Published in final edited form as:

Stat Med. 2015 September 20; 34(21): 2913–2925. doi:10.1002/sim.6530.

Meta-analysis of rate ratios with differential follow-up by treatment arm: Inferring comparative effectiveness of medical devices

Lauren M. Kunz^{a,b,*}, Sharon-Lise T. Normand^{b,c}, Danica Marinac-Dabic^d, and Art Sedrakyan^e

^aOffice of Biostatistics Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892, U.S.A

^bDepartment of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115, U.S.A

^cDepartment of Health Care Policy, Harvard Medical School, Boston, Massachusetts 02115, U.S.A

^dDivision of Epidemiology, Center for Devices and Radiological Health, FDA, Silver Spring, Maryland 20993, U.S.A

^eWeill Cornell Medical College of Cornell University and New York Presbyterian Hospital, New York, N.Y. 10065, U.S.A

Abstract

Modeling events requires accounting for differential follow-up duration, especially when combining randomized and observational studies. Although events occur at any point over a follow-up period and censoring occurs throughout, most applied researchers use odds ratios as association measures, assuming follow-up duration is similar across treatment groups. We derive the bias of the rate ratio when incorrectly assuming equal follow-up duration in the single study binary treatment setting. Simulations illustrate bias, efficiency, and coverage and demonstrate that bias and coverage worsen rapidly as the ratio of follow-up duration between arms moves away from one. Combining study rate ratios with hierarchical Poisson regression models, we examine bias and coverage for the overall rate ratio via simulation in three cases: when average arm-specific follow-up duration is available for all studies, some studies, and no study. In the null case, bias and coverage are poor when the study average follow-up is used and improve even if some arm-specific follow-up information is available. As the rate ratio gets further from the null, bias and coverage remain poor. We investigate the effectiveness of cardiac resynchronization therapy devices compared to those with cardioverter-defibrillator capacity where 3 of 8 studies report arm-specific follow-up duration.

*Correspondence to: 6701 Rockledge Drive, Office of Biostatistics Research, National Heart, Lung, and Blood Institute, Bethesda, Maryland 20892, U.S.A. lauren.kunz@nih.gov.

Keywords

Aggregated data; bayesian; comparative effectiveness

1. Introduction

A common statistical tool used to infer the comparative effectiveness of treatments is through a meta-analysis where study-specific estimates obtained from the literature are combined using standard statistical principles. With an increasing medical literature and wider range of statistical modeling techniques possible through computational advances, the types of meta-analyses have also increased [1]. Researchers combine study effect estimates not only from randomized trials, but also from observational studies [2], from a combination of randomized and observational studies via cross-design synthesis [3], and from trial arms corresponding to different studies via network meta-analysis [4]. The expanding reliance on network or cross-design meta-analyses highlights differential follow-up duration, which must be accounted for when events occur at any point over a follow-up period and censoring occurs throughout that period.

The conventional meta-analysis includes I primary studies using summary information, $\{Y_{ij}, n_{ij}\}$, such as a statistic and sample size, for each treatment arm j , about a parameter, θ_j , with the common objective of making inferences about a population parameter, μ . When follow-up duration varies by treatment arm, we require the exposure in each study arm,

$\bar{e}_j = \sum_{k=1}^{n_j} e_{jk} / n_j$, where e_{jk} is the follow-up for person k in arm j . While use of incident rate models or survival models is common within studies, most applied researchers continue to utilize odds ratios as the primary measures of associations, modeling probabilities of events within a particular time frame to combine study summaries. A BioMed Central article [5] reviewing one year of the Cochrane Library, for example, reported that the majority of cancer-related meta-analyses (63%) employed odds ratios or relative risks rather than hazard ratios.

When the follow-up time is fixed, modeling the probability of death is accomplished using the number of events and the total number of people, while assuming follow-up duration is the same or similar for the treatment arms. For meta-analyses of time to event data, the log hazard ratio can be estimated directly if the observed number of events and the log rank expected number of events in each group for each study are reported, or if the log hazard ratio and its variance from the results of a Cox regression [6] are available. Parmar [7] details other methods when the hazard ratio with confidence interval, p-value for the Mantel-Haenszel version of the log rank statistic, or when the published survival curves are given.

Differential follow-up duration in meta-analysis of observational studies is particularly challenging. For example, in work for the Food and Drug Administration's Medical Device Epidemiology Network (MDEpiNet), we require assessing safety and effectiveness of cardiac resynchronization therapy (CRT) devices compared to CRT devices with cardioverter-defibrillator capacity (CRT-D). Both are implanted pacemakers used to

improve mechanical synchrony in patients with heart failure and involve the placement of three leads (right and left atrium, and right ventricle). The CRT-D has an added defibrillation capability to break fast arrhythmias. The devices differ in costs [8] – average patient costs are higher for CRT-D (\$82,200) compared to CRT-alone (\$59,900). While there is a lack of clinical trials designed to assess the incremental benefit of CRT-D compared to CRT alone, the vast majority of patients receiving therapy with biventricular pacing are now implanted with CRT-D devices. Table 1 provides a summary of the 8 studies that compare CRT-D and CRT-alone. This comparison is part of a larger evidence synthesis project with interest in comparing CRT-D, CRT-alone, and optimal medical therapy. The search strategy was based on a previously published comprehensive review of cardiac resynchronization therapy and implantable cardioverter-defibrillators in left ventricular systolic dysfunction sponsored by the Agency for Healthcare Research and Quality.

A data synthesis of the CRT studies presents a number of complications. First, the average length of follow-up is 25.7 months with a standard deviation of 15.3 months across the studies. The constancy of the log hazard across this time frame is questionable and a full-accounting of the follow-up time is required. Second, not all primary studies use a survival time approach. The information available for all-cause mortality include the number of deaths per arm, the total number of patients per arm, and the average length of follow-up across both arms (rather than average per arm). The typical approach for determining person-months of follow-up per arm involves multiplying the reported average months of follow-up time by the total number of people enrolled in each treatment group. The average all-cause mortality rate, across all primary studies, is 8.83 deaths per 1000 person-months: 7.37 deaths per 1000 person-months in the CRT-D arm and 10.63 deaths per 1000 person-months in the CRT-alone arm. For the observational studies the mortality rate is 8.43 per 1000 person-months, whereas for the RCT studies the mortality rate is higher, at 10.03 deaths per 1000 person-months. Table 1 provides additional follow-up information by study where the evidence suggests differential follow-up time by treatment arm. Only 3 studies provide any information regarding arm-specific follow-up time.

Using theoretical calculations in Section 2, we derive the bias of the rate ratio in the setting of a single study with two treatment groups. We use simulations to illustrate bias, efficiency, and coverage when ignoring variable treatment arm follow-up duration. We present a model to combine rate ratios in the meta-analytic setting. In this setting, we utilize simulation to characterize the operating characteristics of the estimators as a function of the duration of differential follow-up and describe how the availability of follow-up by treatment arm impacts these estimators. In Section 3 we perform a data analysis of the CRT-D and CRT studies. We close with recommendations for performing meta analyses when follow-up varies by treatment arm and when this information is unavailable in Section 4.

2. Methods

2.1. A Single Study

Consider a two-arm study with interest centered on a rate ratio for a control ($j = 0$) and a treatment arm ($j = 1$). Assume the number of events from treatment arm j is $Y_j \sim \text{Pois}(\theta_j)$ where the expected number of events is θ_j . The average length of follow-up for the j th

treatment arm is $\bar{e}_j = \sum_{k=1}^{n_j} e_{jk} / n_j$ where k indexes individuals and n_j indexes the total number of individuals in the j th treatment arm. We write $\theta_j = \lambda_j \times \bar{e}_j \times n_j$, with λ_j as the mortality rate defined as $\lambda_j = \xi \exp(\omega \times j)$. The parameter ξ represents the outcome rate in the $j = 0$ arm and ω is the log rate ratio of the outcome in the $j = 1$ arm compared to the $j = 0$ arm. The maximum likelihood estimator (MLE) of ω is

$$\hat{\omega} = \log \left(\frac{\hat{\lambda}_1}{\hat{\lambda}_0} \right) = \log \left(\frac{Y_1 / \bar{e}_1 n_1}{Y_0 / \bar{e}_0 n_0} \right) \quad (1)$$

because $\hat{\lambda}_j = \frac{Y_j}{\bar{e}_j n_j}$. When average follow-up is the same in each treatment arm the MLE

depends only on the number of subjects in each arm, $\hat{\omega} = \log \left(\frac{Y_1 / n_1}{Y_0 / n_0} \right)$.

The most common approach utilized when follow-up information is not reported separately for each treatment arm but rather reported for the overall study, is to assume follow-up duration is the same in each arm $\bar{e}_1 = \bar{e}_0$ so that the estimator becomes:

$$\hat{\omega}^* = \log \left(\frac{\hat{\lambda}_1^*}{\hat{\lambda}_0^*} \right) = \log \left(\frac{Y_1 / n_1}{Y_0 / n_0} \right) \quad (2)$$

with $\hat{\lambda}_j^* = \frac{Y_j}{\bar{e} n_j}$. Both the correct and incorrect estimators for the rate ratio (RR), defined as $\exp(\omega)$, are biased (see Appendix A.1 for derivations) with

$$E \left\{ \exp(\hat{\omega}) - \exp(\omega) \right\} = \exp(\omega) \left[\frac{\xi \bar{e}_0 n_0}{(\xi \bar{e}_0 n_0)^2 + 3(\xi \bar{e}_0 n_0) + 1} \right] \text{ and} \quad (3)$$

$$E \left\{ \exp(\hat{\omega}^*) - \exp(\omega) \right\} = \exp(\omega) \left[(f-1) + \frac{f \xi \bar{e}_0 n_0}{(\xi \bar{e}_0 n_0)^2 + 3(\xi \bar{e}_0 n_0) + 1} \right]; f = \frac{\bar{e}_1}{\bar{e}_0} > 0. \quad (4)$$

When $\bar{e}_1 = \bar{e}_0$, $f = 1$ so that the bias is identical. When $f < 1$, implying longer follow-up in the $j = 0$ arm, the term $(f-1)$ in Equation 4 is negative so that the incorrect estimator underestimates the true rate ratio. When $f > 1$, a similar argument indicates $\hat{R}^*$ overestimates the true rate ratio.

2.1.1. Single Study Simulations—To illustrate the impact of unequal follow-up we conducted a simulation study using 1000 experiments under a variety of conditions. We assumed $\bar{e}_0 = 24$ months for the CRT arm, varied follow-up in the $j = 1$ arm CRT-D arm using $\bar{e}_1 = f \times \bar{e}_0 = 24f$ and permitted f to range from 0.8 to 1.3 by 0.05 step increments. For instance, the values of f for the Stabile, Ermis, and COMPANION studies reported in Table 1 are 0.945, 0.722, and 0.970 respectively. We assume there are an equal number of people in each arm $n_0 = n_1 = 200$. The baseline rate in the control CRT arm is taken to be $\xi = 0.01$ deaths per person-month. Results are presented in Figure 1 under 3 values of the rate ratio: $RR = 1, 0.7$, and 0.5 (large difference).

The impact of using the study average follow-up rather than the arm-specific follow-up can be large. The simulations confirm the theoretical bias results. In general, when follow-up is shorter in the treated arm, $f < 1.0$, the incorrect method of estimating the rate ratio underestimates the true rate ratio, whereas if follow-up is longer in the treated arm, $f > 1.0$, we overestimate the true rate ratio. The MSE for the incorrect rate ratio is greater than the MSE for the correct rate ratio when $f \neq 1$. As f moves away from one, coverage for the incorrect rate ratio drops from the desired 95%. Moreover, while the bias is larger for all $f \neq 1.0$ with the incorrect estimator, the relative efficiency favors the incorrect estimator when f is slightly less than 1 (around 0.9). We also examined experiments in which the sample sizes in the study arms were unequal and the same pattern of increasing relative bias away from $f = 1.0$ as in Figure 1 (results not shown).

2.2. Multiple Studies

Rather than one study, suppose there are I primary studies such that the number of events from arm j , study i , are $Y_{ij} \sim \text{Poisson}(\theta_{ij})$ with

$$\theta_{ij} = \lambda_{ij} \times \bar{e}_{ij} \times n_{ij} \text{ where } \lambda_{ij} = \xi_i \exp(\omega_i \times j). \quad (5)$$

As before, ξ_i is event rate for the $j = 0$ arm in the i th study; ω_i is the log rate ratio of the event for $j = 1$ versus $j = 0$ in the i th study; λ_{ij} is the event rate; n_{ij} is the total number of individuals; $\bar{e}_{ij}$ is the average person-months of follow-up; and θ_{ij} is the expected number of events. The baseline rate and the log relative rate ratio are assumed to vary across studies to accommodate between-study variation using

$$\xi_i \stackrel{\text{indep.}}{\sim} \text{Gamma}(a, b) \text{ and } \omega_i \stackrel{\text{indep.}}{\sim} \text{Normal}(\mu, \sigma^2). \quad (6)$$

The selection of a Gamma distribution for the event rate in the control arm ensures positivity and is commonly used in hierarchical models with Poisson data [17]. The choice of a normal distribution for the log relative rate accommodates both positive and negative values.

A fully Bayesian approach places proper distributions for all the hyperparameters but we focus here on μ and σ^2 . The model defined by equations (5) – (6) assumes that the $\bar{e}_{ij}$ are known for all i and j . When the average person-months of follow-up per arm and the total number of people per arm are available, then the total follow up by arm

$\sum_{k=1}^{n_{ij}} e_{ijk} = n_{ij} \times \bar{e}_{ij}$. However, as before, we assume $\bar{e}_{ij} = \sum_{k=1}^{n_{ij}} e_{ijk} / n_{ij}$ are unavailable and rather $\bar{e}_i = \frac{\bar{e}_{i1}n_{i1} + \bar{e}_{i0}n_{i0}}{n_{i1} + n_{i0}}$ are reported. Primary interest remains focused on $\exp(\mu)$, the overall rate ratio across all studies. Because we are combining estimates across studies, σ the between-study standard deviation, is also a key parameter. In our motivating study, with such few primary studies, the overall results will be sensitive to this parameter.

2.2.1. Multiple Study Simulations—We generated 1000 experiments under 36 different parameter configurations: 3 relative rates $\times$ 2 standard deviation values $\times$ 6 values of f . We assume a moderate number of studies, $I=20$ studies. We fixed μ , the summary log rate ratio for the I studies, and σ , the between study standard deviation of the log rate ratios. Twenty

individual log rate ratios, ω_i , were drawn from $N(\mu, \sigma)$. We also fixed the shape and scale parameters for simulating ξ_i from a Gamma distribution at $a = 2.55$ and $b = 0.00445$ implying a mean baseline rate of $a \times b = 1.13$ per 100 person-months and sampled 20 baseline rates in $j = 0$ arm. As in the single study simulations, we assume equal sample sizes in the two arms within each study, but permitted the sample sizes across studies to vary. This was accomplished by drawing sample sizes from a Uniform(50,1000).

The average follow-up times in the control arms, $\overline{e_{i0}}$, were generated using a mixture of uniforms. We assumed 9 studies had $\overline{e_{i0}} \sim \text{Uniform}(10, 32)$; 9 studies from Uniform(33, 57); and 2 studies from a Uniform(58,200). Under this scenario, between-study follow-up could range from 10 to 200 months. We let $f = \overline{e_{i1}} / \overline{e_{i0}}$ and generate average number of deaths $\theta_{ij} = \lambda_{ij} \times \overline{e_{ij}} \times n_{ij}$ and use this to simulate the number of deaths assuming the data are Poisson.

We estimated the person-time assuming full knowledge of the follow-up in both arms (e.g., the correct exposures) and then again using average exposure across both arms (e.g., the incorrect method). We fit the data using the WinBUGS software [18] and place non-informative priors for $\xi_i \sim \text{Gamma}(2.5, 224.7)$ and $\omega_i \sim N(\mu, \sigma^2)$ where $\mu \sim N(0, 10^6)$ and $\sigma \sim \text{Half-Normal}(0.26)$. The choice of the prior distribution for the between-study standard deviation is much more influential in a hierarchical model than the choice of the prior distribution for the overall mean. The Half-Normal distribution ensures positivity of the standard deviation and because it has a mode at 0, also permits no differences between studies. A Half-Normal(0.26) indicates that 0.26 is the variance and yields a median value for σ of 0.39 with a 95% quantile of 1.0 for the between study standard deviation of the log rate ratio.

We ran one chain with 20000 iterations, 10000 burn in and thin every 10, resulting in 1000 Markov Chain Monte Carlo (MCMC) iterations. Convergence was assessed using the Geweke diagnostic in order to have a resulting 1000 simulations. Inference for the parameters is based on posterior means of the resulting 1000 MCMC iterations; these posterior means are averaged over the 1000 simulations. Credible intervals are found by taking the 2.5% and 97.5% percentiles of the 1000 iterations sorted. Coverage is found by calculating the frequency out of 1000 that the true value of the parameter falls between the 2.5% and 97.5% percentiles.

As f moves away from 1.0 the bias and coverage are worse for the incorrect RR (Table 2). Bias and coverage of σ are not impacted by f , and are similar for the correct and incorrect methods (Table 2). The relative bias (the percent bias of the incorrect estimator divided by the percent bias of the correct estimator) tends to be larger in magnitude moving away from the null value $RR = 1$. The relative bias is larger in magnitude when σ^2 is smaller (results not shown). The coverage probability breaks down more rapidly when σ^2 is smaller because it is more difficult to contain the true mean for a smaller between-study variation.

2.2.2. Partially Reported Follow-up Times—To address the problem of partially reported follow-up duration information by study arm, we examined two situations. In the first, we assume that some studies do not report follow-up by study arm completely at

random (MCAR). In the second, the “missingness” mechanism is at random (MAR) and is related to whether the study is observational or randomized. As in the previous multiple study setting, data is generated under the Poisson model assuming follow-up is known for each arm. If a study is selected to be missing arm-specific follow-up, the follow-up is replaced by the average follow-up for the study over both arms.

Under MCAR, each of the 20 studies has on average a 17.5% chance of being selected to not have arm-specific follow-up. This resulted in 3.5 studies missing arm-specific follow-up over the 1000 simulations (minimum 0 studies and maximum 11). Under MAR, 10 of the studies were labeled as observational and 10 as randomized. Each of the 10 observational studies has on average a 30% chance of being selected to not have arm-specific follow-up and each of the 10 randomized studies has on average a 5% chance of being selected to not have arm-specific follow up. On average, 0.5 randomized studies are missing arm-specific follow-up (minimum 0 studies and maximum 5) and 3 observational studies are missing arm-specific follow-up over the 1000 simulations (minimum 0 studies and maximum 10). The full data for a study i consists of $Y_{i1}, Y_{i0}, n_{i1} = n_{i0}, i_1, i_0$. For MCAR, $Pr(i_j \text{ is missing}) = Pr(i_j | i_j, n_{ij}, Y_{ij}) = 0.175$. For MAR, let X_i be an indicator for whether study i is observational. Then $Pr(i_j \text{ is missing} | X_i) = Pr(i_j | i_j, n_{ij}, Y_{ij}, X_i)$ and $Pr(i_j \text{ is missing} | X_i = 1) = 0.3$ and $Pr(i_j \text{ is missing} | X_i = 0) = 0.05$, so the probability of missing arm-specific follow-up depends on whether the study was observational or randomized, but within study type, the probability of missing arm-specific follow-up does not depend on the length of arm-specific follow-up.

For the randomized trials, we always assume that there is more extreme follow-up in the treatment arm, such that f is further from the null versus a comparable observational study. If $f < 1$, $f_{RCT} = \frac{3}{4}f_{OBS}$ and if $f > 1$, $f_{RCT} = \frac{4}{3}f_{OBS}$, hence the observational study will always be assumed to have follow-up that is more unbalanced between the two arms of an observational study versus a randomized trial. After the data is generated to reflect this, the models are fit using the same fully Bayesian Poisson model in WinBUGS as previously described. When the simulated experiment reported follow-up by study arm, that information was utilized and when arm-specific follow-up was unavailable, the average study follow-up was used.

In general, use of partially observed follow-up times has a bias for the rate ratio that is between the bias for the correct case of having complete arm-specific follow-up and the incorrect case of having incomplete arm-specific follow-up for all primary studies (Table 4). Coverage of the rate ratio is similar for both cases of missingness. As f moves away from 1.0, the bias and coverage worsen for the RR under both MAR and MCAR, but are more pronounced when σ^2 is larger (Figure 2). Bias and coverage of σ do not follow any trend related to f and again are similar for both types of missingness (Table 4).

3. Data Analysis: Effectiveness of CRT-D vs CRT

We analyzed the mortality data reported in Table 1 using the model described in Equations (5) - (6). Because of the small number of studies for the analysis, we considered a total of nine different sets of prior distributions for μ (the overall log rate ratio) and σ (the between-

study standard deviation) that ranged in terms of informativeness. Models were estimated using the WinBUGS software and ran until convergence as determined by the Geweke score for the between-study standard deviation component. The all-cause mortality rate averaged across all primary studies is 8.83 deaths per 1000 person-months (for the CRT-D arm the mortality rate is 7.37 deaths per 1000 person-months and for the CRT alone arm the mortality rate is 10.63 deaths per 1000 person-months). Seven of the 8 studies show a benefit of CRT-D over CRT alone with a rate ratio less than 1.

Only 3 of the 8 studies report arm-specific follow up, with $f=0.72$ [13], 0.95 [10], and 0.97 [15]. We estimate the overall rate ratio using two main approaches: in the first, we use arm-specific follow-up when available and the study average when it is not (as in Section 2.2.2). In the second, we ignore any arm-specific follow-up information and use the study average follow-up. Additionally, we do some sensitivity analysis to explore how much the estimates change under various follow-up ratios between the two arms.

3.1. Prior Distributions

We assumed the overall log rate ratio arose from a normal distribution centered at the null value of 0 (a rate ratio of 1.0). We selected three different variances for μ (overall log rate ratio): (1) the variance is 2 yielding a 95% interval from -2.77 to 2.77 (0.063 to 15.96 on the rate ratio scale) which is vague; (2) variance is 10 indicating the 95% interval for log rate ratio could range -6.2 to 6.2 which is quite vague; and (3) a variance of 1000000 which is extremely vague.

Three different prior distributions for the between-study standard deviation were selected. Two half-normal distributions permitted the underlying log rate ratio for a study to (1) have a median value of 0.39 with 95% quantile of 1.0 (Half-Normal(0.26)) and (2) have a median value of 0.14 with 95% quantile of 0.36 (Half-Normal(0.03)). A uniform distribution (Uniform(0,0.7)) had a mean and median of 0.35.

3.2. Results

When using the arm-specific follow-up information (for 3 of the 8 studies that provided it), the posterior mean (95% credible interval) of the overall rate ratio was 0.71 (0.49, 0.96) and 0.71 (0.55, 0.89) under the most non-informative pair of priors, $\mu \sim \text{Normal}(0, 1e06)$ and $\sigma \sim \text{Half-Normal}(0.26)$, and most informative pair of priors, $\mu \sim \text{Normal}(0, 2)$ and $\sigma \sim \text{Half-Normal}(0.03)$, respectively. These results indicate a survival benefit of CRT-D compared to CRT-alone, such that there is approximately a 30% lower rate of death in the CRT-D arm (Figure 3(a)). The posterior mean of the between-study standard deviation was estimated as 0.34 (0.08, 0.75) (Figure 3(b)). All other priors resulted in a similar overall rate ratio. The 95% credible interval did not cover 1 for any of the priors (Table 3). Using the Half-Normal(0.03) prior for σ resulted in shorter credible intervals (Figure 3).

Results are similar, but the overall rate ratio ($\exp(\mu)$) is further from the null when ignoring the arm-specific follow-up information reported in the three studies with a posterior mean of 0.69 (see Appendix for more results). The findings from the simulation studies when $f < 1$ (as in the CRT-D vs CRT meta-analysis) suggest that the estimate ignoring arm-specific

information will be further from the null than the estimate using the arm-specific follow-up. It is not surprising that our two posterior means (0.69 and 0.71) do not differ, given only 3 out of 8 studies reported arm-specific follow-up.

In addition, we performed some sensitivity analyses to look at the impact of how much the estimates change under various follow-up ratios between the two arms. The first scenario considered used follow-up by arm for the 3 of 8 studies that reported it separately and for those that did not, we assumed $f \approx 0.9$. The results for the overall rate ratio were 0.76 (0.56, 1.00). Our original analysis with $f = 1$ using arm-specific when available for 3 out of 8 studies and average otherwise was 0.71 (0.49, 0.96). We should be overestimating the true RR when all studies now have $f < 1$. Using a more extreme $f = 0.722$ which was the most extreme we observed in the 8 studies, the RR was 0.86 (0.64, 1.14), so even closer to the null.

4. Remarks

We examined the impact of missing duration of follow-up between two treatment arms in meta-analysis on inference about an overall mean. Although events occur at any point over a follow-up period and censoring occurs throughout that period, most applied researchers continue to use odds ratios and assume similar follow-up across treatment groups. Equal follow-up among treatment groups is unlikely to hold in observational studies. In the single study setting, when longer follow-up occurs in the treatment arm, e.g., $f > 1$, the incorrect estimator underestimates the true rate ratio and overestimates the true rate ratio when $f < 1$. Mean squared error is larger relative to the correct estimator when $f \neq 1$ and coverage is poor. Inferences are impacted with a fairly modest difference in follow-up duration, e.g., when $f = 0.8$ and the null is true the coverage of 95% intervals dropped to 0.71 and the bias neared 20%, implying an estimated rate ratio of 0.8.

We utilized hierarchical Poisson regression models to combine rate ratios across studies and examined operating characteristics of posterior means under a variety of conditions using simulation studies. While it is impossible to determine the direction of the bias, both bias and coverage when there is no true effect are worse when using average follow-up – including the available arm-specific information reduces bias compared to including none. However, there is no way to correct the bias unless more information regarding arm-specific follow-up duration is reported. The analyses could place a symmetric prior probability distribution around 1 for f , even if this uncertainty is vague. For example, we propose a prior distribution for the natural logarithm of f , $\log(f) \sim N(0, \sigma_f)$. This approach may be reasonable when relatively many studies are available, but is questionable in our example that includes only 8 studies given the amount of *observed* information relative to the number of parameters.

Bushman and Wang [19] proposed methods to combine effects when some studies do not report effect estimates using a mixture of the reported effect estimates and vote-counting procedures. Vote-counting procedures require effects to be homogeneous across studies, an assumption not likely to be met in practice. When analyzing the CRT vs. CRT-D studies, there was substantial variability in the duration of follow-up within and between studies, and

the majority of the studies did not report duration information. If differences in follow-up between arms exist, then our estimates may over or underestimate the true rate ratio and are unable to tell in which way.

In this article, we assumed a constant hazard ratio which is a strong assumption at best. Other researchers [20] have cautioned against the use of a single number to summarize the hazard ratio due to its dependence on the duration of follow-up and the inherent bias associated with period-specific hazard ratios (e.g., subjects must survive to each period-specific interval to be included in the calculation for the period). Given these concerns, we expect similar biases to arise in the estimates of period-specific hazards if the arm-specific follow-up durations were not reported in the primary studies. A prudent approach to avoid the issue of bias altogether would ideally have the primary studies summarize adjusted survival curves using discrete-time hazards that include follow-up time as an explanatory variable interacted with treatment arm. Moving forward, it is important that publications contain information on arm-specific follow-up duration as applied researchers increasingly combine information from multiple studies to learn about treatment and safety effectiveness. Network meta-analysis may be even more prone to issues of differential follow-up duration because in such analyses the number of treatments and the number of types of studies being compared are large.

Acknowledgments

The authors thank Jennifer Moon, PhD for performing the data abstraction. The authors thank Francesca Dominici, PhD and Miguel Hernan, MD, DrPH for input and suggestions that improved the paper. This material is based upon work supported by the Research Participation Program, Center for Devices and Radiological Health, administered by the Oak Ridge Institute for Science and Education through an inter-agency agreement between the U.S. Department of Energy and the U.S. Food and Drug Administration (FDA), as well as Contract No. HHSF223201110172C and the Critical Path Initiative (Development of Innovative Methodologies for Medical Devices), both from the FDA.

References

1. Sutton AJ, Higgins JPT. Recent developments in meta-analysis. *Statistics in Medicine*. 2008; 27(5): 625–650. [PubMed: 17590884]
2. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, Moher D, Becker BJ, Sipe TA, Thacker SB. Meta-analysis of observational studies in epidemiology: A proposal for reporting. *Journal of the American Medical Association*. 2000; 283(15):2008–2012. [PubMed: 10789670]
3. Drotocour J, Silberman G, Chelimsky E. Cross-design synthesis: A new form of meta-analysis for combining results from randomized clinical trials and medical-practice databases. *International Journal of Technology Assessment in Health Care*. 1993; 9:440–449. [PubMed: 8340208]
4. Lumley T. Network meta-analysis for indirect treatment comparisons. *Statistics in Medicine*. 2002; 21:2313–2324. [PubMed: 12210616]
5. Tierney JF, Stewart LA, Ghersi D, Burdett S, Sydes MR. Practical methods for incorporating summary time-to-event data into meta-analysis'. *Trials*. 2007; 8(16)
6. Woods BS, Hawkins N, Scott DA. 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*. 2010; 10(54)
7. Parmar MKB, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Statistics in Medicine*. 1998; 17:2815–2834. [PubMed: 9921604]

8. Feldman AM, de Lissovoy G, Bristow MR, Saxon LA, Marco TD, Kass DA, Boehmer J, Singh S, Whellan DJ, Carson P, Boscoe A, Baker TM, Gunderman MR. Cost effectiveness of cardiac resynchronization therapy in the comparison of medical therapy, pacing, and defibrillation in heart failure (companion) trial. *Journal of the American College of Cardiology*. 2005; 46(12):2311–2321. [PubMed: 16360064]
9. Adlbrecht C, Hulsman M, Gwechenberger M, Strunk G, Khazen C, Wiesbauer F, Elhenicky M, Neuhold S, Binder T, Maurer G, Lang IM, Pacher R. Outcome after device implantation in chronic heart failure is dependent on concomitant medical treatment. *European Journal of Clinical Investigation*. 2009; 39(12):1073–1081. [PubMed: 19843157]
10. Stabile G, Solimene F, Bertaglia E, Rocca VL, Accogli M, Scaccia A, Marrazzo N, Zoppo F, Turco P, Iuliano A, Shopova G, Ciardiello C, Simone AD. Long-term outcomes of crt-pm versus crt-d recipients. *Pacing and Clinical Electrophysiology*. 2009; 32(S1):S141–S145. [PubMed: 19250079]
11. Bai R, Biase LD, Elayi C, Ching CK, Barrett C, Philipps K, Lim P, Patel D, Callahan T, Martin DO, Arruda M, Schweikert RA, Saliba WI, Wilkoff B, Natale A. Mortality of heart failure patients after cardiac resynchronization therapy: Identification of predictors. *Journal of Cardiovascular Electrophysiology*. 2008; 19(12):1259–1265. [PubMed: 18631272]
12. Auricchio A, Metra M, Gasparini M, Lamp B, Klersy C, Curnis A, Fantoni C, Gronda E, Vogt J. Long-term survival of patients with heart failure and ventricular conduction delay treated with cardiac resynchronization therapy. *The American Journal of Cardiology*. 2007; 99:232–238. [PubMed: 17223424]
13. Erms C, Lurie KG, Zhu AX, Collins J, Vanheel L, Sakaguchi S, Lu F, Pham S, Benditt DG. Biventricular implantable cardioverter defibrillators improve survival compared with biventricular pacing alone in patients with severe left ventricular dysfunction. *Journal of Cardiovascular Electrophysiology*. 2004; 15(8):862–866. [PubMed: 15333075]
14. Pappone C, Vicedomini G, Augello G, Mazzone P, Nardi S, Rosanio S. Combining electrical therapies for advanced heart failure: The milan experience with biventricular pacing–defibrillation backup combination for primary prevention of sudden cardiac death. *The American Journal of Cardiology*. 2003; 91(9A):74F–80F.
15. Bristow MR, Saxon LA, Boehmer J, Krueger S, Kass DA, Marco TD, Carson P, DiCarlo L, DeMets D, White BG, DeVries DW, Feldman AM. Cardiac-resynchronization therapy with or without an implantable defibrillator in advanced chronic heart failure. *The New England Journal of Medicine*. 2004; 350(21):2140–2150. [PubMed: 15152059]
16. Schuchert A, Muto C, Maounis T, Frank R, Boulogne E, Polauck A, Padeletti L. Lead complications, device infections, and clinical outcomes in the first year after implantation of cardiac resynchronization therapy-defibrillator and cardiac resynchronization therapy-pacemaker. *Europace*. 2013; 15:71–76. [PubMed: 22927665]
17. Carlin, BP.; Louis, TA. Bayes and empirical Bayes methods for data analysis. 2. Chapman and Hall; London: 2001.
18. Lunn DJ, Thomas A, Best N, Spiegelhalter D. Winbugs - a bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*. 2000; 10:325–337.
19. Bushman BJ, Wang MC. A procedure for combining sample standardized mean differences and vote counts to estimate the population standardized mean difference in fixed effects models. *Psychological Methods*. 1996; 1(1):66–80.
20. Hernan MA. The hazards of hazard ratios. *Epidemiology*. 2010; 21(1):13–15. [PubMed: 20010207]

A. Appendix

A.1. Bias of the Single Study Estimator for the Rate Ratio

Assume the number of events from treatment arm j is $s Y_j \sim \text{Pois}(\theta_j)$ where $\theta_j = \lambda_j \times t_j \times n_j$ with t_j the average follow-up in months in arm j and n_j is the number of subjects in arm j . Then the mortality rate is written as

$$\lambda_j = \xi \exp(\omega \times j) \quad (7)$$

where $j = 0$ for the control arm and $j = 1$ for the treatment arm. The maximum likelihood estimator for ω is $\log(\hat{\lambda}_1 / \hat{\lambda}_0) = \log\left(\frac{Y_1 / \bar{e}_1 n_1}{Y_0 / \bar{e}_0 n_0}\right)$ because $\hat{\lambda}_j = \frac{Y_j}{\bar{e}_j n_j}$. When average follow-up is the same in each treatment arm, then $\omega = \log\left(\frac{Y_1 / n_1}{Y_0 / n_0}\right)$. The rate ratio is $\exp(\omega) = \lambda_1 / \lambda_0$.

We are dealing with the expectation and variance of ratios of random variables and by Taylor series expansions, it can be shown that

$$E\left(\frac{Y_1}{Y_0}\right) \approx \frac{EY_1}{EY_0} - \frac{\text{Cov}(Y_1, Y_0)}{E^2 Y_0} + \frac{\text{Var}(Y_0) EY_1}{E^3 Y_0} \quad (8)$$

$$\begin{aligned} \text{Var}\left(\frac{Y_1}{Y_0}\right) &\approx \frac{1}{E^2 Y_0} \text{Var}(Y_1) + 2 \frac{-EY_1}{E^3 Y_0} \text{Cov}(Y_1, Y_0) + \frac{E^2 Y_1}{E^4 Y_0} \text{Var} Y_0 \\ &= \frac{E^2 Y_1}{E^2 Y_0} \left[\frac{\text{Var}(Y_1)}{E^2 Y_1} - 2 \frac{\text{Cov}(Y_1, Y_0)}{EY_0 EY_1} + \frac{\text{Var}(Y_0)}{E^2 Y_0} \right] \end{aligned} \quad (9)$$

when expanding out to 3 terms.

Noting that the moments of the Poisson distribution are $E(Y_j) = \theta_j$ and $E(Y_j^2) = \theta_j + \theta_j^2$, then $\text{Var}(Y_j) = \theta_j$ and $E(Y_j^3) = \theta_j^3 + 3\theta_j^2 + \theta_j$. Because Y_1 and Y_0 are independent, $\text{Cov}(Y_1, Y_0) = 0$. When average follow-up by treatment arm, $\bar{e}_j$, is available, then

$$E(\hat{\lambda}_j) = E\left(\frac{Y_j}{\bar{e}_j n_j}\right) = \frac{1}{\bar{e}_j n_j} E(Y_j) = \frac{\theta_j}{\bar{e}_j n_j}.$$

Under the model,

$$E(\hat{\lambda}_1) = \frac{\xi \exp(\omega) \bar{e}_1 n_1}{\bar{e}_1 n_1} = \xi \exp(\omega) \text{ and } E(\hat{\lambda}_0) = \frac{\xi \bar{e}_0 n_0}{\bar{e}_0 n_0} = \xi$$

Hence,

$$\begin{aligned} E(\exp(\hat{\omega})) &= E\left(\frac{\hat{\lambda}_1}{\hat{\lambda}_0}\right) = \frac{\bar{e}_0 n_0}{\bar{e}_1 n_1} E\left(\frac{Y_1}{Y_0}\right) \\ &\approx \frac{\bar{e}_0 n_0}{\bar{e}_1 n_1} \left(\frac{\theta_1}{\theta_0} - 0 + \frac{\theta_0 * \theta_1}{\theta_0^3 + 3\theta_0^2 + \theta_0} \right) \\ &= \frac{\bar{e}_0 n_0}{\bar{e}_1 n_1} \left(\frac{\xi \exp(\omega) \bar{e}_1 n_1}{\xi \bar{e}_0 n_0} - 0 + \frac{\xi \bar{e}_0 n_0 * \xi \exp(\omega) \bar{e}_1 n_1}{(\xi \bar{e}_0 n_0)^3 + 3(\xi \bar{e}_0 n_0)^2 + (\xi \bar{e}_0 n_0)} \right) \quad (10) \\ &= \exp(\omega) + \bar{e}_0 n_0 \left(\frac{\xi \exp(\omega)}{(\xi \bar{e}_0 n_0)^2 + 3(\xi \bar{e}_0 n_0) + 1} \right) \\ &= \exp(\omega) \left[1 + \frac{\xi \bar{e}_0 n_0}{(\xi \bar{e}_0 n_0)^2 + 3(\xi \bar{e}_0 n_0) + 1} \right]. \end{aligned}$$

Therefore the bias, $E(\exp(\hat{\omega})) - \exp(\omega)$ is $\frac{\xi \bar{e}_0 n_0 \exp(\omega)}{(\xi \bar{e}_0 n_0)^2 + 3(\xi \bar{e}_0 n_0) + 1}$.

In the absence of arm-specific follow-up, and only average follow-up for the study,

$$E(\hat{\lambda}_j^*) = E\left(\frac{Y_j}{\bar{e} n_j}\right) = \frac{1}{\bar{e} n_j} E(Y_j) = \frac{\theta_j}{\bar{e} n_j} \text{ and}$$

$$E(\hat{\lambda}_1^*) = \frac{\xi \exp(\omega) \bar{e}_1 n_1}{\bar{e} n_1} = \frac{\xi \exp(\omega) \bar{e}_1}{\bar{e}} \text{ and } E(\hat{\lambda}_0^*) = \frac{\xi \bar{e}_0 n_0}{\bar{e} n_0} = \frac{\xi \bar{e}_0}{\bar{e}}.$$

Hence,

$$\begin{aligned} E(\exp(\hat{\omega}^*)) &= E\left(\frac{\hat{\lambda}_1^*}{\hat{\lambda}_0^*}\right) = \frac{n_0}{n_1} E\left(\frac{Y_1}{Y_0}\right) \\ &\approx \frac{n_0}{n_1} \left(\frac{\theta_1}{\theta_0} - 0 + \frac{\theta_0 * \theta_1}{\theta_0^3 + 3\theta_0^2 + \theta_0} \right) \\ &= \frac{n_0}{n_1} \left(\frac{\xi \exp(\omega) \bar{e}_1 n_1}{\xi \bar{e}_0 n_0} - 0 + \frac{\xi \bar{e}_0 n_0 * \xi \exp(\omega) \bar{e}_1 n_1}{(\xi \bar{e}_0 n_0)^3 + 3(\xi \bar{e}_0 n_0)^2 + (\xi \bar{e}_0 n_0)} \right) \quad (11) \\ &= \exp(\omega) \frac{\bar{e}_1}{\bar{e}_0} + \frac{n_0 * \xi \exp(\omega) \bar{e}_1}{(\xi \bar{e}_0 n_0)^2 + 3(\xi \bar{e}_0 n_0) + 1} \\ &= \exp(\omega) \left[\frac{\bar{e}_1}{\bar{e}_0} + \frac{n_0 * \xi \bar{e}_1}{(\xi \bar{e}_0 n_0)^2 + 3(\xi \bar{e}_0 n_0) + 1} \right]. \end{aligned}$$

Therefore the bias, $E(\exp(\hat{\omega}^*)) - \exp(\omega)$ is $\exp(\omega) \left[\left(\frac{\bar{e}_1}{\bar{e}_0} - 1 \right) + \frac{n_0 * \xi \bar{e}_1}{(\xi \bar{e}_0 n_0)^2 + 3(\xi \bar{e}_0 n_0) + 1} \right]$.

A.2. Simulation Results: Partially Observed Follow-up Times

A.3. CRT Data Analysis: Ignoring Arm-Specific Follow-up for the 3 Studies Reporting Follow-Up

When ignoring the arm-specific follow-up information reported in the three studies, the posterior mean (95% credible interval) of the overall rate ratio was 0.69 (0.48, 0.93) using the most non-informative pair of priors, $\mu \sim \text{Normal}(0, 1e06)$ and $\sigma \sim \text{Half-Normal}(0.26)$. This result also indicates a survival benefit of CRT-D compared to CRT-alone, but the benefit is slightly further from the null than when including the information (compare to 0.71 [0.49, 0.96] for using arm-specific follow-up). A similar comparison based on the most informative pair of priors, $\mu \sim \text{Normal}(0, 2)$ and $\sigma \sim \text{Half-Normal}(0.03)$, the rate ratio is 0.69 [0.54, 0.86]. All other priors resulted in a similar overall rate ratio with 95% credible intervals that do not contain 1 (Table 5). The posterior mean of the between-study standard deviation was estimated as 0.34 [0.03, 0.74], similar to that obtained using arm-specific follow-up (compare to 0.34 [0.08, 0.75]).

Figure 1. Simulation results for single study as function of relative follow-up in treatment arms

Each experimental condition is based on 1000 simulated datasets; $f = \frac{c_1}{c_0}$.

Percent Bias = $\frac{\hat{\theta} - \theta}{\theta} \times 100$; RB = Relative Bias = $\text{Bias}(\text{RR}^*) / \text{Bias}(\text{RR})$; MSE = Mean Squared Error = $1/1000 \times \sum (\hat{\theta} - \theta)^2$; and RE = Relative Efficiency = $\text{MSE}(\text{RR}^*) / \text{MSE}(\text{RR})$.

Figure 2. Percent Bias for the overall rate ratio via simulation in four cases for various RR and σ^2
 arm-specific follow-up is available for all studies (correct), some studies
 (with "missingness" at random (MAR) and completely at random (MCAR)), and no study
 (average).

(a) **Risk of dying in CRT-D vs. CRT, $\exp \mu$.** Thick vertical line indicates null rate ratio of 1.0.

(b) **Between study standard deviation, σ .**

Figure 3. Posterior densities for parameters in the CRT meta-analysis of 8 primary studies
Solid (dashed) lines represent least (most) informative prior distributions for the hyperparameters. Vertical lines represent the 95% credible intervals. Based on 1000 draws from the joint posterior distribution.

Table 1

CRT-D versus CRT-alone primary studies

Baseline characteristics, detailed follow-up information, and all-cause mortality outcome. IHD = ischemic heart disease; NYHA = New York Heart Association; LVEF = left ventricular ejection fraction; QRS represents the time it takes for depolarization of the ventricles. NA indicate that the data was not reported. The ratio of follow-up by treatment arm is denoted $f = \bar{e}_1 / \bar{e}_0$.

Study	Year	# Patients	Mean Age (years)	% Male	% IHD	% NYHA Class III	Mean (SD) % LVEF	Mean (SD) QRS (milliseconds)
Adlbrecht et al. [9]	2009	205	65 (11)	78	46	83	27.5	158 (31)
Stabile et al. [10]	2009	233	69 (8)	77	49	69	26.5	120
Bai et al. [11]	2008	542	67 (11)	77	67	81	20	162 (24)
Auricchio et al. [12]	1298	2007	64 (9)	76	43	80	24	168 (29)
Ermis et al. [13]	2004	126	69 (11.5)	96	56	87	22	NA
Pappone et al. [14]	2003	135	64 (11)	76	43	100	28	153 (11)
TOTAL/MEAN Observational		2539	66.3	80	50.7	83.3	24.7	160.3
Bristow et al. [15]	2004	1212	67 (NA)	67	55	87	21	160 (NA)
Schuchert et al. [16]	2013	402	68 (9)	80	50	85	25	163 (NA)
TOTAL/MEAN Randomized		1614	67.5	73.5	52.5	86	23	161.5
f = $\frac{\bar{e}_1}{\bar{e}_0}$								
			Follow-up (months)		All-Cause Mortality/N			
Study	Overall	CRT-D	CRT			CRT-D	CRT	
Adlbrecht et al. [9]	16.8 (12.4)	NA	NA			19/110	9/95	
Stabile et al. [10]	58 (15)	56.8	60.1			49/116	53/117	
Bai et al. [11]	26.7 (17.6)	NA	NA			73/395	57/147	
Auricchio et al. [12]	34 (NA)	NA	NA			91/726	119/572	
Ermis et al. [13]	13.5 (12)	13	18			8/62	26/64	
Pappone et al. [14]	27.6 (8.4)	NA	NA			6/88	9/47	
TOTAL/MEAN Observational	NA	NA	NA			246/1497	273/1042	
Bristow et al. [15]	NA	16	16.5			105/595	131/617	
Schuchert et al. [16]	12	NA	NA			20/228	19/174	

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Study	Year	# Patients	Mean Age (years)	% Male	% IHD	% NYHA Class III	Mean (SD) % LVEF	Mean (SD) QRS (milliseconds)
TOTAL/MEAN Randomized		NA	NA				125/823	150/791

Table 2

Percent bias and coverage of the rate ratio, $\exp(\mu)$, and between-study standard deviation, σ , using partially reported follow-up timesSimulation results for 20 primary studies as a function of relative follow-up in treatment arms. Percent bias [(estimated - true)/true $\times$ 100].

f	RR=1				RR=0.7				RR=0.5			
	$\sigma^2 = 0.01$		$\sigma^2 = 0.05$		$\sigma^2 = 0.01$		$\sigma^2 = 0.05$		$\sigma^2 = 0.01$		$\sigma^2 = 0.05$	
	Corr	Incorr	Corr	Incorr	Corr	Incorr	Corr	Incorr	Corr	Incorr	Corr	Incorr
RR: Bias												
0.9	2.00	-8.07	5.32	-4.67	-0.86	-10.49	3.33	-6.84	3.34	-6.58	5.38	-4.82
0.95	3.22	-1.68	0.62	-4.51	1.16	-4.17	3.40	-1.54	-4.20	-9.34	-1.86	-6.70
1.0	2.51	2.52	-0.87	-0.84	2.54	2.54	-4.67	-4.67	0.78	0.78	-0.62	-0.60
1.05	-0.66	4.57	1.87	7.15	2.71	8.41	2.69	8.11	-1.54	3.50	4.16	9.42
1.1	-0.13	9.87	-2.79	6.58	2.71	13.00	-2.91	6.46	-0.58	9.02	-0.14	9.34
1.2	-1.96	17.16	4.49	24.81	-1.01	18.79	-1.10	18.37	1.42	21.58	5.16	25.72
RR: Coverage												
0.9	0.986	0.322	0.987	0.995	0.998	0.044	0.998	0.970	0.956	0.774	0.986	0.996
0.95	0.951	0.990	1.000	0.978	0.990	0.940	0.994	0.991	0.955	0.447	0.992	0.819
1.0	0.963	0.963	1.000	1.000	0.983	0.984	0.998	0.990	0.995	0.994	0.998	0.999
1.05	0.991	0.844	1.000	0.750	0.993	0.651	0.999	0.902	0.986	0.968	0.998	0.897
1.1	0.997	0.346	0.989	0.931	0.973	0.201	1.000	0.992	1.000	0.636	1.000	0.879
1.2	0.984	0.000	0.991	0.001	0.991	0.002	1.000	0.115	0.982	0.000	0.989	0.001
σ : Bias												
0.9	9.90	15.60	12.75	12.75	10.30	10.05	18.92	17.89	47.60	46.30	5.50	6.62
0.95	9.80	10.90	-8.45	-9.66	39.40	42.20	-11.76	-10.73	61.30	63.50	-24.51	-25.00
1.0	-4.80	-4.60	-7.02	-6.98	24.90	25.10	10.64	10.60	48.10	48.10	-9.97	-9.93
1.05	-1.10	-2.90	-34.75	-34.39	53.30	57.40	-3.40	-3.53	50.90	50.90	11.81	12.16
1.1	20.10	17.30	-13.15	-14.22	52.10	52.50	28.17	27.82	52.80	55.20	11.23	9.62

RR=0.5													
RR=0.7							RR=1						
$\sigma^2 = 0.01$		$\sigma^2 = 0.05$		$\sigma^2 = 0.01$		$\sigma^2 = 0.05$		$\sigma^2 = 0.01$		$\sigma^2 = 0.05$		$\sigma^2 = 0.01$	
Corr	Incorr	Corr	Incorr	Corr	Incorr	Corr	Incorr	Corr	Incorr	Corr	Incorr	Corr	Incorr
1.2	16.50	14.10	-8.36	-7.11	32.20	30.90	7.83	8.32	24.70	23.80	-0.98	-0.89	-0.89
σ Coverage													
0.9	0.991	0.985	0.994	0.992	0.988	0.989	0.983	0.986	0.869	0.885	0.998	0.998	0.998
0.95	0.985	0.984	0.996	0.994	0.892	0.872	0.987	0.991	0.684	0.656	0.890	0.885	0.885
1.0	0.983	0.980	0.995	0.996	0.937	0.937	0.989	0.990	0.849	0.842	0.981	0.975	0.975
1.05	0.988	0.991	0.650	0.667	0.774	0.731	1.000	1.000	0.842	0.847	0.994	0.991	0.991
1.1	0.969	0.978	0.962	0.958	0.754	0.757	0.942	0.947	0.807	0.783	0.998	0.998	0.998
1.2	0.980	0.979	0.999	1.000	0.956	0.960	0.999	0.998	0.956	0.961	0.999	1.000	1.000

Table 3

CRT-D vs CRT-alone: posterior mean for the overall rate ratio and 95% credible intervals for 8 primary studies under a variety of prior distributions utilizing arm-specific follow-up when available.

Prior for between-study standard deviation, σ	Prior for overall underlying log rate ratio, μ		
	Normal(0, 2)	Normal(0, 10)	Normal(0, 1e06)
Half-Normal(0.03) ^a	0.71 (0.55, 0.89)	0.71 (0.56, 0.90)	0.71 (0.55, 0.89)
Uniform(0, 0.7) ^b	0.71 (0.49, 0.99)	0.71 (0.51, 0.96)	0.72 (0.52, 0.98)
Half-Normal(0.26) ^c	0.70 (0.50, 0.94)	0.71 (0.51, 0.94)	0.71 (0.49, 0.96)

^aE(σ) = 0.14;

^bE(σ) = 0.35;

^cE(σ) = 0.41.

Table 4
**Bias and coverage of the rate ratio, $\exp(\mu)$, and between-study standard deviation, σ ,
 using partially reported follow-up times**

Simulation results for 20 primary studies as a function of relative follow-up in treatment arms. Percent bias [(estimated - true)/true $\times$ 100].

f	RR=1 and $\sigma^2 = 0.05$		RR=0.5 and $\sigma^2 = 0.05$	
	MCAR	MAR	MCAR	MAR
RR: Bias				
0.9	-2.18	-2.89	-1.33	-1.85
1.0	-0.70	-0.71	-0.54	-0.54
1.1	0.23	1.51	2.00	2.50
1.2	1.19	2.39	2.10	2.53
RR: Coverage				
0.9	0.997	0.990	0.994	0.983
1.0	1.000	1.000	0.999	0.999
1.1	0.999	1.000	0.974	0.968
1.2	0.994	0.995	0.992	0.996
σ: Bias				
0.9	3.46	4.11	4.30	4.83
1.0	-7.28	-7.32	-10.08	-10.04
1.1	3.56	4.35	-2.23	-1.12
1.2	1.86	3.72	29.67	38.01
σ: Coverage				
0.9	0.981	0.972	0.951	0.932
1.0	0.999	0.999	0.988	0.983
1.1	0.993	0.942	0.970	0.982
1.2	0.958	0.897	0.806	0.691

Table 5

CRT-D vs CRT-alone: posterior mean for the overall rate ratio and 95% credible intervals for 8 primary studies under a variety of prior distributions ignoring arm specific follow-up.

Prior for between-study standard deviation, σ	Prior for overall underlying log rate ratio, μ		
	Normal(0, 2)	Normal(0, 10)	Normal(0, 1e06)
Half-Normal(0.03) ^a	0.69 (0.54, 0.86)	0.69 (0.54, 0.86)	0.69 (0.53, 0.86)
Uniform(0, 0.7) ^b	0.69 (0.49, 0.92)	0.69 (0.49, 0.92)	0.68 (0.48, 0.93)
Half-Normal(0.26) ^c	0.69 (0.49, 0.94)	0.69 (0.48, 0.96)	0.69 (0.48, 0.93)

^aE(σ) = 0.14;

^bE(σ) = 0.35;

^cE(σ) = 0.41.