Evaluating Treatment Effect in Multicenter Trials with Small Centers Using Survival Modeling

Usha S. Govindarajulu[1,*] and Elizabeth J. Malloy[2]

[1]Department of Epidemiology and Biostatistics, SUNY Downstate School of Public Health, Brooklyn, NY 11203, USA

[2]Department of Mathematics and Statistics, American University, Washington, DC 20016, USA

Abstract: Clinical trials of rare diseases commonly enlist several centers to achieve recruitment goals. The aim of this study is to examine the estimation of treatment effects for survival outcomes in multicenter clinical trials with varying numbers of centers and few patients per center for rarer disease outcomes (i.e. rare cancers). We modeled the heterogeneity between centers using Cox frailty models to account for the variability in patients and patient care between centers and examined measures of model fit *via* smoothed functions of a prognostic factor. Through a simulation study, we were able to examine the consequence of having only a few centers or a few patients per center on the treatment and prognostic factor effects and model performance indices. Overall, we found it is preferable to have more patients per site and more sites in a multicenter trial as expected. However, having a few patients per site is feasible if there are many sites in a trial.

Keywords: Frailty, survival, clinical trial, prognostic factor, rare disease.

1. INTRODUCTION

For large-scale multi-center clinical trials, hospital-based universities are often chosen as study sites and may only enroll a few patients. Due to the rarity of the disease some studies may need to find patients at as many potential centers as possible. As Matthews [1] noted, defaulting to multicenter often occurs, however, since each center is different due to variations in patients as well as provider care, this can introduce bias into the analysis, even after post-hoc adjustment for center as a fixed effect [2]. Despite the common practice of including many centers with few patients per center, little is known about the combined behavior of number of centers and number of subjects per center on the estimation of treatment effect in the survival model and the effect of this imbalance across centers. This is of particular relevance for cancers, such as breast cancer (inflammatory breast cancer, Paget disease, and phyllodes tumor), and rarer cancers, for which enough events are needed to have sufficient power for survival analysis.

Cox proportional hazards models with random effects terms (frailty) are commonly used for survival data with censored outcomes that account for heterogeneity such as that between centers. It is noted that a traditional Cox model may have problems with estimation when there are few events overall such as in the case of rare diseases, however penalized likelihood methods can aid in estimation [3]. Our model will be estimated by penalized methods, hence reducing the bias in estimation. Also, since the centers will be accounted for by a random effect, we can avoid the issue of fixed effects which can lead to biased estimation of the center effect [4].

What is also of interest is how well the variance of the random effect is estimated in order to assess the model effect on the variability between centers. Duchateau [5, 6] have used a frailty model to model the heterogeneity between centers, as influenced by the number of centers and the number of patients through simulations. They assessed the change in variability by the frailty variance as a measure of the heterogeneity between centers. They found for a certain number of centers, either 15 or 30, there was little bias on treatment effect when increasing patient numbers. Although they varied numbers of centers and patients in those centers, they did not look at center size smaller than 20 nor patient size per center smaller than 15. We have attempted to address this issue for rarer diseases.

Smoothing techniques for modeling nonlinear relationships between one or more covariates and the response have been employed. Jeong and Constantino [7] used smoothing methods to evaluate treatment-prognostic factor effects on breast cancer data obtained from clinical trials. Smoothing has been incorporated in similar applications. Silverman *et al.* [8] used locally weighted scatterplot smoothing with Cox modeling to estimating trends of breast cancer incidence.

*Address correspondence to this author at the Department of Epidemiology and Biostatistics, SUNY Downstate School of Public Health, Brooklyn, NY 11203, USA; Tel: 718-221-6599; Fax: 718-270-2533;
E-mail: usha.govindarajulu@downstate.edu

E-ISSN: 1929-6029/15 © 2015 Lifescience Global

In this study, using comprehensive simulation, we evaluate properties of the Cox frailty model with a smoothed prognostic factor when there is a very small number of patients per center or when there are only a few centers but with more patients per center. We examine bias and inferences about the main treatment effect. Unlike Duchateau *et al.* [5, 6], Ambler *et al.* [3], and Silverman *et al.* [8], we also model a smoothed curve of a prognostic factor, such as age, for each using natural splines, similar to previous work [9], and further assess the model fit. We use a comprehensive simulation study to examine these features: estimation and testing of the treatment effect and the nonlinear prognostic variable relationship under different patient sample sizes and different number of centers. We also demonstrate this on a real dataset.

2. METHODS

2.1. Estimating Model

Survival models, such as the Cox proportional hazards (PH) model, are used to model time to an event. A random effect (frailty) term, incorporated as a shared frailty (frailty modeled between groups), accounts for the heterogeneity between centers. We incorporate a dichotomous treatment effect as a covariate and include a continuous prognostic variable. We use a natural spline smoothing function of the prognostic variable to model potential non-linear relationships.

In a Cox PH regression model this gives:

$$h_{ij}(t) = h_0(t)\exp(x_{ij}\beta + s(y_{ij}) + w_i) \quad (1)$$

where $h_0(t)$ is the baseline hazard rate, $i=1,\ldots,c$ centers and $j=1,\ldots,p_i$ patients per center, x_{ij} is a dichotomous treatment variable, y_{ij} is the prognostic variable, which we will use as age, and $s(y_{ij})$ is a smoothed function of y_{ij}. In addition, we include a log-normal frailty density for the center effect, similar to that employed by Duchateau *et al.* [5], where we estimate a given number of random center effects, with a particular number of patients per center.

While many options exist for smoothing the prognostic variable, such as restricted cubic splines, penalized splines, and fractional polynomials, due to their ease of implementation and behavior in previous simulation studies [9] we opted to use a natural spline (NS) function to model the non-linearity in $s(y_{ij})$. As implemented in the ns function of R software, natural splines are essentially restricted cubic splines which uses B-splines in a basis expansion of $s(y_{ij})$:

$$s(y_{ij}) = \gamma_0 y_{ij} + \sum_{h=1}^{H-2} \gamma_h \cdot B_h(y_{ij})$$

for $B_h(y_{ij})$ the B-spline basis functions. These are described in further detail [9]. We used the ns function in R with a default degrees of freedom (df), i.e. smoothness of df = 4. We employed penalized likelihood to estimate the frailty models [10].

2.2. Data Generation

We generated survival data using methods similar to those in Bender *et al*. [11], and described in more detail [9, 12], but adjusted to incorporate a center effect in the model. We use as the true generating model, the Cox PH model,

$$h(t \mid x_{ij}, y_{ij}, w_i) = h_0(t)\exp(\beta_T x_{ij} + s(y_{ij}) + w_i)\,. \quad (2)$$

As in the previous section, the subscripts *i* and *j* represent the center and patient respectively. The coefficient β_T corresponds to the dichotomous treatment effect x_{ij}, and we examined three treatment scenarios, β_T =1, β_T =0, and β_T = -1. The treatment group assignment, x_{ij}, was generated so that 84% were treated, on average. The $s(y_{ij})=\log(y_{ij})$ to allow for non-linearity in the prognostic factor, y_{ij}, using a true logarithmic association or $s(y_{ij})=y_{ij}$ to allow for linearity. We sampled y_{ij} randomly with replacement from a range of 30 to 66 to mimic a reasonable age distribution of subjects enrolled. The center effect, w_i, was generated from a $N(0,\sigma^2)$ distribution, with values σ^2 = 1 and σ^2= 0.25^2 used in the simulations.

Bender *et al*. [11] discussed the use of different distributions for generating survival times. We allowed for a baseline Weibull hazard [11, 13], so that $h_0(t) = \theta v t^{\theta-1}$, depending on parameters θ and v, are chosen to generate realistic survival data. Survival times are found from this generating distribution by using the relationship between the hazard function, survival function, $S(t)$, and cumulative distribution function (CDF), $F(t)$, of the survival time random variable, T.

Given the baseline hazard function, $h_0(t)$, the CDF is found through the relationship $F(t) = 1 - S(t) = 1 - e^{-H(t)}$ where $H(t)$ is the cumulative hazard function. This in turn is given by $H(t) = e^{g(x,y,w)}\int_0^t h_0(u)du = e^{g(x,y,w)} v t^{\theta}$, where $g(x,y,w)$ = $\beta_T x + s(y) + w$ is the true log hazard

ratio defined in (2) above. Using the probability integral transformation, for a given *t*, the CDF, *F,* has a Uniform distribution on the interval from 0 to 1. Generating *U1* from a Uniform (0,1) distribution and solving for *t* gives the simulated survival time, $t = \left(\frac{-e^{-g(x,y,w)} \log(1-U1)}{v} \right)^{1/\theta}$ for given values of *x*, *y*, *w*, θ and v.

For added realism, we also incorporate a competing risk into the simulations. Competing risk times were generated in a similar manner based on an exponential distribution with scale parameter γ, giving, $t_{cr} = \frac{\log(1-U2)}{\gamma}$ where *U2* ~ Uniform(0,1). The observed survival time was taken to be the minimum of t_0, t_{cr}, and a pre-specified end-of-study time, τ. An observation was considered censored if t_0 was larger than the minimum of t_{cr} and τ. The user-specified parameters, θ, v,γ, and τ were set to generate realistic survival time distributions. We used τ = 20 years of follow-up time and θ=5 which is considered typical of many types of cancer [14, 15]. The parameters, v and γ, were chosen to control the amount of censoring in the simulated data. We had 71 cases on average in the dataset where the average number of cases increased across the scenarios from 47 to 91. The final simulated datasets contain: (*x*, *y*, *t*, *d*, *c*): treatment, the prognostic factor (age), survival time, event indicator, and center number.

2.3. Performance Measures

To assess the performance of the Cox model under the different simulation conditions, we examined a number of measures: 1) bias in the treatment coefficient estimate, 2) bias in the non-linear prognostic effect, 3) power to detect a treatment effect, 4) power to detect the treatment and prognostic effects simultaneously, 5) power to detect the prognostic effect on a nonlinear scale, and 6) the estimated variance between centers. We describe each of these in more detail.

The bias in the treatment coefficient estimate is the difference in the true generating coefficient, β_T, in model (2) and the estimate of the coefficient, β, in model (1). Bias in the estimated smoothed curves of the prognostic variable are summarized using mean-squared error [9, 12]. This computed as $rMSE = \sqrt{\sum_{k=1}^{K} \left[s(y_{ij}) - \hat{s}_k(y_{ij}) \right]^2}$ where $\hat{s}_k(y_{ij})$ is the estimated curve for patient *j* at center *i* in the kth simulated data and $s(y_{ij}) = \log(y_{ij})$ or $s(y_{ij}) = y_{ij}$ is the corresponding true value of curve. The distribution of rMSE is presented in boxplots and the median of the rMSE was used to summarize model fit to reduce the influence of the tails.

We also examined two hypothesis tests related to the natural spline fit. The first is a test of the overall model and the second is a test of a regression effect of age as estimated by NS. Both tests were conducted using likelihood ratio test statistics. We present the proportion of significant tests (p-values less than or equal to 0.05) over all simulated data sets.

3. RESULTS

3.1. Simulations

All simulations were performed in R software Version 3.1.1. Table **1a** gives the number of patients per site and number of sites that were used in the simulation study. These were selected to conform to what has been seen in practice with multicenter trial designs. We simulated K = 1000 data sets per each of 96 scenarios defined by the number of sites, the number of patients per site (as given in Table **1a**), the treatment effect (β_T = -1,1, and 0), the variance between centers (σ^2 = 1 and 0.25^2), and the choice of parameters, v and γ, from the generating distributions described in Section 2.2. The parameters, v and γ, impact the amount of censoring in the simulated survival data and we thus list the number of resulting cases, on average, per scenario. Table **1b** gives these quantities for each of the 96 scenarios.

Overall, the 25th, 50th and 75th percentiles for the prognostic factor, age, were 39, 48, and 57, respectively. Cases were only slightly older than non-cases but in general very close. In Table **2a**, the test of no effect of treatment nor age is rejected most often for the scenarios which consist of five sites or for true linear age with either low number of cases and no treatment or positive treatment or high number of cases and least often for scenarios consisting of ten sites for true nonlinear age. The test for the regression age effect on the natural spline scale was rejected most often for those scenarios where the frailty variance was 1.0 as compared to those with variance of 0.25^2 and the p-value increases as number of sites increase, holding other parameters constant. Once there was a true linear age and high number of cases, then it was always rejected which is expected. NS curves started fitting better for scenarios 17-32 and 65-96; the median and IQR of the rMSE appears to be lower for these scenarios, and therefore, better fitting, while they were worse fitting for true linear age with treatment

Table 1a: Number of Sites and Number of Patients/Site for Simulations Scenarios

Sites	Number of sites			
	3	5	10	20
Number of patients per site for a given scenario	Site 1=16 Site 2=32 Site 3=64	Site1=10 Site2=15 Site3=20 Site4=27 Site5=40	Site1=4 Site2=4 Site3=8 Site4=8 Site5=12 Site6=12 Site7=16 Site8=16 Site9=16 Site10=16	Site1=3 Site2=3 Site3=3 Site4=3 Site5=3 Site6=3 Site7=5 Site8=5 Site9=5 Site10=5 Site11=5 Site12=5 Site13=8 Site14=8 Site15=8 Site16=8 Site17=8 Site18=8 Site19=8 Site20=8

Table 1b: Simulation Parameter Settings for all Scenarios

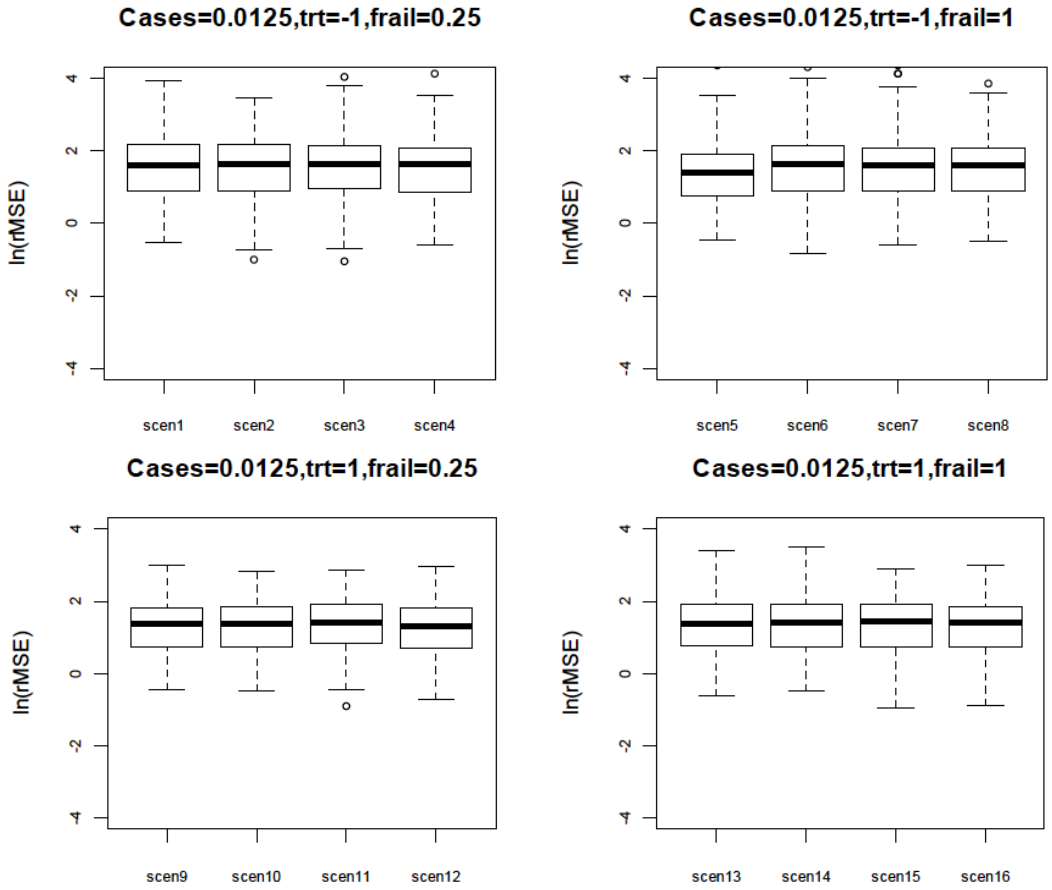
Scenario #	# of sites for each respective scenario	Treatment effect β_T	Frailty variance	Average number of cases across scenarios	True age effect
1,2,3,4	3,5,10,20	-1	0.25^2	31	log
5,6,7,8	3,5,10,20	-1	1	39	log
9,10,11,12	3,5,10,20	1	0.25^2	59	log
13,14,15,16	3,5,10,20	1	1	58	log
17,18,19,20	3,5,10,20	-1	0.25^2	88	log
21,22,23,24	3,5,10,20	-1	1	82	log
25,26,27,28	3,5,10,20	1	0.25^2	98	log
29,30,31,32	3,5,10,20	1	1	95	log
33,34,35,36	3,5,10,20	0	0.25^2	43	log
37,38,39,40	3,5,10,20	0	1	46	log
41,42,43,44	3,5,10,20	0	0.25^2	97	log
45,46,47,48	3,5,10,20	0	1	91	log
49,50,51,52	3,5,10,20	-1	0.25^2	29	linear
53,54,55,56	3,5,10,20	-1	1	32	linear
57,58,59,60	3,5,10,20	0	0.25^2	39	linear
61,62,63,64	3,5,10,20	0	1	41	linear
65,66,67,68	3,5,10,20	1	0.25^2	53	linear
69,70,71,72	3,5,10,20	1	1	53	linear
73,74,75,76	3,5,10,20	-1	0.25^2	79	linear
77,78,79,80	3,5,10,20	-1	1	76	linear
81,82,83,84	3,5,10,20	0	0.25^2	89	linear
85,86,87,88	3,5,10,20	0	1	85	linear
89,90,91,92	3,5,10,20	1	0.25^2	94	linear
93,94,95,96	3,5,10,20	1	1	90	linear

Table 2a: Properties of Estimated Natural Spline Curve in 1000 Simulated Datasets Generated for each of 96 Scenarios

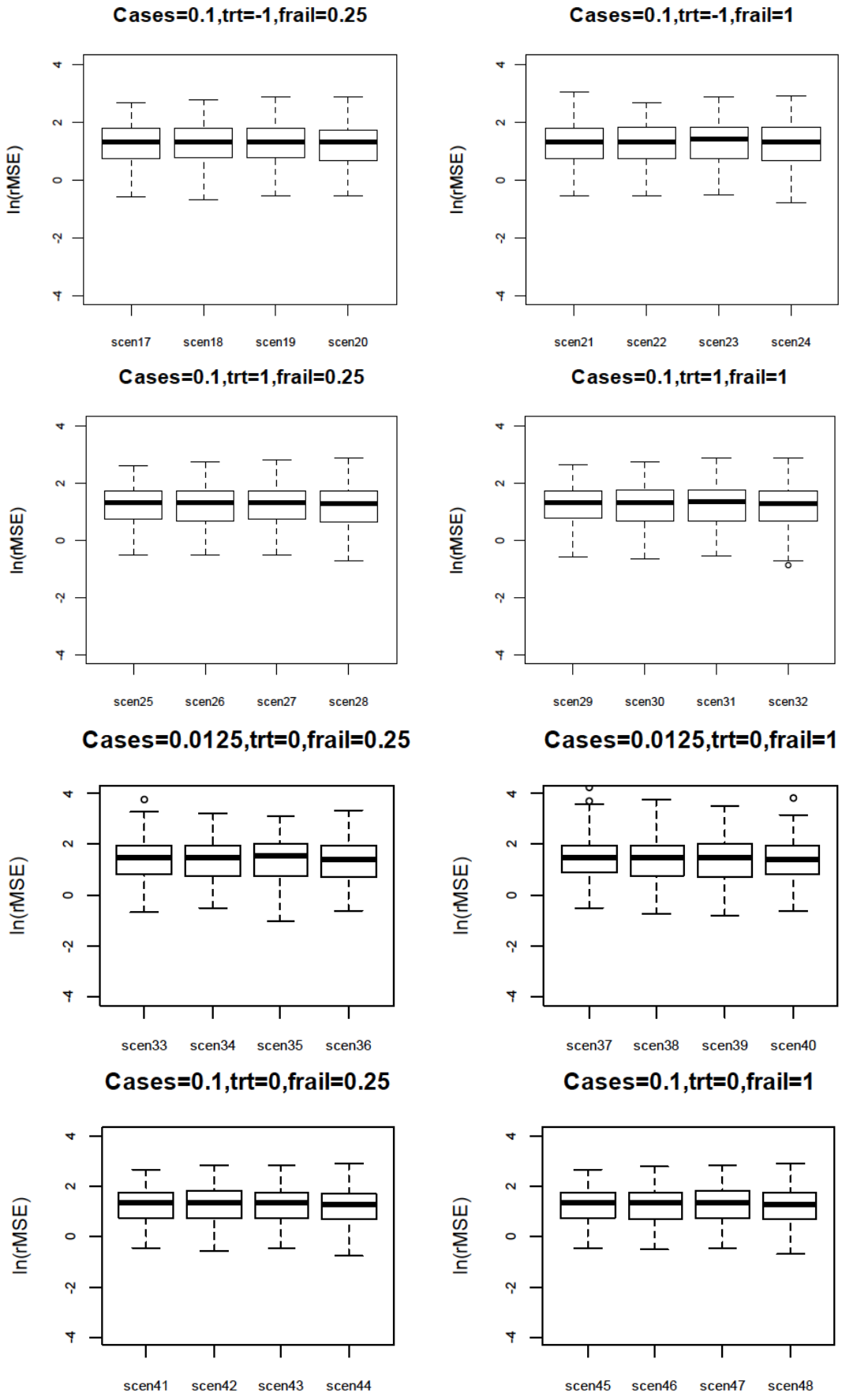
Test	Number of sites			
Scenarios	**3**	**5**	**10**	**20**
proportion of times test for no effect of both treatment and age is rejected				
1-4	0.981	1.000	0.827	0.859
5-8	0.997	1.000	0.982	0.989
9-12	0.997	1.000	0.979	0.983
13-16	0.999	1.000	1.000	0.999
17-20	1.000	1.000	0.992	0.999
21-24	1.000	1.000	1.000	1.000
25-28	1.000	1.000	0.998	0.998
29-32	1.000	1.000	1.000	1.000
33-36	0.867	0.999	0.588	0.648
37-40	0.967	1.000	0.983	0.993
41-44	0.949	1.000	0.819	0.855
45-48	0.995	1.000	0.998	1.000
49-52	0.999	1.000	0.962	0.990
53-56	0.995	1.000	0.997	0.998
57-60	1.000	0.989	0.994	0.999
61-64	1.000	1.000	1.000	1.000
65-68	1.000	1.000	1.000	1.000
69-72	1.000	1.000	1.000	1.000
73-76	1.000	1.000	1.000	1.000
77-80	1.000	1.000	1.000	1.000
81-84	1.000	1.000	1.000	1.000
85-88	1.000	1.000	1.000	1.000
89-92	1.000	1.000	1.000	1.000
93-96	1.000	1.000	1.000	1.000
proportion of times test for age effect from NS fit is rejected				
1-4	0.423	0.512	0.565	0.637
5-8	0.887	0.923	0.961	0.981
9-12	0.593	0.705	0.684	0.750
13-16	0.896	0.979	0.993	0.997
17-20	0.715	0.790	0.815	0.839
21-24	0.952	0.985	0.998	0.999
25-28	0.746	0.815	0.834	0.864
29-32	0.962	0.993	0.998	0.999
33-36	0.495	0.596	0.636	0.678
37-40	0.865	0.961	0.982	0.994
41-44	0.735	0.815	0.833	0.866
45-48	0.964	0.992	0.998	0.999
49-52	0.957	0.974	0.923	0.979
53-56	0.964	0.990	0.995	0.997
57-60	0.992	0.995	0.995	0.996
61-64	0.985	0.996	1.000	1.000
65-68	1.000	0.998	1.000	0.999
69-72	0.997	0.997	1.000	1.000
73-76	1.000	1.000	1.000	1.000
77-80	1.000	1.000	1.000	1.000
81-84	1.000	1.000	1.000	1.000
85-88	1.000	1.000	1.000	1.000
89-92	1.000	1.000	1.000	1.000
93-96	1.000	1.000	1.000	1.000

(Table 2a). Continued.

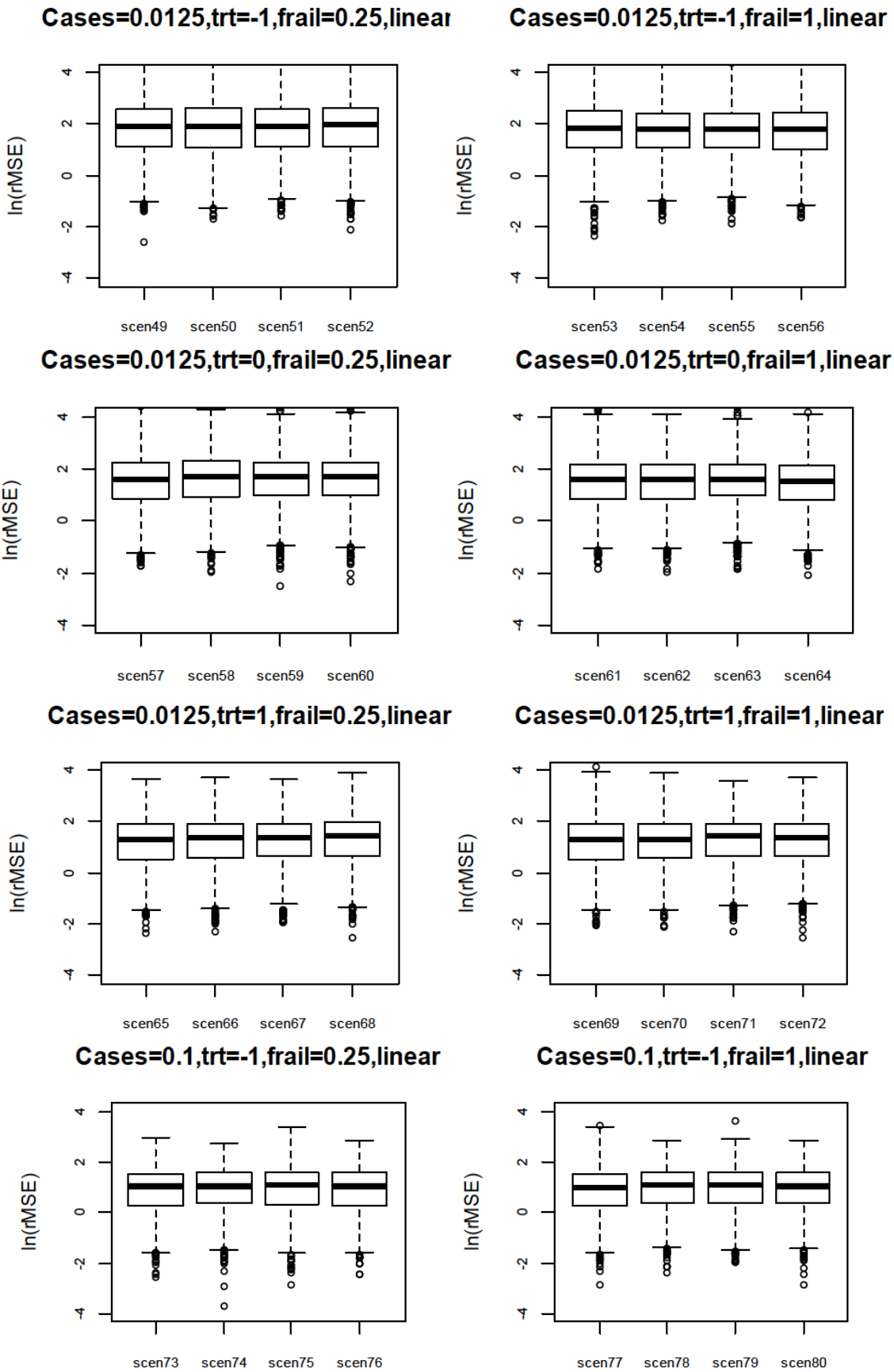
Median (IQR) root mean-square error				
1-4	4.81 (6.17)	5.08 (6.17)	5.15(5.93)	5.11 (5.68)
5-8	4.09 (4.58)	5.11 (6.16)	4.82 (5.63)	4.79 (5.46)
9-12	4.06 (4.28)	4.03 (4.44)	4.23 (4.45)	3.83 (4.22)
13-16	4.03 (4.56)	4.09 (4.57)	4.29 (4.79)	4.13 (4.30)
17-20	3.73 (3.70)	3.75 (3.85)	3.78 (3.80)	3.65 (3.76)
21-24	3.72 (3.74)	3.75 (4.06)	4.03 (4.02)	3.65 (3.99)
25-28	3.71 (3.56)	3.75 (3.85)	3.87 (3.64)	3.65 (3.67)
29-32	3.82 (3.44)	3.73 (3.93)	3.92 (4.02)	3.67 (3.67)
33-36	4.33 (5.08)	4.47 (5.08)	4.58 (5.42)	4.09 (4.90)
37-40	4.35 (4.60)	4.50 (5.11)	4.37 (5.41)	3.76 (3.58)
41-44	3.76 (3.58)	3.71 (3.86)	3.89 (3.61)	3.70 (3.60)
45-48	3.77 (3.59)	3.70 (3.95)	3.91 (3.91)	3.63 (3.86)
49-52	6.81 (9.99)	6.91 (10.86)	6.60 (9.73)	7.11 (10.47)
53-56	6.31 (9.34)	5.89 (8.12)	6.10 (8.14)	6.08 (9.00)
57-60	4.97 (7.27)	5.32 (7.32)	5.40 (6.78)	5.37 (6.81)
61-64	4.84 (6.28)	4.85 (6.46)	4.90 (6.05)	4.70 (6.26)
65-68	3.77 (5.10)	3.97 (5.00)	3.93 (4.81)	4.19 (5.29)
69-72	3.78 (5.08)	3.79 (5.16)	4.16 (5.00)	4.02 (4.86)
73-76	2.79 (3.31)	2.88 (3.36)	2.92 (3.46)	2.77 (3.45)
77-80	2.66 (3.29)	2.97 (3.43)	2.97 (3.48)	2.86 (3.47)
81-84	2.47 (2.93)	2.53 (3.13)	2.47 (2.95)	2.40 (3.05)
85-88	2.49 (2.89)	2.53 (3.11)	2.64 (3.15)	2.60 (3.29)
89-92	2.21 (2.71)	2.32 (2.82)	2.30 (2.75)	2.28 (2.77)
93-96	2.21 (2.80)	2.37 (2.73)	2.50 (2.94)	2.35 (3.01)



(Figure 1). Continued.



(Figure 1). Continued.



(Figure 1). Continued.

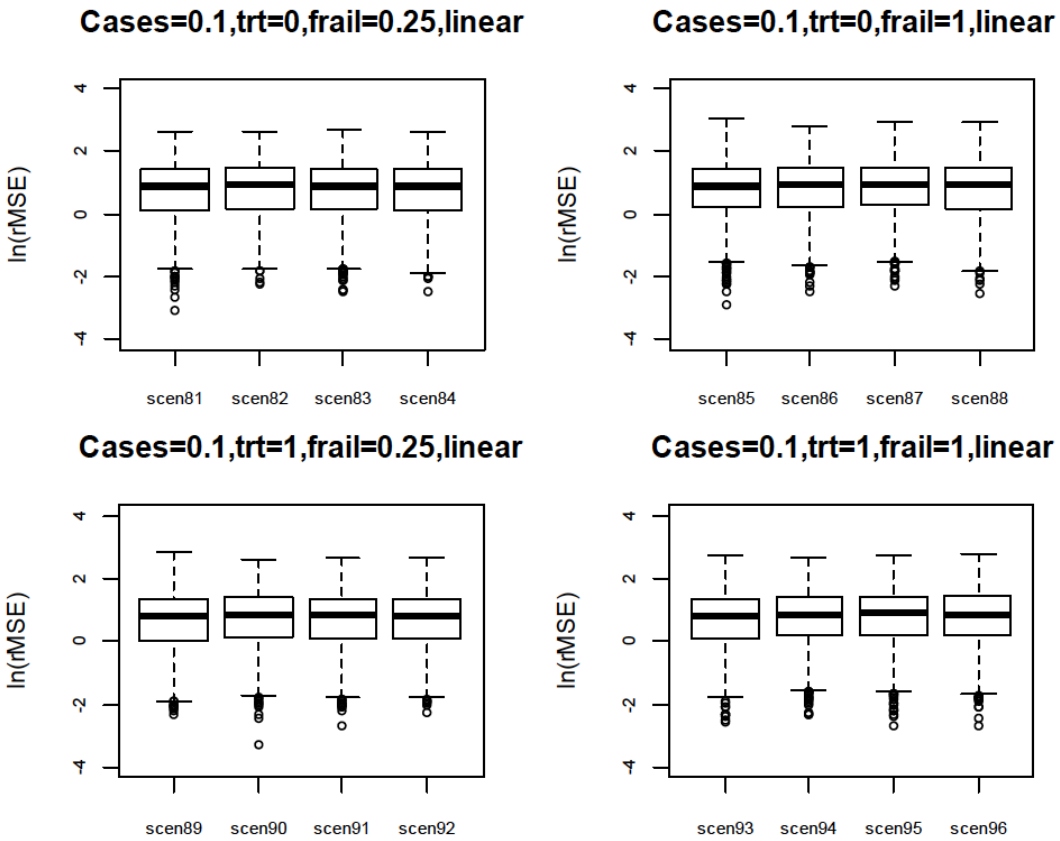


Figure 1: Boxplots of natural logarithm of rMSE of fitted natural spline curve.

coefficients of -1 or 0 and low number of cases. Also, the rMSE appeared lower for scenarios with 20 sites in general. Boxplots in Figure **1** have natural log-transformed MSE for better visualization. According to the boxplots, there is not much variability in rMSE across varying number of sites, although there does appear to be slightly more variability as seen in the boxplots for the early scenarios, 1-4 and 5-8, which have the fewest cases and for all scenarios with true linear age effect.

The proportion of times the treatment effect was rejected does appear to increase across the 96 scenarios and is the highest for the later scenarios but where treatment coefficient was 0, there was still some rejection up to 10% for 20 sites so the type I error rate almost doubled for these (Table **2b**). The median treatment coefficient and the average frailty variance appeared consistent with the corresponding scenarios in terms of parameter settings although the variability (frailty variance) appears to be lower for scenarios with more sites. Across scenarios, the models were better able to fit a frailty variance of 1.0 than of 0.25 and were underestimating the frailty variance in scenarios set for 0.25.

Figures **2a-2f** demonstrate the log-hazard curve that generated the median rMSE as a function of age. For each scenario, the solid line is the truth and the dashed curve is the NS spline curve that represents the one that has the minimum median rMSE amongst the 1000 simulated datasets. It appears that the curves in Figure **2b** in general show the NS curve more closely approximating the truth than in Figure **2a**, where there is a lower proportion of cases. It does not appear that there is a definitive trend by number of sites, although the scenarios with 20 sites and the lowest number of patients per site seemed to have the more deviant curves. This trend appeared to follow for the true linear age affect (Figures **2d-2f**), where the estimated curves were closer to the truth for high number of cases.

3.2. Real Data Example

We used a rare cancer dataset provided by Eastern Cooperative Oncology Group (COG) for which a clinical trial was conducted with a total of 644 subjects at 299 sites for a mix of rare cancers (testicular, ovarian, or extragonadal) in which people on either treatment or placebo were followed forward for time to an event of death during the years of 2007 to 2010. We applied our

Table 2b: Properties of Estimated Treatment Effect in 1000 Simulated Datasets Generated for each of 96 Scenarios

Test	Number of sites			
Scenarios	**3**	**5**	**10**	**20**
Proportion of times test of no treatment effect was rejected				
1-4	0.766	0.741	0.721	0.727
5-8	0.900	0.762	0.748	0.791
9-12	0.952	0.952	0.957	0.923
13-16	0.912	0.946	0.943	0.934
17-20	0.994	0.988	0.984	0.966
21-24	0.986	0.986	0.974	0.979
25-28	0.995	0.995	0.994	0.980
29-32	0.995	0.994	0.996	0.982
33-36	0.064	0.070	0.088	0.101
37-40	0.058	0.063	0.082	0.097
41-44	0.074	0.074	0.083	0.108
45-48	0.071	0.060	0.076	0.095
49-52	0.734	0.717	0.680	0.709
53-56	0.729	0.721	0.725	0.748
57-60	0.061	0.070	0.099	0.106
61-64	0.065	0.072	0.087	0.107
65-68	0.927	0.929	0.915	0.876
69-72	0.895	0.919	0.922	0.895
73-76	0.989	0.977	0.970	0.946
77-80	0.984	0.973	0.968	0.960
81-84	0.072	0.070	0.078	0.090
85-88	0.073	0.067	0.080	0.084
89-92	0.994	0.994	0.992	0.974
93-96	0.991	0.992	0.991	0.982
Median (sd) of treatment coefficient				
1-4	-1.08 (0.44)	-1.06 (0.44)	-1.04 (0.45)	-1.09 (0.44)
5-8	-1.08 (0.37)	-1.06 (0.44)	-1.02 (0.44)	-1.06 (0.43)
9-12	1.03 (0.30)	1.06 (0.31)	1.09 (0.31)	1.07 (0.31)
13-16	1.04 (0.33)	1.06 (0.32)	1.06 (0.32)	1.04 (0.32)
17-20	-1.08 (0.26)	-1.06 (0.26)	-1.03 (0.26)	-1.06 (0.26)
21-24	-1.07 (0.27)	-1.04 (0.27)	-1.02 (0.27)	-1.03 (0.28)
25-28	1.04 (0.25)	1.06 (0.25)	1.07 (0.24)	1.08 (0.25)
29-32	1.03 (0.25)	1.05 (0.26)	1.05 (0.25)	1.05 (0.27)
33-36	-0.01 (0.34)	0.02 (0.34)	0.03 (0.35)	-0.002 (0.33)
37-40	-0.01 (0.36)	0.01 (0.35)	0.03 (0.35)	0.001 (0.34)
41-44	-0.02 (0.24)	0.003 (0.23)	0.02 (0.23)	0.01 (0.24)
45-48	-0.02 (0.24)	0.01 (0.24)	0.02 (0.24)	0.01 (0.26)
49-52	-1.08 (0.47)	-1.07 (0.46)	-1.01 (0.44)	-1.15 (1.44)
53-56	-1.15 (1.31)	-1.06 (0.56)	-1.02 (0.45)	-1.05 (0.47)
57-60	-0.01 (0.35)	0.001 (0.35)	0.03 (0.37)	-0.004 (0.36)
61-64	-0.01 (0.38)	0.01 (0.38)	0.03 (0.38)	0.01 (0.38)
65-68	1.04 (0.32)	1.06 (0.33)	1.09 (0.34)	1.07 (0.33)
69-72	1.04 (0.35)	1.06 (0.34)	1.07 (0.35)	1.04 (0.35)
73-76	-1.07 (0.27)	-1.04 (0.27)	-1.03 (0.27)	-1.05 (0.27)
77-80	-1.07 (0.28)	-1.04 (0.28)	-1.02 (0.28)	-1.04 (0.29)
81-84	-0.02 (0.24)	0.004 (0.24)	0.02 (0.23)	0.01 (0.24)
85-88	-0.03 (0.25)	0.002 (0.25)	0.02 (0.25)	0.01 (0.26)
89-92	1.03 (0.25)	1.05 (0.25)	1.07 (0.25)	1.07 (0.25)
93-96	1.03 (0.26)	1.05 (0.26)	1.05 (0.26)	1.05 (0.27)

(Table 2b). Continued.

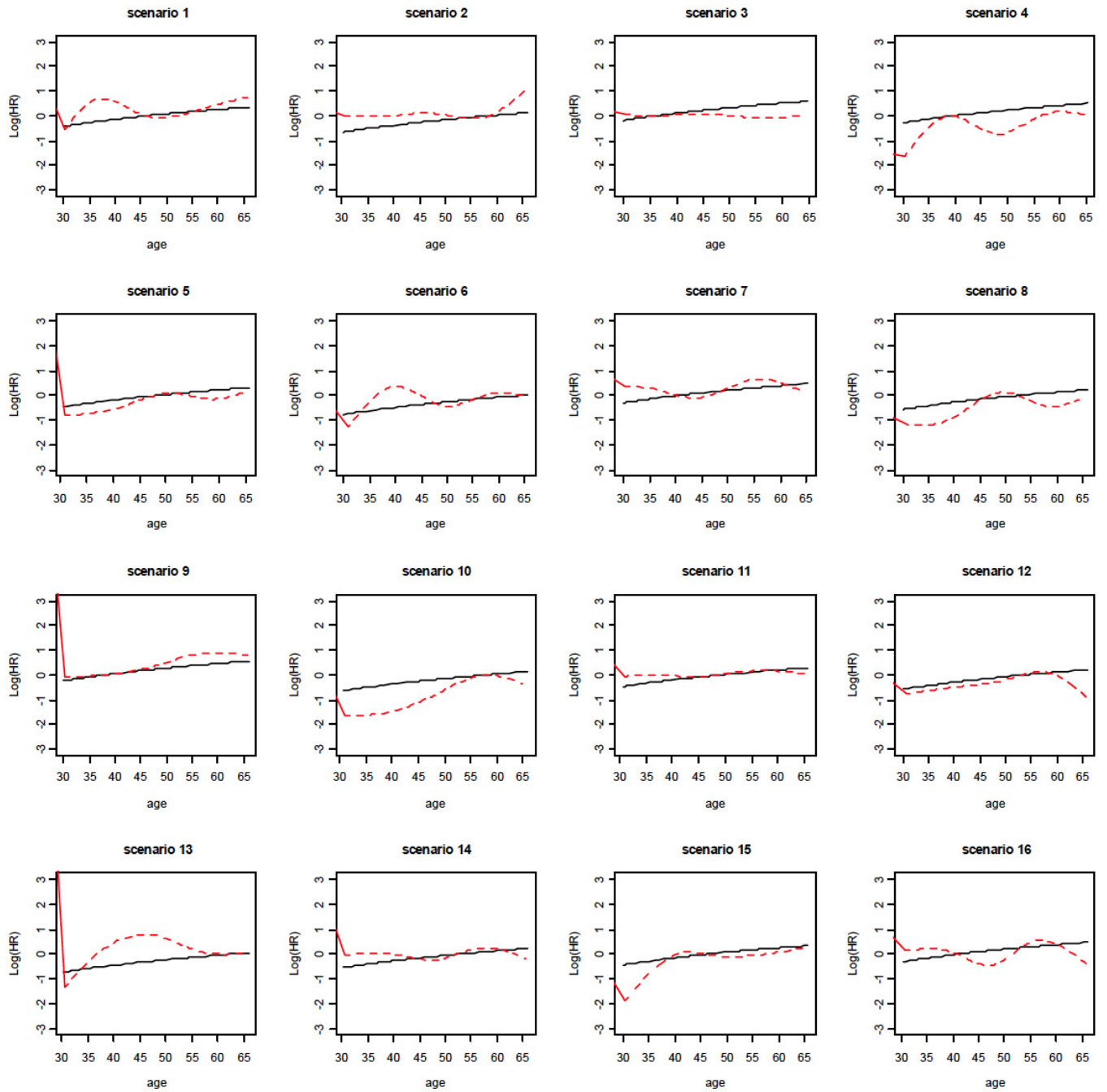
Average frailty variance				
1-4	0.134	0.128	0.126	0.173
5-8	1.185	1.158	0.926	0.958
9-12	0.095	0.101	0.093	0.129
13-16	1.190	1.168	0.975	1.015
17-20	0.093	0.089	0.082	0.112
21-24	1.144	1.166	0.999	1.059
25-28	0.090	0.088	0.081	0.105
29-32	1.148	1.160	1.002	1.071
33-36	0.107	0.106	0.103	0.142
37-40	1.222	1.159	0.955	0.981
41-44	0.092	0.091	0.083	0.111
45-48	1.156	1.165	1.000	1.066
49-52	0.137	0.139	0.110	0.191
53-56	1.248	1.189	0.949	0.966
57-60	0.116	0.119	0.110	0.159
61-64	1.199	1.148	0.961	0.979
65-68	0.101	0.104	0.100	0.136
69-72	1.155	1.167	0.990	1.017
73-76	0.093	0.089	0.083	0.110
77-80	1.130	1.163	1.001	1.054
81-84	0.091	0.088	0.081	0.108
85-88	1.143	1.155	1.004	1.056
89-92	0.088	0.087	0.079	0.104
93-96	1.136	1.154	1.002	1.066

model to the data to predict time to event and we used age at diagnosis as a prognostic factor. Comparing the traditional model fit to our model (Tables **3a**/**3b**), there was some improvement in the treatment effect by employing our modeling. The frailty variance was 0.0007, which showed there was not much heterogeneity between centers. The p-value for the treatment effect was 0.07 so potentially borderline significant (Table **3b**). The p-value for the test for age effect from NS fit being rejected was 0.059. The p-value for the test for no effect of both treatment and age being rejected was 0.035. Therefore, though it appeared there may not be a significant treatment effect, there was some effect of age in the model, and the overall model was statistically significant. Our method allows plotting of age and the log(HR), allowing a smoothed view of the non-linearity of the treatment-prognostic curve (Figure **3**).

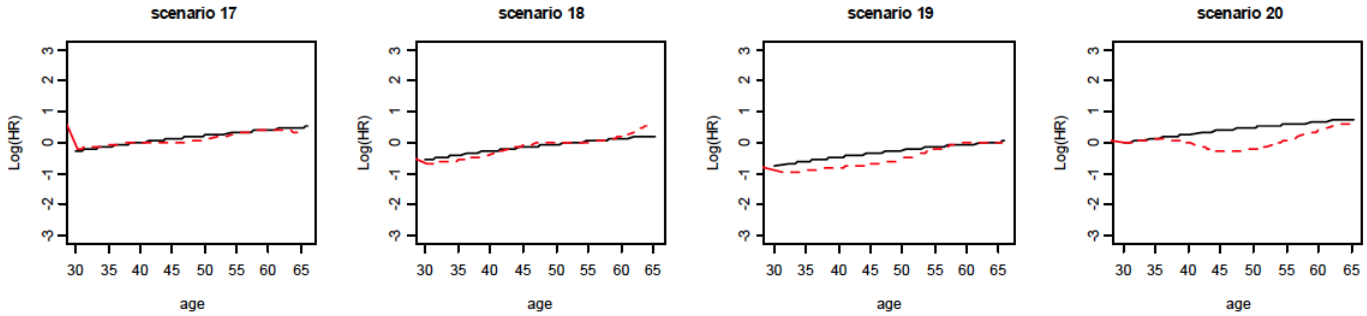
4. DISCUSSION

We have conducted a simulation study on potential patterns of multicenter clinical trials for a rare outcome with a prognostic factor such as rare forms of cancer. Many of these studies are small and able to recruit only a few patients per study site or even involve few centers in the overall study. We analyzed different scenarios in which numbers of sites and numbers of patients per site were varied in light of being able to recruit few patients per center. Using various criteria to judge the scenarios, such as proportion of times treatment effect was rejected, bias between estimated curves, and proportion of times regression age effect from NS fit was rejected, we were able to ascertain model fits from each scenario. From our previous paper and research, in the test of no null effect for a NS fit of the covariate the test of null effect was rejected 84% of the time under the true log scenario and 100% of the time under the linear scenario. Though our current results examine overall model fit due to having additional covariates, it keeps consistent with the previous null model results except for scenarios 35 and 36.

Overall, it is still preferable to have more patients per site and more sites and higher case rate, rather than a few patients per site. However, we have demonstrated that having a few patients per site, as low as 3 patients per site, is feasible if there are many sites in the trial (such as 20 sites) with higher rather than lower case rate. In addition, we were able to demonstrate the use of this modeling technique on an actual dataset of rare cancers which had anywhere from 1 to 20 patients per center. This allowed us to

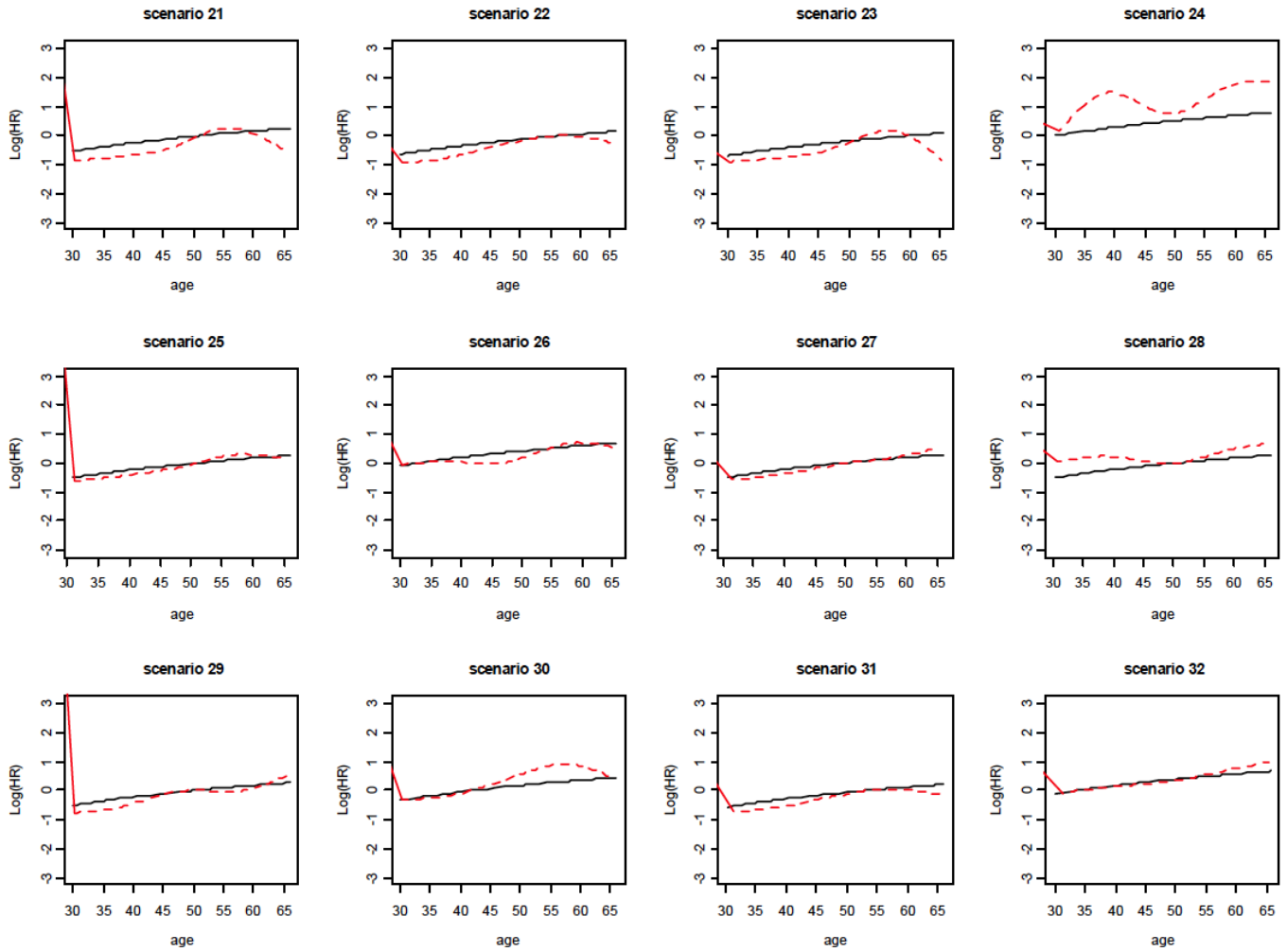

scenario 1
scenario 2
scenario 3
scenario 4
scenario 5
scenario 6
scenario 7
scenario 8
scenario 9
scenario 10
scenario 11
scenario 12
scenario 13
scenario 14
scenario 15
scenario 16
Log(HR)
age

a

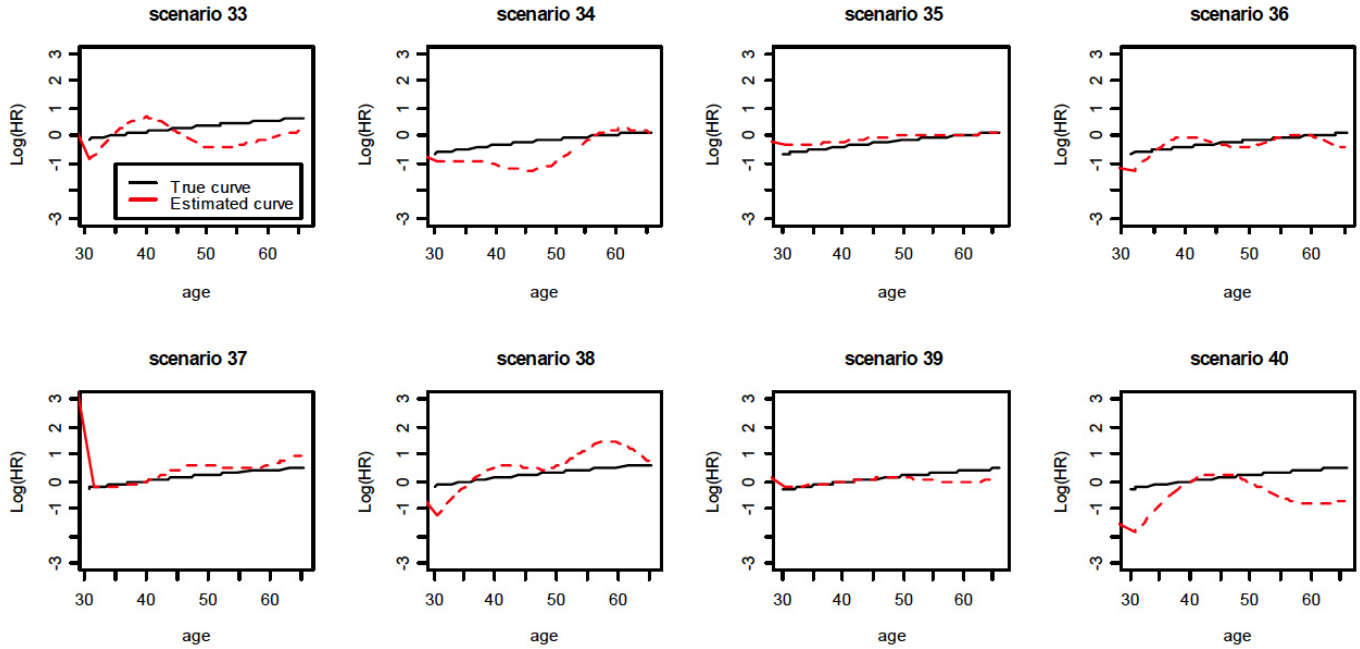

scenario 17
scenario 18
scenario 19
scenario 20
Log(HR)
age

b

(Figure 2b). Continued.

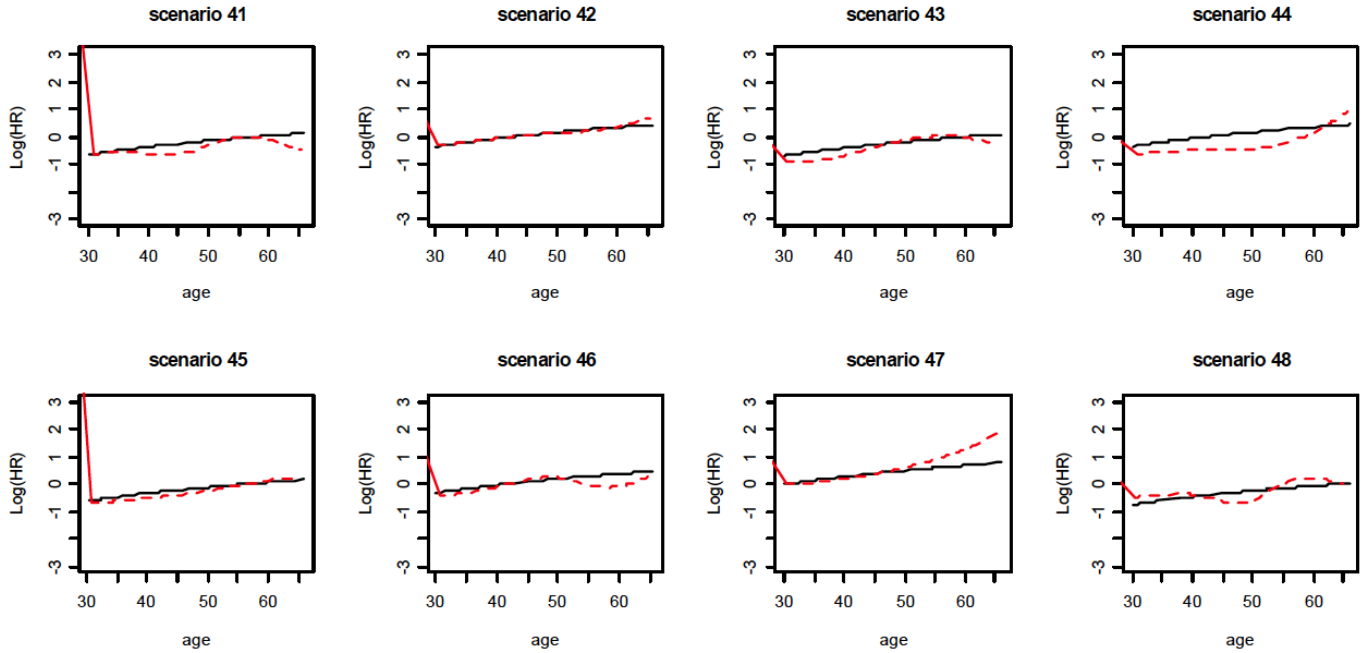


b

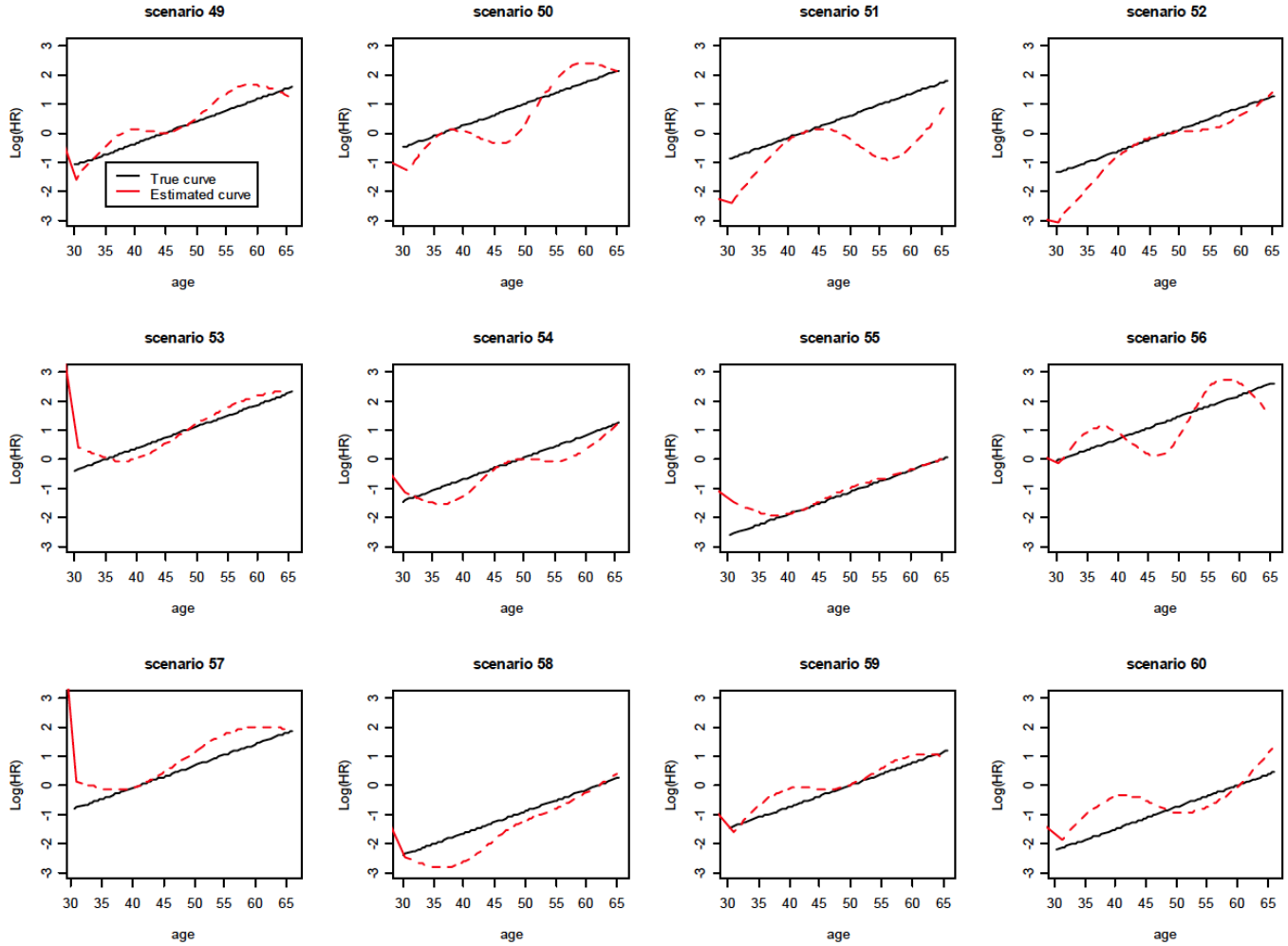


c

(Figure 2c). Continued.

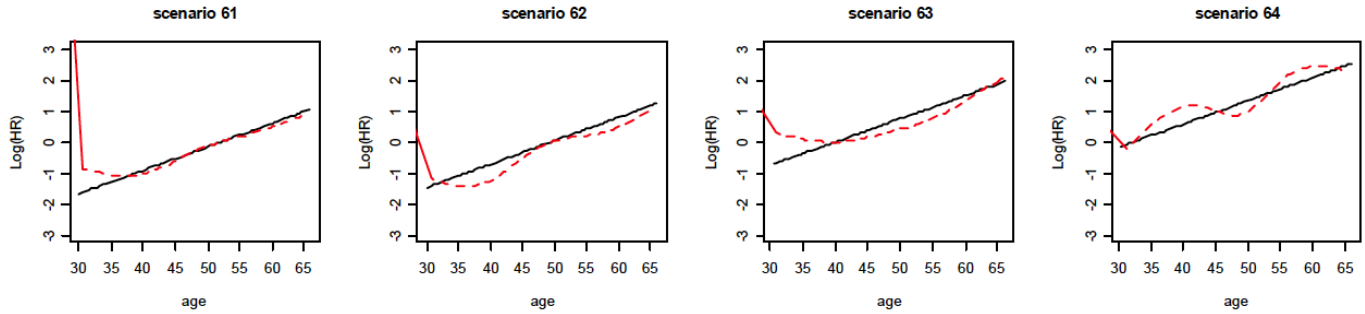


c

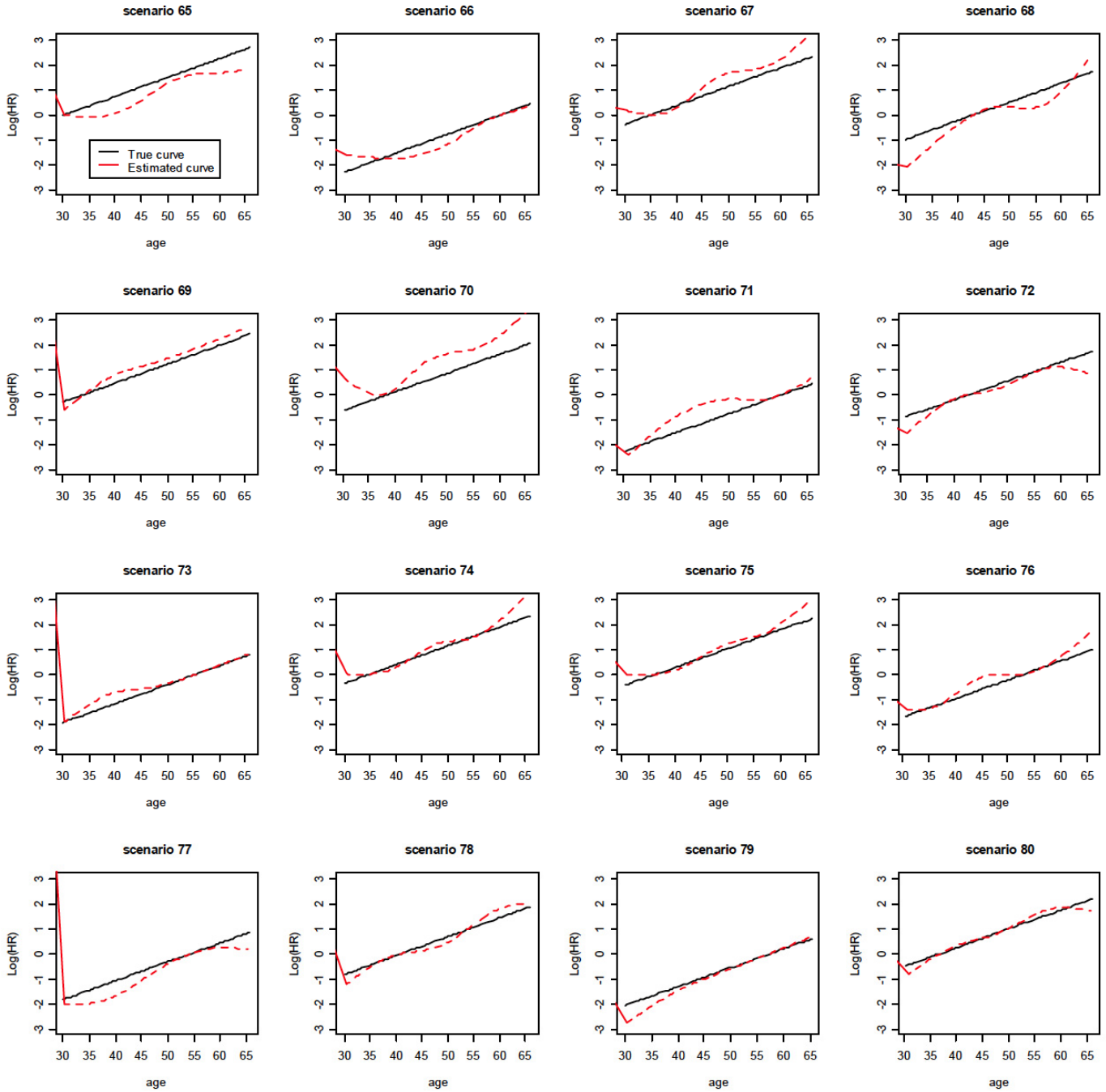


d

(Figure 2d). Continued.

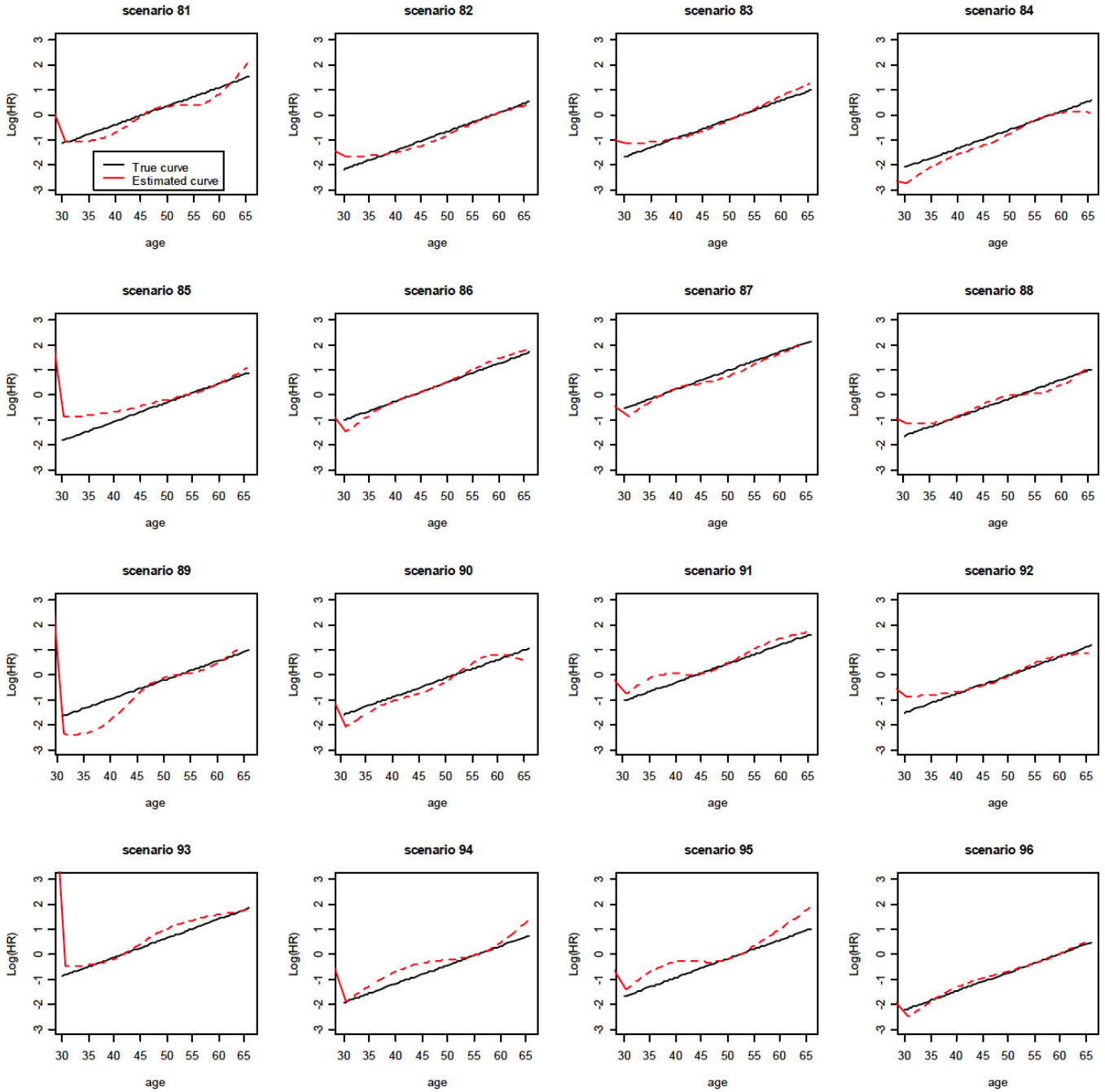


d



e

(Figure 2f). Continued.



f

Figure 2: a: Log hazard ratio of fitted median NS curves at minimum rMSEacross age for scenarios 1-16.

b: Log hazard ratio of fitted median NS curves at minimum rMSE across age for scenarios 17-32.

c: Log hazard ratio of fitted median NS curves at minimum rMSE across age for scenarios 33-48.

d: Log hazard ratio of fitted median NS curves at minimum rMSE across age for scenarios 49-64.

e: Log hazard ratio of fitted median NS curves at minimum rMSE across age for scenarios 65-80.

f: Log hazard ratio of fitted median NS curves at minimum rMSE across age for scenarios 81-96.

model potential heterogeneity between centers as well as the treatment effect and probable non-linearity. We were also able to show this under the scenario when the true underlying effect of the prognostic factor is either linear or non-linear. In fact a potential drawback of this modeling is relying only on the natural spline to

Table 3a: Cox Model Results for Rare Cancer Dataset Example

Variable	Coefficient	Standard error	Chi-square	p-value
Treatment	0.629	0.372	2.86	0.091
age	0.041	0.028	2.16	0.140

Table 3b: Cox Model Results with ns and Frailty for Rare Cancer Dataset Example

Variable	Coefficient	Standard error	Chi-square	p-value
Treatment	0.655	0.374	3.08	0.079
Ns-age1	-3.060	1.802	2.88	0.089
Ns-age2	0.837	1.122	0.56	0.460
Ns-age3	0.703	1.364	0.27	0.610
Ns-age4	0.396	0.962	0.17	0.680
Frailty			0.03	0.600

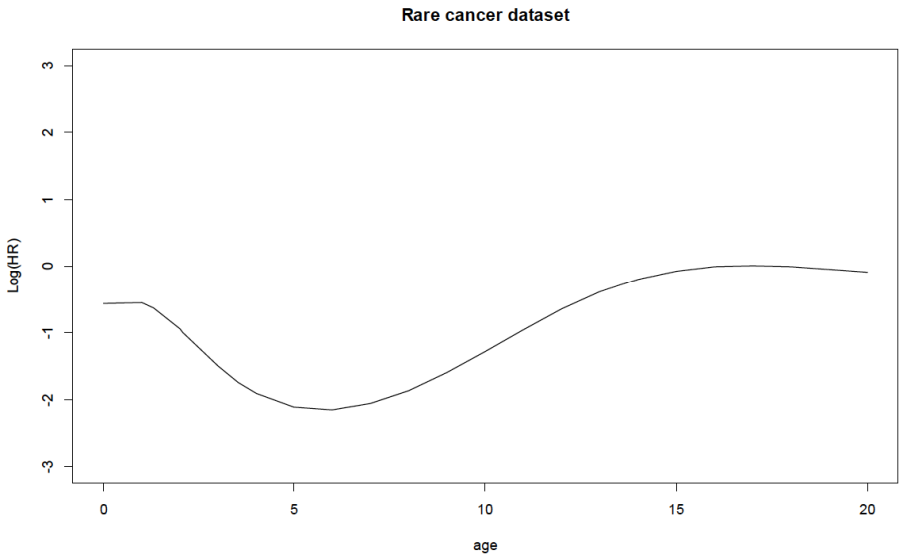


Figure 3: Log hazard ratio of fitted median NS curve at minimum rMSE across age for rare cancer dataset example.

smooth the prognostic factor. Also, our actual dataset example appeared to have a non-linear effect in age so we couldn't see if the age effect had been truly linear in this real example.

We were also able to demonstrate the use of modeling the heterogeneity between centers with frailty, similar to Duchateau [5, 6], but this time with far fewer sites and fewer patients per site. We were also able to incorporate nonlinear effects of the prognostic factor through modeling by smoothing, which aids in visualization of plots of hazard rate by that factor. In addition, the penalized estimation approach is used for both the frailty fit and the nonlinear prognostic factor fit [10]. Our hope is that this modeling can be employed for rare disease datasets from studies conducted at multicenters with survival outcomes.

While the paper provides guidance for Cox proportional hazards regression of a survival outcome, modeling the heterogeneity between centers through the frailty effect and being able to obtain a smoothed estimate for the prognostic factor, it may be of interest to see how other smoothing methods could apply and/or varying the distribution of the frailty effect to see if these have any impact at all. While this was of interest, the scope of the paper was largely to ascertain modeling rare disease survival outcomes in small center settings.

ACKNOWLEDGEMENTS

Carlos Rodriguez-Galinda, M.D. of Eastern Cooperative Oncology Group; James Dziura of Yale University.

CONFLICT OF INTEREST STATEMENT

The authors have declared no conflict of interest.

REFERENCES

[1] Matthews JN. Small clinical trials: are they all bad? Statist Med 1995; 14(2): 115-126. http://dx.doi.org/10.1002/sim.4780140204

[2] Localio A, Berlin J, Ten Have T, Kimmel S. Adjustments for center in multicenter studies: An overview. Ann Inter Med 2007; 135(2): 112-123. http://dx.doi.org/10.7326/0003-4819-135-2-200107170-00012

[3] Ambler G, Seaman S, Omar RZ. An evaluation of penalised survival methods for developing prognostic models with rare events. Statist Med 2012; 31(11-12): 1150-1161. http://dx.doi.org/10.1002/sim.4371

[4] Senn S. Planning and analyzing multi-centre trials. Statist Med 1998; 17: 1753-1765. http://dx.doi.org/10.1002/(SICI)1097-0258(19980815/30)17:15/16<1753::AID-SIM977>3.0.CO;2-X

[5] Duchateau L, Janssen P, Lindsey P, Legrand C, Nguti R, Sylvester R. The shared frailty model and the power for heterogeneity tests in multicenter trials. Comput Statist Data Anal 2002; 40.

[6] Duchateau L, Janssen P. The frailty model. Vol New York 2008.

[7] Jeong J-H, Costantino J. Application of smoothing methods to evaluate treatment-prognostic factor interactions in breast cancer data. Cancer Investig 2006; 24: 288-293. http://dx.doi.org/10.1080/07357900600633841

[8] Silverman BG, Siegelmann-Danieli N, Braunstein R, Kokia ES. Trends in breast cancer incidence associated with reductions in the use of hormone replacement therapy. Cancer Epidemiol 2011; 35(1): 11-16.

[9] Govindarajulu US, Malloy EJ, Ganguli B, Spiegelman D, Eisen EA. The comparison of alternative smoothing methods for fitting non-linear exposure-response relationships with Cox models in a simulation study. Int J Biostatist 2009; 5(1): Article 2.

[10] Therneau T, Grambsch P. Modeling survival data: extending the Cox Model. New York: Springer-Verlag 2002.

[11] Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. Statist Med 2005; 24(11): 1713-1723. http://dx.doi.org/10.1002/sim.2059

[12] Malloy EJ, Spiegelman D, Eisen EA. Comparing measures of model selection for penalized splines in Cox models. Comput Stat Data Anal 2009; 53(7): 2605-2616. http://dx.doi.org/10.1016/j.csda.2008.12.008

[13] Klein J, Moeschberger M. Survival analysis: Techniques for censored and truncated data. New York: Springer 1997.

[14] Armitage P, Doll R, (Ed) IJN. Stochastic models for carcinogenesis. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability; Berkeley (Univ. of California) 1961.

[15] Breslow N, Day N. Statistical methods in cancer research: The design and analysis of cohort studies. Int Agency Res Cancer 1987; 2.

Received on 24-10-2014 Accepted on 12-01-2015 Published on 27-01-2015

http://dx.doi.org/10.6000/1929-6029.2015.04.01.2

© 2015 Govindarajulu and Malloy; Licensee Lifescience Global.
This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/3.0/) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.