Meta-analysis of aggregate data on medical events

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

von Dipl.-Math. Björn Holzhauer

geb. am 20. Januar 1979 in Henstedt-Ulzburg

genehmigt durch die Fakultät für Mathematik

der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Rainer Schwabe

PD Dr. Ekkehard Glimm

Prof. Dr. Tim Friede

eingereicht am: 19. Mai 2017

Verteidigung am: 20. Oktober 2017

Abstract

A meta-analysis combines analysis results from multiple independent sets of data. Meta-analyses to assess whether a treatment affects the occurrence of a medical event are frequently based on published aggregate data from controlled clinical trials. In that setting, the number of patients with an event in each trial is often treated as a binomial random variable. This assumes that censoring times have the same distribution in all treatment groups of a trial, and are independent of event times.

To allow for different drop-out time distributions across treatment groups, we derive a likelihood for commonly available aggregate data that assumes specific event and drop-out time distributions for a number of situations. These include exponentially or Weibull distributed event and drop-out times, event-driven trials, the situation when a patient may experience multiple potentially fatal events, and when individual patient data are available for some trials.

The assumption that parameters of survival distributions are exchangeable between trials is more plausible than for the expected proportion of patients with an event. For this reason the proposed likelihood is more suitable than a binomial likelihood for use in hierarchical meta-analysis models and for incorporating prior information from historical control group data. Hierarchical models and prior information are useful in sparse data settings and to avoid parameter identifiability problems. We use simulations to compare hierarchical Bayesian models with the proposed trial-level likelihood against other meta-analysis methods and to compare methods for using historical control group data. We also demonstrate how conjugate priors may be used to analyze exponentially distributed failure times without the need for Markov chain Monte Carlo methods.

Zusammenfassung

Eine Metaanalyse fasst die Ergebnisse der Analysen von mehreren unabhängigen Datensätzen zusammen. Wenn der Vergleich einer Behandlung mit einer Kontrollgruppe bezüglich des Auftretens eines medizinischen Ereignisses nur auf publizierten Aggregatdaten aus klinischen Studien beruht, wird typischerweise die Likelihood der Binomialverteiligung benutzt, was jedoch voraussetzt, dass die Zensierungszeiten unabhängig von den Ereigniszeiten sind und in allen Behandlungsgruppen einer Studie der gleichen Verteilung folgen. Wir entwickeln eine Likelihood-Funktion für üblicherweise verfügbare Aggregatdaten, die die zweite dieser Annahmen vermeidet, aber spezifische Ereignis- und Zensierungszeitverteilungen annimmt. Wir behandeln dabei die Szenarien mit exponential- und Weibull-verteilten Ereignis- und Zensierungszeiten, als auch mit mehreren potentiell tödlichen Ereignissen pro Patient, mit verfügbaren individuellen Patientendaten und ereignisgesteuerte Studien. Mehrebenmodelle und A-priori-Verteilungen basierend auf Kontrollgruppendaten aus vorherigen Studien sind insbesondere in Situationen mit spärlichen Daten oder Parameteridentifizierbarkeitsproblemen hilfreich. Hierfür ist die vorgeschlagene Likelihood-Funktion besser geeignet als eine binomiale Likelihood-Funktion, weil die Austauschbarkeit von Ereigniszeitverteilungsparametern zwischen verschiedenen Studien plausibler ist als für den erwarteten Anteil von Patienten mit einem Ereignis. Mit Hilfe von Simulationen vergleichen wir auf der vorgeschlagenen Aggregatdaten-Likelihood-Funktion basierende Bayessche Mehrebenenmodelle mit anderen Metaanalysemethoden, als auch Methoden zur Benutzung der Kontrollgruppendaten aus vorherigen Studien. Wir zeigen auch, wie konjugierte A-priori Verteilungen für exponentialverteilte Ausfallzeiten konstruiert und zur Analyse ohne Benutzung des Markov-Chain-Monte-Carlo-Verfahrens genutzt werden können.

Acknowledgments

I would like to thank my advisers Prof. Rainer Schwabe und Dr. Ekkehard Glimm for their support, advice and suggestions. The discussions on many statistical topics with my colleagues at Novartis, especially Björn Bornkamp and Heinz Schmidli, were extremely enlightening and fruitful. I received valuable advice on implementing some of my work in Stan from Sebastian Weber and Bob Carpenter. I am also very grateful to my wife Jenny, who supported me in this project and proof-read my writings. I also indebted to our geese for sitting next to me, offering encouraging comments and only ripping apart one paper.

Contents

1. Introduction

1.1. Background

This work deals with the estimation of hazard ratios in time-to-event ("survival") models when no information on the exact event and censoring times is available for each individual experimental unit. Instead, only aggregate data (AD) data are available at a single observation time. Researchers face this type of problem when they have to rely on publicly available summaries of more detailed data, or when a continuous observation of the system under study is not feasible. This occurs in diverse applications such as educational research, economics, reliability engineering and medical research. We focus on the last of these applications and in particular methods for investigating whether a medical intervention changes the hazard rate for patients to experience a specific medical event.

Randomized controlled clinical trials (RCTs) are considered to provide the best evidence about the effects of a medical intervention (Higgins and Green, 2011, Section 1.3). RCTs compare the outcomes for patients receiving the intervention (the test group) and patients that did not receive it (the control group). The random assignment of concurrently recruited patients to test and control groups ensures that the distribution of outcomes only differs between groups due to the assignment to one of the two groups (Friedman et al., 2015, p. 36).

Some parts of this chapter are reproduced from Holzhauer (2017) (available at http://dx.doi.org/10.1002/sim.7181) with permission from John Wiley & Sons. Copyright ©2016 John Wiley & Sons, Ltd.

Often there are multiple RCTs that provide data relevant to a question and each RCT may not be able to conclusively answer it on its own. As a result, meta-analyses — "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings" (Glass, 1976) — have become a key tool of evidence based medicine for quantitatively combining the results of independent RCTs (Higgins and Green, 2011, Section 1.2.2).

Despite current efforts to increase the availability of individual patient data (IPD) from RCTs (Bertagnolli et al., 2017; Strom et al., 2014; Taichman et al., 2016), methods for AD are important for a number of reasons. Firstly, it is uncertain whether IPD will ever be accessible for historic clinical trials, to which new journal requirements or voluntary disclosure commitments do not apply. Secondly, many sponsors offer access to IPD via a secure server, from which data cannot be downloaded (Strom et al., 2014) so that an analysis cannot be simultaneously performed across RCTs from different sponsors. Finally, competitors are excluded from access to individual patient data (IPD) by most pharmaceutical companies. Thus, there is a need for methods using only AD extracted from the published literature.

In safety meta-analyses, events are often so rare that none occur in one or all treatment groups of some RCTs. It appears to be underappreciated that RCTs with no events in any group do provide information on relative risk measures, if there is prior information on what is expected for the control group.

For example, Senn (2008) pointed out that 6 out of 6 patients with an adverse event on a test drug compared with 0 out of 2 placebo patients may not be statistically significant at the one-sided 2.5% level according to Fisher's exact test, but that the prior knowledge on the rarity of the specific adverse event — a cytokine storm requiring admission to an intensive care unit — in healthy individuals means that these adverse events have been attributed to the test drug in the case of the first-in-human trial of TGN1412.

For this reason, it has been suggested that Bayesian meta-analyses that incorporate prior information would be particularly useful for rare events (Hamra et al., 2013; Lane,

2013; Mao and Xia, 1992; Sutton et al., 2002). In a recent review of safety meta-analyses, Bayesian methods were used only in 4% of papers with no increase (3% of papers) for rare events (Warren et al., 2012). Weaver et al. (2016) suggest that the use of Bayesian analyses of drug safety using informative prior distributions has been limited by concerns about subjectivity in formulating prior distributions.

Such concerns may be partially addressed by basing prior distributions on historical data (Gelman et al., 2014, p. 13) or by avoiding the explicit formulation of an informative prior distribution in favor of implicitly inducing it through "borrowing of strength" across trials using hierarchical models assuming the exchangeability of parameters across trials (Browne and Draper, 2006; Higgins and Whitehead, 1996; U.S. FDA Center for Devices and Radiological Health, 2010). Because these approaches make non-randomized comparisons to control groups that were not concurrently recruited, the reliability of such comparisons and to what extent such data should be discounted based on potential between-trial differences are key questions (Schmidli et al., 2014).

1.2. Structure of this thesis

In Chapter 2 we review existing meta-analysis approaches for AD on medical event occurrence. We also introduce hierarchical meta-analysis models, their underlying assumption of exchangeability of parameters between trials, and how they are related to different approaches for borrowing information across trials and making use of historical data. In Chapter 3 we derive a likelihood based on commonly available AD from trials of a fixed planned duration under the assumption of exponentially distributed event and drop-out times, and assuming that within each treatment group all patients with an event are equally likely to die. We propose to use this likelihood in a Bayesian hierarchical meta-analysis model and to specify informative priors based on historical control group data. At the end of Chapter 3 we compare the proposed model to other AD and IPD meta-analysis approaches using simulations. In Chapter 4 we extend the proposed approach to other frequently encountered settings such as Weibull distributed event and drop-out times, event driven trials, and the situation when a patient may

experience multiple potentially fatal events. In Chapter 5 we demonstrate how conjugate priors may be used for stratified meta-analyses of exponentially distributed failure times. If it were possible to further extend this approach to hierarchical models, it could increase the efficiency of future simulation studies. We also conducted a simulation study comparing different methods for using historical control group data in Chapter 6. We discuss our findings and conclusions in Chapter 7. Additional results and background information are available in the appendices at the end of this work.

1.3. Notation, conventions and assumptions
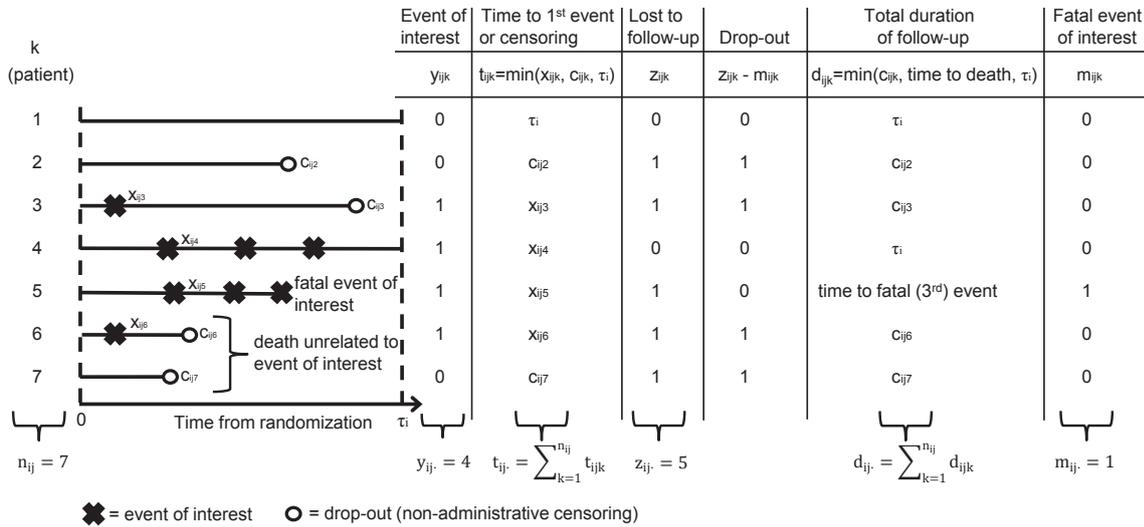
Figure 1.1.: Examples for the data used in the estimation of the event time distribution F_{ij} and the drop-out time distribution G_{ij} of treatment group $j = 0, 1$ of trial $i = 1, \ldots, I$

Unless noted otherwise, we assume that data $\mathcal{D}$ from I RCTs with planned durations $\tau_i > 0$ for $i = 1, \ldots, I$ are available. In addition, historical control group AD $\mathcal{D}'$ may be available from H historical trials $i = I + 1, \ldots, I + H$. The notation for the IPD and AD of an example trial is illustrated in Figure 1.1. We will deal solely with meta-analyses that aim to estimate an effect measure such as a hazard ratio comparing two treatment groups. For the inference on the treatment effect, good properties in terms of power, type I error rate, coverage probability mean squared error and bias are desired.

The number of patients in each of the two treatment groups $j = 0, 1$ is denoted by n_{ij} with n_{i1} patients receiving a test drug ($j = 1$) and n_{i0} patients being part of a

control group ($j = 0$). We denote the data from trial i by $\mathcal{D}_i$ and that from treatment group j of this trial by $\mathcal{D}_{ij}$. We use bold symbols for vectors or matrices, and non-bold symbols for scalar quantities. We use uppercase roman letters such as X_{ijk} for r.v.s and lowercase roman letters such as x_{ijk} for realizations of r.v.s.

We assume that the times X_{ijk} from trial start to event occurrence are random variables (r.v.s) that are independent and identically distributed (i.i.d.) for all patients $k = 1, \ldots, n_{ij}$ within group $j = 0, 1$ of trial i and follow an absolutely continuous distribution with cumulative distribution function (cdf) $F_{ij}(x) := F(x; \boldsymbol{\theta}_{F,ij})$ with $F(0; \boldsymbol{\theta}_{F,ij}) = 0$ and $\lim_{s \to \infty} F(s; \boldsymbol{\theta}_{F,ij}) = 1$ parametrized in terms of a parameter vector $\boldsymbol{\theta}_{F,ij}$ and probability density function (pdf) $f_{ij}(x) := f(x; \boldsymbol{\theta}_{F,ij})$.

Some patients may not have an observed event, because each patient is only observed for the fixed planned trial duration $\tau_i > 0$. For patients that are observed to be event free at time τ_i, the event time X_{ijk} is said to be administratively censored, because we only know that $x_{ijk} > \tau_i$. Some patients may drop out — i.e. no longer be followed for events before time τ_i. We denote the drop-out (non-administrative censoring) times by C_{ijk} and assume that they are i.i.d. for all patients within group $j = 0, 1$ of trial i with an absolutely continuous cdf $G_{ij}(c) := G(c; \boldsymbol{\theta}_{G,ij})$ with $G(0; \boldsymbol{\theta}_{G,ij}) = 0$ and $\lim_{s \to \infty} G(s; \boldsymbol{\theta}_{G,ij}) = 1$ parametrized in terms of a parameter vector $\boldsymbol{\theta}_{G,ij}$ and pdf $g_{ij}(c) := g(c; \boldsymbol{\theta}_{G,ij})$.

While the literature on typical drop-out patterns in RCTs is limited, an exponential distribution is often assumed (Lachin and Foulkes, 1986) and the Weibull distribution has been reported to fit the drop-out data from actual trials well (Gueorguieva et al., 2012; Sun et al., 2013; Veyrat-Follet et al., 2000).

We will assume that the event times X_{ijk} and censoring times C_{ijk} follow a joint distribution with cdf $H_{ij}(x, c) := H(x, c; \boldsymbol{\theta}_{H,ij})$ and pdf $h_{ij}(x, c) := h(x, c; \boldsymbol{\theta}_{H,ij})$. We will generally assume that C_{ijk} and X_{ijk} are independent for each patient within the same treatment group and trial with distinct parameters so that $\boldsymbol{\theta}_{H,ij} = \{\boldsymbol{\theta}_{F,ij}, \boldsymbol{\theta}_{G,ij}\}$ and $h(x, c; \boldsymbol{\theta}_{H,ij}) = f(x; \boldsymbol{\theta}_{F,ij}) g(c; \boldsymbol{\theta}_{G,ij})$.

Further, we define M_{ijk} as the indicator for whether a patient died due to an event

of interest. Different models will assume different processes underlaying the random variable M_{ijk}, which will be described when needed.

Instead of the IPD consisting of realizations of the r.v.s $T_{ijk} := \min(C_{ijk}, X_{ijk}, \tau_i)$, $Y_{ijk} := \mathbb{1}\{X_{ijk} \leq \min(C_{ijk}, \tau_i)\}$ and $Z_{ijk} := \mathbb{1}\{C_{ijk} < \tau_i \text{ or } M_{ijk} = 1\}$, in practice we will often only have the AD that $y_{ij\cdot} := \sum_{k=1}^{n_{ij}} y_{ijk}$ patients experienced an event of interest, that $z_{ij\cdot}$ of the n_{ij} patients in treatment group j of trial i were drop-outs, and that $m_{ij\cdot}$ of these $y_{ij\cdot}$ patients died of an event of interest.

We use "$\cdot$" in a subscript as a short-hand notation for summation over all values of the subscript replaced by the "$\cdot$" — e.g. $y_{i\cdot\cdot} := \sum_{j=0}^{1} \sum_{k=1}^{n_{ij}} y_{ijk}$. Without IPD $t_{ij\cdot}$ is typically unavailable and we do not know how many patients were lost to follow-up after experiencing a non-fatal event — i.e. the overlap in patients between $y_{ij\cdot} - m_{ij\cdot}$ and $z_{ij\cdot} - m_{ij\cdot}$.

Further notation and symbols are introduced as they are needed. Unless they only occur in the same section, in which they are introduced, they are listed in the List of symbols and abbreviations.

2. A review of meta-analysis methods for event occurrence

A key differences between a meta-analysis and the analysis of a single RCT is that differences between groups from different RCTs in the distribution of outcomes may occur for reasons other than the random assignment to a specific treatment group, e.g. due to differences in trial duration, recruited patients or outcome assessment. As a result a naive pooling of the data from multiple trials is susceptible to confounding between trial and treatment effects (Pearson et al., 1899; Simpson, 1951; Yule, 1903). Statistical meta-analysis methods avoid this by allowing either for completely unrelated unconstrained distributional parameters for each trial or assuming that the separate parameters for each trial are related via random effects (Ohlssen, 2014). In the following sections we will discuss meta-analysis approaches starting with those that treat event occurrence as a binomial outcome.

Time-to-event methods using IPD are considered the "gold standard" for analyzing the occurrence of medical events in RCTs (Bennett et al., 2013; Crowther et al., 2012; Lyman and Kuderer, 2005; Stewart and Clarke, 1995) for several reasons. Firstly, the probability of a patient experiencing almost any event increases the longer he is followed. Thus, patients that drop-out are less likely to have an observed event, if drop-out times are independent of event times within each treatment group. This has the potential to bias treatment comparisons, if the drop-out time distributions differ between treatment

groups. Time-to-event methods explicitly account for censoring. Secondly, it is often not just of interest whether an intervention prevents the occurrence of an event, but also whether the intervention delayed the occurrence of an event (or accelerated it). Time-to-event methods gain in power the larger the proportion of patients with an event is, while this is not the case for binomial data methods.

Finally IPD allow us to include patient characteristics affecting a patient's risk for an event into our analysis model. Even though randomization balances observed and unobserved covariates on average within each RCT (Senn, 1994), generalized linear models may give biased estimates of patient-level — as opposed to population-level — effects when covariates $\boldsymbol{b}_{ijk}$ that affect the outcome are omitted from the model. Neuhaus and Jewell (1993) showed that the employed link-function g, where $g(EY_{ijk}) = \boldsymbol{b}_{ijk}^T \boldsymbol{\beta}$, determines whether and in which direction the estimates of the treatment effect from generalized linear models in RCTs are biased when covariates that affect the time-to-event are omitted: when $-1/g'(\cdot)$ is convex (e.g. for the logit-, probit- and complementary log-log-link function) the estimate of the treatment effect is biased towards zero, biased towards inflated effect sizes when this function is concave and unbiased otherwise (e.g. for the linear- and log-link functions). The estimates of log-odds ratios from logistic regression, as well as of log-hazard ratios from exponential — except when there is no censoring — Cox (Gail et al., 1984) and Weibull (Lancaster, 1985) time-to-event models are attenuated towards zero when covariates are omitted.

IPD are often unavailable, because journal articles and clinical trial results databases only report AD. For this reason our discussion of IPD time-to-event methods primarily serves as a background for AD methods. We discuss existing AD time-to-event methods in Section 2.3. Sometimes IPD are available for a subset of trials. Methods for combining a mixture of IPD and AD have been developed (Riley and Steyerberg, 2010; Riley et al., 2008; Sutton et al., 2008) and we describe how to do so for our proposed model.

We discuss hierarchical models and their underlying assumption of exchangeable trial-level parameters in Section 2.4, but we already encounter examples of such models in preceding sections. Finally, we discuss Bayesian hierarchical meta-analysis models and different approaches for incorporating prior information into a meta-analysis.

For a full discussion of all aspects of meta-analyses such as how to select RCTs for inclusion in meta-analyses, issues surrounding the unavailability of the results of some RCTs or indirect treatment comparisons using network meta-analysis, we refer the reader to the Cochrane handbook for systematic reviews of interventions (Higgins and Green, 2011) or the book by Borenstein et al. (2009).

2.1. Fixed and random effects meta-analyses

Whether a fixed treatment effect parameter or one that is allowed to vary across trials should be estimated is a controversial topic. It has recently been suggested that much more heterogeneity in treatment effect estimates across trials is present in meta-analyses than usually identified (Kontopantelis et al., 2013). However, alternative explanations for such an observed heterogeneity have been insufficiently explored. These include differential losses to follow-up combined with withdrawal rates that increase with trial duration or the choice of the effect measure (Whitehead and Whitehead, 1991) such as a fixed odds, or risk ratio, which are inconsistent with a constant effect on a hazard scale except for small hazard ratios and very rare events (Symons and Moore, 2002).

When the events of interest are rare, meta-analyses typically focus on whether there is a treatment effect, at all, for which models that allow the treatment effect to vary across trials are not recommended (Higgins and Green, 2011, Section 16.9.5). For these reasons we focus on estimating a fixed treatment effect across trials in all that follows.

2.2. Event occurrence as a binomial outcome

The most commonly reported AD on medical event occurrence in publications and clinical trial results registries is the number of patients with an event $y_{ij.}$ out of the n_{ij} patients in treatment group $j = 0, 1$ of trial $i = 1, \ldots, I$. A binomial distribution for $Y_{ij.}$ is frequently assumed. In this section we discuss approaches for the meta-analysis of binomial outcomes.

2.2.1. Logistic regression

For meta-analysis purposes a logistic regression model that assumes $Y_{ij\cdot} \sim \text{Bin}(n_{ij}, \pi_{ij})$ with a logit-link function

$$
\begin{aligned}
\text{logit}\, \pi_{ij} &:= \log \frac{\pi_{ij}}{1 - \pi_{ij}} \\
&= \beta_i + j\varphi.
\end{aligned}
\tag{2.2.1}
$$

is often used, where $\beta_i = \log(\frac{\pi_{i0}}{1-\pi_{i0}})$ represents the log-odds of a patient in the control group of trial $i = 1, \ldots, I$ experiencing an event and φ is the log-odds ratio for test compared with control groups. Parameter estimates and asymptotic confidence intervals (CIs) can be obtained by maximum likelihood (Agresti, 2007, pp. 106–110). By using a log- or identity-link function in (2.2.1), φ would represent a log-risk ratio or risk difference, respectively. The odds ratio is more popular than these alternative effect measures, because unlike these alternatives the odds ratio ensures that π_{i1} will lie in $(0, 1)$ for any $\pi_{i0} \in (0, 1)$ and $\varphi \in (-\infty, \infty)$.

In practice there are a number of difficulties with obtaining unconditional maximum likelihood estimates (MLEs) for a logistic regression model. Firstly, no finite MLE exists for some of the model parameters, when $y_{i0\cdot} = 0$ or $y_{i0\cdot} = n_{i0}$ for at least one $i = 1, \ldots, I$, or if $y_{\cdot 1\cdot} = 0$ or $y_{\cdot 1\cdot} = n_{\cdot 1}$. Secondly, point estimates are known to be biased when there are only a few patients with or without an event (Heinze and Schemper, 2002). Thirdly, CIs and p-values based on an asymptotic normal approximation may perform poorly in the small sample cases (Heinze and Schemper, 2002).

For these reasons Firth's penalized likelihood logistic regression and exact conditional logistic regression are popular for the purpose of meta-analyses. Firth's penalized likelihood logistic regression approach removes the first-order term of the asymptotic bias of the MLE by penalizing the likelihood by Jeffreys prior and leads to the existence of a finite maximum-a-posteriori estimate even when a finite MLE does not exist (Heinze and Schemper, 2001). As a meta-analysis method, exact conditional logistic regression uses the exact permutational distributions of the sufficient statistic for the regression

coefficient for the treatment effect after eliminating the trial main effects by conditioning on their sufficient statistic — the number of patients with an event in each trial (Mehta and Patel, 1995). In this manner one can construct tests, CIs and median unbiased estimates that exist even when no MLE exists (Mehta and Patel, 1995). However, non-randomized and non-mid-P exact tests and CIs tend to be conservative in the small sample case (Bradburn et al., 2007).

The beta-binomial model proposed by Kuss (2015) modifies (2.2.1) to logit $\pi_{ij} =$ logit $u_{ij} + j\varphi$, where φ is the log-odds ratio and $U_{ij} \sim \text{Beta}(a, b)$ is a beta-distributed treatment group specific random effect. Alternatively Platt et al. (1999) proposed to let β_i in (2.2.1) be a normally distributed random trial effect.

2.2.2. The inverse variance estimate for odds ratios

Historically meta-analysis methods have focused on combining analysis results across trials based on published estimates for each trial and their standard errors (SEs) (Bagos and Nikolopoulos, 2009). Meta-analysis methods such as the inverse variance method that rely on these inputs are referred to as a two-step approach, because two analysis steps occur: first each individual trial is analyzed and secondly the results of these analyses are combined.

The MLE of the log-odds ratio for logistic regression based on a single trial with $0 < y_{ij\cdot} < n_{ij}$ for $j = 0, 1$ is $\hat{\varphi}_i := \log\left(y_{i1\cdot}(n_{i0} - y_{i0\cdot})(n_{i1} - y_{i1\cdot})^{-1}y_{i0\cdot}^{-1}\right)$ with asymptotic variance estimate $\widehat{\text{SE}}(\hat{\varphi}_i)^2 = y_{i1\cdot}^{-1} + (n_{i0} - y_{i0\cdot})^{-1} + (n_{i1} - y_{i1\cdot})^{-1} + y_{i0\cdot}^{-1}$ (Borenstein et al., 2009, p. 36). Given independent estimates $\hat{\varphi}_i$ and their SEs from each trial $i = 1, \ldots, I$, the inverse variance fixed effects estimate of φ — the minimum variance weighted average of the $\hat{\varphi}_i$ — is

$$\hat{\varphi} = \frac{\sum_{i=1}^{I} \frac{\hat{\varphi}_i}{\widehat{\text{SE}}(\hat{\varphi}_i)^2}}{\sum_{i=1}^{I} \widehat{\text{SE}}(\hat{\varphi}_i)^{-2}}$$

with estimated variance $\widehat{\text{SE}}(\hat{\varphi})^2 = \left(\sum_{i=1}^{I} \widehat{\text{SE}}(\hat{\varphi}_i)^{-2}\right)^{-1}$ (Borenstein et al., 2009, pp. 63–66).

Asymptotically the log-likelihood contribution from each trial is well approximated by $\hat{\varphi}_i$ and $\widehat{\text{SE}}(\hat{\varphi}_i)$ — i.e. a normal approximation. This approximation is improved for finite sample sizes by combining log-odds ratios or log-hazard ratio estimates rather than untransformed ratios (Borenstein et al., 2009, pp. 35–36). However, the approximation performs poorly for the meta-analysis of the occurrence of rare events — i.e. when very few patients experienced an event in at least one treatment group (Higgins and Green, 2011, Section 16.9.1).

In fact, if either $y_{i1\cdot}$, $(n_{i0} - y_{i0\cdot})$, $(n_{i1} - y_{i1\cdot})$ or $y_{i0\cdot}$ is zero for a trial, we do not obtain a finite MLE $\hat{\varphi}_i$ for that trial. In this scenario most statistical software packages use a so-called "continuity correction" by automatically adding 0.5 to all four of these quantities (Higgins and Green, 2011, Section 16.9.2) and omit trials without any events from analyses (Higgins and Green, 2011, Section 16.9.3). However, the normal approximation remains questionable in case of rare events so that inverse variance methods are not recommended in this setting and one-step methods — such as all other analysis methods described in Section 2.2 — that directly analyze the data from all trials in a single step have been recommended instead (Higgins and Green, 2011, Section 16.9.5; Bradburn et al., 2007).

2.2.3. Peto one-step odds ratio method

The Peto one-step estimate of the log-odds ratio given a set of observed data is

$$\hat{\varphi} = \frac{\sum_{i=1}^{I}(y_{i1\cdot} - y_{i\cdot\cdot}n_{i1}/n_{i\cdot})}{\sum_{i=1}^{I}\frac{n_{i1}n_{i0}y_{i\cdot\cdot}(n_{i\cdot}-y_{i\cdot\cdot})}{n_{i\cdot}^2(n_{i\cdot}-1)}}$$

with estimated variance $\widehat{\text{SE}}(\hat{\varphi})^2 = \left(\sum_{i=1}^{I}\frac{n_{i1}n_{i0}y_{i\cdot\cdot}(n_{i\cdot}-y_{i\cdot\cdot})}{n_{i\cdot}^2(n_{i\cdot}-1)}\right)^{-1}$ (Borenstein et al., 2009, pp. 336–337). It is the method recommended in the Cochrane handbook when the proportion of patients with an event is below 1%, but it is known to be biased for large effect sizes and strongly imbalanced allocation ratios within trials (Higgins and Green, 2011, Section 16.9.5), and its limitations are further discussed by Brockhaus et al. (2016), and Hirji (2006, p. 271).

2.2.4. Cochran-Mantel-Haenszel odds-ratio

An alternative estimate of the log-odds ratio without a continuity correction is the stratified Cochran-Mantel-Haenszel (CMH) estimate of the log-odds ratio

$$\hat{\varphi} = \log\left(\frac{\sum_{i=1}^{I} y_{i1\cdot}(n_{i0} - y_{i0\cdot})/n_{i\cdot}}{\sum_{i=1}^{I} y_{i0\cdot}(n_{i1} - y_{i1\cdot})/n_{i\cdot}}\right) \tag{2.2.2}$$

with estimated variance (Lachin, 2000, pp. 95–98)

$$\widehat{\mathrm{SE}}(\hat{\varphi})^2 = \frac{\sum_{i=1}^{I} \frac{(y_{i1\cdot} + n_{i0} - y_{i0\cdot})y_{i1\cdot}(n_{i0} - y_{i0\cdot})}{n_{i\cdot}^2}}{2(\sum_{i=1}^{I} \frac{y_{i1\cdot}(n_{i0} - y_{i0\cdot})}{n_{i\cdot}})^2} + \frac{\sum_{i=1}^{I} \frac{y_{i0\cdot}(n_{i1} - y_{i1\cdot})(y_{i0\cdot} + n_{i1} - y_{i1\cdot})}{n_{i\cdot}^2}}{2(\sum_{i=1}^{I} \frac{y_{i0\cdot}(n_{i1} - y_{i1\cdot})}{n_{i\cdot}})^2}$$
$$+ \frac{\sum_{i=1}^{I} \frac{(y_{i1\cdot} + n_{i0} - y_{i0\cdot})y_{i0\cdot}(n_{i1} - y_{i1\cdot}) + (y_{i0\cdot} + n_{i1} - y_{i1\cdot})y_{i1\cdot}(n_{i0} - y_{i0\cdot})}{n_{i\cdot}^2}}{2(\sum_{i=1}^{I} \frac{y_{i1\cdot}(n_{i0} - y_{i0\cdot})}{n_{i\cdot}})(\sum_{i=1}^{I} \frac{y_{i0\cdot}(n_{i1} - y_{i1\cdot})}{n_{i\cdot}})}.$$

It is more popular to re-write (2.2.2) in terms of a weighted sum of the odds ratios $\frac{y_{i1\cdot}(n_{i0} - y_{i0\cdot})}{y_{i0\cdot}(n_{i1} - y_{i1\cdot})}$ in each study with weights $y_{i0\cdot}(n_{i1} - y_{i1\cdot})/n_{i\cdot}$ for each study and to use continuity corrections as needed (Lachin, 2000, p. 95).

2.3. Event occurrence as a time-to-event outcome

2.3.1. Time-to-event models for individual patient data

When IPD are available for all RCTs to be meta-analyzed, a common approach is to conduct an IPD meta-analysis that uses trial as one factor in the model besides other covariates (Sutton et al., 2008) — much like the analysis of a RCT with data from several centers. Another popular alternative is to stratify the analysis by trial.

In a stratified parametric IPD meta-analysis the contribution of the observed data $\mathcal{D}_{ij}$ from treatment group j of trial i consisting of $(y_{ijk}, \min(x_{ijk}, \tau_i))^T$ for each patient $k = 1, \ldots, n_{ij}$ to the likelihood of the parameters $\boldsymbol{\theta}_{F,ij}$ of the event time distribution

with pdf $f(\cdot; \boldsymbol{\theta}_{F,ij})$ and cdf $F(\cdot; \boldsymbol{\theta}_{F,ij})$ is given by

$$\mathcal{L}(\boldsymbol{\theta}_{F,ij}; \mathcal{D}_{ij}) = \prod_{k=1}^{n_{ij}} f\left(\min(x_{ijk}, \tau_i); \boldsymbol{\theta}_{F,ij}\right)^{y_{ijk}} \left[1 - F\left(\min(x_{ijk}, \tau_i); \boldsymbol{\theta}_{F,ij}\right)\right]^{1-y_{ijk}},$$

if trial completion is the only form of censoring that occurs. When there are also drop-outs before trial completion, then the likelihood contribution of each trial arm for the parameters $\boldsymbol{\theta}_{H_{ij}}$ of the joint distribution of event and drop-out times with pdf $h(\cdot; \boldsymbol{\theta}_{H,ij})$ and cdf $H(\cdot; \boldsymbol{\theta}_{H,ij})$ is

$$\mathcal{L}(\boldsymbol{\theta}_{H,ij}; \mathcal{D}_{ij}) = \prod_{k=1}^{n_{ij}} h\left(t_{ijk}, \min(c_{ijk}, \tau_i); \boldsymbol{\theta}_{H,ij}\right)^{y_{ijk}z_{ijk}} \times \left[\int_{t_{ijk}}^{\infty} h(t_{ijk}, c; \boldsymbol{\theta}_{H,ij}) \, \mathrm{d}c\right]^{y_{ijk}(1-z_{ijk})}$$
$$\times \left[\int_{\min(c_{ijk}, \tau_i)}^{\infty} h(x, \min(c_{ijk}, \tau_i); \boldsymbol{\theta}_{H,ij}) \, \mathrm{d}x\right]^{z_{ijk}(1-y_{ijk})}$$
$$\times \left[1 - H(\tau_i, \tau_i; \boldsymbol{\theta}_{H,ij})\right]^{(1-y_{ijk})(1-z_{ijk})}.$$

Inference is typically focused on the parameters of the event time distribution, while the parameters of the distribution of drop-out times are regarded as nuisance parameters.

We will assume that X_{ijk} and C_{ijk} are independent with distinct parameters so that $\boldsymbol{\theta}_{H,ij}$ can be partitioned into two subsets of parameters: $\boldsymbol{\theta}_{F,ij}$ for the distribution of event times and $\boldsymbol{\theta}_{G,ij}$ for the distribution of drop-out times. In that case the contribution of a trial arm to the joint likelihood for $\boldsymbol{\theta}_{F,ij}$ and $\boldsymbol{\theta}_{G,ij}$ is

$$\mathcal{L}(\boldsymbol{\theta}_{F,ij}, \boldsymbol{\theta}_{G,ij}; \mathcal{D}_{ij}) = \prod_{k=1}^{n_{ij}} f\left(t_{ijk}; \boldsymbol{\theta}_{F,ij}\right)^{y_{ijk}} \left[1 - F(t_{ijk}; \boldsymbol{\theta}_{F,ij})\right]^{1-y_{ijk}}$$
$$\times g(d_{ijk}; \boldsymbol{\theta}_{G,ij})^{z_{ijk}} (1 - G(d_{ijk}; \boldsymbol{\theta}_{G,ij}))^{1-z_{ijk}},$$

where $d_{ijk} = \min(c_{ijk}, \tau_i)$. Thus, the contribution of a trial arm to the marginal likelihood of $\boldsymbol{\theta}_{F,ij}$ is

$$\mathcal{L}(\boldsymbol{\theta}_{F,ij}; \mathcal{D}_{ij}) \propto \prod_{k=1}^{n_{ij}} f(t_{ijk}; \boldsymbol{\theta}_{F,ij})^{y_{ijk}} (1 - F(t_{ijk}; \boldsymbol{\theta}_{F,ij}))^{1-y_{ijk}},$$

even if the drop-out time distributions differ across trials and treatment groups.

2.3.2. Popular parametric and semi-parametric individual-patient-data time-to-event models

2.3.2.1. Proportional hazards models

It is often useful to write a time-to-event model in terms of the log-hazard function $\log \lambda_{ij}(s)$ for arm j of trial i and $s \geq 0$, which given an event time distribution with cdf F and pdf f is defined as (Ibrahim et al., 2001, p. 14)

$$
\begin{aligned}
\log \lambda_{ij}(s) &:= \log \lim_{\Delta s \to 0} \frac{\mathbb{P}(s < X_{ijk} \leq s + \Delta s | X_{ijk} > s)}{\Delta s} \\
&= \log \left(-\frac{\mathrm{d}}{\mathrm{d}s} \log(1 - F(s; \boldsymbol{\theta}_{F,ij})) \right) \\
&= \log f(s; \boldsymbol{\theta}_{F,ij}) - \log(1 - F(s; \boldsymbol{\theta}_{F,ij})).
\end{aligned}
$$

Further we will assume a proportional hazards model so that

$$
\log \lambda_{ij}(s) = \log \lambda_{i0}(s) + \boldsymbol{b_{ijk}}^T \boldsymbol{\beta}, \tag{2.3.1}
$$

where $\boldsymbol{b_{ijk}}$ is a vector of covariates for each subject and $\boldsymbol{\beta}$ is a vector of regression coefficients. $\lambda_{i0}(s)$ in (2.3.1) is referred to as the trial-specific baseline hazard function. We will typically only consider the situation where the only factor or covariate in the model is the treatment effect, so that

$$
\log \lambda_{ij}(s) = \log \lambda_{i0}(s) + j\varphi, \tag{2.3.2}
$$

where φ is defined to be the log-hazard ratio for the test group ($j = 1$) compared with the control group ($j = 0$). Within the framework of proportional hazard models we will focus on the exponential time-to-event model, the Weibull model and the semi-parametric Cox regression model.

We focus on cases with a constant hazard ratio over time, because models with this assumption are the most frequently used and do — under moderate deviations from a constant log-hazard ratio — provide meaningful estimates of an average log-hazard ratio

weighted by the number of events in different time periods (Allison, 1995, pp. 154–155). However, if required the discussed time-to-event models can also be extended to allow for time varying effects (Allison, 1995, pp. 154–157).

2.3.2.2. The exponential time-to-event model

If we assume a constant baseline hazard function $\lambda_{i0}(s) = \lambda_{i0}$ in (2.3.2), we obtain an exponential time-to-event model with $X_{ijk} \sim \text{Exp}(\lambda_{ij})$, where $\log \lambda_{ij} := \log \lambda_{i0} + j\varphi$. The exponential distribution is a convenient one-parameter distribution, but makes the strong assumptions that each patients in group j of trial i has a rate that is constant over time and that all of these hazard rate are the same — unlike e.g. for a gamma-exponential mixture (Lomax distribution). The plausibility of these assumptions depends on the patient population enrolled in each trial. For example, in medically stable patients with a chronic condition there is no reason to expect a meaningful change in the control group hazard rates over the relatively short duration of a typical clinical trial. The less risk factors for events vary across patients, the better the approximation obtained by assuming an identical hazard rate for all patients will be. More complex distributions such as the Weibull or Lomax distributions may be needed to capture the effects of changes in control group hazard rates over time or between subject heterogeneity.

Note that the likelihood of the exponential time-to-event model assuming independent right censoring within each treatment group and trial is

$$
\begin{aligned}
\mathcal{L}(\phi, \lambda_{01}, \ldots, \lambda_{I1} | y_{011}, \ldots, y_{I1n_{I1}}, t_{011}, \ldots, t_{I1n_{I1}}) &= \prod_{i=1}^{I} \prod_{j=0}^{1} \prod_{k=1}^{n_{ij}} (\phi^j \lambda_{i0} e^{-\phi^j \lambda_{i0} t_{ijk}})^{y_{ijk}} \\
&\quad (e^{-\phi^j \lambda_{i0} t_{ijk}})^{1-y_{ijk}} \\
&= \prod_{i=1}^{I} \prod_{j=0}^{1} (\phi^j \lambda_{i0})^{y_{ij\cdot}} e^{-\phi^j \lambda_{ij} t_{ij\cdot}}
\end{aligned}
$$

(2.3.3)

for $\phi := e^{\varphi}$. Up to a normalizing constant that only depends on the observed data this is the same likelihood as for a Poisson distribution having observed $y_{ij\cdot}$ events in

$t_{ij.}$ time units of follow-up. As a result such a model can be fitted using software for Poisson regression with a $\log t_{ij.}$ offset (Crowther et al., 2012; Stijnen et al., 2010).

The exponential model is notably also an AD analysis method, because the likelihood (2.3.3) only requires the AD on the number of patients with an event $y_{ij.}$ amongst the n_{ij} patients in group $j = 0, 1$ of trial $i = 1, \ldots, I$ and the total follow-up to first event or censoring $t_{ij.}$ — the sufficient statistic for the parameters of an exponential time-to-event distribution under random right-censorship (Ibrahim et al., 2012).

Similar issues as those described for logistic regression may occur, if there are no events in some or all arms of a trial. For this reason two alternative models have been used. Firstly, a hierarchical exponential time-to-event model is available in many standard software packages, which uses a Poisson likelihood with gamma distributed trial random effect $\lambda_{i0} \sim \text{Gamma}(a, b)$ with shape parameter $a > 0$ and rate (inverse-scale) parameter $b > 0$, and a $\log t_{ij.}$ offset. Secondly, an exact version of Poisson regression stratified by trial, in which we condition on the total number of cases $y_{i..}$ in each trial $i = 1, \ldots, I$, is also available and can provide tests, CIs, and median unbiased estimates (Hirji, 2006, pp. 263–266).

2.3.2.3. Weibull model

The Weibull distribution is a widely used parametric event time distribution, which can reflect increasing or decreasing hazard functions (Ibrahim et al., 2012, p. 35). We assume the following parametrization of the Weibull pdf (Ibrahim et al., 2012, p. 14)

$$f_{\text{Wb}}(x; a, b) := ab(xa)^{b-1} \exp\left(-(xa)^b\right) \text{ for } a > 0 \text{ and } b > 0, \qquad (2.3.4)$$

which ensures that for shape parameter $b = 1$ the inverse scale parameter a corresponds to an exponential hazard rate. The corresponding cdf is

$$F_{\text{Wb}}(x; a, b) := 1 - \exp\left(-(xa)^b\right)$$

and the hazard function

$$\lambda_{\text{Wb}}(x; a, b) := ba^b x^{b-1}.$$

If we assume that the event times in the control group $j = 0$ of trial $i = 1, \ldots, I$ are i.i.d. a Weibull distributed r.v.s $X_{i0k} \sim \text{Wb}(\lambda_{i0}, \gamma_i)$ for $k = 1, \ldots, n_{i0}$ and assume a proportional hazards model with a log-hazard ratio $\varphi := \log \phi$, then the event times in the test group are i.i.d. Weibull distributed r.v.s $X_{i1k} \sim \text{Wb}(\lambda_{i0} e^{\varphi/\gamma_i}, \gamma_i)$ with the same shape parameter as in the control group.

2.3.2.4. Cox regression

Cox's proportional hazards model (Cox, 1972) has gained considerable popularity, because it does not assume a particular functional form for the baseline hazard function $\lambda_{i0}(s)$ in (2.3.1). If no events occurred at the same time despite any rounding of time units, a stratified Cox regression that allows unrelated baseline hazard functions for each trial uses the partial likelihood

$$\mathcal{PL}(\varphi | \mathcal{D}) = \prod_{i=1}^{I} \prod_{j=0}^{1} \prod_{k=1}^{n_{ij}} \left[\frac{e^{j\varphi}}{\sum_{\ell=0}^{1} \sum_{r=1}^{n_{i\ell}} \mathbb{1}\{t_{i\ell r} \leq t_{ijk}\} e^{\ell \varphi}} \right]^{y_{ijk}}.$$

The test resulting from a Cox regression model is equivalent to the log-rank test, if we assume that there is no underlying ordering of tied event times (Allison, 1995, pp. 127–137). Similar issues as those discussed for logistic regression in Section 2.2.1 can occur with Cox regression and Firth's penalized (partial) likelihood approach can also be applied to Cox regression (Heinze and Schemper, 2001). It has been reported to perform well for meta-analyzing time-to-event IPD (Bennett et al., 2013).

2.3.3. Time-to-event approaches for aggregate data

Three main types of approach for survival meta-analysis when IPD are not available have been previously described. The first type of approach assumes that some type of survival analysis has already been conducted and that the resulting Kaplan-Meier plots, hazard ratios or p-values are available (Arends et al., 2008; Combescure et al., 2016; Guyot et al., 2012; Hoyle and Henley, 2011; Parmar et al., 1998; Tierney et al., 2007; Wei et al., 2015; Williamson et al., 2002). The second type of approach uses

the exponential survival model discussed in Section 2.3.2.2, which can be seen as both an IPD and an AD model. Finally, some authors have proposed to modify binomial data methods to account for differential follow-up.

2.3.3.1. Practical issues with the exponential aggregate data model

The main practical issue with the exponential time-to-event model as an AD method is the availability of the total time to first event or censoring $t_{ij\cdot}$, which the consolidated standards of reporting trials (CONSORT) guidelines (Schulz et al., 2010) do not require to be reported. In a rare events setting, it may be reasonable to use the total duration of follow-up to censoring $d_{ij\cdot}$ instead of $t_{ij\cdot}$, as e.g. Stijnen et al. (2010), Böhning et al. (2015) and Kunz et al. (2015) have done. This is especially appropriate, if the event of interest is always fatal so that $d_{ij\cdot} = t_{ij\cdot}$. However, for non-rare, not necessarily fatal events this leads to an underestimation of the hazard rate in each trial arm. Additionally, $d_{ij\cdot}$ and the duration of exposure to therapies are also typically unavailable. Methods for estimating or imputing these values based on commonly available information have not received attention in the literature so far.

2.3.3.2. Binomial data methods with adjustment for differential follow-up

It has been suggested that one could adjust binomial data methods by e.g. adding the logarithm of the average exposure — or perhaps more appropriately follow-up — as an offset variable in logistic regression (Lane, 2013). If data on the duration of follow-up are available, then this may approximate the relationship of follow-up to the log-odds of an event. Similarly, the Decision Support Unit of the UK National Institute for Health and Clinical Excellence has suggested a log trial duration offset for a binomial model with a complementary log-log link function (Dias et al., 2014). This is motivated by the fact that assuming exponentially distributed event times $X_{ijk} \sim \text{Exp}(\lambda_{i0}e^{\varphi j})$, we obtain $Y_{ijk}|(D_{ijk} = d_{ijk}) \sim \text{Ber}(\pi_{ijk})$ we obtain

$$\log(-\log(1 - \pi_{ijk})) = \log \lambda_{i0} + j\varphi + \log d_{ijk}, \tag{2.3.5}$$

i.e. a generalized linear model for binomial data with a complementary log-log link-function and a $\log d_{ijk}$ offset (Dias et al., 2014). If all patients were followed for the same planned duration τ_i, we can substitute $\log \tau_i$ in (2.3.5). However, when follow-up differs across patients, e.g. due to losses to follow-up, using a $\log \tau_i$ or $\log(d_{ij.}/n_{ij})$ offset is also only an approximation to the IPD relationship (2.3.5). Whether a logit or a complementary log-log link function is used, the quality of these approximations depends on the distribution of failure times and is reasonable for a low drop-out hazard rate that is close to constant over time. Lacking alternatives, these approximations for low and close to constant hazard rates may be preferable to ignoring differential follow-up between treatment groups. However, we are not aware of any advantage of using such adjustments instead of an exponential time-to-event model. For either approach the same AD is needed and exponentially distributed event times are assumed.

2.4. Exchangeability and hierarchical models

The random effects models discussed at the end of Section 2.2.1 and the gamma-exponential model in Section 2.3.2.2 are examples of 2-stage hierarchical (also called "multi-level" or "mixed-effects") meta-analysis models. These are models that allow for a specification of the joint distribution of the data from all trials in terms of successive hierarchical levels of conditional distributions. For example the likelihood for the type of hierarchical model we will primarily discuss throughout this thesis can be specified as

$$\mathcal{L}(\boldsymbol{\zeta}, \boldsymbol{\vartheta}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I | \mathcal{D}) = \prod_{i=1}^{I} p(\boldsymbol{\theta}_i | \boldsymbol{\vartheta}) \prod_{j=0}^{1} p(\mathcal{D}_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\zeta}_j). \qquad (2.4.1)$$

$\boldsymbol{\vartheta}$ denotes the hyperparameters determining the distribution of i.i.d. latent r.v.s $\boldsymbol{\theta}_i$ through the conditional pdf or probability mass function (pmf) $p(\boldsymbol{\theta}_i | \boldsymbol{\vartheta})$. $\boldsymbol{\theta}_i$ and $\boldsymbol{\zeta}$ together determine the distribution of the observable data $\mathcal{D}_{ij}$ in each trial arm. We chose to split the hyperparameter vector into $\boldsymbol{\vartheta}$ and $\boldsymbol{\zeta}$ for the clarity of exposition. In the models we consider, $\boldsymbol{\zeta}$ will have treatment arm specific components $\boldsymbol{\zeta}_j$ for $j = 0, 1$ — e.g. the gamma-exponential model in Section 2.3.2.2 can be parametrized as $\log \lambda_{ij} = \log \lambda_{i0} + \zeta_j$ with $\zeta_0 = 0$ and $\zeta_1 = \varphi$.

We follow Bernardo and Smith (2000, ch. 4) in outlining the assumptions made in specifying a hierarchical model, in particular that of partial exchangeability — an extension of the concept of independence.

Definition 2.1 (Finite exchangeability). A set of r.v.s $X_1, \ldots, X_I$ are finitely exchangeable under a probability measure P, if the joint probability distribution satisfies

$$P(X_1, \ldots, X_I) = P(X_{\rho(1)}, \ldots, X_{\rho(I)})$$

for all permutations ρ defined on the set $\{1, \ldots, I\}$ (Bernardo and Smith, 2000, p. 169).

Definition 2.2 (Infinite exchangeability). The infinite sequence of r.v.s $X_1, X_2, \ldots$ is infinitely exchangeable, if every finite subset of the sequence is finitely exchangeable in the sense of Definition 2.1 (Bernardo and Smith, 2000, p. 171).

Remark 2.1. Exchangeable r.v.s generalize the concept of i.i.d. r.v.s. In either case the r.v.s are marginally identically distributed, but the joint cdf of exchangeable r.v.s is not necessarily the product of the marginal cdfs unlike for i.i.d. r.v.s. Instead, Theorem 2.1 provides a representation in terms of r.v.s that are independent conditional on latent variables.

Theorem 2.1. *If $\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots$ is an infinitely exchangeable sequence of r.v.s $\in \mathbb{R}^d$ for $d \geq 1$, each with a marginal density (i.e. Radon-Nikodym derivative) $p(\boldsymbol{x}_i | \boldsymbol{\theta})$ for $\boldsymbol{\theta} \in \Theta$, then there exists a probability measure Q over Θ so that for any subset $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_I$ of the sequence the joint density of the subset has the form (Bernardo and Smith, 2000, pp. 177–180)*

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_I) = \int_{\Theta} \prod_{i=1}^{I} p(\boldsymbol{x}_i | \boldsymbol{\theta}) \, \mathrm{d}Q(\boldsymbol{\theta}).$$

Remark 2.2. Authors such as Bernardo and Smith (2000, ch. 4) and Gelman et al. (2014, pp. 104–108) give Theorem 2.1 as the justification for why, after a judgment of exchangeability has been made, a hierarchical model can be employed. However, Theorem 2.1 is an existence theorem that does not specify the form of the probability measure Q. I.e. a density $p(\boldsymbol{\theta}_i | \boldsymbol{\vartheta})$ for latent trial-level parameters $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I$ given a finite dimensional hyperparameter $\boldsymbol{\vartheta}$ as in (2.4.1) does not necessarily exist. On

this point Spiegelhalter et al. (2004, p.168) argue that a — possibly multivariate — normal distribution for the latent parameters is reasonable based on the central limit theorem applied to the sum of all the small differences in trials that contributed to the variation of latent parameters across trials. McCulloch and Neuhaus (2011) suggest that inference from hierarchical models is in many settings robust to deviations from this assumption. For this reason we will primarily consider hierarchical models of the form (2.4.1) with normally distributed random trial effects for suitably transformed nuisance parameters. I.e. $p(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I | \boldsymbol{\vartheta}) = \phi(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I | \boldsymbol{\nu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\vartheta}$ consists of a vector of means $\boldsymbol{\nu}$ of dimension $p \geq 1$ and a $p \times p$ covariance matrix $\boldsymbol{\Sigma}$. We will assume that $\boldsymbol{\Sigma}$ is a diagonal matrix with diagonal entries $\sigma_1^2, \ldots, \sigma_p^2$. Friede et al. (2017) consider a between-trial standard deviation of σ_r for $r = 1, \ldots, p$ of 0.125 as small, 0.25 as moderate, 0.5 as substantial, 1 as large and 2 as very large for an log-odds ratio. An example of a non-normally distributed random trial effect is the gamma-exponential model in Section 2.3.2.2.

Remark 2.3. Corresponding results to Theorem 2.1 hold under partial exchangeability — i.e. when the outcomes in each trial are conditionally independent given known covariates in addition to unknown latent parameter vectors $\boldsymbol{\theta}_i$ (Bernardo and Smith, 2000, pp.209–226). Such covariates might be the planned trial duration τ_i, or the number of randomized patients in each of the treatment group n_{i0} and n_{i1}. We refer to this form of partial exchangeability when stating that a parameter vector $\boldsymbol{\theta}_i$ for $i = 1, \ldots, I$ is exchangeable across trials.

In hierarchical models the estimation of each latent trial-level parameter $\boldsymbol{\theta}_i$ influences the estimation of those in other trials through the hyper-distribution $p(\boldsymbol{\theta}_i | \boldsymbol{\vartheta})$ — effectively inducing a prior distribution based on the other trials (Gelman et al., 2014, p. 1). This has often been referred to as "borrowing of information" across trials and hierarchical models have gained considerable popularity in the last quarter century due to this property (Browne and Draper, 2006; Higgins and Whitehead, 1996; U.S. FDA Center for Devices and Radiological Health, 2010). As a consequence of this "borrowing of information" trial-level parameters may be identifiable in such a hierarchical model despite being non-identifiable based on the available data from each trial alone.

The exchangeability assumption made when formulating a hierarchical model implies that there is a-priori no reason to assume the parameters for one trial would be larger or smaller than those for another trial (Gelman et al., 2014, pp. 104–107) (U.S. FDA Center for Devices and Radiological Health, 2010). Yet the plausibility of this assumption is not always given sufficient consideration.

For example, the beta-binomial model of Kuss (2015) assumes the exchangeability of the expected proportion of patients with an event across arms. However, this is a questionable exchangeability assumption, because trials with longer follow-up would be expected to have a higher proportion of cases than shorter trials. Nevertheless, such an assumption has been advocated or used for the meta-analysis of studies of substantially differing duration by several other authors (Cai et al., 2010; Kaizar et al., 2006; Kuss, 2015; Stijnen et al., 2010; Warren et al., 2014; Woods et al., 2010).

More plausible exchangeability assumptions between trials of different duration can be made in time-to-event models on the log-hazard for event occurrence as e.g. in the gamma-exponential model discussed in Section 2.3.2.2.

2.5. Bayesian meta-analysis approaches

2.5.1. Bayesian inference

Bayesian inference combines the likelihood for the observed data — in the Bayesian context denoted by $p(\mathcal{D}|\boldsymbol{\theta})$ instead of $\mathcal{L}(\boldsymbol{\theta}|\mathcal{D})$ — with a prior distribution (or "prior" for short) with density $p(\boldsymbol{\theta})$ for the model parameters $\boldsymbol{\theta}$ to obtain the posterior distribution (or "posterior" for short) of the parameters given the data with density $p(\boldsymbol{\theta}|\mathcal{D})$ using Bayes' theorem (Gelman et al., 2014, pp. 6–7)

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\mathcal{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}).$$

Meta-analyses of two or more trials in the Cochrane database of systematic reviews are based on a median of 3 trials (inter-quartile range 2 to 6) with a median of 91 patients

per trial (inter-quartile range 44 to 210) (Davey et al., 2011). This emphasizes the need for meta-analysis methods that are robust to sparse data situations — both in terms of sparse data within each trial and a low number of trials.

Bayesian approaches — even when using only vague priors — often perform well in terms of repeated sampling properties in sparse data situations (Gelman and Rubin, 1995). This is for example the case for random effects meta-analyses, where standard frequentist methods perform badly in terms of coverage for confidence and prediction intervals in some settings — particularly when the heterogeneity of the treatment effect across trials is low and study sizes are varied (Bodnar et al., 2017; Partlett and Riley, 2017). In contrast, Bayesian methods using vague priors (Bai et al., 2016; Bodnar et al., 2017) or using informative priors for the between-trial standard deviation based on historical data (Higgins and Whitehead, 1996; Turner et al., 2015) perform better.

2.5.2. Markov chain Monte Carlo methods

In practice Bayesian inference is primarily performed using Markov chain Monte Carlo (MCMC) methods. The main reason is that analytical solutions for the posterior are primarily available for a specific choice of prior ("conjugate priors") for the rather limited class of distributions with sufficient statistic of fixed dimension given a large enough sample size (Robert, 2007, pp. 116–121). MCMC methods draw pseudo-random samples from a simulated Markov chain with a limiting distribution that is the posterior of the model parameters. If the Markov chain has converged to its limiting distribution and samples are taken from sufficiently far apart steps ("thinning"), then these pseudo-random samples can be treated as i.i.d. samples from the posterior (Gelman et al., 2014, pp. 282–283). When starting a MCMC algorithm at an arbitrary starting value for the parameters of interest, it is a typical practice to ignore an initial set of samples in the posterior inference. These ignored samples are referred to as "burn-in samples" (Gelman et al., 2014, p. 282). There is a wide range of MCMC algorithms, out of which we primarily employed Hamiltonian Monte Carlo using the No-U-Turn sampler (NUTS) (Hoffman and Gelman, 2014), Gibbs sampling and an adaptive blocked form of the random walk Metropolis algorithm (Gelman et al., 2014, pp. 275–281).

2.5.3. Vague, weakly informative and informative prior distributions

In the absence of prior information it is often considered desirable to specify priors that exert minimal influence on the posterior (Gelman et al., 2014, p. 51). Such default priors are a topic of ongoing research, on which Bhadra et al. (2016) provide a recent overview. We will use the convention of reserving the term "vague prior" for very diffuse proper priors to distinguish these from — possibly improper — priors that are flat with respect to some parametrization of the likelihood and are sometimes referred to as "uninformative". The use of proper priors avoids the possibility of an improper posterior (Robert, 2007, p. 39), which may otherwise occur and cannot always be identified from the output generated by MCMC methods (Robert, 2007, p. 48).

Informative priors are priors that reflect available information on parameters e.g. based on the posterior of a previous analysis. Section 2.5.4 discusses methods for obtaining informative priors from historical data. We will refer to priors as "weakly informative", if available informative prior information is down-weighted in order to avoid too strong an influence of the prior on the posterior, while still giving low prior probability to those parts of the parameter space a-priori considered to contain implausible values (Gelman et al., 2014, pp. 51–52). When using weakly informative priors, posterior inference enjoys many of the same advantages as with vague priors (Gelman et al., 2014, pp. 51–52), but any shrinkage occurs towards more plausible parameter values.

2.5.4. Methods for incorporating historical prior information in Bayesian hierarchical models

The Bayesian models we will consider throughout this thesis will be of the form described in Remark 2.2. We will discuss the choice of vague and weakly informative priors for the mean and scale hyperparameters of these hierarchical models in Section 3.4.1 in the context of a specific proposed model. We discuss the choice of the prior for the treatment effect in a meta-analysis in Section 3.4.2 and Appendix A.

While vague proper priors already confer some desirable frequentist operating charac-
teristics on Bayesian methods, there is considerable promise in making use of available
historical data to formulate informative priors. While prior information about a new
test treatment may be limited, several previous trials studying the control group may be
available to provide information on the nuisance parameters describing the distribution
of control group outcomes (Schmidli et al., 2014). However, while an exchangeability
of the parameters in historical and new trials may often be plausible, there will in-
evitably be concerns that there could be systematic differences that could bias posterior
inference (Schmidli et al., 2014). In this section we review a number of the existing
proposals for making posterior inference more robust against such prior-data conflicts.

While a range of methods has been proposed for this purpose, a systematic evaluation
and comparison between these options has not been conducted, except for a paper
by Li et al. (2016) that compared a limited set of methods. We start by reviewing the
closely related meta-analytic combined (MAC) and meta-analytic predictive (MAP)
approaches that provide a useful background, but are in their basic form not robust
against non-exchangeability between the parameters of historical and new trials. Then
we discuss the robust MAP (rMAP) approach and propose extend the MAC approach
using model averaging with a shrinkage prior. The discussed methods are compared
in a simulation study in Section 6.1, where the employed models are fully specified
for the case of right-censored exponentially distributed failure times and a log-hazard
ratio φ assumed to be constant across trials. There are a number of further methods
that we do not discuss such as the "power prior" (Ibrahim et al., 2015) and the related
"commensurate power prior" (Hobbs et al., 2011) approaches.

2.5.4.1. Meta-analytic combined approach

The MAC approach is to analyze all historical and new data using a single hierarchical
model that assumes that all historical and new trials have parameters that arise from
the same random effects distribution. I.e. no distinction is made between the control
group data from historical and new trials. The term MAC approach aims to contrast
such an analysis against the closely related MAP approach discussed below. To aid the

comparison of different approaches, we spell out the posterior obtained from a MAC analysis given a — typically vague or weakly informative — prior $p(\boldsymbol{\zeta}, \boldsymbol{\vartheta})$ for $\boldsymbol{\zeta}$ and $\boldsymbol{\vartheta}$, which is given by

$$p(\boldsymbol{\zeta}, \boldsymbol{\vartheta}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{I+H} | \mathcal{D}, \mathcal{D}') \propto p(\boldsymbol{\zeta}, \boldsymbol{\vartheta}) \prod_{i=1}^{I+H} p(\boldsymbol{\theta}_i | \boldsymbol{\vartheta}) \prod_{j=0}^{1-h_i} p(\mathcal{D}_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\zeta}_j), \qquad (2.5.1)$$

where $h_i := \mathbb{1}\{i > I\}$ is the indicator for whether trial $i = 1, \ldots, I + H$ is a historical trial.

In the MAC approach we do not explicitly define priors on the distribution of control group parameters across trials based on the historical control group data. Instead, we induce them implicitly through random effects on trial-level parameters. The same effect could also be achieved by fitting model (2.5.1) using maximum likelihood.

2.5.4.2. Meta-analytic predictive approach

The MAP prior approach is a two-step version of the MAC approach. First, we obtain

$$p(\boldsymbol{\zeta}_0, \boldsymbol{\vartheta}, \boldsymbol{\theta}_{I+1}, \ldots, \boldsymbol{\theta}_{I+H} | \mathcal{D}') \propto p(\boldsymbol{\zeta}_0, \boldsymbol{\vartheta}) \prod_{i=I+1}^{I+H} p(\boldsymbol{\theta}_i | \boldsymbol{\vartheta}) p(\mathcal{D}'_{i0} | \boldsymbol{\theta}_i, \boldsymbol{\zeta}_0), \qquad (2.5.2)$$

the posterior from fitting a hierarchical meta-analysis model to the historical data $\mathcal{D}'$ (Schmidli et al., 2014). Secondly, the marginal posterior of the hyperparameters given $\mathcal{D}'$ is used as the hyperprior for a new analysis (Schmidli et al., 2014), i.e.

$$p(\boldsymbol{\zeta}, \boldsymbol{\vartheta}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I | \mathcal{D}, \mathcal{D}') \propto p(\boldsymbol{\zeta}_1) p(\boldsymbol{\zeta}_0, \boldsymbol{\vartheta} | \mathcal{D}') \prod_{i=1}^{I} p(\boldsymbol{\theta}_i | \boldsymbol{\vartheta}) \prod_{j=0}^{1} p(\mathcal{D}_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\zeta}_j). \qquad (2.5.3)$$

If the new data consists of a single new trial so that $I = 1$ and if $p(\boldsymbol{\zeta}, \boldsymbol{\vartheta}) = p(\boldsymbol{\zeta}_1) p(\boldsymbol{\vartheta})$ for some $\boldsymbol{\vartheta} \in \Theta'$, then (2.5.3) corresponds to (Schmidli et al., 2014)

$$p(\boldsymbol{\theta}_1, \boldsymbol{\zeta}_1 | \mathcal{D}, \mathcal{D}') \propto p(\boldsymbol{\zeta}_1) p(\boldsymbol{\theta}_1 | \mathcal{D}') \prod_{j=0}^{1} p(\mathcal{D}_{1j} | \boldsymbol{\theta}_1, \boldsymbol{\zeta}_j)$$

with

$$p(\boldsymbol{\theta}_1|\mathcal{D}') = \int_{\Theta'} p(\boldsymbol{\theta}_1|\boldsymbol{\vartheta}) \underbrace{\int_{\Theta} \dots \int_{\Theta} p(\boldsymbol{\vartheta}, \boldsymbol{\theta}_{I+1}, \dots, \boldsymbol{\theta}_{I+H}|\mathcal{D}') \, \mathrm{d}\boldsymbol{\theta}_{I+1} \dots \, \mathrm{d}\boldsymbol{\theta}_{I+H}}_{=p(\boldsymbol{\vartheta}|\mathcal{D}')} \, \mathrm{d}\boldsymbol{\vartheta}. \quad (2.5.4)$$

(2.5.4) is the marginal MAP distribution for $\boldsymbol{\theta}_1$ given $\mathcal{D}'$, which is why this method is called "MAP" approach. However, when there are $I \geq 2$ new trials the density of the joint MAP distribution of $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_I$ on the left-hand side of (2.5.5) is not equal to the product of the densities of the marginal MAP distributions on the right-hand side of (2.5.5)

$$\int_{\Theta'} p(\boldsymbol{\vartheta}|\mathcal{D}') \prod_{i=1}^{I} p(\boldsymbol{\theta}_i|\boldsymbol{\vartheta}) \, \mathrm{d}\boldsymbol{\vartheta} \neq \prod_{i=1}^{I} \int_{\Theta'} p(\boldsymbol{\vartheta}|\mathcal{D}') p(\boldsymbol{\theta}_i|\boldsymbol{\vartheta}) \, \mathrm{d}\boldsymbol{\vartheta}. \quad (2.5.5)$$

A stratified meta-analysis with marginal MAP distributions $p(\boldsymbol{\theta}_i|\mathcal{D}')$ as independent priors for the trial-level parameters $\boldsymbol{\theta}_i$ for $i = 1, \dots, I$ equates the two sides of (2.5.5) by ignoring the correlations between the parameters. Under what circumstances this naive extension of MAP approach for a single trial provides a good approximation to (2.5.3) has not been investigated.

Note that when a Bayesian analysis is performed using MCMC methods — as is typically the case — we do not need to explicitly perform the integrations in (2.5.4) or those to obtain $p(\boldsymbol{\zeta}, \boldsymbol{\vartheta}|\mathcal{D}')$ from (2.5.2). Instead, we can directly use the values of the pseudo-random samples $\boldsymbol{\vartheta}^{(l)}$ and $\boldsymbol{\zeta}^{(l)}$ with $l = 1, \dots, L$ from the joint posterior $p(\boldsymbol{\zeta}, \boldsymbol{\vartheta}, \boldsymbol{\theta}_{I+1}, \dots, \boldsymbol{\theta}_{I+H}|\mathcal{D}')$ and obtain predictions for $\boldsymbol{\theta}_i$ for $i = 1, \dots, I$ according to $p(\boldsymbol{\theta}_i^{(l)}|\boldsymbol{\vartheta}^{(l)})$ (Gelman et al., 2014, pp. 63–67).

The use of standard MCMC methods requires a specification of the MAP prior in terms of a mathematical function based on the MCMC samples obtained from fitting the hierarchical model to the historical data. One approach is to use a mixture distribution that closely approximates the posterior samples for the hyperparameters (Schmidli et al., 2014). The use of a finite mixture of (conditionally) conjugate priors is particularly convenient in order to obtain — when possible — analytical posteriors or to enable Gibbs sampling (Schmidli et al., 2014). An arbitrarily close approximation to the MAP prior in terms of e.g. pointwise convergence in the L_1-norm, the Kullback-Leibler divergence

or the Prohorov distance can be achieved by using a sufficient number of mixture components (Lijoi, 2003; Schmidli et al., 2014; Robert, 2007, pp. 123–126). Schmidli et al. (2014) have suggested that in practice two or three mixture components will typically be sufficient, if the MAP prior is unimodal. However, a good approximation of the prior does not ensure that the posterior based on the approximate prior will closely approximate the posterior based on the original MAP prior (Robert, 2007, pp. 126–127). The arbitrariness of the choice of distribution for the random effects, as well as of the priors for its parameters, and the approach described in the next section of making the MAP prior more vague through the addition of a weakly informative mixture component may in practice somewhat alleviate concerns around this point.

2.5.4.3. The robust Meta-analytic predictive approach

To make the MAP approach more robust against non-exchangeability between the parameters of historical and new trials Schmidli et al. (2014) have proposed the rMAP method. It makes the tails of the mixture approximation to a MAP distribution heavier by adding vague or weakly informative mixture components that assign weight to a very wide range of potential parameter values. Such a rMAP prior increasingly discards the prior information with increasing prior-data conflict (Schmidli et al., 2014).

However, little research exists on the choice of the weight for the vague or weakly informative mixture component in the rMAP approach. Prior weights of 0.5 to 0.9 for the informative mixture component have been used (Schmidli et al., 2014). Weaver et al. (2016) suggest a weight of 0.5 for binomially distributed outcomes. They also suggested that assigning a sufficiently large weight to the vague mixture component becomes more important the more informative the likelihood is. Note that the weight of the informative mixture component is not simply a factor by which the prior effective sample size is down-weighted. In fact, if the likelihood strongly supports the MAP prior, the informative mixture component may have close to 100% of the posterior weight and the prior effective sample size is close to that of the MAP prior. This is in contrast to the power prior approach of Ibrahim et al. (2015), in which the discounting parameter is directly related to the prior effective sample size.

2.5.4.4. An example of using robust mixture priors

We will use the example of estimating a single exponential rate parameter when there is considerable prior information. Program code for this example is available under https://doi.org/10.6084/m9.figshare.5007833. We assume that there is uncertainty as to whether an exchangeability assumption between the new data under analysis and the source of prior information is truly appropriate.

As the top-left panel of Figure 2.1 illustrates, an informative single-component conjugate prior can lead to the posterior having minimal overlap with both the high-density regions of the likelihood and the prior. As the other panels of the figure illustrate, this is mitigated by using a supposedly vague conjugate prior such as a $\text{Gamma}(\epsilon, \epsilon)$ prior with e.g. $\epsilon = 0.001$, a weakly informative conjugate prior such as a $\text{Gamma}(1, 2.84)$ prior, or a mixture of a weakly informative and an informative conjugate prior.

The posterior weight of the weakly informative component of the mixture is greater than 99.99% so that the informative mixture component has extremely little effect on the posterior despite having a prior weight of 80%. Note that the scale parameter for the $\text{Gamma}(1, 2.84)$ prior was chosen so that the prior mean on the log-scale
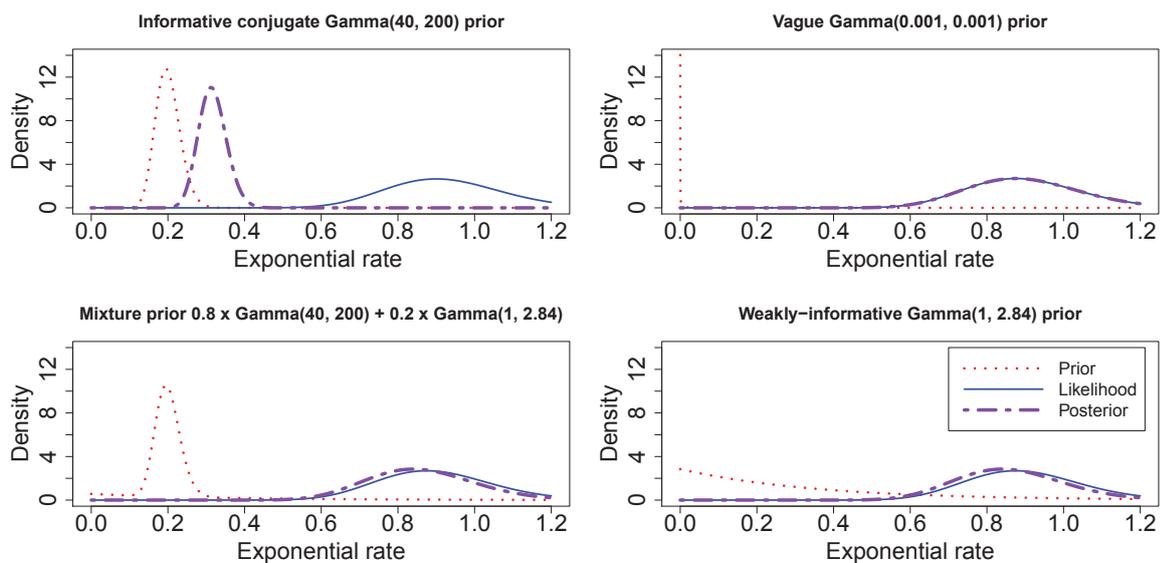
Figure 2.1.: Conjugate analysis of an exponential rate when data (36 patients with an event in 40 patient-years of follow-up) is in conflict with the informative Gamma(40, 200) prior density
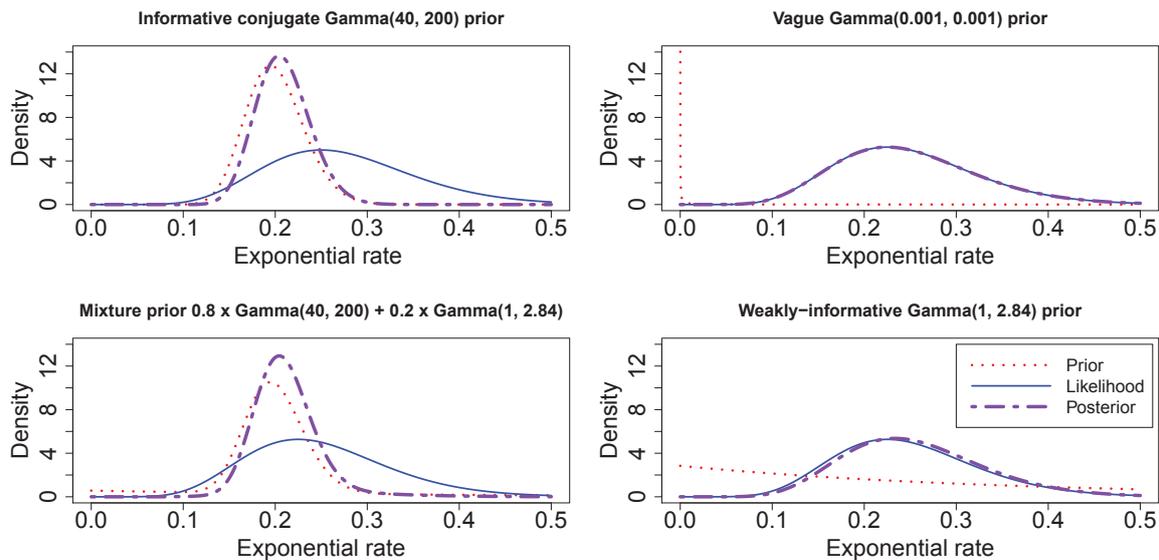
Figure 2.2.: Conjugate analysis of an exponential rate when data (10 patients with an event in 40 patient-years of follow-up) is reasonably consistent with the informative Gamma(40, 200) prior density

matches that of the informative Gamma(40, 200) prior, i.e. so that $\psi(1) - \log(\text{scale}) = \psi(40) - \log(200) \approx 2.84$, where $\psi(x) := \Gamma'(x)/\Gamma(x)$ denotes the digamma function.

While a Gamma(ϵ, ϵ) prior is often considered a default vague prior (Coory et al., 2009; Neil et al., 2008), the choice of ϵ can exert strong influence on posterior inference. E.g. when no events were observed in 1 day of observation time, the upper end of an equal-tailed posterior credible interval (CI) for the hazard rate per patient year is 1.5×10^{-9} for $\epsilon = 0.001$ and 3.7 for $\epsilon = 0.01$.

On the other hand, as Figure 2.2 illustrates, an informative single-component conjugate prior, as well as a mixture of a weakly informative and an informative conjugate prior result in a notably more concentrated posterior density than weakly informative or non-informative priors when the likelihood and the prior have a substantial overlap.

This is the result of prior information consistent with the observed data substantially influencing the posterior. In the analysis using a mixture prior the informative mixture component has a higher posterior (91.7%) than prior weight (80%), because it overlaps more with the likelihood than the more spread-out weakly informative mixture component.

2.5.4.5. Model averaging using a shrinkage prior

Instead of attempting to make the MAP approach more robust against prior-data conflicts, one can attempt to do the same within the MAC framework. This has the advantage of requiring only a single analysis instead of two analyses with an in-between step for approximating the MAP prior by a mixture distribution as in the rMAP approach. We propose to extend the MAC approach for hierarchical models with normally distributed random effects by introducing an additional parameter $\boldsymbol{\beta}$ that describes the extent to which historical and new data differ to obtain

$$p(\boldsymbol{\nu}, \boldsymbol{\Sigma}, \boldsymbol{\zeta}, \boldsymbol{\beta}, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_{I+H} | \mathcal{D}, \mathcal{D}') \propto p(\boldsymbol{\nu}, \boldsymbol{\Sigma}) p(\boldsymbol{\zeta}) p(\boldsymbol{\beta})$$
$$\prod_{i=1}^{I+H} \phi(\boldsymbol{\theta}_i; \boldsymbol{\nu} + h_i \boldsymbol{\beta}, \boldsymbol{\Sigma}) \prod_{j=0}^{1-h_i} p(\mathcal{D}_{ij} | \boldsymbol{\theta}_i, \boldsymbol{\zeta}).$$

In terms of the notation of Section 2.4, $\boldsymbol{\vartheta}$ consists of $\boldsymbol{\nu}$, $\boldsymbol{\Sigma}$ and $\boldsymbol{\beta}$. Specifying the prior $p(\boldsymbol{\beta})$ may be seen as an a-priori assessment of how much the parameters of the new data may differ from the historical data.

Alternatively, Carvalho et al. (2009) have motivated the choice of a horseshoe $\mathrm{HS}(\sigma_a, \sigma_b)$ prior by aiming for a behavior similar to Bayesian model averaging. With a local scale parameter $\sigma_a > 0$ and global scale parameter $\sigma_b > 0$ the horseshoe prior for a vector of regression coefficients $\boldsymbol{\beta}$ is given by

$$p_{\mathrm{HS}}(\boldsymbol{\beta}; \sigma_a, \sigma_b) := \int_0^\infty \ldots \int_0^\infty p_{\mathrm{HCa}}(b; 0, \sigma_b) \prod_{r=1}^p \phi(\beta_r; 0, a_r b) p_{\mathrm{HCa}}(a_r; 0, \sigma_a) \, \mathrm{d}a_1 \ldots \mathrm{d}a_p \, \mathrm{d}b,$$

where $p_{\mathrm{HCa}}(\cdot; 0, \sigma)$ is the pdf of a half-Cauchy distribution with location parameter 0 and scale parameter σ.

The candidate models considered in the model averaging are the simple MAC model (2.5.1) (i.e. $\boldsymbol{\beta} = \boldsymbol{0}$), a model with some hierarchical mean parameters differing for new and historical data (when some components of $\boldsymbol{\beta}$ are non-zero), and a model with an unconstrained $\boldsymbol{\beta}$ — i.e. a model with two completely separate means for the random effects in the historical and new data. The horseshoe prior approximates a "spike-and-slab" prior by having considerable weight in the vicinity of $\boldsymbol{\beta} = \boldsymbol{0}$, where its density is unbounded,

and otherwise having a relatively flat heavy-tailed density centered on zero.

For a single scalar β, the prior simplifies to $\int_0^\infty \int_0^\infty \phi(\beta; 0, ab) p_{\mathrm{HCa}}(a; 0, \sigma_a) p_{\mathrm{HCa}}(b; 0, \sigma_b) \, \mathrm{d}a \, \mathrm{d}b$. Alternative proposals for $p(\beta)$ include the Laplace (double-exponential) and the Cauchy priors (Carvalho et al., 2009). All of these priors are illustrated in Figure 6.1.

The alternative of a full Bayesian model averaging has been investigated by Neuenschwander et al. (2016a) and Neuenschwander et al. (2016b), but efficient MCMC sampling for this approach can be challenging (Carvalho et al., 2009; Neuenschwander et al., 2016b).

3. Time-to-event meta-analysis of commonly available aggregate data

We start this chapter by illustrating the need for AD time-to-event meta-analysis methods that do not rely on the availability of $t_{ij\cdot}$ or $d_{ij\cdot}$. First, we show that meta-analysis methods for binomial data require quite restrictive assumptions on the drop-out distribution. Secondly, we discuss options for imputing $t_{ij\cdot}$ and $d_{ij\cdot}$, which are needed for AD approaches such as using a "Poisson likelihood" or logistic regression with a log duration to first event or censoring offset, or the binomial model with a complementary-log-log link function with a log average follow-up offset. Methods for imputing these quantities have not so far received attention in the literature.

The difficulties we face when attempting to do so illustrate the need for the kind of likelihood, which we derive in Section 3.2. We discuss issues with parameter identifiability and maximum likelihood estimation for the proposed likelihood in Section 3.3, as well as approaches for avoiding these issues. One such approach is to use a Bayesian hierarchical model as proposed in Section 3.4.

We then illustrate a number of the discussed approaches using the example of a meta-analysis of the occurrence of major adverse cardiovascular events (MACE) with the diabetes drug rosiglitazone in Section 3.5. Finally, we evaluate a range of methods in a simulation study in Section 3.6.

3.1. Motivation for aggregate data time-to-event meta-analysis methods

3.1.1. Validity of meta-analysis methods for binomial data

We consider a meta-analysis of I trials with i.i.d. event times X_{ijk} with absolutely continuous pdf f_{ij} and hazard function $\lambda_{ij}(s)$ for $s \geq 0$ for the n_{ij} patients $k = 1, \ldots, n_{ij}$ in group $j = 0, 1$ of trial $i = 1, \ldots, I$ with planned follow-up $\tau_i > 0$ for each patient. We assume i.i.d. drop-out times C_{ijk} with an absolutely continuous pdf g_{ij}, and that the X_{ijk} and C_{ijk} are independent of each other. As before, we define $Y_{ij \cdot} := \sum_{k=1}^{n_{ij}} \mathbb{1}\{X_{ijk} \leq \min(C_{ijk}, \tau_i)\}$ and $\pi_{ij} := \mathbb{E}Y_{ij \cdot}/n_{ij}$.

Theorem 3.1. *If for the drop-out times $g_{i0}(s) = g_{i1}(s) \; \forall s \in [0, \tau_i]$ and $\forall i = 1, \ldots, I$, then we can test the null hypothesis*

$$H_0 : \lambda_{i1}(s) \leq \lambda_{i0}(s) \; \forall i = 1, \ldots, I \; and \; \forall s \in [0, \tau_i]$$

at level $\alpha \in (0, 1)$ with any level α test of the "substitute" null hypothesis $H_0' : \pi_{i1} \leq \pi_{i0}$ $\forall i = 1, \ldots, I$.

Proof. We will demonstrate that H_0 implies H_0' under the conditions given in Theorem 3.1, which proves the statement of the theorem. Since we assume i.i.d. event and drop-out times within a treatment group $j = 0, 1$ of a trial $i = 1, \ldots, I$, all patients within the same group in a trial have the same probability π_{ij} of having an observed event prior to dropping out or reaching the planned end of follow-up at time τ_i and the number of patients with an observed event $Y_{ij \cdot}$ follows a binomial distribution. π_{ij} is given by

$$\begin{aligned}
\pi_{ij} &:= \mathbb{P}\{X_{ijk} \leq \min(C_{ijk}, \tau_i)\} \\
&= \int_0^\infty \int_0^\infty f_{ij}(x) g_{ij}(c) \mathbb{1}\{x \leq \min(c, \tau_i)\} \, \mathrm{d}x \, \mathrm{d}c \\
&= \int_0^\infty \int_0^{\min(c, \tau_i)} f_{ij}(x) g_{ij}(c) \, \mathrm{d}x \, \mathrm{d}c \\
&= \int_0^\infty F_{ij}(\min(c, \tau_i)) g_{ij}(c) \, \mathrm{d}c.
\end{aligned}$$

Since $F_{ij}(s) = 1 - \exp\left(-\int_0^s \lambda_{ij}(u)\,\mathrm{d}u\right)$, $\lambda_{i1}(u) \leq \lambda_{i0}(u)$ for all $u \in [0, s]$ implies $F_{i1}(s) \leq F_{i0}(s)$. Thus, if $\lambda_{i1}(s) \leq \lambda_{i0}(s)$ and $g_{i0}(s) = g_{i1}(s)\ \forall s \in [0, \tau_i]$, then $\pi_{i1} \leq \pi_{i0}$. $\qquad\square$

Remark 3.1. Under the conditions of Theorem 3.1 an unbiased level $0 < \alpha < 1$ test of the alternative hypothesis $H'_A : \pi_{i1} \geq \pi_{i0}$ with strict inequality for some $i = 1, \ldots, I$ against the null hypothesis H'_0 of Theorem 3.1 is an unbiased level α test for the alternative hypothesis

$$H_A : \lambda_{i1}(s) \geq \lambda_{i0}(s)\ \forall i = 1, \ldots, I \text{ and } \forall s \in [0, \tau_i] \text{ with strict inequality for some}$$
$$i = 1, \ldots, I \text{ and for some } s \in [0, \min(\tau_i, G_{i0}^{-1}(1))]$$

against the null hypothesis H_0 of Theorem 3.1. Note that H_A is not the complement of H_0 without additional assumptions such as a proportional hazards assumption. That the power function of such a test is at least α under H_A can be shown in a similar manner to the proof of Theorem 3.1, because $\lambda_{i1}(u) \geq \lambda_{i0}(u)$ for all $u \in [0, s]$ implies $F_{i1}(s) \geq F_{i0}(s)$. Strict inequality occurs, if $\lambda_{i1}(u) > \lambda_{i0}(u)$ for some $u \in [0, s]$, because in this case the hazard functions differ on a subinterval of non-zero length of $[0, s]$ due to their absolute continuity implied by the absolute continuity of the f_{ij}.

While Theorem 3.1 and Remark 3.1 may be obvious, what seems less widely appreciated is that the given sufficient conditions cannot easily be relaxed. Firstly, having the same expected proportion of drop-outs across treatment groups in each trial would not be a sufficient condition. Secondly, an identical expected follow-up is not a sufficient condition. Both of these can be illustrated with two counter-examples illustrated in Figure 3.1. Program code for these counter examples is available under `https://doi.org/10.6084/m9.figshare.5007833`.

Counter-example 3.1. Let $\tau_i = 1$, $g_{i1}(s) := 12(s - 0.5)^2$ for $s \in [0, 1]$, $g_{i0}(s) := 12/5(0.5 - (s - 0.5)^2)$ for $s \in [0, 1]$ and let $f_{ij}(s) := 2s$ for $s \in [0, 1]$. Then the expected proportion of drop-outs is 1 and the expected duration of follow-up is $1/2$ for both treatment groups, but $\pi_{i0} = 0.32$ and $\pi_{i1} = 0.4$ even though the hazard function for events is the same for both treatment groups throughout the trial period.

Counter-example 3.2. Let $\tau_i = 1$, $g_{i0}(s) := 12(s - 0.5)^2$ for $s \in [0, 1]$, $g_{i1}(s) := 12/5(0.5 - (s - 0.5)^2)$ for $s \in [0, 1]$, $f_{i0}(s) := 3s^2$ for $s \in [0, 1]$ and $f_{i1}(s) := 2s$ for
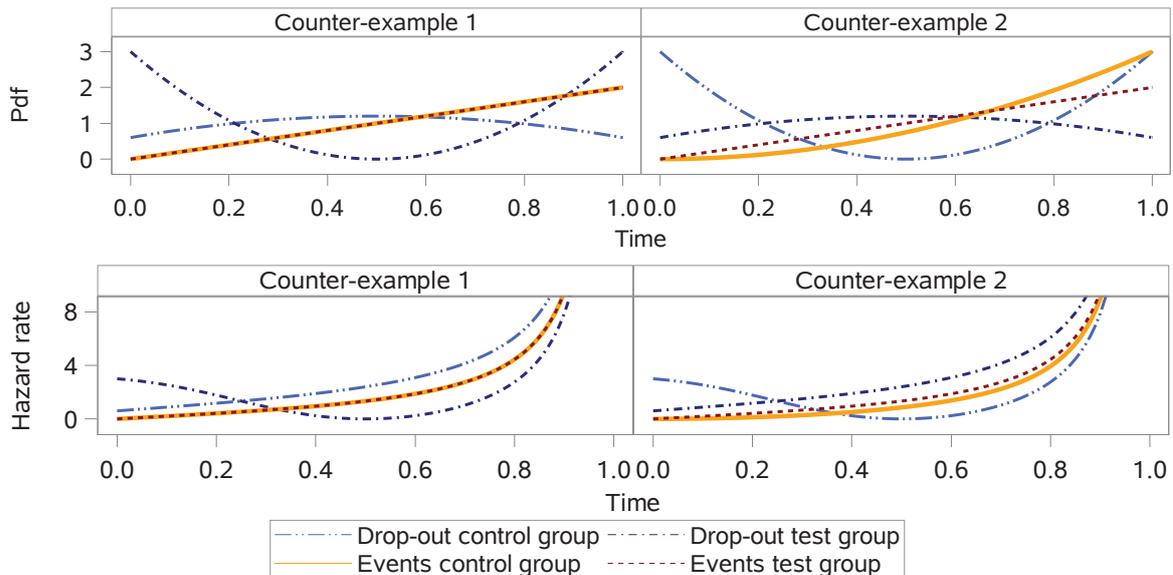
Figure 3.1.: Probability density functions and hazard functions for the two counter-examples

$s \in [0, 1]$. As before the expected proportion of drop-outs is 1 and the expected duration of follow-up is $1/2$ for both treatment groups, but $\pi_{i0} = 0.35$ is larger than $\pi_{i1} = 0.32$ even though the hazard function for the control group $(3x^2)/(1 - x^3)$ is smaller throughout the trial duration than that of the test group $2x/(1 - x^2)$.

Remark 3.2. Buyse and Ryan (1987) have shown that for a meta-analysis of trials of a fixed planned duration for each patients with identical drop-out distributions across groups, with at most 10% drop-outs, with at most 75% of patients experiencing an event and with a Weibull shape parameter ≤ 1, the asymptotic relative efficiency of the CMH test is at least 80% of that of a stratified log-rank test.

3.1.2. Imputing the duration of follow-up, or of follow-up to first event or censoring

3.1.2.1. Imputing the duration of follow-up

The $z_{ij.}$ patients that do not complete a trial have interval censored drop-out times that lie in the interval $[0, \tau_i]$, if we treat deaths due to events of interest like any other drop-out. Similarly, the drop-out times for patients that complete a trial are right-censored to lie in (τ_i, ∞). If we assume that drop-out times follow a $\text{Exp}(\mu_{ij})$

distribution and treat all deaths like any other trial discontinuation, then the MLE based on the log-likelihood

$$\log \mathcal{L}(\mu_{ij}|z_{ij\cdot}) = z_{ij\cdot} \log G(\tau_i; \mu_{ij}) + (n_{ij} - z_{ij\cdot}) \log(1 - G(\tau_i; \mu_{ij}))$$
$$= z_{ij\cdot} \log(1 - e^{-\mu_{ij}\tau_i}) - (n_{ij} - z_{ij\cdot})\mu_{ij}\tau_i + \text{constant}$$

for an exponential drop-out hazard rate μ_{ij} in each treatment arm j of each trial i is

$$\hat{\mu}_{ij} = \tau_i^{-1}\{\log n_{ij} - \log(n_{ij} - z_{ij\cdot})\}, \tag{3.1.1}$$

as long as $z_{ij\cdot} < n_{ij}$, while for $z_{ij\cdot} = n_{ij}$ no finite MLE exists. For $z_{ij\cdot} = 0$ we obtain $\hat{\mu}_{ij} = 0$, but already know that all patients were followed for the planned trial duration so that $d_{ij\cdot} = n_{ij}\tau_i$. Otherwise, we can calculate the expected follow-up duration conditional on $z_{ij\cdot}$ and substituting $\hat{\mu}_{ij} \in (0, \infty)$ for μ_{ij} as

$$\mathbb{E}_{\mu_{ij}=\hat{\mu}_{ij}}(D_{ij\cdot}|Z_{ij\cdot} = z_{ij\cdot}) = (n_{ij} - z_{ij\cdot})\mathbb{E}_{\mu_{ij}=\hat{\mu}_{ij}}(D_{ijk}|D_{ijk} \geq \tau_i)$$
$$+ z_{ij\cdot}\mathbb{E}_{\mu_{ij}=\hat{\mu}_{ij}}(D_{ijk}|D_{ijk} < \tau_i)$$
$$= (n_{ij} - z_{ij\cdot})\tau_i + z_{ij\cdot}\frac{\int_0^{\tau_i} c\hat{\mu}_{ij}e^{-\hat{\mu}_{ij}c}\,\mathrm{d}c}{\int_0^{\tau_i} \hat{\mu}_{ij}e^{-\hat{\mu}_{ij}c}\,\mathrm{d}c}$$
$$= (n_{ij} - z_{ij\cdot})\tau_i + z_{ij\cdot}(\frac{1}{\hat{\mu}_{ij}} + \frac{\tau_i}{1 - e^{\hat{\mu}_{ij}\tau_i}}) \tag{3.1.2}$$
$$= (n_{ij} - z_{ij\cdot})\tau_i - z_{ij\cdot}\tau_i(\frac{n_{ij} - z_{ij\cdot}}{z_{ij\cdot}} + \frac{1}{\log(\frac{n_{ij}-z_{ij\cdot}}{n_{ij}})})$$
$$= (n_{ij} - z_{ij\cdot})\tau_i + z_{ij\cdot}\frac{\tau_i}{2} + \mathcal{O}(\frac{z_{ij\cdot}}{n_{ij}}) \quad \text{as } \frac{z_{ij\cdot}}{n_{ij}} \searrow 0.$$

The Taylor series expansion in the last line of (3.1.2) is for $\tau_i n_{ij} u(1/u - 1 + \frac{1}{\log(1-u)})$ around $u = 0$ for $u = \frac{z_{ij\cdot}}{n_{ij}}$. It suggests that we expect the $z_{ij\cdot}$ patients that did not complete the trial to have approximately half the planned follow-up, if the proportion of drop-outs is low. In case $z_{ij\cdot} = 0$ evidently no approximation is involved, while for $z_{ij\cdot} = n_{ij}$ the approximation is likely poor.

At the end of Section 3.1.2.2 we discuss the limitations and issues of approximating $t_{ij\cdot}$ in a similar fashion and all of the points outlined there also apply to approximation (3.1.2) for $d_{ij\cdot}$.

3.1.2.2. Imputing the duration of follow-up to first event or censoring

If we assume event and censoring times to be independent and exponentially distributed with exponential hazard rates λ_{ij} and μ_{ij}, then the time to first event or drop-out is also exponentially distributed with hazard rate parameter $\xi_{ij} = \lambda_{ij} + \mu_{ij}$. We can estimate ξ_{ij} in analogy to equation (3.1.1) as

$$\hat{\xi}_{ij} = \tau_i^{-1}\{\log n_{ij} - \log(n_{ij} - y_{ij.} - (z_{ij.} - m_{ij.}) + \sum_{k=1}^{n_{ij}} y_{ijk}(z_{ijk} - m_{ij.}))\},$$

if we knew the overlap $\sum_{k=1}^{n_{ij}} y_{ijk}(z_{ijk} - m_{ij.})$ between patients that had non-fatal events and the drop-outs. Analogously to equation (3.1.2) we then obtain the conditional expectation

$$\tilde{t}_{ij.} := \mathbb{E}_{\xi_{ij}=\hat{\xi}_{ij}}(T_{ij.}|Z_{ij.} = z_{ij.} \text{ and } Y_{ij.} = y_{ij.} \text{ and } M_{ij.} = m_{ij.})$$
$$\approx \tau_i(n_{ij} - z_{ij.} + m_{ij.} - y_{ij.}) + \frac{\tau_i}{2}(z_{ij.} - m_{ij.} + y_{ij.}),$$

if both the proportion of censored patients and of patients with an event are low.

On average $t_{ij.}$ will be overestimated by $\tilde{t}_{ij.}$, because without any approximations the expected average event and drop-out times are expected to be below $\tau_i/2$. Thus, if there is differential follow-up across groups, we expect the approximations to bias the hazard ratio between two treatment groups in favor — assuming that event occurrence is an undesirable outcome — of the group with the higher drop-out rate, because using $\tilde{t}_{ij.}$ in the MLE for an exponential hazard rate $\hat{\lambda}_{ij} = y_{ij.}/t_{ij.}$ is expected to underestimate λ_{ij} more in that group. This expected direction of the bias is seen in the simulation results presented in Appendix C. Additionally, such a single value imputation for $t_{ij.}$ ignores the uncertainty about $\tilde{t}_{ij.}$ in two ways: firstly by taking a conditional expectation given the MLE of the hazard rate of events and drop-outs as a single value imputation, and secondly by ignoring uncertainty around this estimated hazard rate. Using the conditional expectation given $\xi_{ij} = \hat{\xi}_{ij}$ as a single value imputation for $t_{ij.}$ ignores the variation in $T_{ij.}$ for a given true ξ_{ij}, as well as the uncertainty around $\hat{\xi}_{ij}$. This will not only result in understated uncertainty, but also in potentially biased inference about

event occurrence due to the non-linear link-function through which $t_{ij\cdot}$ enters analyses such as an exponential AD time-to-event model.
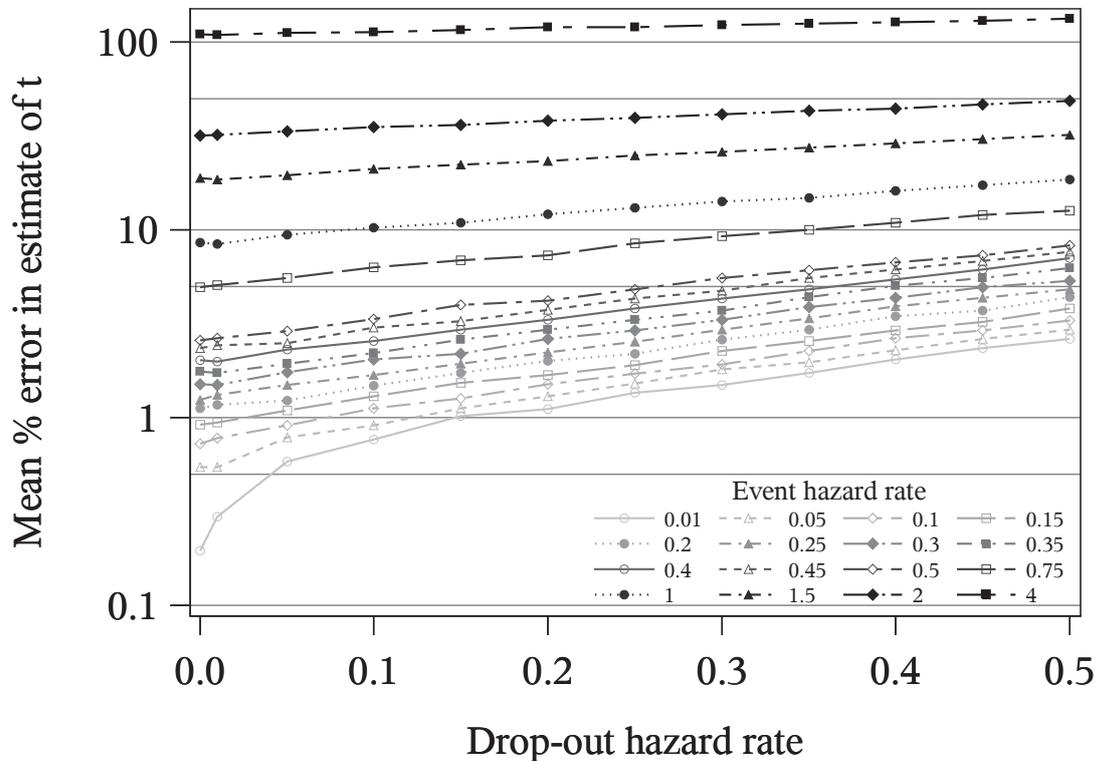
Figure 3.2.: Average percentage error $100 \times |\tilde{t}_{ij\cdot}/t_{ij\cdot} - 1|$ for different exponential event and drop-out time hazard rates based on 10,000 simulations for each scenario

As Figure 3.2 shows the average absolute percentage error when using $\tilde{t}_{ij\cdot}$ is limited to below 5% for many practically relevant event rates (≤ 0.5) for realistic censoring rates below 0.2. On the other hand, for higher event hazard rates the error can be substantial. Program code for this example is available under https://doi.org/10.6084/m9.figshare.5007833.

3.2. Likelihood for commonly available aggregate data

In this section we derive a generally applicable likelihood for $\boldsymbol{\theta}_{F,ij}$ and $\boldsymbol{\theta}_{G,ij}$, as well as q_{ij} — the probability of events being fatal given that they have occurred — given observed AD consisting of $y_{ij\cdot}$, $z_{ij\cdot}$, $m_{ij\cdot}$, n_{ij} and τ_i.

For deriving this likelihood, we will consider five mutually exclusive outcomes a patient may experience and relate these to the observed AD.

Definition 3.1. Let $\boldsymbol{W_{ij}} := (W_{ij1}, \ldots, W_{ij5})^T$ denote the random vector of the number of patients in each of the following five mutually exclusive categories that a patient in group j of trial i may fall into:

1. Patient experienced a fatal event of interest,

2. Patient experienced a non-fatal event and completed the trial,

3. Patient experienced a non-fatal event and then dropped out before completing the trial,

4. Patient completed the study without an event (administratively censored), which in fixed duration studies occurs after the planned trial duration τ_i,

5. Patient dropped out before experiencing any event or completing the trial.

The random vector $\boldsymbol{W_{ij}}$ has a multinomial distribution with pmf

$$f(\boldsymbol{w_{ij}}; p_{ij1}, \ldots, p_{ij5}) = \binom{n_{ij}}{\boldsymbol{w_{ij}}} \prod_{r=1}^{5} p_{ijr}^{w_{ijr}} \qquad (3.2.1)$$

for some probabilities $p_{ij1}, \ldots, p_{ij5}$ that satisfy $\sum_{r=1}^{5} p_{ijr} = 1$, where

$$\binom{n_{ij}}{\boldsymbol{w_{ij}}} := n_{ij}! / \prod_{r=1}^{5} w_{ijr}!$$

denotes the multinomial coefficient.

We treat patients that die for reasons other than the event of interest as censored for the purpose of estimating $\boldsymbol{\theta}_{F,ij}$ and as an event for estimating $\boldsymbol{\theta}_{G,ij}$.

We take into account whether events were fatal in order to correctly reflect that a fatal event truncates further follow-up for drop-out from the trial, while patients continue to be at risk for that outcome after a non-fatal event. We do not simply treat the truncation of follow-up by a fatal event as drop-outs, because then event and drop-out times would no longer be independent.

The probabilities

$$p_{ij1} := \mathbb{P}(X_{ijk} \leq \min(\tau_i, C_{ijk}) \text{ and } M_{ijk} = 1) = q_{ij} \times \mathbb{P}(X_{ijk} \leq \min(\tau_i, C_{ijk})),$$

$$p_{ij2} := \mathbb{P}(X_{ijk} \leq \tau_i < C_{ijk} \text{ and } M_{ijk} = 0) = (1 - q_{ij}) \times \mathbb{P}(X_{ijk} \leq \tau_i < C_{ijk}),$$

$$p_{ij3} := \mathbb{P}(X_{ijk} \leq C_{ijk} \leq \tau_i \text{ and } M_{ijk} = 0) = (1 - q_{ij}) \times \mathbb{P}(X_{ijk} \leq C_{ijk} \leq \tau_i), \quad (3.2.2)$$

$$p_{ij4} := \mathbb{P}(\min(X_{ijk}, C_{ijk}) > \tau_i) \text{ and}$$

$$p_{ij5} := \mathbb{P}(C_{ijk} < X_{ijk} \text{ and } C_{ijk} \leq \tau_i)$$

for group j of trial i can be expressed in terms of τ_i and the parameter vector $\boldsymbol{\theta}_{ij}$ consisting of $\boldsymbol{\theta}_{F,ij}$, $\boldsymbol{\theta}_{G,ij}$, and q_{ij}, if we assume specific parametric distributions with cdfs F for the failure times X_{ijk} and G for the drop-out times C_{ijk} for patient $k = 1, \ldots, n_{ij}$ in group j of trial i.

The second equalities for p_{ij1}, p_{ij2} and p_{ij3} only follow, if we assume that $\mathbb{P}(M_{ijk} = 1 | X_{ijk} \leq \min(\tau_i, C_{ijk})) = q_{ij}$. This may be a reasonable assumption regarding whether the first event itself is fatal, but is a simplifying approximation when the first event is not fatal. In the latter case, the probability of the patient dying of an event of interest depends on the hazard over time of dying of an event of interest following a first non-fatal event, the hazard over time for drop-out prior to dying and on the remaining duration of follow-up until censoring after the first non-fatal event.

By considering follow-up as truncated at the time of the first event, when in fact the fatal event may only have occurred subsequently, we underestimate the time at risk for drop-out. In the absence of the data on recurrent events to fully model this, we consider this approximation reasonable particularly if either first, recurrent or fatal events are rare. In those cases any approximations for q_{ij} will have little influence on inference, because q_{ij} is primarily a nuisance parameter that we account for in order to estimate $\boldsymbol{\theta}_{G,ij}$. Similarly, less of an approximation is involved if recurrent events are rare, because q_{ij} will be primarily driven by the probability of the initial event being fatal and to a much smaller extent by subsequent events being fatal.

If we can specify a joint distribution of event and drop-out times, as well as event fatality, (3.2.2) can also deal with non-independent event and drop-out times, but in

what follows we will assume independence between drop-out times, event times and whether events are fatal.

For exponentially distributed failure times $X_{ijk} \sim \text{Exp}(\lambda_{ij})$ and exponentially distributed drop-out times $C_{ijk} \sim \text{Exp}(\mu_{ij})$ for patients $k = 1, \ldots, n_{ij}$, we have $\boldsymbol{\theta}_{ij} = (\lambda_{ij}, \mu_{ij}, q_{ij})^T$ and obtain

$$
\begin{aligned}
p_{ij1} &= q_{ij} \times \frac{\lambda_{ij}}{\lambda_{ij} + \mu_{ij}} \left(1 - e^{-(\lambda_{ij} + \mu_{ij})\tau_i}\right), \\
p_{ij2} &= (1 - q_{ij}) \times \left(1 - e^{-\lambda_{ij}\tau_i}\right) e^{-\mu_{ij}\tau_i}, \\
p_{ij3} &= (1 - q_{ij}) \times \left[\frac{\lambda_{ij} + \mu_{ij} e^{-(\lambda_{ij}+\mu_{ij})\tau_i}}{\lambda_{ij} + \mu_{ij}} - e^{-\mu_{ij}\tau_i}\right], \\
p_{ij4} &= e^{-(\lambda_{ij}+\mu_{ij})\tau_i}, \text{ and} \\
p_{ij5} &= \frac{\mu_{ij}}{\lambda_{ij} + \mu_{ij}} \left(1 - e^{-(\lambda_{ij}+\mu_{ij})\tau_i}\right).
\end{aligned} \tag{3.2.3}
$$

Note that we can choose to replace the calculation of one of these probabilities by a use of the identity $\sum_{r=1}^{5} p_{ijr} = 1$.

If we would observe realizations $\boldsymbol{w_{ij}}$ of $\boldsymbol{W_{ij}}$ — i.e. if we knew how many patients fall into each of the five mutually exclusive categories in Definition 3.1 — we could substitute the expressions of the probabilities $p_{ij1}, \ldots, p_{ij5}$ in terms of $\boldsymbol{\theta}_{ij}$ into (3.2.1) to conduct inference about the parameters of interest. However, in practice publications will typically only report $w_{ij1} = m_{ij\cdot}$, while $w_{ij2}, \ldots, w_{ij5}$ will not be directly known.

Given $y_{ij\cdot}$, $z_{ij\cdot}$, $m_{ij\cdot}$, n_{ij} and τ_i, we face a coarse data situation, in which we do not know the overlap w_{ij3} between the $y_{ij\cdot} - m_{ij\cdot}$ patients with a non-fatal event and those $z_{ij\cdot}$ patients that did not complete the trial. The data on w_{ij3} is interval censored to lie between $r_1 = \max(0, y_{ij\cdot} + z_{ij\cdot} - m_{ij\cdot} - n_{ij})$ and $r_2 = \min(y_{ij\cdot} - m_{ij\cdot}, z_{ij\cdot} - m_{ij\cdot})$. w_{ij2}, w_{ij4}, and w_{ij5} are also interval censored, but are determined for a given value of w_{ij3}. Thus, the likelihood is obtained as

$$
\begin{aligned}
\mathcal{L}(p_{ij1}, \ldots, p_{ij5} | \text{AD}) = {} &n_{ij}! \frac{p_{ij1}{}^{m_{ij\cdot}}}{m_{ij\cdot}!} \sum_{r=r_1}^{r_2} \frac{p_{ij2}{}^{y_{ij\cdot} - m_{ij\cdot} - r}}{(y_{ij\cdot} - m_{ij\cdot} - r)!} \frac{p_{ij3}{}^{r}}{r!} \\
&\frac{p_{ij4}{}^{n_{ij} - y_{ij\cdot} - z_{ij\cdot} + m_{ij\cdot} + r}}{(n_{ij} - y_{ij\cdot} - z_{ij\cdot} + m_{ij\cdot} + r)!} \frac{p_{ij5}{}^{z_{ij\cdot} - m_{ij\cdot} - r}}{(z_{ij\cdot} - m_{ij\cdot} - r)!}
\end{aligned} \tag{3.2.4}
$$

by summing over the possible data constellations consistent with the observed AD. (3.2.4) can be used as the likelihood for the parameters of interest $\boldsymbol{\theta}_{ij}$ in a straightforward manner, if closed form expressions for $p_{ij1}, \ldots, p_{ij5}$ in terms of $\boldsymbol{\theta}_{ij}$ exist. When this is not the case, numerical integration may be necessary to evaluate the likelihood as is the case for Weibull distributed event and drop-out times (see Section 4.1).

If we assume that $\boldsymbol{\theta}_{F,i0}$, $\boldsymbol{\theta}_{G,i0}$ and $\boldsymbol{\theta}_{G,i1}$ in different trials $i = 1, \ldots, I$ are completely unrelated, while $\boldsymbol{\theta}_{F,i1}$ is specified for each trial in relation to $\boldsymbol{\theta}_{F,i0}$ e.g. by a fixed hazard ratio, we obtain a stratified analysis.

However, as we will discuss in Section 3.3 and Section 4.1.2, these parameters may not always be identified, and even if they are, it may not be possible to obtain a unique MLE.

In an attempt to avoid these problems, we can assume that $\boldsymbol{\theta}_{F,i0}$, $\boldsymbol{\theta}_{G,i0}$ and $\boldsymbol{\theta}_{G,i1}$ are exchangeable across the I trials to be meta-analyzed leading to the hierarchical meta-analysis model described in Section 3.4. Additionally, we will describe in Section 3.4.3 how to obtain priors for the random trial effect in a meta-analysis based on a set of H historical trials $i = I + 1, \ldots, I + H$ in a manner that has some robustness against prior-data conflicts. Another way to avoid the described issues, which we also address in Section 3.4.3, is to use (weakly-)informative priors for $\boldsymbol{\theta}_{F,i0}$, $\boldsymbol{\theta}_{G,i0}$ and $\boldsymbol{\theta}_{G,i1}$ in a stratified analysis.

3.3. Parameter identifiability and maximum likelihood estimation for the proposed likelihood

Definition 3.2. A parameter $\boldsymbol{\theta} \in \Theta$ for a family of distributions with pdf or pmf $f(\boldsymbol{w}; \boldsymbol{\theta})$ is identifiable, if for $\boldsymbol{\theta} \neq \boldsymbol{\theta}'$ the pdf or pmf is not the same (Casella and Berger, 2002, p. 523).

Theorem 3.2. *For independent and exponentially distributed failure times $X_{ijk} \sim Exp(\lambda_{ij})$ and drop-out times $C_{ijk} \sim \mathrm{Exp}(\mu_{ij})$ the parameters $\lambda_{ij} \in (0, \infty)$, $\mu_{ij} \in [0, \infty)$*

and $q_{ij} \in [0, 1]$ of the distribution defined by the pmf (3.2.1) *and* (3.2.3) *are identifiable.*

Proof. $\boldsymbol{p}_{ij} := (p_{ij1}, \dots, p_{ij5})^T$ determines the pmf (3.2.1) of a multinomial distribution. Thus, Theorem 3.2 holds, if any two parameter vectors $(\lambda_{ij}, \mu_{ij}, q_{ij})^T \neq (\lambda'_{ij}, \mu'_{ij}, q'_{ij})^T$ that differ in at least one component lead to different $\boldsymbol{p}_{ij}$.

Two such parameter vectors only result in the same $p_{ij4} = e^{-(\lambda_{ij} + \mu_{ij})\tau_i}$ for $\lambda_{ij} + \mu_{ij} = \lambda'_{ij} + \mu'_{ij}$. However, holding $\xi := \lambda_{ij} + \mu_{ij} = \lambda'_{ij} + \mu'_{ij}$ constant, $p_{ij5} = \mu_{ij}/\xi \times (1 - e^{-\xi\tau_i})$ is a strictly increasing function of μ_{ij} so that $\mu_{ij} \neq \mu'_{ij}$ lead to different distributions. Given a fixed $\mu_{ij} \in [0, \infty)$ the probability p_{ij4} as a function of $\lambda_{ij} \in (0, \infty)$ is a strictly decreasing one-to-one function. Thus, $\lambda_{ij} \neq \lambda'_{ij}$ lead to different distributions. Given fixed values $\mu_{ij} \in [0, \infty)$ and $\lambda_{ij} \in (0, \infty)$ we obtain $\lambda_{ij}/(\lambda_{ij} + \mu_{ij}) \times (1 - e^{-(\lambda_{ij} + \mu_{ij})\tau_i}) > 0$ so that p_{ij1} as a function of q_{ij} is a strictly increasing function. Thus, $q_{ij} \neq q'_{ij}$ lead to different distributions. $\square$

Remark 3.3. We needed to exclude the case $\lambda_{ij} = 0$, because otherwise $p_{ij1} = p_{ij2} = p_{ij3} = 0$ irrespective of the value of q_{ij}. In that case q_{ij} would not be identified.

Remark 3.4. Theorem 3.2 does not guarantee parameter identifiability for the pmf (3.2.4), but when the parameters for the pmf (3.2.1) are non-identifiable, this is also the case for the pmf (3.2.4). For example, in Section 4.1.2 we will prove parameter non-identifiability when event and drop-out times follow Weibull distributions.

Remark 3.5. Parameter identifiability does not ensure the existence of a unique or finite MLE for any particular realization from the distribution. Under the assumptions of Theorem 3.2 a unique finite MLE for all model parameters does not exist when $y_{ij\cdot} = n_{ij}$, $y_{ij\cdot} = 0$ or when $z_{ij\cdot} = n_{ij}$. Additionally, the MLE may lie on the boundary of the parameter space when $m_{ij\cdot} = y_{ij\cdot}$, $m_{ij\cdot} = 0$ or if $z_{ij\cdot} - m_{ij\cdot} = 0$. In the latter case partial derivatives of the likelihood function do not exist so that we cannot obtain SEs and confidence intervals using the usual asymptotic approximations.

Remark 3.6. A unique MLE $\hat{q}_{ij}$ exists — not just for exponentially distributed event and drop-out times — if $\boldsymbol{\theta}_{F,ij}$ and $\boldsymbol{\theta}_{G,ij}$ are distinct from q_{ij}, and $y_{ij\cdot} > 0$. In this case the MLE is $\hat{q}_{ij} = m_{ij\cdot}/y_{ij\cdot}$, because the first partial derivative of the

logarithm of (3.2.1) with respect to q_{ij} is $w_{ij1}/q_{ij} + w_{ij2}/(q_{ij} - 1) + w_{ij3}/(q_{ij} - 1)$, which is equal to zero for the given MLE, and because the second partial derivative $-w_{ij1}/q_{ij}^2 - w_{ij2}/(q_{ij} - 1)^2 - w_{ij3}/(q_{ij} - 1)^2$ is negative at the extremum so that it is a maximum.

The issues with maximum likelihood estimation from the proposed likelihood can be avoided or minimized in at least three ways. Firstly, we can make exchangeability assumptions about trial-level parameters across studies. This regularizes the estimates of the trial-level parameters. Even non-identifiable parameters — such as those of an AD Weibull model — may become identifiable, if trials vary sufficiently in duration to provide information on how the proportion of drop-outs and of patients with events varies with trial duration (Dias et al., 2014). Secondly, if parameter identifiability issues arise in (3.2.4) due to a lack of knowledge of each patient's time to event or censoring, having IPD available from some trials may avoid this issue. Exchangeability assumptions can then be used to borrow the information obtained from trials with available IPD, in order to avoid parameter identifiability issues in trials with AD, even when all trials are of similar planned duration. The likelihood contribution of the IPD of arm j of trial i is

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}_{F,ij}, \boldsymbol{\theta}_{G,ij}, q_{ij}|\text{IPD}) = \prod_{k=1}^{n_{ij}} & q_{ij}^{m_{ijk}}(1 - q_{ij})^{y_{ijk} - m_{ijk}} \\
& \times f(t_{ijk}; \boldsymbol{\theta}_{F,ij})^{y_{ijk}}\{1 - F(t_{ijk}; \boldsymbol{\theta}_{F,ij})\}^{1 - y_{ijk}} \\
& \times g(d_{ijk}; \boldsymbol{\theta}_{G,ij})^{z_{ijk} - m_{ijk}}\{1 - G(d_{ijk}; \boldsymbol{\theta}_{G,ij})\}^{1 - z_{ijk} + m_{ijk}}
\end{aligned}
$$

Thirdly, using proper priors for $\boldsymbol{\theta}_{F,ij}$, $\boldsymbol{\theta}_{G,ij}$ and q_{ij}, or on the hyperparameters $\boldsymbol{\vartheta}$ specifying random effects for the trial-level parameters ensures a proper posterior, although with vague priors it may remain a very diffuse distribution that is difficult to sample from (Gelfand and Sahu, 1999) — an issue that truly informative priors will reduce. Similarly, penalized likelihood approaches could be considered.

Exchangeability assumptions about the parameters of each trial also allow us to analyze data with partially missing information on $y_{ij\cdot}$, $z_{ij\cdot}$ or $m_{ij\cdot}$ without explicit imputation,

if data are missing at random (MAR). A likelihood-based analysis makes assumptions equivalent to imputing the missing data according to its conditional distribution given the observed data, while a Bayesian analysis implicitly does so according to the posterior predictive distribution. In either case missingness must only depend on the observed data for the trial under the MAR assumption, but we will be unable to assess whether the unavailability of data depends on the unreported data ("missing not at random"). E.g. publication authors may not report that no patients withdrew from a trial or that no patient had a particular adverse event.

3.4. Proposed hierarchical aggregate data meta-analysis model

We will use a hierarchical meta-analysis model for the data $y_{ij\cdot}$, $m_{ij\cdot}$, $z_{ij\cdot}$, n_{ij} and τ_i from a set of I trials $i = 1, \ldots, I$ of a test intervention against controls with the trial-level log-likelihood (3.2.4) and random study effects on $\boldsymbol{\theta}_{F,i0}$, $\boldsymbol{\theta}_{G,i0}$ and $\boldsymbol{\theta}_{G,i1}$. We assume that $\boldsymbol{\theta}_{F,i1}$ is defined by $\boldsymbol{\theta}_{F,i0}$ and a log-hazard ratio φ that is fixed across all trials. For each parameter $r = 1, \ldots, p$ out of the parameters $\boldsymbol{\theta}_{F,i0}$, $\boldsymbol{\theta}_{G,i0}$ and $\boldsymbol{\theta}_{G,i1}$ we will assume normally distributed independent $N(\nu_r, \sigma_r)$ random effects with mean ν_r and standard deviation σ_r across the trials for suitably transformed forms of the parameters. For the example of exponentially distributed event and drop-out times we would use the hierarchical meta-analysis model in Figure 3.3.

We will assume that $q_{ij} = q_j$ is identical for all patients in treatment group $j = 0, 1$ across all trials $i = 1, \ldots, I$. While this may not be true in general for trials of differing duration, we consider this a pragmatic choice, because the q_{ij} should typically be the least influential nuisance parameters as discussed in Section 3.2. Additionally, fatal events will often be much more rare than events so that separate parameters q_{ij} for each trial would be difficult to estimate. Alternative approaches include a random trial effect for logit q_{ij} and the approach introduced in Section 4.3. In the absence of any other prior information we will use independent Beta$(1/2, 1/2)$ priors for q_0 and q_1.
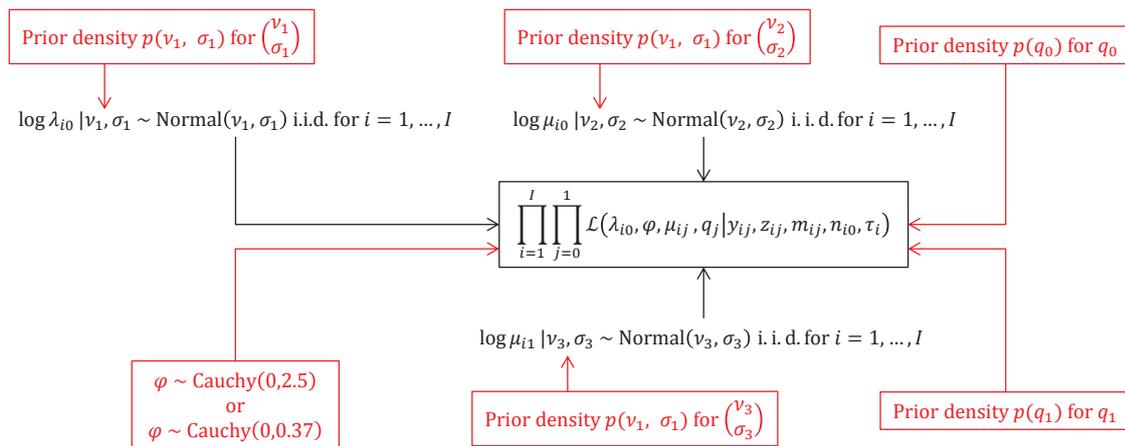
Figure 3.3.: Proposed hierarchical model for a meta-analysis of test compared with control assuming exponentially distributed event and drop-out times (prior distributions indicated in red), the likelihood for each treatment group is given by (3.2.4) with the probabilities (3.2.3)

3.4.1. Choice of priors for hierarchical mean and scale parameters

While we write $p(\nu_r, \sigma_r)$ for $r = 1, \ldots, p$ in Figure 3.3, we assume that vague priors for the hyperparameters are independent so that $p(\nu_r, \sigma_r) = p(\nu_r)p(\sigma_r)$ — except when using informative prior information (see Section 3.4.3). Even when there is a small number of trials, the trial-level data will typically be highly informative about the means of random trial effects (Gelman et al., 2014, p. 115). Thus, independent vague priors — e.g. $N(0, 1000)$ distributions for the means $\nu_1, \ldots, \nu_p$ of the random effects across trials — are a standard choice in the absence of prior information (Gelman et al., 2014, p. 115). A choice of a mean of zero for the priors lacks any justification, but will have extremely limited influence on inference as long as the prior is sufficiently flat across a wide range of possible values. For example, in case of exponentially distributed failure times, values of the log hazard rate of 0 correspond to a hazard rate of 1 indicating that 63% of patients at risk at the start of a year would experience an event by the end of the year, but the prior probability for a wide range of intervals for the log-hazard ratio φ is very similar: $[\log(0.0001), \log(0.001)]$ has a prior probability of 0.091857% and $[\log(0.1), \log(1)]$ of 0.091860%. Alternatively, a normal prior for the hierarchical mean parameter centered on a reasonable prior estimate and a standard deviation of, say, $\log(10)$ is a weakly informative alternative.

In contrast to the means the scale parameters $\sigma_1, \ldots, \sigma_p$ of random effects tend to

be more difficult to estimate (Gelman et al., 2014, pp. 128–132). In a hierarchical model the MLE for the hierarchical scale may be 0 despite the likelihood also providing considerable support for values far away from zero and many procedures for maximum likelihood estimation then proceed, as if the scale parameter were known to be 0 (Chung et al., 2013). This can be avoided by penalizing the likelihood sufficiently to avoid such an estimate, which is equivalent to Bayesian maximum-a-posteriori estimation with a specific prior (Chung et al., 2013). The choice of the prior for $\sigma_1, \ldots, \sigma_3$ will be particularly influential when only a small number of trials is included in a meta-analysis and it is difficult to formulate a truly uninformative prior — e.g. a conditionally conjugate inverse-gamma(ϵ, ϵ) prior for a between-trial variance is problematic due to the strong impact of the choice of the exact value of $\epsilon > 0$ on the posterior inference (Gelman et al., 2014, p. 129–130). For this reason using information on these parameters from historical data as discussed in Section 3.4.3 is attractive. In the absence of historical data, independent half-normal $HN(0, 100)$ with scale parameter 100, half-t $HT_3(0, 100)$ with 3 degrees of freedom and scale parameter 100, $HCa(0, 100)$ or uniform $U(0, 100)$ priors for the between-trial standard deviation may be a reasonable choice. They lead to an overestimation of the between-trial standard deviation, if the number of trials is small or the true between-trial standard deviation is small (Lambert et al., 2005). This results in less borrowing of information and greater uncertainty around the treatment effect, which can be seen as a conservative choice. An alternative may be to explicitly formulate a weakly informative prior that somewhat overstates the likely extent of between-trial heterogeneity. Out of the aforementioned priors Gelman et al. (2014, pp. 128–133) have suggested the half-t family of distributions and in particular the half-Cauchy distribution as default weakly informative priors, because their long tails allow the likelihood to dominate the prior. For such a prior we might choose a scale parameter of $\log(10)$ to express that it is a-priori unlikely that parameters such as an exponential hazard rate would vary by much more than a factor of 10 between trials.

3.4.2. Parametrization and prior for the treatment effect

The parametrization (2.3.2) is particularly suitable, if the prior for the control group hazard rate $\lambda_{i0}(s)$ is based on historical trials and that for φ on other considerations. If

we wish to express equal prior uncertainty about the hazard rates for all treatment groups or treatment effects are assumed to vary across trials in a frequentist random effects model, it is advantageous to parametrize the model in terms of main effects for trial and treatment, i.e. $\log \lambda_{i0}(s) = \log \lambda_i(s) - \varphi/2$ and $\log \lambda_{i1}(s) = \log \lambda_i(s) + \varphi/2$ (Agresti and Hartzel, 2000; Piepho et al., 2012; Smith et al., 1995).

For the rosiglitazone example in Section 3.5 we use weakly informative $\text{Ca}(0, 0.37)$ or $\text{Ca}(0, 2.5)$ Cauchy priors with location parameter 0 and, scale parameter 0.37 or 2.5 for the log-hazard ratio φ. The rationale for these priors is described in Appendix A.

3.4.3. Deriving robust historical meta-analytic predictive priors

To obtain joint rMAP priors for each pair of hyperparameters (ν_r, σ_r) $r = 1, \ldots, p$ from historical control group data $y_{i0\cdot}$, $m_{i0\cdot}$, $z_{i0\cdot}$, n_{i0} and τ_i for $i = I + 1, \ldots, I + H$, we fit a hierarchical model similar to that in Section 3.4 to the historical data. For the example of exponentially distributed event and drop-out times, the hierarchical meta-analysis model that we will also use in the rosiglitazone example in Section 3.5 is illustrated in Figure 3.4. When approximating the posterior from fitting this model to historical data, we approximate the joint posterior of the hyperparameters $(\nu_r, \log \sigma_r)$ $r = 1, \ldots, p$ by a mixture of bivariate normal distributions with a weakly informative mixture component chosen as in subsection 3.4.1.
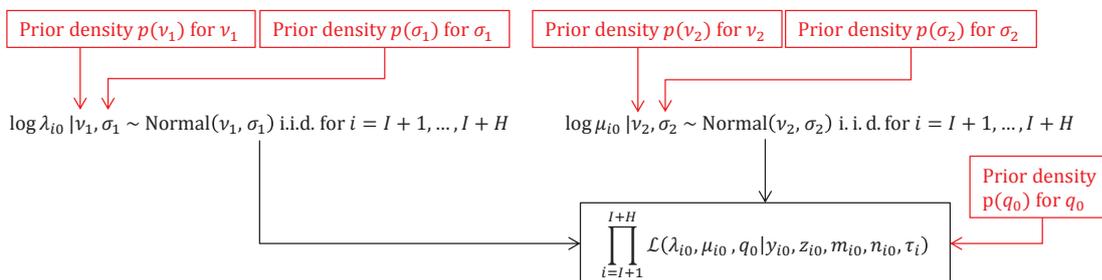
Figure 3.4.: Proposed hierarchical meta-analysis model for the historical control group data assuming exponentially distributed event and drop-out times with prior distributions indicated in red

Similarly, for approximating the predictive distribution for a single trial we approximate the marginal MAP distribution for $\log \lambda_{i0}$ by a mixture of normal or t-distributions.

In the same fashion we obtain rMAP priors for ν_2, $\log \sigma_2$ and $\log \mu_{i0}$, as well as ν_3, $\log \sigma_3$ and $\log \mu_{i1}$ based on the posterior for ν_2 and $\log \sigma_2$. To obtain priors for q_0 and q_1, we approximate the posterior for q_0 given the historical trials by a mixture of beta distributions and add a vague beta$(1/2, 1/2)$ mixture component.

3.5. Rosiglitazone cardiovascular safety example

3.5.1. Rosiglitazone dataset and selected historical data

As an example, we will consider a meta-analysis of type 2 diabetes trials with a duration of 2 months to 2 years presented at a 2010 United States Food and Drug Administration (FDA) advisory committee meeting to discuss the cardiovascular risk of rosiglitazone and the historical control group data from 64 type 2 diabetes trials of other drugs conducted in the same time period. This was the second of several meetings discussing concerns about the cardiovascular safety of rosiglitazone. These concerns were initially brought to prominence by a meta-analysis in 2007 (Nissen and Wolski, 2007). In 15 of the 54 (28%) of the rosiglitazone trials no event occurred in any treatment arm and in an additional 21 (39%) of the rosiglitazone trials events only occurred in one treatment arm. Additionally, in 66% of trials a higher proportion of control group than rosiglitazone patients did not complete the trial. Further details on the selection of data, used data sources, the compiled set of data and program code are available under https://doi.org/10.6084/m9.figshare.5007833.

3.5.2. Used analysis methods

We used a range of the standard analysis methods introduced in Section 2. We also fitted the proposed model in Figure 3.3 for independent exponentially distributed failure and drop-out times with the vague priors described in Section 3.4.1.

For an analysis with weakly informative hyperpriors we used the observational data

reported in the drug sponsor's briefing document for the 2010 rosiglitazone FDA advisory committee meeting(Glaxo-Smith-Kline, 2010), as well as information on typical trial drop-out rates as described in Appendix B. When fitting Bayesian models using MCMC methods we used every 5000th out of 50,000,000 MCMC samples from each of 5 chains after discarding 10,000 burn-in samples. We checked the mixing of each chain by visual inspection of trace and autocorrelation function plots. We assessed the non-convergence of each chain using the Geweke diagnostics and across chains using the Gelman-Rubin statistic.

We also conducted two analyses with rMAP priors based on the model in Figure 3.4. The first one used joint rMAP priors for each pair of hyperparameters (ν_r, σ_r) $r = 1, \ldots, 3$ (see Section 3.4.3) with weights of 20% or 50% for the weakly informative mixture component. The weakly informative mixture component consisted of two independent priors: a $HN(0, \log 10)$ prior for σ_r and a normal distribution with standard deviation $\log 10$ centered on the mean of the MAP prior for each ν_r. The second analysis was a stratified analysis with independent marginal rMAP priors for $\log \lambda_{i0}$, $\log \mu_{i0}$ and $\log \mu_{i1}$ for $i = 1, \ldots, I$. E.g. for the exponential control group MACE hazard rate in each trial $\log \lambda_{i0}$, the informative MAP component was

$$p(\log \lambda_{i0}) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \phi(\log \lambda_{i0}; \nu_1, \sigma_1) p(\nu_1, \sigma_1 | \mathcal{D}') \, \mathrm{d}\sigma_1 \, \mathrm{d}\nu_1.$$

Each of the priors $p(\log \lambda_{10}), \ldots, p(\log \lambda_{I0})$ was assumed to be independent. Each of these rMAP priors additionally had a weakly informative mixture component consisting of a t-distribution with 3 degrees of freedom with scale parameter 100 centered on the prior mean for the respective MAP prior. The limitations of such a stratified analysis with independent marginal rMAP priors were discussed in Section 2.5.4.2.

3.5.3. Results for the rosiglitazone example

In what follows we will concentrate of the hyperparameters for the log-hazard rate for events, but similar considerations apply for the other hyperparameters of the model. Figure 3.5 shows a bivariate histogram for the posterior density of the hierarchical mean
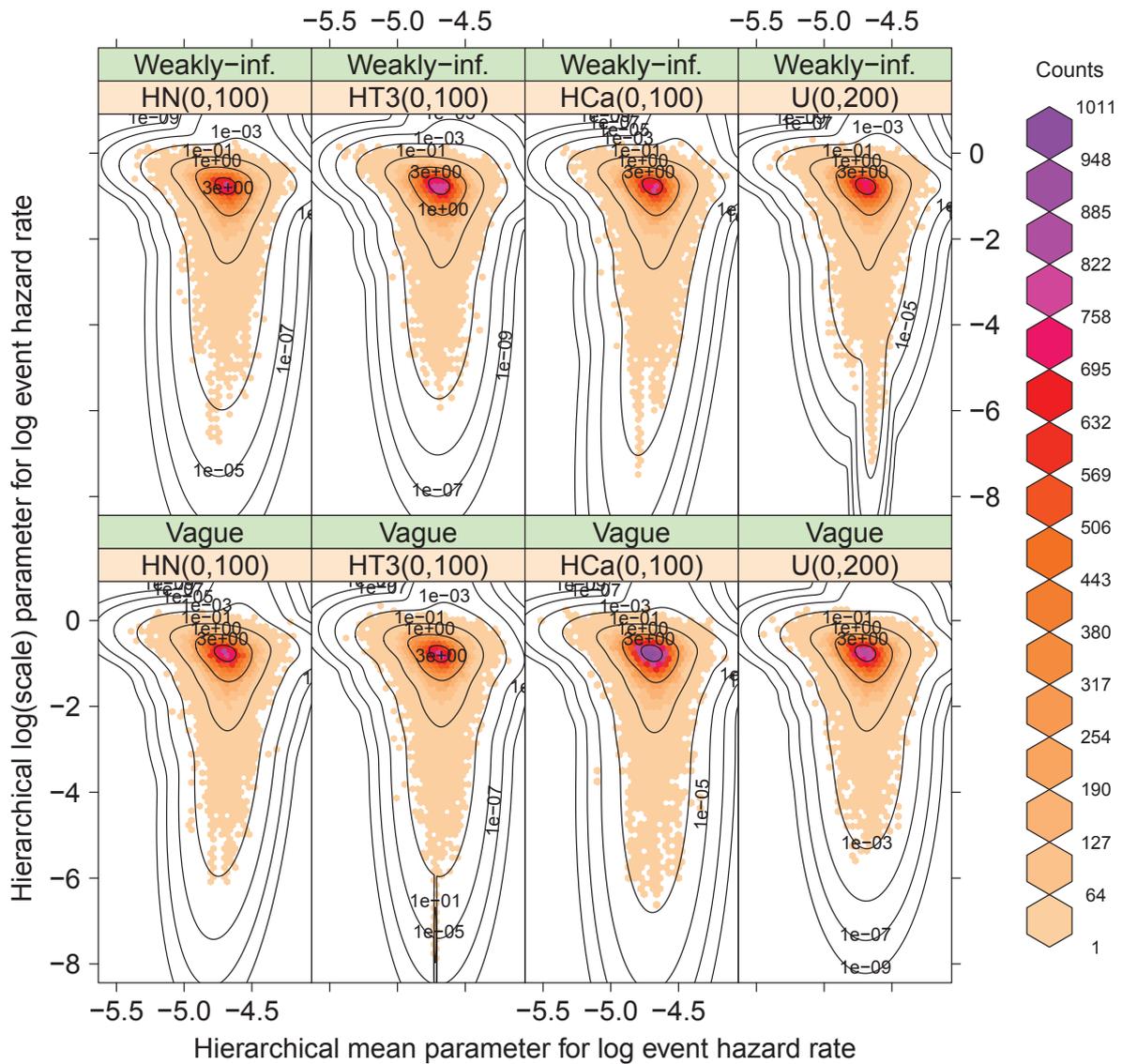
Figure 3.5.: Posterior density $p(\nu_1, \log \sigma_1 | \mathcal{D}')$ for hierarchical mean and log-scale parameters of the random effect for the log-hazard rate for events based on the historical data for different priors for the hierarchical mean (vague or weakly informative) and scale parameters (half-normal, half-T_3, half-Cauchy or uniform) with overlaid mixture approximation to the posterior density for hierarchical mean and log-scale parameters of the random effect for the log-hazard rate for events based on the historical data obtained using the mclust R package (Fraley et al., 2012) assuming a mixture of 5 bivariate normal distributions

and log-scale parameters ν_1 and $\log \sigma_1$ resulting from fitting the proposed model to the historical data. As these figures illustrate, the joint posterior does not differ substantially depending on the priors used for the hierarchical hyperparameters. Figure 3.6 illustrates that the choice of hyperpriors for the hierarchical model for the historical data also had little impact on the marginal MAP distribution for the exponential control group hazard rate for this particular large set of 64 historical trials. This was also the case of

the exponential drop-out rate and the probability of an event being fatal.

For these reasons we subsequently present only the results with a vague half-normal prior on hierarchical scale parameters for the historical data.

Figure 3.5 also shows the mixture approximations using bivariate normal distributions overlaid on top of the bivariate histogram for the MCMC samples. These mixture approximations were used as the basis for the informative part of the MAP hyperpriors used in the analyses labeled as being "with borrowing of information" below.
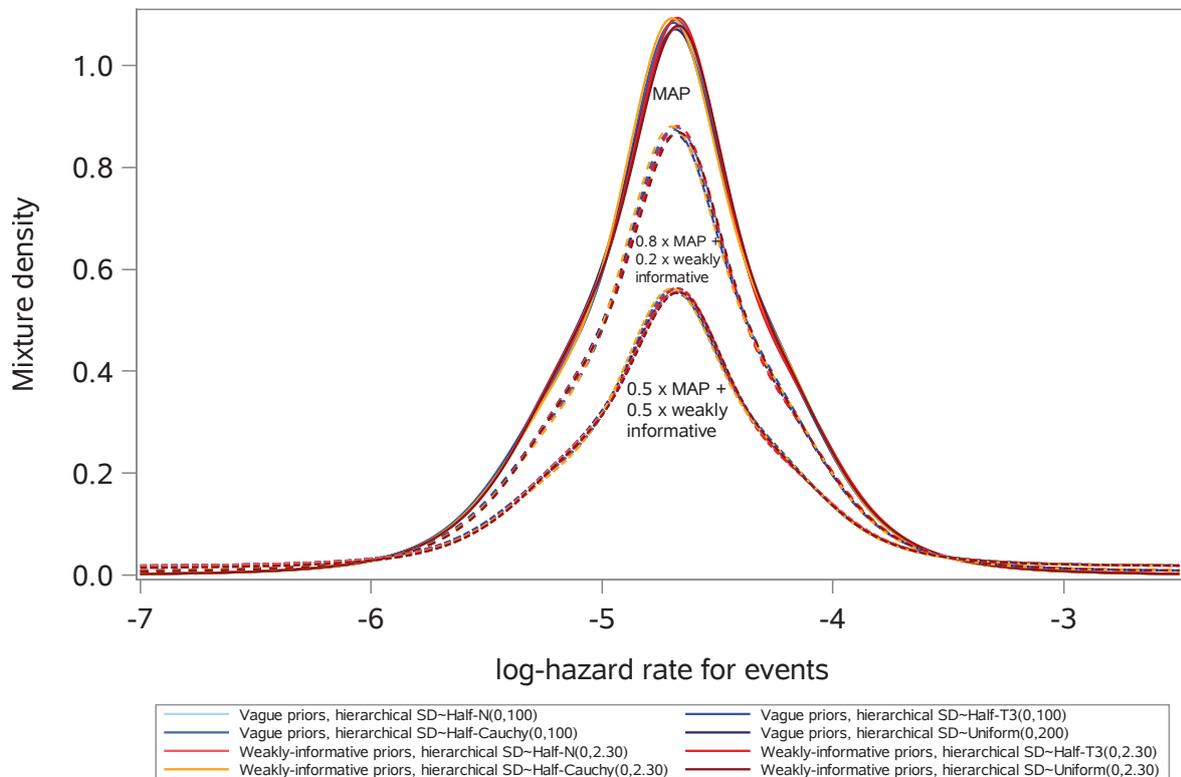
Figure 3.6.: Mixture approximations to the marginal MAP distribution for the exponential control group hazard rate for major adverse cardiovascular events in a new trial based on 64 historical trials for vague and weakly informative hyperpriors for the means of random trial-effects, as well as different hyperpriors for the hierarchical standard deviation (SD). The MAP distribution was approximated by a 3-component mixture of t-distributions with 3 degrees of freedom, while the additional weakly informative mixture component had a scale parameter of 100 and centered on the mean of the MAP distribution.

Remark 3.7. Using the MAP distribution in Figure 3.6 as the prior for the control group hazard rate and a $Ca(0, 0.37)$ prior for the log-hazard ratio, a single trial with no events in 100 patient-years of follow-up for both the group on a new test drug and control group results in a 95% posterior equal-tailed CI for the hazard ratio from

0.00004 to 2.1. I.e. the impact of the prior information on the control group hazard rate is substantial on the upper limit of the CI suggesting that more than a doubling of the hazard rate on the new drug is unlikely, while the prior information does little to bring the lower limit of the CI towards no effect so that an extremely large benefit of the new drug cannot be excluded.
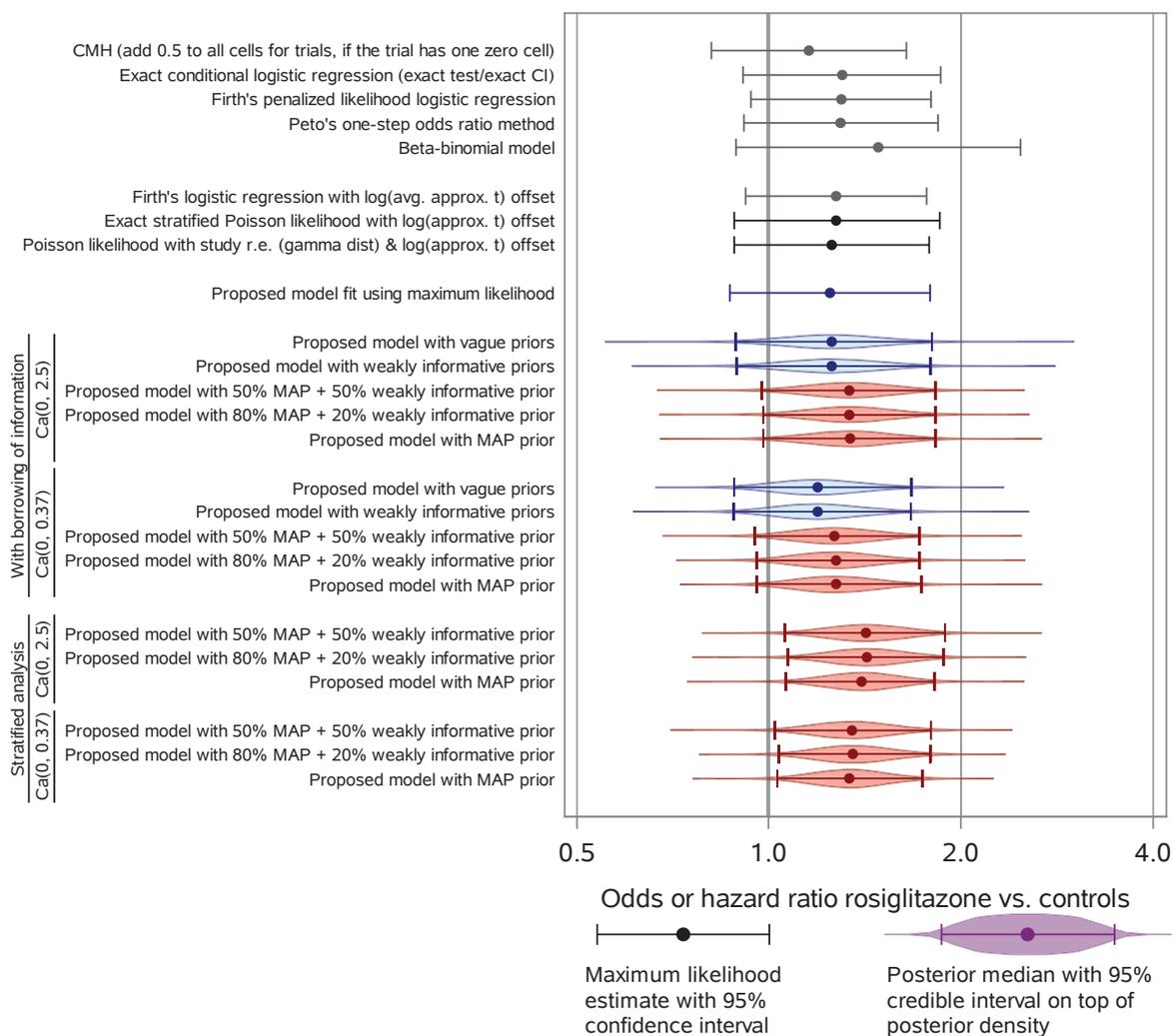
Figure 3.7.: Odds or hazard ratio for rosiglitazone compared with control groups for major adverse cardiovascular events from a meta-analysis of 54 trials. For the 17 analyses using the proposed trial-level likelihood (3.2.4) either maximum-likelihood estimation, a Ca(0, 0.37), or a Ca(0, 2.5) prior on the log-hazard ratio was used. Historical prior information was used to specify to specify independent priors for the parameters of each study ("stratified analysis") or a prior for the hyperparameters of a hierarchical model that allows borrowing of information as described in Section 3.4.3 either with no down-weighting, some down-weighting, only to center weakly informative priors, or not at all.

As shown in Figure 3.7 the proposed hierarchical meta-analysis model fitted using maximum likelihood resulted in very similar point estimates — with slightly greater

uncertainty around the estimate — as fitting the same model using vague or weakly informative priors on nuisance parameters and a $Ca(0, 2.5)$ prior on the log-hazard ratio. The closely related exponential time-to-event model using the imputed duration of follow-up to first event or censoring with a random effect on the control group hazard rates of each trial also resulted in very similar inference, but with narrower CIs. In this example, using a log imputed duration of follow-up to first event or censoring $\log \tilde{t}_{ij}$ offset in a stratified exact exponential model or the Firth's penalized likelihood logistic regression also resulted in very similar results.

Approaches that ignore the differences in follow-up such as Firth's penalized likelihood logistic, exact conditional logistic regression, and Peto's one-step odds ratio method resulted in slightly higher estimated odds-ratios than Firth's penalized likelihood logistic regression with a $\log \tilde{t}_{ij}$ offset. This suggests that under the assumption of independent drop-out and event times any differences in follow-up have only a minor effect on inference for this example.

The results for the CMH odds-ratio using a continuity correction of 0.5 for those studies with at least one treatment group without an event illustrate that considerable shrinkage towards no treatment effect is induced by such a continuity correction, if — as in this example — two thirds of trials have at least one arm with no events and most trials have as many or more patients on the test group than on the control group. The beta-binomial model proposed by Kuss (Kuss, 2015) resulted in a higher point estimate and a much wider CI than all other models.

Using historical prior information on control group hazard rates and other nuisance parameters resulted in a higher point estimate for the hazard ratio for rosiglitazone compared to control groups, because historical control rates were somewhat lower than those observed in the control groups of rosiglitazone trials. The effect of the historical prior information was less pronounced when strength was borrowed across trials within the rosiglitazone meta-analysis than in the corresponding stratified analyses at the bottom of Figure 3.7. Whether the historical prior was given 100%, 80% or 50% weight in the mixture prior for an analysis with borrowing of information made no appreciable difference to the analysis results, while for the stratified model there appeared to be

an increase in the width of CIs with increasing down-weighting of the historical prior information.

3.6. Simulation study

3.6.1. Simulation set-up and conduct

We conducted a simulation study mimicking a typical drug development program involving 1425 patients that compares a test drug with control groups. The simulated program shown in Figure 3.8 consists of two early-stage 12 week trials with 6:1 and 1:1 randomization, respectively, three half-year trials with 2:1 randomization, and a single 1-year trial with 1:1 randomization. Sample sizes range from 25 to 300 patients per arm. For each treatment group we assumed identical exponential event and drop-out time distributions in all trials. We assumed either no increase in the hazard rate for events in the test group compared with control groups or an increased hazard rate chosen to result in approximately 50% power. We chose the drop-out hazard rate to be the same, substantially higher or substantially lower in the test compared with the control groups.

To investigate the effect of informative priors we simulated control group data for 12 historical trials with a mixture of trial durations and sample sizes as shown in Figure 3.8 assuming the same event and drop-out time distributions as for the main meta-analysis. We then fitted the proposed Bayesian hierarchical AD meta-analysis model with borrowing of information described in Section 3.4 either with vague priors, a MAP prior or a rMAP prior for the hyperparameters of the model.

We used SAS/STAT® 14.1 software, Version 9.4 of the SAS System for Linux to simulate IPD, derive AD and to implement the following reference methods: the CMH odds ratio with continuity correction, the beta-binomial model, the Peto one-step odds-ratio method, exact conditional logistic regression stratified by trial, Firth's penalized logistic regression with a model term for each trial (with and without a $\log(\tilde{t}_{ij.}/n_{ij})$ offset) and exponential time-to-event models using a "Poisson likelihood" with a $\log(n_{ij}\tau_i)$, or
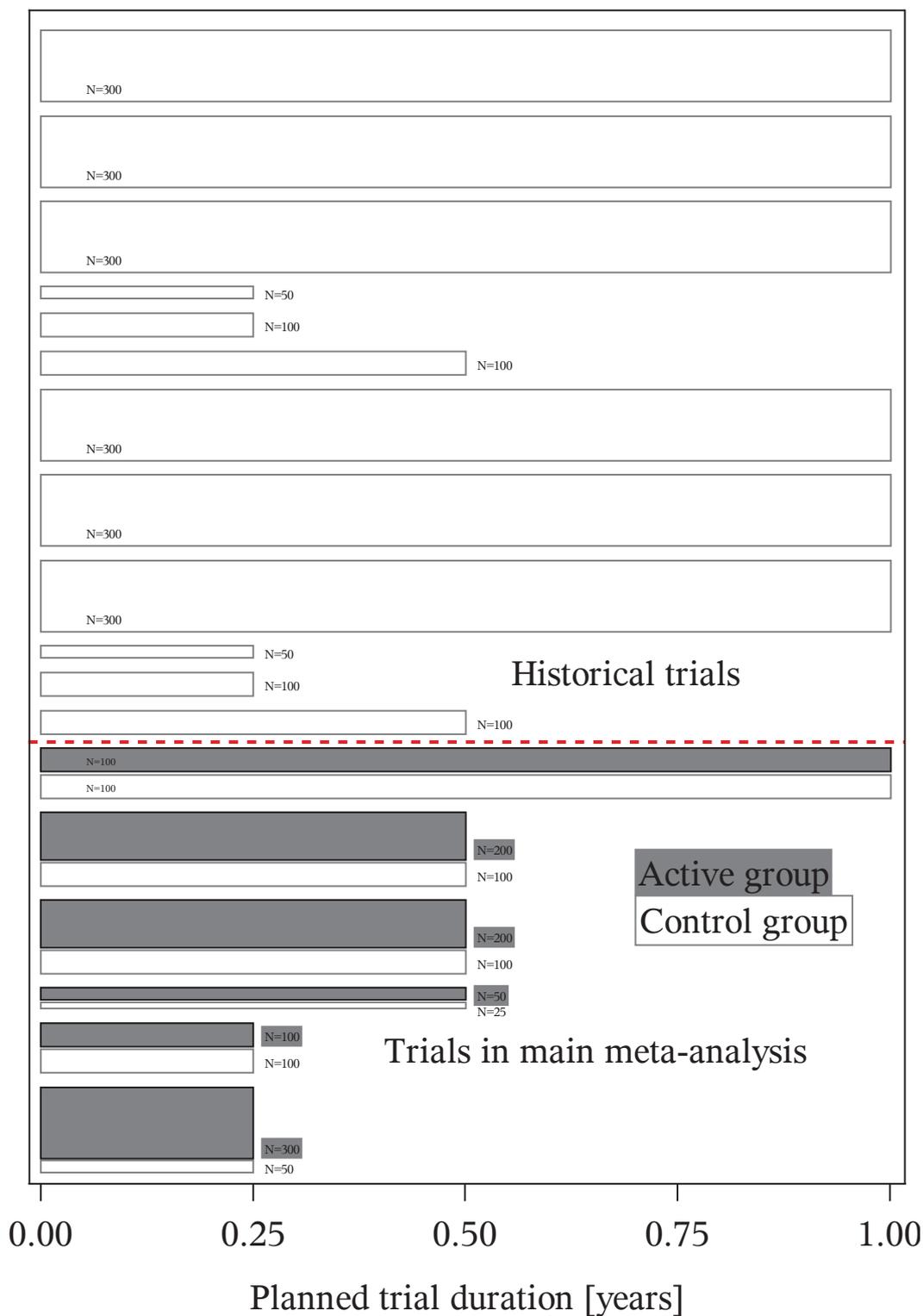
Figure 3.8.: Details of simulated trials in terms of sample size and planned trial duration

a $\log(\tilde{t}_{ij.})$ offset either with a gamma distributed random trial effect, or using software for exact Poisson regression stratified by trial. To optimize runtime, we implemented

all Bayesian models using Stan (Carpenter et al., 2017) via the rstan 2.7.0 package within R 3.0.2 for Linux. We used the NUTS (Hoffman and Gelman, 2014) and a non-centered parametrization that has been recommended for Hamiltonian Monte-Carlo when data are sparse (Betancourt and Girolami, 2015). Program code is available under https://doi.org/10.6084/m9.figshare.5007833.

Table 3.1.: Parameters for the simulation scenarios: placebo event hazard rate λ_{i0}, log-hazard ratio for events on test drug compared with placebo φ, placebo drop-out hazard rate μ_{i0}, test drug drop-out hazard rate μ_{i1}, probability that an event is fatal on placebo q_0 and probability that an event is fatal on test drug q_1 for all trials $i = 1, \ldots, I$ in the main meta-analysis and the historical trial $i = I + 1, \ldots, I + H$

Scenario	I	H	λ_{i0}	φ	μ_{i0}	μ_{i1}	q_0	q_1
1	6	12	0.05	0	0.5	0.5	0.35	0.35
2	6	12	0.05	0	0.5	1.36	0.35	0.35
3	6	12	0.05	0	0.5	0.18	0.35	0.35
4	6	12	0.05	0.7	0.5	0.5	0.35	0.35
5	6	12	0.05	0.7	0.5	1.36	0.35	0.35
6	6	12	0.05	0.7	0.5	0.18	0.35	0.35
7	6	12	0.5	0	0.5	0.5	0.35	0.35
8	6	12	0.5	0	0.5	1.36	0.35	0.35
9	6	12	0.5	0	0.5	0.18	0.35	0.35
10	6	12	0.5	0.25	0.5	0.5	0.35	0.35
11	6	12	0.5	0.25	0.5	1.36	0.35	0.35
12	6	12	0.5	0.25	0.5	0.18	0.35	0.35

The parameters used in the 12 simulation scenarios to simulate data are given in Table 3.1, while the sample size and planned trial duration for the I trials in the main meta-analysis and the H historical trials are shown in Figure 3.8.

The log-hazard ratios of 0.7 and 0.25 — corresponding to hazard ratios of 2.0 and 1.3, respectively — were chosen to result in approximately 50% power for an unstratified log-rank test to ensure differences between analysis methods could be seen clearly. If an event occurred for a patient, the event was considered to have been fatal according to the probabilities q_{ij} given in Table 3.1 — i.e. the simulations exactly followed the assumptions made by the proposed analysis model in this respect.

For evaluating the different analysis methods we present type I error rate and power — defined as the proportion of simulations with two-sided p-value ≤ 0.05 or with the Bayesian 95% CI excluding a log-hazard ratio of 0. In addition, we also investigated CI

coverage, bias $\sum_{r=1}^{R} \frac{\varphi - \hat{\varphi}_r}{R}$, to assess median-unbiasedness also the median value of estimated minus true log-hazard ratio and root mean squared error (RMSE) $\sqrt{\sum_{r=1}^{R} \frac{(\varphi - \hat{\varphi}_r)^2}{R}}$, for which we give full results in Appendix C.

3.6.2. Simulation study results
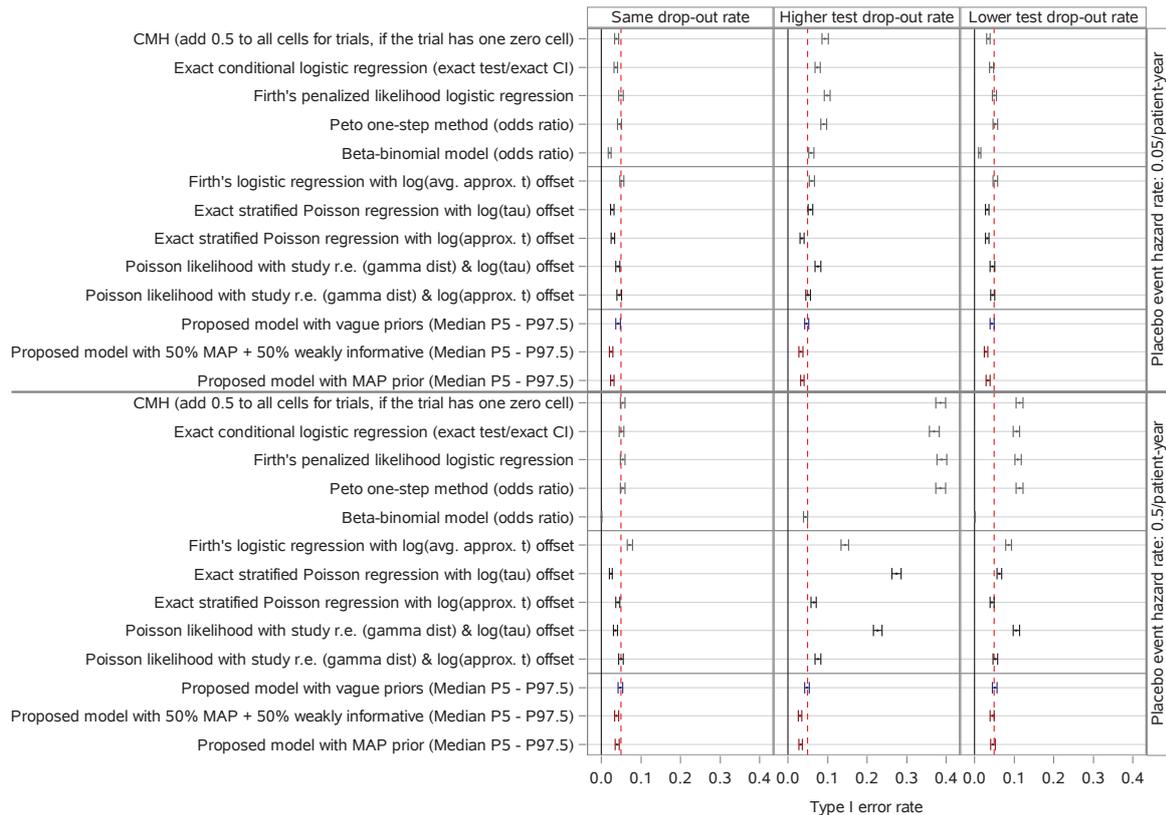
Figure 3.9.: Type I error rate with 99% Clopper-Pearson CIs based on 10,000 simulations under different scenarios: control and test group hazard rate for events (0.05 or 0.5; no treatment effect), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)

Under the null hypothesis of no treatment effect on event times, nearly all methods approximately kept the type I error rate, if the drop-out time distribution was the same across treatment groups (see Figure 3.9). The only exception was Firth's penalized likelihood logistic regression with a $\log \tilde{t}_{ij}$ offset, which only approximates the true relationship between t_{ij} and the probability of an event. The type I error inflation occurred in the scenario of a high hazard rate for events of 0.5/patient-year — i.e. when the approximation is expected to be poor (see Section 2.3.3.2). In contrast, the beta-binomial model had a type I error rate notably below the nominal level.

All methods that assumed a binomial distribution and ignored differences in the number of drop-outs had inflated type I error rates when there was a higher drop-out hazard rate in the test group — this type I error rate inflation was more pronounced for higher event hazard rates.

Using a log estimated average follow-up offset in logistic regression reduced this type I error rate inflation, but in some scenarios still resulted in a type I error rate more than three times above the nominal level. Using a log estimated follow-up offset in combination with a "Poisson likelihood" was effective in this regard for the lower hazard rate for events of 0.05/patient-year, while using a log planned trial duration offset resulted in an increased type I error rate in the presence of differential drop-out.

For the higher exponential hazard rate for events of 0.5/patient-year the "Poisson likelihood" with a log estimated follow-up offset also showed some type I error rate inflation.

In the absence of differential drop-out, there were only small differences in power between the methods that do not use historical data as shown in Figure 3.10. The only exception was the beta-binomial model that had low power in all considered scenarios despite similar point estimates as other methods, because its CIs were wider by a factor of 2.4 to 4.7 compared to the method with the consistently shortest CIs — Firth's penalized likelihood logistic regression.

When drop-out hazard rates differed between treatment groups, the methods that control the type I error rate showed only small changes in power consistent with fewer observed events in case of more drop-outs and more observed events in case of fewer drop-outs.

As expected, using historical control group data generated from the same distribution as the data in the main meta-analysis resulted in a clear increase in power. There was only a small loss in power when down-weighting the informative prior by adding a weakly informative mixture component with 50% weight.

Full results on CI coverage, bias, median-unbiasedness and RMSE are presented in
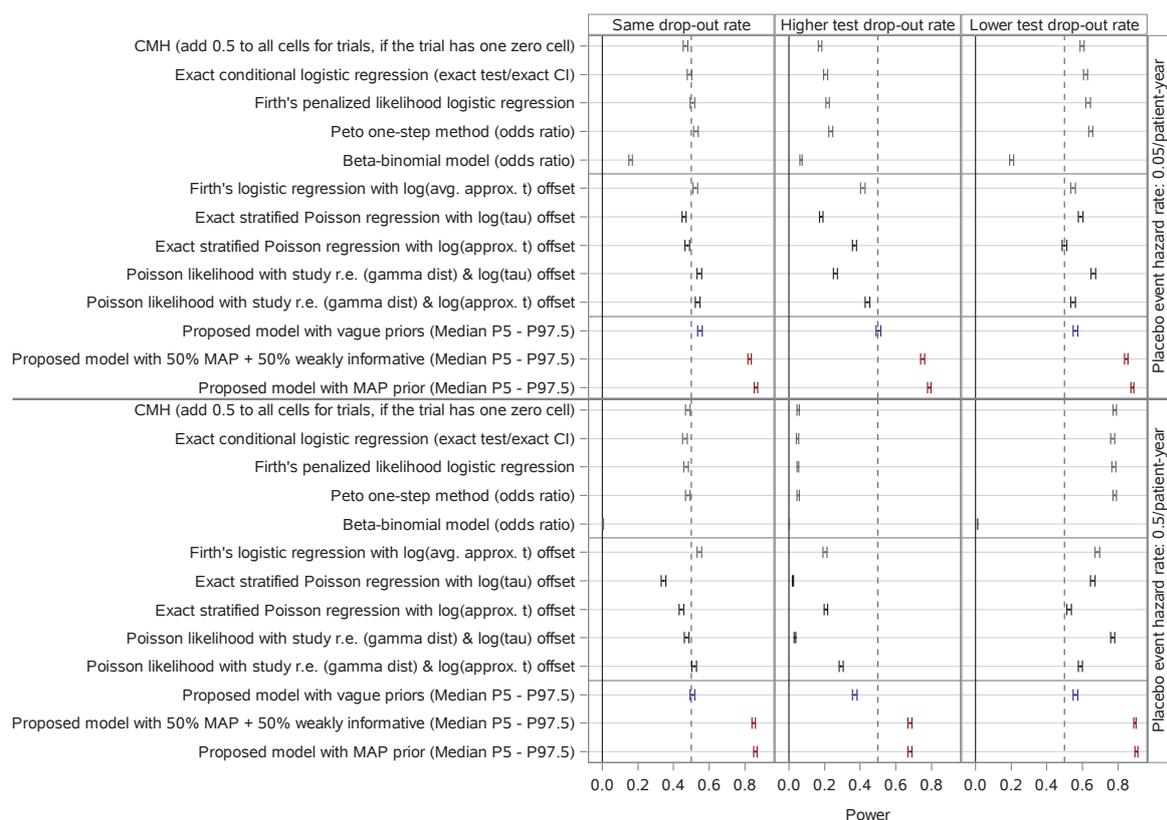
Figure 3.10.: Power to detect an increased event occurrence in the test compared to the control groups with 99% Clopper-Pearson CIs based on 10,000 simulations under different scenarios: control group hazard rate for events (0.05 or 0.5), test group hazard rate for events (0.1 or 0.64), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)

Appendix C. In summary, the proposed Bayesian hierarchical model with or without informative priors provided the least biased estimates across the evaluated scenarios and the same was the case for the median difference between estimate and true value of the log-hazard ratio. As expected, methods that estimate an odds ratio provided biased estimates of the hazard ratio in the presence of a treatment effect.

In terms of RMSE, the proposed Bayesian hierarchical model with robust historical priors performed consistently well across the considered scenarios. As expected — given that the historical data was generated under the same model and parameters as the data for the main meta-analysis — giving greater weight to the historical prior reduced the RMSE.

The proposed Bayesian hierarchical model with vague priors, as well as exponential time-to-event analyses using a Poisson likelihood and a $\log \tilde{t}_{ij.}$ offset, performed relatively

consistently, but in some scenarios other methods had a lower RMSE.

The coverage of 95% confidence or credible intervals (see Appendix C) was largely in line with the results for the type 1 error rate. The proposed Bayesian hierarchical model with vague priors provided coverage at approximately the nominal level in all scenarios, while coverage was generally slightly above the nominal level with informative priors. All other methods had substantially below nominal coverage in some scenarios.

3.6.3. Comparison with IPD time-to-event methods

To investigate the loss of efficiency involved in only having AD available, we also compared the proposed Bayesian hierarchical AD model with standard IPD time-to-event analysis methods in terms of type I error rate and power for the same scenarios as in the preceding section.

As Figure 3.11 shows, all IPD time-to-event methods controlled the type I error rate and had similar power to each other. For the considered scenarios with exponentially distributed failure times and without any covariates to be taken into account, the proposed model with vague priors appeared to have slightly higher power than the IPD time-to-event methods in several of the scenarios and in particular those with a low placebo hazard rate (0.05/patient-year).

The outcome of this comparison was less clear for the higher placebo hazard rate (0.5/patient-year) — IPD time-to-event models clearly outperformed the proposed model with vague priors when there were more drop-outs in the test group than the placebo group, while it was the other way around when there were fewer drop-outs in the test compared to the placebo group.

This matches the theoretical finding that the power advantage of IPD time-to-event analyses over binomial methods increases with an increasing proportion of censored observations (Buyse and Ryan, 1987).

Note that the proposed hierarchical AD model — unlike the considered IPD time-to-
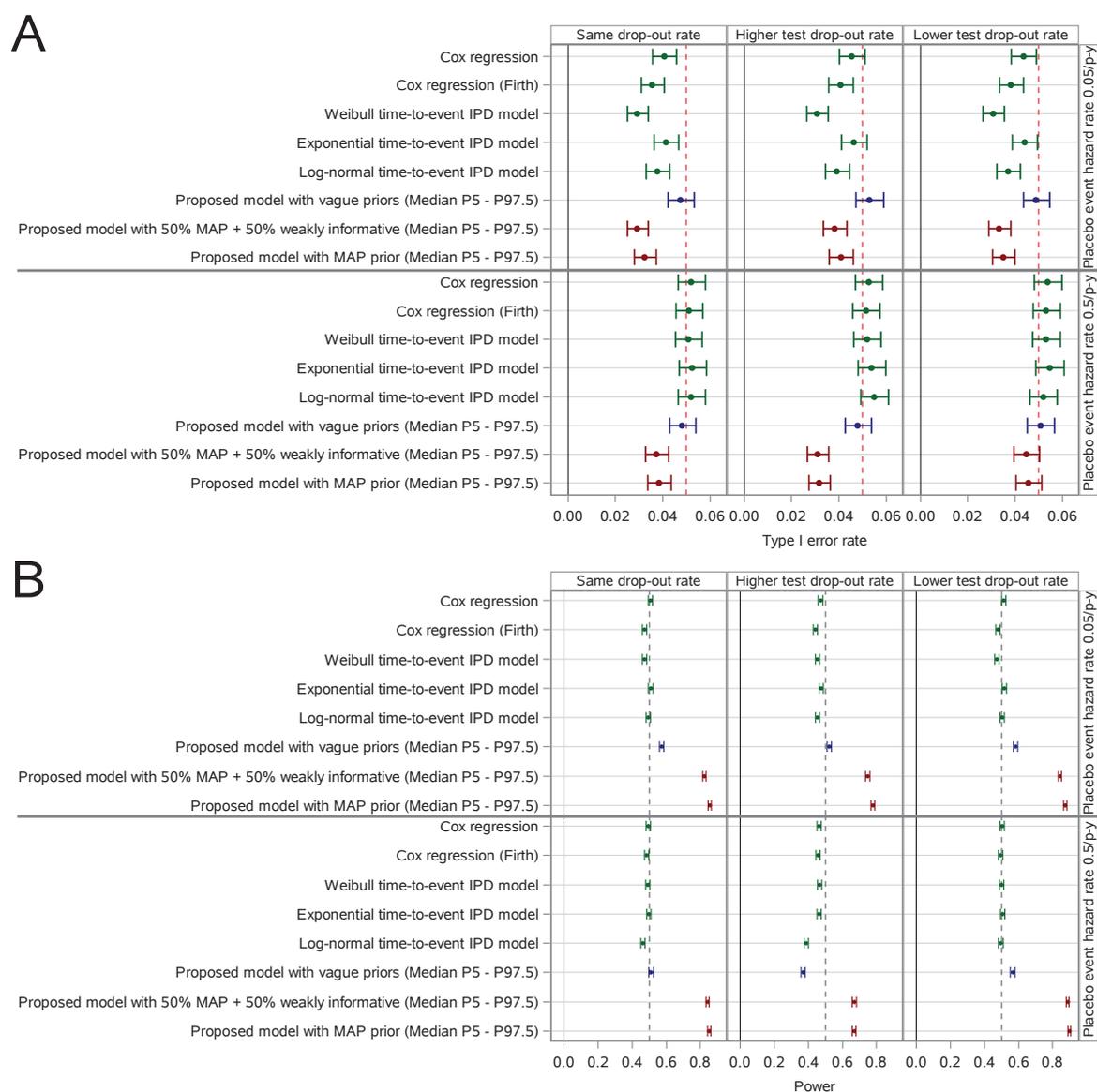
Figure 3.11.: Proportion of simulations with two-sided p-value ≤ 0.05 or with the Bayesian 95% CI excluding a log-hazard ratio of 0 under the null hypothesis (Panel A) or the alternative (Panel B) with 99% Clopper-Pearson CIs based on 10,000 simulations: hazard rate for events 0.05 or 0.5 per patient-year (/p-y) in control groups, and 0.05 or 0.5/p-y (Panel A) and 0.1 or 0.64/p-y (Panel B) for test groups; drop-out hazard rate 0.5/p-y in control groups, and 0.5, 1.36 or 0.18/p-y for test groups.

event models — borrows information across trials. For this reason the proposed model should perhaps be compared to the exponential AD model with a Poisson likelihood and gamma-distributed random trial effect in Figure 3.10, which performed very similarly to the proposed model with vague priors. The proposed Bayesian hierarchical AD model with informative MAP or rMAP priors had higher power than IPD methods.

4. Extensions to the proposed aggregate data model

In Chapter 3 we considered the general principles for deriving the probability of each of the mutually exclusive outcomes for a patient in Definition 3.1. We also derived these probabilities for a trial of fixed planned duration under the assumption of exponentially distributed event and drop-out times, and assuming that within each treatment group all patients with an event are equally likely to die.

In this chapter we demonstrate how each of these assumptions can be relaxed by extending the results to Weibull distributed event and drop-out times, event driven trials and the situation when a patient may have multiple potentially fatal events. The material in this chapter may be skipped by the reader, because it is not a necessary background for subsequent chapters.

4.1. Weibull distributed event and drop-out times

4.1.1. Likelihood for Weibull distributed event and drop-out times

We will assume i.i.d. Weibull distributed failure times $X_{ijk} \sim \mathrm{Wb}(\lambda_{ij}, \gamma_{ij})$ with inverse scale parameter λ_{ij} and shape parameter γ_{ij} using the parametrization (2.3.4), and i.i.d. drop-out times $C_{ijk} \sim \mathrm{Wb}(\mu_{ij}, \omega_{ij})$ that are independent of the event times for patient $k = 1, \ldots, n_{ij}$ in group j of trial i of planned duration τ_i.

Theorem 4.1. *Under the assumptions above, the probabilities of a patient falling into*

each of the five mutually exclusive categories of Definition 3.1 are given by

$$p_{ij1} = q_{ij} \underbrace{\gamma_{ij}\lambda_{ij}^{\gamma_{ij}} \int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}\right) x^{\gamma_{ij}-1}\,\mathrm{d}x}_{=:u_1},$$

$$p_{ij2} = (1 - q_{ij}) \exp\left(-(\tau_i\mu_{ij})^{\omega_{ij}}\right)\left[1 - \exp\left(-(\tau_i\lambda_{ij})^{\gamma_{ij}}\right)\right],$$

$$p_{ij3} = (1 - q_{ij})\left(u_1 - \exp\left(-(\tau_i\mu_{ij})^{\omega_{ij}}\right)\left[1 - \exp\left(-(\tau_i\lambda_{ij})^{\gamma_{ij}}\right)\right]\right),$$

$$p_{ij4} = \exp\left(-(\tau_i\lambda_{ij})^{\gamma_{ij}} - (\tau_i\mu_{ij})^{\omega_{ij}}\right) \quad and$$

$$p_{ij5} = 1 - \underbrace{\exp\left(-(\tau_i\lambda_{ij})^{\gamma_{ij}} - (\tau_i\mu_{ij})^{\omega_{ij}}\right)}_{=p_{ij4}} - u_1.$$

Remark 4.1. p_{ij1}, p_{ij3} and p_{ij5} all require the evaluation of the one-dimensional integral in u_1. We address its numerical evaluation in Appendix D.

Proof of Theorem 4.1. We obtain the probability that a patient has a fatal event as

$$
\begin{aligned}
p_{ij1} &= q_{ij} \int_0^\infty \int_0^\infty f_{\mathrm{Wb}}(x; \lambda_{ij}, \gamma_{ij}) f_{\mathrm{Wb}}(c; \mu_{ij}, \omega_{ij}) \mathbb{1}\{x \leq \min(c, \tau_i)\}\,\mathrm{d}c\,\mathrm{d}x \\
&= q_{ij} \int_0^{\tau_i} f_{\mathrm{Wb}}(x; \lambda_{ij}, \gamma_{ij}) \underbrace{\left[\int_x^{\tau_i} f_{\mathrm{Wb}}(c; \mu_{ij}, \omega_{ij})\,\mathrm{d}c + \int_{\tau_i}^\infty f_{\mathrm{Wb}}(c; \mu_{ij}, \omega_{ij})\,\mathrm{d}c\right]}_{=1 - F_{\mathrm{Wb}}(x; \mu_{ij}, \omega_{ij})}\,\mathrm{d}x \\
&= q_{ij}\gamma_{ij}\lambda_{ij}^{\gamma_{ij}} \int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}\right) x^{\gamma_{ij}-1}\,\mathrm{d}x,
\end{aligned}
\tag{4.1.1}
$$

the probability that the patient experiences a non-fatal event and completes the trial as

$$
\begin{aligned}
p_{ij2} &= (1 - q_{ij}) \int_0^\infty \int_0^\infty f_{\mathrm{Wb}}(x; \lambda_{ij}, \gamma_{ij}) f_{\mathrm{Wb}}(c; \mu_{ij}, \omega_{ij}) \mathbb{1}\{x \leq \tau_i\}\mathbb{1}\{c \geq \tau_i\}\,\mathrm{d}x\,\mathrm{d}c \\
&= (1 - q_{ij}) \int_{\tau_i}^\infty \omega_{ij} c^{\omega_{ij}-1}\mu_{ij}^{\omega_{ij}}\left[\exp\left((\lambda_{ij}\tau_i)^{\gamma_{ij}}\right) - 1\right] \exp\left(-(c\mu_{ij})^{\omega_{ij}} - (\lambda_{ij}\tau_i)^{\gamma_{ij}}\right)\,\mathrm{d}c \\
&= (1 - q_{ij}) \exp\left(-(\tau_i\mu_{ij})^{\omega_{ij}}\right)\left[1 - \exp\left(-(\tau_i\lambda_{ij})^{\gamma_{ij}}\right)\right],
\end{aligned}
$$

the probability that a patient completes the study without an event as

$$
\begin{aligned}
p_{ij4} &= \int_0^\infty \int_0^\infty f_{\mathrm{Wb}}(x; \lambda_{ij}, \gamma_{ij}) f_{\mathrm{Wb}}(c; \mu_{ij}, \omega_{ij}) 1\{\tau_i < \min(x, c)\}\,\mathrm{d}x\,\mathrm{d}c \\
&= \int_{\tau_i}^\infty \omega_{ij} c^{\omega_{ij}-1}\mu_{ij}^{\omega_{ij}} \exp\left(-(c\mu_{ij})^{\omega_{ij}} - (\lambda_{ij}\tau_i)^{\gamma_{ij}}\right)\,\mathrm{d}c \\
&= \exp\left(-(\tau_i\lambda_{ij})^{\gamma_{ij}} - (\tau_i\mu_{ij})^{\omega_{ij}}\right),
\end{aligned}
$$

the probability of a patient having a non-fatal event and dropping out thereafter as

$$
\begin{aligned}
p_{ij3} =& (1-q_{ij}) \int_0^\infty \int_0^\infty f_{\mathrm{Wb}}(x; \lambda_{ij}, \gamma_{ij}) f_{\mathrm{Wb}}(c; \mu_{ij}, \omega_{ij}) 1\{x \le c < \tau_i\} \, \mathrm{d}c \, \mathrm{d}x \\
=& (1-q_{ij}) \gamma_{ij} \lambda_{ij}^{\gamma_{ij}} \exp\left(-(\tau_i \mu_{ij})^{\omega_{ij}}\right) \int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}\right) \\
& \hspace{4cm} \times \left[\exp\left((\tau_i \mu_{ij})^{\omega_{ij}}\right) - \exp\left((x\mu_{ij})^{\omega_{ij}}\right)\right] x^{\gamma_{ij}-1} \, \mathrm{d}x \\
=& (1-q_{ij}) \gamma_{ij} \lambda_{ij}^{\gamma_{ij}} \Bigg[\int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}\right) x^{\gamma_{ij}-1} \, \mathrm{d}x \\
& \hspace{1.5cm} - \exp\left(-(\tau_i \mu_{ij})^{\omega_{ij}}\right) \underbrace{\int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}}\right) x^{\gamma_{ij}-1} \, \mathrm{d}x}_{=\gamma_{ij}^{-1} \lambda_{ij}^{-\gamma_{ij}} \left[1-\exp\left(-(\tau_i \lambda_{ij})^{\gamma_{ij}}\right)\right]} \Bigg] \quad (4.1.2) \\
=& (1-q_{ij}) \Bigg(\underbrace{\gamma_{ij} \lambda_{ij}^{\gamma_{ij}} \int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}\right) x^{\gamma_{ij}-1} \, \mathrm{d}x}_{=u_1 \text{ as defined in Theorem 4.1}} \\
& \hspace{1cm} - \exp\left(-(\tau_i \mu_{ij})^{\omega_{ij}}\right) \left[1 - \exp\left(-(\tau_i \lambda_{ij})^{\gamma_{ij}}\right)\right] \Bigg),
\end{aligned}
$$

and the probability that a patient drops out before any event or the end of the trial as

$$
\begin{aligned}
p_{ij5} =& \int_0^\infty \int_0^\infty f_{\mathrm{Wb}}(x; \lambda_{ij}, \gamma_{ij}) f_{\mathrm{Wb}}(c; \mu_{ij}, \omega_{ij}) 1\{c < x\} 1\{c \le \tau_i\} \, \mathrm{d}c \, \mathrm{d}x \\
=& \exp\left(-(\tau_i \lambda_{ij})^{\gamma_{ij}} - (\tau_i \mu_{ij})^{\omega_{ij}}\right) \left[\exp\left((\tau_i \mu_{ij})^{\omega_{ij}}\right) - 1\right] \\
& + \underbrace{\int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}}\right) \gamma_{ij} x^{\gamma_{ij}-1} \lambda_{ij}^{\gamma_{ij}} \, \mathrm{d}x}_{=1-\exp\left(-(\tau_i \lambda_{ij})^{\gamma_{ij}}\right)} \\
& - \int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}\right) \gamma_{ij} x^{\gamma_{ij}-1} \lambda_{ij}^{\gamma_{ij}} \, \mathrm{d}x \\
=& 1 - \underbrace{\exp\left(-(\tau_i \lambda_{ij})^{\gamma_{ij}} - (\tau_i \mu_{ij})^{\omega_{ij}}\right)}_{=p_{ij4}} \\
& - \underbrace{\gamma_{ij} \lambda_{ij}^{\gamma_{ij}} \int_0^{\tau_i} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}\right) x^{\gamma_{ij}-1} \, \mathrm{d}x}_{=u_1 \text{ as defined in Theorem 4.1}}.
\end{aligned}
$$

$\square$

4.1.2. Parameter identifiability

Result 4.1. *For Weibull distributed failure times $X_{ijk} \sim \mathrm{Wb}(\lambda_{ij}, \gamma_{ij})$ and drop-out times $C_{ijk} \sim \mathrm{Wb}(\mu_{ij}, \omega_{ij})$ the parameters λ_{ij}, γ_{ij}, μ_{ij} and ω_{ij} of the distribution de-*

fined by the pmf (3.2.1) and the probabilities given in Theorem 4.1 are not in general identifiable.

Proof. We will demonstrate that there are multiple parameter vectors that differ in at least one of the first 4 components $(\lambda_{ij}, \gamma_{ij}, \mu_{ij}, \omega_{ij}, q_{ij})^T \neq (\lambda'_{ij}, \gamma'_{ij}, \mu'_{ij}, \omega'_{ij}, q_{ij})^T$, but lead to the same $\boldsymbol{p}_{ij}$. We will without loss of generality consider the case with $\tau_i = 1$, because a re-scaling of the time unit leads to distributions $X_{ijk} \sim \text{Wb}(\lambda'_{ij}, \gamma_{ij})$ with $\lambda'_{ij} := \lambda_{ij}\tau_i$ and $C_{ijk} \sim \text{Wb}(\mu'_{ij}, \omega_{ij})$ with $\mu'_{ij} := \mu_{ij}\tau_i$.

The same value of p_{ij4} and p_{ij2} will result from any parameter vectors that lead to the same $u_2 := \lambda_{ij}^{\gamma_{ij}}$ and $u_3 := \mu_{ij}^{\omega_{ij}}$. Given fixed values for u_2 and u_3, the same value of p_{ij1} will result for all parameter vectors that keep

$$
\begin{aligned}
u_1 &:= \gamma_{ij}\lambda_{ij}^{\gamma_{ij}} \int_0^1 \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}\right) x^{\gamma_{ij}-1} \, \mathrm{d}x \\
&= \int_0^1 u_2 \exp\left(-u_3 s^{\omega_{ij}/\gamma_{ij}} - u_2 s\right) \, ds
\end{aligned}
\tag{4.1.3}
$$

constant. For $\omega_{ij}/\gamma_{ij} = 1$, (4.1.3) simplifies to $u_1 = u_2 \frac{1-e^{-u_2-u_3}}{u_2+u_3}$.

Thus, by noting that $p_{ij3} = (1 - q_{ij})(u_1 - p_{ij2})$ and $p_{ij5} = 1 - p_{ij4} - u_1$, any parameter vectors that satisfy $\omega_{ij}/\gamma_{ij} = 1$ and result in the same values for u_2 and u_3 lead to the same $\boldsymbol{p}_{ij}$. This is achieved for $(\lambda_{ij}, \gamma_{ij}, \mu_{ij}, \gamma_{ij}, q_{ij})^T$ and $(\lambda_{ij}^{\gamma_{ij}/\gamma'_{ij}}, \gamma'_{ij}, \mu_{ij}^{\gamma_{ij}/\gamma'_{ij}}, \gamma'_{ij}, q_{ij})^T$ for $\lambda_{ij} \neq 1$ and $\mu_{ij} \neq 1$. This demonstrates the lack of parameter identifiability, because $\boldsymbol{p}_{ij}$ uniquely determines the pmf of a multinomial distribution (3.2.1). $\qquad\square$

Example 4.1. The two parameter vectors $q_{ij} = 0.5$, $\gamma_{ij} = \omega_{ij} = 1.5$, $\lambda_{ij} = 0.1$ and $\mu_{ij} = 0.15$, and $q'_{ij} = 0.5$, $\gamma'_{ij} = \omega'_{ij} = 0.75$, $\lambda'_{ij} = 0.01$ and $\mu'_{ij} = 0.0225$ both result in the same $\boldsymbol{p}_{ij} \approx (0.0151, 0.0147, 0.0004, 0.9142, 0.0556)^T$.

Remark 4.2. It is also straightforward to demonstrate the lack of parameter identifiability in the absence of drop-outs. E.g. assuming $\tau_i = 1$, $\mu_{ij} = 0$, $\omega_{ij} = 1$ and $q_{ij} = 0.5$, any pair of parameter vectors that satisfy $\gamma_{ij} \log \lambda_{ij} = \gamma'_{ij} \log \lambda'_{ij}$ will lead to the same $\boldsymbol{p}_{ij}$. E.g. for both $\lambda_{ij} = e^{-2}$ and $\gamma_{ij} = 1$, and $\lambda'_{ij} = e^{-3}$ and $\gamma'_{ij} = 2/3$ we obtain $\boldsymbol{p}_{ij} \approx (0.0633, 0.0633, 0, 0.8734, 0)^T$.

Remark 4.3. Due to the lack of parameter identifiability, a unique maximum likelihood estimate for the trial-level parameters λ_{ij}, γ_{ij}, μ_{ij} and ω_{ij} does not typically exist, if they are assumed to be unrelated. As discussed in Section 3.3 approaches for addressing this issue include prior distributions and hierarchical models.

4.1.3. Choice of prior for the control group Weibull parameters

One popular way to specify vague priors for the parameters of a $\mathrm{Wb}(a, b)$ distribution is to assume a vague gamma prior for b, a vague normal or an improper uniform prior for $\log a$, and that $\log a$ and b are a-priori independent (Ibrahim et al., 2012, p. 35–36).

This may be a reasonable choice in the complete absence of any prior knowledge, but prior knowledge may often imply a prior dependence between b and a. In fact, prior information will often not be directly available for the parameters of multi-parameter distributions, but will rather be available about some derived quantity such as the expected proportion of patients with an event after a certain amount of follow-up. For this reason we suggest an alternative way of specifying a prior.

Let us assume that a-priori there is some plausibility to the hazard rate being constant, but considerable uncertainty about it, in which case a marginal $N(0, \sigma_\kappa)$ prior for the logarithm of the shape parameter $\kappa := \log b$ with a large value of σ_κ may be reasonable. Note that the density of this prior is centered on zero — corresponding to a constant hazard rate.

If prior knowledge otherwise comes in the form of information about the expected proportion $\varpi_{s_0} := F_{\mathrm{Wb}}(s_0; a, e^\kappa)$ of patients with an event at a fixed time s_0 in the absence of censoring, then for a given value of κ this implies a prior for $v := \log a$ via the cdf that depends on κ.

Let us assume that we have a $N(u_0, \sigma_u)$ prior for the log-odds of an event $u := \log \varpi_{s_0} - \log(\varpi_{s_0})$ by time s_0, where $u_0 := \log \varpi_0 - \log(1 - \varpi_0)$ given some $\varpi_0 \in (0, 1)$ and $\sigma_u > 0$.

We can either specify independent priors for κ and u, or alternatively specify a joint

prior for κ and $v := \log a$. The latter can be implemented using a change of variables to

$$v = \log \left\{ \log(1 + e^u) \right\} e^{-\kappa} - \log s_0$$

with the inverse function of the transform given by

$$u(v) = \log \left\{ \exp \left[\exp \left(e^\kappa (v + \log s_0) \right) \right] - 1 \right\}$$

for a given value of κ, we obtain the conditional prior density for v given κ as

$$
\begin{aligned}
p(v|\kappa) =& \phi(u(v); u_0, \sigma_u) \times \left| \frac{\mathrm{d}u(v)}{\mathrm{d}v} \right| \\
=& \frac{1}{\sqrt{2\sigma_u^2 \pi}} \exp \left(-\frac{\left(\log \left\{ \exp \left[\exp \left(e^\kappa (v + \log s_0) \right) \right] - 1 \right\} - u_0 \right)^2}{2\sigma_u^2} \right) \\
& \times \frac{\exp \left\{ \exp \left[e^\kappa (v + \log s_0) \right] + e^\kappa (v + \log s_0) + \kappa \right\}}{\left| \exp \left\{ \exp \left[e^\kappa (v + \log s_0) \right] \right\} - 1 \right|}.
\end{aligned}
$$

Thus, a prior can be specified as $p(\kappa)p(v|\kappa)$. As all of the discussed parametrizations involve one-to-one transformations of the original parameters, the parameter identifiability considerations of Section 4.1.2 remain unaffected.

4.1.4. Hierarchical model set-up and priors for the hyperparameters

In analogy to the preceding section, one way to specify a hierarchical Weibull meta-analysis model is to specify six independent random effects

$$u_{\lambda,i0} \sim N(\nu_1, \sigma_1),\ u_{\mu,i0} \sim N(\nu_3, \sigma_3), \qquad u_{\mu,i1} \sim N(\nu_5, \sigma_5),$$
$$\log \gamma_{i0} \sim N(\nu_2, \sigma_2),\ \log \omega_{i0} \sim N(\nu_4, \sigma_4) \text{ and } \log \omega_{i1} \sim N(\nu_6, \sigma_6).$$

We assume that event times in the control group follow a $\mathrm{Wb}(\lambda_{i0}, \gamma_{i0})$ distribution with $\log \lambda_{i0} = \log \left\{ \log(1 + \exp(u_{\lambda,i0})) \right\} / \gamma_{i0} - \log s_0$, control group drop-out times a $\mathrm{Wb}(\mu_{i0}, \omega_{i0})$ distribution with $\log \mu_{i0} = \log \left\{ \log(1 + \exp(u_{\mu,i0})) \right\} / \omega_{i0} - \log s_0$, test group drop-out times a $\mathrm{Wb}(\mu_{i1}, \omega_{i1})$ distribution with $\log \mu_{i1} = \log \left\{ \log(1 + \exp(u_{\mu,i1})) \right\} / \omega_{i0} - \log s_0$ and test group event times a $\mathrm{Wb}(\lambda_{i0} e^{\varphi/\gamma_{i0}}, \gamma_{i0})$ distribution.

To conduct a Bayesian hierarchical meta-analysis we also specify hyperpriors such as vague $N(0, \varsigma)$ priors for some large value of $\varsigma > 0$ for each $\nu_r \ r = 1, \ldots, 6$ and HCa$(0, 100)$ priors for $\sigma_r \ r = 1, \ldots, 6$. In the case of vague priors, it may be simplest to choose $s_0 = 1$. Alternatively, if more informative priors using available prior information are specified, s_0 would be a follow-up duration, for which prior information exists regarding the expected proportion of patients with an event or of drop-outs.

4.2. Extension to event driven trials

Uncertainty about nuisance parameters like the control group hazard leads to uncertainty about the power of a RCT. We can reduce this uncertainty by specifying that the RCT will finish when a specified amount of statistical information has accrued (Friedman et al., 2015, p. 364). In the case of time-to-event outcomes such information-based trials are called "event driven" trials, because a fixed number of patients is followed for a flexible amount of time until a specified number of patients has had an event (Friedman et al., 2015, p. 364). In this section, let τ_i denote the trial duration from start of recruitment until the specified number of patients has had an event.

The CONSORT guidelines require the reporting of the length of recruitment $\rho_i > 0$ and the end of follow-up $\tau_i > \rho_i$ (Schulz et al., 2010). Even though technically both the time to recruit all patients ρ_i and the time until the requisite number of events has accrued τ_i are r.v.s, we will treat their realizations as fixed quantities. While the observation of all participants will not cease instantaneously once the target number of patients with an event has been achieved — instead final visits are often scheduled during a close-out period of one to two months (Friedman et al., 2015, pp. 464–465) — it may still be a reasonable approximation to assume that this is the case, unless details of the close-out period are known.

By assuming a specific distribution for the recruitment times this information may allow us to model the recruitment process in addition to the time-to-event, time-to-censoring and event fatality. However, this adds an additional level of complexity compared to trials of fixed duration as we illustrate for exponentially distributed event and

drop-out times in Section 4.2.2. The same approach can be applied for more complex distributions, but we were unable to derive convenient closed form expressions. When at least $t_{ij\cdot}$ is available, modeling the recruitment process is not necessary for exponentially distributed event times as we outline in Section 4.2.1.

4.2.1. When $t_{ij\cdot}$ and $d_{ij\cdot}$ are available

When the total duration of follow-up to first event or censoring $t_{ij\cdot}$ and the duration of follow-up $d_{ij\cdot}$ are available for each arm $j = 0, 1$ of each trial $i = 1, \ldots, I$, then assuming exponential distributions for event and censoring times with $\boldsymbol{\theta}_{F,ij} = \lambda_{ij}$ and $\boldsymbol{\theta}_{G,ij} = \mu_{ij}$, we obtain

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\theta}_{F,ij}, \boldsymbol{\theta}_{G,ij}, q_{ij}|\text{IPD}) = \prod_{k=1}^{n_{ij}} & q_j^{m_{ijk}}(1 - q_j)^{y_{ijk}-m_{ijk}} \\
& \times f(t_{ijk}; \boldsymbol{\theta}_{F,ij})^{y_{ijk}}\{1 - F(t_{ijk}; \boldsymbol{\theta}_{F,ij})\}^{1-y_{ijk}} \\
& \times g(t_{ijk}; \boldsymbol{\theta}_{G,ij})^{z_{ijk}-m_{ijk}}\{1 - G(t_{ijk}; \boldsymbol{\theta}_{G,ij})\}^{1-z_{ijk}} \\
& \times \{1 - G(t_{ijk}; \boldsymbol{\theta}_{G,ij})\}^{m_{ijk}}, \\
= & q_j^{m_{ij\cdot}}(1 - q_j)^{y_{ij\cdot}-m_{ij\cdot}} \lambda_{ij}^{y_{ij\cdot}} e^{-\lambda_{ij}t_{ij\cdot}} \mu_{ij}^{z_{ij\cdot}-m_{ij\cdot}} e^{-\mu_{ij}d_{ij\cdot}}.
\end{aligned}
\tag{4.2.1}
$$

$t_{ij\cdot}$ is often available for event driven trials, if the event of interest is a primary or secondary outcome of the event driven trial.

Such outcomes are often summarized in terms of the number of patients with an event per 100 (or 1000) patient-years of follow-up to first-event or censoring. Similar summary statistics are often reported for all-cause mortality, in which case $d_{ij\cdot}$ is available.

Remark 4.4. If only $t_{ij\cdot}$, but not $d_{ij\cdot}$, is available, then one option is to ignore the information about the censoring hazard rate the event driven trial contributes. When doing so, then likelihood simplifies to $q_j^{m_{ij\cdot}}(1 - q_j)^{y_{ij\cdot}-m_{ij\cdot}} \lambda_{ij}^{y_{ij\cdot}} e^{-\lambda_{ij}t_{ij\cdot}}$.

Remark 4.5. When $t_{ij\cdot}$ and $d_{ij\cdot}$ are unavailable for an event driven trial, one option is to impute them by assuming a fixed trial duration on the basis of the assumed recruitment pattern. Assuming that the trial entry times of patients have a linearly increasing probability density function throughout the recruitment period $[0, \rho_i]$, the expected time

of recruitment for a patient is $2\rho_i/3$. Assuming that the trial entry times of patients are uniformly distributed across the recruitment period, this expected time would be $\rho_i/2$. By treating an event driven trial as a fixed duration trial as in Section 3.1.2 with duration $\tau_i - 2\rho_i/3$ (or $\tau_i - \rho_i/2$), we obtain $\tilde{t}_{ij.} \approx (n_{ij} - z_{ij.}/2 + m_{ij.}/2 - y_{ij.}/2)(\tau_i - 2\rho_i/3)$ and $\tilde{d}_{ij.} \approx (n_{ij} - z_{ij.}/2)(\tau_i - 2\rho_i/3)$. Note that this involves further approximations in addition to those already pointed out in that section, because we assume that all patients were recruited at the expected time of recruitment. We demonstrate how to avoid these approximations in Section 4.2.2 by deriving the probabilities $p_{ij1}, \ldots, p_{ij5}$ for event driven trials that can be used with the likelihood (3.2.4).

4.2.2. Using information on the recruitment period

Multi-center clinical trials recruit patients in multiple trial centers, which typically start recruitment in a staggered fashion, so that patients are not recruited in a uniform fashion and an increasing recruitment rate is common (Friedman et al., 2015, p. 220). For such a trial a linearly increasing pdf $p_i(r; \rho_i) := \frac{2r}{\rho_i^2}$ on $[0, \rho_i]$ for $\rho_i < \tau_i$ may provide a reasonable approximation to the distribution of recruitment times.

Theorem 4.2. *The probabilities of each of the outcomes of Definition 3.1 assuming a linearly increasing density of recruitment times across the recruitment period, and exponentially distributed event and drop-out times as in (3.2.3) are*

$$p_{ij1} = q_j \frac{\lambda_{ij} e^{-(\lambda_{ij}+\mu_{ij})\tau_i} \left(\rho_i^2 (\lambda_{ij}+\mu_{ij})^2 e^{\tau_i(\lambda_{ij}+\mu_{ij})} - 2e^{\rho_i(\lambda_{ij}+\mu_{ij})}(\rho_i(\lambda_{ij}+\mu_{ij})-1) - 2 \right)}{\rho_i^2(\lambda_{ij}+\mu_{ij})^3},$$

$$p_{ij2} = 2(1-q_j)\left(\frac{\left(e^{\mu_{ij}\rho_i}(\mu_{ij}\rho_i-1)+1\right)e^{-\mu_{ij}\tau_i}}{\rho_i^2\mu_{ij}^2} \right.$$
$$\left. - \frac{\left(1 + e^{\rho_i(\lambda_{ij}+\mu_{ij})}(\rho_i(\lambda_{ij}+\mu_{ij})-1)\right)e^{-\tau_i(\lambda_{ij}+\mu_{ij})}}{\rho_i^2(\lambda_{ij}+\mu_{ij})^2} \right),$$

$$p_{ij3} = \frac{2\mu_{ij}(1-q_j)}{\rho^2}\left(\frac{\rho_i^2}{2\mu_{ij}} - \frac{\left(e^{\mu_{ij}\rho_i}(\mu_{ij}\rho_i-1)+1\right)e^{-\mu_{ij}\tau_i}}{\mu_{ij}^3} \right.$$
$$\left. + \frac{\left(e^{\rho_i(\lambda_{ij}+\mu_{ij})}(\rho_i(\lambda_{ij}+\mu_{ij})-1)+1\right)e^{\tau_i(-(\lambda_{ij}+\mu_{ij}))}}{(\lambda_{ij}+\mu_{ij})^3} - \frac{\rho_i^2}{2(\lambda_{ij}+\mu_{ij})} \right),$$

$$p_{ij4} = 2e^{-(\lambda_{ij}+\mu_{ij})\tau_i}\frac{1 - e^{\rho_i(\lambda_{ij}+\mu_{ij})}\left(1 - \rho_i(\lambda_{ij}+\mu_{ij})\right)}{\rho_i^2(\lambda_{ij}+\mu_{ij})^2} \quad and$$

$$p_{ij5} = \frac{\mu_{ij}}{\lambda_{ij}+\mu_{ij}} - 2\mu_{ij}e^{-\tau_i(\lambda_{ij}+\mu_{ij})}\frac{1 - e^{\rho_i(\lambda_{ij}+\mu_{ij})}\left(1 - \rho_i(\lambda_{ij}+\mu_{ij})\right)}{\rho_i^2(\lambda_{ij}+\mu_{ij})^3}.$$

Proof of Theorem 4.2. We obtain

$$
\begin{aligned}
p_{ij1} &= q_j \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{2r}{\rho_i^2} \lambda_{ij} \mu_{ij} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(x \leq c) \mathbb{1}(x \leq \tau_i) \, \mathrm{d}x \, \mathrm{d}c \, \mathrm{d}r \\
&= \int_0^{\rho_i} \frac{2\mu_{ij} q_j r}{\rho_i^2} \left(\int_r^{\tau_i} e^{\mu_{ij}(r-c)} - e^{(r-c)(\lambda_{ij}+\mu_{ij})} \, \mathrm{d}c \right. \\
&\qquad\qquad \left. + \int_{\tau_i}^{\infty} e^{\mu_{ij}(r-c)} - e^{-c\mu_{ij}-\lambda_{ij}\tau_i+r(\lambda_{ij}+\mu_{ij})} \, \mathrm{d}c \right) \mathrm{d}r \\
&= \int_0^{\rho_i} \frac{2\mu_{ij} q_j}{\rho_i^2} r \left(\frac{1 - e^{\mu_{ij}(r-\tau_i)}}{\mu_{ij}} + \frac{e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)} - 1}{\lambda_{ij}+\mu_{ij}} + \frac{e^{\mu_{ij}(r-\tau_i)}}{\mu_{ij}} - \frac{e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)}}{\mu_{ij}} \right) \mathrm{d}r \\
&= \frac{2\mu_{ij} q_j}{\rho_i^2} \left(\frac{\rho_i^2}{2\mu_{ij}} + \frac{\left(e^{\rho_i(\lambda_{ij}+\mu_{ij})}(\rho_i(\lambda_{ij}+\mu_{ij})-1)+1\right) e^{\tau_i(-(\lambda_{ij}+\mu_{ij}))}}{(\lambda_{ij}+\mu_{ij})^3} - \frac{\rho_i^2}{2(\lambda_{ij}+\mu_{ij})} \right. \\
&\qquad\qquad \left. - \frac{\left(e^{\rho_i(\lambda_{ij}+\mu_{ij})}(\rho_i(\lambda_{ij}+\mu_{ij})-1)+1\right) e^{-\tau_i(\lambda_{ij}+\mu_{ij})}}{\mu_{ij}(\lambda_{ij}+\mu_{ij})^2} \right) \\
&= q_j \frac{\lambda_{ij} e^{-(\lambda_{ij}+\mu_{ij})\tau_i} \left(\rho_i^2(\lambda_{ij}+\mu_{ij})^2 e^{\tau_i(\lambda_{ij}+\mu_{ij})} - 2e^{\rho_i(\lambda_{ij}+\mu_{ij})}(\rho_i(\lambda_{ij}+\mu_{ij})-1) - 2\right)}{\rho_i^2(\lambda_{ij}+\mu_{ij})^3}
\end{aligned}
$$

for the probability that a patient experiences a fatal event of interest,

$$
\begin{aligned}
p_{ij2} &= (1-q_j) \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{2r}{\rho_i^2} \lambda_{ij} \mu_{ij} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(c > \tau_i) \mathbb{1}(x \leq \tau_i) \, \mathrm{d}x \, \mathrm{d}c \, \mathrm{d}r \\
&= (1-q_j) \int_0^{\rho_i} \frac{2\mu_{ij} r}{\rho_i^2} \int_{\tau_i}^{\infty} e^{\mu_{ij}(r-c)} - e^{-c\mu_{ij}-\lambda_{ij}\tau_i+r(\lambda_{ij}+\mu_{ij})} \, \mathrm{d}c \, \mathrm{d}r \\
&= (1-q_j) \int_0^{\rho_i} \frac{2r}{\rho_i^2} \left(e^{\mu_{ij}(r-\tau_i)} - e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)} \right) \mathrm{d}r \\
&= 2(1-q_j) \left(\frac{\left(e^{\mu_{ij}\rho_i}(\mu_{ij}\rho_i - 1)+1\right) e^{-\mu_{ij}\tau_i}}{\rho_i^2 \mu_{ij}^2} \right. \\
&\qquad\qquad \left. - \frac{\left(1 + e^{\rho_i(\lambda_{ij}+\mu_{ij})}(\rho_i(\lambda_{ij}+\mu_{ij})-1)\right) e^{-\tau_i(\lambda_{ij}+\mu_{ij})}}{\rho_i^2(\lambda_{ij}+\mu_{ij})^2} \right)
\end{aligned}
$$

for the probability that a patient experiences a non-fatal event and completes the trial,

$$
\begin{aligned}
p_{ij3} &= (1-q_j) \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{2r}{\rho_i^2} \lambda_{ij} \mu_{ij} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(c \leq \tau_i) \mathbb{1}(x \leq c) \, \mathrm{d}x \, \mathrm{d}c \, \mathrm{d}r \\
&= (1-q_j) \int_0^{\rho_i} \frac{2\mu_{ij} r}{\rho_i^2} \int_r^{\tau_i} e^{\mu_{ij}(r-c)} - e^{(r-c)(\lambda_{ij}+\mu_{ij})} \, \mathrm{d}c \, \mathrm{d}r \\
&= \frac{2\mu_{ij}(1-q_j)}{\rho_i^2} \int_0^{\rho_i} r \left(\frac{1 - e^{\mu_{ij}(r-\tau_i)}}{\mu_{ij}} + \frac{e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)} - 1}{\lambda_{ij}+\mu_{ij}} \right) \mathrm{d}r \\
&= \frac{2\mu_{ij}(1-q_j)}{\rho^2} \left(\frac{\rho_i^2}{2\mu_{ij}} - \frac{\left(e^{\mu_{ij}\rho_i}(\mu_{ij}\rho_i - 1)+1\right) e^{-\mu_{ij}\tau_i}}{\mu_{ij}^3} \right. \\
&\qquad\qquad \left. + \frac{\left(e^{\rho_i(\lambda_{ij}+\mu_{ij})}(\rho_i(\lambda_{ij}+\mu_{ij})-1)+1\right) e^{\tau_i(-(\lambda_{ij}+\mu_{ij}))}}{(\lambda_{ij}+\mu_{ij})^3} - \frac{\rho_i^2}{2(\lambda_{ij}+\mu_{ij})} \right)
\end{aligned}
$$

for the probability that a patient experiences a non-fatal event and then drops out before completing the trial,

$$
\begin{aligned}
p_{ij4} &= \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{2r}{\rho_i^2} \lambda_{ij} \mu_{ij} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(c > \tau_i) \mathbb{1}(x > \tau_i) \, \mathrm{d}x \, \mathrm{d}c \, \mathrm{d}r \\
&= \int_0^{\rho_i} \int_{\tau_i}^{\infty} \frac{2\mu_{ij} r e^{-c\mu_{ij} - \lambda_{ij}\tau_i + r(\lambda_{ij} + \mu_{ij})}}{\rho_i^2} \, \mathrm{d}c \, \mathrm{d}r \\
&= \int_0^{\rho_i} \frac{2r e^{(\lambda_{ij} + \mu_{ij})(r - \tau_i)}}{\rho_i^2} \, \mathrm{d}r \\
&= 2 e^{-(\lambda_{ij} + \mu_{ij})\tau_i} \frac{1 - e^{\rho_i(\lambda_{ij} + \mu_{ij})} \left(1 - \rho_i(\lambda_{ij} + \mu_{ij})\right)}{\rho_i^2 (\lambda_{ij} + \mu_{ij})^2}
\end{aligned}
$$

for the probability that a patient completes the study without an event and

$$
\begin{aligned}
p_{ij5} &= \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{2r}{\rho_i^2} \lambda_{ij} \mu_{ij} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(c \le \tau_i) \mathbb{1}(c < x) \, \mathrm{d}x \, \mathrm{d}c \, \mathrm{d}r \\
&= \int_0^{\rho_i} \int_r^{\tau_i} \frac{2\mu_{ij} r e^{(\lambda_{ij} + \mu_{ij})(r - c)}}{\rho_i^2} \, \mathrm{d}c \, \mathrm{d}r \\
&= \int_0^{\rho_i} \frac{2\mu_{ij} r \left(1 - e^{(\lambda_{ij} + \mu_{ij})(r - \tau_i)}\right)}{\rho_i^2 (\lambda_{ij} + \mu_{ij})} \, \mathrm{d}r \\
&= \frac{\mu_{ij}}{\lambda_{ij} + \mu_{ij}} - 2\mu_{ij} e^{-\tau_i(\lambda_{ij} + \mu_{ij})} \frac{1 - e^{\rho_i(\lambda_{ij} + \mu_{ij})} \left(1 - \rho_i(\lambda_{ij} + \mu_{ij})\right)}{\rho_i^2 (\lambda_{ij} + \mu_{ij})^3}
\end{aligned}
$$

for the probability that a patient is lost to follow-up before experiencing any event. $\qquad \square$

Remark 4.6. In (4.2.1) the part of the likelihood that depends on q_j, λ_{ij} and μ_{ij}, respectively, does not depend on the other parameters. Theorem 4.2 shows that the likelihood no longer factorizes in this manner, when $t_{ij\cdot}$ and $d_{ij\cdot}$ are not available.

Remark 4.7. For a single center trial recruitment times may be uniformly distributed across $[0, \rho_i]$. We provide the corresponding results under this assumption in Appendix E.

4.2.3. EMPA-REG OUTCOME example

In the event driven EMPA-REG OUTCOME trial 7020 patients were randomized, started treatment with placebo or empagliflozin between September 2010 and April 2013, and were followed until at least 691 patients had had a primary endpoint (Zinman et al., 2015). The final data collection date for the primary outcome measure was April 2015 (Boehringer Ingelheim, 2016). The AD for this trial are shown in Table 4.1.

Table 4.1.: EMPA-REG OUTCOME AD: Patient-years of follow-up until the primary endpoint, or censoring $t_{ij.}$ were derived from the reported information on $y_{ij.}/t_{ij.}$, patient-years of follow-up $d_{ij.}$ were derived from the reported number of deaths from any cause per 1000 patient-years and $z_{ij.}$ was taken to be the sum of total deaths, and the number of patients not completing the trial based on the primary results publication (Zinman et al., 2015).

Treatment group	n_{ij}	$y_{ij.}$	$t_{ij.}$	$y_{ij.}/t_{ij.}$	$d_{ij.}$	$m_{ij.}$	$z_{ij.}$
empagliflozin	4,687	490	13,102	0.0374	13,866	172	413
placebo	2,333	282	6,424	0.0439	6,783	137	261

As Table 4.2 shows there were no meaningful differences between the hazard ratios and their CIs obtained by different analysis methods for this example. While the parameter estimates for placebo hazard rate substantially deviated from the MLE when using $\tilde{t}_{ij.}$ and $\tilde{d}_{ij.}$, this happened to a similar degree in both treatment groups so that the estimated hazard ratio remained essentially unchanged. Program code for this example is available under https://doi.org/10.6084/m9.figshare.5007833.

Table 4.2.: Comparing the covariate-adjusted Cox regression reported by Zinman et al. (2015) with alternative AD maximum-likelihood analyses of the EMPA-REG OUTCOME data

Analysis	Hazard ratio (95% CI)
Covariate adjusted Cox regression	0.86 (0.74 − 0.99)
Poisson likelihood (2.3.3) with $\log t_{ij.}$ offset	0.85 (0.74 − 0.99)
Poisson likelihood (2.3.3) with $\log \tilde{t}_{ij.}$ offset[*]	0.86 (0.74 − 0.99)
Poisson likelihood (2.3.3) with $\log \tilde{t}_{ij.}$ offset[†]	0.86 (0.74 − 0.99)
Likelihood (4.2.1) using $t_{ij.}$ and $d_{ij.}$	0.85 (0.74 − 0.99)
Likelihood (4.2.1) using $\tilde{t}_{ij.}$ and $\tilde{d}_{ij.}$ [*]	0.86 (0.74 − 0.99)
Likelihood (4.2.1) using $\tilde{t}_{ij.}$ and $\tilde{d}_{ij.}$ [†]	0.86 (0.74 − 0.99)
Proposed AD likelihood (3.2.4)[*]	0.86 (0.74 − 0.99)
Proposed AD likelihood (3.2.4)[†]	0.86 (0.74 − 0.99)

[*]: assuming uniform recruitment (see Remark 4.5 and Appendix E)

[†]: assuming linearly increasing recruitment (see Remark 4.5 and Section 4.2.2)

4.3. Extension taking into account fatal events arising from recurrent events

In Section 3.2 we made some restrictive assumptions when deriving the formulae for the probabilities $p_{ij1}, \dots, p_{ij5}$ of the following five mutually exclusive outcomes of Definition 3.1. In particular, we assumed that the probability $q_{ij} := \mathbb{P}(M_{ijk} = 1 | Y_{ijk} = 1)$

that patient $k = 1, \dots, n_{ij}$ in group $j = 0, 1$ of trial $i = 1, \dots, I$ dies of an event of interest given that the patient has at least experienced one event is independent of the time when the first event occurs and the remaining follow-up duration after a first event. Since the first event of a patient is not necessarily fatal and further events may subsequently occur, we investigate the effect of the possibility of subsequent events.

4.3.1. General aggregate data likelihood with recurrent events

Let E_{ijk} be the number of events a patient would experience prior to being censored, if there were no fatal events. With this notation, we obtain

$$p_{ij1} = \sum_{r=1}^{\infty} \mathbb{P}(E_{ijk} = r \text{ and at least one of the } r \text{ events was fatal}),$$

$$p_{ij2} = \sum_{r=1}^{\infty} \mathbb{P}(E_{ijk} = r \text{ and none of the } r \text{ events was fatal and } C_{ijk} > \tau_i),$$

$$p_{ij3} = \sum_{r=1}^{\infty} \mathbb{P}(E_{ijk} = r \text{ and none of the } r \text{ events was fatal and } C_{ijk} \leq \tau_i),$$

$$p_{ij4} = \mathbb{P}(E_{ijk} = 0 \text{ and } C_{ijk} > \tau_i) \text{ and}$$

$$p_{ij5} = \mathbb{P}(E_{ijk} = 0 \text{ and } C_{ijk} \leq \tau_i).$$

If we assume that each event has an equal probability of being fatal, then whether a patient dies due to an event of interest is determined by whether at least one of e_{ijk} Bernoulli trials with probability q_j is positive. Alternatively, $M_{ijk}|E_{ijk} = e_{ijk} \sim$ $\text{Ber}(1 - (1 - q_j)^{e_{ijk}})$. Thus, assuming that event and drop-out times are independent, as well as that the drop-out times have a distribution with pdf g_{ij}, we obtain

$$p_{ij1} = \sum_{r=1}^{\infty} \mathbb{P}(E_{ijk} = r)(1 - (1 - q_j)^r)$$

$$= \sum_{r=1}^{\infty} \int_0^{\infty} g_{ij}(c) \mathbb{P}(E_{ijk} = r | C_{ijk} = c)(1 - (1 - q_j)^r) \, \mathrm{d}c,$$

$$p_{ij2} = \sum_{r=1}^{\infty} \mathbb{P}(E_{ijk} = r | C_{ijk} > \tau_i) \mathbb{P}(C_{ijk} > \tau_i)(1 - q_j)^r$$

$$= \sum_{r=1}^{\infty} \int_{\tau_i}^{\infty} g_{ij}(c) \mathbb{P}(E_{ijk} = r | C_{ijk} = c)(1 - q_j)^r \, \mathrm{d}c,$$

$$p_{ij3} = \sum_{r=1}^{\infty} \mathbb{P}(E_{ijk} = r | C_{ijk} \leq \tau_i) \mathbb{P}(C_{ijk} \leq \tau_i)(1 - q_j)^r$$

$$= \sum_{r=1}^{\infty} \int_0^{\tau_i} g_{ij}(c) \mathbb{P}(E_{ijk} = r | C_{ijk} = c)(1 - q_j)^r \, dc,$$

$$p_{ij4} = \mathbb{P}(E_{ijk} = 0 | C_{ijk} > \tau_i) \mathbb{P}(C_{ijk} > \tau_i)$$

$$= \int_{\tau_i}^{\infty} g_{ij}(c) \mathbb{P}(E_{ijk} = 0 | C_{ijk} = c) \, dc \text{ and}$$

$$p_{ij5} = \mathbb{P}(E_{ijk} = 0 | C_{ijk} \leq \tau_i) \mathbb{P}(C_{ijk} \leq \tau_i)$$

$$= \int_0^{\tau_i} g_{ij}(c) \mathbb{P}(E_{ijk} = 0 | C_{ijk} = c) \, dc.$$

4.3.2. AD likelihood with recurrent events assuming exponentially distributed waiting and drop-out times

We will now demonstrate how to derive the probabilities above assuming exponentially distributed waiting times between events. We will investigate both the situation of the same or a different exponential hazard rate before and after a first event.

The latter setting is of particular interest, if the hazard ratio of test compared with control group solely based on the time-to-first-event is of interest. This is because in the former setting, the occurrence of subsequent events would influence the estimation of the hazard rate for first events in each treatment group.

Lemma 4.1. *For $q_{ij} \in (0, 1)$, $\lambda'_{ij} \in (0, \infty)$, $\tau_i > 0$ and $x \in (0, \tau_i)$ we have*

$$\sum_{r=0}^{\infty} (1 - q_{ij})^{r+1} \frac{(\lambda'_{ij}(\tau_i - x))^r e^{-\lambda'_{ij}(\tau_i - x)}}{r!} = (1 - q_{ij}) e^{-\lambda'_{ij} q_{ij}(\tau_i - x)}.$$

Proof. We use that

$$\sum_{r=0}^{\infty} (1 - q_{ij})^r \frac{(\lambda'_{ij}(\tau_i - x))^r e^{-\lambda'_{ij}(\tau_i - x)}}{r!}$$

is by definition the moment generating function (mgf) $M_R(s) := \mathbb{E}(e^{Rs})$ of a Poisson r.v. R with rate parameter $\lambda_{ij}(\tau_i - x)$ and pmf $f(r; \lambda_{ij}(\tau_i - x)) = (\lambda_{ij}(\tau_i - x))^r e^{-\lambda_{ij}(\tau_i - x)}/r!$ evaluated at $s = \log(1 - q_{ij})$ (Johnson and Kotz, 1969, pp. 87–91). Since the mgf for such a Poisson r.v. is $M_R(s) = e^{\lambda_{ij}(\tau_i - x)(e^s - 1)}$ (Johnson and Kotz, 1969, pp. 91), we

obtain

$$\sum_{r=1}^{\infty}(1-q_{ij})^r \frac{(\lambda_{ij}(\tau_i - x))^r e^{-\lambda_{ij}(\tau_i-x)}}{r!} = \exp\{\lambda_{ij}(\tau_i - x)(e^{\log(1-q_{ij})} - 1)\} - e^{-\lambda_{ij}(\tau_i-x)}$$

$$= e^{-\lambda_{ij}q_{ij}(\tau_i-x)} - e^{-\lambda_{ij}(\tau_i-x)}.$$

(4.3.1)

$\square$

Theorem 4.3. *Let the waiting times to the first event be i.i.d. r.v.s following a* $\mathrm{Exp}(\lambda_{ij})$ *distribution with* $\lambda_{ij} \in (0,\infty)$ *for all patients* $k = 1,\ldots,n_{ij}$ *of group* $j = 0,1$ *of trial* $i = 1,\ldots,I$ *with planned duration* $\tau_i > 0$, *let the waiting times to the first event be independent from the i.i.d. waiting times between the first and subsequent events following a* $\mathrm{Exp}(\lambda'_{ij})$ *distribution with* $\lambda'_{ij} \in (0,\infty)$, *as well as the i.i.d. drop-out times* $C_{ijk} \sim \mathrm{Exp}(\mu_{ij})$ *with* $\mu_{ij} \in (0,\infty)$ *and let each event have the same probability* $q_{ij} \in (0,1)$ *of being fatal. Then we obtain the following probabilities for the 5 mutually exclusive outcomes of Definition* 3.1

$$p_{ij1} = \begin{cases} \frac{\lambda_{ij}q_{ij}}{\mu_{ij}+\lambda'_{ij}q_{ij}} \left(\frac{\lambda'_{ij}+\mu_{ij}}{\lambda_{ij}+\mu_{ij}} \left(1 - e^{-\tau_i(\lambda_{ij}+\mu_{ij})}\right) \right. \\ \qquad \left. - \frac{(1-q_{ij})\lambda'_{ij}}{\lambda'_{ij}q_{ij}-\lambda_{ij}} \left(e^{-\tau_i(\lambda_{ij}+\mu_{ij})} - e^{-\tau_i(\mu_{ij}+\lambda'_{ij}q_{ij})} \right) \right) & \text{if } \lambda'_{ij}q_{ij} \neq \lambda_{ij} \\[2ex] \frac{\lambda_{ij}q_{ij}}{\mu_{ij}+\lambda'_{ij}q_{ij}} \left(\frac{\lambda'_{ij}+\mu_{ij}}{\lambda_{ij}+\mu_{ij}} \left(1 - e^{-\tau_i(\lambda_{ij}+\mu_{ij})}\right) - (1-q_{ij})\lambda'_{ij}\tau_i e^{-\tau_i(\lambda_{ij}+\mu_{ij})} \right) & \text{if } \lambda'_{ij}q_{ij} = \lambda_{ij} \end{cases}$$

$$p_{ij2} = \begin{cases} \frac{\lambda_{ij}(1-q_{ij}) \left(e^{\tau_i(\lambda'_{ij}q_{ij}-\lambda_{ij})} - 1 \right) e^{-\tau_i(\mu_{ij}+\lambda'_{ij}q_{ij})}}{\lambda'_{ij}q_{ij}-\lambda_{ij}} & \text{if } \lambda'_{ij}q_{ij} \neq \lambda_{ij} \\[2ex] \lambda_{ij}(q_{ij}-1)\tau_i \left(-e^{-\tau_i(\lambda_{ij}+\mu_{ij})} \right) & \text{if } \lambda'_{ij}q_{ij} = \lambda_{ij} \end{cases}$$

$$p_{ij3} = \frac{\lambda_{ij}\mu_{ij}(1-q_{ij})}{\mu_{ij}+\lambda'_{ij}q_{ij}} \left(\frac{1 - e^{-\tau_i(\lambda_{ij}+\mu_{ij}+\lambda'_{ij}q_{ij})}}{\lambda_{ij}+\mu_{ij}+\lambda'_{ij}q_{ij}} + \frac{\left(1-e^{\lambda_{ij}\tau_i}\right)e^{-\tau_i(\lambda_{ij}+\mu_{ij}+\lambda'_{ij}q_{ij})}}{\lambda_{ij}} \right),$$

while p_{ij4} *and* p_{ij5} *remain the same as in* (3.2.3).

Proof. The probabilities p_{ij4} and p_{ij5} remain the same as in (3.2.3), because they only depend on the timing of the first event.

In order to derive the probabilities p_{ij1}, p_{ij2} and p_{ij3} we need to distinguish the time $\min(X_{ijk}, C_{ijk}, \tau_i)$ to the first event or censoring of a patient, during which the expo-

nential hazard rate is λ_{ij}, from the remaining time at risk $\max(0, \min(C_{ijk}, \tau_i) - X_{ijk})$ after the first event conditional on $X_{ijk} \leq \min(C_{ijk}, \tau_i)$, during which the hazard rate is λ'_{ij}. Conditional on there having been an observed first non-fatal event at time x_{ijk} and $\min(c_{ijk}, \tau_i) > x_{ijk}$, the number of additional events after the first event R_{ijk} is distributed as

$$R_{ijk} | (X_{ijk} = x_{ijk} \text{ and } C_{ijk} = c_{ijk} \text{ and } X_{ijk} \leq \min(C_{ijk}, \tau_i)) \sim \text{Poisson}(\lambda'_{ij}(\min(c_{ijk}, \tau_i) - x_{ijk}))$$

assuming that there are no fatal events.

If a patient has experienced a first event and a number $R_{ijk} = r_{ijk}$ of additional events, the probability that this patient does not die due to an event is $(1 - q_{ij})^{r_{ijk}+1}$. Taking this into account, we obtain

$$
\begin{aligned}
p_{ij2} &= \int_0^\infty \int_0^\infty \mu_{ij} e^{-\mu_{ij}c} \mathbb{1}(c \geq \tau_i) \lambda_{ij} e^{-\lambda_{ij}x} \mathbb{1}(x \leq \tau_i) \\
&\qquad \underbrace{\sum_{r=0}^\infty (1 - q_{ij})^{r+1} \frac{(\lambda'_{ij}(\tau_i - x))^r e^{-\lambda'_{ij}(\tau_i - x)}}{r!}}_{=(1-q_{ij})e^{-\lambda'_{ij}q_{ij}(\tau_i-x)} \text{ using Lemma } 4.1} \, \mathrm{d}x \, \mathrm{d}c \\
&= \int_0^{\tau_i} \lambda_{ij}(1 - q_{ij}) e^{\lambda_{ij}qx - \lambda_{ij}x - \lambda_{ij}q\tau_i - \mu_{ij}\tau_i} \, \mathrm{d}x \\
&= \begin{cases} \dfrac{\lambda_{ij}(1-q_{ij})\left(e^{\tau_i(\lambda'_{ij}q_{ij} - \lambda_{ij})} - 1\right)e^{-\tau_i(\mu_{ij}+\lambda'_{ij}q_{ij})}}{\lambda'_{ij}q_{ij} - \lambda_{ij}} & \text{if } \lambda'_{ij}q_{ij} \neq \lambda_{ij} \\[2ex] \lambda_{ij}(q_{ij} - 1)\tau_i \left(-e^{-\tau_i(\lambda_{ij}+\mu_{ij})}\right) & \text{if } \lambda'_{ij}q_{ij} = \lambda_{ij} \end{cases}.
\end{aligned}
$$

In a similar manner we obtain

$$
\begin{aligned}
p_{ij3} &= \int_0^\infty \int_0^\infty \mu_{ij}\lambda_{ij} e^{-\mu_{ij}c - \lambda_{ij}x} \mathbb{1}(x \leq c < \tau_i) \underbrace{\sum_{r=0}^\infty (1 - q_{ij})^{r+1} \frac{(\lambda'_{ij}(c - x))^r e^{-\lambda'_{ij}(c-x)}}{r!}}_{=(1-q_{ij})e^{-\lambda'_{ij}q_{ij}(c-x)} \text{ using Lemma } 4.1} \, \mathrm{d}c \, \mathrm{d}x \\
&= \frac{\lambda_{ij}\mu_{ij}(1 - q_{ij})}{\mu_{ij} + \lambda'_{ij}q_{ij}} \int_0^{\tau_i} e^{-x(\lambda_{ij}+\mu_{ij}+\lambda'_{ij}q_{ij})} - e^{-\lambda'_{ij}q_{ij}\tau_i - \mu_{ij}\tau_i - \lambda_{ij}x} \, \mathrm{d}x \\
&= \frac{\lambda_{ij}\mu_{ij}(1 - q_{ij})}{\mu_{ij} + \lambda'_{ij}q_{ij}} \left(\frac{1 - e^{-\tau_i(\lambda_{ij}+\mu_{ij}+\lambda'_{ij}q_{ij})}}{\lambda_{ij} + \mu_{ij} + \lambda'_{ij}q_{ij}} + \frac{\left(1 - e^{\lambda_{ij}\tau_i}\right) e^{-\tau_i(\lambda_{ij}+\mu_{ij}+\lambda'_{ij}q_{ij})}}{\lambda_{ij}} \right).
\end{aligned}
$$

4.3. Extension taking into account fatal events arising from recurrent events

We use

$$
\begin{aligned}
&\sum_{r=0}^{\infty}(1-(1-q_{ij})^{r+1})\frac{(\lambda'_{ij}(\min(c,\tau_i)-x))^r e^{-\lambda'_{ij}(\min(c,\tau_i)-x)}}{r!} \\
&=\underbrace{\sum_{r=0}^{\infty}\frac{(\lambda'_{ij}(\min(c,\tau_i)-x))^r e^{-\lambda'_{ij}(\min(c,\tau_i)-x)}}{r!}}_{=1\ \text{(Poisson pmf summed over all outcomes)}} \\
&\quad-(1-q_{ij})\underbrace{\sum_{r=0}^{\infty}(1-q_{ij})^r\frac{(\lambda'_{ij}(\min(c,\tau_i)-x))^r e^{-\lambda'_{ij}(\min(c,\tau_i)-x)}}{r!}}_{=e^{-\lambda'_{ij}q_{ij}(\min(c,t)-x)}\ \text{(definition of the mgf of a Poisson r.v.)}} \\
&=1-(1-q_{ij})e^{-\lambda'_{ij}q_{ij}(\min(c,t)-x)}
\end{aligned}
\tag{4.3.2}
$$

to derive that

$$
\begin{aligned}
p_{ij1}&=\int_0^{\infty}\int_0^{\infty}\mu_{ij}e^{-\mu_{ij}c}\lambda_{ij}e^{-\lambda_{ij}x}\mathbb{1}(x\le\min(c,\tau_i)) \\
&\quad\times\underbrace{\sum_{r=0}^{\infty}\left(1-(1-q_{ij})^{r+1}\right)\frac{(\lambda'_{ij}(\min(c,\tau_i)-x))^r e^{-\lambda_{ij}(\min(c,\tau_i)-x)}}{r!}}_{=1-(1-q_{ij})e^{-\lambda'_{ij}q_{ij}(\min(c,t)-x)}\ \text{using (4.3.2)}}\,\mathrm{d}c\,\mathrm{d}x \\
&=\frac{\lambda_{ij}q_{ij}}{\mu_{ij}+\lambda'_{ij}q_{ij}}\int_0^{\tau_i}(\lambda'_{ij}+\mu_{ij})e^{-x(\lambda_{ij}+\mu_{ij})}-(1-q_{ij})\lambda'_{ij}e^{\lambda'_{ij}q_{ij}(x-\tau_i)-\mu_{ij}\tau_i-\lambda_{ij}x}\,\mathrm{d}x \\
&=\begin{cases}
\frac{\lambda_{ij}q_{ij}}{\mu_{ij}+\lambda'_{ij}q_{ij}}\left(\frac{\lambda'_{ij}+\mu_{ij}}{\lambda_{ij}+\mu_{ij}}\left(1-e^{-\tau_i(\lambda_{ij}+\mu_{ij})}\right)\right. \\
\qquad\left.-\frac{(1-q_{ij})\lambda'_{ij}}{\lambda'_{ij}q_{ij}-\lambda_{ij}}\left(e^{-\tau_i(\lambda_{ij}+\mu_{ij})}-e^{-\tau_i(\mu_{ij}+\lambda'_{ij}q_{ij})}\right)\right) & \text{if }\lambda'_{ij}q_{ij}\ne\lambda_{ij} \\[2ex]
\frac{\lambda_{ij}q_{ij}}{\mu_{ij}+\lambda'_{ij}q_{ij}}\left(\frac{\lambda'_{ij}+\mu_{ij}}{\lambda_{ij}+\mu_{ij}}\left(1-e^{-\tau_i(\lambda_{ij}+\mu_{ij})}\right)-(1-q_{ij})\lambda'_{ij}\tau_i e^{-\tau_i(\lambda_{ij}+\mu_{ij})}\right) & \text{if }\lambda'_{ij}q_{ij}=\lambda_{ij}.
\end{cases}
\end{aligned}
$$

$\square$

Remark 4.8. As long as some patients had at least one event and some patients had no event, we have information to estimate the parameter λ_{ij}. However, λ'_{ij} and q_{ij} may be poorly identified, because the likelihood given an observed number of fatal events may be the same or similar for different combinations of λ'_{ij} and q_{ij}. I.e. it may be hard to distinguish a scenario with a low rate of events that are frequently fatal and a scenario with a high rate of events that are more rarely fatal. This can be resolved by IPD from some trials (see Section 4.3.3), by trials having substantially different duration, or by having informative prior information on q_{ij}, λ'_{ij} or the ratio of λ'_{ij} to λ_{ij}.

Corollary 4.1. *Let the waiting times to the first event and between subsequent events be i.i.d. r.v.s following a* $\mathrm{Exp}(\lambda_{ij})$ *distribution with* $\lambda_{ij} \in (0, \infty)$ *for all patients* $k = 1, \ldots, n_{ij}$ *of group* $j = 0, 1$ *of trial* $i = 1, \ldots, I$ *of planned duration* $\tau_i > 0$, *let the waiting times be independent from the i.i.d. drop-out times* $C_{ijk} \sim \mathrm{Exp}(\mu_{ij})$ *with* $\mu_{ij} \in (0, \infty)$ *and let each event have the same probability* $q_{ij} \in (0, 1)$ *of being fatal. Then we obtain*

$$p_{ij1} = \frac{\lambda_{ij}q_{ij}\left(1 - e^{-\tau_i(\mu_{ij} + \lambda_{ij}q_{ij})}\right)}{\mu_{ij} + \lambda_{ij}q_{ij}},$$

$$p_{ij2} = e^{-\tau_i(\mu_{ij} + \lambda_{ij}q_{ij})} - e^{-\tau_i(\lambda_{ij} + \mu_{ij})} \qquad and$$

$$p_{ij3} = \frac{\lambda_{ij}\mu_{ij}(1 - q_{ij})}{\mu_{ij} + \lambda_{ij}q_{ij}}\left(\frac{1 - e^{-\tau_i(\lambda_{ij} + \mu_{ij} + \lambda_{ij}q_{ij})}}{\lambda_{ij} + \mu_{ij} + \lambda_{ij}q_{ij}} + \frac{\left(1 - e^{\lambda_{ij}\tau_i}\right)e^{-\tau_i(\lambda_{ij} + \mu_{ij} + \lambda_{ij}q_{ij})}}{\lambda_{ij}}\right)$$

for the probabilities of the 5 mutually exclusive outcomes of Definition 3.1, while p_{ij4} *and* p_{ij5} *remain the same as in* (3.2.3).

Proof. By setting $\lambda'_{ij} = \lambda_{ij}$ in Theorem 4.3 and noting that for $q_{ij} \in (0, 1)$ we only need to consider the case with $q_{ij}\lambda_{ij} \neq \lambda_{ij}$, we obtain Corollary 4.1. $\qquad\square$

4.3.3. IPD likelihood with recurrent events assuming exponentially distributed waiting and drop-out times

Let t_{ijkr} for $r = 1, \ldots, e_{ijk} + 1$ be the waiting times until the e_{ijk} events or until censoring for patient $k = 1, \ldots, n_{ij}$ in group $j = 0, 1$ of trial $i = 1, \ldots, I$. Note that $t_{ijk1} = t_{ijk}$, that $t_{ijk(e_{ijk}+1)}$ is always a time to censoring and that $t_{ijk(e_{ijk}+1)} = 0$, if the last event was fatal, or occurred at time $\min(c_{ijk}, \tau_i)$. With this notation the IPD likelihood corresponding to the AD model with a different exponential hazard rate before and after the first event is

$$\mathcal{L}(\lambda_{ij}, \lambda'_{ij}, \mu_{ij}, q_{ij}|\mathcal{D}) = \prod_{j=0}^{1}\prod_{i=1}^{I}(1 - q_{ij})^{e_{ij\cdot} - m_{ij\cdot}}q_{ij}^{m_{ij\cdot}}\mu_{ij}^{z_{ij\cdot} - m_{ij\cdot}}e^{\mu_{ij}d_{ij\cdot}}$$

$$\lambda_{ij}^{y_{ij\cdot}}e^{\lambda_{ij}t_{ij\cdot}}\lambda'^{e_{ij\cdot} - y_{ij\cdot}}_{ij}e^{\lambda'_{ij}(d_{ij\cdot} - t_{ij\cdot})}$$

$$= \prod_{j=0}^{1}(1 - q_j)^{e_{\cdot j\cdot} - m_{\cdot j\cdot}}q_j^{m_{\cdot j\cdot}}\prod_{i=1}^{I}\mu_{ij}^{z_{ij\cdot} - m_{ij\cdot}}e^{\mu_{ij}d_{ij\cdot}}$$

$$\lambda_{ij}^{y_{ij\cdot}}e^{\lambda_{ij}t_{ij\cdot}}\lambda'^{e_{ij\cdot} - y_{ij\cdot}}_{ij}e^{\lambda'_{ij}(d_{ij\cdot} - t_{ij\cdot})}, \text{ if } q_{ij} = q_j.$$

Note that if the AD on $y_{ij\cdot}$, $e_{ij\cdot}$, $m_{ij\cdot}$, $z_{ij\cdot}$, $d_{ij\cdot}$ and $t_{ij\cdot}$ are available, this likelihood can also be considered an AD likelihood.

If $\lambda'_{ij} = \lambda_{ij}$, the likelihood simplifies to

$$
\mathcal{L}(\lambda_{ij}, \mu_{ij}, q_{ij} | \mathcal{D}) = \prod_{j=0}^{1} \prod_{i=1}^{I} (1 - q_{ij})^{e_{ij\cdot} - m_{ij\cdot}} q_{ij}^{m_{ij\cdot}} \mu_{ij}^{z_{ij\cdot} - m_{ij\cdot}} e^{\mu_{ij} d_{ij\cdot}} \lambda_{ij}^{e_{ij\cdot}} e^{\lambda_{ij} d_{ij\cdot}}.
$$

$$
= \prod_{j=0}^{1} (1 - q_j)^{e_{\cdot j\cdot} - m_{\cdot j\cdot}} q_j^{m_{\cdot j\cdot}} \prod_{i=1}^{I} \mu_{ij}^{z_{ij\cdot} - m_{ij\cdot}} e^{\mu_{ij} d_{ij\cdot}} \lambda_{ij}^{e_{ij\cdot}} e^{\lambda_{ij} d_{ij\cdot}}, \text{ if } q_{ij} = q_j.
$$

5. Conjugate updating for exponential time-to-event regression models

Definition 5.1 (Conjugate family of prior distributions)**.** A conjugate family of prior distributions for a statistical model results in a posterior distribution from the same family of distributions as the prior distribution (Robert, 2007, p. 114). We will only refer to distributions as conjugate, if they are parametrized in terms of a fixed finite number of parameters.

Conjugate priors result in tractable posterior inference as Theorem 5.1 illustrates.

Theorem 5.1. *A conjugate prior density $p(\boldsymbol{\theta}; \boldsymbol{\vartheta})$ for a exponential family distribution with parameter vector $\boldsymbol{\theta}$ in canonical form and sufficient statistic $\boldsymbol{T}$ of dimension d can be parametrized in terms of a d dimensional parameter vector $\boldsymbol{\vartheta}$ so that the posterior distribution $p(\boldsymbol{\theta}|\boldsymbol{T} = \boldsymbol{t})$ after observing a realization $\boldsymbol{t}$ of the sufficient statistic is $p(\boldsymbol{\theta}; \boldsymbol{\vartheta} + \boldsymbol{t})$ (Bernardo and Smith, 2000, pp. 265–269).*

There are a number of reasons why (conditionally) conjugate priors remain important despite the availability of fast implementations of general purpose MCMC methods and continuing increases in computing power. Firstly, we typically need to analytically summarize the MAP distribution obtained e.g. using MCMC methods from a previous analysis in order to use it as the prior for a future analysis. For this purpose mixtures of (conditionally) conjugate priors are one possible choice that simplifies subsequent analyses. Secondly, (mixtures of) conjugate priors allow for analytic solutions and (mixtures of) conditionally conjugate priors allow for Gibbs sampling. While such

approaches are no longer required for conducting Bayesian analyses in practice, there are applications such as simulation studies to evaluate Bayesian methods, for which computational efficiency remains important.

For analyzing exponentially distributed event times with right censoring using a "Poisson likelihood", (Zhu et al., 2016) have proposed separate conjugate gamma priors for test and control group hazard rates.

However, this proposal has several limitations. Firstly, conjugate gamma priors constitute a limited class of distributions that may not be able to reflect the available prior information. For example, we often wish to use wide priors symmetric around 0 for the log-hazard ratio, but it is unclear how to construct such a prior when there is considerable prior information about the control group hazard rate and substantially less information about a new test intervention. Secondly, when we use informative priors e.g. for the control group hazard rate, conjugate priors are not robust against prior-data conflicts. Thirdly, independent priors for the test group hazard rates in multiple trials cannot capture that in a meta-analysis a similar effect of the test intervention compared with control groups across trials would be expected.

For many generalized linear models in the exponential family there is no closed form conjugate prior (Chen and Ibrahim, 2003). However, for comparing exponentially distributed failure times with independent right censoring between two treatment groups in a RCT, a closed form analytic solution is available, as we will show. In the case of meta-analyzing data from multiple trials, the normalizing constant for the prior and posterior densities needs to be evaluated using one-dimensional numeric integration.

5.1. Conjugate prior and updating rule for a single trial

Using the results of Chen and Ibrahim (2003) for constructing conjugate priors in generalized linear regression models for exponential family distributions, we obtain the conjugate prior

$$p(\lambda_0, \phi) \propto \lambda_0^{a_0} e^{-\lambda b_0} (\lambda_0 \phi)^{a_1} e^{-\lambda_0 \phi b_1}$$

with constants $a_0 > 0$, $b_0 > 0$, $a_1 > -1$ and $b_1 > 0$ for an exponential time-to-event model under independent right-censoring comparing a control group with hazard rate λ_0 versus a test group with hazard rate $\lambda_1 = \phi\lambda_0$, where ϕ is the hazard ratio. We obtain

$$
\begin{aligned}
\gamma^{-1} &= \int_0^\infty \int_0^\infty \lambda_0^{a_0} e^{-\lambda_0 b_0} (\lambda_0\phi)^{a_1} e^{-\lambda_0 \phi b_1} \, \mathrm{d}\lambda_0 \, \mathrm{d}\phi \\
&= \frac{\Gamma(a_0)\Gamma(a_1 + 1)}{b_0^{a_0} b_1^{a_1 + 1}}
\end{aligned}
$$

for the normalizing constant γ of the prior and thus, the prior density is given by

$$
p(\lambda_0, \phi) = \frac{b_0^{a_0} b_1^{a_1 + 1}}{\Gamma(a_0)\Gamma(a_1 + 1)} \lambda_0^{a_0} e^{-\lambda_0 b_0} (\lambda_0\phi)^{a_1} e^{-\lambda_0 \phi b_1}. \tag{5.1.1}
$$

Theorem 5.2. (5.1.1) *is a conjugate prior for* $\mathrm{Exp}(\lambda_0\phi^j)$ *distributed failure times from two groups* $j = 0, 1$ *with independent right-censoring.*

Proof. This is a conjugate prior when an exponential time-to-event model is parametrized in terms of control group rate λ_0 and hazard ratio ϕ, because given observed data consisting of y_0 patients with an event in a control group during t_0 time units of follow-up to first event or censoring, and y_1 patients with an event in the test group during t_1 time units of follow-up to first event or censoring, we obtain

$$
\begin{aligned}
p(\lambda_0, \phi | y_0, t_0, y_1, t_1) &\propto p(\lambda_0, \phi) \times \mathcal{L}(\lambda_0, \phi | y_0, t_0, y_1, t_1) \\
&= \frac{\lambda_0^{a_0} b_0^{a_0} b_1^{a_1 + 1} (\lambda_0\phi)^{a_1} e^{-\beta_0 \lambda_0 - \beta_1 \lambda_0 \phi}}{\Gamma(a_0)\Gamma(a_1 + 1)} \times \lambda_0^{y_0} e^{-\lambda_0 t_0} (\lambda_0\phi)^{y_1} e^{-\lambda_0 t_1 \phi} \\
&= \frac{b_0^{a_0} b_1^{a_1 + 1}}{\Gamma(a_0)\Gamma(a_1 + 1)} \lambda_0^{a_0 + y_0} e^{-\lambda_0(\beta_0 + t_0)} (\lambda_0\phi)^{a_1 + y_1} e^{-\lambda_0 \phi(b_1 + t_1)}.
\end{aligned}
$$
$$\tag{5.1.2}$$

This expression integrates to

$$
\begin{aligned}
&\frac{b_0^{a_0} b_1^{a_1 + 1}}{\Gamma(a_0)\Gamma(a_1 + 1)} \int_0^\infty \int_0^\infty \lambda_0^{a_0 + y_0} e^{-\lambda_0(\beta_0 + t_0)} (\lambda_0\phi)^{a_1 + y_1} e^{-\lambda_0 \phi(b_1 + t_1)} \, \mathrm{d}\lambda_0 \, \mathrm{d}\phi \\
&= \frac{b_0^{a_0} b_1^{a_1 + 1} \Gamma(a_0 + y_0)\Gamma(a_1 + y_1 + 1)}{(b_0 + t_0)^{a_0 + y_0} (b_1 + t_1)^{a_1 + y_1 + 1} \Gamma(a_0)\Gamma(a_1 + 1)}
\end{aligned}
$$

so that the posterior is given by

$$p(\lambda_0, \phi | y_0, t_0, y_1, t_1) = \frac{(b_0 + t_0)^{a_0 + y_0} (b_1 + t_1)^{a_1 + y_1 + 1} \Gamma(a_0) \Gamma(a_1 + 1)}{\Gamma(a_0 + y_0) \Gamma(a_1 + y_1 + 1) b_0^{a_0} b_1^{a_1 + 1}}$$

$$\times \frac{b_0^{a_0} b_1^{a_1 + 1}}{\Gamma(a_0) \Gamma(a_1 + 1)} \lambda_0^{a_0 + y_0} e^{-\lambda_0 (\beta_0 + t_0)} (\lambda_0 \phi)^{a_1 + y_1} e^{-\lambda_0 \phi (b_1 + t_1)}$$

$$= \frac{(b_0 + t_0)^{a_0 + y_0} (b_1 + t_1)^{a_1 + y_1 + 1}}{\Gamma(a_0 + y_0) \Gamma(a_1 + y_1 + 1)} \lambda_0^{a_0 + y_0} e^{-\lambda_0 (\beta_0 + t_0)} (\lambda_0 \phi)^{a_1 + y_1} e^{-\lambda_0 \phi (b_1 + t_1)}.$$

This is the same distribution as the conjugate prior, just with parameters $a_0 + y_0$, $b_0 + t_0$, $a_1 + y_1$ and $b_1 + t_1$. $\qquad\square$

5.2. Properties of the proposed conjugate prior and its posterior

Proposition 5.1. *Using the conjugate prior 5.1.1 is equivalent to using two independent conjugate priors for the control group hazard rate λ_0 and the test group hazard rate λ_1.*

Proof. The conjugate prior for a single rate in an exponential time-to-event model with random right-censoring is a gamma distribution (Ibrahim et al., 2001, pp. 30–32). If we assume the two priors to be independent, we thus obtain

$$p(\lambda_0, \lambda_1) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \lambda_0^{\alpha_0 - 1} e^{-\beta_0 \lambda_0} \times \frac{\beta_1^{\alpha_1}}{\Gamma(\alpha_1)} \lambda_1^{\alpha_1 - 1} e^{-\beta_1 \lambda_1} \tag{5.2.1}$$

with shape parameter $\alpha_0 > 0$, rate parameter $\beta_0 > 0$, shape parameter $\alpha_1 > 0$ and rate parameter $\beta_1 > 0$. We obtain the posterior by independently updating the two priors after observing y_0 patients with an event in a control group during t_0 time units of follow-up to first event or censoring, and y_1 patients with an event in the test group during t_1 time units using the updating rule for the conjugate prior for a single rate (Ibrahim et al., 2001, pp. 30–32) as

$$p(\lambda_0, \lambda_1 | y_0, t_0, y_1, t_1) = \frac{(\beta_0 + t_0)^{\alpha_0 + y_0}}{\Gamma(\alpha_0 + y_0)} \lambda_0^{\alpha_0 + y_0 - 1} e^{-(\beta_0 + t_0)\lambda_0}$$
$$\frac{(\beta_1 + t_1)^{\alpha_1 + y_1}}{\Gamma(\alpha_1 + y_1)} \lambda_1^{\alpha_1 + y_1 - 1} e^{-(\beta_1 + t_1)\lambda_1}. \tag{5.2.2}$$

If we conduct a change of variables for the prior (5.2.1), we obtain

$$p(\lambda_0, \phi) = \frac{\beta_0^{\alpha_0} \beta_1^{\alpha_1}}{\Gamma(\alpha_0)\Gamma(\alpha_1)} \lambda_0^{\alpha_0} (\lambda_0 \phi)^{\alpha_1 - 1} e^{-\lambda_0(\beta_0 + \beta_1 \phi)},$$

which for $\alpha_0 = a_0$, $\alpha_1 = a_1 + 1$, $\beta_0 = b_0$ and $\beta_1 = b_1$ is exactly the prior (5.1.1). Similarly, conducting a change of variables for the posterior (5.2.2) we obtain

$$
\begin{aligned}
p(\lambda_0, \phi | y_0, t_0, y_1, t_1) = & \frac{(\beta_0 + t_0)^{\alpha_0 + y_0}}{\Gamma(\alpha_0 + y_0)} \lambda_0^{\alpha_0 + y_0 - 1} e^{-(\beta_0 + t_0)\lambda_0} \\
& \times \frac{(\beta_1 + t_1)^{\alpha_1 + y_1}}{\Gamma(\alpha_1 + y_1)} (\lambda_0 \phi)^{\alpha_1 + y_1 - 1} e^{-(\beta_1 + t_1)\lambda_0 \phi} \times |\lambda_0|,
\end{aligned}
$$

which for $\alpha_0 = a_0$, $\alpha_1 = a_1 + 1$, $\beta_0 = b_0$ and $\beta_1 = b_1$ is exactly the posterior (5.1.2). $\square$

Remark 5.1. The conjugate priors and posteriors for the two parametrizations are identical in terms of the implied prior beliefs about λ_0 and ϕ. We therefore parametrize the model in terms of the parameters of interest to avoid the single change of variables that would otherwise be necessary. However, once a meta-analysis of multiple trials $i = 1, \ldots, I$ is being conducted, a parametrization in terms of independent priors for the λ_{ij} is no longer feasible, if we wish to respect the constraint $\lambda_{i1} = \phi \lambda_{i0}$ for all $i = 1, \ldots, I$. We will derive the conjugate prior for the parametrization of this problem in terms of the control group hazard rates $\lambda_{10}, \ldots, \lambda_{I0}$ and the hazard ratio ϕ in subsection 5.5.

Remark 5.2. Note that the marginal prior for λ_0

$$
\begin{aligned}
p(\lambda_0) &= \int_0^\infty \frac{b_0^{a_0} b_1^{a_1 + 1}}{\Gamma(a_0)\Gamma(a_1 + 1)} \lambda_0^{a_0} e^{-\lambda b_0} (\lambda_0 \phi)^{a_1} e^{-\lambda \phi b_1} \, \mathrm{d}\phi \\
&= \frac{b_0^{a_0}}{\Gamma(a_0)} \lambda_0^{a_0 - 1} e^{-b_0 \lambda_0}
\end{aligned}
$$

is a gamma distribution with parameters a_0 and b_0 irrespective of the values for $a_1 > -1$ and $b_1 > 0$. It has prior mean $(a_0 - 1)/b_0$ for $a_0 \geq 1$ and the prior mean for $\log \lambda_0$ is $\psi(a_0) - \log(b_0)$ Johnson et al. (1994, pp. 337–349).

The implied marginal prior $p(\phi)$ for the hazard ratio is

$$p(\phi) = \frac{\Gamma(a_0 + a_1 + 1) b_0^{a_0} b_1 (b_1 \phi)^{a_1}}{\Gamma(a_0)\Gamma(a_1 + 1)(b_0 + b_1 \phi)^{a_0 + a_1 + 1}} \tag{5.2.3}$$

with mode $\arg\max p(\phi) = \frac{a_1 b_0}{b_1(a_0+1)}$. This can be seen by solving

$$\frac{\mathrm{d}}{\mathrm{d}\phi}\log p(\phi) = \frac{a_1}{\phi} - \frac{b_1(a_0 + a_1 + 1)}{b_0 + \phi b_1} = 0$$

for ϕ. The extremum constitutes a maximum, because $\frac{\mathrm{d}^2}{\mathrm{d}\phi^2}\log p(\phi) = (a_0+1)^2 b_1^2/b_0^2(1/(a_0 + a_1 + 1) - 1/a_1)$ is negative, if $a_0 > -1$, $a_1 > 0$, $b_0 > 0$ and $b_1 > 0$. The mean of the marginal prior is

$$\int_0^\infty \phi p(\phi)\,\mathrm{d}\phi = \frac{b_0(a_1 + 1)}{b_1(a_0 - 1)} \tag{5.2.4}$$

for $a_0 > 1$.

Corollary 5.1. *Given a prior of the form* (5.1.1) *the marginal conjugate prior for the log-hazard ratio* $\varphi := \log\phi$ *is*

$$p(\varphi) = e^\varphi \frac{\Gamma(a_0 + a_1 + 1)b_0^{a_0}b_1(b_1 e^\varphi)^{a_1}}{\Gamma(a_0)\Gamma(a_1 + 1)(b_0 + b_1 e^\varphi)^{a_0+a_1+1}}. \tag{5.2.5}$$

The prior mode of the log-hazard ratio φ *is* $\arg\max p(\varphi) = \log\left(\frac{b_0(a_1+1)}{a_0 b_1}\right)$, *the prior mean is*

$$\mathbb{E}\varphi = \psi(a_0) - \psi(a_1 + 1) + \log(b_1/b_0), \tag{5.2.6}$$

and the variance is $Var(\varphi) = \psi_1(a_0) + \psi_1(a_1 + 1)$, *where* $\psi_1(x) := \frac{\mathrm{d}\psi(x)}{\mathrm{d}x}$ *denotes the trigamma function. For large values of* a_1 *and* a_0 $Var(\varphi) \approx \frac{a_0+a_1+1}{(a_1+1)a_0} \approx \frac{a_0+a_1}{a_1 a_0}$.

Proof. After a change of variables in (5.2.3) to the log-hazard ratio $\varphi := \log\phi$, we obtain (5.2.5) from

$$p(\lambda_0, \varphi) = p(\lambda_0, \phi = e^\varphi)\left|\frac{\mathrm{d}e^\varphi}{\mathrm{d}\varphi}\right|.$$

The first derivative of the logarithm of (5.2.5) is

$$\frac{-a_0 b_1 e^\varphi + a_1 b_0 + b_0}{b_0 + b_1 e^\varphi},$$

which is zero when $-a_0 b_1 e^\varphi + a_1 b_0 + b_0 = 0$. This is the case for $\varphi = \log\left(\frac{b_0(a_1+1)}{a_0 b_1}\right)$ and the second derivative evaluated at this point is $-a_0(a_1 + 1)/(a_0 + a_1 + 1) < 0$, while $p(\varphi) \to 0$ for $\varphi \to \infty$ or $\varphi \to -\infty$, so that the point is the global maximum of $p(\varphi)$.

Using that $\varphi := \log \lambda_1 - \log \lambda_0$, Proposition 5.1 and the properties of the logarithm of gamma distributed r.v.s given by Johnson et al. (1994, pp. 337–349), we obtain

$$\mathbb{E}\varphi := \underbrace{\mathbb{E}(\log \lambda_1)}_{\psi(a_0) - \log(b_0)} - \underbrace{\mathbb{E}(\log \lambda_0)}_{\psi(a_1 + 1) - \log(b_1)}$$

$$= \psi(a_0) - \psi(a_1 + 1) + \log(b_1/b_0)$$

and

$$\mathrm{Var}(\varphi) = \mathrm{Var}(\log \lambda_1) + \mathrm{Var}(\log \lambda_0)$$
$$= \psi_1(a_1 + 1) + \psi_1(a_0). \tag{5.2.7}$$

For large values of a_1 and a_0 (5.2.7) can be approximated by $\frac{1}{a_1+1} + \frac{1}{a_0} = \frac{a_0+a_1+1}{(a_1+1)a_0} \approx \frac{a_0+a_1}{a_1 a_0}$ Johnson et al. (1994, p. 347). $\qquad \square$

5.3. Construction of a conjugate prior capturing desired prior information

With (5.2.1) and (5.2.2) in mind, the prior information implied by the prior (5.1.1) can be interpreted to being equivalent to having observed historic data having the same parameters λ_0 and λ_1, in which a_0 control group patients had an event during b_0 time units of follow-up to first event or censoring and $a_1 + 1$ test group patients had an event during b_1 time units of follow-up to first event or censoring. If prior information is available in this form, then the specification of the conjugate prior is straightforward.

However, in practice the prior belief about the control group hazard rate λ_0 may not be available in this form and may not be well approximated by a single gamma distribution. If a $\mathrm{Gamma}(a_0, b_0)$ distribution captures the prior information about λ_0, one could fix the prior mean — or mode as suggested by Chen and Ibrahim (2003) — to correspond to a hazard ratio of 1 by setting $a_1 = b_1/b_0(a_0 - 1) - 1$ for any choice of b_1 that satisfies $b_1 > b_0/(a_0 - 1)$ based on (5.2.4). However, the resulting marginal prior density $p(\varphi)$ will in general not be symmetric (Chen and Ibrahim, 2003) and will not correspond

to typically recommended default priors for regression coefficients such as a uniform distribution, a normal distribution, a t-distribution, or a Cauchy distribution.

We can use a mixture of conjugate priors to more closely approximate desired marginal priors for λ_0 and φ. One way of constructing such a mixture prior is to first construct a mixture prior that approximates the marginal prior information on the control group hazard rate λ_0. This can be done irrespective of our prior beliefs about λ_1, because the marginal prior density $p_r(\lambda_0; a_{r,0}, b_{r,0})$ of each of these mixture components $r = 1, \ldots, R$ will not depend on the choice of $a_{r,1}$ and $b_{r,1}$. Using a mixture prior also gives us the option of making the prior more robust against prior-data conflicts by adding a weakly informative component to the mixture. This could be a unit information prior that has the same prior control group mean log-hazard rate as the informative part of the prior by choosing $a_{R+1,0} = 1$ and $b_{R+1,0} = \exp\left(\psi(1) - \sum_{r=1}^{R} w_r \left(\log(b_{r,0}) - \psi(a_{r,0})\right)\right)$, where the $w_r \in (0,1)$ with $\sum_{r=1}^{R} w_r = 1$ are the mixture weights of each component. After fixing $a_{r,0}$ and $b_{r,0}$ for each mixture component $r = 1, \ldots, R$, we then choose $a_{r,1}$ and $b_{r,1}$ for each mixture component so that the mixture of marginal distributions $p_r(\varphi)$ for the log-hazard ratio approximates the desired marginal prior for the log-hazard ratio.

If no mixture is needed to represent the prior information on λ_0, then R mixture components with identical values for $a_{r,0}$ and $b_{r,0}$ can be used — each with the same marginal prior for λ_0. These mixture components can then differ in terms of the values of $a_{r,1}$ and $b_{r,1}$ to represent the prior information on the hazard ratio. In analogy to kernel density estimation we can choose these values in a somewhat ad-hoc fashion by first choosing quantiles $\xi_1, \ldots, \xi_R$ of the desired prior for φ that correspond to equally spaced percentiles, and choosing $a_{r,1}$ and $b_{r,1}$ for each mixture component so that $p_r(\varphi)$ has a mean of ξ_r and a standard deviation approximately proportional to the distance to the adjacent quantiles ξ_{r-1} and ξ_{r+1}. Thus, $b_{r,1} = \exp\left(\psi(a_{r,1} + 1) - \psi(a_{0,r}) + \log b_{r,0} + \xi_r\right)$ based on (5.2.6) and

$$
a_{r,1} = \begin{cases} 0 & \text{for } (\xi_{r+1} - \xi_{r-1})^2 a_{r,0} = 1 \\ \frac{a_{r,0}}{(\xi_{r+1} - \xi_{r-1})^2 a_{r,0} - 1} & \text{for } (\xi_{r+1} - \xi_{r-1})^2 a_{r,0} \neq 1 \end{cases},
$$

where $\xi_0 := 2\xi_1 - \xi_2$ and $\xi_{R+1} := 2\xi_R - \xi_{R-1}$. We use $\mathrm{Var}(\varphi) \approx \frac{a_0 + a_1}{a_1 a_0}$ instead of $\frac{a_0 + a_1 + 1}{(a_1 + 1)a_0}$,

because it reduces numerical issues when $b_{r,1}$ is very close to zero in the tail area of the distribution. We assign all mixture components the same prior weights $w_r = 1/R$.

If a mixture with R_1 components is needed to represent the marginal prior information on λ_0, then the same process with R_2 mixture components for each of the R_1 mixtures can be applied to obtain a mixture with $R = R_1 \times R_2$ components.

Given a mixture of conjugate priors with R components

$$p(\lambda_0, \phi) = \sum_{r=1}^{R} w_r p_r(\lambda_0, \phi; a_{r,0}, b_{r,0}, a_{r,1}, b_{r,1})$$

that closely approximates any desired marginal priors for λ_0 and ϕ, the posterior is obtained as (Diaconis and Ylvisaker, 1985)

$$p(\lambda_0, \phi | y_0, t_0, y_1, t_1) = \sum_{r=1}^{R} w'_r p_r(\lambda_0, \phi; a_{r,0} + y_0, b_{r,0} + t_0, a_{r,1} + y_1, b_{r,1} + t_1).$$

The updated posterior mixture weights for a mixture of such priors is obtained based on the consistency of the observed data with the prior predictive distribution as

$$w'_r := \frac{w_r}{w'} \int_0^\infty \int_0^\infty p_r(\lambda_0, \phi; a_{r,0}, b_{r,0}, a_{r,1}, b_{r,1}) \underbrace{p(y_0, t_0, y_1, t_1 | \lambda_0, \phi)}_{:=\lambda_0^{y_0} e^{-\lambda_0 t_0} (\phi\lambda_0)^{y_1} e^{-\phi\lambda_0 t_1}} \, \mathrm{d}\lambda_0 \, \mathrm{d}\phi$$

$$= \frac{w_r}{w'} \frac{b_{r,0}^{a_{r,0}} b_{r,1}^{a_{r,1}+1} \Gamma(a_{r,0} + y_0) \Gamma(a_{r,1} + y_1 + 1)}{(b_{r,0} + t_0)^{a_{r,0}+y_0} (b_{r,1} + t_1)^{a_{r,1}+y_1+1} \Gamma(a_{r,0}) \Gamma(a_{r,1} + 1)},$$

where w' is chosen so that $\sum_{r=1}^{R} w'_r = 1$ (Diaconis and Ylvisaker, 1985).

5.4. Example: TGN1412

During a first-in-man trial of the monoclonal antibody TGN1412 conducted in March 2006, all 6 of 6 the healthy volunteers in the test group were admitted to a critical care unit within 12 to 16 hours (total follow-up to first event or censoring $t_1 = 91.5$ hours) after receiving TGN1412 due to cytokine storms, while this occurred for none of the
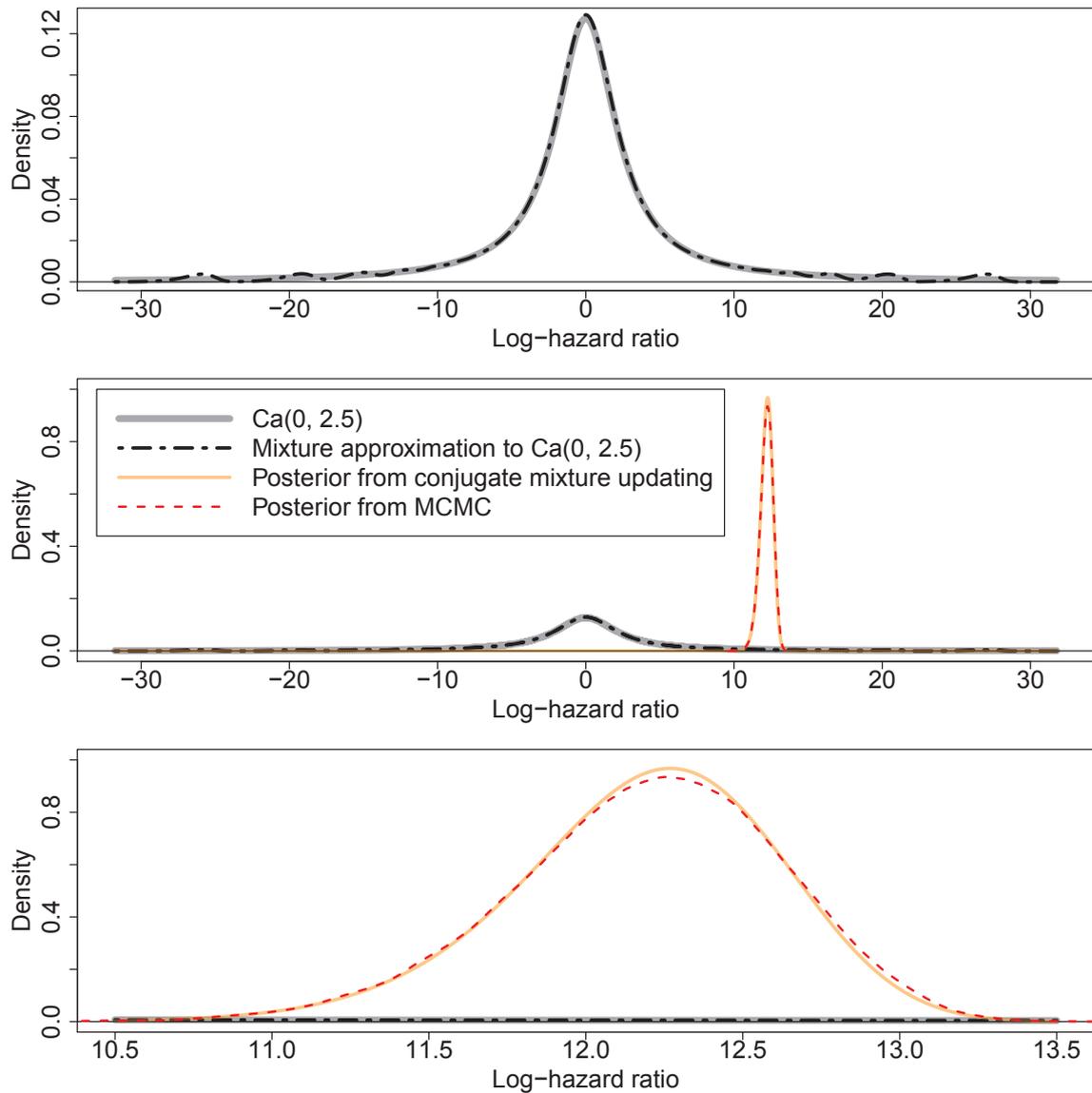
Figure 5.1.: Marginal prior and posterior distributions for the log-hazard ratio for the TGN1412 example showing the analytical $Ca(0, 2.5)$ prior density, an approximation to this prior using a mixture of 99 conjugate prior components constructed in the manner described in Section 5.3 with $\upsilon = 4/3$, a kernel density estimate of the posterior density obtained using MCMC sampling with a $Ca(0, 2.5)$ prior for the log-hazard ratio and the posterior obtained using conjugate updating of the mixture of conjugate priors

2 subjects treated with a placebo (t_0 presumably at least 32 hours) (Royal Statistical Society working party on statistical issues in first-in-man studies, 2007). We formalize the informal comparison with historical rates made by Senn (2008) for this example, which we had discussed on page 2 of the Introduction. Fisher's exact test results in a one-sided p-value of 0.0357 and an exponential time-to-event model using exact Poisson regression software leads to a median unbiased estimate of 1.05 (95% CI -0.62 to ∞)

for the log-hazard ratio with a two-sided p-value of 0.3308.

In order to obtain a prior for the placebo hazard rate, we used historical intensive care unit admission rates. In 2001, there were 127'614 discharges including due to death from intensive care units in England amongst a total population of 48'867'538 above 1 year of age (Wunsch et al., 2009). This can be seen as providing a conservative upper bound for the expected rate of intensive care admissions for healthy young male adults like those enrolled in the TGN1412 first-in-man trial, because they would be expected to be less likely to be hospitalized than the average member of the population. On this basis we used a Gamma(127'614, 48'867'538) prior for the control group hazard rate λ_0. For the log-hazard ratio φ of the TGN1412 group compared with the placebo group we used the Cauchy(0, 0.25) prior introduced in Section 3.4.2. With these priors, we obtained posterior median log-hazard ratio of 12.2 with equal-tailed 95% CI from 11.3 to 13.0 with a posterior probability in excess of 99.999% that TGN1412 increased the hazard rate for the admission to critical care using MCMC as shown in Figure 5.1. Applying conjugate updating to the mixture prior approximation to the Ca(0, 2.5) prior shown in Figure 5.1, we obtained closely matching results with a posterior median log-hazard ratio of 12.2 with equal-tailed 95% CI from 11.3 to 12.9.

Note that the cause of hospitalization in all patients was the same potential side effect of the drug, which was even considered a-priori possible, if unlikely (Royal Statistical Society working party on statistical issues in first-in-man studies, 2007). The calculated posterior probability ignores this further evidence of a causal drug effect, but on the other hand it also ignores the multiplicity issue of the many safety outcomes that could have been compared between groups. Program code for this example is available under https://doi.org/10.6084/m9.figshare.5007833.

5.5. Extending the results to a stratified meta-analysis

We will now extend the results from the previous subsections to a time-to-event meta analysis setting. We assume that we have prior information regarding the event rate

within each study, but are not making an assumption of exchangeability across studies.

Using the results of Chen and Ibrahim (2003), we obtain a conjugate prior for an exponential time-to-event model under independent right-censoring comparing the control groups in trials $i = 1, \ldots, I$ with hazard rate λ_{i0} versus test groups with hazard rates $\lambda_{i1} = \phi\lambda_{i0}$, where ϕ is the hazard ratio, as

$$p(\lambda_{10}, \ldots, \lambda_{I0}, \phi) \propto \exp\left(\sum_{i=1}^{I} \left(a_{i0} \log\lambda_{i0} - b_{i0}\lambda_{i0} + a_{i1}\log(\lambda_{i0}\phi) - b_{i1}\lambda_{i0}\phi\right)\right).$$

To obtain the prior density we need to calculate the normalizing constant γ, which is given by

$$\gamma^{-1} = \int_0^\infty \cdots \int_0^\infty \int_0^\infty e^{\sum_{i=1}^{I}(a_{i0}\log\lambda_{i0} - b_{i0}\lambda_{i0} + a_{i1}\log(\lambda_{i0}\phi) - b_{i1}\lambda_{i0}\phi)} \, \mathrm{d}\lambda_{10} \, \ldots \, \mathrm{d}\lambda_{I0} \, \mathrm{d}\phi$$
$$= \int_0^\infty \phi^{\sum_{i=1}^{I} a_{i1}} \prod_{i=1}^{I} \frac{\Gamma(a_{i0} + a_{i1} + 1)}{(b_{i0} + b_{i1}\phi)^{a_{i0} + a_{i1} + 1}} \, \mathrm{d}\phi.$$

Thus, calculating γ only requires a one-dimensional numeric integration for any given values of a_{ij}, b_{ij} for $i = 1, \ldots, I$ and $j = 0, 1$ so that the prior becomes

$$p(\lambda_{10}, \ldots, \lambda_{I0}, \phi) = \gamma \exp\left(\sum_{i=1}^{I} \left(a_{i0} \log\lambda_{i0} - b_{i0}\lambda_{i0} + a_{i1}\log(\lambda_{i0}\phi) - b_{i1}\lambda_{i0}\phi\right)\right).$$

The marginal prior for the hazard ratio ϕ is

$$p(\phi) = \gamma\phi^{\sum_{i=1}^{I} a_{i1}} \prod_{i=1}^{I} \frac{\Gamma(a_{i0} + a_{i1} + 1)}{(b_{i0} + b_{i1}\phi)^{a_{i0} + a_{i1} + 1}},$$

while the marginal prior for the control group hazard rate λ_{i0} for any study i is given by

$$p(\lambda_{i0}) = \gamma \int_0^\infty e^{-\lambda_{i0}(b_{i0} + b_{i1}\phi)} \lambda_{i0}^{a_{i0} + a_{i1}} \phi^{\sum_{r=1}^{I} a_{r1}}$$
$$\left(\prod_{r \in 1, \ldots, I \neq i} \frac{\Gamma(a_{r0} + a_{r1} + 1)}{(b_{r0} + b_{r1}\phi)^{a_{r0} + a_{r1} + 1}}\right) \mathrm{d}\phi.$$

The posterior after observing the data $\mathcal{D}$ consisting of $y_{ij\cdot}$ events in treatment group j of study i during $t_{ij\cdot}$ time units of follow-up to first event or censoring is

$$
p(\lambda_{10}, \ldots, \lambda_{I0}, \phi | \mathcal{D}) = \gamma' \exp \left(\sum_{i=1}^{I} \Big((a_{i0} + y_{i0\cdot}) \log \lambda_{i0} - (b_{i0} + t_{i0\cdot}) \lambda_{i0} \right.
$$
$$
\left. + (a_{i1} + y_{i1\cdot}) \log(\lambda_{i0}\phi) - (b_{i1} + t_{i1\cdot}) \lambda_{i0}\phi \Big) \right)
$$

with normalizing constant

$$
\gamma'^{-1} := \int_0^{\infty} \phi^{\sum_{i=1}^{I}(a_{i1}+y_{i1\cdot})} \prod_{i=1}^{I} \frac{\Gamma(a_{i0} + y_{i0\cdot} + a_{i1} + y_{i1\cdot} + 1)}{(b_{i0} + y_{i0\cdot} + (b_{i1} + y_{i1\cdot})\phi)^{a_{i0}+y_{i0\cdot}+a_{i1}+y_{i1\cdot}+1}} \, \mathrm{d}\phi.
$$

The updated posterior weight w'_r for a mixture component $r = 1, \ldots, R$ in a mixture of such priors after conjugate updating is obtained as (Diaconis and Ylvisaker, 1985)

$$
w'_r \propto w_r \int_0^{\infty} \int_0^{\infty} \ldots \int_0^{\infty} p_r(y_{i0\cdot}, t_{i0\cdot}, \ldots, y_{I1\cdot}, t_{I1\cdot} | \phi, \lambda_{10}, \ldots, \lambda_{I0})
$$
$$
p_r(\phi, \lambda_{10}, \ldots, \lambda_{I0}) \, \mathrm{d}\phi \, \mathrm{d}\lambda_{10} \ldots \, \mathrm{d}\lambda_{I0}
$$
$$
= \gamma_r^{-1} \int_0^{\infty} \phi^{\sum_{i=1}^{I}(a_{r,i1}+y_{i1\cdot})} \prod_{i=1}^{I} \frac{\Gamma(a_{r,i0} + a_{r,i1} + y_{i0\cdot} + y_{i1\cdot} + 1)}{(b_{r,i0} + t_{i0\cdot} + (b_{r,i1} + t_{i1\cdot})\phi)^{a_{r,i0}+a_{r,i1}+y_{i0\cdot}+y_{i1\cdot}+1}} \, \mathrm{d}\phi,
$$

which can be evaluated by one-dimensional numeric integration.

6. Comparison of methods for incorporating historical data

In Chapters 3 and 5 we used historical prior information through rMAP priors. As discussed in Section 2.5.4, there is limited previous research into the optimal choice of weight for the weakly informative mixture component in rMAP priors and how the rMAP approach compares to other methods for incorporating historical data into meta-analyses. We conducted a simulation study to investigate these questions.

6.1. Evaluated methods

The evaluated methods, that are also summarized in Table 6.1, were

1. a Bayesian hierarchical model with vague priors fitted only to the trials in the main meta-analysis ignoring the historical prior information,

2. the MAC approach (see Section 2.5.4.1) with a single Bayesian hierarchical model with vague priors fitted to all trials without any down-weighting of the historical prior information other than through the estimated between-trial variability,

3. the rMAP prior approach (see Section 2.5.4.3) with the MAP prior approximated by a mixture distribution and with different choices for the weight for the weakly informative mixture component, and

4. a Bayesian hierarchical meta-regression model with variable selection on the historical trial effect using a shrinkage prior (see Section 2.5.4.5). As shrinkage

priors we used the horseshoe, double-exponential (Laplace) and Cauchy priors with different choices for the scale parameters.

As Table 6.1 shows, we used a Poisson-likelihood based on simulated values of $y_{ij\cdot}$ and $t_{ij\cdot}$ for each trial arm. Compared to using the AD likelihood (3.2.4) this substantially reduced the time required for the simulations given that we fitted all models using MCMC methods. We did not use conjugate updating, because we were unable to find a reliable approach for performing conjugate updating in models with borrowing of information between trials.

Figure 6.1 illustrates several of the shrinkage priors used in the fourth of the approaches outlined above for parameter values that performed well in the simulation results presented in Section 6.3. As can be seen the horseshoe prior is substantially more peaked around zero than the other illustrated priors, while at the same time having higher density in the tail area outside of, say, $(-3.5, 3.5)$.
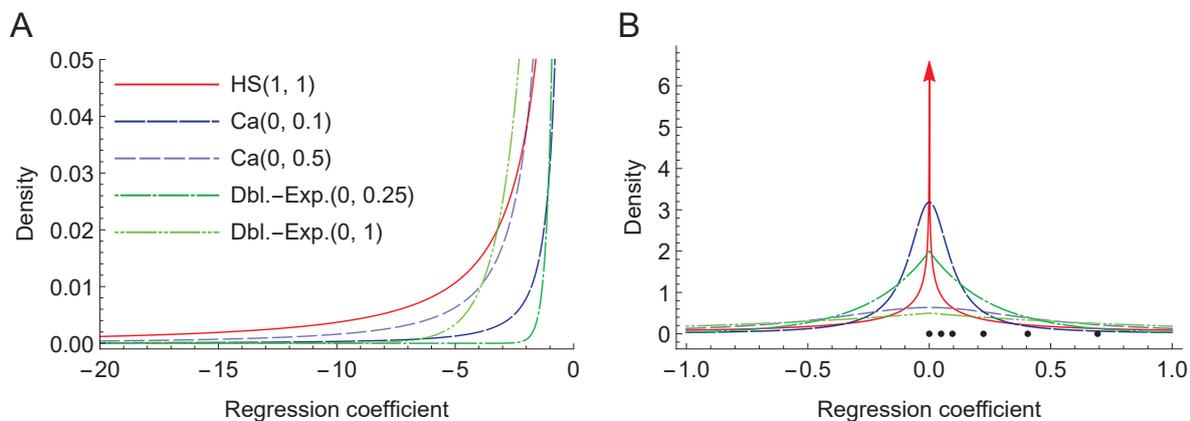
Figure 6.1.: Illustration of several of the considered shrinkage priors for the meta-regression coefficient for historical trials: density in the tail area (Panel A) and near the origin (Panel B) with the arrow indicating that the density of the horseshoe prior is unbounded at zero and with the black dots indicating the degrees of discrepancy between the placebo hazard rates in the historical and new trials considered in the simulations described in Section 6.2

6.2. Simulation study setup

The simulation study evaluated the scenarios shown in Table 6.2. Within each scenario, we considered 7 different sub-scenarios, and within these the situation with and without

Table 6.1.: Investigated methods for incorporating historical data, abbreviations used for them in results figures and priors used: a $Ca(0, 2.5)$ prior was used for the log-hazard ratio φ in all methods, $rMAP(\mathcal{D}', w)$ denotes the robust MAP prior based on the historical data $\mathcal{D}'$ with weight $w = 0, 0.005, \ldots, 0.995, 1$ for the informative mixture component; $h_i := \mathbb{1}\{i > I\}$

Approach	Historical data likelihood	Likelihood for new data	Parameter	Prior used
Hierarchical model with vague priors ("Vague")	not used	$\prod_{i=1}^{I} \phi(\log \lambda_{i0}; \nu_1, \sigma_1) \prod_{j=0}^{1} (\lambda_{i0} e^{\varphi j})^{y_{ij\cdot}} e^{-\lambda_{i0} e^{\varphi j} t_{ij\cdot}}$	ν_1 σ_1	$N(0, 1000)$ $HCa(0, 100)$
Meta-analytic combined (MAC)	$\prod_{i=1}^{I+H} \phi(\log \lambda_{i0}; \nu_1, \sigma_1) \prod_{j=0}^{1-h_i} (\lambda_{i0} e^{\varphi j})^{y_{ij\cdot}} e^{-\lambda_{i0} e^{\varphi j} t_{ij\cdot}}$		ν_1 σ_1	$N(0, 1000)$ $HCa(0, 100)$
Model averaging using shrinkage priors ("Horseshoe", "Cauchy", "Dbl.-Exp.")	$\prod_{i=1}^{I+H} \phi(\log \lambda_{i0}; \nu_1, \sigma_1) \prod_{j=0}^{1-h_i} (\lambda_{i0} e^{\beta h_i + \varphi j})^{y_{ij\cdot}} e^{-\lambda_{i0} e^{\beta h_i + \varphi j} t_{ij\cdot}}$		ν_1 σ_1 β	$N(0, 1000)$ $HCa(0, 100)$ $HS(\sigma_{\lambda\tau}, \sigma_{\lambda\tau})$ $Ca(0, \sigma_{Ca})$ $Dbl.-Exp.(0, \sigma_L)$
Robust meta-analytic predictive (rMAP)	$\prod_{i=I+1}^{I+H} (\phi(\log \lambda_{i0}; \nu_1', \sigma_1')$ $\lambda_{i0}^{y_{i0\cdot}} e^{-\lambda_{i0} t_{i0\cdot}})$	$\prod_{i=1}^{I} \phi(\log \lambda_{i0}; \nu_1, \sigma_1) \prod_{j=0}^{1} (\lambda_{i0} e^{\varphi j})^{y_{ij\cdot}} e^{-\lambda_{i0} e^{\varphi j} t_{ij\cdot}}$	ν_1' σ_1' ν_1, σ_1	$N(0, 1000)$ $HCa(0, 100)$ $rMAP(\mathcal{D}', w)$

a treatment effect. As a basic reference scenario we considered 10 historical and 10 new trials of 1 year duration, each with 100 patients in each treatment group. In addition, we investigated 6 other scenarios with only 5 new trials, with 20 new trials, with a higher sample size in each trial (1000 instead of 100 per arm), with more historical trials (100 instead of 10), with between-trial variability in control group hazard rates, and with this between-trial variability differing between historical and new trials.

Table 6.2.: Parameters for the simulation scenarios for the trials in the main meta-analysis $i = 1, \ldots, I$ and the historical trials $i = I+1, \ldots, I+H$: sample size per group n_{ij}, placebo event hazard rate λ_{i0}, log-hazard ratio for events on test drug compared with placebo φ, dispersion parameter κ for variation around λ_{i0} across trials, placebo drop-out hazard rate μ_{i0}, test drug drop-out hazard rate μ_{i1} and planned trial duration τ_i

	Main meta-analysis					Historical trials				All trials		
#	I	n_{ij}	λ_{i0}	φ	κ	H	n_{i0}	λ_{i0}	κ	μ_{i0}	μ_{i1}	τ_i
1	10	100	0.1	0	0	10	100	$0.1 - 0.2$	0	0.15	0.15	1
2	10	100	0.1	$\log(1.75)$	0	10	100	$0.1 - 0.2$	0	0.15	0.15	1
3	5	100	0.1	0	0	10	100	$0.1 - 0.2$	0	0.15	0.15	1
4	5	100	0.1	$\log(1.75)$	0	10	100	$0.1 - 0.2$	0	0.15	0.15	1
5	20	100	0.1	0	0	10	100	$0.1 - 0.2$	0	0.15	0.15	1
6	20	100	0.1	$\log(1.75)$	0	10	100	$0.1 - 0.2$	0	0.15	0.15	1
7	10	1000	0.1	0	0	10	1000	$0.1 - 0.2$	0	0.15	0.15	1
8	10	1000	0.1	$\log(1.75)$	0	10	1000	$0.1 - 0.2$	0	0.15	0.15	1
9	10	100	0.1	0	0	100	100	$0.1 - 0.2$	0	0.15	0.15	1
10	10	100	0.1	$\log(1.75)$	0	100	100	$0.1 - 0.2$	0	0.15	0.15	1
11	10	100	0.1	0	0.75	10	100	$0.1 - 0.2$	0.75	0.15	0.15	1
12	10	100	0.1	$\log(1.75)$	0.75	10	100	$0.1 - 0.2$	0.75	0.15	0.15	1
13	10	100	0.1	0	0.75	10	100	$0.1 - 0.2$	0.1	0.15	0.15	1
14	10	100	0.1	$\log(1.75)$	0.75	10	100	$0.1 - 0.2$	0.1	0.15	0.15	1

Within each scenario, we considered sub-scenarios in terms of different placebo hazard rates in the historical data chosen to be 0.1, 0.105, 0.11, 0.125, 0.15 or 0.2 per patient-year. Program code for the conducted simulations is available under https://doi.org/10.6084/m9.figshare.5007833 and the simulation results under https://doi.org/10.6084/m9.figshare.5027927.

6.3. Simulation study results

Figure 6.2 shows the RMSE across all scenarios for different degrees of expected prior-data conflict. The results illustrate that the direct MAC approach and the MAP

Figure 6.2.: Difference in root-mean-square-error for the log-hazard ratio when including historical information via different methods (with the parameters for each method given at the top of the graph) compared with an analysis with a vague prior (blue=lower, white=similar, red=higher) across 14 different scenarios and sub-scenarios regarding the discrepancy in placebo hazard rates between present (0.1) and historical (0.1 to 0.2) trials

approach give very similar results despite the approximation of the posterior of the hyperparameters by a mixture of bivariate normals.

Figure 6.3 shows the average RMSE (aRMSE) across different degrees of expected prior data conflict — i.e. sub-scenarios with the same placebo hazard rate in the historical data as in the new data to those with a doubled placebo hazard rate — in each scenario. It illustrates the inherent bias-variance trade-off that occurs when it is a-priori suspected that there may be some degree of prior-data conflict. In terms of aRMSE it is neither optimal to totally disregard prior information, nor to use it without any down-weighting, but rather to allow some degree of borrowing of information.
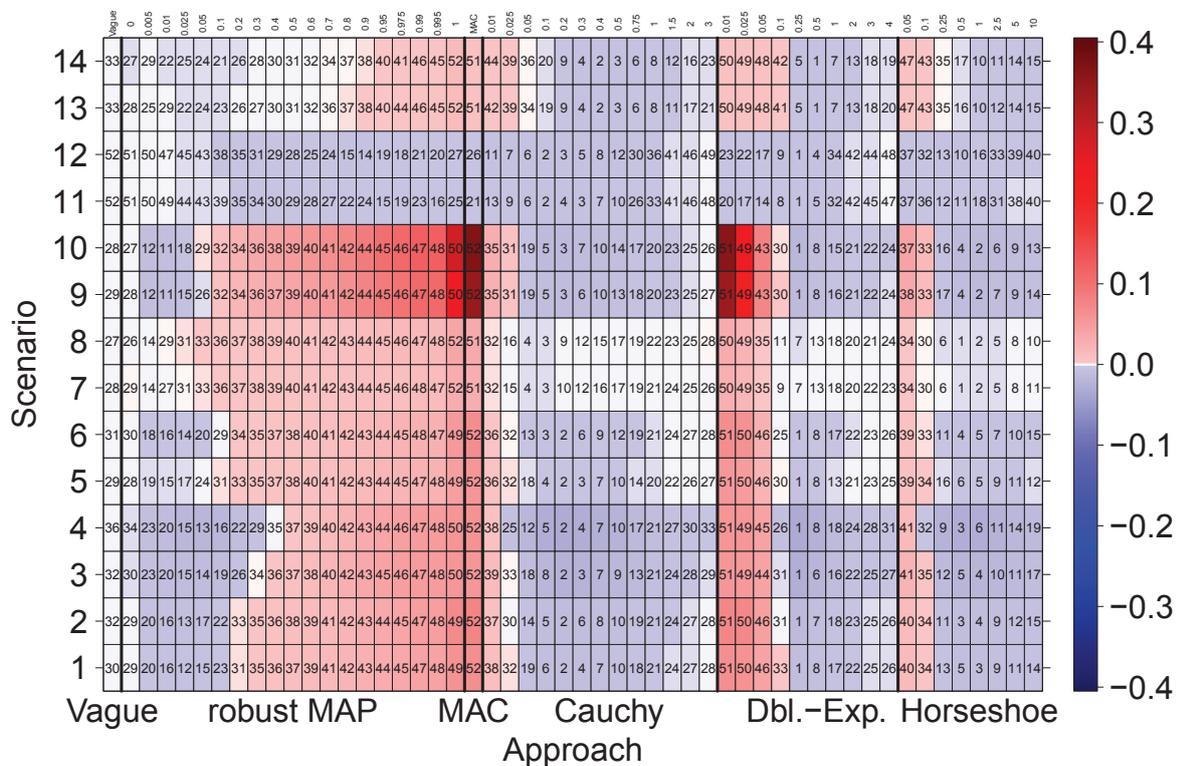
Figure 6.3.: Difference in average root-mean-square-error for the log-hazard ratio when including historical information via different methods (with the parameters for each method given at the top of the graph) compared with an analysis with a vague prior (blue=lower, white=similar, red=higher) across the 14 different scenarios: numbers indicate the rank of the method within the scenario from 1 (best) to 52 (worst)

Out of the investigated priors for inducing shrinkage on the coefficient for historical trials in hierarchical meta-regression models the double-exponential and Cauchy priors performed well in terms of aRMSE across the evaluated scenarios. The horseshoe prior also performed better than the rMAP approach, but appeared to be outperformed by

the other priors in scenarios with limited prior information.

It is not clear that there is a generally recommendable default class of priors for inducing shrinkage on the coefficient for historical trials in hierarchical meta-regression models. In fact, all priors performed reasonably well and the choice of parameters of the priors may be more crucial. Depending on the expected degree of prior data conflict a more or less aggressive parameter choice may be preferable: while a scale parameter of 0.25 for a double-exponential prior performed extremely well on average, a scale parameter of 0.5 performed nearly as well on average and performed better in case of more substantial expected prior data conflict.

It was perhaps to be expected that these prior parameter choices would perform well on average across different degrees of discrepancy between the placebo hazard rates in the historical and new trials. This is because the resulting prior approximately resembles the distribution of the considered discrepancies as illustrated in Figure 6.1.

As Figure 6.4 shows the prior choices that performed well in terms of aRMSE also performed reasonably well in terms of the coverage probability of 95% CIs for the treatment effect. E.g. for a double-exponential prior with scale parameter of 0.5 the coverage was very close to the nominal level, while it was more notably below or above the nominal level for a more aggressive scale parameter value of 0.25.

As expected, when the parameters for generating the historical data and the data in the main meta-analysis are identical, RMSE improves as increasingly more weight is given to the historical prior information in the rMAP approach. However, the largest reductions in RMSE were achieved when increasing the weight from close to 0 to, say, 0.2 and in some scenarios 0.5, while further improvements in RMSE beyond that point were smaller. Previously suggested weights for the informative MAP prior component of the rMAP prior such as 0.5, 0.8 or 0.9 performed less well in terms of aRMSE than substantially lower weights such as 0.01 to 0.05. The only exceptions were scenarios #11 and #12, in which there was substantial between-trial variability in placebo hazard rates, in which case one would expect a wider and more dispersed rMAP prior. In these scenarios weights of between 0.8 to 0.95 performed best. In contrast to this finding
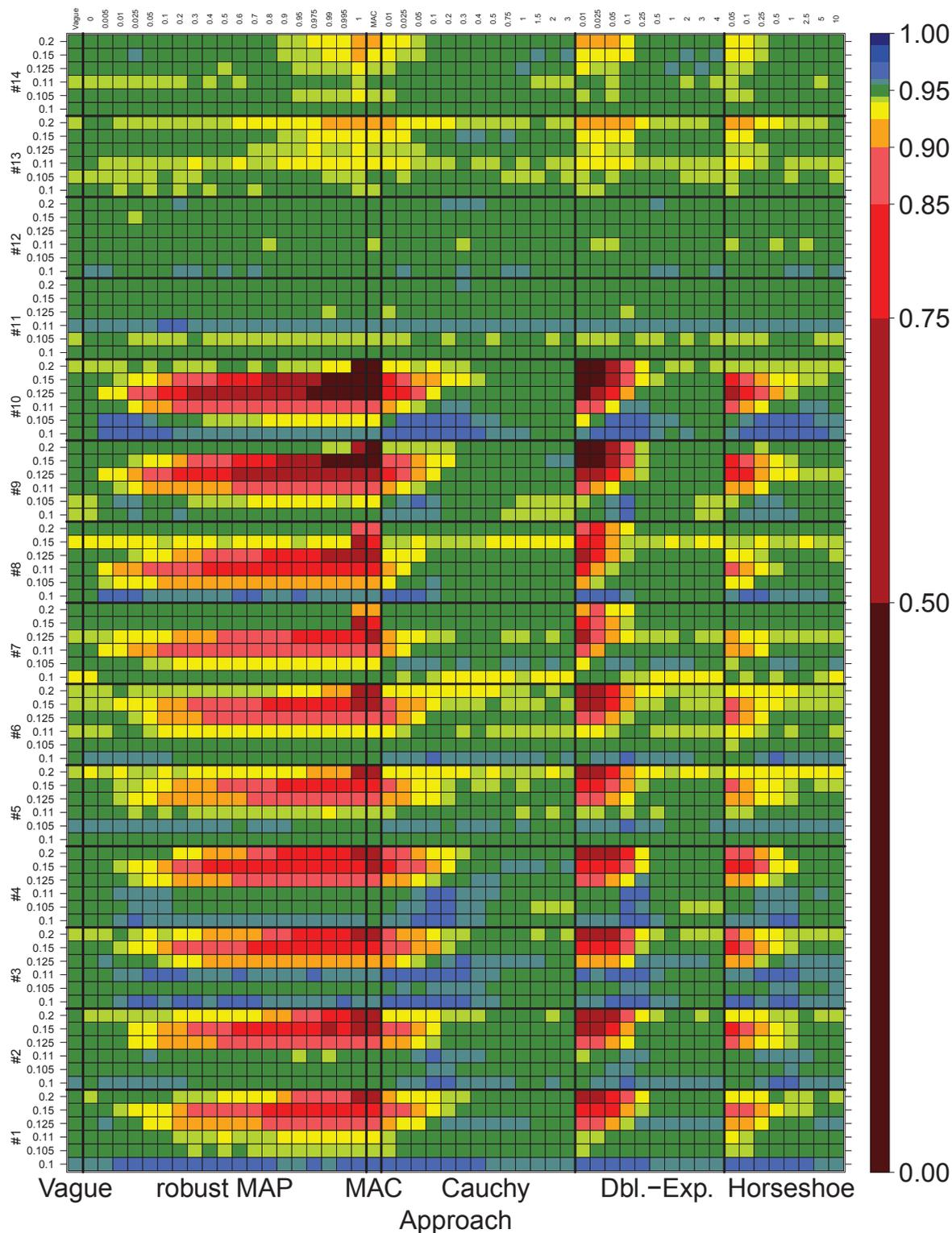
Figure 6.4.: Coverage probability of 95% CIs for the log-hazard ratio from different approaches for including historical information via different methods (with the parameters for each method given at the top of the graph) compared with an analysis with a vague prior across the 14 different scenarios and sub-scenarios regarding the discrepancy in placebo hazard rates between present (0.1) and historical (0.1 to 0.2) trials

for the rMAP approach, there was no notable difference between the optimal scale parameter choice for the considered shrinkage priors across scenarios.

When there is a moderate conflict between historical and present model parameters ($\lambda_{i0} = 0.125$ for $i = I + 1, \ldots, I + H$ versus $\lambda_{i0} = 0.1$ for $i = 1, \ldots, I$), the RMSE still initially improves slightly as the informative historical prior component is given more weight. In contrast a relatively strong conflict ($\lambda_{i0} = 0.15$ versus 0.1) was the most unfavorable situation in most scenarios and RMSE uniformly increased with increasing weight for the informative mixture component. While a hazard ratio of 1.5 between the historical and the present data is quite large, the historical and the present data were not informative enough to entirely discard the informative prior in the inference in most scenarios. The only exceptions are scenarios #7 and #8, in which there is substantially more data (1000 instead of 100 patients per trial), but in these scenarios the same issue occurs at a lower level of expected prior data conflict ($\lambda_{i0} = 0.11$ to 0.125 versus λ_{i0} = 0.1). In contrast for the scenario with the largest expected prior-data conflict (λ_{i0} = 0.2 versus 0.1), the weakly informative mixture component more clearly dominates the inference and the RMSE even with a weight of 0.9 for the informative mixture component remains below the RMSE for substantially lower weights in the scenario with a strong conflict.

These findings illustrates that the relative amount of information in the historical and the present data is a key factor in the optimal weight given to the historical data. This can also be seen from the greater gain in terms of aRMSE in scenarios #3 and #4, where there is fewer new trials (5 trials) compared to scenarios #1, #2 (10 trials), #3 and #4 (20 trials). However, it should be noted that this overall reduction in aRMSE is the result of a greater reduction in terms of RMSE when there is no or little expected prior-data conflict, but that this is partially offset by a greater increase in aRMSE when there is a strong expected prior-data conflict. The same effects can be seen when there is a larger number of historical trials (100 instead of 10) in scenarios #9 and #10.

7. Discussion and conclusions

7.1. Discussion

While meta-analyses typically treat the occurrence of medical events as binomial data, it is rarely considered that this makes the strong assumption that in each trial drop-out times follow the same distribution in all treatment groups and are independent of event times. In the absence of IPD, the AD time-to-event meta-analysis model proposed in Chapter 3 requires several — arguably more plausible — assumptions, most importantly that both event and drop-out times are independent and follow specific distributions. Our simulations show that the proposed model — unlike binomial distribution based methods such as the Peto one-step odds ratio, logistic regression, the CMH odds ratio or the beta-binomial model of Kuss — avoids type I error rate inflation due to differences between treatment groups in drop-out time distributions. We assumed an exponential distribution for event and drop-out times in both the rosiglitazone case study and for our simulations, but the event and drop-out time distributions can be chosen to be sufficiently flexible to provide a good approximation to the expected distributions for each application. For example, in Section 4.1 we derived the AD likelihood assuming Weibull distributed event and drop-out times.

The necessary inputs for the proposed AD meta-analysis model are typically reported in the publications of clinical trials or clinical trial results databases, but an imputation of

missing values from the predictive distribution under a MAR assumption is also easily accomplished. We have derived both the AD and the IPD likelihood for the proposed model so that a mixture of both types of data can be used, which may also ameliorate the parameter identifiability issues that arise with AD, if we assume flexible event and drop-out time distributions such as the Weibull distribution. In Sections 4.2 and 4.3 we derived the AD likelihood for event driven trials and demonstrated how to account for the risk of fatal events due to recurrent events.

There are a number of alternatives to using the proposed AD likelihood. For the exponential time-to-event model using the imputed follow-up to first event or censoring $\tilde{t}_{ij.}$ — assuming half the planned follow-up for those that drop-out or experience an event — in a "Poisson likelihood" also performed reasonably well in our simulations, especially for low event rates. While $\tilde{t}_{ij.}$ is a biased approximation to $t_{ij.}$, it appears that by having a similar bias in all treatment groups of a trial the bias often approximately cancels out when it comes to the estimation of the hazard ratio. Further alternatives include approximate Bayesian computations (Marin et al., 2012), which would avoid the need to derive and evaluate an AD likelihood.

Borrowing of information in meta-analyses using hierarchical models relies on an assumption of exchangeability of parameters between trials. Instead of assuming the exchangeability of the expected proportion of cases across the control groups of trials of very different lengths, we suggest that it is more appropriate to assume that parameters of survival distributions are exchangeable — e.g. the control group log-hazard rate in an exponential model. In fact, we feel that the choice of parameters about which exchangeability assumptions are made deserves more careful consideration in many settings. Additionally, between-trial differences in explanatory variables expected to affect trial-level parameters — e.g. the percentage of patients with a key risk factor for events — should be accounted for, instead of being absorbed into random effects as unexplained between-trial heterogeneity.

Criticism of Bayesian methods frequently focuses on the subjectivity of priors. This may be partially addressed by basing priors on historical data chosen according to pre-specified objective criteria. Our simulations illustrate that using priors based

on historical data can result in a substantial gain in power. However, the historical data will often not fulfill several of the criteria for selecting "acceptable historical controls" suggested by Pocock (1976) and the validity of an exchangeability assumption for the parameters underlying a present set of data, and historical information will often be questioned. Widening the MAP prior and using long-tailed priors that offer more robustness against a prior-data conflict (Schmidli et al., 2014) acknowledges this uncertainty. Our simulations in Chapter 6 support the suggestion by Senn (2008) that the inherent bias-variance trade-off involved in using historical controls is especially favorable in situations with sparse data. Our simulations in Section 3.6 and Chapter 6 also suggest that the loss in power due to using a rMAP instead of a MAP prior is small over a wide range of weights given to the informative mixture component. In fact, substantially lower weights than previously suggested appear to optimize RMSE under a plausible extent of differences in parameter values underlying the present and historical data. In our simulations, model averaging via shrinkage priors outperformed the rMAP prior approach for a moderate degree of expected prior-data conflict. We assume that this occurs, because both approaches can completely discount the historical data or give it full weight, but only the shrinkage prior approach explicitly allows for the possibility of a somewhat broadened version of the MAP prior that may be optimal in terms of RMSE in case of a moderate prior-data conflict. It may be possible to achieve a similar behavior with a type of rMAP prior that includes mixture components with different degrees of informativeness. Such an approach could be seen as a discrete mixture approximation to the power prior approach of (Ibrahim et al., 2015) or a more formal basis for the empirical MAP prior approach of Li et al. (2016).

The use of historical data and borrowing information across trials in a meta-analysis step out of the relatively safe bounds of randomized comparisons. In the rosiglitazone example, it would be necessary to further investigate whether differences in populations explain the differences in control rates between historical and new trials before relying on historical prior information for strengthening the case for an increased risk of cardiovascular events with rosiglitazone. In any case only the stratified Bayesian analysis resulted in 95% CIs excluding a hazard ratio of 1.0. This analysis used

independent informative priors based on historical data for each of the trials in the main meta-analysis. As a result parameters are assumed to be exchangeable between the historical trials, as well as between all historical trials and any individual trial in the main meta-analysis, but not between the trials in the main meta-analysis. By making these somewhat inconsistent assumptions that do not allow borrowing of information across trials in the main meta-analysis the informative priors become more influential. For these reasons we prefer the analysis with borrowing of information between the trial in the main meta-analysis. The question of how to introduce "borrowing of information" is also the greatest remaining challenge — besides the practical problem of the availability of $t_{ij.}$ — regarding the conjugate prior derived in Section 5 for the exponential time-to-event model. If this challenge can be overcome, then conjugate updating would allow for substantially faster simulations to explore questions such as the value of prior information on event rates and the operating characteristics of different methods of using such prior information.

The proposed Bayesian hierarchical meta-analysis model with borrowing of information performed well in the scenarios investigated in our simulation study. When control group parameters were exchangeable between historical and new trials, using historical prior information — as expected — lead to increased power under the alternative hypothesis without an inflation of the type I error rate under the null hypothesis.

Methods for binomially distributed data performed well in simulations with identical drop-out time distributions across all treatment groups with the exception of the beta-binomial model of Kuss (2015). This last model with an independent beta-distributed random effect on each trial arm performed badly with respect to most operating characteristics in our simulations and produced excessively wide CIs. This is because even for a test and control group from the same trial, this model attributes a part of the differences between arms to the between-arm variability rather than the treatment effect. This effect could be seen in the rosiglitazone case study and was even more pronounced in our simulations, because in both cases there were substantial between-trial differences in trial duration that caused substantial between-arm differences in the expected proportion of patients with an event. In such a setting we recommend not to use this particular beta-binomial model.

While we focused on the meta-analysis of time-to-event data from clinical trials, there are a number of other possible uses of the discussed methods for using historical data in the clinical trials setting including blinded monitoring of clinical trials (Gould, 2016), interim decision making and trial planning.

7.2. Conclusions

Current methods for the meta-analysis of aggregate data on the occurrence of medical events have several limitations with respect to how drop-outs are accounted for, how "borrowing of information" across trials in a meta-analysis is done and how information from historical trials is used. We proposed a trial-level time-to-event likelihood that accounts for differences in follow-up between groups and only requires commonly available AD. We also demonstrated that exchangeability assumptions should be made about parameters of survival distributions rather than for the probability of an event. The investigated methods for borrowing of information in meta-analyses and using historical data are attractive for the analysis of sparse data.

Bibliography

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd edition. Wiley, Hoboken.

Agresti, A. and Hartzel, J. (2000). Strategies for comparing treatments on a binary response with multi-centre data. *Statistics in Medicine*, **19**(8):1115–1139.

Allison, P. D. (1995). *Survival analysis using SAS: A practical guide*. SAS Institute, Cary.

Arends, L. R., Hunink, M., and Stijnen, T. (2008). Meta-analysis of summary survival curve data. *Statistics in Medicine*, **27**(22):4381–4396.

Bagos, P. G. and Nikolopoulos, G. K. (2009). Mixed-effects Poisson regression models for meta-analysis of follow-up studies with constant or varying durations. *The International Journal of Biostatistics*, **5**(1).

Bai, O., Chen, M., and Wang, X. (2016). Bayesian estimation and testing in random effects meta-analysis of rare binary adverse events. *Statistics in Biopharmaceutical Research*, **8**(1):49–59.

Bennett, M. M., Crowe, B. J., Price, K. L., Stamey, J. D., and Seaman Jr, J. W. (2013). Comparison of Bayesian and frequentist meta-analytical approaches for analyzing time to event data. *Journal of Biopharmaceutical Statistics*, **23**(1):129–145.

Bernardo, J. M. and Smith, A. F. (2000). *Bayesian theory*, revised edition. Wiley, Chichester.

Bertagnolli, M. M., Sartor, O., Chabner, B. A., Rothenberg, M. L., Khozin, S., Hugh-Jones, C., Reese, D. M., and Murphy, M. J. (2017). Advantages of a truly open-access data-sharing model. *New England Journal of Medicine*, **376**(12):1178–1181.

Betancourt, M. and Girolami, M. (2015). Hamiltonian Monte Carlo for hierarchical models. In Upadhyay, S. K., Singh, U., Dey, D. K., and Loganathan, A., editors, *Current trends in Bayesian methodology with applications*, pages 79–101. CRC Press, Boca Raton.

Bhadra, A., Datta, J., Polson, N. G., and Willard, B. (2016). Default Bayesian analysis with global-local shrinkage priors. *Biometrika*, **103**(4):955–969.

Bodnar, O., Link, A., Arendacká, B., Possolo, A., and Elster, C. (2017). Bayesian estimation in random effects meta-analysis using a non-informative prior. *Statistics in Medicine*, **36**(2):378–399.

Boehringer Ingelheim (2016). ClinicalTrials.gov registry posting for BI 10773 (empagliflozin) cardiovascular outcome event trial in type 2 diabetes mellitus patients (EMPA-REG OUTCOME). Available: `https://clinicaltrials.gov/ct2/show/NCT01131676` [accessed: 9 August 2016].

Böhning, D., Mylona, K., and Kimber, A. (2015). Meta-analysis of clinical trials with rare events. *Biometrical Journal*, **57**:633–648.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., and Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley, Chichester.

Bradburn, M. J., Deeks, J. J., Berlin, J. A., and Russell Localio, A. (2007). Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine*, **26**(1):53–77.

Brockhaus, A. C., Grouven, U., and Bender, R. (2016). Performance of the Peto odds ratio compared to the usual odds ratio estimator in the case of rare events. *Biometrical Journal*, **58**(6):1428–1444.

Browne, W. J. and Draper, D. (2006). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, **1**(3):473–514.

Bibliography

Burden, R. L. and Faires, J. D. (2011). *Numerical analysis*, 9th edition. Brooks/Cole, Boston.

Buyse, M. and Ryan, L. M. (1987). Issues of efficiency in combining proportions of deaths from several clinical trials. *Statistics in Medicine*, **6**(5):565–576.

Cai, T., Parast, L., and Ryan, L. (2010). Meta-analysis for rare events. *Statistics in Medicine*, **29**(20):2078–2089.

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, **76**(1):1–32.

Carvalho, C. M., Polson, N. G., and Scott, J. G. (2009). Handling sparsity via the horseshoe. In van Dyk, D. and Welling, M., editors, *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, pages 73–80, Microtome Publishing, Brookline.

Casella, G. and Berger, R. L. (2002). *Statistical inference*, 2nd edition. Duxbury, Pacific Grove.

Chen, M.-H. and Ibrahim, J. G. (2003). Conjugate priors for generalized linear models. *Statistica Sinica*, **13**(2):461–476.

Chung, Y., Rabe-Hesketh, S., and Choi, I.-H. (2013). Avoiding zero between-study variance estimates in random-effects meta-analysis. *Statistics in Medicine*, **32**(23): 4071–4089.

Combescure, C., Courvoisier, D. S., Haller, G., and Perneger, T. V. (2016). Meta-analysis of two-arm studies: Modeling the intervention effect from survival probabilities. *Statistical Methods in Medical Research*, **25**(2):857–871.

Coory, M. D., Wills, R. A., and Barnett, A. G. (2009). Bayesian versus frequentist statistical inference for investigating a one-off cancer cluster reported to a health department. *BMC Medical Research Methodology*, **9**:30.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**(2):187–220.

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 Medical Research Methodology*, **12**:34.

Davey, J., Turner, R. M., Clarke, M. J., and Higgins, J. P. (2011). Characteristics of meta-analyses and their component studies in the Cochrane Database of Systematic Reviews: A cross-sectional, descriptive analysis. *BMC Medical Research Methodology*, **11**:160.

Diaconis, P. and Ylvisaker, D. (1985). Quantifying prior opinion. In Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and Smith, A. F. M., editors, *Bayesian statistics 2: Proceedings of the second Valencia International Meeting, September 6-10, 1983*, pages 133–156, North-Holland, Amsterdam.

Dias, S., Welton, N. J., Sutton, A. J., and Ades, A. (2014). NICE DSU technical support document 2: A generalised linear modelling framework for pairwise and network meta-analysis of randomised controlled trials. Technical report, NICE Decision Support Unit. Available: `http://www.nicedsu.org.uk/TSD2%20General%20meta%20analysis%20corrected%2015April2014.pdf` [accessed: 30 September 2014].

Djulbegovic, B., Kumar, A., Glasziou, P. P., Perera, R., Reljic, T., Dent, L., Raftery, J., Johansen, M., Di Tanna, G. L., Miladinovic, B., Soares, H. P., Vist, G. E., and Chalmers, I. (2012). New treatments compared to established treatments in randomized trials. *Cochrane Database of Systematic Reviews*, **2012**(10):1–50.

Fraley, C., Raftery, A., Murphy, T., and Scrucca, L. (2012). mclust Version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation. technical report No. 597. Technical report, Department of Statistics, University of Washington. Available: `https://www.stat.washington.edu/research/reports/2012/tr597.pdf` [accessed: 7 January 2016].

Friede, T., Röver, C., Wandel, S., and Neuenschwander, B. (2017). Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods*, **8**(1):79–91.

Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M., and Granger, C. B. (2015). *Fundamentals of clinical trials*, 5th edition. Springer, Cham.

Gail, M. H., Wieand, S., and Piantadosi, S. (1984). Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*, **71**(3):431–444.

Gelfand, A. E. and Sahu, S. K. (1999). Identifiability, improper priors, and Gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, **94**(445):247–253.

Gelman, A. and Rubin, D. B. (1995). Avoiding model selection in Bayesian social research. *Sociological Methodology*, **25**:165–173.

Gelman, A., Jakulin, A., Pittau, M. G., and Su, Y.-S. (2008). A weakly informative default prior distribution for logistic and other regression models. *The Annals of Applied Statistics*, **2**(4):1360–1383.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2014). *Bayesian data analysis*, 3rd edition. CRC Press, Boca Raton.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, **5**(10):3–8.

Glaxo-Smith-Kline (2010). GSK briefing information, Avandia (rosiglitazone maleate) tablets, for the July 13-14, 2010 joint meeting of the endocrinologic and metabolic drugs advisory committee and drug safety and risk management advisory committee. Available: `http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/Drugs/EndocrinologicandMetabolicDrugsAdvisoryCommittee/UCM218492.pdf` [accessed: 14 November 2013].

Gould, A. L. (2016). Control charts for monitoring accumulating adverse event count frequencies from single and multiple blinded trials. *Statistics in Medicine*, **35**(30): 5561–5578.

Greenland, S. (2000). Small-sample bias and corrections for conditional maximum-likelihood odds-ratio estimators. *Biostatistics*, **1**(1):113–122.

Gueorguieva, R., Rosenheck, R., and Lin, H. (2012). Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **175** (2):417–433.

Guyot, P., Ades, A., Ouwens, M. J., and Welton, N. J. (2012). Enhanced secondary analysis of survival data: Reconstructing the data from published Kaplan-Meier survival curves. *BMC Medical Research Methodology*, **12**:9.

Hahn, G. J. and Meeker, W. Q. (1991). *Statistical intervals: A guide for practitioners.* Wiley, New York.

Hamra, G. B., MacLehose, R. F., and Cole, S. R. (2013). Sensitivity analyses for sparse-data problems — using weakly informative Bayesian priors. *Epidemiology*, **24** (2):233–239.

Heinze, G. and Schemper, M. (2001). A solution to the problem of monotone likelihood in Cox regression. *Biometrics*, **57**(1):114–119.

Heinze, G. and Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, **21**(16):2409–2419.

Higgins, J. and Whitehead, A. (1996). Borrowing strength from external trials in a meta-analysis. *Statistics in Medicine*, **15**(24):2733–2749.

Higgins, J. P. and Green, S., editors (2011). *Cochrane handbook for systematic reviews of interventions, Version 5.1.0 (updated March 2011)*. The Cochrane Collaboration. Available: `http://handbook.cochrane.org` [accessed: 23 May 2016].

Bibliography

Hirji, K. F. (2006). *Exact analysis of discrete data.* Chapman & Hall/CRC, Boca Raton.

Hobbs, B. P., Carlin, B. P., Mandrekar, S. J., and Sargent, D. J. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics*, **67**(3):1047–1056.

Hoffman, M. D. and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, **15**:1593–1623.

Holzhauer, B. (2017). Meta-analysis of aggregate data on medical events. *Statistics in Medicine*, **36**(5):723–737.

Hoyle, M. W. and Henley, W. (2011). Improved curve fits to summary survival data: Application to economic evaluation of health technologies. *BMC Medical Research Methodology*, **11**:139.

Ibrahim, J. G., Chen, M.-H., and Sinha, D. (2001). *Bayesian survival analysis*. Springer, New York.

Ibrahim, J. G., Chen, M.-H., Xia, H. A., and Liu, T. (2012). Bayesian meta-experimental design: Evaluating cardiovascular risk in new antidiabetic therapies to treat type 2 diabetes. *Biometrics*, **68**(2):578–586.

Ibrahim, J. G., Chen, M.-H., Gwon, Y., and Chen, F. (2015). The power prior: Theory and applications. *Statistics in Medicine*, **34**(28):3724–3749.

Johnson, N. L. and Kotz, S. (1969). *Univariate discrete distributions*. Wiley, New York.

Johnson, N. L., Kotz, S., and Balakrishnan, N. (1994). *Continuous univariate distributions Volume 1*, 2nd edition. Wiley, New York.

Kaizar, E. E., Greenhouse, J. B., Seltman, H., and Kelleher, K. (2006). Do antidepressants cause suicidality in children? A Bayesian meta-analysis. *Clinical Trials*, **3**(2): 73–98.

Kault, D. and Kault, S. (2015). From p-values to objective probabilities in assessing medical treatments. *PLOS ONE*, **10**(11):1–19.

Kontopantelis, E., Springate, D. A., and Reeves, D. (2013). A re-analysis of the Cochrane library data: The dangers of unobserved heterogeneity in meta-analyses. *PLoS ONE*, **8**(7):e69930.

Kunz, L. M., Normand, S.-L. T., and Sedrakyan, A. (2015). Meta-analysis of rate ratios with differential follow-up by treatment arm: Inferring comparative effectiveness of medical devices. *Statistics in Medicine*, **34**(21):2913–2925.

Kuss, O. (2015). Statistical methods for meta-analyses including information from studies without any events — add nothing to nothing and succeed nevertheless. *Statistics in Medicine*, **34**(7):1097–1116.

Lachin, J. M. (2000). *Biostatistical methods: The assessment of relative risks*. Wiley, New York.

Lachin, J. M. and Foulkes, M. A. (1986). Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics*, **42**(3):507–519.

Lambert, P. C., Sutton, A. J., Burton, P. R., Abrams, K. R., and Jones, D. R. (2005). How vague is vague? A simulation study of the impact of the use of vague prior distributions in MCMC using WinBUGS. *Statistics in Medicine*, **24**(15):2401–2428.

Lancaster, T. (1985). Generalised residuals and heterogeneous duration models: With applications to the Weibull model. *Journal of Econometrics*, **28**(1):155–169.

Lane, P. W. (2013). Meta-analysis of incidence of rare events. *Statistical Methods in Medical Research*, **22**(2):117–132.

Li, J. X., Chen, W.-C., and Scott, J. A. (2016). Addressing prior-data conflict with empirical meta-analytic-predictive priors in clinical studies with historical information. *Journal of Biopharmaceutical Statistics*, **26**(6):1056–1066.

Lijoi, A. (2003). Approximating priors by finite mixtures of conjugate distributions for an exponential family. *Journal of Statistical Planning and Inference*, **113**(2):419–435.

Lyman, G. H. and Kuderer, N. M. (2005). The strengths and limitations of meta-analyses based on aggregate data. *BMC Medical Research Methodology*, **5**:14.

Mao, S. and Xia, J. (1992). The hierarchical Bayesian analysis of the zero-failure data. *Applied Mathematics — A Journal of Chinese Universities*, **7**:411–421.

Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2012). Approximate Bayesian computational methods. *Statistics and Computing*, **22**(6):1167–1180.

McCulloch, C. E. and Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science*, **26**(3): 388–402.

Mehta, C. R. and Patel, N. R. (1995). Exact logistic regression: Theory and examples. *Statistics in Medicine*, **14**(19):2143–2160.

Neil, M., Tailor, M., Marquez, D., Fenton, N., and Hearty, P. (2008). Modelling dependable systems using hybrid Bayesian networks. *Reliability Engineering & System Safety*, **93**(7):933–939.

Neuenschwander, B., Roychoudhury, S., and Schmidli, H. (2016a). On the use of co-data in clinical trials. *Statistics in Biopharmaceutical Research*, **8**(3):345–354.

Neuenschwander, B., Wandel, S., Roychoudhury, S., and Bailey, S. (2016b). Robust exchangeability designs for early phase clinical trials with multiple strata. *Pharmaceutical Statistics*, **15**(2):123–134.

Neuhaus, J. M. and Jewell, N. P. (1993). A geometric approach to assess bias due to omitted covariates in generalized linear models. *Biometrika*, **80**(4):807–815.

Nissen, S. E. and Wolski, K. (2007). Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. *New England Journal of Medicine*, **356**(24):2457–2471.

Ohlssen, D. (2014). Points for consideration in Bayesian mixed treatment comparison of safety endpoints: An expert viewpoint. Paper presented at the American Conference of Pharmacometrics 5, Las Vegas, October 14th, 2014.

Parmar, M. K., Torri, V., and Stewart, L. (1998). Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine*, **17**(24):2815–2834.

Partlett, C. and Riley, R. D. (2017). Random effects meta-analysis: Coverage performance of 95% confidence and prediction intervals following REML estimation. *Statistics in Medicine*, **36**(2):301–317.

Pearson, K., Lee, A., and Bramley-Moore, L. (1899). Mathematical contributions to the theory of evolution. VI. Genetic (reproductive) selection: Inheritance of fertility in man, and of fecundity in thoroughbred racehorses. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **192**:257–330.

Piepho, H.-P., Williams, E. R., and Madden, L. V. (2012). The use of two-way linear mixed models in multitreatment meta-analysis. *Biometrics*, **68**(4):1269–1277.

Platt, R. W., Leroux, B. G., and Breslow, N. (1999). Generalized linear mixed models for meta-analysis. *Statistics in Medicine*, **18**(6):643–654.

Pocock, S. J. (1976). The combination of randomized and historical controls in clinical trials. *Journal of Chronic Diseases*, **29**(3):175–188.

Riley, R. D. and Steyerberg, E. W. (2010). Meta-analysis of a binary outcome using individual participant data and aggregate data. *Research Synthesis Methods*, **1**(1): 2–19.

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. *Statistics in Medicine*, **27**(11):1870–1893.

Robert, C. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation*, 2nd edition. Springer, New York.

Royal Statistical Society working party on statistical issues in first-in-man studies (2007). Report of the working party on statistical issues in first-in-man studies. Available: http://www.rss.org.uk/Images/PDF/publications/rss-reports-statistical-issues-first-in-man-studies-2007.pdf [accessed: 15 July 2016].

Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., and Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. *Biometrics*, **70**(4):1023–1032.

Schulz, K. F., Altman, D. G., and Moher, D. (2010). CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials. *Annals of Internal Medicine*, **152**(11):726–732.

Senn, S. (1994). Testing for baseline balance in clinical trials. *Statistics in Medicine*, **13**(17):1715–1726.

Senn, S. (2007). Trying to be precise about vagueness. *Statistics in Medicine*, **26**(7): 1417–1430.

Senn, S. (2008). Lessons from TGN1412 and TARGET: Implications for observational studies and meta-analysis. *Pharmaceutical Statistics*, **7**(4):294–301.

Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, **13**(2):238–241.

Smith, T. C., Spiegelhalter, D. J., and Thomas, A. (1995). Bayesian approaches to random-effects meta-analysis: A comparative study. *Statistics in Medicine*, **14**(24): 2685–2699.

Spiegelhalter, D. J., Abrams, K. R., and Myles, J. P. (2004). *Bayesian approaches to clinical trials and health-care evaluation.* Wiley, Chichester.

Stewart, L. A. and Clarke, M. J. (1995). Practical methodology of meta-analyses (overviews) using updated individual patient data. *Statistics in Medicine*, **14**(19): 2057–2079.

Stijnen, T., Hamza, T. H., and Özdemir, P. (2010). Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Statistics in Medicine*, **29**(29):3046–3067.

Strom, B. L., Buyse, M., Hughes, J., and Knoppers, B. M. (2014). Data sharing, year 1 — Access to data from industry-sponsored clinical trials. *New England Journal of Medicine*, **371**(22):2052–2054.

Sun, W., Laughren, T. P., Zhu, H., Hochhaus, G., and Wang, Y. (2013). Development of a placebo effect model combined with a dropout model for bipolar disorder. *Journal of Pharmacokinetics and Pharmacodynamics*, **40**(3):359–368.

Sutton, A. J., Kendrick, D., and Coupland, C. A. (2008). Meta-analysis of individual-and aggregate-level data. *Statistics in Medicine*, **27**(5):651–669.

Sutton, A. J., Cooper, N. J., Lambert, P. C., Jones, D. R., Abrams, K. R., and Sweeting, M. J. (2002). Meta-analysis of rare and adverse event data. *Expert Review of Pharmacoeconomics & Outcomes Research*, **2**(4):367–379.

Symons, M. J. and Moore, D. (2002). Hazard rate ratio and prospective epidemiological studies. *Journal of Clinical Epidemiology*, **55**(9):893–899.

Taichman, D. B., Backus, J., Baethge, C., Bauchner, H., de Leeuw, P. W., Drazen, J. M., Fletcher, J., Frizelle, F. A., Groves, T., Haileamlak, A., James, A., Laine, C., Peiperl, L., Pinborg, A., Sahni, P., and Wu, S. (2016). Sharing clinical trial data — A proposal from the international committee of medical journal editors. *New England Journal of Medicine*, **374**(4):384–386.

Tierney, J. F., Stewart, L. A., Ghersi, D., Burdett, S., and Sydes, M. R. (2007). Practical methods for incorporating summary time-to-event data into meta-analysis. *Trials*, **8**: 16.

Turner, R. M., Jackson, D., Wei, Y., Thompson, S. G., and Higgins, J. P. T. (2015). Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis. *Statistics in Medicine*, **34**(6):984–998.

Bibliography

U.S. FDA Center for Devices and Radiological Health (2010). Guidance for the use of Bayesian statistics in medical device clinical trials. Available: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071072.htm [accessed: 16 July 2014].

Veyrat-Follet, C., Bruno, R., Olivares, R., Rhodes, G. R., and Chaikin, P. (2000). Clinical trial simulation of docetaxel in patients with cancer as a tool for dosage optimization. *Clinical Pharmacology & Therapeutics*, **68**(6):677–687.

Warren, F. C., Abrams, K. R., Golder, S., and Sutton, A. J. (2012). Systematic review of methods used in meta-analyses where a primary outcome is an adverse or unintended event. *BMC Medical Research Methodology*, **12**:64.

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. *Statistics in Medicine*, **33**(14):2449–2466.

Weaver, J., Ohlssen, D., and Li, J. X. (2016). Strategies on using prior information when assessing adverse events. *Statistics in Biopharmaceutical Research*, **8**(1):106–115.

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. *Statistics in Medicine*, **34**(21):2881–2898.

Whitehead, A. and Whitehead, J. (1991). A general parametric approach to the meta-analysis of randomized clinical trials. *Statistics in Medicine*, **10**(11):1665–1677.

Williamson, P. R., Smith, C. T., Hutton, J. L., and Marson, A. G. (2002). Aggregate data meta-analysis with time-to-event outcomes. *Statistics in Medicine*, **21**(22):3337–3351.

Woods, B. S., Hawkins, N., and Scott, D. A. (2010). Network meta-analysis on the log-hazard scale, combining count and hazard ratio statistics accounting for multi-arm trials: A tutorial. *BMC Medical Research Methodology*, **10**:54.

Wunsch, H., Linde-Zwirble, W. T., Harrison, D. A., Barnato, A. E., Rowan, K. M., and Angus, D. C. (2009). Use of intensive care services during terminal hospitalizations in England and the United States. *American Journal of Respiratory and Critical Care Medicine*, **180**(9):875–880.

Yule, G. U. (1903). Notes on the theory of association of attributes in statistics. *Biometrika*, **2**:121–134.

Zhu, L., Yao, B., Xia, H. A., and Jiang, Q. (2016). Statistical monitoring of safety in clinical trials. *Statistics in Biopharmaceutical Research*, **8**(1):88–105.

Zinman, B., Wanner, C., Lachin, J. M., Fitchett, D., Bluhmki, E., Hantel, S., Mattheus, M., Devins, T., Johansen, O. E., Woerle, H. J., Broedl, U. C., and Inzucchi, S. E. (2015). Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *New England Journal of Medicine*, **373**(22):2117–2128.

List of symbols and abbreviations

$\mathbb{1}\{\ldots\}$	indicator function
$a_j,\ b_j$	parameters of the conjugate prior for exponentially distributed failure times from two treatment groups ($j = 0, 1$)
$a_{r,j},\ b_{r,j}$	parameters of mixture component r of a mixture of conjugate priors for exponentially distributed failure times for group j
$a_{ij},\ b_{ij}$	parameters of the conjugate prior for exponentially distributed failure times for trial $i = 1, \ldots, I$ and treatment group $j = 0, 1$
$a_{r,ij},\ b_{r,ij}$	parameters of mixture component r of a mixture of conjugate priors for exponentially distributed failure times for trial i and group j
AD	aggregate data
aRMSE	average RMSE across different (sub-)scenarios
as $x \searrow x_0$	as x decreases in value approaching x_0
$\boldsymbol{b}_{ijk}$	covariates for patient k in group j of trial i to be used in modeling events times
$\mathrm{Ber}(\pi)$	Bernoulli distribution with parameter π
$\beta,\ \boldsymbol{\beta}$	symbol for a generic regression coefficient
$\mathrm{Beta}(a, b)$	Beta distribution with parameter a and b
$\mathrm{Bin}(n, \pi)$	binomial distribution with parameters n and π
$C_{ijk},\ c_{ijk}$	drop-out time for patient k in treatment group j of study i
$\mathrm{Ca}(\mu, \sigma)$	Cauchy distribution with location parameter μ and scale parameter σ

cdf	cumulative distribution function
CI	confidence or credible interval
CMH	Cochran-Mantel-Haenszel (test, statistic or odds ratio estimate)
CONSORT	consolidated standards of reporting trials
$\mathcal{D}$	observed data for a meta-analysis
$\mathcal{D}'$	observed historical control group data
$\mathcal{D}_i$	observed data for trial i
$\mathcal{D}_{ij}$	observed data for treatment group j of trial i
$\mathcal{D}'_{i0}$	observed control group data for historical trial i
D_{ijk}, d_{ijk}	follow-up until censoring for patient k in treatment group j of study i
$\tilde{d}_{ij.}$	approximation to $d_{ij.}$ (see Section 3.1.2.1)
Dbl.-Exp.	double-exponential or Laplace (distribution)
δ	risk or rate difference
$\mathbb{E}$	expectation
$\mathbb{E}_{\theta=\theta_0}(Y)$	expectation of Y when the parameter θ of the distribution of Y has value θ_0
$\mathbb{E}(Y\|X = x)$	conditional expectation of Y given $X = x$
E_{ijk}, e_{ijk}	total number of events for patient k in group j of study i
$\mathrm{Exp}(\lambda)$	exponential distribution with rate parameter λ
F,G,H	cumulative distribution function, usually for event times, drop-out times and their joint distribution, respectively, when a single distribution is assumed that only differs across treatment groups $j = 0, 1$ and trials $i = 1, \dots, I$ with respect to parameters $\boldsymbol{\theta}_{F,ij}$, $\boldsymbol{\theta}_{G,ij}$ and $\boldsymbol{\theta}_{H,ij}$
F_{ij},G_{ij},H_{ij}	cumulative distribution function, usually for event times, drop-out times and their joint distribution, respectively, for group j of trial i

f,g,h	pdf (or pmf), usually for event times, drop-out times and their joint distribution, respectively, when a single distribution is assumed that only differs across groups $j = 0, 1$ and trials $i = 1, \ldots, I$ with respect to parameters $\boldsymbol{\theta}_{F,ij}$, $\boldsymbol{\theta}_{G,ij}$ and $\boldsymbol{\theta}_{H,ij}$
f_{ij},g_{ij},h_{ij}	pdf (or pmf), usually for event times, drop-out times and their joint distribution, respectively, for group j of trial i
f_{Wb}, F_{Wb}	pdf and cdf of the Weibull distribution
FDA	United States Food and Drug Administration
γ_{ij}	Weibull shape parameter for event times in group j of trial i
Γ	Γ function
H	number of historical trials
h_i	indicator for whether trial $i = 1, \ldots, I + H$ is a historical trial
$\mathrm{HCa}(0, \sigma)$	half-Cauchy distribution with location 0 and scale parameter σ
$\mathrm{HN}(0, \sigma)$	half-normal distribution with mean 0 and standard deviation σ
HS	horseshoe (prior) distribution
$\mathrm{HT}_\nu(0, \sigma)$	half-t distribution with ν degrees of freedom, location 0 and scale parameter σ
i	index for study
I	number of trials in meta-analysis
i.i.d.	independent and identically distributed
$\int_A f(\mathbf{x})\,\mathrm{d}\mathbf{x}$	integration for a measurable function f with respect to the Lebesgue measure on $A \subseteq \mathbb{R}^D$
IPD	individual patient data (e.g. y_{ijk} and t_{ijk} for each patient)
j	index for treatment group with (0 = control and 1 = test group)
k	index for patient number
$\mathcal{L}(\theta|\mathcal{D})$	likelihood of parameter θ given data $\mathcal{D}$
λ_{ij}	exponential rate parameter or inverse Weibull scale parameter for event times in treatment group j of trial i

$\lambda_{ij}(s)$	hazard rate in treatment group j of trial i at time $s \in [0, \tau_i]$
M_{ijk}, m_{ijk}	indicator for whether patient k in treatment group j of study i has an observed fatal event
MLE	maximum likelihood estimate
MAC	meta-analytic combined (approach or prior)
MACE	major adverse cardiovascular events
MAP	meta-analytic predictive (distribution, approach or prior)
MAR	missing at random
MCMC	Markov chain Monte Carlo
mgf	moment generating function
μ_{ij}	exponential rate parameter or inverse Weibull scale parameter for drop-out times in treatment group j of trial i
$N(\nu, \sigma)$	normal distribution with mean ν and standard deviation σ
$N(\boldsymbol{\nu}, \boldsymbol{\Sigma})$	multivariate normal distribution with mean vector $\boldsymbol{\nu}$ and covariance matrix $\boldsymbol{\Sigma}$
n_{ij}	number of patients in treatment group j of study i
ν	mean of a normally distributed random effect
NUTS	no-U-turn (Hamiltonian MCMC) sampler
$f(x) = \mathcal{O}(g(x))$ as $x \to x_0$	there exist $\epsilon > 0$ and $\delta > 0$ so that for all x that satisfy $0 < |x - x_0| < \epsilon$ we have $|f(x)| \leq \delta |g(x)|$
ODE	ordinary differential equation
ω_{ij}	Weibull shape parameter for drop-out times in group j of trial i
p	used to denote the number of random effects or the dimension of a (hyper-)parameter vector
$p(\theta)$	prior density of parameter θ
$p(\theta|\mathcal{D})$	posterior density of parameter θ given the data $\mathcal{D}$
$p(\mathcal{D}|\theta)$	pdf or pmf for data $\mathcal{D}$ given parameter θ, alternative notation for likelihood $\mathcal{L}(\theta|\mathcal{D})$ used in Bayesian context

$p(\boldsymbol{\theta}_i\|\boldsymbol{\vartheta})$	pdf or pmf for the conditional distribution of the latent trial-level variable $\boldsymbol{\theta}_i$ given the hyperparameter $\boldsymbol{\vartheta}$
$p(\boldsymbol{\theta};\boldsymbol{\vartheta})$	prior density for the parameter vector $\boldsymbol{\theta}$ with hyperparameter vector $\boldsymbol{\vartheta}$
$\mathbb{P}$	probability
pdf	probability density function
ϕ	hazard or odds ratio
$\phi(x;\nu,\sigma)$	pdf of a $N(\nu,\sigma)$ r.v.
$\phi(\boldsymbol{x};\boldsymbol{\nu},\boldsymbol{\Sigma})$	pdf of a $N(\boldsymbol{\nu},\boldsymbol{\Sigma})$ r.v.
π_{ij}	expected proportion of patients with an event in group j of trial i
π_{ijk}	probability of an observed event for patient k in group j of trial i
$\boldsymbol{p}_{ij}$	vector with the five components p_{ijr} for $r = 1,\ldots,5$
p_{ijr}	probability that a patient in group j of trial i will have the rth of the outcomes in Definition 3.1
pmf	probability mass function
$\psi(x)$	the digamma function $\psi(x) := \Gamma'(x)/\Gamma(x)$
$\psi_r(x)$	$\frac{\mathrm{d}^r}{\mathrm{d}x^r}\psi(x)$ polygamma function of order $r = 1, 2, \ldots$
$q_{ij},\, q_j$	probability that an event in group j of trial i is fatal given that it has occurred (q_j when the probability is assumed to be the same across all trials)
$\mathbb{R}$	the real numbers $(-\infty,\infty)$
$\mathbb{R}^+$	the non-negative real numbers $[0,\infty)$
RCT	randomized controlled trial
ρ_i	length of the recruitment period of trial $i = 1,\ldots,I$
rMAP	robust meta-analytic predictive
RMSE	root mean squared error
r.v.	random variable

SE	standard error
σ	standard deviation of a normally distribution random effect ("hierarchical scale parameter")
$\boldsymbol{\Sigma}$	covariance matrix of a multivariate normal random effect
$\square^T$	transposition of the matrix or vector $\square$
T_{ijk}, t_{ijk}	follow-up to first event or censoring for patient k in treatment group j of study i
$\tilde{t}_{ij.}$	approximation to $t_{ij.}$ (see Section 3.1.2.2)
θ, $\boldsymbol{\theta}$	symbol for a generic parameter or parameter vector
$\boldsymbol{\theta}_{F,ij}$	parameters of event time distribution F for group j of trial i
$\boldsymbol{\theta}_{G,ij}$	parameters of drop-out time distribution G for group j of trial i
$\boldsymbol{\theta}_{H,ij}$	parameters of joint distribution H of event and drop-out times for group j of trial i
$\boldsymbol{\theta}_i$	vector of latent trial-level parameters, for the proposed Bayesian hierarchical model $\boldsymbol{\theta}_{F,i0}$, $\boldsymbol{\theta}_{G,i0}$ and $\boldsymbol{\theta}_{G,i1}$
$\boldsymbol{\theta}_{ij}$	trial arm specific parameters, for the proposed model typically $\boldsymbol{\theta}_{F,ij}$, $\boldsymbol{\theta}_{G,ij}$ and q_{ij}
τ_i	planned duration of trial i (time of administrative censoring); time from start of recruitment to trial end in an event driven trial
$\mathrm{U}(a,b)$	uniform distribution on $[a,b]$
Var	variance
φ	log-hazard or log-odds ratio $\varphi := \log \phi$
$\boldsymbol{\vartheta}$	hyperparameters determining the distribution of $\boldsymbol{\theta}_i$ across trials
$\mathrm{Wb}(a,b)$	Weibull distribution with inverse scale parameter a and shape parameter b
W_{ijr}, w_{ijr}	The number of patients in group j of trial i with the rth of the outcomes in Definition 3.1
X_{ijk}, x_{ijk}	event time for patient k in treatment group j of study i

Y_{ijk}, y_{ijk}	indicator for whether patient k in treatment group j of study i has an observed event
Z_{ijk}, z_{ijk}	indicator for whether patient k in treatment group j of study i is a drop-out
$\boldsymbol{\zeta}$	hyperparameters that do not define random effects
$\boldsymbol{\zeta}_j$	component of the hyperparameter vector $\boldsymbol{\zeta}$ corresponding to treatment group j

List of Figures

List of Tables

Appendices

Appendix A.

Prior for the log-hazard ratio

When evaluating whether one intervention is superior to another it is usually seen as inappropriate to use a prior distribution that favors one of the interventions (Gelman et al., 2014, p. 56). Similarly, when evaluating whether an intervention is non-inferior to another, a prior that strongly favors a conclusion of no treatment difference is problematic. With this in mind, one possible approach is to use a weakly informative prior that is symmetrical around a "neutral" parameter value representing "no treatment effect" and is more vague than we consider a-priori plausible based on either the historical distribution of treatment effects, or based on considerations of biological plausibility.

A Cauchy$(0, 0.37)$ prior puts 90% of the prior probability mass within $(-\log(10), \log(10))$ reflecting a mildly skeptical prior belief that extremely large drug effects with relative risk measures such as hazard, rate, risk, or odds ratios above 10 or below 0.1 are rare (Greenland, 2000; Hamra et al., 2013). In fact, this prior puts more probability in the tail area than the distribution of the estimated hazard ratios for new drugs versus controls from several sets of historical trials (Djulbegovic et al., 2012) shown in Figure A.1. This distribution reflects not just the distribution of actual treatment effect sizes, but also of sampling variation. For this reason priors developed on this basis have been criticized for depending on the sampling variation in historical trials (Senn, 2007). Kault and Kault (2015) have attempted to estimate the distribution of intrinsic drug effects free of sampling variation in the Cochrane database of systematic reviews. On

the other hand, a prior centered around no effect that is as wide or wider than the distribution of estimated treatment effects in actual trials — such as a $\mathrm{Cauchy}(0, 0.37)$ prior — can be considered as a weakly informative prior or as appropriately reflecting the uncertainty around the exchangeability of the relative risk measures between the historical comparisons and the comparison under consideration. As a sensitivity analysis, one may consider the default weakly informative $\mathrm{Ca}(0, 2.5)$ prior proposed by Gelman et al. (Gelman et al., 2008) with 90% of the prior probability mass within $(-\log(108), \log(108))$. $\mathrm{Ca}(0, 0.37)$ and $\mathrm{Ca}(0, 2.5)$ distributions are the main priors we consider throughout this thesis. Program code for this appendix is available under https://doi.org/10.6084/m9.figshare.5007833.
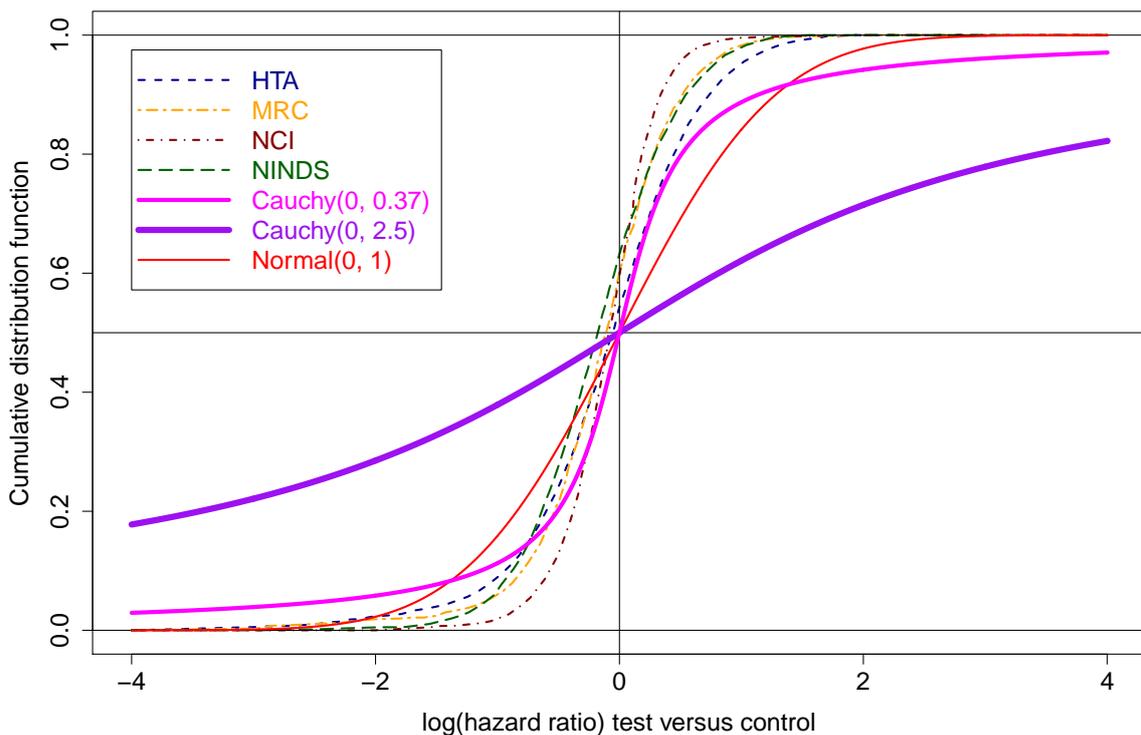
Figure A.1.: Cumulative distribution function for the estimated hazard ratios for test groups compared with controls in all 624 trials from 1955 to 2006 with 781 comparisons by the US National Cancer Institute (NCI), all 63 superiority trials from 1999 to 2008 with 94 comparisons by the UK Health Technology Assessment (HTA) Programme, all 28 trials from 1973 to 1994 with 31 comparisons by the United Kingdom Medical Research Council (MRC), all 28 trials from 1984 to 2003 with 32 comparisons by the US National Institute for Neurological Disorders and Stroke (NINDS) (Djulbegovic et al., 2012), two Cauchy distributions with location 0 and scale parameter 0.37 or 2.5, and the $\mathrm{Normal}(0, 1)$ distribution

Appendix B.

Weakly informative hyperpriors for the rosiglitazone example

For weakly informative hyperpriors $p(\nu_1)$ and $p(\sigma_1)$ for the placebo group MACE hazard rate we used the observational data reported in the drug sponsor's briefing document for the 2010 rosiglitazone FDA advisory committee meeting(Glaxo-Smith-Kline, 2010). The sponsor provided three estimates from different observational databases for the MACE hazard rate per patient-year in type 2 diabetes patients on control therapies: 0.0142 based on 38 patients with an event, 0.0135 based on 13 patients with an event and 0.0140 based on 51 patients with an event(Glaxo-Smith-Kline, 2010). The inverse variance weighted average of these estimates combined on the log-scale is -4.27 corresponding to a rate of 0.0140 per patient-year.

For weakly informative hyperpriors $p(\nu_2)$, $p(\sigma_2)$, $p(\nu_3)$ and $p(\sigma_3)$ regarding drop-out hazard rates in the control and test groups, we considered 20% of drop-outs by the end of 1 year as a plausible average proportion of drop-outs corresponding to an average exponential hazard rate for drop-out of 0.22.

As it seemed unlikely that these estimates would be incorrect by a factor substantially above 10, we used a normal prior for the mean of the random effects with these means and a standard deviation of log(10). We considered this as a weakly informative prior that understates our prior beliefs. This is illustrated in Figure B.1 for the MACE hazard
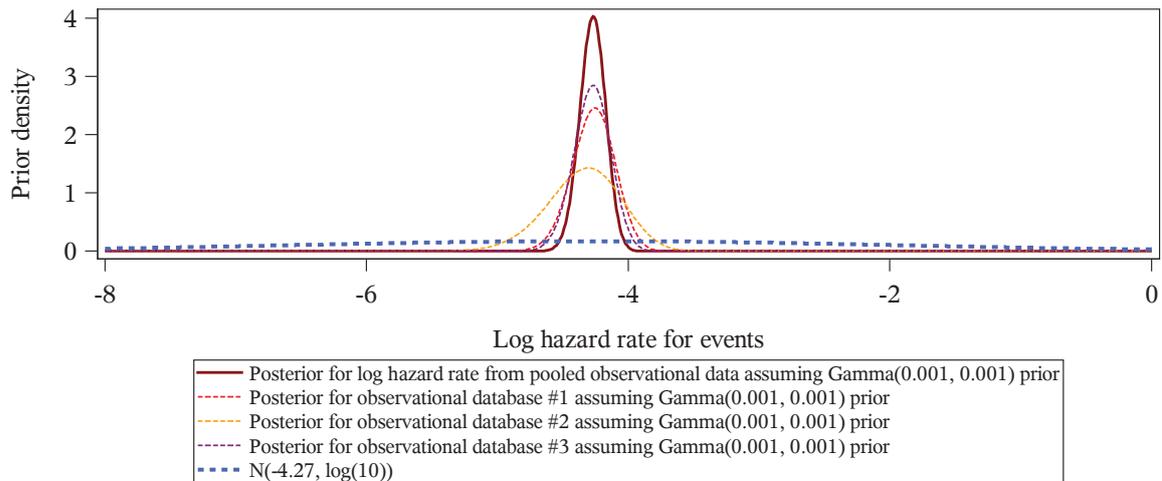
Figure B.1.: Weakly informative prior compared to informative priors based on possible analyses of the observational databases

rate that shows that for the MACE log-hazard rate, the employed prior is substantially less informative than the posterior from an analysis of the observational data, which can be considered reasonable given the potential wide range of differences between observational data and data from RCTs including in terms of patient characteristics, outcome definitions and outcome reporting.

Program code for the results in this Appendix is available under https://doi.org/10.6084/m9.figshare.5007833.

Appendix C.

Additional results from the simulations in Section 3.6
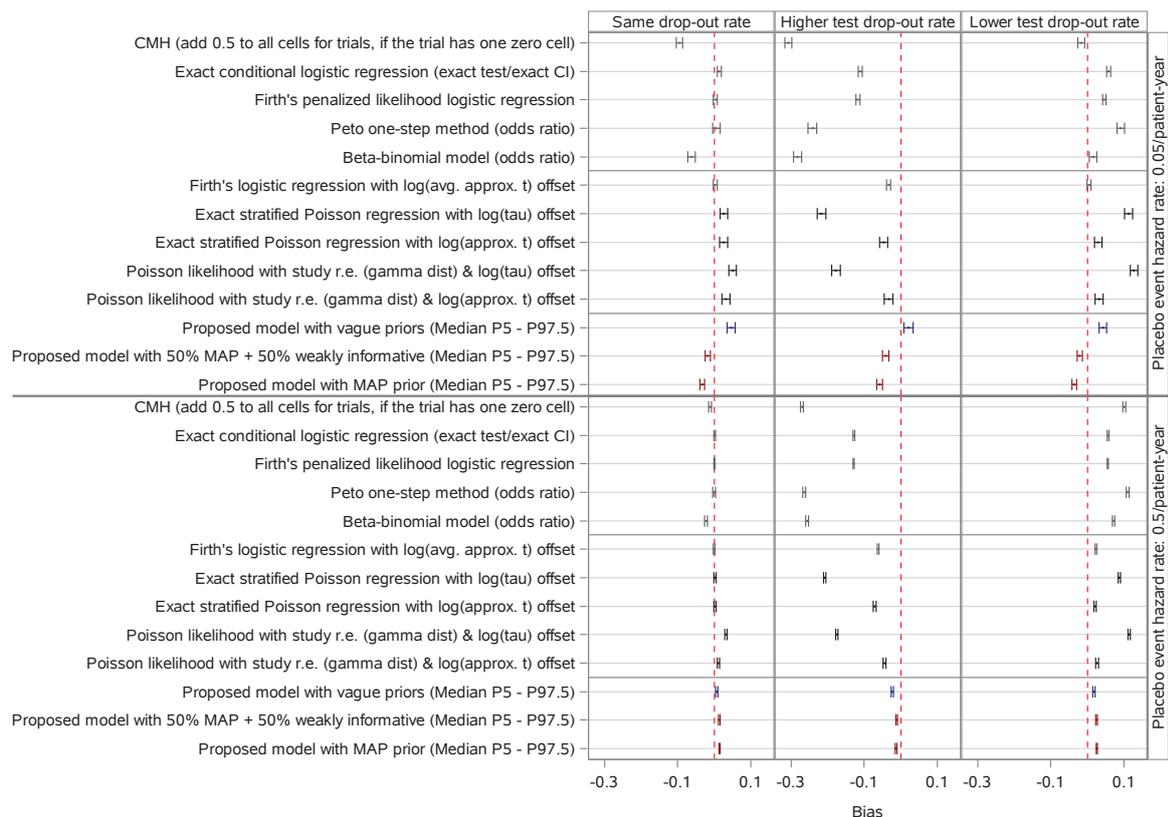
Figure C.1.: Bias in the estimated log-hazard ratio in the absence of a treatment effect with 99% CIs based on 10,000 simulations under different scenarios: control and test group hazard rate for events (0.05 or 0.5; no treatment effect), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)

For the lower placebo event hazard rate (0.05 per patient-year) nearly all simulations

(>90%) included at least one trial arm and often (10 to 30% of simulations) at least a whole trial without any events, while this was hardly ever the case for the higher placebo event hazard rate of 0.5 per patient-year. There was no simulation without any events and a single case of complete separation (no events in one treatment group across all trials of a simulation) for scenario 2. For 14 simulated meta-analyses the potential scale reduction factor for the proposed Bayesian hierarchical model with vague priors remained > 1.1 for some variables despite repeated runs of the model for these scenarios. We used the estimates and CIs obtained from the Bayesian hierarchical model with vague priors for these 14 cases for the evaluation of the method despite their questionable convergence status. We do not expect this to affect the overall conclusions of the simulation study, because only 14 out of 1,440,000 ($< 0.001\%$) simulations were affected.
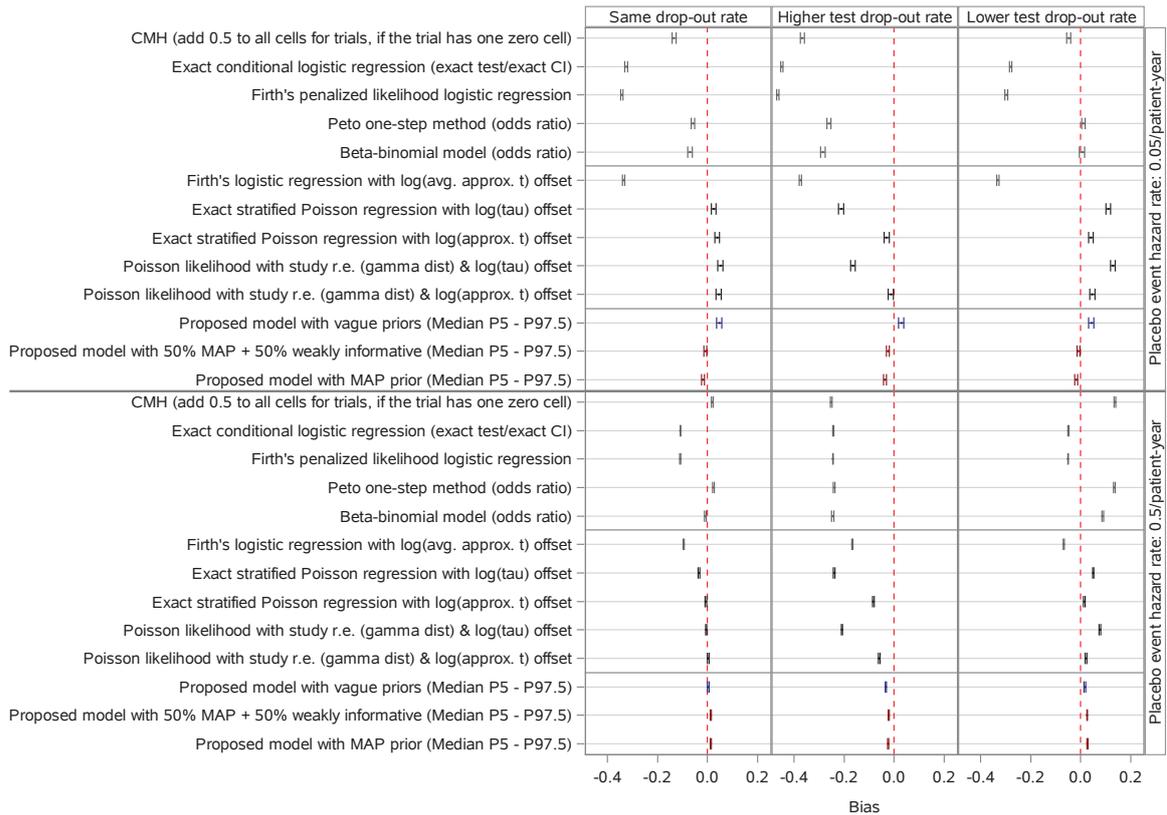
Figure C.2.: Bias in the estimated log-hazard ratio in the presence of a treatment effect with 99% CIs based on 10,000 simulations under different scenarios: control group hazard rate for events (0.05 or 0.5), test group hazard rate for events (0.1 or 0.64), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)

As shown in Figure C.1 and Figure C.2, the proposed Bayesian hierarchical model with

or without informative priors provided the least biased estimates across the evaluated scenarios. As seen in Figure C.2, exponential time-to-event analyses using a Poisson likelihood and a $\log \tilde{t}_{ij.}$ offset primarily performed worse than the proposed Bayesian hierarchical model in the presence of differential drop-out. Binomial-distribution-based methods with the exception of the CMH odds ratio with continuity correction and the beta-binomial model performed relatively well in the absence of differential drop-out under the null hypothesis of no treatment effect as seen in Figure C.1, while as seen in Figure C.2 in the presence of a differential drop-out or a treatment effect all methods that estimate an odds ratio were clearly biased. This considerable bias of methods that estimate an odds ratio in the presence of a treatment effect reflects the inconsistent effect measures — i.e. that were are attempting to estimate a hazard ratio by an odds ratio. The CMH odds ratio with continuity correction and the beta-binomial model resulted in biased estimates even in the absence of a treatment effect and in the absence
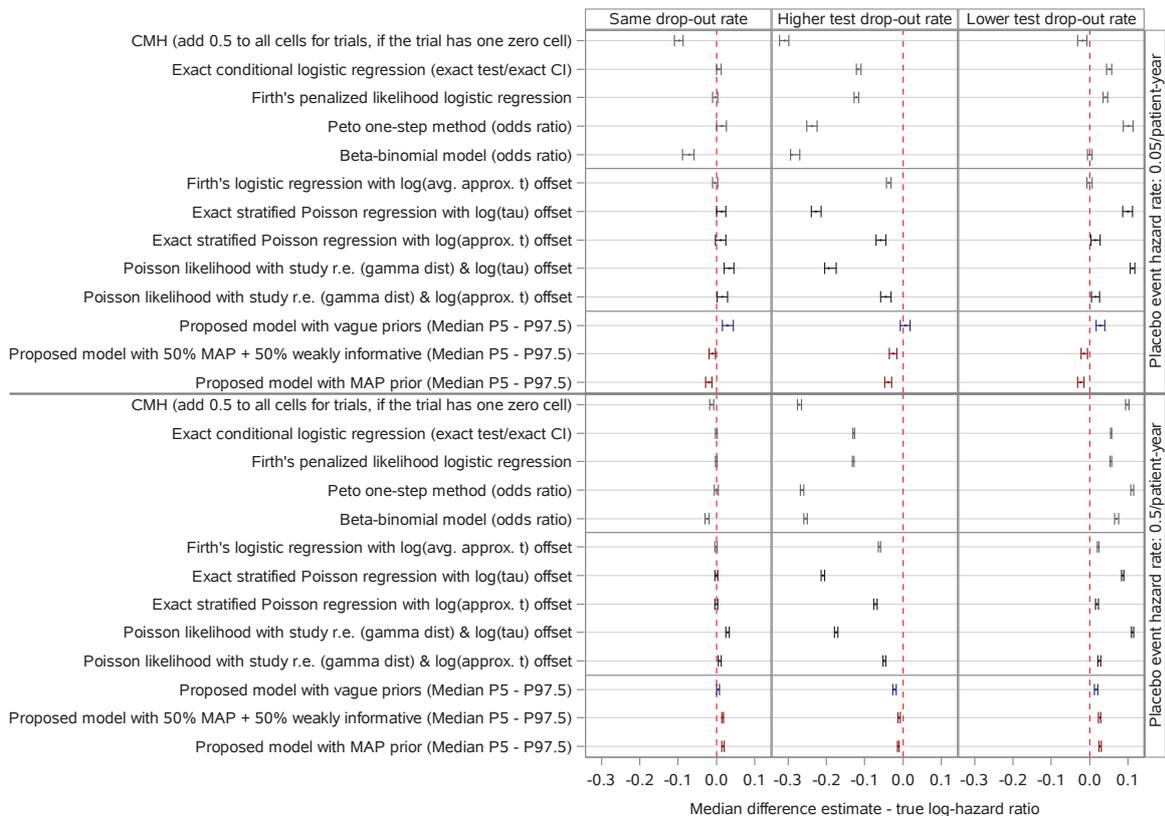
Figure C.3.: Median difference between estimate and true value of log-hazard ratio in the absence of a treatment effect with 99% distribution-free CIs based on 10,000 simulations under different scenarios: control and test group hazard rate for events (0.05 or 0.5; no treatment effect), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)
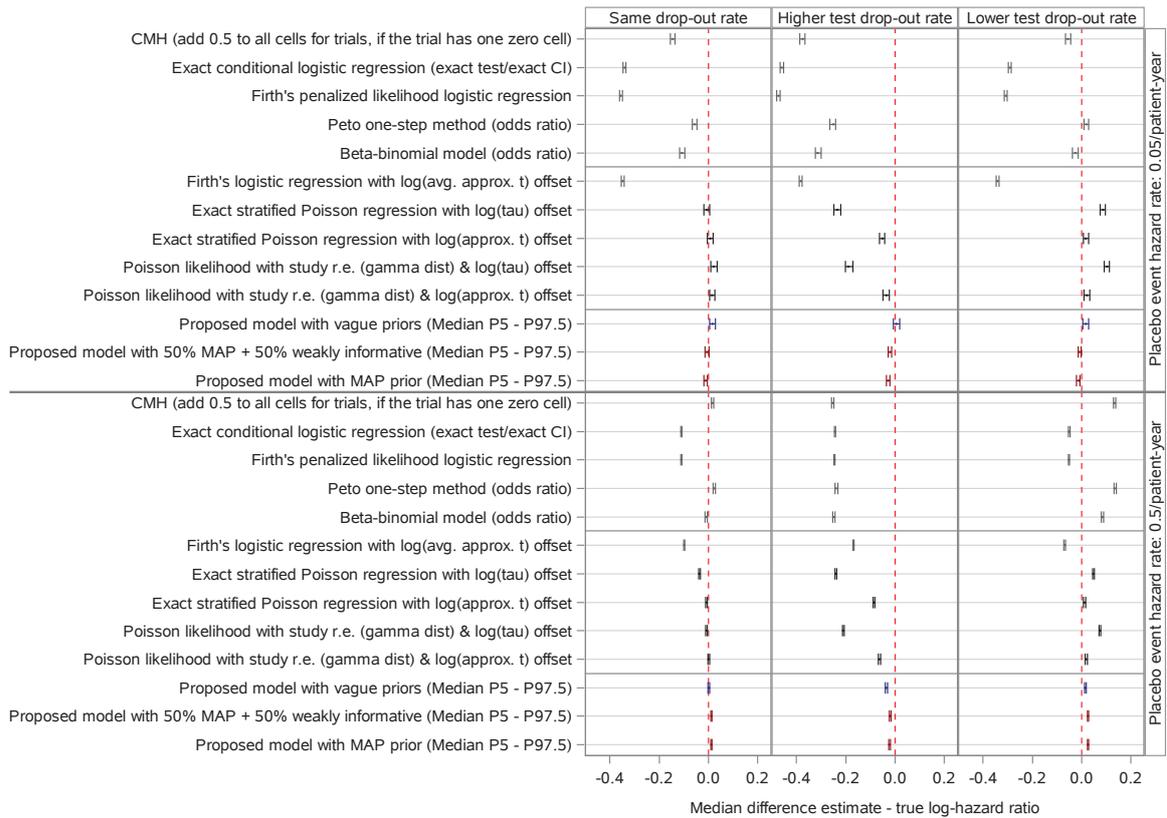
Figure C.4.: Median difference between estimate and true value of log-hazard ratio in the presence of a treatment effect with 99% distribution-free CIs based on 10,000 simulations under different scenarios: control group hazard rate for events (0.05 or 0.5), test group hazard rate for events (0.1 or 0.64), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)

of differential drop-out.

To assess median-unbiasedness, Figure C.3 and Figure C.4 show the median difference between the estimated and true log-hazard ratio with distribution-free 99% confidence limits (Hahn and Meeker, 1991, pp. 82–89). In general, the patterns seen for bias were also seen for the median difference between estimate and true value of log-hazard ratio.

In terms of RMSE, the proposed Bayesian hierarchical model with robust historical priors performed consistently well across the considered scenarios as seen in Figure C.5 and Figure C.6. As expected — given that the historical data was generated under the same model and parameters as the data for the main meta-analysis — giving greater weight to the historical prior reduced the RMSE. The proposed Bayesian hierarchical model with vague priors, as well as exponential time-to-event analyses using a Poisson likelihood and a $\log \tilde{t}_{ij}$ offset performed relatively consistently, but in some scenarios
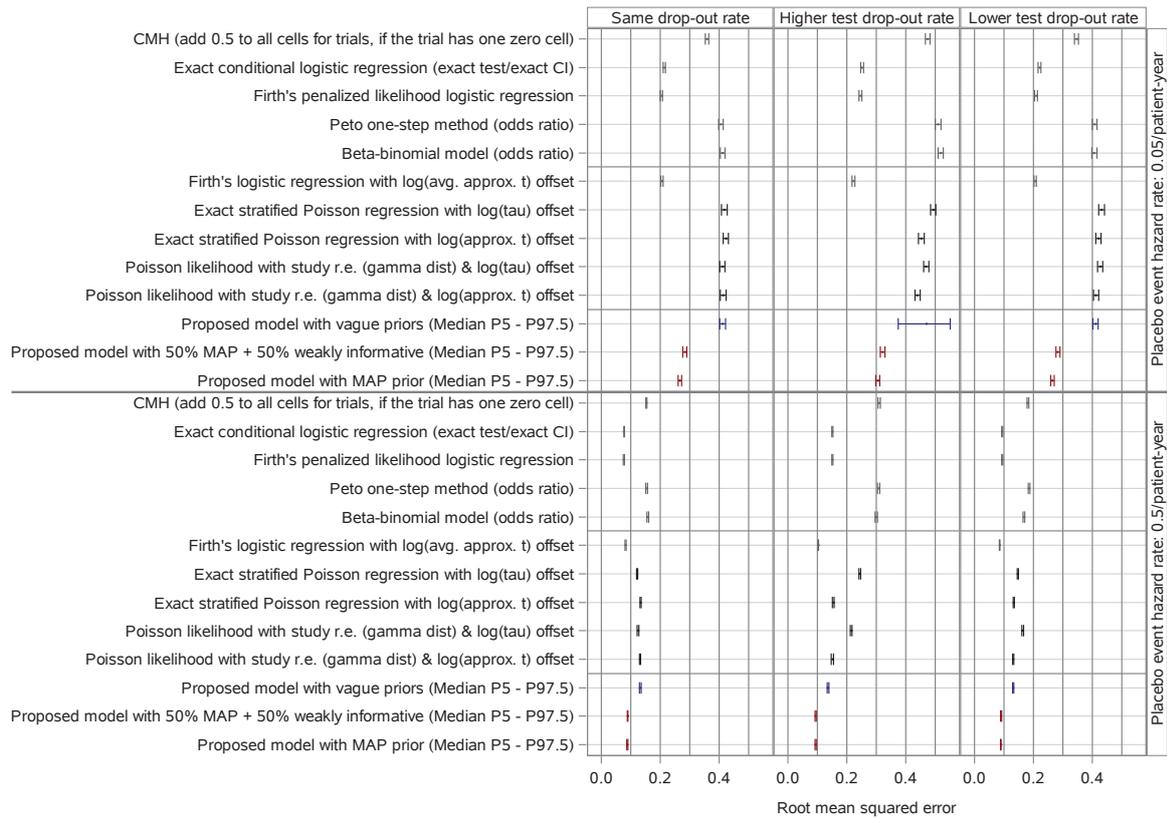
Figure C.5.: Root mean squared error in the estimated log-hazard ratio in the absence of a treatment effect with 99% CIs based on 10,000 simulations under different scenarios: control and test group hazard rate for events (0.05 or 0.5; no treatment effect), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)

other methods had a lower RMSE. In particular, exact and Firth's penalized likelihood logistic regression performed well under the null hypothesis as seen in Figure C.5, while in the presence of a treatment effect this was not the case (Figure C.6). This again partly reflects the inconsistent effect measure — i.e. that we are attempting to estimate a hazard ratio by an odds ratio. For the lower event hazard rate scenario (0.05 events/patient-year) and in the absence of differential drop-out, the Peto on-step method performed as well or better than the the proposed Bayesian hierarchical model with vague priors and exponential time-to-event methods, but less well for the higher event hazard rate scenario (0.5 events/patient-year) or in the presence of differential drop-out. The wide CIs for the proposed Bayesian hierarchical model with vague priors in the second scenario of Figure C.5 are due to an extreme estimate in a single simulation with complete separation.

The coverage of 95% confidence or credible intervals shown in Figure C.7 is in line with
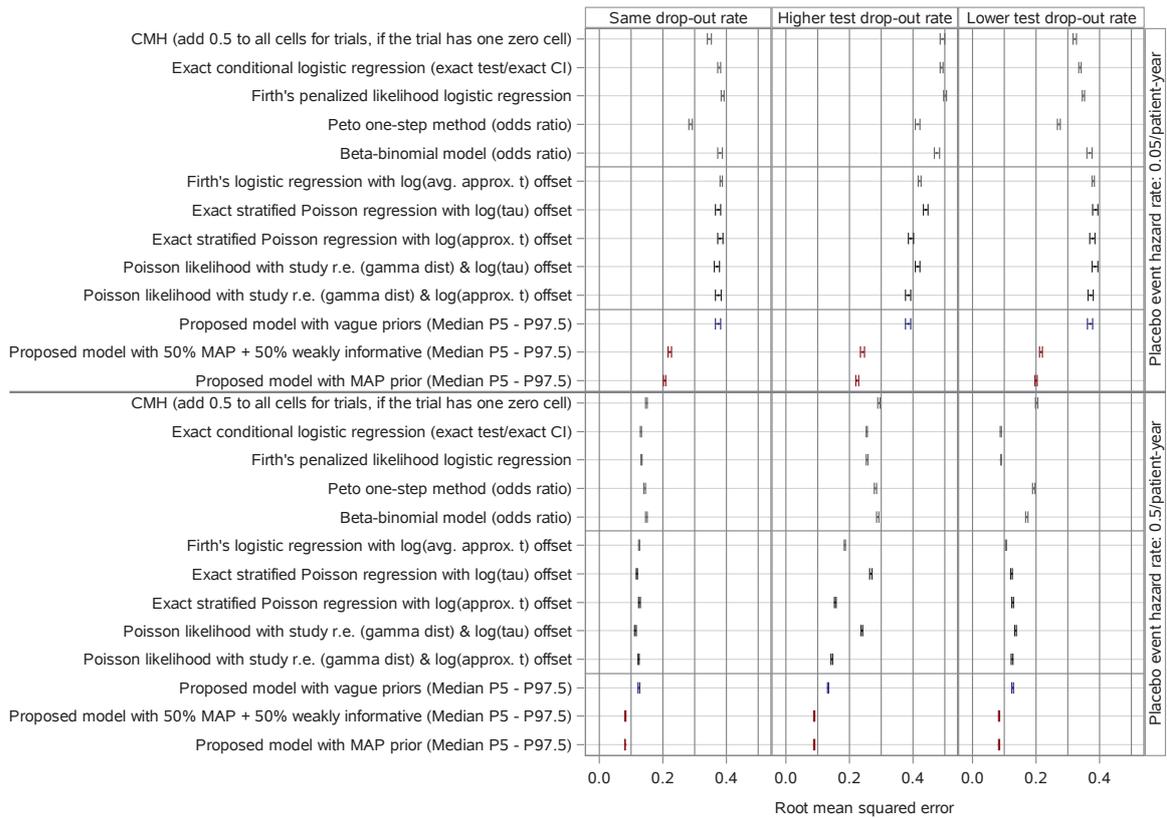
Figure C.6.: Root mean squared error in the estimated log-hazard ratio in the presence of a treatment effect with 99% CIs based on 10,000 simulations under different scenarios: control group hazard rate for events (0.05 or 0.5), test group hazard rate for events (0.1 or 0.64), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)

the results for the type 1 error rate. In line with the findings regarding bias, methods that estimate an odds ratio achieve poor coverage of 95% CIs in the presence of a treatment effect as seen in Figure C.8. Coverage probabilities for the beta-binomial model of Kuss (2015) are hard to see in some panels of Figures C.7 and C.8, because they are close to 100% due to very wide 95% CIs provided by this method. The proposed Bayesian hierarchical model with vague priors provided coverage at approximately the nominal level in all scenarios, while coverage was generally slightly above the nominal level with informative priors. Exponential time-to-event models using a Poisson likelihood and a log estimated follow-up to first event or censoring offset had below nominal coverage in the scenarios with a high hazard rate for events and a higher drop-out hazard rate on test treatment than in the control groups, while a log planned trial duration offset resulted in below nominal coverage in most scenarios with differential censoring.
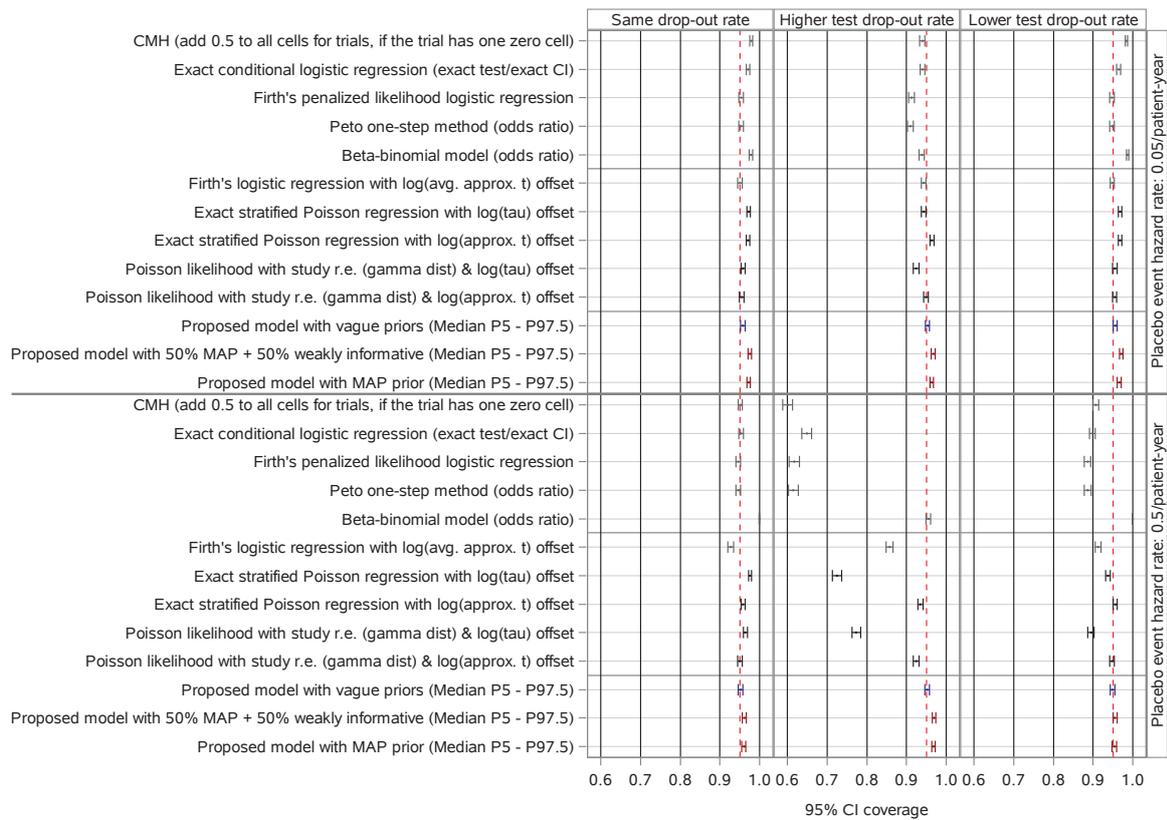
Figure C.7.: Coverage of 95% confidence or credible intervals for the log-hazard ratio in the absence of a treatment effect with 99% Clopper-Pearson CIs based on 10,000 simulations under different scenarios: control and test group hazard rate for events (0.05 or 0.5; no treatment effect), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)
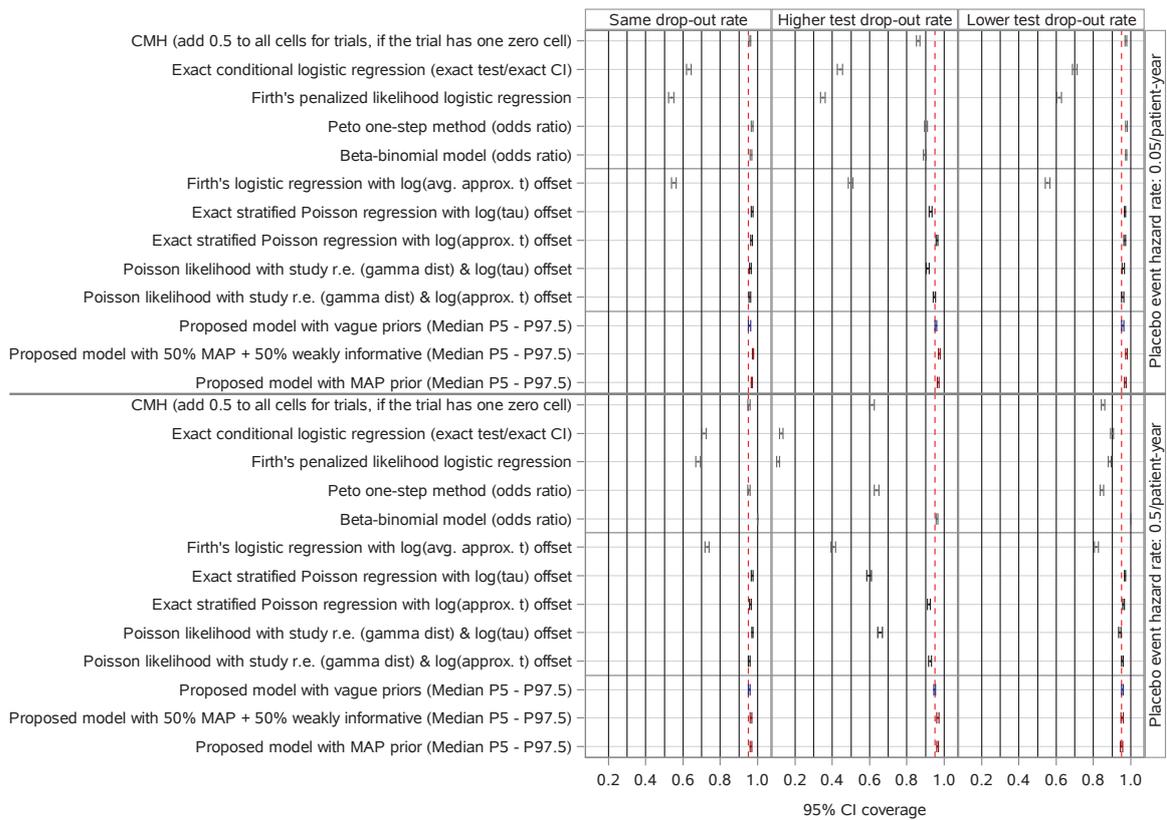
Figure C.8.: Coverage of 95% confidence or credible intervals for the log-hazard ratio in the presence of a treatment effect with 99% Clopper-Pearson CIs based on 10,000 simulations under different scenarios: control group hazard rate for events (0.05 or 0.5), test group hazard rate for events (0.1 or 0.64), control group hazard rate for drop-out (0.5) and test group hazard rate for drop-out (0.5, 1.36 or 0.18)

Appendix D.

Likelihood evaluation for Weibull distributed event and drop-out times

The numerical evaluation of the likelihood (3.2.4) with the probabilities p_{ij1}, p_{ij2}, p_{ij3}, p_{ij4} and p_{ij5} given in Theorem 4.1 when either event times, drop-out times or both follow a Weibull distribution requires numerical integration.

For the numerical evaluation of the definite integral in (4.1.1), one approach is to use the composite Simpson's rule that is based on Lagrange polynomial interpolation

$$\int_0^\tau f(x)\, \mathrm{d}x \approx \frac{\tau}{3R}\left(f(x_0) + f(x_R) + 4\sum_{r=1}^{R/2} f(x_{2r-1}) + 2\sum_{r=1}^{R/2-1} f(x_{2r})\right), \qquad \text{(D.1)}$$

where $R \geq 4$ is an even integer and the $x_0, \ldots, x_R$ denote evenly spaced points on $[0, \tau]$ with $x_0 = 0$ and $x_R = \tau$ (Burden and Faires, 2011, p. 206). To use this approach it is necessary to remove the singularity in the integrand that occurs at $x = 0$ for values of the parameter γ_{ij} that lie in $(0, 1)$. This can be achieved by integration by substitution using a change of variables to $u = x^{\gamma_{ij}}$

$$\lambda_{ij}^{\gamma_{ij}} \int_0^{\tau_i} \gamma_{ij} x^{\gamma_{ij}-1} \exp\left(-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\nu_{ij}}\right)\, \mathrm{d}x$$

$$=\lambda_{ij}^{\gamma_{ij}} \int_0^{\tau_i^{\gamma_{ij}}} \exp\left(-u\lambda_{ij}^{\gamma_{ij}} - u^{\nu_{ij}/\gamma_{ij}}\mu_{ij}^{\nu_{ij}}\right)\, \mathrm{d}u.$$

For the integral (D.1), it is challenging to derive error bounds for fixed values of R. Instead, one can use an adaptive Simpson's rule (Burden and Faires, 2011, p. 220–226). However, the use of this approach in Stan is problematic, because choice of R that depends on the parameter values causes issues with the automatic differentiation of the log-likelihood that Stan performs. An ad-hoc alternative may be to fit the model for different choices of R (e.g. 5 MCMC chains with $R = 25$ and 5 MCMC chains with $R = 50$), in which case the Rhat statistic across all chains can be used to diagnose whether the MCMC samples from all chains come from the same distribution (Sebastian Weber, personal communication, November 3, 2016).

Alternatively, we can represent the integral as the solution to an ordinary differential equation (ODE) and use the ODE solver provided in Stan to numerically evaluate the integral. For using the ODE solver, we would re-express the integral

$$\int_0^{\tau_i} e^{-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}}} x^{\gamma_{ij}-1} \, \mathrm{d}x \tag{D.2}$$

as the ODE

$$f'(x) = \exp\left[-(x\lambda_{ij})^{\gamma_{ij}} - (x\mu_{ij})^{\omega_{ij}} + (\gamma_{ij} - 1)\log(x)\right]$$

with the boundary condition $f(0) = 0$. Then the solution to the integral (D.2) is the solution of the ODE evaluated at τ_i (Burden and Faires, 2011, p. 304).

Appendix E.

Event driven trials with uniformly distributed recruitment times

Assuming that the trial entry times of patients are uniformly distributed across the recruitment period $[0, \rho_i]$ with pdf $p_i(r; \rho_i) := 1/\rho_i$, where $\rho_i < \tau_i$, we obtain

$$
\begin{aligned}
p_{ij1} &= q_j \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{\lambda_{ij}\mu_{ij}}{\rho_i} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(x \leq c)\mathbb{1}(x \leq \tau_i)\, \mathrm{d}x\, \mathrm{d}c\, \mathrm{d}r \\
&= q_j \Bigg(\int_0^{\rho_i} \underbrace{\int_r^{\tau_i} \frac{\mu_{ij}\left(e^{c\lambda_{ij}} - e^{\lambda_{ij}r}\right) e^{\mu_{ij}r - c(\lambda_{ij}+\mu_{ij})}}{\rho_i}\, \mathrm{d}c}_{= \frac{\mu_{ij}\left(e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)}-1\right)}{\rho_i(\lambda_{ij}+\mu_{ij})} - \frac{e^{\mu_{ij}(r-\tau_i)}-1}{\rho_i}}\, \mathrm{d}r \\
&\quad + \int_0^{\rho_i} \underbrace{\int_{\tau_i}^{\infty} \frac{\mu_{ij}\left(e^{\lambda_{ij}\tau_i} - e^{\lambda_{ij}r}\right) e^{-c\mu_{ij}-\lambda_{ij}\tau_i+\mu_{ij}r}}{\rho_i}\, \mathrm{d}c}_{= \frac{e^{\mu_{ij}(r-\tau_i)}-e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)}}{\rho_i}}\, \mathrm{d}r \Bigg) \\
&= \frac{q_j\lambda_{ij}}{\rho_i(\lambda_{ij}+\mu_{ij})} \int_0^{\rho_i} 1 - e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)}\, \mathrm{d}r \\
&= \frac{q_j\lambda_{ij}}{\lambda_{ij}+\mu_{ij}} \left(1 - \frac{\left(e^{\rho_i(\lambda_{ij}+\mu_{ij})} - 1\right) e^{-\tau_i(\lambda_{ij}+\mu_{ij})}}{\rho_i(\lambda_{ij}+\mu_{ij})} \right)
\end{aligned}
$$

for the probability that a patient experiences a fatal event of interest,

$$p_{ij2} = (1 - q_j) \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{\lambda_{ij}\mu_{ij}}{\rho_i} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(c > \tau) \mathbb{1}(x \leq \tau) \, dx \, dc \, dr$$

$$= \frac{(1 - q_j)}{\rho_i} \int_0^{\rho_i} \left(e^{\mu_{ij}r} - e^{(\lambda_{ij}+\mu_{ij})r - \lambda_{ij}\tau_i} \right) \int_{\tau_i}^{\infty} \mu_{ij} e^{-c\mu_{ij}} \, dc \, dr$$

$$= \frac{(1 - q_j)}{\rho_i} \int_0^{\rho_i} e^{\mu_{ij}(r-\tau_i)} - e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)} \, dr$$

$$= \frac{(1 - q_j)}{\rho_i} \left(\frac{\left(e^{\mu_{ij}\rho_i} - 1 \right) e^{-\mu_{ij}\tau_i}}{\mu_{ij}} - \frac{\left(e^{\rho_i(\lambda_{ij}+\mu_{ij})} - 1 \right) e^{-\tau_i - (\lambda_{ij}+\mu_{ij})}}{\lambda_{ij} + \mu_{ij}} \right)$$

for the probability that a patient experiences a non-fatal event and completes the trial,

$$p_{ij3} = (1 - q_j) \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{\lambda_{ij}\mu_{ij}}{\rho_i} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(c \leq \tau_i) \mathbb{1}(x \leq c) \, dx \, dc \, dr$$

$$= \frac{\mu_{ij}(1 - q_j)}{\rho_i} \int_0^{\rho_i} \int_r^{\tau_i} e^{\mu_{ij}(r-c)} - e^{(\lambda_{ij}+\mu_{ij})(r-c)} \, dc \, dr$$

$$= \frac{\mu_{ij}(1 - q_j)}{\rho_i} \int_0^{\rho_i} \frac{1 - e^{\mu_{ij}(r-\tau_i)}}{\mu_{ij}} - \frac{1 - e^{(\lambda_{ij}+\mu_{ij})(r-\tau_i)}}{\lambda_{ij} + \mu_{ij}} \, dr$$

$$= \mu_{ij}(1 - q_j) \left(\frac{\lambda_{ij}}{\mu_{ij}(\lambda_{ij}+\mu_{ij})} - \frac{\left(e^{\mu_{ij}\rho_i} - 1 \right) e^{-\mu_{ij}\tau_i}}{\rho_i \mu_{ij}^2} + \frac{\left(e^{\rho_i(\lambda_{ij}+\mu_{ij})} - 1 \right) e^{-\tau_i(\lambda_{ij}+\mu_{ij})}}{\rho_i(\lambda_{ij}+\mu_{ij})^2} \right)$$

for the probability that a patient experiences a non-fatal event and then drops out before completing the trial,

$$p_{ij4} = \int_0^{\rho_i} \int_r^{\infty} \int_r^{\infty} \frac{\lambda_{ij}\mu_{ij}}{\rho_i} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(c > \tau_i) \mathbb{1}(x > \tau_i) \, dx \, dc \, dr$$

$$= \int_0^{\rho_i} \frac{e^{-\lambda_{ij}\tau_i + r(\lambda_{ij}+\mu_{ij})}}{\rho_i} \int_{\tau_i}^{\infty} \mu_{ij} e^{-c\mu_{ij}} \, dc \, dr$$

$$= \frac{e^{-(\lambda_{ij}+\mu_{ij})\tau_i}}{\rho_i} \int_0^{\rho_i} e^{(\lambda_{ij}+\mu_{ij})r} \, dr$$

$$= \frac{e^{-(\lambda_{ij}+\mu_{ij})\tau_i} \left(e^{\rho_i(\lambda_{ij}+\mu_{ij})} - 1 \right)}{\rho_i(\lambda_{ij}+\mu_{ij})}$$

for the probability that a patient completes the study without an event and

$$
\begin{aligned}
p_{ij5} &= \int_0^{\rho_i} \int_r^\infty \int_r^\infty \frac{\lambda_{ij}\mu_{ij}}{\rho_i} e^{-\mu_{ij}(c-r)} e^{-\lambda_{ij}(x-r)} \mathbb{1}(c \le \tau_i) \mathbb{1}(c < x) \,\mathrm{d}x\,\mathrm{d}c\,\mathrm{d}r \\
&= \frac{\mu_{ij}}{\rho_i} \int_0^{\rho_i} e^{r(\lambda_{ij}+\mu_{ij})} \int_r^{\tau_i} e^{-c(\lambda_{ij}+\mu_{ij})} \,\mathrm{d}c\,\mathrm{d}r \\
&= \frac{\mu_{ij}}{\rho_i(\lambda_{ij}+\mu_{ij})} \int_0^{\rho_i} 1 - e^{-\tau_i(\lambda_{ij}+\mu_{ij})} e^{r(\lambda_{ij}+\mu_{ij})} \,\mathrm{d}r \\
&= \frac{\mu_{ij}}{(\lambda_{ij}+\mu_{ij})} \left(1 - \frac{\left(e^{\rho_i(\lambda_{ij}+\mu_{ij})} - 1\right) e^{-\tau_i(\lambda_{ij}+\mu_{ij})}}{\rho_i(\lambda_{ij}+\mu_{ij})} \right)
\end{aligned}
$$

for the probability that a patient drops out before trial completion.