One-Step Individual Participant Data Network Meta-Analysis of Time-to-Event Data

Suzanne Claire Freeman

A thesis presented for the degree of

Doctor of Philosophy

Institute of Clinical Trials & Methodology

University College London

March 2017

Declaration

I, Suzanne Claire Freeman confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the   thesis.

Abstract

Network meta-analysis (NMA) combines direct and indirect evidence from trials to calculate and rank treatment effect estimates. While modelling approaches for continuous and binary outcomes are relatively well developed, less work has been done with time-to-event outcomes. Such outcomes have usually been analysed using Cox proportional hazard (PH) models, but in oncology, with longer follow-up of trials and time-dependent effects of targeted treatments, this may no longer be appropriate. Alongside this, NMA conducted in the Bayesian setting has been increasing in popularity. In this thesis I extend the work of Royston and Parmar to the NMA setting, showing that Royston-Parmar models, fitted in WinBUGS, provide a flexible, practical approach for Bayesian NMA with time-to-event data and can accommodate non-PH.

Inconsistency in NMA occurs when the direct and indirect evidence are not in agreement with each other and can result in biased treatment effect estimates. It is therefore important that attempts are made to identify, understand and, where appropriate, adjust for inconsistency. In this thesis I consider four increasingly complex methods of assessing inconsistency in NMA, proposed (relatively) recently in the literature. Motivated by individual participant data (IPD) from 42 trials comparing radiotherapy, sequential and concomitant chemotherapy from 7531 people with lung cancer, I illustrate why one of these approaches may be misleading and propose an alternative approach.

Stratified medicine aims to identify groups of patients most likely to respond to treatment. However, many trials are underpowered to detect clinically meaningful differences in subgroups. NMA models fitted with treatment-covariate interactions potentially have greater power to identify such differences. In the final part of this thesis I extend the one-step IPD NMA Royston-Parmar model to include treatment-covariate interactions, providing practical guidance on how to deal with missing covariate data and how to combine or separate within and across trial information.

Acknowledgements

Contents

List of Figures

List of Tables

Glossary

AD	Aggregate data
ADT	Androgen Deprivation Therapy
CI	Confidence interval
Con CT	Concomitant chemotherapy
CP	Carboplatin-paclitaxel chemotherapy
CrI	Credible interval
CTRT	Chemoraditaion
CT+RT	Chemotherapy plus radiotherapy
CT+S	Chemotherapy plus surgery
DIC	Deviance information criterion
FTE	Fixed treatment effect
HR	Hazard ratio
IPD	Individual participant data
IGR	Institut Gustave-Roussy
LogHR	Log hazard ratio
MA	Meta-analysis
MRC CTU at UCL	Medical Research Council Clinical Trials Unit at University College London
NMA	Network meta-analysis
PH	Proportional hazards
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PS	Performance status
RCS	Restricted cubic spline
RCTs	Randomised Clinical Trials
RT	Radiotherapy
RTE	Random treatment effect
Seq CT	Sequential chemotherapy
WHO	World Health Organisation

1 General Introduction to Network Meta-Analysis

1.1 What is network meta-analysis?

Network meta-analysis (NMA) is the extension of pairwise meta-analysis (MA) methods to the setting where there are more than two treatments. A network can range in complexity with the simplest situation being a network of three treatments (Figure 1.1) but a network could include many more treatments (Figure 1.2). NMA uses a single statistical model to combine both direct and indirect evidence from all the trials in a network to calculate treatment effect estimates for every treatment comparison, regardless of whether two treatments have been compared directly within an individual trial, and thus permits ranking the treatments. Therefore within a network of clinical trials, where each trial compares at least two treatments, NMA combines direct, randomised evidence with indirect, non-randomised evidence.
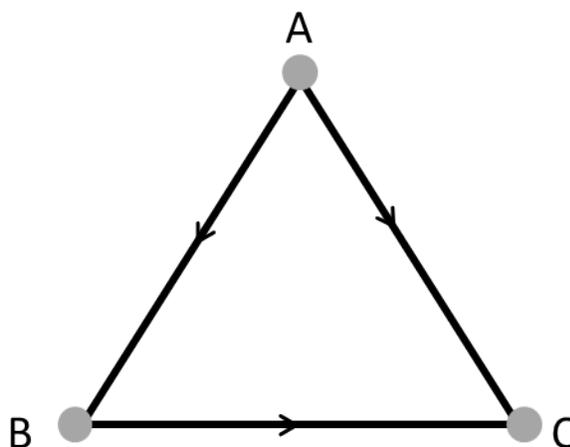
Figure 1.1: Example network diagram for three treatments forming a closed loop. Solid black lines represent randomised trial evidence directly comparing treatments A, B and C. The arrows indicate the direction of the treatment effects.

NMA can be conducted using **individual participant data (IPD)** or **aggregate data (AD)** with IPD considered the gold standard for MA, and here NMA, since it offers many advantages over AD (Chalmers, 1993; Stewart and Tierney, 2002; Stewart et al., 2015). IPD is particularly useful when individual trials analyse data in different ways, as it allows for all trials to be re-analysed using the same method. IPD allows for the standardisation
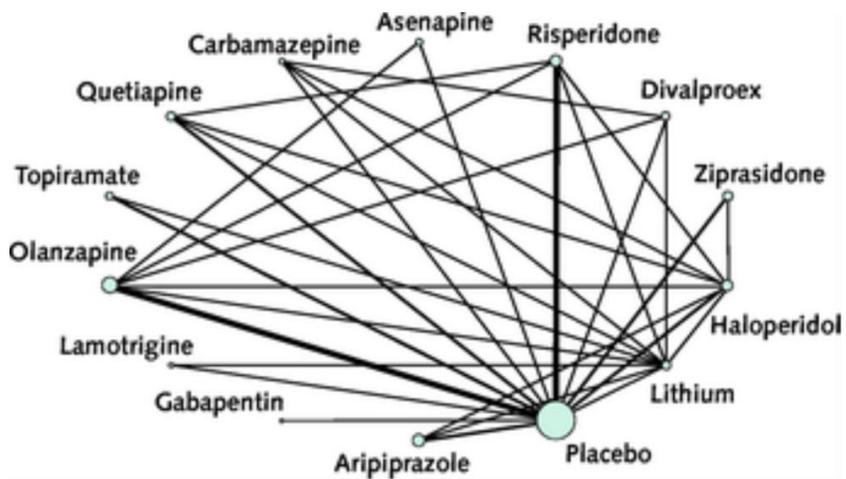
Figure 1.2: Example network diagram of thirteen treatments and placebo for acute mania (Cipriani et al., 2011). The solid lines represent direct comparisons between treatments with the width of the lines proportional to the number of trials directly comparing the two treatments. The area of each treatment node is proportional to the number of patients randomised to the treatment.

of which patients are included or excluded from analyses, the possibility of recoding of outcomes and covariates to a standardised definition, the inclusion of patients who were excluded from the original trial analyses and the use of up-to-date follow-up information. One of the main benefits of IPD over AD is that IPD allows for detailed checking of the data against the published results and checks for the integrity and quality of randomisation and follow-up (Stewart and Tierney, 2002). Another major benefit of IPD is in the analysis of the data. IPD is highly desirable and provides greater statistical power for investigating interactions between treatment and patient-level covariates and allows for multiple patient-level factors to be considered in combination (Jansen, 2012; Simmonds et al., 2005). However, obtaining IPD can be difficult and time-consuming as there are often agreements that need to be put in place before data can be transferred. Once the IPD is received re-analysing each trial to standardise the analysis takes time which in turn increases the costs associated with obtaining IPD. Additionally, if some trials identified for a MA or NMA can provide IPD and other trials do not then there can be bias in the obtained data. This thesis will focus on conducting NMA with IPD.

NMA models can either assume **fixed treatment effects (FTE)** or **random treatment effects (RTE)**. A FTE model estimates one treatment effect for each treatment comparison, including treatment comparisons only informed by indirect evidence, which is assumed to be the same for all trials comparing the same two treatments. A RTE model assumes that there is no single underlying treatment effect but that the mean treatment effect comes from a common distribution (Donegan et al., 2012; Lu and Ades, 2004). This thesis will consider both FTE and RTE NMA models.

NMA can be conducted in a one-step or a two-step process. In a two-step process the first step involves obtaining point estimates of treatment effects along with a measure of uncertainty from each trial. In the second step the estimates are then combined using either a FTE or RTE model. In a one-step process the two-steps are combined and conducted by fitting one single statistical model. Advantages of the one-step approach when using IPD include reducing the number of parameters to be estimated, gaining efficiency and a wider choice of models that can be fitted. Furthermore, in a one-step model all relevant parts of the data can be modelled simultaneously. This thesis will focus on conducting NMA in a one-step process.

1.2 Statistical concepts

This section introduces some key statistical concepts which will be used throughout this thesis. These concepts will be developed further in Chapter 2.

1.2.1 Consistency (& Inconsistency)

A network is considered to be consistent when the treatment effect estimates from the direct comparisons are in agreement with the treatment effect estimates from the indirect comparisons. For example, in a network of three treatments A, B and C, such as Figure 1.1 where the arrows indicate the direction of treatment effects, with treatment effect estimates μ_{AB}, μ_{AC} and μ_{BC}, where μ_{AB} is the treatment effect of treatment B compared to treatment A, μ_{AC} is the treatment effect of treatment C compared to treatment A and μ_{BC} is the treatment effect of treatment C compared to treatment B, the network is considered to be consistent if the following set of equations are satisfied:

$$\mu_{BC} = \mu_{AC} - \mu_{AB}$$

$$\mu_{AB} = \mu_{AC} - \mu_{BC} \tag{1.1}$$

$$\mu_{AC} = \mu_{AB} + \mu_{BC}$$

These equations are known as the consistency equations and when they are not satisfied the network is considered to be inconsistent i.e. the direct and indirect evidence are not in agreement (Higgins and Whitehead, 1996).

1.2.2 Heterogeneity

Methodological heterogeneity refers to differences in study populations, methods, settings, outcome measurements or anything else that makes trials different, which may or may not lead to differences in treatment effect (statistical heterogeneity). Methodological heterogeneity should be assessed both within and between direct and indirect treatment comparisons and can also result in differences between direct and indirect evidence (inconsistency) (Mills et al., 2013).

Statistical heterogeneity is the extent of disagreement between trial-specific treatment effects, in trials of the same design, which exceeds that which would be expected by chance

alone (Ciprani et al., 2013; Higgins et al., 2012; Lu and Ades, 2006; Lumley, 2002; White et al., 2012). As is typical in the NMA literature, throughout this thesis 'design' will refer to the treatments being compared within a trial and not to design characteristics such as parallel groups or cross-over (Higgins et al., 2012). For example, two trials both comparing treatment A to treatment B will be considered to be of the same design whereas a third trial comparing treatment A to treatment B and treatment C will be considered to be a different design. Furthermore, in this case direct evidence comparing treatment A and treatment B comes from more than one design. Throughout this thesis the term 'heterogeneity' will be used to refer to statistical heterogeneity.

1.2.3 Frequentist framework

In a frequentist framework, the process giving rise to, or generating, the data is considered to be repeatable, while model parameters are assumed to be fixed. Clinical trials are conducted on samples of patients which are assumed to be representative of the general population (Lunn et al., 2013). Results are often presented as point estimates with 95% confidence intervals. A 95% confidence interval means that with 100 repeated samples we would expect the point estimate of the treatment effect to lie within the 95% confidence interval 95 times.

1.2.4 Bayesian framework

In a Bayesian framework probability is used to represent uncertainty. The observed data is considered fixed whilst the parameters are unknown random variables with associated probability distributions. A Bayesian analysis requires two things: prior information about what could be considered a plausible treatment effect and the likelihood of the treatment effect based on the trial data. These two are then combined using Bayes Theorem to obtain the posterior distribution which provides the final estimate of the treatment effect (Lunn, 2014). All inference will be based on the posterior distribution. Results are often presented as point estimates with 95% credible intervals. A 95% credible interval means that the probability of the point estimate lying within that interval is 95%.

1.3 Motivation for thesis

The Medical Research Council Clinical Trials Unit at University College London (MRC CTU at UCL) has been conducting oncology trials and MA for more than 30 years. In particular two currently recruiting oncology trials in ovarian (ICON8B) and prostate cancer (STAMPEDE) motivate the development of appropriate methodology for conducting NMA with time-to-event outcomes.

ICON8B was designed as a trial in women with high-risk ovarian cancer to compare standard carboplatin-paclitaxel chemotherapy (CP) once every 3 weeks plus bevacizumab with dose dense CP plus bevacizumab and dose dense CP (without bevacizumab). The trial is powered to consider superiority of dose dense CP plus bevacizumab compared to either of the other two treatment regimens. If dose dense CP plus bevacizumab is not superior then the question of whether dose dense CP (without bevacizumab) is non-inferior to standard CP plus bevacizumab would be of interest. However to power a trial for this non-inferior comparison would require a large number of patients and long recruitment and follow-up periods. Therefore ICON8B was designed to be incorporated within a NMA allowing the direct evidence from ICON8B to be supplemented with indirect evidence from other trials.

STAMPEDE was designed as a multi-arm multi-stage trial for men with prostate cancer. The trial was designed to compare five new treatment regimens with the androgen deprivation therapy (ADT) control arm, but was not powered to compare new treatment regimens to each other. When the trial started in 2005 the new treatment regimens considered were ADT plus zoledronic acid, ADT plus docetaxel, ADT plus celecoxib, ADT plus zoledronic acid and docetaxel and ADT plus zoledronic acid and celecoxib. Since the trial started four new treatment regimens have been added to the trial (ADT plus abiraterone, ADT plus enzalutamide and abiraterone, ADT plus radiotherapy to the prostate for men with metastatic disease and ADT plus metformin) and the original five new treatment regimens have all completed recruitment. Therefore the treatment regimens have not all been recruiting patients at the same time. Direct comparisons will only be made between the ADT control arm and the new treatment regimens in patients recruited concurrently. Since STAMPEDE opened other trials with overlapping treatment comparisons have either been started or completed. A NMA based on some of the treatment regimens being considered in STAMPEDE, with the addition of other prostate can-
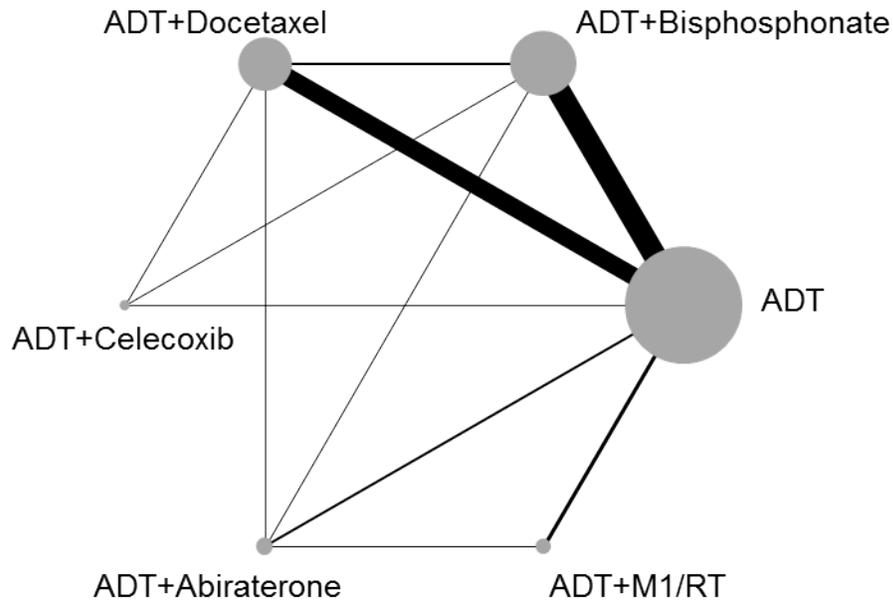
Figure 1.3: Network diagram based on the STAMPEDE trial. Node size is proportional to the number of patients randomised to the treatment. Line thickness is proportional to the number of trials involved in each direct comparison. ADT = androgen deprivation therapy, M1/RT = radiotherapy to the prostate for men with metastatic disease.

cer trials, is currently being planned by the MRC CTU at UCL. In this NMA the zolendronic acid treatment regimen from the STAMPEDE trial will be included within the ADT plus bisphosphonate treatment node. A network diagram of the data available to inform this network (up to January 2015) is presented in Figure 1.3.

In oncology cure may be unfeasible, particularly in more advanced stages or poorer prognosis cancers. Therefore trials are often interested in answering the question 'how much longer does this new treatment extend a patient's life or time to progression?' and generate time-to-event data. The literature for conducting NMA with time-to-event data is relatively sparse (see Section 2.8). Traditionally time-to-event outcomes have been analysed using the semi-parametric Cox proportional hazards (PH) model (Cox, 1972), but in oncology with longer follow-up of trials, and time-dependent effects of targeted treatments, there is increasing evidence of non-PH so this may no longer be appropriate (Royston and Parmar, 2016; Trinquart et al., 2016). One of the aims of this thesis is to show that when using the Royston-Parmar model with restricted cubic splines, the Bayesian framework provides a natural, practical and flexible approach for NMA of time-to-event data.

1.4 Structure of thesis

This thesis continues in **Chapter 2** with a review of the methodology for conducting NMA with continuous, binary and time-to-event data summarising some of the most common NMA models and looking specifically at methods of accounting for and incorporating heterogeneity, consistency, bias and treatment-covariate interactions. I will also consider methods for modelling time-to-event data.

Chapter 3 will conduct a thorough exploration of two networks of trials by examining the data, conducting exploratory analyses and performing a two-step MA. **Chapter 4** describes the Royston-Parmar approach to one-step IPD NMA, how it can be implemented in a Bayesian setting using WinBUGS, how the results can be presented and extensions to test for and accommodate departures from the non-PH assumption. **Chapter 5** assesses inconsistency in a lung cancer network, takes a closer look at the net heat approach for assessing inconsistency, explains how the net heat approach can be misleading and proposes and outlines an alternative approach. **Chapter 6** extends the one-step IPD Royston-Parmar NMA model to include treatment-covariate interactions. Finally in **Chapter 7** this thesis concludes with a review and discussion of the work conducted and consideration of what future research is needed.

2 Literature Review

2.1 Literature search

The primary aim of this literature review was to identify important developments in methodology for conducting NMA since 1995, with binary, continuous or time-to-event data and relate them to each other. I conducted a literature review using Embase, Ovid MEDLINE and Ovid MEDLINE In-Process & Other Non-Indexed Citations electronic databases and through manual searching of reference lists of relevant articles found. A search strategy was used to identify combinations of keywords (network, meta, analysis, mixed, multiple, treatment, comparison) in the titles or abstracts of articles. Databases were searched on the 1st June 2015 using the following search strategy:

1. (network adj2 meta?analys*).ab, ti

2. (indirect adj2 comparis*).ti

3. (mixed adj2 treatment adj2 comparis*).ab, ti

4. (multiple adj2 treatment adj2 comparis*).ab, ti

5. 1 OR 2 OR 3 OR 4

6. Remove duplicates from 5

7. Limit 6 to 1995-current

* wildcard to allow for multiple endings e.g. comparison and comparisons; ab indicates that abstracts were searched for the relevant terms; ti indicates that titles were searched for the relevant terms; adj2 retrieves words occurring within 2 words of each other, in any order; ? is a substitute for one character or none.

1192 results were returned through searching of the electronic databases and 194 articles were obtained based on the titles of the papers. Based on a review of the abstracts 103 papers were obtained in full. Manual searching of reference lists identified a further 67 papers. This resulted in a total of 170 papers considered for this literature review.

This chapter will start with a review of conducting NMA with continuous outcomes in Section 2.2. I will then look at the development of methodology for NMA over the past 20 years beginning with the most common models in Section 2.3. I will consider models which specifically estimate and account for consistency in Section 2.4 and heterogeneity in Section 2.5. It is important to consider potential sources of bias within a network and Section 2.6 explores the different types of bias which may be present in a NMA and methods of adjusting for bias. NMA models for continuous and binary outcomes are widely available but are less common for time-to-event data. In Section 2.8 I consider alternative methods to the Cox PH model which could be used in a NMA of time-to-event data. In Section 2.9 I highlight other methodological advances in NMA. Section 2.10 compares Bayesian and frequentist methods of NMA. Section 2.11 considers the recent increase in NMA and what guidance is available for authors and readers. Finally, Section 2.12 summarises the literature review and highlights areas where more research is needed.

2.2   NMA with continuous outcomes

Let A, B and C be three treatments in a network comprised of three two-arm trials so that each trial compares two of the three treatments forming a triangular network of the same form as Figure 1.1, where the arrows inform the parameterisation of the network. Parameterisation of the network id discussed in Section 4.4. Applying the approach of Higgins (2001) to the NMA setting let the response, y_{ij}, of patient i in trial j be a continuous outcome. Let γ parameters represent fixed treatment effects, β parameters represent random treatment effects and let v be random variables and ε the residual error. Then the random treatment effect NMA model is:

$$y_{ij} = \gamma_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \varepsilon_{ij} \tag{2.1}$$

$$\beta_{1j} = \gamma_1 + v_{1j}$$

$$\beta_{2j} = \gamma_2 + v_{2j}$$

$$\varepsilon_{ij} \sim N(0, \sigma_j^2)$$

$$\begin{pmatrix} v_{1j} \\ v_{2j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau^2 & \frac{\tau^2}{2} \\ \frac{\tau^2}{2} & \tau^2 \end{pmatrix} \right)$$

where γ_{0j} is the fixed intercept for each trial j, and x_{1ij} and x_{2ij} are dummy variables representing treatment allocation such that:

$$x_{1ij} = \begin{cases} 1 & \text{if treatment=B for patient } i \text{ in trial } j \text{ from an AB trial} \\ \text{-1} & \text{if treatment=C for patient } i \text{ in trial } j \text{ from a BC trial} \\ 0 & \text{otherwise} \end{cases}$$

$$x_{2ij} = \begin{cases} 1 & \text{if treatment=C for patient } i \text{ in trial } j \\ 0 & \text{otherwise} \end{cases}$$

Note that in a usual random effect model v_{1j} and v_{2j} would have an unstructured covariance matrix. However, here, a simpler approach with one parameter, τ, is adopted because there is (typically) relatively little information to estimate an unstructured covariance matrix.

In this model $\hat{\beta}_{1j}$ is an estimate of the network treatment effect μ_{AB} in trial j and $\hat{\beta}_{2j}$ is an estimate for the network treatment effect μ_{AC} in trial j. The model assumes consistency across the network so that an estimate of the network treatment effect μ_{BC} can be obtained from the estimates of μ_{AC} and μ_{AB}.

$$\mu_{BC} = \mu_{AC} - \mu_{AB} = \hat{\beta}_2 - \hat{\beta}_1$$

Let h_{ij} be a continuous covariate for patient i in trial j that is centered around the trial mean $\bar{h}_j$, then introducing a fixed treatment-covariate interaction into (2.1) we get:

$$y_{ij} = \gamma_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \gamma_1(h_{ij} - \bar{h}_j) + \gamma_2(h_{ij} - \bar{h}_j)x_{1ij} + \gamma_3(h_{ij} - \bar{h}_j)x_{2ij} + \varepsilon_{ij} \quad (2.2)$$

with $\gamma_1, \gamma_2, \gamma_3$ the coefficients of the fixed treatment-covariate interaction terms, and γ_{0j}, $\beta_{1j}, \beta_{2j}, \varepsilon_{ij}, v_{1j}, v_{2j}, x_{1ij}$ and x_{2ij} as defined previously.

With the inclusion of a treatment-covariate interaction in (2.2) the interpretation of β_1 and β_2 changes. For example, β_1 could be very significant in the model without a treatment-covariate interaction (2.1) but not at all significant, and in fact even zero, in the model with a treatment-covariate interaction (2.2). However, this does not mean that there is no treatment effect. In the presence of a treatment-covariate interaction it does not make sense to interpret β_1 and β_2 on their own and treatment effects only make sense when presented for each level of the covariate. Consider Figure 2.1, in diagram A there is no treatment-covariate interaction and the treatment effect is the same at all values of the covariate. When treatment-covariate interactions are included, the covariate should be centered on

the mean value which becomes the reference value. Therefore, in diagram B, which in-cludes a treatment-covariate interaction, at the mean value of the covariate the treatment effect is zero. However, it is clear that there is an effect of treatment, which differs, at differ-ent values of the covariate.
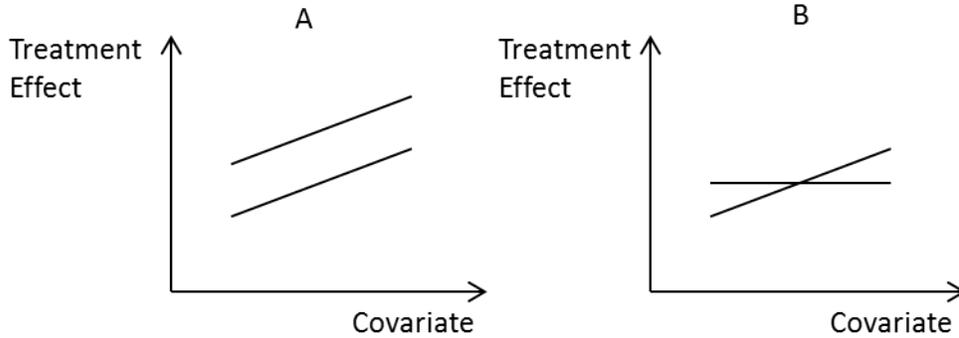
Figure 2.1: Diagram explaining how a treatment-covariate interaction changes the treat-ment effects. A: No treatment-covariate interaction - the treatment effect is the same at all values of the covariate. B: Treatment-covariate interaction present - the treatment effect is different at each value of the covariate.

Now consider the addition of a three-arm trial comparing all three treatments A, B and C to the network. There is potential for inconsistency in treatment comparisons between dif-ferent designs. For example, the treatment effect of B versus A (or C versus A) may differ between the two-arm trial comparing A and B (or A and C) and the three-arm trial. To ad-dress this, we can extend (2.1) by adding inconsistency parameters. Therefore, a random treatment effect NMA model with random inconsistency parameters (α) takes the form:

$$y_{ij} = \gamma_{0j} + \beta_{1j}x_{1ij} + \beta_{2j}x_{2ij} + \alpha_{1j}x_{1ij}\mathsf{design}_{ij} + \alpha_{2j}x_{2ij}\mathsf{design}_{ij} + \varepsilon_{ij}$$

$$\beta_{1j} = \gamma_1 + v_{1j}$$

$$\beta_{2j} = \gamma_2 + v_{2j}$$

$$\varepsilon_{ij} \sim N(0, \sigma_j^2)$$

$$\begin{pmatrix} \alpha_{1j} \\ \alpha_{2j} \end{pmatrix} \sim N(0, \tau_\alpha^2)$$

$$\begin{pmatrix} v_{1j} \\ v_{2j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_v^2 & \frac{\tau_v^2}{2} \\ \frac{\tau_v^2}{2} & \tau_v^2 \end{pmatrix} \right)$$

where x_{1ij} and x_{2ij} are as defined previously and design$_{ij}$ is a dummy variable taking the value 1 if patient i from trial j is a member of the three-arm trial and 0 otherwise. Three-arm trials must be internally consistent. Therefore, by including inconsistency parameters this model effectively removes the three-arm trial from contributing to the network estimates.

2.3   Description of the main models for NMA and their development

Most NMA methods have developed as a result of extending MA methods for two treatments to three or more treatments to take advantage of the indirect evidence. Higgins and Whitehead (1996) set out the consistency equations (1.1) for a set of trials each comparing two of three treatments. They showed that the relative effects of different treatments could be jointly estimated in a single MA model to improve power by 'borrowing strength' from direct comparisons to inform indirect comparisons. In this section, I describe NMA models assuming consistency and explore this assumption in further detail in Section 2.4.

In three-arm trials any assumptions about heterogeneity will impact on the treatment effect parameters. Higgins and Whitehead (1996) showed that under the assumption of equal heterogeneity parameters (τ^2) for each treatment effect there is a covariance of $\frac{\tau^2}{2}$ between any two treatment effects, an assumption still commonly used. This assumption may be implausible in some situations. However, it offers a simple approach requiring the estimation of only one parameter when there is often relatively little information available to estimate an unstructured covariance matrix.

The first NMA model to consider data from more than three treatments was proposed by Lumley (2002) for use with AD. Lumley proposed a hierarchical model that could account for sampling variability, heterogeneity and inconsistency through the inclusion of heterogeneity and inconsistency parameters for each treatment comparison and used the method of maximum likelihood to estimate the model parameters. A major limitation of this model was that it could be applied to two-arm trials only.

Let y_{jkl} be the outcome for treatment k compared to treatment l in trial j and let μ_k and μ_l be the true treatment effects of treatments k and l respectively, then the Lumley (2002) FTE model can be written as:

$$y_{jkl} = \mu_k + \eta_{jk} - \mu_l + \eta_{jl} + \xi_{kl}$$

35

$$\eta_{jk} \sim N(0, \tau^2)$$

$$\xi_{kl} \sim N(0, \omega^2)$$

where η_{jk} and η_{jl} are random effects with variance τ^2 that represent the difference between the average effects of treatment k and l and the effect of treatment k and l in trial j. ξ_{kl} is a random effect representing the change in the effect of treatment k when compared to treatment l (ξ_{kl} is an inconsistency parameter).

Arguably one of the most popular models for NMA is that of Lu & Ades (2004) which has been influential in the work of many other authors (Figure 2.2). This was the first NMA model which had the ability to include multi-arm trials. Lu & Ades (2004) extended the Bayesian hierarchical model for a MA of two treatments, first proposed by Smith, Spiegelhalter and Thomas (1995), to a general framework for an AD NMA of k treatments. Multi-arm trials must be internally consistent. The treatment effects from a multi-arm trial will always be correlated because they share a common control arm. Lu & Ades (2004) used the Higgins & Whitehead (1996) method of 'borrowing strength' from direct comparisons, involving one of the treatments of interest, to inform the (indirect) comparison of interest to allow multi-arm trials to be incorporated into the model.

Unlike Lumley, who modelled treatment contrasts (e.g. log hazard ratios), Lu & Ades model the treatment effect for each treatment arm. Let y_{jk} be the outcome for treatment k in trial j where b is the trial-specific baseline treatment, and let μ_{jb} be the effect of the baseline treatment b in trial j. Then the Lu & Ades (2004) RTE model can be written as:

$$y_{jk} = \begin{cases} \mu_{jb} & b = \text{ A, B, C, ... if } k = b \\ \mu_{jb} + \delta_{jbk} & k = \text{ B, C, D, ... if } k \text{ alphabetically after } b \end{cases}$$

$$\delta_{jbk} \sim N(d_{bk}, \sigma^2)$$

$$d_{bk} = d_{Ak} - d_{Ab}.$$

2.4   Consistency

NMA models can be based on the assumption of consistency (Caldwell et al., 2005; Higgins and Whitehead, 1996; Lu and Ades, 2004; Nixon et al., 2007) or inconsistency (Lu and Ades, 2006; Lumley, 2002; Salanti et al., 2007). Potential sources of inconsistency in
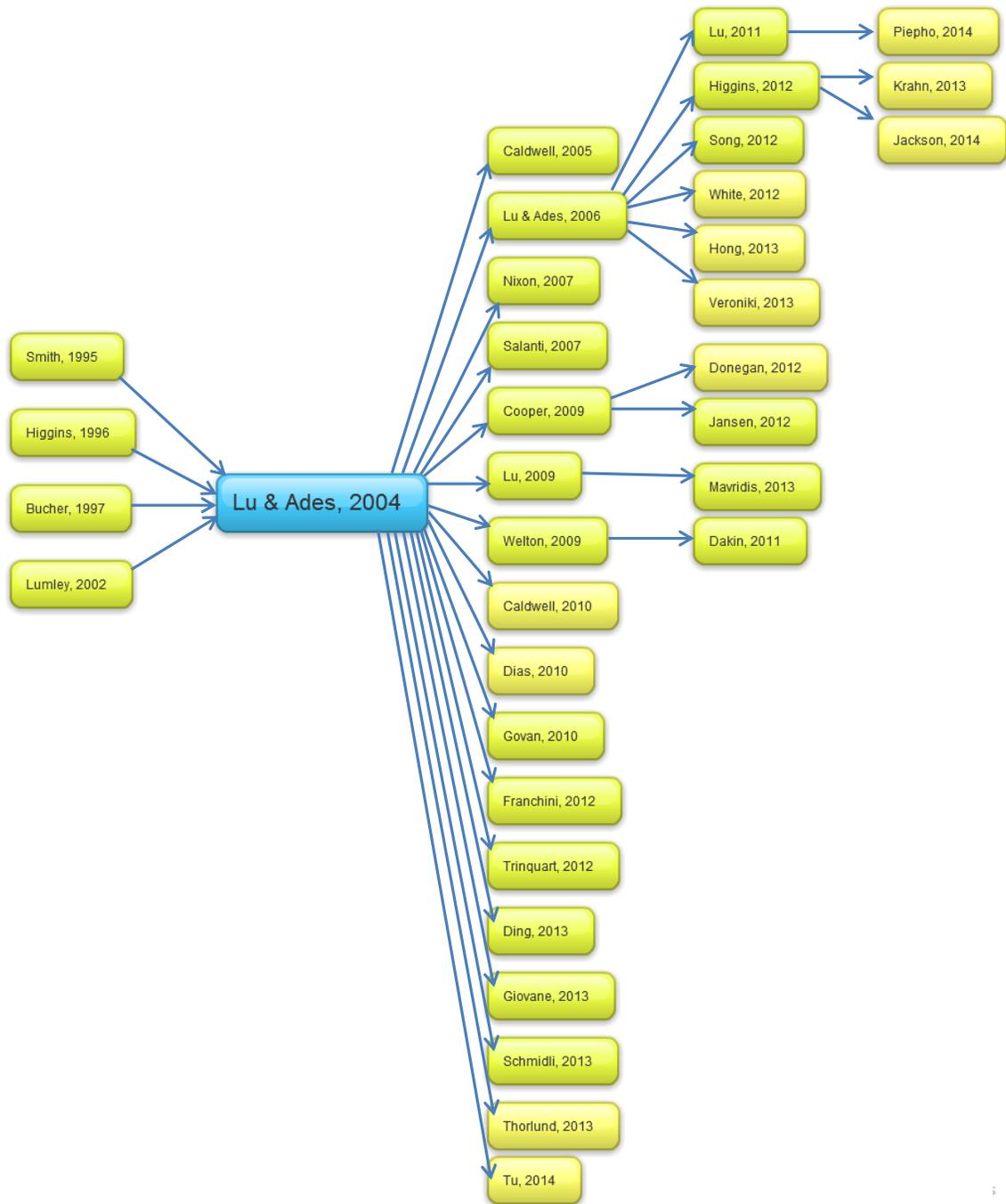
Figure 2.2: Influence of Lu & Ades NMA framework in the development of other NMA models. Journal articles at the arrow heads use or extend the methods from the journal articles at the arrow bases.

a network may arise from bias in direct comparisons (e.g. optimism bias, publication bias, selective outcome reporting, quality of randomisation, allocation concealment and masking), sponsorship bias and even genuine diversity in treatment effects across comparisons in a network (Ioannidis, 2009; Salanti et al., 2007). The power of tests for inconsistency is generally considered to be low because indirect evidence is a weaker component of most treatment estimates in NMA. Therefore, failure to reject the null hypothesis does not mean that the entire network is inconsistent (Veroniki et al., 2013). It is therefore important that attempts are made to identify, understand and, where appropriate, adjust for inconsistency. Because of this, I argue later that when presenting the results of a NMA the direct, indirect and combined evidence should be separated out and presented alongside each other.

There are many approaches to assessing inconsistency in a network. A 2013 review of methods for assessing consistency in a NMA identified ten different approaches (Donegan et al., 2013b). Some of these approaches are discussed below. One of the simplest approaches is the method of Bucher for evaluating inconsistency in all closed loops of three treatments (Bucher et al., 1997). A different and more complex approach is through the use of a design-by-treatment interaction model (Higgins et al., 2012). Other methodologies for evaluating inconsistency are outlined by Lu (2011), Dias (2010b), Krahn (2013) and Piepho (2014). I will return to these approaches later on in this section. In a Bayesian framework consistency can also be assessed by comparing a model assuming consistency with a model not assuming consistency using the deviance information criterion (DIC).

Bucher (1997) developed a method for assessing loop inconsistency in loops of three treatments within a network. The approach involved calculating the difference between the direct and indirect evidence for a treatment comparison and testing it against the null hypothesis of consistency by referring the test statistic to the normal distribution. It is important to note that in this case inconsistency can only be identified as being present in a particular treatment loop and can not be attributed to a specific design.

In a loop of three treatments — denoted A, B and C — the direct evidence of treatment C versus treatment B, $\hat{d}_{BC}^{\text{dir}}$, is compared to the indirect evidence, $\hat{d}_{BC}^{\text{ind}}$, where $\hat{d}_{BC}^{\text{ind}} = \hat{d}_{AC}^{\text{dir}} - \hat{d}_{AB}^{\text{dir}}$ and $\text{Var}(\hat{d}_{BC}^{\text{ind}}) = \text{Var}(\hat{d}_{AC}^{\text{dir}}) + \text{Var}(\hat{d}_{AB}^{\text{dir}})$. Following the method of Bucher (1997) estimates of the inconsistency parameter, ω_{BC}, and its variance can be formed by subtracting the direct and indirect estimates:

$$\hat{\omega}_{BC} = \hat{d}_{BC}^{\text{dir}} - \tilde{\hat{d}}_{BC}^{\text{ind}}$$

38

$$\text{Var}(\hat{\omega}_{BC}) = \text{Var}(\hat{d}^{\text{dir}}_{BC}) + \text{Var}(\hat{d}^{\text{ind}}_{BC}) = \text{Var}(\hat{d}^{\text{dir}}_{BC}) + \text{Var}(\hat{d}^{\text{dir}}_{AB}) + \text{Var}(\hat{d}^{\text{dir}}_{AC})$$

An approximate test of the null hypothesis of consistency is conducted by referring the test statistic $z_{BC} = \frac{\hat{\omega}_{BC}}{\sqrt{\text{Var}(\hat{\omega}_{BC})}}$ to the normal distribution. It is clear that the test statistic can have low power. If the direct evidence has variance ω^2 then the test statistic will have variance $3\omega^2$. Published in 1997 this was one of the first papers to make use of indirect comparisons while still preserving the randomisation of treatment groups. However, in a large network with a lot of treatment loops this approach can be both cumbersome and time-consuming (Dias et al., 2013b).

An indirect comparison may be estimated through several independent treatment loops within a network. Caldwell (2010) used a method similar to Bucher, still within triangular networks, in which a chi-squared test was used to provide a composite test of the null hypothesis that the indirect treatment effect estimates were consistent across all treatment loops. Meanwhile Dias (2010b) extended the Bucher method to apply to any network, not just triangular networks, and constructed a test statistic which could also be compared to a normal distribution.

One of the most popular models to account for inconsistency in a network is the Bayesian hierarchical model of Lu & Ades (2006). This AD model relaxes the consistency assumption by including an extra random effect in each loop in which inconsistency could occur. These additional random effects are referred to as inconsistency parameters and are assumed to come from a common normal distribution with mean 0 and variance σ^2_ω. Models with and without inconsistency parameters can be compared to assess whether a network is consistent.

Song (2012) conducted a simulation study to evaluate three approaches for detecting inconsistency between direct and indirect comparisons, in a simple network of three treatments and two-arm trials (such as Figure 1.1). The three methods considered were Bucher (1997), consistency NMA (Lu and Ades, 2004) and random inconsistency NMA (Lu and Ades, 2006). Song concluded that although the three methods were, on average, unbiased for estimating inconsistency the random inconsistency NMA had much wider 95% intervals and therefore, had lower power to detect inconsistency in a network.

Higgins (1996) and Veroniki (2013) both distinguish between two types of inconsistency in a network: loop inconsistency and design inconsistency. Loop inconsistency is when the

direct and indirect treatment effect estimates in a loop are not consistent with each other. Design inconsistency is when there are differences in effect sizes between studies involving different sets of treatments.

A more recent method of assessing consistency is the design-by-treatment interaction model proposed by Higgins (2012). The design-by-treatment interaction model is an extension of the Lu & Ades (2006) hierarchical model and assesses both loop and design inconsistency simultaneously. The model evaluates inconsistency in the network as a whole through an extension of multivariate meta-regression in which different treatment effects are allowed for trials with different designs. This follows the approach of Lumley (2002) which is a design-by-treatment interaction model for two-arm trials only.

Two computational strategies for fitting the design-by-treatment interaction model and estimating inconsistency in a network including multi-arm trials were provided by White (2012). Both strategies fitted consistency and inconsistency models as multivariate meta-regressions. These two fitting strategies were termed estimation by the standard approach (assuming no treatment is common to all trials) and estimation by data augmentation (assuming one treatment is common to all trials). Both methods are frequentist two-stage estimation procedures. Based on these two strategies for fitting the design-by-treatment interaction model, Stata code (White et al., 2012; White, 2015) was made available which has helped popularise the approach. A global Wald test statistic, of all the inconsistency parameters estimated in the model, can be used to test the assumption of consistency in the network.

Estimation by the standard approach involves identifying a reference treatment for each individual trial and then estimating the contrasts (e.g. log odds ratio) between the reference and non-reference treatments for each trial. Estimation by data augmentation involves identifying a treatment that will be considered the reference treatment for all trials regardless of whether the treatment was included in each trial. The data augmentation technique then introduces the reference treatment as an arm containing a very small amount of data in the trials that did not include the reference treatment. Both approaches can be fitted in a two-step procedure using the Stata (StataCorp, 2015) command *network* (White, 2015), which itself uses *mvmeta* (White et al., 2012).

A special case of the design-by-treatment interaction model, introduced by Jackson (2014), includes the inconsistency parameters as random effects so that the inconsistency across

different sources is considered as additional variation. If the inconsistency parameters equate to zero then the consistency assumption is satisfied. Unlike Higgins (2012), the inconsistency parameters are modelled as random effects, which facilitates the ranking of treatments under inconsistency. This model has the added benefit of incorporating both between-study heterogeneity and inconsistency.

An alternative method splits the NMA approach into two stages (Lu et al., 2011). The first step performs MA amongst groups of trials comparing the same two treatments and the second stage aims to find an estimate of the treatment effect that fulfils the consistency equations. Various methods could be used for the first stage including fixed effect, random effect or Bayesian hierarchical models. The second stage uses the Lu & Ades (2006) method to test the consistency of the network by deriving the likelihood ratio statistic.

An alternative method of assessing inconsistency within a Bayesian framework is node-splitting. For the node of interest, the direct evidence (e.g. trials directly comparing treatments X and Y) and indirect evidence (e.g. trials that do not directly compare treatments X and Y) are split into separate independent components. A posterior distribution is generated for the mean treatment effect from the direct evidence, (e.g. d_{XY}^{dir}), and for the mean treatment effect from the indirect evidence, (e.g. d_{XY}^{ind}). The amount of agreement between these two sources is measured by examining the posterior distribution of the inconsistency parameter (e.g. $\omega_{XY} = d_{XY}^{\text{dir}} - d_{XY}^{\text{ind}}$). Agreement between the direct and indirect evidence is calculated by counting the number of MCMC iterations where $\omega_{XY} > 0$ (Dias et al., 2010b). Essentially, this method compares a model where the consistency assumption is relaxed for one treatment comparison to the model assuming consistency across the entire network, which can highlight inconsistent treatment comparisons within the network. One of the advantages of this method is that it allows each treatment comparison to be considered separately and one at a time for evidence of possible inconsistency. Additionally it can also be applied to networks of any size (Dias et al., 2013b). However, similarly to Bucher, this method can be both cumbersome and time-consuming in a network with a large number of treatments because each treatment node is considered one at a time.

So far I have discussed several methods which include inconsistency parameters, in a design or loop, as a way of identifying inconsistency in a network. When an inconsistency parameter (design or loop) is added to a model, the design or loop is effectively removed from contributing to the estimate of the indirect evidence, whichever method is used.

To aid the identification of inconsistency within a network Krahn (2013) developed a method, known as a net heat plot, which could be used as a visual aid for locating and identifying any inconsistency within a network of randomised controlled trials (RCTs). They use a FTE model for NMA within the generalised linear models framework. For designs where a treatment loop is formed the net heat plot is constructed by detaching each design one at a time and assessing the contribution of each design to the inconsistency. The net heat plot takes the form of a matrix in which the colouring of component squares indicates designs which increase or decrease inconsistency within the network. This approach is considered further in Chapter 5.

Piepho (2014) considered factorial analysis-of-variance models for conducting NMA and assessing inconsistency within a network. Within the factorial analysis-of-variance framework they showed that their approach could detach each design one at a time and assess the contribution of the design to the inconsistency across the whole network as well as the inconsistency that remains in the network when the design is detached. The factorial analysis-of-variance approach models treatment means, rather than the baseline contrasts of the Krahn (2013) approach. However, the two approaches to assessing inconsistency in a network can give the same results.

If covariates are distributed unevenly between trials then one method of reducing inconsistency between treatment comparisons is to adjust for covariates using methods such as those from Cooper (2009) and Donegan (2012), described in Section 2.7.

2.5 Heterogeneity

Conventionally, Cochran's Q statistic has been used to identify heterogeneity in MA (Cochran, 1954). However, there are problems with this method including low power to determine heterogeneity, so that a non-significant result does not mean there is no heterogeneity in the network (Higgins et al., 2003). Higgins (2003) suggested I^2 as an alternative method of quantifying the amount of heterogeneity present in a MA. However, as I^2 is calculated from Cochran's Q statistic it also has low power to detect heterogeneity. I^2 calculates the percentage of variation due to heterogeneity across trials rather than chance and can be helpful in investigating the causes and types of heterogeneity, such as methodological subgroups, effect measures or clinically important subgroups. However as the number of

patients included in the trials in a MA increases so does I^2 (Rücker et al., 2008). Rücker (2008) presented H^2, R^2 and τ^2 as alternative heterogeneity statistics. However, H^2 and R^2 have the same problem as I^2, where they increase as the number of patients increase. Rücker concluded that τ^2, the between-study variance, may be a better measure of heterogeneity because it has a clinical meaning. In addition, Jackson (2012) extended I^2, H^2 and R^2 from the univariate MA case to the multivariate NMA case. I^2, H^2 and R^2 do not solve the low power problem of Cochran's Q statistic, instead they provide an alternative interpretation. The generalised Cochran's Q statistic for multivariate meta-analyses (Gasparrini et al., 2012) can be used in the context of NMA to quantify heterogeneity across the whole network, within trial designs and between trial designs (also known as inconsistency). However, it also suffers from the same problem of low power to detect heterogeneity.

Random effects MA can be used to incorporate the additional uncertainty associated with heterogeneity by assuming that the underlying treatment effects from trials comparing the same treatments come from a common distribution (Ciprani et al., 2013; Higgins et al., 2012; Mills et al., 2013). While this allows for heterogeneity it does not explain it (Cooper et al., 2009). Cooper (2009) and Jansen (2012) both considered including covariates in a NMA model to explore potential sources of heterogeneity and reduce inconsistency. Inclusion of covariates in a model allows the treatment effects to vary with the covariates potentially explaining any systematic variability between trials (Cooper et al., 2009). Therefore it is important to check whether heterogeneity can be explained by any variability in the treatment-covariate interactions (Cooper et al., 2009).

A four-part process for assessing the feasibility of a NMA was proposed by Cope (2014) as a way to assess heterogeneity and inconsistency within a network. The process involves visualising the clinical heterogeneity in terms of treatment and outcome characteristics and also in terms of study and patient characteristics before assessing differences within and across observed treatment effects and within and across baseline risk for all direct comparisons.

The design-by-treatment interaction model, first proposed by Higgins (2012) but extended by Jackson (2014) to include the inconsistency parameters as random effects, can incorporate both inconsistency (Section 2.4) and between-study heterogeneity in the model. The inconsistency parameters are design-specific and allow the effect of two treatments to differ between different designs. For example, the effect of treatment A versus treatment

B can differ between a design including treatments A, B and C to a design including treatments A and B. The between-study heterogeneity, τ^2, is assumed to be the same for all treatment comparisons in each trial so that the variance of the heterogeneity parameters is a square matrix with entries τ^2 on the leading diagonal and all other entries $\frac{\tau^2}{2}$. As mentioned previously, this assumption may be implausible in some situations. However, it offers a simpler approach requiring the estimation of only one parameter when there is often relatively little information available to estimate an unstructured covariance matrix. The impact of the between-study heterogeneity and inconsistency can be quantified using the multivariate R^2 and I^2 statistics (Jackson et al., 2012).

2.6 Bias

There are many biases which can operate within a NMA including publication bias, selection bias, selective reporting, optimism bias and sponsorship bias (Salanti et al., 2007). In head-to-head trials, bias can be introduced through lack of allocation concealment, inadequate blinding and imbalanced withdrawals (Song et al., 2008); and within a network, bias can cause inconsistency.

Song (2008) used the method of Bucher (1997) to compare direct and adjusted indirect estimates of treatment effects in three different networks. Song hypothesised that there might be situations where adjusted indirect treatment estimates are less biased than direct estimates particularly when there is bias within a network. Song concluded that where treatment effects from direct estimates are different to the adjusted indirect estimate there are several possibilities which could explain this phenomenon including chance, bias in head-to-head comparisons, bias in adjusted indirect comparisons (which are generally more unpredictable and extreme than direct estimates) and clinically meaningful heterogeneity across trials. Therefore it is important to consider potential sources of bias within a network.

Optimism bias, also known as novelty bias, where the new treatment is favoured over the older treatment can be caused by inadequate allocation concealment (Dias et al., 2010c). Dias (2010c) proposed a model which extended the standard NMA model to incorporate a trial-specific bias term. The model could then adjust the treatment effect estimates for any bias and estimate the overall mean bias simultaneously. The risk of bias in a trial is characterised as a probability. Trials known to be at risk of bias, e.g. because they

have inadequate allocation concealment, are given a probability of one. Trials known not to be at risk of bias are given a probability of zero. Trials where the risk of bias is unknown, e.g. because the quality of allocation concealment can not be determined from the trial publication, are given a probability of bias, which is drawn from a Bernoulli distribution. The model can be extended further to take into account whether bias exists between active and control treatments or between active treatments only.

Optimism bias can be confounded with sponsorship bias, where the magnitude and direction of treatment effects are influenced by who the trial sponsor is. Quality-related biases (allocation concealment, blinding, randomisation) can also be related to optimism bias and can therefore also be confounded with sponsorship bias (Dias et al., 2010a). To include industry sponsored and non-industry sponsored trials in the same network Cipriani (2009) included a covariate for sponsor effect in their NMA model.

Salanti (2010) introduced the idea of 'novel agent effects' which incorporate optimism bias and other reasons beyond bias for why newer treatments may be preferred. They estimated novel agent effects by extending the NMA meta-regression framework (Higgins and Whitehead, 1996; Lu and Ades, 2004; Salanti et al., 2007) to include adjustment for novel agent effects and used Bayesian methods.

Small studies may differ from larger studies due to characteristics that might affect the effectiveness of the treatment (e.g. small studies have higher-risk patients). 'Small study effects' is the term used when treatment effects from small studies are exaggerated in a NMA (or MA). However publication bias and selective reporting are also possible explanations for an association between study size and treatment effect (Chaimani and Salanti, 2012).

A network meta-regression model where the treatment effect size is allowed to depend on its standard error can be used to account for small study effects. The regression slope reflects the magnitude of the association of effect size and precision, the small study effect (Chaimani and Salanti, 2012; Trinquart et al., 2012). The direction of the effects needs to be considered and Chaimani (2012) suggested two approaches. In the first approach the small study effects exaggerate the treatment effect in the active treatment when compared to the placebo only, so that no treatment is favoured in a trial of two active treatments. In the second approach the newer treatment is always exaggerated by the small study effects. A third approach suggested by Dias (2010c) involves applying the prob-

ability that a treatment is favoured to each comparison. The model can be applied so that the small study effects are either identical or exchangeable across treatment comparisons. Alternatively Trinquart (2012) suggested using the NMA model of Lu & Ades (2004) and including a selection model which adjusts for publication bias and assumes that the small study effects are exchangeable across the network.

Aggregation of IPD to obtain a study-level variable which is then used to define different groups of patients within a study or by risk factors associated with the group variable can result in ecological bias (Greenland and Morgenstern, 1989). Covariates not associated with treatment, because of randomisation, can still cause ecological bias (Govan et al., 2010). Ecological bias can be a problem in trials which collapse covariate categories. Govan (2010) described an approach which estimated treatment and covariate effects whilst simultaneously controlling for ecological bias and allowed the covariate distributions within trials to be estimated. They then extended the Lu & Ades (2004) model for NMA to obtain the collapsed categories.

There are many methods for assessing publication bias including funnel plots, formal statistical tests, the non-parametric trim-and-fill method and selection models. The Copas selection model, in which the probability of a trial being published is related to the trial's effect size and precision, is superior to trim-and-fill however it may not fully eliminate bias (Mavridis et al., 2013). Mavridis (2013) considered a fully Bayesian implementation of the Copas selection model for a star-shaped network of two-arm trials in which all treatments were compared to the reference treatment only.

Assessing the quality of evidence from a NMA has become important in recent years and can be considered as a form of bias assessment. The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group described a four-step approach for NMA to determining the quality of treatment effect estimates which considered five components: study limitations, inconsistency, indirectness, imprecision and publication bias (Puhan et al., 2014). Salanti (2014) took this four-step approach and assessed each component to be of high, moderate, low or very low quality. Each component was ranked separately for two outcomes; pairwise comparisons and overall ranking. These were then summarised across all five components to obtain a confidence in each (pairwise) effect size and a confidence in ranking of treatments. Caution should be exhibited when using trial reports only to assess risk of bias. A review using the Cochrane risk of

bias tool to compare the reliability of trial reports only with trial reports with additional information collected for IPD MA, concluded that using trial publications alone, to assess risk of bias in a trial, could be unreliable (Vale et al., 2013).

Finally, both Cooper (2009) and Jansen (2012) included treatment-covariate interactions as a method for reducing inconsistency in a network. Benefits of this approach include reducing inconsistency and adjusting for confounding bias if there are systematic differences in covariates across trials (Cooper et al., 2009; Jansen and Cope, 2012).

2.7 Treatment-covariate interactions

Treatment effects can be modified by covariates. The inclusion of treatment-covariate interactions in a NMA model seeks to estimate how much the covariate modifies the treatment effects (Donegan et al., 2012). Including treatment-covariate interactions can reduce both inconsistency and heterogeneity in a network.

The first paper to consider including study-level covariates within the NMA framework was reported by Nixon (2007). Nixon introduced a method based on the NMA approach of Lu & Ades (2004) and meta-regression techniques (Thompson and Higgins, 2002). This incorporated covariates in a model where all treatments were compared to the same reference treatment, and demonstrated its application to a rheumatoid arthritis dataset. Cooper (2009) then extended this approach to a more general framework in which any treatment could be considered as the reference treatment, and introduced three different ways to model random effects for treatment-covariate interactions: common, independent or exchangeable. Common random effects assume that the regression coefficients are the same for all treatment-covariate interactions so that the covariate effect is the same for each treatment compared to the control. Independent random effects assume that all treatment-by-covariate interactions are different and unrelated for each treatment versus the control so that a separate regression coefficient for each treatment is included in the model. Exchangeable random effects assume that all treatment-covariate interactions are different from each other but similar enough that they can be sampled from the same distribution.

The methods of both Nixon (2007) and Cooper (2009) considered AD only. Therefore

Donegan (2012) extended the models of Cooper (2009) to use IPD and allow for patient-level outcomes and treatment-covariate interactions, which could be independent, exchangeable or common. They then applied the methods of Riley (2008), previously used in MA, to separate the within and across trial information for the treatment-covariate interactions. Donegan (2012) specified models which were only applicable to two-arm trials however they note that the models could be adapted to include multi-arm trials.

Jansen (2012) extended the methods of Jansen (2011) and Ouwens (2010) for random effect NMA of aggregate survival data using fractional polynomials by including either treatment-specific covariate interactions or a constant covariate interaction. Including treatment-covariate interactions allows the model to adjust for confounding bias if there are systematic differences in covariates across trials.

2.8 Modelling time-to-event data

NMA models are widely available for continuous and binary outcomes but are considerably less common for time-to-event outcomes. Traditionally censored survival time data has been analysed using the Cox PH model (Cox, 1972) and MA (and NMA) is conducted using a two-step method in which the hazard ratio and a measure of uncertainty for each trial are pooled together in a MA (or NMA). The Cox PH model is a semi-parametric model which makes no assumption about the baseline hazard rate and requires the hazard rates between treatment arms to be proportional over time (Cox, 1972). In oncology trials with longer follow-up of trials, and time-dependent effects of targeted treatments, there is increasing evidence of non-PH so this may no longer be appropriate (Royston and Parmar, 2016; Trinquart et al., 2016). Models which could offer an alternative approach to modelling time-to-event data and could be used in a one-step NMA have been suggested by Royston (2002), Crowther (2012; 2014), Jansen (2011) and Saramago (2014).

One approach to modelling time-to-event data which provides a parametric alternative to the Cox PH model is the Royston-Parmar model (Royston and Parmar, 2002). The Royston-Parmar model uses a restricted cubic spline function of log time to model the (baseline) log cumulative hazard rate (Royston and Parmar, 2002). The complexity and flexibility of the model is determined by the restricted cubic spline and model parameters are estimated using maximum likelihood (Lambert and Royston, 2009). The Royston-

Parmar model provides a flexible framework which can be easily extended to assess non-PH and incorporate covariates and treatment-covariate interactions.

Crowther (2014) proposed a one-step multilevel mixed effects parametric survival model for IPD MA which extended the Royston-Parmar (Royston and Parmar, 2002) approach, of modelling the (baseline) log cumulative hazard rate using a restricted cubic spline function of log time, to include random treatment effects and time dependent effects. The inclusion of time dependent effects relaxes the assumption of PH. Gauss-Hermite quadrature was used to estimate the model parameters using maximum likelihood.

Jansen (2011) used the fractional polynomials approach of Royston & Altman (1994) to model the log hazard function. Jansen showed that for NMA the consistency equations must hold on the log hazard scale and went on to specify the two-step random effects NMA model for time-to-event data using a fractional polynomial. Jansen (2012) then extended this method and the work of Ouwens (2010) to include treatment-covariate interactions which allowed the model to adjust for confounding.

Crowther (2012) considered the use of Poisson regression models as an alternative to the Cox model for analysing IPD time-to-event data in a MA. Crowther used Poisson generalised linear survival models in a one-step analysis using IPD. These models can be implemented with either fixed or random treatment effects and with the baseline hazard stratified by trial. Crowther extended these models to include treatment-covariate interactions and to allow non-PH of the treatment effects.

Saramago (2014) proposed a fixed effect NMA model for time-to-event outcomes which jointly synthesised IPD and AD and controlled for baseline covariates, under the assumption that event times were Weibull distributed, using a Bayesian framework. Saramago showed that their basic model could be extended and applied to other situations, for example, by including treatment-covariate interactions or by changing a number of assumptions such as random treatment effects or the shape parameter.

Finally, an alternative to the hazard ratio for quantifying treatment effects with time-to-event data is restricted mean survival time (Royston and Parmar, 2011). Wei (2015) considered conducting a MA of restricted mean survival data. As of October 2016, this is yet to be applied to a NMA.

2.9 Further approaches to estimation

The main focus of this literature review was to identify the main methods of dealing with inconsistency, heterogeneity, bias and treatment-covariate interactions. However there are plenty of other methodological advances that have been made in NMA since the publication of the first NMA model by Lumley (2002). Since 2002 models have been developed for NMA that consider accounting for prior knowledge, repeated measurements, combining IPD and AD, definitions of treatment nodes and many others.

Conducting NMA in a Bayesian framework requires the specification of prior distributions for parameters which are typically chosen to be non-informative. Lu & Ades (2009) built on their previous work on Bayesian hierarchical models (Lu and Ades, 2004, 2006; Lu et al., 2007) to model the between-trial variance in a way that was both compatible with the assumption of consistency and could incorporate prior knowledge of the correlation between treatment arms. Thorlund (2013) considered prior distributions for variance parameters in a random effect NMA. Four weakly informative priors and two moderately informative priors were applied to two datasets and a range of models with differing variance assumptions were considered. Models considered included homogeneous variance, unrestricted heterogeneous variances and exchangeable variances. The authors found that precision can be gained by incorporating informative variance priors.

Methods for adapting NMA models to estimate treatment effects for repeated measurements have been proposed by both Dakin (2011) and Ding (2013). Dakin (2011) proposed a model, assuming a normal likelihood, which has been adapted to allow for repeated measurements of a continuous outcome (Ades et al., 2006; Higgins and Whitehead, 1996; Salanti et al., 2009; Welton et al., 2009). By contrast, Ding (2013) proposed a longitudinal Bayesian hierarchical model which automatically modelled correlations across different time points and was equivalent to the Lu & Ades (2004) model. Ding also extended the 'integrated two-component prediction' model of Fu (2010) to handle different shapes of longitudinal profiles.

Jansen (2012) and Saramago (2012) both considered extending the one-step MA methods of Sutton (2008), for combining IPD and AD, to improve the precision of treatment effect estimates over IPD only models. Saramago (2012) also considered the inclusion of patient-level covariates.

One property of multi-arm trials is that they must be internally consistent. Previously, Lu & Ades (2006) showed that the correlation between treatment arms in multi-arm trials must be taken into account in a NMA which includes multi-arm trials when contrast-based summaries are used. Franchini (2012) extended this approach to show that in a network with multi-arm trials the results of a NMA of arm-level and contrast-level summaries will differ unless an adjustment to the likelihood is also made when using contrast-level summaries.

Del Giovane (2013) considered models for NMA that account for variations in the definition of treatment nodes e.g. sub nodes for different doses. Their basic NMA model followed the method of Lu & Ades (2004) and Dias (2010b) but accounted for multi-arm trials using the method of Higgins and Whitehead (1996).

NMA has also been considered for use within the cost-effectiveness research setting and for use in competing risks analysis (Achana et al., 2013; Ades et al., 2006, 2010; Cooper et al., 2011; Dias et al., 2013a).

Furthermore, methods for conducting NMA within the framework of Markov models, generalised linear mixed models, using non-inferiority trials, incorporating evidence from different trial designs (e.g. RCTs and cohorts) and using adverse event outcomes have been considered by Price (2011), Tu (2014), Schmidli (2013), Schmitz (2013) and Warren (2014) respectively.

2.10 Bayesian & Frequentist Methods

Since 2009 there has been a large increase in the number of published NMA articles, in both the Bayesian and frequentist settings (Lee, 2014; Nikolakopoulou et al., 2014; Sobieraj et al., 2013). NMA conducted in the Bayesian setting is often implemented through the statistical software package WinBUGS (Lunn et al., 2000). The Bayesian framework naturally handles random effects (avoiding awkward numerical integration) and can offer many other benefits including treatment effect estimates for treatments never compared directly, easy assessment of network consistency, a natural framework for ranking treatments, incorporation of prior knowledge and the ability to adjust for correlations which arise from the inclusion of multi-arm trials (Ades et al., 2006; Dominici et al., 1999; Ioannidis, 2009; Lu and Ades, 2006). The rise in popularity of frequentist methods has been

fuelled by the creation of the Stata suite of programs *mvmeta* (White, 2011) and *network* (White, 2015) and a series of graphical tools for NMA (Chaimani et al., 2013).

Hong (2013) compared Bayesian and frequentist approaches for multiple outcomes NMA to a dataset on incontinence and concluded that the Bayesian approach was preferred and more straightforward conceptually, used all available data, had easier interpretability of results, was easier to extend simple models to more complex models and allowed a natural method of ranking treatments.

2.11 Reporting of NMA

In line with the increase in the number of published NMA articles there have also been several articles published, in recent years, which provide an overview of NMA, including the basic assumptions of consistency and heterogeneity and instructing the reader on how to interpret the results (Belsey, 2015; Berlin and Cepeda, 2012; Bhatnagar et al., 2014; Biondi-Zoccai et al., 2015; Caldwell, 2014; Ciprani et al., 2013; Mills et al., 2012).

Laws (2014) compared the national guidelines on conducting and reporting NMA from nine countries (Australia, Belgium, Canada, England and Wales, France, Germany, Scotland, Spain and South Africa) focusing on the design, conduct and reporting of the trial search, selection of databases, study selection, bias assessment and conduct of NMA. Currently no transnational guidelines for NMA exist and the guidelines that do exist for the nine countries considered had a few main points in common and then differed in their requirements. Most countries required specification of search terms for a systematic review, MEDLINE, Embase and Cochrane (CENTRAL) to be searched as part of the systematic review, inclusion and exclusion criteria predefined, assessment of homogeneity between trial populations in terms of prognostic factors, assessment of blinding adequacy, rationale for and description of sensitivity analyses and presentation of results as relative risk, hazard ratios or odds ratios. In reporting a NMA, Germany was the only country to require a description of the design and methodology of each trial according to the CONSORT guidelines. They also required specification of model code and the software package used for analysis. The guideline for England and Wales was the only one to require treatment effect modifiers to be identified before comparing study results from a NMA. France was the only country to require an analysis of the hypothesis of consistency.

The first set of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were published in 2009 updating the 1999 QUOROM Statement (Quality of Reporting of Meta-analyses). In 2015 two new sets of PRISMA guidelines were published (Hutton et al., 2015; Stewart et al., 2015). The first focuses on reporting of IPD analyses (Stewart et al., 2015), while the second focuses on reporting of NMA (Hutton et al., 2015). The NMA guidelines include two new items for the methods section and three new items for the results section. In the methods section a description of the methods for assessing the network geometry and inconsistency are required. In the results section a graphical representation of the network structure, an overview of the characteristics of the network and the results from assessing inconsistency are required. In addition some other items were modified or extended to be appropriate for NMA and a detailed explanation and example of all the checklist items is presented. NMA has the potential to inform the future research agenda. However, to maximise this potential appropriate reporting of NMA is necessary.

2.12 Summary

This literature review has demonstrated that NMA models have been around for the last 15-20 years, particularly for binary and continuous outcomes. In this literature review I identified a wide range of papers concerning NMA methodology. There were some papers that considered an array of methods for one particular issue, such as inconsistency or bias and there were some papers which attempted to provide an overview of the field for a non-technical audience. However, the greatest proportion of papers were technical exploring one particular modelling approach for NMA. I found no systematic review of the development of the methodology for NMA. Therefore, this literature review could also be helpful to other researchers new to the field of NMA.

The methodology for binary and continuous outcomes is reasonably well developed and can incorporate a wide range of scenarios allowing the analysis to be adjusted for inconsistency, heterogeneity, bias and treatment-covariate interactions. In contrast, Section 2.8 highlights a need for continued development of methodology to conduct NMA with time-to-event data. To date, little work has been done on using the Royston-Parmar model and exploring its inherent flexibility for NMA in a one-step approach which could,

at least potentially, include extensions to consider handling of non-PH, assessment of inconsistency, inclusion of treatment-covariate interactions and handling of missing covariate data. I aim to address these issues throughout this thesis. In Chapter 4 I describe the methodology for the one-step IPD Royston-Parmar NMA model and illustrate its potential through application to two networks introduced in Chapter 3. In Chapter 4 I also consider how the one-step IPD Royston-Parmar NMA model can be extended to handle non-PH and assess inconsistency. A detailed examination of assessing inconsistency in NMA will follow in Chapter 5. In Chapter 6 I return to the one-step IPD Royston-Parmar NMA model and consider how to handle missing covariate data before describing an extension of the model to include treatment-covariate interactions.

3 Datasets & Initial Analyses

3.1 Introduction

In Chapter 2 I reported a literature review of the principal approaches to, and models for, NMA. In practice, conducting a NMA requires careful examination of the data before any NMA model is fitted. In this chapter I will conduct a thorough exploration of two networks of RCTs examining the data, conducting exploratory analyses and performing two-step MA.

This chapter is structured as follows. In Section 3.2 I introduce two datasets which will be used throughout the remainder of this thesis to apply and illustrate the methods described. Both datasets provide IPD and consider the time-to-event outcome overall survival. In Section 3.3 I will look at the baseline characteristics and pattern of follow-up across all trials, assess the PH assumption in each trial and check for evidence of publication bias. In Section 3.4 I will synthesise the evidence from the individual trials using a two-step MA approach for each pairwise treatment comparison and examine each comparison for evidence of statistical heterogeneity. Finally, Section 3.5 will summarise the findings of this chapter.

3.2 Description of the datasets

3.2.1 Cervical cancer

The first network is formed of trials from three meta-analyses of RCTs in cervical cancer performed by the Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration (2008) and the Neoadjuvant Chemotherapy for Cervical Cancer Meta-Analysis Collaboration (2003). These data were obtained from the MRC CTU at UCL. The three meta-analyses considered four different treatments: radiotherapy (RT), chemoradiation (CTRT), neoadjuvant chemotherapy plus radiotherapy (CT+RT) and neoadjuvant chemotherapy plus surgery (CT+S, Figure 3.1) using four different designs: RT v CTRT, RT v CT+RT, RT v CT+S and RT v CT+RT v CT+S. The Neoadjuvant Chemotherapy for Cervical Cancer Meta-Analysis Collaboration (2003) conducted one systematic review to consider two related but separate treatment comparisons: RT v CT+RT and RT v CT+S. Trial accrual periods ranged from 1982 to 1999. The Chemoradiotherapy for Cervical Cancer

Meta-Analysis Collaboration (2008) conducted one systematic review to compare RT and CTRT. Trial accrual periods ranged from 1987 to 2006. Both systematic reviews were completed following detailed pre-specified protocols.
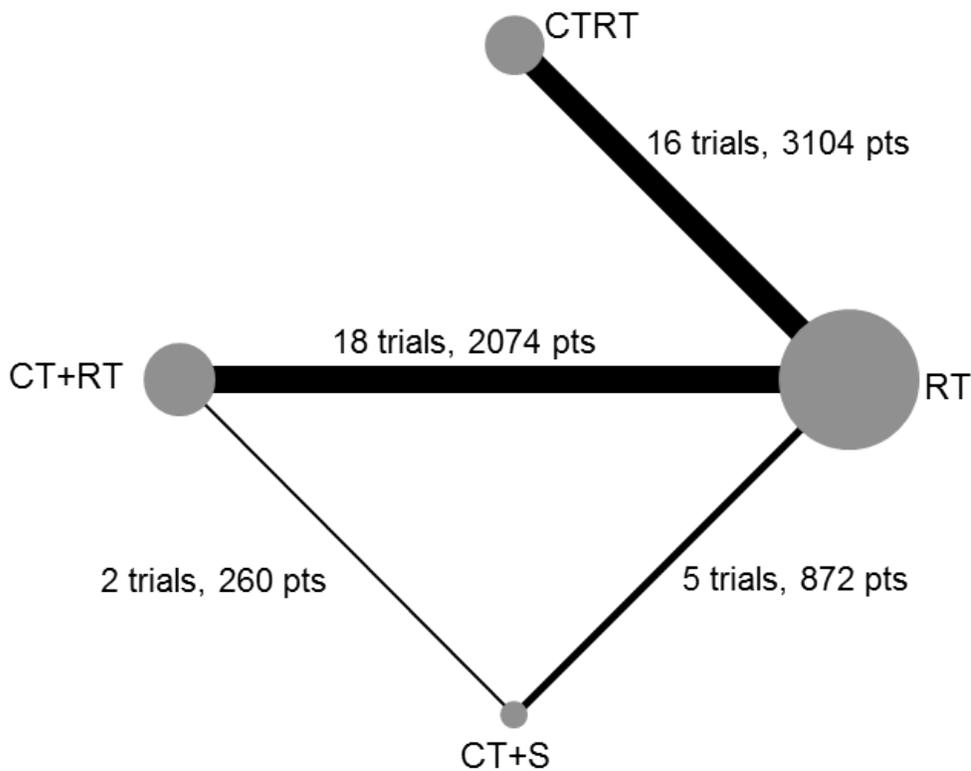
Figure 3.1: Cervical cancer network diagram. Node size is proportional to the number of patients randomised to each treatment and line thickness is proportional to the number of studies involved in each direct comparison. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery. Note that this network diagram includes the main set of 13 RT v CTRT trials only, and the number of patients for each treatment arm does not add up to the total number of patients included in the network, because multi-arm patients are counted twice. There are a total of 37 trials in this network. However, in the figure the two multi-arm trials are counted three times each as they are included in the number of trials for each pairwise comparison.

In the three original meta-analyses (Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration, 2008; Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration, 2003) all analyses were intention-to-treat and the individual times to event were used to obtain log-rank hazard ratio estimates of treatment effect for the individual trials,

which were then pooled across trials using a stratified-by-trial FTE model. Chi-squared heterogeneity tests were used to assess statistical heterogeneity across trials. Subgroup analyses were conducted using stratified log-rank analyses in which hazard ratio estimates were obtained for each predefined subgroup within each trial before being pooled across trials. Chi-squared tests of interaction or trend were used to test for differences in treatment effectiveness between subsets of trials or subgroups of patients. This is not an ideal approach for subgroup analyses as the pooling of within and across trial information can result in ecological bias. Subgroup analyses and ecological bias are considered in more detail in Chapter 6.

The RT v CTRT comparison was published in 2008 and included a total of 18 RCTs and 4818 patients (Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration, 2008). However, in the original publication five trials were excluded from the main analysis as patients on the control arm received additional treatment or the comparison was otherwise confounded. In two trials (Kantazardic et al., 2004; Peters et al., 2000) patients received additional adjuvant chemotherapy alongside CTRT. In two trials (Rose et al., 1999; Whitney et al., 1999) patients in the RT arm received additional hydroxyurea and in one trial (Morris et al., 1999) patients on the RT control arm received extended field RT as well as standard RT. Therefore, a subset of 13 trials (3104 patients) was identified and used for the main analysis. As in the original publication, I will only consider the main set of 13 trials in the remainder of this thesis. Within the RT v CTRT comparison, two three-arm trials combined two different forms of CTRT and compared them with a single control arm and three four-arm trials were split into two unconfounded comparisons of RT v CTRT for analysis as separate trials. The data will be treated in the same way throughout this thesis giving a total of 16 RCTs of RT v CTRT in the network (3104 patients). Figure 3.2 recreates the forest plot of log hazard ratio and 95% confidence intervals for each trial from the original publication (Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration, 2008).

The RT v CTRT MA identified a 19% reduction in the risk of death for CTRT (hazard ratio=0.81, 95% confidence interval: 0.71, 0.91) in the main set of 13 trials which represented an absolute survival benefit of 6% at 5 years. No evidence of a difference in effect size was found when trials were grouped according to type of chemotherapy (platinum-based or non-platinum based), planned RT dose or total planned duration of RT. In trials using cisplatin-based CTRT no evidence of a difference in effect size was found when trials were

grouped according to cycle length or dose intensity of cisplatin. Patient subgroup analyses found no evidence to suggest that the effect of CTRT differed in groups of patients defined by age (test for trend p=0.436), histology (test for interaction p=0.992), tumour grade (test for interaction p=0.961) or pelvic node involvement (test for interaction p=0.483). There was a suggestion of trend in the effect of CTRT by stage of disease (χ^2=5.65, p=0.017). The direction of treatment effect was consistent in the analysis of disease-free survival however the p-value did not reach statistical significance (χ^2=3.21, p=0.073).

The RT v CT+RT comparison was published in 2003 and included 18 RCTs and 2074 patients (Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration, 2003). Two trials included in this comparison were originally three-arm trials comparing RT, CT+RT and CT+S. However, only the patients receiving RT and CT+RT are included in the 2074 patients in this comparison. The forest plot of log hazard ratios and 95% confidence intervals from each trial, from the original publication (Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration, 2003), is re-created in Figure 3.2. A large amount of heterogeneity was identified within this comparison, which appeared to be best accounted for by grouping trials by chemotherapy cycle length. Trials with chemotherapy cycle lengths longer than 14 days had a hazard ratio (HR) of 1.25 (95% CI: 1.07, 1.46) and for shorter cycle lengths the HR was 0.83 (95% CI: 0.69, 1.00). The overall HR was 1.05 (95% CI: 0.94, 1.19). Subgroup analyses were conducted within the trial groups defined by chemotherapy cycle length, and there was no evidence to suggest a difference in effect size in groups of patients defined by age, stage, histology, tumour grade or performance status.
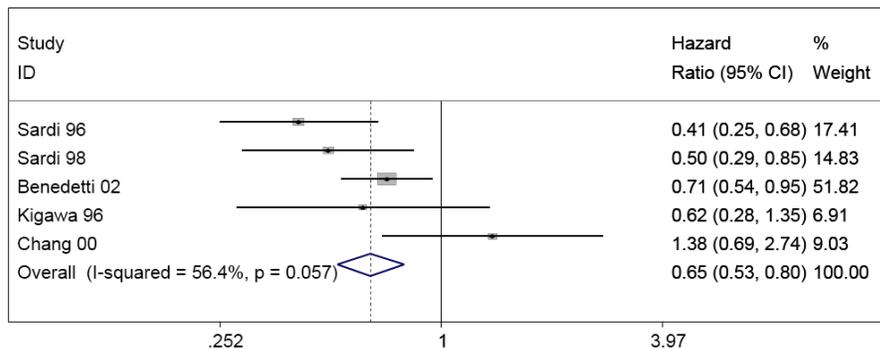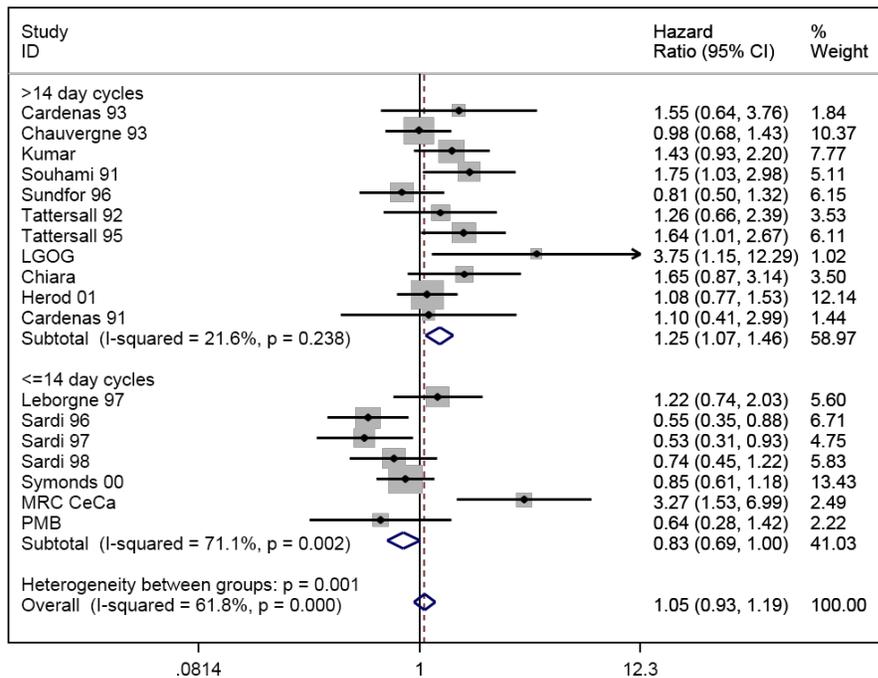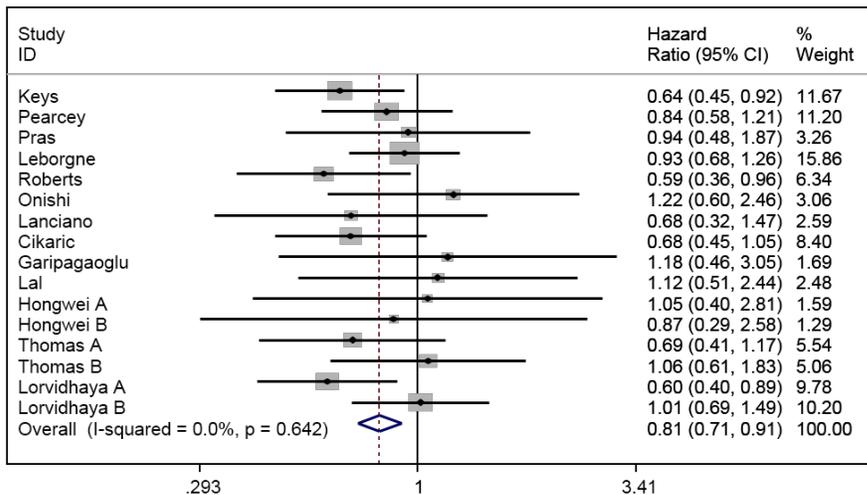
Figure 3.2: Cervical cancer forest plots. Results come from a Mantel-Haenszel FTE model. Top: RT v CTRT, Middle: RT v CT+RT, Bottom: RT v CT+S. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chmeotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery, CI = confidence interval.

The RT v CT+S comparison, also published in 2003, included five RCTs and a total of 872 patients (Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration, 2003). This comparison also included the two trials which were originally three-arm trials comparing RT, CT+RT and CT+S. Only patients receiving RT and CT+S are included in the 872 patients analysed in this comparison. Figure 3.2 recreates the forest plot of log hazard ratios and 95% confidence intervals for each trial from the original publication (Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration, 2008). The HR in favour of CT+S was 0.65 (95% CI: 0.53, 0.80) indicating a 35% reduction in the risk of death with CT+S. A small amount of heterogeneity was identified between the five trials, mainly due to one trial with a result quite different to the other four trials. There was no evidence to suggest that the effect of CT+S varied across groups of patients defined by age (test for trend p=0.363), stage of disease (test for trend p=0.258), histology (test for interaction p=0.082), tumour grade (test for trend p=0.781) or performance status (test for interaction p=0.713).

In total, I used overall survival data for 5922 patients from 37 RCTs. Covariate data were provided for age at randomisation and stage of disease. Age at randomisation was provided for all trials however stage of disease was missing for all patients in one trial.

3.2.2 Lung cancer

The second network is formed of trials from three meta-analyses of RCTs in lung cancer performed by the Non-Small-Cell Lung Cancer Collaborative Group. The data was obtained from the Institut Gustave-Roussy (IGR) in Paris. The three meta-analyses considered three different treatments: radiotherapy (RT), radiotherapy plus sequential chemotherapy (Seq CT) and radiotherapy plus concomitant chemotherapy (Con CT, Figure 3.3) using four different designs: RT v Seq CT, RT v Con CT, Seq CT v Con CT and RT v Seq CT v Con CT.

Similarly to the cervical cancer network, in the three original meta-analyses (Non-small Cell Lung Cancer Collaborative Group, 1995; Auperin et al., 2006, 2010) all analyses were intention-to-treat and the individual times to event were used to obtain log-rank hazard ratio estimates of treatment effect for the individual trials, which were then pooled across trials using a stratified-by-trial FTE model. Chi-squared heterogeneity tests were used to assess statistical heterogeneity across trials. Subgroup analyses were conducted using stratified
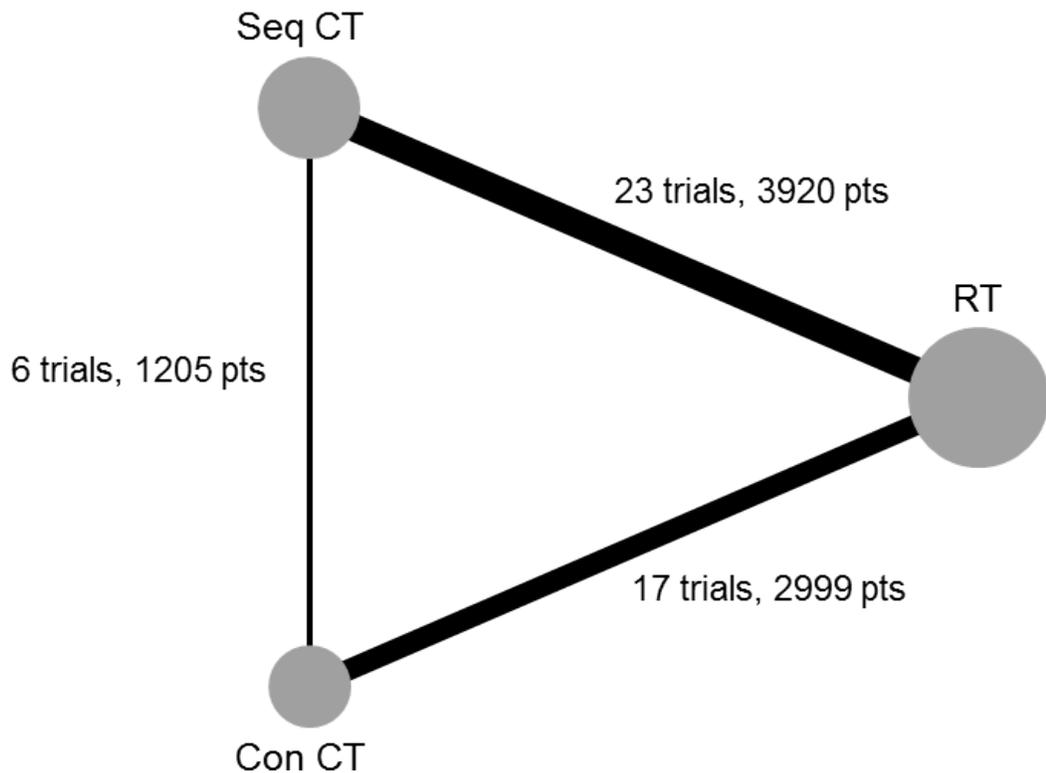
Figure 3.3: Lung cancer network diagram. Node size is proportional to the number of patients randomised to each treatment and line thickness is proportional to the number of studies involved in each direct comparison. RT = radiotherapy, Seq CT = radiotherapy plus sequential chemotherapy, Con CT = radiotherapy plus concomitant chemotherapy. Note that the number of patients for each treatment arm does not add up to the total number of patients included in the network as multi-arm patients are counted twice. There are a total of 44 trials in this network. However, in the figure one multi-arm trial is counted three times as it is included in the number of trials for each pairwise comparison.

log-rank analyses in which hazard ratio estimates were obtained for each predefined subgroup within each trial before being pooled across trials. Chi-squared tests of interaction or trend were used to test for differences in treatment effectiveness between subsets of trials or subgroups of patients. As with the cervical cancer network, this is not an ideal approach for subgroup analyses as the pooling of within and across trial information can result in ecological bias. Subgroup analyses and ecological bias are considered in more detail in Chapter 6.

The MA of RT compared to Seq CT, published in 1995, included 3033 patients from 22 RCTs and identified a 10% reduction in the risk of death with Seq CT (HR=0.90, p=0.006) (Non-small Cell Lung Cancer Collaborative Group, 1995). The chi-squared test for heterogeneity suggested no evidence of heterogeneity between the trials (p=0.56) and a test for interaction between chemotherapy categories was not statistically significant (p=0.59). The current dataset was updated by the IGR to exclude some trials using older forms of chemotherapy and to include some newer trials. Two RCTs were split into two separate comparisons. One trial (Fairlamb et al., 2005) was split into two comparisons based on the timing of chemotherapy (neoadjuvant or adjuvant). A second trial (Mira et al., 1990) was split into two comparisons based on whether patients received prophylactic cranial irradiation or not. These trials will be treated in the same way throughout this thesis. This comparison now includes 23 RCTs and 3920 patients. A forest plot of the log hazard ratios and 95% confidence intervals from each trial is presented in Figure 3.4.

The MA of RT compared to Con CT, published in 2006, included 1764 patients from nine RCTs (Auperin et al., 2006). The MA identified an 11% reduction in the risk of death with Con CT (HR=0.89, 95% CI: 0.81, 0.98). There was evidence of a small amount of heterogeneity (I^2=32%) in the comparison; however the chi-squared heterogeneity test was not significant (p=0.16). Several sensitivity analyses were performed excluding two small trials, two old trials and two trials with incomplete data. Excluding the old trials or the trials with incomplete data reduced the I^2 heterogeneity statistic and in all three cases the treatment effect was reduced and no longer statistically significant. There was no evidence of a difference in the effect of Con CT when trials were grouped by chemotherapy schedule, platinum agent, radiotherapy schedule, radiotherapy dose and whether induction chemotherapy was used or not. There was a smaller effect of Con CT in patients receiving single agent cisplatin or carboplatin (HR=0.93) compared to the two-drug regimen of

platinum plus etoposide (HR=0.72; test for interaction p=0.05). Patient subgroup analyses found no evidence to suggest that the effect of Con CT differed in groups of patients defined by gender, performance status, pathological type or weight loss. There was a suggestion of trend in the effect of Con CT by age (p=0.001). There was also evidence that the effect of Con CT was greater in stage IIIA patients (HR=0.81) compared to stage IIIB (HR=1.01; test for interaction p=0.053). These results were similar in the analysis of event-free survival. The IGR updated this comparison to give a total of 2999 patients from 17 RCTs. The log hazard ratios and 95% confidence intervals for each trial are presented in Figure 3.4.

The Seq CT v Con CT MA included 6 RCTs and 1205 patients (Auperin et al., 2010). The forest plot of log hazard ratios and 95% confidence intervals from each trial, from the original publication (Auperin et al., 2010), is re-created in Figure 3.4. The HR in favour of Con CT was 0.85 (95% CI: 0.74, 0.95) suggesting a 15% reduction in the risk of death with Con CT. No heterogeneity was identified between the six trials. There was no evidence of a difference in the effect of Con CT when trials were grouped by whether the same chemotherapy was used in both arms or not, induction or consolidation chemotherapy and doublet/triplet regimen or single agent. Patient subgroup analyses found no evidence to suggest the effect of Con CT differed in groups of patients defined by age (test for trend p=0.24), gender (test for interaction p=0.84), performance status (test for interaction p=0.88), histology (test for interaction p=0.62) or stage (test for interaction p=0.21). For the current dataset the data remains the same as in the original publication.

All three comparisons included one multi-arm trial which was split into multiple pairwise comparisons for inclusion in the individual meta-analyses. Therefore in total, overall survival data was available for 8079 patients from 44 RCTs.

Covariate data was provided for all trials on age at randomisation, histology, performance status and stage of disease. However performance status was missing for all patients from two trials, histology was missing for all patients from three trials and stage of disease was missing for all patients from eight trials.

Top plot: RT v Seq CT

Study ID	Hazard Ratio (95% CI)	% Weight
Brussels	1.46 (0.85, 2.51)	1.50
Essen	1.24 (0.66, 2.31)	1.13
SLCSG	0.85 (0.68, 1.07)	8.99
WSLCRG-FI	0.91 (0.57, 1.45)	2.06
Perugia	0.74 (0.45, 1.23)	1.71
CALGB 8433	0.69 (0.51, 0.95)	4.44
EORTC 08842	0.86 (0.54, 1.36)	2.07
CEBI 138	0.76 (0.61, 0.94)	9.77
SWOG 8300a	0.90 (0.63, 1.29)	3.48
SWOG 8300b	1.10 (0.77, 1.57)	3.51
MIC1 (CRC TU LU3001)	0.86 (0.71, 1.04)	12.15
RTOG 8808 - ECOG 458	0.73 (0.58, 0.91)	8.57
New Delhi	0.50 (0.41, 0.62)	10.57
Buenos Aires	0.82 (0.52, 1.30)	2.11
FLCSG 2	0.96 (0.75, 1.23)	7.17
Gwent 3	1.00 (0.65, 1.54)	2.34
Tax SI009	0.84 (0.62, 1.14)	4.77
Gwent 1	0.61 (0.35, 1.07)	1.44
SWOG 7635	1.78 (0.97, 3.27)	1.20
NCCTG 822451	0.96 (0.67, 1.40)	3.23
BLT4 (Adjuvant)	1.13 (0.77, 1.65)	3.03
BLT4 (Neo-adjuvant)	1.06 (0.75, 1.48)	3.95
GMMA Ankara 1995	1.00 (0.48, 2.10)	0.81
Overall (I-squared = 55.8%, p = 0.001)	0.83 (0.77, 0.88)	100.00

.306   1   3.27

Middle plot: RT v Con CT

Study ID	Hazard Ratio (95% CI)	% Weight
EORTC 08844	0.78 (0.61, 0.99)	10.13
HOG LUN 86 1	0.89 (0.68, 1.16)	8.36
Aviano	0.96 (0.71, 1.29)	6.65
PMCI 88 C091	0.96 (0.73, 1.26)	7.92
CALGB-ECOG	1.00 (0.78, 1.27)	10.34
NKB-CKVO 94 11	1.17 (0.85, 1.61)	5.93
NPC IIIB 96-01	0.87 (0.73, 1.04)	19.88
Kragujevac 88	0.62 (0.43, 0.89)	4.45
Kragujevac 90	0.63 (0.43, 0.93)	3.97
NCCTG 90 24 51	1.10 (0.68, 1.79)	2.54
ACR LAMP 427	1.09 (0.79, 1.50)	5.81
Uludag	0.75 (0.41, 1.36)	1.65
Brocat Study Group CT/RT 99/97	0.77 (0.56, 1.05)	6.18
GMMA Ankara 1997	0.50 (0.27, 0.93)	1.57
Tax 206	0.86 (0.51, 1.43)	2.29
JCOG 9812	0.71 (0.35, 1.43)	1.22
GMMA Ankara 1995	0.87 (0.42, 1.81)	1.12
Overall (I-squared = 16.2%, p = 0.264)	0.87 (0.81, 0.95)	100.00

.271   1   3.69

Bottom plot: Seq CT v Con CT

Study ID	Hazard Ratio (95% CI)	% Weight
WJLCG	0.78 (0.61, 0.99)	25.59
RTOG 9410	0.80 (0.65, 0.98)	34.61
GLOT-GFPC NPC 95-01	0.80 (0.60, 1.07)	17.11
EORTC 08972	0.98 (0.69, 1.39)	12.10
CALGB 8831	1.12 (0.73, 1.72)	7.93
GMMA Ankara 1995	0.87 (0.41, 1.82)	2.65
Overall (I-squared = 0.0%, p = 0.663)	0.85 (0.74, 0.95)	100.00

.413   1   2.42

Figure 3.4: Lung cancer forest plots. Results come from a Mantel-Haenszel FTE model. Top: RT v Seq CT, Middle: RT v Con CT, Bottom: Seq CT v Con CT. RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy, CI = confidence interval.

3.3 Exploratory analyses of individual trials

In this section I start by comparing the baseline characteristics across all trials and assessing the pattern of follow-up time. Inverse Kaplan-Meier graphs of follow-up time in which patients who die are censored at their time of death were plotted for each trial to check that the pattern of follow-up was similar between the two treatment arms. The median and interquartile ranges for follow-up time in each treatment arm were calculated and 'interesting' patterns were investigated.

Each trial was assessed individually for evidence of PH. The Nelson-Aalen estimate of the log cumulative hazard was plotted against log time for all trials as a visual aid. If the PH assumption is met then the lines for each treatment should be parallel. The Schoenfeld residuals were plotted and a chi-squared test based on these residuals was also conducted. A p-value>0.05 from the chi-squared test suggests that there is no evidence to reject the null hypothesis of proportional hazards.

Visual assessment of publication bias was conducted through the use of contour enhanced funnel plots. In the absence of reporting bias a funnel plot will look approximately symmetrical in shape. Egger's test was also performed. This tests the null hypothesis that the funnel plot is symmetrical and a p-value>0.1 suggests that there is no evidence to reject the null hypothesis resulting in the conclusion that symmetry exists in the funnel plot.

3.3.1 Cervical cancer

Trials in the cervical cancer network ranged in size from 27 patients to 575 patients. Stage of disease varied by trial with one trial including patients with cervical cancer ranging from stage IA to stage IVB disease and other trials being more selective. Median age at randomisation was similar between treatment arms in all trials. Across all trial arms median age at randomisation ranged from 39 years to 76 years. Across all patients age at randomisation ranged from 19 to 91 years. Median follow-up time ranged from 1.7 to 11.4 years (Table A.1).

The plots of log cumulative hazard against log time showed eighteen trials where the treatment lines intersected. However there were only two trials for which the p-value of the chi-squared test based on the Schoenfeld residuals was less than 0.05 (Keys et al., 1999;
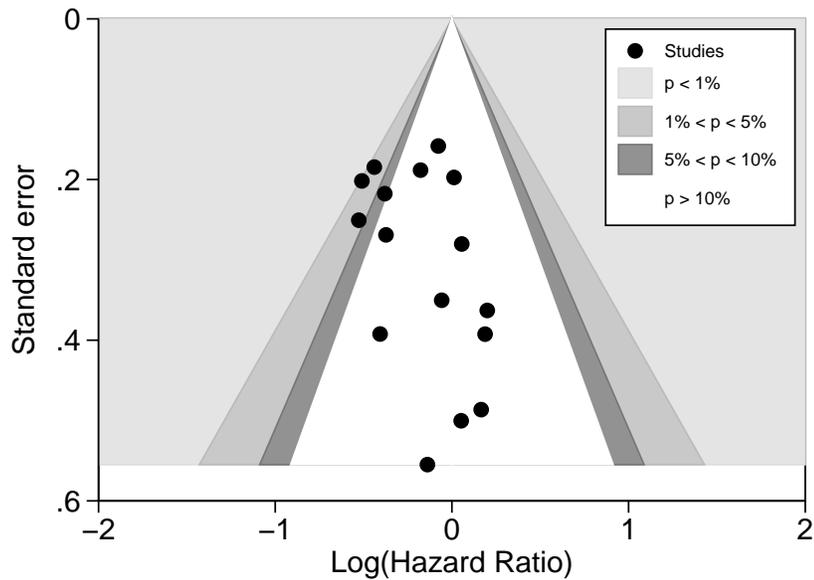
Figure 3.5: Funnel plot for RT versus CTRT trials.

Lorvidhaya et al., 2003). Both trials compared RT with CTRT (Table A.1). I will continue with the assumption of PH for each trial throughout the rest of this chapter and will examine it further in Chapter 4.

There was no suggestion of publication bias in the RT v CTRT and RT v CT+RT comparisons (Figure 3.5, Figure 3.6). The RT v CT+S funnel plot (Figure 3.7) suggested possible publication bias through the absence of some medium-sized trials with inconclusive results (i.e. log hazard ratios of approximately 0). However, this may be an effect of there only being a small number of trials in the comparison. The p-values from Egger's test were 0.262, 0.140 and 0.995 for the RT v CTRT, RT v CT+RT and RT v CT+S comparisons, respectively, suggesting no evidence of funnel plot asymmetry for any of the comparisons.
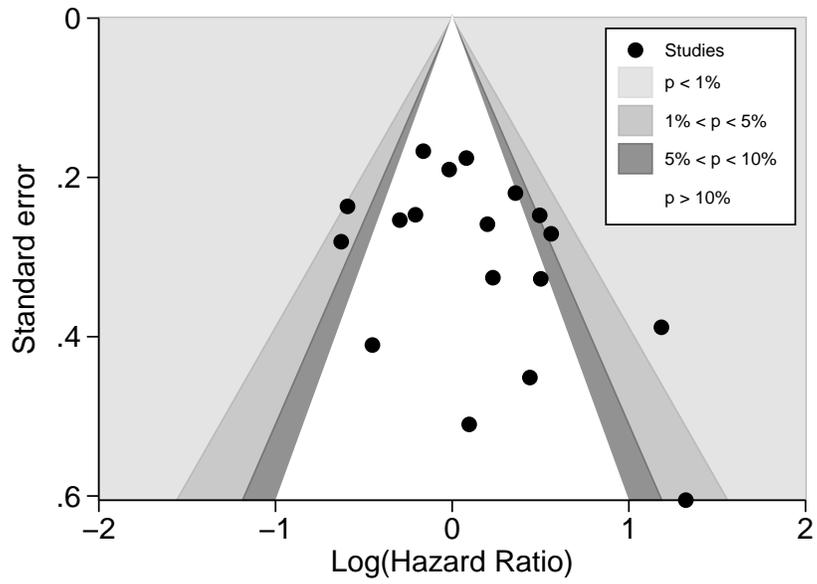
Figure 3.6: Funnel plot for RT versus CT+RT trials.
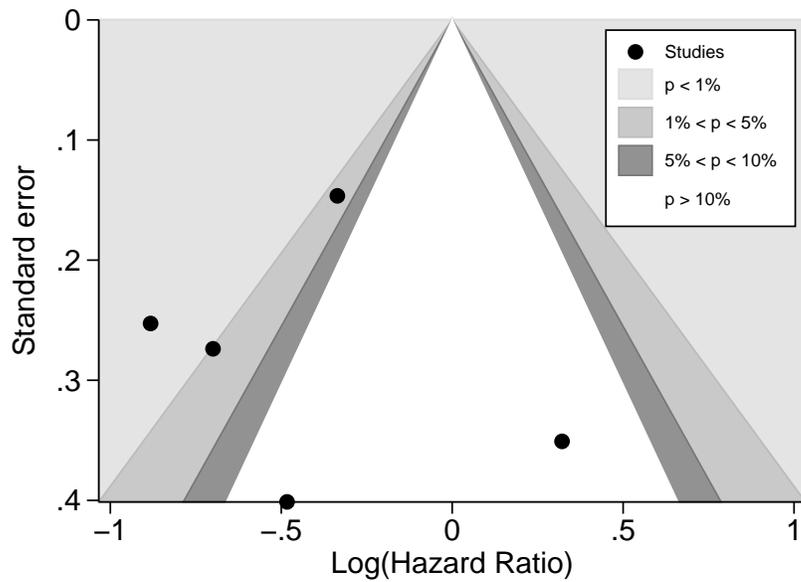
Figure 3.7: Funnel plot for RT versus CT+S trials.

3.3.2 Lung cancer

Trials in the lung cancer network ranged in size from 45 patients to 584 patients. Stage of disease varied by trial with some trials being more restrictive than others. Median age at randomisation was similar between treatment arms in all trials. Across all trial arms median age at randomisation ranged from 52 years to 66 years. Across all patients age at randomisation ranged from 27 to 84 years. Median follow-up time ranged from 0.5 to 1.8 years (Table B.1). For two trials (Blanke et al., 1995; Brodin et al., 1996) follow-up time appeared to be different between the trial arms. However, this is likely due to the small number of patients still alive in either arm at the median survival time.

Plots of the log cumulative hazard against log time showed twenty-two trials where the treatment lines intersected. However, the chi-squared test of the Schoenfeld residuals indicated only two trials with possible evidence of non-PH (Clamon et al., 1994; Sharma et al., 2003). Plots of the Schoenfeld residuals also hint at possible non-PH in two other trials (Alberti et al., 1990; Scagliotti et al., 2006). I will continue with the assumption of PH for each trial throughout the rest of this chapter and examine this further in Chapter 4.

There was no suggestion of publication bias in the RT v Con CT and Seq CT v Con CT comparisons in the funnel plots (Figure 3.8, Figure 3.9) or from Egger's test (p=0.287, p=0.182 respectively). There was some evidence of publication bias in the RT v Seq CT comparison from both an asymmetrical funnel plot (Figure 3.10) and Egger's test (p=0.015). Excluding the one trial with a small standard error and large treatment effect (Sharma et al., 2003) improved the symmetry of the funnel plot. However, Egger's test (p=0.034) suggested that some bias remained in this comparison.
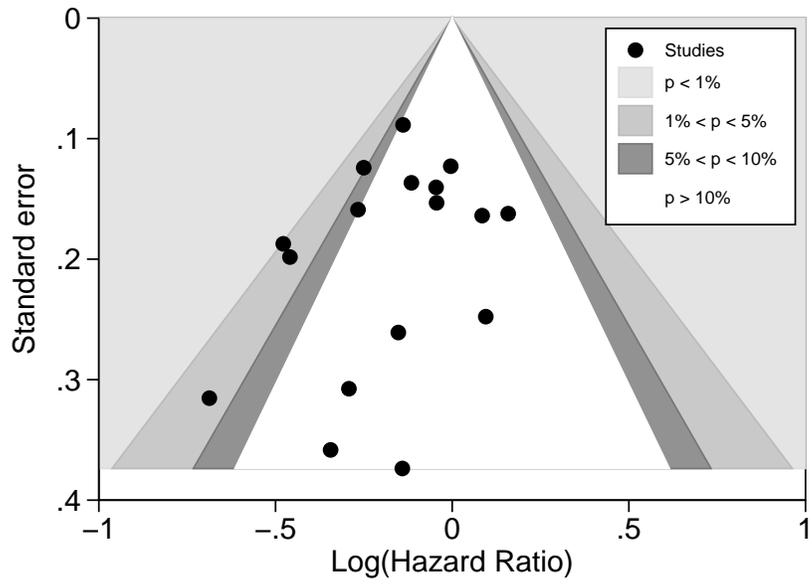
Figure 3.8: Funnel plot for RT versus Con CT trials.
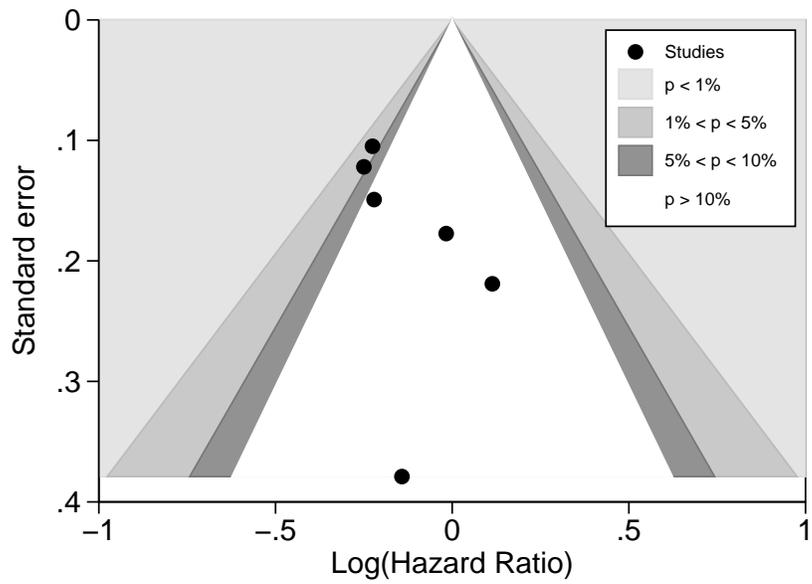
Figure 3.9: Funnel plot for Seq CT versus Con CT trials.
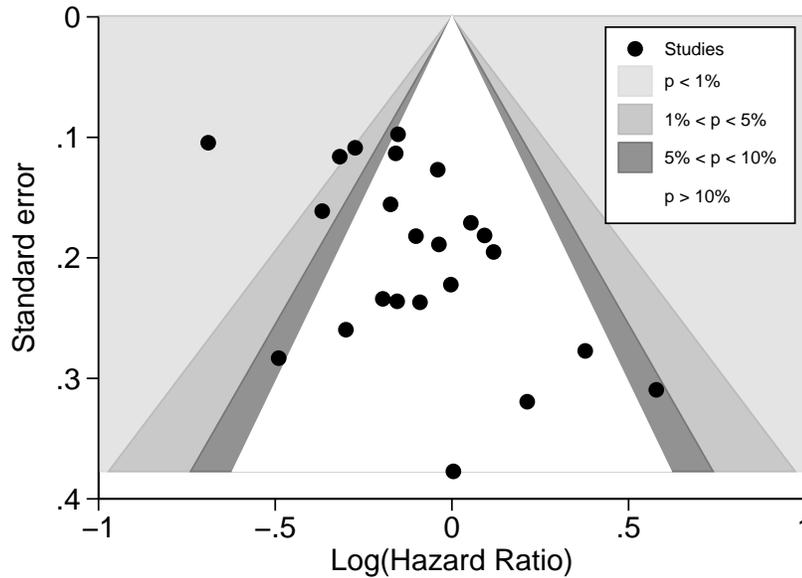
Figure 3.10: Funnel plot for RT versus Seq CT trials.

3.4 Two-step meta-analysis

The log-rank test is one of the most popular methods for comparing survival between two groups of patients. The test assumes that the survival probabilities are the same for all patients randomised throughout the trial and that censoring of a patient is unrelated to their prognosis (Bland and Altman, 2004). Hazard ratios are obtained using the individual participant survival times for each trial. The observed (O) and expected (E) number of overall survival events on the experimental arm were calculated and along with the variance (V) from the log-rank test were used in (3.1) and (3.2) to calculate the log hazard ratio (LogHR) and standard error (SE) for each trial. These estimates were then combined using the FTE Mantel-Haenszel method implemented using the *metan* command (Harris et al., 2008) in Stata (StataCorp, 2015).

$$LogHR = \frac{O - E}{V} \tag{3.1}$$

$$SE(LogHR) = \frac{1}{\sqrt{V}} \tag{3.2}$$

Fixed and random treatment effect meta-analyses were also conducted using the *ipdmetan* command (Fisher, 2015) in Stata (StataCorp, 2015) in which a Cox PH model (Cox, 1972) was fitted to each trial and the results combined in a MA, using either an inverse vari-

ance weighted FTE or DerSimonian and Laird (1986) RTE model, through the one command. To investigate possible sources of heterogeneity in the cervical cancer network the *ipdmetan* command was used to fit Cox PH models stratified by trial and including the covariates stage of disease, histology and tumour grade, one at a time, to the individual trials. Results were then combined using an inverse variance weighted FTE or DerSimonian and Laird (1986) RTE model. For the lung cancer network the covariates histology, stage of disease and sex were considered in the same way. Where patients had missing covariate data multiple imputation was not considered and patients with missing covariate data were excluded from the model. Heterogeneity was assessed in RTE models through the I^2 statistic and the between-study variance τ^2. Values of I^2>50% indicate some evidence of heterogeneity (Higgins et al., 2003).

To check the influence of each trial on the overall treatment effect I used the *metaninf* command (Palmer et al., 2016) in Stata (StataCorp, 2015). This command works by excluding each trial one at a time and re-calculating the overall treatment effect using a Mantel-Haenszel FTE model. The new treatment effects are then shown on a forest plot which shows how the treatment effect changes when each trial is excluded. A log hazard ratio of zero indicates a null effect and a log hazard ratio less than zero indicates a beneficial effect relative to the reference treatment.

3.4.1   Cervical   cancer

Initially the numbers of patients and events in each trial were compared to those presented in the original publications. All discrepancies in numbers between the published analyses and IPD analyses had previously been identified by the MRC CTU at UCL Meta-Analysis Group.

Table 3.1 presents the results from the log-rank and Cox models. There is close agreement between the FTE and RTE point estimates for the RT v CTRT and RT v CT+S comparisons. There is a small increase in the treatment effect for the RT v CT+RT comparison when using the RTE model. However the confidence intervals indicate that the conclusion of no treatment effect remains the same. The treatment effects following the inclusion of the covariates histology, stage of disease and tumour grade remain very similar and the conclusions   unchanged.

There was no evidence of heterogeneity in the RT v CTRT comparison (I^2=0%, $\hat{\tau}^2 = 0$). In the RT v CT+RT comparison I^2=61.8% ($\hat{\tau}^2 = 0.333^2$) and in the RT v CT+S comparison I^2=56.4% ($\hat{\tau}^2 = 0.294^2$) so there was some evidence of heterogeneity in these two comparisons. This was noted in the original publications and is also supported in the RT v CT+RT comparison by the small change in treatment effect from the FTE to the RTE model.

In the RT v CTRT comparison the overall LogHR from the log-rank Mantel-Haenszel FTE model was -0.214 (95% CI: -0.338, -0.091) and when each trial was excluded one at a time the LogHR remained very similar. The overall LogHR from the log-rank Mantel-Haenszel FTE model for the RT v CT+RT comparison was 0.052 (95% CI: -0.068, 0.172) and the biggest departure from this was Sardi 96 (Sardi et al., 1996) which when excluded from the MA resulted in a LogHR of 0.099 (95% CI: -0.026, 0.223).

The funnel plot for the RT v CT+S comparison (Figure 3.7) indicated that there was quite a lot of variation in the treatment effect estimates amongst the small number of trials for this comparison. Therefore when each trial was excluded one at a time there was more change in the treatment effect estimates than for either of the other two comparisons. The overall treatment effect from the log-rank Mantel-Haenszel FTE model for CT+S compared to RT was -0.436 (95% CrI: -0.642, -0.229). The greatest change in this treatment effect was seen with the exclusion of the Benedetti trial, which is the largest trial in this comparison with more than twice the number of patients as any other trial. With the exclusion of the Benedetti trial, the LogHR was reduced to -0.543 (95% CrI: -0.841, -0.245; Table A.2).

Table 3.1: Two-step meta-analysis results for the cervical cancer network. Values are log hazard ratios and 95% confidence intervals. * Patients with missing values for histology, stage and grade are excluded from Cox models as appropriate. FTE = fixed treatment effect, RTE = random treatment effect, RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery, N = number of patients, MH = Mantel-Haenszel, IV = inverse variance weighted, DL = DerSimonian & Laird.

	N	RT v CTRT	N	RT v CT+RT	N	RT v CT+S
Log-rank MH FTE	3104	−0.214 (−0.338, −0.091)	2074	0.052 (−0.068, 0.172)	872	−0.436 (−0.642, −0.229)
Log-rank DL RTE	3104	−0.214 (−0.338, −0.091)	2074	0.104 (−0.101, 0.309)	872	−0.444 (−0.797, −0.092)
Cox IV FTE	3104	−0.215 (−0.339, −0.091)	2074	0.047 (−0.074, 0.167)	872	−0.432 (−0.640, −0.223)
Cox DL RTE	3104	−0.215 (−0.339, −0.091)	2074	0.093 (−0.108, 0.295)	872	−0.442 (−0.789, −0.094)
Cox IV FTE + Histology*	3040	−0.245 (−0.373, −0.118)	2061	0.046 (−0.076, 0.168)	872	−0.427 (−0.635, −0.218)
Cox DL RTE + Histology*	3040	−0.245 (−0.373, −0.118)	2061	0.101 (−0.106, 0.307)	872	−0.428 (−0.793, −0.062)
Cox IV FTE + Stage*	3033	−0.222 (−0.350, −0.095)	1929	0.007 (−0.122, 0.136)	872	−0.445 (−0.653, −0.236)
Cox DL RTE + Stage*	3033	−0.222 (−0.350, −0.095)	1929	0.065 (−0.146, 0.276)	872	−0.448 (−0.791, −0.105)
Cox IV FTE + Grade*	1862	−0.289 (−0.461, −0.118)	1422	−0.068 (−0.213, 0.077)	834	−0.448 (−0.660, −0.235)
Cox DL RTE + Grade*	1862	−0.286 (−0.461, −0.111)	1422	−0.029 (−0.237, 0.179)	834	−0.464 (−0.790, −0.138)

3.4.2 Lung cancer

With the cervical cancer network I started by checking the numbers of patients and events in each trial against those presented in the original publications. For the lung cancer network this was not possible because the dataset has been updated since the three original publications.

Table 3.2 presents the results from the log-rank and Cox models. There is close agreement between the FTE and RTE point estimates for the RT v Con CT and Seq CT v Con CT comparisons and there was no evidence of heterogeneity in either of these comparisons. In the RT v Con CT comparison I^2=16% ($\hat{\tau}^2 = 0.073^2$) and in the Seq CT v Con CT comparison I^2=0% ($\hat{\tau}^2 = 0$). In the RT v Seq CT comparison the differences between the point estimates for the FTE and RTE models suggest that there could be heterogeneity in this comparison. The I^2 and τ^2 statistics also suggest heterogeneity in this comparison (I^2=56%, $\hat{\tau}^2 = 0.185^2$). The treatment effects following the inclusion of the covariates histology, stage of disease and sex remained very similar to the models without covariates therefore the heterogeneity in the RT v Seq CT comparison was not explained by these covariates. Excluding the trial with the largest treatment effect (Sharma et al., 2003) reduced the amount of heterogeneity in the comparison (I^2=13%, $\hat{\tau}^2 = 0.066^2$) and resulted in a similar LogHR from the FTE and the RTE models. This trial is discussed further below.

Table 3.2: Two-step meta-analysis results for the lung cancer network. Values are log hazard ratios and 95% confidence intervals. * Patients with missing values for histology, stage and sex are excluded from Cox models as appropriate. FTE = fixed treatment effect, RTE = random treatment effect, RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy, N = number of patients, MH = Mantel-Haenszel, IV = inverse variance weighted, DL = DerSimonian & Laird.

	N	RT v Seq CT	N	RT v Con CT	N	Seq CT v Con CT
Log-rank MH FTE	3905	-0.189 (-0.256, -0.123)	2999	-0.134 (-0.211, -0.056)	1205	-0.177 (-0.298, -0.056)
Log-rank DL RTE	3905	-0.138 (-0.246, -0.030)	2999	-0.137 (-0.225, -0.049)	1205	-0.177 (-0.298, -0.056)
Cox IV FTE	3905	-0.188 (-0.254, -0.122)	2999	-0.132 (-0.209, -0.056)	1205	-0.176 (-0.297, -0.055)
Cox DL RTE	3905	-0.137 (-0.242, -0.032)	2999	-0.134 (-0.218, -0.051)	1205	-0.176 (-0.297, -0.055)
Cox IV FTE + Histology*	3880	-0.196 (-0.263, -0.130)	2105	-0.092 (-0.184, 0.001)	1200	-0.160 (-0.282, -0.038)
Cox DL RTE + Histology*	3880	-0.151 (-0.254, -0.048)	2105	-0.092 (-0.184, 0.001)	1200	-0.160 (-0.282, -0.038)
Cox IV FTE + Stage*	2474	-0.158 (-0.242, -0.075)	2575	-0.106 (-0.189, -0.024)	1197	-0.151 (-0.273, -0.029)
Cox DL RTE + Stage*	2474	-0.146 (-0.241, -0.051)	2575	-0.107 (-0.205, -0.009)	1197	-0.151 (-0.273, -0.029)
Cox IV FTE + Sex*	3896	-0.187 (-0.254, -0.121)	2997	-0.143 (-0.219, -0.066)	1201	-0.171 (-0.292, -0.050)
Cox DL RTE + Sex*	3896	-0.139 (-0.242, -0.036)	2997	-0.148 (-0.243, -0.053)	1201	-0.171 (-0.292, -0.050)

In the RT v Con CT and Seq CT v Con CT comparisons excluding each trial one at a time had very little impact on the overall treatment effect. In the comparison of RT v Seq CT excluding most trials had little impact on the overall treatment effect however the exclusion of one trial (Sharma et al., 2003) changed the overall LogHR from the log-rank Mantel-Haenszel FTE model from -0.189 to -0.130 (Table B.2). This was the same trial, which when excluded from the network, reduced the amount of heterogeneity in the RT v Seq CT comparison. Further investigation into this trial showed that some of the trial baseline characteristics differed to the rest of the network population. In particular the mean age in this trial was eight years younger than the mean across the rest of the network. There also appeared to be a higher proportion of patients with poor performance status than compared to the rest of the network. Stage of disease was unknown for all participants in this trial. However, the trial publication identified the eligibility criteria for the trial to be stage IIIA-IIIB disease. The radiotherapy and chemotherapy regimens used in this trial were comparable to the other trials in the RT v Seq CT comparison. When the trial characteristics from the IPD were compared to the published paper some discrepancies were noticed. Contact was made with the Institut Gustave-Roussy to query these discrepancies. The Institut Gustave-Roussy was aware of these discrepancies but unable to resolve them. Based on these discrepancies I felt that the IPD from this trial was unreliable and therefore I decided to exclude this trial from the lung cancer network. Throughout the rest of this thesis the lung cancer network will contain 43 RCTs and 7576 patients.

3.5   Summary

In this chapter I have conducted a thorough review of data from two networks of clinical trials in oncology. By examining the cervical cancer network for evidence of non-PH, heterogeneity and sources of bias, I found that there is some suggestion of non-PH in two RT v CTRT trials, heterogeneity in one of the direct comparisons (RT v CT+RT) and possible publication bias in another direct comparison (RT v CT+S). For now I will continue to assume PH and will explore this further in Chapter 4. Heterogeneity will also be assessed further in Chapter 4. In five RT v CTRT trials additional treatment was received by patients and therefore these trials were excluded from the dataset and will remain excluded throughout the rest of this thesis.

One trial (Sharma et al., 2003) in the lung cancer network was identified as an influential trial and later excluded. The process of excluding this trial was initially driven by finding a large amount of heterogeneity in the RT v Seq CT comparison. Exploratory analyses comparing trial characteristics across the network, tabulated in Table B.1, did not identify the trial as a potential cause for concern. However, the trial was identified as potentially causing heterogeneity, publication bias and influencing the treatment effect estimate. Further investigation into this trial showed that the data did not reflect the trial publication or the network population. In particular, I identified a difference in the mean age and distribution of performance status for this trial compared to the rest of the network. It is possible that as an alternative to excluding this trial, age and performance status could have been included in the NMA model as covariates. Excluding a trial based purely on the fact that it causes heterogeneity is not a valid process. However, in this case the IPD differed to the published trial results. Therefore, I felt that the data for the trial was unreliable and would result in unreliable treatment effect estimates. As a result I made the decision to exclude the trial from the lung cancer network throughout the rest of this thesis.

In the lung cancer network I identified some evidence of heterogeneity in the RT v Seq CT comparison however this was no longer present following the removal of the influential trial (Sharma et al., 2003). I also identified some trials which suggested that the PH assumption may not be appropriate. For now I will continue to assume PH and will explore this further in Chapter 4.

Chapter 4 describes the Royston-Parmar approach to one-step IPD NMA, how it can be implemented in a Bayesian setting using WinBUGS and extensions to test for and accommodate departures from the PH assumption, identify inconsistency and assess heterogeneity.

4 One-Step IPD Network Meta-Analysis

4.1 Introduction

Chapter 3 introduced the cervical and lung cancer networks, and conducted a thorough review of both datasets including checking the PH assumption in each trial, looking for evidence of publication bias, heterogeneity and performing two-step MA.

Traditionally time-to-event outcomes have been analysed using semi-parametric Cox PH models (Cox, 1972). However, in oncology with longer follow-up of trials, and time-dependent effects of targeted treatments, there is increasing evidence of non-PH. Therefore the PH assumption may no longer be appropriate (Royston and Parmar, 2016; Trinquart et al., 2016). NMA conducted in the Bayesian setting has been increasing in popularity in recent years (Sobieraj et al., 2013). The Bayesian framework naturally handles random effects, avoiding awkward numerical integration, and offers many other benefits. These include easy assessment of network consistency, a natural ranking method and the ability to adjust for correlations which arise from the inclusion of multi-arm trials (Ades et al., 2006; Dominici et al., 1999; Lu and Ades, 2006). Additionally covariates with missing data can be readily handled in a Bayesian framework (Section 6.4). Bayesian NMA models are commonly fitted in WinBUGS. However fitting the Cox PH model in the Bayesian setting is computationally intensive, as each individual's data has to be repeated for each risk set they belong to, making it unsuitable for even moderately sized datasets, such as the cervical cancer network introduced in Section 3.2. Therefore alternative methods for time-to-event data are needed.

Royston-Parmar models are a parametric alternative to the Cox model which use a restricted cubic spline (RCS) function of log time to model the (baseline) log cumulative hazard rate (Royston and Parmar, 2002). The complexity and flexibility of the model is determined by the RCS and model parameters are estimated using maximum likelihood (Lambert and Royston, 2009). Fitted in WinBUGS (Lunn et al., 2000) they provide a flexible parametric Bayesian alternative to the Cox PH model for analysing time-to-event data, which extends naturally to allow for non-PH and has the flexibility to allow the inclusion of covariates and associated random effects.

In this chapter I will describe and apply the Royston-Parmar model for conducting a one-

step IPD NMA of time-to-event data to the networks of clinical trials in cervical and lung cancer introduced in Chapter 3. To the best of my knowledge, this is the first time one-step IPD NMA has been conducted using Royston-Parmar models. I will show how including a treatment-ln(time) interaction can be used to conduct a global test for PH, illustrate testing for consistency of direct and indirect evidence, and assessing within design heterogeneity. NMA combines randomised evidence with non-randomised evidence, and the latter relies on the assumption of no (unmeasured) confounding. When presenting the results, I therefore propose, and illustrate, presenting the direct and indirect treatment estimates alongside the combined estimate.

This chapter is structured as follows. In Section 4.2 I review the Royston-Parmar model, an extension to test for non-PH and how to implement the model in WinBUGS. In Section 4.3 the Royston-Parmar model is applied in the MA setting to both the cervical and lung cancer networks. In Section 4.4 the Royston-Parmar model is extended to the NMA setting and the assessment of PH, inconsistency and heterogeneity are explored. In Section 4.5 I consider the prior distributions required for fitting the Royston-Parmar model using the Bayesian framework. Section 4.6 and Section 4.7 present the results of the Royston-Parmar NMA model applied to the cervical and lung cancer networks, respectively. This chapter finishes with a discussion in Section 4.8.

Throughout this chapter and the rest of this thesis I will use the subscript i to denote patient and the subscript j to denote trial.

Work from this chapter was first presented at the International Clinical Trials Methodology Conference in November 2015. A journal paper based on this chapter has been submitted for publication in Research Synthesis Methods.

4.2 Royston-Parmar model for the log cumulative hazard rate

In this section I describe the Royston-Parmar model (Royston and Parmar, 2002) for the log cumulative hazard, an extension of the Royston-Parmar model to test for non-PH and the process for implementing the approach in the MA setting to estimate treatment effects (Lambert and Royston, 2009). In this section I describe the work of Royston & Parmar (2002) and Lambert & Royston (Lambert and Royston, 2009) in applying the Royston-

Parmar model in the MA setting which is necessary in preparation for Section 4.4 where I show how I extended the Royston-Parmar model to the NMA setting.

To implement the Royston-Parmar model in the NMA setting, a RCS is used to model the log baseline cumulative hazard rate for each trial. A RCS has a number of interior knots as well as boundary knots at the minimum and maximum of the uncensored survival times. The fitted spline is continuous with continuous 1st and 2nd derivatives (Lambert and Royston, 2009).

The spline function for the log cumulative hazard from trial j at time t with p interior knots can be written as:

$$s_j\big(\ln(t)\big) = \gamma_1 + \gamma_2 u_0\big(\ln(t)\big) + \gamma_3 u_1\big(\ln(t)\big) + \cdots + \gamma_{p+2} u_p\big(\ln(t)\big), \qquad (4.1)$$

where $\ln(t)$ is the natural logarithm of event time for patient i in trial j, $u_0\big(\ln(t)\big), \ldots, u_p\big(\ln(t)\big)$ are the orthogonalised basis functions and γ's their coefficients. Basis functions are defined in Subsection 4.2.2.

The RCS for the log cumulative hazard can be incorporated into a PH flexible parametric model with x_{ij} the treatment indicator for patient i from trial j and β the coefficient,

$$\ln\{H(t|x_i j)\} = \eta_{ij} = s_j\big(\ln(t)\big) + \beta x_{ij}. \qquad (4.2)$$

Covariates can also be included in (4.2) as adjustment factors if necessary. To fit this flexible parametric model (4.2) the log likelihood of the observed data must be calculated. In order to derive the log likelihood the derivative of η_{ij} is required,

$$d\eta_{ij} = \gamma_2 du_0\big(\ln(t)\big) + \gamma_3 du_1\big(\ln(t)\big) + \cdots + \gamma_{p+2} du_p\big(\ln(t)\big) \qquad (4.3)$$

where du_p is the derivative with respect to $\ln(t)$ of u_p.

The likelihood, l_{ij}, for patient i in trial j is then:

$$\log(l_{ij}) = \begin{cases} \log(d\eta_{ij}) + \eta_{ij} - \exp(\eta_{ij}) & \text{for an observed event,} \\ -\exp(\eta_{ij}) & \text{for a censored observation.} \end{cases}$$

WinBUGS can be used for Bayesian inference with this likelihood. WinBUGS does not have an appropriate inbuilt distribution for the Royston-Parmar model. Therefore, the "zeros trick" is required to enable a general likelihood to be specified (Royston and Lambert, 2011). The "zeros trick" works because an observed value of zero has a Poisson

likelihood of $\exp(-\lambda)$. If λ is set equal to the negative log likelihood for patient i then the correct likelihood contribution will be obtained (Royston and Lambert, 2011). As a Bayesian approach, WinBUGS has the added advantage of the flexibility to extend models (e.g. to include multiple random effects and covariates) avoiding the complexities of the numerical integration needed for maximum likelihood estimation. Thus, the fixed effect of treatment in (4.2) can be readily replaced by a random effect if desired.

4.2.1 Testing for non-proportional hazards

Lambert & Royston (2009) show that non-PH can be assessed by including a treatment-ln(time) interaction in (4.2):

$$\ln\{H(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta x_{ij} + \alpha x_{ij}\left(\ln\left(t\right)\right) \tag{4.4}$$

where $x_{ij}\left(\ln\left(t\right)\right)$ is the treatment-ln(time) interaction term for patient i from trial j and α the coefficient. The treatment-ln(time) interaction should in fact be an interaction between treatment and the spline function, $s_j(\ln(t))$. However, the notation used here is a good approximation as the basis function $v_0(\ln(t))$ is defined as $\ln(t)$, see Subsection 4.2.2. In (4.3) the derivative of (4.2) is calculated with respect to $\ln(t)$ therefore (4.3) must be updated appropriately when treatment-ln(time) interactions are included. The null hypothesis states that there is no evidence of non-PH in the MA (i.e. α=0). A further extension is to allow α to be random across (groups of) trials; see Subsection 4.4.1.

Before conducting MA or NMA each trial should be assessed individually for evidence of non-PH. An alternative method of assessing non-PH is to consider the χ^2 tests of the Schoenfeld residuals for each trial. As each trial is independent of each other, in each MA, the values of the χ^2 statistics can be added together to provide an overall test statistic with degrees of freedom equal to the number of trials in the MA.

The Schoenfeld residual test, applied to each trial in turn, looks for any evidence of a different trend in the Schoenfeld residuals between the treatment groups (Grambsch and Therneau, 1994). It highlights any trials which show a marked departure from PH. Such departures may be due to quirks of the design or follow-up. By contrast, testing the null hypothesis that $\alpha = 0$ in (4.4) provides a more powerful test of the specific hypothesis

that the log cumulative hazard has a different linear trend in $\log(t)$ in the different treatment groups. If, across the MA, $\alpha = 0$, is rejected then summarising treatment effects by a single hazard ratio is inappropriate. It is useful to apply the Schoenfeld residual test to each trial, in turn, to identify any trials which could potentially be the cause of non-PH in the MA. However, I recommend that the decision on whether the MA exhibits evidence of non-PH is based on the more powerful test of $\alpha = 0$ in (4.4). As far as I am aware, at the moment this approach is not widely used in practice.

4.2.2 Implementation

For a patient i with survival (or censoring time) $\ln(t)$ and a spline model with p interior knots where the location of the knots are $k_0, k_1, \ldots, k_p, k_{p+1} (k_0 < k_1 < \ldots < k_p < k_{p+1})$ and k_0 and k_{p+1} are the boundary knots and where $(x)_+$ equals x if $x > 0$ and 0 otherwise, the basis functions $v_0\big(ln(t)\big), v_1\big(ln(t)\big), v_2\big(ln(t)\big), \ldots, v_{p-1}\big(ln(t)\big), v_p\big(ln(t)\big)$ are calculated as follows:

$$v_0\big(\ln(t)\big) = \ln(t)$$

$$v_1\big(\ln(t)\big) = \big(\ln(t) - k_1\big)_+^3 - \phi_1\big(\ln(t) - k_0\big)_+^3 - (1 - \phi_1)\big(\ln(t) - k_{p+1}\big)_+^3,$$

$$\text{where } \phi_1 = \frac{(k_{p+1} - k_1)}{(k_{p+1} - k_0)}$$

$$v_2\big(\ln(t)\big) = \big(\ln(t) - k_2\big)_+^3 - \phi_2\big(\ln(t) - k_0\big)_+^3 - (1 - \phi_2)\big(\ln(t) - k_{p+1}\big)_+^3,$$

$$\text{where } \phi_2 = \frac{(k_{p+1} - k_2)}{(k_{p+1} - k_0)}$$

$$\vdots$$

$$v_{p-1}\big(\ln(t)\big) = \big(\ln(t) - k_{p-1}\big)_+^3 - \phi_{p-1}\big(\ln(t) - k_0\big)_+^3 - (1 - \phi_{p-1})\big(\ln(t) - k_{p+1}\big)_+^3,$$

$$\text{where } \phi_{p-1} = \frac{(k_{p+1} - k_{p-1})}{(k_{p+1} - k_0)}$$

$$v_p\big(\ln(t)\big) = \big(\ln(t) - k_p\big)_+^3 - \phi_p\big(\ln(t) - k_0\big)_+^3 - (1 - \phi_p)\big(\ln(t) - k_{p+1}\big)_+^3,$$

$$\text{where } \phi_p = \frac{(k_{p+1} - k_p)}{(k_{p+1} - k_0)}$$

Untransformed spline basis functions can be highly correlated. Therefore, Gram-Schmidt orthogonalisation is used to linearly transform the basis functions $v_0\big(\ln(t)\big), \ldots, v_p\big(\ln(t)\big)$ into $u_0\big(\ln(t)\big), \ldots, u_p\big(\ln(t)\big)$ to improve numerical stability when fitting the model. This process results in uncorrelated basis functions with mean zero and standard deviation one

(Royston and Lambert, 2011). The uncorrelated basis functions $u_0\big(\ln(t)\big), \ldots, u_p\big(\ln(t)\big)$ can then be used in equations (4.1), (4.2), (4.3) and (4.5). Let $u_0 = u_0\big(\ln(t)\big)$, $u_1 = u_1\big(\ln(t)\big)$, $un_1 = un_1\big(\ln(t)\big)$ etc., then

$$u_0 = \frac{v_0 - mean(v_0)}{sd(v_0)}$$

$$un_1 = v_1 - \frac{<v_1, u_0>}{<u_0, u_0>}u_0, \text{ and normalising } u_1 = \frac{un_1 - mean(un_1)}{sd(un_1)}$$

$$un_2 = v_2 - \frac{<v_2, u_0>}{<u_0, u_0>}u_0 - \frac{<v_2, u_1>}{<u_1, u_1>}u_1, \text{ and normalising } u_2 = \frac{un_2 - mean(un_2)}{sd(un_2)}$$

$$\vdots$$

$$un_p = v_p - \frac{<v_p, u_0>}{<u_0, u_0>}u_0 - \frac{<v_p, u_1>}{<u_1, u_1>}u_1 - \cdots - \frac{<v_p, u_{p-1}>}{<u_{p-1}, u_{p-1}>}u_{p-1}, \text{ and normalising}$$

$$u_p = \frac{un_p - mean(un_p)}{sd(un_p)}$$

where $<u\big(\ln(t)\big), v\big(\ln(t)\big)> = \sum_i u\big(\ln(t)\big)v\big(\ln(t)\big)$ i.e. the inner product of the vectors $\mathbf{u}$ and $\mathbf{v}$.

4.2.3 Estimation

Having calculated the orthogonalised basis functions $u_0\big(\ln(t)\big), \ldots, u_p\big(\ln(t)\big)$ and their derivatives $du_0\big(\ln(t)\big), \ldots, du_p\big(\ln(t)\big)$ these are then passed to WinBUGS to fit the one-step MA (4.2) or NMA model (4.5) in which the logarithm of the baseline cumulative hazard function is modelled as a natural cubic spline function of log time (Royston and Parmar, 2002).

The default knot locations for RCS are based on percentiles of the uncensored survival times with additional boundary knots placed at the minimum and maximum values of the uncensored survival times. Royston & Lambert (2011) do not recommend models with more than three knots as the resulting curves can be unstable; however, they do acknowledge that in larger datasets a larger number of knots may be required. It has been shown recently that parameter estimates are generally robust to knot locations (Rutherford et al., 2015); however it is also possible to choose knot locations. With the cervical cancer network I chose the knot locations because I wanted to ensure that the log cumulative hazard resulting from the Royston-Parmar model (4.2), fitted in WinBUGS, was as similar to the

non-parametric Nelson-Aalen estimate of the log cumulative hazard as possible for each trial. Starting with the default knot locations I plotted the log cumulative hazard resulting from the Royston-Parmar model with one, two and three knots against log time alongside the Nelson-Aalen estimate of the log cumulative hazard and its 95% confidence intervals. For each trial I chose the model with the number of knots that showed the best agreement between the Royston-Parmar model and the Nelson-Aalen estimate. Knot locations were then tweaked where necessary to improve the agreement between the Royston-Parmar model and the Nelson-Aalen estimate and to ensure the log cumulative hazard from the Royston-Parmar model fell within the 95% confidence intervals of the Nelson-Aalen estimate. A table of knot locations for the cervical cancer network can be found in Table A.3. For the lung cancer network I fitted all trials with two interior knots placed at the 33rd and 67th percentiles of the uncensored survival times.

Models were run in WinBUGS version 1.4.3 (Lunn et al., 2000). The Stata suite of commands *winbugs* (Thompson et al., 2006) was used to control all aspects of model fitting in WinBUGS through Stata version 14 (StataCorp, 2015). Models were run with 20,000 burn-in and then 20,000 iterations and with two sets of initial values. Convergence was checked by examining the trace and histograms of the posterior distribution. Models were compared using the Deviance Information Criteria (DIC) statistic (Lunn et al., 2013; Spiegelhalter et al., 2002).

4.3   Results of pairwise MA using the Royston-Parmar method

In this section, in both networks, each treatment comparison is treated as an independent pairwise MA and all three-arm trials were split into two-arm comparisons. Initially each trial was fitted with a FTE Royston-Parmar model to assess the range of treatment effects for overall survival across all trials. Then for each comparison a one-step MA was conducted using a FTE Royston-Parmar MA model (4.2). Each treatment comparison was assessed for evidence of statistical heterogeneity using Cochran's Q statistic (Higgins et al., 2003).

4.3.1 Cervical cancer

The log hazard ratio for CTRT compared to RT ranged from -0.535 in the Roberts trial (Roberts et al., 2000) to 0.230 in the Onishi trial (Onishi et al., 1999). The log hazard ratio for CT+RT compared to RT ranged from -0.655 in the Sardi 97 trial (Sardi et al., 1997) to 1.776 in the LGOG trial (unpublished). The log hazard ratio for CT+S compared to RT ranged from -0.910 in the Sardi 96 trial (Sardi et al., 1996) to 0.338 in the Chang trial (Chang et al., 2000). Table A.4 contains the log hazard ratios for the treatment effect in each trial.

Figure 4.1 shows forest plots of log hazard ratios and 95% credible intervals (CrI) from each trial and overall for each pairwise comparison. The resulting fixed treatment effects are presented in Table 4.1. Results of the pairwise MA suggest CTRT improves overall survival by 19% compared to RT (LogHR=-0.215, 95% CrI: -0.336, -0.086), CT+S improves overall survival by 36% compared to RT (LogHR=-0.447, 95% CrI: -0.654, -0.243) and CT+S also improves overall survival by 36% compared to CT+RT (LogHR=-0.444, 95% CrI: -0.830, -0.061). When the RT v CT+RT comparison is split into two comparisons based on chemotherapy cycle length, short cycles of CT+RT improve overall survival by 17% (LogHR=-0.191, 95% CrI: -0.375, -0.007) and long cycles reduce overall survival by 25% (LogHR=0.227, 95% CrI: 0.073, 0.385).

There was no evidence of heterogeneity within the RT v CTRT (p=0.625, Table 4.1), RT v CT+S (p=0.065) and CT+RT v CT+S (p=0.939) comparisons while there was some evidence of statistical heterogeneity in the RT v CT+RT comparison (p<0.001, also noted in the original publication (Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration, 2003)). The RT v CT+RT comparison was split into two comparisons based on whether the chemotherapy cycle in a trial was greater than 14 days. Throughout the rest of this thesis trials with chemotherapy cycles less than or equal to 14 days will be referred to as 'short cycles' and trials with chemotherapy cycles greater than 14 days will be referred to as 'long cycles'. When the RT v CT+RT comparison was split into subgroups based on length of chemotherapy cycles no evidence of heterogeneity in the trials was found with long cycles (p=0.263). However there was evidence of heterogeneity in the trials with short cycles (p=0.002). Combinations of the same chemotherapy drugs were used in trials with long and short cycles suggesting that chemotherapy cy-

Top left panel

Trial ID	LogHR (95% CrI)
Keys	-0.446 (-0.826, -0.083)
Hongwei A	0.061 (-0.981, 1.091)
Hongwei B	-0.131 (-1.295, 1.001)
Pearcey	-0.179 (-0.559, 0.194)
Pras	-0.009 (-0.714, 0.684)
Leborgne CTRT	-0.079 (-0.382, 0.233)
Thomas A	-0.373 (-0.910, 0.149)
Thomas B	0.052 (-0.498, 0.616)
Lorvidhaya A	-0.515 (-0.926, -0.110)
Lorvidhaya B	0.009 (-0.372, 0.398)
Roberts	-0.535 (-1.061, -0.026)
Onishi	0.230 (-0.486, 0.956)
Lanciano	-0.396 (-1.102, 0.348)
Cikaric	-0.388 (-0.824, 0.037)
Garipagaglu	0.197 (-0.786, 1.165)
Lal	0.127 (-0.651, 0.915)
Overall (I-squared = 0.0%, p = 0.625)	-0.215 (-0.336, -0.086)

-1.5   0   1.5
Favours CTRT   Favours RT

Top right panel

Trial ID	LogHR (95% CrI)
>14 days	
Cardenas 93	0.521 (-0.373, 1.479)
Chauvergne	-0.016 (-0.386, 0.353)
Kumar	0.367 (-0.066, 0.810)
Souhami	0.538 (0.019, 1.045)
Sundfor	-0.195 (-0.683, 0.291)
Tattershall 92	0.214 (-0.424, 0.847)
Tattershall 95	0.500 (0.003, 0.996)
LGOG	1.776 (0.271, 3.814)
Chiara	0.519 (-0.128, 1.179)
Herod	0.081 (-0.262, 0.429)
Cardenas 91	0.099 (-0.914, 1.113)
Subtotal (I-squared = 18.9%, p = 0.263)	0.227 (0.073, 0.385)
<=14 days	
Leborgne NeoCT+RT	0.194 (-0.310, 0.692)
Sardi 96	-0.578 (-1.043, -0.112)
Sardi 97	-0.655 (-1.236, -0.080)
Sardi 98	-0.298 (-0.801, 0.199)
Symonds	-0.168 (-0.493, 0.156)
MRC CeCa	1.280 (0.510, 2.128)
PMB	-0.462 (-1.328, 0.370)
Subtotal (I-squared = 71.0%, p = 0.002)	-0.191 (-0.375, -0.007)

-1.5   0   1.5
Favours CT+RT   Favours RT

Bottom left panel

Trial ID	LogHR (95% CrI)
Sardi 96	-0.910 (-1.428, -0.418)
Sardi 98	-0.673 (-1.232, -0.131)
Benedetti	-0.340 (-0.628, -0.054)
Kigawa	-0.534 (-1.376, 0.275)
Chang	0.338 (-0.376, 1.091)
Overall (I-squared = 54.8%, p = 0.065)	-0.447 (-0.654, -0.243)

-1.5   0   1.5
Favours CT+S   Favours RT

Bottom right panel

Trial ID	LogHR (95% CrI)
Sardi 96	-0.427 (-0.961, 0.096)
Sardi 98	-0.457 (-1.013, 0.091)
Overall (I-squared = 0.0%, p = 0.939)	-0.444 (-0.830, -0.061)

-1.5   0   1.5
Favours CT+S   Favours CT+RT

Figure 4.1: Cervical cancer meta-analysis plots. Trial results come from a FTE Royston-Parmar model. Overall results come from a one-step IPD FTE Royston-Parmar MA model. Top left: RT v CTRT, Top right: RT v CT+RT, Bottom left: RT v CT+S, Bottom right: CT+RT v CT+S. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery, LogHR = log hazard ratio, CrI = credible interval.

cle length was not confounded with type of chemotherapy in the RT v CT+RT comparison. Chemotherapy cycle length was not considered for the other comparisons as these did not exhibit evidence of heterogeneity and a difference in treatment effect based on chemotherapy cycle length was not identified in the original MAs. There was no evidence globally of non-PH in any of the treatment comparisons (Table 4.1, column 4), however the Schoenfeld residuals indicate that there may be some trials in the RT v CTRT comparison which are at risk of non-PH (Table 4.1, column 5).

Table 4.1: Cervical cancer meta-analysis results using Royston-Parmar models. FTE = fixed treatment effect. * Values are log hazard ratios and 95% credible intervals.

Comparison	FTE*	Cochran's Q	Global non-PH test	Schoenfeld residuals
RT v CTRT	−0.215 (−0.336, −0.086)	12.71, 15df, p=0.625	χ^2=0.161, 1df, p=0.688	χ^2=25.64, 16df, p=0.059
RT v CT+RT short cycles	−0.191 (−0.375, −0.007)	20.69, 6df, p=0.002	χ^2=2.522, 1df, p=0.112	χ^2=10.34, 7df, p=0.170
RT v CT+RT long cycles	0.227 (0.073, 0.385)	12.34, 10df, p=0.263	χ^2=0.006, 1df, p=0.944	χ^2=7.65, 11df, p=0.744
RT v CT+S	−0.447 (−0.654, −0.243)	8.85, 4df, p=0.065	χ^2=0.118, 1df, p=0.731	χ^2=8.65, 5df, p=0.124
CT+RT v CT+S	−0.444 (−0.830, −0.061)	0.01, 1df, p=0.939	χ^2=0.164, 1df, p=0.686	χ^2=0.49, 2df, p=0.783

4.3.2 Lung cancer

The log hazard ratio for Seq CT compared to RT alone ranged from −0.606 in the Gwent 1 trial (Anderson et al., 1981) to 0.607 in the SWOG 7635 trial (White et al., 1982). The log hazard ratio for RT v Con CT ranged from −0.640 in the GMMA Ankara 1997 trial (Cüneyt Ulutin and Pak, 2003) to 0.141 in the NKB-CKVO trial (Groen, 2004). The log hazard ratio for Seq CT v Con CT ranged from −0.250 in the WJLCG trial (Furuse et al., 1999) to 0.117 in the CALGB 8831 trial (Clamon et al., 1994). The three-arm trial was split into three pairwise comparisons and log hazard ratios for the treatment effect in each trial are displayed in Table B.3.

Figure 4.2 shows forest plots displaying the log hazard ratios and 95% credible intervals
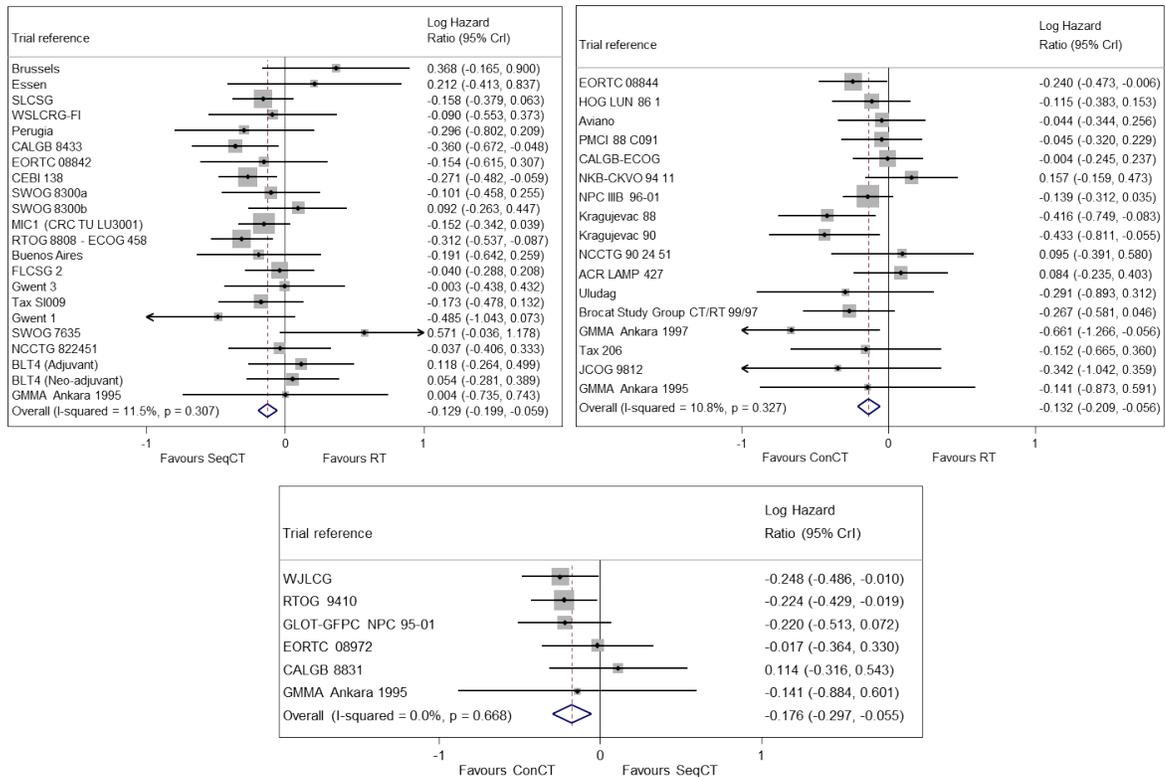
Figure 4.2: Lung cancer meta-analysis plots. Trial results come from a FTE Royston-Parmar model. Overall results come from a one-step IPD FTE Royston-Parmar MA model. Top left: RT v Seq CT, Top right: RT v Con CT, Bottom: Seq CT v Con CT. RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy, CrI = credible interval.

from each trial and overall for each pairwise comparison. The overall treatment effects are presented in Table 4.2. From the pairwise comparisons Seq CT improves overall survival by 12% compared to RT (LogHR=-0.129, 95% CrI: -0.199, -0.059). Con CT also improves overall survival by 12% compared to RT (LogHR=-0.132, 95% CrI: -0.209, -0.056). Meanwhile Con CT improves overall survival by 16% compared to Seq CT (LogHR=-0.176, 95% CrI: -0.297, -0.055).

Cochran's Q statistic provided no evidence of statistical heterogeneity in the RT v Seq CT (p=0.984), RT v Con CT (p=0.907) or Seq CT v Con CT (p=0.147) comparisons (Table 4.2, column 3). The global test for non-PH and the Schoenfeld residuals suggested no evidence of non-PH in any of the comparisons (Table 4.2, columns 4 & 5).

Table 4.2: Lung cancer meta-analysis results using Royston-Parmar models. FTE = fixed treatment effect. * Values are log hazard ratios and 95% credible intervals.

Comparison	FTE*	Cochran's Q	Global non-PH test	Schoenfeld residuals
RT v Seq CT	-0.129	23.72, 21df,	χ^2=0.158, 1df,	χ^2=10.28, 22df,
	$(-0.199, -0.059)$	p=0.307	p=0.691	p=0.984
RT v Con CT	-0.132	17.94, 16df,	χ^2=1.151, 1df,	χ^2=9.93, 17df,
	$(-0.209, -0.056)$	p=0.327	p=0.283	p=0.907
Seq CT v Con CT	-0.176	3.21, 5df,	χ^2=0.236, 1df,	χ^2=9.51, 6df,
	$(-0.297, -0.055)$	p=0.668	p=0.627	p=0.147

4.4   One-step IPD NMA model for time-to-event data

In this section I extend (4.2) to the NMA setting. As far as I aware this is the first time the Royston-Parmar model has been considered in the NMA setting.

The one-step IPD NMA model models the log cumulative hazard individually for each trial with its own spline function (4.1) and location of knots. For patient i in trial j in a network of $q + 1$ treatments the FTE model takes the following form:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1\text{trt1}_{ij} + \cdots + \beta_q\text{trtq}_{ij} \tag{4.5}$$

where trtq_{ij} is a treatment contrast variable, $\beta_1, \ldots, \beta_q$ the treatment effects and $s_j\big(\ln(t)\big)$

is the spline function for trial j as defined in (4.1). In a network of $q+1$ treatments only q treatment contrasts are defined. Some care is needed in defining the treatment contrasts to ensure they are in the right direction. This is necessary for the model to be properly defined. The treatment contrasts are patient level variables which can take the value 0, 1 or -1. Where there are treatment loops in the network, the treatment contrasts represent the consistency equations. For example, in a three-treatment network consisting of treatments A, B and C, where μ_{AB} is the treatment effect of treatment B compared to treatment A, the treatment effect for treatment C compared to treatment B can be calculated as $\mu_{BC} = \mu_{AC} - \mu_{AB}$. This means that only two treatment contrast variables (representing the coefficients of μ_{AB} and μ_{AC}) need defining.

In the cervical cancer network where there are four treatments I chose to define three treatment contrast variables. I chose to define the treatment contrast variables for RT v CTRT, RT v CT+RT and RT v CT+S. In Figure 4.3 the arrows indicate the direction of the treatment effects in the cervical cancer network. RT is the reference treatment for trials comparing RT and CTRT, RT and CT+RT and RT and CT+S. For trials comparing CT+RT and CT+S, CT+RT is the reference treatment and the treatment contrasts need to reflect this. For patients in a CT+RT v CT+S trial receiving CT+S there must be a '-1' for the coefficient of RT v CT+RT and a '1' for the coefficient of RT v CT+S. For patients in a CT+RT v CT+S trial receiving CT+RT the coefficients of RT v CT+RT and RT v CT+S must both be '0'.

In addition, in Section 4.3.1, heterogeneity was identified in the RT v CT+RT comparison and was addressed by splitting the comparison based on the length of chemotherapy cycles. This was incorporated into the NMA model through the inclusion of an additional parameter for cycle length. Through the use of an indicator variable, the additional parameter, can only contribute to the hazard in trials with long chemotherapy cycles. By doing this CT+RT short cycles and CT+RT long cycles are treated as two separate treatments, explaining the heterogeneity.

Consider Figure 4.3 and let $\hat{\beta}_1$ represent the treatment effect for CTRT compared to RT, $\hat{\beta}_2$ represent the treatment effect for CT+RT compared to RT and $\hat{\beta}_3$ represent the treatment effect for CT+S compared to RT. Let $\hat{\beta}_4$ represent the additional parameter which accounts for long chemotherapy cycles, effectively treating short and long cycles as two separate treatments to explain the heterogeneity in the RT v CT+RT comparison. Assuming the network is consistent, and the treatment contrasts are correctly defined, the treatment
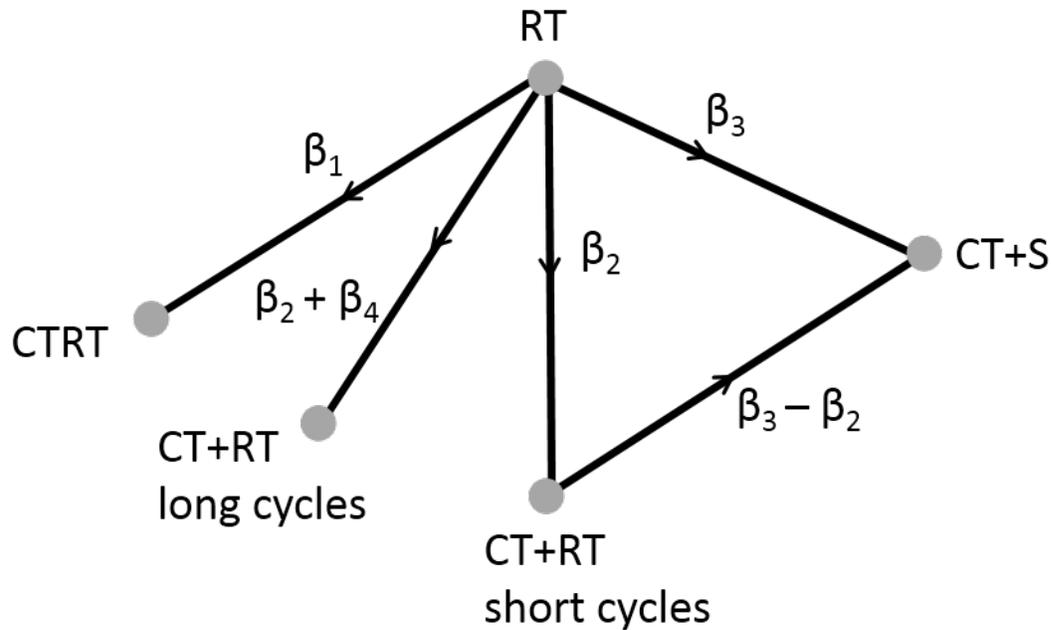
Figure 4.3: Cervical cancer network with treatment effects under the assumption of consistency. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery. $\beta_1, \beta_2, \beta_3$ represent treatment effects with arrows denoting the direction of the treatment effects in NMA models. β_4 denotes the additional effect of long chemotherapy cycles in trials comparing RT and CT+RT.

effect for CT+S compared to CT+RT short cycles can be calculated as $\hat{\beta}_3 - \hat{\beta}_2$.

In the lung cancer network RT is the reference treatment for trials comparing RT and Seq CT and RT and Con CT. For trials comparing Seq CT and Con CT, Seq CT is the reference treatment and the treatment contrasts need to reflect this. For patients in a Seq CT v Con CT trial receiving Con CT there must be a '−1' for the coefficient of RT v Seq CT and a '1' for the coefficient of RT v Con CT. For patients in a Seq CT v Con CT trial receiving Seq CT the coefficients for RT v Seq CT and RT v Con CT must both be '0'. Let $\hat{\beta}_1$ represent the treatment effect for Seq CT compared to RT and $\hat{\beta}_2$ represent the treatment effect for Con CT compared to RT then assuming the network is consistent, and the treatment contrasts are correctly defined, the treatment effect for Con CT compared to Seq CT can be calculated as $\hat{\beta}_2 - \hat{\beta}_1$. Annotated code based on the lung cancer network is provided in Appendix C.

The corresponding RTE model takes the form:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_{1j}trt1_{ij} + \cdots + \beta_{qj}trtq_{ij} \qquad (4.6)$$

$$\begin{pmatrix} \beta_{1j} \\ \vdots \\ \beta_{qj} \end{pmatrix} \sim MVN(\mu, \mathbf{T})$$

where $\mathbf{T}$ is the unstructured inverse between-study variance-covariance matrix. Previously in Section 2.3 I described the Higgins & Whitehead (1996) approach to estimating the between-study variance-covariance matrix. This is a simple approach which requires the estimation of only one parameter, and so is particularly popular when there is relatively little information available to estimate an unstructured covariance matrix. However, this assumption is not always appropriate. Here, I use an unstructured covariance matrix because the cervical and lung cancer networks are simple networks with lots of data which can support the estimation of an unstructured covariance matrix.

4.4.1 Global test for non-proportional hazards

In this section, I propose two approaches for testing the assumption of PH. I first propose a network test for non-PH in a NMA which can be conducted by including an interaction between treatment and ln(time) in a FTE or RTE model. This is done in a similar way to including a treatment-ln(time) interaction in a MA (4.4) and extends the MA approach of Lambert & Royston (2009) to the NMA setting. The FTE model (4.5) can be extended to include treatment-ln(time) interaction parameters in this way:

$$\begin{aligned} \ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1 trt1_{ij} + \cdots + \beta_q trtq_{ij} \\ + \beta_{(q+1)} trt1_{ij}\,(\ln(t)) + \cdots + \beta_{(2q)} trtq_{ij}\,(\ln(t)) \end{aligned} \qquad (4.7)$$

As before (Subsection 4.2.1) the derivative (4.3) of the log cumulative hazard must also be updated. After fitting the model, an approximate global Wald test can be performed on the treatment-ln(time) interaction terms to determine whether there is, on average, any evidence of non-PH within the network. The null hypothesis states that the treatment-ln(time) interactions are simultaneously equal to zero so there is no evidence of non-PH in the network. The Wald test determines whether the treatment-ln(time) interactions could be removed from the model without harming the fit of the model. Details on conducting a Wald test can be found in Subsection 4.4.2.

A second approach which gives more insight into which trials are driving any non-proportionality is to allow the interaction terms to vary by trial. To the best of my knowledge this is the first time this has been proposed in this context. The FTE model (4.5) can be extended in this way:

$$
\begin{aligned}
\ln\{H_j(t|x_{ij})\} = {} & s_j\big(\ln(t)\big) + \beta_1 trt1_{ij} + \cdots + \beta_q trtq_{ij} \\
& + \big(\beta_{(q+1)} + u_j\big) trt1_{ij}\left(\ln(t)\right) + \cdots + \big(\beta_{(2q)} + u_j\big) trtq_{ij}\left(\ln(t)\right)
\end{aligned}
\tag{4.8}
$$

$$
u_j \sim N(0, \sigma_u^2)
$$

As before, an approximate global Wald test of the fixed treatment-ln(time) and variance parameters can then be conducted to determine whether there is any evidence of non-PH within the network. By allowing a random effect of treatment-ln(time) by trial a shrinkage estimate of the departures from PH in each trial is obtained. This can be displayed graphically by plotting the values of the u_j parameters along with an interval of $u_j \pm 1.96$ sd$_j$, where sd$_j$ is the standard deviation of u_j for trial j. Model code, for both approaches, based on the lung cancer network is provided in Appendix C.2.

Non-PH in some or all of the trials can be accommodated by re-fitting (4.7) or (4.8) and restricting the treatment-ln(time) interaction terms to apply only to the trials exhibiting evidence of non-PH. The time-scale could then be divided up and the log hazard ratios assessed within each time interval. Alternatively a spline that allows the treatment effect to vary over time could be added.

4.4.2 Global Wald test

An approximate global Wald test on the parameter estimates can be obtained by importing the WinBUGS output into Stata (StataCorp, 2015) and using the programming language Mata which is accessible through Stata. Let p be the number of interaction terms in the model and n be the number of MCMC updates of the model performed in WinBUGS. Then the following algorithm can be used to conduct the Wald test:

1. Let the matrix $\mathbf{B}$ contain the results from WinBUGS of the n iterations for each interaction term. $\mathbf{B}$ will be a $n \times p$ matrix.

2. Calculate the mean value for each p and store as a vector called $\mathbf{M}$. $\mathbf{M}$ will be a $1 \times p$ vector. Duplicate the row of mean values to get a $n \times p$ matrix in which each row is identical and call this $\mathbf{MM}$. This makes $\mathbf{B}$ and $\mathbf{MM}$ matrices of the same dimension (which is necessary for the next step).

3. Calculate $\mathbf{C} = \mathbf{B} - \mathbf{MM}$. $\mathbf{C}$ will be a $n \times p$ matrix.

4. Calculate $\mathbf{A} = \frac{\mathbf{C}'\mathbf{C}}{n}$. $\mathbf{A}$ will be a $p \times p$ matrix.

5. The Wald test is equal to $\mathbf{M}$ multiplied by the inverse of $\mathbf{A}$ multiplied by the transpose of $\mathbf{M}$ ($\chi^2 = \mathbf{M}\mathbf{A}^{-1}\mathbf{M}'$).

6. Compare χ^2 to a distribution with p degrees of freedom.

As an additional check against the standard deviation estimates from the WinBUGS output, taking the diagonal elements of $\mathbf{A}$ and square rooting them will obtain the standard deviation estimates for each interaction term. An example of Stata code to conduct the Wald test can be found in Appendix D.

4.4.3 Assessment of inconsistency

To assess inconsistency, a fixed effect inconsistency parameter can be introduced to (4.5) following the method of Lu & Ades (Lu and Ades, 2006). This allows estimates of the direct and indirect information to be obtained for each comparison within the treatment loop formed by RT, CT+RT and CT+S in the cervical cancer network and RT, Seq CT and Con CT in the lung cancer network. In a network containing one three-treatment loop between treatments A, B and C, let ω_{ABC} represent the inconsistency parameter for this loop. Then (4.5) becomes:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1\text{trt1}_{ij} + \beta_2\text{trt2}_{ij} - \omega_{ABC}\text{trt1}_{ij}\text{trt2}_{ij}. \tag{4.9}$$

where $s_j\big(\ln(t)\big)$, trt1_{ij}, trt2_{ij}, β_1 and β_2 are as defined in (4.5).

When an inconsistency parameter is included the treatment effects for each comparison in a treatment loop are estimated from the direct evidence only. For example, consider Figure 4.4, a network consisting of three treatments A, B and C where the arrows indicate the direction of treatment effects. The NMA model including an inconsistency parameter will take the same form as (4.9). In this case, as before, the direct
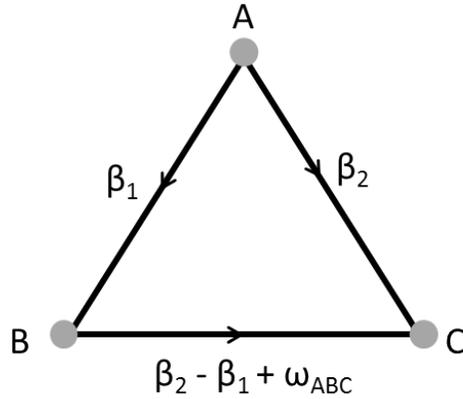
Figure 4.4: Three-treatment network diagram with inconsistency parameter. A, B and C represent treatments. β_1, β_2 and ω_{ABC} represent NMA model parameters.

evidence for A v B will be estimated by $\hat{\beta}_1$ and the direct evidence for A v C will be estimated by $\hat{\beta}_2$. The difference here to (4.5) is that the direct evidence for B v C will be estimated by $\hat{\beta}_2 - \hat{\beta}_1 + \omega_{ABC}$ rather than through the consistency equation. If the direct evidence for all three treatment comparisons can be calculated then the indirect evidence for all three treatment comparisons can also be calculated. The indirect evidence for A v B will be estimated by $\hat{\beta}_2 - (\hat{\beta}_2 - \hat{\beta}_1 + \omega_{ABC}) = \hat{\beta}_1 + \omega_{ABC}$ and the indirect evidence for A v C will be estimated by $\hat{\beta}_1 + (\hat{\beta}_2 - \hat{\beta}_1 + \omega_{ABC}) = \hat{\beta}_2 + \omega_{ABC}$. The indirect evidence for B v C will be estimated by $\hat{\beta}_2 - \hat{\beta}_1$.

A single model can estimate (and test) all inconsistency parameters, because only one model containing an inconsistency parameter for each treatment loop, in which inconsistency could occur, in the network needs to be fitted to separate out the direct and indirect evidence for all the treatment loops in the network. The inclusion of an inconsistency parameter allows testing of inconsistency between two-arm trials only as by definition multi-arm trials are internally consistent. Model code is provided in Appendix C.3 for the lung cancer network. Appendix E details the treatment parameterisation in the presence of an inconsistency parameter.

4.4.4 Assessment of heterogeneity

Cochran's Q statistic can be used to assess heterogeneity within a network. The overall Q statistic from the FTE NMA model can be decomposed into within-design heterogene-

ity (Q^{het}) and between-design heterogeneity representing inconsistency between designs (Q^{inc}). Let $\hat{\theta}_{jk}$ be the treatment effect estimate for trial j of design k, $\hat{\theta}_k$ be the treatment effect from the direct evidence for design k only and $\hat{\theta}_{Nk}$ be the network estimate of the treatment effect for design k then:

$$Q = \sum_k \sum_j \left\{ \frac{\hat{\theta}_{jk} - \hat{\theta}_{Nk}}{\hat{\sigma}_{jk}} \right\}^2$$

$$Q^{\text{inc}} = \sum_k \left\{ \frac{\hat{\theta}_k - \hat{\theta}_{Nk}}{\hat{\sigma}_k} \right\}^2$$

$$Q^{\text{het}} = \sum_k \sum_j \left\{ \frac{\hat{\theta}_{jk} - \hat{\theta}_k}{\hat{\sigma}_{jk}} \right\}^2,$$

with $Q = Q^{\text{inc}} + Q^{\text{het}}$. As written, this applies to two-arm trials only. However, a corresponding matrix decomposition also holds for networks containing multi-arm trials (Gasparrini et al., 2012).

4.4.5 Ranking of treatments

To rank the treatments I took each MCMC iteration in turn and ranked the treatments from most effective to least effective. The most effective treatment had the smallest log hazard ratio value and the least effective treatment had the largest log hazard ratio value. In a network of q treatments I then counted how many times each treatment was considered to be the $1^{st}, 2^{nd}, \ldots, q^{th}$ most effective treatment and expressed these as percentages.

4.5 Prior distributions

In the FTE model, parameters representing the spline function for the baseline log cumulative hazard function, treatment effects, inconsistency parameters and treatment-ln(time) interactions were fitted with non-informative positive half-normal prior distributions ($\boldsymbol{\gamma} \sim N(0, 10000), \boldsymbol{\beta} \sim N(0, 1000), \omega \sim N(0, 10)$). For model (4.8) $\boldsymbol{\sigma_u} \sim N(0, 1000)$ which was restricted to be positive.

In the RTE model, $\boldsymbol{\beta} \sim MVN(\boldsymbol{\mu}, \mathbf{T})$ with $\boldsymbol{\mu} \sim (0, \boldsymbol{\sigma})$ and $\boldsymbol{\sigma}$ a matrix with 0.001 on the diagonal and 0 elsewhere. The prior distribution for $\mathbf{T}$ is an inverse Wishart distribution $\mathbf{T} \sim$

$IW(\mathbf{V}, k)$ where $\mathbf{V}$ is a p x p scale matrix with 0.1 on the diagonal and 0.005 otherwise and the degrees of freedom, $k = p$, are as small as possible to reflect vague prior knowledge. Prior distributions for all other parameters remain the same as for the FTE model.

For all parameters prior distributions were chosen to be non-informative so that the posterior distribution would be driven by the data. Sensitivity analysis considering alternative prior distributions may be useful, but I did not explore this as it was not the main focus of my work.

4.6 Cervical cancer results

In this section I present the results of using the one-step IPD Royston-Parmar approach for NMA to the cervical cancer network. Parameter estimates are presented as log hazard ratios and 95% credible intervals (CrI) for the posterior mean. A log hazard ratio of zero indicates a null effect and a log hazard ratio less than zero indicates a beneficial effect relative to the reference treatment, RT. The MA reported in Subsection 4.3.1 identified heterogeneity in the RT v CT+RT comparison and presented results with the trials split by chemotherapy cycle length. I will treat the data in the same way throughout the rest of this thesis.

Figure 4.5 shows the direct, indirect and network treatment effects for the cervical cancer network. The direct and indirect treatment effects are estimated through the inclusion of an inconsistency parameter as described in Subsection 4.4.3. In this network there is limited indirect evidence so that the network treatment effects are fairly close to the direct effects. The results of the FTE and RTE models are consistent with each other. The DIC provides only weak evidence in favour of the RTE model (difference in DIC of 5, Table 4.3).

Assuming consistency, the network treatment effect for CTRT compared to RT is statistically significant in both the FTE and RTE model with the FTE model suggesting a 19% improvement in overall survival with CTRT (LogHR=-0.211, 95% CrI: -0.337, -0.087, Table 4.3). There was only a statistically significant improvement in overall survival for CT+S compared to RT in the FTE model which suggested a 33% improvement in overall survival (LogHR=-0.396, 95% CrI: -0.611, -0.185). There was no evidence to suggest an effect of CT+RT short cycles compared to RT in either model. However, there was evidence to suggest that CT+RT long cycles reduces overall survival, in the FTE model,

by 25% compared to RT (LogHR=0.223, 95% CrI: 0.065, 0.380).

Table 4.3: NMA results for the cervical cancer network. Values are log hazard ratios and 95% credible intervals. FTE = fixed treatment effect, RTE = random treatment effect, RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery, p_D = effective number of parameters, $\bar{D}$ = posterior mean residual deviance, DIC = deviance information criteria.

	FTE	RTE
RT	0	0
CTRT	-0.211 (-0.337, -0.087)	-0.207 (-0.374, -0.046)
CT+RT short cycles	0.028 (-0.164, 0.220)	0.086 (-0.229, 0.428)
CT+RT long cycles	0.223 (0.065, 0.380)	0.273 (0.031, 0.538)
CT+S	-0.396 (-0.611, -0.185)	-0.333 (-0.701, 0.011)
p_D	138.6	152.2
$\bar{D}$	12182.9	12163.6
DIC	12321.5	12315.8

Comparison	Log Hazard Ratio (95% CrI)		Comparison	Log Hazard Ratio (95% CrI)

RT v CTRT

Direct	-0.215 (-0.340, -0.091)		Direct	-0.198 (-0.362, -0.029)
Network (= direct)	-0.211 (-0.337, -0.087)		Network (= direct)	-0.207 (-0.374, -0.046)

RT v CT+RT short cycles

Direct	-0.074 (-0.289, 0.154)		Direct	-0.026 (-0.446, 0.352)
Indirect	0.357 (-0.056, 0.768)		Indirect	0.458 (-0.309, 1.233)
Network	0.028 (-0.164, 0.220)		Network	0.086 (-0.229, 0.428)

RT v CT+RT long cycles

Direct	0.226 (0.070, 0.383)		Direct	0.283 (0.040, 0.566)
Network (= direct)	0.223 (0.065, 0.380)		Network (= direct)	0.273 (0.031, 0.538)

RT v CT+S

Direct	-0.272 (-0.522, -0.030)		Direct	-0.183 (-0.690, 0.258)
Indirect	-0.704 (-1.112, -0.306)		Indirect	-0.667 (-1.438, 0.109)
Network	-0.396 (-0.611, -0.185)		Network	-0.333 (-0.701, 0.011)

CT+RT short cycles v CT+S

Direct	-0.630 (-0.967, -0.308)		Direct	-0.640 (-1.324, 0.024)
Indirect	-0.198 (-0.531, 0.132)		Indirect	-0.157 (-0.716, 0.394)
Network	-0.425 (-0.654, -0.202)		Network	-0.419 (-0.854, -0.018)

Axis (left): -1.5   0   1.3

Axis (right): -1.5   0   1.3

Figure 4.5: NMA results for the cervical cancer network. Left = fixed treatment effect, Right = random treatment effect. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

4.6.1 Global test for non-proportional hazards

After fitting (4.7), the Wald test for non-PH from the FTE model with fixed treatment-ln(time) interactions gave χ^2=0.932 on 3 degrees of freedom (p=0.818) providing no evidence of non-PH within the network. For the RTE model with random treatment-ln(time) interactions the Wald test for non-PH gave χ^2=0.324 on 3 degrees of freedom (p=0.955) also suggesting that, on average, there is no evidence of non-PH within the network.

Using (4.8), where the treatment-ln(time) interaction parameters are allowed to vary by trial, the Wald test for the FTE model gave χ^2=1.189 on 4 degrees of freedom (p=0.880) and for the RTE model gave χ^2=0.663 on 4 degrees of freedom (p=0.956) suggesting that, on average, there is no evidence of non-PH within the network.

In Section 3.3.1 the Schoenfeld resiudual test identified evidence of non-PH in some trials. The Schoenfeld residual test looks for evidence of a difference in trend in the Schoenfeld residuals between the treatment groups in each trial. A more powerful test of the specific hypothesis that the log cumulative hazard has a different linear trend in $\log(t)$ in the different treatment groups can be conducted by testing the null hypothesis that $\alpha = 0$ in (4.7) and (4.8). In the cervical cancer network testing $\alpha = 0$ in both (4.7) and (4.8) suggested that there was no evidence of non-PH in the network.

Figure 4.6 displays the trial specific deviations from the overall treatment-ln(time) coefficients from the RTE model with random treatment-ln(time) interactions. The deviation in each trial from the overall treatment-ln(time) coefficients are small supporting the conclusion, from the Wald test, that there is no evidence of non-PH in the cervical cancer network. If PH was not an appropriate assumption for the cervical cancer network then Figure 4.6 would show some trials with much larger differences from the overall treatment-ln(time) coefficients and the deviations would not look so uniform across all the trials.

4.6.2 Assessment of inconsistency

In Figure 4.5 the direct and indirect evidence for each treatment comparison are separated out and displayed alongside the network estimates. It can be seen that the direct and indirect treatment effects differ from each other with the network estimates balancing out these two sources of information. By fitting (4.9), the inconsistency parame-
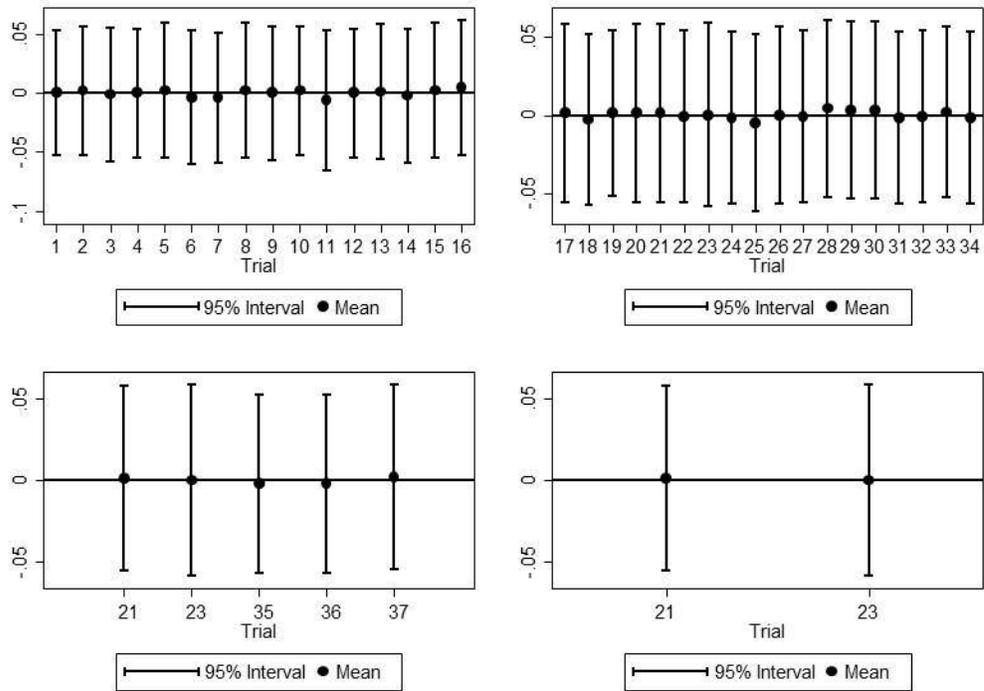
Figure 4.6: Variation in treatment-ln(time) interactions for assessment of non-PH from the RTE NMA model including random treatment-ln(time) interactions applied to the cervical cancer network. Top left: RT v CTRT, top right: RT v CT+RT, bottom left: RT v CT+S, bottom right: CT+RT v CT+S. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

ter quantifies the difference between the direct and the indirect treatment effects for the treatment loop formed by RT, CT+RT short cycles and CT+S. Estimating the inconsistency parameter in this way allows a credible interval to be calculated to quantify the uncertainty surrounding the estimate of the inconsistency parameter. The credible interval can then be used to determine whether the difference between the direct and indirect evidence is statistically significantly different from zero.

Although not statistically significant the inconsistency parameters do hint at evidence of inconsistency within the network. From the FTE model the inconsistency parameter was estimated as -0.432 (95% CrI: -0.905, 0.036) and from the RTE model the inconsistency parameter was estimated as -0.484 (95% CrI: -1.314, 0.354). This suggests that in the cervical cancer network there is an average difference of 0.4, on the log hazard scale, between the direct and indirect treatment effect estimates.

4.6.3 Assessment of heterogeneity

From the FTE model there was evidence of statistically significant heterogeneity in the whole network (Q=56.86 on 35 df, p=0.011) and between designs (Q=10.32, 2 df, p=0.006). There was some evidence of heterogeneity within each design (Q=46.21 on 33 df, p=0.063) which was largely driven by the heterogeneity within the RT v CT+RT short cycles comparison (Q=16.74, 6 df, p=0.010), as previously identified in Subsection 4.3.1. The heterogeneity between designs was driven by the Sardi 96 trial (Sardi et al., 1996). The Sardi 96 trial was a three-arm trial comparing RT, CT+RT and CT+S. Particularly for the RT v CT+S comparison (Figure 4.1), the Sardi 96 trial had a treatment effect estimate more extreme than the other trials. Sensitivity analysis excluding the Sardi 96 trial reduced the overall Q to borderline significance (Q=47.98 on 33 df, p=0.044) and removed the inconsistency between designs (Q=2.53 on 2df, p=0.282). Treatment effect estimates for RT v CT+RT short cycles and RT v CT+S were slightly reduced in both the FTE and RTE models and remained consistent with each other.

4.6.4 Ranking of treatments

The ranking of treatments in order of most effective to least effective was consistent between the FTE and RTE models. In both models CT+S came out as the most effective treatment, CTRT the second most effective treatment, CT+RT short cycles the third most effective treatment, RT the fourth most effective treatment and CT+RT long cycles as the least effective treatment (Figure 4.7).
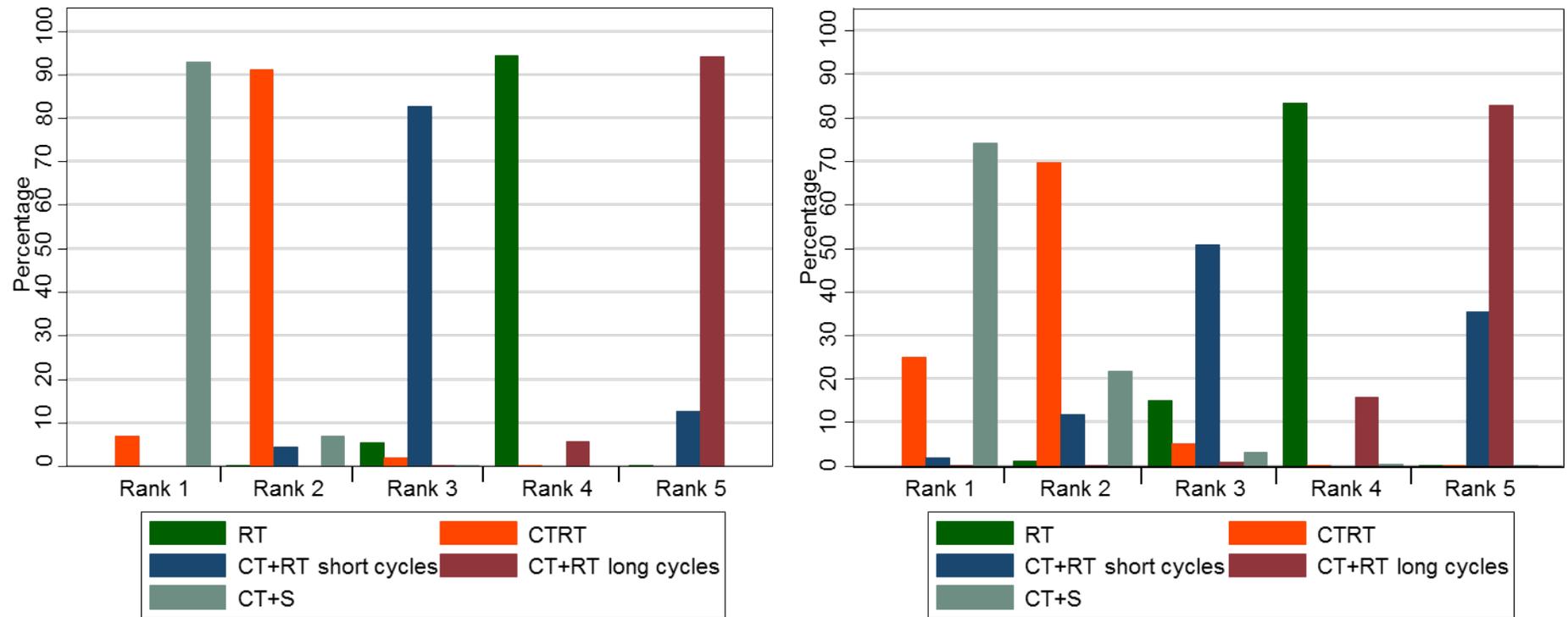
Figure 4.7: Ranking of treatments in the cervical cancer network. Left = fixed treatment effect NMA model, right = random treatment effect NMA model. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

4.7 Lung cancer results

In this section I present the results of applying the one-step IPD Royston-Parmar NMA model to the lung cancer network, which consists of forty-three trials. As before, the parameter estimates are log hazard ratios with corresponding 95% credible intervals. A log hazard ratio of zero indicates a null effect and a log hazard ratio less than one indicates a beneficial effect relative to the reference treatment.

Figure 4.8 shows the direct, indirect and network treatment effects for the lung cancer network. The direct and indirect treatment effects are estimated through the inclusion of an inconsistency parameter as described in Subsection 4.4.3. The results of the FTE and RTE models are consistent with each other and the DIC provides no evidence that the RTE model is a better fit to the data than the FTE model (difference in DIC of 1.5, Table 4.4).

Assuming consistency, the network treatment effect for Seq CT compared to RT is statistically significant in both the FTE and RTE models with the FTE model suggesting a 10% improvement in overall survival with Seq CT (LogHR$=-0.102$, 95% CrI: -0.164, -0.041, Table 4.4). The network treatment effect for Con CT compared to RT is also statistically significant in both the FTE and RTE models with the FTE model suggesting a 16% improvement in overall survival with ConCT (LogHR$=-0.179$, 95% CrI: -0.248, -0.111). The network treatment effect for Con CT compared to Seq CT was not statistically significant in both the FTE and RTE models.

Table 4.4: NMA results for the lung cancer network. Values are log hazard ratios and 95% credible intervals. FTE = fixed treatment effect, RTE = random treatment effect, RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy, p_D = effective number of parameters, $\bar{D}$ = posterior mean residual deviance, DIC = deviance information criteria.

	FTE	RTE
RT v Seq CT	-0.102 (-0.164, -0.041)	-0.100 (-0.191, -0.004)
RT v Con CT	-0.179 (-0.248, -0.111)	-0.169 (-0.279, -0.059)
Seq CT v Con CT	-0.077 (-0.158, 0.002)	-0.070 (-0.200, 0.064)
p_D	167.1	184.5
$\bar{D}$	21368.1	21349.2
DIC	21535.2	21533.7
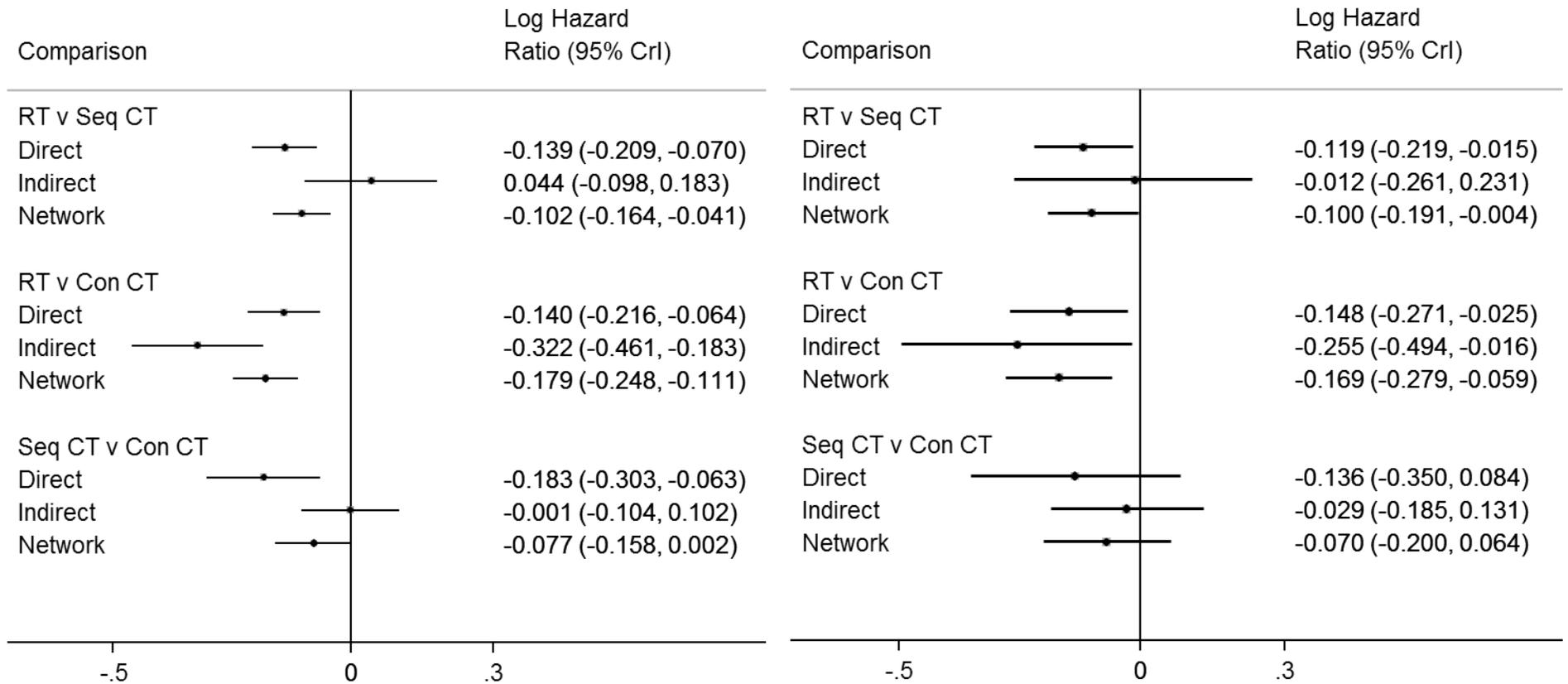
Figure 4.8: NMA results for the lung cancer network. Left = fixed treatment effect, Right = random treatment effect. RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy, CrI = credible interval.

4.7.1 Global test for non-proportional hazards

After fitting (4.7), the approximate global Wald test for non-PH from the FTE model with fixed treatment-ln(time) interactions gave χ^2=1.330 on 2 degrees of freedom (p=0.514) suggesting that, on average, there is no evidence of non-PH within the network. For the RTE model with random treatment-ln(time) interactions the Wald test for non-PH gave χ^2=0.374 on 2 degrees of freedom (p=0.829) also suggesting that, on average, there is no evidence of non-PH within the network.

When the treatment-ln(time) interaction parameters are allowed to vary by trial (4.8) the Wald test for the FTE model gave χ^2=1.646 on 3 degrees of freedom (p=0.649) and for the RTE model gave χ^2=0.907 on 3 degrees of freedom (p=0.824) suggesting that, on average, there is no evidence of non-PH within the network.

In Section 3.3.2 the Schoenfeld resiudual test identified evidence of non-PH in some trials. The Schoenfeld residual test looks for evidence of a difference in trend in the Schoenfeld residuals between the treatment groups in each trial. A more powerful test of the specific hypothesis that the log cumulative hazard has a different linear trend in $\log(t)$ in the different treatment groups can be conducted by testing the null hypothesis that $\alpha = 0$ in (4.7) and (4.8). In the lung cancer network testing $\alpha = 0$ in both (4.7) and (4.8) suggested that there was no evidence of non-PH in the network.

Figure 4.9 displays the trial specific deviations from the overall treatment-ln(time) coefficients from the RTE model with random treatment-ln(time) interactions. Across trials, there is little variation in the deviation from the overall treatment-ln(time) coefficients which supports the conclusion, from the Wald test, that there is no evidence of non-PH in the lung cancer network. If PH was not an appropriate assumption for the lung cancer network then Figure 4.9 would show some trials with much larger differences from the overall treatment-ln(time) coefficients and the deviations would not look so uniform across all the trials.

4.7.2 Assessment of inconsistency

In Figure 4.8 the direct and indirect evidence for each treatment comparison are separated out and displayed alongside the network estimates. There is a suggestion of inconsistency in this network. The point estimates of the direct and indirect evidence for each compar-
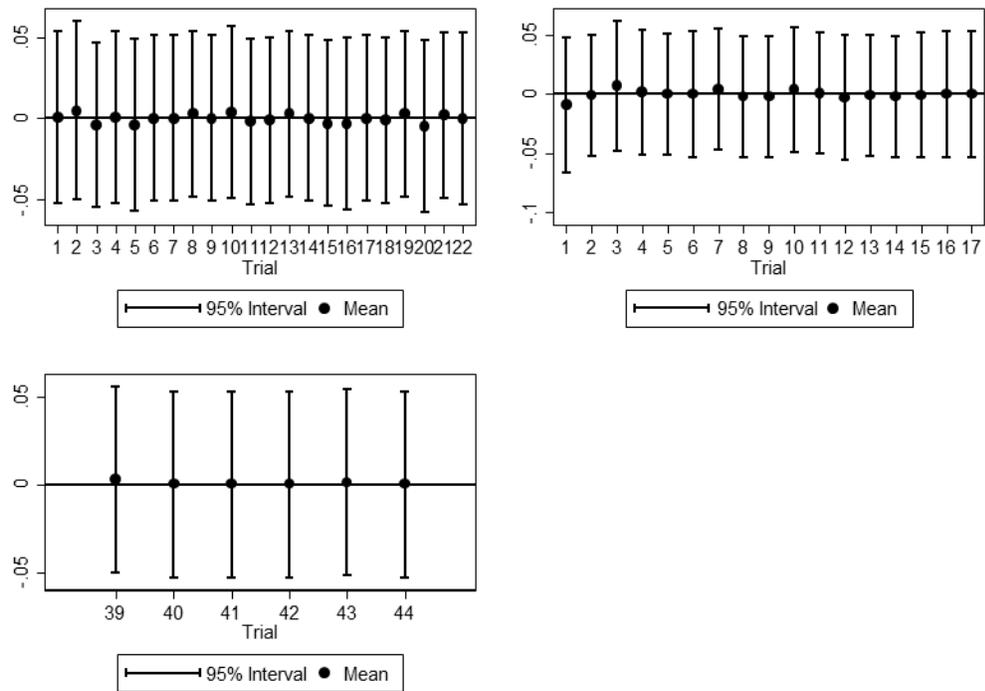
Figure 4.9: Variation in treatment-ln(time) interactions for assessment of non-PH in RTE NMA model including random treatment-ln(time) interactions applied to the lung cancer network. Top left: RT v Seq CT, top right: RT v Con CT, bottom left: Seq CT v Con CT. RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy.

ison differ and the credible intervals only slightly overlap (particularly in the FTE model). Estimating the inconsistency parameter and credible interval allows the uncertainty surrounding the estimate of the inconsistency parameter to be quantified. The credible interval can be used to determine whether the difference between the direct and indirect evidence is statistically significantly different from zero. From the FTE model the inconsistency parameter was estimated as -0.182 (95% CrI: -0.338, 0.024) and from the RTE model the inconsistency parameter was estimated as -0.107 (95% CrI: -0.372, 0.156), suggesting no evidence of inconsistency in the lung cancer network.

4.7.3 Assessment of heterogeneity

From the FTE model there was borderline evidence of statistically significant heterogeneity in the whole network (Q=56.75 on 42 df, p=0.064) and within each design (Q=52.07 on 39 df, p=0.079). There was no evidence of inconsistency between designs (Q=4.68, 3 df, p=0.197). This is in agreement with the DIC which showed little change between the FTE and RTE models.

4.7.4 Ranking of treatments

The ranking of treatments in order of most effective to least effective was consistent between the FTE and RTE models. In both models Con CT came out as the most effective treatment, Seq CT as the second most effective treatment and RT as the least effective treatment (Figure 4.10).
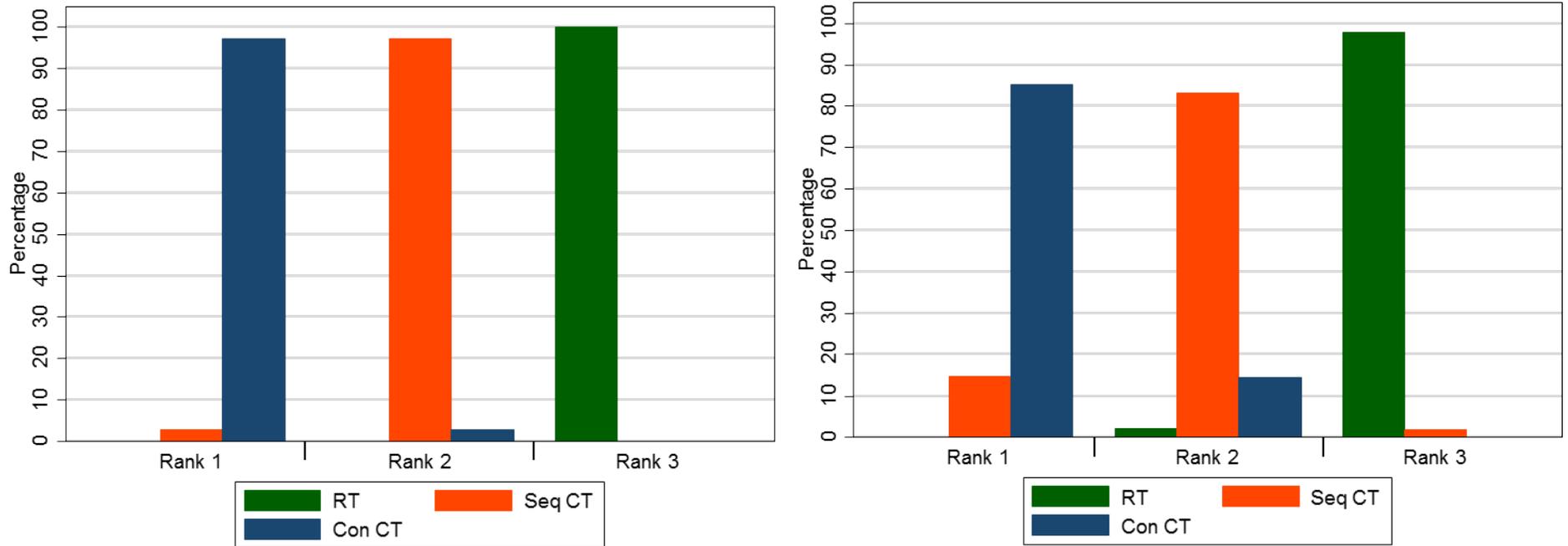
Figure 4.10: Ranking of treatments in the lung cancer network. Left = fixed treatment effect NMA model, right = random treatment effect NMA model. RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy.

4.8 Discussion

In this chapter I have presented the methodology behind fitting a one-step IPD NMA of time-to-event data using the Royston-Parmar model. I have illustrated the potential of this approach by successfully applying the methodology to the cervical cancer and lung cancer networks to assess overall survival.

For the cervical cancer network, the treatment effect estimates in Table 4.3 indicated that both CTRT and CT+S significantly improved overall survival compared to RT. The treatment effect estimate for CT+S compared to RT (33% improvement in overall survival) was larger than the treatment effect estimate for CTRT v RT (19% improvement in overall survival). However, due to the smaller amount of direct evidence available to inform the RT v CT+S comparison the credible intervals for this comparison were much wider and failed to reach statistical significance in the RTE model. The cervical cancer network did not contain any direct evidence comparing CTRT and CT+S. However, by conducting a NMA and ranking the treatments in terms of efficacy the two treatments could be compared to each other. The rankings suggested that CT+S was the most effective treatment in the network. This NMA suggests that direct evidence comparing CTRT and CT+S is needed and it could be used in support of developing trials to directly compare these two treatments.

Globally, cervical cancer is most prevalent in resource poor countries as there is often little or no access to routine screening. In addition one of the difficulties with CTRT in countries, such as India, is that many women live too far away from hospitals, with appropriate RT facilities, to be able to complete the intensive RT schedule required for CTRT. Therefore, CT+S could provide an alternative effective treatment option for women with limited access to RT. My NMA could be used to inform the design for a trial comparing CTRT and CT+S. Both CTRT and CT+S have been shown to be more effective than RT therefore this NMA suggests that a head-to-head trial of CTRT compared to CT+S would be most appropriate. My NMA could also be used to inform the sample size calculation and the choice of prior distributions for any future trials directly comparing CTRT and CT+S. I am aware of two phase III trials in India (ClinicalTrials.gov, 2015a,b) currently recruiting patients to directly compare CTRT and CT+S. Once completed these two trials could be used to update this NMA and hopefully provide a more definitive answer as to which treatment is most effective.

In Chapter 3, two trials from the cervical cancer network were identified as showing pos-

sible evidence for non-PH. However, there was no global evidence of non-PH in this network. Therefore, throughout this thesis I analyse this network assuming hazards are proportional. There was no evidence of statistically significant inconsistency in the network. There was a suggestion of heterogeneity in the RT v CT+RT comparison which was reduced by splitting the comparison into two based on the length of chemotherapy cycles. Some heterogeneity remained in the RT v CT+RT short cycles comparison.

Cochran's Q statistic found some evidence of heterogeneity within the cervical cancer network but sensitivity analysis showed this could be reduced with the exclusion of the Sardi 96 trial (Sardi et al., 1996). The Sardi 96 trial was a three-arm trial comparing RT, CT+RT and CT+S. Looking back at Figure 4.1 it can be seen that for the RT v CT+S comparison the most extreme treatment effect estimate is from the Sardi 96 trial. Therefore, excluding the Sardi 96 trial reduced the amount of heterogeneity in the network. Excluding the Sardi 96 trial slightly reduced the treatment effect estimates for RT v CT+RT short cycles and RT v CT+S however the conclusions remained the same. The more extreme treatment effect in the Sardi 96 trial could be explained by the baseline characteristics of the trial compared to the other trials in the RT v CT+S comparison. In this comparison stage of disease ranges from IB up to IIIB. However, all patients in the Sardi 96 trial have stage IIIB disease making up 57% of the total number of patients in the RT v CT+S comparison with stage IIIB disease. In addition, the Sardi 96 trial also has the smallest proportion of patients with a performance status of 0 in the RT v CT+S comparison. Therefore, the treatment effect in the Sardi 96 trial could be explained by the trial having a greater proportion of patients with more advanced disease and worse health than the other trials in the RT v CT+S comparison. However, an extreme treatment effect on it's own is not a valid reason to exclude a trial from the network therefore the Sardi 96 trial remains included in the cervical cancer network throughout the rest of this thesis.

The NMA of the lung cancer network illustrated that both Seq CT and Con CT improved overall survival compared to RT, in both the FTE and RTE models. Seq CT improved overall survival by 10% compared to RT, Con CT improved overall survival by 16% compared to RT and there was no evidence of a statistically significant improvement in overall survival for Con CT compared to Seq CT. In Figure 4.8 the direct, indirect and network evidence all suggest that Con CT improves overall survival compared to RT. The direct evidence for RT v Seq CT suggests that Seq CT improves overall survival compared to

RT however the indirect evidence is inconclusive. Similarly the direct evidence for Seq CT v Con CT suggests that Con CT improves overall survival compared to Seq CT however the indirect evidence is inconclusive. In both cases, the uncertainty of the indirect evidence results in uncertainty in the network estimate. After ranking the treatments it is clear that Con CT is the most effective treatment followed by Seq CT. There was no evidence of statistically significant inconsistency in the network. Similarly to the cervical cancer network, Chapter 3 identified evidence of non-PH in some trials. However, there was no global evidence of non-PH in the lung cancer network. Therefore throughout this thesis I analyse this network assuming hazards are proportional.

This chapter adds to the small pool of literature for analysing time-to-event data by extending the work of Royston and Parmar (Royston and Parmar, 2002) to the NMA setting and showing that Royston-Parmar models, fitted in WinBUGS, provide a flexible, practical approach for Bayesian NMA with time-to-event data. Royston-Parmar models avoid the computational issues that beset a Bayesian implementation of the Cox PH model. To fit the Cox PH model in the Bayesian setting the data for each individual has to be repeated for each risk set they belong to. This makes it extremely cumbersome and computationally intractable for even moderately sized datasets such as the cervical cancer network. The full power of the Bayesian framework was not utilised in this chapter, as there was no evidence for non-PH or inconsistency, in either network. Nevertheless this chapter provides a convincing proof-of-concept and opens the door to better predictions, which can be naturally done in the Bayesian framework. An appropriate baseline hazard is required for predictions. An advantage of the Royston-Parmar approach is that, an estimate of the baseline hazard, pooled across trials, can be obtained. To do this, the coefficients for the restricted cubic spline are made random across trials, which requires the knots to be in the same position for all studies. The Bayesian approach also provides a computationally straightforward and inferentially natural framework for ranking treatments. Ranking treatments can be extremely useful in networks containing many treatments and can also prove particularly useful when comparing treatments not previously compared in head-to-head trials.

The inclusion of treatment-ln(time) interactions allows for, tests and accommodates departures from the PH assumption in some or all of the trials within a network. The extent of departure from PH can be assessed in each trial by making the treatment-ln(time) interaction terms random (4.8). This results in Bayesian shrinkage estimates which reduce the likeli-

hood of over interpreting departures from proportionality in smaller studies. Where proportionality is not appropriate, the Royston-Parmar approach naturally allows for — for example — effect estimation using restricted mean survival time as an estimate of treatment efficacy (Royston and Parmar, 2011), which has so far been considered only in the MA setting (Wei et al., 2015).

To fit the one-step IPD Royston-Parmar NMA model a RCS must be calculated for each trial. The NMA model itself is fitted in WinBUGS, therefore it makes sense to use either Stata or R to calculate the RCS, because they can both easily interact with WinBUGS. As described in Subsection 4.2.2 the basis functions are calculated using the survival time for each patient. If this is done in Stata or R, then along with the treatment contrast variables, the data can be passed over to WinBUGS where the model is fitted before being passed back to Stata or R. The advantage of this is that calculating the RCS for each trial can be an automated process. Another approach to NMA which, like the Royston-Parmar approach, can account for treatment effects changing over time, is to model the log hazard rate using fractional polynomials (Jansen, 2011). However, compared to the Royston-Parmar approach fractional polynomials are harder to fit in an automated way because, as they are polynomials, they are prone to end effects. The RCS which is linear at each end is much more robust in this regard.

NMA combines both direct and indirect evidence. Since the latter requires much stronger assumptions it is important to check they are consistent. In this chapter I used a model-based version of the Bucher method (Bucher et al., 1997) to assess consistency in the cervical and lung cancer networks. The Bucher method can also be applied to more complex networks. One inconsistency parameter is required for each treatment loop, and the model can simply be re-fitted with all these parameters included. This allows the direct and indirect contributions to each treatment effect to be separated so they can be presented alongside the network estimate (as in Figure 4.5 and Figure 4.8), providing a visual display of the agreement between the direct and indirect evidence and the extent to which conclusions might be based on the indirect evidence. This is important as researchers should be aware of the extent to which conclusions rely on the indirect evidence, with its additional assumption of no unmeasured confounding. In both networks, a small amount of evidence for inconsistency came from the NMA model including an inconsistency parameter which was close to the boundary of $p=0.05$ for statistical significance. Conclusions based on the

network treatment effects remained the same as conclusions based on the direct treatment effects. Therefore, I will continue with the assumption of consistency in both networks but, going forward, will be cautious about the possibility of inconsistency in the networks.

In summary, Bayesian NMA of IPD offers many practical advantages, but is computationally problematic with the Cox PH model, even with moderately sized datasets. In this chapter I have shown that the Royston-Parmar model provides a flexible, computationally practical, way forward which can extend to accommodate issues such as non-PH which are increasingly arising in oncology studies. This chapter provides a base from which extensions to the one-step IPD Royston-Parmar NMA model, such as the inclusion of patient-level covariates and treatment-covariate interactions, can be considered. There is the potential to address many issues within NMA including assessing inconsistency (Chapter 5) and treatment-covariate interactions (Chapter 6). In Chapter 5 I will explore the assessment of inconsistency further using the lung cancer network.

5 Assessing Inconsistency in Network Meta-Analysis

5.1 Introduction

Inconsistency in NMA occurs when the direct and indirect evidence are not in agreement with each other. In this case, combining them can result in biased treatment effect estimates. Potential sources of inconsistency within a network may arise from bias in direct comparisons, for example optimism bias and publication bias, sponsorship bias and even different trial populations across comparisons in a network (Ioannidis, 2009; Salanti et al., 2007). The power of tests for inconsistency is generally considered to be low because indirect evidence is typically a relatively weak component of most treatment estimates in NMA. Therefore, failure to reject the null hypothesis of no inconsistency does not mean that the entire network is consistent (Veroniki et al., 2013). Nevertheless the increasing use of NMA in health decision modelling means that it is important that attempts are made to identify, understand and, where appropriate, adjust for inconsistency.

As is typical in the NMA literature, throughout this chapter 'design' will refer to the treatments being compared within a trial (Higgins et al., 2012). For example, two trials both comparing treatment A to treatment B will be considered to be of the same design, whereas a third trial comparing treatment A to treatment B and treatment C will be considered to be a different design. In addition throughout this chapter the shorthand *dir* represents direct evidence, *ind* represents indirect evidence and *net* represents network evidence (i.e. the combination of the direct and indirect evidence).

There are several approaches for assessing inconsistency in a network, including Cochran's Q statistic (1954), the loop inconsistency approach (Bucher et al., 1997), the inconsistency parameter approach (Lu and Ades, 2006) and the net heat approach (Krahn et al., 2013). Between them, these four methods offer a range of increasingly complex methods for identifying inconsistency in a network. Cochran's Q statistic (Cochran, 1954) and the loop inconsistency approach of Bucher (1997) are relatively simple methods which aim to identify inconsistency through one test statistic and a p-value. The inconsistency parameter approach of Lu & Ades (2006) allows for inconsistency in a Bayesian hierarchical model, which allows the amount of inconsistency to be quantified and a credible interval calculated. Krahn (2013) also use a modelling approach; however the results

are displayed graphically allowing inconsistency to be identified, located and quantified. All four methods are described in more detail in Chapter 2. In this chapter I will apply these four methods to the lung cancer network, critique the approaches and make a suggestion for an alternative approach to assessing inconsistency.

This chapter continues in Section 5.2 by assessing inconsistency in the lung cancer network using Cochran's Q statistic, the loop inconsistency approach and the inconsistency parameter approach. In Section 5.3, I provide an overview of the net heat approach. In Section 5.4 I present a detailed explanation of the net heat approach illustrated using the lung cancer network. In Section 5.5 I apply the net heat principle to several scenarios, showing why the net heat approach can be misleading when assessing inconsistency. In Section 5.6 I propose an alternative method for assessing inconsistency in NMA. I finish in Section 5.7 with a discussion.

Work from this chapter was first presented in an oral presentation at the International Society of Clinical Biostatistics conference in August 2016 and a journal paper will be submitted to Statistics in Medicine following approval from the data provider.

5.2 Exploring inconsistency in the lung cancer network

In this section I will assess inconsistency in the lung cancer network, introduced in Subsection 3.2.2, using the methods proposed by Cochran (1954), Bucher (1997) and Lu & Ades (2006). To simplify calculations the multi-arm trial, GMMA Ankara 1995 (Ulutin et al., 2000), is excluded from the network throughout the rest of this chapter. Therefore in this chapter the lung cancer network consists of 42 two-arm trials comparing three different treatments (RT, Seq CT, Con CT).

The treatment effects for the three different comparisons were estimated in a number of ways. Network estimates combining both direct and indirect treatment effects were obtained by fitting the one-step IPD Royston-Parmar NMA model for time-to-event data (4.5) and by fitting a two-step NMA using the R package netmeta (Rücker et al., 2014). An estimate of the direct evidence was obtained by fitting the one-step IPD Royston-Parmar MA model (4.2) to trials directly comparing the treatments of interest only. Indirect treatment effects were also calculated using the one-step IPD Royston-Parmar MA model, where all trials directly comparing the two treatments of interest were excluded from the model leaving only the indirect evidence to be synthesised. Finally, to assess inconsistency and estimate both the direct and indirect evidence simultaneously I conducted a NMA using the one-step IPD Royston-Parmar time-to-event model including a fixed effect inconsistency parameter (4.9) following the method of Lu & Ades (2006). Throughout this chapter all models are fitted with fixed effects assuming no heterogeneity in any of the direct comparisons. The one-step IPD Royston-Parmar NMA model was fitted in the same way as in Chapter 4 with the same non-informative prior distributions.

The forest plot of treatment effects for each pairwise comparison, using the methods described above, is presented in Figure 5.1, which clearly shows a difference between the direct and indirect evidence for each pairwise comparison, suggesting a strong possibility of inconsistency within this network of trials.

Following the approach of Lu & Ades (2006) the inconsistency parameter from the one-step IPD Royston-Parmar NMA model was estimated as -0.176 (95% CrI: -0.337, -0.016), giving an approximate p-value of 0.032 and suggesting evidence of network inconsistency.

In a loop of three treatments A, B and C, the Bucher (1997) approach compares the direct

Treatment comparison & model	Log Hazard Ratio (95% CrI)
RT v RT + Seq CT:	
Royston-Parmar NMA	-0.098 (-0.161, -0.035)
netmeta	-0.098 (-0.161, -0.036)
Direct only	-0.132 (-0.202, -0.062)
Direct (IP)	-0.134 (-0.202, -0.061)
Indirect (IP)	0.043 (-0.100, 0.184)
Indirect only	0.041 (-0.106, 0.182)
RT v RT + Con CT:	
Royston-Parmar NMA	-0.177 (-0.244, -0.110)
netmeta	-0.177 (-0.244, -0.110)
Direct only	-0.139 (-0.215, -0.063)
Direct (IP)	-0.138 (-0.217, -0.063)
Indirect (IP)	-0.314 (-0.455, -0.174)
Indirect only	-0.311 (-0.455, -0.171)
RT + Seq CT v RT + Con CT:	
Royston-Parmar NMA	-0.079 (-0.159, -0.001)
netmeta	-0.079 (-0.157, -0.000)
Direct only	-0.179 (-0.302, -0.054)
Direct (IP)	-0.181 (-0.304, -0.061)
Indirect (IP)	-0.004 (-0.110, 0.098)
Indirect only	-0.007 (-0.108, 0.095)

Figure 5.1: Forest plot of various analyses of the lung cancer data. All models are fitted with FTE. RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy, NMA = Network meta-analysis, IP = Inconsistency parameter, CrI = credible interval (except netmeta models where confidence intervals are presented).

evidence of treatment C versus treatment B, $\hat{d}_{BC}^{\text{dir}}$, to the indirect evidence, $\hat{d}_{BC}^{\text{ind}}$, where $\hat{d}_{BC}^{\text{ind}} = \hat{d}_{AC}^{\text{dir}} - \hat{d}_{AB}^{\text{dir}}$ and $\text{Var}(\hat{d}_{BC}^{\text{ind}}) = \text{Var}(\hat{d}_{AC}^{\text{dir}}) + \text{Var}(\hat{d}_{AB}^{\text{dir}})$. Estimates of the inconsistency parameter, $\hat{\omega}_{BC}$, and its variance can be formed by subtracting the direct and indirect estimates:

$$\hat{\omega}_{BC} = \hat{d}_{BC}^{\text{dir}} - \hat{d}_{BC}^{\text{ind}} \tag{5.1}$$

$$\text{Var}(\hat{\omega}_{BC}) = \text{Var}(\hat{d}_{BC}^{\text{dir}}) + \text{Var}(\hat{d}_{BC}^{\text{ind}}) = \text{Var}(\hat{d}_{BC}^{\text{dir}}) + \text{Var}(\hat{d}_{AB}^{\text{dir}}) + \text{Var}(\hat{d}_{AC}^{\text{dir}})$$

An approximate test of the null hypothesis of consistency is conducted by referring the test statistic $z_{BC} = \frac{\hat{\omega}_{BC}}{\sqrt{\text{Var}(\hat{\omega}_{BC})}}$ to the normal distribution.

Letting A=RT, B=Seq CT, C=Con CT, for the lung cancer network I have:

$$\hat{d}_{AB}^{\text{dir}} = -0.132, \text{Var}(\hat{d}_{AB}^{\text{ind}}) = 0.036^2$$

$$\hat{d}_{AC}^{\text{dir}} = -0.139, \text{Var}(\hat{d}_{AC}^{\text{ind}}) = 0.039^2$$

$$\hat{d}_{BC}^{\text{dir}} = -0.179, \text{Var}(\hat{d}_{BC}^{\text{ind}}) = 0.062^2$$

$$\hat{d}_{BC}^{\text{ind}} = -0.139 - -0.132 = -0.006$$

$$\text{Var}(\hat{d}_{BC}^{\text{ind}}) = 0.039^2 + 0.036^2 = 0.053^2$$

$$\hat{\omega}_{BC} = -0.179 - -0.006 = -0.173$$

$$\text{Var}(\hat{\omega}_{BC}) = 0.062^2 + 0.036^2 + 0.039^2 = 0.006594$$

$$z_{BC} = \frac{-0.173}{\sqrt{0.006594}} = -2.126, p = 0.0335$$

Therefore, Bucher's method provides similar evidence of inconsistency within the lung cancer network, as expected.

Cochran's Q statistic can also be used to assess heterogeneity within the network. The overall Q statistic from the one-step IPD Royston-Parmar NMA model can be decomposed into within-design heterogeneity (Q^{het}) and between-design heterogeneity representing inconsistency between designs (Q^{inc}). Let $\hat{\theta}_{jk}$ be the treatment effect estimate for the comparison of treatments in design k with corresponding standard error $\hat{\sigma}_{jk}$, $\hat{\theta}_k$ be the treatment effect from the direct evidence for design k only and $\hat{\theta}_{Nk}$ be the network estimate of the treatment effect for design k then:

$$Q = \sum_k \sum_j \left\{ \frac{\hat{\theta}_{jk} - \hat{\theta}_{Nk}}{\hat{\sigma}_{jk}} \right\}^2$$

$$Q^{\text{het}} = \sum_{k} \sum_{j} \left\{ \frac{\hat{\theta}_{jk} - \hat{\theta}_k}{\hat{\sigma}_{jk}} \right\}^2$$

$$Q^{\text{inc}} = \sum_{k} \left\{ \frac{\hat{\theta}_k - \hat{\theta}_{Nk}}{\hat{\sigma}_k} \right\}^2,$$

with $Q = Q^{\text{het}} + Q^{\text{inc}}$.

At the 5% level, in the lung cancer network the one-step IPD Royston-Parmar NMA model showed evidence of statistically significant heterogeneity in the whole network (Q=56.59, 40 df, p=0.0428) and between designs (Q^{inc}=4.52, 1 df, p=0.0335). Heterogeneity within designs was also borderline significant (Q^{het}=52.07, 39 df, p=0.0786).

The lung cancer network consists of three treatments and two-arm trials only. In this setting the results of Q^{inc}, Bucher's test and the inconsistency parameter are all equal. Clearly in a network like this, only one of these approaches is needed. However, in the case of larger networks or networks including multi-arm trials, one approach could be to calculate Q^{inc} first and then, if this shows evidence of inconsistency, use the inconsistency parameter approach. The inconsistency parameter approach allows for explicit statistical tests and can focus on particular areas of the network. The inconsistency parameter approach will also allow for differentiation between design and loop inconsistency. For example, in a three-treatment network consisting of two and three-arm trials, if two-arm trials are not consistent with three-arm trials then there is design inconsistency. However, if the two-arm trials are not comparable with each other then there is loop inconsistency. It is feasible for the two-arm trials to be consistent with each other but inconsistent with the three-arm trials.

In summary, testing the inconsistency parameter, Bucher's test and the between designs Q statistic all suggest evidence of inconsistency between the direct and indirect evidence in the lung cancer network.

5.3 Net heat approach

Krahn (2013) introduced the net heat plot as a method for identifying and locating inconsistency within a network of RCTs. In a network of RCTs with at least one treatment loop the net heat plot is constructed by detaching each design one at a time and assessing the contribution of each design to the inconsistency of the whole network.

Krahn (2013) propose the use of a design-by-treatment interaction approach, whereby one of the designs is saturated so that the remaining inconsistency across the network can be calculated. In practice, this is computationally simple because it is equivalent to a 'leave one out' approach in which Q^{inc} is simply recalculated from scratch after the (temporary) removal of each design in turn. Designs which do not contribute to a treatment loop or when removed would split the network into two distinct parts are excluded from the net heat plot. I now give an overview of the Krahn approach, before illustrating it on the lung cancer network and then exploring it in further detail.

The between design Q^{inc} statistic is calculated as the sum of the inconsistency in each design c, i.e. $Q^{\text{inc}} = \sum_c Q_c^{\text{inc}}$. Let Q_c^{inc} represent the inconsistency in the network for design c before any designs are detached, $Q_{c(d)}^{\text{inc}}$ the inconsistency remaining in the network for design c when design d is detached and $Q_{c,d}^{\text{diff}}$ denote the change in inconsistency for design c resulting from detaching design d. Then:

$$Q_{c,d}^{\text{diff}} = Q_c^{\text{inc}} - Q_{c(d)}^{\text{inc}}$$

The values of $Q_{c,d}^{\text{diff}}$ form the basis of the net heat plot. The net heat plot is constructed as a matrix in which each off-diagonal square represents the contribution of the row design to the total inconsistency across the network when the column design is detached (i.e. the consistency assumption is removed for the column design). The leading diagonal, running from the top left to the bottom right corner, displays the contribution of each design c to the between design statistic, Q^{inc}.

Additionally, in each net heat plot the area of the grey squares within each matrix cell are proportional to the absolute values of the hat matrix. These are interpretable as the contribution of the direct estimate of the column design to the network estimate of the row design. As proposed by Krahn, the net heat plot is coloured on a scale according to values of $Q_{c,d}^{\text{diff}}$. The colouring varies in intensity with the maximum intensity (i.e. the brightest
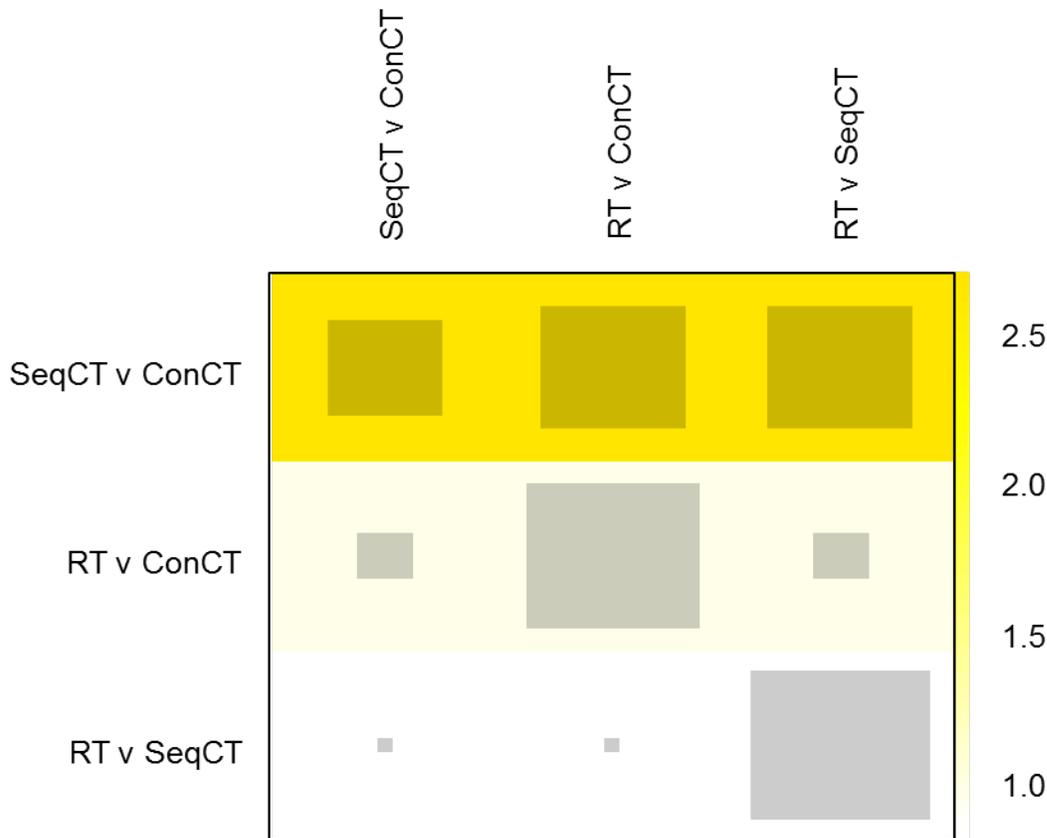
Figure 5.2: Net heat plot for the lung cancer network. RT = radiotherpy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy.

colours) representing absolute values of $Q_{c,d}^{\text{diff}}$ greater than or equal to eight. The scale runs from red for values of $Q_{c,d}^{\text{diff}} = 8$ through yellow shades, down to white for $Q_{c,d}^{\text{diff}} = 0$ down to bright blue for $Q_{c,d}^{\text{diff}} = -8$. Red and yellow colours indicate that the evidence for the row design from the column design is inconsistent with the other evidence in the network. Blue colours indicate that the evidence for the row design from the column design is consistent (Schwarzer et al., 2015). This enables the reader to identify which designs are most likely to be responsible for the inconsistency in the network.

Net heat plots can be produced with the package netmeta (Rücker et al., 2014) in R (R Core Team, 2014). This package was used to construct a net heat plot for the lung cancer network, shown in Figure 5.2.

The net heat plot can be awkward to interpret as it is unclear how vibrant the colours need to be to conclude that there is statistically significant or clinically important evidence of inconsistency within a network. This is because there is no reference sampling

distribution for Q^{diff}. Figure 5.2 contains varying shades of yellow indicating $Q_{c,d}^{\text{diff}} > 0$ and suggesting that there could be inconsistency in the network. The difference in the shades of yellow suggests that inconsistency is most likely in the Seq CT v Con CT treatment comparison; however in a three-treatment network inconsistency can only be identified and not actually located. The Seq CT v Con CT comparison has the least amount of direct evidence and therefore the decomposition of Q has attributed the inconsistency mainly to this comparison. There are no areas of vibrant red colour in the plot so it would seem reasonable, according to the guidelines of Krahn (2013), to conclude that there is no meaningful inconsistency in the lung cancer network. This appears to contradict the findings from the previous analyses in Section 5.2 resulting in a paradox. In Section 5.4 I investigate this paradox in more detail.

5.4   How is the net heat plot formed?

Section 5.3 culminated in a paradox in which the net heat plot suggested no evidence of inconsistency in the lung cancer network contradicting the findings from earlier analyses in Section 5.2 which concluded that there was evidence of inconsistency in the lung cancer network. Therefore, in this section I investigate this paradox by presenting an expanded and illustrated explanation of the extremely condensed discussion in the Krahn paper (Krahn et al., 2013), which is a necessary pre-amble to understand what is going on and resolve the paradox. It is unclear why Krahn did not expand their explanation and instead chose to focus on explaining the calculations required to create the net heat plot rather than explaining the methodology underpinning these calculations. Possibly they wanted the focus to be on a practical example of the net heat plot in action. They may also have been limited by the journal in terms of article length. In this section I use the lung cancer network to illustrate the methodology behind the R package netmeta (Rücker et al., 2014). Once again, in this section I use only the two-arm trials from the lung cancer network to illustrate the net heat approach.

The net heat plot is the result of a two-step approach to conducting NMA using AD. Each trial in a NMA contributes an estimate of the treatment effect and its corresponding standard error. In the lung cancer network an estimate of the treatment effect and standard error for each trial were calculated using the Royston-Parmar model. To begin, an estimate of
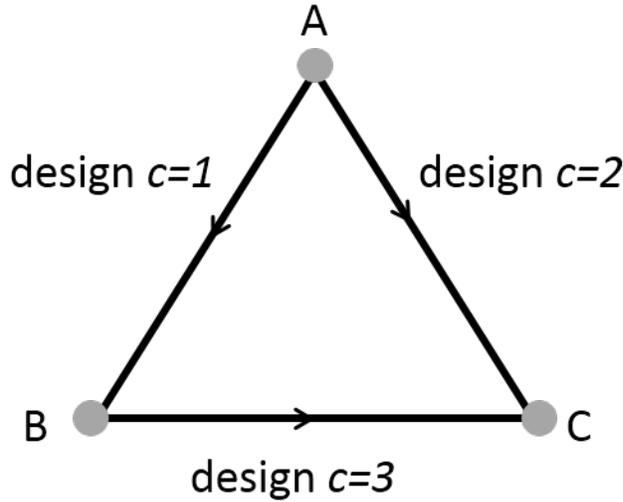
Figure 5.3: Example of a three treatment network. A, B and C represent the three treatments being compared.

the direct treatment effect, $\hat{\theta}_c^{\text{dir}}$, and its variance, $\hat{V}_c^{\text{dir}}$, must be calculated for each design c.

In the lung cancer network there are three designs. In Figure 5.3, let A=RT, B=Seq CT, C=Con CT and $c = 1$ represent the set of trials directly comparing A v B, $c = 2$ the set of trials directly comparing A v C and $c = 3$ the set of trials directly comparing B v C. Let $\mathcal{C}$ denote the set of all trials of design c. The direct treatment effect for each design c can be calculated as:

$$\hat{\theta}_c^{\text{dir}} = \left(\sum_{j \in \mathcal{C}} \hat{V}_{jc}^{-1} \right)^{-1} \sum_{j \in \mathcal{C}} \hat{V}_{jc}^{-1} Y_{jc}$$

$$\hat{V}_c^{\text{dir}} = \left(\sum_{j \in \mathcal{C}} \hat{V}_{jc}^{-1} \right)^{-1}$$

where Y_{jc} is a column vector containing the direct treatment effect estimate from the Royston-Parmar model for each trial j of design c, and $\hat{V}_{jc}$ is the estimated variance matrix of Y_{jc}.

In the lung cancer network there are 21 clinical trials of design $c = 1$ so Y_1 is a column vector with 21 entries and V_1^{dir} is a matrix with 21 rows and 21 columns, while in design $c = 2$ there are 16 trials and in design $c = 3$ there are 5 trials. For the lung cancer network the direct treatment effect estimates are:

$$\hat{\theta}_1^{\text{dir}} = 0.1316, \hat{V}_1^{\text{dir}} = 0.0013$$

$$\hat{\theta}_2^{\text{dir}} = 0.1378, \hat{V}_2^{\text{dir}} = 0.0015$$

$$\hat{\theta}_3^{\text{dir}} = 0.1787, \hat{V}_3^{\text{dir}} = 0.0038$$

Having obtained the direct effect estimates and their variances, the estimates of $\hat{\theta}_1^{\text{dir}}$, $\hat{\theta}_2^{\text{dir}}$ and $\hat{\theta}_3^{\text{dir}}$ are put into a column vector so that:

$$\hat{\theta}^{\text{dir}} = \begin{pmatrix} \hat{\theta}_1^{\text{dir}} \\ \hat{\theta}_2^{\text{dir}} \\ \hat{\theta}_3^{\text{dir}} \end{pmatrix} = \begin{pmatrix} 0.1316 \\ 0.1378 \\ 0.1787 \end{pmatrix}$$

The estimates of $\hat{V}_1^{\text{dir}}$, $\hat{V}_2^{\text{dir}}$ and $\hat{V}_3^{\text{dir}}$ are put into a diagonal matrix so that:

$$\hat{V}_a = \begin{pmatrix} \hat{V}_1^{\text{dir}} & 0 & 0 \\ 0 & \hat{V}_2^{\text{dir}} & 0 \\ 0 & 0 & \hat{V}_3^{\text{dir}} \end{pmatrix} = \begin{pmatrix} 0.0013 & 0 & 0 \\ 0 & 0.0015 & 0 \\ 0 & 0 & 0.0038 \end{pmatrix}$$

In a consistent three-treatment network there are two basic parameters, for example $\hat{\theta}_{AB}^{\text{net}}$ and $\hat{\theta}_{AC}^{\text{net}}$. From these two basic parameters the third network effect, $\hat{\theta}_{BC}^{\text{net}}$, can be calculated:

$$\hat{\theta}_{BC}^{\text{net}} = \hat{\theta}_{AC}^{\text{net}} - \hat{\theta}_{AB}^{\text{net}}$$

The design matrix, X, contains the structure of the network at the study level and links the observed treatment effects with the basic parameters. Therefore, the number of columns in the design matrix is one less than the number of treatments in the network. The full design matrix X contains one row for each study, while the compressed design matrix X_a contains one row for each design in the network.

In the lung cancer network the compressed design matrix, X_a, in which each row corresponds to one design, consists of 3 rows and 2 columns (to represent the two basic parameters) and $\hat{\theta}^{\text{net}}$ consists of the two basic parameters $\hat{\theta}_{AB}^{\text{net}}$ and $\hat{\theta}_{AC}^{\text{net}}$.

$$X_a = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}$$

Because the lung cancer network consists of two-arm trials only in a closed loop, $\hat{\theta}^{\text{net}}$ can be calculated as:

$$\hat{\theta}^{\text{net}} = (X_a' \hat{V}_a^{-1} X_a)^{-1} X_a' \hat{V}_a^{-1} \hat{\theta}^{\text{dir}} = \begin{pmatrix} \hat{\theta}_{AB}^{\text{net}} \\ \hat{\theta}_{AC}^{\text{net}} \end{pmatrix} = \begin{pmatrix} 0.0985 \\ 0.1773 \end{pmatrix}$$

Equivalently, estimates of the network treatment effects can be calculated over the whole network. In this case, for each design c, the Y_{jc} are stacked to form a column vector, Y, of the direct treatment effect estimates. Similarly the $\hat{V}_{jc}$ form the blocks in a block-diagonal matrix, $\hat{V}$, representing all the trials. The design matrix X for the whole network is formed by repeating the first row of the compressed design matrix X_a p times, where p is the number of trials j in design $c = 1$. The second row of the compressed design matrix is repeated q times, where q is the number of trials j in design $c = 2$. This process is repeated until there is a row in the design matrix corresponding to each trial in the network. Estimates of the network treatment effects can then be calculated as:

$$\hat{\theta}^{\text{net}} = \left(X' \hat{V}^{-1} X \right)^{-1} X' \hat{V}^{-1} Y \tag{5.2}$$

In the lung cancer network the variance matrix $\hat{V}$ for all trials, is a matrix with dimension 42 x 42, Y is a column vector of length 42 and X is a matrix with dimension 42 x 2 in which each row corresponds to one trial and each column corresponds to one of the basic parameters.

Having obtained the NMA point estimates, the between-designs Q statistic for the network is calculated as:

$$Q^{\text{inc}} = \left(\hat{\theta}^{\text{dir}} - X_a \hat{\theta}^{\text{net}} \right)' \hat{V}_a^{-1} \left(\hat{\theta}^{\text{dir}} - X_a \hat{\theta}^{\text{net}} \right), \tag{5.3}$$

with $\hat{\theta}^{\text{dir}}$, $\hat{\theta}^{\text{net}}$, $\hat{V}_a$ and X_a as defined before. In the lung cancer network $Q^{\text{inc}} = 4.519$ on 1df so p=0.0335.

An alternative approach to calculating the total inconsistency for the whole network involves first calculating the inconsistency for each design. To do this, let $c = 1, \ldots, D$ where $D = $ total number of designs in the network. Later I use d to denote the detached design. The inconsistency across all trials of design c can be calculated as:

$$Q_c^{\text{inc}} = \left(\hat{\theta}_c^{\text{dir}} - X_c \hat{\theta}^{\text{net}} \right)' \hat{V}_c^{-1} \left(\hat{\theta}_c^{\text{dir}} - X_c \hat{\theta}^{\text{net}} \right)$$

where $\hat{\theta}_c^{\text{dir}}$ is the value of $\hat{\theta}^{\text{dir}}$ corresponding to design c, X_c is the design matrix for design c, $\hat{V}_c$ is the value of $\hat{V}_a$ corresponding to design c and $\hat{\theta}^{\text{net}}$ is as defined above. Then, the total inconsistency for the whole network is equal to the sum of the inconsistency for each design c:

$$Q^{\text{inc}} = \sum_{c=1}^{D} Q_c^{\text{inc}}$$

Returning to the lung cancer network the components of Q^{inc} for each design c are:

$$\begin{pmatrix} Q_1^{inc} \\ Q_2^{inc} \\ Q_3^{inc} \end{pmatrix} = \begin{pmatrix} 0.8687 \\ 1.0359 \\ 2.6147 \end{pmatrix}$$

The net heat approach identifies inconsistency in a network by detaching each design, one at a time, and recalculating the Q statistic. Let d represent the design that is detached, then $Q_{c(d)}^{inc}$ is the amount of inconsistency across the network for design c when design d is detached. The difference between this and Q_c^{inc}, the total amount of inconsistency for design c (i.e. without detaching design d), is given by $Q_{c,d}^{diff}$. This difference identifies if a particular design (or designs) is responsible for increasing or decreasing the inconsistency across the network.

To detach a design, an additional column is added to the compressed design matrix X_a. The number of columns added is equal to the number of treatments in design d minus 1. This column consists of '1' in the position for the design which is being detached and '0' elsewhere. In a network of two-arm trials only, one column is added for each design. This new design matrix is denoted by $X_{b,d}$. This is equivalent to adding a treatment-design interaction for design d to the NMA model. The network treatment effects with design d detached are calculated in the same way as above with X_a replaced by $X_{b,d}$ so that:

$$\hat{\theta}_d^{net} = (X_{b,d}\hat{V}_a^{-1}X_{b,d})^{-1}X_{b,d}'\hat{V}_a^{-1}\hat{\theta}^{dir}$$

In the lung cancer network there are two treatments in each design so when a design is detached one column will be added to the design matrix. In a three-treatment network, detaching a design is equivalent to fitting an inconsistency parameter. When the AB treatment comparison is detached the design matrix, $X_{b,d}$, is:

$$X_{b,1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 1 & 0 \end{pmatrix}$$

When the AC treatment comparison is detached the design matrix, $X_{b,d}$, is:

$$X_{b,2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ -1 & 1 & 0 \end{pmatrix}$$

When the BC treatment comparison is detached the design matrix, $X_{b,d}$, is:

$$X_{b,3} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1 & 1 & 1 \end{pmatrix}$$

The residuals, which are equivalent to a consistency model fitted when design d is detached, are then calculated as:

$$R_d = \hat{\theta}^{\text{dir}} - X_{b,d}\hat{\theta}_d^{\text{net}}$$

For each design d in the lung cancer network, the result is the zero column vector of length three since the model is saturated.

$$R_d = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

A three-treatment network with all two-arm trials is a special case and in networks with more treatment loops this column vector would not consist solely of zeros. However you always obtain a zero in the position of the detached design. This is because the consistency assumption is relaxed for design d by adding the treatment-by-design interaction which means that $X_{b,d}\hat{\theta}_d^{\text{net}} = \hat{\theta}_d^{\text{dir}}$ for design d.

The residuals, R_d, are used in the calculation of $Q_{c(d)}^{\text{inc}}$, which represents the reduction in inconsistency for design c due to the detachment of design d. To calculate $Q_{c(d)}^{\text{inc}}$, the matrix of residuals is squared and multiplied by the corresponding inverse variance entry of $\hat{V}_a$ for design d.

$$Q_{c,d}^{\text{inc}} = \frac{R_d^2}{\hat{V}_a}$$

In the lung cancer network, where the residuals are all equal to zero, this results in a matrix of zeros for $Q_{c(d)}^{\text{inc}}$.

Finally, inconsistency in the network is located by comparing the inconsistency for design c after detachment of design d to the inconsistency for design c before detachment of any designs i.e.

$$Q_{c,d}^{\text{diff}} = Q_c^{\text{inc}} - Q_{c(d)}^{\text{inc}}$$

130

In other words, $Q_{c,d}^{\text{diff}}$ are the residuals resulting from fitting a consistency model which relaxes the consistency assumption for design d through the inclusion of a treatment-design interaction for design d. Where $Q_{c,d}^{\text{diff}} < 0$ the net heat plot is coloured blue and where $Q_{c,d}^{\text{diff}} > 0$ the net heat plot is coloured red with the greatest intensity of colour at absolute values of $Q_{c,d}^{\text{diff}} = 8$.

For the lung cancer network, the resulting Q^{diff} matrix is:

$$Q^{\text{diff}} = Q^{\text{inc}} - Q^{\text{inc}} = \begin{pmatrix} 0.8687 & 0.8687 & 0.8687 \\ 1.0359 & 1.0359 & 1.0359 \\ 2.6147 & 2.6147 & 2.6147 \end{pmatrix} - \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0.8687 & 0.8687 & 0.8687 \\ 1.0359 & 1.0359 & 1.0359 \\ 2.6147 & 2.6147 & 2.6147 \end{pmatrix}$$

The values of this matrix are responsible for the colours displayed in Figure 5.2.

5.5 A closer look at the net heat approach

As a form of regression I would expect any diagnostic useful in the NMA case to be meaningful in simpler cases. Therefore, in order to try and understand the findings from the net heat plot for the lung cancer network, I now look in more detail at the calculation underlying the net heat plot. As explained before (Section 5.4), the net heat plot is a version of a 'leave-one-out' approach, where for design c we calculate $Q_{c,d}^{\text{diff}}$ for each $d \neq c$. Here, I calculate explicitly what this quantity is. I begin with the simpler setting of just one design, which with IPD is essentially a one-way ANOVA model. I then return to the network case and consider three, four and five treatment networks before generalising the result. This section finishes by illustrating the results using data from the lung cancer network. Throughout this section I assume FTE.

5.5.1 One-way ANOVA: one design with IPD

As discussed above, the net heat approach is based on the decomposition of Cochran's Q statistic which can be used to measure the overall heterogeneity within a network of tri-

als. In this section, the argument I will apply at the network-design level is illustrated at the design-trial level. In the net heat setting we have:

$$Q = Q^{\text{het}} + Q^{\text{inc}}$$

where Q^{het} is a sum of within-design Q statistics (i.e. heterogeneity within each design) and Q^{inc} is a between-design Q statistic (i.e. heterogeneity between designs). Q is the sum of the within design and between design Q statistics, and reflects the total heterogeneity of the network. This is a generalisation of the one-way analysis of variance (ANOVA) which partitions the total variability (total sum of squares) within a dataset into that due to systematic differences between groups and variation between individuals within each group (residual sum of squares).

Consider a series of n clinical trials with common two-arm design, observed treatment effects $\hat{y}_1, \ldots, \hat{y}_n$, and common known standard error s. Thus $\hat{y}_c$ is the observed treatment effect from trial c where $c = 1, \ldots, n$. Let Q_c^{inc} measure the inconsistency between $\hat{y}_c$ and the mean, $\bar{y}$, over all n trials and let $Q_{c(d)}^{\text{inc}}$ measure the inconsistency between $\hat{y}_c$ and the mean, $\bar{y}_{(d)}$ over the remaining $n-1$ trials after excluding trial d. More specifically, define the following expressions:

$$Q_c^{\text{inc}} = \frac{(\hat{y}_c - \bar{y})^2}{s^2}, \tag{5.4}$$

$$Q_{c(d)}^{\text{inc}} = \frac{(\hat{y}_c - \bar{y}_{(d)})^2}{s^2}, \tag{5.5}$$

where $\bar{y}_{(d)}$ is defined as the mean of $\hat{y}_1, \ldots, \hat{y}_n$ excluding d, so

$$\bar{y}_{(d)} = \frac{\sum_c y_c - y_d}{n-1}.$$

Inconsistency across all trials can be calculated by adding together Q^{inc} for all n trials, so $Q^{\text{inc}} = \sum_{c=1}^n Q_c^{\text{inc}}$.

To simplify the calculation below, δ_c is defined as the residual from the grand mean $\bar{y}$ for a trial c:

$$\delta_c = \hat{y}_c - \bar{y}.$$

Then I can re-write $\bar{y}_{(d)}$ as follows:

$$\bar{y}_{(d)} = \frac{\sum_c y_c - y_d}{n-1} = \frac{n\bar{y} - y_d}{n-1} = \frac{\bar{y} + (n-1)\bar{y} - y_d}{n-1}$$

$$= \frac{\bar{y}}{n-1} + \frac{(n-1)\bar{y}}{n-1} - \frac{y_d}{n-1} = \frac{\bar{y} - y_d}{n-1} + \bar{y} = \bar{y} - \frac{\delta_d}{n-1}.$$

Now I can re-write (5.5) as:

$$Q_{c(d)}^{\text{inc}} = \frac{(y_c - \bar{y} + \frac{\delta_d}{n-1})^2}{s^2} = \frac{(\delta_c + \frac{\delta_d}{n-1})^2}{s^2}.$$

(5.6)

In this setting, the argument behind the net heat plot tells us to calculate the change in inconsistency when one trial d is excluded from the MA of all trials $c = 1, \ldots, n$. Therefore, in this setting $Q_{c,d}^{\text{diff}}$ corresponds to

$$Q_{c,d}^{\text{diff}} = Q_c^{\text{inc}} - Q_{c(d)}^{\text{inc}},$$

(5.7)

where Q_c^{inc} and $Q_{c(d)}^{\text{inc}}$ are defined by (5.4) and (5.6), respectively.

$Q_{c,d}^{\text{diff}}$ can therefore be written as:

$$Q_{c,d}^{\text{diff}} = \frac{(y_c - \bar{y})^2}{s^2} - \frac{(\delta_c + \frac{\delta_d}{n-1})^2}{s^2} = \frac{1}{s^2}\left\{\delta_c^2 - \left(\delta_c + \frac{\delta_d}{n-1}\right)^2\right\}$$

$$= \frac{1}{s^2}\left(\frac{-2\delta_c\delta_d}{n-1} - \frac{\delta_d^2}{(n-1)^2}\right) = \frac{-\delta_d}{s^2(n-1)}\left(2\delta_c + \frac{\delta_d}{n-1}\right).$$

Consider the interpretation of $Q_{c,d}^{\text{diff}}$ as a diagnostic. First, notice that it will tend to be small, even if δ_d is relatively large, when δ_c is small. It can also be small if δ_c and δ_d are large, if $\delta_c = \frac{-\delta_d}{2(n-1)}$. Furthermore, it has no readily computable sampling distribution under the null hypothesis of no inconsistency, to allow evaluation of whether the values are larger than we would expect if the null hypothesis is true. Neither is $Q_{c,d}^{\text{diff}}$ a simple residual, showing whether $\hat{y}_d$ is consistent with the underlying model or not. In other words, application of the principle behind the net heat plot to the one-way ANOVA setting leads to a $Q_{c,d}^{\text{diff}}$ which is both unlike any diagnostic that has been proposed for ANOVA, and does not have a clear interpretation. I will now use the insights obtained from this relatively simple case and consider further the interpretation of the net heat plot in increasingly complex networks.

5.5.2 Three treatment network

I now consider a three-treatment network, consisting of two-arm trials only, as pictured in Figure 5.3, page 126. I assume the same number of trials, and the same number of patients per trial, in each comparison and a common variance estimate of s^2. I assume an equal weight of $\frac{1}{s^2}$ for each of the direct comparisons in the network so each indirect comparison has weight $\frac{1}{2s^2}$. In this network I am interested in the direct and network evidence for c which represents the design comparing treatments B and C. The aim

here is to look at what happens to the inconsistency for design c when I detach design d. There are two possible scenarios: $c \neq d$ and $c = d$.

Considering Figure 5.3, there is only one pathway of indirect evidence for the comparison BC. This pathway goes via treatment A and I denote this indirect treatment effect by $\hat{\theta}_c^{\mathsf{ind}(1)}$. Applying these definitions to the network displayed in Figure 5.3:

$$\hat{\theta}_c^{\mathsf{ind}} = \hat{\theta}_{AC}^{\mathsf{dir}} - \hat{\theta}_{AB}^{\mathsf{dir}} = \hat{\theta}_c^{\mathsf{ind}(1)}, \text{ with variance } \mathsf{Var}(\hat{\theta}_{AC}^{\mathsf{dir}}) + \mathsf{Var}(\hat{\theta}_{AB}^{\mathsf{dir}}) = 2s^2.$$

First consider the scenario where $c \neq d$. The network estimate of c is equal to the inverse variance weighted average of all the direct and indirect evidence combined:

$$\hat{\theta}_c^{\mathsf{net}} = \frac{\left(\hat{\theta}_c^{\mathsf{dir}} s^2 + \frac{1}{2} \hat{\theta}_c^{\mathsf{ind}(1)} s^2 \right)}{\frac{3}{2} s^2} = \frac{1}{3} \left(2 \hat{\theta}_c^{\mathsf{dir}} + \hat{\theta}_c^{\mathsf{ind}(1)} \right) = \frac{2}{3} \hat{\theta}_c^{\mathsf{dir}} + \frac{1}{3} \hat{\theta}_c^{\mathsf{ind}(1)}$$

In a three-treatment network, when $c \neq d$, there is only one pathway of indirect evidence which must include design d. Therefore, considering Figure 5.3, the network estimate of the BC treatment effect when the AB or AC designs are detached is:

$$\hat{\theta}_{c(d)}^{\mathsf{net}} = \hat{\theta}_c^{\mathsf{dir}}$$

The inconsistency Q statistics, (5.4) and (5.6), are then defined in the network case as:

$$Q_c^{\mathsf{inc}} = \frac{1}{s^2} \left(\hat{\theta}_c^{\mathsf{dir}} - \hat{\theta}_c^{\mathsf{net}} \right)^2 \tag{5.8}$$

$$Q_{c(d)}^{\mathsf{inc}} = \frac{1}{s^2} \left(\hat{\theta}_c^{\mathsf{dir}} - \hat{\theta}_{c(d)}^{\mathsf{net}} \right)^2 \tag{5.9}$$

As before, Q_c^{inc} represents the difference between the direct and network evidence for design c across the whole network and $Q_{c(d)}^{\mathsf{inc}}$ represents the difference between the direct and network evidence for design c when design d is detached.

In the NMA setting, $Q_{c,d}^{\mathsf{diff}}$, represents the change in inconsistency for design c when design d is excluded from the network so that:

$$Q_{c,d}^{\mathsf{diff}} = Q_c^{\mathsf{inc}} - Q_{c(d)}^{\mathsf{inc}} \tag{5.10}$$

In a three-treatment network with only one pathway of indirect evidence and $\hat{\theta}_{c(d)}^{\mathsf{net}} = \hat{\theta}_c^{\mathsf{dir}}$ then

$Q_{c(d)}^{\text{inc}} = 0$. Therefore (5.10) can be re-written as:

$$Q_{c,d}^{\text{diff}} = Q_c^{\text{inc}} - Q_{c(d)}^{\text{inc}}$$

$$= \frac{1}{s^2}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{net}}\right)^2 - 0$$

$$= \frac{1}{s^2}\left(\hat{\theta}_c^{\text{dir}} - \frac{2}{3}\hat{\theta}_c^{\text{dir}} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)}\right)^2$$

$$= \frac{1}{s^2}\left(\frac{1}{3}\hat{\theta}_c^{\text{dir}} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)}\right)^2$$

$$= \frac{1}{s^2}\cdot\frac{1}{9}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}(1)}\right)^2 \tag{5.11}$$

Now consider the second scenario where $c = d$. The network estimate of c is the same as the first scenario:

$$\hat{\theta}_c^{\text{net}} = \frac{1}{3}\left(2\hat{\theta}_c^{\text{dir}} + \hat{\theta}_c^{\text{ind}(1)}\right) = \frac{2}{3}\hat{\theta}_c^{\text{dir}} + \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)}$$

When the direct evidence for design c is excluded the network estimate for design c is equal to the indirect evidence for design c:

$$\hat{\theta}_{c(c)}^{\text{net}} = \hat{\theta}_c^{\text{ind}(1)}$$

Therefore $Q_{c,c}^{\text{diff}}$ is calculated as:

$$Q_{c,c}^{\text{diff}} = Q_c^{\text{inc}} - Q_{c(c)}^{\text{inc}} = \frac{1}{s^2}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{net}}\right)^2 - \frac{1}{s^2}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_{c(c)}^{\text{net}}\right)^2$$

$$= \frac{1}{s^2}\left[\left(\hat{\theta}_c^{\text{dir}} - \frac{2}{3}\hat{\theta}_c^{\text{dir}} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}(1)}\right)^2\right]$$

$$= \frac{1}{s^2}\left[\left(\frac{1}{3}\hat{\theta}_c^{\text{dir}} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}(1)}\right)^2\right]$$

$$= \frac{1}{s^2}\left[\frac{1}{9}\left(\hat{\theta}_c^{\text{dir}}\right)^2 - \frac{2}{9}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{9}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \left(\hat{\theta}_c^{\text{dir}}\right)^2 - \left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 + 2\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)}\right]$$

$$= \frac{1}{s^2}\left[\frac{-8}{9}\left(\hat{\theta}_c^{\text{dir}}\right)^2 - \frac{8}{9}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 + \frac{16}{9}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)}\right]$$

$$= \frac{-1}{s^2}\cdot\frac{8}{9}\left[\left(\hat{\theta}_c^{\text{dir}}\right)^2 + \left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 - 2\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)}\right]$$

$$= \frac{-1}{s^2}\cdot\frac{8}{9}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}(1)}\right)^2 \tag{5.12}$$

In both cases, $c \neq d$ and $c = d$, (5.11) and (5.12) are scaled and squared versions of the inconsistency parameter. This shows that the net heat approach makes some sense
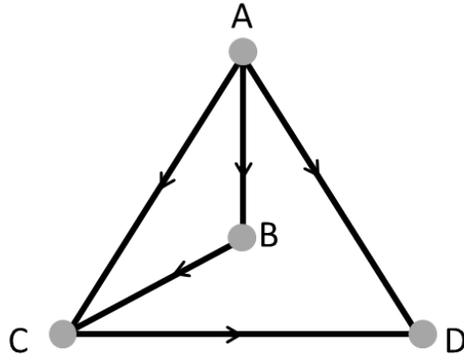
Figure 5.4: Example of a four treatment network. A, B, C, D represent the treatments being compared.

in this setting however these scaled versions of the inconsistency parameter do not have easy to calculate distributions so it is unclear why they would be preferred to (5.1). The net heat plot for a three-treatment network is an unintuitive representation of the inconsistency, the extent of which can be assessed and tested using (5.1).

5.5.3 Four treatment network

In this section I consider a four-treatment network, consisting of two-arm trials only, such as that pictured in Figure 5.4. I assume the same number of patients per trial and the same number of trials in each comparison and a common variance estimate of s^2. I assume an equal weight of $\frac{1}{s^2}$ for each of the direct comparisons in the network so each indirect comparison has weight $\frac{1}{2s^2}$. As before, the aim here is to look at what happens to the inconsistency for design c when I detach design d. I let c represent the set of trials comparing treatments A and C so that there are two possible scenarios to consider: $c \neq d$ and $c = d$.

Considering Figure 5.4 there are two pathways of indirect evidence from treatment A to treatment C via either treatment B or treatment D. I denote the indirect treatment effect estimate from the first pathway (via treatment B) as $\hat{\theta}_c^{\text{ind}(1)}$ and the indirect treatment effect estimate from the second pathway (via treatment D) as $\hat{\theta}_c^{\text{ind}(2)}$. I define $\hat{\theta}_c^{\text{ind}}$ as the weighted average of all the indirect evidence for design c, $\hat{\theta}_c^{\text{net}}$ as the weighted average of all the direct and indirect evidence for design c and $\hat{\theta}_{c(d)}^{\text{net}}$ as the weighted average of all the direct and indirect evidence for design c when design d is excluded. Applying these

definitions to the network displayed in Figure 5.4 and assuming $c \neq d$:

$$\hat{\theta}_c^{\text{ind}(1)} = \hat{\theta}_{AB}^{\text{dir}} + \hat{\theta}_{BC}^{\text{dir}}$$

$$\hat{\theta}_c^{\text{ind}(2)} = \hat{\theta}_{AD}^{\text{dir}} - \hat{\theta}_{CD}^{\text{dir}}$$

$$\hat{\theta}_c^{\text{ind}} = \hat{\theta}_c^{\text{ind}(1)} + \hat{\theta}_c^{\text{ind}(2)}$$

$$\hat{\theta}_c^{\text{net}} = \frac{1}{4}\left(\hat{\theta}_c^{\text{dir}} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)}\right) = \frac{1}{4}\left(2\hat{\theta}_c^{\text{dir}} + \hat{\theta}_c^{\text{ind}(1)} + \hat{\theta}_c^{\text{ind}(2)}\right)$$

$$\hat{\theta}_{c(d)}^{\text{net}} = \frac{1}{3}\left(2\hat{\theta}_c^{\text{dir}} + \hat{\theta}_c^{\text{ind}(2)}\right) = \frac{2}{3}\hat{\theta}_c^{\text{dir}} + \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)}$$

Another quantity needed in the calculation below is $\hat{\theta}_{c(d/2)}^{\text{net}}$ which is the average of all the network evidence for design c and the network evidence for design c that remains when design d is excluded:

$$\hat{\theta}_{c(d/2)}^{\text{net}} = \frac{1}{2}\left(\hat{\theta}_c^{\text{net}} + \hat{\theta}_{c(d)}^{\text{net}}\right) = \frac{1}{2}\left(\frac{1}{2}\hat{\theta}_c^{\text{dir}} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(2)} + \frac{2}{3}\hat{\theta}_c^{\text{dir}} + \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)}\right)$$

$$= \frac{7}{12}\hat{\theta}_c^{\text{dir}} + \frac{1}{8}\hat{\theta}_c^{\text{ind}(1)} + \frac{7}{24}\hat{\theta}_c^{\text{ind}(2)}$$

$$= \frac{1}{2}\left[\frac{7}{6}\hat{\theta}_c^{\text{dir}} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)} + \frac{7}{12}\hat{\theta}_c^{\text{ind}(2)}\right] \tag{5.13}$$

In the four-treatment network, and more generally (5.10) can be re-written using (5.8) and (5.9) as:

$$Q_{c,d}^{\text{diff}} = Q_c^{\text{inc}} - Q_{c(d)}^{\text{inc}} = \frac{1}{s^2}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{net}}\right)^2 - \frac{1}{s^2}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_{c(d)}^{\text{net}}\right)^2$$

$$= \frac{1}{s^2}\left[\left(\hat{\theta}_c^{\text{net}}\right)^2 - \left(\hat{\theta}_{c(d)}^{\text{net}}\right)^2 - 2\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{net}} + 2\hat{\theta}_c^{\text{dir}}\hat{\theta}_{c(d)}^{\text{net}}\right]$$

$$= \frac{1}{s^2}\left[\left(\hat{\theta}_c^{\text{net}} + \hat{\theta}_{c(d)}^{\text{net}}\right)\left(\hat{\theta}_c^{\text{net}} - \hat{\theta}_{c(d)}^{\text{net}}\right) + 2\hat{\theta}_c^{\text{dir}}\left(\hat{\theta}_{c(d)}^{\text{net}} - \hat{\theta}_c^{\text{net}}\right)\right]$$

$$= \frac{1}{s^2}\left[\left(\hat{\theta}_{c(d)}^{\text{net}} - \hat{\theta}_c^{\text{net}}\right)\left\{2\hat{\theta}_c^{\text{dir}} - \left(\hat{\theta}_{c(d)}^{\text{net}} + \hat{\theta}_c^{\text{net}}\right)\right\}\right]$$

$$= \frac{2}{s^2}\left(\hat{\theta}_{c(d)}^{\text{net}} - \hat{\theta}_c^{\text{net}}\right)\left[\hat{\theta}_c^{\text{dir}} - \frac{1}{2}\left(\hat{\theta}_{c(d)}^{\text{net}} + \hat{\theta}_c^{\text{net}}\right)\right] \tag{5.14}$$

Also:

$$\hat{\theta}_{c(d)}^{\text{net}} - \hat{\theta}_c^{\text{net}} = \frac{2}{3}\hat{\theta}_c^{\text{dir}} + \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{2}\hat{\theta}_c^{\text{dir}} - \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{4}\hat{\theta}_c^{\text{ind}(2)}$$

$$= \frac{1}{6}\hat{\theta}_c^{\text{dir}} - \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{12}\hat{\theta}_c^{\text{ind}(2)} \tag{5.15}$$

Therefore, (5.14) can be re-written using (5.13) and (5.15) as:

$$
\begin{aligned}
Q_{c,d}^{\text{diff}} &= \frac{2}{s^2}\left(\hat{\theta}_{c(d)}^{\text{net}} - \hat{\theta}_c^{\text{net}}\right)\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_{c(d/2)}^{\text{net}}\right) \\
&= \frac{2}{s^2}\left(\frac{1}{6}\hat{\theta}_c^{\text{dir}} - \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{12}\hat{\theta}_c^{\text{ind}(2)}\right)\left(\hat{\theta}_c^{\text{dir}} - \frac{1}{2}\left[\frac{7}{6}\hat{\theta}_c^{\text{dir}} + \frac{7}{12}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)}\right]\right) \\
&= \frac{2}{s^2}\left(\frac{1}{6}\hat{\theta}_c^{\text{dir}} - \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{12}\hat{\theta}_c^{\text{ind}(2)}\right)\left(\frac{5}{12}\hat{\theta}_c^{\text{dir}} - \frac{7}{24}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{8}\hat{\theta}_c^{\text{ind}(1)}\right) \\
&= \frac{1}{s^2}\left[\frac{5}{36}\left(\hat{\theta}_c^{\text{dir}}\right)^2 - \frac{7}{144}\left(\hat{\theta}_c^{\text{ind}(2)}\right)^2 + \frac{1}{16}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \frac{1}{36}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{4}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)}\right. \\
&\qquad\left. + \frac{1}{8}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(2)}\right] \\
&= \frac{1}{4s^2}\left[\frac{5}{9}\left(\hat{\theta}_c^{\text{dir}}\right)^2 - \frac{7}{36}\left(\hat{\theta}_c^{\text{ind}(2)}\right)^2 + \frac{1}{4}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \frac{1}{9}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(2)} - \hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)}\right. \\
&\qquad\left. + \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(2)}\right] \\
&= \frac{1}{4s^2}\left[\frac{2}{3}\hat{\theta}_c^{\text{dir}} + \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)} - \hat{\theta}_c^{\text{ind}(1)}\right]\left[\frac{5}{6}\hat{\theta}_c^{\text{dir}} - \frac{7}{12}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)}\right] \\
&= \frac{1}{4s^2}\left[\frac{1}{3}(2\hat{\theta}_c^{\text{dir}} + \hat{\theta}_c^{\text{ind}(2)}) - \hat{\theta}_c^{\text{ind}(1)}\right]\left[2\hat{\theta}_c^{\text{dir}}\left(1 - \frac{7}{12}\right) - \frac{1}{4}\left(\frac{7}{3}\hat{\theta}_c^{\text{ind}(2)} + \hat{\theta}_c^{\text{ind}(1)}\right)\right]
\end{aligned}
$$

In the second scenario where $c = d$ I have:

$$
\hat{\theta}_c^{\text{ind}(1)} = \hat{\theta}_{AB}^{\text{dir}} + \hat{\theta}_{BC}^{\text{dir}}
$$

$$
\hat{\theta}_c^{\text{ind}(2)} = \hat{\theta}_{AD}^{\text{dir}} - \hat{\theta}_{CD}^{\text{dir}}
$$

$$
\hat{\theta}_c^{\text{ind}} = \hat{\theta}_c^{\text{ind}(1)} + \hat{\theta}_c^{\text{ind}(2)}
$$

$$
\hat{\theta}_c^{\text{net}} = \frac{1}{4}\left(\hat{\theta}_c^{\text{dir}} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)}\right) = \frac{1}{2}\hat{\theta}_c^{\text{dir}} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(2)}
$$

$$
\hat{\theta}_{c(c)}^{\text{net}} = \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)}.
$$

Therefore $Q_{c,c}^{\text{diff}}$ can be written as:

$$
\begin{aligned}
Q_{c,c}^{\text{diff}} &= Q_c^{\text{inc}} - Q_{c,c}^{\text{inc}} = \frac{1}{s^2}\left[\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{net}}\right)^2 - \left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_{c(c)}^{\text{net}}\right)^2\right] \\
&= \frac{1}{s^2}\left[\left(\hat{\theta}_c^{\text{dir}} - \frac{1}{2}\hat{\theta}_c^{\text{dir}} - \frac{1}{4}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{4}\hat{\theta}_c^{\text{ind}(2)}\right)^2 - \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)}\right)^2\right] \\
&= \frac{1}{s^2}\left[\frac{-3}{4}\left(\hat{\theta}_c^{\text{dir}}\right)^2 - \frac{3}{16}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \frac{3}{16}\left(\hat{\theta}_c^{\text{ind}(2)}\right)^2 + \frac{3}{4}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)} + \frac{3}{4}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(2)}\right. \\
&\qquad\left. - \frac{3}{8}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(2)}\right] \\
&= \frac{-3}{4s^2}\left(\hat{\theta}_c^{\text{dir}} - \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)}\right)^2
\end{aligned}
$$

I defer discussion of this until after consideration of a more general network below.

5.5.4 A more general network

In Subsection 5.5.2 and Subsection 5.5.3 I considered three and four treatment networks, respectively. I followed the same process for a five-treatment network and the calculations are set out in full in Appendix F.1. When considered together — see Appendix F.2 — it can be seen that $Q_{c,d}^{\text{diff}}$ and $Q_{c,c}^{\text{diff}}$ for three, four and five treatment networks all have common formats and components. This led to a more general formula suggested by my co-supervisor David Fisher and confirmed by me, which I now present. The work in this section has not yet been published but forms part of the publication that will be submitted to Statistics in Medicine once approval for use of the lung cancer data is received. In this section the same principles as Subsection 5.5.2 and Subsection 5.5.3 are followed to develop a more general formula for $Q_{c,d}^{\text{diff}}$ and $Q_{c,c}^{\text{diff}}$ which can be applied to networks of any size, in which at least two of the treatments are both directly compared with other treatments, and assuming two-arm trials are considered only.

In this section I assume a network of two-arm trials in which the two treatments making up the design of interest c, e.g. A and B, are both linked directly with a number of other treatments (e.g. $X_1, X_2, \ldots, X_k$). I make the same assumption as before, each trial has the same number of patients and each comparison has the same number of trials. I assume an equal weight $\frac{1}{s^2}$ for each of the direct comparisons in the network so that each indirect comparison has weight $\frac{1}{2s^2}$. Then, as before, let c be the design of interest, with direct estimate $\hat{\theta}_c^{\text{dir}}$. Suppose there are k possible indirect pathways, each involving a single additional node. Each additional node adds one loop to the network. Therefore there are a total of $k + 2$ treatments relevant to design c. Denote the indirect estimates by $\hat{\theta}_c^{\text{ind}(i)}, i = 1, ..., k$. The network estimate of c is equal to the weighted average of all the direct and indirect evidence combined; that is:

$$\hat{\theta}_c^{\text{net}} = \frac{1}{k+2} \left\{ 2\hat{\theta}_c^{\text{dir}} + \sum_{i=1}^{k} \hat{\theta}_c^{\text{ind}(i)} \right\}. \tag{5.16}$$

To test the effect of detaching design d there are two scenarios: $c \neq d$ and $c = d$. Assume first that $c \neq d$; then without loss of generality let the effect size for design d be $\theta_c^{\text{ind}(d)}$. Then when design d is detached the remaining network evidence on c is:

$$\hat{\theta}_{c(d)}^{\text{net}} = \frac{1}{k+1} \left\{ 2\hat{\theta}_c^{\text{dir}} + \sum_{i,i \neq d} \hat{\theta}_c^{\text{ind}(i)} \right\}. \tag{5.17}$$

If instead the direct comparison, $c = d$, is detached the network evidence remaining for design c is:

$$\hat{\theta}_{c(c)}^{\text{net}} = \frac{1}{k} \sum_{i=1}^{k} \hat{\theta}_c^{\text{ind}(i)}.$$

When $c \neq d$, $\hat{\theta}_c^{\text{net}}$ can be re-written in terms of $\hat{\theta}_c^{\text{ind}(i)}$ as follows:

$$\hat{\theta}_c^{\text{net}} = \frac{1}{k+2} \left\{ 2\hat{\theta}_c^{\text{dir}} + \hat{\theta}_c^{\text{ind}(d)} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right\}.$$

The inconsistency Q statistics (Q_c^{inc}, $Q_{c(d)}^{\text{inc}}$, $Q_{c,d}^{\text{diff}}$) were defined for the network case in (5.8), (5.9) and (5.10) (Subsection 5.5.2). From (5.14) we see:

$$Q_{c,d}^{\text{diff}} = \frac{2}{s^2} \left(\hat{\theta}_{c(d)}^{\text{net}} - \hat{\theta}_c^{\text{net}} \right) \left[\hat{\theta}_c^{\text{dir}} - \frac{1}{2} \left(\hat{\theta}_{c(d)}^{\text{net}} + \hat{\theta}_c^{\text{net}} \right) \right]$$

As before, define $\hat{\theta}_{c(d/2)}^{\text{net}}$ as the average of all the network evidence for design c and the network evidence that remains for design c when design d is excluded so that:

$$\hat{\theta}_{c(d/2)}^{\text{net}} = \frac{1}{2} \left(\hat{\theta}_{c(d)}^{\text{net}} + \hat{\theta}_c^{\text{net}} \right).$$

Next, write the difference between the network evidence on c when d is excluded and the network evidence on c (i.e. (5.17) - (5.16)) in terms of $\hat{\theta}_c^{\text{ind}(i)}$, as follows:

$$\hat{\theta}_{c(d)}^{\text{net}} - \hat{\theta}_c^{\text{net}} = \left(\frac{1}{k+1} - \frac{1}{k+2} \right) \left\{ 2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right\} - \frac{1}{k+2} \hat{\theta}_c^{\text{ind}(d)}$$

$$= \frac{1}{k+2} \left\{ \frac{1}{k+1} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) - \hat{\theta}_c^{\text{ind}(d)} \right\}.$$

and similarly

$$\hat{\theta}_{c(d/2)}^{\text{net}} = \frac{1}{2} \left(\hat{\theta}_{c(d)}^{\text{net}} + \hat{\theta}_c^{\text{net}} \right)$$

$$= \frac{1}{2} \left(\frac{1}{k+1} + \frac{1}{k+2} \right) \left\{ 2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right\} + \frac{1}{2(k+2)} \hat{\theta}_c^{\text{ind}(d)}$$

$$= \frac{1}{2(k+2)} \left\{ \frac{2k+3}{k+1} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) + \hat{\theta}_c^{\text{ind}(d)} \right\}.$$

Finally, putting it all together:

$$Q_{c,d}^{\text{diff}} = \frac{1}{s^2} \cdot \frac{1}{k+2} \left\{ \frac{1}{k+1} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) - \hat{\theta}_c^{\text{ind}(d)} \right\}$$

$$\times \left[2\hat{\theta}_c^{\text{dir}} \left(1 - \frac{2k+3}{(k+1)(k+2)} \right) - \frac{1}{k+2} \left(\frac{2k+3}{k+1} \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} + \hat{\theta}_c^{\text{ind}(d)} \right) \right]. \quad (5.18)$$

Or, if the direct comparison is detached:

$$Q_{c,c}^{\text{diff}} = -\frac{1}{s^2} \cdot \frac{4(k+1)}{(k+2)^2} \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{k} \sum_{i=1}^{k} \hat{\theta}_c^{\text{ind}(i)} \right)^2.$$

Let $k = 1$, a 3-treatment network, then as in Subsection 5.5.2, the resulting $Q_{c,d}^{\text{diff}}$ and $Q_{c,c}^{\text{diff}}$ statistics for a three-treatment network are scaled and squared versions of the inconsistency parameter:

$$Q_{c,d}^{\text{diff}} = \frac{1}{s^2} \cdot \frac{1}{9} \left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}(1)} \right)^2$$

Or, if the direct comparison is detached:

$$Q_{c,c}^{\text{diff}} = -\frac{1}{s^2} \cdot \frac{8}{9} \left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}(1)} \right)^2$$

which match what was found in Subsection 5.5.2.

5.5.5 Interpretation of equation (5.18)

Suppose k is large so that $k + 1 \approx k$ then we can approximate (5.18) by:

$$Q_{c,d}^{\text{diff}} \approx \frac{1}{s^2} \cdot \frac{1}{k} \left\{ \frac{1}{k} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i \neq d} \hat{\theta}_c^{\text{ind}(i)} \right) - \hat{\theta}_c^{\text{ind}(d)} \right\}$$
$$\times \left[2\hat{\theta}_c^{\text{dir}} \left(1 - \frac{2}{k} \right) - \frac{1}{k} \left(2 \sum_{i,i \neq d} \hat{\theta}_c^{\text{ind}(i)} + \hat{\theta}_c^{\text{ind}(d)} \right) \right].$$

Let $1 - \frac{2}{k} \approx 1$, then:

$$Q_{c,d}^{\text{diff}} \approx \frac{1}{s^2} \left\{ \frac{1}{k} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i \neq d} \hat{\theta}_c^{\text{ind}(i)} \right) - \hat{\theta}_c^{\text{ind}(d)} \right\}$$
$$\times \left[\frac{1}{k} \left\{ 2\hat{\theta}_c^{\text{dir}} - \frac{2}{k} \sum_{i,i \neq d} \hat{\theta}_c^{\text{ind}(i)} - \frac{1}{k} \hat{\theta}_c^{\text{ind}(d)} \right\} \right]$$

Now $\hat{\theta}_c^{\text{ind}}$ is the average of all the indirect evidence across the whole network:

$$\hat{\theta}_c^{\text{ind}} = \frac{1}{k} \left(\sum_{i,i \neq d} \hat{\theta}_c^{\text{ind}(i)} + \hat{\theta}_c^{\text{ind}(d)} \right).$$

Then:

$$Q_{c,d}^{\text{diff}} \approx \frac{1}{s^2} \left\{ \frac{1}{k} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) - \hat{\theta}_c^{\text{ind}(d)} \right\}$$
$$\times \left[\frac{1}{k} \left\{ 2\hat{\theta}_c^{\text{dir}} - \frac{1}{k} \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} - \hat{\theta}_c^{\text{ind}} \right\} \right]$$
$$\approx \frac{1}{s^2} \left\{ \frac{1}{k} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) - \hat{\theta}_c^{\text{ind}(d)} \right\}$$
$$\times \left[\frac{1}{k} \left\{ \left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}} \right) + \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{k} \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) \right\} \right] \qquad (5.19)$$

Essentially, (5.19) is a scaled product of two terms. Let

$$A \equiv \frac{1}{s^2} \left\{ \frac{1}{k} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) - \hat{\theta}_c^{\text{ind}(d)} \right\}$$

$$B \equiv \left[\frac{1}{k} \left\{ \left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}} \right) + \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{k} \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) \right\} \right].$$

Term A is the difference between the approximate inverse variance network estimate of the effect for design c excluding design d, and the indirect evidence from design d. While the square of this is a plausible measure of the difference between the evidence coming from the loop including design d and the rest of the network (including the direct evidence), it is not specific to design d but to the loop including design d. Further it can be positive or negative, depending on whether the estimate from the loop including design d is above or below the estimate from the remaining network. The sign does not affect the consistency, yet the authors of the net heat plot argue that negative values of the statistic show more consistency (these are shown in the net heat plot by blue colours).

Term B is the sum of the difference between the direct and indirect evidence including the loop with design d and the difference between the direct and indirect evidence excluding the loop with design d. This multiplier can therefore be large if the direct and indirect evidence differ (regardless of the effect of the loop including design d) and will tend to be small if the indirect and direct evidence are similar (with little regard for the effect of the loop including design d). It is therefore a poor choice of multiplier for term A. Further this statistic does not have a readily describable sampling distribution; it is not normal, or chi-squared (as it can take negative values). Thus, despite its promising definition, the close inspection I have given to $Q_{c,d}^{\text{diff}}$ shows that it is an unintuitive, imprecise measure of the inconsistency due to design d.

5.5.6 Example based on lung cancer network

The formula for $Q_{c,d}^{\text{diff}}$ from a three-treatment network, (5.11), is a scaled and squared version of the inconsistency parameter. To explore what this actually means in this section I take the lung cancer network, from Section 5.2, and modify it by forcing each design to have a common variance and calculate $Q_{c,d}^{\text{diff}}$.

In Section 5.2 I calculated the observed treatment effect estimates for each design in the lung cancer network. In this section, I use a modified version of the lung cancer network data. The parameter estimates for each design from the NMA model without an inconsistency parameter were kept the same as in Section 5.2, but each design was fixed to have a common variance of $s^2 = 0.002$. To achieve this, the treatment effects and variances for each trial were systematically changed to ensure that the treatment effect for each design remained the same as in Section 5.2. I then applied the formulas from Subsection 5.5.2 to this modified network.

Let A = RT, B = Seq CT and C = Con CT then after fitting the one-step IPD Royston-Parmar NMA model, (4.5), to the modified lung cancer network the following direct treatment effects were obtained:

$$\theta_{\text{AB}}^{\text{dir}} = 0.1316$$

$$\theta_{\text{AC}}^{\text{dir}} = 0.1378$$

$$\theta_{\text{BC}}^{\text{dir}} = 0.1787$$

with common variance $s^2 = 0.002$, as desired.

Let c = AC be the comparison of interest then there is one pathway of indirect evidence for c in the network which is denoted by $\hat{\theta}_c^{\text{ind}(1)}$. The indirect evidence for design c, the network evidence for design c and the network evidence for design c when design d is detached can be calculated as follows:

$$\hat{\theta}_c^1 = \hat{\theta}_{AC}^{\text{ind}} = \hat{\theta}_{AB}^{\text{dir}} + \hat{\theta}_{BC}^{\text{dir}} = 0.1316 + 0.1787 = 0.3103$$

$$\hat{\theta}_c^{\text{net}} = \frac{2}{3}\hat{\theta}_{AC}^{\text{dir}} + \frac{1}{3}\hat{\theta}_c^1 = \left(\frac{2}{3} \times 0.1378\right) + \left(\frac{1}{3} \times 0.3103\right) = 0.1953$$

$$\hat{\theta}_{c(d)}^{\text{net}} = \hat{\theta}_{AC}^{\text{dir}} = 0.1378$$

The Q statistics, Q_c^{inc}, $Q_{c(d)}^{\text{inc}}$ and $Q_{c,d}^{\text{diff}}$, can then be calculated from (5.8), (5.9) and (5.10), as follows:

$$Q_c^{\text{inc}} = \frac{1}{s^2}(\hat{\theta}_{AC}^{\text{dir}} - \hat{\theta}_c^{\text{net}})^2 = \frac{1}{0.002}(0.1378 - 0.1953)^2 = 1.65$$

$$Q_{c(d)}^{\text{inc}} = \frac{1}{s^2}(\hat{\theta}_{AC}^{\text{dir}} - \hat{\theta}_{c(d)}^{\text{net}})^2 = \frac{1}{0.002}(0.1378 - 0.1378)^2 = 0$$

$$Q_{c,d}^{\text{diff}} = Q_c^{\text{inc}} - Q_{c(d)}^{\text{inc}} = 1.65$$

which gives the same result as (5.11).

A single loop, such as a three treatment network, with common variance of on all designs is a special case so that when repeated for all possible combinations of c and d the resulting $Q_{c,d}^{\text{diff}}$ matrix consists of just one value. For the modified lung cancer network this results in 1.65 in all squares. This would result in a pale yellow colour on the net heat plot leading to the conclusion that there is no meaningful inconsistency in the modified lung cancer network.

Q^{inc} can be calculated for this modified lung cancer network by first calculating $\hat{\theta}^{\text{net}}$ using (5.2). This gives:

$$\hat{\theta}^{\text{net}} = \begin{pmatrix} 0.0741 \\ 0.1953 \end{pmatrix}$$

I set up the matrices for $\hat{\theta}^{\text{dir}}$, V_a and X_a as follows:

$$\hat{\theta}^{\text{dir}} = \begin{pmatrix} 0.1316 \\ 0.1378 \\ 0.1787 \end{pmatrix}$$

$$V_a = \begin{pmatrix} 0.002 & 0 & 0 \\ 0 & 0.002 & 0 \\ 0 & 0 & 0.002 \end{pmatrix}$$

$$X_a = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 1 \end{pmatrix}$$

Then using (5.3):

$$Q^{\text{inc}} = \left(\hat{\theta}^{\text{dir}} - X_a\hat{\theta}^{\text{net}}\right)' V_a^{-1} \left(\hat{\theta}^{\text{dir}} - X_a\hat{\theta}^{\text{net}}\right) = 4.966$$

In this example, $Q^{\text{inc}} = 4.966$ (p=0.026) suggests inconsistency in the network however the net heat plot suggests consistency and would be misleading. The problem is the scaling of $\frac{1}{9}$ in (5.11) and the corresponding misinterpretation of the resulting colours in the net heat plot. As far as I am aware, this is the first time this problem has been identified.

5.6 Developing an alternative method of assessing inconsistency

In Section 5.5 I showed that the net heat approach is an unintuitive measure of the difference between the direct and the indirect evidence which does not specifically target what I am interested in and may be misleading. As far as I am aware this problem with the net heat plot has not been identified previously. In this section I consider the development of an alternative method of assessing inconsistency in NMA.

Although the net heat plot can be misleading it does have some useful aspects which I think should, if possible, be incorporated into an alternative approach. These are the grey boxes and the colour scheme. The grey boxes display the proportion of information, for each design, that comes from the direct evidence and are useful for highlighting to what extent conclusions might be based on the indirect evidence. I also like the idea of using colours to highlight the designs in the network which could be causing any inconsistency. The three key things that I think an alternative approach should include are: an indicator of the direct evidence for each design, an indicator of any heterogeneity present in pairwise comparisons and a colour scale which relates to the extent of inconsistency between the direct evidence and the evidence from each indirect loop in the network.

Within a network, each treatment loop has the potential for inconsistency and therefore the NMA model can include an inconsistency parameter for each treatment loop. A network can contain multiple pathways between treatments and therefore inconsistency can be a linear combination of the inconsistency parameters. As an alternative approach I propose to set out, for each treatment comparison, all possible treatment loops in the network contributing to this comparison, and the extent of inconsistency within each of these loops. One of the big limitations to assessing inconsistency using a graphical approach is that for large networks graphs can become complicated, messy and hard to interpret. Therefore this alternative approach, based on the use of inconsistency parameters, takes the form of a table which — for each treatment contrast — sets out all the treatment loops within the

145

network and can be coloured to highlight statistically significant inconsistency within treatment loops and any heterogeneity within pairwise comparisons. I consider inconsistency parameters to be statistically significant if they produce p-values<0.05. The proportion of direct evidence coming from each treatment loop can be shown by a square box in each table cell, with shading indicating the proportion of the total evidence coming from that source.

There are four main steps to this alternative approach:

1. Set up a table listing all possible pairwise comparisons in the network and complete the table considering all possible treatment loops in the network

2. Fit a NMA model including an inconsistency parameter for each treatment loop in which inconsistency could occur

3. Colour the table based on the statistical significance of the p-values for the inconsistency parameters and heterogeneity tests.

4. Decide on the inconsistency parameters to retain in the NMA model (i.e. the sources of evidence to accept for each comparison)

I now explain these four steps in more detail through the use of an example to illustrate this alternative approach.

Example

To illustrate the idea behind the proposed tabular approach to assessing inconsistency in NMA, I consider an example based on the network diagram in Figure 5.5 which consists of the five treatments A, B, C, D and E. The solid lines indicate direct evidence between treatments. For simplicity I assume that all trials in this network are two-arm trials.
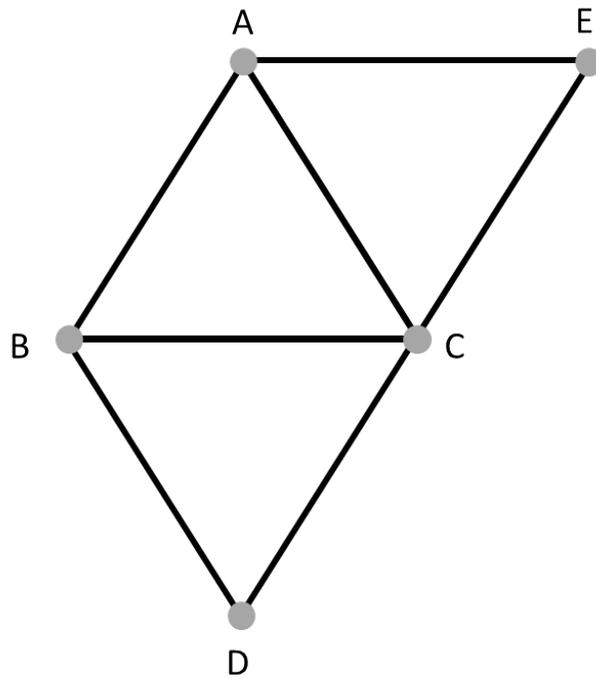
Figure 5.5: Example network for illustrating the alternative approach to assessing incon-
sistency. The network consists of five treatments: A, B, C, D and E. Solid lines indicate
direct evidence between treatments.

Table 5.1: Example of alternative approach to assessing inconsistency

Comparison	Direct evidence	Loop with 1 additional treatment	Loop with 2 additional treatments	Loop with 3 additional treatments
AB	✓	C	CD, EC	ECD
AC	✓	B, E	BD	×
AD	×	×	×	×
AE	✓	C	BC	BDC
BC	✓	A, D	AE	×
BD	✓	C	AC	AEC
BE	×	×	×	×
CD	✓	B	AB	EAB
CE	✓	A	BA	DBA
DE	×	×	×	×

The first step, setting up and completing the table, is illustrated in Table 5.1. In the first column of Table 5.1 I have listed all the possible pairwise treatment comparisons for the network. With five treatments there are ten possible pairwise comparisons. In the second column I use '✓' to indicate that there is direct evidence comparing the two treatments and '×' otherwise. The rest of the table considers loops of increasing size which include the two treatments of interest and other treatments. Throughout this section I will refer to the direct comparison of two treatments as a no-treatment loop, e.g. AB. I will refer to a direct comparison with one additional treatment as a one-treatment loop, e.g. ACB, a direct comparison with two additional treatments as a two-treatment loop, e.g. ACDB and so on.

In the third column I start by considering one-treatment loops (i.e. the direct comparison and one additional treatment). In this column I place the additional treatment which makes up the one-treatment loop or I place a '×' to indicate that there are no one-treatment loops in the network which include the direct comparison of interest. If there is more than one one-treatment loop I use a comma to separate the different treatment loops. In the fourth column I look for any two-treatment loops (i.e. the direct comparison and two additional treatments) and in the fifth column I look for any three-treatment loops (i.e. the direct comparison and three additional treatments). If treatment loops are present I add the additional treatments to the column and if there are no treatment loops I place a '×' in the column. In larger networks this process continues until treatment loops of all possible sizes have been considered.

Taking the AB comparison as a starting point. Figure 5.5 shows direct evidence between treatments A and B so I start by putting a '✓' in the 'direct evidence' column. A one-treatment loop can be formed through treatment C so I place 'C' in the 'loop with 1 additional treatment' column. There are two possible two-treatment loops: ABDC and ABCE so I add 'DC, CE' to the 'loop with 2 additional treatments' column. Finally there is one loop which features the AB comparison within a loop containing three additional treatments so I add 'DCE' to the final column.

Step two involves fitting a NMA model including an inconsistency parameter for each treatment loop in which it is possible for inconsistency to occur, such as that described in Subsection 4.4.3. Before fitting the NMA model in step two it is important to think about the parameterisation of the network and the inconsistency parameters. The smallest size loop in which inconsistency can occur is a three-treatment loop. Therefore I want the NMA

model to include an inconsistency parameter for each three-treatment loop. Considering the example network in Figure 5.5 I will add an inconsistency parameter for each of the three-treatment loops: ABC, BCD and ACE. In order to parameterise the network correctly it is important to know the direction of the treatment effects. I assume the treatment effects are in the direction from the lowest alphabetical treatment to the highest alphabetical treatment. In each three-treatment loop I need to choose one comparison for which the treatment effect will be calculated through the consistency equations (1.1) with the addition of an inconsistency parameter. For example, in the ABC loop, where θ_{AB} represents the treatment effect for treatment B compared to treatment A, I can calculate θ_{BC} as:

$$\theta_{BC} = \theta_{AC} - \theta_{AB} + I_{ABC}$$

where I_{ABC} represents the inconsistency parameter for the ABC loop.

Similarly, loop BCD can be parameterised as:

$$\theta_{CD} = \theta_{BD} - \theta_{BC} + I_{BCD}$$

where I_{BCD} represents the inconsistency parameter for the BCD loop.

Finally, loop ACE can be parameterised as:

$$\theta_{CE} = \theta_{AE} - \theta_{AC} + I_{ACE}$$

where I_{ACE} represents the inconsistency parameter for the ACE loop.

Step three requires the p-values from the inconsistency parameters to be used as the basis for colouring each cell in the columns based on the different sized treatment loops. The cells in the direct evidence column can be coloured based on the p-value for heterogeneity from either the I^2 statistic or equivalently Cochran's Q statistic, which should come from a MA of the direct evidence (rather than the network Q). For the colouring system I suggest using one colour, e.g. red, and allowing the intensity of the colour to vary depending on the statistical significance of the inconsistency parameter or heterogeneity test. For example, bright shades of red could indicate comparisons with statistically significant inconsistency or heterogeneity and act as a warning to the researcher that they should stop and consider how to account for inconsistency and/or heterogeneity before continuing with their NMA. Lighter shades of red, e.g. ranging from white for a p-value of one to pink for borderline statistical significance, could indicate non-statistically significant inconsistency or heterogeneity and suggest that researchers can proceed with their NMA.

Step four involves deciding on the inconsistency parameters that will remain in the NMA model. One approach could be to remove all inconsistency parameters that were not statistically significant in step two. For these treatment loops you are choosing to combine the direct and indirect evidence to draw inference from the network. In the treatment loops where inconsistency parameters remain only the direct evidence will be used for inference.

For the AB comparison there is one one-treatment loop which involves treatment C therefore the colour of the third column would be determined by the significance of I_{ABC}. The AB comparison can be part of two two-treatment loops (ACDB and AECB). To assess the inconsistency for the AB comparison from the two-treatment loops I have to consider the significance of the inconsistency parameters present in each loop, ACDB and AECB.

First, start by calculating the direct and indirect evidence for the AB comparison from the ACDB loop:

$$\theta_{AB}^{\text{dir}} = \theta_{AB}$$
$$\theta_{AB}^{\text{ind}} = \theta_{AC} + \theta_{CD} - \theta_{BD}$$
$$= \theta_{AC} + (\theta_{BD} - \theta_{BC} + I_{BCD}) - \theta_{BD}$$
$$= \theta_{AC} - (\theta_{AC} - \theta_{AB} + I_{ABC}) + I_{BCD}$$
$$= \theta_{AB} - I_{ABC} + I_{BCD}$$

The inconsistency for the ACDB loop can be calculated as:

$$\theta_{AB}^{\text{dir}} - \theta_{AB}^{\text{ind}} = \theta_{AB} - (\theta_{AB} - I_{ABC} + I_{BCD}) = I_{ABC} - I_{BCD}$$

Therefore the colour of the fourth column for the ACDB loop for the AB comparison will be determined by the significance of $I_{ABC} - I_{BCD}$.

Next, calculate the direct and indirect evidence for the AB comparison from the AECB loop:

$$\theta_{AB}^{\text{dir}} = \theta_{AB}$$
$$\theta_{AB}^{\text{ind}} = \theta_{AE} + \theta_{CE} - \theta_{BC}$$
$$= \theta_{AE} - (\theta_{AE} - \theta_{AC} + I_{ACE}) - \theta_{BC}$$
$$= \theta_{AC} - I_{ACE} - (\theta_{AC} - \theta_{AB} + I_{ABC})$$
$$= \theta_{AB} - I_{ABC} - I_{ACE}$$

The inconsistency for the AECB loop can be calculated as:

$$\theta_{AB}^{\text{dir}} - \theta_{AB}^{\text{ind}} = \theta_{AB} - (\theta_{AB} - I_{ABC} - I_{ACE}) = I_{ABC} + I_{ACE}$$

Therefore the colour of the fourth column for the AECB loop for the AB comparison will be determined by the significance of $I_{ABC} + I_{ACE}$.

The AB comparison is part of one three-treatment loop so there is only one route to consider:

$$\theta_{AB}^{\text{dir}} = \theta_{AB}$$

$$\theta_{AB} = \theta_{AE} - \theta_{CE} + \theta_{CD} - \theta_{BD}$$

$$= \theta_{AC} + \theta_{CE} - I_{ACE} - \theta_{CE} - \theta_{BC} + \theta_{BD} + I_{BCD} - \theta_{BD}$$

$$= \theta_{AC} - I_{ACE} - \theta_{AC} + \theta_{AB} - I_{ABC} + I_{BCD}$$

$$= \theta_{AB} - I_{ABC} - I_{ACE} + I_{BCD}$$

The inconsistency for the AECDB loop can be calculated as:

$$\theta_{AB}^{\text{dir}} - \theta_{AB}^{\text{ind}} = \theta_{AB} - (\theta_{AB} - I_{ABC} - I_{ACE} + I_{BCD}) = I_{ABC} + I_{ACE} - I_{BCD}$$

Therefore the colour of the fifth column for the AB comparison will be determined by the significance of $I_{ABC} + I_{ACE} - I_{BCD}$.

As the size of the network increases the complexity and the number of columns in the table will also increase however the colouring system and the detailed nature of where treatment loops exist means that the table should remain easy to interpret. However, there are several limitations to my proposed approach including multiple testing.

Multiple testing could be a problem with my proposed approach particularly in large, well connected networks. Methods for dealing with multiple testing, such as the Bonferroni correction, could be used to adjust the p-value used to determine statistically significant results. Another limitation of my approach is that the colouring system focuses on shading based on p-values. This means that a decision has to be made as to what constitutes a statistically significant p-value for evidence of inconsistency. A better approach might be to find a way of displaying the confidence interval around the estimate of the inconsistency parameter so that the uncertainty around the inconsistency parameter can be taken into account rather than basing decisions solely on p-values. I suggest that inconsistency parameters are removed based on statistical significance. Without any adjustment for multiple

testing a p-value less than 0.05 is generally considered to show statistically significant evidence of inconsistency. However, due to the low power of tests for inconsistency a p-value greater than 0.05 does not necessarily mean that a treatment loop is consistent.

Inconsistency parameters quantify the difference between the direct and the indirect evidence for a particular comparison within a treatment loop. Therefore, as described in Section 4.4.3, if inconsistency parameters are present in a treatment loop then this allows inference to be based on the direct evidence only. Inference in treatment loops without inconsistency parameters will be informed by both the direct and indirect evidence. By including inconsistency parameters in treatment loops exhibiting evidence of inconsistency I make the assumption that the direct evidence in the treatment loop is more reliable than the indirect evidence. As Song (2008) argues this is not always the case. For example, in a three-treatment loop where the direct evidence for one comparison is informed by only one trial considered to be of low methodological quality (i.e. exhibits evidence of breaking allocation concealment or inadequate randomisation) and the indirect evidence is informed by several trials of high quality then the indirect evidence is likely to be more reliable than the direct evidence.

5.7   Discussion

In this chapter, I have considered four of the most popular methods for assessing inconsistency in NMA and applied them to the lung cancer network. I have taken a closer look at the net heat approach (Krahn et al., 2013) and derived formulas for $Q_{c,d}^{\text{diff}}$ for a range of networks, to get a detailed understanding of what the quantity represents given the amount of inconsistency in the network. This shows that $Q_{c,d}^{\text{diff}}$ is an unintuitive and imprecise measure of inconsistency which could be misleading. In the special case of three-treatment networks, this is approximately a scaled version of the difference between the direct and the indirect evidence which explains why, in the lung cancer example, the net heat plot did not highlight the statistically significant inconsistency.

The net heat plot was developed as a visual aid for identifying and locating inconsistency in NMA. As well as assessing inconsistency the plot also visualises the hat matrix which highlights which direct comparisons contribute towards the network estimate for each treatment comparison. The net heat approach aims to identify a specific design (or designs)

that drive inconsistency in a network. However, locating inconsistency to a specific design (or even a pair of designs) is a difficult task, since inconsistency arises from comparisons between at least three designs. Thus any attempt to locate inconsistency within designs is potentially misleading, in particular because (as I have shown) it may tend to attribute inconsistency to areas with less evidence. Inconsistency does not make sense at the level of an individual design and it may be that locating inconsistency within a network depends on the structure of the network and that no method works for all networks.

The net heat approach uses Cochran's Q statistic (1954) in a FTE framework and decomposes it into within-trial and between-trial heterogeneity. This reflects the fact that heterogeneity and inconsistency can be considered as different aspects of heterogeneity where inconsistency is the discrepancy between results of single studies and predictions based on a consistency model (Krahn et al., 2013). The lung cancer example showed little evidence of heterogeneity and therefore it was appropriate to use FTE models, which assumed that there was no heterogeneity within designs, throughout this chapter. However, although more complex the calculations in Section 5.5 could be conducted using RTE models and this may be more appropriate when heterogeneity is present in a network.

Unlike Q, Q^{het} and Q^{inc} which follow chi-squared distributions, $Q^{\text{diff}}_{c,d}$ as the difference between two approximately chi-square distributed, correlated components, has a non-standard distribution and is therefore hard to interpret. Complex calculations would be required to calculate the sampling distribution and obtain a p-value. One possibility would be to use bootstrapping, but since $Q^{\text{diff}}_{c,d}$ does not have a natural interpretation I did not pursue this further. It is unclear then why these scaled versions of the inconsistency parameter, which do not have straight forward distributions, would be preferred to the inconsistency parameter.

Caution is advised when interpreting the net heat plot. If a component MA is identified as deviating or identified as a source of heterogeneity it may or may not provide the more reliable part of the whole body of evidence. Song (2008) argues that sometimes the indirect evidence can be more reliable than the direct evidence. Furthermore, of concern with the net heat approach is how the intensity of colour in the net heat plot relates to statistically significant and clinically meaningful inconsistency. The net heat plot appears to have limited utility for assessing inconsistency.

Throughout this chapter, all networks were assumed to contain two-arm trials only and the

indirect evidence for a design was assumed to come from pathways involving one additional treatment only. While this is unlikely to be true in larger networks, the weighting of the indirect evidence gets smaller as more additional treatments are involved so the contribution of longer pathways to the indirect evidence is minimal. Furthermore, I have shown that the net heat approach can be misleading when only considering two arm trials. Therefore given the added complexity of including multi-arm trials in a network, it is likely that the net heat approach will only become more problematic with increasing network complexity.

Identifying inconsistency in a network will depend to some extent on the network connectedness and the number of treatments and designs. If more than one design deviates from the true effect then it is possible that inconsistency might be masked. Similarly, inconsistency might be harder to spot in a fully connected network, where there are numerous pathways of indirect evidence, than in a network with fewer direct (and indirect) connections. In MA, forest plots can be used to check for outlying single studies and highly weighted studies which can both be influential. In NMA where evidence for a treatment comparison comes from several sources a forest plot may not provide all the information necessary for assessing influential trials or designs. Additional complexity arises when a network includes multi-arm trials. Therefore, careful exploratory work plus presenting the results as in Chapter 4 are the key, rather than the net heat plot. Alternative graphical methods to the net heat approach and forest plot which appropriately assess the amount of inconsistency within a network and display the results graphically, clearly highlighting influential and inconsistent designs, are needed.

There are many methods for assessing inconsistency in NMA. In this chapter I considered four of the most popular methods from the simple method of Bucher (1997) to more complex models such as Lu & Ades (2006) and the graphical net heat approach (Krahn et al., 2013). Using Bucher's approach to test for inconsistency within each loop leads to problems with multiple testing and can be cumbersome in large networks with many loops whereas using the net heat approach in a large network is straight forward. The Lu & Ades approach is straight forward to incorporate within the one-step IPD Royston-Parmar NMA model and quantifies inconsistency but does not provide a straight forward way for locating the inconsistency. Other methods of assessing inconsistency which have not been considered in this chapter include node-splitting (Dias et al., 2010b), alternative inconsistency models (e.g. design-by-treatment interaction model (Higgins et al., 2012))

and the two-stage approach (Lu et al., 2011). All methods to assess inconsistency should be interpreted cautiously, taking the clinical context into account.

In this chapter I proposed an alternative approach to assessing inconsistency in a NMA. This approach draws on the useful aspects of the net heat plot such as the use of a colour range to indicate differing levels of inconsistency. Other strengths of this approach include the detailed consideration of the network structure and the use of standard statistical tests to inform the decision about the extent of inconsistency in the network. Clearly this approach needs developing further. The biggest disadvantage of this approach is that I have not been able to fully illustrate the approach as I do not have an appropriate dataset to do so. The cervical and lung cancer networks only contain one three-treatment loop each and therefore only require the inclusion of one inconsistency parameter. To fully illustrate this approach I will need to obtain a network with at least two treatment loops but ideally with three or more treatment loops. This would then allow me to test my proposed alternative approach on a network which has the potential for inconsistency to arise from more than one source. Furthermore, the proposed alternative approach also needs to be applied to a number of networks of varying size and structure. This should help determine the generalisability of the approach and also identify any areas in which users might require more detailed guidance. In addition, I have proposed that the amount of direct information available for each comparison is displayed by a box in each table cell which is shaded to indicate the proportion of the total evidence coming from the direct evidence for each comparison. I need to identify where best to place these boxes within the table cells. I will also need to consider what to do if removing all the non-statistically significant inconsistency parameters and re-fitting the NMA model results in changes to the significance of the inconsistency parameters. This alternative approach provides an ideal starting point for developing a method for assessing inconsistency which starts by considering the structure of the network and uses common statistical tests which can be easily interpreted to provide an easy visual assessment of inconsistency in a network.

Inconsistency in a network can lead to biased treatment effect estimates therefore it is important that attempts are made to identify, understand and adjust for inconsistency. Inconsistency models can be used to adjust for inconsistency in a network however if covariates are distributed unevenly between trials then one method of reducing inconsistency is to adjust for covariates. Consistency models extended to include patient-level covariates or

treatment-covariate interactions could be used to both explain and adjust for inconsistency.

In Chapter 6 I will extend the one-step IPD Royston-Parmar NMA model, from Chapter 4, to include patient-level covariates and treatment-covariate interactions before proposing a practical framework for conducting one-step IPD NMA with treatment-covariate interactions.

6 Interactions

6.1 Introduction

Stratified medicine has increased in popularity in recent years as it aims to target treatments at subgroups of patients who are most likely to benefit from treatment (Medical Research Council, 2016). Single RCTs are often underpowered to be able to detect interactions between treatment effects and subgroups. However, MA and NMA — which pool information from multiple RCTs — can provide greater power for identifying subgroups of patients most likely to respond to treatment (Rothwell, 2005; Stewart and Tierney, 2002). Treatment-covariate interactions can only be investigated thoroughly when IPD has been collected and if the treatment effect estimates are affected by a covariate, or covariates, it is important to consider including treatment-covariate interactions as they can help explain both heterogeneity and inconsistency. By including treatment-covariate interactions in a NMA we can explore the identification of any covariates which may modify the treatment effect estimates and estimate how much the covariate modifies the treatment effects (Donegan et al., 2012). For example, a NMA model including a treatment-age interaction might establish whether older patients benefit more (or less) from treatment.

Chapter 4 introduced the Royston-Parmar model as a computationally practical and flexible alternative to the Cox PH model for analysing time-to-event data and detailed how it can be used in a one-step approach to conduct NMA. Section 2.7 reviewed how some of the methodology for analysing treatment-covariate interactions in NMA has developed since the first paper was published on the subject in 2007. For pairwise MA, Riley (2008) and Fisher (2011) described how patient-level interactions may contain both within and across trial information, and described processes for separating them across a range of models. In this chapter I will apply these methods to the NMA setting showing how the one-step IPD Royston-Parmar NMA model can be easily extended to incorporate both patient-level covariates and treatment-covariate interactions, and how the within and across trial information can be separated. The inclusion of covariates inevitably raises the problem of how to deal with missing covariate data and I will show that this can be accommodated using a Bayesian framework and a one-step approach.

This chapter continues in Section 6.2 by reviewing the current practice for estimating

treatment-covariate interactions in MA. In Section 6.3 I consider the methodology for the inclusion of both continuous and categorical patient-level covariates into the one-step IPD Royston-Parmar NMA model. Section 6.4 will explore how missing covariate data can be handled in WinBUGS. In Section 6.5 I apply the models described in Section 6.3 to the cervical cancer network using the methods from Section 6.4 to deal with any missing covariate data. Section 6.6 explains the methodology for including treatment-covariate interactions in the one-step IPD Royston-Parmar NMA model including how the within and across trial information can be combined or separated. These models are applied to the cervical cancer network and the results are presented in Section 6.7. Building on this, in Section 6.8, a practical framework for conducting one-step IPD NMA with treatment-covariate interactions is proposed. Section 6.9 illustrates the use of the practical framework for conducting one-step IPD NMA with treatment-covariate interactions to the lung cancer network as an example of the framework in action. I conclude with a discussion in Section 6.10.

6.2 Treatment-covariate interactions in meta-analysis

One of the big advantages of IPD MA over AD MA is the greater power that it affords for investigating treatment and patient-level covariate interactions (Tierney et al., 2015). To explore how the treatment effect may vary in relation to a patient-level covariate, a treatment-covariate interaction can be fitted. In the MA setting (as opposed to the single trial setting) this results in two sources of information: within trial information and across trial information. Hence, there are three possible ways of analysing treatment-covariate interactions: using across trial information only, using within trial information only and combining the two (Fisher, Carpenter, Morris, Freeman and Tierney, 2016). Throughout this section continuous effects of covariates are assumed.

An analysis using across trial information only considers how the treatment effect varies across trials in relation to the trial mean value of the covariate and fails to use the patient-level information. This requires the assumption of no unmeasured confounding between the outcome and the covariate; in particular, that there is no ecological bias (Fisher et al., 2016). Ecological bias arises when conclusions about individuals are made using group data (Greenland and Morgenstern, 1989). For example, if the mean age of patients within each trial is used to predict treatment outcomes, the effect may be confounded by other co-

variates associated with the trials (e.g. aspects of protocol) rather than with the individual patients. Hence, conclusions about how age affects the treatment outcomes of individual patients will be subject to ecological bias. Unfortunately it is typically not possible to identify such confounders, nor hence to test whether the inclusion of across-trial information will induce bias.

An analysis using within trial information only more closely parallels the underlying principles of MA. The treatment effect is calculated within each trial for each level of the covariate. The treatment effects for each level of the covariate are compared within each trial to give an interaction effect estimate for each trial, which are then pooled together using MA methods. Any trials where all patients have the same covariate value will not contribute to this analysis as they do not provide any within trial information (Fisher et al., 2016). Recommendations on the presentation and analysis of treatment-covariate interactions using this approach are proposed by Fisher (2016).

In an analysis which combines the within and across trial information, the treatment effect is calculated within each trial for each level of the covariate. The treatment effects for each level of the covariate are combined across all trials, using standard MA techniques, resulting in an overall effect for each level of the covariate which are then compared to each other. Any trials where all patients have the same covariate value will not contribute to the within trial information but can contribute to the across trial information. This is a common approach used in IPD MA. However, the within trial information can be exaggerated or masked by the across trial information, which as described is already at risk of ecological bias. Therefore, this approach is also at risk of ecological bias (Fisher et al., 2011).

The preferred reporting items for systematic reviews and meta-analyses based on IPD (PRISMA-IPD) guidelines recommend including a description of the methods to be used for exploring treatment-covariate interactions (Stewart et al., 2015; Tierney et al., 2015). This includes pre-specification of whether across trial information is to be combined with within trial information. The PRISMA-IPD guidelines also include an additional item which requires authors to report whether an interaction is consistent across trials (Stewart et al., 2015).

In pairwise MA combining within trial information with across trial information is seen as controversial by some, with Fisher (2016) and Riley (2008) recommending use of the within

trial information only. However, Fisher (2016) also acknowledges that inclusion of across-trial information may be of value for exploratory analysis, provided the two sources of information are also presented separately. NMA cannot add to within trial information but could add to the across trial information. Therefore, if across trial information is disregarded completely then theoretically there is nothing to be gained from conducting NMA with treatment-covariate interactions. Therefore, the approach I have taken in this thesis is to conduct NMA with treatment-covariate interactions fitting models which combine the within and across trial information and models which separate out the within and across trial information. This gives researchers the option to use the methods and decide whether there is value in combining the within and across trial information.

There are some additional decisions for NMA which also need to be considered before a NMA with treatment-covariate interactions can be fitted. These include the parameterisation of covariate effects, parameterisation of interaction effects (including separation of within and across trial information) and consistency of covariate and interaction effects. In addition, an added complication, frequently encountered in practice, is how to handle missing patient-level covariate data and how to handle trials which do not have patients at all levels of the covariate. This chapter will explore these areas illustrating the methodology with the cervical cancer network. Following on from this, a practical framework for how to explore treatment-covariate interactions in a one-step IPD NMA is developed and a definitive analysis of the lung cancer network presented.

6.3 Covariate effects in NMA

In this section I extend the one-step IPD Royston-Parmar NMA model (4.5) introduced in Section 4.4 to include patient-level covariates. I consider continuous covariates first and then move on to consider categorical covariates. As an important practical aside note that continuous covariates should be centered to aid convergence of the MCMC algorithm.

Three different assumptions about a patient-level covariate can be made. Firstly, a common effect of the covariate can be fitted, which assumes that the covariate has the same effect in all trials. Secondly, a fixed trial-level effect of the covariate can be fitted, which estimates a trial effect of the covariate for each trial. Thirdly, a random trial-level effect of the covariate can be fitted, which allows the effect of the covariate to differ in each trial, but

assumes that the coefficients for each trial come from a common (typically normal) distribution.

In this section I describe the one-step IPD Royston-Parmar NMA model including a patient-level covariate assuming there is no missing covariate data. I will return to the issue of missing covariate data in Section 6.4. Model code, based on the lung cancer network, for continuous covariates is provided in Appendix C.4.

6.3.1 Continuous covariates

As before, throughout this chapter the subscript i denotes patient and the subscript j denotes trial.

Adding a common effect of the continuous covariate, z_{ij}, to the FTE model (4.5) gives:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1 \text{trt1}_{ij} + \cdots + \beta_q \text{trtq}_{ij} + \alpha z_{ij}, \tag{6.1}$$

where z_{ij} is the covariate value for patient i in trial j and α the coefficient. $s_j\big(\ln(t)\big)$, $\beta_1, \ldots, \beta_q$ and trtq_{ij} are as defined in (4.5). For example, in the cervical cancer network if z_{ij} represents age, then a common effect of age, α, implies that the effect of age is the same whatever the log cumulative baseline hazard rate is for each trial. In other words, α represents the pooled effect of age across all trials. Therefore for patient i in trial j, receiving treatment q at time t their log cumulative baseline hazard rate will be equal to $s_j\big(\ln(t)\big) + \beta_q + \alpha z_{ij}$.

The FTE model (4.5) including a fixed trial-level effect of a continuous covariate, z_{ij}, becomes:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1 \text{trt1}_{ij} + \cdots + \beta_q \text{trtq}_{ij} + \alpha_j z_{ij}, \tag{6.2}$$

where z_{ij} is the covariate value for patient i in trial j and α_j is the coefficient for trial j. $s_j\big(\ln(t)\big)$, $\beta_1, \ldots, \beta_q$ and trtq_{ij} are as defined in (4.5). Note that as the model estimates a trial effect of the covariate for each trial, this model does not give an overall effect for the covariate.

The RTE model (4.6) including a random trial-level effect of the continuous covariate, z_{ij}, becomes:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_{1j} \text{trt1}_{ij} + \cdots + \beta_{qj} \text{trtq}_{ij} + \alpha_j z_{ij} \tag{6.3}$$

$$\boldsymbol{\beta_j} \sim MVN(\boldsymbol{\mu}, \boldsymbol{T})$$

$$\alpha_j \sim N(\theta, \sigma^2)$$

where z_{ij} is the covariate value for patient i in trial j and α_j the coefficient for trial j. $s_j\big(\ln(t)\big)$, $\beta_1, \ldots, \beta_q$ and trtq_{ij} are as defined in (4.5). For example, in the cervical cancer network if z_{ij} represents age, for patient i in trial j at time t their log cumulative baseline hazard rate will be equal to $s_j\big(\ln(t)\big)$ plus their trial specific estimate of the effect of age α_j. Note that this model also estimates an overall effect of age across all trials, represented by θ. In a standard random effects MA, with sufficient information, an unstructured correlation matrix over β and α would be used. However, in the cervical and lung cancer networks, there was insufficient data to estimate a sparse covariance matrix. Therefore instead, I allow the random effects for β to be correlated and the random effects for α to be correlated.

If a common effect of a covariate is used when a trial-level effect would be more appropriate then this can result in a poorly fitting model. This could affect convergence and suppress the differences between trials, which could affect the treatment effect estimates.

6.3.2 Categorical covariates

In this section I assume a categorical covariate with levels $1, \ldots, s$ where level 1 is considered as the reference level for the covariate.

Adding a common effect of a categorical covariate with s levels to the FTE model (4.5) gives:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1\text{trt1}_{ij} + \cdots + \beta_q\text{trtq}_{ij}$$
$$+ \alpha_1 1[\text{level}_{ij} = 2] + \cdots + \alpha_{(s-1)} 1[\text{level}_{ij} = s],$$

where $1[\text{level}_{ij} = s]$ is an indicator variable taking the value 1 if $\text{level}_{ij} = s$ for patient i from trial j and 0 otherwise, and $\alpha_{(s-1)}$ the coefficient. $s_j\big(\ln(t)\big)$, $\beta_1, \ldots, \beta_q$ and trtq_{ij} are as defined in (4.5).

The FTE model (4.5) including a fixed trial-level effect of a categorical covariate with s levels becomes:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1\text{trt1}_{ij} + \cdots + \beta_q\text{trtq}_{ij}$$
$$+ \alpha_{1j} 1[\text{level}_{ij} = 2] + \cdots + \alpha_{(s-1)j} 1[\text{level}_{ij} = s], \tag{6.4}$$

where $1[\text{level}_{ij} = s]$ is an indicator variable taking the value 1 if $\text{level}_{ij} = s$ for patient i from trial j and 0 otherwise, and $\alpha_{(s-1)j}$ the coefficient. $s_j\big(\ln(t)\big)$, $\beta_1, \ldots, \beta_q$ and trtq_{ij} are as defined in (4.5). Note that as the model estimates a trial effect of the covariate for each trial, this model does not give an overall effect for the covariate.

The RTE model (4.6) including a random trial-level effect of a categorical covariate with s levels becomes:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_{1j}\text{trt1}_{ij} + \cdots + \beta_{qj}\text{trtq}_{ij}$$
$$+ \alpha_{1j}1[\text{level}_{ij} = 2] + \cdots + \alpha_{(s-1)j}1[\text{level}_{ij} = 2]$$

$$\boldsymbol{\beta_j} \sim MVN(\boldsymbol{\mu_\beta}, \boldsymbol{T_\beta})$$

$$\boldsymbol{\alpha_j} \sim MVN(\boldsymbol{\mu_\alpha}, \boldsymbol{T_\alpha})$$

where $1[\text{level}_{ij} = s]$ is an indicator variable taking the value 1 if $\text{level}_{ij} = s$ for patient i from trial j and 0 otherwise, and $\alpha_{(s-1)j}$ the coefficient. $s_j\big(\ln(t)\big)$, $\beta_1, \ldots, \beta_q$ and trtq_{ij} are as defined in (4.5).

By design a trial may not have all levels of the covariate present. To minimise this problem, the level of the covariate which is included in the greatest number of trials should be chosen as the reference level for the covariate. However, some trials may remain in which all levels of the covariate are not present. In this case, care is needed in defining the covariate effects. For example, when a covariate has three levels and level one is the reference level let α_1 represent the covariate effect for level 2 compared to level 1 and let α_2 represent the covariate effect for level 3 compared to level 1. Then, in a trial in which only level 2 and level 3 are present then the covariate effect for level 3 compared to level 2 is $\alpha_2 - \alpha_1$.

6.4   Missing covariate data

One of the advantages of conducting NMA within a Bayesian framework using WinBUGS is that missing covariate data can be naturally handled in WinBUGS. Missing covariate data can be accommodated within the NMA model by including a distribution for the covariate with missing values. By doing this, patients with missing covariate data will be included in the model and therefore the observed information from these patients will increase the precision of the estimates for the other model parameters (Lunn et al., 2013). In the case of

the models presented in Section 6.3, the patients with missing covariate data will increase the precision of the treatment effects. By incorporating missing covariate data in this way two things can happen: missing covariate values are imputed which allows a patient to be included in the NMA model and this in turn increases the precision of the treatment effects which themselves inform the imputation of the missing values.

To illustrate how this works, consider a simple example. Suppose a NMA model is written, generically, as

$$f(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta}),$$

where $\mathbf{y}$ is the vector of outcome data, $\mathbf{x}$ the vector of covariates and $\boldsymbol{\theta}$ the corresponding vector of parameters.

When a model is fitted in the Bayesian framework, a prior distribution for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta})$, is specified which is typically uninformative. Using MCMC, we then draw from the posterior distribution of $\boldsymbol{\theta}$, which is proportional to $f(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\boldsymbol{\theta})$.

When some individuals have missing values for the covariate $\mathbf{x}$ we choose an appropriate marginal distribution for $\mathbf{x}$, say $g(x)$, and then extend the MCMC procedure as follows:

1. Draw missing covariate values from the posterior distribution of $\mathbf{x}$ given $\boldsymbol{\theta}$

2. Draw from the posterior distribution of $\boldsymbol{\theta}$ given $\mathbf{x}$ (including the current draws of any missing values) and $\mathbf{y}$.

The MCMC process iterates between (1) and (2) giving a posterior distribution for $\boldsymbol{\theta}$ which now draws on the information available from the patients with missing covariate values.

Throughout this chapter the missing covariate data is assumed to be missing at random. This means that given the observed data the data are missing independently of their actual (unobserved) value. By putting a prior distribution on the covariate in the Bayesian model, missing individual participant covariate data is imputed consistent with that model. The results will be biased if there are substantial departures from the (untestable) missing at random model, or if the covariate distribution is markedly mis-specified.

For continuous covariates, missing data can be imputed from a specific distribution, most commonly the normal distribution. In practice it is a good idea to centre continuous covariates to aid convergence of the MCMC algorithm. For example, the continuous covariate

age will have non-zero mean. However, if age is centered on its mean value, a distributional assumption can be introduced which imputes missing values of age from a normal distribution with mean zero. Alternatively, missing data can also be imputed from a normal distribution in which the mean and precision are also drawn from distributions. Categorisation of continuous covariates can result in lost information and a reduction in the power to detect an interaction (Altman and Royston, 2006). In some cases the covariate may be restricted to a certain range. For example, the eligibility criteria of RCTs often restricts entry into trials to patients aged 18 or over. In such cases, the 'I' function, in WinBUGS, can be used to truncate the distribution. Missing covariates can be imputed at both the network and trial level. At the trial level, a distribution for the covariate with missing values is specified for each trial with at least one missing covariate value. At the network level, one distribution for the covariate with missing values is specified and the missing covariate values from all trials are imputed from this one distribution. Throughout this chapter, covariates will be imputed at the network level as this is the easiest approach to implement.

Categorical covariates raise two issues: the choice of distributional assumption and how to parameterise the model when some trials by design do not have all levels of the covariate present. The parameterisation requires some care because it may be that in some trials the reference level for a covariate is not actually observed. For example, the eligibility criteria for some trials might restrict entry to the trial based on stage of disease. With categorical covariates missing data can be imputed for each level of the covariate from a multinomial distribution in which the probabilities of the multinomial distribution sum to one. For example, if the covariate has three levels then the missing data can be imputed from a multinomial distribution with three probabilities summing to one. These three probabilities can be equal, e.g. for a covariate with three levels they can all be $\frac{1}{3}$, or they can be allowed to differ, e.g. for a covariate with three levels they could be 0.2, 0.3 and 0.5.

6.5 Cervical cancer NMA including patient-level covariates

In this section I present the results of applying the one-step IPD Royston-Parmar NMA model including the covariates age and stage of disease to the cervical cancer network to assess overall survival. Models (6.2) and (6.4) were not considered in this section as they do not provide an overall effect of the covariate. In this section age was considered as a

continuous covariate and stage of disease was considered as both a continuous and categorical covariate.

Age was missing for twelve patients (0.2%). Age was centered and imputed from a normal distribution with mean zero and variance ten. The variance could not be estimated from the data as the model would not fit. Therefore, a variance of ten was chosen by looking at the distribution of the data. There was no evidence of an effect of age in either the FTE with common age effect (LogHR=0, 95% CrI: −0.003, 0.004, Table 6.1) or RTE with random trial level age effect model (LogHR=0, 95% CrI: −0.004, 0.005).

Stage of disease was missing for 405 patients (6.8%). When stage of disease was included as a continuous covariate missing values were imputed from a normal distribution with mean one and variance one. Again, the variance could not be estimated from the data as the model would not fit. Therefore, a variance of one was chosen from looking at the distribution of the data. A linear effect of stage was assumed which could take the values 0 = stages 1A-2A, 1 = stage 2B, 2 = stages 3A-4A so that the reference value, 0, was representing stages 1A-2A. When included as a continuous covariate in both the FTE with common effect of stage and RTE with random trial-level effect of stage models, the parameter estimate for stage suggests that overall survival is reduced as stage of disease increases (RTE model with random trial-level effect of stage: LogHR=0.561, 95% CrI: 0.475, 0.641). This means that the risk of death is higher for patients with stages 3A-4A disease than for patients with stages 1A-2A disease, as expected.

In the categorical models stage of disease was divided into three categories: stages 1A-2A, stage 2B, stages 3A-4A with stages 3A-4A taken to be the reference group. Note, the reference group in this model differs to the model with stage treated as a continuous covariate therefore I would expect the treatment effects to go in the opposite direction. The missing data parameters were allowed to vary by trial and were drawn from a multinomial distribution with mean μ and precision T where μ was drawn from a multivariate normal distribution and T from a Wishart distribution. In both models there was evidence to suggest that overall survival was increased in patients with stage 1A-2A and stage 2B disease compared to stage 3A-4A disease. In the RTE model with random-trial level effect of stage there was evidence to suggest that overall survival was improved by 65% in patients with stage 1A-2A disease compared to patients with stage 3A-4A disease (LogHR=−1.041, 95% CrI: −1.327, −0.749) and by 46% in patients with stage 2B disease compared to

patients with stage 3A-4A disease (LogHR=−0.609, 95% CrI: −0.750, −0.461).

The treatment effect estimates, presented in Table 6.1, for the NMA models including a patient-level covariate are consistent with those presented in Table 4.3 for the FTE and RTE models with no covariates. Despite the inclusion of age or stage of disease as a covariate, the treatment effect for CTRT compared to RT remained statistically significant in all models, the treatment effect for CT+S compared to RT was statistically significant in the FTE models only and the treatment effect for RT v CT+RT short cycles was not statistically significant in any of the models. In Table 4.3 the treatment effect for RT v CT+RT long cycles was statistically significant in both the FTE and RTE models whereas in Table 6.1 the treatment effect for RT v CT+RT long cycles was not statistically significant in the RTE model with random trial-level effect of stage when stage was categorical.

Assuming a common effect of stage might not be appropriate given that some trials in the cervical cancer network restricted entry to the trial based on stage of disease. For example, the Keys (1999) trial includes stage 1B patients only, the Cardenas 93 (Cardenas et al., 1993) trial includes stage 3B patients only and the Pearcey (2002) trial includes patients from stage 1B up to 4A. Clearly it would be inappropriate to assume that the effect of stage of disease was the same across these trials. Likewise with age, unless the distribution of age within each trial was the same it would be inappropriate to assume that the effect of age would be the same in all trials. The cervical cancer network showed some evidence of heterogeneity. Therefore, the RTE model with random trial-level effect of the covariate would be more appropriate, for both age and stage, as it increases the variability of the log hazard ratio to reflect the heterogeneity.

Another consideration should be the size of trials. In the cervical cancer network the size of the trials varies quite considerably. It is well known that smaller studies can give more extreme parameter estimates (Chaimani and Salanti, 2012). Therefore, in networks where trials vary in size, such as the cervical cancer network, it might be more appropriate to use the RTE model with random trial-level effect of the covariate.

Table 6.1: Cervical cancer covariate model results. RT = radiotherapy, CT+RT = neodadjuavnt chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

	RT v CTRT	RT v CT+RT short cycles	RT v CT+RT long cycles	RT v CT+S	Covariate effect
FTE common age	−0.218 (−0.343, −0.095)	0.024 (−0.166, 0.221)	0.227 (0.066, 0.382)	−0.396 (−0.603, −0.186)	0 (−0.003, 0.004)
RTE random trial level age	−0.203 (−0.374, −0.027)	0.044 (−0.328, 0.440)	0.297 (0.041, 0.591)	−0.370 (−0.803, 0.032)	0 (−0.004, 0.005)
FTE common stage (continuous)	−0.210 (−0.332, −0.085)	0.001 (−0.193, 0.204)	0.183 (0.022, 0.340)	−0.426 (−0.640, −0.207)	0.556 (0.471, 0.642)
RTE random trial level stage (continuous)	−0.198 (−0.346, −0.031)	0.005 (−0.320, 0.328)	0.254 (0.008, 0.540)	−0.372 (−0.803, 0.056)	0.561 (0.475, 0.641)
FTE common stage (categorical)	−0.208 (−0.333, −0.083)	0.008 (−0.187, 0.200)	0.174 (0.020, 0.329)	−0.427 (−0.645, −0.214)	Stage 1A-2A: −1.057 (−1.282, −0.844); Stage 2B: −0.617 (−0.727, −0.509)
RTE random trial level stage (categorical)	−0.190 (−0.350, −0.026)	0.098 (−0.231, 0.465)	0.206 (−0.063, 0.486)	−0.323 (−0.725, 0.057)	Stage 1A-2A: −1.041 (−1.327, −0.749); Stage 2B: −0.609 (−0.750, −0.461)

6.6 Treatment-covariate interactions

In this section I extend the one-step IPD Royston-Parmar NMA model (4.5) to include treatment-covariate interactions.

As in Section 6.3 there are three possible assumptions for treatment-covariate interactions: common, fixed trial-level and random trial-level. In this section I consider common and random trial-level effects of the treatment-covariate interactions only. A common effect of the treatment-covariate interaction assumes that the treatment-covariate interaction has the same effect in all trials. A random trial-level effect of the treatment-covariate interaction allows the effect of the treatment-covariate interaction to differ in each trial but assumes that the coefficients for each trial come from a common (typically normal) distribution.

Following work by Riley (2008) and Fisher (2011) for pairwise MA, I consider two ways of including a treatment-covariate interaction in a model. Firstly, as a single effect which combines within and across trial information; and secondly, as two effects which separate out the within and across trial information. In this section, I show how these methods can be applied to the NMA setting using the one-step IPD Royston-Parmar NMA model including how the within and across trial information can be combined or separated when treatment-covariate interactions are included in the model. These models are then applied to the cervical cancer network and the results are presented in Section 6.7.

In this section I assume that ordered categorical covariates can be considered as continuous covariates represented by a linear effect across all values of the covariate. Model code for the RTE models described in this section are presented in Appendix C.5.

6.6.1 Combining within and across trial information

The FTE model with common treatment-covariate interaction and fixed trial-level effect of the covariate, z_{ij}, becomes:

$$
\begin{aligned}
\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1\mathsf{trt1}_{ij} + \cdots + \beta_q\mathsf{trtq}_{ij} \\
+ \alpha_j z_{ij} \\
+ \delta_1\mathsf{trt1}_{ij}z_{ij} + \cdots + \delta_q\mathsf{trtq}_{ij}z_{ij}
\end{aligned}
\tag{6.5}
$$

where $s_j\big(\ln(t)\big)$ and trtq$_{ij}$ are as defined in (4.5). In this model z_{ij} is the covariate value for patient i in trial j, $\beta_1, \ldots, \beta_q$ are common treatment effects, α_j is a fixed trial level effect of the covariate z_{ij} for trial j and $\delta_1, \ldots, \delta_q$ are common treatment-covariate interaction effects. Going forward this model will be referred to as the 'FTE combined' model.

An approximate global Wald test can be conducted on the treatment-covariate interaction parameter estimates to assess whether the interaction terms could be removed from the model without harming the fit of the model. The null hypothesis states that the interaction terms are simultaneously equal to zero. For example, the null hypothesis would test that $\delta_1 = \cdots = \delta_q = 0$. For details on conducting an approximate global Wald test from the posterior distribution see Subsection 4.4.2 and Appendix D.

The RTE model with random trial-level treatment-covariate interaction and random trial-level effect of the covariate, z_{ij}, becomes:

$$
\begin{aligned}
\ln\{H_j(t|x_{ij})\} = {} & s_j\big(\ln(t)\big) + \beta_{1j}\text{trt1}_{ij} + \cdots + \beta_{qj}\text{trtq}_{ij} \\
& + \alpha_j z_{ij} \\
& + \delta_{1j}\text{trt1}_{ij}z_{ij} + \cdots + \delta_{qj}\text{trtq}_{ij}z_{ij}
\end{aligned}
\tag{6.6}
$$

$$
\boldsymbol{\beta_j} \sim MVN(\boldsymbol{\mu_\beta}, \boldsymbol{T_\beta})
$$

$$
\boldsymbol{\delta_j} \sim MVN(\boldsymbol{\mu_\delta}, \boldsymbol{T_\delta})
$$

$$
\boldsymbol{\alpha_j} \sim N(\theta, \sigma^2)
$$

where $s_j\big(\ln(t)\big)$ and trtq$_{ij}$ are as defined in (4.5). In this model z_{ij} is the covariate value for patient i in trial j, $\beta_{1j}, \ldots, \beta_{qj}$ are random treatment effects, α_j is a random trial level effect of the covariate z_{ij} for trial j and $\delta_{1j}, \ldots, \delta_{qj}$ are random treatment-covariate interaction effects. Going forward this model will be referred to as the 'RTE combined' model. With standard random effects, with sufficient information, an unstructured correlation matrix over β, α, and δ would be used. However, in the cervical and lung cancer networks, there was insufficient data to estimate such a sparse covariance matrix. Therefore instead, I allow the random effects for β to be correlated, the random effects for α to be correlated and the random effects for δ to be correlated.

In the treatment-covariate interaction model a trial-level effect of the covariate is included which allows for a separate effect of the covariate by trial. If a common effect of the covariate is used instead this could suppress the differences between trials, biasing the

interaction terms and leaning the results towards suggesting evidence of an interaction. Figure 6.1 illustrates the addition of a treatment-covariate interaction for the direct comparison of treatment A versus treatment B when a continuous covariate, z_{ij}, which can take one of three values, is included in the model. The model includes a random treatment effect (β_j), fixed trial-level covariate effect (α_j) and random treatment-covariate interaction effect (δ_j). Making the treatment-covariate interaction random allows the effect to differ in each trial but assumes that there is a mean overall effect. This means that in Figure 6.1 δ does not need to be exactly the same for each trial. Note that when treatment-covariate interactions are present the treatment effects (β) are expected to head towards zero because the covariates are centered on the mean.

$$\eta_{ij} = s_j\big(ln(t)\big) + \beta_j \text{trt}_{ij} + \alpha_j z_{ij} + \delta_j \text{trt}_{ij} z_{ij}$$

Where $trt_{ij} = \begin{cases} 0 \; if \; trt = A \\ 1 \; if \; trt = B \end{cases}$

Log cumulative hazard at time t
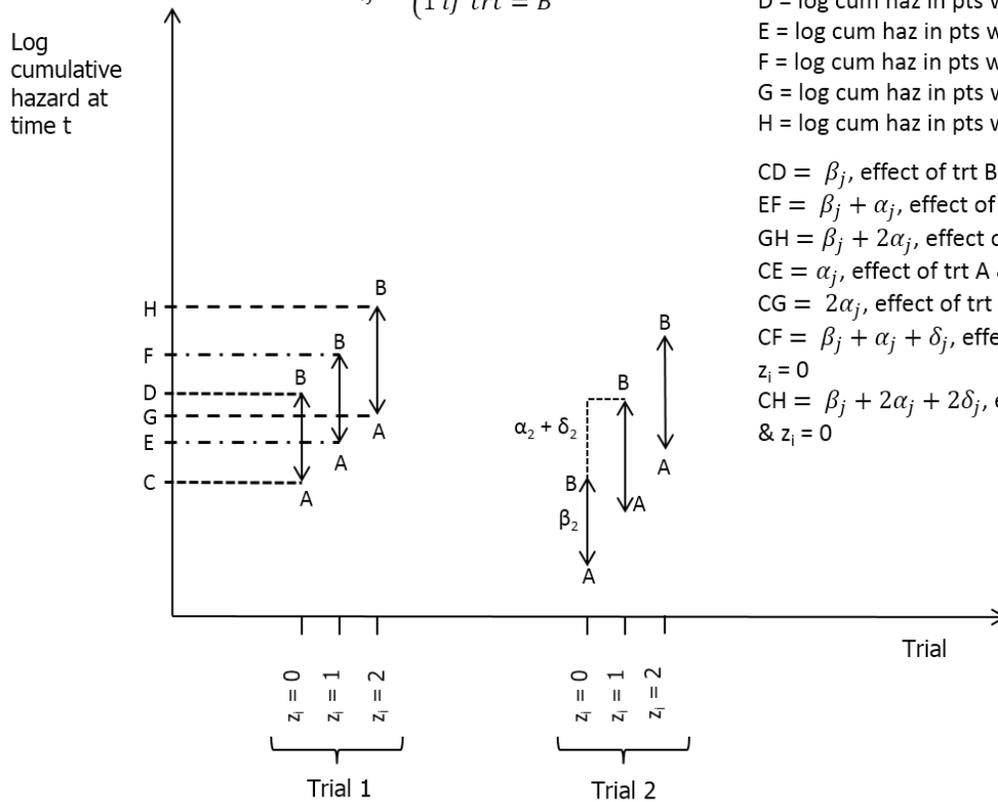
C = log cum haz in pts with z_i = 0 and trt A
D = log cum haz in pts with z_i = 0 and trt B
E = log cum haz in pts with z_i = 1 and trt A
F = log cum haz in pts with z_i = 1 and trt B
G = log cum haz in pts with z_i = 2 and trt A
H = log cum haz in pts with z_i = 2 and trt B

CD = β_j, effect of trt B versus trt A when z_i = 0
EF = $\beta_j + \alpha_j$, effect of trt B versus trt A when z_i = 1
GH = $\beta_j + 2\alpha_j$, effect of trt B versus trt A when z_i = 2
CE = α_j, effect of trt A & z_i = 1 versus trt A & z_i = 0
CG = $2\alpha_j$, effect of trt A & z_i = 2 verus trt A & z_i = 0
CF = $\beta_j + \alpha_j + \delta_j$, effect of trt B & z_i = 1 versus trt A & z_i = 0
CH = $\beta_j + 2\alpha_j + 2\delta_j$, effect of trt B & z_i = 2 versus trt A & z_i = 0

H

B

F

B

D

B

G

E

A

A

C

A

Trial

B

$\alpha_2 + \delta_2$

B

A

B

β_2

A

A

$z_i = 0$ $z_i = 1$ $z_i = 2$

$z_i = 0$ $z_i = 1$ $z_i = 2$

Trial 1

Trial 2

Figure 6.1: Schematic of random treatment-covariate interactions for a comparison of treatment A versus treatment B including a continuous covariate, z_{ij} which can take one of three values. A fixed trial-level effect of the covariate z_{ij} is fitted where i represents patient and j trial. The random treatment-covariate interaction allows δ_j to differ between trials (but assumes that all δ_j come from a common distribution). Trt = treatment, log cum haz = log cumulative hazard, pts = patients.

Example

Consider a consistent three-treatment network with treatments A, B and C, consisting of three two-arm trials, i.e. $j = 1, 2, 3$. Let $j = 1$ compare treatments A and B, $j = 2$ compare treatments A and C and $j = 3$ compare treatments B and C. Let treatment A be the reference treatment and β_1 represent the treatment effect for treatment B compared to treatment A and β_2 represent the treatment effect for treatment C compared to treatment A. Then:

$$
\text{trt1}_{ij} = \begin{cases} 1 & \text{if } j = 1 \text{ and patient receives treatment B} \\ -1 & \text{if } j = 3 \text{ and patient receives treatment C} \\ 0 & \text{otherwise} \end{cases}
$$

$$
\text{trt2}_{ij} = \begin{cases} 1 & \text{if } j = 2 \text{ or } j = 3 \text{ and patient receives treatment C} \\ 0 & \text{otherwise} \end{cases}
$$

The log cumulative hazard from the FTE model with common treatment-covariate interaction and fixed trial-level effect of the covariate (6.5) when $z_{ij} = 1$ for patient i receiving treatment B in trial $j = 1$ is:

$$
\ln\{H_j(t|x_{ij})\} = s_1\big(\ln(t)\big) + \beta_1 + \alpha_1 + \delta_1.
$$

The log cumulative hazard for patient i receiving treatment C in trial $j = 2$ is:

$$
\ln\{H_j(t|x_{ij})\} = s_2\big(\ln(t)\big) + \beta_2 + \alpha_2 + \delta_2.
$$

The log cumulative hazard for patient i receiving treatment B in trial $j = 3$ is:

$$
\ln\{H_j(t|x_{ij})\} = s_3\big(\ln(t)\big) + \alpha_3.
$$

The log cumulative hazard for patient i receiving treatment C in trial $j = 3$ is:

$$
\ln\{H_j(t|x_{ij})\} = s_3\big(\ln(t)\big) - \beta_1 + \beta_2 + \alpha_3 - \delta_1 + \delta_2.
$$

Hence, the treatment-covariate interaction term in $j = 3$ is estimated by $\delta_2 - \delta_1$.

6.6.2 Separating within and across trial information

Across trial information informing a treatment-covariate interaction is present in all trials where more than one value of the covariate is represented. In Section 6.6.1 I combined the

across trial information with the within trial information to estimate the treatment-covariate interaction terms. In this section I separate out these two sources of information to determine how much of the interaction effect is driven by the across trial information.

The FTE model with common treatment-covariate interaction and fixed trial-level effect of the covariate, z_{ij}, becomes:

$$\ln\{H_j(t|x_{ij})\} = s_j\big(\ln(t)\big) + \beta_1\text{trt1}_{ij} + \cdots + \beta_q\text{trtq}_{ij}$$
$$+ \alpha_j z_{ij}$$
$$+ \delta_{A1}\text{trt1}_{ij}z_{ij} + \cdots + \delta_{Aq}\text{trtq}_{ij}z_{ij} \tag{6.7}$$
$$+ \delta_{B1}\text{trt1}_{ij}\bar{z}_j + \cdots + \delta_{Bq}\text{trtq}_{ij}\bar{z}_j$$

where z_{ij} is the covariate value for patient i in trial j and $\bar{z}_j$ is the mean value of z_{ij} for trial j. In this model $\beta_1, \ldots, \beta_q$ are common treatment effects and α_j is a fixed trial-level effect of the covariate z_{ij} for trial j. The within trial information is represented by the $\delta_{A1}, \ldots, \delta_{Aq}$ parameters whilst the across trial information is equal to $\delta_A + \delta_B$. $s_j\big(\ln(t)\big)$ and trtq_{ij} are as defined in (4.5). Going forward this model will be referred to as the 'FTE separate' model.

In Figure 6.2 the top half of the diagram shows how the treatment effect varies across different values of the covariate for four individual trials (i, ii, iii and iv). This is the within trial information. A treatment-covariate interaction is present where the straight lines representing the treatment effects for treatment A and treatment B are not parallel to each other. For each trial the arrow indicates the treatment effect at the mean value of the covariate. In the bottom half of the diagram the treatment effect at the mean value of the covariate is plotted for each trial. The across trial information comes from fitting a meta-regression to this data.

Within trial information
— Treatment A treatment effect with treatment-covariate interaction
– – – Treatment B treatment effect with treatment-covariate interaction
–··–··– Treatment B treatment effect without treatment-covariate interaction
⟷ Treatment effect at mean value of covariate

Across trial & network information
⬤ Treatment effect, size proportional to inverse variance of trial
▬ Across trial effect

Figure 6.2: Schematic of within and across trial information for a simple comparison of two treatments, A and B.

The RTE model with random trial-level treatment-covariate interaction and random trial-level effect of the covariate, z_{ij}, becomes:

$$
\begin{aligned}
\ln\{H_j(t|x_{ij})\} = {} & s_j\big(\ln(t)\big) + \beta_{1j}\mathsf{trt1}_{ij} + \cdots + \beta_{qj}\mathsf{trtq}_{ij} \\
& + \alpha_j z_{ij} \\
& + \delta_{A1j}\mathsf{trt1}_{ij}z_{ij} + \cdots + \delta_{Aqj}\mathsf{trtq}_{ij}z_{ij} \\
& + \delta_{B1j}\mathsf{trt1}_{ij}\bar{z}_j + \cdots + \delta_{Bqj}\mathsf{trtq}_{ij}\bar{z}_j \\
& \boldsymbol{\beta_j} \sim MVN(\boldsymbol{\mu_\beta}, \boldsymbol{T_\beta}) \\
& \boldsymbol{\alpha_j} \sim N(\theta, \sigma^2) \\
& \boldsymbol{\delta_{Aj}} \sim MVN(\boldsymbol{\mu_{\delta_A}}, \boldsymbol{T_{\delta_A}}) \\
& \boldsymbol{\delta_{Bj}} \sim MVN(\boldsymbol{\mu_{\delta_B}}, \boldsymbol{T_{\delta_B}})
\end{aligned}
\tag{6.8}
$$

where z_{ij} is the covariate value for patient i in trial j and $\bar{z}_j$ is the mean value of z_{ij} for each trial j. In this model $\beta_{1j}, \ldots, \beta_{qj}$ are random treatment effects and α_j is a random trial level effect of the covariate z_{ij} for trial j. The within trial information is estimated by μ_{δ_A} whilst the across trial information is estimated by $\mu_{\delta_A} + \mu_{\delta_B}$. $s_j\big(\ln(t)\big)$ and trtq_{ij} are as defined in (4.5). Going forward this model will be referred to as the 'RTE separate' model. In a standard random effects MA, with sufficient information, an unstructured correlation matrix over β, α, δ_A and δ_B would be used. However, in the cervical and lung cancer networks, there was insufficient data to estimate such a sparse covariance matrix. Therefore instead, I allow the random effects for β to be correlated, the random effects for α to be correlated, the random effects for δ_A to be correlated and the random effects for δ_B to be correlated.

In NMA models including treatment-covariate interactions the β parameters on their own will not have a useful interpretation and treatment effects will need to be presented separately for each level of the covariate. In this case the treatment effects would differ between the combined and separated models. For example, from the combined model for a patient with covariate value zero the treatment effect will be β. Whereas, from the separated model the treatment effect for a patient with covariate value zero will be $\beta + \delta_B$.

Example

Consider a consistent three-treatment network with treatments A, B and C consisting of three two-arm trials, i.e. $j = 1, 2, 3$. Let $j = 1$ compare treatments A and B, $j = 2$ compare

treatments A and C and $j = 3$ compare treatments B and C. Let treatment A be the reference treatment and β_1 represent the treatment effect for treatment B compared to treatment A and β_2 represent the treatment effect for treatment C compared to treatment A. Then:

$$\text{trt1}_{ij} = \begin{cases} 1 & \text{if } j = 1 \text{ and patient receives treatment B} \\ -1 & \text{if } j = 3 \text{ and patient receives treatment C} \\ 0 & \text{otherwise} \end{cases}$$

$$\text{trt2}_{ij} = \begin{cases} 1 & \text{if } j = 2 \text{ or } j = 3 \text{ and patient receives treatment C} \\ 0 & \text{otherwise} \end{cases}$$

The log cumulative hazard from the FTE model with common treatment-covariate interaction and fixed trial-level effect of the covariate (6.5) when $z_{ij} = 1$ for patient i receiving treatment B in trial $j = 1$ is:

$$\ln\{H_j(t|x_{ij})\} = s_1\big(\ln(t)\big) + \beta_1 + \alpha_1 + \delta_{A1} + \delta_{B1}\bar{z}_1.$$

The log cumulative hazard for patient i receiving treatment C in trial $j = 2$ is:

$$\ln\{H_j(t|x_{ij})\} = s_2\big(\ln(t)\big) + \beta_2 + \alpha_2 + \delta_{A2} + \delta_{B2}\bar{z}_2.$$

The log cumulative hazard for patient i receiving treatment B in trial $j = 3$ is:

$$\ln\{H_j(t|x_{ij})\} = s_3\big(\ln(t)\big) + \alpha_3.$$

The log cumulative hazard for patient i receiving treatment C in trial $j = 3$ is:

$$\ln\{H_j(t|x_{ij})\} = s_3\big(\ln(t)\big) - \beta_1 + \beta_2 + \alpha_3 - \delta_{A1} + \delta_{A2} - \delta_{B1}\bar{z}_3 + \delta_{B2}\bar{z}_3.$$

Hence, the within trial information for the treatment-covariate interaction term in $j = 3$ is estimated by $\delta_{A2} - \delta_{A1}$ and the across trial information is estimated by $\delta_{A2} - \delta_{A1} + \delta_{B2} - \delta_{B1}$.

6.6.3 Calculating the mean value with missing covariate data

When treatment-covariate interactions are included in a NMA model, to separate out the within and across trial information, the mean value of the covariate is included in the model (Subsection 6.6.2). To calculate the mean covariate value for a trial where some or all

patients have missing covariate data, first fit the model with the within and across trial information combined (Subsection 6.6.1) including the appropriate distributional assumption for missing covariates. Within this model monitor and retain the mean value of the imputed values for each trial where there are any missing covariate values. WinBUGS will only monitor the mean value of the imputed values so in a trial where only some of the patients have missing values of the covariate a weighted average of the observed and imputed values of the covariate is needed. This weighted average can then be used as the trial mean value in the model with the within and across trial information separated. In trials where all patients have missing covariate values the mean value from the imputed values, taken directly from the WinBUGS output, can be used as the trial mean value in the model with the within and across trial information separated.

In the cervical cancer network stage of disease was missing for all patients from one trial. In the lung cancer network stage of disease was missing for all patients from eight trials and performance status was missing for all patients from two trials. Throughout this chapter missing covariate data was imputed at the network level. This assumes that the patient characteristics of the individual trials follows the same distribution as the network population.

6.7 Cervical cancer NMA with treatment-covariate interactions

In this section, I present the results of applying the models described in Subsection 6.6.1 and Subsection 6.6.2 to the cervical cancer network, considering stage of disease as the covariate of interest. Code for the models presented in this section, based on the lung cancer network, can be found in Appendix C.5.

As in Section 6.5, stage of disease was considered as a continuous covariate with missing data imputed from a normal distribution. A linear effect of stage was assumed taking the values 0 = stages 1A-2A, 1 = stage 2B, 2 = stages 3A-4A. Table 6.2 presents the results from FTE and RTE models, which both combine and separate the within and across trial information.

The treatment effect for CTRT at the reference value zero, representing stages 1A-2A, remains statistically significant in the FTE combined and RTE combined models despite

the presence of treatment-stage interactions. There is a statistically significant interaction between stage and CTRT in the FTE combined model which suggests that CTRT is more effective in patients with lower stage disease (LogHR=0.194, 95% CrI: 0.032, 0.357). However, this is no longer statistically significant when RTE are used or when the within and across trial information are separated.

In the presence of treatment-stage interactions, with within and across trial information combined, the treatment effect for patients with stage 1A-2A disease receiving CT+RT long cycles compared to RT is reduced, compared to Table 6.1, and no longer statistically significant. There is no evidence of a treatment effect, in stage 1A-2A patients, for CT+RT short cycles compared to RT. There is no evidence of an interaction between CT+RT or CT+S and stage of disease in any of the models. Assuming a mean value of zero the same conclusions are also reached for the FTE and RTE separate models.

Looking back at Table 4.3 to the FTE and RTE models without covariates the treatment effect for CT+S compared to RT was only statistically significant in the FTE model. This was also the case when stage of disease was added as a covariate to the FTE with common effect of stage model (Table 6.1). However, in the presence of treatment-stage interactions the treatment effect for CT+S compared to RT is no longer statistically significant (Table 6.2). The credible intervals are much wider for this comparison, relative to the other treatment comparisons, possibly reflecting the small amount of within-trial information. In this comparison, there are only two trials which have patients distributed over more than one value of stage and can therefore contribute to the within trial information. Both the combined estimate of the within and across trial information from the FTE and RTE combined models and the estimates of the within and across trial information from the FTE and RTE separate models have wide credible intervals.

An approximate global Wald test on the treatment-stage parameter estimates from the FTE combined model provided χ^2=8.079 on 3 df with p=0.0444 and from the RTE combined model χ^2=8.602 on 3 df with p=0.0351. This provides some evidence that at least one of the interaction terms is statistically different from zero suggesting that the treatment effect differs according to stage of disease for at least one of the treatments.

The treatment effects for RT v CT+RT and RT v CT+S from the FTE separate and RTE separate models have changed considerably compared to the FTE combined and RTE

combined models. This is because the within and across trial information are not consistent with each other. From the FTE separate model the log hazard ratio of the within trial information for CT+RT is -0.133 (95% CrI: -0.412, 0.168) and the log hazard ratio for the across trial information is 0.062 (95% CrI: -0.182, 0.305). In this model the across trial information could be subject to ecological bias and the within and across trial information should remain separated.

Figures 6.3, 6.4 and 6.5 display the parameter estimates for the treatment-stage interactions from the FTE and RTE models, separating out the within and across trial information and presenting them alongside a combined estimate, for the three main treatment comparisons. Alongside the NMA estimates, the MA estimates from a FTE model are also presented. I propose a relatively strict criterion for agreement between the within and across trial information from the FTE model. I suggest that agreement is shown if the within trial information is within half a standard error of the across trial information. For example, for the RT v CTRT comparison in the cervical cancer network (Figure 6.3), the within trial information mean is 0.174 with standard error of 0.125. The across trial information mean is 0.208. Therefore the within trial information mean is within half a standard error of the across trial information mean so agreement is suggested. This is in line with my expectations as the RT v CTRT comparison was a branch of the network without any indirect evidence informing the comparison. In this case it would be reasonable to present the combined estimates of the treatment-stage interactions. The agreement between the within and across trial information in Figure 6.3 also suggests that this comparison is not subject to ecological bias. However, in the case of RT v CT+RT and RT v CT+S (Figure 6.4 and Figure 6.5) the within and across trial information do not agree and the combined estimate averages out the within and across trial information. In this case it would be more appropriate to separate out the within and across trial information. For the RT v CT+RT and RT v CT+S comparisons further investigation into the difference between the within and across trial information may be required as these comparisons could be subject to ecological bias.
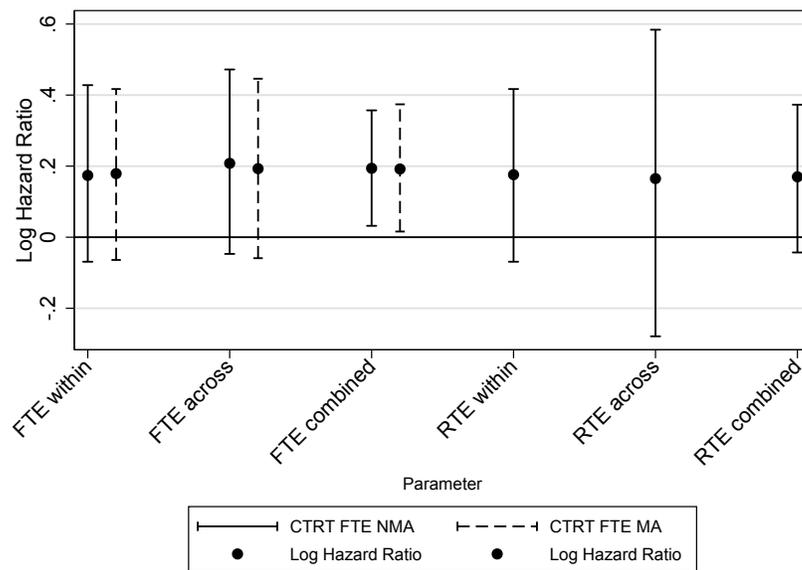
Figure 6.3: Treatment-stage interaction parameter estimates for RT v CTRT. FTE = fixed treatment effect, RTE = random treatment effect, NMA = network meta-analysis, MA = meta-analysis. Solid lines represent NMA estimates. Dashed lines represent MA estimates.

Table 6.2: Cervical cancer treatment-stage interaction model results. Reference level is stages 1A-2A. RT = radiotherapy, CT+RT = neodadjuavnt chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery, FTE = fixed treatment effect, RTE = random treatment effect.

	FTE Combined	FTE Separate	RTE Combined	RTE Separate
RT v CTRT	-0.471 (-0.720, -0.222)	-0.489 (-0.834, -0.159)	-0.428 (-0.738, -0.114)	-0.421 (-0.910, 0.101)
RT v CT+RT short cycles	-0.021 (-0.318, 0.280)	-0.096 (-0.467, 0.260)	0.118 (-0.273, 0.596)	-0.007 (-0.519, 0.550)
RT v CT+RT long cycles	0.182 (-0.068, 0.426)	0.169 (-0.093, 0.433)	0.099 (-0.426, 0.613)	0.100 (-0.551, 0.670)
RT v CT+S	-0.092 (-0.575, 0.369)	0.103 (-0.468, 0.660)	-0.195 (-0.855, 0.380)	0.332 (-0.593, 1.102)
CTRT - stage combined	0.194 (0.032, 0.357)		0.170 (-0.043, 0.373)	
CT+RT - stage combined	0.001 (-0.176, 0.171)		0.006 (-0.234, 0.212)	
CT+S - stage combined	-0.259 (-0.592, 0.079)		-0.120 (-0.635, 0.415)	
CTRT - stage within		0.174 (-0.069, 0.428)		0.176 (-0.069, 0.417)
CT+RT - stage within		-0.133 (-0.412, 0.168)		-0.035 (-0.285, 0.204)
CT+S - stage within		0.192 (-0.423, 0.846)		0.172 (-0.459, 0.776)
CTRT - stage across		0.208 (-0.047, 0.472)		0.165 (-0.279, 0.584)
CT+RT - stage across		0.062 (-0.182, 0.305)		0.110 (-0.315, 0.545)
CT+S - stage across		-0.409 (-0.818, 0.009)		-0.563 (-1.319, 0.230)
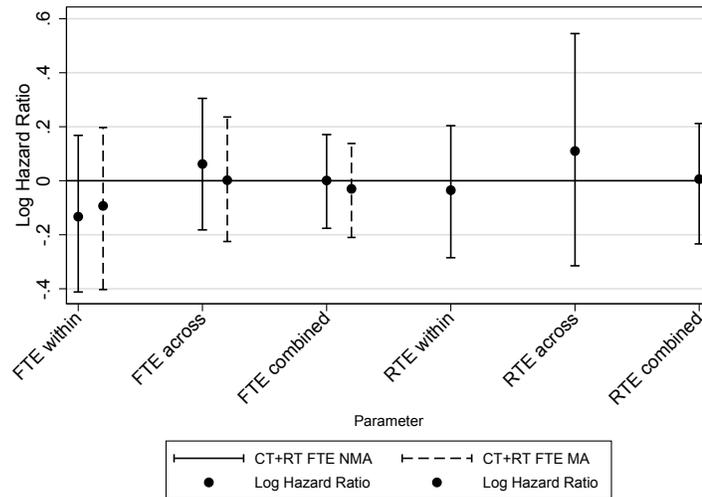
Figure 6.4: Treatment-stage interaction parameter estimates for RT v CT+RT. FTE = fixed treatment effect, RTE = random treatment effect, NMA = network meta-analysis, MA = meta-analysis. Solid lines represent NMA estimates. Dashed lines represent MA estimates.
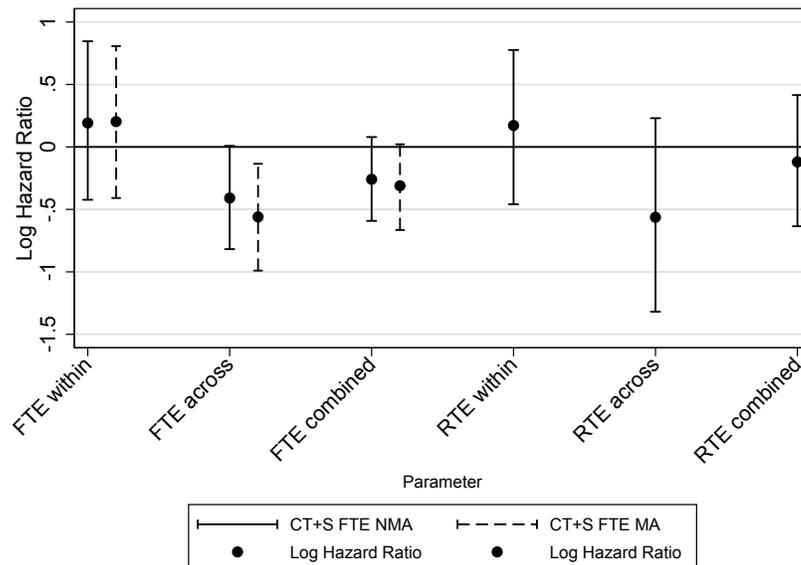
Figure 6.5: Treatment-stage interaction parameter estimates for RT v CT+S. FTE = fixed treatment effect, RTE = random treatment effect, NMA = network meta-analysis, MA = meta-analysis. Solid lines represent NMA estimates. Dashed lines represent MA estimates.

6.8 Practical framework for one-step IPD NMA of time-to-event data with treatment-covariate interactions

In this section, based on the experience gained from using the methods described in the previous sections of this thesis, I develop a twelve-step framework for conducting one-step IPD NMA of time-to-event data with treatment-covariate interactions. I start by outlining all twelve steps before going on to explain each step in more detail. I only consider treatment-covariate interactions for comparisons where there is direct evidence as in this case the network provides both within and across trial information. Where there is no direct evidence the network only provides across trial information. Following Riley (2008) and Fisher (2011), I consider the use of across-trial information only to be ill-advised, and hence do not explore it further in this thesis. This framework considers continuous covariates only. However, ordered-categorical covariates which are treated as continuous could also be used within this framework.

This framework was developed based on NMA being conducted using the one-step IPD Royston-Parmar NMA model to analyse a time-to-event outcome. The framework itself adopts a more general approach and will be applicable to other situations. However, the guidance that follows assumes that NMA is conducted using the one-step IPD Royston-Parmar NMA model to assess overall survival within a Bayesian framework. The framework can be applied to networks including multi-arm trials. Steps 1-6 are applicable for any NMA whether covariates are considered for inclusion or not. From step 7 onwards the framework specifically considers the inclusion of covariates and treatment-covariate interactions. The aim of this framework is to provide, for researchers new to NMA, a useful guide to the steps that need to be considered before a NMA model with treatment-covariate interactions can be fitted. The framework starts by assuming that a systematic review has been conducted in which all eligible trials for inclusion in the NMA have been identified.

Twelve-step framework for one-step IPD NMA of time-to-event data with treatment-covariate interactions

1. Draw a network diagram - Do the treatments form one connected network?

    (a) If yes, move on to the next step

(b) If no,

    i. Can some treatments be split up or combined together to result in a connected network?

    ii. If the treatments form two or more distinct smaller networks then each smaller network will need to be considered separately as it's own NMA

    iii. If there are two treatments connected to each other that are separate from the rest of the network they will have to be excluded from the NMA and a standard pairwise MA with treatment-covariate interaction can be conducted

    iv. If there is one treatment unconnected to the rest of the network then the treatment will have to be excluded from the NMA

2. Assess all pairwise treatment comparisons for evidence of heterogeneity and violation of the PH assumption

(a) If heterogeneity is present, explore the baseline characteristics of all trials. Can the heterogeneity be explained by differences in baseline characteristics across trials?

    i. If yes, covariates which may be causing heterogeneity should be considered in Steps 7 to 9

    ii. If no, it could be unsuitable to combine the pairwise comparison in a NMA

(b) If PH assumption is violated, add treatment-ln(time) interaction to comparison and move on to the next step

(c) If neither are present, move on to the next step

3. Decide on treatment parameterisation

4. Fit NMA model without covariates

5. Assess network for evidence of inconsistency - if inconsistency is present, use an inconsistency parameter in all further models

6. Confirm correct assumption about PH was made in step 2 by fitting NMA model with treatment-ln(time) interaction - Is there evidence of non-PH in the network?

(a) If yes, all further models will need to include the treatment-ln(time) interaction so that non-PH is taken into account

      (b) If no, move on to the next step

7. Investigate patterns of missing data for the covariate of interest

8. Modelling considerations for including covariate in NMA model (FTE or RTE? Common, fixed-trial or random-trial effect of covariate?)

9. Fit NMA model including covariate and assess model results

10. Fit NMA model including treatment-covariate interaction with within and across trial information combined

11. Fit NMA model including treatment-covariate interaction with within and across trial information separated

12. Present results with within and across trial information combined and separated

Guidance on implementing the framework

NMA often starts with a systematic review being conducted to identify all treatments and trials to be considered in the network. As part of the review and in discussion with appropriate clinicians, consideration of any covariates which could be included in a NMA model with treatment-covariate interactions should be discussed before any NMA models are fitted.

Step 1

Start by drawing a network diagram based on the direct evidence identified from the systematic review. The diagram should consist of treatment nodes connected by solid lines where there is direct evidence informing the comparison. The thickness of the lines connecting treatments should be proportional to the number of trials directly comparing the two treatments and the size of the treatment nodes should be proportional to the number of patients randomised to each treatment, such as in Figure 3.1. It can also be useful to draw on arrows indicating the direction of treatment effects, such as in Figure 4.3.

NMA requires all treatments to be connected through direct evidence to at least one other treatment in the network. If the network diagram shows one or more treatments that are not connected to the rest of the network then, consider whether any treatments can be

'split' into several nodes, such as by treatment dose, or whether any treatment nodes could be 'lumped' together, such as combining drugs that act in the same way (Del Giovane et al., 2013). The idea here is to include as much of the direct evidence as possible in the network, however care should be taken to ensure that the comparisons that remain in the network are still appropriate. For example, combining drugs that act in the same way would not be appropriate if they are used in different patient populations. If the network contains two or more smaller clusters of trials which themselves form networks unconnected to each other then each cluster will need to be analysed as it's own NMA. If the network contains two treatments connected to each other but unconnected from the rest of the network then they will have to be excluded from the NMA. A standard pairwise MA with treatment-covariate interaction can be conducted on excluded comparisons. If the network contains one treatment unconnected to the rest of the network then this treatment will have to be excluded from the NMA.

Step 2

Before a NMA model is fitted, in a preliminary step, all pairwise treatment comparisons in the network should be explored for evidence of heterogeneity and violation of the PH assumption. Heterogeneity can be assessed through the I^2, τ^2 and Cochran's Q statistics. If heterogeneity is present, explore the baseline characteristics of all trials. This can be done through tabulating characteristics, such as Table A.1 for the cervical cancer network. If one trial, or a subgroup of trials, are found to be causing the heterogeneity then exploring the baseline characteristics can identify what is different about this trial, or trials, and the impact this might have on the treatment effect. For example, consider the cervical cancer network, from Subsection 4.3.1. Here, the RT v CT+RT comparison was split into two comparisons based on chemotherapy cycle length and this successfully accounted for the heterogeneity present in this pairwise comparison. The identification of heterogeneity, at this stage, in one or more pairwise comparisons can determine whether FTE or RTE models are used in step 4. If the source of heterogeneity cannot be identified or accounted for then either RTE will need to be used or it will be unsuitable to use this comparison in a NMA, particularly if removing the pairwise comparison exhibiting heterogeneity means that a FTE model can be used. Any covariates identified, during this step, as potentially causing heterogeneity should be considered in steps 7 to 9.

The PH assumption can be assessed in each trial by plotting the log cumulative hazard against log time for each treatment and looking for parallel lines, and also by plotting the Schoenfeld residuals. A chi-squared test based on the Schoenfeld residuals for each trial can be conducted to determine if there is any evidence of non-PH. Additionally, if each trial is independent of each other than adding together the value of the chi-squared statistic for each trial will provide an overall test statistic with degrees of freedom equal to the number of trials in the MA. Alternatively, the PH assumption can be assessed across all trials in the pairwise comparison by fitting a MA model including a treatment-ln(time) interaction.

In addition, publication bias can also be assessed, in this step, using Egger's test (if enough trials are included) and visually through the use of contour enhanced funnel plots. The aim of this step is to identify whether any adjustments for heterogeneity, such as RTE, or violation of the PH assumption will be required when fitting the NMA model. Once this is achieved, move on to the next step.

Step 3

As part of the process of fitting a NMA model it is important that the treatment parameterisation satisfies the consistency equations. The number of treatment parameters should be one less than the number of treatments in the network (see Section 4.4 and Appendix E for more details). The network diagram can help you decide which treatment parameters need to be directly estimated in the NMA model and which will be calculated as a contrast through the consistency equation. If the network consists of three treatments with all treatments directly compared, such as the lung cancer network, it is recommended that the largest treatment node is chosen as the reference treatment. Alternatively, the network diagram might identify a common reference treatment, such as in the cervical cancer network where RT was the only treatment directly compared with all the other treatments in the network. Including the arrows showing the direction of treatment effects on the network diagram can also indicate a common reference treatment. Arrows showing the direction of treatment effects can be particularly useful when a network does not have an overall reference treatment or includes multi-arm trials.

Step 4

A NMA model without any covariates can be fitted using both FTE and RTE and monitoring the DIC. The choice of FTE or RTE can be informed by the presence of heterogeneity or publication bias in step 2. If heterogeneity is present the RTE model should be used as this increases the variability around the point estimate to reflect the heterogeneity. Alternatively, both FTE and RTE models can be fitted. The RTE model gives more weight to smaller studies than the FTE model. Therefore, a difference in the treatment effect estimates between the FTE and RTE models can indicate publication bias and small study effects. Comparison of the DIC between models will determine which model fits the data the best although small differences (i.e. < 10) should not be over-interpreted and simpler models should be chosen where they can be.

A graphical method of comparing all the treatments in a network is to rank the treatments in terms of efficacy and display the results in a graph (such as Figure 4.7). This is readily done in WinBUGS. Alternatively, the treatment effect estimates for each pairwise comparison in the network can be presented as log hazard ratios and 95% credible intervals in a table. The size and complexity of the table will depend on the size of the network and for very large networks this might not be feasible.

Step 5

The network should be assessed for evidence of inconsistency. To visualise this, it is useful to present the model results as a forest plot with the network, direct and indirect evidence separated out (as shown in Figure 4.5). If both FTE and RTE models were fitted then this immediately shows whether the FTE and RTE models are consistent with each other. Different treatment effects between the FTE and RTE models can indicate small study effects or publication bias. Inconsistency can also be assessed by including an inconsistency parameter for each treatment loop and re-fitting the model including the additional parameters (Subsection 4.4.3). It is recommended that initially a NMA model is fitted with inconsistency parameters for all treatment loops. Any non-significant inconsistency parameters can then be removed and the NMA model re-fitted. If inconsistency is present in the network then an inconsistency parameter can be used in all further models. Treatment loops with inconsistency parameters are reduced to the direct evidence only

and therefore do not contribute to the across trial information in the network.

Step 6

In step 2 the assumption of PH was assessed within each trial and within each pairwise comparison. To check the correct conclusion about the PH assumption was made in step 2 the network can be assessed for evidence of non-PH. The PH assumption can be assessed through the inclusion of a treatment-ln(time) interaction in the NMA model. Allowing the treatment-ln(time) interactions to vary by trial results in Bayesian shrinkage estimates of the departures from PH in each trial which reduces the likelihood of over-interpreting departures from PH (Subsection 4.4.1). Plotting the treatment-ln(time) interactions for each trial allows visual assessment of the variation between trials (such as Figure 4.6). If the treatment-ln(time) interactions are found to be statistically significant then the assumption of PH is violated.

If the PH assumption is violated all future models will need to take non-PH into account. Fitting a NMA model with a treatment-covariate interaction and a treatment-ln(time) interaction is not straight forward. The log cumulative hazard is calculated by integrating the hazard function over all values of time t. However, when a treatment-ln(time) interaction is included in the NMA model the hazard function itself now includes the treatment-ln(time) interaction so integrating over all values of time t is not as straight forward. The result of this is that the log hazard ratio will no longer be a suitable effect measure and restricted mean survival time should be considered instead (Wei et al., 2015). This aspect has not been explored in detail in this thesis and further work is required to understand how best to carry out, interpret and present such analyses. Therefore, the rest of this framework assumes that the PH assumption is appropriate.

Step 7

Prior to including a covariate in the NMA model, consider the distribution of the covariate of interest in each trial. Are there any trials where some patients have missing covariate data? Are there any trials where all patients have missing covariate data? Is the covariate continuous or categorical? Can a linear effect between the groups of an ordered categorical

covariate be assumed? If there are trials with missing covariate data a distributional assumption will be required to impute the missing covariate data. In trials where the covariate is observed the distribution of the observed values can inform the decision of which distributional assumption to use. In this step, you should also determine which covariate value will be considered as the reference value. With a continuous covariate it is recommended that the covariate is centered on its mean value, so that this becomes the reference value.

Step 8

A covariate can be included in a NMA model as a common effect, a fixed trial-level effect or random trial-level effect. A common effect of the covariate assumes that the covariate has the same effect in all trials. A fixed trial-level effect of the covariate estimates an effect of the covariate for each trial. A random trial-level effect of the covariate allows the effect of the covariate to differ in each trial, but assumes that the coefficients for each trial come from a common (typically normal) distribution. At this stage in the framework it is important to think about which assumption makes the most sense for the network. Assuming a common effect of a covariate is only appropriate if the distribution of the covariate is the same in every trial in the network. Smaller studies can give more extreme estimates therefore a fixed trial-level or random trial-level effect of the covariate will be more appropriate, than a common effect, in networks with trials of varying size. The choice of FTE or RTE should be informed by previous steps such as the presence of heterogeneity from step 2 or the DIC from step 4. By the end of this step, you should know which assumptions about patient-level covariates you want to make and how to deal with missing covariate data.

Step 9

In this step, first fit the NMA models including patient-level covariate, identified in step 8. It is important to consider the results of these models, as this will help decide whether it is worth fitting a NMA with treatment-covariate interactions and what assumptions might be sensible.

The treatment-covariate interactions can be fitted as either common effects or random trial-level effects. A common effect of the treatment-covariate interaction assumes that

the treatment-covariate interaction has the same effect in all trials. A random trial-level effect of the treatment-covariate interaction allows the effect of the treatment-covariate interaction to differ in each trial but assumes that the coefficients for each trial come from a common (typically normal) distribution. The choice of assumption for treatment-covariate interactions should be based on what makes the most sense given the distribution of the covariate within and across trials and how similar the trials are in terms of the patient populations. For example, a common treatment-covariate interaction should be used if the assumption that the effect of the covariate on treatment is the same across all trials is reasonable. This is most likely to be reasonable when the distribution of the covariate is similar across all trials and all trials have the same patient population.

Once again, the choice of FTE or RTE should be informed by previous steps such as the presence of heterogeneity from step 2, the DIC from step 4 and the results of the NMA models including patient-level covariates. The choice of common effect, fixed trial-level effect or random trial-level effect for the covariate can be decided based on the considerations from step 8.

By the end of this step important decisions from previous steps, such as the parameterisation of the NMA model and how to deal with missing covariate data, combined with the decisions on the appropriate assumptions for the covariate and treatment-covariate interactions should make fitting a NMA model with treatment-covariate interactions in steps 10 and 11 a straight forward process.

Step 10

Treatment-covariate interactions can be fitted in two ways: within and across trial information combined or within and across trial information separated. NMA uses both within and across trial information therefore it is recommended that both of these models are fitted so that the consistency between the within and across trial information can be assessed. This can also allow for the assessment of ecological bias.

First, fit the treatment-covariate interaction with the within and across trial information combined using the appropriate distribution assumption for the missing covariate data. Models can be fitted as described in Subsection 6.6.1. Step 11 requires the mean value of the covariate for each trial. It is easy and simple to calculate, monitor and retain this whilst fitting

the NMA treatment-covariate interaction model with the within and across trial information combined. Within this model monitor the mean value of the imputed values for each trial where there are any missing covariate values. WinBUGS will only monitor the mean value of the imputed values so in a trial where only some of the patients have missing values of the covariate a weighted average of the mean observed and mean imputed value of the covariate is needed.

Second, a sensitivity analysis in which patients with missing covariate data are excluded should be conducted. The results of this model should be compared for consistency with the results of the model including the patients with missing covariate data to check the imputation of the missing covariate data has been handled correctly.

Step 11

To fit a NMA model with treatment-covariate interactions where the within and across trial information is separated, the mean value of the covariate must be calculated for all trials. In step 10 when the NMA model with treatment-covariate interactions with the within and across trial information combined was fitted the mean value of the covariate should have been calculated and retained for each trial (with some missing covariate data). For trials with only some missing covariate data the weighted average of the mean observed value and mean imputed value of the covariate can be used as the trial mean value in the NMA model with the within and across trial information separated. In trials where all patients have missing covariate values the mean covariate value from the imputed values, taken directly from the WinBUGS output, can be used as the trial mean value in the model with the within and across trial information separated. Models can be fitted as described in Subsection 6.6.2.

A sensitivity analysis in which patients with missing covariate data are excluded should be conducted. The results of this model should be compared for consistency with the results of the model including the patients with missing covariate data to check the imputation of the missing covariate data has been handled correctly.

Step 12

After fitting the NMA models including treatment-covariate interactions it is important that the results are displayed clearly so that an assessment of the agreement between the within and across trial information can be made. This can be achieved by plotting the treatment-covariate interaction parameter estimates from the combined and separate models alongside each other. For the cervical cancer network this is illustrated in Figure 6.3. This provides an easy visual assessment of the agreement between the estimates of the interactions from the two models. Based on this it should be clear whether combining the within and across trial information is a sensible idea. For example, for the cervical cancer network in Figure 6.5, the difference between the within and across trial information makes it easy to decide that, in this network, the within and across trial information should remain separated. As a useful rule of thumb, agreement is suggested between the within and across trial information from the FTE model, if the within trial information is within half a standard error of the across trial information.

Log hazard ratios along with 95% credible intervals can be presented in tables. Note that in a NMA model with treatment-covariate interactions the β parameters on their own do not have a useful interpretation. Treatment effects should be presented separately for each level of the covariate. It is useful to present the combined estimates of the within and across trial information alongside the separated estimates of the within and across trial information. This is illustrated for the cervical cancer network in Table 6.2. If treatment-covariate interactions are present, graphs ranking the treatments for each level of the covariate can be used as a visual aid for determining the most effective treatment for each level of the covariate. The use of appropriate graphs and tables will aid the comparison of the NMA models including treatment-covariate interactions with the pairwise MA models to ensure that the results are sensible and in line with what might have been expected.

Guidance on improving convergence in WinBUGS

Additional practical things to consider to improve convergence if using WinBUGS to fit NMA models:

1. Set sensible initial values for all parameters and use parameter estimates (i.e. the

posterior distribution means) from previous models as starting values in more complex models. This makes the model more likely to run and more likely to return appropriate estimates.

2. Run at least two parallel chains and check they give comparable results. This provides confidence that the MCMC process has converged.

3. Use a random number seed, so you can re-run the model and obtain the same results at a later date.

4. Check the distribution of parameter estimates using histograms. If the model has converged properly then the posterior distributions of the parameter estimates should produce normally distributed histograms.

5. Check the parameter estimates and credible intervals to ensure convergence and sensible results. Wide credible intervals can indicate that the model has not converged properly. Convergence can also be assessed by comparing the mean and the median of the posterior distributions. Disagreement can indicate a skewed posterior distribution which itself suggests that more MCMC iterations may be required for convergence.

6.9 Application of practical framework for one-step IPD NMA with treatment-covariate interactions to the lung cancer network

In this section I illustrate the application of my proposed practical framework for treatment-covariate interactions described in Section 6.8 to the lung cancer network. In this section, I include all 43 lung cancer trials as I did previously in Chapter 4.

Step 1: The network diagram for the lung cancer network is presented in Figure 3.3. The line thickness shows that the greatest amount of direct evidence is available for the RT v Seq CT comparison and the least amount of direct evidence for the Seq CT v Con CT comparison. All treatments are connected to at least one other treatment in the network. Based on the availability of the IPD it was decided that both stage of disease and performance status would be considered for inclusion in a NMA model with treatment-covariate interactions.

Step 2: Following the steps suggested in the framework, all pairwise treatment comparisons in the lung cancer network were assessed for evidence of heterogeneity and violation of the PH assumption in Subsection 3.3.2 and Subsection 3.4.2. Heterogeneity was identified in the RT v Seq CT comparison. Baseline characteristics of all trials were compared in Table B.1. The average age at randomisation in one trial was identified as being younger than the rest of the network. Further investigation into this trial showed that other baseline characteristics such as performance status also differed to the rest of the network population. Discrepancies between the IPD provided and the trial publication, which could not be resolved, resulted in the exclusion of the trial from all further analyses (Subection 3.4.2). Excluding this trial removed the heterogeneity from the RT v Seq CT comparison. Therefore, I concluded that there was no evidence of heterogeneity or non-PH in any of the pairwise comparisons (Subsection 4.3.2).

Step 3: Treatment parameterisation for the lung cancer network was discussed in Section 4.4. The network diagram in Figure 3.3 indicated that the treatment with the greatest number of randomised patients was RT and therefore I chose this as the reference treatment. I included the RT v Seq CT and RT v Con CT treatment contrasts as parameters in the NMA model resulting in the treatment effect for Seq CT v Con CT being estimated through the consistency equation.

Step 4: A one-step IPD Royston-Parmar NMA model was fitted to the lung cancer network using both FTE and RTE with the results presented in Table 4.4. The DIC provided no evidence that the RTE model was a better fit to the data than the FTE model (Section 4.7). There was no difference between the treatment effect estimates from the FTE and RTE models confirming no evidence of publication bias or small study effects in this network. In both the FTE and RTE models the treatment effects for Con CT compared to RT and Seq CT compared to RT were statistically significant with the FTE model suggesting an improvement in overall survival of 16% (LogHR=-0.179, 95% CrI: -0.248, -0.111) and 10% (LogHR=-0.102, 95% CrI: -0.164, -0.041), respectively. The treatment effect for Con CT compared to Seq CT was not statistically significant. Overall, Con CT was ranked as the most effective treatment, Seq CT the second most effective treatment and RT as the least effective treatment.

Step 5: Inconsistency was assessed in Subsection 4.7.2. There was a suggestion of inconsistency in the lung cancer network. Figure 4.8 showed differing point estimates for

the direct and indirect evidence with the credible intervals only just overlapping. However, when an inconsistency parameter was included in the NMA model this was not statistically significant. Cochran's Q statistic showed no evidence of inconsistency between designs. Note, Cochran's Q statistic here differs from that presented in Section 5.2 as the multi-arm trial was excluded throughout Chapter 5.

Step 6: In step 2 there was no evidence of non-PH in any of the pairwise comparisons. The non-PH assumption was assessed across the lung cancer network in Subsection 4.7.1. The treatment-ln(time) interaction terms and the approximate global Wald test were not statistically significant confirming that the PH assumption is appropriate in the lung cancer network.

Step 7: Two covariates, stage of disease and performance status, were considered for inclusion in the NMA model separately. Both stage of disease and performance status were considered as continuous covariates.

A linear effect of stage of disease was assumed which could take the values 0 = stages 1A-2A, 1 = stage 2B, 2 = stages 3A-4A. 1383 patients (18.3%) had missing data. Thirty-two of forty-three trials had at least one patient with missing data. Eight of these trials had missing data for all patients.

Performance status is an indicator of how well a patient is and how much mobility they have. The World Health Organisation (WHO) performance status scale runs from 0 to 5 with 0 representing a patient who is fully active and able to carry out all pre-disease activities without restriction and 5 representing a patient who is dead. Performance status is subjective and recorded by a clinician at the time of randomisation to the trial. A linear effect of performance status was assumed which could take the values 0 = WHO PS 0, 1 = WHO PS=1, 2 = WHO PS 2-3. 236 patients (3.1%) had missing data. Twenty of forty-three trials had at least one patient with missing data. Two of these trials had missing data for all patients.

In both cases missing data was imputed from a normal distribution, with mean μ and precision T, which was truncated through the use of the 'I' function in WinBUGS to restrict missing covariates to take values between 0 and 2. For μ a vague normal prior distribution with mean 0 and precision 0.001 was chosen. For T a uniform distribution ranging from 0.1 to 10 was chosen.

Step 8: In step 4, the DIC provided no evidence that the RTE model was a better fit to the data than the FTE model therefore both FTE and RTE models were considered when stage of disease and performance status were considered for inclusion as covariates. A common effect of either covariate appears to be inappropriate as the distribution of the covariates varies across trials and the network includes trials of varying sizes. However, to illustrate the models in action all three assumptions for the covariate were included. For each covariate three models were fitted: FTE with common effect of the covariate, FTE with fixed trial-level effect of the covariate and RTE with random trial-level effect of the covariate.

Step 9: For stage and PS, following the methods described in Subsection 6.3.1, three models were fitted to the lung cancer network: FTE with common effect of the covariate, FTE with fixed trial-level effect of the covariate and RTE with random trial-level effect of the covariate. The results are presented in Table 6.3.

Stage was statistically significant in both the FTE with common effect of stage and RTE with random effect of stage models. In the FTE with common effect of stage model each increase in stage of disease represented an 18% increase in the risk of death (LogHR=0.167, 95% CrI: 0.115, 0.219). In the RTE with random trial-level effect of stage model each increase in stage of disease represented a 22% increase in the risk of death (LogHR=0.197, 95% CrI: 0.129, 0.267). The difference between the FTE with common effect of stage model and the RTE with random trial-level effect of stage model could be, in part, evidence of larger effects in the smaller studies. In all three models the treatment effect for Seq CT compared to RT and Con CT compared to RT remained statistically significant with the inclusion of stage of disease as a covariate.

Performance status was also statistically significant in both the FTE with common effect of stage and RTE with random trial-level effect of stage models. As performance status increased, indicating a deterioration in health status, the risk of death increased by 36% (LogHR=0.311, 95% CrI: 0.268, 0.354) in the FTE with common effect of stage model and 41% (LogHR=0.347, 95% CrI: 0.278, 0.418) in the RTE with random trial-level effect of stage model. In all three models, despite the presence of performance status as a covariate, the treatment effect for Seq CT compared to RT and Con CT compared to RT remained statistically   significant.

In all models presented in Table 6.3 the treatment effects for all three pairwise treatment

comparisons remained consistent with the results of the FTE and RTE models without co-variates presented in Table 4.4.

The assumption of a common effect of stage or PS appeared to be inappropriate as the distribution of these covariates varies across trials. Therefore going forward I will consider the covariate effect to be a fixed trial-level effect. From step 4 there was no evidence of a difference in DIC between the FTE and RTE models. Therefore, I will continue to consider both FTE and RTE in the treatment-covariate interaction models.

Treatment-covariate interactions can be fitted as either common effects or random trial-level effects. To illustrate the models from Subsection 6.6.1 and Subsection 6.6.2 both assumptions will be considered. For the lung cancer network two models will be fitted for each covariate: FTE with fixed trial-level effect of covariate and common effect of treatment-covariate interactions and RTE with fixed trial-level effect of covariate and random trial-level effect of treatment-covariate interactions.

Table 6.3: Lung cancer covariate model results. Values are log hazard ratios and 95% credible intervals. PS = performance status, RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy, FTE = fixed treatment effect, RTE = random treatment effect, N/A = not applicable.

	RT v Seq CT	RT v Con CT	Seq CT v Con CT	Covariate effect
FTE + Common Stage	-0.105 (-0.169, -0.043)	-0.180 (-0.247, -0.113)	-0.075 (-0.154, 0.004)	0.167 (0.115, 0.219)
FTE + Fixed Stage	-0.108 (-0.176, -0.038)	-0.177 (-0.247, -0.107)	-0.069 (-0.153, 0.014)	N/A
RTE + Random Stage	-0.105 (-0.197, -0.007)	-0.169 (-0.274, -0.066)	-0.063 (-0.190, 0.065)	0.197 (0.129, 0.267)
FTE + Common PS	-0.098 (-0.161, -0.034)	-0.182 (-0.250, -0.113)	-0.084 (-0.162, -0.005)	0.311 (0.268, 0.354)
FTE + Fixed PS	-0.101 (-0.165, -0.036)	-0.198 (-0.266, -0.133)	-0.097 (-0.178, -0.017)	N/A
RTE + Random PS	-0.098 (-0.192, -0.002)	-0.179 (-0.288, -0.072)	-0.081 (-0.212, 0.050)	0.347 (0.278, 0.418)

Step 10: NMA models with treatment-covariate interaction with within and across trial information combined were fitted using stage and performance status as the covariates of interest using the methods described in Subsection 6.6.1. In all models a fixed trial-level effect of the covariate was used. The results are presented in the first and third columns of Table 6.4.

When stage takes the value zero, representing stage 1A-2A, the treatment effect for Con CT compared to RT remains statistically significant in both the FTE and RTE combined models. The RTE combined model suggests a 30% improvement in overall survival for a patient with stage 1A-2A disease receiving Con CT compared to a patient with stage 1A-2A disease receiving RT (LogHR=-0.356, 95% CrI: -0.588, -0.123). When stage is zero, the treatment effect for Seq CT compared to RT is not statistically significant in either model. There is no evidence of an interaction between Seq CT and stage of disease or Con CT and stage of disease in either of the models.

A global Wald test on the treatment-stage parameter estimates from the FTE combined model provided χ^2=1.365 on 2 df with p=0.505 and from the RTE combined model χ^2=3.059 on 2 df with p=0.217. This suggests that none of the interaction terms are statistically different from zero and therefore that the effect of stage of disease does not vary by treatment.

When PS takes the reference value 0, the treatment effect for Con CT compared to RT remains statistically significant in both the FTE and RTE combined models. The RTE combined model suggests a 16% improvement in overall survival for a patient with PS 0 receiving Con CT compared to a patient with PS 0 receiving RT (LogHR=-0.176, 95% CrI: -0.326, -0.037). When PS is zero, the treatment effect for Seq CT compared to RT remains statistically significant in the FTE combined model only. There is no evidence of an interaction between Seq CT and performance status or Con CT and performance status in either of the models.

A global Wald test on the treatment-performance status parameter estimates from the FTE combined model provided χ^2=2.191 on 2 df with p=0.334 and from the RTE combined model χ^2=0.427 on 2 df with p=0.808. This suggests that none of the interaction terms are statistically different from zero and therefore that the effect of performance status does not vary by treatment.

Table 6.4: Lung cancer treatment-covariate interaction model results. Reference level for stage is stages 1A-2A. Reference level for PS is WHO PS 0. RT = radiotherapy, Seq CT = sequential chemotherapy, Con CT = concomitant chemotherapy, FTE = fixed treatment effect, RTE = random treatment effect, PS = performance status.

	FTE Combined	FTE Separate	RTE Combined	RTE Separate
RT v Seq CT	−0.115 (−0.269, 0.042)	−0.307 (−0.581, −0.034)	−0.126 (−0.298, 0.047)	−0.262 (−0.738, 0.229)
RT v Con CT	−0.282 (−0.484, −0.096)	−0.122 (−0.493, 0.222)	−0.356 (−0.588, −0.123)	−0.171 (−0.919, 0.752)
Seq CT v Con CT	−0.167 (−0.411, 0.065)	0.184 (−0.238, 0.611)	−0.230 (−0.475, 0.004)	0.091 (−0.763, 0.859)
Seq CT - stage combined	0.007 (−0.109, 0.120)		0.042 (−0.093, 0.182)	
Con CT - stage combined	0.076 (−0.046, 0.205)		0.137 (−0.017, 0.291)	
Seq CT - stage within		−0.024 (−0.160, 0.104)		0.026 (−0.118, 0.164)
Con CT - stage within		0.106 (−0.047, 0.245)		0.124 (−0.048, 0.278)
Seq CT - stage across		0.184 (−0.082, 0.436)		0.133 (−0.325, 0.558)
Con CT - stage across		−0.126 (−0.417, 0.161)		−0.114 (−0.736, 0.368)
RT v Seq CT	−0.097 (−0.193, −0.004)	−0.196 (−0.379, 0.014)	−0.074 (−0.200, 0.052)	−0.261 (−0.563, 0.062)
RT v Con CT	−0.136 (−0.231, −0.038)	−0.136 (−0.343, 0.092)	−0.176 (−0.326, −0.037)	0.014 (−0.359, 0.385)
Seq CT v Con CT	−0.039 (−0.153, 0.078)	0.060 (−0.149, 0.284)	−0.102 (−0.270, 0.060)	0.276 (−0.127, 0.682)
Seq CT - PS combined	0.008 (−0.097, 0.113)		−0.030 (−0.151, 0.092)	
Con CT - PS combined	−0.085 (−0.205, 0.034)		0.024 (−0.143, 0.197)	
Seq CT - PS within		−0.007 (−0.117, 0.105)		−0.043 (−0.167, 0.084)
Con CT - PS within		−0.077 (−0.205, 0.050)		0.047 (−0.125, 0.225)
Seq CT - PS across		0.163 (−0.141, 0.434)		0.247 (−0.210, 0.672)
Con CT - PS across		−0.097 (−0.454, 0.244)		−0.327 (−0.939, 0.300)

Step 11: NMA models with treatment-covariate interaction with within and across trial information separated were fitted using stage and performance status as the covariates of interest using the methods described in Subsection 6.6.2. The results are presented in the second and fourth columns of Table 6.4.

When the reference value zero for stage is taken and the mean stage value is zero then the treatment effect for Con CT compared to RT is not statistically significant in either the FTE or RTE separate models. However, the treatment effect for Seq CT compared to RT remains statistically significant in the FTE separate model only (LogHR=−0.307, 95% CrI: −0.581, −0.034). There is no evidence of an interaction between Seq CT and stage of disease or Con CT and stage of disease in either of the models indicating that the effect of stage of disease does not vary by treatment.

The treatment effects for RT v Seq CT and RT v Con CT have changed considerably compared to the FTE combined and RTE combined models. This is because the within and across trial information are not consistent with each other. From the FTE separate model the log hazard ratio of the within trial information for Con CT is 0.106 (95% CrI: −0.047, 0.245) and the log hazard ratio for the across trial information is −0.126 (95% CrI: −0.417, 0.161). For this network the within and across trial information should remain separated and the across trial information could be subject to ecological bias.

When the reference value zero for PS is taken and the mean PS value is zero then the treatment effect for Con CT compared to RT is not statistically significant in either the FTE or RTE separate models. The treatment effect for Seq CT compared to RT is not statistically significant in either model. There is no evidence of an interaction between Seq CT and performance status or Con CT and performance status in any of the models indicating that the effect of performance status does not vary by treatment.

The treatment effects for RT v Seq CT and RT v Con CT have changed compared to the FTE combined and RTE combined models. This is because the within and across trial information are not consistent with each other. From the FTE separate model the log hazard ratio of the within trial information for Seq CT is −0.007 (95% CrI: −0.117, 0.105) and the log hazard ratio for the across trial information is 0.169 (95% CrI: −0.157, 0.458). For this network the within and across trial information should remain separated. The exception to this is the RT v Con CT comparison in the FTE separate model. The
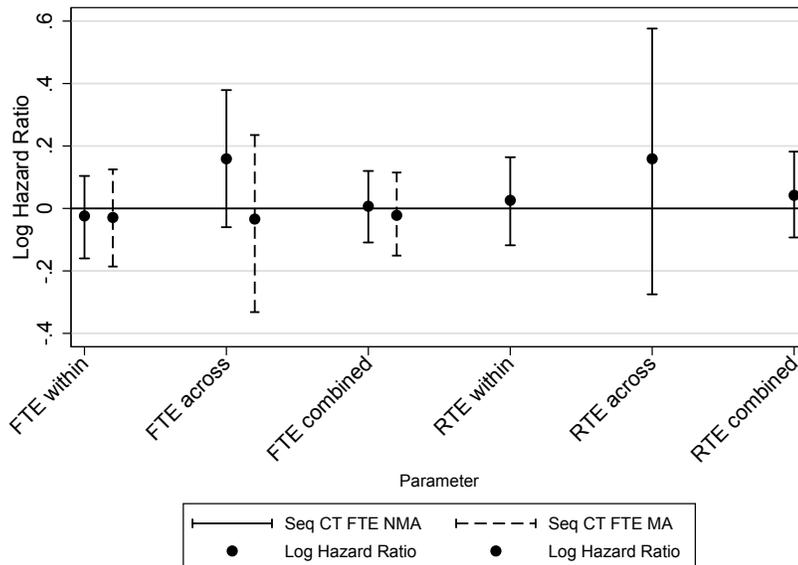
Figure 6.6: Treatment-stage interaction parameter estimates for RT v Seq CT. FTE = fixed treatment effect, RTE = random treatment effect, NMA = network meta-analysis, MA = meta-analysis. Solid lines represent NMA estimates. Dashed lines represent MA estimates.

point estimate of the treatment effect at the reference value zero and mean PS value zero is similar to the FTE and RTE combined models with the RTE combined model having a slightly larger credible interval, as expected. Therefore, there is only a small difference between the within and across trial information for the RT v Con CT comparison in the FTE separate model, as illustrated in Figure 6.9.

Step 12: Following the framework advice, to aid the decision making process for whether the within and across trial information from the treatment-covariate interactions should be combined or separated I plotted the parameter estimates for the treatment-stage interactions from the FTE and RTE models for the two main treatment comparisons, RT v Seq CT and RT v Con CT, in Figure 6.6 and Figure 6.7. Alongside the NMA estimates, the MA estimates from a FTE model are also presented. In Figure 6.6 and Figure 6.7 the within and across trial information differ to each other. In both of these cases it may be more appropriate to present the treatment-stage interactions with the within and across trial information separated. Note that the difference in the across trial information between the MA and NMA model, particularly for the RT v Seq CT comparison, suggests that some across trial information is gained from the network for this comparison.
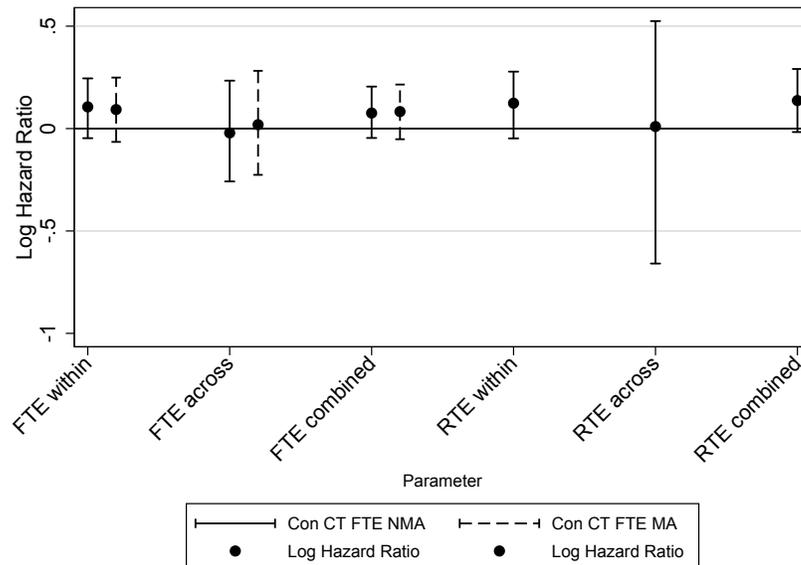
Figure 6.7: Treatment-stage interaction parameter estimates for RT v Con CT. FTE = fixed treatment effect, RTE = random treatment effect, NMA = network meta-analysis, MA = meta-analysis. Solid lines represent NMA estimates. Dashed lines represent MA estimates.

The parameter estimates for the treatment-performance status interactions from the FTE and RTE models for the two main treatment comparisons, RT v Seq CT and RT v Con CT, are presented in Figure 6.8 and Figure 6.9. Alongside the NMA estimates, the MA estimates from a FTE model are also presented. In Figure 6.8, the RT v Seq CT comparison, there is a clear difference between the within and across trial information suggesting that the within and across trial information should be separated for this comparison and it may be at risk of ecological bias. In Figure 6.9, the RT v Con CT comparison, the within and across trial information are in agreement with each other and the combined estimate could be suitable for this comparison. In both graphs there is a difference between the across trial information from the NMA and the MA models. This difference suggests that some across trial information has been gained from the network for both comparisons.
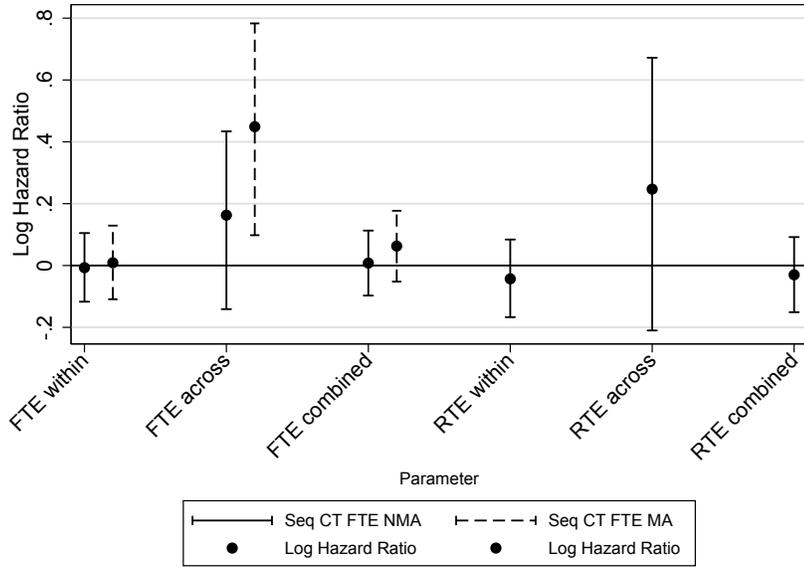
Figure 6.8: Treatment-performance status interaction parameter estimates for RT v Seq CT. FTE = fixed treatment effect, RTE = random treatment effect, NMA = network meta-analysis, MA = meta-analysis. Solid lines represent NMA estimates. Dashed lines represent MA estimates.
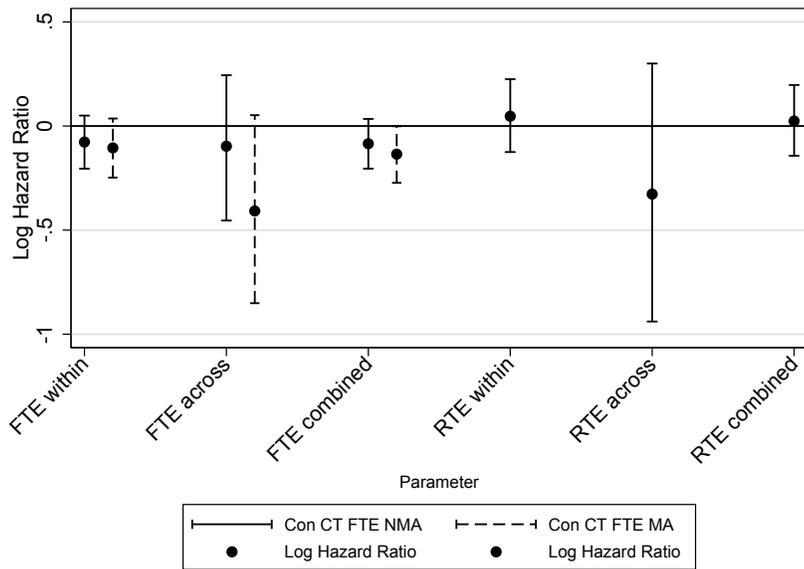
Figure 6.9: Treatment-performance status interaction parameter estimates for RT v Con CT. FTE = fixed treatment effect, RTE = random treatment effect, NMA = network meta-analysis, MA = meta-analysis. Solid lines represent NMA estimates. Dashed lines represent MA estimates.

6.10   Discussion

In this chapter I have presented the methodology for extending the one-step IPD Royston-Parmar NMA model, initially described in Chapter 4, to include patient-level covariates and treatment-covariate interactions. I have successfully applied this methodology to the cervical and lung cancer networks. In doing so, I have developed a practical framework for implementing this methodology which will be of use to other researchers looking to conduct their own NMA.

In the cervical and lung cancer networks I showed that stage of disease had a statistically significant effect on overall survival with advanced disease increasing the risk of death. However, despite the suggestion of a CTRT-stage interaction in the FTE combined model, neither network suggested any evidence for a treatment-stage interaction leading to the conclusion that stage did not modify the treatment effect. In the lung cancer network, the same conclusions were also reached with performance status. Worsening health increased the risk of death, but the effect of performance status did not differ by treatment. The treatment-covariate interaction models, for both networks, showed a difference between the within and across trial information for some of the comparisons and it was therefore most appropriate to separate out the within and across trial information. In particular, when PS was included in the treatment-covariate interaction model the difference in the NMA and MA estimates of the across trial information suggested that across trial information was gained from the network.

The cervical and lung cancer networks are small, well-connected networks with a lot of direct evidence. Despite this I was still able to show that some across trial information is gained when conducting a NMA. Information can only be gained from the network, relative to the direct estimate, where there is consistency. This is because the presence of an inconsistency parameter removes a treatment loop from contributing to the across trial information. In practice, not all networks will contain as much direct evidence as the cervical and lung cancer networks. Therefore, I would expect NMA to contribute a greater amount of across trial information in a consistent network where some treatment comparisons are only informed by a small amount of direct evidence. Further investigation into this could be conducted by removing some of the direct evidence from one of the comparisons in the lung cancer network and re-fitting the treatment-covariate interaction models.

This chapter provides a practical framework for conducting one-step IPD NMA of time-to-event data with treatment-covariate interactions. The framework was developed as a way of providing guidance to researchers, like myself before starting my PhD, who wish to embark on performing NMA and would like to include treatment-covariate interactions. There is currently no other framework available detailing the process for conducting a one-step IPD NMA with treatment-covariate interactions. The aim of this framework was not only to outline the process but to also provide useful and specific guidance. The framework is applicable to a number of situations but the guidance is specific to conducting one-step IPD NMA of time-to-event data using the Royston-Parmar approach. I think alongside the guidance this framework could result in more researchers conducting NMA with treatment-covariate interactions and also speed up the process for those already partly familiar with the process. The framework recommends that issues such as exploration of heterogeneity and inconsistency are considered early on in the process and are used to inform the decision making process surrounding which modelling assumptions are appropriate. Therefore, I think this framework has the potential to improve the conduct and analysis of NMA with treatment-covariate interactions.

The framework has not yet been evaluated in any formal process. It will be important to assess the user-friendliness and acceptability of the framework before encouraging widespread use. Asking other users to evaluate the framework would provide useful feedback and could potentially identify new steps that should be considered as part of the framework and areas where more guidance is required. This framework could be developed further by providing specific guidance for other situations such as binary or continuous outcomes. Furthermore, following the framework could, theoretically, result in a model that includes both inconsistency parameters to account for inconsistency and treatment-ln(time) interactions to account for non-PH. Both of these have been explored in varying detail throughout this thesis as singular items and, theoretically, a model could be specified to account for both simultaneously. However, in practice, I have not considered them together and therefore it is possible that there will be some practical limitations of whether such a complex model fitted in WinBUGS would actually result in convergence and also how it would be interpreted. Further work is needed here, particularly around the interpretation.

Conducting a NMA including treatment-covariate interactions requires the use of both within and across trial information. Each trial contributes to the within trial information

which is estimated using patient-level covariates. Meanwhile across trial information is estimated through the relation of the trial-level aggregated covariates (Hong et al., 2015). Combining within trial and across trial information can result in confounding and ecological bias. Ecological bias arises as the across trial information uses trial averages which are then used to draw conclusions about individuals. Therefore a large difference between the within and across trial information can suggest evidence of ecological bias (Donegan et al., 2013a; Hong et al., 2015). A large difference could also be due to different ranges of patient-level covariates within and across trials. For example, with a continuous covariate such as age it is feasible that average values of age could differ across trials, but within each trial the interaction between age and treatment is similar therefore the across trial information will contribute more to the interaction than the within trial information (Hong et al., 2015). By using across trial information the assumption has to be made that there is no unmeasured confounding in the network, but unfortunately this assumption will always be hard to test. There was evidence of ecological bias in both the cervical cancer and the lung cancer networks. Further investigation into this is needed.

In this chapter I have shown that the Royston-Parmar approach naturally allows the inclusion of continuous and categorical patient-level covariates and treatment-covariate interactions. The Bayesian setting allows covariates and interactions to have random effects while avoiding the awkward numerical integration needed to maximise the corresponding likelihoods, and naturally handles missing covariate data.

In Chapter 7 I provide a summary of this thesis, discuss problems that occurred and consider future work arising from this thesis.

7   Discussion

7.1   Motivation for thesis

This thesis was motivated by two ongoing trials (STAMPEDE and ICON8B) at the MRC CTU at UCL, which will both require the use of NMA methods in the future. The STAMPEDE trial is a multi-arm multi-stage trial assessing overall survival in men with prostate cancer. The trial started by considering five new treatment regimens but has since added more treatments whilst recruitment to the original new treatment regimens has stopped. The control arm will be compared directly to the new treatment regimens in patients recruited concurrently. New treatment regimens will not be compared directly. Therefore, NMA methods will be required to compare overall survival between new treatment regimens. ICON8B is a trial comparing overall survival in women with high-risk ovarian cancer. The trial is interested in answering a non-inferiority question comparing two treatment regimens. A trial directly answering this non-inferiority question was not possible because of the large number of patients required and the time that would be required for recruitment. Therefore ICON8B was designed to be incorporated in a NMA along with other ovarian cancer trials which will provide indirect evidence to supplement the direct evidence from ICON8B.

One of the big benefits of using NMA is that it can compare treatments never directly compared and allow for the ranking of treatments. Therefore, the most effective treatment for a particular disease area can be estimated. To be able to compare treatments never directly compared NMA uses both direct and indirect evidence. Treatment effect estimates will be robust if the direct and indirect evidence are in agreement. Inconsistency between the direct and indirect evidence can result in biased treatment effect estimates. Therefore, it is important that the consistency of the direct and indirect evidence is considered, and if inconsistency is present that it is accounted for within the NMA model. Hence, it is important that researchers are clear about the direct and indirect evidence, and whether they wish to combine them.

As survival rates in oncology trials increase patients are living longer after their initial diagnosis. Therefore there will be a greater number of patients still alive who could relapse. As patients relapse and receive second, third and fourth line treatments the PH assumption is

likely to be violated as overall survival will now be dependent not only on the randomised treatment but the second, third and fourth line treatments, which could differ between treatment arms. Therefore, the PH assumption in oncology trials with long term follow-up is not always appropriate.

In an era where targeted treatments are more widely available, identifying subgroups of patients most likely to respond to treatment is becoming increasingly important. At the design stage most single RCTs are not powered to detect clinically meaningful treatment-covariate interactions. Therefore, NMA which combines multiple trials has the potential to identify subgroups of patients most likely to respond to treatment. In order to address this issue NMA models which can explore treatment-covariate interactions, and a practical framework for applying them, are needed.

In recent years, NMA conducted in the Bayesian setting has been increasing in popularity (Sobieraj et al., 2013). Alongside this, the development of new user-friendly Bayesian software options could have the potential to further increase the popularity and practical utility of Bayesian NMA. Traditionally, time-to-event data has been analysed using the Cox PH model. However, fitting the Cox PH model using a Bayesian framework is computationally extremely (and sometimes infeasibly) intensive and the PH assumption may no longer be appropriate for oncology trials with long term follow-up. Therefore alternative methods for NMA of time-to-event data which can be implemented in a Bayesian setting are needed.

This thesis set out to show that when using the Royston-Parmar model with RCS, the Bayesian framework provides a natural, practical and flexible approach for NMA of time-to-event data.

7.2 Overview of thesis

This thesis started with a literature review of the statistical methodology available for conducting NMA with binary, continuous and time-to-event data. Methods for conducting NMA with binary or continuous data have been developed over the last 15-20 years with specific areas such as consistency, heterogeneity, bias and treatment-covariate interactions well developed. By contrast there is only a small pool of literature for conducting NMA with time-to-event data. To the best of my knowledge a Bayesian approach for the Royston-Parmar

model using natural cubic splines has not been considered previously in the NMA setting. Therefore this thesis will add to the small pool of methodology for conducting NMA of time-to-event data.

Before fitting a NMA model careful consideration of the data is required. In Chapter 3, I set out to conduct a thorough exploration of two networks of RCTs: one in cervical cancer and one in lung cancer. In Subsection 3.4.1 some evidence of heterogeneity in the RT v CT+RT comparison in the cervical cancer network was identified. In the lung cancer network, I identified one influential trial (Sharma et al., 2003) which was causing heterogeneity and publication bias. Further investigation into this trial showed that the trial population was inconsistent with the network population. The mean age for this trial was eight years younger and there were a higher proportion of patients with poor performance status. The reporting of this trial was also inconsistent with the IPD received. Therefore, I made the decision to exclude this trial from the lung cancer network.

In Chapter 4 I explored the implementation of the Royston-Parmar model in the Bayesian setting. The Royston-Parmar model is fitted using a RCS for each trial, which models the baseline log cumulative hazard individually for each trial. The RCS includes a number of interior knots which allow flexibility in the shape of the baseline log cumulative hazard. In the cervical cancer network I chose the location of the knots for each trial to ensure that the Royston-Parmar model was as close to the Nelson-Aalen estimate as possible. However, the parameter estimates were robust to the location of knots. Therefore in the lung cancer network I placed the interior knots at the 33rd and 67th percentiles of the uncensored survival times. Log hazard ratios arising from the Royston-Parmar model for each trial were compared to log hazard ratios from the Cox model to verify the choice of knot locations.

Initially I fitted a MA to each pairwise comparison in the two networks using the Royston-Parmar approach. All pairwise comparisons were assessed for evidence of heterogeneity and non-PH. Evidence of heterogeneity was first identified in the cervical cancer network in Chapter 3. In Subsection 4.3.1, this was successfully addressed by splitting the comparison into two comparisons based on the length of chemotherapy cycles. Despite this a small amount of heterogeneity did remain in the comparison with short chemotherapy cycles. Therefore, when fitting NMA models to the cervical cancer network I considered both FTE and RTE models and checked for consistency between the results of the two models.

I successfully applied the Royston-Parmar model to the NMA setting. In the NMA setting care must be taken to ensure that the model is parameterised correctly and treatment effects are in the right direction (Appendix E). Using arrows on the network diagram to illustrate the direction of treatment effects is helpful. By making the assumption of consistency across the network, the parameter estimates from the model allow estimation of the ranking of treatments in order of effectiveness. The Bayesian setting provides a natural method for doing this (Subsection 4.4.5). Heterogeneity was assessed in each network using Cochran's Q statistic. The Royston-Parmar model can be extended to assess the PH assumption through the inclusion of treatment-ln(time) interactions, and to assess inconsistency through the inclusion of an inconsistency parameter. NMA models including inconsistency parameters allow the direct and indirect treatment effect estimates to be separated out. Plotting them alongside the network treatment effect estimates provides an easy visual assessment of inconsistency in the network. Taken together with the fact that a Bayesian approach is a natural approach for random effect modelling, it is clear that this is a very flexible framework for NMA.

Applying the one-step IPD FTE Royston-Parmar NMA model to the cervical cancer network suggested a 33% improvement in overall survival (LogHR=-0.396, 95% CrI: -0.611, -0.185) for patients receiving CT+S compared to RT and a 19% improvement in overall survival (LogHR=-0.211, 95% CrI: -0.337, -0.087) for patients receiving CTRT compared to RT. The FTE model ranked CT+S as the most effective treatment with a probability of 93%, followed by CTRT which had a 91% probability of being the second most effective treatment. CT+S had 99.96% chance and CTRT 98.03% chance of being one of the top two most effective treatments in the network. As indicated earlier, there was some evidence of heterogeneity in the cervical cancer network. Therefore, both FTE and RTE models were considered when patient-level covariates were included. There was no evidence of non-PH or inconsistency.

Applying the one-step IPD FTE Royston-Parmar NMA model to the lung cancer network suggested 16% (LogHR=-0.179, 95% CrI: -0.248, -0.111) and 10% (LogHR=-0.102, 95% CrI: -0.164, -0.041) improvements in overall survival for Con CT and Seq CT compared to RT, respectively. The FTE model ranked Con CT as the most effective treatment with a probability of 97.22% and Seq CT as the second most effective treatment with a probability of 97.17%. Con CT had 100% chance and Seq CT 99.95% chance of being

one of the two most effective treatments. There was some evidence of heterogeneity in the lung cancer network. Therefore, both FTE and RTE models were considered when patient-level covariates were included. There was no evidence of non-PH or inconsistency.

In Chapter 5 I set out to explore methods for assessing inconsistency in NMA using the lung cancer network. In this chapter the lung cancer network included two-arm trials only. I started by exploring four methods for assessing inconsistency: Cochran's Q statistic (1954), the loop inconsistency approach (Bucher et al., 1997), the inconsistency parameter approach (Lu and Ades, 2006) and the net heat approach (Krahn et al., 2013). The net heat approach is a relatively new approach for assessing inconsistency in a network compared to the other three methods. It provides a graphical approach to assessing inconsistency, which uses a colour scale to indicate areas of inconsistency within the network. I applied all four methods to the lung cancer network. Cochran's Q statistic, the loop inconsistency approach and the inconsistency parameter approach all gave the same result. The net heat plot contradicted the other three approaches. Therefore I investigated the net heat approach in more detail.

I applied the underlying principles of the net heat approach to networks of varying size before considering a more generalised approach. For a three-treatment network, I showed that the net heat plot draws conclusions about inconsistency from a quantity that is a scaled and squared version of the inconsistency parameter. For a more general network, I further showed that this quantity was an unintuitive and imprecise measure of inconsistency which could be misleading. I proposed an alternative method for assessing inconsistency which I based on the significance of inconsistency parameters. I combined inference from formal statistical tests with displaying the results in a way which offers a visual aid for assessing inconsistency in a network.

An attraction of the Royston-Parmar model is that it can be easily extended to incorporate patient-level covariates and treatment-covariate interactions. The Royston-Parmar model can be easily extended to incorporate patient-level covariates, which can be modelled as common effects, fixed trial-level effects or random trial-level effects. However, in practice, there is often missing covariate data. In Section 6.4 I show that by using a Bayesian approach, missing covariate data can be readily accommodated by adding a distributional assumption to the NMA model to impute the missing covariate data. In both datasets, when stage of disease was included as a continuous and a categorical covariate overall

survival was shown to be lower in patients with the highest stage of disease. Therefore, stage of disease was considered for inclusion as a treatment-covariate interaction in the cervical cancer network. In the lung cancer network overall survival was also reduced in patients with increasing performance status. Therefore, both stage of disease and PS were considered for inclusion as treatment-covariate interactions in the lung cancer network.

Extending the one-step IPD Royston-Parmar NMA model to include treatment-covariate interactions requires combining within and across trial information. This enables information to be gained from the network and used to inform the treatment-covariate interactions. Treatment-covariate interactions were considered for inclusion as common or random trial-level effects. Models were presented in which the within and across trial information were combined and separated. Separating out the within and across trial information allowed for agreement between the two sources of information to be assessed. Displaying the within and across trial information parameter estimates graphically provides a visual assessment of the agreement between the within and across trial information. This can also be used to determine the extent to which the treatment-covariate interactions may be subject to ecological, or other, biases. Chapter 6 finished by pulling together all the aspects of this thesis into a twelve-step framework aimed at providing researchers new to the field of NMA with a step-by-step guide to conducting NMA with treatment-covariate interactions. Alongside the framework guidance is provided on the key considerations and assumptions required for each step. I illustrated this guidance through application to the lung cancer network.

In the cervical cancer network, with within and across trial information combined, there was some evidence that the effect of CTRT differed by stage of disease (treatment-stage interaction LogHR=0.194, 95% CrI: 0.032, 0.357). However, once the within and across trial information were separated this was no longer the case. For both networks, in some comparisons, the within and across trial information differed suggesting possible evidence of ecological bias. Therefore, in both networks, it was most appropriate to separate out the within and across trial information. After separating the within and across trial evidence, I concluded that in the cervical cancer network there was no evidence to suggest any treatment-stage interactions. Likewise in the lung cancer network there was no evidence to suggest any treatment-stage or treatment-PS interactions.

7.3 Methodological considerations

The first problem encountered during this thesis was fitting a Cox PH model in WinBUGS. To fit the Cox PH model in the Bayesian setting, each participant's row of data has to be repeated for each risk set they belong to. For example, in a trial of 1000 patients with 200 deaths there are 200 risk sets. If each patient on average belongs to 50 risk sets then the size of the data set is multiplied by 50. The result is a very large dataset, which makes the computations so intensive that they were infeasible even for the moderately sized cervical cancer network. To avoid this problem I moved on to consider the Royston-Parmar model using RCS for the log cumulative baseline hazard. RCS have the advantage over fractional polynomials that they are linear at each end, and so avoid the unexpected, and undesirable, end effects of fractional polynomials. The shape of a fractional polynomial at each end of the dataset, where there is often less information, tends to be dictated by what happens in the middle of the dataset. For example, the shape of a quadratic polynomial over the central 50% of data points is quite different to the shape for the 25% of data points at either end. RCS are linear at each end, avoiding these end effects and are therefore more likely to be appropriate for most networks. Using the Royston-Parmar model with RCS avoids both the computational issues of the Cox PH model and the problem of end effects which are encountered by fractional polynomials. Furthermore, the Royston-Parmar model can be extended to assess and take into account non-PH. Therefore, the Royston-Parmar model provides a flexible and practical method for conducting NMA of time-to-event data.

Conducting NMA with treatment-covariate interactions involves combining within and across trial information. Alongside my PhD, I contributed to a MRC CTU at UCL project which used a systematic review to support previous arguments (e.g. Riley (2008)) recommending using within trial information and only combining within and across trial information in an exploratory analysis. An invited revision of this manuscript is under review with the BMJ (Fisher et al., 2016). By definition, NMA uses both the within and across trial information. Using across trial information requires the assumption of no unmeasured confounding. Making this assumption allows information to be gained from the network to inform both treatment estimates and treatment-covariate interactions. Differences between the estimates of the within and across trial information can suggest ecological bias. It is therefore important that the within and across trial information for treatment-covariate interactions can be separated out. Separating out the within and across trial informa-

tion allows the influence of the across-trial information on the treatment-covariate interactions to be assessed and allows researchers to identify which data source is driving the treatment-covariate interactions. Then if researchers are satisfied they are consistent, the information can be combined for final inference. In this thesis I have shown (for the first time) how a Royston-Parmar model can be used in the Bayesian framework with IPD to allow separation of the within and across trial information.

The RTE models throughout this thesis were fitted using an inverse Wishart prior for the between-study variance-covariance matrix. It has been highlighted by Burke and Wei that a Wishart prior may not be the most appropriate choice of prior distribution (Burke et al., 2016; Wei and Higgins, 2013), however in the NMA setting where we have multiple treatments, and hence multiple random effects, there are few alternatives. A Wishart prior can become influential in the estimation of the between-study variance-covariance matrix and can lead to the overestimation of heterogeneity parameters particularly when the true heterogeneity is close to zero (Wei and Higgins, 2013). Instead of worrying about whether the Wishart prior is appropriate when true heterogeneity is close to zero, it's better to use empirical information to inform the prior, as argued by Turner (2012). Including empirical evidence in the prior distribution could result in a more realistic prior distribution for the between-study variance-covariance matrix, particularly when small numbers of trials are available (Turner et al., 2012).

In Chapter 5 two-arm trials were sufficient to understand and demonstrate the limitations of the net heat approach. In reality, there will be many situations in which a multi-arm trial is present. However, the presence of multi-arm trials generally adds another level of complexity. In the case of Bucher (1997) the method is not suitable for use with multi-arm trials because it assumes independence between treatment comparisons. However, the presence of multi-arm trials can sometimes allow us to identify the source of inconsistency in a network. Cochran's Q statistic (1954), Lu & Ades (2006) and the net heat approach (Krahn et al., 2013) can all be used in the presence of multi-arm trials. Furthermore, graphical methods, such as the net heat approach, offer potential visual aids which help identify, locate and quantify inconsistency. However, only the design-by-treatment interaction model (Higgins et al., 2012) and Cochran's Q statistic (1954) produce results that are independent of the parameterisation of the model (Efthimiou et al., 2016). When assessing inconsistency, the key difference is between the observed data and the fitted values from

the network. If there are two parameterisations of the network which give the same fitted values then they must both fit the data equally well. Therefore, if the two parameterisations say different things about the location of the inconsistency then there is no basis from the data to be able to choose between the two network parameterisations.

In this thesis I showed that the net heat approach can be misleading when only two-arm trials are considered. Given the additional complexity of including multi-arm trials, it is likely that the net heat approach would also be misleading in more complex networks. Therefore, alternative methods for visually displaying information which can help identify and locate inconsistency in networks are needed.

Although this thesis has focused on fitting a one-step IPD Royston-Parmar NMA model it would also be possible to use the Royston-Parmar approach in a two-step process. The first step would require fitting a Royston-Parmar model to each trial individually to obtain an estimate of the log hazard ratio and its corresponding standard error, which are then pooled together in the second step using standard MA methods. However, there are several benefits to using a one-step approach, particularly when IPD is available. A pooled effect of a covariate can only be considered in a one-step model: either by a fixed weighting of the trial specific baseline hazards, or by using a random effect model for the baseline hazards (i.e. the coefficients of the spline are random across trials). If non-PH are present in a network it is important that they are accounted for. IPD is essential for including treatment-ln(time) interactions in a NMA model.

An advantage of the Bayesian setting is that it allows for better predictions. However to do this it requires an appropriate baseline hazard. By setting the knot locations to be the same for all trials, the Royston-Parmar model can estimate the baseline hazard pooled across all trials. Another advantage of the Bayesian setting is that FTE models can easily be extended to RTE avoiding awkward numerical integration.

7.4 Literature review update

The literature search conducted in Section 2.1 was repeated on the 18th May 2016. Embase, Ovid MEDLINE and Ovid MEDLINE In-Process & Other Non-Indexed Citations electronic databases were searched for articles published since January 2015. This resulted in

some duplicate results but also ensured any publications not indexed on the 1st June 2015 were picked up. In addition PubMed was also searched for articles from 1995 using the same search strategy.

123 titles were identified from the search of Embase, Ovid MEDLINE and Ovid MEDLINE In-Process & Other Non-Indexed Citations electronic databases. Of these titles eight abstracts were reviewed and eight papers obtained in full. Searching PubMed resulted in 1054 titles of which 165 abstracts were reviewed. 117 papers had already been identified through the other databases leaving 48 abstracts to be reviewed. Of these abstracts 26 papers were obtained in full. A further update on the 30th September 2016 identified an additional 12 papers to obtain in full from the search of Embase, Ovid MEDLINE and Ovid MEDLINE In-Process & Other Non-Indexed Citations electronic databases and an additional 16 papers to obtain in full from the search of Pubmed.

Throughout my thesis I have used a contrast based approach to NMA which models the treatment effect for treatment A compared to treatment B. NMA can also be conducted using an arm based approach, such as that proposed by Hong (2016) and Zhang (2014). An arm based approach models the absolute effect of each treatment rather than the relative effect. To assess inconsistency in the arm based model Zhao (2016) proposed computing the posterior distribution of the discrepancy factor. To calculate the discrepancy factor between treatment A and treatment B the trials in the network are split into four groups based on whether they include treatment A, treatment B, both or neither. In trials including both treatments the direct evidence for A v B is calculated as the difference between the treatment effect for treatment A and the treatment effect for treatment B. The difference between the treatment effect for treatment A in trials only including treatment A and the treatment effect for treatment B in trials only including treatment B is calculated. In a three-treatment network this is the indirect evidence for A v B. The discrepancy factor is then calculated as the difference between these two quantities (Zhao et al., 2016). This is a similar concept to the contrast-based approach of node-splitting (Dias et al., 2010b).

In the setting of multiple continuous outcomes Hong (2015) introduced an arm-based IPD NMA model which can be extended to include treatment-covariate interactions. As I have done, in Chapter 6, they applied the framework of Riley (2008), for pairwise MA, to allow for the separation of the treatment-covariate interactions into within and across trial information, so that ecological bias can be monitored.

In practice, the scientific focus in a MA and NMA is often on comparing treatments to each other and therefore the contrast based approach offers a more natural way of conducting MA and NMA. The number of events observed in a trial often depends on the length of follow-up. Therefore when comparing two trials with different lengths of follow-up it makes more sense to compare the odds or hazards of events rather than additive changes in the actual probabilities, or hazards, of events. For example, the underlying hazard rates for treatment A and treatment B can be very different but the hazard ratio may not show such a large difference.

A method of identifying inconsistency in NMA that was not explored in Chapter 5 is the method of node-splitting (described in Section 2.4). The development of a programme implemented in R which automates the process of node-splitting and implements a decision rule which determines which comparisons require splitting could increase the popularity of this method (van Valkenhoef et al., 2016). However the authors acknowledge that assessment of heterogeneity and inconsistency remains a challenge as there is no clear distinction between the two concepts and it is always possible that one model may detect an inconsistency whereas another model detects high heterogeneity but not inconsistency.

NMA is commonly performed in a Bayesian framework using WinBUGS. As NMA has grown in popularity so have the number of software options available (Chambers et al., 2015; Lee, 2014; Nikolakopoulou et al., 2014; Sobieraj et al., 2013). A 2015 survey looking at alternative Bayesian software options to WinBUGS concluded that Stan (Carpenter et al., 2016) is the most promising alternative (Stephenson et al., 2015). The survey concluded that Stan allows model flexibility, user specification and integrates with R for high quality graphics. It also has a user manual and, in contrast to WinBUGS, provides useful error messages. Alongside this, the introduction of two new suites of commands for conducting NMA (White, 2015) and graphically displaying the evidence base, the assumptions and the results (Chaimani and Salanti, 2015) has made NMA in Stata (StataCorp, 2015) a much more attractive option for researchers who wish to use frequentist methods.

NMA can be used to inform the future research agenda identifying areas of the network that require further investigation in new RCTs (Nikolakopoulou et al., 2016). The quality of a NMA is only as good as the quality of the RCTs contributing to it. Therefore, as I have argued in this thesis, it is important that MA is conducted thoroughly exploring issues such as heterogeneity, non-PH and consistency before a NMA is performed. A recent

guidance paper on IPD MA provides guidance on how to recognise a well-designed and conducted IPD MA (Tierney et al., 2015). However, a recent review of indirect comparison methods using IPD found that better reporting was needed (Veroniki et al., 2016). The two sets of PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines published in 2015 focusing on reporting of IPD analyses (Stewart et al., 2015) and NMA (Hutton et al., 2015) should help to continue to improve reporting standards. To be able to inform the future research agenda the best possible reporting standards are required and it is therefore important that reporting guidelines are adhered to by researchers.

7.5   Future   research

Following on from my thesis and considering recent developments in NMA there are a number of areas in which I think future research should be focused. The first one involves developing new graphical methods for assessing inconsistency in NMA. The net heat approach plots Q^{diff} which, for three-treatment networks, is basically a scaled and squared version of the inconsistency parameter. Therefore a new graphical method could revolve around the inconsistency parameter from the Lu & Ades inconsistency model. As proposed in Chapter 5, the p-value of the inconsistency parameter for a treatment loop could be used to quantify the amount of inconsistency in a treatment loop. The approach proposed in Chapter 5 needs to be developed further to consider, visually, the best location for the square boxes which indicate the proportion of direct evidence for each treatment loop. The proposed approach also needs to be tested on a variety of network structures, including larger networks and networks with fewer direct comparisons. For large networks the development of graphical methods is more complicated and can potentially result in messy plots that are hard to understand. Perhaps a more novel approach, such as 3D graphs, should be considered.

In the lung cancer network, in Chapter 4 there was only a small amount of evidence to suggest inconsistency in the lung cancer network. When the multi-arm trial was excluded from the network in Chapter 5 inconsistency was identified. In Chapter 6, with the multi-arm trial included, there was a difference in the NMA estimate of the across trial information for the RT v Seq CT comparison and the estimate arising from a pairwise MA. Formal tests have low power for detecting inconsistency in a network. Instead, the cervical and lung cancer examples illustrate that graphical presentation shows the impact of

the non-randomised information more clearly. The difference in the MA and NMA across trial information in the lung cancer network leads to several questions which need to be addressed. Is it inconsistency in the network that drives the difference in the across trial information? How should inconsistency in a network be dealt with when fitting treatment-covariate interactions? What do the treatment-covariate interactions mean when inconsistency is present? How should the model be parameterised?

In my thesis the cervical and lung cancer networks did not show any evidence of non-PH. Throughout this thesis I have argued that the Royston-Parmar approach can be extended to include treatment-ln(time) interactions to account for non-PH. In the presence of non-PH the log hazard ratio would no longer be an appropriate effect measure and an alternative, such as the restricted mean survival time, would need to be considered. Restricted mean survival time has been considered for use in MA (Wei et al., 2015) but to date, it has not been considered in NMA. Further work is needed to put into practise the methods suggested within this thesis for dealing with non-PH and to extend these to scenarios which deal with multiple issues. For example, future work should consider the practicalities, parameterisation and interpretation of models that account for all or combinations of non-PH, inconsistency and treatment-covariate interactions.

IPD can be difficult, time-consuming and expensive to collect whereas AD can often be extracted from journal papers. Donegan (2013a) showed that collecting IPD for just a few trials can be beneficial for a NMA. For example, when exploring treatment-covariate interactions in a NMA, combining IPD and AD can lead to increased precision of the credible intervals. In recent years there has been a drive towards greater data sharing which could make it easier to obtain IPD in the future. However, until this becomes more widespread it seems likely that methods for combining IPD and AD will become increasingly popular. Donegan (2013a) and Saramago (2014) have shown that IPD and AD can be synthesised together. In both cases covariates can be included, with patient-level values used for IPD trials and trial mean values used for AD trials. However, PH can only be assessed using the equations presented in Subsection 4.4.1 in IPD trials. Synthesis of IPD and AD is particularly natural in the Bayesian framework, where random effects can be naturally included to accommodate the inevitable heterogeneity. Therefore, the Royston-Parmar approach, fitted using RCS, could provide a flexible method of synthesising IPD and AD for time-to-event outcomes, which avoids distributional as-

sumptions, and allows for a greater range of models to be fitted.

7.6 Conclusion

This thesis contributes to the development of methodology for NMA of time-to-event data in a number of ways. Firstly, by outlining the one-step IPD Royston-Parmar NMA model using RCS, and demonstrating that, when fitted in WinBUGS, it provides a flexible, practical approach for Bayesian NMA with time-to-event data. The one-step IPD Royston-Parmar NMA model can be easily extended to assess and account for non-PH and inconsistency. The one-step IPD Royston-Parmar NMA model can be easily extended to include patient-level covariates and treatment-covariate interactions. Missing covariate data can be accommodated in a straight forward manner within a Bayesian framework. Plotting direct, indirect and network treatment effect estimates alongside each other is a useful visual assessment of inconsistency. Plotting within and across trial information parameter estimates alongside each other is a useful visual assessment of ecological bias. Finally, this thesis provides a practical framework for applying the methodology offering advice and guidance to researchers to enable them to conduct their own NMA.

References

Achana, F., Dequen, P., Gray, L., Cooper, N. J., Abrams, K. R. and Owen, R. K. (2013) Network meta-analysis of multiple outcomes: A simulation study and application. *Value in Health*, **16**, A609.

Ades, A. E., Mavranezouli, I., Dias, S., Welton, N. J., Whittington, C. and Kendall, T. (2010) Network meta-analysis with competing risk outcomes. *Value Health*, **13**, 976–83.

Ades, A. E., Sculpher, M., Sutton, A., Abrams, K., Cooper, N., Welton, N. and Lu, G. (2006) Bayesian methods for evidence synthesis in cost-effectiveness analysis. *Pharmacoeconomics*, **24**, 1–19.

Alberti, W., Niederle, N., Budach, V., Konietzko, N. and Sack, H. (1990) Prospective randomized study comparing immediate radiotherapy, chemo- plus radiotherapy or delayed radiotherapy in non-small cell lung cancer. *Journal of Cancer Research and Clinical Oncology*, **116(Part I)**, 503.

Altman, D. G. and Royston, P. (2006) The cost of dichotomising variables. *BMJ*, **332**, 1080.

Anderson, G., Deeley, T. J., Smith, C. and Jani, J. (1981) Comparison of radiotherapy alone and radiotherapy with chemotherapy using adriamycin and 5-fluorouracil in bronchogenic carcinoma. *Thorax*, **36**, 190–193.

Atagi, S., Kawahara, M., Tamura, T., Noda, K., Watanabe, K., Yokoyama, A., Sugiura, T., Senba, H., Ishikura, S., Ikeda, H., Ishizuka, N., Saijo, N. and Japan Clinical Oncology Group (2005) Standard thoracic radiotherapy with or without concurrent daily low-dose carboplatin in elderly patients with locally advanced non-small cell lung cancer: A phase III trial of the Japan Clinical Oncology Group (JCOG9812). *Jpn J Clin Oncol*, **35**, 195–201.

Auperin, A., Le Pechoux, C., Pignon, J. P., Koning, C., Jeremic, B., Clamon, G., Einhorn, L., Ball, D., Trovo, M. G., Groen, H. J., Bonner, J. A., Le Chevalier, T. and Arriagada, R. O. (2006) Concomitant radio-chemotherapy based on platin compounds in patients with locally advanced non-small cell lung cancer (NSCLC): A meta-analysis of individual data from 1764 patients. *Ann Oncol*, **17**, 473–83.

Auperin, A., Le Pechoux, C., Rolland, E., Curran, W. J., Furuse, K., Fournel, P., Belderbos, J., Clamon, G., Ulutin, H. C., Paulus, R., Yamanaka, T., Bozonnat, M. C., Uitterhoeve, A.,

Wang, X., Stewart, L., Arriagada, R., Burdett, S. and Pignon, J. P. (2010) Meta-analysis of concomitant versus sequential radiochemotherapy in locally advanced non-small-cell lung cancer. *J Clin Oncol*, **28**, 2181–90.

Ball, D., Bishop, J., Smith, J., O'Brien, P., Davis, S., Ryan, G., Olver, I., Toner, G., Walker, Q. and Joseph, D. (1999) A randomised phase III study of accelerate or standard fraction radiotherapy with or without concurrent carboplatin in inoperable non-small cell lung cancer: Final report of an Australian multi-centre trial. *Radiotherapy and Oncology*, **52**, 129–136.

Belani, C. P., Choy, H., Bonomi, P., Scott, C., Travis, P., Haluschak, J. and Curran, W. J., J. (2005) Combined chemoradiotherapy regimens of paclitaxel and carboplatin for locally advanced non-small-cell lung cancer: A randomized phase II locally advanced multi-modality protocol. *J Clin Oncol*, **23**, 5883–91.

Belderbos, J., Uitterhoeve, L., van Zandwijk, N., Belderbos, H., Rodrigus, P., van de Vaart, P., Price, A., van Walree, N., Legrand, C., Dussenne, S., Bartelink, H., Giaccone, G., Koning, C. and EORTC LCG and RT Group (2007) Randomised trial of sequential versus concurrent chemo-radiotherapy in patients with inoperable non-small cell lung cancer (EORTC 08972-22973). *Eur J Cancer*, **43**, 114–21.

Belsey, J. D. (2015) Appropriate use of information in therapeutic decision-making: Reflections on indirect comparisons. *Curr Med Res Opin*, **31**, 343–6.

Benedetti-Panici, P., Greggi, S., Colombo, A., Amoroso, M., Smaniotto, D., Giannarelli, D., Amunni, G., Raspagliesi, F., Zola, P., Mangioni, C. and Landoni, F. (2002) Neoadjuvant chemotherapy and radical surgery versus exclusive radiotherapy in locally advanced squamous cell cervical cancer: Results from the Italian multicenter randomized study. *Journal of Clinical Oncology*, **20**, 179–188.

Berlin, J. A. and Cepeda, M. S. (2012) Some methodological points to consider when performing systematic reviews in comparative effectiveness research. *Clinical Trials*, **9**, 27–34.

Bhatnagar, N., Lakshmi, P. V. and Jeyashree, K. (2014) Multiple treatment and indirect treatment comparisons: An overview of network meta-analysis. *Perspect Clin Res*, **5**, 154–8.

Biondi-Zoccai, G., Abbate, A., Benedetto, U., Palmerini, T., D'Ascenzo, F. and Frati, G. (2015) Network meta-analysis for evidence synthesis: What is it and why is it posed to dominate cardiovascular decision making? *Int J Cardiol*, **182C**, 309–314.

Bland, J. M. and Altman, D. G. (2004) The logrank test. *BMJ*, **328**, 1073.

Blanke, C., Ansari, R., Mantravadi, R., Gonin, R., Tokars, R., Fisher, W., Pennington, K., O'Connor, T., Rynard, S., Miller, M. and Einhorn, L. (1995) Phase III trial of thoracic irradiation with or without cisplatin for locally advanced unresectable non-small-cell lung cancer: A Hoosier Oncology Group protocol. *J Clin Oncol*, **13**, 1425–1429.

Bonner, J. A., McGinnis, W. L., Stella, P. J., Marschke Jr, R. F., Sloan, J. A., Shaw, E. G., Mailliard, J. A., Creagan, E. T., Ahuja, R. K. and Johnson, P. A. (1998) The possible advanatge of hyperfractionated thoracic radiotherapy in the treatment of locally advanced non small cell lung carcinoma. *Cancer*, **82**, 1037–1048.

Brodin, O., Nou, E., Mercke, C., Linden, C. J., Lundstrom, R., Arwidi, A., Brink, J. and Ringborg, U. (1996) Comparison of induction chemotherapy before radiotherapy with radiotherapy only in patients with locally advanced squamous cell carcinoma of the lung. *European Journal of Cancer*, **32A**, 1893–1900.

Bucher, H. C., Guyatt, G. H., Griffith, L. E. and Walter, S. D. (1997) The results of direct and indirect treatment comparisons in meta-analysis of randomised controlled trials. *J Clin Epidemiol*, **50**, 683–691.

Burke, D. L., Bujkiewicz, S. and Riley, R. D. (2016) Bayesian bivariate meta-analysis of correlated effects: Impact of the prior distributions on the between-study correlation, borrowing of strength, and joint inferences. *Stat Methods Med Res*.

Caldwell, D. M. (2014) An overview of conducting systematic reviews with network meta-analysis. *Syst Rev*, **3**, 109.

Caldwell, D. M., Ades, A. E. and Higgins, J. P. T. (2005) Simultaneous comparison of multiple treatments: Combining direct and indirect evidence. *BMJ*, **331**, 1897–900.

Caldwell, D. M., Welton, N. J. and Ades, A. E. (2010) Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency. *J Clin Epidemiol*, **63**, 875–82.

Cardenas, J., Olugin, A., Figueroa, F., Becerra, F. and Huizar, R. M. (1991) Randomized neoadjuvant chemotherapy in cervical carcinoma stage IIB PEC+RT vs RT. In *ASCO*.

Cardenas, J., Olugin, A., Figueroa, F., Pena, J., Becerra, F. and Huizar, R. (1993) A randomized trial of chemotherapy (CT) followed by radiotherapy (RT) vs radiotherapy alone in stage IIIB cervical carcinoma: Preliminary results. In *Fourth-international-congress-on-anti-cancer-chemotherapy*.

Cardiello, C., Blanco Villalba, J., Anac, S., Francheri Wilson, C., Trodler, C., Huñis, A. and Alvarez Bermudez, C. (1985) Combined radiochemotherapy (RTCT) versus radiotherapy (RT) in limited inoperable non-small cell carcinoma of the lung (NSCLC). *Proceedings American Society of Clinical Oncology*, **4**, 177.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M. A., Guo, J., Li, P. and Riddell, A. (2016) Stan: A probabilistic programming laguage. *Journal of Statistical Software*. In press.

Chaimani, A., Higgins, J. P. T., Mavridis, D., Spyridonos, P. and Salanti, G. (2013) Graphical tools for network meta-analysis in Stata. *PLOS One*, **8**, e76654.

Chaimani, A. and Salanti, G. (2012) Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. *Research Synthesis Methods*, **3**, 161–176.

— (2015) Visualizing assumptions and results in network meta-analysis: The network graphs package. *The Stata Journal*, **15**, 905–950.

Chalmers, I. (1993) The Cochrane Collaboration: Preparing, maintaining and disseminating systematic reviews of the effects of health care. *Annals of New York Academy of Sciences*, **703**, 156–165.

Chambers, J. D., Naci, H., Wouters, O. J., Pyo, J., Gunjal, S., Kennedy, I. R., Hoey, M. G., Winn, A. and Neumann, P. J. (2015) An assessment of the methodological quality of published network meta-analyses: A systematic review. *PLoS One*, **10**, e0121715.

Chang, T. C., Lai, C. H., Hong, J. H., Hsueh, S., Huang, K. G., Chou, H. H., Tseng, C. J., Tsai, C. S., Chang, J. T., Lin, C. T., Chang, H. H., Chao, P. J., Ng, K. K., Tang, S. G. J. and Soong, Y. K. (2000) Randomized trial of neoadjuvant cisplatin, vincristine, bleomycin and

radical hysterectomy versus radiation therapy for bulky stage IB and IIA cervical cancer. *Journal of Clinical Oncology*, **18**, 1740–1747.

Chauvergne, J., Lhommé, C., Rohart, J., Héron, J. F., Ayme, Y., Goupil, A., Fargeot, P. and David, M. (1993) Chimiothérapie néoadjuvante des cancer du col utérin aux stades IIb e III. résultats éloignés d'un essai randomisé pluricentrique portant sur 151 patients. *Bull Cancer (Paris)*, **80**, 1069–1079.

Chemoradiotherapy for Cervical Cancer Meta-Analysis Collaboration (2008) Reducing uncertainties about the effects of chemoradiotherapy for cervical cancer: A systematic review and meta-analysis of individual patient data from 18 randomized trials. *J Clin Oncol*, **26**, 5802–12.

Chen, H.-W., Jun-Jie, F. and Wei, L. (1997) A randomized trial of hyperthermo-radiochemotherapy for uterine cervix cancer. *Chineese Journal of Oncology*, **24**, 249–251.

Chiara, S., Bruzzone, M., Merlini, L., Bruzzi, P., Rosso, R., Franzone, P., Orsatti, M., Vitale, V., Foglia, G., Odicino, F., Ragni, N., Rugiati, S. and Conte, P. (1994) Randomized study comparing chemotherapy plus radiotherapy versus radiotherapy alone in FIGO stage IIB-III cervical carcinoma. *Am J Clin Oncol*, **17**, 294–297.

Cikaric, S., Petrovic-Stupar, S., Marjanov, I., Rudan, L., Colakovic, S., Tomasevic, A. and Rakovic, B. (2005) Radiotherapy vs radiotherapy+chemotherapy of advanced cervical cancer (IIB-IVA): Regression of tumour, early and late sequelaes. In *European Cancer Conference*.

Ciprani, A., Higgins, J. P. T., Geddes, J. R. and Salanti, G. (2013) Conceptual and technical challenges in network meta-analysis. *Annals of Internal Medicine*, **159**, 130–137.

Cipriani, A., Barbui, C., Salanti, G., Rendell, J., Brown, R., Stockton, S., Purgato, M., Spineli, L. M., Goodwin, G. M. and Geddes, J. R. (2011) Comparative efficacy and acceptability of antimanic drugs in acute mania: A multiple treatments meta-analysis. *Lancet*, **378**, 1306–15.

Cipriani, A., Furukawa, T. A., Salanti, G., Geddes, J. R., Higgins, J. P. T., Churchill, R., Watanabe, N., Nakagawa, A., Omori, I. M., McGuire, H., Tansella, M. and Barbui, C.

(2009) Comparative efficacy and acceptability of 12 new-generation antidepressants: A multiple-treatments meta-analysis. *Lancet*, **373**, 746–758.

Clamon, G., Herndon, J., Cooper, R., Chang, A. Y., Rosenman, J. and Green, M. R. (1999) Radiosensitization with carboplatin for patients with unresectable stage III non-small-cell lung cancer: A phase III trial of the Cancer and Leukemia Group B and the Eastern Cooperative Oncology Group. *J Clin Oncol*, **17**, 4–11.

Clamon, G., Herndon, J., Eaton, W., Rosenman, J., Maurer, L. H., Cooper, M. R. and Green, M. R. (1994) A feasibility study of extended chemotherapy for locally advanced non-small cell lung cancer: A phase II trial of Cancer and Leukemia Group B. *Cancer Invest*, **12**, 273–282.

ClinicalTrials.gov (2015a) Neoadjuvant chemotherapy followed by surgery versus concurrent chemoradiation in carcinoma of the cervix (NACTcervix). `https://clinicaltrials.gov/ct2/show/study/NCT00193739?term=chemoradiation+cervix&rank=1#desc`. Accessed 19/03/2017.

— (2015b) Study on early stage bulky cervical cancers. `https://clinicaltrials.gov/ct2/show/NCT01917695?term=chemoradiation+cervix&rank=8`. Accessed 19/03/2017.

Cochran, W. G. (1954) The combination of estimates from different experiments. *Biometrics*, **10**, 101–129.

Cooper, N. J., Peters, J., Lai, M. C., Juni, P., Wandel, S., Palmer, S., Paulden, M., Conti, S., Welton, N. J., Abrams, K. R., Bujkiewicz, S., Spiegelhalter, D. and Sutton, A. J. (2011) How valuable are multiple treatment comparison methods in evidence-based health-care evaluation? *Value Health*, **14**, 371–80.

Cooper, N. J., Sutton, A. J., Morris, D., Ades, A. E. and Welton, N. J. (2009) Addressing between-study heterogeneity and inconsistency in mixed treatment comparisons: Application to stroke prevention treatments in individuals with non-rheumatic atrial fibrillation. *Stat Med*, **28**, 1861–81.

Cope, S., Zhang, J., Saletan, S., Smiechowski, B., Jansen, J. P. and Schmid, P. (2014) A process for assessing the feasibility of a network meta-analysis: A case study of

everolimus in combination with hormonal therapy versus chemotherapy for advanced breast cancer. *BMC Medicine*, **12**, 93.

Cox, D. R. (1972) Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

Crino, L., Latini, P., Meacci, M., Corgna, E., Maranzano, E., Darwish, S., Minotti, V., Santucci, A. and Tonato, M. (1993) Induction chemotherapy plus high-dose radiotherapy versus radiotherapy alone in locally advanced unresectable non-small-cell lung cancer. *Annals of Oncology*, **4**, 847–851.

Crowther, M. J., Look, M. P. and Riley, R. D. (2014) Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Stat Med*, **33**, 3844–58.

Crowther, M. J., Riley, R. D., Staessen, J. A., Wang, J., Gueyffier, F. and Lambert, P. C. (2012) Individual patient data meta-analysis of survival data using Poisson regression models. *BMC Med Res Methodol*, **12**, 34.

Cullen, M. H., Billingham, L. J., Woodroffe, C. M., Chetiyawardana, A. D., Gower, N. H., Joshi, R., Ferry, D. R., Rudd, R. M., Spiro, S. G., Cook, J. E., Trask, C., Bessell, E., Connolly, C. K., Tobias, J. and Souhami, R. L. (1999) Mitomycin, Ifosfamide and Cisplatin in unresectable non-small-cell lung cancer: Effects on survival and quality of life. *J Clin Oncol*, **17**, 3188–3194.

Cüneyt Ulutin, H. and Pak, Y. (2003) Preliminary results of radiotherapy with or without weekly paclitaxel in locally advanced non-small cell lung cancer. *J Cancer Res Clin Oncol*, **129**, 52–6.

Curran Jr, W. J., Paulus, R., Langer, C. J., Komaki, R., Lee, J. S., Hauser, S., Movsas, B., Wasserman, T., Rosenthal, S. A., Gore, E., Machtay, M., Sause, W. and Cox, J. D. (2011) Sequential vs. concurrent chemoradiation for stage III non-small cell lung cancer: Randomized phase III trial RTOG 9410. *J Natl Cancer Inst*, **103**, 1452–60.

Dakin, H. A., Welton, N. J., Ades, A. E., Collins, S., Orme, M. and Kelly, S. (2011) Mixed treatment comparison of repeated measurements of a continuous endpoint: An example using topical treatments for primary open-angle glaucoma and ocular hypertension. *Stat Med*, **30**, 2511–35.

Del Giovane, C., Vacchi, L., Mavridis, D., Filippini, G. and Salanti, G. (2013) Network meta-analysis models to account for variability in treatment definitions: Application to dose effects. *Stat Med*, **32**, 25–39.

DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Controlled Clinical Trials*, **7**, 177.

Dias, S., Sutton, A. J., Welton, N. J. and Ades, A. E. (2013a) Evidence synthesis for decision making 6: Embedding evidence synthesis in probabilistic cost-effectiveness analysis. *Med Decis Making*, **33**, 671–8.

Dias, S., Welton, N. J. and Ades, A. E. (2010a) Study designs to detect sponsorship and other biases in systematic reviews. *J Clin Epidemiol*, **63**, 587–8.

Dias, S., Welton, N. J., Caldwell, D. M. and Ades, A. E. (2010b) Checking consistency in mixed treatment comparison meta-analysis. *Stat Med*, **29**, 932–44.

Dias, S., Welton, N. J., Marinho, V. C. C., Salanti, G. and Higgins, J. P. T. (2010c) Estimation and adjustment of bias in randomized evidence by using mixed treatment comparison meta-analysis. *Journal of the Royal Statistical Society Series A*, **173**, 613–629.

Dias, S., Welton, N. J., Sutton, A. J., Caldwell, D. M., Lu, G. and Ades, A. E. (2013b) Evidence synthesis for decision making 4: Inconsistency in networks of evidence based on randomized controlled trials. *Med Decis Making*, **33**, 641–56.

Dillman, R. O., Seagren, S. L., Propert, K. J., Guerra, J., Eaton, W. L., Perry, M. C., Carey, R. W., Frei, E. F. and Green, M. R. (1990) A randomized trial of induction chemotherapy plus high-dose radiation versus radiation alone in stage III non-small-cell lung cancer. *N Engl J Med*, **323**, 940–945.

Ding, Y. and Fu, H. (2013) Bayesian indirect and mixed treatment comparisons across longitudinal time points. *Stat Med*, **32**, 2613–28.

Dominici, F., Parmigiani, G., Wolpert, R. L. and Hasselblad, V. (1999) Meta-analysis of migraine headache treatments: Combining information from heterogeneous designs. *Journal of the American Statistical Association*, **94**, 16–28.

Donegan, S., Williamson, P., D'Alessandro, U., Garner, P. and Tudur Smith, C. (2013a) Combining individual patient data and aggregate data in mixed treatment comparison

meta-analysis: Individual patient data may be beneficial if only for a subset of trials. *Stat Med*, **32**, 914–30.

Donegan, S., Williamson, P., D'Alessandro, U. and Tudur Smith, C. (2012) Assessing the consistency assumption by exploring treatment by covariate interactions in mixed treatment comparison meta-analysis: Individual patient-level covariates versus aggregate trial-level covariates. *Stat Med*, **31**, 3840–57.

— (2013b) Assessing key assumptions of network meta-analysis: A review of methods. *Research Synthesis Methods*, **4**, 391–323.

Douillard, J., Gervais, R., Quoix, E., Chevalier, T. L., Groumellec, A. L., Lemarie, E., Haller, M., Bardet, E., Lapachoux, C. and Henry-Amar, M. (2005) Randomized phase III trial for stage III unresectable non samll cell lung cancer: Induction chemotherapy [vinorelbine-cisplatin] followed by conventional radiation without or with daily carboplatin. *Lung Cancer*, **49 (suppl 2)**, S16 (abstract O–040).

Efthimiou, O., Debray, T. P., van Valkenhoef, G., Trelle, S., Panayidou, K., Moons, K. G., Reitsma, J. B., Shang, A., Salanti, G. and GetReal Methods Review Group (2016) GetReal in network meta-analysis: A review of the methodology. *Res Synth Methods*, **7**, 236–263.

F, L., Leborgne, J. H., Doldan, R., Zubizarreta, E., Ortega, B., Maisonneuve, J., Musetti, E., Hekimian, L. and Mezzera, J. (1997) Induction chemotherapy and radiotherapy of advanced cancer of the cervix: A pilot study and phase III randomized trial. *Int J Radiation Oncology Biol Phys*, **37**, 343–350.

Fairlamb, D., Milroy, R., Gower, N., Parmar, M., Peake, M., Rudd, R., Souhami, R., Spiro, S., Stephens, R., Waller, D. and on behalf of all BLT participants (2005) A randomised comparison of radical radiotherapy with or without chemotherapy for patients with non-small cell lung cancer: Results from the Big Lung Trial. *Radiotherapy and Oncology*, **75**, 134–140.

Fisher, D., Carpenter, J., Morris, T., Freeman, S. and Tierney, J. (2016) Identifying who benefits most from treatments: Daft, deluded or deft meta-analysis? *Submitted for publication.*

Fisher, D. J. (2015) Two-stage individual participant data meta-analysis and generalized forest plots. *The Stata Journal*, **15**, 369–396.

Fisher, D. J., Copas, A. J., Tierney, J. F. and Parmar, M. K. (2011) A critical review of methods for the assessment of patient-level interactions in individual participant data meta-analysis of randomized trials, and guidance for practitioners. *Journal of Clinical Epidemiology*, **64**, 949–967.

Fournel, P., Robinet, G., Thomas, P., Souquet, P. J., Lena, H., Vergnenegre, A., Delhoume, J. Y., Le Treut, J., Silvani, J. A., Dansin, E., Bozonnat, M. C., Daures, J. P., Mornex, F. and Perol, M. (2005) Randomized phase III trial of sequential chemoradiotherapy compared with concurrent chemoradiotherapy in locally advanced non-small-cell lung cancer: Groupe Lyon-Saint-Etienne d'Oncologie Thoracique-Groupe Francais de Pneumo-Cancerologie NPC 95-01 study. *J Clin Oncol*, **23**, 5910–7.

Franchini, A. J., Dias, S., Ades, A. E., Jansen, J. P. and Welton, N. J. (2012) Accounting for correlation in network meta-analysis with multi-arm trials. *Research Synthesis Methods*, **3**, 142–160.

Fu, H. and Manner, D. (2010) Bayesian adaptive dose-finding studies with delayed responses. *J Biopharm Stat*, **20**, 1055–70.

Furuse, K., Fukuoka, M., Kawahara, M., Nishikawa, H., Takada, Y., Kudoh, S., Katagami, N. and Ariyoshi, Y. (1999) Phase III study of concurrent versus sequential thoracic radiotherapy in combination with mitomycin, vindesine and cisplatin in unresectable stage III non-small cell lung cancer. *J Clin Oncol*, **17**.

Garipagaoglu, M. (2004) Adding concurrent low dose continuous infusion of cisplatin to radiotherapy in locally advanced cervical carcinoma: A prospective randomized pilot study. *British Journal of Radiology*, **77**, 581–587.

Gasparrini, A., Armstrong, B. and Kenward, M. G. (2012) Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med*, **31**, 3821–39.

Govan, L., Ades, A. E., Weir, C. J., Welton, N. J. and Langhorne, P. (2010) Controlling ecological bias in evidence synthesis of trials reporting on collapsed and overlapping covariate categories. *Stat Med*, **29**, 1340–56.

Grambsch, P. M. and Therneau, T. M. (1994) Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, **81**, 515.

Greenland, S. and Morgenstern, H. (1989) Ecological bias, confounding and effect modification. *International Journal of Epidemiology*, **18**, 269–274.

Gregor, A., Macbeth, F. R., Paul, J., Cram, L. and Hansen, H. H. (1993) Radical radiotherapy and chemotherapy in localized inoperable non-small-cell lung cancer: A randomized trial. *Journal of the National Cancer Institute*, **85**, 997–999.

Groen, H. J. M. (2004) Continuously infused carboplatin used as radiosensitizer in locally unresectable non-small-cell lung cancer: A multicenter phase III study. *Annals of Oncology*, **15**, 427–432.

Harris, R. J., Bradburn, M. J., Deeks, J. J., Harbord, R. M., Altman, D. G. and Sterne, J. A. C. (2008) metan: Fixed and random-effects meta-analysis. *The Stata Journal*, **8**, 3–28.

Herod, J., Burton, A., Buxton, J., Tobias, J., Luesley, D., Jordan, S., Dunn, J. and Poole, C. J. (2000) A randomised, prospective, phase III clinical trial of primary bleomycin, ifosfamide and cisplatin (BIP) chemotherapy followed by radiotherapy versus radiotherapy alone in operable cancer of the cervix. *Annals of Oncology*, **11**, 1175–1181.

Higgins, J. P. T., Jackson, D., Barrett, J. K., Lu, G., Ades, A. E. and White, I. R. (2012) Consistency and inconsistency in network meta-analysis: Concepts and models for multi-arm studies. *Research Synthesis Methods*, **3**, 98–110.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J. and Altman, D. G. (2003) Measuring inconsistency in meta-analyses. *BMJ*, **327**, 557.

Higgins, J. P. T. and Whitehead, A. (1996) Borrowing strength from external trials in a meta-analysis. *Statistics in medicine*, **15**, 2733–2749.

Higgins, J. P. T., Whitehead, A., Turner, R. M., Omar, R. Z. and Thompson, S. G. (2001) Meta-analysis of continuous outcome data from individual patients. *Stat Med*, **20**, 2219–2241.

Hong, H., Carlin, B. P., Shamliyan, T. A., Wyman, J. F., Ramakrishnan, R., Sainfort, F. and Kane, R. L. (2013) Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. *Med Decis Making*, **33**, 702–14.

Hong, H., Chu, H., Zhang, J. and Carlin, B. P. (2016) A Bayesian missing data framework for generalized multiple outcome mixed treatment comparisons (with discussion and rejoinder). *Res Synth Methods*, **7**, 6–33.

Hong, H., Fu, H., Price, K. L. and Carlin, B. P. (2015) Incorporation of individual-patient data in network meta-analysis for multiple continuous endpoints, with application to diabetes treatment. *Stat Med*, **34**, 2794–819.

Huber, R. M., Flentje, M., Schmidt, M., Pollinger, B., Gosse, H., Willner, J. and Ulm, K. (2006) Simultaneous chemoradiotherapy compared with radiotherapy alone after induction chemotherapy in inoperable stage IIIA or IIIB non-small-cell lung cancer: Study CTRT99/97 by the Bronchial Carcinoma Therapy Group. *J Clin Oncol*, **24**, 4397–404.

Hutton, B., Salanti, G., Caldwell, D. M., Chaimani, A., Schmid, C. H., Cameron, C., Ioannidis, J. P., Straus, S., Thorlund, K., Jansen, J. P., Mulrow, C., Catala-Lopez, F., Gotzsche, P. C., Dickersin, K., Boutron, I., Altman, D. G. and Moher, D. (2015) The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. *Ann Intern Med*, **162**, 777–84.

Ioannidis, J. P. (2009) Integration of evidence from multiple meta-analyses: A primer on umbrella reviews, treatment networks and multiple treatments meta-analyses. *CMAJ*, **181**, 488–93.

Jackson, D., Barrett, J. K., Rice, S., White, I. R. and Higgins, J. P. (2014) A design-by-treatment interaction model for network meta-analysis with random inconsistency effects. *Stat Med*, **33**, 3639–54.

Jackson, D., White, I. R. and Riley, R. D. (2012) Quantifying the impact of between-study heterogeneity in multivariate meta-analyses. *Stat Med*, **31**, 3805–20.

Jansen, J. P. (2011) Network meta-analysis of survival data with fractional polynomials. *BMC Med Res Methodol*, **11**, 61.

— (2012) Network meta-analysis of individual and aggregate level data. *Research Synthesis Methods*, **3**, 177–190.

Jansen, J. P. and Cope, S. (2012) Meta-regression models to address heterogeneity and inconsistency in network meta-analysis of survival outcomes. *BMC Med Res Methodol*, **12**, 152.

Jeremic, B., Shibamoto, Y., Acimovic, L. and Djuric, L. (1995) Randomized trial of hyperfractionated radiation theapy with or without concurrent chemotherapy for stage III non-small-cell lung cancer. *J Clin Oncol*, **13**.

Jeremic, B., Shibamoto, Y., Acimovic, L. and Milsavljevic, S. (1996) Hyperfractionated radiation therapy with or without concurrent low-dose daily carboplatin/etoposide for stage III non-small-cell lung cancer: A randomized study. *J Clin Oncol*, **14**.

Kantazardic, N., Beslija, S. and Begic, D. (2004) Comparation (comparison) of parameters of mielotoxicity in patients treated with concomitant chemotherapy and radiotherapy versus radiotherapy alone. *MedArh*, **58**, 19–22.

Keys, H. M., Bundy, B. N., Stehman, F. B., Muderspach, L. I., Chafe, W. E., Suggs, C. L., Walker, J. L. and Gersell, D. (1999) Cisplatin, radiation and adjuvant hysterectomy compared with radiation and adjuvant hysterectomy for bulky stage IB cervical carcinoma. *The New England Journal of Medicine*, **340**, 1154–1161.

Kigawa, J., Minagawa, Y., Ishiara, H., Itamochi, H., Kanamori, Y. and Terakawa, N. (1996) The role of neoadjuvant intraarterial infusion chemotherapy with cisplatin and bleomycin for locally advanced cervical cancer. *Am J Clin Oncol*, **19**, 255–259.

Krahn, U., Binder, H. and König, J. (2013) A graphical tool for locating inconsistency in network meta-analyses. *BMC Med Res Methodol*, **13**, 35.

Kumar, L., Kaushal, R., Biswal, B. M., Kumar, S., Kriplani, A., Singh, R., Rath, G. K. and Kochupillai, V. (1994) Chemotherapy followed by radiotherapy versus radiotherapy alone in locally advanced cervical cancer: A randomized study. *Gynecologic Oncology*, **54**, 307–315.

Kumar, S., Lal, P., Tiwari, A., Kumar, S., Dimri, K., Rastogi, N. and Ayyagari, S. (2004) Chemo-irradiation versus radiotherapy alone in locally advanced carcinoma of the uterine cervix - an ongoing phase III trial. In *ESTRO*.

Lambert, P. C. and Royston, P. (2009) Further development of flexible parametric models for survival analysis. *The Stata Journal*, **9**, 265–290.

Lanciano, R., Calkins, A., Bundy, B. N., Parham, G., Lucci 3rd, J. A., Moore, D. H., Monk, B. J. and O'Connor, D. M. (2005) Randomized comparison of weekly cisplatin or protracted venous infusion of fluorouracil in combination with pelvic radiation in advanced cervix cancer: A Ggynecologic Oncology Group study. *J Clin Oncol*, **23**, 8289–95.

Laws, A., Kendall, R. and Hawkins, N. (2014) A comparison of national guidelines for network meta-analysis. *Value Health*, **17**, 642–54.

Le Chevalier, T., Arriagada, R., Quoix, E., Ruffie, P., Martin, M., Tarayre, M., Lacombe-Terrier, M. J., Douillard, J. Y. and Laplanche, A. (1991) Radiotherapy alone versus combined chemotherapy and radiotherapy in nonresectable non-small-cell lung cancer: First analysis of a randomized trial in 353 patients. *J Natl Cancer Inst*, **83**, 417–423.

Lee, A. W. (2014) Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. *J Clin Epidemiol*, **67**, 138–43.

Lorvidhaya, V., Chitapanarux, I., Sangruchi, S., Lertsanguansinchai, P., Kongthanarat, Y., Tangkaratt, S. and Visetsiri, E. (2003) Concurrent mitomycin C, 5-fluorouracil, and radiotherapy in the treatment of locally advanced carcinoma of the cervix: A randomized trial. *Int J Radiation Oncology Biol Phys*, **55**, 1226–1232.

Lu, G. and Ades, A. (2009) Modeling between-trial variance structure in mixed treatment comparisons. *Biostatistics*, **10**, 792–805.

Lu, G. and Ades, A. E. (2004) Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*, **23**, 3105–24.

— (2006) Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, **101**, 447–459.

Lu, G., Ades, A. E., Sutton, A. J., Cooper, N. J., Briggs, A. H. and Caldwell, D. M. (2007) Meta-analysis of mixed treatment comparisons at multiple follow-up times. *Stat Med*, **26**, 3681–99.

Lu, G., Welton, N. J., Higgins, J. P. T., White, I. R. and Ades, A. E. (2011) Linear inference for mixed treatment comparison meta-analysis: A two-stage approach. *Research Synthesis Methods*, **2**, 43–60.

Lumley, T. (2002) Network meta-analysis for indirect treatment comparisons. *Stat Med*, **21**, 2313–24.

Lunn, D. (2014) Introduction to Bayesian analysis using WinBUGS. Short course.

Lunn, D., Jackson, C., Best, N., Thomas, A. and Spiegelhalter, D. (2013) *The BUGS Book. A practical introduction to Bayesian Analysis*. Texts in Statistical Science. Boca Raton, FL, USA: CRC Press.

Lunn, D. J., Thomas, A., Best, N. and Spiegelhalter, D. (2000) WinBUGS - a Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–337.

Mattson, K., Holsti, L. R., Holsti, P., Jakobsson, M., Kajanti, M., Luppo, K., Mantyla, M., Nutamo-Korhonen, S., Nikkanen, V., Nordman, E., Platin, L. H., Pyrhonen, S., Romp-panen, M. L., Salmi, R., Tammilehto, L. and Taskinen, P. J. (1988) Inoperable non-small cell lung cancer: Radiation with or without chemotherapy. *European Journal of Clinical Oncology*, **24**, 477–482.

Mattson, K. V. (2003) Docetaxel as neoadjuvant therapy for radically treatable stage III non-small-cell lung cancer: A multinational randomised phase III study. *Annals of Oncology*, **14**, 116–122.

Mavridis, D., Sutton, A., Cipriani, A. and Salanti, G. (2013) A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. *Stat Med*, **32**, 51–66.

Medical Research Council (2016) Stratified medicine - Our research - Medical Research Council. `http://www.mrc.ac.uk/research/initiatives/stratified-medicine`. Accessed 15/08/2016.

Mills, E. J., Ioannidis, J. P. A., Thorlund, K., Schünemann, H. J., Puhan, M. A. and Guyatt, G. H. (2012) How to use an article reporting a multiple treatment comparison meta-analysis. *JAMA*, **308**, 1246–1253.

Mills, E. J., Thorlund, K. and Ioannidis, J. P. (2013) Demystifying trial networks and network meta-analysis. *BMJ*, **346**, f2914.

Mira, J. G., Miller, T. P. and Crowley, J. J. (1990) Chest irradiation versus chest irradiation plus chemotherapy with or without cranial prophylatctic brain radiotherapy in localized non-small cell lung cancer: A Southwest Oncology Group randomized study. *Int J Radiation Oncology Biol Phys*, **19**, 145.

Morris, M., Eifel, P. J., Lu, J., Grigsby, P. W., Levenback, C., Stevens, R. E., Rotman, M., Gershenson, D. M. and Mutch, D. G. (1999) Pelvic radiation with concurrent chemotherapy compared with pelvic and para-aortic radiation for high-risk cervical cancer. *The New England Journal of Medicine*, **340**, 1137.

Morton, R. F., Jett, J. R., McGinnis, W. L., Earle, J. D., Therneau, T. M., Krook, J. E., Elliott, T. E., Mailliard, J. A., Nelimark, R. A. and Maksymiuk, A. W. (1991) Thoracic radiation therapy alone compared with combined chemoradiotherapy for locally unresectable non-small cell lung cancer. A randomized, phase III trial. *Ann Intern Med*, **115**, 681–686.

Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration (2003) Neoadjuvant chemotherapy for locally advanced cervical cancer. *European Journal of Cancer*, **39**, 2470–2486.

Nikolakopoulou, A., Chaimani, A., Veroniki, A. A., Vasiliadis, H. S., Schmid, C. H. and Salanti, G. (2014) Characteristics of networks of interventions: A description of a database of 186 published networks. *PLoS One*, **9**, e86754.

Nikolakopoulou, A., Mavridis, D. and Salanti, G. (2016) Planning future studies based on the precision of network meta-analysis results. *Stat Med*, **35**, 978–1000.

Nixon, R. M., Bansback, N. and Brennan, A. (2007) Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Stat Med*, **26**, 1237–54.

Non-small Cell Lung Cancer Collaborative Group (1995) Chemotherapy in non-small cell lung cancer: A meta-analysis using updated data on individual patients from 52 randomised clinical trials. *BMJ*, **311**, 899–909.

Onishi, H., Yamaguchi, M., Kuriyama, K., Tsukamoto, T., Ishigame, K., Ichikawa, T., Aoki, S., Yoshikawa, T., Araki, T., Nambu, A., Araki, T., Hashi, A., Yasumizu, T., Hoshi, K. and Ito, H. (1999) Effect of concurrent intra-arterial infusion of platinum drugs for patients

with stage III or IV uterine cervical cancer treated with radical radiation therapy. *Cancer J Sci Am*, **5**, 40–45.

Ouwens, M. J. N. M., Philips, Z. and Jansen, J. P. (2010) Network meta-analysis of parametric survival curves. *Research Synthesis Methods*, **1**, 258–271.

Palmer, T. M., Sterne, J. A. C., Newton, H. J. and Cox, N. J. (2016) *Meta-Analysis in Stata: An updated collection from the Stata Journal*. College Station, Texas, USA: Stata Press.

Pearcey, R., Brundage, M., Drouin, P., Jeffrey, J., Johnston, D., Lukka, H., MacLean, G., Souhami, L., Stuart, G. and Tu, D. (2002) Phase III trial comparing radical radiotherapy with and without cisplatin chemotherapy in patients with advanced squamous cell cancer of the cervix. *Journal of Clinical Oncology*, **20**, 966–972.

Peters, W. A., Liu, P. Y., Barrett, R. J., Stock, R. J., Monk, B. J., Berek, J. S., Souhami, L., Grigsby, P., Gordon Jr, W. and Alberts, D. S. (2000) Concurrent chemotherapy and pelvic radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix. *Journal of Clinical Oncology*, **18**, 1606–1613.

Piepho, H. P. (2014) Network-meta analysis made easy: Detection of inconsistency using factorial analysis-of-variance models. *BMC Med Res Methodol*, **14**, 61.

Planting, A., Helle, P., Drings, P., Dalesio, O., Kirkpatrick, A., McVie, G. and Giaccone, G. (1996) A randomized study of high-dose split course radiotherapy preceded by high-dose chemotherapy versus high-dose radiotherapy only in locally advanced non-small-cell lung cancer. *Annals of Oncology*, **7**.

Price, M. J., Welton, N. J. and Ades, A. E. (2011) Parameterization of treatment effects for meta-analysis in multi-state Markov models. *Stat Med*, **30**, 140–51.

Puhan, M. A., Schunemann, H. J., Murad, M. H., Li, T., Brignardello-Petersen, R., Singh, J. A., Kessels, A. G., Guyatt, G. H. and the GRADE Working Group (2014) A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ*, **349**, g5630.

R Core Team (2014) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URLhttp://www.R-project.org.

Riley, R. D., Lambert, P. C., Staessen, J. A., Wang, J., Gueyffier, F., Thijs, L. and Boutitie, F. (2008) Meta-analysis of continuous outcomes combining individual patient data and aggregate data. *Stat Med*, **27**, 1870–93.

Roberts, K. B., Urdaneta, N., Vera, R., Vera, A., Gutierrez, E., Aguilar, Y., Ott, S., Medina, I., Sempere, P., Rockwell, S., Sartorelli, A., Fischer, D. B. and Fischer, J. J. (2000) Interim results of a randomized trial of mitomycin C as an adjunct to radical radiotherapy in the treatment of locally advanced squamous-cell carcinoma of the cervix. *Int J Cancer*, **90**, 206–223.

Rose, P. G., Bundy, B. N., Watkins, E. B., Thigpen, T., Deppe, G., Maiman, M. A., Clarke-Pearson, D. L. and Insalaco, S. (1999) Concurrent cisplatin-based radiotherapy and chemotherapy for locally advanced cervical cancer. *The New England Journal of Medicine*, **340**, 1144–1153.

Rothwell, P. M. (2005) Subgroup analysis in randomised controlled trials: Importance indications, and interpretation. *The Lancet*, **365**, 176–186.

Royston, P. and Altman, D. G. (1994) Regression using fractional polynomials of continuous covariates: Parsimonious parametric modelling (with discussion). *Applied Statistics*, **43**, 429–467.

Royston, P. and Lambert, P. C. (2011) *Flexible parametric survival analysis using Stata: Beyond the Cox model*. College Station, Texas, USA: Stata Press.

Royston, P. and Parmar, M. K. (2002) Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Stat Med*, **21**, 2175–97.

— (2016) Augmenting the logrank test in the design of clinical trials in which non-proportional hazards of the treatment effect may be anticipated. *BMC Med Res Methodol*, **16**, 16.

Royston, P. and Parmar, M. K. B. (2011) The use of restricted mean survival time to estimate the treatment effect in randomized clinical trial when the proportional hazards assumption is in doubt. *Stat Med*, **30**, 2409–2421.

Rücker, G., Schwarzer, G., Carpenter, J. R. and Schumacher, M. (2008) Undue reliance on I^2 in assessing heterogeneity may mislead. *BMC Med Res Methodol*, **8**, 79.

Rücker, G., Schwarzer, G. and Krahn, U. (2014) *netmeta: Network meta-analysis with R.* URL`http://CRAN.R-project.org/package=netmeta`. R package version 0.4-2.

Rutherford, M. J., Crowther, M. J. and Lambert, P. C. (2015) The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: A simulation study. *Journal of Statistical Computation and Simulation*, **85**, 777–793.

Salanti, G., Dias, S., Welton, N. J., Ades, A. E., Golfinopoulos, V., Kyrgiou, M., Mauri, D. and Ioannidis, J. P. (2010) Evaluating novel agent effects in multiple-treatments meta-regression. *Stat Med*, **29**, 2369–83.

Salanti, G., Giovane, C. D., Chaimani, A., Caldwell, D. M. and Higgins, J. P. T. (2014) Evaluating the quality of evidence from a network meta-anlaysis. *PLoS One*, **9**, e99682.

Salanti, G., Higgins, J. P. T., Ades, A. E. and Ioannidis, J. P. A. (2007) Evaluation of networks of randomized trials. *Statistical methods in medical research*, **17**, 279–301.

Salanti, G., Marinho, V. and Higgins, J. P. (2009) A case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. *J Clin Epidemiol*, **62**, 857–64.

Saramago, P., Chaung, L. and Soares, M. O. (2014) Network meta-analysis of (individual patient) time to event data alongside (aggregate) count data. *BMC Med Res Methodol*, **14**.

Saramago, P., Sutton, A. J., Cooper, N. J. and Manca, A. (2012) Mixed treatment comparisons using aggregate and individual participant level data. *Stat Med*, **31**, 3516–36.

Sardi, J., Giaroli, A., Sananes, C., Rueda, N. G., Vighi, S., Ferreira, M., Bastardas, M., Paniceres, G. and di Paola, G. (1996) Randomized trial with neoadjuvant chemotherapy in stage IIIB squamous carcinoma cervix uteri: An unexpected therapeutic management. *Int J Gynecol Cancer*, **6**, 85–93.

Sardi, J. E., Giaroli, A., Sananes, C., Ferreira, M., Soderini, A., Bermudez, A., Snaidas, L., Vighi, S., Rueda, N. G. and di Paola, G. (1997) Long-term follow-up of the first randomized trial using neoadjuvant chemotherapy in stage IB squamous carcinoma of the cervix: The final results. *Gynecologic Oncology*, **67**, 61–69.

Sardi, J. E., Sananes, C. E., Giaroli, A., Bermudez, A., Ferreira, M. H., Soderini, A. H., Snaidas, L., Guardado, N., Anchezar, P., Ortiz, O. C. and di Paola, G. R. (1998) Neoadjuvant chemotherapy in cervical carcinoma stage IIB: A randomized controlled trial. *Int J Gynecol Cancer*, **8**, 441.

Sarihan, S., Kayisogullari, U., Ercan, I. and Engin, K. (2004) Randomized phase II study of radiotherapy alone versus radiotherapy with paclitaxel in non-small cell lung cancer. *Journal of International Medical Research*, **32**, 375–383.

Sause, W., Kolesar, P., Taylor IV, S., Johnson, D., Livingston, R., Komaki, R., Emami, B., Curran Jr, W., Byhardt, R., Rashid Dar, A. and Turrisi, A. (2000) Final results of phase III trial in regionally advanced unresectable non-small cell lung cancer. *Chest*, **117**, 358–64.

Scagliotti, G. V., Szczesna, A., Ramlau, R., Cardenal, F., Mattson, K., Van Zandwijk, N., Price, A., Lebeau, B., Debus, J. and Manegold, C. (2006) Docetaxel-based induction therapy prior to radiotherapy with or without docetaxel for non-small-cell lung cancer. *Br J Cancer*, **94**, 1375–82.

Schaake-Koning, C., van den Bogaert, W., Dalesio, O., Festen, J., Hoogenhout, J., van Houtte, P., Kirkpatrick, A., Koolen, M., Maat, B., Nijs, A., Renaud, A., Rodrigus, P., Schuster-Uitterhoeve, L., Sculier, J. P., van Zandwijk, N. and Bartelink, H. (1992) Effects of concomitant cisplatin and radiotherapy on inoperable non-small-cell lung cancer. *N Engl J Med*, **326**, 524–530.

Schmidli, H., Wandel, S. and Neuenschwander, B. (2013) The network meta-analytic-predictive approach to non-inferiority trials. *Stat Methods Med Res*, **22**, 219–40.

Schmitz, S., Adams, R. and Walsh, C. (2013) Incorporating data from various trial designs into a mixed treatment comparison model. *Stat Med*, **32**, 2935–49.

Schwarzer, G., Carpenter, J. R. and Rücker, G. (2015) *Meta-Analysis with R*. Use R! Switzerland: Springer.

Sharma, S., Sharma, R. and Bhowmik, K. T. (2003) Sequential chemoradiotherapy versus radiotherapy in the management of locally advanced non-small-cell lung cancer. *Advances in Therapy*, **20**, 14–19.

Simmonds, M. C., Higgins, J. P. T., Stewart, L. A., Tierney, J. F., Clarke, M. J. and Thompson, S. G. (2005) Meta-analysis of individual patient data from randomized trials: A review of methods used in practice. *Clinical Trials*, **2**, 209–17.

Smith, T. C., Spiegelhalter, D. J. and Thomas, S. L. (1995) Bayesian approaches to random-effects meta-analysis: A comparative study. *Stat Med*, **14**, 2685–2699.

Sobieraj, D. M., Cappelleri, J. C., Baker, W. L., Phung, O. J., White, C. M. and Coleman, C. I. (2013) Methods used to conduct and report Bayesian mixed treatment comparisons published in the medical literature: A systematic review. *BMJ Open*, **3**, e003111.

Song, F., Clark, A., Bachmann, M. O. and Maas, J. (2012) Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. *BMC Med Res Methodol*, **12**, 138.

Song, F., Harvey, I. and Lilford, R. (2008) Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *J Clin Epidemiol*, **61**, 455–63.

Souhami, L., Gil, R. A., Allan, S. E., Canary, P. C. V., Araujo, C. M. M., Pinto, L. H. J. and Silveira, T. R. P. (1991) A randomized trial of chemotherapy followed by pelvic radiation therapy in stage IIIB carcinoma of the cervix. *Journal of Clinical Oncology*, **9**, 970–977.

Spiegelhalter, D. J., Best, N. G. and van der Linde, A. (2002) Bayesian measures of model complexity and fit. *J R Statist Soc B*, **64**, 583–639.

StataCorp (2015) *Stata Statistical Software: Release 14.* College Station, TX: StataCorp LP.

Stephenson, M., Fleetwood, K. and Yellowlees, A. (2015) Alternatives to WinBUGS for network meta-analysis. *Value Health*, **18**, A720.

Stewart, L. A., Clarke, M., Rovers, M., Riley, R. D., Simmonds, M., Stewart, G., Tierney, J. F. and PRISMA-IPD Development Group (2015) Preferred reporting items for systematic review and meta-analyses of individual participant data: The PRISMA-IPD statement. *JAMA*, **313**, 1657–65.

Stewart, L. A. and Tierney, J. F. (2002) To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Evaluation & The Health Professions*, **25**, 76–97.

Sundfor, K., Trope, C. G., Hogberg, T., Onsrud, M., Koern, J., Simonsen, E., Bertelsen, K. and Westberg, R. (1996) Radiotherapy and neoadjuvant chemotherapy for cervical carcinoma. *Cancer*, **77**, 2371–2378.

Sutton, A. J., Kendrick, D. and Coupland, C. A. (2008) Meta-analysis of individual- and aggregate-level data. *Stat Med*, **27**, 651–69.

Symonds, R. P., Habeshaw, T., Reed, N. S., Paul, J., Pyper, E., Yosef, H., Davis, J., Hunter, R., Davidson, S. E., Stewart, A., Cowie, V. and Sarkar, T. (2000) The Scottish and Manchester randomised trial of neo-adjuvant chemotherapy for advanced cervical cancer. *European Journal of Cancer*, **36**, 994–1001.

Tattersall, M. H. N., Lorvidhaya, V., Vootiprux, V., Cheirsilpa, A., Wong, F., Azhar, T., Lee, H. P., Kang, S. B., Manalo, A., Yen, M. S., Kampono, N. and Aziz, F. (1995) Randomized trial of epirubicin and cisplatin chemotherapy followed by pelvic radiation in locally advanced cervical cancer. *Journal of Clinical Oncology*, **13**, 444–451.

Tattersall, M. H. N., Ramirez, C. and Coppleson, M. (1992) A randomized trial comparing platinum-based chemotherapy followed by radiotherapy vs radiotherapy alone in patients with locally advanced cervical cancer. *Int J Gynecol Cancer*, **2**, 244–251.

Thomas, G., Dembo, A., Ackerman, I., Franssen, E., Balogh, J., Fyles, A. and Levin, W. (1998) A randomized trial of standard versus partially hyperfractionated radiation with or without concurrent 5-fluorouracil in locally advanced cervical cancer. *Gynecologic Oncology*, **69**, 137–145.

Thompson, J., Palmer, T. and Moreno, S. (2006) Bayesian analysis in Stata with WinBUGS. *The Stata Journal*, **6**, 530–549.

Thompson, S. G. and Higgins, J. P. (2002) How should meta-regression analyses be undertaken and interpreted? *Stat Med*, **21**, 1559–73.

Thorlund, K., Thabane, L. and Mills, E. J. (2013) Modelling heterogeneity variances in multiple treatment comparison meta-analysis - Are informative priors the better solution? *BMC Med Res Methodol*, **13**, 2.

Tierney, J. F., Vale, C., Riley, R., Tudur Smith, C., Stewart, L., Clarke, M. and Rovers, M. (2015) Individual participant data (IPD) meta-analyses of randomised controlled trials: Guidance on their use. *PLoS Medicine*, **12**, e1001855.

Trinquart, L., Chatellier, G. and Ravaud, P. (2012) Adjustment for reporting bias in network meta-analysis of antidepressant trials. *BMC Med Res Methodol*, **12**, 150.

Trinquart, L., Jacot, J., Conner, S. C. and Porcher, R. (2016) Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *J Clin Oncol*, **34**, 1813–9.

Trovo, M. G., Minatel, E., Franchin, G., Boccieri, M. G., Nascimben, O., Bolzicco, G., Pizzi, G., Torretta, A., Veronesi, A., Gobitti, C., Zanelli, D. J. and Monfardini, S. (1992) Radiotherapy versus radiotherapy enhanced by cisplatin in stage III non-small cell lung cancer. *Int J Radiation Oncology Biol Phys*, **24**, 11–15.

Tu, Y. K. (2014) Use of generalized linear mixed models for network meta-analysis. *Med Decis Making*, **34**, 911–8.

Turner, R. M., Davey, J., Clarke, M. J., Thompson, S. G. and Higgins, J. P. (2012) Predicting the extent of heterogeneity in meta-analysis, using empirical data from the cochrane database of systematic reviews. *Int J Epidemiol*, **41**, 818–27.

Ulutin, H. C., Güden, M., Oysul, K., Sürenkök, S. and Pak, Y. (2000) Split-course radiotherapy with or without concurrent or sequential chemotherapy in non-small cell lung cancer. *Radiat Med*, **18**, 93–96.

Vale, C. L., Tierney, J. F. and Burdett, S. (2013) Can trial quality be reliably assessed from published reports of cancer trials: Evaluation of risk of bias assessments in systematic reviews. *BMJ*, **346**, f1798.

van Valkenhoef, G., Dias, S., Ades, A. E. and Welton, N. J. (2016) Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. *Res Synth Methods*, **7**, 80–93.

Van Houtte, P., Klastersky, J., Renaud, A., Michel, J., Vandermoten, G., Nguyen, H., Sculier, J. P., Devriendt, J. and Mommen, P. (1988) Induction chemotherapy with cisplatin, etoposide and vindesine before radiation therapy with non-small-cell lung cancer. *Antibiot Chemother*, **41**.

Veroniki, A. A., Straus, S. E., Soobiah, C., Elliott, M. J. and Tricco, A. C. (2016) A scoping review of indirect comparison methods and applications using individual patient data. *BMC Med Res Methodol*, **16**, 47.

Veroniki, A. A., Vasiliadis, H. S., Higgins, J. P. and Salanti, G. (2013) Evaluation of inconsistency in networks of interventions. *Int J Epidemiol*, **42**, 332–45.

Warren, F. C., Abrams, K. R. and Sutton, A. J. (2014) Hierarchical network meta-analysis models to address sparsity of events and differing treatment classifications with regard to adverse outcomes. *Stat Med*, **33**, 2449–66.

Wei, Y. and Higgins, J. P. (2013) Bayesian multivariate meta-analysis with multiple outcomes. *Stat Med*, **32**, 2911–34.

Wei, Y., Royston, P., Tierney, J. F. and Parmar, M. K. (2015) Meta-analysis of time-to-event outcomes from randomized trials using restricted mean survival time: Application to individual participant data. *Stat Med*, **34**, 2881–98.

Welton, N. J., Caldwell, D. M., Adamopoulos, E. and Vedhara, K. (2009) Mixed treatment comparison meta-analysis of complex interventions: Psychological interventions in coronary heart disease. *Am J Epidemiol*, **169**, 1158–65.

White, I. R. (2011) Multivariate random-effects meta-regression: Updates to mvmeta. *The Stata Journal*, **11**, 255–270.

— (2015) Network meta-analysis. *The Stata Journal*, **15**, 951–985.

White, I. R., Barrett, J. K., Jackson, D. and Higgins, J. P. T. (2012) Consistency and inconsistency in network meta-analysis: Model estimation using multivariate meta-regression. *Research Synthesis Methods*, **3**, 111–125.

White, J. E., Chen, T., Reed, R., Mira, J., Stuckey, W. J., Weatherall, Y., O'Bryan, R., Samson, M. K. and Seydel, H. G. (1982) Limited squamous cell carcinoma of the lung: A Southwest Oncology Group randomized study of radiation with or without doxorubicin chemotherapy and with or without levamisole immunotherapy. *Cancer Treat Rep*, **66**, 1113–1120.

Whitney, C. W., Sause, W., Bundy, B. N., Malfetano, J. H., Hannigan, E. V., Fowler Jr, W. C., Clarke-Pearson, D. L. and Liao, S. (1999) Randomized comparison of fluorouracil

plus cisplatin versus hydroxyurea as an adjunct to radiation therapy in stage IIB-IVA carcinoma of the cervix with negative para-aortic lymph nodes: A Gynecologic Oncology Group and Southwest Oncology Group study. *Journal of Clinical Oncology*, **17**, 1339–1348.

Zhang, J., Carlin, B., Neaton, J., soon, G. G., Nie, L., Kane, R., Virnig, B. A. and Chu, H. (2014) Network meta-analysis of randomized clinical trials: Reporting the proper summaries. *Clinical Trials*, **11**, 246–262.

Zhao, H., Hodges, J. S., Ma, H., Jiang, Q. and Carlin, B. P. (2016) Hierarchical Bayesian approaches for detecting inconsistency in network meta-analysis. *Stat Med*, **35**, 3524–3536.

A Cervical Cancer Network

This appendix contains additional tables relating to the cervical cancer network.

Table A.1: Cervical cancer trial baseline characteristics. Values for age are mean (min-max). Values for follow-up time are median (interquartile range). RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

Trial	Treatments	Number of Patients	Number of OS events (%)	Stage of disease	Age at randomisation	Follow-up time (years)	P-value from χ^2 PH test
Keys	RT	189	69 (36.5%)	1B	42 (22-78)	8.4 (5.8-10.0)	0.0475
	CTRT	185	49 (26.5%)		40 (20-81)	8.7 (5.4-10.1)	
Hongwei A	RT	30	8 (26.7%)	2B-3B	48.5 (31-69)	5.6 (5.4-5.7)	0.5486
	CTRT	30	8 (26.7%)		46.5 (31-64)	5.5 (5.4-5.7)	
Hongwei B	RT	30	7 (23.3%)	2B-3B	45.5 (30-60)	5.7 (5.5-5.7)	0.2064
	CTRT	30	6 (20.0%)		43.5 (25-63)	5.5 (5.4-5.6)	
Pearcey	RT	129	60 (46.5%)	1B-4A	46 (25-74)	10.1 (9.0-11.8)	0.5134
	CTRT	130	53 (40.8%)		45 (25-75)	10.4 (9.0-12.3)	
Pras	RT	26	16 (61.5%)	Unknown	47 (28-65)	5.1 (3.5-7.1)	0.9632
	CTRT	28	17 (60.7%)		46.5 (34-73)	3.7 (2.8-6.6)	
Leborgne (CTRT)	RT	170	85 (50.0%)	1A-4B	47 (22-70)	4.5 (2.6-6.9)	0.1750
	CTRT	170	75 (44.1%)		45 (21-71)	4.1 (2.4-5.6)	
Thomas A	RT	58	32 (55.2%)	1B-4A	47.5 (27-83)	6.2 (4.6-7.1)	0.0846
	CTRT	57	34 (59.6%)		47 (23-75)	6.8 (4.9-7.8)	
Thomas B	RT	60	25 (41.7%)	1B-4A	49 (26-84)	5.7 (3.7-7.2)	0.5691
	CTRT	58	26 (44.8%)		46.5 (24-81)	5.5 (4.3-7.6)	
Lorvidhaya A	RT	242	59 (24.4%)	2B-4A	52 (24-68)	4.2 (1.5-6.4)	0.0451
	CTRT	233	40 (17.2%)		49 (24-67)	5.2 (2.0-6.7)	
Lorvidhaya B	RT	221	49 (22.2%)	2B-4A	51 (24-68)	4.4 (1.4-6.5)	0.7051
	CTRT	230	54 (23.5%)		51 (28-68)	5.4 (1.9-7.2)	
Roberts	RT	124	39 (31.5%)	1B-4A	44.5 (26-75)	4.4 (2.0-7.3)	0.1688
	CTRT	124	25 (20.2%)		44.5 (21-76)	3.9 (1.6-6.6)	
Onishi	RT	23	15 (65.2%)	2B-4B	76 (38-91)	7.9 (4.4-8.5)	0.1054
	CTRT	26	16 (61.5%)		64.5 (28-83)	6.5 (5.8-7.2)	

Table A.1: Cervical cancer trial baseline characteristics. Values for age are mean (min-max). Values for follow-up time are median (interquartile range). RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

Trial	Treatments	Number of Patients	Number of OS events (%)	Stage of disease	Age at randomisation	Follow-up time (years)	P-value from χ^2 PH test
Lanciano	RT	24	12 (50.0%)	2B-4A	45.4 (29-70)	7.3 (6.7-7.5)	0.1259
	CTRT	53	19 (35.8%)		44 (28-83)	6.5 (4.1-7.4)	
Cikaric	RT	100	48 (48.0%)	2B-3B	48 (30-73)	3.5 (3.2-4.0)	0.6753
	CTRT	100	37 (37.0%)		49.5 (29-74)	3.6 (3.1-4.1)	
Garipagaoglu	RT	22	8 (36.4%)	2B-3B	49.5 (33-62)	4.3 (4.2-5.0)	0.9689
	CTRT	22	9 (40.9%)		49.5 (38-68)	4.7 (4.2-5.2)	
Lal	RT	86	12 (14.0%)	1B-4B	52 (25-76)	2.8 (1.1-4.2)	0.0834
	CTRT	94	14 (14.9%)		50 (24-70)	2.4 (1.0-3.4)	
Cardenas 93	RT	16	8 (50.0%)	3B	45.5 (30-60)	6.1 (4.0-6.1)	0.6022
	CT+RT	14	12 (85.7%)		48 (27-60)	11.4 (5.6-11.4)	
Chauvergne	RT	90	54 (60.0%)	2B-3B	53.5 (27-74)	8.6 (5.9-14.0)	0.4818
	CT+RT	92	57 (62.0%)		54.5 (28-74)	9.2 (7.0-11.0)	
Kumar	RT	85	34 (40.0%)	1B-3A	47 (21-65)	8.8 (2.2-10.0)	0.3625
	CT+RT	88	49 (55.7%)		45 (30-65)	9.8 (6.6-10.3)	
Leborgne (NeoCT)	RT	49	28 (57.1%)	1B-4A	44 (19-63)	7.3 (6.3-7.7)	0.6543
	CT+RT	48	32 (66.7%)		44 (24-71)	7.6 (6.7-8.4)	
Sardi 96	RT	54	41 (75.9%)	3B	49 (28-68)	7.3 (6.5-8.3)	0.1206
	CT+RT	54	34 (63.0%)		47 (27-68)	7.6 (7.0-8.0)	
	CT+S	53	25 (47.2%)		49 (31-69)	7.5 (6.4-8.4)	
Sardi 97	RT	106	32 (30.2%)	1B	41.5 (24-63)	7.9 (6.2-9.1)	0.4390
	CT+RT	104	19 (18.3%)		39 (24-64)	8.0 (6.7-9.1)	
Sardi 98	RT	74	33 (44.6%)	2B	40 (27-65)	7.0 (5.8-8.2)	0.0738
	CT+RT	73	30 (41.1%)		41 (27-66)	7.3 (6.2-8.6)	
	CT+S	80	22 (27.5%)		45.5 (25-66)	7.2 (6.4-7.9)	

Table A.1: Cervical cancer trial baseline characteristics. Values for age are mean (min-max). Values for follow-up time are median (interquartile range). RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

Trial	Treatments	Number of Patients	Number of OS events (%)	Stage of disease	Age at randomisation	Follow-up time (years)	P-value from χ^2 PH test
Souhami	RT	55	31 (56.4%)	3B	49 (26-69)	4.3 (3.6-4.6)	0.2965
	CT+RT	48	29 (60.4%)		47 (23-69)	4.1 (3.7-5.0)	
Sundfor	RT	48	35 (72.9%)	3B-4A	52 (26-69)	5.4 (4.1-5.7)	0.4476
	CT+RT	48	31 (64.6%)		51.5 (25-68)	4.3 (3.9-5.8)	
Symonds	RT	110	76 (69.1%)	2B-4A	48 (24-70)	6.0 (5.0-7.5)	0.4843
	CT+RT	105	68 (64.8%)		49 (25-69)	6.8 (5.6-7.6)	
Tattershall 92	RT	37	18 (48.6%)	2B-4A	56 (23-74)	4.3 (3.5-5.0)	0.2406
	CT+RT	34	20 (58.8%)		54.5 (33-70)	4.5 (4.1-5.3)	
Tattershall 95	RT	131	28 (21.4%)	2B-4A	52 (27-77)	1.8 (0.8-2.8)	0.5173
	CT+RT	129	38 (29.5%)		47 (25-74)	1.7 (0.7-2.4)	
LGOG	RT	12	2 (16.7%)	1B-3B	46 (27-72)	6.1 (6.0-7.6)	0.5866
	CT+RT	15	9 (60.0%)		41 (21-63)	6.5 (6.3-7..5)	
MRC CeCa	RT	24	9 (37.5%)	1B-3A	50 (26-75)	4.2 (3.7-5.2)	0.8222
	CT+RT	24	19 (79.2%)		48 (29-72)	7.1 (4.1-7.1)	
PMB	RT	19	15 (78.9%)	1B-4A	50 (30-73)	9.8 (9.1-10.2)	0.5882
	CT+RT	16	9 (56.3%)		46 (25-67)	9.6 (4.4-10.5)	
Chiara	RT	32	16 (50.0%)	2B-4A	59.5 (32-75)	8.7 (7.7-9.6)	0.6144
	CT+RT	32	22 (68.8%)		59 (32-76)	9.5 (7.9-9.9)	
Herod	RT	88	62 (70.5%)	1B-4A	46 (27-73)	9.5 (7.7-10.8)	0.1540
	CT+RT	89	68 (76.4%)		48 (24-74)	9.1 (6.5-11.1)	
Cardenas 91	RT	18	9 (50.0%)	2B	43.5 (24-61)	9.1 (3.7-10.1)	0.9242
	CT+RT	13	7 (53.8%)		42 (35-56)	3.6 (3.2-10.1)	
Benedetti	RT	214	101 (47.2%)	1B-3B	53 (27-999)	4.0 (2.9, 5.0)	0.1491
	CT+S	227	88 (38.8%)		49 (25-999)	4.4 (3.5-5.2)	

Table A.1: Cervical cancer trial baseline characteristics. Values for age are mean (min-max). Values for follow-up time are median (interquartile range). RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

Trial	Treatments	Number of Patients	Number of OS events (%)	Stage of disease	Age at randomisation	Follow-up time (years)	P-value from χ^2 PH test
Kigawa	RT	25	15 (60.0%)	2B-3B	61 (43-69)	9.0 (8.7-9.6)	0.3616
	CT+S	25	10 (40.0%)		57 (41-67)	9.0 (8.1-9.6)	
Chang	RT	52	12 (23.1%)	1B-2A	45.5 (32-70)	5.3 (3.7-6.8)	0.5745
	CT+S	68	21 (30.9%)		44 (30-69)	5.6 (4.5-8.1)	

Table A.2: Radiotherapy versus neoadjuvant chemotherapy plus surgery *metaninf* results. Values are log hazard ratios and 95% confidence intervals.

Trial omitted	LogHR (95% CI)
Sardi 96 (1996)	-0.342 $(-0.569, -0.114)$
Sardi 98 (1998)	-0.390 $(-0.614, -0.166)$
Benedetti (2002)	-0.543 $(-0.841, -0.245)$
Kigawa (1996)	-0.432 $(-0.646, -0.217)$
Chang (2000)	-0.511 $(-0.728, -0.294)$
All trials	-0.436 $(-0.642, -0.229)$

Table A.3: Knot locations from restricted cubic spline for each cervical cancer trial. Values are ln(time). Three trials had one interior knot only which is represented by n/a in the 2nd knot column.

Trial ID	Minimum	Percentile	1st Knot	Percentile	2nd Knot	Percentile	Maximum	Percentile
Keys	−0.96	0%	−0.5	5%	1	63%	2.41	100%
Hongwei A	−0.69	0%	−0.5	21%	1	86%	1.05	100%
Hongwei B	−0.02	0%	0.9	44%	1.1	57%	1.66	100%
Pearcey	−0.76	0%	0.2	32%	1	69%	2.67	100%
Pras	−0.92	0%	−0.2	11%	1	74%	1.84	100%
Leborgne CTRT	−3.60	0%	1.5	96%	n/a	n/a	2.05	100%
Thomas A	−2.77	0%	0.2	34%	1	75%	2.00	100%
Thomas B	−1.06	0%	0	35%	1.5	94%	1.77	100%
Lorvidhaya A	−3.13	0%	−1.5	6%	0.5	72%	1.81	100%
Lorvidhaya B	−2.40	0%	−0.1	34%	0.6	67%	2.04	100%
Roberts	−1.70	0%	0	35%	1	75%	1.95	100%
Onishi	−0.94	0%	−0.5	15%	1	65%	2.10	100%
Lanciano	−3.82	0%	−0.5	12%	1	70%	1.89	100%
Cikaric	−2.53	0%	−0.2	20%	0.8	80%	1.52	100%
Garipagaoglu	−0.52	0%	0.4	30%	1	67%	1.50	100%
Lal	−2.16	0%	−0.5	26%	0.8	92%	0.94	100%
Cardenas 93	−0.81	0%	1	78%	1.5	85%	2.09	100%
Chauvergne	−3.19	0%	−0.5	14%	0.5	51%	2.84	100%
Kumar	−1.77	0%	0	45%	1	83%	2.28	100%
Leborgne CT+RT	−2.14	0%	−1	6%	0	31%	1.99	100%

Table A.3: Knot locations from restricted cubic spline for each cervical cancer trial. Values are ln(time). Three trials had one interior knot only which is represented by n/a in the 2nd knot column.

Trial ID	Minimum	Percentile	1st Knot	Percentile	2nd Knot	Percentile	Maximum	Percentile
Sardi 96	−1.79	0%	−1	4%	1	88%	1.88	100%
Sardi 97	−0.26	0%	1	55%	1.5	83%	1.99	100%
Sardi 98	−0.09	0%	0.5	21%	1	53%	2.19	100%
Souhami	−2.03	0%	−0.5	30%	0.5	76%	1.26	100%
Sundfor	−2.81	0%	−0.5	10%	1	86%	1.78	100%
Symonds	−2.77	0%	−1	9%	1	82%	2.07	100%
Tattershall 92	−2.24	0%	0	39%	1	80%	1.34	100%
Tattershall 95	−3.34	0%	−1	12%	1	98%	1.02	100%
LGOG	−0.44	0%	0.6	68%	n/a	n/a	1.16	100%
MRC CeCa	−0.53	0%	1	83%	1.5	94%	1.72	100%
PMB	−1.97	0%	1	85%	n/a	n/a	2.24	100%
Chiara	−0.97	0%	1	62%	1.5	81%	2.31	100%
Herod	−5.21	0%	1	79%	1.5	93%	2.13	100%
Cardenas 91	−0.57	0%	0	23%	1	83%	2.23	100%
Benedetti	−2.53	0%	−1	7%	1	81%	1.73	100%
Kigawa	−1.47	0%	−1	5%	1.5	85%	1.97	100%
Chang	−0.75	0%	−0.3	22%	0.5	60%	1.81	100%

Table A.4: Log hazard ratios (95% credible intervals) for cervical cancer trials from a fixed treatment effect Royston-Parmar model fitted in WinBUGS. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

Trial	Treatment	LogHR (95% CrI)
Cikaric (2005)	RT v CTRT	-0.388 (-0.824, 0.037)
Garipagaoglu (2004)	RT v CTRT	0.197 (-0.786, 1.165)
Hongwei A (1997)	RT v CTRT	0.061 (-0.981, 1.091)
Hongwei B (1997)	RT v CTRT	-0.131 (-1.295, 1.001)
Keys (1999)	RT v CTRT	-0.446 (-0.826, -0.083)
Lal (2004)	RT v CTRT	0.127 (-0.651, 0.915)
Lanciano (2005)	RT v CTRT	-0.396 (-1.102, 0.348)
Leborgne CTRT (unpublished)	RT v CTRT	-0.079 (-0.382, 0.233)
Lorvidhaya A (2003)	RT v CTRT	-0.515 (-0.926, -0.110)
Lorvidhaya B (2003)	RT v CTRT	0.009 (-0.372, 0.398)
Onishi (1999)	RT v CTRT	0.230 (-0.486, 0.956)
Pearcey (2002)	RT v CTRT	-0.179 (-0.559, 0.194)
Pras (unpublished)	RT v CTRT	-0.009 (-0.714, 0.684)
Roberts (2000)	RT v CTRT	-0.535 (-1.061, -0.026)
Thomas A (1998)	RT v CTRT	-0.373 (-0.910, 0.149)
Thomas B (1998)	RT v CTRT	0.052 (-0.498, 0.616)
Cardenas 91 (1991)	RT v CT+RT	0.099 (-0.914, 1.113)
Cardenas 93 (1993)	RT v CT+RT	0.521 (-0.373, 1.479)
Chauvergne (1993)	RT v CT+RT	-0.016 (-0.386, 0.353)
Chiara (1994)	RT v CT+RT	0.519 (-0.128, 1.179)
Herod (2000)	RT v CT+RT	0.081 (-0.262, 0.429)
Kumar (1994)	RT v CT+RT	0.367 (-0.066, 0.810)
Leborgne CT+RT (1997)	RT v CT+RT	0.194 (-0.310, 0.692)
LGOG (unpublished)	RT v CT+RT	1.776 (0.271, 3.814)
MRC CeCa (unpublished)	RT v CT+RT	1.280 (0.510, 2.128)
PMB (unpublished)	RT v CT+RT	-0.462 (-1.328, 0.370)
Sardi 96 (1996)	RT v CT+RT	-0.578 (-1.043, -0.112)
Sardi 97 (1997)	RT v CT+RT	-0.655 (-1.236, -0.080)
Sardi 98 (1998)	RT v CT+RT	-0.298 (-0.801, 0.199)
Souhami (1991)	RT v CT+RT	0.538 (0.019, 1.045)
Sundfor (1996)	RT v CT+RT	-0.195 (-0.683, 0.291)
Symonds (2000)	RT v CT+RT	-0.168 (-0.493, 0.156)
Tattershall 92 (1992)	RT v CT+RT	0.214 (-0.424, 0.847)
Tattershall 95 (1995)	RT v CT+RT	0.500 (0.003, 0.996)

Table A.4: Log hazard ratios (95% credible intervals) for each cervical cancer trial from a fixed treatment effect model. RT = radiotherapy, CTRT = chemoradiation, CT+RT = neoadjuvant chemotherapy plus radiotherapy, CT+S = neoadjuvant chemotherapy plus surgery.

Trial	Treatment	LogHR (95% CrI)
Benedetti (2002)	RT v CT+S	-0.340 (-0.628, -0.054)
Chang (2000)	RT v CT+S	0.338 (-0.376, 1.091)
Kigawa (1996)	RT v CT+S	-0.534 (-1.376, 0.275)
Sardi 96 (1996)	RT v CT+S	-0.910 (-1.428, -0.418)
Sardi 98 (1998)	RT v CT+S	-0.737 (-1.286, -0.204)

B Lung Cancer Network

This appendix contains additional tables relating to the lung cancer network.

Table B.1: Lung cancer trial baseline characteristics. Values for age are mean (min-max). Values for follow-up time are median (interquartile range). RT = radiotherapy, SeqCT = sequential chemotherapy, ConCT = concomitant chemotherapy.

Trial	Treatments	Number of Patients	Number of OS events (%)	Stage of disease	Age at randomisation	Follow-up time (years)	P-value from χ^2 PH test
Brussels	RT	34	31 (91.2%)	1 - 3A	60.3 (45-75)	0.9 (0.4-1.6)	0.6404
	SeqCT	31	27 (87.1%)		63.5 (46-73)	0.6 (0.3-1.2)	
Essen	RT	26	22 (84.6%)	1-3B	60.3 (46-71)	1.4 (1.0-2.6)	0.1718
	SeqCT	22	21 (95.5%)		57.6 (42-67)	1.6 (1.2-2.1)	
SLCSG	RT	164	161 (98.2%)	1-4	63.6 (38-76)	0.8 (0.4-1.5)	0.5246
	SeqCT	163	159 (97.6%)		64.1 (37-76)	0.8 (0.5-1.5)	
WSLCRG-FI	RT	39	35 (89.7%)	Unknown	62.3 (43-70)	0.9 (0.5-1.4)	0.9429
	SeqCT	40	37 (92.5%)		59.6 (39-70)	1.0 (0.5-1.7)	
Perugia	RT	33	32 (97.0%)	2-4	60.4 (45-70)	0.7 (0.4-1.7)	0.2934
	SeqCT	33	32 (97.0%)		60.2 (47-69)	0.8 (0.4-2.2)	
CALGB 8433	RT	91	83 (91.2%)	1-3B	61.5 (45-77)	0.8 (0.3-1.4)	0.2235
	SeqCT	89	78 (87.6%)		58.6 (38-78)	1.0 (0.6-1.9)	
EORTC 08842	RT	37	37 (100%)	2-4	60.7 (53-67)	0.9 (0.5-1.4)	0.996
	SeqCT	38	37 (97.4%)		58.9 (39-68)	0.9 (0.6-1.7)	
CEBI 138	RT	177	177 (100%)	1-4	58.9 (36-73)	0.8 (0.5-1.4)	0.9392
	SeqCT	176	174 (98.9%)		57.8 (32-76)	1.0 (0.6-1.7)	
SWOG 8300a	RT	64	62 (96.9%)	3A-4	61.5 (44-78)	0.9 (0.4-1.7)	0.5787
	SeqCT	64	62 (96.9%)		58.9 (30-82)	0.9 (0.5-1.6)	
SWOG 8300b	RT	63	63 (100%)	0-4	60.5 (37-78)	0.7 (0.4-1.2)	0.4176
	SeqCT	63	63 (100%)		61.3 (42-77)	0.7 (0.4-1.3)	
MIC1	RT	232	216 (93.1%)	Unknown	62.5 (35-75)	0.7 (0.4-1.5)	0.5389
	SeqCT	229	211 (92.1%)		61.9 (37-75)	1.0 (0.4-1.6)	
RTOG 8808-ECOG 458	RT	162	155 (95.7%)	2-4	61.0 (35-78)	0.9 (0.5-1.6)	0.9502
	SeqCT	164	153 (93.3%)		60.9 (36-84)	1.1 (0.7-2.3)	
New Delhi	RT	252	213 (84.5%)	Unknown	52.0 (39-69)	1.0 (0.8-1.4)	<0.0001

Table B.1: Lung cancer trial baseline characteristics. Values for age are mean (min-max). Values for follow-up time are median (interquartile range). RT = radiotherapy, SeqCT = sequential chemotherapy, ConCT = concomitant chemotherapy.

Trial	Treatments	Number of Patients	Number of OS events (%)	Stage of disease	Age at randomisation	Follow-up time (years)	P-value from χ^2 PH test
	SeqCT	251	185 (73.7%)		52.5 (39-70)	1.4 (1.1-1.7)	
Buenos Aires	RT	38	35 (92.1%)	1-3B	61.7 (39-77)	0.6 (0.4-1.2)	0.5436
	SeqCT	43	43 (100%)		57.0 (33-74)	0.9 (0.5-1.3)	
FLCSG 2	RT	127	126 (99.2%)	1-4	62.0 (39-71)	0.8 (0.5-1.4)	0.7321
	SeqCT	125	124 (99.2%)		61.8 (46-72)	0.9 (0.4-1.6)	
Gwent 3	RT	44	43 (97.7%)	Unknown	63.6 (54-74)	1.0 (0.5-1.8)	0.7839
	SeqCT	41	40 (97.6%)		63.0 (47-73)	0.6 (0.3-2.3)	
Tax SI009	RT	105	88 (83.8%)	1-3B	60.8 (32-81)	0.9 (0.4-1.6)	0.3304
	SeqCT	103	79 (76.7%)		61.4 (37-78)	1.0 (0.5-1.6)	
Gwent 1	RT	30	30 (100%)	Unknown	63.7 (47-76)	0.6 90.2-1.4)	0.6421
	SeqCT	26	23 (88.5%)		61.7 (45-73)	1.0 (0.3-1.9)	
SWOG 7635	RT	32	23 (71.9%)	Unknown	62.0 (51-72)	0.8 (0.1-1.6)	0.52
	SeqCT	30	25 (83.3%)		61.5 (49-75)	0.5 (0.3-1.1)	
NCCTG 822451	RT	63	59 (93.7%)	1-4	62.2 (39-76)	0.8 (0.4-1.9)	0.955
	SeqCT	58	54 (93.1%)		61.3 (38-77)	0.9 (0.5-1.8)	
BLT4 (Adjuvant)	RT	60	53 (83.3%)	1-4	62.6 (40-77)	1.0 (0.4-2.1)	0.3187
	SeqCT	59	53 (89.8%)		62.7 (45-77)	0.9 (0.4-1.6)	
BLT4 (Neo-adjuvant)	RT	86	68 (79.1%)	1-4	64.6 (47-80)	1.1 (0.7-1.9)	0.8422
	SeqCT	83	70 (84.4%)		65.0 (50-77)	1.2 (0.5-2.1)	
EORTC 08844	RT	114	111 (97.4%)	1-4	57.9 (39-70)	1.0 (0.5-1.4)	0.4357
	ConCT	217	211 (97.2%)		58.9 (37-70)	1.0 (0.5-1.8)	
HOG LUN 86 1	RT	120	113 (94.2%)	1-4	60.4 (31-81)	0.8 (0.4-1.3)	0.497
	ConCT	177	105 (89.7%)		62.4 (37-77)	0.7 (0.4-1.5)	
Aviano	RT	88	88 (100%)	2-3B	61.7 (43-70)	0.8 (0.5-1.5)	0.7501
	ConCT	85	85 (100%)		60.8 (36-69)	0.8 (0.6-1.4)	

Table B.1: Lung cancer trial baseline characteristics. Values for age are mean (min-max). Values for follow-up time are median (interquartile range). RT = radiotherapy, SeqCT = sequential chemotherapy, ConCT = concomitant chemotherapy.

Trial	Treatments	Number of Patients	Number of OS events (%)	Stage of disease	Age at randomisation	Follow-up time (years)	P-value from χ^2 PH test
PMCI 88 C091	RT	101	100 (99.0%)	1-3B	63.7 (40-78)	1.2 (0.7-2.2)	0.1124
	ConCT	107	107 (100%)		65.9 (46-79)	1.4 (0.6-2.6)	
CALGB-ECOG	RT	136	128 (94.1%)	2-4	61.9 (35-80)	1.2 (0.7-2.4)	0.964
	ConCT	146	139 (95.2%)		61.5 (38-79)	1.1 (0.7-2.5)	
NKB-CKVO 94 11	RT	83	81 (97.6%)	2-3B	60.4 (38-75)	1.0 (0.6-2.2)	0.7842
	ConCT	77	76 (98.7%)		59.6 (38-75)	1.0 (0.6-1.5)	
NPC IIIB 96-01	RT	292	263 (90.0%)	1-3B	59.1 (29-80)	0.9 (0.5-1.8)	0.3832
	ConCT	292	250 (85.6%)		60.1 (34-80)	1.2 (0.6-2.0)	
Kragujevac 88	RT	61	58 (95.1%)	3A-3B	56.1 (38-70)	0.7 (0.5-1.9)	0.7575
	ConCT	108	88 (81.5%)		56.4 (41-68)	1.3 (0.5-2.1)	
Kragujevac 90	RT	66	60 (90.9%)	3A-3B	58.4 (46-65)	1.2 (0.8-2.0)	0.7034
	ConCT	65	50 (76.9%)		58.5 (42-67)	1.8 (0.9-2.9)	
NCCTG 90 24 51	RT	36	32 (88.9%)	3A-3B	63.2 (47-81)	1.0 (0.5-3.5)	0.4809
	ConCT	38	34 (89.5%)		63.0 (47-82)	0.9 (0.4-2.2)	
ACR LAMP 427	RT	97	85 (87.6%)	1-4	60.1 (40-79)	1.0 (0.7-2.2)	0.2724
	ConCT	80	69 (86.3%)		62.1 (27-78)	1.0 (0.5-2.0)	
Uludag	RT	23	23 (100%)	Unknown	62.4 (36-77)	0.8 (0.4-2.3)	0.9979
	ConCT	22	20 (90.9%)		54.3 (36-68)	1.2 (0.9-2.3)	
Brocat Study Group	RT	113	90 (79.7%)	Unknown	60.7 (36-77)	1.2 (0.7-2.1)	0.7705
	ConCT	99	70 (70.7%)		61.9 (34-76)	1.4 (0.7-3.0)	
GMMA Ankara 1997	RT	26	24 (92.3%)	3A-3B	62.5 (41-75)	0.8 (0.6-1.1)	0.2688
	ConCT	25	20 (80.0%)		62.1 (43-75)	1.3 (0.8-2.2)	
Tax 206	RT	46	33 (71.7%)	3A-3B	58.9 (38-74)	1.2 (0.7-1.5)	0.1397
	ConCT	43	27 (62.8%)		58.8 (42-73)	1.1 (0.7-1.6)	
JCOG 9812	RT	23	17 (73.9%)	Unknown	76.9 (72-84)	1.2 (0.5-1.9)	0.8796

Table B.1: Lung cancer trial baseline characteristics. Values for age are mean (min-max). Values for follow-up time are median (interquartile range). RT = radiotherapy, SeqCT = sequential chemotherapy, ConCT = concomitant chemotherapy.

Trial	Treatments	Number of Patients	Number of OS events (%)	Stage of disease	Age at randomisation	Follow-up time (years)	P-value from χ^2 PH test
	ConCT	23	15 (65.2%)		76.7 (71-83)	1.5 (0.8-2.2)	
WJLCG	SeqCT	158	142 (89.9%)	3A-3B	62.3 (39-75)	1.1 (0.6-2.1)	0.9323
	ConCT	156	131 (84.0%)		61.4 (40-75)	1.4 (0.8-2.8)	
RTOG 9410	SeqCT	203	189 (93.1%)	1-4	61.1 (33-79)	1.2 (0.6-2.3)	0.6076
	ConCT	204	180 (88.2%)		59.8 (33-79)	1.4 (0.7-3.2)	
GLOT-GFPC NOC	SeqCT	103	96 (93.2%)	3A-3B	56.5 (38-70)	1.1 (0.6-2.3)	0.1266
	ConCT	102	87 (85.3%)		55.9 (38-69)	1.3 (0.5-2.8)	
EORTC 08972	SeqCT	78	66 (84.6%)	1-3B	63.5 (46-79)	1.4 (0.9-2.4)	0.3078
	ConCT	80	63 (78.8%)		62.1 (36-79)	1.4 (0.7-2.8)	
CALGB 8831	SeqCT	45	39 (86.7%)	1-4	61.7 (34-82)	1.0 (0.5-2.3)	0.0182
	ConCT	46	45 (97.8%)		59.1 (39-75)	1.1 (0.7-2.1)	
GMMA Ankara 1995	RT	15	15 (100%)	3A-3B	62.5 (39-74)	0.8 (0.4-1.3)	0.7545
	SeqCT	15	15 (100%)		61.3 (33-75)	1.0 (0.6-1.2)	
	ConCT	15	15 (100%)		62.7 (35-75)	0.8 (0.4-1.4)	

Table B.2: Radiotherapy versus sequential chemotherapy *metaninf* results. Values are log hazard ratios and 95% confidence intervals.

Trial omitted	LogHR (95% CI)
Brussels (Van Houtte et al., 1988)	$-0.198\ (-0.265, -0.131)$
Essen (Alberti et al., 1990)	$-0.194\ (-0.261, -0.127)$
SLCSG (Brodin et al., 1996)	$-0.192\ (-0.262, -0.122)$
WSLCRG-FI (Gregor et al., 1993)	$-0.191\ (-0.259, -0.124)$
Perugia (Crino et al., 1993)	$-0.187\ (-0.254, -0.120)$
CALGB 8433 (Dillman et al., 1990)	$-0.181\ (-0.249, -0.113)$
EORTC 08842 (Planting et al., 1996)	$-0.190\ (-0.257, -0.123)$
CEBI 138 (Le Chevalier et al., 1991)	$-0.180\ (-0.250, -0.110)$
SWOG 8300a (Mira et al., 1990)	$-0.192\ (-0.260, -0.125)$
SWOG 8300b (Mira et al., 1990)	$-0.199\ (-0.267, -0.132)$
MIC1 (CRC TU LU3001) (Cullen et al., 1999)	$-0.194\ (-0.265, -0.123)$
RTOG 8808 - ECOG 458 (Sause et al., 2000)	$-0.177\ (-0.247, -0.108)$
New Delhi (Sharma et al., 2003)	$-0.130\ (-0.200, -0.060)$
Buenos Aires (Cardiello et al., 1985)	$-0.189\ (-0.256, -0.122)$
FLCSG 2 (Mattson et al., 1988)	$-0.201\ (-0.270, -0.132)$
Gwent 3 (unpublished)	$-0.194\ (-0.261, -0.126)$
Tax SI009 (Mattson, 2003)	$-0.190\ (-0.258, -0.122)$
Gwent 1 (Anderson et al., 1981)	$-0.185\ (-0.252, -0.118)$
SWOG 7635 (White et al., 1982)	$-0.199\ (-0.266, -0.132)$
NCCTG 822451 (Morton et al., 1991)	$-0.194\ (-0.262, -0.127)$
BLT4 (Adjuvant) (Fairlamb et al., 2005)	$-0.199\ (-0.266, -0.131)$
BLT4 (Neo-adjuvant) (Fairlamb et al., 2005)	$-0.199\ (-0.267, -0.131)$
GMMA Ankara 1995 (Ulutin et al., 2000)	$-0.191\ (-0.258, -0.124)$
All trials	$-0.189\ (-0.256, -0.123)$

Table B.3: Log hazard ratios (95% credible intervals) for lung cancer trials from a fixed treatment effect Royston-Parmar model fitted in WinBUGS. RT = radiotherapy, SeqCT = sequential chemotherapy, ConCT = concomitant CT.

Trial	Treatment	LogHR (95% CrI)
BLT4 (Adjuvant) (Fairlamb et al., 2005)	RT v SeqCT	0.124 (−0.255, 0.503)
BLT4 (Neo-adjuvant) (Fairlamb et al., 2005)	RT v SeqCT	0.050 (−0.275, 0.390)
Brussels (Van Houtte et al., 1988)	RT v SeqCT	0.374 (−0.162, 0.895)
Buenos Aires (Cardiello et al., 1985)	RT v SeqCT	−0.165 (−0.609, 0.283)
CALGB 8433 (Dillman et al., 1990)	RT v SeqCT	−0.396 (−0.704, −0.089)
CEBI 138 (Le Chevalier et al., 1991)	RT v SeqCT	−0.281 (−0.489, −0.076)
EORTC 08842 (Planting et al., 1996)	RT v SeqCT	−0.193 (−0.656, 0.266)
Essen (Alberti et al., 1990)	RT v SeqCT	0.268 (−0.353, 0.881)
FLCSG 2 (Mattson et al., 1988)	RT v SeqCT	−0.039 (−0.287, 0.204)
GMMA Ankara 1995 (Ulutin et al., 2000)	RT v SeqCT	0.040 (−0.685, 0.765)
Gwent 1 (Anderson et al., 1981)	RT v SeqCT	−0.606 (−1.191, −0.035)
Gwent 3 (unpublished)	RT v SeqCT	0.027 (−0.402, 0.449)
MIC1 (Cullen et al., 1999)	RT v SeqCT	−0.157 (−0.349, 0.029)
NCCTG 822451 (Morton et al., 1991)	RT v SeqCT	−0.026 (−0.391, 0.344)
Perugia (Crino et al., 1993)	RT v SeqCT	−0.315 (−0.817, 0.173)
RTOG 8808-ECOG 458 (Sause et al., 2000)	RT v SeqCT	−0.317 (−0.536, −0.091)
SLCSG (Brodin et al., 1996)	RT v SeqCT	−0.158 (−0.377, 0.059)
SWOG 7635 (White et al., 1982)	RT v SeqCT	0.607 (−0.002, 1.232)
SWOG 8300a (Mira et al., 1990)	RT v SeqCT	−0.085 (−0.427, 0.260)
SWOG 8300b (Mira et al., 1990)	RT v SeqCT	0.124 (−0.228, 0.479)
Tax SI009 (Mattson, 2003)	RT v SeqCT	−0.174 (−0.475, 0.123)
WSLCRG-FI (Gregor et al., 1993)	RT v SeqCT	−0.065 (−0.529, 0.395)
ACR LAMP 427 (Belani et al., 2005)	RT v ConCT	0.099 (−0.216, 0.410)
Aviano (Trovo et al., 1992)	RT v ConCT	−0.014 (−0.314, 0.285)
Brocat Study Group (Huber et al., 2006)	RT v ConCT	−0.245 (−0.566, 0.069)
CALGB-ECOG (Clamon et al., 1999)	RT v ConCT	0.003 (−0.244, 0.238)
EORTC 08844 (Schaake-Koning et al., 1992)	RT v ConCT	−0.254 (−0.480, −0.019)
GMMA Ankara 1995 (Ulutin et al., 2000)	RT v ConCT	−0.104 (−0.814, 0.628)

Table B.3: Log hazard ratios (95% credible intervals) for each lung cancer trial from a fixed treatment effect model. RT = radiotherapy, SeqCT = sequential chemotherapy, ConCT = concomitant CT.

Trial	Treatment	LogHR (95% CrI)
GMMA Ankara 1997 (Cüneyt Ulutin and Pak, 2003)	RT v ConCT	-0.640 (-1.239, -0.050)
HOG LUN 86 1 (Blanke et al., 1995)	RT v ConCT	-0.118 (-0.381, 0.148)
JCOG 9812 (Atagi et al., 2005)	RT v ConCT	-0.389 (-1.091, 0.316)
Kragujevac 88 (Jeremic et al., 1995)	RT v ConCT	-0.476 (-0.800, -0.146)
Kragujevac 90 (Jeremic et al., 1996)	RT v ConCT	-0.465 (-0.844, -0.076)
NCCTG 90 24 51 (Bonner et al., 1998)	RT v ConCT	0.104 (-0.379, 0.595)
NKB-CKVO 94 11 (Groen, 2004)	RT v ConCT	0.141 (-0.170, 0.447)
NPC IIIB 96-01 (Douillard et al., 2005)	RT v ConCT	-0.141 (-0.313, 0.027)
PMCI 88 C091 (Ball et al., 1999)	RT v ConCT	-0.056 (-0.317, 0.213)
Tax 206 (Scagliotti et al., 2006)	RT v ConCT	-0.158 (-0.664, 0.340)
Uludag (Sarihan et al., 2004)	RT v ConCT	-0.402 (-1.010, 0.206)
CALGB 8831 (Clamon et al., 1994)	SeqCT v ConCT	0.117 (-0.307, 0.533)
EORTC 08972 (Belderbos et al., 2007)	SeqCT v ConCT	-0.028 (-0.368, 0.311)
GLOT-GFPC NOC (Fournel et al., 2005)	SeqCT v ConCT	-0.231 (-0.516, 0.059)
GMMA Ankara 1995 (Ulutin et al., 2000)	SeqCT v ConCT	-0.144 (-0.870, 0.604)
RTOG 9410 (Curran Jr et al., 2011)	SeqCT v ConCT	-0.222 (-0.426, -0.021)
WJLCG (Furuse et al., 1999)	SeqCT v ConCT	-0.250 (-0.488, -0.013)

C WinBUGS model code

In this appendix I include the WinBUGS model code from the models fitted in this thesis and include a description of how the data is set up. The models presented in this section were all applied to the lung cancer network and the index i represents patient and j trial.

In the lung cancer network there are three treatments so two treatment contrasts were parameterised and estimated directly in the model code whilst the third one was calculated through the consistency equations. RT is the reference treatment for trials comparing RT and Seq CT and RT and Con CT. For trials comparing Seq CT and Con CT, Seq CT is the reference treatment and the treatment contrasts need to reflect this. For patients in a Seq CT v Con CT trial receiving Con CT there must be a '-1' for the coefficient of RT v Seq CT and a '1' for the coefficient of RT v Con CT. For patients in a Seq CT v Con CT trial receiving Seq CT the coefficients for RT v Seq CT and RT v Con CT must both be '0'. In the model code I let β_1 be the treatment effect estimate for Seq CT compared to RT and β_2 be the treatment effect estimate for Con CT compared to RT. The treatment effect estimate for Con CT compared to Seq CT can then be calculated as $\beta_2 - \beta_1$. Let trt1$[i]$ be an indicator variable for β_1 for patient i and trt2$[i]$ be an indicator variable for β_2 for patient i where:

$$
\text{trt1}[i] = \begin{cases} 1 & \text{if patient was randomised to Seq CT and is from a trial comparing} \\ & \text{RT and Seq CT} \\ -1 & \text{if patient was randomised to Con CT and is from a trial comparing} \\ & \text{Seq CT and Con CT} \\ 0 & \text{otherwise} \end{cases}
$$

$$
\text{trt2}[i] = \begin{cases} 1 & \text{if patient was randomised to Con CT and is from a trial comparing} \\ & \text{Seq CT and Con CT} \\ 0 & \text{otherwise} \end{cases}
$$

For patients from multi-arm trials:

$$
\text{trt1}[i] = \begin{cases} 1 & \text{if patient was randomised to Seq CT} \\ 0 & \text{otherwise} \end{cases}
$$

$$\text{trt2}[i] = \begin{cases} 1 & \text{if patient was randomised to Con CT} \\ 0 & \text{otherwise} \end{cases}$$

Throughout the model code lines that begin with # represent comments and are ignored by WinBUGS when compiling the model. In WinBUGS normal distributions are specified by the mean and the precision where precision $= \frac{1}{\text{variance}}$.

C.1 One-step IPD Royston-Parmar NMA model

In this section I present the WinBUGS model code for the one-step IPD Royston-Parmar NMA model fitted with RTE (4.6).

```
Model  {

for(j in 1:Ntrials) {

for(i in offset[j]+1:offset[j+1]) {

zeros[i] <- 0

# Spline and treatment parameters
eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1[i]
+ gamma[4, j]*u2[i]
+ beta[j, 1]*trt1[i] + beta[j, 2]*trt2[i]

# Derivative with respect to ln(t)
d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1[i] + gamma[4, j]*du2[i]

# Likelihood
lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001 ) )
+ eta[i, j]- exp(eta[i, j]) ) - (1-d[i])*exp(eta[i, j]) ), 0.000001)

# Use zeros trick to maximise likelihood
zeros[i] ~ dpois(lnL[i])
```

```
}


# Prior Distributions


#Spline parameters
for(p in 1:4) {
gamma[p, j] ~ dnorm(0,0.0001)
}


# Treatment parameters
      beta[j, 1:2 ] ~ dmnorm(mu[1:2 ], T[1:2 ,1:2 ])


}


# Hyper-priors:
   mu[1:2] ~ dmnorm(pmu[1:2 ], pT[1:2 ,1:2 ])
   T[1:2 ,1:2 ] ~ dwish(R[1:2 ,1:2 ], 2)


# Calculate log hazard ratio for Con CT compared to Seq CT
lhr.trt1vtrt2 <- mu[2] - mu[1]
}
```

C.2 Assessing PH assumption

In this section I present model code for the one-step IPD Royston-Parmar NMA model including treatment-ln(time) interactions for the assessment of non-proportional hazards. In Subsection 4.4.1 I present two models for doing this. In the first model, (4.7), treatment-ln(time) interactions are included as common effects (Appendix C.2.1). In the second model, (4.8), the treatment-ln(time) interactions can vary by trial (Appendix C.2.2).

C.2.1 Common treatment-ln(time) interactions

```
Model  {
```

```
for(j in 1:Ntrials) {

for(i in offset[j]+1:offset[j+1]) {

zeros[i] <- 0

# Spline, treatment parameters and treatment-ln(time) interactions
eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1[i]
+ gamma[4, j]*u2[i]
+ beta[j, 1]*trt1[i] + beta[j, 2]*trt2[i]
+ beta[j, 3]*trt1[i]*lnt[i] + beta[j, 4]*trt2[i]*lnt[i]

# Derivative with respect to ln(t)
d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1[i] + gamma[4, j]*du2[i]
 + beta[3]*trt1[i] + beta[4]*trt2[i]

# Likelihood
lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001 ) )
+ eta[i, j]- exp(eta[i, j]) ) - (1-d[i])*exp(eta[i, j]) ), 0.000001)

# Use zeros trick to maximise likelihood
zeros[i] ~ dpois(lnL[i])


}


# Prior Distributions


# Spline parameters
for(p in 1:4) {
gamma[p, j] ~ dnorm(0,0.0001)
}


# Treatment and treatment-ln(time) parameters
      beta[j, 1:4 ] ~ dmnorm(mu[1:4 ], T[1:4 ,1:4 ])
```

```
}


# Hyper-priors:

   mu[1:4] ~ dmnorm(pmu[1:4 ], pT[1:4 ,1:4 ])

   T[1:4 ,1:4 ] ~ dwish(R[1:4 ,1:4 ], 4)


# Calculate log hazard ratio for Con CT compared to Seq CT

lhr.trt1vtrt2 <- mu[2] - mu[1]

}
```

C.2.2 Treatment-ln(time) interactions varying by trial

```
Model  {


for(j in 1:Ntrials) {


for(i in offset[j]+1:offset[j+1]) {


zeros[i] <- 0


# Spline, treatment parameters and treatment-ln(time) interactions

eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1[i]

+ gamma[4, j]*u2[i]

+ beta[j, 1]*trt1[i] + beta[j, 2]*trt2[i]

+ (beta[j, 3] + u[j])*trt1[i]*lnt[i] + (beta[j, 4] + u[j])*trt2[i]*lnt[i]


# Derivative with respect to ln(t)

d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1[i] + gamma[4, j]*du2[i]

+ (beta[3]+u[j])*trt1[i] + (beta[4]+u[j])*trt2[i]


# Likelihood

lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001 ) )

+ eta[i, j]- exp(eta[i, j]) ) - (1-d[i])*exp(eta[i, j]) ), 0.000001)
```

```
# Use zeros trick to maximise likelihood
zeros[i] ~ dpois(lnL[i])


}


# Prior Distributions


# Spline parameters
for(p in 1:4) {
gamma[p, j] ~ dnorm(0,0.0001)
}


# Treatment and treatment-ln(time) parameters
      beta[j, 1:4 ] ~ dmnorm(mu[1:4 ], T[1:4 ,1:4 ])


}


# Hyper-priors:
   mu[1:4] ~ dmnorm(pmu[1:4 ], pT[1:4 ,1:4 ])
   T[1:4 ,1:4 ] ~ dwish(R[1:4 ,1:4 ], 4)


# Calculate log hazard ratio for Con CT compared to Seq CT
lhr.trt1vtrt2 <- mu[2] - mu[1]
}
```

C.3 Including an inconsistency parameter

In this section model code is presented for the one-step IPD Royston-Parmar NMA model
including an inconsistency parameter (4.9). In this model consider the lung cancer network
(Figure 3.3) to be coded as: A = RT, B = Seq CT, C = Con CT. Appendix E shows that when
$\theta_{AB} = \beta_1$ and $\theta_{AC} = \beta_2$ then $\theta_{BC} = \beta_2 - \beta_1 + \gamma_{BC}$ where γ_{BC} is the inconsistency parame-
ter. In the model code below the inconsistency parameter is represented by $inconBC$. This
parameter is only included in the spline when trt1$[i] = -1$ and trt2$[i] = 1$ which only oc-
curs for patients who were randomised to Con CT and are from a trial which compared Seq
CT and Con CT. In this model I make $inconBC$ negative because when trt1$[i] = -1$ and

$trt2[i] = 1$ then $trt1[i] * trt2[i] = -1$ so this results in adding the inconsistency parameter to the BC comparison (see Appendix E for more details on parameterisation).

```
Model  {

for(j in 1:Ntrials) {

for(i in offset[j]+1:offset[j+1]) {

zeros[i] <- 0

# Spline, treatment parameters and inconsistency parameter
eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1[i]
+ gamma[4, j]*u2[i]
+ beta[j, 1]*trt1[i] + beta[j, 2]*trt2[i]
- inconBC*trt1[i]*trt2[i]

# Derivative with respect to ln(t)
d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1[i] + gamma[4, j]*du2[i]

# Likelihood
lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001 ) )
+ eta[i, j]- exp(eta[i, j]) ) - (1-d[i])*exp(eta[i, j]) ), 0.000001)

# Use zeros trick to maximise likelihood
zeros[i] ~ dpois(lnL[i])

}

# Prior Distributions

# Spline parameters
for(p in 1:4) {
gamma[p, j] ~ dnorm(0,0.0001)
}
```

```
# Treatment parameters

    beta[j, 1:2 ] ~ dmnorm(mu[1:2 ], T[1:2 ,1:2 ])


}


# Hyper-priors:

   mu[1:2] ~ dmnorm(pmu[1:2 ], pT[1:2 ,1:2 ])

   T[1:2 ,1:2 ] ~ dwish(R[1:2 ,1:2 ], 2)


# Inconsistency parameter
inconBC ~ dnorm(0, 0.1)


# Calculate direct effect for BC (Seq CT v Con CT)
mu4 <- mu[3] - mu[2] + inconBC


# Calculate indirect treatment effects
thetaAB.ind <- mu[3] - mu4
thetaAC.ind <- mu[2] + mu4
thetaBC.ind <- mu[3] - mu[2]
}
```

C.4   Covariate   effects

In this section model code is presented for the one-step IPD Royston-Parmar NMA model
including a patient-level covariate. A patient-level covariate can be included in three ways:
common effect (6.1), fixed trial-level effect (6.2) and random trial-level effect (6.3). The
models in this section are all fitted with RTE. These models are fitted with stage$[i]$ repre-
senting the value of the continuous covariate stage of disease for patient i. In these models
missing stage data is imputed at the network level from a normal distribution with mean one
and variance one.

C.4.1   Common effect of covariate

```
Model  {
```

```
for(j in 1:Ntrials) {

for(i in offset[j]+1:offset[j+1]) {

zeros[i] <- 0

# Spline
eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1n[i]
+ gamma[4, j]*u2n[i]
+ beta[1]*trt1[i] + beta[2]*trt2[i]
+ alpha*stage[i]

# Derivative with repect to ln(t)
d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1n[i] + gamma[4, j]*du2n[i]

# Likelihood
lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001) )
+  eta[i, j]- exp(eta[i, j])  )  - (1-d[i])*exp(eta[i, j]) ), 0.000001)

# Use zeros trick to maximise likelihood
zeros[i] ~ dpois(lnL[i])

# Distributional assumption for imputing missing data
stage[i] ~ dnorm(1, 1)


}


# Prior Distributions


# Spline parameters
for(p in 1:4) {
gamma[p, j] ~ dnorm(0,0.0001)
}
```

```
}

# Covariate
alpha ~ dnorm(0, 0.1)

# Treatment parameters
for(p in 1:2) {
beta[p] ~ dnorm(0, 0.001)
}

# Calculate log hazard ratio for Con CT compared to Seq CT
lhr.trt1vtrt2 <- beta[2] - beta[1]
}
```

C.4.2 Fixed trial-level effect of covariate

```
Model  {

for(j in 1:Ntrials) {

for(i in offset[j]+1:offset[j+1]) {

zeros[i] <- 0

# Spline, treatment parmaters and covariate
eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1n[i]
+ gamma[4, j]*u2n[i]
+ beta[1]*trt1[i] + beta[2]*trt2[i]
+ alpha[j]*stage[i]

# Derivatiave with respect to ln(t)
d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1n[i] + gamma[4, j]*du2n[i]

# Likelihood
lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001) )
```

```
+  eta[i, j]- exp(eta[i, j])  )  - (1-d[i])*exp(eta[i, j]) ), 0.000001)


# Use zeros trick to maximise likelihood

zeros[i] ~ dpois(lnL[i])


# Distributional assumption for imputing missing data

stage[i] ~ dnorm(1, 1)



}


# Prior Distributions


# Spline parameters

for(p in 1:4) {

gamma[p, j] ~ dnorm(0,0.0001)

}



}


# Covariate - one for each trial

for(p in 1:43) {

alpha[p] ~ dnorm(0, 0.1)

}


# Treatment parameters

for(p in 1:2) {

beta[p] ~ dnorm(0, 0.001)

}


# Calculate log hazard ratio for Con CT compared to Seq CT

lhr.trt1vtrt2 <- beta[2] - beta[1]

}
```

C.4.3 Random trial-level effect of covariate

```
Model {

for(j in 1:Ntrials) {

for(i in offset[j]+1:offset[j+1]) {

zeros[i] <- 0

# Spline, treatment parameters and covariate
eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1n[i]
+ gamma[4, j]*u2n[i]
+ beta[j, 1]*trt1[i] + beta[j, 2]*trt2[i]
+ alpha[1, j]*stage[i]

# Derivative with respect to ln(t)
d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1n[i] + gamma[4, j]*du2n[i]

# Likelihood
lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001) )
+  eta[i, j]- exp(eta[i, j])  )  - (1-d[i])*exp(eta[i, j]) ), 0.000001)

# Use zeros trick to maximise likelihood
zeros[i] ~ dpois(lnL[i])

# Distributional assumption for imputing missing data
stage[i] ~ dnorm(1, 1)

}

# Prior Distributions

# Spline parameters
for(p in 1:4) {
```

```
gamma[p, j] ~ dnorm(0,0.0001)

}


# Covariate

alpha[1, j] ~ dnorm(theta, prec)


# Treatment parameters

beta[j, 1:2 ] ~ dmnorm(mu[1:2 ], T[1:2 ,1:2 ])


}


# Covariate

theta ~ dnorm(0, 0.0001)

prec <- 1/var

var <- pow(sd, 2)

sd ~ dnorm(0, 1000)


# Hyper-priors

mu[1:2] ~ dmnorm(pmu[1:2 ], pT[1:2 ,1:2 ])

T[1:2 ,1:2 ] ~ dwish(R[1:2 ,1:2 ], 2)


# Calculate log hazard ratio for Con CT compared to Seq CT

lhr.trt1vtrt2 <- mu[2] - mu[1]

}
```

C.5 Treatment-covariate interactions

In this section model code is presented for the one-step IPD Royston-Parmar NMA model including treatment-covariate interactions. Treatment-covariate interaction models can be fitted with the within and across trial information either combined (6.6) or separate (6.8). The models in this section are all fitted with RTE. These models are fitted with stage$[i]$ representing the value of the continuous covariate stage of disease for patient i. In the combined model I monitor the mean of the imputed values for stage of disease in trials with any missing stage data. These values can then be used in the model with within

and across trial information separated. See Section 6.4 for full details. In these models missing data is imputed at the network level from a normal distribution where the mean is drawn from a normal distribution and precision from a uniform distribution.

C.5.1 Combined RTE

```
Model {

for(j in 1:Ntrials) {

for(i in offset[j]+1:offset[j+1]) {

zeros[i] <- 0

# Spline, treatment, covariate and treatment-covariate interactions
eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1n[i]
+ gamma[4, j]*u2n[i]
+ beta[j, 1]*trt1[i] + beta[j, 2]*trt2[i]
+ alpha[j]*stage[i]
+ delta[j, 1]*trt1[i]*stage[i] + delta[j, 2]*trt2[i]*stage[i]


# Derivative with respect to ln(t)
d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1n[i] + gamma[4, j]*du2n[i]

# Likelihood
lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001) )
+  eta[i, j]- exp(eta[i, j])  )  - (1-d[i])*exp(eta[i, j]) ), 0.000001)

# Use zeros trick to maximise likelihood
zeros[i] ~ dpois(lnL[i])

# Distributional assumption for dealing with missing data
stage[i] ~ dnorm( zeta, phi)I(0, 2)
```

```
}


# Prior Distributions


# Spline parameters
for(p in 1:4) {
gamma[p, j] ~ dnorm(0,0.0001)
}


# Treatment and treatment-covariate parameters
beta[j, 1:2 ] ~ dmnorm(mu[1:2 ], T[1:2 ,1:2 ])
alpha[j, 1:2 ] ~ dmnorm(mu2[1:2 ], T2[1:2 ,1:2 ])


# Covariate
delta[j] ~ dnorm(theta, prec)


}


theta ~ dnorm(0, 0.0001)
prec <- 1/var
var <- pow(sd, 2)
sd ~ dnorm(0, 1000)


# Missing data
zeta ~ dnorm(0, 0.001)
phi ~ dunif(0.1, 10)


# Hyper-priors:
   mu[1:2] ~ dmnorm(pmu[1:2 ], pT[1:2 ,1:2 ])
   T[1:2 ,1:2 ] ~ dwish(R[1:2 ,1:2 ], 2)


   mu2[1:2] ~ dmnorm(pmu2[1:2 ], pT2[1:2 ,1:2 ])
   T2[1:2 ,1:2 ] ~ dwish(R2[1:2 ,1:2 ], 2)


# Calculate log hazard ratio for Con CT compared to Seq CT
```

```
lhr.trt1vtrt2 <- mu[2] - mu[1]


# In the model with within and across trial information separated values of
# mean stage for each trial are needed so here I monitor mean stage in all
# trials that have some missing data.
# This returns the mean value of the imputed values (not the observed values)
mstage.trial.1 <- mean(stage_cont[1:65])
mstage.trial.3 <- mean(stage_cont[114:440])
mstage.trial.4 <- mean(stage_cont[441:519])
mstage.trial.5 <- mean(stage_cont[520:585])
mstage.trial.6 <- mean(stage_cont[586:765])
mstage.trial.7 <- mean(stage_cont[766:840])
mstage.trial.9 <- mean(stage_cont[1194:1321])
mstage.trial.10 <- mean(stage_cont[1322:1447])
mstage.trial.11 <- mean(stage_cont[1448:1908])
mstage.trial.12 <- mean(stage_cont[1909:2234])
mstage.trial.13 <- mean(stage_cont[2235:2315])
mstage.trial.14 <- mean(stage_cont[2316:2567])
mstage.trial.15 <- mean(stage_cont[2568:2652])
mstage.trial.16 <- mean(stage_cont[2653:2860])
mstage.trial.17 <- mean(stage_cont[2861:2916])
mstage.trial.18 <- mean(stage_cont[2917:2978])
mstage.trial.19 <- mean(stage_cont[2979:3099])
mstage.trial.20 <- mean(stage_cont[3100:3218])
mstage.trial.21 <- mean(stage_cont[3219:3387])
mstage.trial.22 <- mean(stage_cont[3388:3718])
mstage.trial.23 <- mean(stage_cont[3719:3955])
mstage.trial.24 <- mean(stage_cont[3956:4128])
mstage.trial.25 <- mean(stage_cont[4129:4336])
mstage.trial.26 <- mean(stage_cont[4337:4618])
mstage.trial.28 <- mean(stage_cont[4779:5362])
mstage.trial.32 <- mean(stage_cont[4957:5004])
mstage.trial.33 <- mean(stage_cont[5737:5913])
mstage.trial.34 <- mean(stage_cont[5914:5958])
mstage.trial.37 <- mean(stage_cont[6311:6356])
```

```
mstage.trial.39 <- mean(stage_cont[6671:7077])

mstage.trial.40 <- mean(stage_cont[7078:7282])

mstage.trial.41 <- mean(stage_cont[7283:7440])

}
```

C.5.2   Separate RTE

When the within and across trial information is separated as in (6.7) and (6.8) in addition to the treatment-covariate interaction terms, treatment-mean stage interaction terms are also included in the model. In the code below mstage$[i]$ represents the mean value of stage of disease for patient i from trial j. The within trial information is represented by the δ_1 parameters and the across trial information is equal to $\delta_1+\delta_2$. See Subsection 6.6.2 for full details.

```
Model  {

for(j in 1:Ntrials) {

for(i in offset[j]+1:offset[j+1]) {

zeros[i] <- 0

# Spline, treatment, covariate, treatment-covariate and treatment-mean covariate
eta[i, j] <- gamma[1, j] + gamma[2, j]*u0[i] + gamma[3, j]*u1n[i]
+ gamma[4, j]*u2n[i]
+ beta[j, 1]*trt1[i] + beta[j, 2]*trt2[i]
+ alpha[j]*stage[i]
+ delta1[j, 1]*trt1[i]*stage[i] + delta1[j, 2]*trt2[i]*stage[i]
+ delta2[j, 1]*trt1[i]*mstage[i] + delta2[j, 2]*trt2[i]*mstage[i]

# Derivative with respect to ln(t)
d.sp[i, j] <- gamma[2, j]*du0[i] + gamma[3, j]*du1n[i] + gamma[4, j]*du2n[i]

# Likelihood
lnL[i] <- max( - ( d[i] * ( log( max( d.sp[i, j],0.0001) )
+  eta[i, j]- exp(eta[i, j]) )  - (1-d[i])*exp(eta[i, j]) ), 0.000001)
```

286

```
# Use zeros trick to maximise likelihood
zeros[i] ~ dpois(lnL[i])


# Distributional assumption for imputing missing data
stage[i] ~ dnorm(zeta, phi)I(0, 2)


}


# Prior Distributions


# Spline parameters
for(p in 1:4) {
gamma[p, j] ~ dnorm(0,0.0001)
}


# Treatment, treatment-covariate, treatment-mean covariate
beta[j, 1:2 ] ~ dmnorm(mu[1:2 ], T[1:2 ,1:2 ])
delta1[j, 1:2 ] ~ dmnorm(mu2[1:2 ], T2[1:2 ,1:2 ])
delta2[j, 1:2 ] ~ dmnorm(mu3[1:2 ], T3[1:2 ,1:2 ])


# Covariate
alpha[j] ~ dnorm(theta, prec)


}


# Covariate
theta ~ dnorm(0, 0.0001)
prec <- 1/var
var <- pow(sd, 2)
sd ~ dnorm(0, 1000)


# Missing data
zeta ~ dnorm(0, 0.001)
phi ~ dunif(0.1, 10)
```

```
# Hyper-priors:

   mu[1:2] ~ dmnorm(pmu[1:2 ], pT[1:2 ,1:2 ])

   T[1:2 ,1:2 ] ~ dwish(R[1:2 ,1:2 ], 2)


   mu2[1:2] ~ dmnorm(pmu2[1:2 ], pT2[1:2 ,1:2 ])

   T2[1:2 ,1:2 ] ~ dwish(R2[1:2 ,1:2 ], 2)


   mu3[1:2] ~ dmnorm(pmu3[1:2 ], pT3[1:2 ,1:2 ])

   T3[1:2 ,1:2 ] ~ dwish(R3[1:2 ,1:2 ], 2)


# Calculate log hazard ratio for Con CT compared to Seq CT
lhr.trt1vtrt2 <- mu[2] - mu[1]


# Calculating within and across trial information
delta.across.trt1 <- mu2[1] + mu3[1]
delta.within.trt1 <- mu2[1]


delta.across.trt2 <- mu2[2] + mu3[2]
delta.within.trt2 <- mu2[2]
}
```

D Global Wald Test

This appendix presents Stata code to conduct a Wald test on the treatment-covariate interaction parameter estimates. Let p1 and p2 be the parameter estimates and let there be 40000 iterations.

```
/* Calculate the mean of each treatment-covariate parameter estimate */
su p1
gen p1_mean = 'r(mean)'


su p2
gen p2_mean = 'r(mean)'



/*Conduct Wald test using Mata */

mata

/* Step 1: Set up B as a matrix containing the WinBUGS estimates */
B = J(40000, 2, .)
st_view(B , . , "p1 p2", )



/* Step 2: Set up MM as a matrix containing the mean values of the
parameter estimates */
MM = J(40000, 2, .)
st_view(MM , . , "p1_mean p2_mean", )

/* Step 3: Calculate C */
C = B-MM


/* Step 4: Calculate A */
A = C'*C/(40000)

/* Step 5: Take the column means of B and store as matrix called M */
```

```
M = mean(B)


/* Step 6: Calculate Wald test statistic */
chi_2 = M*(invsym(A))*M'


/* Additional check: take the diagonal elements of A and square root to
obtain standard deviation estimates */
D = diagonal(A)
E = sqrt(D)


/* Exit Mata */
end
```

E Treatment Parameterisation for Inconsistency Parameters

In this appendix I explain how a three treatment loop can be parametrised when fitting a NMA model including an inconsistency parameter and provide a numerical example.

E.1 Treatment parameterisation in a three treatment loop

In a triangular network consisting of three-treatments A, B and C (as pictured in Figure 5.3, page 126), let γ be the inconsistency parameter. Then let μ_B^{AB} be the treatment effect for treatment B in a A v B trial, μ_A^{AB} be the treatment effect for treatment A in a A v B trial, μ_C^{AC} be the treatment effect for treatment C in a A v C trial, μ_A^{AC} be the treatment effect for treatment A in a A v C trial, μ_C^{BC} be the treatment effect for treatment C in a B v C trial and μ_B^{BC} be the treatment effect for treatment B in a B v C trial.

Let $\theta_{AB}, \theta_{AC}, \theta_{BC}$ be the direct estimates of the treatment effects for treatment A v B, treatment A v C and treatment B v C respectively. Then:

$$\mu_B^{AB} - \mu_A^{AB} = \theta_{AB}$$

$$\mu_C^{AC} - \mu_A^{AC} = \theta_{AC}$$

$$\mu_C^{BC} - \mu_B^{BC} = \theta_{BC}$$

With three treatment effects to be estimated the network can be parameterised in 3 different ways. In parameterisation 1:

$$\theta_{AB} = \beta_1$$

$$\theta_{AC} = \beta_2$$

$$\theta_{BC} = \beta_2 - \beta_1 + \gamma_{BC}$$

In parameterisation 2:

$$\theta_{AB} = \beta_1$$

$$\theta_{BC} = \beta_3$$

$$\theta_{AC} = \beta_1 + \beta_3 + \gamma_{AC}$$

In parameterisation 3:

$$\theta_{AC} = \beta_2$$

$$\theta_{BC} = \beta_3$$

$$\theta_{AB} = \beta_2 - \beta_3 + \gamma_{AB}$$

Overall:

$$\beta_1 = \beta_2 - \beta_3 + \gamma_{AB}$$

$$\beta_2 = \beta_1 + \beta_3 + \gamma_{AC}$$

$$\beta_3 = \beta_2 - \beta_1 + \gamma_{BC}$$

β_2 can be re-arranged so that:

$$\beta_1 = \beta_2 - \beta_3 - \gamma_{AC}$$

Therefore:

$$\gamma_{AB} = -\gamma_{AC}$$

β_3 can be re-arranged so that:

$$\beta_2 = \beta_1 + \beta_3 - \gamma_{BC}$$

Therefore:

$$\gamma_{AC} = -\gamma_{BC}$$

β_1 can be re-arranged so that:

$$\beta_3 = \beta_2 - \beta_1 + \gamma_{AB}$$

Therefore:

$$\gamma_{BC} = \gamma_{AB}$$

Therefore overall:

$$\gamma_{AB} = \gamma_{BC} = -\gamma_{AC}$$

E.2 Numerical Example

In this section I use a numerical example to illustrate the results presented in Appendix E.1.

Let $\mu_B^{AB} = 0.3$, $\mu_A^{AB} = 0.6$, $\mu_C^{AC} = 0.2$, $\mu_A^{AC} = 0.4$, $\mu_C^{BC} = 2$, $\mu_B^{BC} = 1.3$.

Then:

$$\theta_{AB} = \mu_B^{AB} - \mu_A^{AB} = 0.3 - 0.6 = -0.3$$

$$\theta_{AC} = \mu_C^{AC} - \mu_A^{AC} = 0.2 - 0.4 = -0.2$$

$$\theta_{BC} = \mu_C^{BC} - \mu_B^{BC} = 2 - 1.3 = 0.7$$

Parameterisation 1:

$$\theta_{AB} = \beta_1 = -0.3$$

$$\theta_{AC} = \beta_2 = -0.2$$

$$\theta_{BC} = \beta_2 - \beta_1 + \gamma_{BC} = -0.2 - -0.3 + \gamma_{BC} = 0.1 + \gamma_{BC}$$

Therefore:

$$0.7 = 0.1 + \gamma_{BC}$$

$$\gamma_{BC} = 0.6$$

Parameterisation 2:

$$\theta_{AB} = \beta_1 = -0.3$$

$$\theta_{BC} = \beta_3 = 0.7$$

$$\theta_{AC} = \beta_1 + \beta_3 + \gamma_{AC} = -0.3 + 0.7 + \gamma_{AC} = 0.4 + \gamma_{AC}$$

Therefore:

$$-0.2 = 0.4 + \gamma_{AC}$$

$$\gamma_{AC} = -0.6$$

Parameterisation 3:

$$\theta_{AC} = \beta_2 = -0.2$$

$$\theta_{BC} = \beta_3 = 0.7$$

$$\theta_{AB} = \beta_2 - \beta_3 + \gamma_{AB} = -0.2 - 0.7 + \gamma_{AB} = -0.9 + \gamma_{AB}$$

Therefore:

$$-0.3 = -0.9 + \gamma_{AB}$$

$$\gamma_{AB} = 0.6$$

Therefore overall:

$$\gamma_{AB} = \gamma_{BC} = -\gamma_{AC}$$
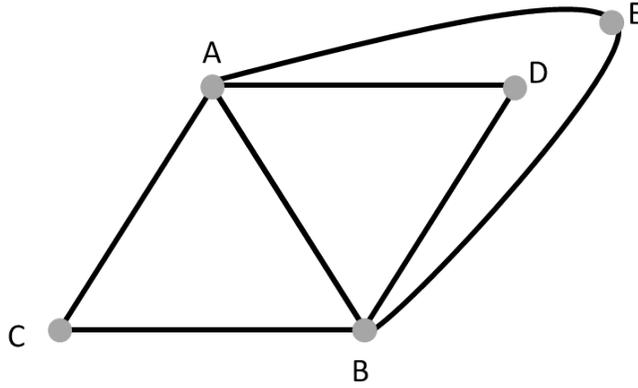
Figure F.1: Example of a five treatment network. A, B, C, D, E represent the five treatments being compared.

F Inconsistency in NMA

This appendix extends the methods developed in Section 5.5 for three and four treatment networks to a five treatment network in Section F.1. The results from three, four and five treatment networks are then displayed alongside each other in Section F.2.

F.1 Five treatment network

In this section I apply the same principles as in Subsection 5.5.2 (three treatment networks) and Subsection 5.5.3 (four treatment networks) to a five treatment network, such as that pictured in Figure F.1.

Considering Figure F.1 there are three pathways of indirect evidence between treatment A and treatment B. I denote the treatment effect from the first pathway (via treatment C) by $\hat{\theta}_c^{\text{ind}(1)}$, the treatment effect from the second pathway (via treatment D) by $\hat{\theta}_c^{\text{ind}(2)}$ and the treatment effect from the third pathway (via treatment E) by $\hat{\theta}_c^{\text{ind}(3)}$. I define $\hat{\theta}_c^{\text{ind}}$ as the weighted average of all the indirect evidence for design c, $\hat{\theta}_c^{\text{net}}$ as the weighted average of all the direct and indirect evidence for design c, $\hat{\theta}_{c(d)}^{\text{net}}$ as the weighted average of all the direct and indirect evidence for design c when design d is excluded and $\hat{\theta}_{c(d/2)}^{\text{net}}$ as the average of all the network evidence for design c and the network evidence for design c that remains when design d is excluded. I assume an equal weight of $\frac{1}{s^2}$ for each of the direct comparisons in the network so each indirect comparison has weight $\frac{1}{2s^2}$. I also assume that FTE are used.

Applying these definitions to the five-treatment network displayed in Figure F.1 and assuming $c \neq d$:

$$\hat{\theta}_c^{\text{ind}(1)} = \hat{\theta}_{AC}^{\text{dir}} - \hat{\theta}_{BC}^{\text{dir}}$$

$$\hat{\theta}_c^{\text{ind}(2)} = \hat{\theta}_{AD}^{\text{dir}} - \hat{\theta}_{BD}^{\text{dir}}$$

$$\hat{\theta}_c^{\text{ind}(3)} = \hat{\theta}_{AE}^{\text{dir}} - \hat{\theta}_{BE}^{\text{dir}}$$

$$\hat{\theta}_c^{\text{ind}} = \hat{\theta}_c^{\text{ind}(1)} + \hat{\theta}_c^{\text{ind}(2)} + \hat{\theta}_c^{\text{ind}(3)}$$

$$\begin{aligned}
\hat{\theta}_c^{\text{net}} &= \frac{1}{5}\left(\hat{\theta}_c^{\text{dir}} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(3)} \right) \\
&= \frac{2}{5}\hat{\theta}_c^{\text{dir}} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(3)}
\end{aligned}$$

$$\begin{aligned}
\hat{\theta}_{c(d)}^{\text{net}} &= \frac{1}{4}\left(\hat{\theta}_c^{\text{dir}} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(3)} \right) \\
&= \frac{1}{2}\hat{\theta}_c^{\text{dir}} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(3)}
\end{aligned}$$

$$\begin{aligned}
\hat{\theta}_{c(d/2)}^{\text{net}} &= \frac{1}{2}\left(\hat{\theta}_c^{\text{net}} + \hat{\theta}_{c(d)}^{\text{net}} \right) \\
&= \frac{1}{2}\left[\frac{2}{5}\hat{\theta}_c^{\text{dir}} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(3)} + \frac{1}{2}\hat{\theta}_c^{\text{dir}} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(3)} \right] \\
&= \frac{1}{2}\left[\frac{9}{10}\hat{\theta}_c^{\text{dir}} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)} + \frac{9}{20}\hat{\theta}_c^{\text{ind}(2)} + \frac{9}{20}\hat{\theta}_c^{\text{ind}(3)} \right] \\
&= \frac{9}{20}\hat{\theta}_c^{\text{dir}} + \frac{1}{10}\hat{\theta}_c^{\text{ind}(1)} + \frac{9}{40}\hat{\theta}_c^{\text{ind}(2)} + \frac{9}{40}\hat{\theta}_c^{\text{ind}(3)}
\end{aligned}$$

Then (5.10) can be written as:

$$
Q_{c,d}^{\text{diff}} = \frac{1}{s^2}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{net}}\right)^2 - \frac{1}{s^2}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_{c(d)}^{\text{net}}\right)^2 = \frac{2}{s^2}\left(\hat{\theta}_{c(d)}^{\text{net}} - \hat{\theta}_c^{\text{net}}\right)\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_{c(d/2)}^{\text{net}}\right)
$$

$$
= \frac{2}{s^2}\left[\frac{1}{2}\hat{\theta}_c^{\text{dir}} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(3)} - \frac{2}{5}\hat{\theta}_c^{\text{dir}} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(3)}\right]
$$

$$
\left[\hat{\theta}_c^{\text{dir}} - \frac{9}{20}\hat{\theta}_c^{\text{dir}} - \frac{1}{10}\hat{\theta}_c^{\text{ind}(1)} - \frac{9}{40}\hat{\theta}_c^{\text{ind}(2)} - \frac{9}{40}\hat{\theta}_c^{\text{ind}(3)}\right]
$$

$$
= \frac{2}{s^2}\left[\frac{1}{10}\hat{\theta}_c^{\text{dir}} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{20}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{20}\hat{\theta}_c^{\text{ind}(3)}\right]\left[\frac{11}{20}\hat{\theta}_c^{\text{dir}} - \frac{1}{10}\hat{\theta}_c^{\text{ind}(1)}\right.
$$

$$
\left. - \frac{9}{40}\hat{\theta}_c^{\text{ind}(2)} - \frac{9}{40}\hat{\theta}_c^{\text{ind}(3)}\right]
$$

$$
= \frac{1}{s^2}\left[\frac{1}{5}\hat{\theta}_c^{\text{dir}} - \frac{2}{5}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{10}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{10}\hat{\theta}_c^{\text{ind}(3)}\right]\left[\frac{11}{20}\hat{\theta}_c^{\text{dir}} - \frac{1}{10}\hat{\theta}_c^{\text{ind}(1)} - \frac{9}{40}\hat{\theta}_c^{\text{ind}(2)}\right.
$$

$$
\left. - \frac{9}{40}\hat{\theta}_c^{\text{ind}(3)}\right]
$$

$$
= \frac{1}{s^2}\left[\frac{11}{100}\left(\hat{\theta}_c^{\text{dir}}\right)^2 + \frac{1}{25}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \frac{9}{400}\left(\hat{\theta}_c^{\text{ind}(2)}\right)^2 - \frac{9}{400}\left(\hat{\theta}_c^{\text{ind}(3)}\right)^2 - \frac{6}{25}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^1\right.
$$

$$
\left. + \frac{1}{100}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{100}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(3)} + \frac{2}{25}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(2)} + \frac{2}{25}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(3)} - \frac{9}{200}\hat{\theta}_c^{\text{ind}(2)}\hat{\theta}_c^{\text{ind}(3)}\right]
$$

$$
= \frac{1}{5s^2}\left[\frac{11}{20}\left(\hat{\theta}_c^{\text{dir}}\right)^2 + \frac{1}{5}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \frac{9}{80}\left(\hat{\theta}_c^{\text{ind}(2)}\right)^2 - \frac{9}{80}\left(\hat{\theta}_c^{\text{ind}(3)}\right)^2 - \frac{6}{5}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^1\right.
$$

$$
\left. + \frac{1}{20}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{20}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(3)} + \frac{2}{5}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(2)} + \frac{2}{5}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(3)} - \frac{9}{40}\hat{\theta}_c^{\text{ind}(2)}\hat{\theta}_c^{\text{ind}(3)}\right]
$$

$$
= \frac{1}{5s^2}\left\{\frac{1}{2}\hat{\theta}_c^{\text{dir}} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{4}\hat{\theta}_c^{\text{ind}(3)} - \hat{\theta}_c^{\text{ind}(1)}\right\}
$$

$$
\times\left[\frac{11}{10}\hat{\theta}_c^{\text{dir}} - \frac{9}{20}\hat{\theta}_c^{\text{ind}(2)} - \frac{9}{20}\hat{\theta}_c^{\text{ind}(3)} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)}\right]
$$

$$
= \frac{1}{5s^2}\left\{\frac{1}{4}\left(2\hat{\theta}_c^{\text{dir}} + \hat{\theta}_c^{\text{ind}(2)} + \hat{\theta}_c^{\text{ind}(3)}\right) - \hat{\theta}_c^{\text{ind}(1)}\right\}
$$

$$
\times\left[2\hat{\theta}_c^{\text{dir}}\left(1 - \frac{9}{20}\right) - \frac{1}{5}\left\{\frac{9}{4}\left(\hat{\theta}_c^{\text{ind}(2)} + \hat{\theta}_c^{\text{ind}(3)}\right) + \hat{\theta}_c^{\text{ind}(1)}\right\}\right]
$$

In the second scenario where $c = d$ define:

$$
\hat{\theta}_c^{\text{ind}(1)} = \hat{\theta}_{AC}^{\text{dir}} - \hat{\theta}_{BC}^{\text{dir}}
$$

$$
\hat{\theta}_c^{\text{ind}(2)} = \hat{\theta}_{AD}^{\text{dir}} - \hat{\theta}_{BD}^{\text{dir}}
$$

$$
\hat{\theta}_c^{\text{ind}(3)} = \hat{\theta}_{AE}^{\text{dir}} - \hat{\theta}_{BE}^{\text{dir}}
$$

$$
\hat{\theta}_c^{\text{ind}} = \hat{\theta}_c^{\text{ind}(1)} + \hat{\theta}_c^{\text{ind}(2)} + \hat{\theta}_c^{\text{ind}(3)}
$$

$$
\hat{\theta}_c^{\text{net}} = \frac{1}{5}\left(\hat{\theta}_c^{\text{dir}} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{2}\hat{\theta}_c^{\text{ind}(3)}\right) = \frac{2}{5}\hat{\theta}_c^{\text{dir}} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{5}\hat{\theta}_c^{\text{ind}(3)}
$$

$$
\hat{\theta}_{c(c)}^{\text{net}} = \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)} + \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)} + \frac{1}{3}\hat{\theta}_c^{\text{ind}(3)}
$$

Then:

$$
Q_{c,c}^{\text{diff}} = Q_c^{\text{inc}} - Q_{c,c}^{\text{inc}} = \frac{1}{s^2}\left[\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{net}}\right)^2 - \left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_{c(c)}^{\text{net}}\right)^2\right]
$$

$$
= \frac{1}{s^2}\left[\left(\hat{\theta}_c^{\text{dir}} - \frac{2}{5}\hat{\theta}_c^{\text{dir}} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(3)}\right)^2 - \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)}\right.\right.
$$

$$
\left.\left. - \frac{1}{3}\hat{\theta}_c^{\text{ind}(3)}\right)^2\right]
$$

$$
= \frac{1}{s^2}\left[\left(\frac{3}{5}\hat{\theta}_c^{\text{dir}} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{5}\hat{\theta}_c^{\text{ind}(3)}\right)^2 - \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)}\right.\right.
$$

$$
\left.\left. - \frac{1}{3}\hat{\theta}_c^{\text{ind}(3)}\right)^2\right]
$$

$$
= \frac{1}{s^2}\left[\frac{-16}{25}\left(\hat{\theta}_c^{\text{dir}}\right)^2 - \frac{16}{225}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 - \frac{16}{225}\left(\hat{\theta}_c^{\text{ind}(2)}\right)^2 - \frac{16}{225}\left(\hat{\theta}_c^{\text{ind}(3)}\right)^2 + \frac{32}{75}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)}\right.
$$

$$
\left. + \frac{32}{75}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(2)} + \frac{32}{75}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(3)} - \frac{32}{225}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(2)} - \frac{32}{225}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(3)} - \frac{32}{225}\hat{\theta}_c^{\text{ind}(2)}\hat{\theta}_c^{\text{ind}(3)}\right]
$$

$$
= \frac{-16}{25s^2}\left[\left(\hat{\theta}_c^{\text{dir}}\right)^2 + \frac{1}{9}\left(\hat{\theta}_c^{\text{ind}(1)}\right)^2 + \frac{1}{9}\left(\hat{\theta}_c^{\text{ind}(2)}\right)^2 + \frac{1}{9}\left(\hat{\theta}_c^{\text{ind}(3)}\right)^2 - \frac{2}{3}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(1)}\right.
$$

$$
\left. - \frac{2}{3}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(2)} - \frac{2}{3}\hat{\theta}_c^{\text{dir}}\hat{\theta}_c^{\text{ind}(3)} + \frac{2}{9}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(2)} + \frac{2}{9}\hat{\theta}_c^{\text{ind}(1)}\hat{\theta}_c^{\text{ind}(3)} + \frac{2}{9}\hat{\theta}_c^{\text{ind}(2)}\hat{\theta}_c^{\text{ind}(3)}\right]
$$

$$
= \frac{-16}{25s^2}\left(\hat{\theta}_c^{\text{dir}} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(3)}\right)^2
$$

F.2 Summary of formulas for networks

In this section I display the formula for Q^{diff} from the three, four and five treatment networks alongside each other and alongside the general formula.

From a three treatment network with $c \neq d$:

$$
Q_{c,d}^{\text{diff}} = \frac{1}{s^2}\cdot\frac{1}{9}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}(1)}\right)^2
$$

From a three treatment network with $c = d$:

$$
Q_{c,c}^{\text{diff}} = \frac{-1}{s^2}\cdot\frac{8}{9}\left(\hat{\theta}_c^{\text{dir}} - \hat{\theta}_c^{\text{ind}(1)}\right)^2
$$

From a four treatment network with $c \neq d$:

$$
Q_{c,d}^{\text{diff}} = \frac{1}{4s^2}\left[\frac{1}{3}(2\hat{\theta}_c^{\text{dir}} + \hat{\theta}_c^{\text{ind}(2)}) - \hat{\theta}_c^{\text{ind}(1)}\right]\left[2\hat{\theta}_c^{\text{dir}}\left(1 - \frac{7}{12}\right) - \frac{1}{4}\left(\frac{7}{3}\hat{\theta}_c^{\text{ind}(2)} + \hat{\theta}_c^{\text{ind}(1)}\right)\right]
$$

From a four treatment network with $c = d$:

$$Q_{c,c}^{\text{diff}} = \frac{-3}{4s^2} \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{2}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{2}\hat{\theta}_c^{\text{ind}(2)} \right)^2$$

From a five treatment network with $c \neq d$:

$$Q_{c,d}^{\text{diff}} = \frac{1}{5s^2} \left\{ \frac{1}{4} \left(2\hat{\theta}_c^{\text{dir}} + \hat{\theta}_c^{\text{ind}(2)} + \hat{\theta}_c^{\text{ind}(3)} \right) - \hat{\theta}_c^{\text{ind}(1)} \right\}$$
$$\times \left[2\hat{\theta}_c^{\text{dir}} \left(1 - \frac{9}{20} \right) - \frac{1}{5} \left\{ \frac{9}{4} \left(\hat{\theta}_c^{\text{ind}(2)} + \hat{\theta}_c^{\text{ind}(3)} \right) + \hat{\theta}_c^{\text{ind}(1)} \right\} \right]$$

From a five treatment network with $c = d$:

$$Q_{c,c}^{\text{diff}} = \frac{-16}{25s^2} \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(1)} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(2)} - \frac{1}{3}\hat{\theta}_c^{\text{ind}(3)} \right)^2$$

General formula with $c \neq d$:

$$Q_{c,d}^{\text{diff}} = \frac{1}{s^2} \cdot \frac{1}{k+2} \left\{ \frac{1}{k+1} \left(2\hat{\theta}_c^{\text{dir}} + \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} \right) - \hat{\theta}_c^{\text{ind}(d)} \right\}$$
$$\times \left[2\hat{\theta}_c^{\text{dir}} \left(1 - \frac{2k+3}{(k+1)(k+2)} \right) - \frac{1}{k+2} \left(\frac{2k+3}{k+1} \sum_{i,i\neq d} \hat{\theta}_c^{\text{ind}(i)} + \hat{\theta}_c^{\text{ind}(d)} \right) \right]$$

General formula with $c = d$:

$$Q_{c,c}^{\text{diff}} = -\frac{1}{s^2} \cdot \frac{4(k+1)}{(k+2)^2} \left(\hat{\theta}_c^{\text{dir}} - \frac{1}{k} \sum_{i=1}^{k} \hat{\theta}_c^{\text{ind}(i)} \right)^2$$