

A Network Meta-Analysis of Treatments for Locally Advanced/Metastatic Pancreatic Cancer

Matthew Connor Knowles



The
University
Of
Sheffield.

School of Health and Related Research,
The University of Sheffield,
United Kingdom

Supervisor: Dr. Kate Ren
August 2024
Word Count: 14780
60 Credits

*A dissertation submitted in partial fulfillment of the requirements for the degree of Master of Science in
Statistics with Medical Applications*

Abstract

Pancreatic cancer is a particularly nasty cancer associated with exceptionally poor long term survival. Several Network Meta-Analyses (NMAs) have been conducted to analyse the treatment landscape, but none of these NMAs used the Multilevel Network Meta-Regression (ML-NMR) method. ML-NMR is a relatively new method for performing an NMA by combining individual patient-level data (IPD) and aggregate data.

An ML-NMR was performed on data from eight studies comparing gemcitabine (GEM) with six combination therapies and one standalone therapy. The comparative efficacy of the combination therapies was assessed in terms of the median overall survival (OS) and restricted mean survival time (RMST) estimates. FOLFIRINOX provides improved median OS and RMST estimates for the treatment of locally advanced/metastatic pancreatic cancer compared to GEM monotherapy. GEM in combination with nab-paclitaxel (GEM-NAB) also provides improved median OS and RMST compared to GEM and other treatments.

A brief literature review was conducted to validate the results of this ML-NMR. An ML-NMR has not previously been conducted for treatments of pancreatic cancer, but the results of this dissertation aligned with the results of the other meta-analyses found in the literature review. This dissertation showed the suitability of the ML-NMR method for use in pancreatic cancer, but lacked real IPD, meaning extensive adjustment of covariates was not possible. In addition, some conclusions about the ML-NMR method itself were able to be drawn.

Acknowledgements

I would like to thank Dr. Kate Ren for her invaluable supervision and insightful discussions throughout this project. In addition, to my colleagues at OPEN Health who I have learnt so much from, and am continually inspired by. I would also like to thank Dr. David Phillippe, the creator of the `multinma` package, and author of most of the theory on which this dissertation is based. Finally, thank you to my friends and family for putting up with me talking about network meta analysis for the past two years.

Ethical Considerations

This dissertation used anonymised data from clinical trials. No ethical approval was therefore required for this dissertation. The letter confirming this research was able to go ahead without need for ethics approval is available in the appendix.

Computing Environment

All work in this dissertation was conducted on a desktop computer running Ubuntu 20.04.4 on an AMD Ryzen 5 1600 6-core 3.2GHz CPU and NVIDIA GeForce GTX 1060 3GB GPU.

All R code was developed in R version 4.4.1.

For my late grandfather, Norman Fredrick Taylor (July 1943 - April 2015).



"But they that wait upon the Lord shall renew their strength; they shall mount up with wings as eagles; they shall run, and not be weary; and they shall walk, and not faint." Isaiah 40:31

CONTENTS

1 Background	8
1.1 Meta-Analysis and Network Meta-Analysis	8
1.2 Project Aim	8
1.3 Pancreatic Cancer	9
1.4 Treatment Landscape	9
1.4.1 Locally Advanced Pancreatic Cancer	10
1.4.2 Metastatic Pancreatic Cancer	11
1.5 Dissertation Structure	11
2 Survival Analysis Bakground	13
2.1 Survival Functions	13
2.2 The Kaplan-Meier Estimator and Reconcstructing Patient-Level Data	13
2.2.1 The Kaplan-Meier Estimator	13
2.3 Reconstructing Survival Data from Published Curves	14
2.4 Regression Models for Survival	14
2.4.1 Accelerated Failure Time Models	15
2.4.2 Proportional Hazards Models	15
2.5 Key Survival Metrics	15
2.5.1 The Hazard Ratio	15
2.5.2 Median Survival	16
2.5.3 Restricted Mean Survival Time	16
2.6 Parametric Models for Survival Analysis	16
2.6.1 Model Setup	16
2.6.2 Fitting Models	17
2.6.2.1 Right Censoring	17
2.6.2.2 Interval Censoring	17
2.6.3 Model Selection Considerations	17
3 Network Meta-Analysis Theory	20
3.1 Building a Network of Evidence	20
3.2 Standard NMA Model	21
3.3 The General Multilevel Network Meta-Regression	21
3.4 Survival ML-NMR	23
3.5 Quasi Monte-Carlo	24
3.5.1 Background and Low-Discrepancy Sequences	24
3.5.2 The Joint-Covariate Distribution	25
3.6 Population-average estimates	27
3.6.1 Survival Function	27
3.6.2 Hazard Function	27
3.6.3 RMST	27
3.6.4 Median OS	27
3.7 Treatment Ranking	27
3.8 Model Selection and Convergence	28

4 Included Studies	29
4.1 Overview of Studies	29
4.2 Study Eligibility Criteria	30
4.2.1 Eligibility by Study	30
4.2.2 Analysis of Eligibility	32
4.3 Covariates	33
4.4 Parametric Model Fitting to KM Curves	34
5 A Review of Meta-Analyses for Pancreatic Cancer	40
5.1 An Overview of MAs	40
5.2 Implications of the Literature Review	41
6 NMA of Pancreatic Cancer Trials	42
6.1 Base Case	42
6.1.1 Network of Evidence	42
6.1.2 Model Fitting and Selection	42
6.1.3 Results	43
6.2 Sensitivity Analyses	55
6.2.1 Sensitivity Analysis 1	55
6.2.1.1 Model Fitting and Selection	55
6.2.1.2 Results	55
6.2.2 Sensitivity Analysis 2	66
6.2.2.1 Model Fitting and Selection	66
6.2.2.2 Results	66
6.3 Considerations for the ISPOR Good Practice Task Force	68
7 Conclusion and Discussions	79
7.1 Conclusion	79
7.2 Discussion	79
Bibliography	80
8 Appendices	85
8.1 Appendix A: Additional NMA Results	85
8.1.1 Base Case	85
8.1.2 Sensitivity Analysis 1	87
8.2 Sensitivity Analysis 2	88
8.3 Appendix B: ISPOR Good Practice Questions	88
8.3.1 Evidence Base	88
8.3.2 Analysis	89
8.3.3 Reporting Quality and Transparency	90
8.3.4 Interpretation	90
8.3.5 Conflicts of Interest	90
8.4 Appendix C: The PCNMA Package	90
8.4.1 Survival Functions	90
8.4.2 NMA Functions	91

LIST OF FIGURES

1.1 Pancreatic cancer incidence by sex, 1993 to 2018	10
3.1 Visualisation of combining trials into a network of evidence	21
3.2 Random sample of the unit square from a normal distribution	25
3.3 Random sample of the unit square from a Sobol sequence	26
4.1 Forest plot for median OS of GEM in each study	31
4.2 Forest plot for median OS of the comparator in each study	32
4.3 KM curves for each study	33
4.4 Log-cumulative hazard plot for each study	34
4.5 Conroy (2011) parametric model extrapolations	35
4.6 Cunningham (2009) parametric model extrapolations	36
4.7 Goldstein (2015) parametric model extrapolations	37
4.8 Goncalves (2012) parametric model extrapolations	37
4.9 Kindler (2011) parametric model extrapolations	38
4.10 Oettle (2005) parametric model extrapolations	38
4.11 Rocha Lima (2004) parametric model extrapolations	39
4.12 Spano (2008) parametric model extrapolations	39
6.1 Network of evidence	42
6.2 Trace plot for the FE log-logistic model ML-NMR	44
6.3 Trace plot for the RE log-logistic model ML-NMR	44
6.4 Parallel coordinates plot of the FE log-logistic model	44
6.5 Parallel coordinates plot of the RE log-logistic model	44
6.6 Pairs plot for the FE log-logistic model	45
6.7 Pairs plot for the RE log-logistic model	46
6.8 Prior versus posterior distribution for each treatment	47
6.9 OS of each treatment in the Conroy population	48
6.10 OS of each treatment in the Goldstein population	49
6.11 RMST of each treatment in each population	50
6.12 Median OS of each treatment in each population	51
6.13 Relative treatment effects for all treatments versus GEM	52
6.14 Pairwise relative treatment effects for all treatments	53
6.15 Cumulative rank probability for each treatment	54
6.16 Network of evidence	55
6.17 Trace plot for the FE log-normal model ML-NMR	56
6.18 Trace plot for the RE log-normal model ML-NMR	56
6.19 Parallel coordinates plot of the FE log-normal model	57
6.20 Parallel coordinates plot of the RE log-normal model	57
6.21 Pairs plot for the FE log-normal model	57
6.22 Pairs plot for the RE log-normal model	58
6.23 Prior versus posterior distribution for each treatment	59
6.24 OS of each treatment in the Goldstein population	60
6.25 RMST of each treatment in each population	61
6.26 Median OS of each treatment in each population	62

6.27	Relative treatment effects for all treatments versus GEM	63
6.28	Pairwise relative treatment effects for all treatments	64
6.29	Cumulative rank probability for each treatment	65
6.30	Network of evidence	66
6.31	Trace plot for the FE log-normal model ML-NMR	67
6.32	Trace plot for the RE log-normal model ML-NMR	67
6.33	Parallel coordinates plot of the FE log-normal model	68
6.34	Parallel coordinates plot of the RE log-normal model	68
6.35	Pairs plot for the FE log-normal model	69
6.36	Pairs plot for the RE log-normal model	70
6.37	Prior versus posterior distribution for each treatment	71
6.38	OS of each treatment in the Goldstein population	72
6.39	OS of each treatment in the Conroy population	73
6.40	RMST of each treatment in each population	74
6.41	Median OS of each treatment in each population	75
6.42	Relative treatment effects for all treatments versus GEM	76
6.43	Pairwise relative treatment effects for all treatments	77
6.44	Cumulative rank probability for each treatment	78
8.1	OS of each treatment in each population	85
8.2	Rank probabilities for each treatment	86
8.3	OS of each treatment in each population - SA1	87
8.4	OS of each treatment in each population - SA2	88
8.5	Fitted Models as in Flexsurv	104
8.6	Fitted Models with <i>fit_distribution</i>	104

LIST OF TABLES

4.1	Included studies with summary statistics	30
4.2	AIC scores in the Conroy, Cunningham, Goldstein, and Goncalves studies	35
4.3	AIC scores in the Kindler, Oettle, Rocha Lima and Spano studies	36
6.1	Model selection statistics for each model	43
6.2	Model selection statistics for each model	56
6.3	Model selection statistics for each model	67

BACKGROUND

1.1 Meta-Analysis and Network Meta-Analysis

Meta-Analysis (MA) is a statistical technique for synthesising evidence from similar studies to obtain a quantitative overview of the effectiveness of multiple treatments. The first MA was published by [Smith and Glass, 1977], concerning psychotherapy outcomes. The disadvantage with MAs is that it can only make pairwise comparisons of interventions compared in head-to-head (HtH) trials. A Network Meta-Analysis (NMA) extends the idea of standard MAs by allowing a comparison to be made between multiple treatments by combining both direct and indirect evidence. Standard NMAs rely on the assumption that any treatment effect modifying (TEM) covariates are balanced across the populations of the included trials. This assumption can be relaxed if individual patient-level data (IPD) is available for every study in the network, in which case a Network Meta-Regression (NMR) model can be used to adjust for any TEMs.

It is often the case in Health Technology Assessment (HTA) submissions that IPD is only available for a company's own trial, and aggregate data (AgD) is available for the remaining studies. An IPD NMR is therefore not possible in this situation. Multilevel Network Meta-Regression (ML-NMR) relaxes the requirement for IPD to be present for every study by defining an individual-regression model for the IPD and then fitting the AgD by integrating over the covariate distribution to obtain an aggregate likelihood. This enables IPD and AgD to be combined in one network, and accurately adjusting for differing populations. ML-NMR is a relatively new method, first outlined in [Phillippo et al., 2020], and recently shown to extend easily to survival outcomes [Phillippo et al., 2024].

As ML-NMR, particularly for survival outcomes, is a relatively new method for conducting an NMA, there is still some room for investigation on the application of this method.

1.2 Project Aim

As an ML-NMR has not previously been used to conduct an NMA in pancreatic cancer, the primary objective was to assess the comparative efficacy of treatments for locally advanced/metastatic pancreatic cancer using an ML-NMR. The outcome of interest was Overall Survival (OS). The focus of this project was not on a comprehensive assessment of the treatment landscape, but on the application of this method, so inclusion criteria for the included studies were not particularly strict. Studies had to be a phase II or phase III trial, and contain published Kaplan-Meier (KM) curves with numbers at risk. In addition, studies had to report the proportion of male patients on each treatment arm.

The secondary objective was to corroborate the findings of current guidance and published literature (see Chapter 5). The National Institute for Health and Clinical Excellence (NICE) TA476 [National Institute for Health and Care Excellence, 2017b] described some uncertainty in the comparison between GEM-NAB and GEM-CAP, so this NMA also aimed to provide some clarification on the pairwise efficacy of GEM-NAB and GEM-CAP.

The final objective of this dissertation was to provide some more insight on the use of ML-NMR

in the common HTA scenario of having one IPD study and several AgD studies. By performing two sensitivity analyses (SAs), some insight could be gained on the method itself.

1.3 Pancreatic Cancer

Pancreatic cancer is the 10th most common cancer in the UK, accounting for around 3% of cancer cases in the UK [Cancer Research UK, 2024]. Pancreatic cancer is associated with a particularly poor prognosis. In 2017-2019, there were 9,558 deaths from 10,786 cases. 9663 of these cases were in England and Wales, 856 and 267 were in Scotland and Northern Ireland respectively. The percentage of pancreatic cancer cases of all cancer cases ranged from between 2.53% in Scotland to 2.84% in England. The poor prognosis is clear from the long-term survival statistics. Approximately 3% of patients survive pancreatic cancer for more than five years [National Institute for Health and Care Excellence, 2018].

Part of the reason for the poor prognosis is that pancreatic cancer is hard to detect at early stages, meaning most people who present with symptoms already have advanced-stage pancreatic cancer by the time they present. Often, patients only notice symptoms when the tumour has spread to surrounding tissues, or metastasises to other organs [Kelsen et al., 1997]. The liver is the most common site of pancreatic cancer metastases [Deeb et al., 2015]. Common symptoms of pancreatic cancer include indigestion, stomach or back pain, loss of appetite and jaundice [Pancreatic Cancer UK, 2022]. These symptoms are common in other illnesses, which contributes to patients overlooking the fact their symptoms are consistent with pancreatic cancer.

Risk factors of pancreatic cancer include smoking, diabetes, obesity, and high-fat diets. Smoking is the dominant risk factor, with around 20% of cases being caused by cigarette smoking. In addition, cancers from smokers contain more genetic mutations when compared to cancers from non-smokers [Blackford et al., 2009]. An older paper published in 1999 identified that roughly 25% of patients have symptoms compatible with upper abdominal disease up to six months prior to diagnoses. When these patients present with such symptoms, the symptoms were erroneously attributed to problems such as Irritable Bowel Syndrome (IBS) [Dimagno, 1999]. This is of course a problem that contributes to the poor prognosis of pancreatic cancer. The authors suggest that screening of overweight persons with newly-diagnosed diabetes may result in earlier diagnoses of asymptomatic pancreatic cancer. It is likely that at this stage the cancer would be resectable. This is important because the same study suggested that the five-year survival is between 20% – 30% for patients with a resectable tumour that is smaller than 2cm. Of course, this information is now over 20 years old, so some of the statistics may be outdated, but the issue of the late diagnoses is still relevant today.

Around 90% of pancreatic cancer cases are pancreatic adenocarcinomas [Pishvaian and Brody, 2017]. An adenocarcinoma is characterised by excessive growth of epithelial tissue. These cancers often begin in the exocrine component of the pancreas, made up of the head and tail of the pancreas. The exocrine component is primarily responsible for producing digestive enzymes and bicarbonate and carrying them away from the pancreas. Further, these adenocarcinomas begin in the ducts, and are known as pancreatic ductal adenocarcinoma (PDACs) [Neoptolemos et al., 2010].

Figure 1.1 presents the incidence of pancreatic cancer in the UK by sex from 1993 to 2018. The incidence rates were relatively flat until the early 2000s, at which point incidence began to increase. In the 2011-2018, the overall incidence (red line) plateaued and began to drop slightly, primarily driven by a declining incidence of pancreatic cancer in females from around 2013. The incidence has continued to rise in males in this time period. Some of this increase in incidence can be accounted for by advances in diagnostic technologies, such as Positron Emission Tomography (PET) scans and Endoscopic Ultrasound (EUS), which are now becoming more routine and readily available.

1.4 Treatment Landscape

Gemcitabine (GEM) ($C_9H_{11}F_2N_3O_4$) is a standard first-line treatment for pancreatic cancer [National Institute for Health and Care Excellence, 2017a]. GEM is also used to treat other types of cancer, including breast cancer, bladder cancer and non-small-cell lung cancer [Wong et al., 2009].

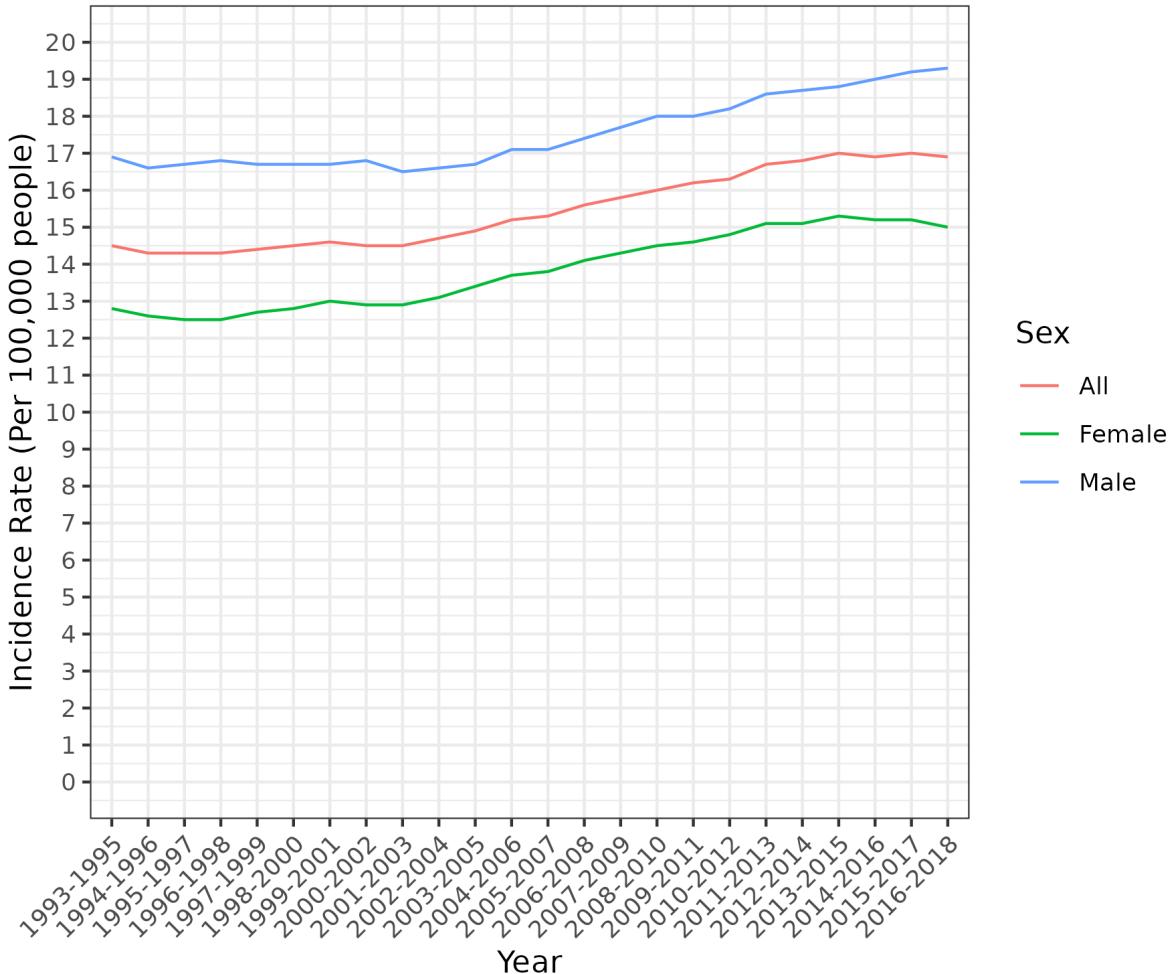


Figure 1.1: Pancreatic cancer incidence by sex, 1993 to 2018

GEM can be administered alone or in combination with another medication. GEM is generally well tolerated by most patients, but can of course lead to fatal complications [Hryciuk et al., 2018].

This dissertation considered six treatments that were given in combination with GEM: capecitabine (CAP) ($C_{15}H_{22}FN_3O_6$), axitinib (AXI) ($C_{22}H_{18}N_4OS$), pemetrexed (PEM) ($C_{20}H_{21}N_5O_6$), sorafenib (SOR) ($C_{21}H_{16}ClF_3N_4O_3$), nab-paclitaxel (NAB)¹ and irinotecan (IRI) ($C_{33}H_{38}N_4O_6$). In addition, one standalone treatment, FOLFIRINOX (FOL), was included. FOL is a combination of oxaliplatin ($C_8H_{14}N_2O_4Pt$), leucovorin ($C_{20}H_{23}N_7O_7$), fluorouracil ($C_4H_3FN_2O_2$), and IRI, which is currently the recommended first-line treatment for metastatic pancreatic cancer in the UK. GEM-NAB is currently recommended as an option for treating previously untreated metastatic adenocarcinoma only if other combination chemotherapies are unsuitable, meaning the patient would normally receive GEM monotherapy.

NICE last updated their guidance on the diagnosis and management of pancreatic cancer in 2018 [National Institute for Health and Care Excellence]. The guidance for treatment is split for patients with locally advanced cancer and metastatic cancer.

1.4.1 Locally Advanced Pancreatic Cancer

Systemic combination chemotherapy is offered to patients who are well enough to tolerate it. For those who are not well enough to tolerate combination therapy, GEM monotherapy is offered. CAP should be considered as the radiosensitiser².

¹ As nab-paclitaxel is a mixture of paclitaxel ($C_{47}H_{51}NO_{14}$) with albumin protein, it does not have a standard chemical formula.

² A radiosensitiser is a treatment that makes cancer cells more susceptible to radiotherapy.

1.4.2 Metastatic Pancreatic Cancer

In first-line treatment, if a patient has an Eastern Cooperative Oncology Group (ECOG) performance status of 0 to 1 and is well enough, they are offered FOL. If the patient is not well enough to tolerate FOL, then GEM combination therapy is offered. Further, if the patient is not well enough to tolerate GEM combination therapy, then GEM monotherapy is offered.

In second-line treatment, oxaliplatin-based chemotherapy is offered to patients who did not receive first-line oxaliplatin. GEM combination therapy is offered in the second-line treatment case for patients who progressed after first-line FOL.

1.5 Dissertation Structure

Chapter 2 outlines some concepts in survival analysis that are important to this dissertation. The survival and hazard functions are introduced along with key metrics used for comparing the survival of different treatments. In particular, the median survival time and Restricted Mean Survival Time (RMST) are introduced. The KM estimator is derived using Maximum Likelihood Estimation (MLE), and an algorithm for reconstructing patient-level data from KM curves is presented. There is some discussion on parametric survival models, which the later chapters rely on. The process by which these models are fit is outlined along with the relevant R packages for doing so. Finally, the selection criteria for survival models are explored.

Chapter 3 discusses the theory of NMAs. First, the idea of a network of evidence is outlined using graph theory. The standard NMA model is then discussed, looking at the relative effect of treatments in a network of evidence. The notion of different types of effects NMAs are introduced, and some discussion on types of variables considered in NMAs is presented. This is then expanded on to derive the ML-NMR model for general likelihoods, which is in turn expanded on to describe the ML-NMR model for survival outcomes. There is some discussion of the numerical integration techniques used when implementing the survival ML-NMR model. Following this, the outcome measures are introduced in the ML-NMR context, building on the initial definitions in Chapter 2. Finally, the model selection criteria and metrics are introduced.

Chapter 4 presents the clinical trials that were included in the NMA. A summary table presents the treatments, number of patients, and summary covariate information in each trial. There is some discussion on the similarity of the studies based on the data in the table. The discussion of similarity is aided by two forest plots presenting the variation of the median OS in each study. The KM curves for each study are also presented. Finally, the extrapolation plots of the parametric models fit to each treatment arm of each study are presented. The models are discussed to inform which survival models would be used in the NMA.

Chapter 5 presents the (relevant) findings of several NMAs in pancreatic cancer. This is primarily used for validation of the results of this NMA. Some discussion is given to the findings of the literature, and where this NMA stands within that wider context.

Chapter 6 presents the NMA itself. For the base case and two SAs, the network of evidence is presented along with some discussion. The model fitting and selection is then presented, with discussion on which model provided the best fit and was therefore selected. A table presenting the information criterion scores for each model is used to compare each model. The results of the ML-NMR are then presented. Simulated KM curves of each treatment in two study populations are presented. Further to this, forest plots of RMST and median OS for each treatment in each population are presented. Pairwise comparisons of each treatment are used to assess treatments with significant differences. Finally, cumulative rank probabilities are presented to determine the ranking of treatments. In addition, there is some discussion on how this NMA aligns with the good practice framework laid out by The Professional Society for Health Economics and Outcomes Research (ISPOR).

Chapter 7 discusses the results of Chapter 6 and the interpretation thereof. The aims presented in Section 1.2 are addressed before some more general discussion and comments on potential future work.

Appendix 8.1 presents some additional results from the NMA that were not required for drawing any conclusions, but may be of value to the interested reader. Appendix 8.3 details the individual ISPOR good practice questions, and the answers to them in the context of this NMA. Appendix 8.4 outlines the development of the PCNMA R package that was used for conducting this NMA. While not relevant to understanding any of the results, this appendix may be interesting to the more programming-minded reader.

SURVIVAL ANALYSIS BAKGROUND

2.1 Survival Functions

Given a population of individuals, assume the time of death for each individual is drawn from a continuous random variable $T > 0$ with probability density function $f(t)$ and distribution function $F(t) = \int_0^t f(\tau)d\tau$. Survival analysis is concerned with estimating the distribution T from Time-To-Event (TTE) data. There are two functions central to survival analysis, the *survival function* and *hazard function*.

Definition 2.1.1: Survival Function

The **Survival function** $S(t)$, gives the probability of an individual surviving longer than time t .

$$S(t) = P(T \geq t) = 1 - F(t) = \int_t^\infty f(\tau)d\tau \quad (2.1)$$

Definition 2.1.2: Hazard Function

The **Hazard function** gives the risk of death at time t , given that the individual has survived up to time t .

$$h(t) = -\frac{d}{dt} \log S(t) \quad (2.2)$$

2.2 The Kaplan-Meier Estimator and Reconcstructing Patient-Level Data

2.2.1 The Kaplan-Meier Estimator

The KM estimator is a non-parametric method for estimating the survival function from event and censoring times. TTE data contains, at a bare minimum, a subject identifier, an event time, and a censoring indicator. The censoring indicator is given in Equation 2.3. The time column of data gives the time that either the event (i.e. death due to the disease being investigated, in overall survival), or censoring (i.e a chemotherapy patient dies due to an adverse event unrelated to their cancer) occurs.

$$c_i = \begin{cases} 1 & \text{If individual } i \text{ has an event} \\ 0 & \text{If individual } i \text{ is censored} \end{cases} \quad (2.3)$$

Let d_i and n_i be the number of events and total individuals at risk at the i^{th} timepoint, t_i . Define the discrete hazard rate h_i as the probability that individual experiences an event at time t_i . The survival rate is then defined as in Equation 2.4, and the likelihood function for the hazard function up to time t_i

is given by Equation 2.5.

$$S(t) = \prod_{i:t_i \leq t} (1 - h_i) \quad (2.4)$$

$$\mathcal{L}(h_{j \leq i} | d_{j:j \leq i}, n_{j:j \leq i}) = \prod_{j=1}^i h_j^{d_j} (1 - h_j)^{n_j - d_j} \binom{n_j}{d_j} \quad (2.5)$$

The KM estimator can be derived by MLE of the discrete hazard function. By obtaining an maximum likelihood estimate of h_i , \hat{h}_i , and substituting it into Equation 2.4, the resulting estimate of the survival function, $\hat{S}(t)$, will be the KM estimator. Taking the logarithm of both sides of Equation 2.5 gives Equation 2.6.

$$\log(\mathcal{L}) = \sum_{j=1}^i \left(d_j \log(h_j) + (n_j - d_j) \log(1 - h_j) + \log \binom{n_j}{d_j} \right) \quad (2.6)$$

By taking the derivative of Equation 2.6 with respect to h_i , and setting the resulting fraction equal to zero, \hat{h}_i is obtained as in Equation 2.9.

$$\frac{\partial \log(\mathcal{L})}{\partial h_i} = \frac{\partial}{\partial h_i} \left(\sum_{j=1}^i \left(d_j \log(h_j) + (n_j - d_j) \log(1 - h_j) + \log \binom{n_j}{d_j} \right) \right) \quad (2.7)$$

$$= \frac{d_i}{\hat{h}_i} - \frac{n_i - d_i}{1 - \hat{h}_i} \quad (2.8)$$

$$\Rightarrow \frac{\partial \log(\mathcal{L})}{\partial h_i} = 0 \Rightarrow \hat{h}_i = \frac{d_i}{n_i} \quad (2.9)$$

Definition 2.2.1: Kaplan-Meier Estimator

The **Kaplan-Meier Estimator**, $\hat{S}(t)$, estimates the survival probability that an individual survives longer than time t , and is given by Equation 2.10.

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{d_i}{n_i} \right) \quad (2.10)$$

2.3 Reconstructing Survival Data from Published Curves

The ML-NMR method requires aggregate survival data to be in the form of the TTE data previously described. Often, the actual patient level event and censoring times/indicators are not published, and therefore must be reconstructed from the published KM curves, which are usually readily available. Indeed, one of the study selection criteria outlined in Section 1.2 was that studies must publish KM curves with numbers at risk. This was motivated by the method for reconstructing patient-level data from KM curves, the Guyot algorithm [Guyot et al., 2012] requiring numbers at risk.

The digitisation step in reconstructing survival data requires manually obtaining the values of the KM curves as coordinates of an image. These coordinates are then of the form $(time, event)$ or $(time, censoring)$. The Guyot algorithm obtains numerical solutions to the inverted KM equations using these digitised survival times, the numbers at risk, and the number of events.

2.4 Regression Models for Survival

The survival time of patients may be dependent on several explanatory variables such as age, sex, the presence of a genetic mutation, etc. We wish to incorporate these variables into our survival functions. There are two forms of models for survival data: Accelerated Failure Time (AFT) models, and Proportional-Hazards (PH) models).

2.4.1 Accelerated Failure Time Models

Let x be a vector of explanatory variables for each individual in a trial. The survival function can be extended to include this,

$$S(t, x) = S_0(t\Psi(x)).$$

Here, $S_0(t) = S(t, x = 0)$, i.e the survival function at baseline. We define the density and hazard functions accordingly,

$$\begin{aligned} f(t, x) &= f_0(t\Psi(x))\Psi(x) \\ h(t, x) &= h_0(t\Psi(x))\Psi(x). \end{aligned}$$

This is equivalent to defining a random variable T such that

$$T = T_0/\Psi(x).$$

Here, T_0 has survival function S_0 . It is required that $\Psi(x) \geq 0$ and $\Psi(0) = 1$, leading to the natural choice

$$\Psi(x) = \exp(-\beta'x).$$

We can then write

$$\begin{aligned} T &= T_0/\Psi(x) \\ \implies E(T) &= E(T_0)/E(e^{-\beta'x}) \\ &= E(T_0)/e^{-\beta'x} \\ &= E(T_0)e^{\beta'x} \end{aligned}$$

In practice, we assume a distribution for T , and estimate parameters using MLE.

2.4.2 Proportional Hazards Models

Let h_0 represent the hazard function for an individual at baseline. In addition, let x be a vector of explanatory variables. The proportional hazards model, also known as the Cox model [Cox, 1972] is then given by

$$h(t, x) = \exp(\beta'x) h_0(t) \quad (2.11)$$

The Cox model is semi-parametric model (Definition 2.4.1) then β is of finite dimension and $h_0(t)$ is infinite-dimensional and does not need to be specified.

Definition 2.4.1: Semi Parametric Model

A statistical model is a parameterised family of distributions $\{P_\theta : \theta \in \Theta\}$. For a parametric model, $\Theta \subseteq \mathbb{R}^k$ for $k \in \mathbb{N}$. Similarly, for a non-parametric model, $\Theta \subseteq V$, where V is some (possibly infinite) dimensional space V . A **Semi-parametric** model is a statistical model with both parametric and non-parametric components. For a semi-parametric model we have $\Theta \subseteq \mathbb{R}^k \times V$.

2.5 Key Survival Metrics

2.5.1 The Hazard Ratio

The Hazard Ratio (HR), often denoted by φ , follows from Equation 2.11. Consider two treatments, $i = 1, 2$, then $h_1(t, x) = \exp(\beta'x) h_0(t)$ and $h_2(t, x) = \exp(\beta'x) h_0(t)$. The HR is obtained as in Equation 2.12.

$$\varphi = \frac{h_1}{h_2} = \exp(\beta'x) \quad (2.12)$$

In practice, the HR is a useful endpoint in performing NMAs on survival outcomes. However, in order to conduct a HR-based NMA, the proportional hazards assumption (PHA), must be satisfied. The PHA is the assumption that the HR remains constant throughout the observation period of a trial. It can be tested by, for example, visual-inspection of a log-cumulative hazards plot. The log-cumulative hazard plot comes from the cumulative hazard function (Definition 2.5.1), which gives the cumulated risk of experiencing an event. Given the cumulative nature of the cumulative hazard function, and the nature of $S(t)$ being monotonically decreasing, the cumulative hazard function is monotonically increasing.

Definition 2.5.1: Cumulative Hazard Function

The **Cumulative Hazard Function**, $H(t)$ is given by

$$H(t) = \int_0^x h(t)dt = -\log(S(t)) \quad (2.13)$$

By extension, the log-cumulative hazard function is given by $\log(-\log(S(t)))$. When plotting this for both arms of a clinical trial, if the curves remain roughly parallel, the PHA is likely satisfied. If the curves cross and/or diverge, it indicates PHA has not been satisfied. Whether the PHA is satisfied has implications later on in the analysis. For example, if the PHA is not satisfied, then an individual model must be used for each treatment arm. If the PHA is satisfied, a model which does not include a treatment parameter may be appropriate. This fits a single model to one treatment arm, and then uses a linear coefficient used for obtaining estimates of survival for the non-reference treatment. It is clear to see why this would be inappropriate if the PHA is not violated, since the difference between survival curves will not be linear.

2.5.2 Median Survival

Median survival is simply the earliest timepoint after which 50% of patients have died.

2.5.3 Restricted Mean Survival Time

The RMST is an alternative measure to the (log) HR in NMAs. RMST is the mean survival time up to a pre-specified time, τ . This measure can be thought of visually as the area under the survival curve. Definition 2.5.2 presents the formal definition. When no time τ is specified, it is taken such that $S(\tau) = 0$.

Definition 2.5.2: RMST

For a survival function $S(t)$, the **RMST** for some pre-specified time $x > 0$,

$$RMST = \int_0^\tau S(t)dt \quad (2.14)$$

2.6 Parametric Models for Survival Analysis

This section discusses the parametric models commonly used in survival analysis. In particular, the seven parametric models recommended by the NICE in Technical Support Document (TSD) 14 [Latimer, 2011]. These models are the exponential, gamma, generalised gamma, Gompertz, log-logistic, log-normal, and Weibull models. All of these models except the exponential models are AFT models. All parametric model fitting for this project was performed in R using the **flexsurv** package [Jackson, 2016]. Section 2.6.1 outlines how the **flexsurv** package works.

2.6.1 Model Setup

The general model of a **flexsurv** survival model takes the form

$$f(t|\mu(z), \alpha(z)). \quad (2.15)$$

Equation 2.15 gives the probability density for death at time $t \geq 0$. The *mean* or *location* of the distribution is given by $\mu = \alpha_0$. The remaining parameters, $\alpha^1 = (\alpha_1, \dots, \alpha_R)$ are called *ancillary* parameters.

Section 2.4 discussed AFT and PH models. Under the **flexsurv** framework, if the hazard function, $h(t) = \frac{f(t)}{S(t)}$, can be factorised as in Equation 2.16, then the model is a PH model. Alternatively, if the survival function is written as in Equation 2.17, then the model is an AFT model.

$$h(t|\alpha, \mu(z)) = \mu(z)h_0(t|\alpha). \quad (2.16)$$

$$S(t|\mu(z), \alpha) = S_0(\mu(z)t/\alpha). \quad (2.17)$$

All parameters may depend on z , a vector of covariates. This is done through the link-transformed linear models in Equation 2.18 and Equation 2.19.

$$g_0(\mu(z)) = \gamma_0 + \beta_0^T z \quad (2.18)$$

$$g_r(\alpha_r(z)) = \gamma_r + \beta_\gamma^T z \quad (2.19)$$

In Equation 2.18 and Equation 2.19, g is usually chosen to be $\log()$ if the parameter is positive, or the identity function if the parameter is unrestricted.

2.6.2 Fitting Models

Let $t_i, i \in \{1, \dots, n\}$, be a sample of times from n individuals. Define c_i as in Equation 2.3. Introduce s_i , which are delayed-entry times. This means for an individual i who is delayed-entry, the survival time is only observed conditionally on individual i having survived up to time s_i . Therefore, $s_i = 0$ when there is no delayed-entry.

2.6.2.1 Right Censoring

In the case of right-censoring and nothing else, the likelihood for the parameters $\theta = \{\gamma, \beta\}$ required in Equation 2.18-2.19 is given by

$$l(\theta|t, c, s) = \frac{\prod_{i:c_i=1} f_i(t_i) \prod_{i:c_i=0} S_i(t_i)}{\prod_i S_i(s_i)} \quad (2.20)$$

2.6.2.2 Interval Censoring

In the case of interval-censoring, where the survival time is censored on (t_i^{\min}, t_i^{\max}) , the likelihood for $\theta = \{\gamma, \beta\}$ is

$$l(\theta|t^{\min}, t^{\max}, c, s) = \frac{\prod_{i:c_i=1} f_i(t_i) \prod_{i:c_i=0} (S_i(t_i^{\min}) - S_i(t_i^{\max}))}{\prod_i S_i(s_i)} \quad (2.21)$$

MLE is performed in R using the analytic derivatives of Equation 2.20 and/or Equation 2.21 to obtain the required survival models.

2.6.3 Model Selection Considerations

Once a set of models has been fit to the data from a trial, the best model must be selected. NICE Decision Support Unit (DSU) TSD 14 [Latimer, 2011] gives guidance on the model selection process.

The first thing to consider is the visual fit. By overlaying the fitted model(s) over the observed data, it is easy to see how well the two align. Indeed, this is not a particularly robust method of selection, but does provide a quick way to discount certain models.

The second, more robust, method is to examine the Akaike's Information Criterion (AIC) score [Akaike, 1974] (Definition 2.6.1). The values of the AIC scores for each model do not intrinsically reflect the best fitting model, but instead the model with the lowest value is considered to be the best fitting. TSD 14 discusses the use of the Bayesian Information Criterion (BIC) score [Schwarz, 1978] (Definition 2.6.2) in addition

to the AIC score due to the ability to account for the number of model parameters. The model with the lowest AIC (or BIC) score is deemed to be the best fitting model. Further to this, models that score within a score of two of the best fitting model are also appropriate models. The AIC and BIC scores do not give any information about the quality of the models, so further assessment of the bestfitting models is therefore required.

Definition 2.6.1: AIC Score

onsider a statistical model with k estimated parameters and let \hat{L} be the maximised value of the likelihood function of the model. The Akaike's Information Criterion (AIC) score for this model is given as in Equation 2.22.

$$AIC = 2(k - \log(\hat{L})) \quad (2.22)$$

Definition 2.6.2: BIC Score

onsider a statistical model with k estimated parameters and let \hat{L} be the maximised value of the likelihood function of the model. Further, let n be the number of data points in the observed data to which the model was fitted. The Akaike's Information Criterion (AIC) score for this model is given as in Equation 2.23.

$$BIC = k \log(n) - 2 \log(\hat{L}) \quad (2.23)$$

In the context of parametric survival models, the BIC score may be surplus to requirements for two reasons.

1. Five of the seven parametric models recommended by NICE have the same number of parameters, $k = 2$, so no extra information is gained by using the BIC score.
2. The AIC score is preferable for models being used in a predictive context, which is of course the case for survival analysis.

Consider item (1). This can be further reduced based on whether or not the PHA holds. If the PHA has been deemed to be violated by sufficient testing, then the exponential model is not an appropriate model, leaving five models with $k = 2$ and one model, the generalised gamma model with $k = 3$. Let $k_A = 3$ and $k_B = 2$ be such that model A is the generalised gamma model, and model B is one of the other five models. The AIC score difference is given in Equation 2.24-2.26, and the BIC score difference is given in Equation 2.27-2.30, the difference between the difference of the differences reduces to Equation 2.32, which shows that each comparison of the generalised gamma model to another model will be offset by the same amount. Further to this, since five of the models all have $k = 2$, Equations 2.24-2.26 and Equations 2.27-2.30 both reduce to $\log\left(\frac{\hat{L}_B}{\hat{L}_A}\right)$. Since the difference between BIC scores for each $k = 2$ model and the generalised gamma model is offset by a linear amount, the same conclusions would be drawn whether using AIC or BIC scores.

If the PHA holds, then exponential model is of course acceptable, then the reasoning the PHA-violated case holds for comparing $k = 2$ models to the exponential model, and the generalised gamma model to the $k = 2$ models, however, the difference between the generalised gamma model and exponential model will replace $\log(n)$ with $2 \log(n)$ in Equation 2.30 and replace 2 with 4 in Equation 2.26, but the same argument as before follows, since the likelihood is still the main driver of difference. As the number of model parameters increases, this argument breaks down, since the $2(k_A - k_A)$ term in the AIC difference will grow much quicker than $(k_B - k_A) \log(n)$ term, leading to unnecessarily large AIC values. However, when the maximum difference between the k terms is 2, as in the case of parametric survival models, little to no extra information is gained by using the BIC score over the AIC score.

$$\delta_{AIC} = AIC_A - AIC_B \quad (2.24)$$

$$= 6 - 2 \log(\hat{L}_A) - (4 - 2 \log(\hat{L}_B)) \quad (2.25)$$

$$= 2 + 2 \log \left(\frac{\hat{L}_B}{\hat{L}_A} \right) \quad (2.26)$$

$$\delta_{BIC} = BIC_A - BIC_B \quad (2.27)$$

$$= 3 \log(n) - 2 \log(\hat{L}_A) - (2 \log(n) - 2 \log(\hat{L}_B)) \quad (2.28)$$

$$= \log(n) + 2 \log(\hat{L}_B) - 2 \log(\hat{L}_A) \quad (2.29)$$

$$= \log(n) + 2 \log \left(\frac{\hat{L}_B}{\hat{L}_A} \right) \quad (2.30)$$

$$\delta_{IC} = \delta_{BIC} - \delta_{AIC} \quad (2.31)$$

$$= \log(n) - 2 \quad (2.32)$$

Item (2), which is the more important, and slightly less specific to this particular scenario, reason for prioritising the AIC score comes from the fact that survival models are inherently predictive. The purpose of survival models is to assess how the survival of a population will evolve over an unseen period without having to keep an RCT running for an indefinite amount of time until all patients have left the study. The AIC method is *minimaxrate optimal* [Ding et al., 2018], meaning the the maximum loss of information reduces at an optimal rate as the sample size increases. These definitions are information theoretic, and beyond the scope of this dissertation, but it does mean that the AIC model is more appropriate for predictive models.

For these reasons, only the AIC score was used for survival model selection in this dissertation. The BIC score would be used if the AIC scores gave inconclusive estimates.

NETWORK META-ANALYSIS THEORY

3.1 Building a Network of Evidence

Consider a set of J two-arm randomised-controlled trials (RCTs). In each trial $j \in 1, \dots, J$, the patients are randomised to receive a treatment A_j , or a placebo P_j . This can be represented as J graphs with two nodes, A_j and P_j , connected by an edge representing the trial comparing A_j and P_j . It is useful at this stage to recall the formal definition of an (undirected) graph, as presented in Definition 3.1.1.

Definition 3.1.1: Graph

A **Graph** is an ordered triple $G = (V, E, \varphi)$. Where V is a set of nodes, E is a set of edges, and $\varphi : E \rightarrow \{\{x, y\} | x, y \in V \text{ such that } x \neq y\}$ is an **incidence function** mapping every edge to a pair of vertices.

We can construct J graphs under the formal definition. Namely, for trial T_i , we have $G_i = (V_i, E_i, \varphi_i)$ where $V_i = \{A_i, P_i\}$, $E_i = \{T_i\}$ and $\varphi_i : E_i \rightarrow \{\{x, y\} | x, y \in V_i \text{ such that } x \neq y\}$. For construction of the graphs, we drop the subscript on P_i , and take placebo (or any common treatment, not necessarily placebo) as a reference treatment. This is done under the assumption that the effect of placebo is constant across all trials. This is a strong assumption, and implications of this are discussed later. Under this assumption however, each V_i now contains a common element, P .

Let

$$V_{trts} = \bigcup_{j=1}^J V_j \quad (3.1)$$

$$E_{trials} = \bigcup_{j=1}^J E_j. \quad (3.2)$$

The incidence function is given in Equation 3.3.

$$\varphi : E_{trials} \rightarrow \{\{x, y\} | x, y \in V_{trts} \text{ such that } x = P\} \quad (3.3)$$

Then the ordered triple $G = (V_{trts}, E_{trials}, \varphi)$ is the network of evidence given by these two arm trials that forms the basis of an NMA. This process expands to trials that compare more than two treatments by weighting the edges by the number of trials making that particular comparison. Figure 3.1 illustrates combining treatments into a network of evidence.

In some scenarios, a HtH trial may be included in an NMA. Including a HtH trial involves special consideration, since these trials do not compare a treatment versus a placebo, but against another active treatment. From a graph-theoretic perspective, this can create a loop. A loop is an edge that connects a vertex to itself, and makes a graph non-simple.

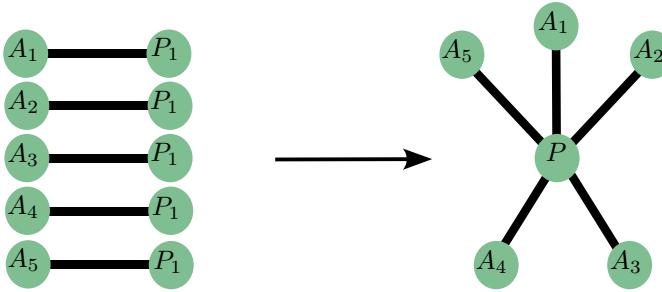


Figure 3.1: Visualisation of combining trials into a network of evidence

3.2 Standard NMA Model

Let d_{ab} denote the relative effect of treatment b versus treatment a . Suppose we have summary outcomes y_{jk} of treatment k in study j . This summary outcome may be, for example, HRs, or RMST values. The standard NMA model is written as in Equation 3.4-3.5. In Equation 3.4, π_{Agg} is a suitable likelihood for the aggregate data, and θ_{jk} represents the expected summary outcome of treatment k in study j . The link function g serves to transform θ_{jk} onto the linear predictor scale. In Equation 3.5, μ_j and δ_{jk} are study-specific intercepts and study-specific relative effect of treatment k versus the reference treatment.

$$y_{jk} \sim \pi_{Agg}(\theta_{jk}) \quad (3.4)$$

$$g(\theta_{jk}) = \mu_j + \delta_{jk} \quad (3.5)$$

There are two types of NMAs: fixed effect (FE) and relative effects (RE) NMAs. In an FE NMA, $\delta_{jk} = d_k$, with $d_1 = 0$. In an RE NMA, $\delta_{jk} \sim N(d_k, \tau^2)$ for the heterogeneity variance τ^2 , with $\delta_{j1} = d_1 = 0$.

As previously mentioned the standard NMA model assumes that any TEMs, i.e covariates that alter the relative effect on a given scale of an active treatment versus control, are balanced across populations. While this can often be a valid assumption, methods such as Matching-Adjusted-Indirect-Comparisons (MAICs), Simulated Treatment Comparisons (STCs), and ML-NMR have sought to relax this assumption by using IPD from at least one of the studies in a population. ML-NMR is a more generalisable method, as MAICs and STCs are only able to perform pairwise comparisons. ML-NMR is also able to estimate treatment effects in a target population for which there was no direct evidence.

3.3 The General Multilevel Network Meta-Regression

The derivation in this chapter outlines the work of [Phillippo et al., 2024]. Under an NMA framework, there are J RCTs investigating a subset $K_j \subset K$ ($j = 1, \dots, J$) treatments. In this project, $|K_j| = 2 \forall j$. Ideally, all IPD from the RCTs in a network would be available, however this is seldom the case. Certainly in a survival analysis context, often published KM curves accompanied by aggregate data are the only available data. From the perspective of a pharmaceutical company working on a Health Technology Assessment (HTA) submission, the likely scenario is that they have data from their own RCT comparing their drug with a key comparator, and then published KM curves and covariate summaries. Consider first the standard IPD network meta-regression (NMR), as given in Definition 3.3.1.

Definition 3.3.1: General IPD Meta-Regression Model

Let y_{ijk} be the IPD outcome for individual $i = 1, \dots, N_{kj}$ in study j receiving treatment $k \in K_j$ given the likelihood distribution $\pi_{Ind}(\theta_{ijk})$, where N_{kj} is the number of patients on treatment k in study j .

$$\begin{aligned} y_{ijk} &\sim \pi_{Ind}(\theta_{ijk}) \\ g(\theta_{ijk}) &= \mu_j + x_{ijk}^T(\beta_1 + \beta_{2,k}) + \gamma_k \\ &= \eta_{jk}(x_{ijk}) \end{aligned}$$

Here, g links the likelihood parameter θ_{ijk} to $\eta_{jk}(x_{ijk})$. The μ_j are study-specific intercepts, and $\beta_1, \beta_{2,k}$ are regression coefficients for prognostic and effect-modifying covariates respectively. Additionally, the γ_k are individual-level treatment effects. For the reference treatment, $\beta_{2,1} = \gamma_1 = 0$.

It is clear to see how the model in Definition 3.3.1 extends Equations 3.4-3.5. Indeed, the model in Definition 3.3.1 is based only on IPD. But the aim is to expand the model in order to incorporate aggregate data. By integrating the individual conditional likelihood function over the joint covariate distribution from an aggregate study, we obtain an individual marginal likelihood function. Compared to the conditional likelihood, the marginal likelihood describes the likelihood where outcomes are known, but covariates are not. In the context of survival analysis, the outcomes are the event and censoring times.

Let $\xi = \{\mu_j, \beta_1, \beta_{2,k}, \gamma_k | \forall j, k\}$ be the parameter space. Denote the individual conditional likelihood function by $L_{ijk|x}^{\text{Con}}(\xi; y_{ijk}, x_{ijk})$. The form of $L_{ijk|x}^{\text{Con}}$ depends on π_{Ind} , g , and η_{jk} .

Integrating $L_{ijk|x}^{\text{Con}}(\xi; y_{ijk}, x)$ over f_{jk} , we obtain the desired individual marginal likelihood function, as given in Equation 3.6.

$$L_{ijk}^{\text{Mar}}(\xi; y_{ijk}) = \int_{\mathfrak{X}} L_{ijk|x}^{\text{Con}}(\xi; y_{ijk}, x) f_{jk}(x) dx \quad (3.6)$$

Here, \mathfrak{X} is the support of x . It is clear from Equation 3.6 that L_{ijk}^{Mar} does not depend on x once the integral on the RHS of Equation 3.6 is evaluated. Let i be an individual on treatment k in study j with outcome y_{ijk} . If we do not know the covariate vector for i , x_{ijk} , but we do know f_{jk} , then we know that the likelihood contribution of i is given by Equation 3.6.

It is likely that a closed-form of Equation 3.6 does not exist. Quasi Monte-Carlo methods can be used to evaluate the integral. We take a set of N integration points, \hat{x} from f_{jk} , and approximate the integral using Monte-Carlo methods (see Section 3.5), giving Equation 3.7.

$$L_{ijk}^{\text{Mar}}(\xi; y_{ijk}) \approx \frac{1}{N} \sum_{\hat{x}} L_{ijk|x}^{\text{Con}}(\xi; y_{ijk}, \hat{x}) \quad (3.7)$$

Consider now a summary outcome y_{jk}^* aggregated over all individuals on treatment k in study j . Each individual i in on treatment k in study j contributes to the aggregate likelihood. Let y_{ijk} denote the observed value of this summary measure for individual i . The aggregate marginal likelihood function is then the product of these y_{ijk} up to a normalising constant, as in Equation 3.8.

$$L_{jk}^{\text{Mar}} \propto \prod_{i=1}^{N_{jk}} L_{ijk}^{\text{Mar}}(\xi; y_{ijk}) \quad (3.8)$$

The full, general ML-NMR model is then given by combining the individual and aggregate level components, as in Definition 3.3.2.

Definition 3.3.2: General ML-NMR Model

Individual:

$$L_{ijk|x}^{\text{Con}}(\xi; y_{ijk}, x_{ijk}) = \pi_{\text{Ind}}(y_{ijk}|\theta_{ijk}) \quad (3.9)$$

$$g(\theta_{ijk}) = \eta_{jk}(x_{ijk}) = \mu_j + x_{ijk}^T(\beta_1 + \beta_{2,k}) + \gamma_k \quad (3.10)$$

Aggregate:

$$L_{ijk}^{\text{Mar}}(\xi; y_{ijk}) = \int_{\mathcal{X}} L_{ijk|x}^{\text{Con}}(\xi; y_{ijk}, x) f_{jk}(x) dx \quad (3.11)$$

$$L_{jk}^{\text{Mar}} \propto \prod_{i=1}^{N_{jk}} L_{ijk}^{\text{Mar}}(\xi; y_{ijk}) \quad (3.12)$$

Under a Bayesian framework, priors are placed on μ_j , β_1 , $\beta_{2,l}$, and γ_k . Uninformative priors were used for this dissertation.

3.4 Survival ML-NMR

The ML-NMR method lends itself well to survival outcomes. Each study in the network reports a pair $y_{ijk} = \{t_{ijk}, c_{ijk}\}$ consisting of outcome times t_{ijk} and censoring indicators c_{ijk} , either from IPD from an RCT or reconstructed IPD. For IPD studies, the covariates x_{ijk} will naturally be available, but for aggregate studies (those for which pseudo-IPD has been re-created), only the joint covariate distribution of covariates at baseline, f_{jk} is available. The idea for deriving the survival ML-NMR model is to include the $y_{ijk} = \{t_{ijk}, c_{ijk}\}$ in Definition 3.3.2.

With some covariates, the pseudo-IPD can include a recreated covariate column for each individual. For example, if the proportion of male patients is given for a particular study, pseudo-patients can be randomly assigned to be male or female such that the number matches that reported in the trial. This is not possible with covariates such as age, where a distribution is usually reported. The proportion of male patients is somewhat unique in this case. It would likely not be statistically sound to do this with a variable such as the site of any metastases. Despite this being a count variable in the same way, the site of metastases may have more influence on survival times than the sex of a patient, and therefore randomly assigning patients to have a given metastases based on the aggregate data may not align with what the original data found.

The censoring indicator for patient i in study j on treatment k is defined as in Equation 2.3. In practice, the censoring indicator can be defined as 0 meaning the individual is censored, or 1 denotes censoring, as in Equation 2.3. For this project, as all the data was digitised, the censoring indicator was already consistent across the data in the network, but when using an IPD study, one must take care to ensure the correct definition of censoring is consistent across all the studies.

Let $S_{jk}(t|x)$ and $h_{jk}(t|x)$ be the survival and hazard functions at time t conditional on the covariates x . Then the individual conditional likelihood contributions for each time t_{ijk} in the IPD studies are given by Equation 3.13.

$$L_{ijk|x}^{\text{Con}}(\zeta; t_{ijk}, c_{ijk}, x_{ijk}) = S_{jk}(t_{ijk}|x_{ijk}) h_{jk}(t_{ijk}|x_{ijk})^{c_{ijk}} \quad (3.13)$$

The forms of S and h depend on the specific survival models chosen. In this dissertation, only parametric models were considered for S and h . Starting from Equation 3.6, the marginal likelihood equations for each event/censoring time in the aggregate data studies can be derived. Substituting $y_{ijk} = \{t_{ijk}, c_{ijk}\}$ into $L_{ijk|x}^{\text{Con}}(\xi; y_{ijk}, x)$ results in Equation 3.16.

$$L_{ijk}^{\text{Mar}}(\xi; y_{ijk}) = \int_{\mathfrak{X}} L_{ijk|x}^{\text{Con}}(\xi; y_{ijk}, x) f_{jk}(x) dx \quad (3.14)$$

$$L_{ijk}^{\text{Mar}}(\xi; t_{ijk}, c_{ijk}) = \int_{\mathfrak{X}} L_{ijk|x}^{\text{Con}}(\xi; t_{ijk}, c_{ijk}, x) f_{jk}(x) dx \quad (3.15)$$

$$= \int_{\mathfrak{X}} S_{jk}(t_{ijk}|x_{ijk}) h_{jk}(t_{ijk}|x_{ijk})^{c_{ijk}} f_{jk}(x) dx \quad (3.16)$$

As with Equation 3.7, Equation 3.16 can be evaluated with quasi-Monte Carlo integration to obtain Equation 3.17.

$$L_{ijk}^{\text{Mar}}(\xi; t_{ijk}, c_{ijk}) = \frac{1}{N} \sum_{\hat{x}} S_{jk}(t_{ijk}|\hat{x}) h_{jk}(t_{ijk}|\hat{x})^{c_{ijk}} \quad (3.17)$$

3.5 Quasi Monte-Carlo

3.5.1 Background and Low-Discrepancy Sequences

The full Quasi Monte-Carlo (QMC) procedure is beyond the scope of this dissertation, and is discussed at length in [Phillippo, 2019], but this section serves as an overview of QMC. QMC is a method for efficient numerical integration. Equation 3.18 gives the form of a Monte-Carlo integration problem for a real integrable function f over the s -dimensional hypercube $I^s = [0, 1]^s$. QMC differs from standard Monte-Carlo (MC) integration in the way the x_i are chosen for which f is evaluated at.

$$\int_{[0,1]^s} f(u) du \approx \frac{1}{N} \sum_{i=1}^N f(x_i) \quad (3.18)$$

Regular MC integration uses random numbers¹, whereas QMC uses a low-discrepancy sequence, which is a sequence such that for all N , the set $P = \{x_1, \dots, x_N\}$ has low discrepancy, as defined in Definition 3.5.1.

Definition 3.5.1: Low-Discrepancy Sequence

Let $P = \{x_1, \dots, x_N\}$. The **discrepancy** of P , $D_N(P)$ is defined as in Equation 3.19.

$$D_N(P) = \sup_{B \in J} \left| \frac{A(B; P)}{N} - \lambda_s(B) \right| \quad (3.19)$$

Here, $A(B; P)$ is the number of points in P that fall into B . λ_s is the s -dimension Lebesgue measure, and J is the set of s -dimensional intervals as in Equation 3.20, where $0 \leq a_i < b_i \leq 1$.

$$\prod_{i=1}^s [a_i, b_i] = \{x \in \mathbb{R}^s | a_i \leq x_i < b_i\} \quad (3.20)$$

In particular, `multinma` uses a Sobol sequence [Sobol', 1967] to sample x_{jk} from the covariate distribution f_{jk} [Phillippo et al., 2020]. For each covariate in the analysis, the Sobol sequence generates points in $I^s = [0, 1]^s$ with a dimension for each covariate. The Smirnov transform (Definition 3.5.2) is then used to transform the points generated by the Sobol sequence to the required distribution. Figure 3.2 presents the sample points in $[0, 1]^2$ generated from sampling the normal distribution. Figure 3.3 presents the sample points from the Sobol sequence in the same region. The Sobol sequence clearly samples the space in a more efficient way. There are some fairly large gaps with no points in the random plot, but these are not present in the Sobol plot.

¹Strictly speaking, these are pseudo-random numbers when considering the implementation of this method.

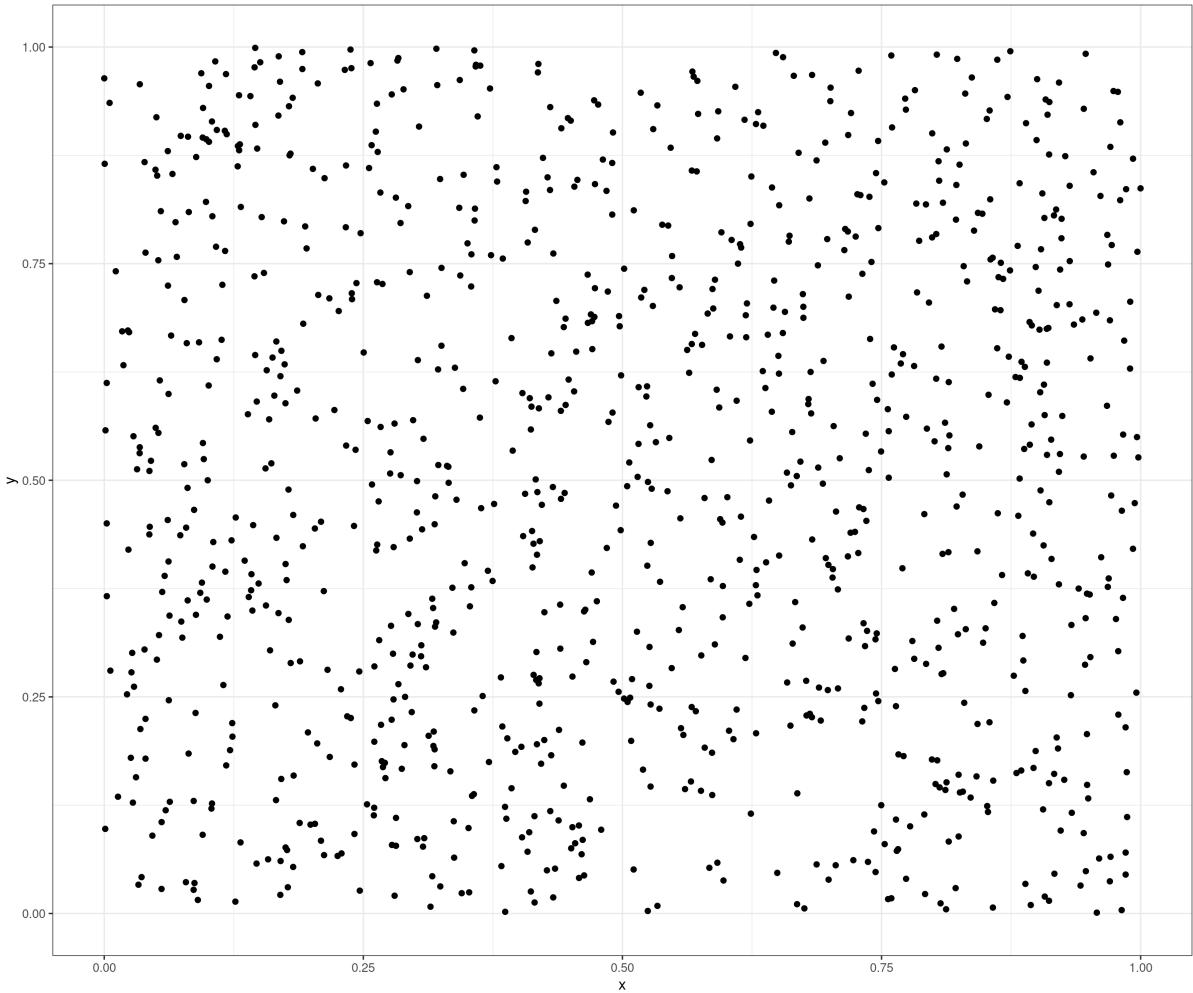


Figure 3.2: Random sample of the unit square from a normal distribution

Definition 3.5.2: Smirnov Transformation

Let $X \in \mathbb{R}$ be any random variable. Then the random variable $F_X^{-1}(U)$ has the same distribution as X . Here, F_X^{-1} is the generalised inverse of the cumulative distribution function F_X of X , as defined in Equation 3.21 for all $p \in [0, 1]$. Further, $U \sim \text{Uniform}[0, 1]$.

$$F^{-1}(p) = \inf\{x \in \mathbb{R} | F(x) \geq p\} \quad (3.21)$$

3.5.2 The Joint-Covariate Distribution

f_{jk} is constructed by the L marginal distributions $f_{jk;l}$ $l \in \{1, \dots, L\}$, where each l is a covariate. By Skalr's Theorem [Sklar, 1959], f_{jk} can be written as in Equation 3.22.

$$f_{jk} = C(f_{jk;1}, \dots, f_{jk,L}) \quad (3.22)$$

Here, C is a *copula*. There are many forms of copula available, but `multinma` uses a Gaussian copula by default. It should be noted that the term “Gaussian” does not refer to the form of the marginal distribution, but of the copula itself.

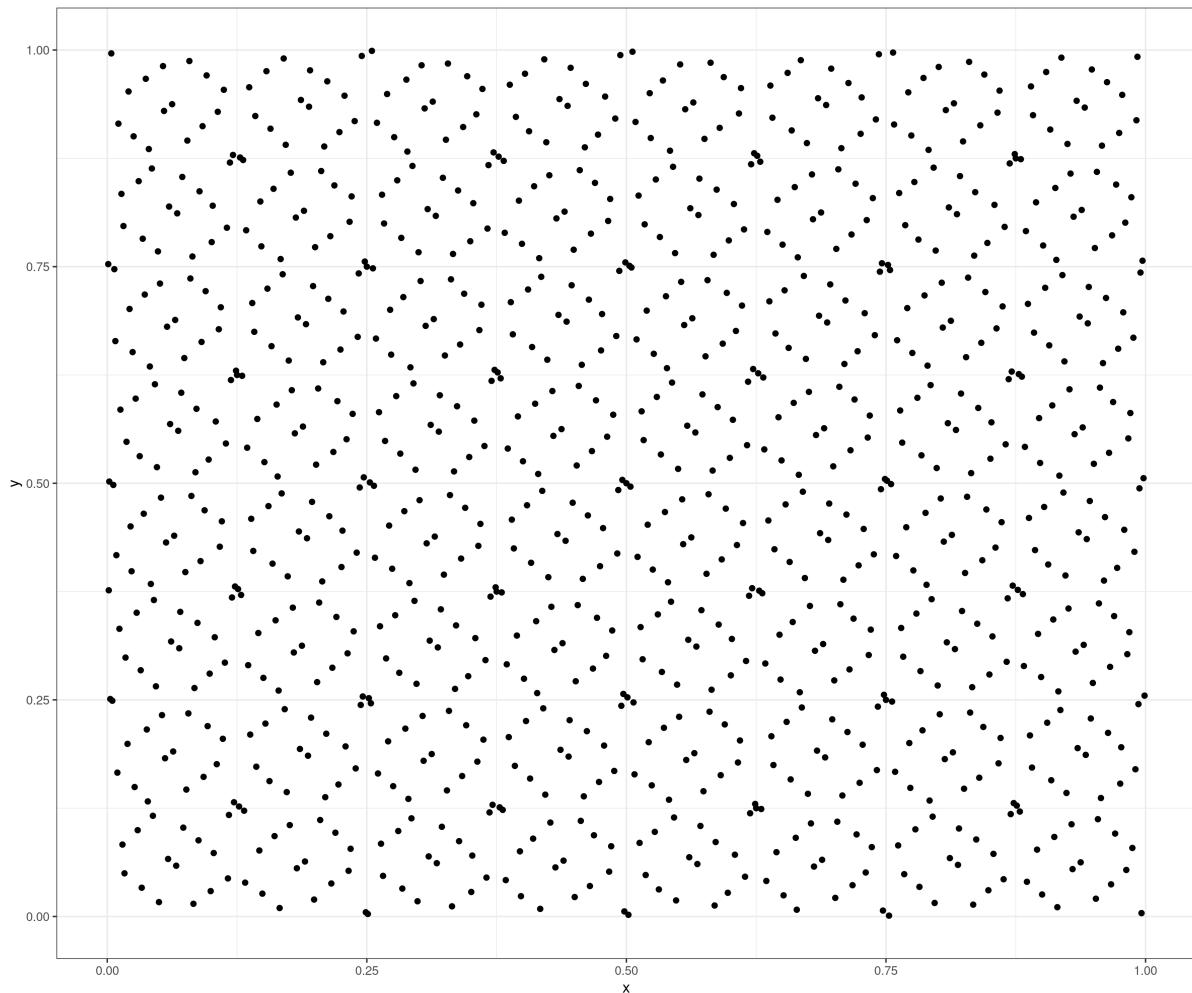


Figure 3.3: Random sample of the unit square from a Sobol sequence

3.6 Population-average estimates

Returning to ML-NMR, recall d_{ab} is the relative effect of treatment b versus treatment a. Let $d_{ab(P)}$ be the population-average relative effect of b versus a in population P. $d_{ab(P)}$ can be calculated as in Equation 3.24.

$$d_{ab(P)} = \int_{\mathfrak{X}} (\eta_{(P)b}(x) - \eta_{(P)a}(x)) f_{(P)}(x) dx \quad (3.23)$$

$$= \gamma_b - \gamma_a + \bar{x}_{(P)}^T (\beta_{2,b} - \beta_{2,a}) \quad (3.24)$$

For this dissertation, The treatment effect measures considered were the RMST and median OS. The survival and hazard functions were also analysed but were not primary outcomes for the NMA.

3.6.1 Survival Function

Let $\bar{S}_{(P)k}(t)$ be the population-average marginal survival probability of treatment k in population P at time t. $\bar{S}_{(P)k}(t)$ is obtained by integrating $S_{(P)k}(t|x)$ over $f_{(P)}(x)$, as in Equation 3.25.

$$\bar{S}_{(P)k}(t) = \int_{\mathfrak{X}} S_{(P)k}(t|x) f_{(P)}(x) dx \quad (3.25)$$

3.6.2 Hazard Function

The population-average marginal hazard function and cumulative hazard function are given by Equation 3.26, and Equation 3.27, respectively.

$$\bar{h}_{(P)k}(t) = \frac{\int_{\mathfrak{X}} S_{(P)k}(t|x) h_{(P)k}(t|x) f_{(P)k}(x)}{\bar{S}_{(P)k}(t)} \quad (3.26)$$

$$\bar{H}_{(P)k}(t) = -\log(\bar{S}_{(P)k}(t)) \quad (3.27)$$

3.6.3 RMST

Let x be some time horizon. The population-average marginal RMST follows from Definition 2.5.2. By default, τ is the time such that $S(\tau) = 0$.

$$RMST_{(P)k}(\tau) = \int_0^\tau \bar{S}_{(P)k}(t) dt. \quad (3.28)$$

3.6.4 Median OS

In general, the $100\alpha\%$ quantile is obtained by solving Equation 3.29 for $\alpha \in [0, 1]$.

$$\bar{S}_{(P)k}(t) = 1 - \alpha. \quad (3.29)$$

Since the median OS is a special case of this with $\alpha = \frac{1}{2}$, the population-average marginal median OS, m is estimated by Equation 3.30.

$$\bar{S}_{(P)k}(m) = \frac{1}{2} \quad (3.30)$$

3.7 Treatment Ranking

The other result considered by this NMA was the treatment ranking. This was done through the Surface Under the Cumulative Ranking Curve (SUCRA) values for each treatment to determine an overall ranking of each treatment. SUCRA values are in the range 0% (worst treatment) to 100% (best treatment) [Mbuagbaw et al.,].

3.8 Model Selection and Convergence

Model selection was based primarily on the Leave-One-Out Information Criterion (LOOIC) score (Defintion 3.8.1). The LOOIC was calculated using the `loo` R package [Vehtari et al., 2024] for each model. The model with the lowest LOOIC score was selected as the best fitting model. The Deviation Information Criterion (DIC) score (Definition 3.8.2) was used a secondary selection criterion. Again, models with lower DIC indicated better fit.

Definition 3.8.1: Leave-One-Out Information Criterion

Begin by defining the expected log pointwise predictive density (elpd) as in Equation 3.31 for data y ,

$$\text{elpd}_{\text{loo}} = \sum_{i=1}^n \log(p(y_i|y_{-i})) \quad (3.31)$$

Here, $p(y_i|y_{-i})$ is the LOO predictive density given the data y with the i^{th} data point removed.

$$p(y_i|y_{-i}) = \int p(y_i|\theta)p(\theta|y_{-i})d\theta \quad (3.32)$$

Define the **Leave One Out Information Criterion (LOOIC)** score as in Equation 3.33.

$$\text{LOOIC} = -2\text{elpd}_{\text{loo}} \quad (3.33)$$

Definition 3.8.2: Deviation Information Criterion (DIC)

Let y be the data on which a model was fitted. Further, let θ and $p(y|\theta)$ be the unknown parameters and likelihood function respectively. Let $E(\theta) = \bar{\theta}$ be the expectation of θ . The deviance is defined as in Equation 3.34.

$$D(\theta) = -2 \log(p(y|\theta)) + C \quad (3.34)$$

Here, C is an unknown constant. C cancels in all model comparisons so is of no importance, but is included here for completeness. Further, the effective number of parameters is given by Equation 3.35.

$$p_D = D(\bar{\theta}) - D(\bar{\theta}) \quad (3.35)$$

The **Deviation Information Criterion (DIC)** is then given by Equation 3.36.

$$\text{DIC} = p_D + D(\bar{\theta}) \quad (3.36)$$

The LOOIC score was preferred over the DIC score as the DIC score is based on a point estimate, and this NMA was performed in a Bayesian framework. As the LOOIC score is using the posterior distribution, it is a fully Bayesian statistic. Traditionally, the DIC score has been used due to the extra computational steps required for the LOOIC score. However, [Vehtari et al., 2017] implemented Pareto-Smoothed Importance Sampling (PSIS) in the R package `loo`, which enables easy, and stable, computation of the LOOIC score from the NMA models.

INCLUDED STUDIES

4.1 Overview of Studies

There was no systematic literature review (SLR) conducted for this study. Studies were searched using Google Scholar to identify Phase II/III studies that reported OS KM curves for GEM and a comparator. In addition, studies had to include the proportion of male patients on each treatment arm. The KM curves were required to present KM curves that could be digitised using WebPlotDigitizer [Rohatgi, 2022]. This meant clear curves with labelling and numbers at risk presented at regular intervals. Studies for both locally advanced and/or metastatic pancreatic cancer were included. Studies could not include GEM-refractory patients.

Table 4.1 presents the studies used in this NMA. In total, there were seven studies comparing GEM with one of six combination therapies. The studies were comparable in terms of median age and proportion male. The [Cunningham et al., 2009] study, compared GEM and GEM-CAP. The [Goldstein et al., 2015] study compared GEM and GEM-NAB. The [Gonçalves et al., 2012] study compared GEM and GEM-SOR. The [Kindler et al., 2011] and [Spano et al., 2008] studies compared GEM and GEM-AXI. The [Oettle et al., 2005] study compared GEM and GEM-PEM. The [Rocha Lima et al., 2004] study compared GEM and GEM-IRI. All studies except Spano were phase III trials, however Spano was included as the OS data was quite mature. The [Conroy et al., 2011] study was the only study to not compare GEM with a combination therapy; comparing GEM and FOL in a phase II-III trial.

Figure 4.1 and Figure 4.2 present forest plots of the median OS of the GEM arm in each study, and the comparator arm in each study, respectively. In particular, the GEM arms in the Conroy, Cunningham, Oettle, Goldstein, and Rocha Lima studies were similar, while the Spano study was noted for having a large 95% CI. This was to be expected given there were only 34 patients in the GEM arm. For the comparator arms, there was more variation in the reported median OS. Large 95% CIs were present for the Spano and Goncalves studies, again due to the comparatively low number of patients in these studies. The efficacy of FOL was clear from Figure 4.2. The median OS of FOL was above the upper bound of 95% credible interval of all comparators.

Figure 4.3 presents the KM curves for each treatment arm in each study. It was clear that the PHA would need to be relaxed when fitting NMA models from the shape of the curves in Figure 4.3, due to the amount of crossing. Most of the studies had mature data, however the Kindler and Goncalves studies were noted for only dropping to an OS of around 0.25 at the end of the observation period. Conroy, and Goldstein were the only studies that showed a clearly higher OS for the comparator compared to GEM. The curves in the Cunningham, Goncalves, and Spano studies showed slight improvement in the comparator compared to GEM. The curves in the Kindler, Oettle, and Rocha Lima studies showed little to no benefit between the comparator and GEM curves.

To assess the PHA more formally, the log-cumulative hazard plot was inspected. If the curves crossed, this indicated violation of the PHA. Figure 4.4 presents the log-cumulative hazard function for each study. All of the studies had crossing log-cumulative hazard curves. In particular, the Kindler and Oettle studies had log-cumulative hazard curves that were almost identical.

Study	Treatment	N	Median Age	Proportion Male	Median OS (Months)
Conroy 2011	GEM	171	61.0 (34, 75)	0.620	6.8 (5.5, 7.6)
Conroy 2011	FOL	171	61.0 (25, 76)	0.614	11.1 (9.0, 13.1)
Cunningham 2009	GEM	266	62.0 (26, 83)	0.580	6.2 (5.5, 7.2)
Cunningham 2009	GEM-CAP	267	62.0 (37, 82)	0.570	7.1 (6.2, 7.8)
Goldstein 2015	GEM	430	63.0 (32, 88)	0.600	6.6 (6.0, 7.2)
Goldstein 2015	GEM-NAB	431	62.0 (27, 86)	0.570	8.7 (7.9, 9.7)
Goncalves 2012	GEM	52	64.0 (40, 82)	0.620	9.2 (7.7, 11.6)
Goncalves 2012	GEM-SOR	52	61.0 (42, 85)	0.580	8.0 (6.0, 10.8)
Kindler 2011	GEM	316	61.0 (35, 89)	0.590	8.3 (6.9, 10.3)
Kindler 2011	GEM-AXI	314	61.0 (34, 84)	0.610	8.5 (6.9, 9.5)
Oettle 2005	GEM	282	63.0 (28, 82)	0.535	6.3 (5.4, 6.9)
Oettle 2005	GEM-PEM	283	63.0 (27, 82)	0.604	6.2 (5.4, 6.9)
Rocha Lima 2004	GEM	180	60.2 (32, 83)	0.533	6.6 (5.2, 7.8)
Rocha Lima 2004	GEM-IRI	180	63.2 (39, 81)	0.572	6.3 (4.7, 7.5)
Spano 2008	GEM	34	61.0 (36, 78)	0.470	5.6 (3.9, 8.8)
Spano 2008	GEM-AXI	69	65.0 (44, 81)	0.510	6.9 (5.3, 10.1)

Table 4.1: Included studies with summary statistics

4.2 Study Eligibility Criteria

4.2.1 Eligibility by Study

Patients in the conroy study had to be aged 18 or over with cytologically confirmed, previously untreated, metastatic PACD. Patients had to have an ECOG score of 0 or 1. Patients were required to have bone marrow such that granulocyte count was $\geq 1500/mm^3$ and platelet count was $\geq 100,000/mm^3$. In addition to bone marrow, the renal and liver function had to be adequate. For the liver function of patients, patients had to have bilirubin ≤ 1.5 times the upper limit of the normal range.

The Conroy study excluded patients who were aged 76 or over. In addition, endocrine or acinar pancreatic carcinoma, previous radiotherapy for measurable lesions, cerebral metastases, a history of another major cancer, active infection, chronic diarrhea, a clinically significant history of cardiac disease, and pregnancy or breast-feeding were all exclusion criteria.

The eligibility criteria in the Cunningham study were similar to Conroy, however patients with locally advanced disease were eligible in Cunningham. Patients could not have had previous chemotherapy or investigative treatment. Adequate bone marrow was required. There were no exclusion criteria.

The Goldstein study had eligibility in terms of the type of cancer closer related to Conroy than Cunningham. Eligible patients had metastatic cancer. In terms of the liver function, the requirement was a bilirubin level below or equivalent to the upper limit of the normal range. In addition, eligible patients had an absolute neutrophil count of $\geq 1.5 \times 10^9/L$ and a hemoglobin level $\geq 9g/dl$. The prior treatment eligibility was more relaxed in the Goldstein study. Treatment with either gemcitabine or fluorouracil as a radion sensitizer was allowed if it had been at least six months before randomisation.

As in Conroy, there were exclusion criteria in the Goldstein study. Previous chemotherapy for metastatic disease was an exclusion criterion. In addition, patients with islet cell neoplasms or locally advanced PACD were also excluded. Patients who received cytotoxic doses of any systemic chemotherapy, including adjuvant GEM were excluded.

The eligibility criteria for the Goncalves study closer aligned with the Cunningham study. Patients with either locally advanced or metastatic cancer were eligible. An ECOG score between 0 and 2 was required, and patients had to be older than 18. In addition, the granulocyte and platelet counts were $> 1.5 \times 10^9/l$ and $> 100 \times 10^9/l$ respectively. Adequate bone marrow, liver, and renal function were required. The total bilirubin was required to be < 1.5 times the upper limit of the normal range.

Exclusion criteria for the Goncalves study were brain metastases, intestinal obstruction, a history of

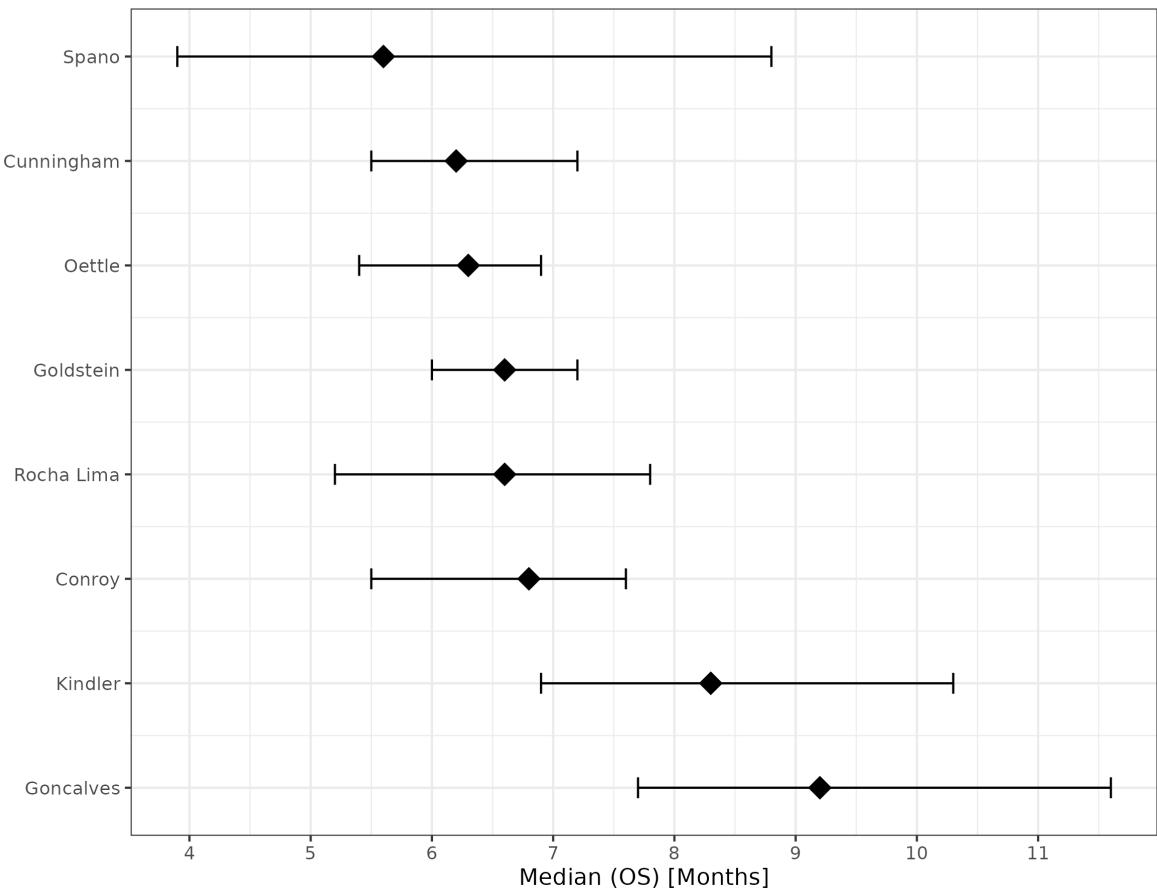


Figure 4.1: Forest plot for median OS of GEM in each study

inflammatory bowel disease or extended small bowel resection. Patients could not have had any major surgery or radiotherapy within 28 days of randomisation.

In the Kindler study, patients had to be at least 18, and could have either locally advanced or metastatic pancreatic cancer. The ECOG score needed to be either 0 or 1. Adequate bone marrow, renal, and liver function was required, but no strict criteria was given. Adjuvant therapy was allowed provided it did not contain GEM, and at least four weeks had passed since the last dose.

Patients were excluded in the Kindler study if they had previous systemic chemotherapy for locally advanced or metastatic disease. In addition recent haemoptysis, myocardial infarction, symptomatic congestive heart failure, cerebrovascular accident, transient ischaemic attack, deepvein thrombosis, or pulmonary embolism in the past 12 months, peptic ulcer disease needing treatment in the past six months, active seizures or gastrointestinal bleeding were all exclusion criteria.

The eligibility criteria for the Oettle study were patients aged at least 18 with either locally advanced or metastatic pancreatic cancer. Prior radiotherapy was allowed if it was completed at least four weeks before entry into the Oettle study. ECOG scores between 0 and 2 were acceptable. Adequate bone marrow, renal, and liver function was required, but no performance criteria given. The Kindler study also required patients to have a life expectancy of at least 12 weeks.

Similar to the Goncalves study, patients with brain metastases were excluded. In addition, patients with significant weight loss were also excluded. This was defined as the loss of > 10% of bodyweight in the previous six weeks. Patients who were unable to interrupt non-steroidal anti-inflammatory drugs for a 5- to 8-day period around PEM administration, or those unable to take folic acid or vitamin B_{12} were also ineligible.

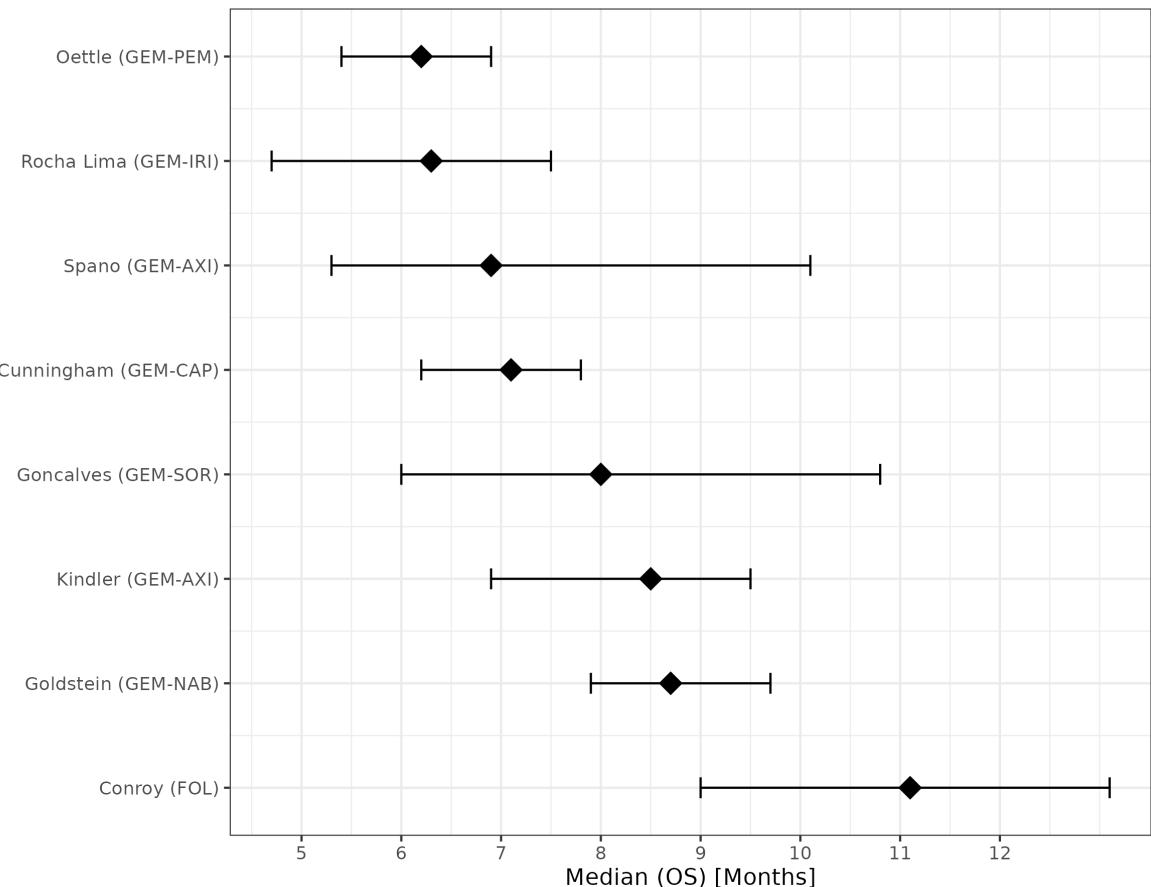


Figure 4.2: Forest plot for median OS of the comparator in each study

The Rocha Lima study allowed for patients aged at least 18 with locally advanced or metastatic pancreatic cancer. The required ECOG score was between 0 and 2. The absolute neutrophil count, platelet count, and bilirubin levels were $\geq 1500/\mu\text{L}$, $\geq 100,00/\mu\text{L}$ and ≤ 1.5 times the upper limit of normal, respectively. These were in line with other studies included.

Patients were excluded from the Rocha Lima study if they had received prior systemic therapy either in an adjuvant setting or for the treatment of advanced pancreatic cancer. Patients could not be pregnant or breastfeeding, have active inflammatory bowel disease, significant bowel obstruction, chronic diarrhea, known brain disease, or myocardial infarction within the previous six months, uncontrollable high blood pressure, unstable angina, congestive heart failure, uncontrolled cardiac arrhythmia, HIV/AIDS or psychiatric illness that prevented the patient giving informed consent.

In the Spano study, patients who were aged 18 or older with locally advanced or metastatic pancreatic cancer were eligible. An ECOG score between 0 and 2 was required. The absolute neutrophil count, platelet count, and hemoglobin levels were $\geq 1500/\mu\text{L}$, $\geq 100,00/\mu\text{L}$ and $\geq 90\text{g/L}$, respectively.

Patients were excluded from Spano if they received prior treatment for metastatic disease or treatment with GEM. Pregnancy or breast feeding, prior cerebrovascular accident, major surgery within the previous four weeks, brain metastases, active second malignancy, uncontrolled intercurrent illness, urine protein of 500mg or more in a 24 hour period, or ongoing uncontrolled hypertension were all exclusion criteria.

4.2.2 Analysis of Eligibility

There was a lot of crossover in the eligibility criteria across the included studies. The two main differences is that the Conroy and Goldstein studies did not include patients with locally advanced disease, and that the Conroy study had an upper age limit for eligible patients. Based on this, it

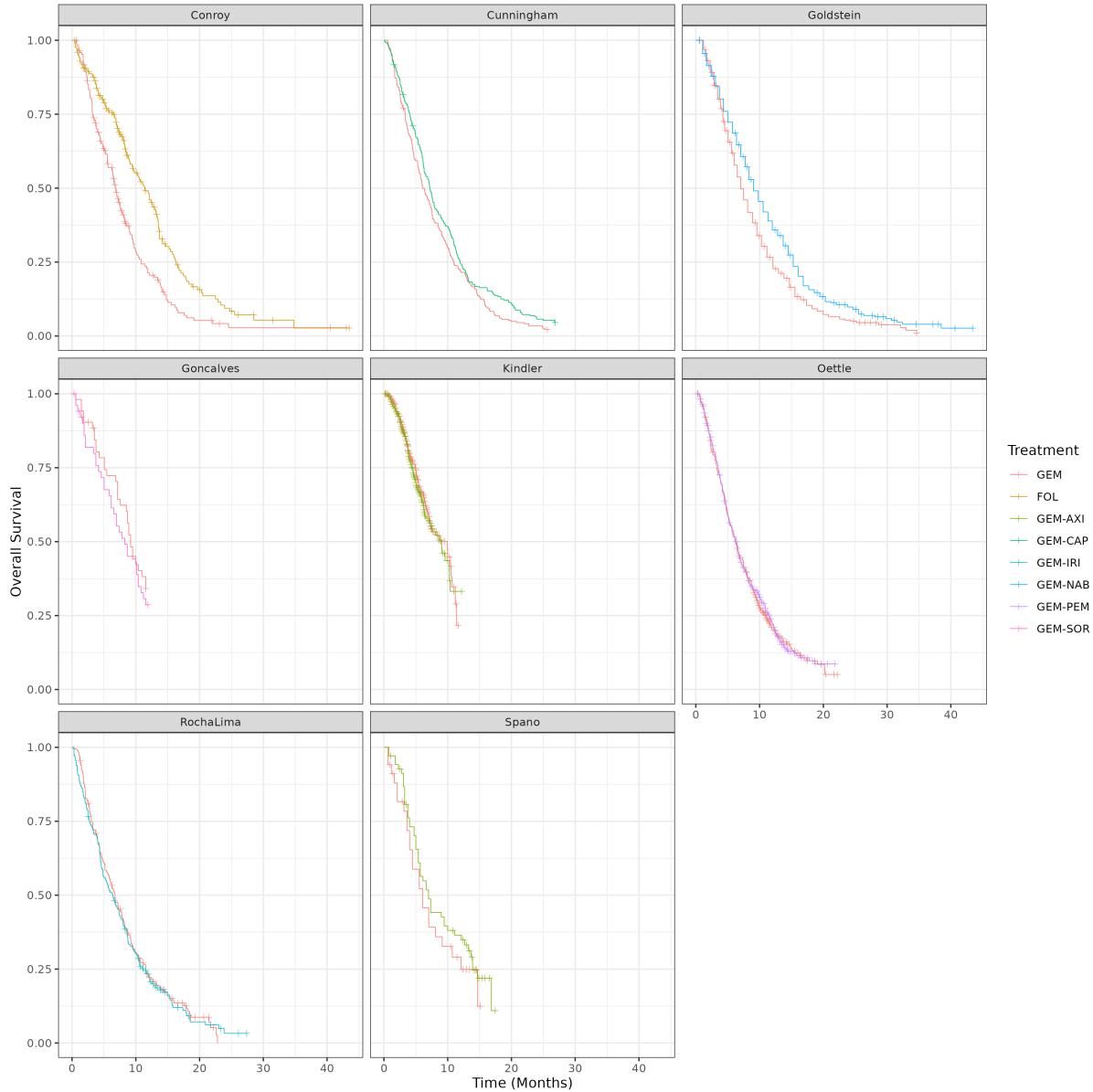


Figure 4.3: KM curves for each study

was deemed to be likely that the populations were homogeneous. Without IPD it is impossible to fully assess homogeneity, but assuming the homogeneity based on the eligibility criteria, and summary statistics presented in Table 4.1 was deemed reasonable.

4.3 Covariates

No IPD was available for any study in the network. Therefore, in order to simulate the common HTA scenario of one IPD study and several AgD studies, the sex of patients in the Goldstein study was simulated based on the reported proportion of male patients in each treatment arm. For the GEM and GEM-NAB treatment arms, 60% and 57% of patients were reported to be male, respectively. These values were used to randomly assign reconstructed patients to be male or female such that the reconstructed data was reflective of that of the original trial. The Goldstein study was chosen as it had very mature data with quite a lot of patients, and was also included in the base case and both SAs.

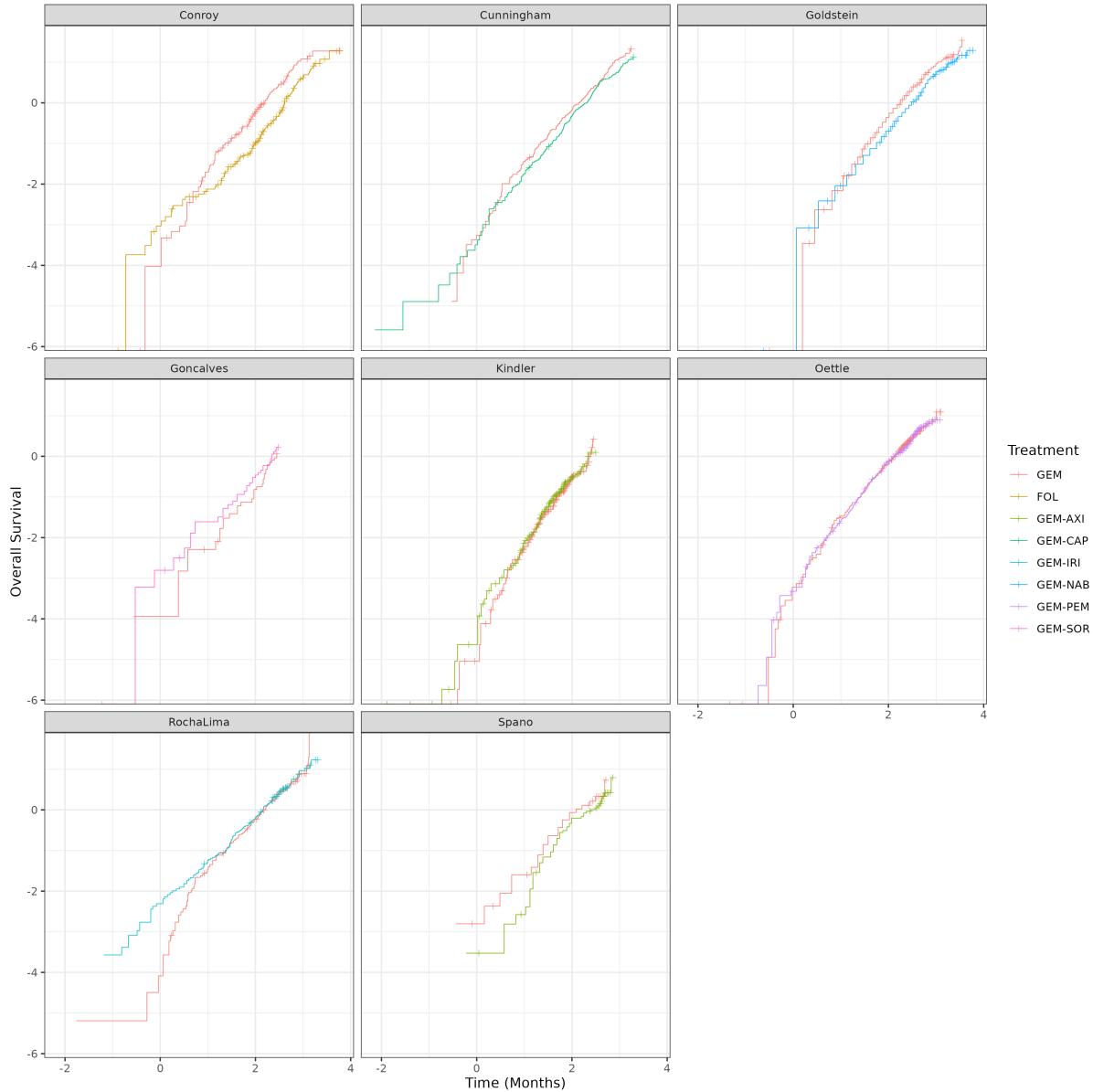


Figure 4.4: Log-cumulative hazard plot for each study

4.4 Parametric Model Fitting to KM Curves

Figure 4.6 to Figure 4.12 present the extrapolation plots for each treatment arm in each study. The data was mature in all studies except the Kindler and Goncalves studies, which meant there was more variation in the survival models for treatments in these populations. The exponential model was noted for presenting poor visual fit in both treatments across both of these studies. For this reason, and due to the NMA not assuming the PHA held, the exponential model was left out of the NMA. This was further supported by the AIC scores as in Table 4.2-4.3, in which the exponential model performs poorly compared to all other models in terms of AIC score in each study.

Table 4.2 and Table 4.3 present the AIC scores of all models in each treatment arm for each study. In the Conroy study, the log-normal model and Weibull model performed best in the GEM and FOL arms respectively. In the Cunningham study, the generalised-gamma and gamma models performed best in the GEM and GEM-CAP arms respectively. In the Goldstein study, the log-normal and generalised gamma model performed best in the GEM and GEM-NAB arms respectively. In the Goncalves study, the Weibull and Gompertz models performed best in the GEM and GEM-SOR arms respectively. In the Kindler and Spano studies, the log-normal models performed best in the GEM and GEM-AXI arms

respectively. In the Oettle study, the log-normal and generalised gamma models performed best in the GEM and GEM-IRI models, respectively. In the Rocha-Lima study, the generalised gamma and gamma model performed best in the GEM and GEM-IRI arm, respectively. The log-normal model performed particularly well in the GEM treatment arms. The log-logistic model scored within a score of two of the log-normal model in every treatment arm except the Oettle GEM arm in Table 4.3, but was further apart in the studies in Table 4.2. The Weibull and generalised gamma models also performed well in a few treatment arms.

Due to the long runtime of the ML-NMR models, it was not deemed feasible to include all the models in the NMA. From Figure 4.5-Figure 4.12, the log-logistic, log-normal, and Weibull models provided consistently reasonable fit to the observed KM curves. These models, in particular the log-normal model, all performed well in terms of AIC score too. The generalised gamma model performed well in terms of AIC in a few studies, but preliminary testing lead to divergent transitions and poor fit when using the generalised gamma model in the NMA. Therefore, the log-logistic, log-normal, and Weibull models were included in the NMA.

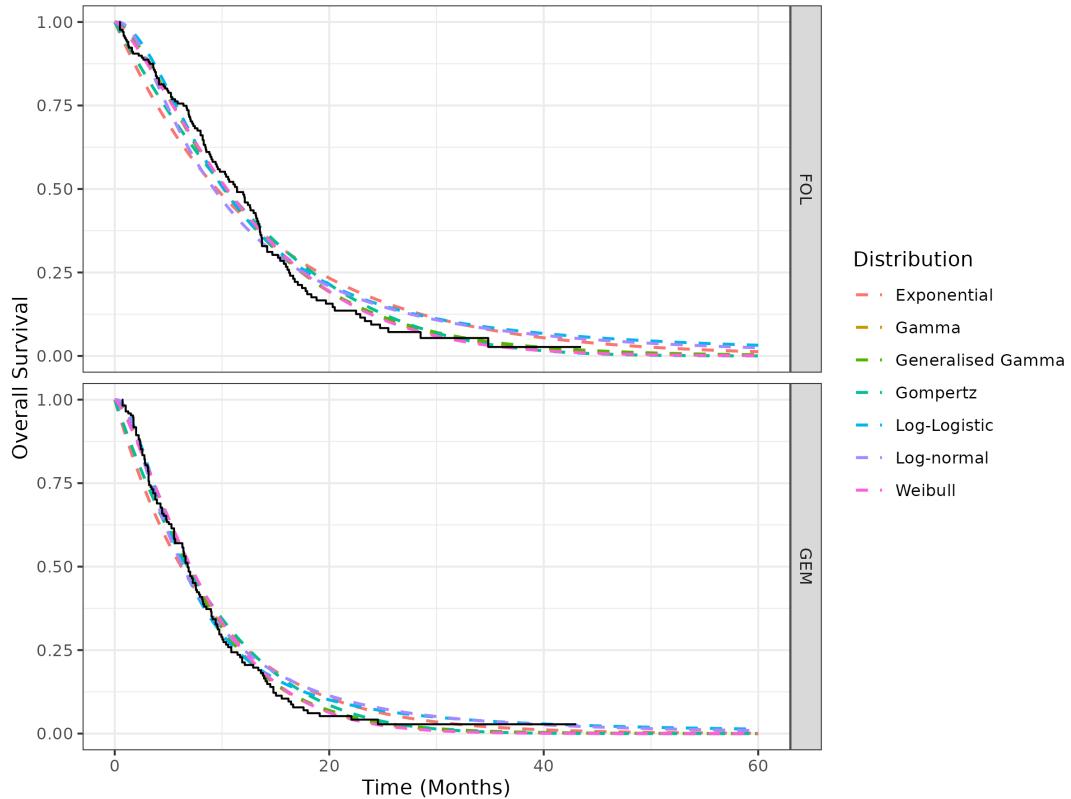


Figure 4.5: Conroy (2011) parametric model extrapolations

Distribution	Conroy		Cunningham		Goldstein		Goncalves	
	GEM	FOL	GEM	GEM-CAP	GEM	GEM-NAB	GEM	GEM-SOR
Exponential	907.972	914.205	1603.030	1644.664	2293.908	2474.915	235.672	238.206
Gamma	886.935	901.431	1571.884	1620.928	2213.465	2426.651	229.716	236.692
Gen. gamma	880.810	900.895	1571.503	1621.730	2200.958	2424.823	229.933	238.145
Gompertz	908.673	903.763	1591.597	1638.689	2274.688	2461.935	228.423	236.361
Log-Logistic	880.525	913.903	1583.669	1625.987	2206.054	2432.977	231.154	238.465
Log-normal	879.209	924.412	1576.771	1637.253	2199.876	2432.134	233.065	239.090
Weibull	893.895	899.301	1576.493	1624.699	2230.390	2434.184	228.917	236.447

Table 4.2: AIC scores in the Conroy, Cunningham, Goldstein, and Goncalves studies

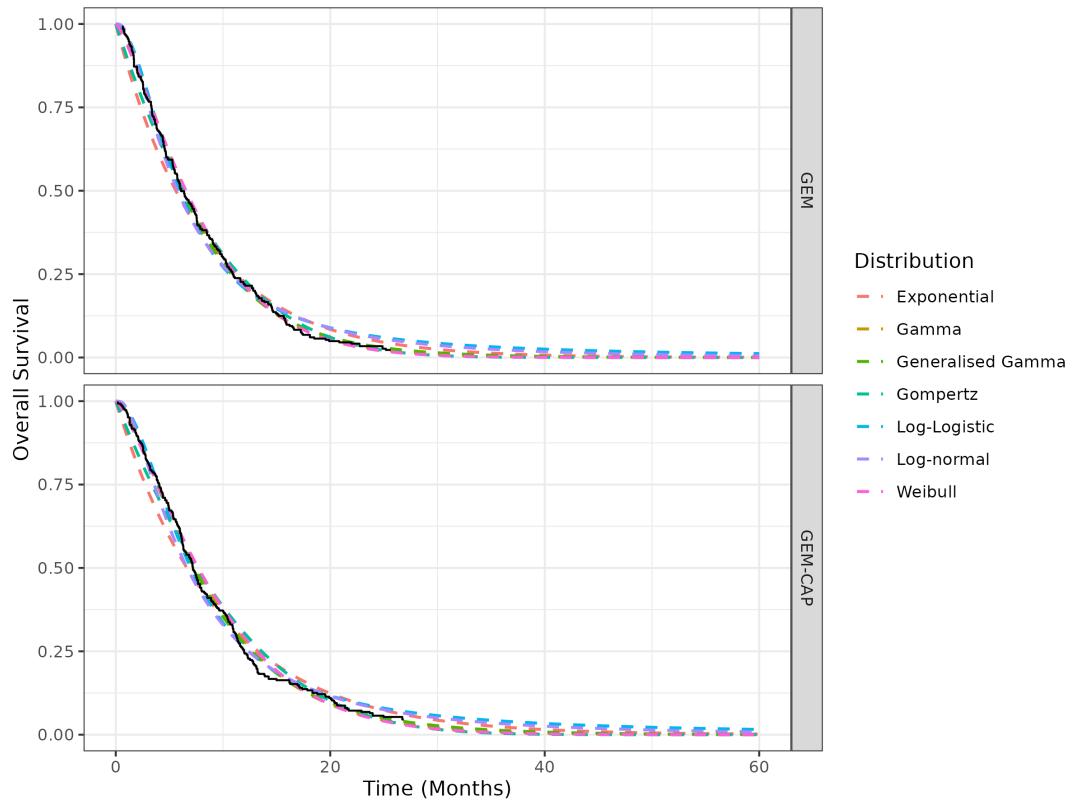


Figure 4.6: Cunningham (2009) parametric model extrapolations

Distribution	Kindler		Oettle		Rocha Lima		Spano	
	GEM	GEM-AXI	GEM	GEM-PEM	GEM	GEM-IRI	GEM	GEM-AXI
Exponential	812.560	791.128	1452.727	1452.816	998.183	996.717	156.984	343.492
Gamma	774.152	760.757	1422.005	1423.437	982.324	994.286	157.256	337.054
Gen. gamma	774.842	761.238	1418.603	1420.779	982.183	996.231	158.536	336.439
Gompertz	788.471	775.646	1444.122	1445.120	992.905	996.677	158.635	342.885
Log-Logistic	774.915	760.016	1421.247	1421.733	985.566	1004.260	156.674	335.106
Log-normal	773.184	759.797	1418.315	1421.474	983.702	1005.959	156.652	334.439
Weibull	776.307	763.113	1427.514	1428.696	985.016	994.628	157.630	338.687

Table 4.3: AIC scores in the Kindler, Oettle, Rocha Lima and Spano studies

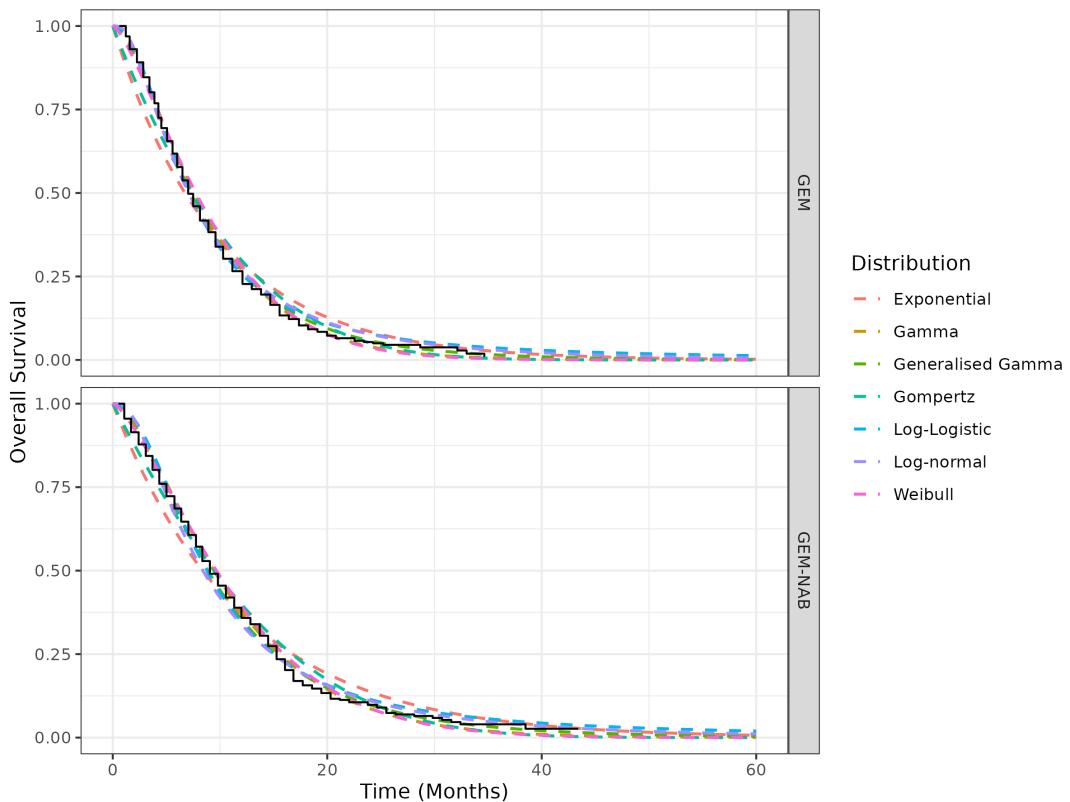


Figure 4.7: Goldstein (2015) parametric model extrapolations

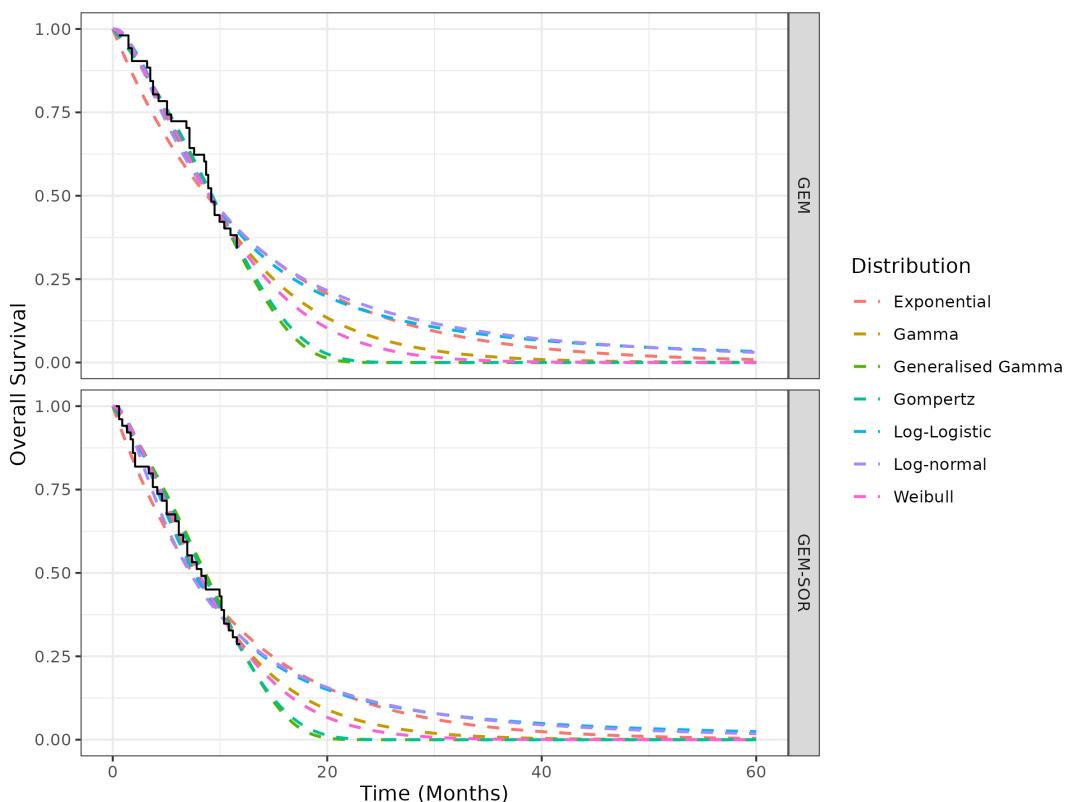


Figure 4.8: Goncalves (2012) parametric model extrapolations

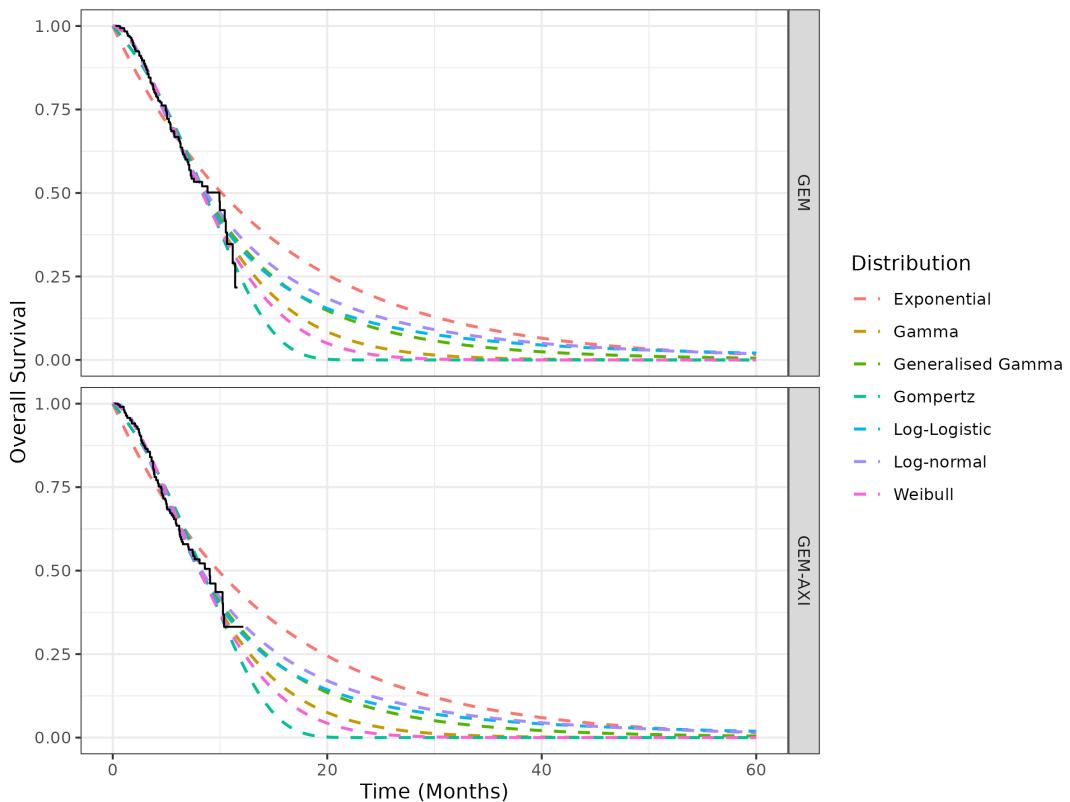


Figure 4.9: Kindler (2011) parametric model extrapolations

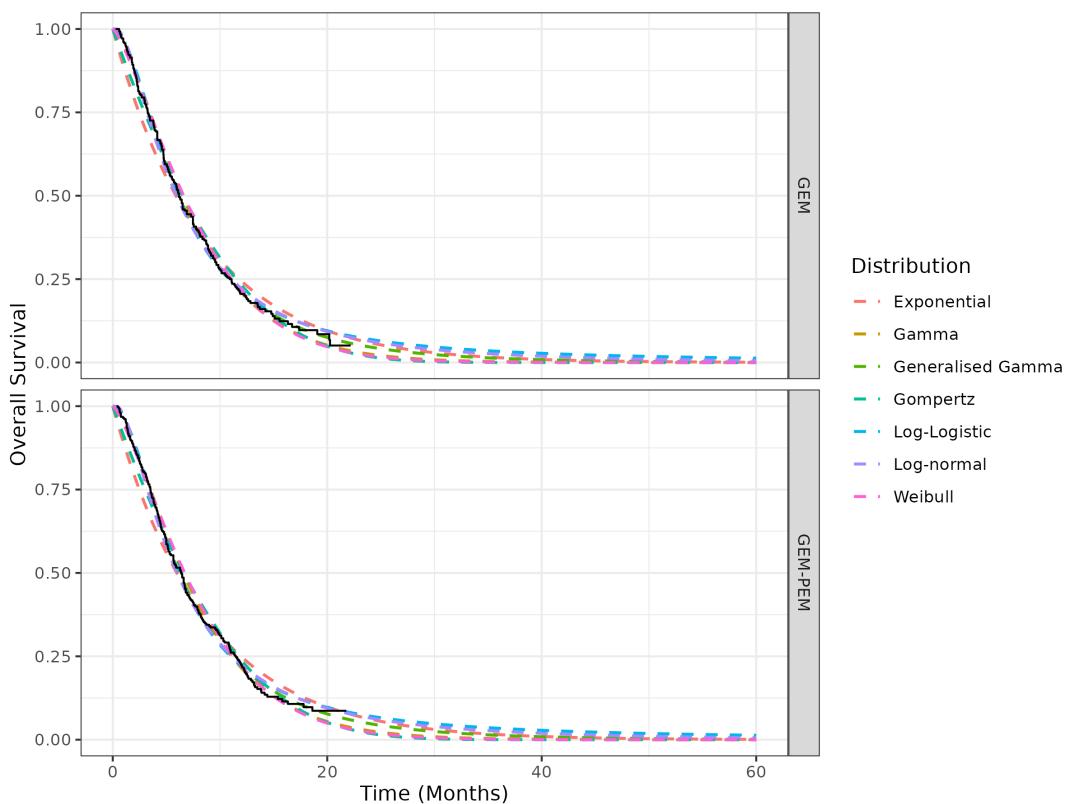


Figure 4.10: Oettle (2005) parametric model extrapolations

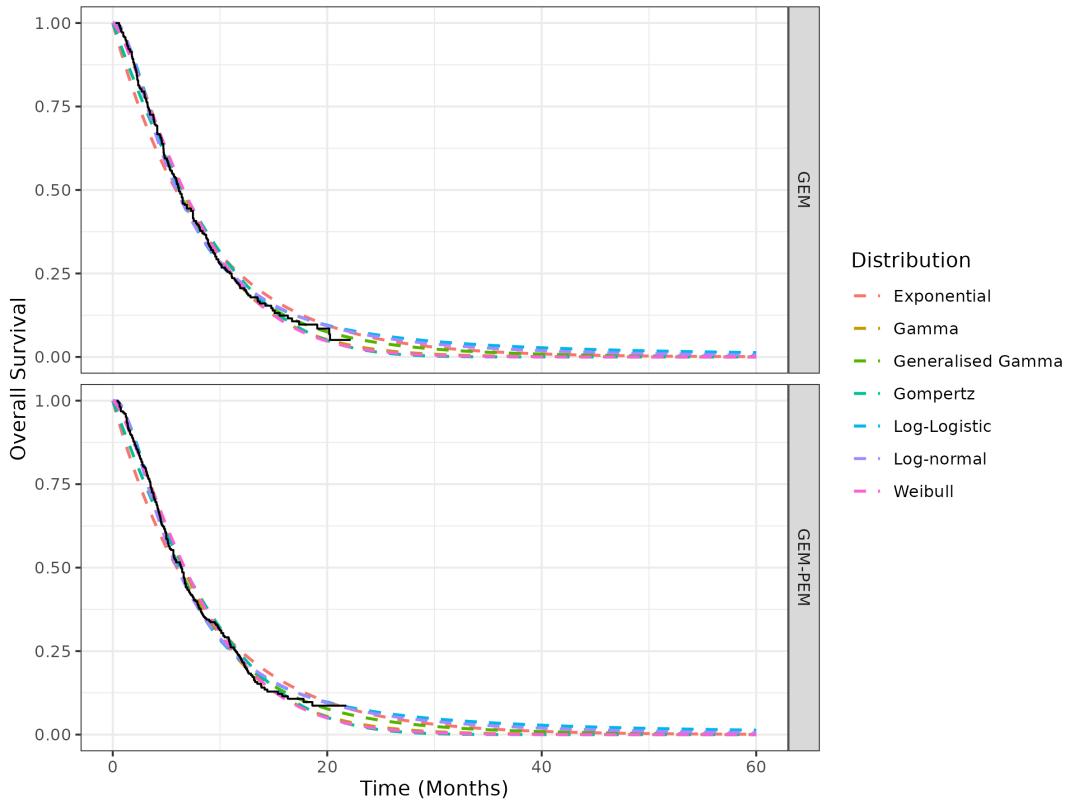


Figure 4.11: Rocha Lima (2004) parametric model extrapolations

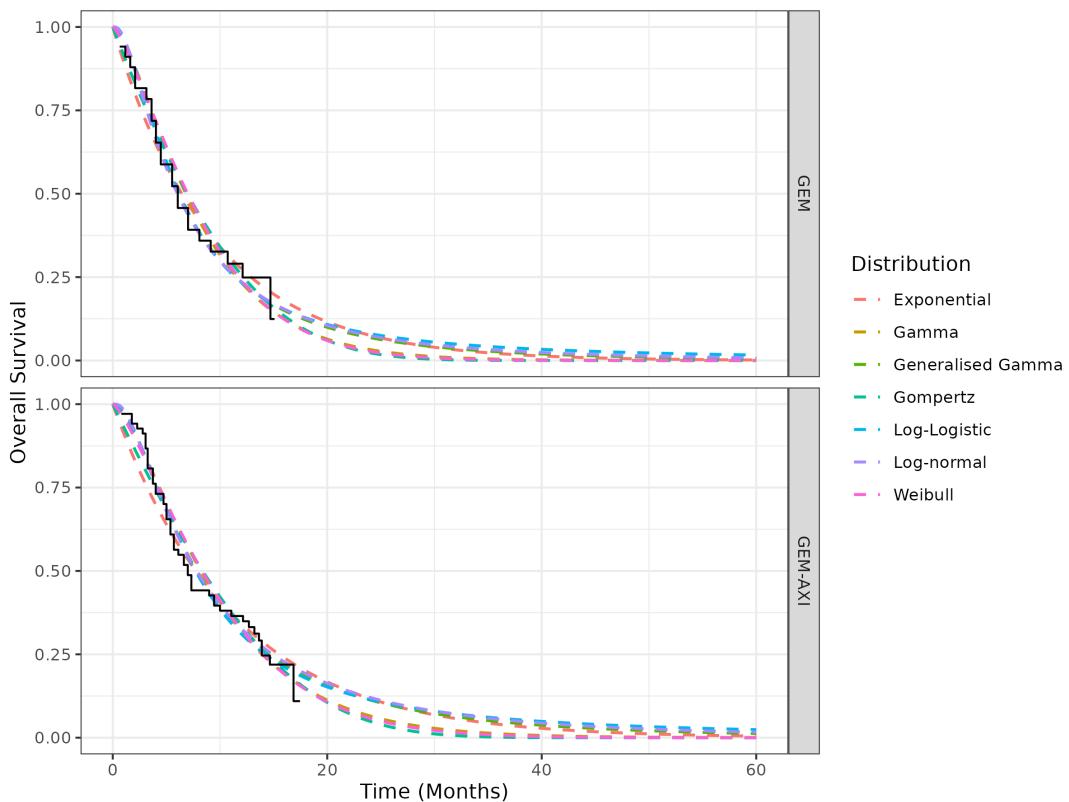


Figure 4.12: Spano (2008) parametric model extrapolations

A REVIEW OF META-ANALYSES FOR PANCREATIC CANCER

Prior to conducting this NMA, the literature was searched to identify MAs (ideally NMAs) comparing treatments for pancreatic cancer that included either all, or a subset of, the treatments included in this NMA. The literature on meta-analyses, and in particular NMAs is quite spares. Unfortunately, as a result of this, an NMA using an ML-NMR approach could not be found. Six meta-analyses were found, using a range of methods.

5.1 An Overview of MAs

In [Chen et al., 2021], the authors used a frequentist model to analyse both survival and toxicity data between modified FOLFIRINOX, regular FOLFIRINOX, and GEM-NAB. Data from 22 studies were included. The results of this NMA suggest that GEM-NAB and FOLFIRINOX are similarly efficacious, but FOLFIRINOX provides superior OS compared to GEM than GEM-NAB does. The results suggested that the improved efficacy of FOLFIRINOX and GEM-NAB were statistically significant compared to GEM, but GEM-NAB and FOLFIRINOX were not significantly different. The Chen NMA also showed the toxicity profile of FOLFIRINOX and GEM-NAB is similar. In particular, nausea diarrhea were more frequently observed in patients treated with FOLFIRINOX than GEM-NAB. In addition, patients treated with GEM-NAB had slightly higher risk of fatigue and anemia.

The NMA conducted in [Lin et al., 2019] included 31 studies in an efficacy meta-analysis and 32 studies in a safety meta-analysis. Naturally, only the efficacy analysis was relevant for this NMA. As with the Chen NMA, the authors used a frequentist NMA, and assessed the HRs. The focus of the Lin NMA was the first-line treatment of advanced pancreatic cancer. The only treatment included in the Lin NMA that was also relevant to this NMA was FOLFIRINOX. They found that FOLFIRINOX-based therapy was the best treatment.

The Nichetti meta-analysis [Nichetti et al., 2024] was not an NMA, but instead reconstructed KM data from seven studies, and then pooled the reconstructed IPD to compare the efficacy of FOLFIRINOX, NALIRIFOX, and GEM-NAB. Only phase III studies were included in this NMA. They found that FOLFIRINOX provided superior OS to GEM-NAB. They did validate their results with an NMA, although this was not reported in the main paper. The median OS of all treatments in this study was below 12 months. In terms of toxicity, the Nichetti NMA found NALIRIFOX had a favourable safety profile compared to FOLFIRINOX and GEM-NAB.

An NMA conducted by [Gresham et al., 2014], which included FOLFIRINOX, GEM-NAB, and GEM-CAP, found all three treatments to be associated with statistically significant improvements in OS relative to GEM and several other treatments. The Gresham NMA was a Bayesian NMA for calculating survival outcomes. The primary result outcomes of their NMA were the HR and survival gain, as defined as in Equation 5.1. The authors concluded that GEM combination therapies had more risk of grade 3 or 4 adverse events. This was to be expected given that NICE does not recommend GEM combination therapies for patients who may not be able to tolerate it.

$$\frac{\frac{\text{GEM Median OS}}{HR} - \text{GEM Median OS}}{\frac{\text{GEM Median PFS}}{HR} - \text{GEM Median PFS}} \quad (5.1)$$

The Takumoto [Takumoto et al., 2022] study considered only first line patients in Japan. In total, 25 studies were included comparing 22 treatments for OS were included in the Takumoto NMA. They found FOLFIRINOX and GEM-NAB offered improved OS compared to GEM. The reported HRs indicated that FOLFIRINOX was more efficacious than GEM-NAB. The Takumoto NMA was performed in a Bayesian framework, using an RE model. Although conducted with the Japanese clinical guidelines in mind, the included studies were two-arm studies not necessarily performed in Japan. Interestingly, the authors compared their results to the Gresham NMA, and found their results to be in line with those of Gresham.

The [Zhang et al., 2018] NMA included several treatments that were also included in this NMA. They found FOLFIRINOX was the best treatment in the network. They used an RE model and summarised the results using the Surface Under the Cumulative Ranking Score (SUCRA) values. In terms of SUCRA values of 12-month OS, FOLFIRINOX performed best, followed by GEM+S1 (not included in this NMA) and GEM-NAB.

5.2 Implications of the Literature Review

The literature showed that FOLFIRINOX and GEM-NAB are the best performing treatments for pancreatic cancer, which was to be expected given the treatment landscape. The papers consistently stated that more trials were needed, which would of course be ideal, but FOLFIRINOX and GEM-NAB are both very expensive treatments, so this is unlikely. Toxicity was considered in the majority of these NMAs, but was not considered here.

None of the NMAs performed any covariate adjustment, meaning there is a gap in the literature that could be filled using an ML-NMR. The issue with this is that IPD is of course required to do any meaningful covariate population adjustment. This NMA could therefore not be used to plug this gap, as only the proportion male was considered as a covariate. It was only included to facilitate using the ML-NMR method.

NMA OF PANCREATIC CANCER TRIALS

6.1 Base Case

6.1.1 Network of Evidence

Figure 6.1 presents the network of evidence for this NMA. There were two studies comparing GEM with GEM-AXI, but only one study for each other comparison. As indicated by the size of each node, GEM-SOR was the treatment with the lowest sample size, and GEM-NAB was the comparator with the highest. The GEM → GEM-NAB edge is a different colour due to being an IPD trial.

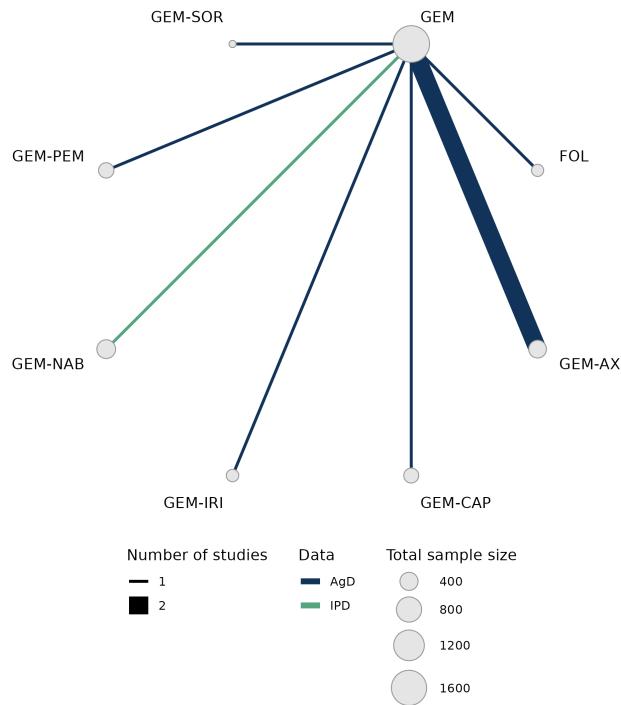


Figure 6.1: Network of evidence

6.1.2 Model Fitting and Selection

Both FE and RE models were fit using log-logistic, log-normal, and Weibull likelihoods. Vague priors were used for each model. Namely, the intercept prior was $N(0, 100)$, the treatment prior was $N(0, 10)$, the auxiliary prior was $hN(0, 5)$, and auxiliary regression prior was $N(0, 10)$. Here, hN denotes a *half-normal distribution*. For each model, sampling was done using 1000 iterations on four chains. The first

500 iterations were warmup iterations. In total, it took approximately six days to run all the models using 64 integration points for the numerical integration.

Table 6.1 presents the selection statistics for each model. The RE log-logistic model performed best in terms of LOOIC and DIC score. The FE log-logistic was very similar to the RE log-logistic model. Indeed, the scores were only separated by 0.0398. The RE log-normal model was only 0.9148 above the LOOIC score of the RE log-logistic model. Given the similarity in the LOOIC scores for the FE and RE log-logistic models, the final model was selected based on model convergence.

Likelihood	Type	DIC	LOOIC
Log-logistic	FE	16974.3668	16972.9184
Log-logistic	RE	16972.2638 ←	16972.8786 ←
Log-normal	FE	107813403.9532	48652.0393
Log-normal	RE	16977.7101	16973.7934
Weibull	FE	16989.2670	16992.9722
Weibull	RE	3.1937e42	5.8355e21

Table 6.1: Model selection statistics for each model

Figure 6.2 and Figure 6.3 present the trace plots for the FE and RE log-logistic models respectively. The FE model demonstrated much better convergence than the RE model. In the RE model, the traces for each treatment were very thin with lots of spikes. In particular, for GEM-AXI, there was a clear issue with convergence just before iteration 300. Conversely, the FE traces for each treatment showed few spikes, and the chains were well mixed, indicating good convergence.

Figure 6.4 presents the (standardized) parallel coordinates plot for the FE log-logistic model. Each green line represents an iteration, connecting the values of the parameter corresponding to each treatment. The value of each parameter has been standardized by the transformation given in Equation 6.1. The purpose of a parallel coordinates plot is to show divergent transitions. Figure 6.4 shows, by nature of all the iterations being green, that there were no divergent transitions. Figure 6.5 shows the parallel coordinates plot for the RE log-logistic model, for which the divergent transitions are clear.

$$f(x) = \frac{x - \mu_x}{\sigma_x} \quad (6.1)$$

Figure 6.6 and Figure 6.7 present the univariate histograms and bivariate scatter plots for the parameters of each treatment in the FE and RE log-logistic models respectively. The FE log-logistic histograms were much wider than their RE model counterparts. The scatter plots for each parameter were much sparser and scattered in the RE model than in the FE model. The divergent transitions are shown as red points in the scatter plot.

Given the clinical plausibility of the FE model and the demonstrated superior convergence compared to the RE model, the FE log-logistic model was selected.

6.1.3 Results

Figure 6.8 presents a density plot of the prior versus posterior distribution for each treatment. The posterior distribution is shown in green, and the prior in red. The priors were of course uninformative priors. There were two key takeaways from Figure 6.8. Firstly, the good convergence of this model was demonstrated by the shape of the posterior histogram through the consistent and unimodal shape of the distributions. Secondly, the shape of the posterior distributions being significantly different from the shape of the prior showed that the data were the primary influence on the parameter estimates. If the posterior were a similar shape to the prior, it would indicate the data has not contributed much to the model. Some inference about the effectiveness of these treatments compared to GEM could already be drawn from this plot. For example, FOL, GEM-NAB, and GEM-CAP are all centered above zero, indicating superior OS compared to GEM. Figure 6.8 also demonstrates the uncertainty for each interval. Indeed, GEM-SOR has a particularly wide posterior distribution, and GEM-NAB had quite a thin posterior distribution. This was to be expected as there is more uncertainty in the GEM-SOR arm due

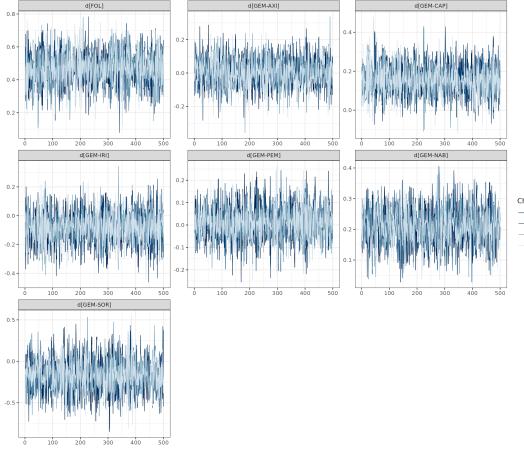


Figure 6.2: Trace plot for the FE log-logistic model ML-NMR

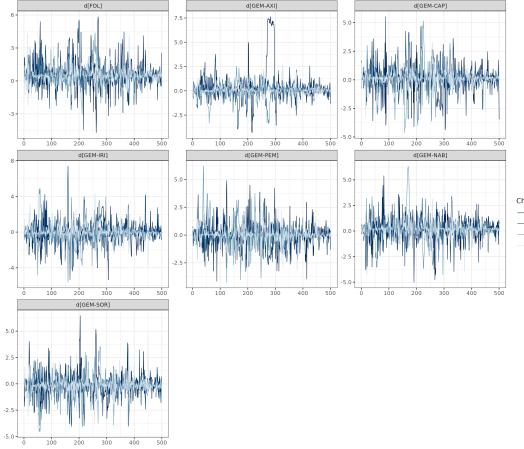


Figure 6.3: Trace plot for the RE log-logistic model ML-NMR

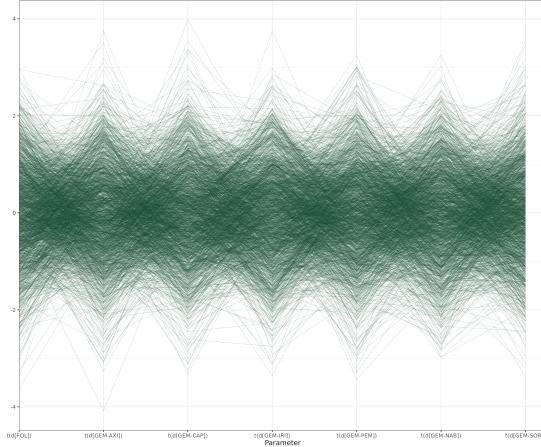


Figure 6.4: Parallel coordinates plot of the FE log-logistic model

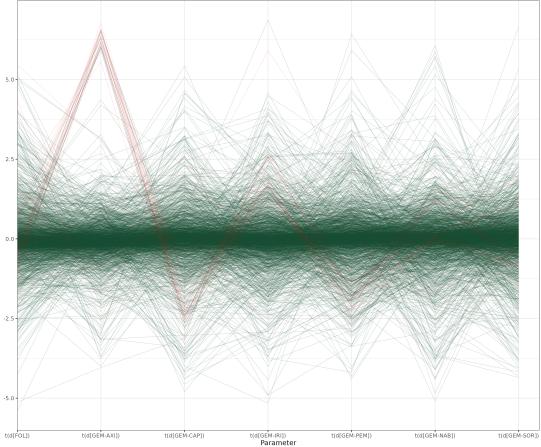


Figure 6.5: Parallel coordinates plot of the RE log-logistic model

to fewer patients in the GEM-SOR arm of the network.

Figure 6.9 and Figure 6.10 present the KM curves of each treatment in the Conroy and Goldstein populations, respectively. Figure 8.1, available in Appendix 8.1 presents the KM curves in each population. As Conroy and Goldstein had the most mature data of all studies, and contained the two main treatments, these plots were given more priority in the analysis. Both Figure 6.9 and Figure 6.10 showed superior OS for the FOL compared to all other treatments. GEM-NAB provides superior OS to GEM-CAP, but worse OS than FOL in the Conroy population. The same was observed in the Goldstein population. In both populations, GEM-SOR had the worst OS until between 12 and 15 months, but then crosses the GEM and GEM-AXI curves to give a higher OS at the end of the extrapolation period.

Figure 6.11 presents the estimated RMST of each treatment in each population. The credible intervals for GEM-SOR were very wide, likely due to the low number of patients in the GEM-SOR arm. The thin line for each treatment represents the 95% credible interval. Based on this interval, FOL was significantly better than GEM, GEM-AXI, GEM-PEM and GEM-SOR in the Conroy, Goldstein and Oettle studies. FOL was also significantly better than GEM-IRI in the Conroy, Cunningham, Goldstein, Kindler, Oettle and Rocha Lima studies. FOL was not significantly better than FOL in any study population. GEM-CAP and GEM-NAB gave similar RMST estimates in each study population. Indeed, GEM-NAB provided a higher RMST estimate in each population, but this improvement was not significant. In addition, GEM, GEM-AXI and GEM-PEM gave similar estimates in each treatment arm, with GEM-PEM providing slightly better RMST than GEM and GEM-AXI in each case. GEM-IRI gave higher RMST estimates than GEM-SOR in each population, but worse RMST estimates than GEM,

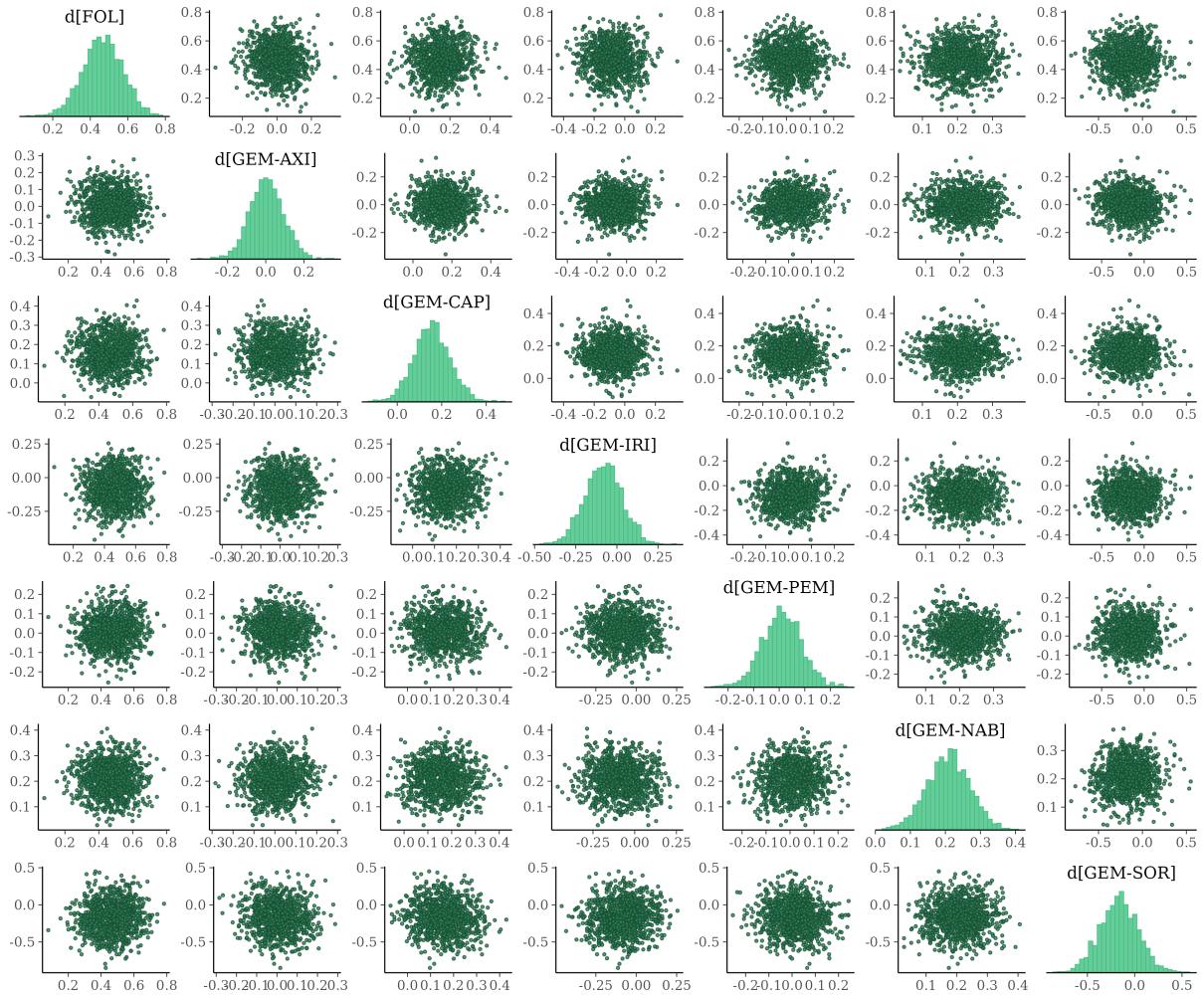


Figure 6.6: Pairs plot for the FE log-logistic model

GEM-AXI, and GEM-PEM.

Figure 6.12 presents the estimated median OS of each treatment in each population. The median OS estimates follows the same pattern as the RMST estimates. Namely, GEM-SOR and GEM-IRI gave the lowest and second-lowest estimates for median OS in each study population, GEM, GEM-AXI, and GEM-PEM gave similar estimates, and FOL and GEM-NAB gave the highest and second-highest estimates of median OS. Compared to the RMST estimates, the credible interval for FOL was quite wide in the Goldstein, Goncalves, Kindler, and Spano studies. GEM-NAB provides higher median OS estimates than GEM-CAP in each population, but does not provide significant improvements. The similarity between GEM, GEM-AXI, and GEM-PEM that was observed in the RMST plots was also observed in the median OS plots.

Figure 6.13 presents the population-average relative treatment effects of each treatment in terms of the log survival-time from the median OS estimates. FOL and GEM-NAB were both significantly better than GEM, but all other treatments crossed 0, indicating no significance. GEM-CAP did not provide a significant improvement to OS compared to GEM. In addition, Figure 6.14 presents the relative treatment effects for all constraints in the network. As expected, all studies except GEM-CAP and GEM-NAB are significantly worse than FOL.

Figure 6.15 presents the cumulative rank probability for each treatment. FOL had a considerably higher probability of being the best treatment. FOL had a probability of being the best treatment of 0.97. GEM-NAB and GEM-CAP had probabilities of being the best treatment of 0.02 and 0.01 respectively. The cumulative probability of GEM-NAB increased quicker than GEM-CAP, having a

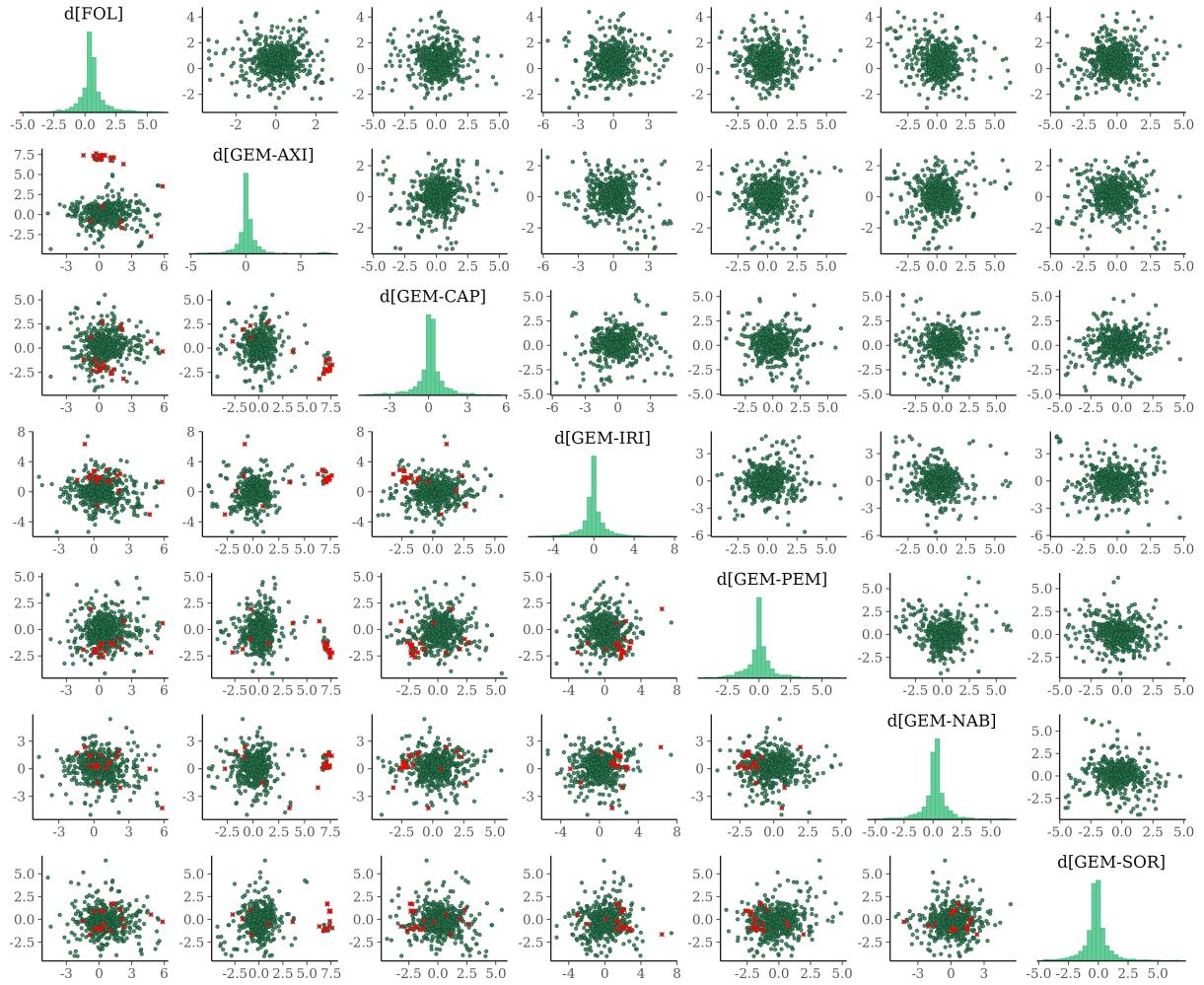


Figure 6.7: Pairs plot for the RE log-logistic model

cumulative probability of being the second best treatment of 0.66, compared to GEM-CAP having a cumulative probability of 0.28. FOL achieves a cumulative probability of 1 by rank three, and GEM-NAB and GEM-CAP achieved a cumulative probability of 1 at rank five and seven respectively. Figure 8.2, available in Appendix 8.1 presents the non-cumulative posterior ranks. The SUCRA value of FOL, GEM-NAB, and GEM-CAP was 99.5%, 80.3%, and 71.0% respectively. The SUCRA values of GEM-SOR and GEM-IRI were 15.8% and 20.7%, respectively which were the lowest SUCRA values in the network. GEM-AXI, GEM-PEM, and GEM had SUCRA values of 37.8%, 39.0%, and 36.0%, respectively.

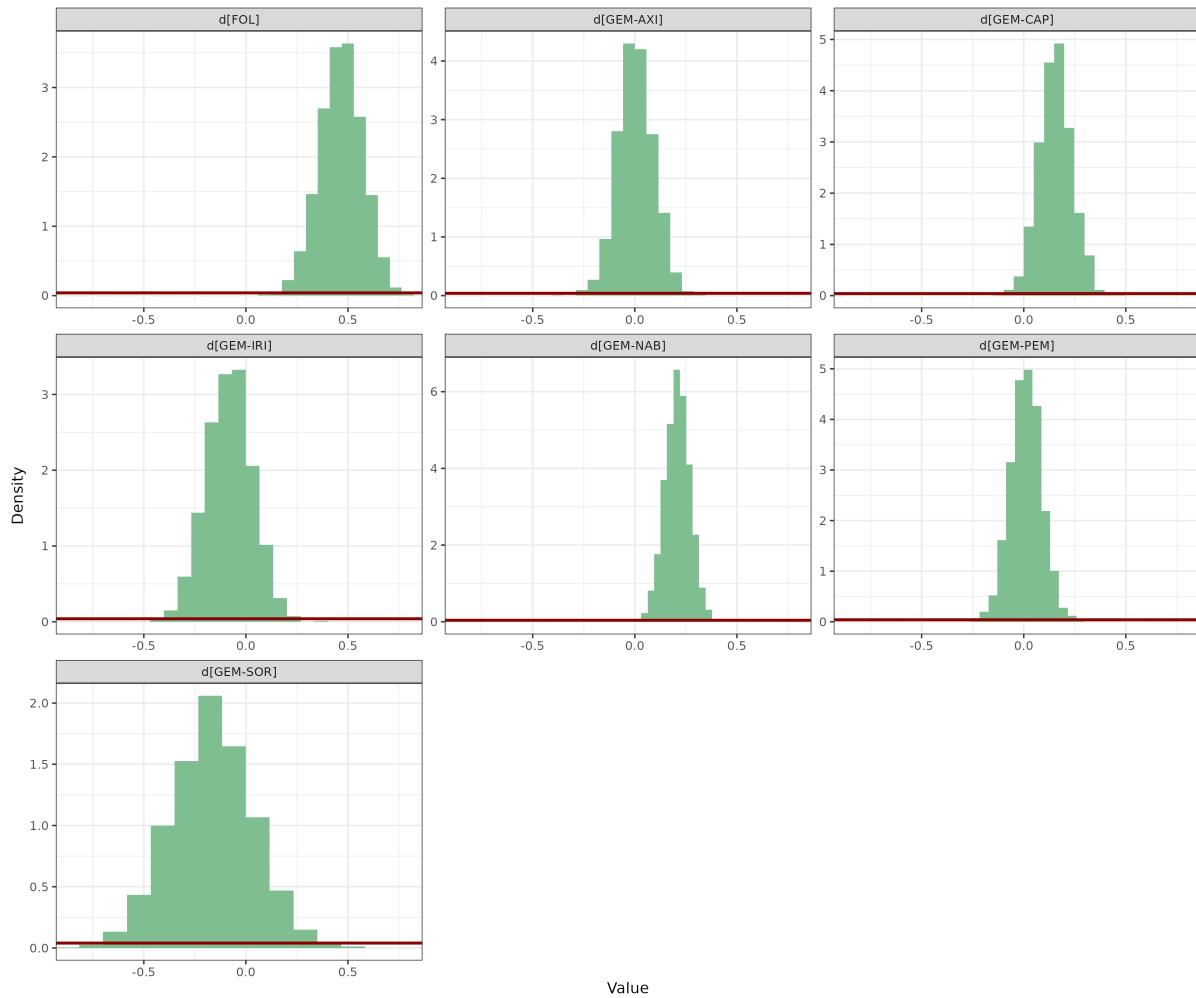


Figure 6.8: Prior versus posterior distribution for each treatment

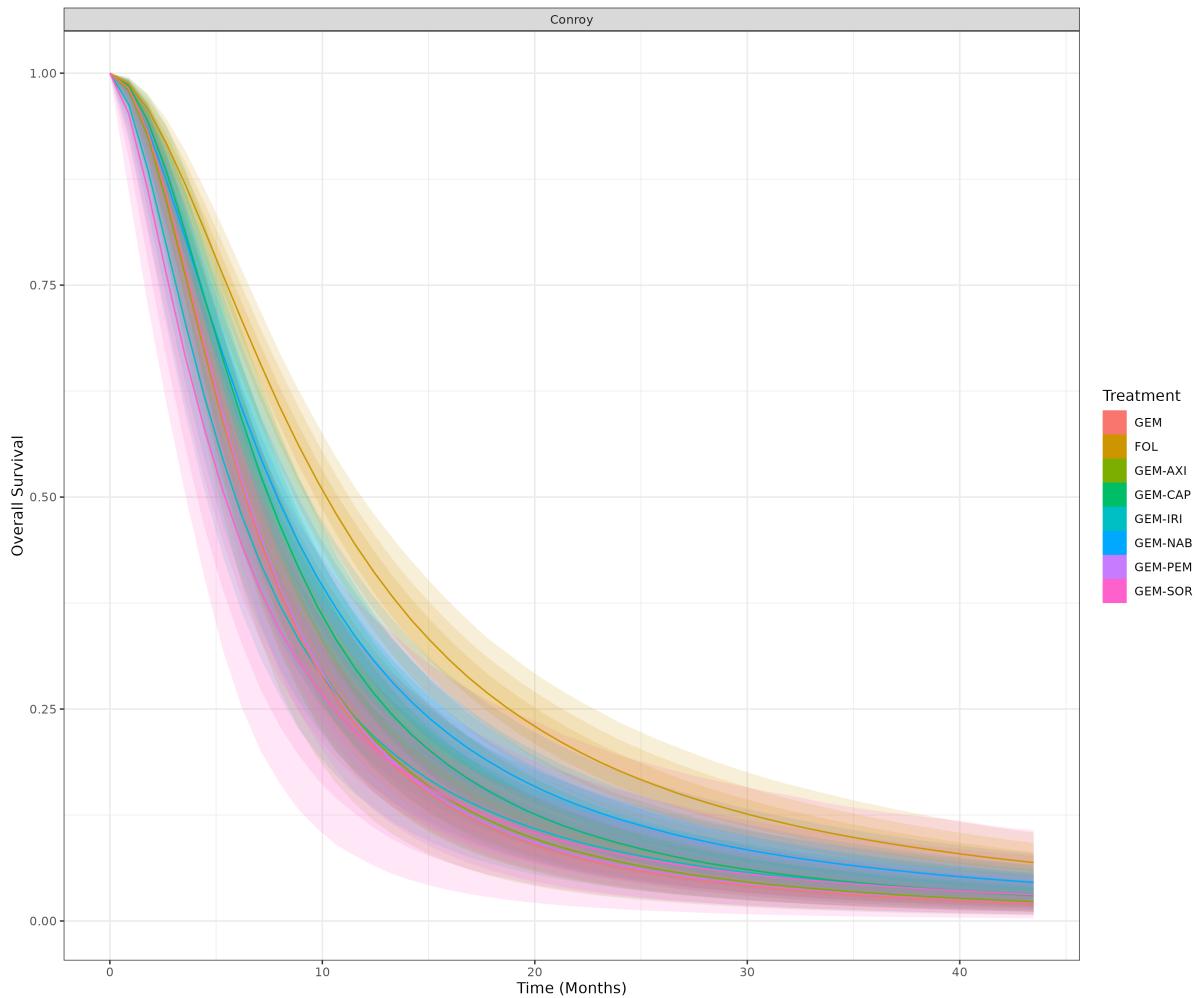


Figure 6.9: OS of each treatment in the Conroy population

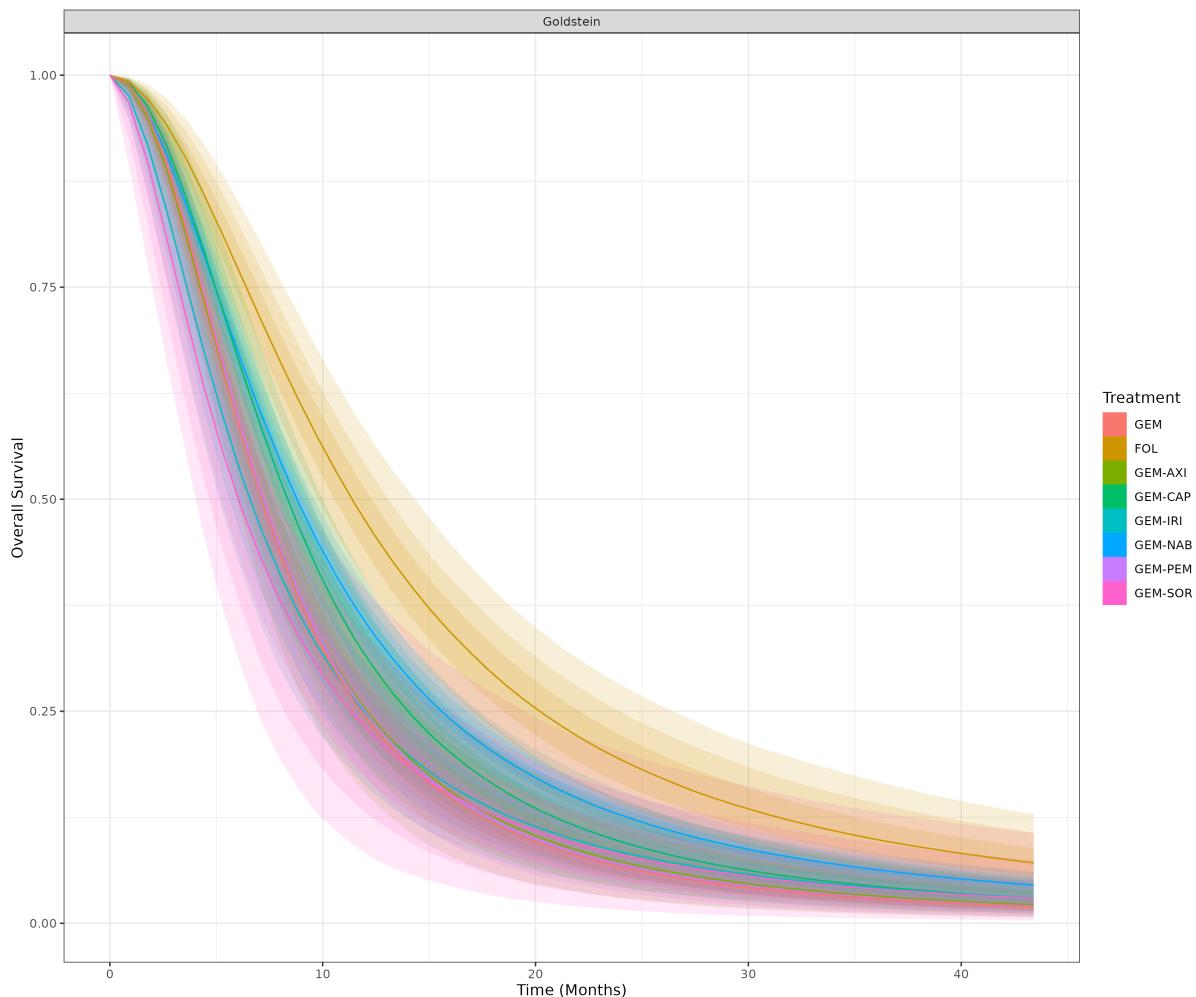


Figure 6.10: OS of each treatment in the Goldstein population

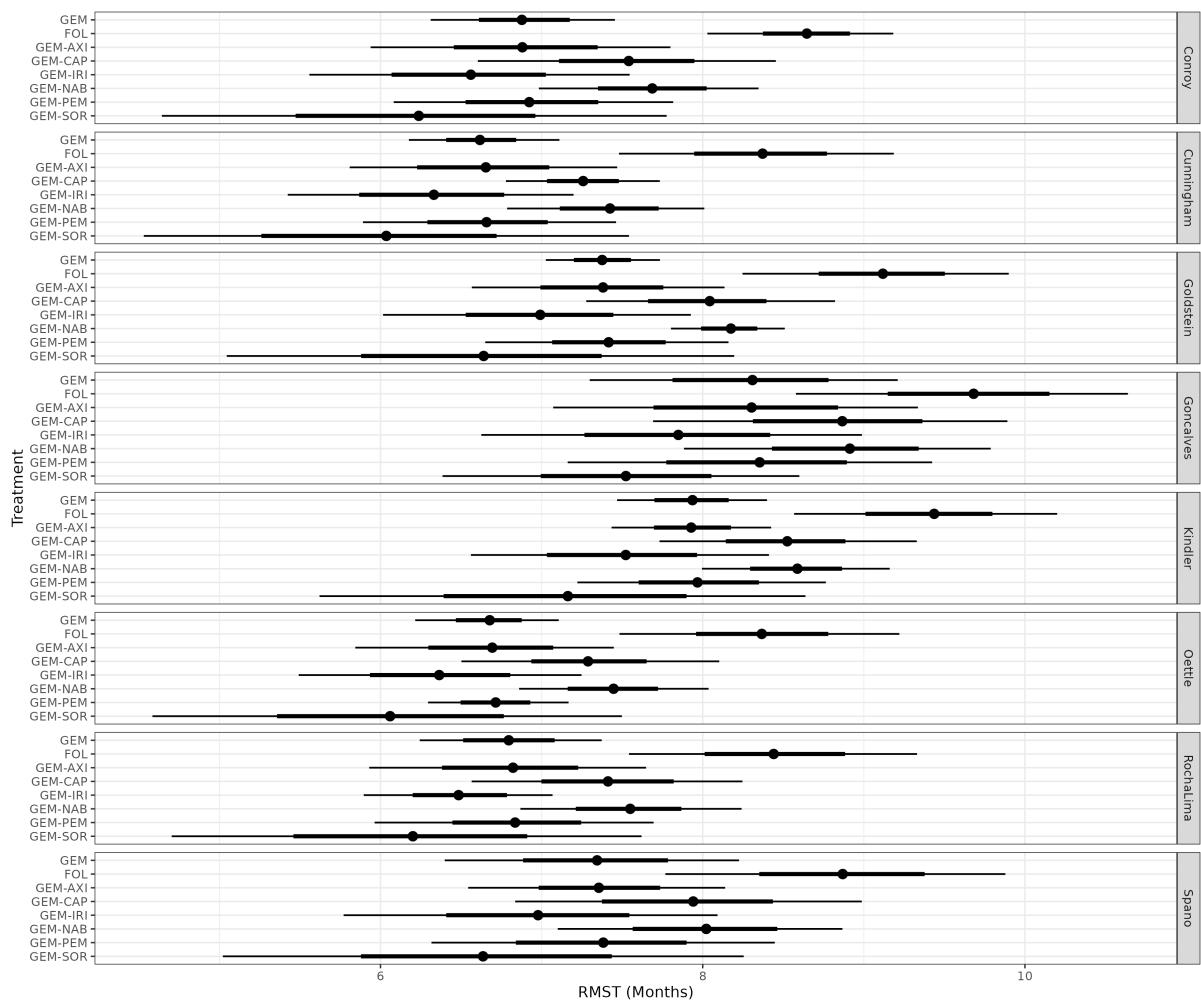


Figure 6.11: RMST of each treatment in each population

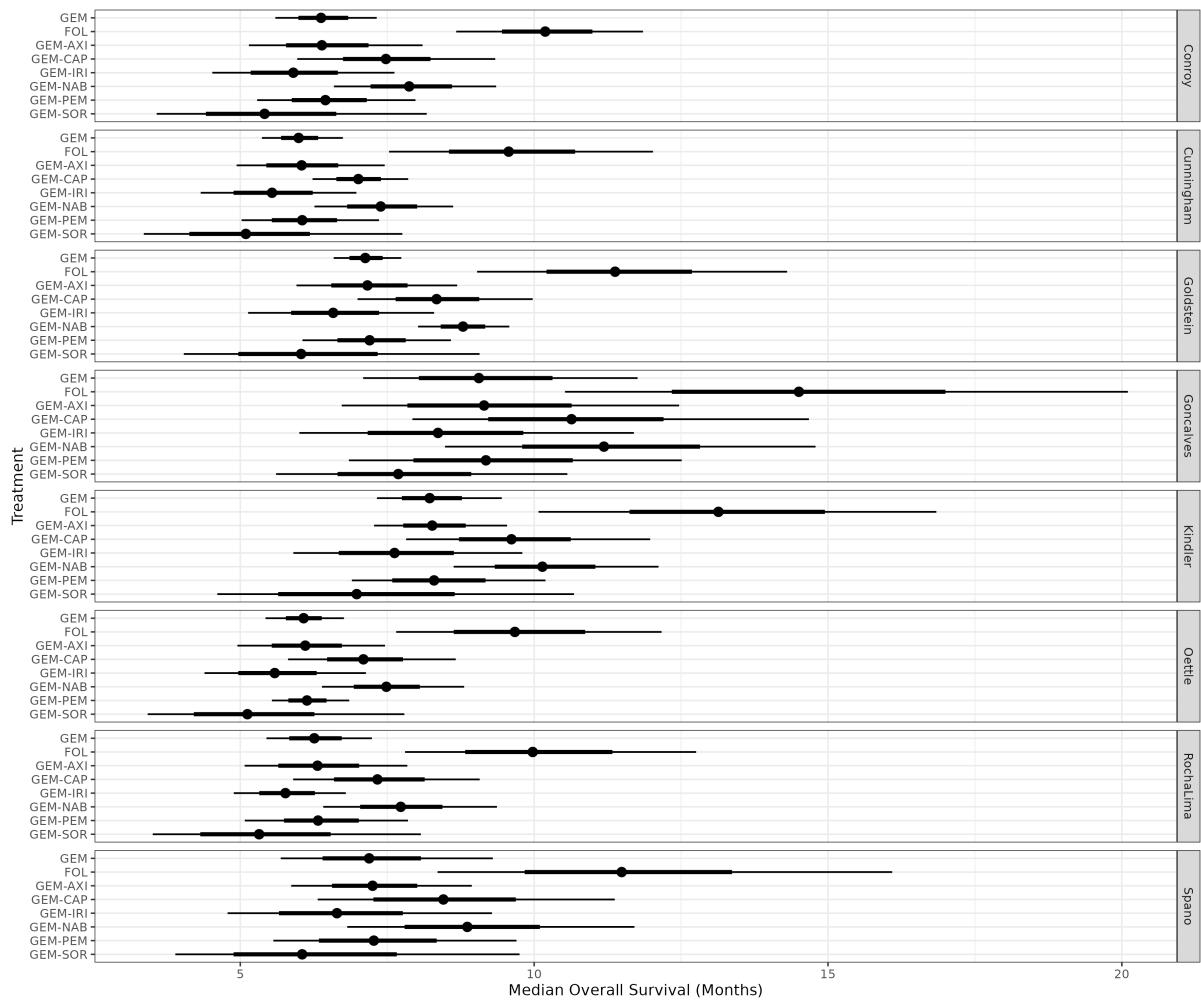


Figure 6.12: Median OS of each treatment in each population

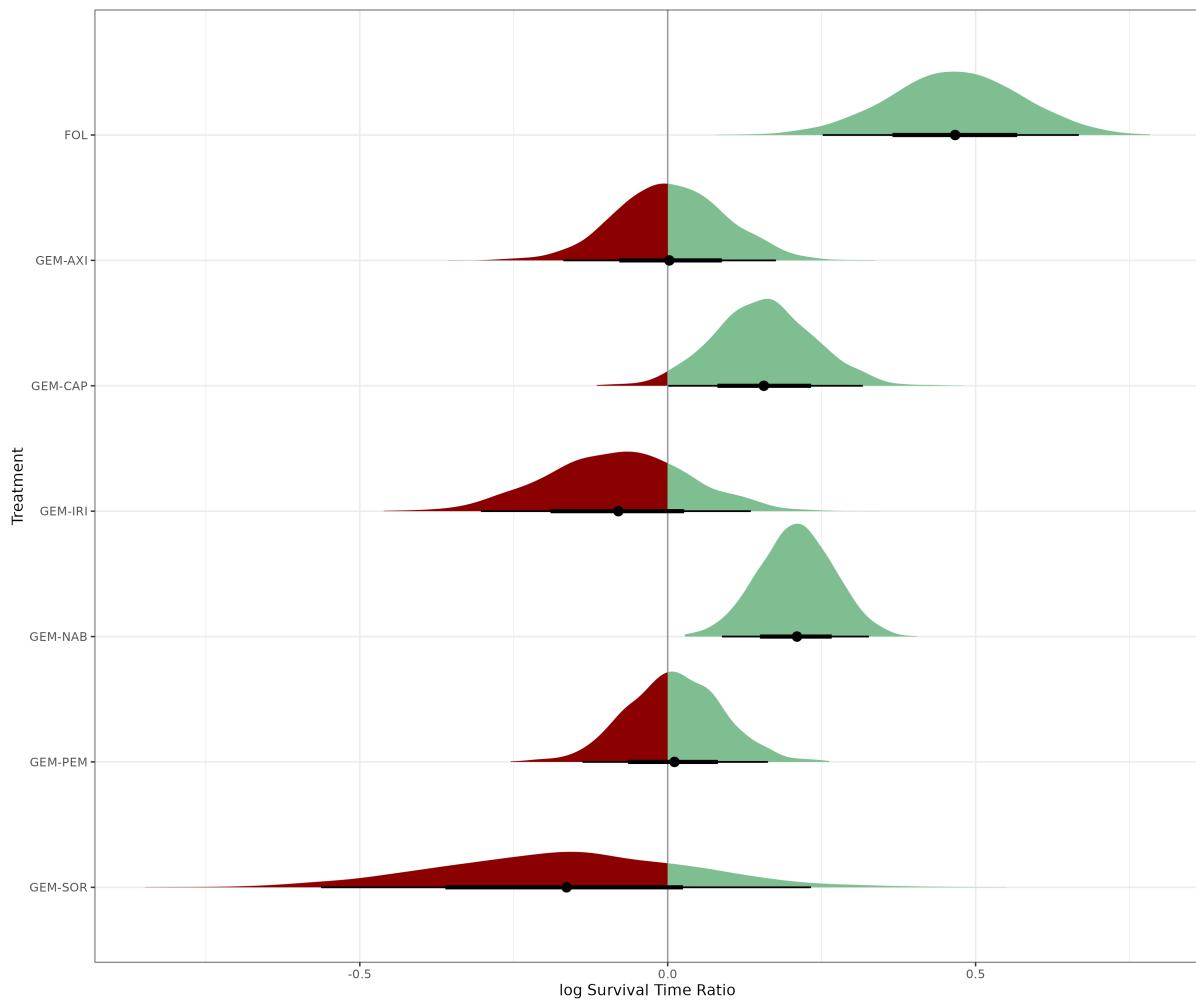


Figure 6.13: Relative treatment effects for all treatments versus GEM

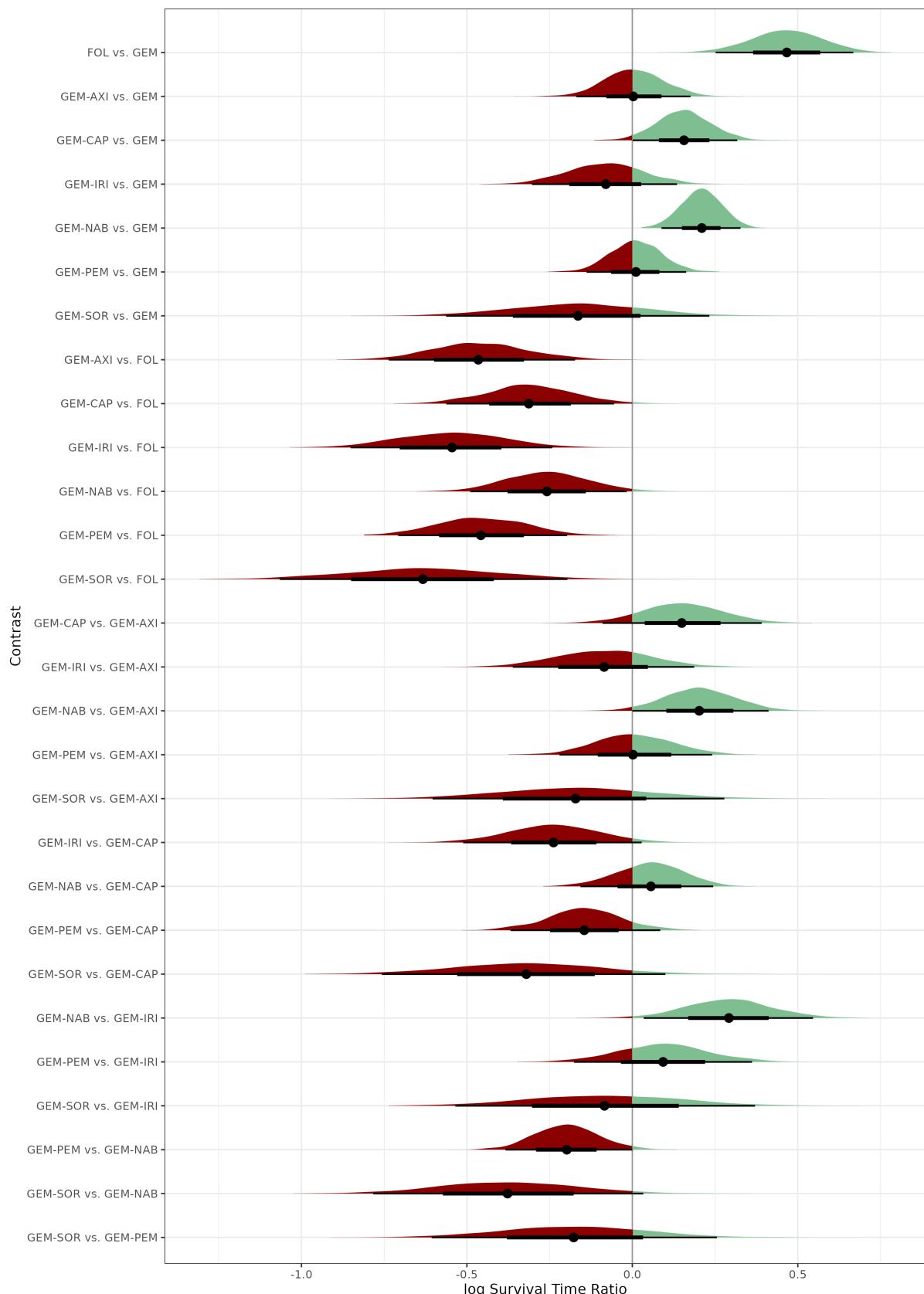


Figure 6.14: Pairwise relative treatment effects for all treatments

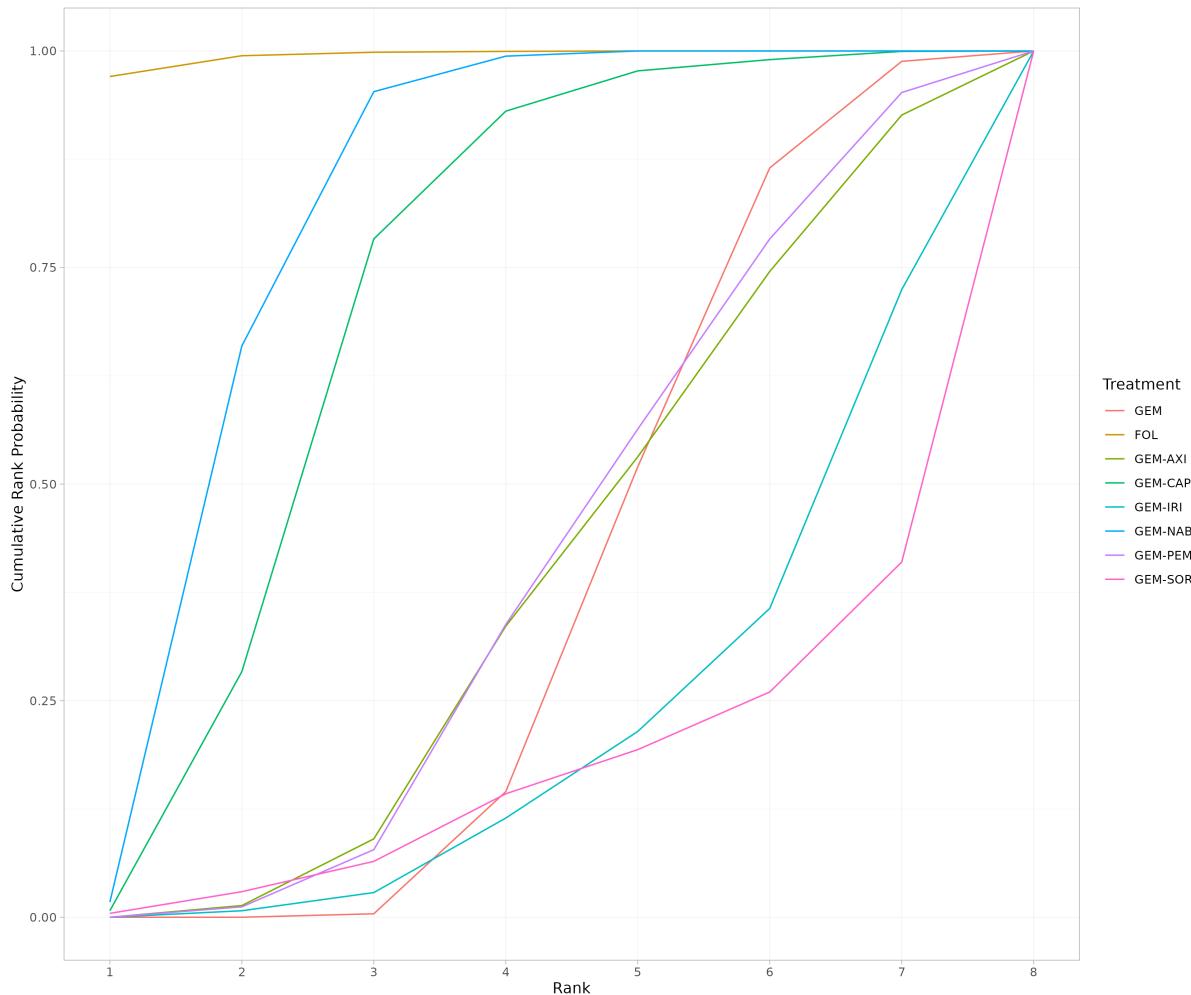


Figure 6.15: Cumulative rank probability for each treatment

6.2 Sensitivity Analyses

6.2.1 Sensitivity Analysis 1

The first SA was used to account for the fact that the Conroy study had a stricter age eligibility criteria than the other studies. By removing FOL, this SA also accounted for the fact that FOL can only be given to patients fit enough to take it, and therefore decision makers may not always have FOL as a treatment option. Figure 6.16 presents the network of evidence for this SA. The only difference between this and the base case evidence network was the absence of a GEM → FOL comparison.

6.2.1.1 Model Fitting and Selection

Table 6.2 presents the model selection statistics. The RE log-normal and FE-log normal scored best in terms of LOOIC and DIC score, respectively. As in the base case, both models were assessed to determine which one provided a better fit. Figure 6.17 and Figure 6.18 present the trace plots for the FE and RE model respectively. The FE model had better traces for each treatment parameter than the RE model. In addition, Figure 6.19 and Figure 6.20 present the normalised parallel coordinates plot for the FE and RE model, respectively. The FE model showed no divergent transitions, with intermingling chains. The RE model showed many divergent transitions. In particular, several chains had a completely different pattern to the remainder, indicating that the sampling may have gotten stuck in a local minima or maxima. The RE model was therefore not deemed to have achieved acceptable convergence, and the FE log-normal model was used. The pairs plot for the FE and RE model are presented in Figure 6.21 and Figure 6.22.

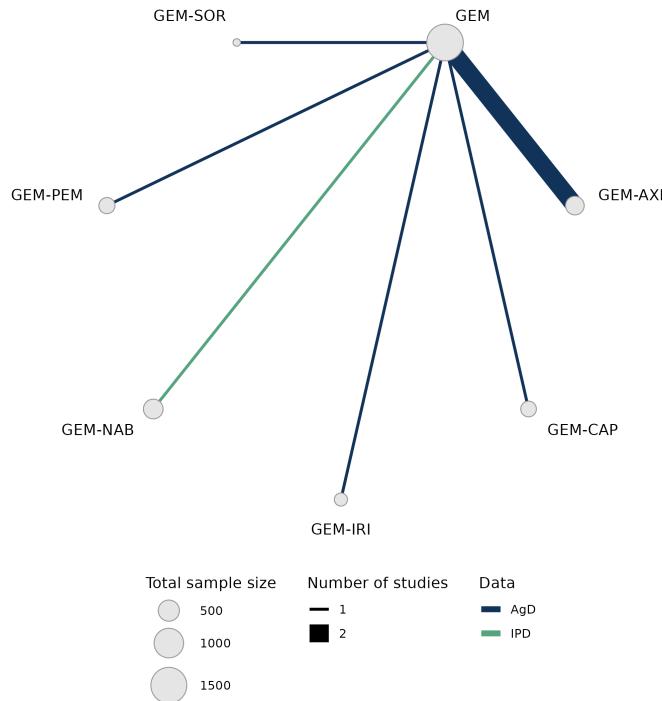


Figure 6.16: Network of evidence

6.2.1.2 Results

Figure 6.23 presents the prior versus posterior distribution for each treatment. The posterior distributions were significantly different from the priors, indicating the data was primary influence on the

Likelihood	Effect	DIC	LOOIC
Log-logistic	Fixed	15176.1347	15177.3745
Log-logistic	Random	15178.0459	15177.7154
Log-normal	Fixed	15172.7599 ←	15174.3034
Log-normal	Random	15172.9396	15173.1961 ←
Weibull	Fixed	15203.4555	15203.4465
Weibull	Random	Inf	23994604.6763

Table 6.2: Model selection statistics for each model

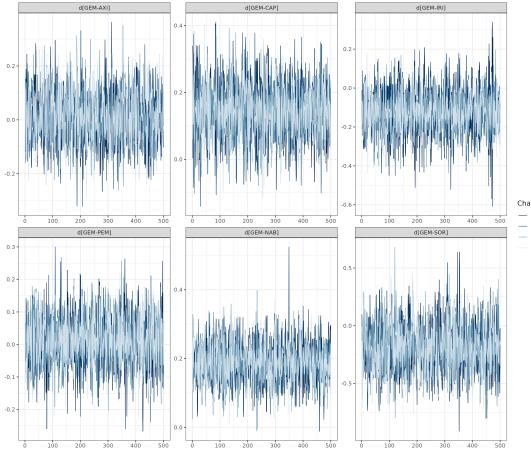


Figure 6.17: Trace plot for the FE log-normal model ML-NMR

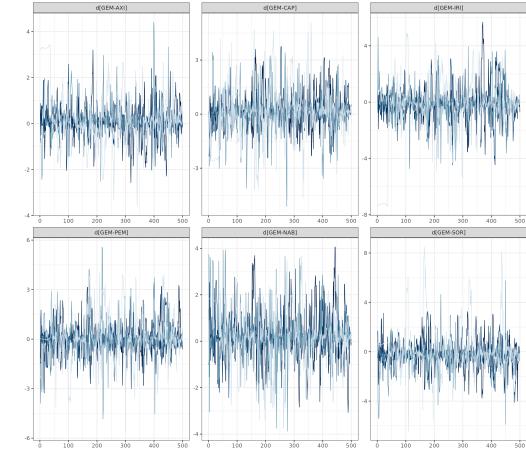


Figure 6.18: Trace plot for the RE log-normal model ML-NMR

parameter estimates. The GEM-SOR parameter had a wider posterior than the other treatments due to less available information for the GEM-SOR treatment arm.

Figure 6.24 presents the simulated KM curves of each treatment in the Goldstein population. GEM-SOR was the worst treatment in this population until around 25 months, where it becomes comparable to GEM. GEM-NAB had the highest OS throughout the entire observation period. GEM-CAB and GEM-NAB had similar OS until around five months, after which the GEM-CAP curve declines quicker than GEM-NAB.

Figure 6.25 presents the RMST estimates for each treatment in each population. GEM-NAB had the highest RMST estimate in each population. GEM-CAP had the second highest RMST of each treatment in each population. In the Goncalves and Kindler studies, the RMST estimates of GEM-CAP and GEM-NAB were quite similar. GEM, GEM-AXI, and GEM-PEM had very similar RMST estimates in each population. GEM-SOR and GEM-IRI had the lowest and second-lowest RMST estimates in each population. Figure 6.26 presents the median OS estimates of each treatment in each population. The same pattern was observed in the median OS estimates as in the RMST estimates. The gap between GEM-NAB and GEM-CAP was slightly more pronounced in the median OS estimates than in the RMST estimates.

Figure 6.27 presents the relative effects of each treatment versus GEM in terms of the log survival time ratio. GEM-AXI, GEM-PEM, GEM-CAP, and GEM-NAB were all better than GEM. Only GEM-NAB was significantly better than GEM. Figure 6.28 presents the relative effects for all comparisons. This plot showed that GEM-NAB was better than GEM-CAP, but not significantly, which was consistent with the base case analysis. GEM-NAB was significantly better than GEM-IRI.

Figure 6.29 presents the cumulative rank probability plot. GEM-NAB and GEM-CAP had the highest probabilities of being the best treatment. The SUCRA values of GEM-NAB and GEM-CAP were 60.8% and 30.0% respectively. GEM-SOR had a higher probability of being the best, second best, third best, and fourth best treatment than GEM-IRI, and a higher probability of being the best and second best treatment than GEM. GEM-PEM and GEM-AXI had similar probabilities for each ranking.

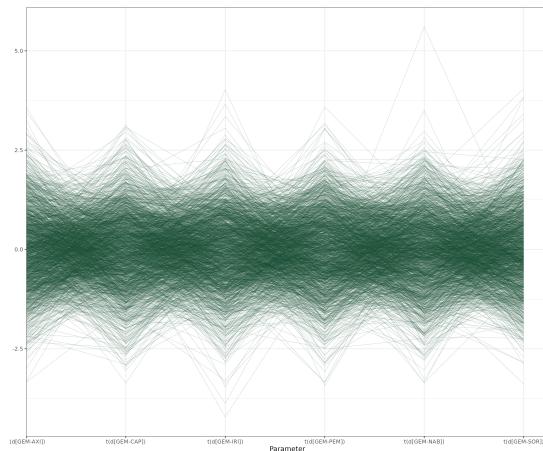


Figure 6.19: Parallel coordinates plot of the FE log-normal model

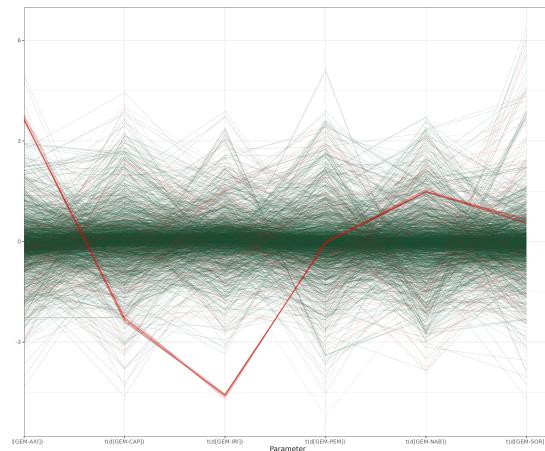


Figure 6.20: Parallel coordinates plot of the RE log-normal model

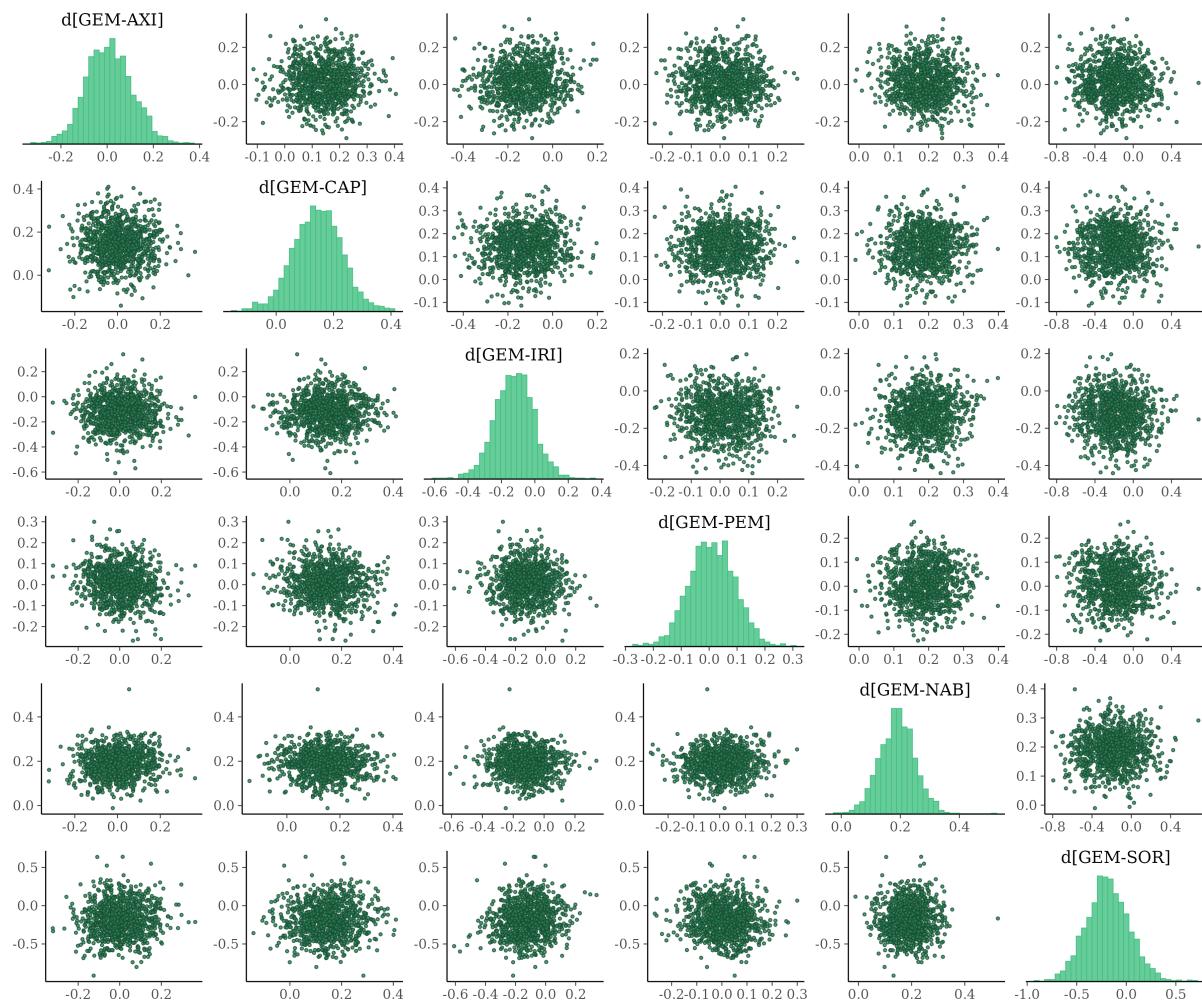


Figure 6.21: Pairs plot for the FE log-normal model

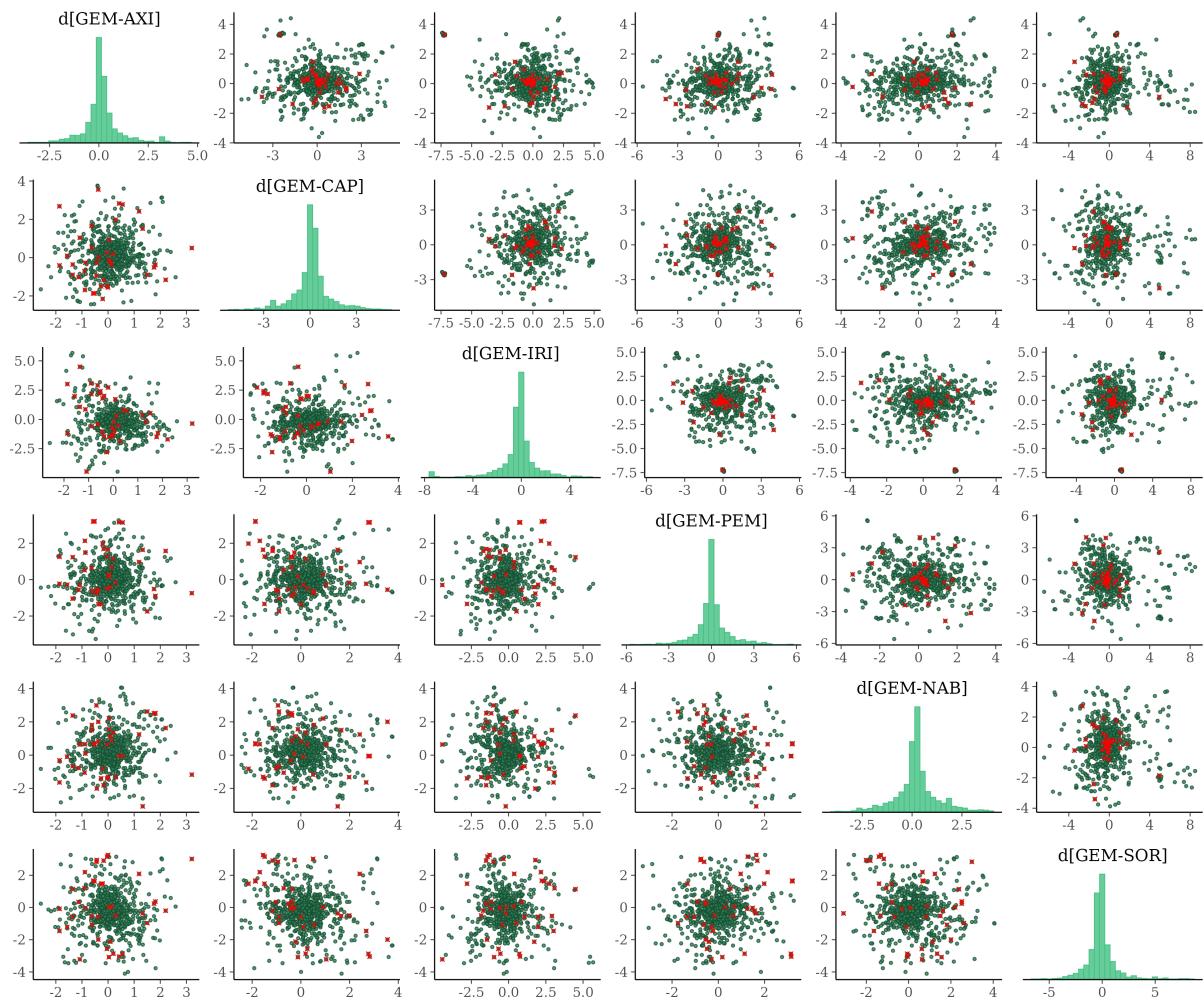


Figure 6.22: Pairs plot for the RE log-normal model

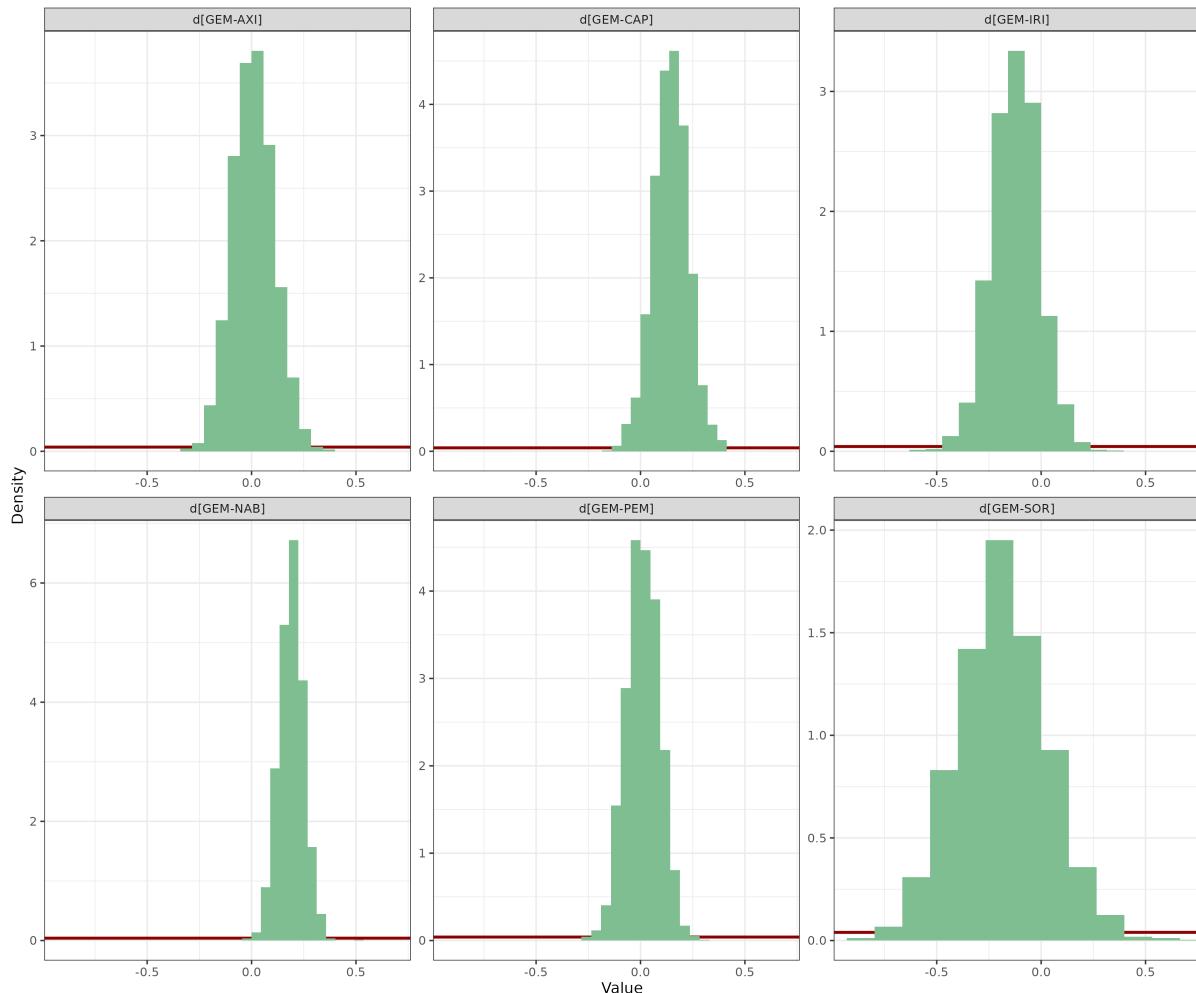


Figure 6.23: Prior versus posterior distribution for each treatment

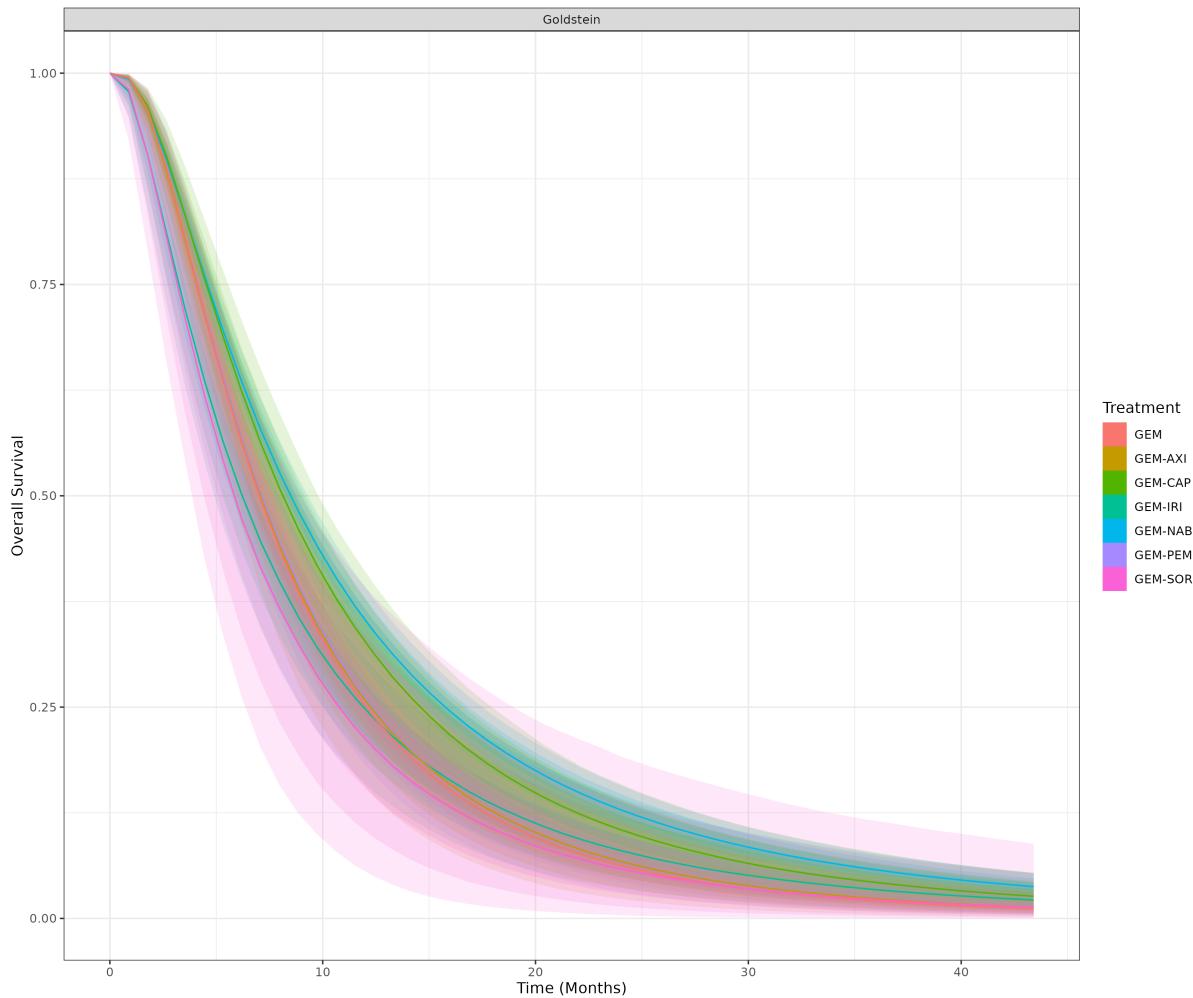


Figure 6.24: OS of each treatment in the Goldstein population

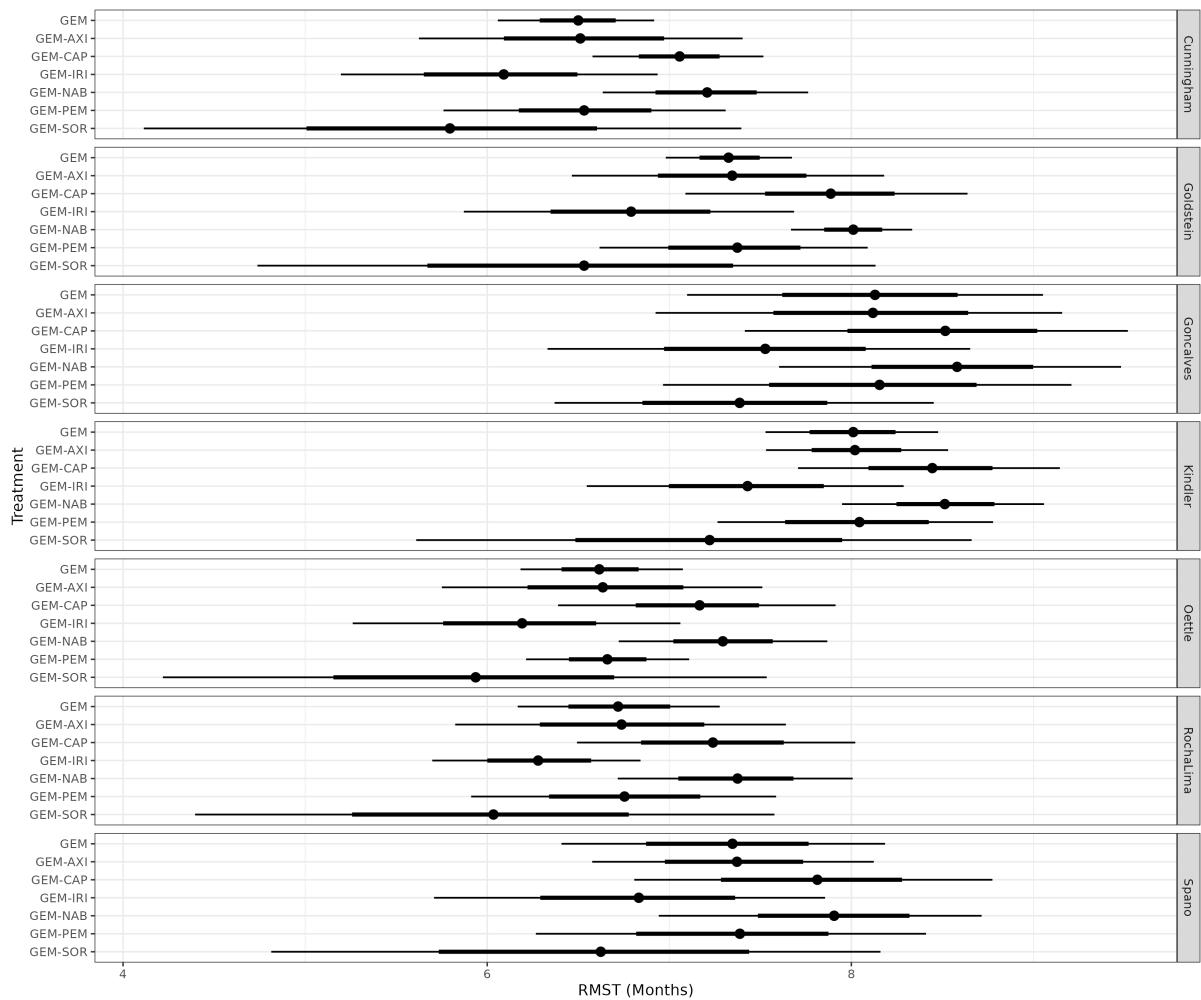


Figure 6.25: RMST of each treatment in each population

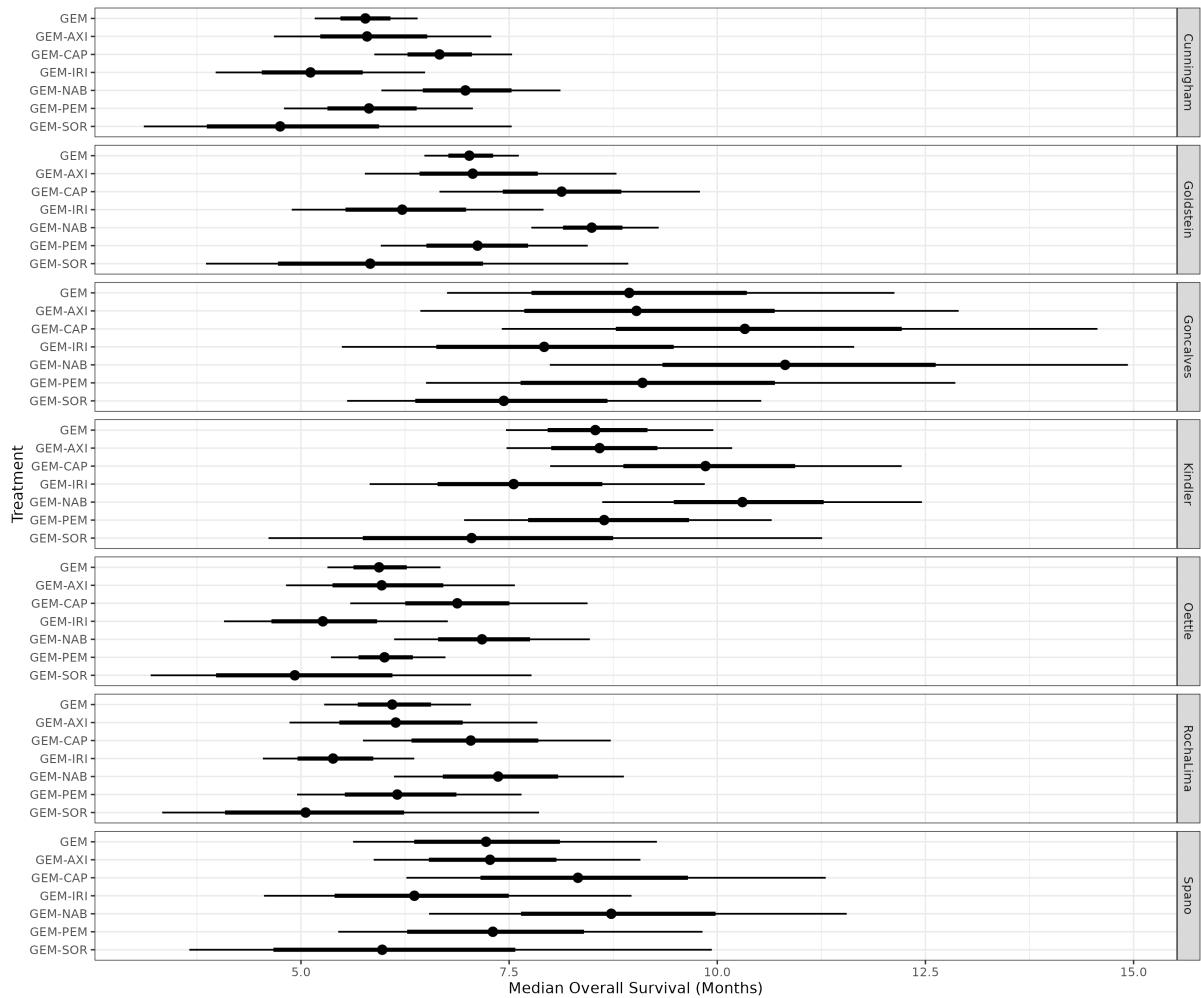


Figure 6.26: Median OS of each treatment in each population

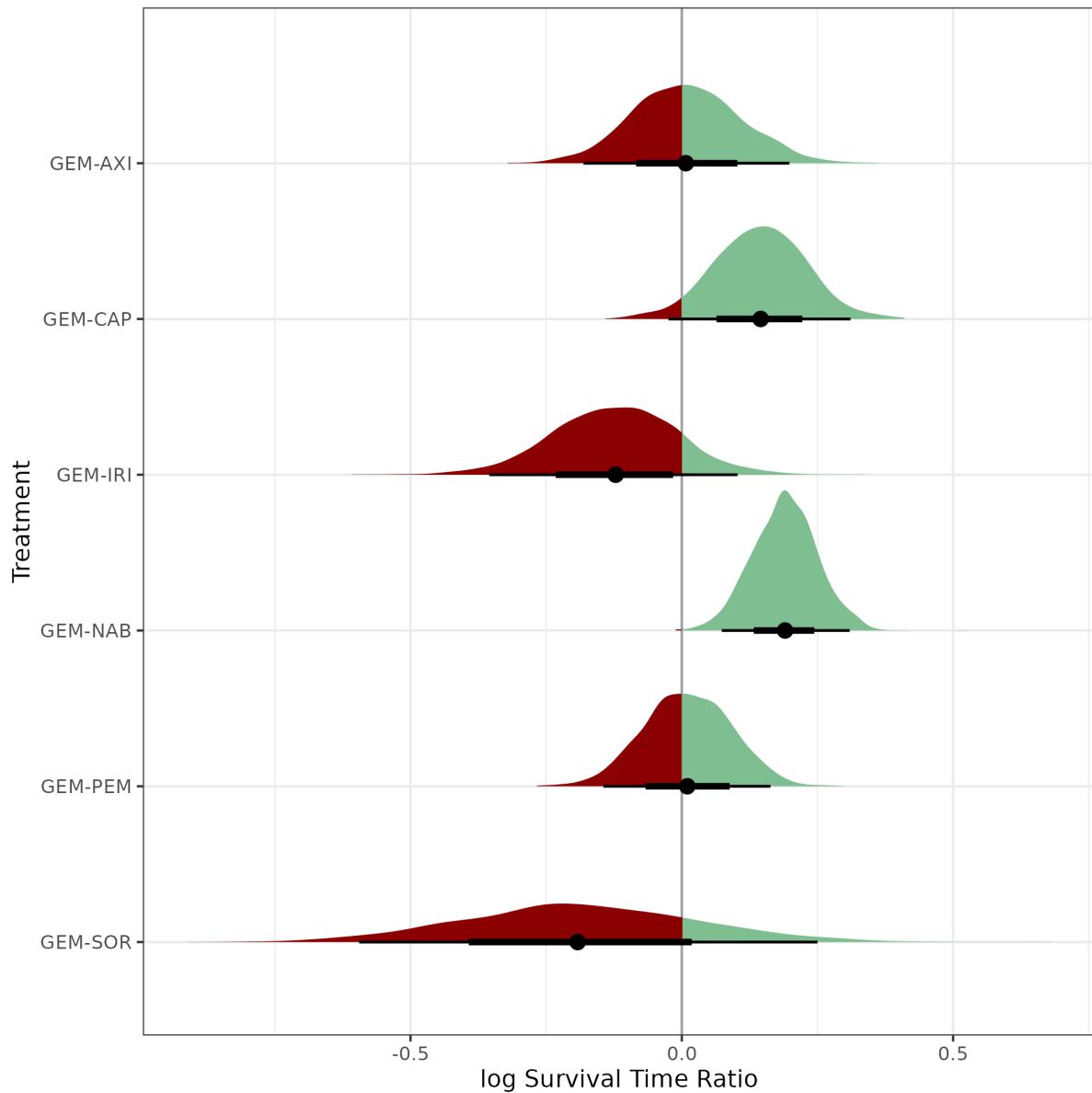


Figure 6.27: Relative treatment effects for all treatments versus GEM

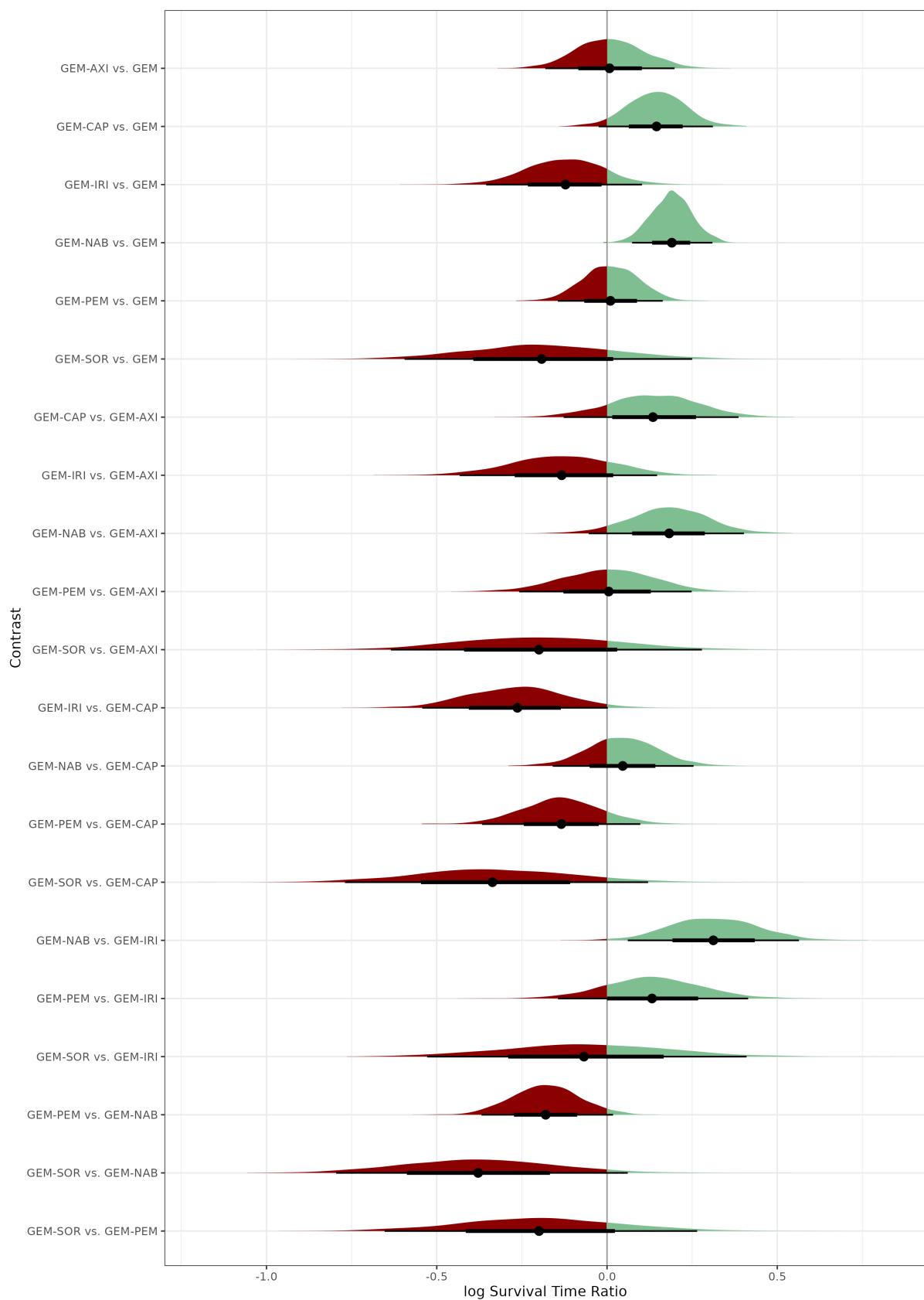


Figure 6.28: Pairwise relative treatment effects for all treatments

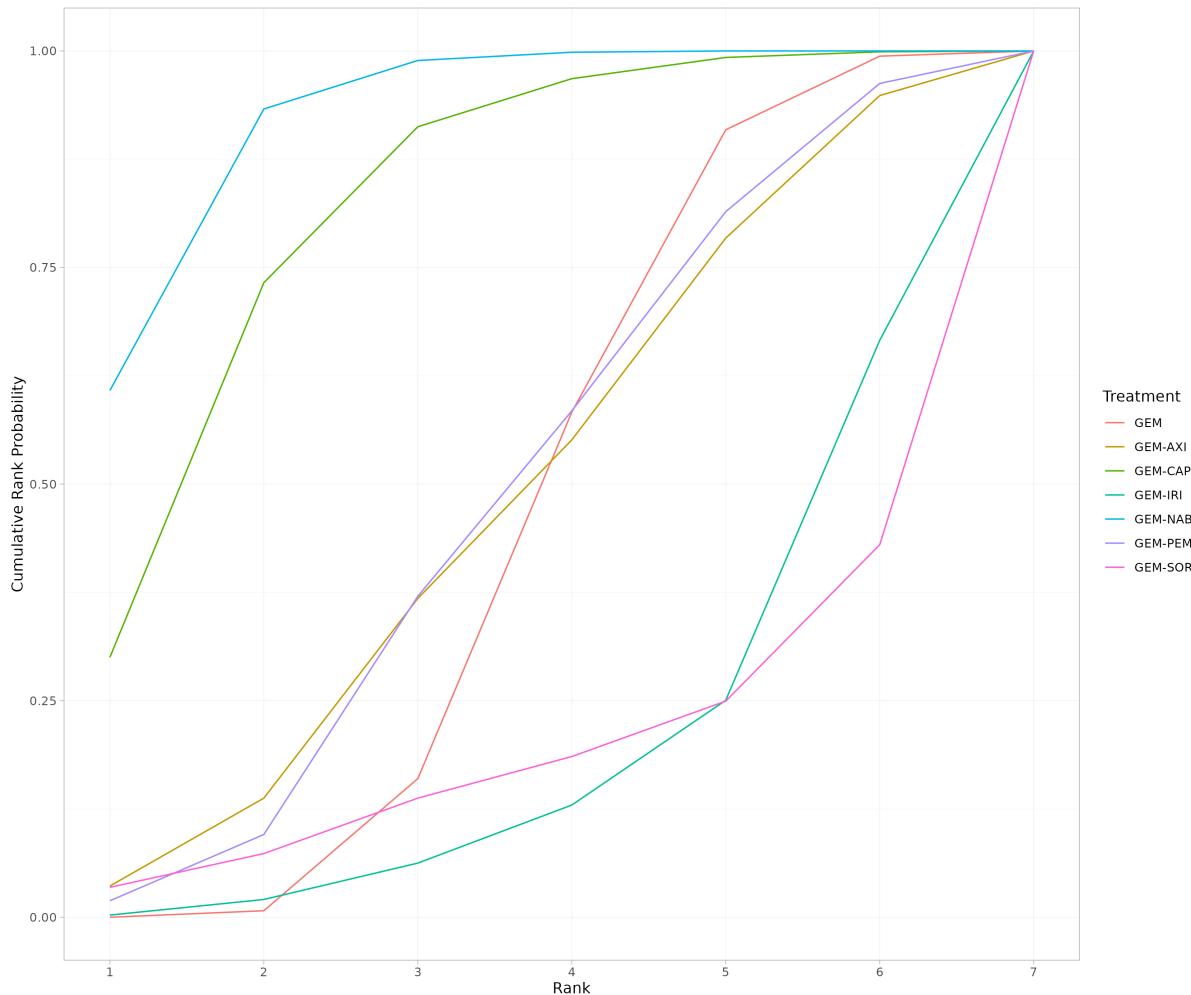


Figure 6.29: Cumulative rank probability for each treatment

6.2.2 Sensitivity Analysis 2

The second SA excluded considered trials for which the OS in both treatment arms dropped below 0.25. By including studies with more mature data, this SA aimed to assess any uncertainty in the base case by removing these trials with immature data. The Goncalves and Kindler studies were therefore removed.

6.2.2.1 Model Fitting and Selection

Figure 6.30 presents the network of evidence for this SA. The differences between this network and the base case were the loss of one GEM → GEM-AXI comparison and the loss of the GEM → GEM-SOR comparison. Given that GEM-SOR was the worst ranked treatment in the base case, dropping this comparison was not deemed to be an issue. The RE log-logistic model scored best in terms of both LOOIC and DIC scores. Figure 6.31 and Figure 6.32 present the trace plots for the FE and RE model respectively. The FE model had better traces for each treatment parameter than the RE model. In addition, Figure 6.19 and Figure 6.20 present the normalised parallel coordinates plot for the FE and RE model, respectively. Unlike the base case and SA 1, the RE model did not have any divergent transitions. In order to determine which model was used, the pairs plots were examined. Figure 6.35 and Figure 6.36 present the pairs plots for the FE and RE models respectively. The shapes of the distributions in the RE model were thinner than the FE model. The FE model was therefore selected as the better fitting model.

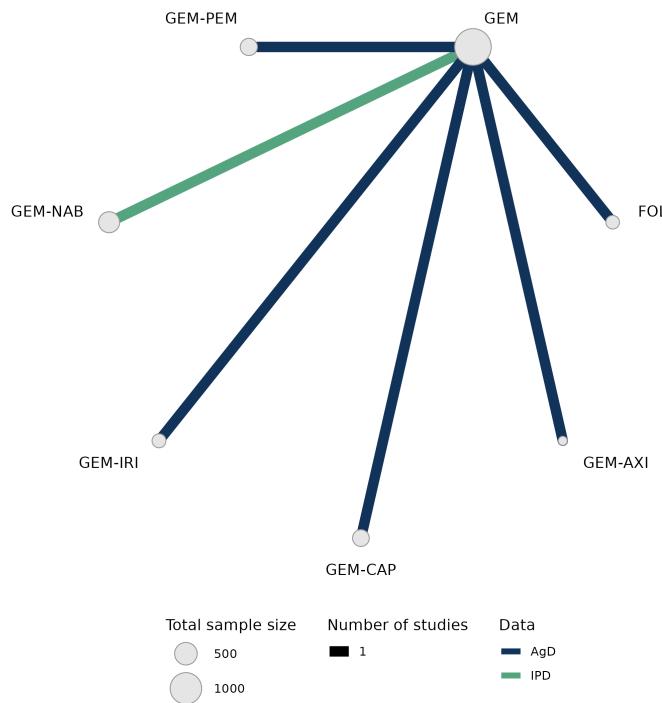


Figure 6.30: Network of evidence

6.2.2.2 Results

Figure 6.37 presents the prior versus posterior distribution for each treatment. The posterior distributions were significantly different from the priors, indicating the data was primary influence on the parameter estimates. The GEM-AXI parameter had a wider posterior than the other treatments.

Figure 6.38 and Figure 6.39 presents the simulated KM curves of each treatment in the Goldstein and Conroy population respectively. FOL had the highest OS in both populations. GEM-NAB had the

Likelihood	Effect	DIC	LOOIC
Log-logistic	Fixed	14970.7099	14968.6730
Log-logistic	Random	14970.2523 ←	14968.1193 ←
Log-normal	Fixed	14971.6948	14973.3886
Log-normal	Random	14972.6324	14973.0334
Weibull	Fixed	14992.7850	14996.2477
Weibull	Random	14992.2891	14995.4664

Table 6.3: Model selection statistics for each model

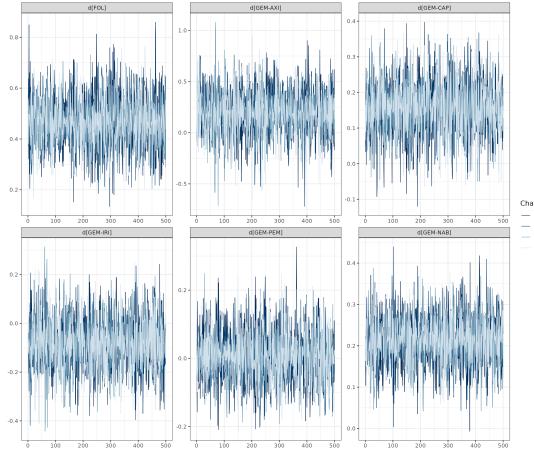


Figure 6.31: Trace plot for the FE log-normal model ML-NMR

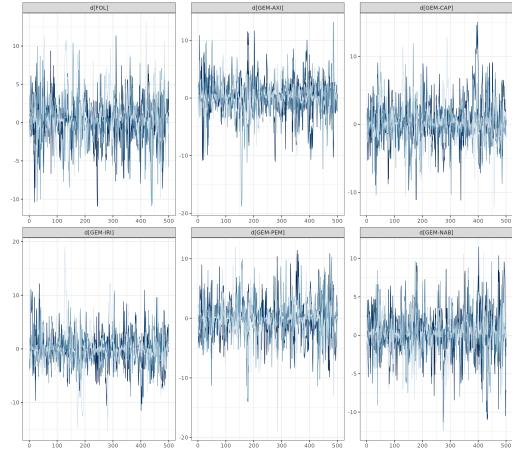


Figure 6.32: Trace plot for the RE log-normal model ML-NMR

second highest OS in both treatment arms. GEM-AXI had a higher OS than GEM-NAB until just before ten months. GEM-IRI had the lowest OS in the Goldstein population until just after ten months, when GEM and GEM-PEM became the worst performing treatment. GEM and GEM-PEM had similar OS curves in the Goldstein population. GEM-IRI had the lowest OS in the Conroy population until around ten months, when GEM and GEM-PEM again become the worst performing treatments.

Figure 6.40 presents the RMST estimates for each treatment in each population. The RMST estimates were more spread out in each population than in the base case and first SA results. FOL had the highest RMST in each population. GEM-NAB had the second highest RMST estimate in each population. GEM-CAP did not have RMST estimates as close to GEM-NAB as in the base case and first SA. GEM-AXI had higher RMST estimates than GEM-CAP in each population, but was noted for having very wide credible intervals. GEM-IRI had the lowest RMST estimate in each population. GEM and GEM-PEM had similar RMST estimates in each population. Figure 6.41 presents the median OS estimates of each treatment in each population. GEM-AXI had similar median OS estimates to GEM-NAB, but did not have higher median OS than GEM-NAB in any treatment population. FOL had the highest median OS in each population. GEM-IRI had the lowest median OS estimates in each population.

Figure 6.42 presents the relative effects of each treatment versus GEM in terms of the log survival time ratio. FOL, GEM-AXI, GEM-CAP, GEM-NAB, and GEM-PEM were all better than GEM. Only FOL and GEM-NAB was significantly better than GEM. GEM-AXI had a higher log survival time ratio than GEM-CAP, and a very similar estimate to GEM-NAB. Figure 6.43 presents the relative effects for all comparisons. This plot showed that GEM-NAB was better than GEM-CAP, but not significantly, which was consistent with the base case analysis. GEM-NAB was not significantly better than GEM-IRI, as it was in the first SA. GEM-IRI and GEM-PEM were significantly worse than FOL. GEM-NAB and GEM-AXI were very similar.

Figure 6.44 presents the cumulative rank probability plot. FOL and GEM-AXI had the highest probabilities of being the best treatment. FOL and GEM-AXI had SUCRA values of 84.2% and 14.3%, respectively. GEM-NAB did not have a higher SUCRA value than GEM-CAP until after the second best treatment. GEM-IRI had the lowest probability of being the fourth to seventh treatment.

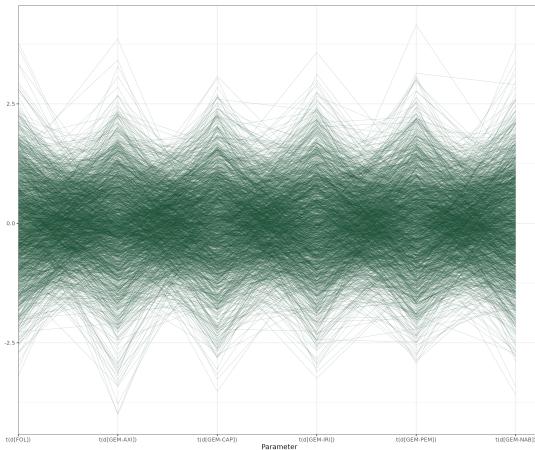


Figure 6.33: Parallel coordinates plot of the FE log-normal model

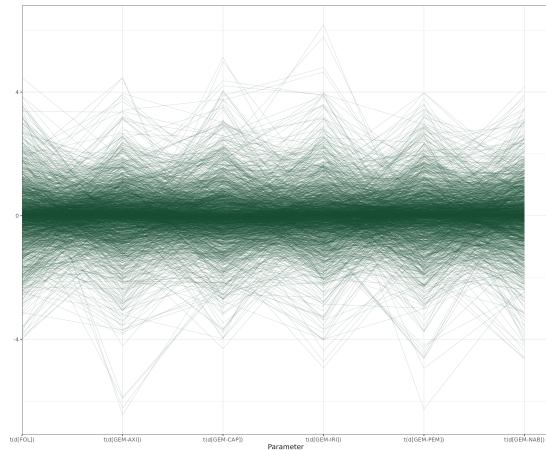


Figure 6.34: Parallel coordinates plot of the RE log-normal model

6.3 Considerations for the ISPOR Good Practice Task Force

ISPOR developed a 26-item questionnaire for assessing the credibility of an NMA [Jansen et al., 2014]. This NMA was performed inline with these practices. While each question is not answered individually here, the themes of the guidance, and how this NMA aligns with it, are discussed. An answer to each question individually is available in Appendix 8.3.

The first set of questions in the guidance concerns the evidence base. This NMA was performed on a fully-connected network of evidence (Figure 6.1), and included no poor-quality studies. Indeed, the study populations and trial characteristics were similar, meaning there was no systematic differences in treatment effect modifiers across the comparisons. The only aspect of this NMA that could be considered not to follow these guidelines was that not all available RCTs were included. The Greshem study, for example included 23 studies obtained by searching several databases. This NMA was not conducted based on results of a systematic literature review or database search. Studies were selected for this NMA based on a brief literature search for trials comparing GEM with another therapy. Since all the KM curves from published papers had to be digitised, which takes a considerable amount of time, there was always to be a limit on how many studies could be included. Since pancreatic cancer treatments have not changed much, the fact the studies included in this NMA were reported between 2004 and 2015 was not deemed to render the results of the NMA inconclusive in today's treatment landscape.

The second set of questions concerns the analysis. No naïve comparisons were made, which preserve within-study randomisation. As there were no cases of both direct and indirect evidence for any treatments, questions eight and nine were not deemed relevant. Question ten concerns imbalance of the distribution of effect modifiers, and how this was accounted for. Since the ML-NMR is a meta-regression model, this was directly accounted for. In terms of FE and RE models, both were fit, and the best fitting model selected in terms of robust selection statistics. Since the studies included in this NMA were not diverse in terms of methodology, FE models were deemed to be clinically appropriate. The guidance generally recommends RE models, but it was deemed clinically appropriate to consider FE models in this case. Were more trials to be included, more consideration would need to be given to the similarity assessment to determine the suitability of FE models.

The third set of questions relates to the reporting quality. While all the studies used, and indeed the associated KM curves were presented, the actual TTE data was not presented. This is due to the form of the data, although it is available within PCNMA R package. Individual study results were provided in Figures 8.1- 6.12. Considerations did not have to be made for direct and indirect comparisons since there were no closed loops. Rankings were reported to address the main project aim, and pairwise comparisons were reported. In particular, the pairwise comparison between GEM-CAP and GEM-NAB was important to clarify the uncertainty mentioned by NICE in NG85. No consideration was given to the effect of important patient characteristics due to the homogeneity in the trial populations and further because of a lack of IPD available for this study. It is not sound for those involved in the study to assess

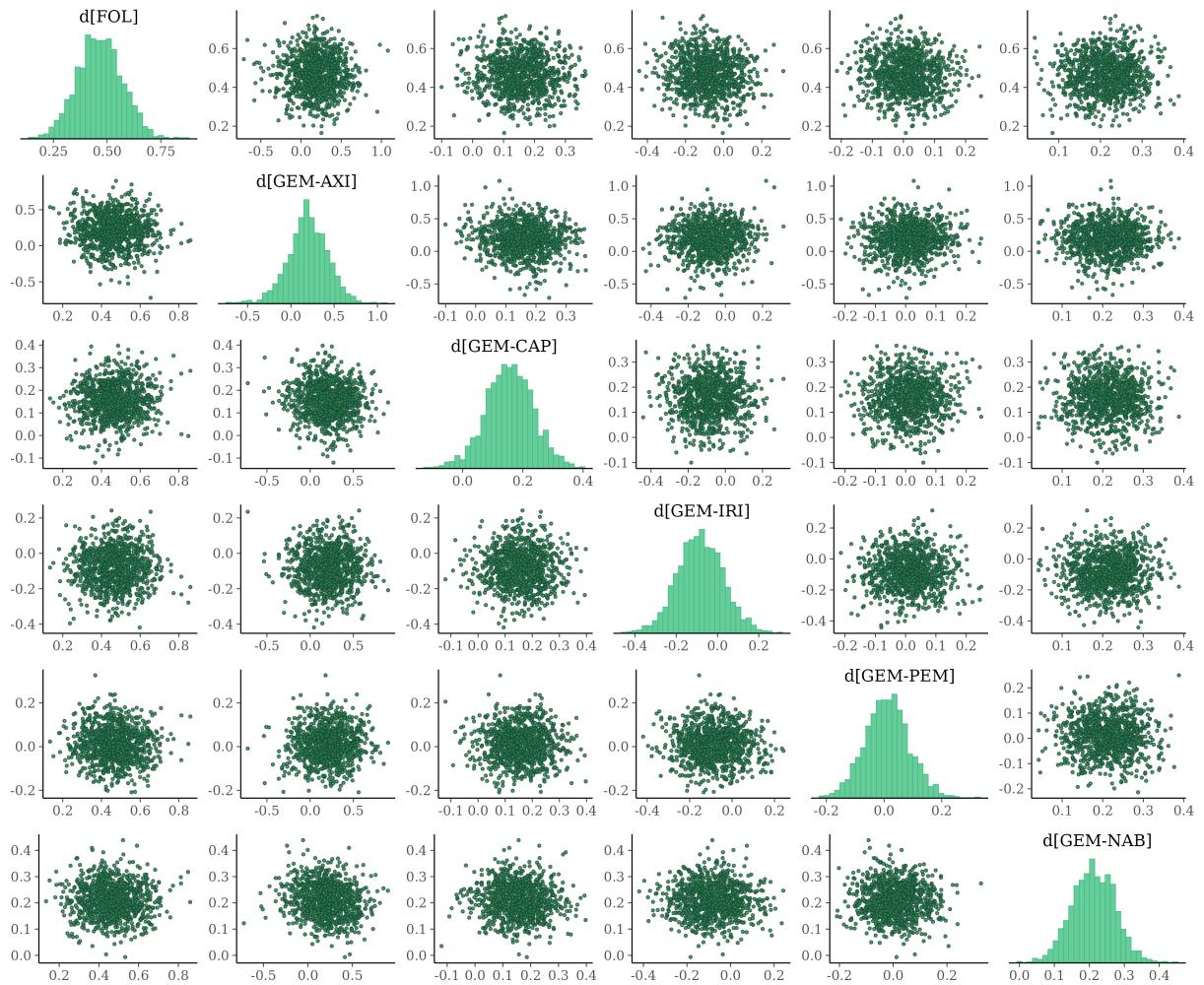


Figure 6.35: Pairs plot for the FE log-normal model

the fairness of the conclusions and interpretation, but every attempt was made to perform this NMA with integrity and interpret the results in line with the evidence.

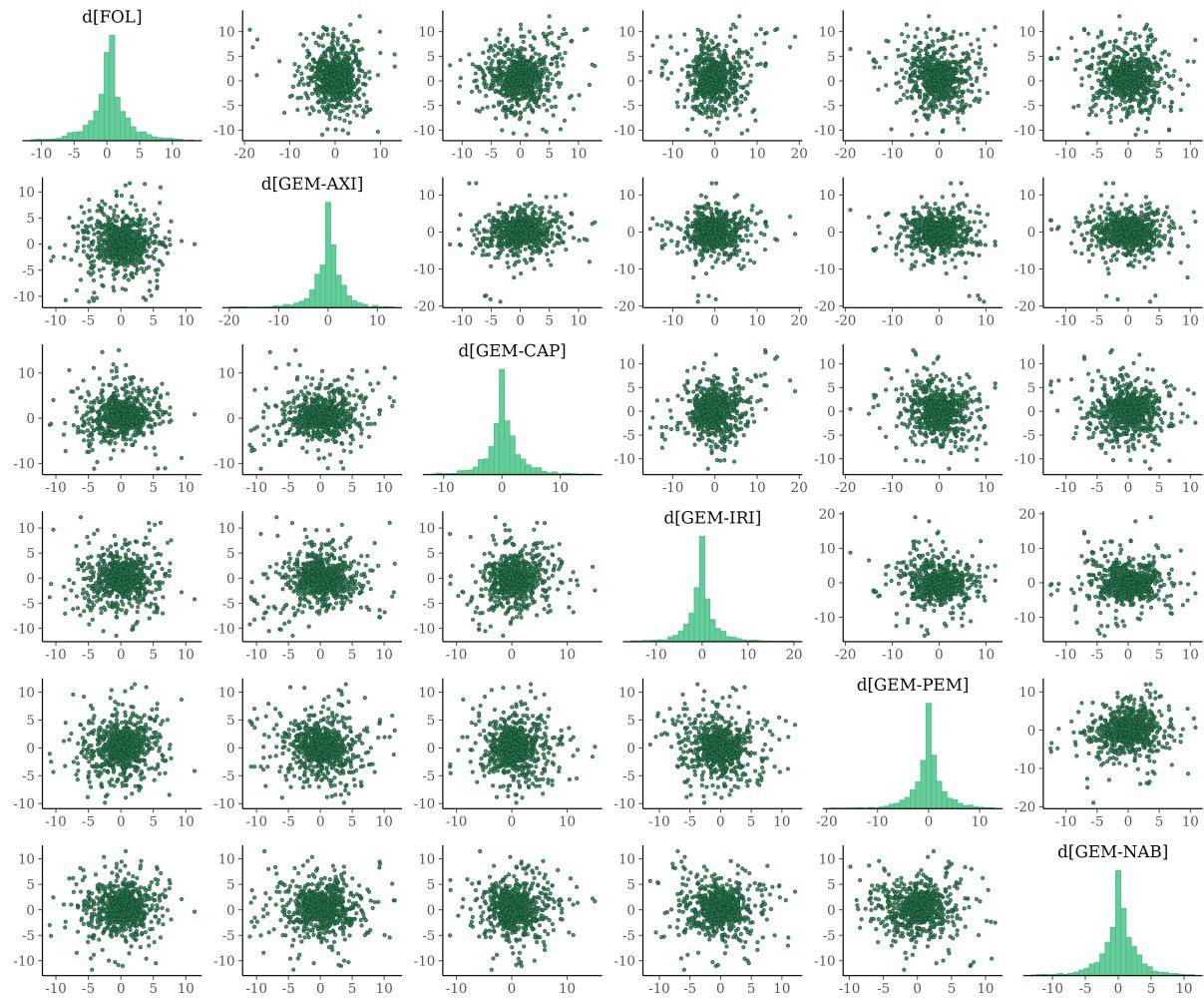


Figure 6.36: Pairs plot for the RE log-normal model

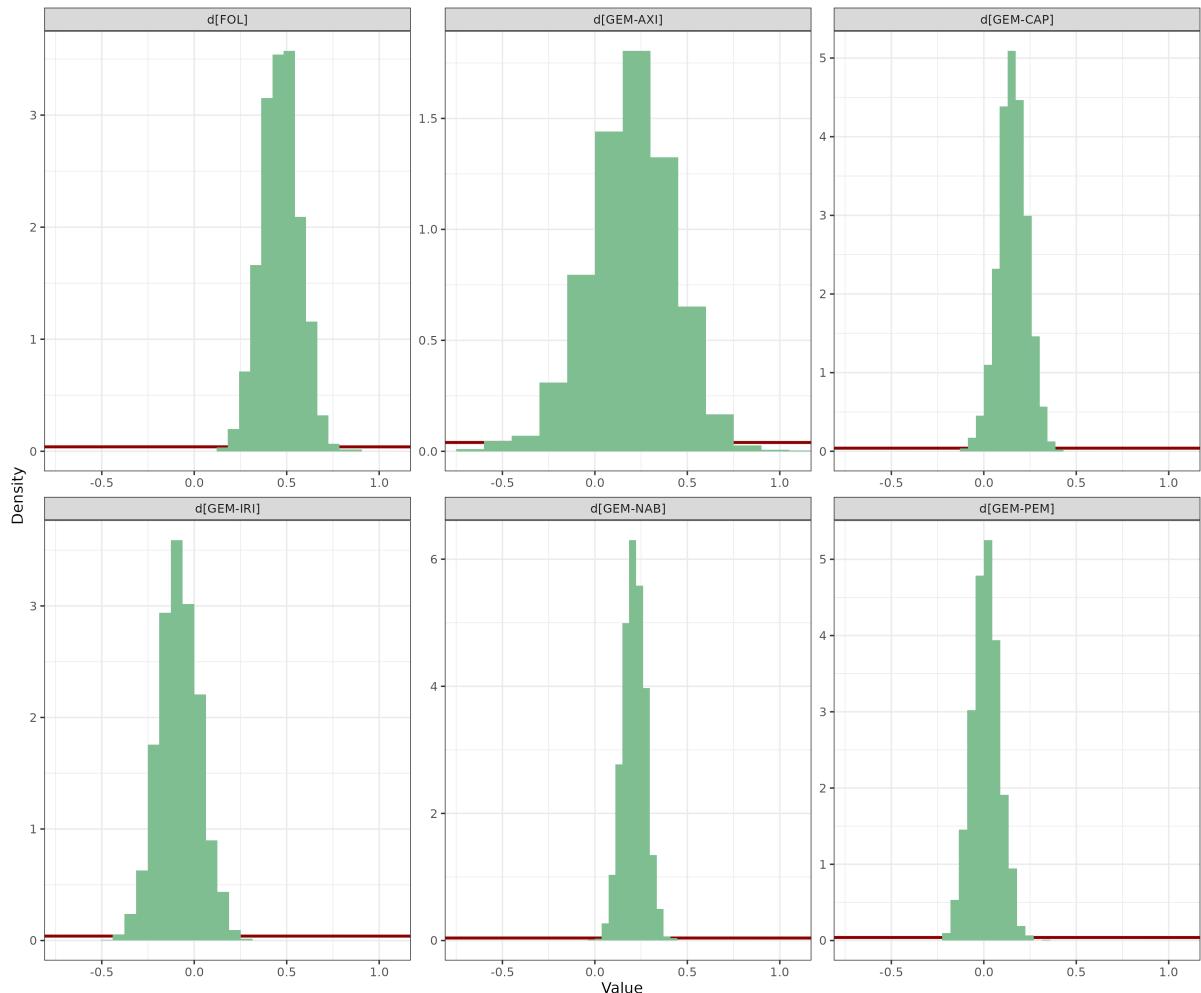


Figure 6.37: Prior versus posterior distribution for each treatment

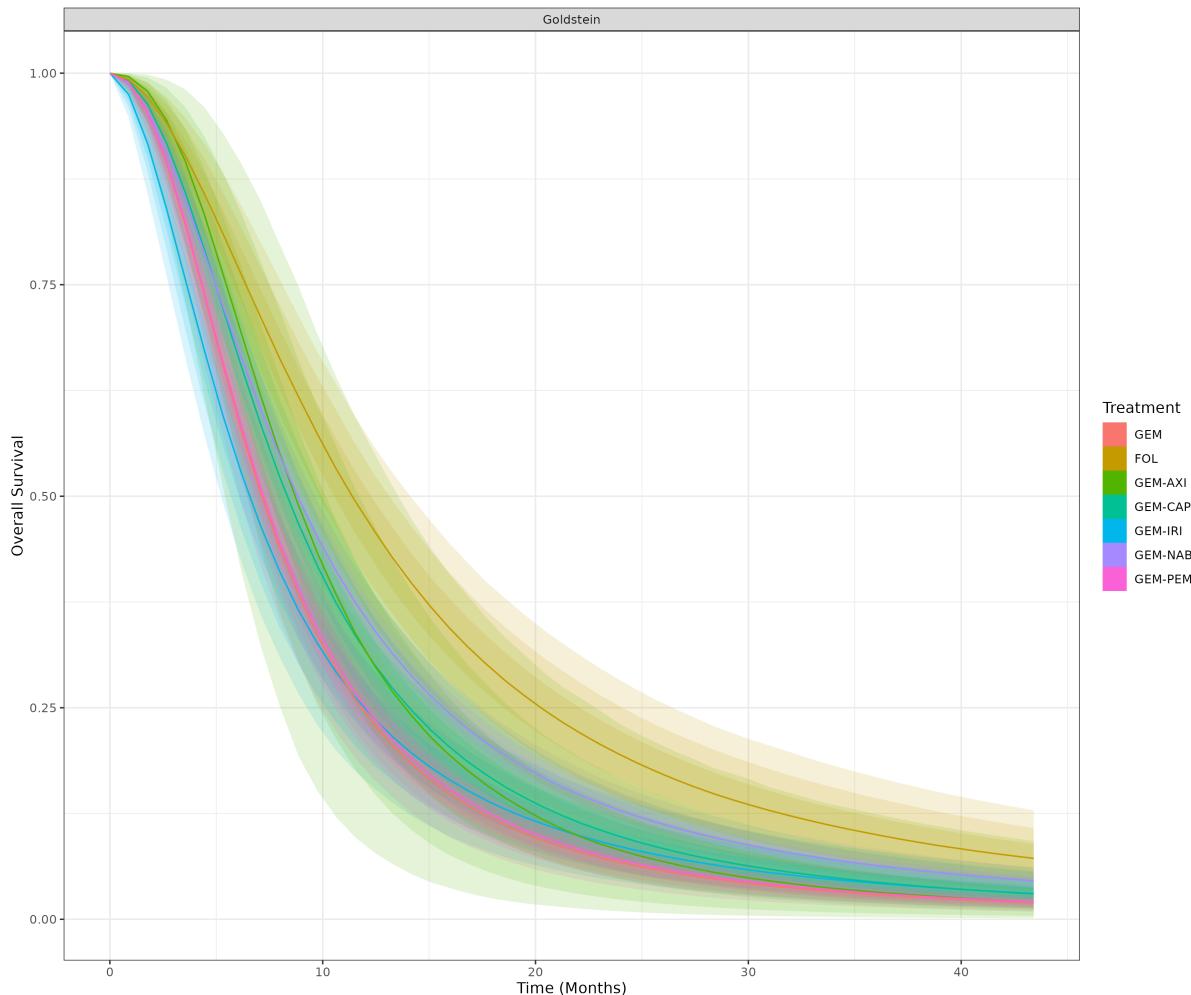


Figure 6.38: OS of each treatment in the Goldstein population

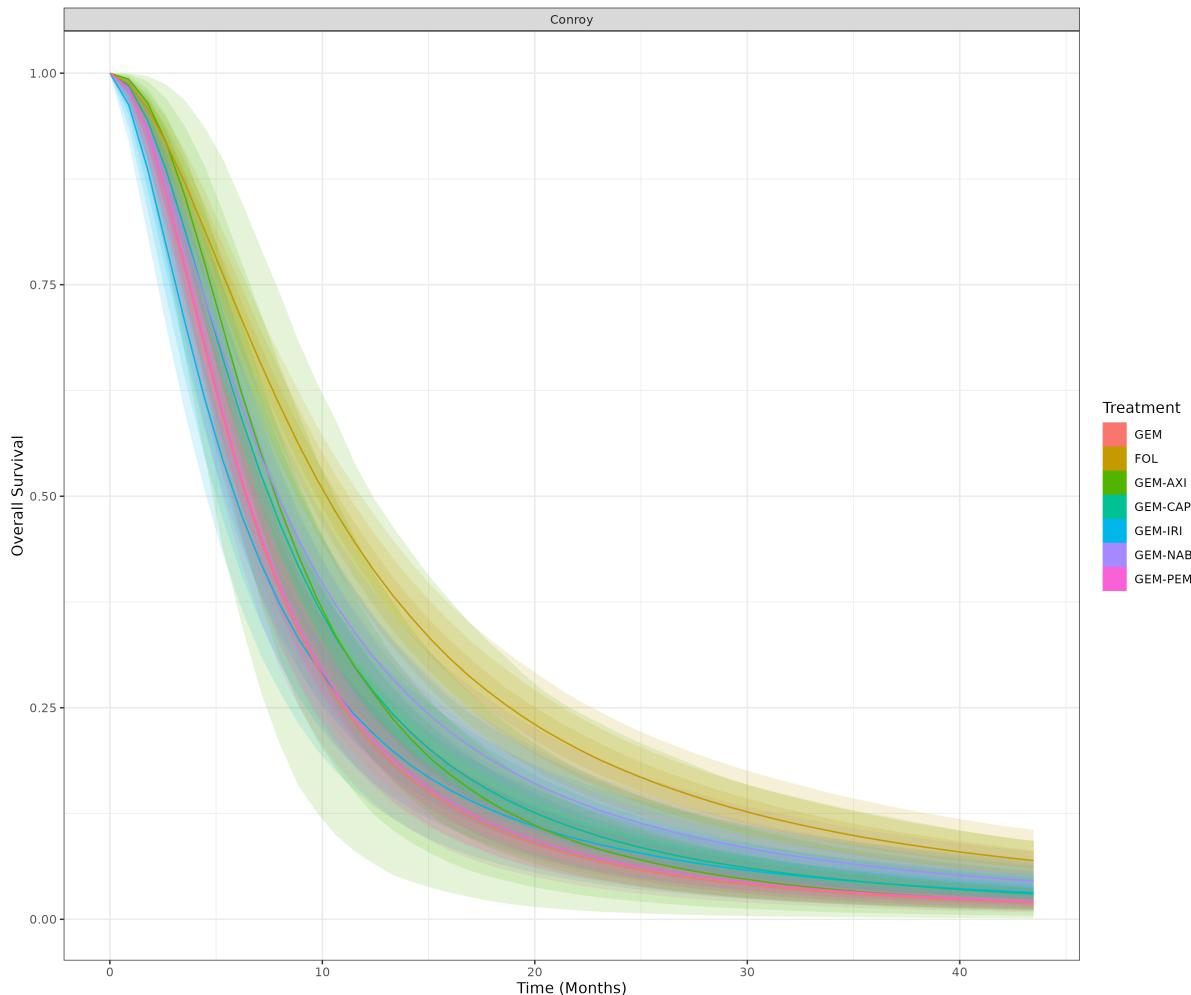


Figure 6.39: OS of each treatment in the Conroy population

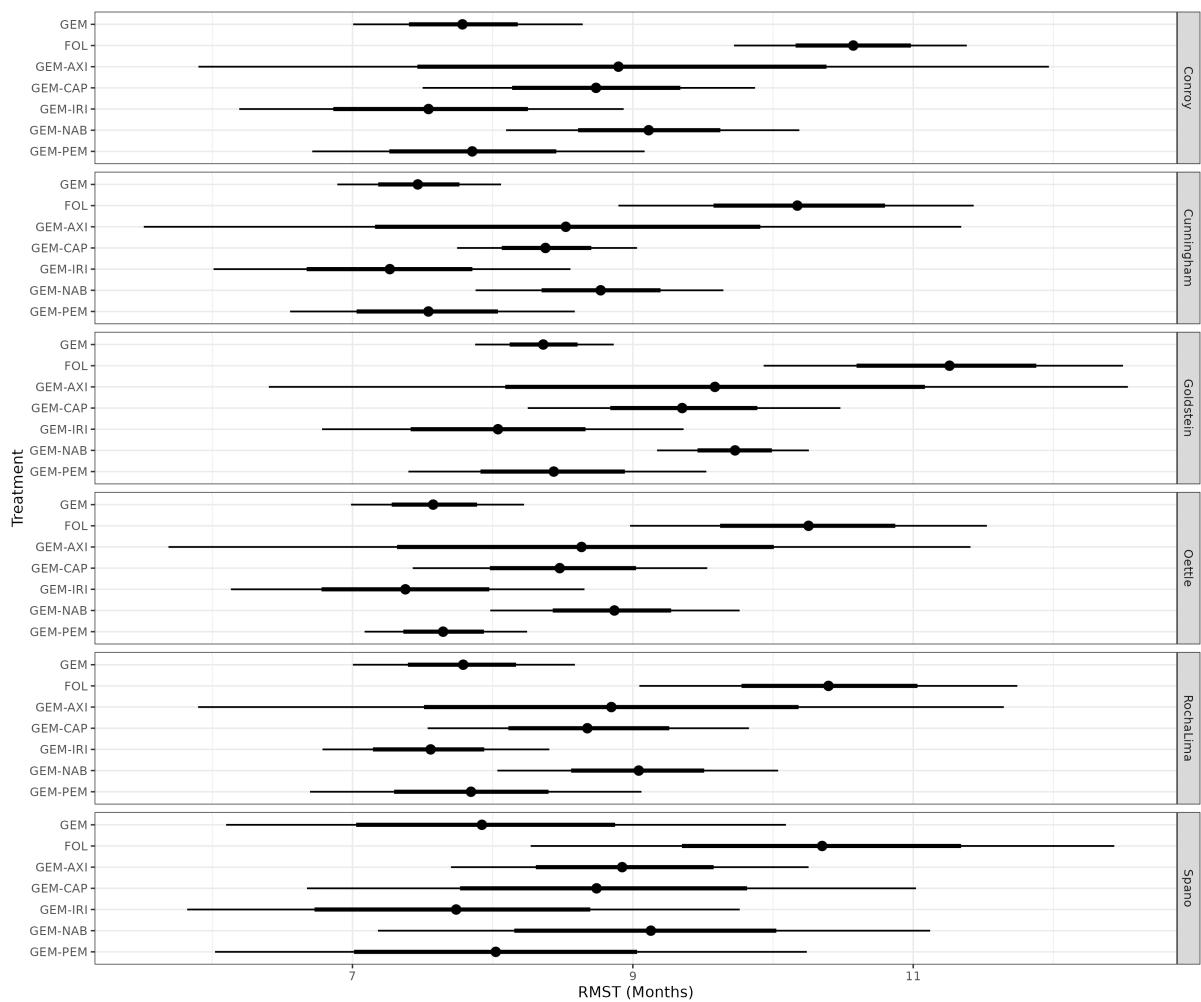


Figure 6.40: RMST of each treatment in each population

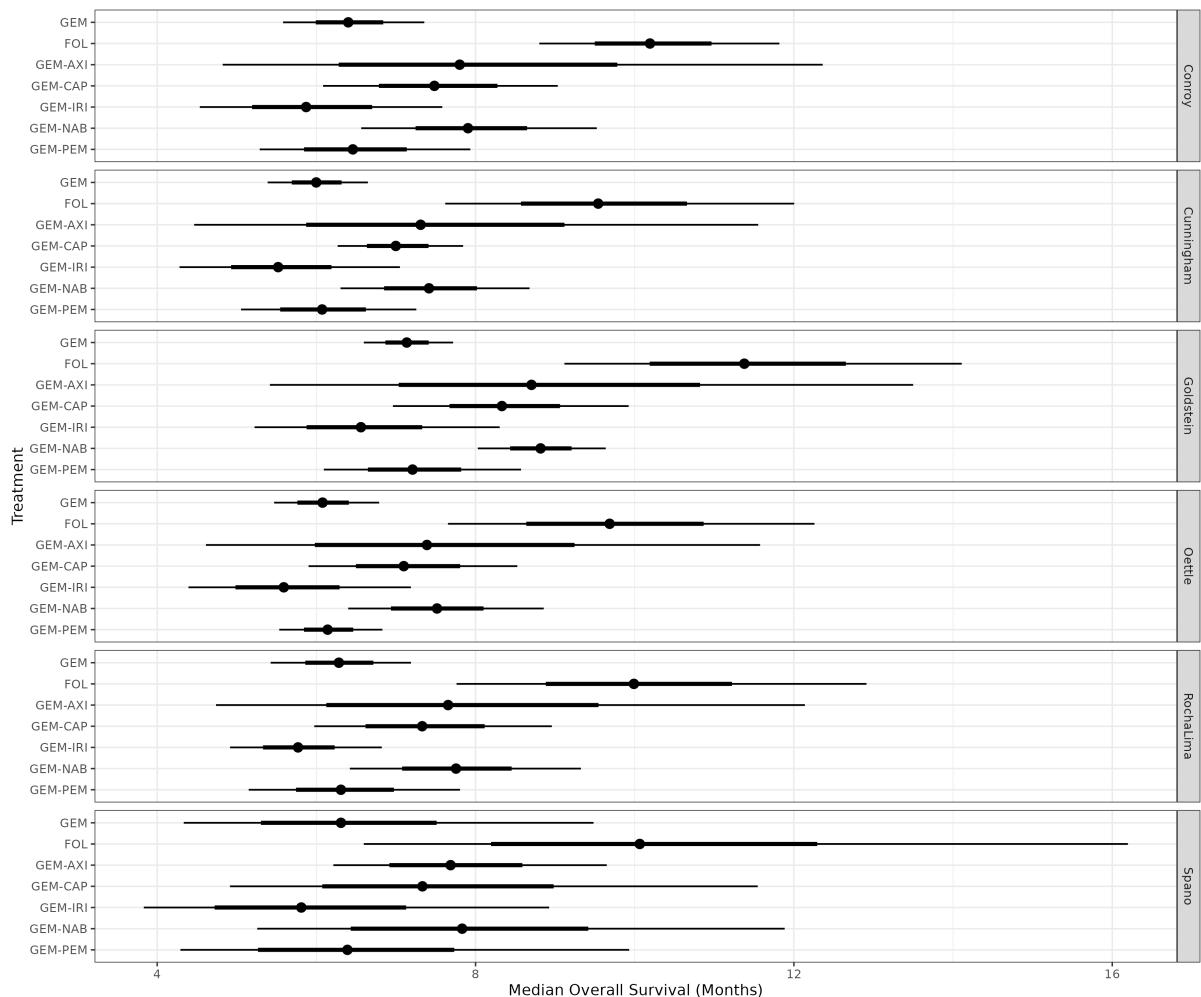


Figure 6.41: Median OS of each treatment in each population

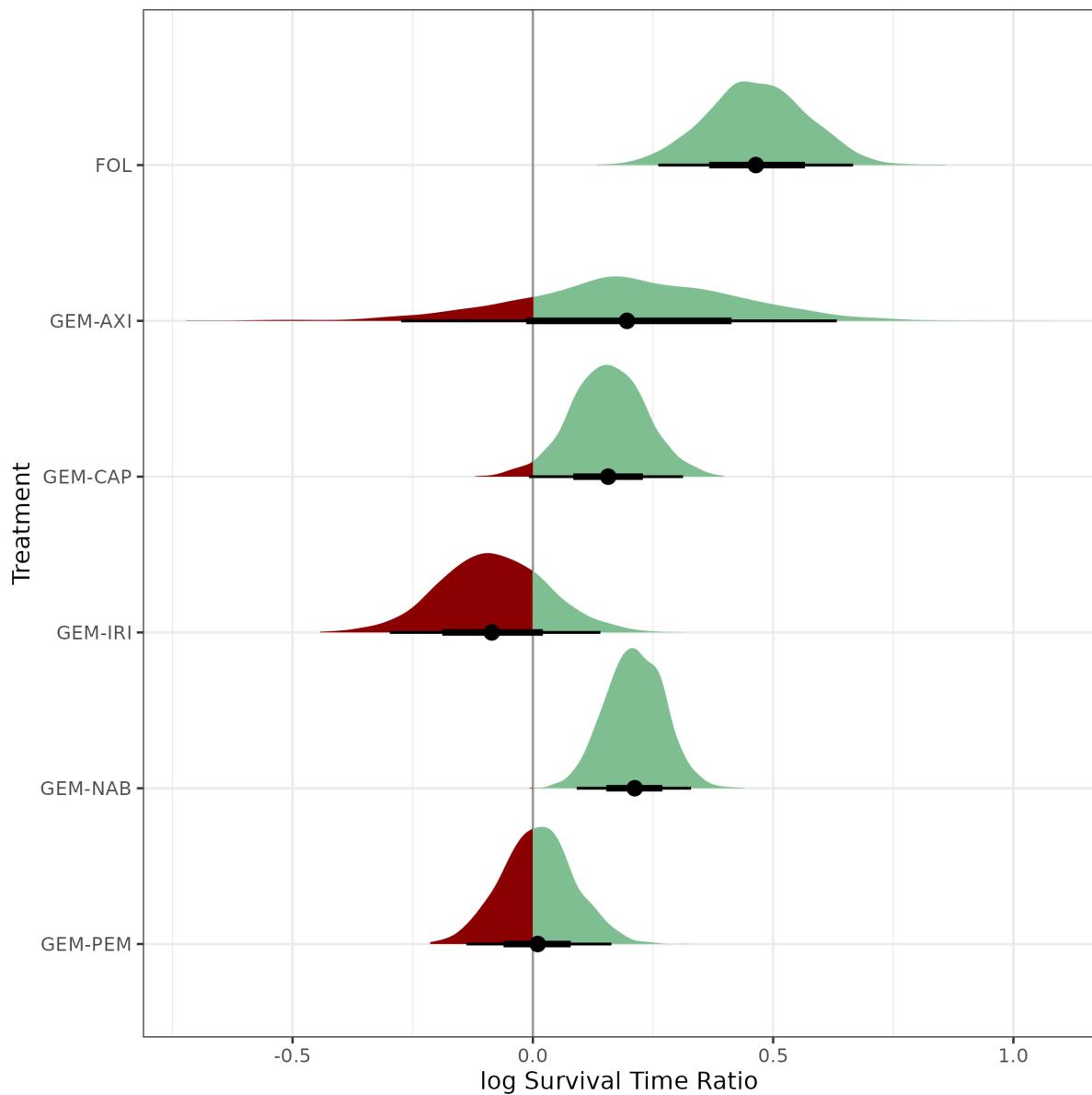


Figure 6.42: Relative treatment effects for all treatments versus GEM

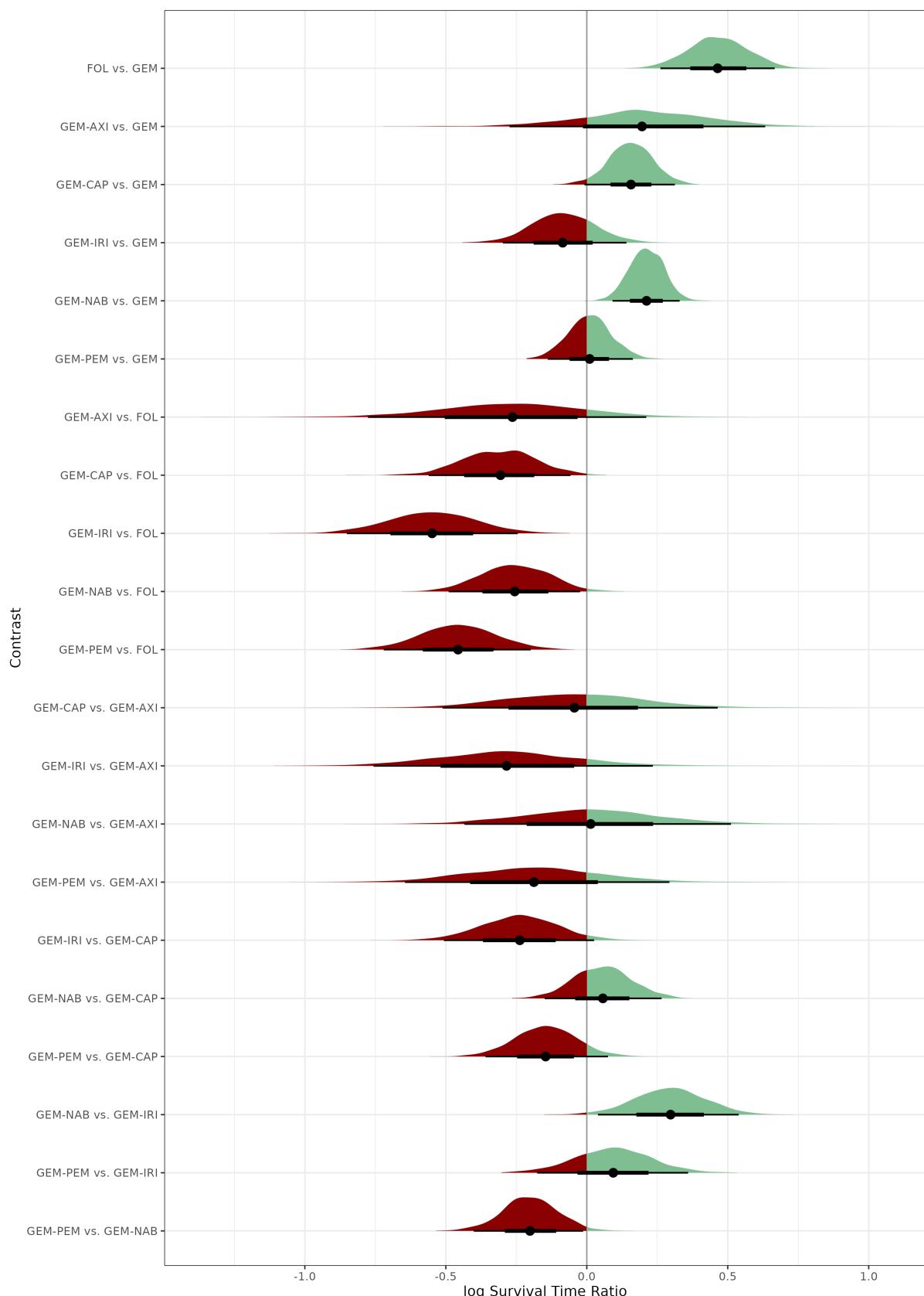


Figure 6.43: Pairwise relative treatment effects for all treatments

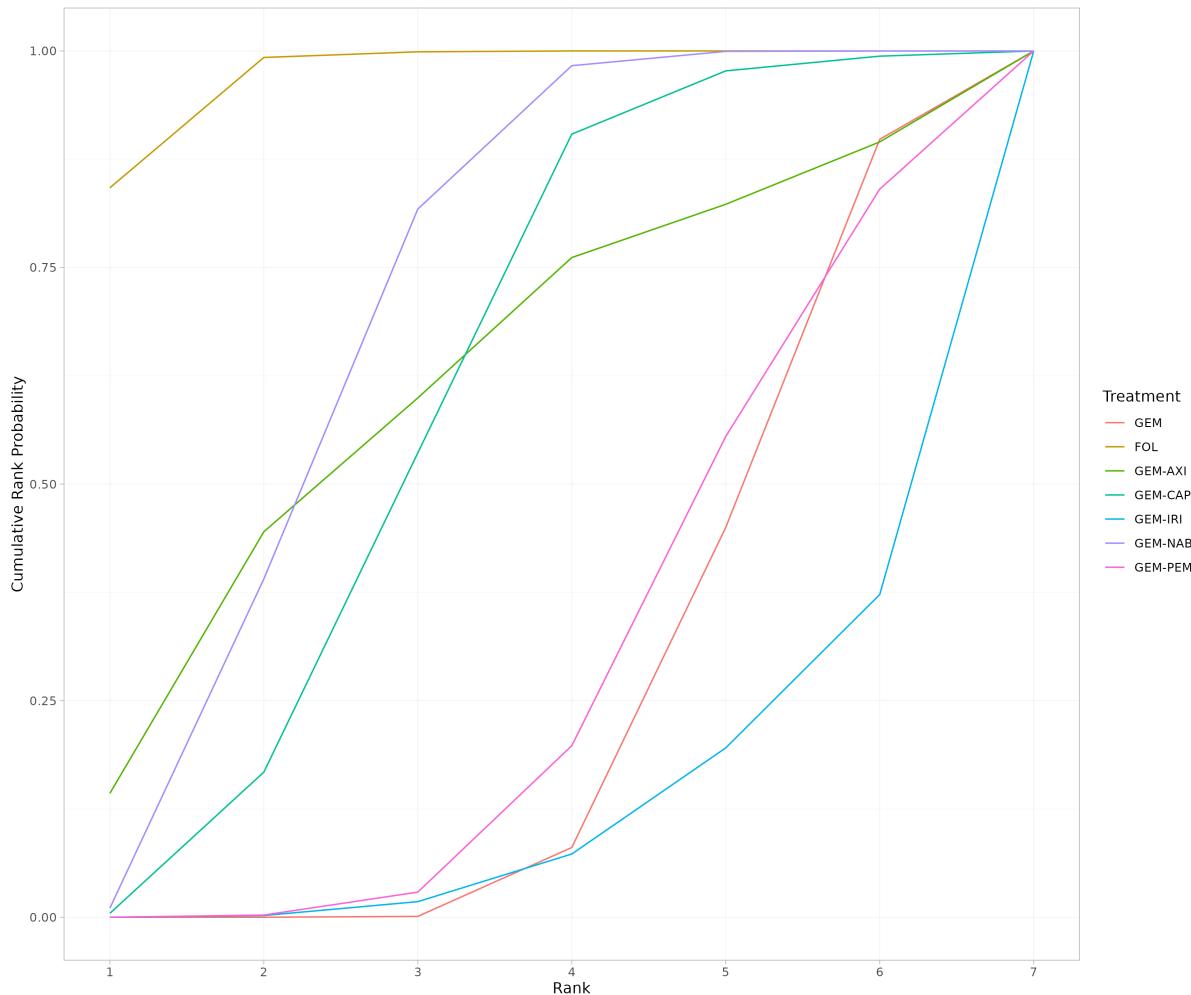


Figure 6.44: Cumulative rank probability for each treatment

CONCLUSION AND DISCUSSIONS

7.1 Conclusion

FOL provides improved OS compared to GEM-NAB and GEM-CAP for the treatment of locally advanced/metastatic pancreatic cancer. FOL outperformed all other treatments in terms of both RMST and median OS in all study populations considered. Based on the survival time ratio, FOL and GEM-NAB provided statistically significant OS benefit compared to GEM. GEM-CAP also provided better OS benefit than GEM, but not significantly. Conversely, GEM-SOR and GEM-IRI provided worse OS benefit compared to GEM, but not significantly worse.

From a clinical perspective, this means FOL should always be considered first for patients who are able to tolerate it, followed by GEM-NAB, and GEM-CAP. GEM should be used for patients who cannot tolerate FOL or GEM-NAB, as GEM-AXI, GEM-IRI, and GEM-PEM do not provide better OS, and may be too toxic to make the treatment worthwhile compared to GEM. While this NMA did not consider toxicity as an outcome, GEM is known to be less toxic than combination therapies.

The second conclusion of this NMA was that GEM-NAB does provide better OS for advanced/metastatic pancreatic cancer than GEM-CAP, but the difference is not significant. This was a consistent conclusion reached in the base case and both sensitivity analyses. The choice of GEM-NAB or GEM-CAP for a patient should be at the discretion of the patient's oncologist.

7.2 Discussion

In 2017, NICE published TA476 [National Institute for Health and Care Excellence, 2017b], which recommended GEM-NAB as an option for untreated metastatic pancreatic cancer in adults only if other combination chemotherapies are unsuitable and the patient would otherwise receive GEM monotherapy. The TA claimed that GEM-NAB was more effective in increasing OS than GEM monotherapy, similarly effective to GEM-CAP, but was less effective than FOL. The comparison to GEM-CAP was stated as "uncertain". This NMA found this to indeed be the case, but with more clarity on the comparison of GEM-NAB and GEM-CAP. GEM-NAB is more effective than GEM-CAP, but not significantly.

While the results of this NMA were conclusively in favour of FOL, the fact that patients in the only trial comparing GEM and FOL were capped at age 76 should not be forgotten. The median age was similar to other studies, and so it is not likely that the age had a significant effect on the performance of FOL, but it would be useful to include an RCT in the network that did not have any exclusion criteria. No such trial appears to exist at present. Two studies: [Matsumoto et al., 2020], and [Chung et al., 2018] were found, but these studies considered patients who had not responded to GEM (Chung, Matsumoto) and GEM or GEM-NAB (Chung). It would therefore not have been appropriate to include these studies in the NMA. This eligibility criteria on age in the Conroy study was not deemed to detract from the clinical validity of these results. One thing that should be considered, which was also pointed out in the discussion of the Nichetti NMA, is that GEM-NAB is administered to a wider patient population. This NMA has proved that the ML-NMR method could plug this gap in the literature if an IPD study was available. If IPD were available, these covariates could easily be included.

It is recommended for any future work to either increase the size of the network of evidence, or include more studies for current comparisons. This NMA is based on fewer assumptions than a standard HR-based survival NMA, but corroborated the findings of several meta analyses, as discussed in Chapter 5. Indeed, FOL, followed by GEM-NAB are clearly the best treatments for advanced/metastatic pancreatic cancer. Based on the results of this NMA and the literature explored in Chapter 5, it is likely that the current treatment landscape for pancreatic cancer is well understood, and there is little to no debate to be had over the best treatments. It is perhaps more wise to invest time and energy on exploring other options for treating this horrible disease, such as surgery. Alternatively, given the number of excess deaths coming from the inability to diagnose pancreatic cancer early, some form of screening process may result in catching cases earlier, giving clinicians more options on how to treat patients. This was alluded to in 1999 by [Dimagno, 1999], as discussed in Chapter 1, and should be given full consideration by NICE and the NHS. This would of course require full economic evaluation.

In terms of the economics of GEM-NAB and FOL, the Incremental Cost Effectiveness Ratio (ICER) for GEM-NAB compared to GEM is between £41,000 and £46,000 per Quality-Adjusted Life Year (QALY) gained [National Institute for Health and Care Excellence, 2017b]. The same NICE TA noted that FOL dominated GEM-NAB¹, making it the preferred treatment. In turn, GEM-CAP dominated GEM-NAB. This NMA casts some doubt on this as GEM-NAB demonstrated improved efficacy to GEM-CAP. A Cost-Effectiveness Model (CEM) of neoadjuvant FOL compared to GEM-NAB from 2021 found the ICER of FOL compared to GEM-NAB to be \$60,856.47 (~ £47,392 based on the July 2024 USD/GBP exchange rate) [Ingram et al., 2022]. A different CEM that took a US perspective and found FOL that the increase in effectiveness associated with FOL was also associated with a \$40,831 increase in the cost when compared to GEM-NAB [Kharat et al., 2021]. The authors found the ICER of FOL compared to GEM-NAB to be \$226,841 (~ £176,614 based on the July 2024 USD/GBP exchange rate), making GEM-NAB a more cost-effective treatment. It should be noted that the Kharat CEM was based on resected pancreatic cancer patients, and the Ingram study considered borderline resectable/locally advanced pancreatic cancer patients, making it more appropriate for this analysis.

The ML-NMR method is clearly a powerful method for conducting population-adjusted NMAs. With R packages like `multinma`, it is easy to implement and conduct statistically and clinically valid analyses. In the second SA, the convergence of the RE model was much better than the analyses that included more immature data, suggesting that the maturity of the survival data may have an impact on the convergence of the models. However, given that the model was only fit with one covariate, it would be wise to assess this in the presence of more IPD. A simulation study may be appropriate for assessing this situation. One thing that was not assessed was the sensitivity of the method to the choice of distribution for the likelihood, but this would also be interesting to assess in the future. The ML-NMR method works well with spline-based likelihood models, but these were not considered here due to the maturity of the data meaning good model fit was able to be obtained with parametric survival models.

Pancreatic cancer remains a nasty disease, despite the advancements in treatments. More focus should be given to national screening programs in the future. Operable pancreatic cancer has much better long term survival, but it is only a “fortunate” few patients who find their cancer early enough to be operable. Given the poor prognosis of pancreatic cancer, even a small improvement in the long-term OS of a couple of months is of immense emotional value to patients and their families when removal of the tumour is not possible. The small improvement in OS given by the treatments evaluated in this NMA should not be overlooked for looking like small improvements compared with treatments for other cancers.

¹i.e., FOL was both more effective, and cheaper

BIBLIOGRAPHY

- [Akaike, 1974] Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, 19(6):716–723.
- [Blackford et al., 2009] Blackford, A., Parmigiani, G., Kensler, T. W., Wolfgang, C., Jones, S., Zhang, X., Parsons, D. W., Lin, J. C.-H., Leary, R. J., Eshleman, J. R., et al. (2009). Genetic mutations associated with cigarette smoking in pancreatic cancer. *Cancer research*, 69(8):3681–3688.
- [Cancer Research UK, 2024] Cancer Research UK (2024). Pancreatic cancer statistics.
- [Chen et al., 2021] Chen, J., Hua, Q., Wang, H., Zhang, D., Zhao, L., Yu, D., Pi, G., Zhang, T., and Lin, Z. (2021). Meta-analysis and indirect treatment comparison of modified folfirinox and gemcitabine plus nab-paclitaxel as first-line chemotherapy in advanced pancreatic cancer. *BMC cancer*, 21:1–9.
- [Chung et al., 2018] Chung, M. J., Kang, H., Kim, H. G., Hyun, J. J., Lee, J. K., Lee, K. H., Noh, M. H., Kang, D. H., Lee, S. H., Bang, S., et al. (2018). Multicenter phase ii trial of modified folfirinox in gemcitabine-refractory pancreatic cancer. *World journal of gastrointestinal oncology*, 10(12):505.
- [Conroy et al., 2011] Conroy, T., Desseigne, F., Ychou, M., Bouché, O., Guimbaud, R., Bécouarn, Y., Adenis, A., Raoul, J.-L., Gourgou-Bourgade, S., De La Fouchardière, C., et al. (2011). Folfirinox versus gemcitabine for metastatic pancreatic cancer. *New England journal of medicine*, 364(19):1817–1825.
- [Cox, 1972] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202.
- [Cunningham et al., 2009] Cunningham, D., Chau, I., Stocken, D. D., Valle, J. W., Smith, D., Steward, W., Harper, P. G., Dunn, J., Tudur-Smith, C., West, J., et al. (2009). Phase iii randomized comparison of gemcitabine versus gemcitabine plus capecitabine in patients with advanced pancreatic cancer. *Journal of Clinical Oncology*, 27(33):5513–5518.
- [Deeb et al., 2015] Deeb, A., Haque, S.-U., and Olowokure, O. (2015). Pulmonary metastases in pancreatic cancer, is there a survival influence? *Journal of gastrointestinal oncology*, 6(3):E48.
- [Dimagno, 1999] Dimagno, E. P. (1999). Pancreatic cancer: clinical presentation, pitfalls and early clues. *Annals of oncology*, 10:S140–S142.
- [Ding et al., 2018] Ding, J., Tarokh, V., and Yang, Y. (2018). Model selection techniques: An overview. *IEEE Signal Processing Magazine*, 35(6):16–34.
- [Goldstein et al., 2015] Goldstein, D., El-Maraghi, R. H., Hammel, P., Heinemann, V., Kunzmann, V., Sastre, J., Scheithauer, W., Siena, S., Tabernero, J., Teixeira, L., et al. (2015). nab-paclitaxel plus gemcitabine for metastatic pancreatic cancer: long-term survival from a phase iii trial. *Journal of the National Cancer Institute*, 107(2):dju413.
- [Gonçalves et al., 2012] Gonçalves, A., Gilabert, M., François, E., Dahan, L., Perrier, H., Lamy, R., Re, D., Largillier, R., Gasmi, M., Tchiknavorian, X., et al. (2012). Baypan study: a double-blind phase iii randomized trial comparing gemcitabine plus sorafenib and gemcitabine plus placebo in patients with advanced pancreatic cancer. *Annals of oncology*, 23(11):2799–2805.
- [Gresham et al., 2014] Gresham, G. K., Wells, G. A., Gill, S., Cameron, C., and Jonker, D. J. (2014). Chemotherapy regimens for advanced pancreatic cancer: a systematic review and network meta-analysis. *BMC cancer*, 14:1–13.

- [Guyot et al., 2012] Guyot, P., Ades, A., Ouwens, M. J., and Welton, N. J. (2012). Enhanced secondary analysis of survival data: reconstructing the data from published kaplan-meier survival curves. *BMC medical research methodology*, 12:1–13.
- [Hryciuk et al., 2018] Hryciuk, B., Szymanowski, B., Romanowska, A., Salt, E., Wasag, B., Grala, B., Jassem, J., and Duchnowska, R. (2018). Severe acute toxicity following gemcitabine administration: A report of four cases with cytidine deaminase polymorphisms evaluation. *Oncology Letters*, 15(2):1912–1916.
- [Ingram et al., 2022] Ingram, M. A., Lauren, B. N., Pumpalova, Y., Park, J., Lim, F., Bates, S. E., Kastrinos, F., Manji, G. A., Kong, C. Y., and Hur, C. (2022). Cost-effectiveness of neoadjuvant folfirinox versus gemcitabine plus nab-paclitaxel in borderline resectable/locally advanced pancreatic cancer patients. *Cancer Reports*, 5(9):e1565.
- [Jackson, 2016] Jackson, C. H. (2016). flexsurv: a platform for parametric survival modeling in r. *Journal of statistical software*, 70.
- [Jansen et al., 2014] Jansen, J. P., Trikalinos, T., Cappelleri, J. C., Daw, J., Andes, S., Eldessouki, R., and Salanti, G. (2014). Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ispor-amcp-npc good practice task force report. *Value in Health*, 17(2):157–173.
- [Kelsen et al., 1997] Kelsen, D. P., Portenoy, R., Thaler, H., Tao, Y., and Brennan, M. (1997). Pain as a predictor of outcome in patients with operable pancreatic carcinoma. *Surgery*, 122(1):53–59.
- [Kharat et al., 2021] Kharat, A. A., Nelson, R., Au, T., and Biskupiak, J. (2021). Cost-effectiveness analysis of folfirinox vs gemcitabine with nab-paclitaxel as adjuvant treatment for resected pancreatic cancer in the united states based on prodige-24 and apact trials. *Journal of Managed Care & Specialty Pharmacy*, 27(10):1367–1375.
- [Kindler et al., 2011] Kindler, H. L., Ioka, T., Richel, D. J., Bennouna, J., Létourneau, R., Okusaka, T., Funakoshi, A., Furuse, J., Park, Y. S., Ohkawa, S., et al. (2011). Axitinib plus gemcitabine versus placebo plus gemcitabine in patients with advanced pancreatic adenocarcinoma: a double-blind randomised phase 3 study. *The lancet oncology*, 12(3):256–262.
- [Latimer, 2011] Latimer, N. (2011). Nice dsu technical support document 14: survival analysis for economic evaluations alongside clinical trials-extrapolation with patient-level data. *Report by the Decision Support Unit*.
- [Lin et al., 2019] Lin, K.-I., Yang, J.-L., Lin, Y.-C., Chou, C.-Y., Chen, J.-H., and Hung, C.-C. (2019). Network meta-analysis of efficacy and safety of chemotherapy and target therapy in the first-line setting of advanced pancreatic cancer. *Cancers*, 11(11):1746.
- [Matsumoto et al., 2020] Matsumoto, T., Kurioka, Y., Okazaki, U., Matsuo, Y., Kimura, S., Miura, K., Tsuduki, T., Takagi, S., Takatani, M., and Morishita, H. (2020). Folfirinox for advanced pancreatic cancer patients after nab-paclitaxel plus gemcitabine failure. *Pancreas*, 49(4):574–578.
- [Mbuagbaw et al.,] Mbuagbaw, L., Rochwerg, B., Jaeschke, R., Heels-Andsell, D., Alhazzani, W., Thabane, L., and Guyatt, G. Approaches to interpreting and choosing the best treatments in network meta-analyses. syst rev 2017; 6: 79.
- [National Institute for Health and Care Excellence, 2017a] National Institute for Health and Care Excellence (2017a). Guidance on the use of gemcitabine for the treatment of pancreatic cancer.
- [National Institute for Health and Care Excellence, 2017b] National Institute for Health and Care Excellence (2017b). Paclitaxel as albumin-bound nanoparticles with gemcitabine for untreated metastatic pancreatic cancer.
- [National Institute for Health and Care Excellence, 2018] National Institute for Health and Care Excellence (2018). Pancreatic cancer in adults: diagnosis and management.
- [Neoptolemos et al., 2010] Neoptolemos, J., Urrutia, R., Abbruzzese, J., and Büchler, M. (2010). *Pancreatic Cancer*. 2010 Springer E-Books. Springer.

- [Nichetti et al., 2024] Nichetti, F., Rota, S., Ambrosini, P., Pircher, C., Gusmaroli, E., Busset, M. D. D., Pusceddu, S., Sposito, C., Coppa, J., Morano, F., et al. (2024). Nalirifox, folfirinox, and gemcitabine with nab-paclitaxel as first-line chemotherapy for metastatic pancreatic cancer: a systematic review and meta-analysis. *JAMA Network Open*, 7(1):e2350756–e2350756.
- [Oettle et al., 2005] Oettle, H., Richards, D., Ramanathan, R., Van Laethem, J.-L., Peeters, M., Fuchs, M., Zimmermann, A., John, W., Von Hoff, D., Arning, M., et al. (2005). A phase iii trial of pemetrexed plus gemcitabine versus gemcitabine in patients with unresectable or metastatic pancreatic cancer. *Annals of Oncology*, 16(10):1639–1645.
- [Pancreatic Cancer UK, 2022] Pancreatic Cancer UK (2022). Signs and symptoms of pancreatic cancer.
- [Phillippo, 2019] Phillippe, D. (2019). Calibration of treatment effects in network meta-analysis using individual patient data [phd thesis]. university of bristol; 2019.
- [Phillippo et al., 2020] Phillippe, D. M., Dias, S., Ades, A., Belger, M., Brnabic, A., Schacht, A., Saure, D., Kadziola, Z., and Welton, N. J. (2020). Multilevel network meta-regression for population-adjusted treatment comparisons. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 183(3):1189–1210.
- [Phillippo et al., 2024] Phillippe, D. M., Dias, S., Welton, N. J., and Ades, A. (2024). Multilevel network meta-regression for general likelihoods: synthesis of individual and aggregate data with applications to survival analysis. *arXiv preprint arXiv:2401.12640*.
- [Pishvaian and Brody, 2017] Pishvaian, M. J. and Brody, J. R. (2017). Therapeutic implications of molecular subtyping for pancreatic cancer. *Oncology (Williston Park, NY)*, 31(3):159–66.
- [Rocha Lima et al., 2004] Rocha Lima, C. M., Green, M. R., Rotche, R., Miller Jr, W. H., Jeffrey, G. M., Cisar, L. A., Morganti, A., Orlando, N., Gruia, G., and Miller, L. L. (2004). Irinotecan plus gemcitabine results in no survival advantage compared with gemcitabine monotherapy in patients with locally advanced or metastatic pancreatic cancer despite increased tumor response rate. *Journal of Clinical Oncology*, 22(18):3776–3783.
- [Rohatgi, 2022] Rohatgi, A. (2022). Webplotdigitizer: Version 4.6.
- [Schwarz, 1978] Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, pages 461–464.
- [Sklar, 1959] Sklar, M. (1959). Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, volume 8, pages 229–231.
- [Smith and Glass, 1977] Smith, M. L. and Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American psychologist*, 32(9):752.
- [Sobol', 1967] Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4):784–802.
- [Spano et al., 2008] Spano, J.-P., Chodkiewicz, C., Maurel, J., Wong, R., Wasan, H., Barone, C., Létourneau, R., Bajetta, E., Pithavala, Y., Bycott, P., et al. (2008). Efficacy of gemcitabine plus axitinib compared with gemcitabine alone in patients with advanced pancreatic cancer: an open-label randomised phase ii study. *The Lancet*, 371(9630):2101–2108.
- [Takumoto et al., 2022] Takumoto, Y., Sasahara, Y., Narimatsu, H., and Akazawa, M. (2022). Comparative outcomes of first-line chemotherapy for metastatic pancreatic cancer among the regimens used in japan: a systematic review and network meta-analysis. *JAMA Network Open*, 5(1):e2145515–e2145515.
- [Vehtari et al., 2024] Vehtari, A., Gabry, J., Magnusson, M., Yao, Y., Bürkner, P.-C., Paananen, T., and Gelman, A. (2024). loo: Efficient leave-one-out cross-validation and waic for bayesian models. R package version 2.7.0.
- [Vehtari et al., 2017] Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, 27:1413–1432.
- [Wickham and Henry, 2023] Wickham, H. and Henry, L. (2023). *purrr: Functional Programming Tools*. R package version 1.0.2, <https://github.com/tidyverse/purrr>.

[Wong et al., 2009] Wong, A., Soo, R. A., Yong, W.-P., and Innocenti, F. (2009). Clinical pharmacology and pharmacogenetics of gemcitabine. *Drug metabolism reviews*, 41(2):77–88.

[Zhang et al., 2018] Zhang, S.-H., Liu, G.-F., Li, X.-F., Liu, L., and Yu, S.-N. (2018). Efficacy of different chemotherapy regimens in treatment of advanced or metastatic pancreatic cancer: a network meta-analysis. *Journal of Cellular Physiology*, 233(4):3352–3374.

CHAPTER
EIGHT

APPENDICES

8.1 Appendix A: Additional NMA Results

This appendix presents some additional NMA results for the FE log-normal ML-NMR model.

8.1.1 Base Case

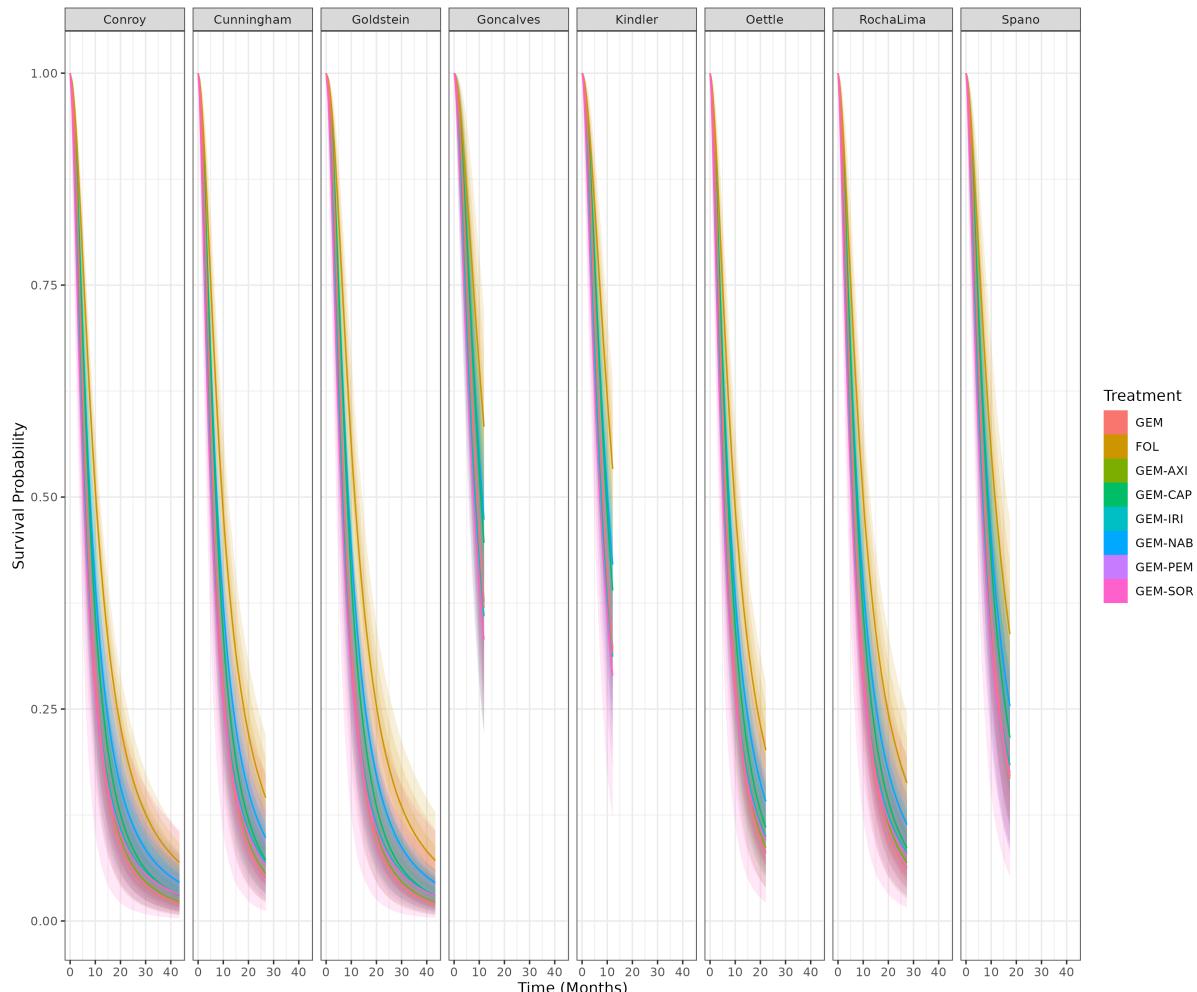


Figure 8.1: OS of each treatment in each population

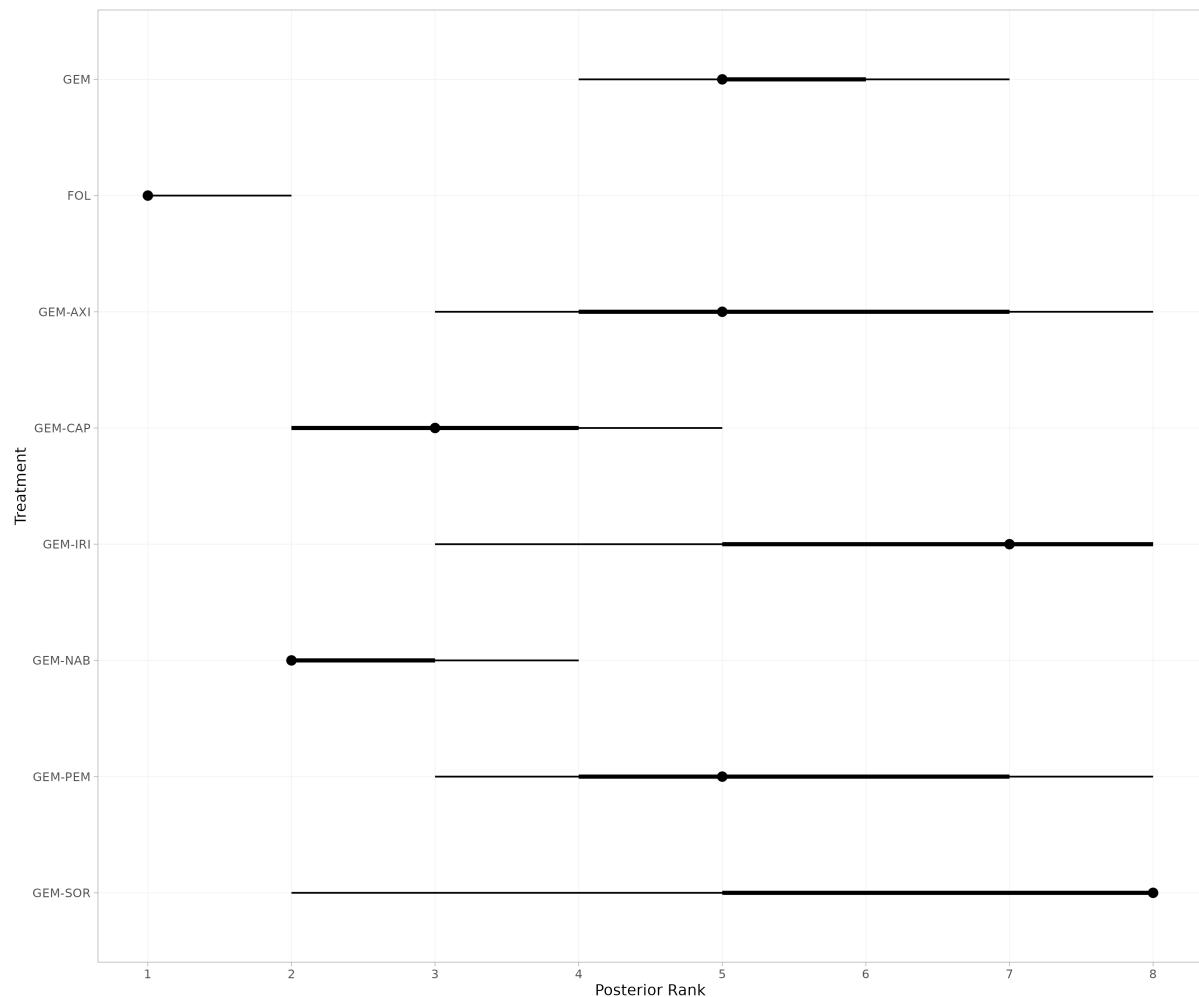


Figure 8.2: Rank probabilities for each treatment

8.1.2 Sensitivity Analysis 1

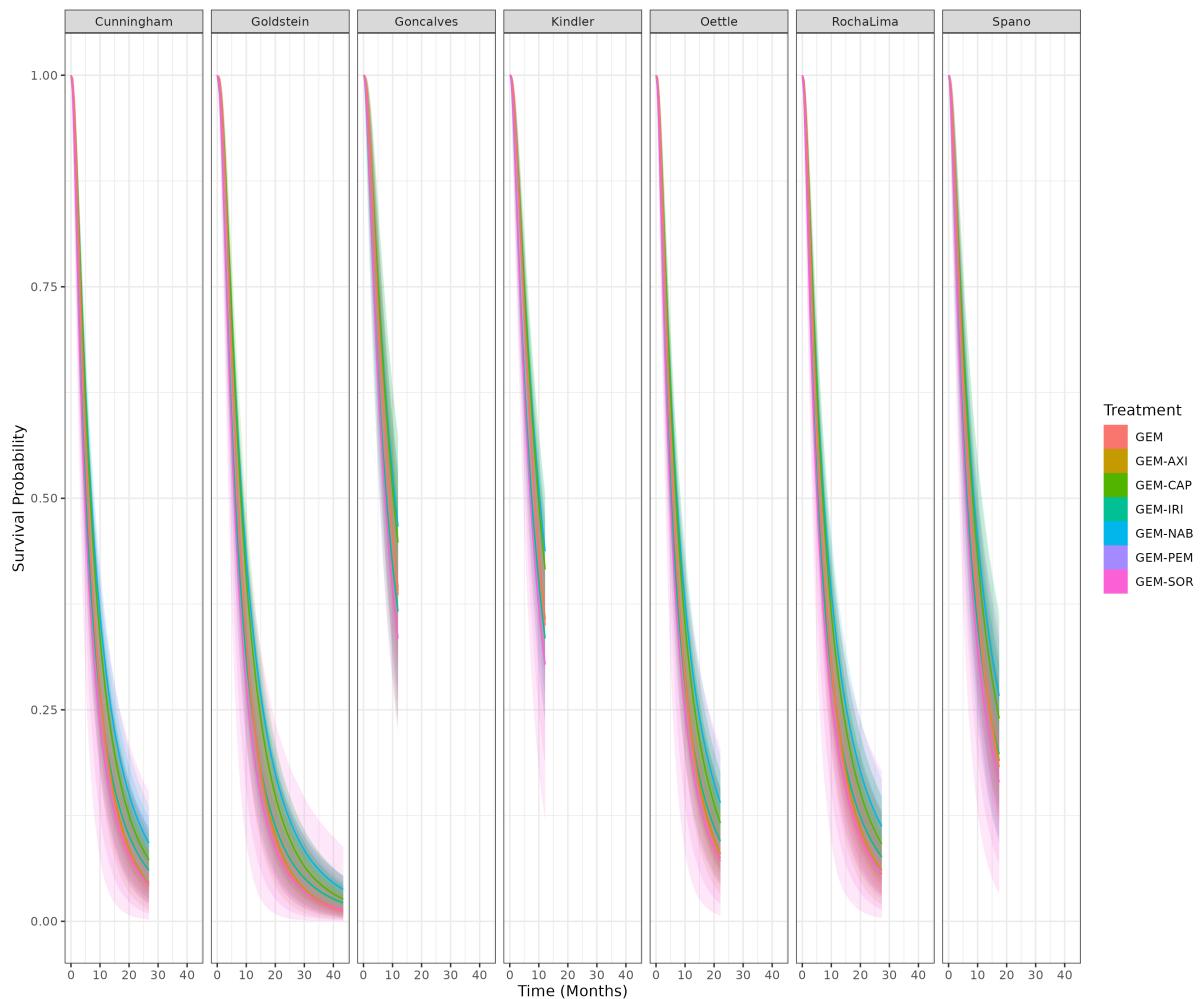


Figure 8.3: OS of each treatment in each population - SA1

8.2 Sensitivity Analysis 2

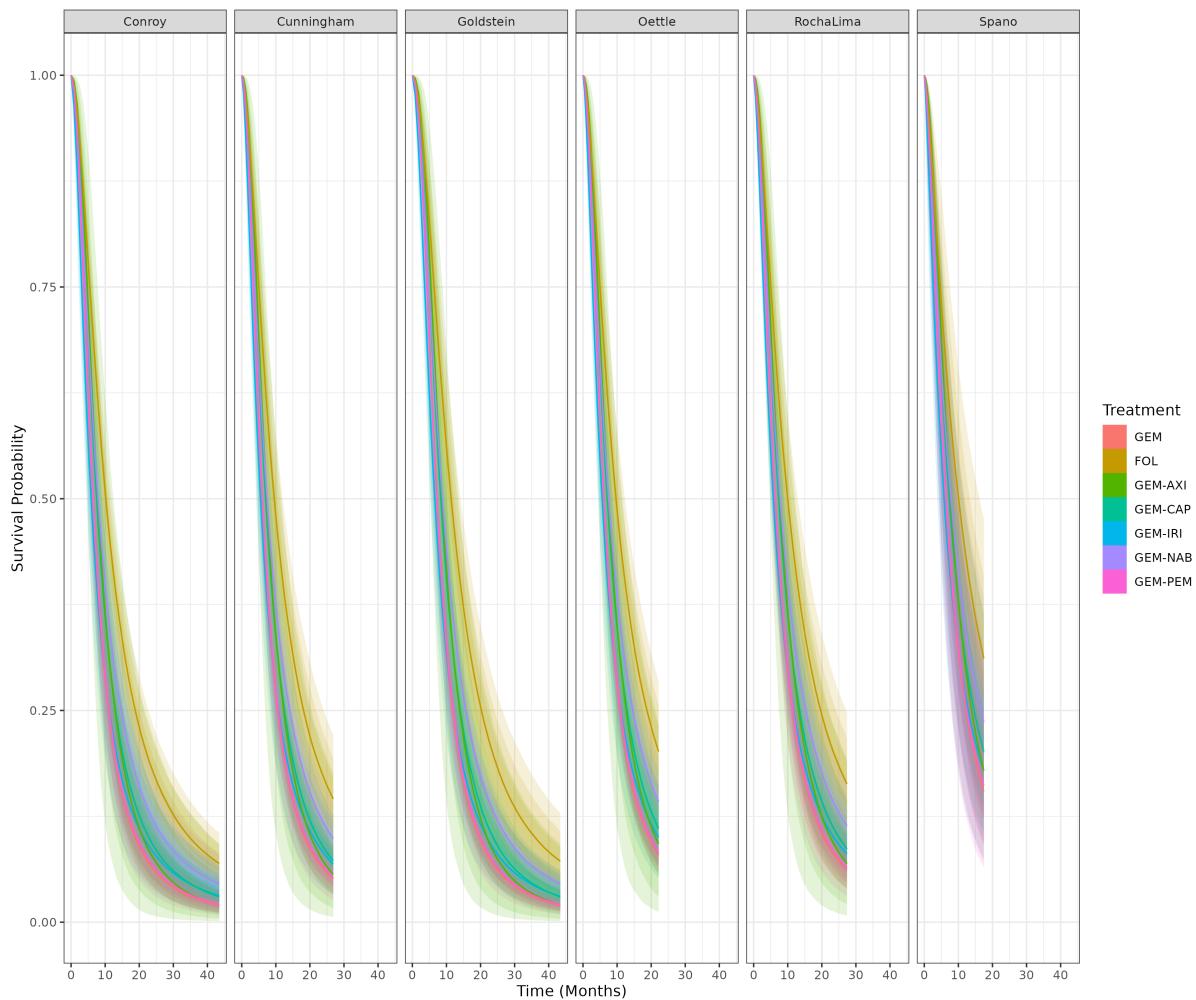


Figure 8.4: OS of each treatment in each population - SA2

8.3 Appendix B: ISPOR Good Practice Questions

8.3.1 Evidence Base

Is the population relevant? Yes. The populations in these trials were reflective of people most likely to have advanced/metastatic pancreatic cancer. The only slight deviation was the Conroy study having a maximum age eligibility requirement, meaning that study likely had fitter patients.

Are any relevant interventions missing? Potentially. The Gresham NMA included 19 treatments in total. This NMA included all of the best-performing treatments from the Gresham study, but could be expanded to include more studies in the future. The treatments in this NMA however were chosen as the ones most likely to be of interest to a decision maker in a clinical setting.

Are any relevant outcomes missing? No PFS data was included. The purpose of this NMA was for assessing the OS endpoint, but the same methods would apply with PFS data.

Is the context (settings and circumstances) applicable? Treatment for pancreatic cancer has not changed much over the last 25 years. Indeed, while the studies used in this NMA were reported between 2004 and 2015, all the treatments are still relevant in 2024. If surgical studies had been included, there may be some discussion required, as more attempts are being made to operate on pancreatic cancer

nowadays.

Did the researchers attempt to identify and include all relevant RCTs? No thorough literature was performed for this NMA. Studies were selected based on how well they were reported, due to considerations with digitising the KM curves, and whether the comparators were relevant. As mentioned previously, the best-performing treatments from the Gresham study were of primary interest for this NMA.

Do the trials for the interventions of interest form one connected network of RCTs? Yes, see Figure 6.1.

Is it apparent that poor quality studies were included, thereby leading to bias? No. The only potential comment here is the low sample size of the Spano or Goncalves studies. These studies were not deemed to be of poor quality despite this.

Is it likely that bias was induced by selective reporting of outcomes in the studies? As only the OS endpoint was considered, this was not deemed to be a potential influencer of any bias in the NMA.

Are there systematic differences in treatment effect modifiers (i.e., baseline patient or study characteristics that have an impact on the treatment effects) across the different treatment comparisons in the network? No. The populations in the included studies were more-or-less identical. We included the proportion of male patients as a covariate in order to use the ML-NMR method.

If yes (i.e., there are such systematic differences in treatment effect modifiers), were these imbalances in effect modifiers across the different treatment comparisons identified before comparing individual study results? NA.

8.3.2 Analysis

Were statistical methods used that preserve within-study randomization? (No naive comparisons) Yes. The NMA was based on relative treatment effects.

If both direct and indirect comparisons are available for pairwise contrasts (i.e., closed loops), was agreement in treatment effects (i.e., consistency) evaluated or discussed? No closed loops. Question not applicable.

In the presence of consistency between direct and indirect comparisons, were both direct and indirect evidence included in the network meta-analysis? No closed loops. Question not applicable.

With inconsistency or an imbalance in the distribution of treatment effect modifiers across the different types of comparisons in the network of trials, did the researchers attempt to minimize this bias with the analysis? This was not deemed relevant due to the similarity of the trials.

Was a valid rationale provided for the use of random-effects or fixed-effect models? Given the similarity of the included trials, FE models were not deemed to be clinically inappropriate. This is why both FE and RE models were fit for each likelihood. The best performing models were selected based on the LOOIC and DIC scores, rather than any clinical considerations.

If a random-effects model was used, were assumptions about heterogeneity explored or discussed? Question not applicable as the RE model was not used.

If there are indications of heterogeneity, were subgroup analyses or meta-regression analysis with prespecified covariates performed? Yes, by nature of using an ML-NMR.

8.3.3 Reporting Quality and Transparency

Is a graphical or tabular representation of the evidence network provided with information on the number of RCTs per direct comparison? Yes, see Figure 6.1. The thickness of lines denotes the number of RCTs available per comparison.

Are the individual study results reported? Yes.

Are results of direct comparisons reported separately from results of the indirect comparisons or network meta-analysis? Question not applicable.

Are all pairwise contrasts between interventions as obtained with the network meta-analysis reported along with measures of uncertainty? Yes, see Figure 6.14.

Is a ranking of interventions provided given the reported treatment effects and its uncertainty by outcome? Yes, see Figure 6.15.

Is the effect of important patient characteristics on treatment effects reported? No, due to lack of IPD.

8.3.4 Interpretation

Are the conclusions fair and balanced? Yes. Every attempt was made to do this.

8.3.5 Conflicts of Interest

Were there any potential conflicts of interest? This project was somewhat personal to the author, but that personal experience did not influence the results. There was no inherent bias to a particular therapy. The author is employed by a health economics consultancy, but has no interest in a particular therapy from a commercial perspective either.

If yes, were steps taken to address these? Not required.

8.4 Appendix C: The PCNMA Package

To facilitate the analysis conducted in this project, an R package was constructed. Performing the analysis in this way ensures easy reproducibility for further analysis in the future. The package is roughly split into two parts: survival code, and NMA code.

8.4.1 Survival Functions

The central function in the package for survival analysis is the *fit_distribution* function. This function is a wrapper around the *flexsurvreg* function from *flexsurv*. It is designed to take a list object containing the names of distributions and the associated argument to pass that distribution to *flexsurv*. For example,

```
1 distributions <- list("Weibull" = "weibull")
```

The reason for defining the distributions in this way is because the default *flexsurv* arguments look untidy by default when plotting. For example, the exponential distribution is passed as just “exp”. The beauty of the *fit_distribution* function is that several distributions can be passed with a tidier name, which *flexsurvreg* never sees. This purely aesthetic change can be seen by comparing Figure 8.5, and Figure 8.6.

Any default *flexsurv* distribution will work with *fit_distribution*, it just needs a suitable name in the list object . The *fit_distribution* function itself doesn’t actually fit any distributions. There is a sub-function, *.fit_distribution* which takes a single distribution as an argument, and some data on which to fit that distribution. This function takes the “weibull” element of the above list and passes it to

flexsurv. The `fit_distribution` function maps `.fit_distribution` across the list of distributions, using the `purrr::map` function [Wickham and Henry, 2023]. After some data cleaning, the object that is returned by `fit_distribution` is given the class “`fitted_distribution`”.

Several S3 methods exist for objects of class “`fitted_distribution`”, these are `plot.fitted_distribution`, `summary.fitted_distribution`, and `coef.fitted_distribution`. These functions allow for plotting fitted models, accessing information such as AIC scores, and accessing the model coefficients respectively.

8.4.2 NMA Functions

The NMA part of the PCNMA package is a wrapper around `multinma`. In particular, the `fit_model` function is a wrapper around `multinma::nma`.

Once a network has been created using the `multinma` functions `set_ipd` and `set_agd`, the `fit_model` function, with arguments such as iterations and chains can be used to fit an ML-NMR method for a given likelihood. Models fit in this way are given the class `fitted_model`. An object of class `fitted_model` has two S3 methods: `plot.fitted_distribution`, used for plotting trace plots and posterior results, such as predicted KM curves for a given study population, and `summary.fitted_distribution`, which is used for obtaining the LOOIC and DIC scores for the model.



Downloaded: 15/07/2024

Approved: 15/07/2024

Matthew Knowles
Registration number: 220209395
Population Health
Programme: MSc Statistics with Medical Applications

Dear Matthew

PROJECT TITLE: A Network Meta Analysis of Treatments for Locally Advanced/Metastatic Pancreatic Cancer

APPLICATION: Reference Number 064029

This letter confirms that you have signed a University Research Ethics Committee-approved self-declaration to confirm that your research will involve only existing research, clinical or other data that has been robustly anonymised. You have judged it to be unlikely that this project would cause offence to those who originally provided the data, should they become aware of it.

As such, on behalf of the University Research Ethics Committee, I can confirm that your project can go ahead on the basis of this self-declaration.

If during the course of the project you need to [deviate significantly from the above-approved documentation](#) please inform me since full ethical review may be required.

Yours sincerely

Email Scharr Rec
Departmental Ethics Administrator

Package ‘PCNMA’

July 15, 2024

Type Package

Title Parametric and Non-Parametric Survival Models in Pancreatic Cancer Trials

Version 0.1

Date 2023-06-27

Author Matthew Knowles

Maintainer Matthew Knowles <mattknowles314@gmail.com>

Description This package contains the functions and data for my MSc Thesis, concerning the use of parametric and non-parametric models for survival in pancreatic cancer

License MIT + file LICENSE

RoxygenNote 7.3.2

Encoding UTF-8

Contents

.fit_distribution	2
coef.fitted_distribution	2
fit_distribution	3
fit_model	3
gen_network	4
gen_network_data	4
hr	5
km_estimates	5
plot.fitted_distribution	6
plot.fitted_model	7
plot.km_obj	7
rmst	8
summary.fitted_distribution	9
summary.fitted_model	9
summary.km_obj	10

Index

11

<code>.fit_distribution</code>	<i>Fit a single survival distribution</i>
--------------------------------	---

Description

This function is a wrapper on [flexsurv::flexsurvreg], and fits a single distribution to the specified data.

Usage

```
.fit_distribution(distribution, data, strata = "Treatment")
```

Arguments

<code>distribution</code>	A single distribution
<code>data</code>	An IPD dataset
<code>strata</code>	Strata for the RHS of the [survival::Surv] function

Value

A [flexsurv::flexsurvreg] object

<code>coef.fitted_distribution</code>	<i>Coefficients of fitted models</i>
---------------------------------------	--------------------------------------

Description

Returns the model coefficients for a given survival model

Usage

```
## S3 method for class 'fitted_distribution'
coef(fit, studies, ...)
```

Arguments

<code>fit</code>	A [PCNMA::fitted_distribution] object
<code>studies</code>	Which studies to include
<code>...</code>	for S3 consistency

fit_distribution *Fit survival distributions to a dataset.*

Description

This function extends the [PCNMA:::fit_distribution] function, by fitting a given set of distributions to a TTE dataset.

Usage

```
fit_distribution(  
  distributions = nice_parametric_dists,  
  data,  
  strata = 1,  
  maxT = 60  
)
```

Arguments

distributions	A list of distributions
data	An IPD dataset
strata	Stratification variables
maxT	maximum time to calculate fitted values at

Value

A dataframe with fitted values

fit_model *Run an NMA*

Description

Run an NMA

Usage

```
fit_model(  
  network,  
  effects,  
  seed = 1,  
  chains = 4,  
  iter = 1000,  
  llhood = "weibull"  
)
```

Arguments

<code>network</code>	A [multinma::nma_data] object
<code>effects</code>	A string specifying the effects, either "fixed" or "random"
<code>seed</code>	Random number generating seed
<code>chains</code>	Number of chains used by Stan
<code>iter</code>	Number of iterations
<code>llhood</code>	Character string specifying a likelihood function
<code>...</code>	Other parameters to pass to [multinma::nma]

gen_network*Generate a network of evidence***Description**

Generate a network of evidence

Usage

```
gen_network(net_data, ref, covs)
```

Arguments

<code>net_data</code>	A data set created by [PCNMA::gen_network_data]
<code>ref</code>	A character reference treatment
<code>covs</code>	Any covariates to be included in the ML-NMR

Value

A [mutlinma::nma_data] object

gen_network_data*Generate network data***Description**

Generate network data

Usage

```
gen_network_data(data, ref)
```

Arguments

data	A date extraction dataset
ref	A character reference treatment

Value

A dataframe

hr	<i>Hazard Ratio</i>
----	---------------------

Description

Hazard Ratio

Usage

```
hr(TTE, strata = "1")
```

Arguments

TTE	A TTE dataframe
strata	Strata for the RHS of the [survival::Surv] function

km_estimates	<i>Generate KM estimates</i>
--------------	------------------------------

Description

Generate Kaplan-Meier estimates from a TTE object. Uses [ggsurvfit::survfit2] to generate KM estimates. Only requires strata to be specified.

Usage

```
km_estimates(TTE, strata = "1")
```

Arguments

TTE	A TTE dataframe
Strata	for the RHS of the [survival::Surv] function

Value

A [PCNMA::km_obj] object

plot.fitted_distribution
Plot a fitted distributions object

Description

Creates a plot for a result of ‘PCNMA::fit_distribution‘.

Usage

```
## S3 method for class 'fitted_distribution'
plot(
  fit,
  CI = FALSE,
  km = FALSE,
  km_alpha = 1,
  linewidth = 0.75,
  linetype = "dashed",
  theme = "bw",
  facet_by = NULL,
  ...
)
```

Arguments

fit	A ‘PCNMA::fitted_distribution‘ object
CI	Include a confidence interval?
km	Add the original KM curve?
km_alpha	How transparent to make the KM curve in the plot
linewidth	how thick to make the model lines
linetype	What type of line to use for the models
theme	Ggplot theme
facet_by	Create a faceted plot
...	For S3 consistency

plot.fitted_model *Plots for an NMA modelr*

Description

Plots for an NMA modelr

Usage

```
## S3 method for class 'fitted_model'  
plot(  
  model,  
  type = "trace",  
  pars = parsForStan,  
  prob = 0.95,  
  ordered = FALSE,  
  xLims = NULL,  
  study  
)
```

Arguments

model	A [PCNMA::fitted_model] object
type	Type of plot to produce
pars	Which model parameters to include in the plot
prob	Significance level
ordered	Order the plots?
xLims	Limits on the x-axis
study	Which study to plot for

plot.km_obj *Plot a KM curve*

Description

Plot a KM curve

Usage

```
## S3 method for class 'km_obj'
plot(
  fit,
  type = "survival",
  risk.table = TRUE,
  break.x.by = 5,
  xMax = 40,
  risktable.height = 0.3,
  ...
)
```

Arguments

<code>fit</code>	A 'PCNMA::km_obj' object
<code>type</code>	See 'ggsurvfit::ggsurvfit' for details.
<code>risk.table</code>	Add numbers at risk?
<code>break.x.by</code>	A numeric value for splitting x axis
<code>xMax</code>	The maximum time value to plot
<code>risktable.height</code>	The proportion of the figure to be taken up by the risk table
<code>...</code>	For S3 consistency

Value

A plotted km curve

rmst

RMST of fitted survival models

Description

Returns the RMST estimate for a fitted survival model up to a time x.

Usage

```
rmst(fit, x, ...)
```

Arguments

<code>fit</code>	A [PCNMA::fitted_distribution] object
<code>x</code>	Time to calculate RMST for
<code>...</code>	For S3 consistency

```
summary.fitted_distribution
```

Summary of a set of fitted models

Description

Summary of a set of fitted models

Usage

```
## S3 method for class 'fitted_distribution'  
summary(fit, AIC = FALSE, median = FALSE)
```

Arguments

fit	A [PCNMA::fitted_distribution] object.
AIC	Returns the AIC scores for a set of models
median	Returns a table of median estimates for a set of models

```
summary.fitted_model    Summary of an NMA model
```

Description

Summary of an NMA model

Usage

```
## S3 method for class 'fitted_model'  
summary(model, likelihood, effect)
```

Arguments

model	A [PCNMA::fitted_model] object
likelihood	A vector of likelihoods
effect	FE or RE

summary.km_obj *Summarise KM data*

Description

Summarise KM data

Usage

```
## S3 method for class 'km_obj'  
summary(fit, ...)
```

Arguments

fit A [PCNMA::km_obj] object

Value

A summary table of the KM data

Index

.fit_distribution, 2
coef.fitted_distribution, 2
fit_distribution, 3
fit_model, 3
gen_network, 4
gen_network_data, 4
hr, 5
km_estimates, 5
plot.fitted_distribution, 6
plot.fitted_model, 7
plot.km_obj, 7
rmst, 8
summary.fitted_distribution, 9
summary.fitted_model, 9
summary.km_obj, 10

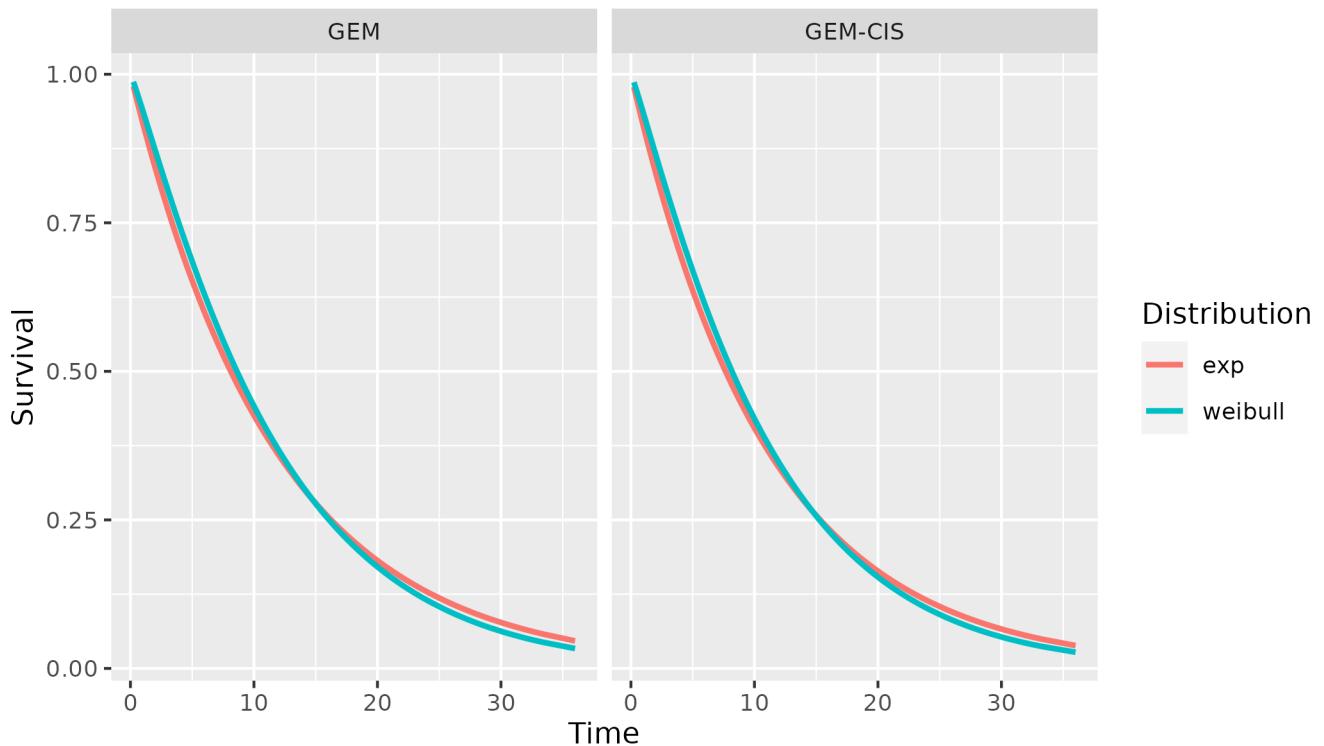


Figure 8.5: Fitted Models as in Flexsurv

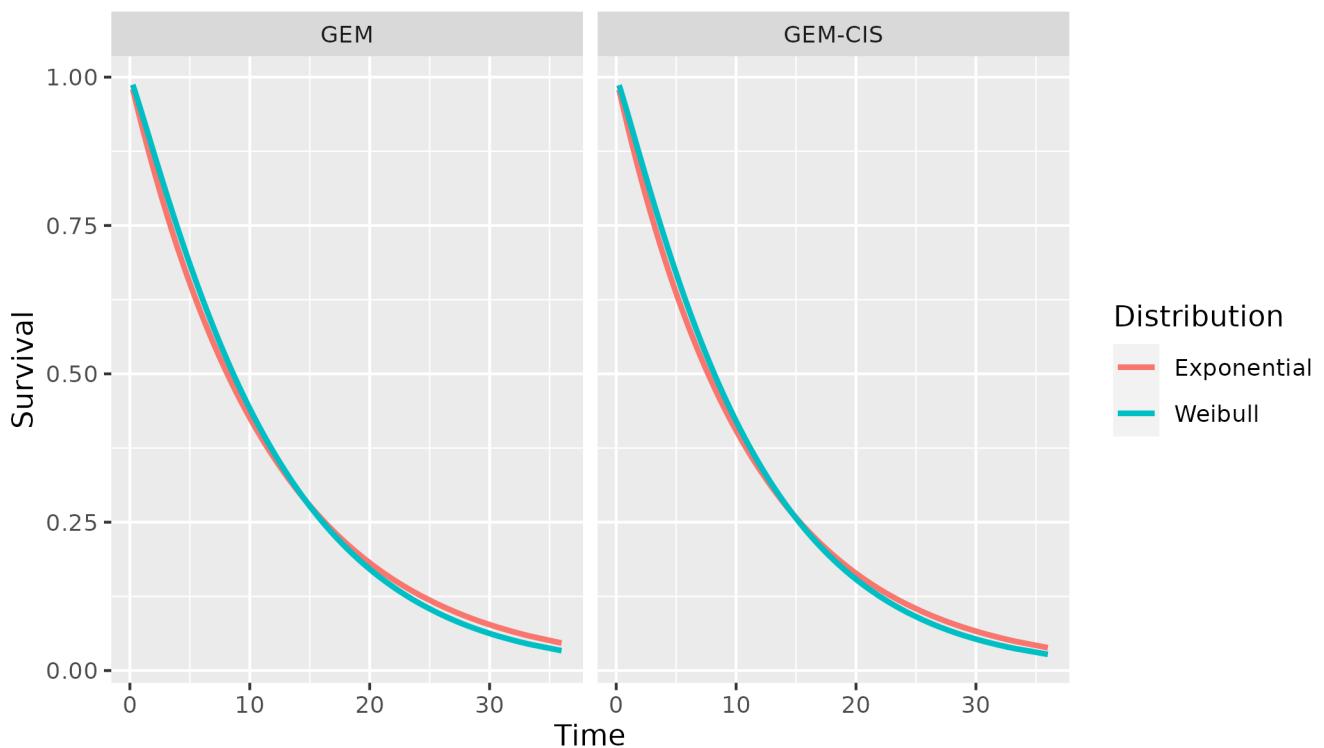


Figure 8.6: Fitted Models with *fit_distribution*