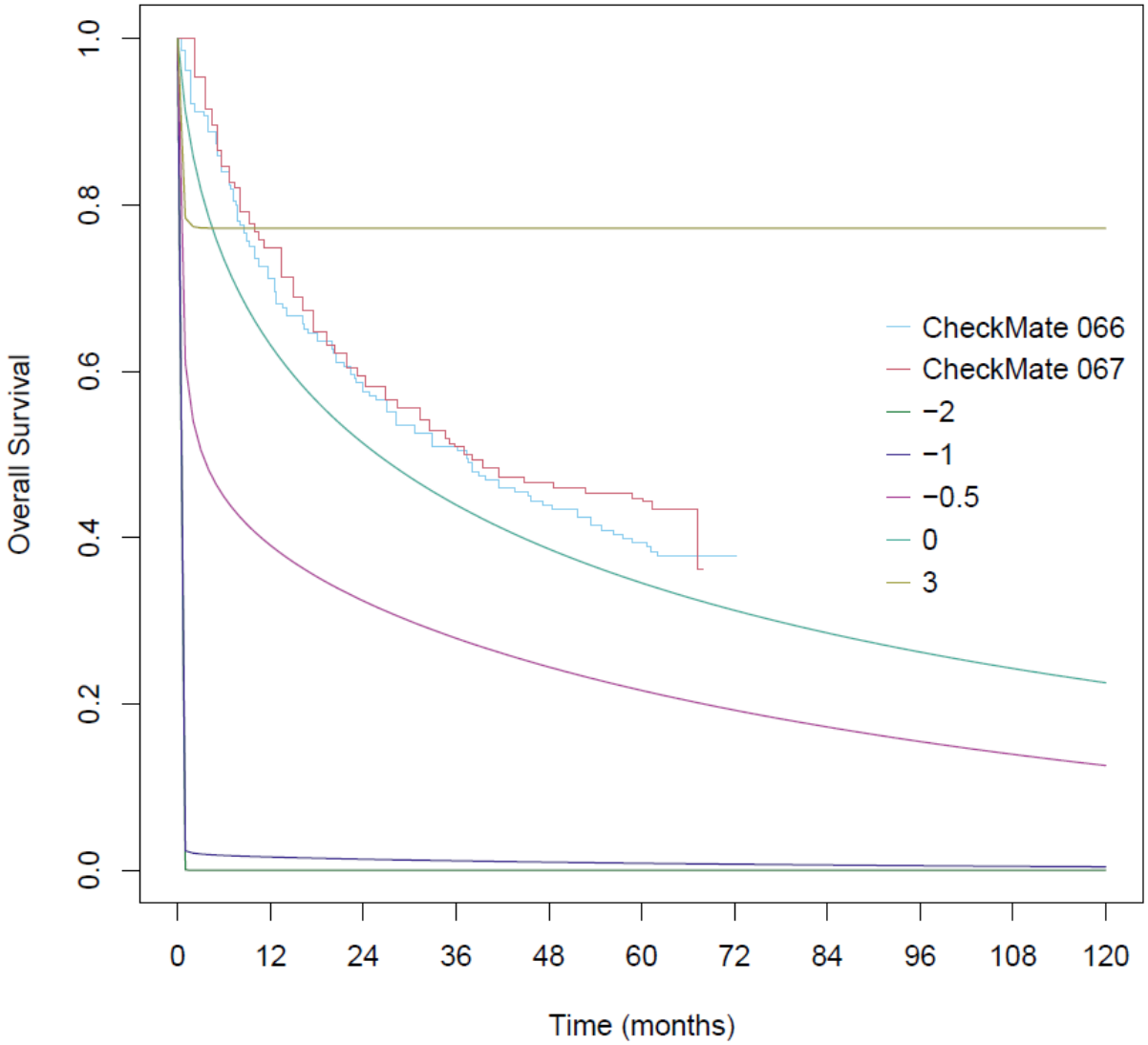


Figure 117: Predicted OS from FP models that converged compared with observed KM data for nivolumab

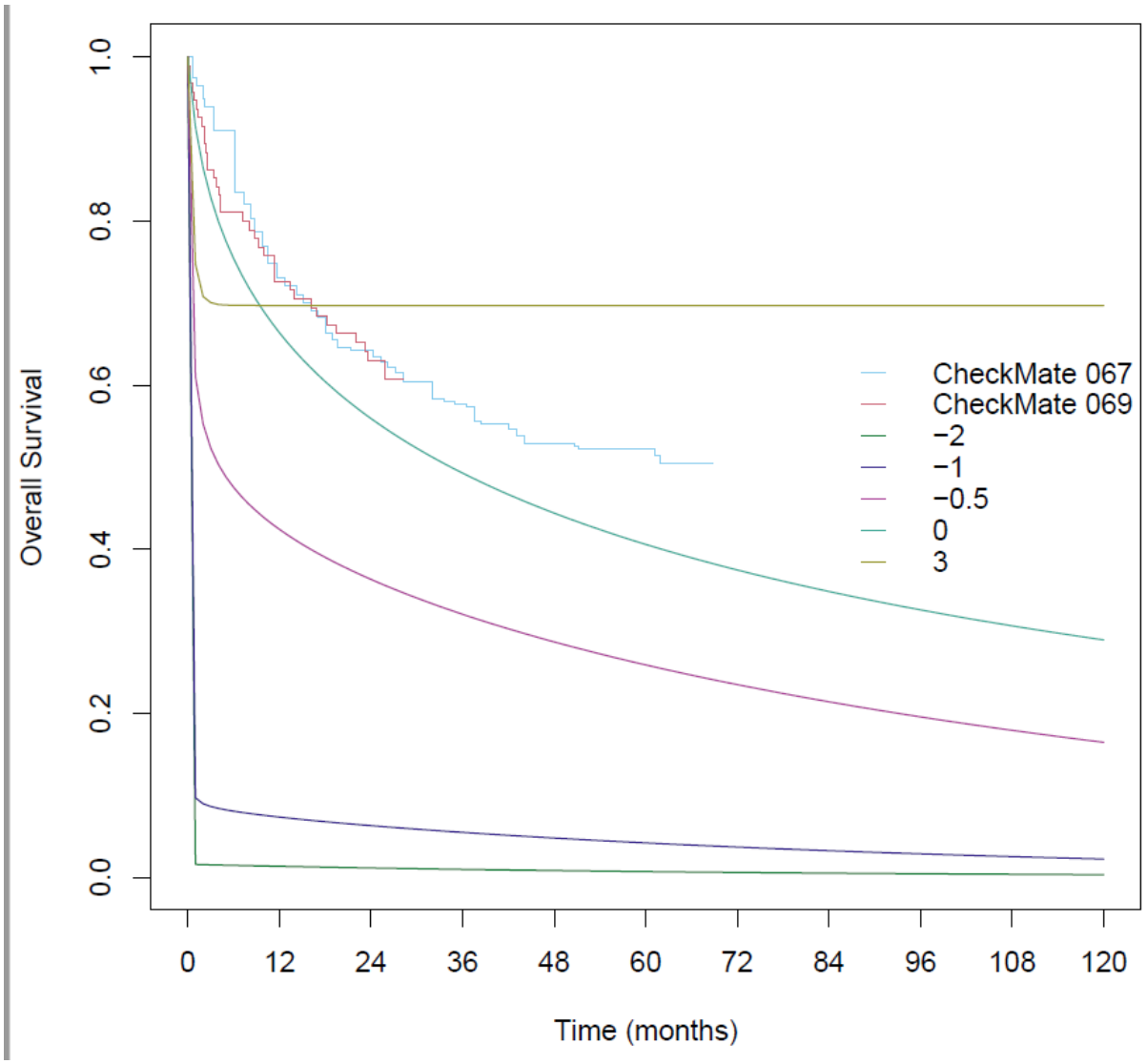


Figure 118: Predicted OS from FP models that converged compared with observed KM data for nivolumab + ipilimumab

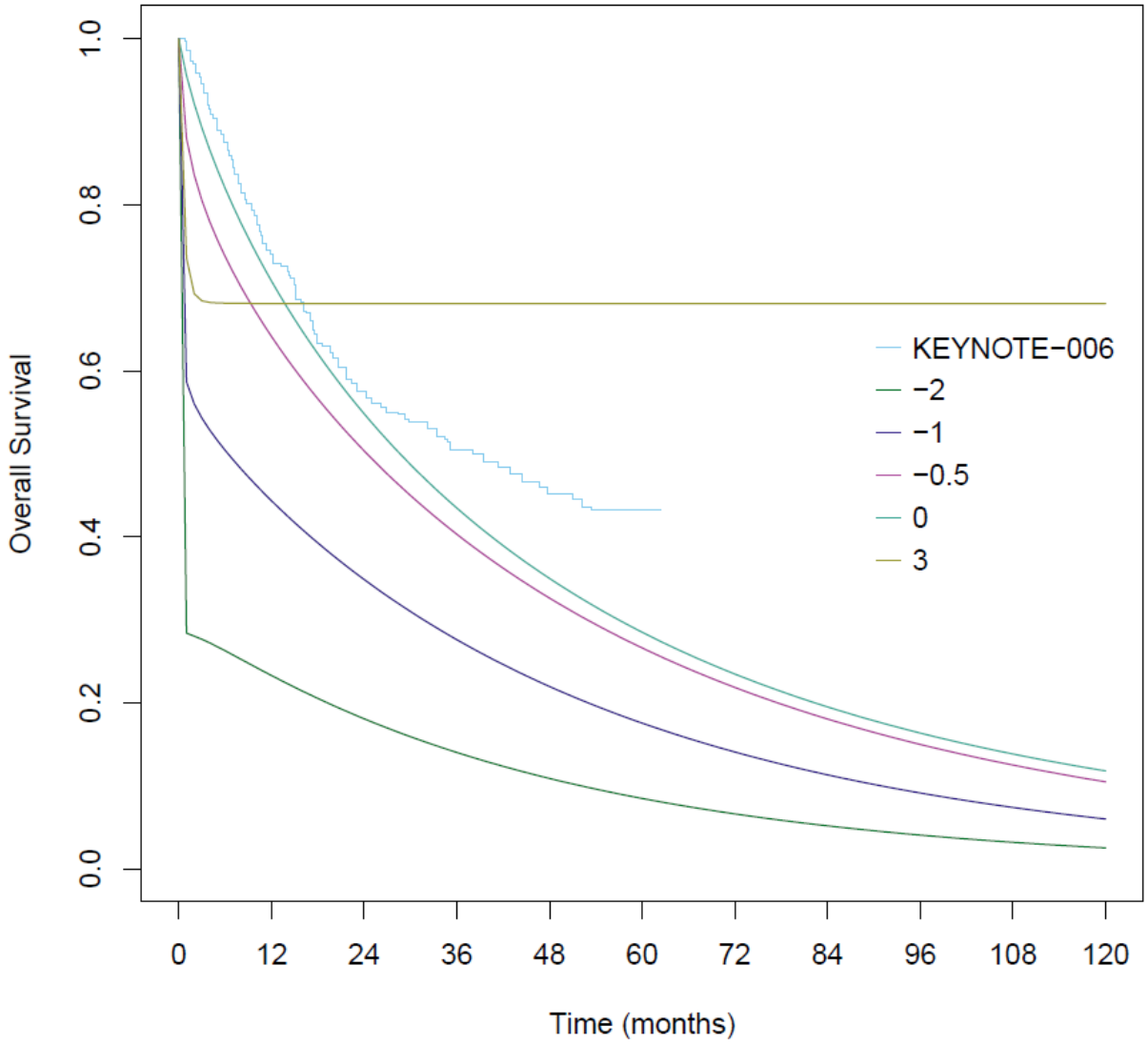


Figure 119: Predicted OS from FP models that converged compared with observed KM data for pembrolizumab

Appendix G: Prisma NMA Checklist (35)

PRISMA NMA Checklist of Items to Include When Reporting A Systematic Review Involving a Network Meta-analysis

Section/Topic	Item #	Checklist Item	Reported on Page #
TITLE			
Title	1	Identify the report as a systematic review *incorporating a network meta-analysis (or related form of meta-analysis).*	***1***
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: **Background:** main objectives **Methods:** data sources; study eligibility	This was not an academic publication so does not have an abstract. The report is

		criteria, participants, and interventions; study appraisal; and *synthesis methods, such as network meta-analysis.* **Results:** number of studies and participants identified; summary estimates with corresponding confidence/credible intervals; *treatment rankings may also be discussed. Authors may choose to summarize pairwise comparisons against a chosen treatment included in their analyses for brevity.* **Discussion/Conclusions:** limitations; conclusions and implications of findings. **Other:** primary source of funding; systematic review registration number with registry name.	structured into these categories.
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known*, including mention of why a network meta-analysis has been conducted.*	***A1***
Objectives	4	Provide an explicit statement of questions being addressed, with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	A1
METHODS			
Protocol and registration	5	Indicate whether a review protocol exists and if and where it can be accessed (e.g., Web address); and, if available, provide registration information, including registration number.	A2.1.2
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale. *Clearly describe eligible treatments included in the treatment network, and note whether any have been clustered or merged into the same node (with justification).*	***A2.1.3***
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	A2.1.2
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Evidence review linked in A2.1.2
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	A2.1.3
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	A2.3, A2.4

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	A1
Geometry of the network	**S1**	Describe methods used to explore the geometry of the treatment network under study and potential biases related to it. This should include how the evidence base has been graphically summarized for presentation, and what characteristics were compiled and used to describe the evidence base to readers.	***A4.2***
Risk of bias within individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	Evidence review linked in A2.1.2
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means). *Also describe the use of additional summary measures assessed, such as treatment rankings and surface under the cumulative ranking curve (SUCRA) values, as well as modified approaches used to present summary findings from meta-analyses.*	A2.5.3-A2.5.5
Planned methods of analysis	14	Describe the methods of handling data and combining results of studies for each network meta-analysis. This should include, but not be limited to: • *Handling of multi-arm trials;* • *Selection of variance structure;* • *Selection of prior distributions in Bayesian analyses; and* • *Assessment of model fit.*	A3.7
Assessment of Inconsistency	**S2**	Describe the statistical methods used to evaluate the agreement of direct and indirect evidence in the treatment network(s) studied. Describe efforts taken to address its presence when found.	***A3.6*** ***A5.2 and A5.3 – discussion for each network***
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	Evidence review linked in A2.1.2
Additional analyses	16	Describe methods of additional analyses if done, indicating which were pre-specified. This may include, but not be limited to, the following: • Sensitivity or subgroup analyses; • Meta-regression analyses; • *Alternative formulations of the treatment network; and* • *Use of alternative prior distributions for Bayesian analyses (if applicable).*	***A2.2.2, A2.2.3***

RESULTS†			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	Evidence review linked in A2.1.2
Presentation of network structure	**S3**	Provide a network graph of the included studies to enable visualization of the geometry of the treatment network.	***A4.2***
Summary of network geometry	**S4**	Provide a brief overview of characteristics of the treatment network. This may include commentary on the abundance of trials and randomized patients for the different interventions and pairwise comparisons in the network, gaps of evidence in the treatment network, and potential biases reflected by the network structure.	***A4.2***
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	A4.1
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment.	Evidence review linked in A2.1.2
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: 1) simple summary data for each intervention group, and 2) effect estimates and confidence intervals. *Modified approaches may be needed to deal with information from larger networks.*	***A4.3***
Synthesis of results	21	Present results of each meta-analysis done, including confidence/credible intervals. *In larger networks, authors may focus on comparisons versus a particular comparator (e.g. placebo or standard care), with full findings presented in an appendix. League tables and forest plots may be considered to summarize pairwise comparisons.* If additional summary measures were explored (such as treatment rankings), these should also be presented.	***A5***
Exploration for inconsistency	**S5**	Describe results from investigations of inconsistency. This may include such information as measures of model fit to compare consistency and inconsistency models, *P* values from statistical tests, or summary of inconsistency estimates from different parts of the treatment network.	***A5.2 and A5.3 – discussion for each network***
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies for the evidence base being studied.	Evidence review linked in A2.1.2
Results of additional analyses	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression analyses*, alternative network geometries studied, alternative choice of prior*	***Appendices D, E and F*** ***Alternative networks discussed throughout***

		distributions for Bayesian analyses, and so forth).	
DISCUSSION			***A6***
Summary of evidence	24	Summarize the main findings, including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy-makers).	A6.1, A6.2, A6.3
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review level (e.g., incomplete retrieval of identified research, reporting bias). *Comment on the validity of the assumptions, such as transitivity and consistency. Comment on any concerns regarding network geometry (e.g., avoidance of certain comparisons).*	A6.4
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	A7
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review. This should also include information regarding whether funding has been received from manufacturers of treatments in the network and/or whether some of the authors are content experts with professional conflicts of interest that could affect use of treatments in the network.	***This research was conducted as part of a guideline update commissioned by the National Institute for health and Care Excellence (NICE).***

PICOS = population, intervention, comparators, outcomes, study design.

* Text in italics indicates wording specific to reporting of network meta-analyses that has been added to guidance from the PRISMA statement.