SHARED-PARAMETER JOINT MODELS OF FATIGUE AND TIME UNTIL DEATH OF NON-RESECTABLE LUNG CANCER PATIENTS

by

David Phillips

A Thesis Submitted to the Faculty of the

MEL AND ENID ZUCKERMAN COLLEGE OF PUBLIC HEALTH

In Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

WITH A MAJOR IN BIOSTATISTICS

In the Graduate College

THE UNIVERSITY OF ARIZONA

2018

STATEMENT BY AUTHOR

The thesis titled *Shared-Parameter Joint Models of Fatigue and Time Until Death of Non-Resectable Lung Cancer Patients* prepared by *David Phillips* has been submitted in partial fulfillment of requirements for a master's degree at the University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

SIGNED: *David Phillips*

APPROVAL BY THESIS DIRECTOR

This thesis has been approved on the date shown below:

Melanie Bell
Professor of Biostatistics

April 25, 2018
Date

Table of Contents

List of Tables

Abstract:

With researchers showing greater interest in the relationship between longitudinal and survival outcomes, joint models are being used with greater frequency. Joint models of longitudinal and time to event outcomes offer distinct advantages. First, joint models can reduce bias in estimates of the relationship between surrogate markers and survival endpoints. Second, this class of model can provide sensitivity analysis of longitudinal estimates in the presence of potential missing data when the longitudinal outcome and survival outcome are related. The aim of this paper is to demonstrate the usefulness of this methodology when dealing with potentially related outcomes. Using a data set from a clinical trial aimed at reducing fatigue with physical activity amongst non-resectable lung cancer patients, several joint models and conventional models such as a linear mixed model and Cox proportional hazards model were generated. Both the longitudinal and survival estimates from these models were compared to demonstrate the utility of joint models. Furthermore, the implementation of joint models is discussed as a result of the analysis.

Introduction:

One critical interest of researchers is the relationship between longitudinal outcomes with survival outcomes. This is evident in the evaluation of potential surrogate markers of disease where researchers attempt to evaluate treatment effectiveness on a survival outcome through a biological mediator as in the case of clinical trials in early treatments of AIDS (Prentice, 1989). However, conventional survival analysis can be biased when evaluating such relationships. Conventional methods such as Cox proportional hazards treat surrogate markers as time-dependent covariates that are measured intermittently during clinical trials (Tsiatis et al., 1995). As a result, joint models that utilize both longitudinal and event time data have been developed. The aim of joint models is to better capture relationships between these outcomes resulting in better estimates of predictors.

As stated, joint models can provide unbiased estimates of the relationship between surrogate markers of survival endpoints. Surrogate markers as defined by Prentice, must capture the effects of treatment through the marker, have different levels of the markers by treatment group, and must be related to the outcomes (Prentice, 1989). To evaluate such surrogate markers, researchers turn to extended Cox regression using recorded values of the surrogate marker as a time-dependent covariate (Andersen & Gill, 1982). While such treatment of the covariates would be appropriate for exogenously related variables, for endogenously related outcomes bias has been demonstrated to occur (Sweeting & Thompson, 2011). Since clinical trials usually measure covariates intermittently at follow-up times, endogenously related markers will only take on the value the last recorded measurement when treated as time-dependent covariates in proportional hazards models. This makes the assumption that marker trajectories follow a step function only changing value at follow-up times. This creates last value carried

forward bias in between follow-up times because of the assumption that the value of the marker is static between time points. This assumption may deviate from the true trajectory of the marker, which may be considerably dynamic.

Another utilization of joint models is for sensitivity analysis in the presence of potentially informative dropout. Many clinical studies will encounter missing values for longitudinal studies. Depending upon the mechanism of missingness, some methods such as linear mixed models when well-specified, will be able to provide unbiased estimates for missing completely at random (MCAR) and missing at random data(MAR) (Bell et al., 2014). However, the presence of missing not at random (MNAR) data will induce bias using linear mixed models (Wu & Carroll, 1988). For example, when sicker participants drop out of a clinical trial, quality of life (QOL) can be overestimated. As a result, sensitivity analysis is recommended to explore the robustness of the primary analysis when MNAR data is suspected. Methods that are sensitive to MNAR data include selection models, pattern mixture models and shared parameter models. Joint models, as a class of shared-parameter models, can address MNAR by using random effects to capture the relationships between measurement and missingness processes(Rizopoulos, 2012b).

Background:

The origin of joint models can be traced to concerns regarding estimation of effects of continuous covariates on survival outcomes. Prentice and Nakamura both examined the relationship of erroneous measurements on radiation dose levels from atomic bomb survivors on survival outcomes. The authors demonstrated measurement errors in dose levels resulted in bias on hazard ratio estimates. Attempts to mitigate the described bias were made by utilizing corrections to the partial likelihood formulation or score function (Nakamura, 1992; Prentice,

1982). Measurement error in the evaluation of surrogate markers can lead to misidentification of potential surrogate markers due to this bias. As a result, the development of joint models was an attempt to alleviate bias due to measurement error of longitudinal covariates (Tsiatis et. al., 1995).

In clinical trials, the history of the longitudinal marker trajectory is usually unknown except only at recorded time points. Such markers can be treated as time-dependent covariates in Cox regression where the last recorded value is carried forward for purposes of estimation. The proportional hazards model with a time-dependent longitudinal marker can be shown for the ith subject as

$$h_i(t) = h_0(t)\exp(\theta^T W_i + \phi y_i(t)), \qquad (1)$$

where $h_0(t)$ is the unspecified baseline hazard, W_i is the vector for baseline covariates and θ^T is the vector of coefficients of the baseline covariates. ϕ is the coefficient of effect of the recorded value of the time-dependent longitudinal marker $y_i(t)$ at time t for the ith subject. As a time-dependent covariate, the model assumes that $y_i(t)$ does not change between follow up times. The result could be substantially biased due to interpolation between recorded observations of $y_i(t)$. During the clinical trials of AIDS patients in the 1990s, several groups attempted to examine the relationship between CD4 cell count and various HIV/AIDS outcomes. Tsiatis et al. developed what is described as the two-stage model in an attempt to address measurement error and interpolation of CD4 cell count (Tsiatis et al., 1995). In the model, the CD4 trajectory was estimated using linear mixed effects models. In the more general case for the ith subject, let

$$m_i(\mathrm{t}) = \beta \mathrm{X}_i(\mathrm{t}) + b_i \mathrm{Z}_i(\mathrm{t}), \qquad (2)$$

where $m_i(\mathrm{t})$ is the predicted trajectory of the longitudinal marker at time t, with $\mathrm{X}_i(\mathrm{t})$ and $\mathrm{Z}_i(\mathrm{t})$ as vectors for fixed effects β and random effects $\mathrm{b_i}$, respectively. Each subject's best linear

unbiased predictor is then estimated at each event time in a proportional hazards model resulting in less biased hazard ratio estimates when compared to the time-dependent Cox model. The hazard for the ith subject of this two-stage model approach can be shown as

$$h_i(t) = h_0(t)\exp(\theta^T W_i + \alpha m_i(t)), \qquad (3)$$

where W_i is the vector for baseline covariates, θ^T is the vector of coefficients of the baseline covariates and α is the effect of the predicted value of the longitudinal marker on the hazard. In this case, the baseline hazard $h_0(t)$ is left unspecified.

Another approach to providing less biased longitudinal estimates is by using survival data of participants when MNAR data is suspected. In the previously mentioned AIDS trials, the presence of informative dropout could also bias estimates of longitudinal models of CD4 cell counts (Wu & Carroll, 1988). As a result, joint models utilizing both survival data and longitudinal data were developed. One approach of joint models outlined by Self and Pawitan was by conditioning the distribution of the longitudinal outcome on the survival data (Self & Pawitan, 1993). The authors utilized a likelihood formulation to jointly estimate T4 cell trajectory, time to infection from HIV, and time to AIDS. The likelihood is shown as

$$L(\theta|T_{HIV}, T_{AIDS}) = \mathrm{p}_{HIV}(T_{HIV}|\mathrm{X}_{HIV}, \theta_{HIV})\mathrm{p}_{AIDS}(T_{AIDS}|\mathrm{X}_{AIDS}, \theta_{AIDS})\mathrm{p}_y\big(y(t)\big|t_{HIV}, t_{AIDS}, \theta_y\big)$$

where $\mathrm{p}_{HIV}(\cdot), \mathrm{p}_{AIDS}(\cdot)$ and $\mathrm{p}_y(\cdot)$ are the probability density functions for HIV infection time, time to AIDS and the longitudinal marker, respectively. T_{HIV}, T_{AIDS}, and $y(t)$ are HIV infection time, time to AIDS and the longitudinal marker, respectively. Let $\mathrm{X} = (\mathrm{X}_{HIV}, \mathrm{X}_{AIDS})$ and be the covariate vector for HIV infection time and time to AIDS while $\theta = (\theta_{HIV}, \theta_{AIDS}, \theta_y)$ is the unknown parameter vector for each process. Using this joint formulation, the authors utilized survival data to provide less biased estimates on the longitudinal process than what would be achieved using a linear mixed model.

An alternative approach to joint models is the shared random effects approach. In these models, the joint probability distribution for survival and longitudinal outcomes can be shown for the ith subject as

$$\mathrm{p}(\mathrm{T}_i, \delta_i, \mathrm{y}_i(\mathrm{t})|b_i; \theta) = \mathrm{p}(\mathrm{T}_i, \delta_i|b_i; \theta)p(y_i(\mathrm{t})|b_i; \theta), \quad (4)$$

where $\mathrm{T_i}$ is the event time, δ_i is the event indicator, and $\mathrm{y}_i(\mathrm{t})$ is the recorded longitudinal marker at time t. By conditioning on the subject-specific random effects b_i, the distributions of survival and longitudinal outcomes are independent. Joint models can provide estimates for longitudinal outcomes in the presence of informative dropout through maximization of the joint likelihood of survival outcomes and longitudinal outcomes (Wulfsohn & Tsiatis, 1997). In this approach, when longitudinal outcomes are endogenously related to survival outcomes, the use of survival data can account for bias from MNAR data. Unlike the method described by Self and Pawitan, this joint model can provide estimates for the relationship between the longitudinal marker and the survival outcome, thereby allowing for screening of potential surrogate markers.

Alternatively, other approaches have been developed in the joint modeling framework. Henderson et al. utilized a joint model where the relationship between failure time and longitudinal markers are captured via a latent Gaussian process. When conditioned on the latent Gaussian process, both outcomes become independent (Henderson et al., 2000). Another approach of joint models outlined by Lin et. al. specifies K latent classes with each class having a unique pattern of event time and longitudinal responses. Joint models described by Wulfsohn and Tsiatis specify a single trajectory for an entire population. Latent class joint models can capture unique trajectories for possible heterogeneous subgroups (Lin et. al., 2002).

Study Data:

To demonstrate such joint modeling methods, a dataset from a randomized controlled trial of advanced small cell and non-small cell lung cancer patients in Australia was utilized (Dhillon et al., 2012). Patients for the study were recruited from 4 hospitals in Sydney, Australia. Major inclusion criteria for participants were as follows:

- Diagnosis with non-resectable lung cancer (small cell and non-small cell).
- Completion of a course of chemo-radiotherapy in the 4 weeks prior to randomization.
- Medically fit to participate in a physical activity program, as assessed by the Physical Activity Readiness Questionnaire.

Major exclusion criteria were as follows:

- Eastern Cooperative Oncology Group performance status greater than 3.
- life expectancy less than 6 months.
- insufficient English to complete questionnaires.

Patients were randomized into an intervention consisting of 2 months of a physical activity program with general health education or into the control group of general health education. The main outcome of the study was fatigue, which was measured by the Functional Assessment of Cancer Therapy: Fatigue (FACT-F). FACT-F is a self-reported survey consisting of 41 items measured on a Likert scale of 0 ("not at all") to 4 ("very much so"). The FACT-F consists of the Functional Assessment of Cancer Therapy: General (FACT-G) with 13 fatigue related items (Cella & Nowinski, 2002). Patients completed FACT-F questionnaires at 0, 2, 4, and 6 months. An exploratory outcome was time until death from lung cancer, however, the study had limited power for this outcome. Due to substantial missing data and a potential relationship between

time until death and fatigue, statistical analysis using joint models could be beneficial for reasons outlined earlier.

Objectives of Thesis:

The objective of this thesis is to demonstrate the usefulness of joint models on the described data set by comparing estimates generated by multiple methods used to analyze survival and longitudinal outcomes. Namely, the data set will be analyzed by using conventional methods such as Cox proportional hazards and linear mixed models, the two-stage model proposed by Tsiatis et al., and several formulations of joint models (Tsiatis et al., 1995; Wulfsohn & Tsiatis, 1997). By comparing estimates generated by the several methods, the relationship of fatigue and death in advanced lung cancer patients can be evaluated and the potential for a MNAR mechanism can be investigated. The rest of the thesis is organized as section 2 detailing methods of evaluating the clinical trial data, section 3 examines the results of the each model, and section 4 discusses the implications of the results on the primary and secondary objectives of the study.

Methods:

In order to model the relationship between the longitudinal outcome, fatigue and the survival outcome, time until death from lung cancer, several methods were executed and the results were compared by examining multiple estimates of parameters of both the longitudinal models for FACT-F and survival models of time to death. Namely, conventional methods such as the Cox model and linear mixed models, the two-stage model, and 3 different joint models were generated. The formulation of the joint models as well as the joint likelihood for these joint models will be outlined. In addition, estimation of the joint likelihood and its complex integrals

will be detailed briefly. Distinctive issues regarding baseline hazard function of joint models and diagnostics for joint models will also be discussed.

Conventional/Standard Models:

Method 1 use the standard methods of Cox Proportional Hazards for the survival outcome and a linear mixed model for FACT-F (Laird & Ware, 1982). In method 1, the recorded fatigue scores were utilized as an external time-dependent covariate in the extended Cox model (Andersen & Gill, 1982). The survival model is a proportional hazards model shown for subject i as

$$h_i(t) = h_0(t)\exp(\theta a_i + \alpha y_i(t)) \quad (5)$$

where $h_i(t)$ is the subject-specific hazard for subject i with unspecified baseline hazard $h_0(t)$. The intervention group indicator and corresponding coefficient is a_i and θ, respectively. The time-dependent variable $y_i(t)$ is the recorded FACT-F score for subject i at time t with α as the effect of recorded FACT-F. The longitudinal model for FACT-F for the ith subject at the jth time will be

$$y_{ij}(t) = \beta_0 + \beta_1 t_{ij} + \beta_2(a_i * t_{ij}) + b_{0i} + b_{1i} t_{ij} + \varepsilon_{ij} \quad (6)$$

where random effects are $\boldsymbol{b} = (b_0, b_1) \sim N(0, \boldsymbol{\Sigma})$ with covariance matrix $\boldsymbol{\Sigma}$, the error term is $\varepsilon_{ij} \sim N(0, \sigma^2)$, and the intervention variable is a_i coded as 1 for the intervention group and 0 for the group receiving standard treatment. The intervention group and control group were assumed to have no difference in fatigue at the start of the study. Thus, a_i only contributes to the model as an interaction with time. The coefficients for fixed effects are β_0 for population intercept, β_1 for the effect of time on fatigue, and β_2 for the interaction between treatment and time. In order to execute these models, SAS 9.4 procedures **PROC PHREG** and **PROC MIXED** were used to generate estimates for the Cox model and linear mixed model, respectively.

Two-Stage Model:

Method 2, known as the two-stage model, utilizes the results of the linear mixed model with the Best Linear Unbiased Predictors (BLUPs) to provide a predicted fatigue score for subject i at each event time t. The longitudinal model for the two-stage model will have the same estimates as the conventional linear mixed effects model from method 1. The predicted fatigue score $m_i(t)$ for subject i at time t was used as a time-dependent covariate in the Cox model (Tsiatis et al., 1995). The hazard for the ith subject of the two-stage model is defined as

$$h_i(t) = h_0(t)\exp(\theta a_i + \alpha m_i(t)) \quad (7)$$

where $h_i(t)$ is the subject-specific hazard with unspecified baseline hazard $h_0(t)$ and intervention group indicator a_i. θ and α are the effect of intervention and effect of the predicted FACT-F score for each subject i at time t, respectively. The two-stage model was expected to deliver less biased results in terms of survival estimates while offering no advantages in longitudinal estimates compared to conventional linear mixed models. BLUPs were used to construct the subject-specific trajectory $m_i(t)$ that was used as a time-dependent covariate in **PROC PHREG**.

Parametric Joint Models:

Method 3 utilized joint modeling of longitudinal and survival outcomes using maximum likelihood estimation with shared random effects (Wulfsohn & Tsiatis, 1997). Method 3A-3C included several different parameterizations of joint models to illustrate the diversity of this method. The survival sub-models of methods 3A-3C are defined as a parametric proportional hazards model. The longitudinal sub-model of methods 3A-3C is a linear mixed model with estimated random slope and random intercept. Method 3A is defined as a trajectory model where the predicted value of FACT-F for subject i, $m_i(\mathrm{t})$, is estimated at each event time t

(Wulfsohn & Tsiatis, 1997). The advantage of the trajectory model is the straight-forward interpretation of the estimates. Namely, researchers can directly estimate the magnitude of the relationship of a longitudinal outcome on the survival outcome. The survival sub-model for the trajectory model for subject i is

$$h_i(t) = h_0(t)\exp(\theta a_i + \alpha m_i(\mathrm{t})) \quad (8)$$

where $h_i(t)$ is the subject-specific hazard with parametric baseline hazard $h_0(t)$ and intervention group indicator a_i. θ and α are the effect of intervention and effect of the trajectory of FACT-F score for each subject i at time t, respectively. It is important to note that equations 7 and 8, while similar in notation, will have estimates generated from different likelihood formulations. Equation 7 will have estimates from the partial likelihood while equation 8 will have estimates from the joint likelihood of the survival and longitudinal processes. Method 3B is defined as a shared random effects model where the values of the random effects of the longitudinal sub-model are used as covariates in the survival model (Henderson, Diggle, & Dobson, 2000). The proportional hazards for subject i is

$$h_i(t) = h_0(t)\exp(\theta a_i + \alpha_1 b_{0i} + \alpha_2 \mathrm{b}_{1i}) \quad (9)$$

where $h_i(t)$ is the subject-specific hazard with parametric baseline hazard $h_0(t)$ and intervention group indicator a_i with coefficient θ. α_1 is the effect of subject-specific random intercept b_{0i} and α_2 is the effect of subject-specific random slope b_{1i}. Method 3C is similar to 3A with the exception that the value of rate of change of mean fatigue, $m_i'(\mathrm{t})$, for subject i at time t was used as a covariate in the survival sub-model (Rizopoulos, 2010). The proportional hazards for subject i is

$$h_i(t) = h_0(t)\exp(\theta a_i + \alpha m_i'(\mathrm{t})) \quad (10)$$

where $h_i(t)$ is the subject-specific hazard with parametric baseline hazard $h_0(t)$ and intervention group indicator a_i with coefficient θ. α is the effect of the value of the rate of change of mean fatigue $m_i'(\mathrm{t})$. In order to generate estimates for these three joint models, SAS Macro **JM Macro v2.01** by Garcia-Hernandez and R package **JM 1.4-5** by Rizopoulos were used(Garcia-Hernandez & Rizopoulos, 2015; Rizopoulos, 2010).

Joint Likelihood Formulation:

The log of the joint likelihood, p, for the ith subject can be formalized as

$$\log p(T_i, \delta_i, y_i(t_{ij})|b_i; \phi) = \log \mathrm{p}(\mathrm{T}_i, \delta_i|b_i;\ \phi_S)p(y_i(t_{ij})|b_i; \phi_y)p(b_i; \phi_b) \quad (12)$$

where $T_i, \delta_i, and\ y_i(t_{ij})$ are time until event, death indicator and fatigue score at time t_{ij} where j is the index of recorded FACT-F scores, $y_i(t_{ij})$ for each subject i, respectively. $\boldsymbol{\phi}$ = (ϕ_S, ϕ_y, ϕ_b) is the vector of parameters for the survival models, longitudinal fixed effects and longitudinal random effects, respectively (Rizopoulos, 2012b). For the ith subject, the likelihood of the survival process of the trajectory model is

$$\mathrm{p}\big(\mathrm{T}_i, \delta_i\big|b_i; \phi_S\big) = \ [h_0(\mathrm{T}_i)\exp(\theta a_i + \alpha m_i(T_i)]^{\,\delta_i} \times \exp[-\textstyle\int_0^{T_i} h_0(\mathrm{s})\exp\{\theta a_i + \alpha m_i(s)\}\,ds] \quad (13)$$

The likelihood for the longitudinal outcome for the ith subject is specified as

$$p\big(y_i\big(t_{ij}\big)\big|b_i; \phi_y\big)p(b_i; \phi_b) =$$
$$(2\pi\sigma^2)^{-k_i/2}\exp\{-\tfrac{\|y_i-\beta \mathrm{X}_i-\mathrm{b}_i\mathrm{Z}_i\|^2}{2\sigma^2}\} \times (2\pi)^{-\frac{q_b}{2}}\det(\boldsymbol{\Sigma})^{-\frac{1}{2}}\exp(-\tfrac{{b_i}'\boldsymbol{\Sigma}^{-1}b_i}{2}) \quad (14)$$

where X_i and Z_i are vectors for fixed and random effects, respectively. $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the random effects b_i and q_b is the dimensionality of the random effects vector.

Numerical Integration:

To evaluate likelihood estimates and standard errors for the joint models, numerical integration must be applied since the integrals with respect to time in the survival function of the

joint likelihood and the score vector with respect to the random effects do not have analytical solutions. It has been established that researchers can use Gaussian quadrature methods to generate solutions to complex survival functions in joint likelihood formulation. One such method that has been used is the 7-point or 15-point Gauss-Konrod rule. The computational demands of evaluating the score vector increases with each additional random effect. In order to evaluate the integrals with respect to the random effects, standard Gaussian-Hermite quadrature can be used to approximate such integrals(Press et al., 2007). Additionally, adaptive Gaussian-Hermite quadrature has been used to further reduce computational demands as well as pseudo – adaptive Gaussian-Hermite (Rizopoulos, 2012a). The **JM** package in R and **JM-Macro v2.01** in SAS use pseudo-adaptive Gaussian-Hermite and adaptive Gaussian-Hermite, respectively (Garcia-Hernandez & Rizopoulos, 2015; Rizopoulos, 2010).

Baseline Hazards Selection:

For methods 3A-3C, a baseline hazard function was selected after evaluation and comparison using AIC, BIC, and likelihood ratio tests when models are nested. Unlike semi-parametric methods, utilizing a nonparametric baseline for the joint models would result in an excessive number of parameters to be estimated since the full likelihood approach must account for the numerous step functions that occur at each event time with a nonparametric baseline. Hsieh et. al. demonstrated profile likelihood approaches to using unspecified baseline in joint models but the result is that standard errors are usually underestimated (Hsieh et al., 2006). Although parametric joint models lose the flexibility of an unspecified baseline, flexible approaches such as Weibull splines or piecewise exponential baselines can be used. However, in order to prevent overfitting, the Harrell rule of thumb of 10 to 20 events per estimated parameter

was considered when determining the adequate number of knots for the piecewise exponential or Weibull spline baseline hazard for the survival sub-models (Harrell, 2001).

Unique Diagnostics for Parametric Joint Models:

Goodness of fit of the survival sub-model was formally tested using the method described by Park et. al. (Park & Qiu, 2014). Using this method, subjects were ranked by baseline values of the longitudinal outcome, fatigue, and grouped according to ranking from low to high. Indicator variables were created for each of the ranked groups and added to the joint models. The null model and model with ranked group indicator variables were tested for difference using the likelihood ratio test at the 5% significance level. In addition, Cox-Snell plots, Martingale residuals and deviance residuals are to be inspected as well for model fit and outliers. The validity of traditional residual plots for the longitudinal model are suspect with the possible presence of nonrandom dropout. Within the joint modeling framework, the subject-specific profile of the longitudinal outcome is assumed to be related to the survival outcome. Thus, the occurrence of events is assumed to be linked with nonrandom dropout of the longitudinal process. Traditional residual plots under this scenario will not have expected characteristics such as zero mean or independence. As a result, multiply-imputed standardized residuals were used to inspect the longitudinal model specification but were not used for sensitivity analysis of model inferences (Rizopoulos et al., 2010). The process of generating multiply-imputed residuals as described by Rizopoulos et. al., utilizes the posterior distribution of the unobserved longitudinal response, y_i^m. The distribution for subject i is

$$p(y_i^m | y_i^o, T_i, \delta_i) = \int p(y_i^m | y_i^o, T_i, \delta_i; \phi) p(\phi | y_i^o, T_i, \delta_i) d\phi \quad (15)$$

where y_i^o is the observed longitudinal responses, T_i is the event time, δ_i is the event indicator, and ϕ is the vector of parameters. Using the conditional independence of the survival outcome and longitudinal outcome of the joint likelihood, this distribution can be further simplified to

$$p(y_i^m|y_i^o, T_i, \delta_i) = \int p(y_i^m|b_i; \phi)p(b_i|y_i^o, T_i, \delta_i; \phi)db_i \text{ (16).}$$

The authors base the simulation of the lth imputed residual for subject i on the following scheme:

1. Draw $\phi^{(l)}$ from N$\{\hat{\phi}, \widehat{var}(\hat{\phi})\}$ where $\hat{\phi}$ is the maximum likelihood estimates with covariance matrix, $\widehat{var}(\hat{\phi})$.

2. Draw $b_i^{(l)} \sim \{b_i|y_i^o, T_i, \delta_i, \phi^{(l)}\}$.

3. Draw $y_i^{m(l)}(t_{ij}) \sim N\left\{\hat{m}_i^{(l)}(t_{ij}), \hat{\sigma}^{2,(l)}\right\}, where\ \hat{m}_i^{(l)}(t_{ij})$ is the predicted longitudinal response at t_{ij} for $t_{ij} \geq T_i$, j = 1, …, n_i' for unobserved follow up times.

4. Repeat steps 1-3 for each subject for L number of imputations with l = 1, …, L.

To draw from the nonstandard distribution of step 2, the authors use a Metropolis-Hastings algorithm with proposal distribution, $t_4\{\hat{b}_i, var(\hat{b}_i)\}$ where $\hat{b}_i$ are the empirical Bayes estimates for subject i. 50 imputations for each subject were generated to create multiply-imputed residual plots using the **JM** package.

Sensitivity Analysis:

A concern regarding the validity of the estimates generated by joint models with the data set was due to the significant time gap between recorded follow-up times and time until event for most participants. This time gap was suspected to generate possible bias due to gross extrapolation. Thus, additional sensitivity analyses were performed by analyzing the data set with a censoring time at 365 days. The results of the sensitivity analysis were compared to the estimates generated from the original, unaltered data set.

Results:

Baseline Characteristics:

Overall, there were 55 and 56 participants randomized to the control and intervention groups, respectively. Characteristics between the two arms were well-balanced after randomization according to table 1. It should be noted that there was considerable missingness in FACT-F over the 6 months of follow up. For the control group, missingness ranged from 3.6% to 54.5% while the treatment group ranged from 0% to 39.3%. Furthermore, only 19 participants from both arms experienced death within the 6 months of follow up while overall the study experienced 83 deaths between the two groups. Median survival times, according to the Kaplan-Meier estimator, showed similar results of 417 days and 485 days for the control and intervention groups, respectively.

Table 1: Baseline Characteristics

	Treatment Group	
Variable	**Control**	**Intervention**
Age (years) Mean (SD)	61.2 (10.2)	62.8 (9.7)
Months since Diagnosis Median (Min, Max)	7.7 (0.7, 190.5)	8.6 (1.6, 80.5)
Resting HR Mean (SD)	85.5 (14.5)	82.8 (13.8)
Simplified Comorbidity Score (Colinet) Median (Min, Max)	9 (1, 19)	8 (1, 18)
Cycles of Completed Chemotherapy Median (Min, Max)	2.5 (1, 17)	3 (1, 6)
ECOG N(%)		
0	32 (58.2 %)	29 (51.8 %)
1	21 (38.2 %)	25 (44.6 %)
2	2 (3.6 %)	1.8 (3.6 %)
Smoking History N(%)		
No	19 (34.6 %)	23 (41.1)

Yes	36 (65.4 %)		33 (58.9 %)	
Advanced Disease N(%)				
No	2 (3.6 %)		3 (5.4 %)	
Yes	53 (98.4 %)		53 (94.6 %)	
Small-Cell Lung Cancer N(%)				
No	3 (5.4 %)		2 (3.6 %)	
Yes	52 (94.6 %)		54 (96.4 %)	
Currently on Chemotherapy N(%)				
No	28 (50.9 %)		27 (48.2 %)	
Yes	27 (49.1 %)		29 (51.8 %)	
Status N(%)				
Censored	13 (23.6 %)		15 (26.8 %)	
Death	42 (76.4 %)		41 (73.2 %)	
Survival Time* Median (95% CI), (Min, Max)	417 (350, 582)	(26, 1799)	485 (355, 734)	(55, 2018)
Fatigue Mean (SD), N Missing (%)				
0 Months	36.4 (12)	2 (3.6 %)	38.4 (11)	0 (0 %)
2 Months	37.3 (10.7)	14 (25.5 %)	38 (11.8)	10 (17.9 %)
4 Months	37.3 (9.4)	22 (40 %)	41.1 (9.3)	18 (32.1 %)
6 Months	36.3 (11.6)	30 (54.5 %)	39.3 (11.3)	22 (39.3 %)

*= Kaplan-Meier Estimator

Longitudinal Estimates:

The estimates generated by the several models fitted varied in magnitude but not directionality for the longitudinal parameters. From table 2, the linear mixed model, which does not use survival data to generate estimates, did not show significant evidence for a relation between fatigue and time or the interaction with treatment. For the joint models, there was significant evidence for a relationship between time and fatigue with values ranging from -1.26 to -1.43. This relationship between time and fatigue suggests that participants generally experienced more fatigue as time progressed. The interaction between time and the intervention for most models did not present significant evidence of differing patterns of change over time with the exception of the trajectory model generated from R. It should be noted that both the

macro used in SAS and package utilized in R were programmed by different authors using different algorithms. The magnitude of the interaction between intervention and time ranged from 0.4 to 0.81 depending upon the linkage of the longitudinal sub-model and survival sub-model.

Table 2: Longitudinal Estimates and 95% CI with Complete Survival Data

Method	1 and 2	3A	3A	3B	3C
Model	LMM (SAS)	TJ (SAS)	TJ(R)	SRE (SAS)	RC (R)
Intercept	37.53	37.79	38.02	37.59	37.72
	(35.44, 39.61)	(35.74, 39.84)	(36, 40.03)	(35.52, 39.65)	(35.8, 39.64)
Time (by 60 days)	-0.93	-1.26	-1.42	-1.33	-1.43
	(-2.16, 0.3)	(-2.26, -0.26)	(-2.16, -0.68)	(-2.56, -0.09)	(-2.67, -0.18)
Interaction of Time and Intervention (by 60 days)	0.79	0.51	0.81	0.68	0.40
	(-0.76, 2.34)	(-0.68, 1.71)	(0.08, 1.54)	(-0.9, 2.25)	(-1.28, 2.08)

LMM = Linear Mixed Model, TJ = Trajectory, SRE = Shared Random Effects, RC = Rate of Change

The longitudinal estimates of the joint models fitted using survival data that were censored at 365 days were generated in table 3. Using this abbreviated data set, estimates experienced shifts in magnitude compared to those in table 2. For instance, the trajectory model generated in R from table 3 reported 0.73 (-0.82, 2.28) for the estimate of the interaction while the same model in table 2 reported 0.81 (0.08, 1.54). In addition, the trajectory model (SAS), shared random effects model, and rate of change model demonstrated shifts of increased magnitude when compared to the estimates from the unabbreviated survival data set.

Table 3: Longitudinal Estimates and 95% CI with Censored Survival Data at 365 Days

Model	3A	3A	3B	3C
Method	TJ (SAS)	TJ(R)	SRE (SAS)	RC (R)
Intercept	37.77 (35.70, 39.85)	37.78 (35.71, 39.85)	37.58 (35.51, 39.66)	37.68 (35.74, 39.61)
Time (by 60 days)	-1.41 (-2.61, -0.21)	-1.41 (-2.61, -0.21)	-1.34 (-2.60, -0.08)	-1.34 (-2.61, -0.07)
Interaction of Time and Intervention (by 60 days)	0.73 (-0.82, 2.28)	0.73 (-0.82, 2.28)	0.72 (-0.87, 2.31)	0.52 (-1.12, 2.17)

TJ = Trajectory, SRE = Shared Random Effects, RC = Rate of Change

Survival Estimates:

The survival estimates in table 4 show that none of the survival sub-models found significant evidence for an effect of treatment on participant survival. In fact, all but the Cox proportional hazards model and rate of change model estimate a deleterious effect on survival due to intervention compared to the control. While the Cox proportional hazards model may have shown a non-significant effect of the intervention of 0.84 (0.46, 1.55), this model does not account for the effect of intervention on fatigue as found in the joint models and two-stage model. There was significant evidence in support of an effect of the trajectory of participant fatigue on survival by the two trajectory models, the two-stage model, and the proportional hazards model. The joint models reported the effect of a 1 unit increase of FACT-F with hazard ratios of 0.92 (0.89, 0.96) and 0.9 (0.63, 0.94) for **SAS** and **R**, respectively. This indicates that those with higher predicted values of FACT-F had a protective effect in relation to time until death. The proportional hazards model and two-stage model reported a weaker effect of FACT-F at 0.95 (0.93, 0.98) and 0.96 (0.94, 0.98), respectively. The shared random effects model confirms that those individuals with higher random intercepts and slopes will have more

favorable survival outcomes with hazard ratios of 0.93 (0.90, 0.97) and 0.84 (0.73, 0.98), respectively. This translates into those participants that experienced less fatigue as reported by higher initial FACT-F measurements or smaller decreases in FACT-F over time had better outcomes. The rate of change model also demonstrates significant evidence for a relationship between the rate of change of fatigue and time until death.

Table 4: Survival Estimates of Hazard Ratios and 95% CI with Complete Survival Data

Model	1	2	3A	3A	3B	3C
Method	PH (SAS)	TS (SAS)	TJ (SAS)	TJ (R)	SRE (SAS)	RC (R)
Intervention vs Control (REF)	0.84 (0.46, 1.55)	1.12 (0.71, 1.78)	1.36 (0.73, 2.53)	1.49 (0.77, 2.88)	1.02 (0.60, 1.74)	1 (0.55, 1.81)
Recorded Fatigue or Trajectory of Fatigue	0.95 (0.93, 0.98)	0.96 (0.94, 0.98)	0.92 (0.89, 0.96)	0.9 (0.63, 0.94)		
Rate of Change in Fatigue (per .01 unit)						0.72 (0.62, 0.82)
Random Intercept					0.93 (0.90, 0.97)	
Random Slope (per .01 unit)					0.84 (0.73, 0.98)	

PH = Proportional Hazards, TS = Two-Stage , TJ = Trajectory, SRE = Shared Random Effects, RC = Rate of Change

The results in table 5 show changes in the survival estimates according to the censored data set at 365 days. Under this abbreviated data set, there are 43 individuals that experienced the event. Like table 4, none of the models using the 365 day-censored data set demonstrated significant evidence for the effect of intervention vs the control. The proportional hazards model showed a change in estimate from a hazard ratio of 0.84 (0.46, 1.55) according to table 4 to 0.96

(0.38, 2.44) with the 365 day-censored data set. The estimates of the two-stage model using the abbreviated data set show that the effect of the trajectory of fatigue reduced towards the null with a ratio of 0.99 (0.97, 1.00) from the estimate of 0.96 (0.94, 0.98) yielded from the complete data set estimates. The two trajectory models estimated from **SAS** and **R** yielded similar results of the hazard ratios of intervention vs control of 1.33 and 1.34, respectively. In addition, the effect of the trajectory of fatigue was identical for both the trajectory models at 0.91 (0.88, 0.95). The rate of change model shows a shift in magnitude of the effect of intervention to 1.26 (0.5, 3.15) from 1 (0.55, 1.81).

Table 5: Survival Estimates of Hazard Ratios and 95% CI with Censored Survival Data at 365 Days

Model	1	2	3A	3A	3B	3C
Method	PH (SAS)	TS (SAS)	TJ (SAS)	TJ (R)	SRE (SAS)	RC (R)
Intervention vs Control (REF)	0.96 (0.38, 2.44)	1.11 (0.60, 2.04)	1.33 (0.67, 2.65)	1.34 (0.67, 2.66)	1.07 (0.54, 2.14)	1.26 (0.50, 3.15)
Recorded Fatigue or Trajectory of Fatigue	0.93 (0.90, 0.96)	0.99 (0.97, 1.00)	0.91 (0.88, 0.95)	0.91 (0.88, 0.95)		
Rate of Change in Fatigue (per .01 unit)						0.75 (0.63, 0.90)
Random Intercept					0.92 (0.88, 0.96)	
Random Slope (per .01 unit)					0.83 (0.71, 0.97)	

PH = Proportional Hazards, TS = Two-Stage , TJ = Trajectory, SRE = Shared Random Effects, RC = Rate of Change

Discussion:

In this demonstration of joint modeling of fatigue and time until death of lung cancer patients, several models were fitted to show contrasts in estimates between joint models and conventional methods. For the conventional methods, a linear mixed model was fitted for FACT-

F scores and a Cox proportional hazards model was fitted for time until death to show what potentially biased methods researchers may use to assess this data. The early iteration of the joint model, known as the two-stage model, was fitted to demonstrate what researchers could use if joint model software is unavailable. Lastly, three joint models using three different assumptions of the linkage between the two outcomes were fitted to show how joint models could alleviate potential bias. One of which was the trajectory model where subject-specific predicted fatigue is used to link the predicted trajectory of fatigue for each subject with a parametric proportional hazards sub-model. The shared random effects joint model linked both sub-models by way of the individual random intercepts and random slopes. The third joint model fitted was the rate of change model, where subject-specific rate of change of fatigue is used as a covariate within the proportional hazards sub-model. To explore the validity of the estimates of the initial analysis, the several models were reevaluated using a data set where subjects were censored at 365 days resulting in roughly half the number of events observed compared to the original data set. The estimates of this second analysis were then compared with the intial analysis.

Major Findings:

The longitudinal estimates generated by the different models may demonstrate the possible presence of a missing not at random mechanism. As noted earlier, the linear mixed model will account for missing at random and missing completely at random data when properly formulated. With the joint models demonstrating evidence for an association between fatigue and time until death, the differences between the longitudinal estimates of the linear mixed model and joint models raise doubts regarding the validity of the linear mixed model for this study. Even with the varying formulations of the joint models, the survival sub-models

demonstrated significant evidence for the linkage between fatigue related outcomes and time until death from lung cancer.

Possible Concerns with Study:

The execution and design of the clinical trial of the PAL data introduced possible concerns for joint models. The most obvious is that the study did not record any FACT-F readings beyond 180 days from randomization. The result of the lack of information means that all models executed were extrapolating fatigue beyond 180 days from a limited data set. It must be noted that additional follow ups beyond 180 days may not be feasible due to the deteriorating condition of cancer patients. Consequentially, the estimates produced have to be interpreted with caution even with the data set censored at 365 days. As noted in the introduction, the study was not powered for the survival outcome and thus, the effect of intervention upon death from lung cancer is inconclusive. Furthermore, the application of the intervention, namely, how patients received physical activity training could be contributing factors to the results of the analysis. One, the physical activity instruction amongst those randomized into the intervention group was not homogeneous. For example, some participants were part of group classes while others received instruction individually. This occurrence leaves open the question whether some groups of participants may have different responses to intervention respective to what type of instruction they received. The second contributing factor was that the physical activity instruction only proceeded for 2 months from the start of the study. Another factor that must be considered is the possibility of selection bias of during enrollment. It is plausible that those potential participants that were healthier and better able to tolerate exercise were more likely to enroll in the trial.

Concerns of Joint Models:

Since software packages such as those utilized in this study often have single authors, the packages are not fully validated such as standard software modules or packages found in **R** or **SAS**. As such, the packages utilized in this study have been recommended to be used for academic purposes by their respective authors. Furthermore, as shown in the results, different software packages can produce different estimates of the same model. Also, since individuals in this study survived considerably longer than the last follow up at 180 days, there is the possibility of bias for the estimation of effect of the different linkages for the joint models presented here. Lastly, both of the packages used in this analysis make the assumption of that the random effects are multivariate normally distributed. Such an assumption can be a concern for joint models since the random effects are used to capture the association between the longitudinal and survival outcome. Misspecification of the random effects distribution can have effects on inferences of the joint model. However, it should be noted that robustness to misspecification of the random effects distribution has been demonstrated in published literature (Rizopoulos, 2010).

Maximizing Usefulness of Joint Models:

In order to maximize the usefulness of joint models, studies must be properly designed to capture the benefits of the method. One such way proposed by Rizopoulos is to randomize participants to different pre-determined follow up times. For instance, longitudinal studies require participants to follow up with a set of pre-determined times $t_1 \ldots t_n$. However, the likelihood of the survival process requires an estimation of the trajectory of the longitudinal outcome. Having only one set of pre-determined follow up times $t_1 \ldots t_n$ will result in interpolation of the trajectory $m_i(t)$ between these follow up times. Instead, randomizing

participants into several different sets of n follow up times would limit the interpolation between follow up times for the trajectory of the longitudinal outcome (Rizopoulos, 2012b). Furthermore, when the survival outcome and longitudinal outcome are suspected to be associated, studies can be powered based on the joint modeling framework (Chen et al., 2011). Alternatively, when there are concerns about the ability to record an adequate number of measurements of the longitudinal marker respective to the expected survival time of the participants, researchers should consider other methods to examine the validity of primary analysis under the suspicion of non-ignorable missingness such as pattern mixture models (Little & Rubin, 1987)

References:

Andersen, P. K., & Gill, R. D. (1982). Cox's Regression Model for Counting Processes: A Large Sample Study. *The Annals of Statistics*, *10*(4), 1100–1120. http://doi.org/10.2307/2240714

Bell, M. L., Fiero, M., Horton, N. J., & Hsu, C.-H. (2014). Handling missing data in RCTs; a review of the top medical journals. *BMC Medical Research Methodology*, *14*(1), 118. http://doi.org/10.1186/1471-2288-14-118

Cella, D., & Nowinski, C. J. (2002). Measuring quality of life in chronic illness: The Functional Assessment of Chronic Illness Therapy measurement system. *Archives of Physical Medicine and Rehabilitation*, *83*(12 SUPPL. 2), 63–74. http://doi.org/10.1053/apmr.2002.36959

Chen, L. M., Ibrahim, J. G., & Chu, H. (2011). Sample size and power determination in joint modeling of longitudinal and survival data. *Statistics in Medicine*, *30*(18), 2295–2309. http://doi.org/10.1002/sim.4263

Dhillon, H. M., van der Ploeg, H. P., Bell, M. L., Boyer, M., Clarke, S., & Vardy, J. (2012). The impact of physical activity on fatigue and quality of life in lung cancer patients: a randomised controlled trial protocol. *BMC Cancer*, *12*(1), 572. http://doi.org/10.1186/1471-2407-12-572

Garcia-Hernandez, A., & Rizopoulos, D. (2015). % JM Macro Reference Manual. Retrieved from http://www.jm-macro.com/reference-manual.html

Harrell, F. E. (2001). Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. *Springer*. http://doi.org/10.1007/978-1-4757-3462-1

Henderson, R., Diggle, P., & Dobson, A. (2000). Joint modelling of longitudinal measurements

and event time data. *Biostatistics (Oxford, England)*, *1*(4), 465–480. http://doi.org/10.1093/biostatistics/1.4.465

Hsieh, F., Tseng, Y., & Wang, J. (2006). Joint Modeling of Survival and Longitudinal Data : Likelihood Approach Revisited. *Biometrics*, *62*(December), 1037–1043. http://doi.org/10.1111/j.1541-0420.2006.00570.x

Laird, N. M., & Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, *38*(4), 963–974. http://doi.org/10.2307/2529876

Lin, H., Turnbull, B. W., McCulloch, C. E., & Slate, E. H. (2002). Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data:Application to Longitudinal Prostate-Specific Antigen Readings and Prostate Cancer. *Journal of the American Statistical Association*, *97*(457), 53–65.

Little, R. J., & Rubin, D. B. (1987). Statistical Analysis with Missing Data. *Chichester: Wiley. Pearson, K*, *6thser2*. http://doi.org/10.2307/1165119

Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics*, *48*(3), 829–838. http://doi.org/10.2307/2532348

Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, *69*(2), 331–342. http://doi.org/10.1093/biomet/69.2.331

Prentice, R. L. (1989). Surrogate endpoints in clinical trials: Definition and operational criteria. *Statistics in Medicine*, *8*(4), 431–440. http://doi.org/10.1002/sim.4780080407

Press, W. H., Teukolsky, S. a, Vetterling, W. T., & Flannery, B. P. (2007). *Numerical Recipes: The Art of Scientific Computing*. *Technometrics* (Vol. 29). http://doi.org/10.2307/1269484

Rizopoulos, D. (2010). JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. *Journal of Statistical Software*, *35*(9), 1–33. http://doi.org/10.1.1.182.3299

Rizopoulos, D. (2012a). Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Computational Statistics and Data Analysis*, *56*(3), 491–501. http://doi.org/10.1016/j.csda.2011.09.007

Rizopoulos, D. (2012b). Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. Boca Raton, Florida. Chapman and Hall/CRC.

Rizopoulos, D., Verbeke, G., & Molenberghs, G. (2010). Multiple-Imputation-Based Residuals and Diagnostic Plots for Joint Models of Longitudinal and Survival Outcomes. *Biometrics*, *66*(1), 20–29. http://doi.org/10.1111/j.1541-0420.2009.01273.x

Self, S., & Pawitan, Y. (1993). Modeling Disease Marker Processes in AIDS Modeling Disease Marker Processes in AIDS. *Journal of America Statistical Association*, *88*(423), 719–726. http://doi.org/10.1080/01621459.1993.10476332

Sweeting, M. J., & Thompson, S. G. (2011). Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture.

Biometrical Journal, *53*(5), 750–763. http://doi.org/10.1002/bimj.201100052

Tsiatis, A. A., Degruttola, V., & Wulfsohn, M. S. (1995). Modeling the Relationship of Survival to Longitudinal Data Measured with Error - Applications to Survival and Cd4 Counts in Patients with Aids. *American Statistical Association. Journal*, *90*(429), 27–37. http://doi.org/10.1080/01621459.1995.10476485

Wu, M. C., & Carroll, R. J. (1988). Estimation and Comparison of Changes in the Presence of Informative Right Censoring by Modeling the Censoring Process. *Biometrics*, *44*(1), 175–188. http://doi.org/10.2307/2531905

Wulfsohn, M. S., & Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics*, *53*(1), 330–339. http://doi.org/10.1111/j.1541-0420.2006.00719.x