12-4-2015

A Bayesian Analysis of the Impact of Hereditary Attributes on the Diagnosis of ALS Patients

Kathryn Dee Coleman

A BAYESIAN ANALYSIS OF THE IMPACT OF HEREDITARY ATTRIBUTES ON THE DIAGNOSIS OF ALS PATIENTS

by

Kathryn D. Coleman

A Thesis

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Master of Science

Major: Mathematical Sciences

The University of Memphis

December 2015

Dedication

To Mrs. Liz Bias—a mother figure to most, a best friend to many, and a role model to me. My prayer for the continuation of ALS research resonates from a wish within—a hope to minimize a deficit of ALS knowledge for the future in honor of a woman that did not deserve any deficits in her life. Thank you for being my motivation for this analysis—you've done more for me than you will ever know. ALS is no longer able to keep you from laughing that contagious laugh. May you rest in peace knowing the fight against ALS and the race to find a cure will never cease.

Additionally, this research is dedicated to Dr. Dale Bowman's father, James Bowman. Our smartest statisticians shouldn't have to speculate about the progression and hereditary aspects of the disease in which their parents were diagnosed.

Acknowledgements

With the encouragement (and patience) of my advisor and editor Dr. Dale, as well as the support from my parents, family, friends, and fiancé Ross, the light at the end of my research tunnel is closer and within reach. Thanks to all.

This thesis was written to offer a different Bayesian approach on the research of the recognizably and often swift progression of ALS in diagnosed patients. One day, families and patients themselves will not be as overwhelmed by the disease's aspects as they are today. Only continued research will determine how to ease those adjustments, and hopefully discover methods to better suit each patient's treatment and therapy after diagnosis.

"Not only does God definitely play dice, but He sometimes confuses us by throwing them where they can't be seen."

Stephen Hawking

Abstract

The purpose of this study is to determine the relationship between hereditary attributes of patients that have been diagnosed with ALS and the change in their respective ALSFRS and/or ALSFRS-R scores over the length of a clinical trial in which they participated. These scores assess the patient's ability to perform everyday tasks, as well as describe the capability to breathe and eat. Data for each of the 8,600 patients was collected by a group called PRO-ACT, and was de-identified prior to access. Bayesian methods are used to estimate the mean of the posterior distributions of parameters in a simple linear model. Gibbs sampling is used to estimate posterior distributions of regression parameters in order to determine which covariates may affect the progression of ALS as measured by change in ALSFRS(-R) score.

Table of Contents

Introduction

Amyotrophic lateral sclerosis, more commonly "ALS", is identified as a disease that debilitates a diagnosed patient by steadily deteriorating the brain's motor neurons, according to a 2010 report by the National Institute of Neurological Disorders and Stroke (NINDS). These neurons control voluntary muscle movement, and are found in the brain, its stem, and the body's spinal cord (NINDS, 2010). Once a motor neuron is dead, it can no longer send signals from the brain to the muscles throughout the body, most noticeably the legs, arms, and face after symptom onset. Eventually, enough motor neurons die, causing more muscles to weaken, thus limiting mobility and normal bodily functions. This, in turn, limits respiratory ability, which is the most common cause of death among ALS patients (NINDS, 2010). ALS-caused respiratory failure victims tend to pass away within three to five years, according to the National Institute of Neurological Disorders and Stroke (2010), but roughly ten percent of diagnosed patients live ten or more years after symptoms begin. Although muscular degeneration occurs, the five senses of the human body are seemingly unbothered (NINDS, 2010). Even though a patient may experience depression, the functions of the mind, emotions, and intellect remain virtually unaffected.

When this research first launched, I alleged that there would not be enough chronological research on ALS, and definitely not enough history of the disease. I presupposed that there would be too many unanswered questions still remaining that would hinder my ability to work toward a personal contribution to research in this area. I assumed that if we, as a community, were not learning about new ALS treatment advances as often as cancer cure breakthroughs, then it must follow that extensive

research and detailed histories must not exist either. I was only slightly mistaken—the research is happening, but the underlying questions are not being resolved.

Many researchers have conducted genetic analyses on the DNA of ALS patients, using a scientific approach based solely on whether or not a patient's family member had been previously diagnosed. There are numerous statistics that are generally known when it comes to ALS. According to NINDS (2010), one of these includes "ALS is more common among white males, non-Hispanics, and persons aged 60-69 years" (p. 2). Unfortunately for both researchers and patients, NINDS also reports that "in 90 to 95 percent of all ALS cases, the disease occurs apparently at random" (p. 2). Should this "randomness" not be a concern of the general public? Are we not diagnosing ourselves with a surprise disease by taking the back seat when it comes to research? Uncertainty abounds.

Many people have heard of "Lou Gehrig's disease" and its namesake's famous slugger of an athlete. Most have acknowledged the name "Stephen Hawking" as the most brilliant scientist since Einstein, having literally written the book on our universe's black holes (Biography.com, Hawking). Still, the fundamental paradox between these two men continues to beg for an answer. Why did Lou Gehrig not live as long as Stephen Hawking after onset of symptoms and diagnosis? Both are males, diagnosed in early age, but Hawking is now 73 years old, while Gehrig passed away at age 38 (Biography.com, Gehrig). Was it a hereditary trait that allowed Hawking to prolong his life? Did Gehrig's disease overcome him more quickly because his parents were German (Biography.com, Gehrig)? The probing questions linger.

Upon mentioning the term “ALS” in conversation, most everyone knows at least one person in his or her community that has been diagnosed. In my childhood, my mother had a church choir friend that would watch us after school. Having known this woman for so long, it was difficult to decipher why, now all of a sudden, she was tripping on her own feet. She had also begun taking notice that her hands weren’t working like they had before. The diagnosis was ALS, and within the year she had lost her ability to walk, as well as the capacity to complete a plethora of daily tasks. Within two years of her foretelling fall in the choir loft, ALS had taken her life. Why did ALS progress so quickly in her? Her family, like many others, yearns to find answers.

The intention of this paper and accompanying research analysis is not to necessarily answer these vexing questions posed. Instead, it is to offer another approach, an additional line of attack, in an attempt to further the development of ALS prognoses and predictions.

Area of Focus

The purpose of this research is to introduce a simplified analysis of ALS data using the methods of Bayesian modeling. Intentions for an end result are to predict the likelihood of survival and length of life after the diagnosis of ALS using distinct patient characteristics such as family history, demographics, and ALSFRS(-R) data. Bayesian methods allow analysts to consider both prior and posterior distributions. Posterior distributions of the parameters of a linear model relating hereditary data, family history of the disease, and patient demographics to ALSFRS(-R) scores, a measure of the progression of the disease, will be derived. These posteriors will be useful to determine what factors affect the progression of ALS as measured by their ALSFRS(-R) scores.

Review of Related Literature and Research

Since the cataloging of ALS began, multiple studies have been conducted in order to predict various factors that "cause" ALS. Researchers, like patients, tend to want to know the basics: who, when, and why. Making inferences on these basics will assist in the further analysis of the complexities of the disease, such as survival time and effective treatment options that may vary from patient to patient.

The Who

Scientists wish to determine who will be diagnosed with the disease in the future by studying the genetic and hereditary attributes of patients that are already diagnosed. Various analyses have been conducted in order to find mutations in genes, enzymes, and other patient diagnosis identifiers. In the United States today, there are approximately 30,000 people diagnosed with ALS, which amounts to the death of two people for every 100,000 deaths annually (ALSA.org, 2015).

To this date, there has been no definite flag of the disease for sporadic cases, such as an enzyme in the blood or lesion on the skin like other diseases. However, there have been many distinctive genetic characteristics of ALS, especially those of familial ALS, also termed *FALS* (Gros-Louis, Gaspar, & Rouleau, 2006). In 2006, an article was published in the journal *ScienceDirect* and stated that at that time, eight FALS *loci* and six ALS-related genes had been discovered (Gros-Louis et al., 2006). *Loci*, plural of *locus*, are locations on a certain chromosome for a particular gene (Loci, 2015). Familial ALS accounts for 5-10% of all ALS diagnoses (Kortebein & Means, 2013), and the only "classifier" of further FALS classification rather than regular ALS is that a family member was diagnosed as well. These discoveries seemed to have only led to additional

patterns and more questions. The ALS-related genes are not necessarily spot-on identifiers or predictors of the disease, but rather mutations in them. Accounting for approximately 15% of the small population of inherited FALS, a mutation in the gene called SOD1 (Cu/Zn superoxide dismutase) has been a prevalent characteristic between patients (Gros-Louis et al., 2006). If a parent has been identified as having the mutation of SOD1 in their genetic makeup, every offspring of that parent is 50% likely to carry the gene, which is later attributed to the diagnosis and development of ALS (Mitsumoto, 2009). Gros-Louis et al.'s article on FALS states (2006),

> Despite intensive research efforts, the mapping and the identification of genes responsible for classical form of FALS has met with limited success. This difficulty arises in part because large families with sufficient statistical power for linkage analysis are hard to come by due to the late onset and age-dependant penetrance of the disease, and the relative short survival time of affected ALS patients. (p. 956)

The When

Naturally, recently diagnosed patients of any disease may have the inclination to ask, "When will I die?" or, "How much time do I have?" While this is a tough question to answer, we revert back to the fact that physicians are usually able to confidently answer this question when it arises with other diseases. For example, doctors are almost always capable of predicting a cancer patient's life expectancy after diagnosis. This prediction, however, hasn't commonly become accurately available to ALS patients.

In Hiroshi Mitsumoto's patient and family guide to ALS, contributing author Valerie Cwik (2009) makes the following statement:

> Although researchers are looking for bio-markers (blood tests, imaging studies, and other tests) that will allow for more precise and earlier diagnosis, at the present time there is no single laboratory test that allows one to make a diagnosis of ALS with 100% certainty, particularly in its early stages. (p. 27)

As researchers, we look for patterns, signs, and indicators that lead to a breakthrough in the development of a disease when studying its life cycle and prognosis. Have we failed or not kept ahead of the curve in the instance of ALS? Have we not been able to provide physicians with adequate information to accurately diagnose a life-threatening disease early enough? If doctors are unable to classify a person's clumsy feet and slowing hand skills as ALS in the early stages, how will we ever get to the point of productive therapy and possible cure for this rapidly progressive disease?

Furthermore, the forecast for a diagnosed patient seems too vague. Patients hold on to hope that their type of ALS is the same as Stephen Hawking's in that they are still able to live a long life. This is especially the case when physicians are still in the elementary stages of understanding and diagnosing the disease. The median length of the disease—from diagnosis until death—ranges from 23 to 52 months (Mitsumoto, 2009). According to Mitsumoto, in one of the latest studies in 2009, it was concluded that if the patient maintains a certain mental health, that the patient's prognosis is extended.

The Why

Researchers would like to know more about the disease itself in general. Since specific genes have been identified as being related to ALS, the possibility exists that there are more unidentified ALS-connected genes. Since its first discovery of familial mutation over 22 years ago, SOD1 has been the only genetic link to almost 20% of the

FALS cases, which is only approximately 10% of the ALS-diagnosed population (Gros-Louis et al., 2006). This, in turn, leads researchers to believe that another gene or mutation must exist that are definite characteristics of FALS, and possibly ALS.

Data Collection and Sources

Although ALS is still considered a rarely diagnosed disease, there have been a large number of records obtained from over 8,600 patients that participated in various clinical trials. The PRO-ACT (Pooled Resource Open-access ALS Clinical Trials) database contains case-specific information from various trials on the progression and treatment of ALS. To date, it is the largest compilation of de-identified patient data according to the Neurological Clinical Research Institute, or NCRI (2015).

In order to protect privacy of ALS patients within its database, all personal identifiers were removed, and "SubjectID" numbers were assigned. For each observation or clinic visit, an alpha-numeric record locator was assigned randomly (NCRI, 2015). The data were collected from varying trials, and their unique trial identifiers were also discarded.

Data from all the experimental tests were compiled and then separated into PRO-ACT's individual assessment files. In total, there exist 11 type files. In order to focus on the research questions posed, the five files and their descriptions used in this analysis are as follows:

- ALSFRS(-R): The ALS Functional Rating Scale (ALSFRS) and its revised version ALSFRS-R are scores that assess the symptom severity during each clinic visit for each patient. The score data are separated into 10 symptom assessments,

with ratings ranging from 0 (complete loss of function) to 4 (normal function) for a maximum score of 40:

1. Speech
2. Salivation
3. Swallowing
4. Handwriting
5. Cutting food and handling utensils (with or without gastrostomy)
6. Dressing and hygiene
7. Turning in bed and adjusting bed clothes
8. Walking
9. Climbing stairs
10. Breathing

Some ALS patients are diagnosed with dysphagia, which causes problems in processing food and sustaining nutritional nourishment, and can sometimes lead to decrease in weight as well as breathing difficulties while eating ("Gastrostomy in patients," 2015). When this occurs, the physician may elect to carry out a *gastrostomy* procedure. This is a way to provide lasting dietary sustenance to the patient without choking or prolonged feeding times due to dysphagia.

With ALSFRS-R, the score for the "Breathing" assessment was broken down into three additional responses in order to more accurately assess respiratory function. Those three responses include respiratory difficulties including dyspnea, orthopnea, and respiratory insufficiency (NCRI, 2015).

- Death Report: This data simply displays one statistic—whether death occurred during the clinical trial. If so, the file also lists the number of days (also called the *delta*) since the patient began the trial.
- Demographics: This assessment file's data is collected in the beginning of the trial and includes patient physical characteristics such as age, gender, race, and age at onset (NCRI, 2015).
- Family History: As mentioned previously, about 5% of ALS cases are considered familial, or hereditary. This file outlines which family member of the diagnosed patient was also diagnosed with ALS.
- Subject ALS History: The history of the progression of ALS within a patient is vital to determining what is next in the duration of the disease. As ALS begins to take over the body, the patient will experience the loss of different parts of the body at varying times. This file contains data that identifies the site of onset, as well as the different sites of muscle loss as the disease wears on.

In each file, there is a delta listed for each assessment by patient record. The number defined to be the delta for each visit/assessment is equivalent to the number of days since the patient began the respective trial; negative values for delta indicate that the trait being assessed occurred previous to trial start.

Interestingly enough, the PRO-ACT organization has also recognized that ALS is only diagnosed given its symptoms, and that there is no unique test that will show that a patient is unquestionably suffering from ALS. According to PRO-ACT data and NCRI (2015), "the gap in time between onset of symptoms and diagnosis is on average more than a year." A person may be led to believe that by the time a year or more passes, that

ALS would continue to progress without treatment. This could cause a quicker progression of the disease, or even worse, death before diagnosis.

Organizing the Data

After acquiring the 26 datasets from the PRO-ACT database, manually coded links had to be made from each downloaded zipped file in order to compile data for each patient. The 13 zipped folders in the list pictured Figure 4 in the "Figures" section of the Appendix each contained two files: a "Data" file containing the observational recorded data by patient and Record ID, as well as a "Data Dictionary" which was a general explanation of the coding that was used in compiling the data during the clinical trials (see Table 3 in the Appendix for Data Dictionary codes). To get an idea of what the messy original files looked like directly after download, see Figures 5 and 6 in the Appendix. Since downloaded datasets were not previously separated into fields specified by their headers, the "Text to Column" feature in Microsoft Excel was implemented and columns were separated (see Appendix, Figures 7 and 8). Then, in the Data Files, reference formulas were written to lookup each respective Data Dictionary code to create a fuller picture for each patient.

Unnecessary data was removed from further consideration, as we were interested in the ALSFRS and ALSFRS-R scores, patient demographics, family history, patient history, and death data. For this analysis we are interested in only the patients with complete records. Although only 20.6% of all the compiled patients had what we considered "complete" records, it still left us with 1,753 patient records that contained information regarding gender, age, race, location onset, onset delta days, death of the patient, and either a corresponding ALSFRS or ALSFRS-R score.

Variable Selection and Methods Used

Given the PRO-ACT datasets, the goal is to predict the expected prognosis of an ALS patient after diagnosis. In the PRO-ACT files, we have ALSFRS and ALSFRS-R scores (further referred to as ALSFRS(-R) scores) which measure the advancement of the disease for each patient. The scores are assessed at each clinic visit over the period of the clinical trial, and correspond to respective delta values. The deltas are measured in days since the trial began. These scores will represent our response variable. In a Bayesian analysis performed by Gilks, Richardson, and Spigelhalter (1998) on Gambian infants infected with Hepatitis B (HB) called the "Gambian Hepatitis Intervention Study" (GHIS), two blood samples were taken from each observed infant over a length of time after vaccination was administered. Our data for the observed ALS patients is similar in that over the period of a clinical trial, ALSFRS(-R) scores are assigned by assessors (assumed clinicians, nurses, or physicians). Thus, we will model our analysis similar to the GHIS analysis.

First, we wish to develop a simple linear model which describes the change in the ALSFRS(-R) score over time. We will be interested in the ALSFRS(-R) score assigned at the initial visit, the final visit, and the difference of the two. We will assume that each clinical trial lasted an equal number of days. However, we cannot assume that each patient's final assessment was on the last day of the trial. Thus, the delta in days from initial to final score may differ for each patient. We will also assume that if the final score is equivalent to zero, that the patient passed away before the end of the trial.

The data provided include score records, demographics, patient history, family history, and death data. PRO-ACT data was collected worldwide, varying in notation

from trial to trial, so portions of the datasets are rather messy. The datasets contain observations from over 8,600 patients, but some observations or patients as a whole were omitted due to lack of score data.

Let S_{i0} represent the ALSFRS(-R) score given to patient *i* at the initial clinic visit, where *t = 0*. This will be considered a baseline score, and we are interested in the change from this initial score to the final score assigned to patient *i* over their respective number of *delta* days from initial visit until final assessment, t_i. The change in the score over time t_i will be defined as $Y_i = S_{i0} - S_{i\delta_i}$, where δ_i represents the final visit for patient i. The differences in scores may be distributed as $Y_i \sim N(\mu_i, \sigma^2)$, where σ is unknown.

Let t_i represent the maximum delta for each patient, since the corresponding ALSFRS(-R) score is assumed to be the final score, and is subtracted from the initial score. Every patient's t_i value differs since scores were assessed at various times, and some patients expired before the trial ended. Since we are considering the impact of covariates, we will assign identifier variables, $X_{\rho i}$, to represent patient attributes:

$$X_{2i} = \begin{cases} 1, & if\ male \\ 0, & \text{if female} \end{cases}$$

$$X_{3i} = \text{age of patient } i$$

$$X_{4i} = \begin{cases} 1, & if\ bulbar\ onset \\ 0, & \text{otherwise} \end{cases}$$

$$X_{5i} = \begin{cases} 1, & if\ limb\ onset \\ 0, & otherwise \end{cases}$$

$$X_{6i} = \begin{cases} 1, & if\ death\ during\ trial \\ 0, & \text{otherwise} \end{cases}$$

Let

$$\mu_i = \beta_0 + \beta_1 t_i + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + \beta_6 X_{6i}$$

for $i = 1, \ldots, N$ where $N = 1{,}753$ is the total number of observed patients in our sample, and where β_ρ are regression coefficients distributed as

$$\beta_\rho \sim N\left(\mu_{\beta_\rho}, \sigma^2_{\beta_\rho}\right).$$

The priors on β_ρ are non-informative priors that are normally distributed with mean zero and large variance, since we are unaware of any previous information that their respective parameter β_ρ or attribute $X_{\rho i}$ may contain.

Before beginning the analysis, we look at the data using exploratory methods. By plotting the raw data for difference in final and initial scores versus the number of days passed when the final score was assessed, we are able to see in Figure 1 that the longer a patient is observed, the larger the difference becomes between a patient's initial score and their final observed ALSFRS(-R) score, which is expected. This is analogous to stating that a patient's score will decline incrementally if given more days to be measured. Figure 1 simply shows the relationship between delta days and score decrease for each patient. It does not serve as our predictor model since it does not take into consideration our covariates. We may also use a function in R software called "predict()" in order to investigate the predicted score change in a patient given certain attributes for this raw data linear model. Based solely on the linear model in R defined as

lm (Y ~ time + sex + age + bulbar onset + limb onset + death), (LM.1)

we can create confidence intervals for predicting the decrease in score for each patient. For example, we have a 55-year-old female diagnosed with ALS and having had limb onset symptoms, and death at 730 days. We are able to predict that her decrease in score would have been 22.525 points at 97.5% confidence. This, however, does not seem to be

a true fit. We would expect a larger decrease in points since given the information that death occurred.

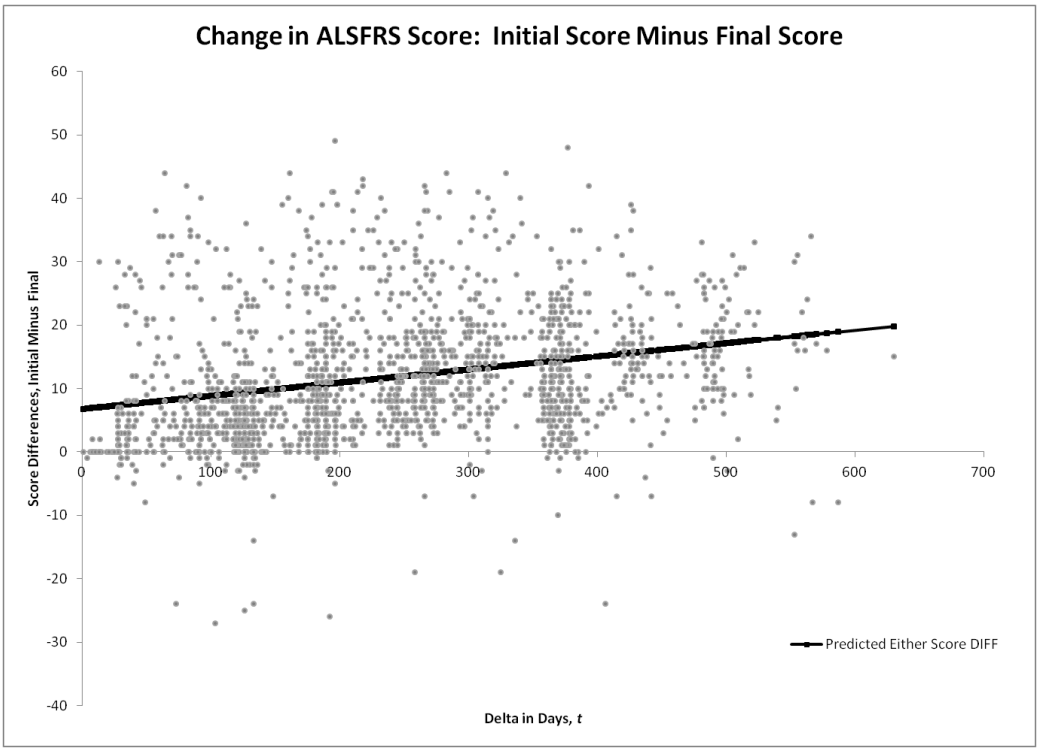

Figure 1: Delta Days versus Difference in ALSFRS Score

We will assume all individual Y_i differences in scores are independent of each other between patients, conditional on their respective means, μ_i, as well as an unknown σ parameter, which directs the sampling error (Gilks, Richardson, & Spiegelhalter, 1998).

A *directed acyclic graph,* or DAG (see Figure 2) is illustrative of the "parent nodes" of each parameter. Thus, a DAG allows us to assume that "the joint distribution of all the random quantities is fully specified in terms of the conditional distribution of each node given its parents" (Gilks et al., 1998). It also gives us a better picture of the nodes of our model, its parameters, and their prior non-informative distributions.

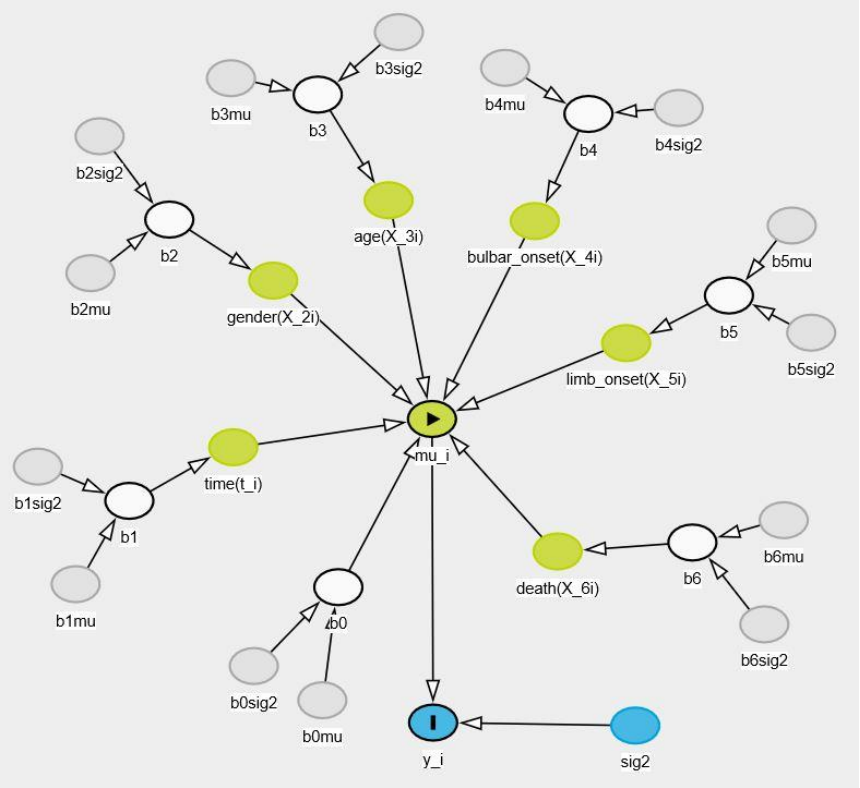

Figure 2: DAG illustrating our model's nodes and priors on parameter nodes.

For the set of priors without parent nodes, μ_{β_ρ} and $\sigma^2_{\beta_\rho}$, we will assume they are normally distributed with mean equivalent to zero and large variance. Using similar priors as Gilks et al. (1998) and the study on GHIS infants, we let the variance be distributed as large as possible, and since we are programming in BUGS (and later in R), we use the precision *tau* (τ) instead of the variance where $\tau = 1/\sigma^2$. Since we want the variance to be large, we let $\tau \sim gamma(0.001, 0.001)$.

Now, we are interested in the mean, μ_i, of the posterior distribution. We will use Gibbs sampling to fit our model, and implement the sampling via statistical software referred to as BUGS. Known for its ability to easily work with Bayesian modeling, our programming choice was BUGS, which is an acronym for "Bayesian Inference Using

Gibbs Sampling" (Lunn, Jackson, Best, Thomas, & Spiegelhalter, 2013). We will also implement R programming software in order to further verify our model.

To set up our process, we first provide initial values for all priors on parameter nodes that we want to observe. We want the Gibbs sampler to overlook its starting values since they are unimportant. We will perform cycles, or runs, with initial values shown in Table 2 in the Appendix. We also implement one chain for each run since we have separated our initial values into three separate vectors. In our first run, we choose starting values based on the least squares estimate of each parameter. In our second and third runs, we set our initial values to be considerably different than those in the first run. We hope that our sampler will eventually "forget" its initial states and begin to converge to one estimated value.

According to Gilks et al. (1998), Gibbs sampling iteratively draws samples from the "full conditional distributions of unobserved nodes in the graph" (p.29). After preparing our beginning values for each run, we may construct a fully conditional distribution for each node in the DAG. Thus, referring back to for any node, say *x*, in our Bayesian model, we can represent the other nodes (not node *x*) as Y_{-x} and rewrite our model as the proportion

$$P(x|Y_{-x}) \propto P(x, Y_{-x})$$

where

$$P(x, Y_{-x}) = P(x|\text{parents}[x]) \times \prod_{z\,\epsilon\,\text{descendants}[x]} P(z|\text{parents}[z]). \quad \text{(P.1)}$$

As demonstrated in (P.1), the fully conditional distribution for *x* has a prior probability $P(x|\text{parents}[x])$.

In monitoring the output, the values generated by the Gibbs sampler must be checked for convergence and intermingling, and can be done so visually in a graph called a trace plot, as well as checked through the statistical output (Gilks et al., p.31). We will run 5,000 iterations and 500 "burn-ins" for each run separately using our starting values in Table 2 in the Appendix. We will also limit our sample to the first 1,000 patient records in our dataset in order to quickly receive an output. Using the full set 1,753 patients at 5,000 iterations slows the software, and will eventually shut down the software, giving a runtime error. From Run 1, Run 2, and Run 3, we display the trace graphs as well as the densities for each parameter in Figures 9a, 9b, and 9c, respectively. We are able to see that each of the three runs is converging to the same posterior distribution (see Output, Appendix). We also have posterior means for each β_ρ in each run.

With our output, we can investigate point estimates of the mean of each parameter, as well as the *highest posterior density* (HPD) intervals. These intervals are created using our simulated data, so in turn consider the fact that the data the interval is based on is simulated (Ghosh, Delampady, & Samanta, 2006). These intervals also work well when the distribution is bimodal (Ghosh et al., 2006). In Table 1, we list the estimates and 97.5% HPD intervals for Run 1. The other two runs gave very similar values, so we will investigate the first. From Table 1, it appears that the variables of time (delta days) and death, t_i and X_6 respectively, impact our response the most. Thus, we are able to conclude that these greatly impact the mean of Y_i, and are significant in predicting the decrease in ALSFRS(-R) scores for each patient.

Table 1: Posterior means, standard deviation, and confidence intervals for each β_ρ.

Parameters	Mean	SD	2.5%	97.5%
β_0	2.562	2.311	-0.312	7.527
β_1	0.021	0.002	**0.016**	**0.025**
β_2	-0.058	0.342	-0.896	0.626
β_3	-0.012	0.030	-0.074	0.037
β_4	0.884	1.003	-0.530	2.967
β_5	-0.692	1.008	-2.942	0.803
β_6	7.315	0.697	**5.940**	**8.696**
σ^2	9.506	0.214	9.114	9.929

As with most statistical modeling, we are not only interested in the sample we have obtained, but also what the sample may represent about the population from which it was drawn. This will allow us to extrapolate results for future patients. We have developed a model using Bayesian methods with non-informative priors on the parameters. Now, we are able to better predict the decrease in score given delta days, patient attributes, and death data.

In Figure 3, our models have been plotted using R software. The black line and gray points are representative of our sample data and simple regression model (LM.1) from page 13. For our three runs, we have plotted the simulated data given the

parameters and priors, as well as the dashed-line regression models, “reg1”, “reg2”, and “reg3”.

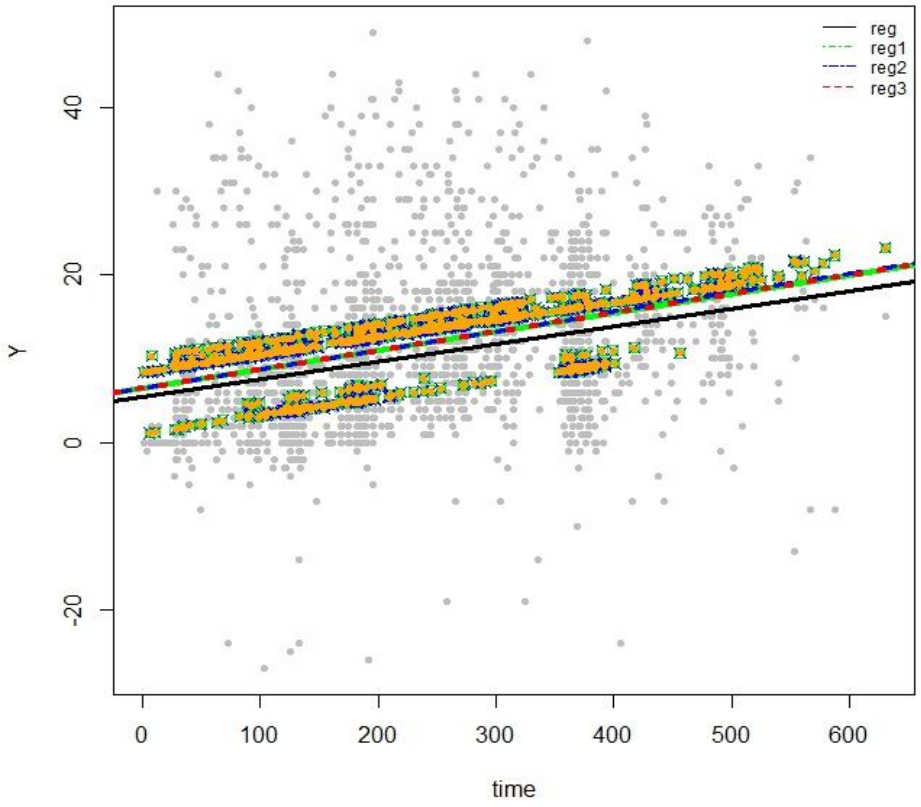

Figure 3: Linear regression models: Black = lm(Y ~ time + sex + age + bulbar + limb + death); Green = Run #1; Blue = Run #2; Red/Orange = Run #3.

Conclusion

After analyzing our data and outputs, we developed a Bayesian model for the change in ALSFRS(-R) scores using non-informative priors. We have applied Gibbs sampling in order to compute the posterior distributions of the parameters of our linear model and in the process, obtained point estimates of those parameters. We were also

able to conclude from our model that an ALS-diagnosed patient with a longer survival time tends to have a larger difference between the initial and final ALSFRS(-R) scores than those patients with shorter survival times.

Interpretation and Recommendations

From this study, many things have been learned, both from a statistical point of view and a research point of view. From the data standpoint, we have learned that even if we are provided a large dataset with multiple variables and factors, it does not imply that all aspects need to be used in formulating a statistical model, and many observations (patients) may be excluded if complete records are required. In beginning this research, we had hoped to use as many variables as possible to describe the patients, even considering multiple multivariate regression. However, as data organization became a hindrance, we decided to limit the number of covariates to gender, age, onset symptom(s), days since onset, and whether or not the patient died during the clinical trial.

From a research standpoint, it was very interesting to learn about the small percentage of familial ALS and how genetic data has already been identified. As for the sporadic cases of ALS, the vast emptiness of research developments was astounding. However, this deficit simply signifies an opportunity for future researchers on the prognosis of ALS.

As a recommendation for continuation of this study, a more sophisticated approach that includes the patient records that do not contain all the covariates may be developed. Due to the various clinical trials and their respective directors, unfortunately not all data was collected uniformly. Further, additional covariates that are available in the Data Files may be considered, including clinical data regarding the administering of a

placebo drug versus the clinical drug being tested, as well as lab results. It is also possible that the ALSFRS(-R) scores may not be a good predictor of disease progression.

Although tobacco use data, drug use data (including prescription and non-prescription drugs), and alcohol use data are not collected in the PRO-ACT Data Files, it would be interesting to see the effects of these covariates on a patient's ALS prognosis. Another factor that would be interesting to consider would be patient environmental factors. Examples might include the city's air quality index in which the patient resides, or the quality of drinking water in the area. Also, genomic information could be of particular use if it were collected.

In an article published in 2010, Anthony Hardie recaps a presentation made by researcher Dr. Ronnie Horner from the University of Cincinnati discussing the prevalence of ALS diagnoses among Gulf War veterans that served in 1991, deployed and non-deployed service members. Dr. Horner goes on to examine the fact that over half of the ALS-diagnosed veterans were 25-years-old or younger, and 98% younger than 55-years-old (Hardie, 2010), which is the average age of diagnosis. While searching for a common theme among these diagnoses, Dr. Horner considers environmental factors such as exposure to heavy metallic substances in the land where they served, substances in the vaccine for anthrax, as well as significant damage to the head during training or in battle. His hypothesis states that these men and women were exposed to chemical agents during the Khamisiya demolition in 1991, and that this exposure triggered an elevated risk to developing ALS (Hardie, 2010).

This project was begun in hopes of submitting the analysis and results to a national research competition called the "DREAM ALS Stratification Prize4Life

Challenge" (Prize4Life, 2015). This contest asks the contestants to predict either the progression of ALS using the ALSFRS(-R) scores, or to predict the probability of survival within a given time period post-diagnosis. We plan to continue this work into next year in order to submit a predictive model before the anticipated September 2016 deadline.

References

ALS Association (2015). *Quick Facts About ALS & the ALS Association.* Retrieved from http://www.alsa.org/news/media/quick-facts.html

Cady, J., Allred, P., Bali, T., Pestronk, A., Goate, A., Miller, T. M., & ... Baloh, R. H. (2015). Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes. *Annals of Neurology*, *77*(1), 100-113 14p. doi:10.1002/ana.24306

Biography.com Editors. (n.d.). Lou Gehrig Biography. Retrieved from http://www.biography.com/people/lou-gehrig-9308266

Biography.com Editors. (n.d.). Stephen Hawking Biography. Retrieved from http://www.biography.com/people/stephen-hawking-9331710

Gastrostomy in patients with amyotrophic lateral sclerosis (ProGas): a prospective cohort study. *The Lancet Neurology* [serial online]. July 1, 2015; 14:702-709. Available from: ScienceDirect, Ipswich, MA. Accessed September 27, 2015.

Ghosh, J., Delampady, M., & Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods.* Location: Springer Science+Business Media, LLC.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. J. (1998). *Markov chain Monte Carlo in practice*. Boca Raton, FL: Chapman & Hall.

Gros-Louis, F., Gaspar, C., & Rouleau, G. A. (2006). Review: Genetics of familial and sporadic amyotrophic lateral sclerosis. *BBA - Molecular Basis of Disease*, *1762*(Molecular Basis of Amyotrophic Lateral Sclerosis), 956-972. doi:10.1016/j.bbadis.2006.01.004

Hardie, A. (2010). *ALS outbreak among Gulf War veterans likely has common cause, says leading researcher* [online]. Retrieved from http://www.91outcomes.com/2010/06/study-gulf-war-chemicals-cause-mood.html

Hoff, B. (November 6, 2015). *DREAM ALS Stratification Prize4Life Challenge* [website]. Retrieved from https://www.synapse.org/#!Synapse:syn2873386/wiki/266768

Kortebein, P. M., & Means, K. M. (2013). *Geriatrics*. New York: Springer Publishing Company, Inc.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2013). *The BUGS Book: A Practical Introduction to Bayesian Analysis.* Boca Raton, FL. Taylor & Francis Group.

Miller, R. G., Gelinas, D. F., & O'Connor, P. (2005). *Amyotrophic Lateral Sclerosis*. New York: Springer Publishing Company, Inc.

Mitsumoto, H. (2009). *Amyotrophic Lateral Sclerosis: A Guide for Patients and Families*. New York: Springer Publishing Company, Inc.

National Institute of Neurological Disorders and Stroke (U.S.). (2010). *ALS (Amyotrophic lateral sclerosis) fact sheet.* Bethesda, MD: National Institutes of Health.

Neurological Clinical Research Institute. (2013). *ALL FORMS* [Data File]. Retrieved from https://nctu.partners.org/ProACT/(F(tInsAVJEdGYbA56mgyWvIFcJw56gysUvK35ZVIOuRbnzhCS-925_SyiNkfmcJiLxxcxnxxnt3goVOwkZDL-KbCKpSRAZ3DuFcoXGuShn72klYOaz0uGnBoKOmZnoyPCrzxgr0e-

ykWpEEWldvdSXig2))/Data/Index?Length=0&LongLength=0&Rank=1&SyncRoot=System.Type%5B%5D&IsReadOnly=FalseedSize=True&IsSynchronized=False

Neurological Clinical Research Institute. (2015). *How to Use PRO-ACT.* Retrieved from https://nctu.partners.org/ProACT/(F(tInsAVJEdGYbA56mgyWvIFcJw56gysUvK35ZVIOuRbnzhCS-925_SyiNkfmcJiLxxcxnxxnt3goVOwkZDL-KbCKpSRAZ3DuFcoXGuShn72klYOaz0uGnBoKOmZnoyPCrzxgr0e-ykWpEEWldvdSXig2))/Document/DisplayLatest/2

Loci [Def.1]. 1993-2015. In R. A. Pagon, M. P. Adam, H. H. Ardinger, et al. (Eds.), *GeneReviews* [Internet]. Seattle, WA: University of Washington, Seattle. Retrieved October 31, 2015, from http://www.ncbi.nlm.nih.gov/books/NBK5192/

Simpson, C. L., & Al-Chalabi, A. (2006). Review: Amyotrophic lateral sclerosis as a complex genetic disease. *BBA - Molecular Basis of Disease*, *1762*(Molecular Basis of Amyotrophic Lateral Sclerosis), 973-985. doi:10.1016/j.bbadis.2006.08.001

Textor, J., Hardt, J., & Knuppel. (2010). DAGitty: A Graphical Tool for Analyzing Casual Diagrams. *Epidemiology*. Retrieved from http://dagitty.net/dags.html

Appendix

Tables

Table 1: Posterior means, standard deviation, and confidence intervals for each β_ρ.

Parameters	Mean	SD	2.5%	97.5%
β_0	2.562	2.311	-0.312	7.527
β_1	0.021	0.002	**0.016**	**0.025**
β_2	-0.058	0.342	-0.896	0.626
β_3	-0.012	0.030	-0.074	0.037
β_4	0.884	1.003	-0.530	2.967
β_5	-0.692	1.008	-2.942	0.803
β_6	7.315	0.697	**5.940**	**8.696**
σ^2	9.506	0.214	9.114	9.929

Table 2: Gibbs sampler beginning values (inits) for parameters

Parameter	Run #1	Run #2	Run #3
b0mu	3	60	0
b1mu	0.019	20	0
b2mu	-0.264	-20	0
b3mu	-0.024	-20	0
b4mu	1.256	30	0
b5mu	-0.803	-20	0
b6mu	7.048	90	0
b0tau	0.001	0.00001	10
b1tau	0.001	0.00001	10
b3tau	0.001	0.00001	10
b4tau	0.001	0.00001	10
b5tau	0.001	0.00001	10
b6tau	0.001	0.00001	10
Tau	0.001	0.00001	10

Table 3: Compiled list of Data Dictionaries and coding of areas of interest in Data Files—Demographics, ALSFRS and ALSFRS-R Scores, Family History, Patient History, and Death Data.

FormID	FormName	Form_CDISC_Name	FieldID	Field_Name	Field_CDISC_Name	Field_Description	Value
144	Demographics	NULL	1203	Demographics Delta	DEMDLTA	NULL	NULL
144	Demographics	NULL	1204	Ethnicity	ETHNIC	NULL	Hispanic or Latino
144	Demographics	NULL	1204	Ethnicity	ETHNIC	NULL	Non-Hispanic or Latino
144	Demographics	NULL	1204	Ethnicity	ETHNIC	NULL	Unknown
144	Demographics	NULL	1205	Sex	Sex	NULL	Female
144	Demographics	NULL	1205	Sex	Sex	NULL	Male
144	Demographics	NULL	1206	Race - American Indian/Alaska Native	RCEAMIND	NULL	NULL
144	Demographics	NULL	1207	Race - Asian	RCEASIAN	NULL	NULL
144	Demographics	NULL	1208	Race - Black/African American	RCEBLACK	NULL	NULL
144	Demographics	NULL	1209	Race - Hawaiian/Pacific Islander	RCEHAW	NULL	NULL
144	Demographics	NULL	1210	Race - Unknown	RCEUNK	NULL	NULL
144	Demographics	NULL	1211	Race - Caucasian	RCEWHITE	NULL	NULL
144	Demographics	NULL	1257	Age	Age	NULL	NULL
144	Demographics	NULL	1393	Race - Other	RCEOTH	NULL	NULL
144	Demographics	NULL	1394	Race Other Specify	RCEOTHSP	NULL	NULL
145	ALSFRS(R)	NULL	1213	1. Speech	ALSFRS1	NULL	NULL
145	ALSFRS(R)	NULL	1214	10. Respiratory	ALSFRS10	NULL	NULL
145	ALSFRS(R)	NULL	1215	2. Salivation	ALSFRS2	NULL	NULL
145	ALSFRS(R)	NULL	1216	3. Swallowing	ALSFRS3	NULL	NULL
145	ALSFRS(R)	NULL	1217	4. Handwriting	ALSFRS4	NULL	NULL
145	ALSFRS(R)	NULL	1218	5a. Cutting without Gastrostomy	ALSFRS5A	NULL	NULL
145	ALSFRS(R)	NULL	1219	5b. Cutting with Gastrostomy	ALSFRS5B	NULL	NULL
145	ALSFRS(R)	NULL	1220	6. Dressing and Hygiene	ALSFRS6	NULL	NULL
145	ALSFRS(R)	NULL	1221	7. Turning in Bed	ALSFRS7	NULL	NULL
145	ALSFRS(R)	NULL	1222	8. Walking	ALSFRS8	NULL	NULL
145	ALSFRS(R)	NULL	1223	9. Climbing Stairs	ALSFRS9	NULL	NULL
145	ALSFRS(R)	NULL	1225	ALSFRS Delta	ALSFDLT	NULL	NULL
145	ALSFRS(R)	NULL	1228	ALSFRS Total	ALSFRST	NULL	NULL
145	ALSFRS(R)	NULL	1229	ALSFRS-R Total	ALSFRSRT	NULL	NULL
145	ALSFRS(R)	NULL	1230	R-1. Dyspnea	ALSFRSR1	NULL	NULL
145	ALSFRS(R)	NULL	1231	R-2. Orthopnea	ALSFRSR2	NULL	NULL
145	ALSFRS(R)	NULL	1232	R-3. Respiratory Insufficiency	ALSFRSR3	NULL	NULL
147	Family History	NULL	1236	Family History Delta	FAMHXDLT	NULL	NULL
147	Family History	NULL	1287	Family History of Neurological Disease	NULL	NULL	ALS
147	Family History	NULL	1287	Family History of Neurological Disease	NULL	NULL	DAT (Dementia Alzheimer's Type)
147	Family History	NULL	1287	Family History of Neurological Disease	NULL	NULL	Other
147	Family History	NULL	1287	Family History of Neurological Disease	NULL	NULL	Parkinson's Disease
147	Family History	NULL	1288	Aunt	Aunt	NULL	NULL
147	Family History	NULL	1289	Aunt (Maternal)	AUNTMAT	NULL	NULL
147	Family History	NULL	1290	Aunt (Paternal)	AUNTPAT	NULL	NULL
147	Family History	NULL	1291	Cousin	COUS	NULL	NULL
147	Family History	NULL	1292	Cousin (Paternal)	COUSPAT	NULL	NULL
147	Family History	NULL	1293	Cousin (Maternal)	COUSMAT	NULL	NULL
147	Family History	NULL	1294	Father	FATHER	NULL	NULL
147	Family History	NULL	1295	Grandfather	GFATH	NULL	NULL
147	Family History	NULL	1296	Grandfather (Maternal)	GFATHMAT	NULL	NULL
147	Family History	NULL	1297	Grandfather (Paternal)	GFATHPAT	NULL	NULL
147	Family History	NULL	1298	Grandmother	GMOTH	NULL	NULL
147	Family History	NULL	1299	Grandmother (Maternal)	GMOTHMAT	NULL	NULL
147	Family History	NULL	1300	Grandmother (Paternal)	GMOTHPAT	NULL	NULL
147	Family History	NULL	1301	Mother	MOTHER	NULL	NULL
147	Family History	NULL	1302	Nephew	NEPH	NULL	NULL
147	Family History	NULL	1305	Niece	NIECE	NULL	NULL
147	Family History	NULL	1309	Sibling	SIBLING	NULL	NULL
147	Family History	NULL	1311	Uncle	UNCLE	NULL	NULL
147	Family History	NULL	1312	Uncle (Maternal)	UNCLEMAT	NULL	NULL
147	Family History	NULL	1313	Uncle (Paternal)	UNCLEPAT	NULL	NULL
147	Family History	NULL	1419	Neurological Disease	NeurDis	NULL	NULL
147	Family History	NULL	1420	Neurological Disease Other Specify	NeurDSp	NULL	NULL
147	Family History	NULL	1424	Son	SON	NULL	NULL
147	Family History	NULL	1425	Daughter	DAUGHTER	NULL	NULL
147	Family History	NULL	1426	Sister	SISTER	NULL	NULL
147	Family History	NULL	1427	Brother	BROTHER	NULL	NULL
148	Subject ALS History	NULL	1193	Site of Onset - Bulbar	ALSONBLB	NULL	NULL
148	Subject ALS History	NULL	1194	Site of Onset - Limb	ALSONLM	NULL	NULL
148	Subject ALS History	NULL	1200	Subject ALS History Delta	ALSHXDLT	NULL	NULL
148	Subject ALS History	NULL	1247	Symptom	ALSSYMP	NULL	NULL
148	Subject ALS History	NULL	1248	Symptom - Other (Specify)	ALSSOSP	NULL	NULL
148	Subject ALS History	NULL	1249	Location	SYMPLOC	NULL	NULL
148	Subject ALS History	NULL	1416	Site of Onset	STONSET	NULL	Onset: Bulbar
148	Subject ALS History	NULL	1416	Site of Onset	STONSET	NULL	Onset: Limb
148	Subject ALS History	NULL	1416	Site of Onset	STONSET	NULL	Onset: Limb and Bulbar
148	Subject ALS History	NULL	1416	Site of Onset	STONSET	NULL	Onset: Other
148	Subject ALS History	NULL	1416	Site of Onset	STONSET	NULL	Onset: Spine
148	Subject ALS History	NULL	1417	Onset Delta	ONSTDLT	NULL	NULL
148	Subject ALS History	NULL	1418	Diagnosis Delta	DIAGDLT	NULL	NULL
219	Death Report	NULL	1465	Subject Died	SUBDIE	Subject Died	No
219	Death Report	NULL	1465	Subject Died	SUBDIE	Subject Died	Yes
219	Death Report	NULL	1466	Death Days	NULL	Days after screening	NULL

Table 4: Regression: Delta Days vs. Initial Minus Final ALSFRS(-R) Scores. From this table, we can see that our expected value of the set of intercepts is approximately 6.755, and the expected value of the set of slopes is 0.02076. The plot of this data is displayed in Figure 1.

SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.264245058							
R Square	0.069825451							
Adjusted R Square	0.069294226							
Standard Error	9.783591954							
Observations	1753							
ANOVA								
	df	*SS*	*MS*	*F*	*Significance F*			
Regression	1	12581.4908	12581.4908	131.4423884	2.15361E-29			
Residual	1751	167603.3938	95.71867151					
Total	1752	180184.8846						
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	6.755260155	0.494238165	13.66802614	1.74179E-40	5.785901123	7.724619187	5.785901123	7.724619187
Either Delta	0.020760385	0.001810788	11.46483268	2.15361E-29	0.01720885	0.02431192	0.01720885	0.02431192

Figures

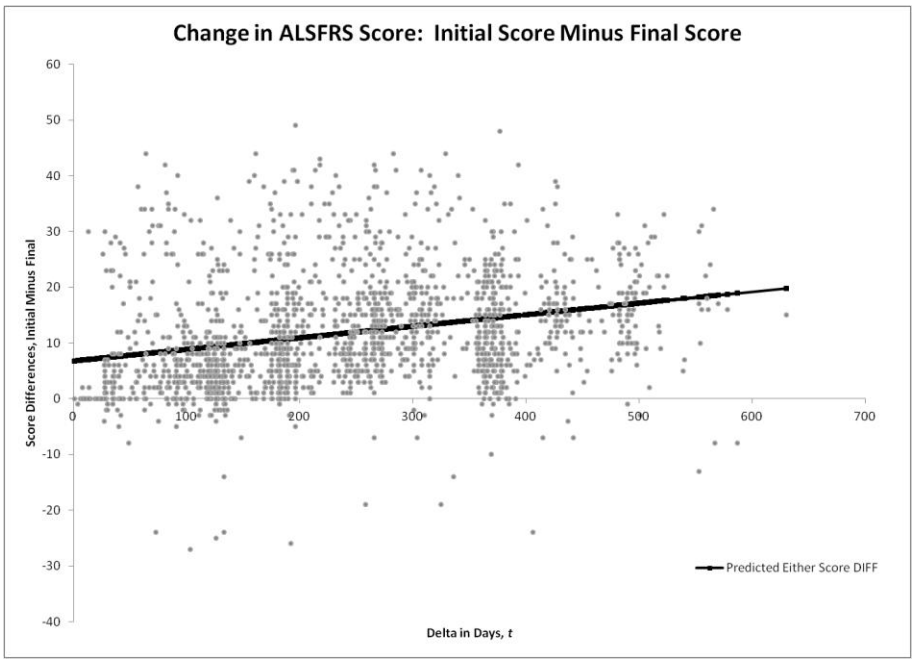

Figure 1: Change in ALSFRS(-R) Score: Initial Score minus Final Score. This plot validates the assumption that as the days continue (Delta), the difference in the initial and the final ALSFRS(-R) score shows growth. Parameters are given in Table 4.

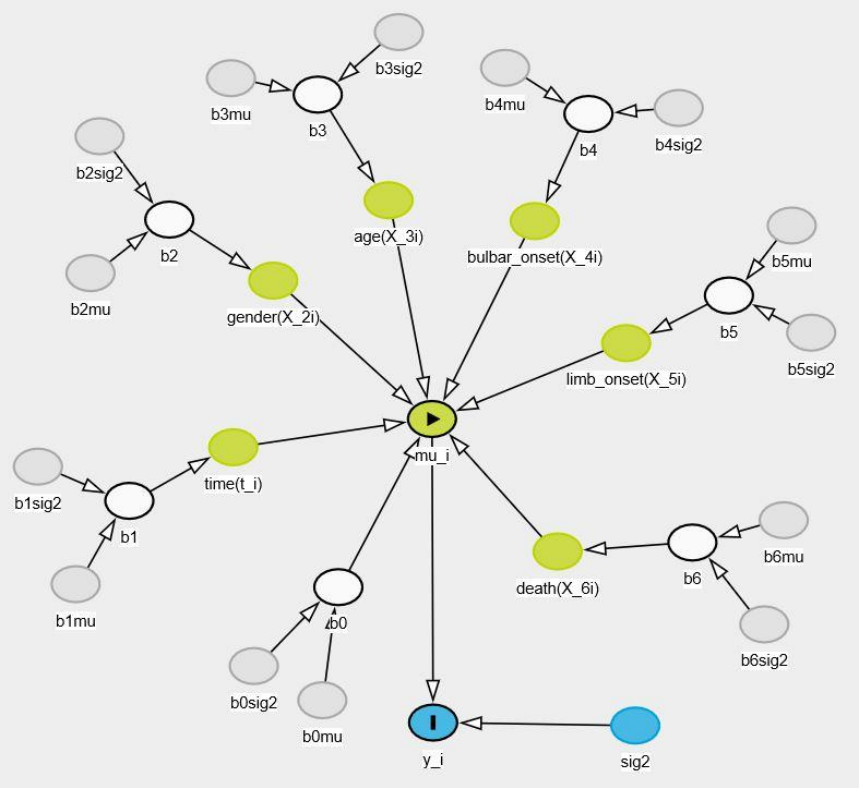

Figure 2: DAG illustrating our model's nodes and priors on parameter nodes.

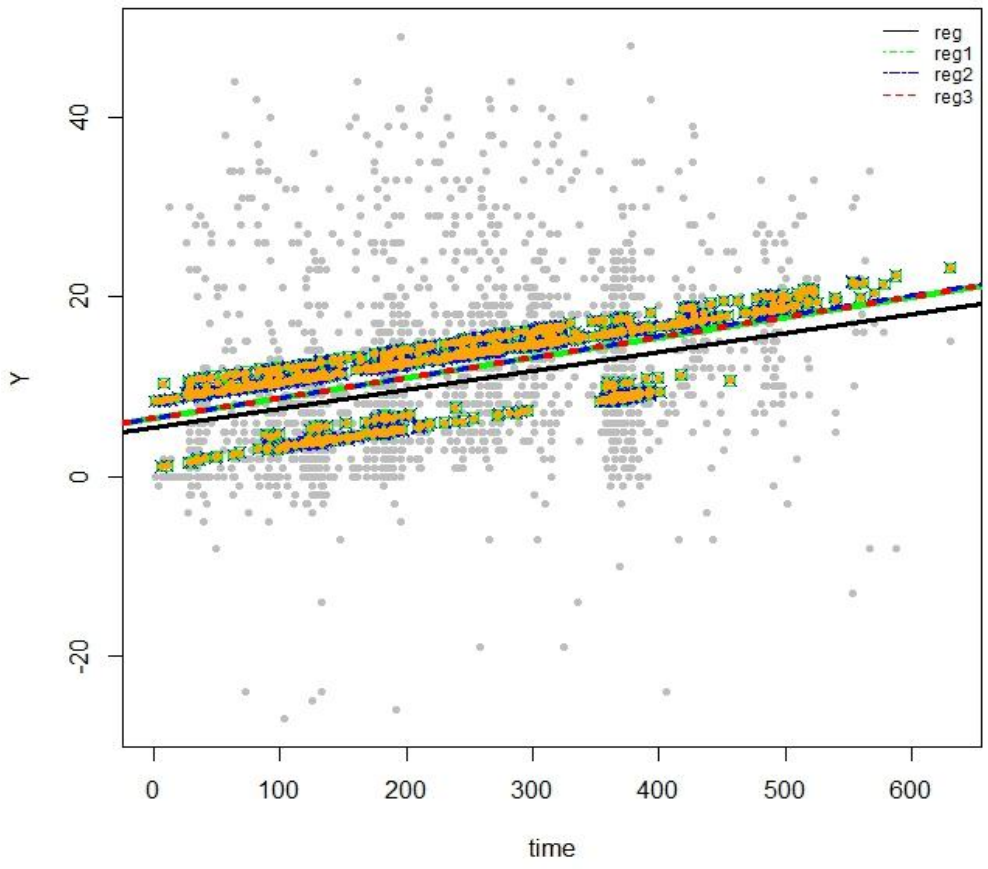

Figure 3: Linear regression models: Black = lm(Y~time + sex + age + bulbar + limb + death); Green = Run #1; Blue = Run #2; Red/Orange = Run #3.

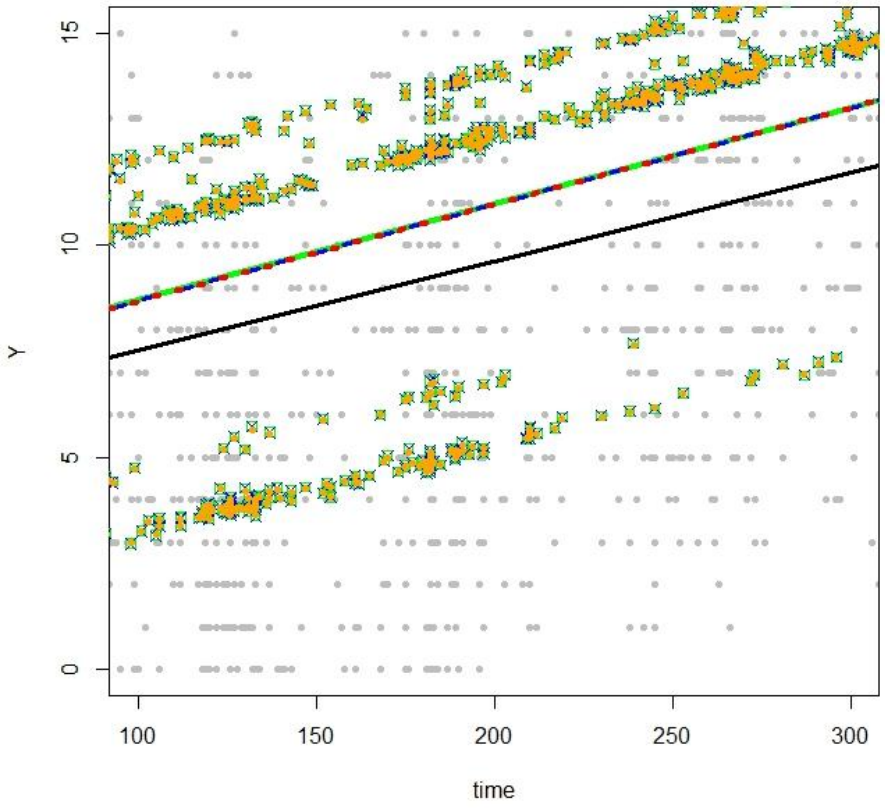

Figure 3a: Zoom-in on left side of regression models from Figure 3.

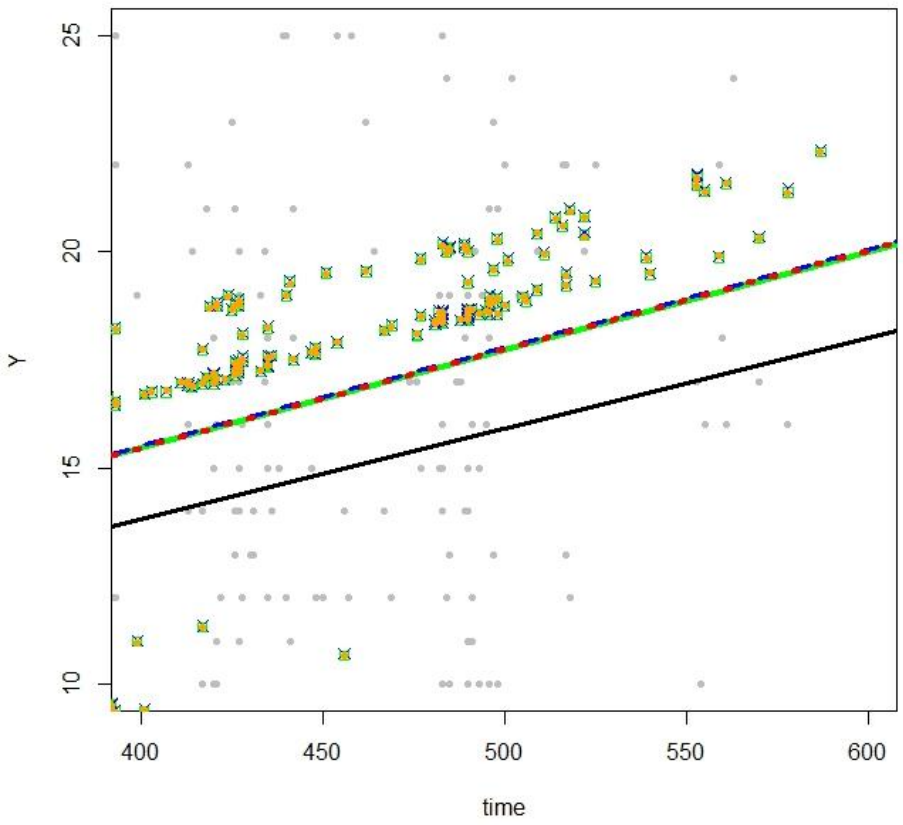

Figure 3b: Zoom-in on right side of regression models from Figure 3.

PROACT_2013_01_23_ALSFRSR_CSV	6/13/2015 10:27 PM	Compressed (zipp...	4,217 KB
PROACT_2013_01_23_Demographics_CSV	6/13/2015 10:50 PM	Compressed (zipp...	1 KB
PROACT_2013_01_23_FamHx_CSV	6/13/2015 10:50 PM	Compressed (zipp...	1 KB
PROACT_2013_01_23_Subject_ALS_Hx_CSV	6/13/2015 10:48 PM	Compressed (zipp...	1 KB
PROACT_2013_08_12_DEATH-XLSX	9/20/2015 1:21 PM	Compressed (zipp...	90 KB
PROACT_2013_08_12_FVC-XLSX	9/20/2015 1:22 PM	Compressed (zipp...	1,525 KB
PROACT_2013_08_12_SVC-XLSX	9/20/2015 1:22 PM	Compressed (zipp...	191 KB
PROACT_2013_08_12_VITALS-XLSX	9/20/2015 1:20 PM	Compressed (zipp...	3,283 KB
PROACT_2013_08_22_TREATMENT	9/20/2015 1:23 PM	Compressed (zipp...	174 KB
PROACT_2013_08_27_LABS	9/20/2015 1:20 PM	Compressed (zipp...	8,055 KB
PROACT_2015_01_09_Adverse_Events-XLSX	9/20/2015 1:21 PM	Compressed (zipp...	6,479 KB
PROACT_2015_02_23_ConMed-XLSX	9/20/2015 1:21 PM	Compressed (zipp...	5,991 KB
PROACT_2015_03_23_RILUZOLE-XLSX	9/20/2015 1:22 PM	Compressed (zipp...	383 KB

Figure 4: List of original zipped data files from PRO-ACT database.

1	FormIDFormNameForm_CDISC_NameFieldIDField_NameField_CDISC_NameField_DescriptionValue
2	144DemographicsNULL1203Demographics DeltaDEMDLTANULLNULL
3	144DemographicsNULL1204EthnicityETHNICNULLHispanic or Latino
4	144DemographicsNULL1204EthnicityETHNICNULLUnknown
5	144DemographicsNULL1204EthnicityETHNICNULLNon-Hispanic or Latino
6	144DemographicsNULL1205SexSexNULLMale
7	144DemographicsNULL1205SexSexNULLFemale
8	144DemographicsNULL1206Race - American Indian/Alaska NativeRCEAMINDNULLNULL
9	144DemographicsNULL1207Race - AsianRCEASIANNULLNULL
10	144DemographicsNULL1208Race - Black/African AmericanRCEBLACKNULLNULL
11	144DemographicsNULL1209Race - Hawaiian/Pacific IslanderRCEHAWNULLNULL
12	144DemographicsNULL1210Race - UnknownRCEUNKNULLNULL
13	144DemographicsNULL1211Race - CaucasianRCEWHITENULLNULL
14	144DemographicsNULL1257AgeAgeNULLNULL
15	144DemographicsNULL1393Race - OtherRCEOTHNULLNULL
16	144DemographicsNULL1394Race Other SpecifyRCEOTHSPNULLNULL

Figure 5: Raw data example of a PRO-ACT Data Dictionary (e.g. Demographics)

1	SubjectIDFormIDRecord_IDFieldIDValue
2	89144B17A1C73-0A2B-4091-842C-D5841EF339FD12030
3	89144B17A1C73-0A2B-4091-842C-D5841EF339FD1205Male
4	329144704CA13F-C1FA-49EC-A3BB-E6B00FA075FA12030
5	329144704CA13F-C1FA-49EC-A3BB-E6B00FA075FA1205Female
6	329144704CA13F-C1FA-49EC-A3BB-E6B00FA075FA1206
7	329144704CA13F-C1FA-49EC-A3BB-E6B00FA075FA1207
8	329144704CA13F-C1FA-49EC-A3BB-E6B00FA075FA12081
9	329144704CA13F-C1FA-49EC-A3BB-E6B00FA075FA1211
10	329144704CA13F-C1FA-49EC-A3BB-E6B00FA075FA125738
11	329144704CA13F-C1FA-49EC-A3BB-E6B00FA075FA1393
12	329144608AB3A7-BC96-4309-AA2D-08F6B7B2BB841204
13	40614485B17364-59B9-4150-AF96-9921742AAC7212030
14	40614485B17364-59B9-4150-AF96-9921742AAC721205Male
15	4111446224DD6E-2B6E-40FB-ABFD-EE694EDAD45D12030
16	4111446224DD6E-2B6E-40FB-ABFD-EE694EDAD45D1205Male
17	5331449C6407A1-6D2D-446D-AD59-0C3A07E8DFA612030
18	5331449C6407A1-6D2D-446D-AD59-0C3A07E8DFA61205Female
19	5331449C6407A1-6D2D-446D-AD59-0C3A07E8DFA61206
20	5331449C6407A1-6D2D-446D-AD59-0C3A07E8DFA61207
21	5331449C6407A1-6D2D-446D-AD59-0C3A07E8DFA61208
22	5331449C6407A1-6D2D-446D-AD59-0C3A07E8DFA612111
23	5331449C6407A1-6D2D-446D-AD59-0C3A07E8DFA6125765
24	5331449C6407A1-6D2D-446D-AD59-0C3A07E8DFA61393
25	5331443E1EDF58-47C6-428C-8A99-B75FB6CCE4F11204
26	6491445AC60165-78AA-4E1D-8CCF-F1A21B944A8B12030
27	6491445AC60165-78AA-4E1D-8CCF-F1A21B944A8B1204
28	6491445AC60165-78AA-4E1D-8CCF-F1A21B944A8B1205Female
29	6491445AC60165-78AA-4E1D-8CCF-F1A21B944A8B1207
30	6491445AC60165-78AA-4E1D-8CCF-F1A21B944A8B1208
31	6491445AC60165-78AA-4E1D-8CCF-F1A21B944A8B12111
32	6491445AC60165-78AA-4E1D-8CCF-F1A21B944A8B125748
33	6491445AC60165-78AA-4E1D-8CCF-F1A21B944A8B1393
34	7081448D0D26E9-7E4C-4DF1-94A0-98FD337D782312030
35	7081448D0D26E9-7E4C-4DF1-94A0-98FD337D78231205Male
36	7081448D0D26E9-7E4C-4DF1-94A0-98FD337D78231206
37	7081448D0D26E9-7E4C-4DF1-94A0-98FD337D78231207
38	7081448D0D26E9-7E4C-4DF1-94A0-98FD337D78231208
39	7081448D0D26E9-7E4C-4DF1-94A0-98FD337D782312111
40	7081448D0D26E9-7E4C-4DF1-94A0-98FD337D7823125746
41	7081448D0D26E9-7E4C-4DF1-94A0-98FD337D78231393
42	708144EB569509-0C7C-49A1-A9E7-ECD1C61DCDE61204
43	90214457A44406-D2C5-4663-9DA4-AE0873E4F03612030
44	90214457A44406-D2C5-4663-9DA4-AE0873E4F0361204Non-Hispanic or Latino
45	90214457A44406-D2C5-4663-9DA4-AE0873E4F0361205Female

Figure 6: Raw data example of a PRO-ACT Data File (e.g. Demographics).

	FormID	FormName	Form_CDISC_Name	FieldID	Field_Name	Field_CDISC_Name	Field_Description	Value
2	144	Demographics	NULL	1203	Demographics Delta	DEMDLTA	NULL	NULL
3	144	Demographics	NULL	1204	Ethnicity	ETHNIC	NULL	Hispanic or Latino
4	144	Demographics	NULL	1204	Ethnicity	ETHNIC	NULL	Unknown
5	144	Demographics	NULL	1204	Ethnicity	ETHNIC	NULL	Non-Hispanic or Latino
6	144	Demographics	NULL	1205	Sex	Sex	NULL	Male
7	144	Demographics	NULL	1205	Sex	Sex	NULL	Female
8	144	Demographics	NULL	1206	Race - American Indian/Alaska Native	RCEAMIND	NULL	NULL
9	144	Demographics	NULL	1207	Race - Asian	RCEASIAN	NULL	NULL
10	144	Demographics	NULL	1208	Race - Black/African American	RCEBLACK	NULL	NULL
11	144	Demographics	NULL	1209	Race - Hawaiian/Pacific Islander	RCEHAW	NULL	NULL
12	144	Demographics	NULL	1210	Race - Unknown	RCEUNK	NULL	NULL
13	144	Demographics	NULL	1211	Race - Caucasian	RCEWHITE	NULL	NULL
14	144	Demographics	NULL	1257	Age	Age	NULL	NULL
15	144	Demographics	NULL	1393	Race - Other	RCEOTH	NULL	NULL
16	144	Demographics	NULL	1394	Race Other Specify	RCEOTHSP	NULL	NULL

Figure 7: Modified version of Demographics Data Dictionary after implementing "Text to Columns" function in Microsoft Excel.

	SubjectID	FormID	Record_ID	FieldID	Value
2	89	144	B17A1C73-0A2B-4091-842C-D5841EF339FD	1203	0
3	89	144	B17A1C73-0A2B-4091-842C-D5841EF339FD	1205	Male
4	329	144	704CA13F-C1FA-49EC-A3BB-E6B00FA075FA	1203	0
5	329	144	704CA13F-C1FA-49EC-A3BB-E6B00FA075FA	1205	Female
6	329	144	704CA13F-C1FA-49EC-A3BB-E6B00FA075FA	1206	
7	329	144	704CA13F-C1FA-49EC-A3BB-E6B00FA075FA	1207	
8	329	144	704CA13F-C1FA-49EC-A3BB-E6B00FA075FA	1208	1
9	329	144	704CA13F-C1FA-49EC-A3BB-E6B00FA075FA	1211	
10	329	144	704CA13F-C1FA-49EC-A3BB-E6B00FA075FA	1257	38
11	329	144	704CA13F-C1FA-49EC-A3BB-E6B00FA075FA	1393	
12	329	144	608AB3A7-BC96-4309-AA2D-08F6B7B2BB84	1204	
13	406	144	85B17364-59B9-4150-AF96-9921742AAC72	1203	0
14	406	144	85B17364-59B9-4150-AF96-9921742AAC72	1205	Male
15	411	144	6224DD6E-2B6E-40FB-ABFD-EE694EDAD45D	1203	0
16	411	144	6224DD6E-2B6E-40FB-ABFD-EE694EDAD45D	1205	Male
17	533	144	9C6407A1-6D2D-446D-AD59-0C3A07E8DFA6	1203	0
18	533	144	9C6407A1-6D2D-446D-AD59-0C3A07E8DFA6	1205	Female
19	533	144	9C6407A1-6D2D-446D-AD59-0C3A07E8DFA6	1206	
20	533	144	9C6407A1-6D2D-446D-AD59-0C3A07E8DFA6	1207	
21	533	144	9C6407A1-6D2D-446D-AD59-0C3A07E8DFA6	1208	
22	533	144	9C6407A1-6D2D-446D-AD59-0C3A07E8DFA6	1211	1
23	533	144	9C6407A1-6D2D-446D-AD59-0C3A07E8DFA6	1257	65
24	533	144	9C6407A1-6D2D-446D-AD59-0C3A07E8DFA6	1393	
25	533	144	3E1EDF58-47C6-428C-8A99-B75FB6CCE4F1	1204	
26	649	144	5AC60165-78AA-4E1D-8CCF-F1A21B944A8B	1203	0
27	649	144	5AC60165-78AA-4E1D-8CCF-F1A21B944A8B	1204	
28	649	144	5AC60165-78AA-4E1D-8CCF-F1A21B944A8B	1205	Female
29	649	144	5AC60165-78AA-4E1D-8CCF-F1A21B944A8B	1207	
30	649	144	5AC60165-78AA-4E1D-8CCF-F1A21B944A8B	1208	
31	649	144	5AC60165-78AA-4E1D-8CCF-F1A21B944A8B	1211	1
32	649	144	5AC60165-78AA-4E1D-8CCF-F1A21B944A8B	1257	48
33	649	144	5AC60165-78AA-4E1D-8CCF-F1A21B944A8B	1393	
34	708	144	8D0D26E9-7E4C-4DF1-94A0-98FD337D7823	1203	0
35	708	144	8D0D26E9-7E4C-4DF1-94A0-98FD337D7823	1205	Male
36	708	144	8D0D26E9-7E4C-4DF1-94A0-98FD337D7823	1206	
37	708	144	8D0D26E9-7E4C-4DF1-94A0-98FD337D7823	1207	
38	708	144	8D0D26E9-7E4C-4DF1-94A0-98FD337D7823	1208	
39	708	144	8D0D26E9-7E4C-4DF1-94A0-98FD337D7823	1211	1
40	708	144	8D0D26E9-7E4C-4DF1-94A0-98FD337D7823	1257	46
41	708	144	8D0D26E9-7E4C-4DF1-94A0-98FD337D7823	1393	
42	708	144	EB569509-0C7C-49A1-A9E7-ECD1C61DCDE6	1204	
43	902	144	57A44406-D2C5-4663-9DA4-AE0873E4F036	1203	0
44	902	144	57A44406-D2C5-4663-9DA4-AE0873E4F036	1204	Non-Hispanic or Latino
45	902	144	57A44406-D2C5-4663-9DA4-AE0873E4F036	1205	Female

Figure 8: Modified version of Demographics Data File after implementing "Text to Columns" function in Microsoft Excel (first 40 observations).

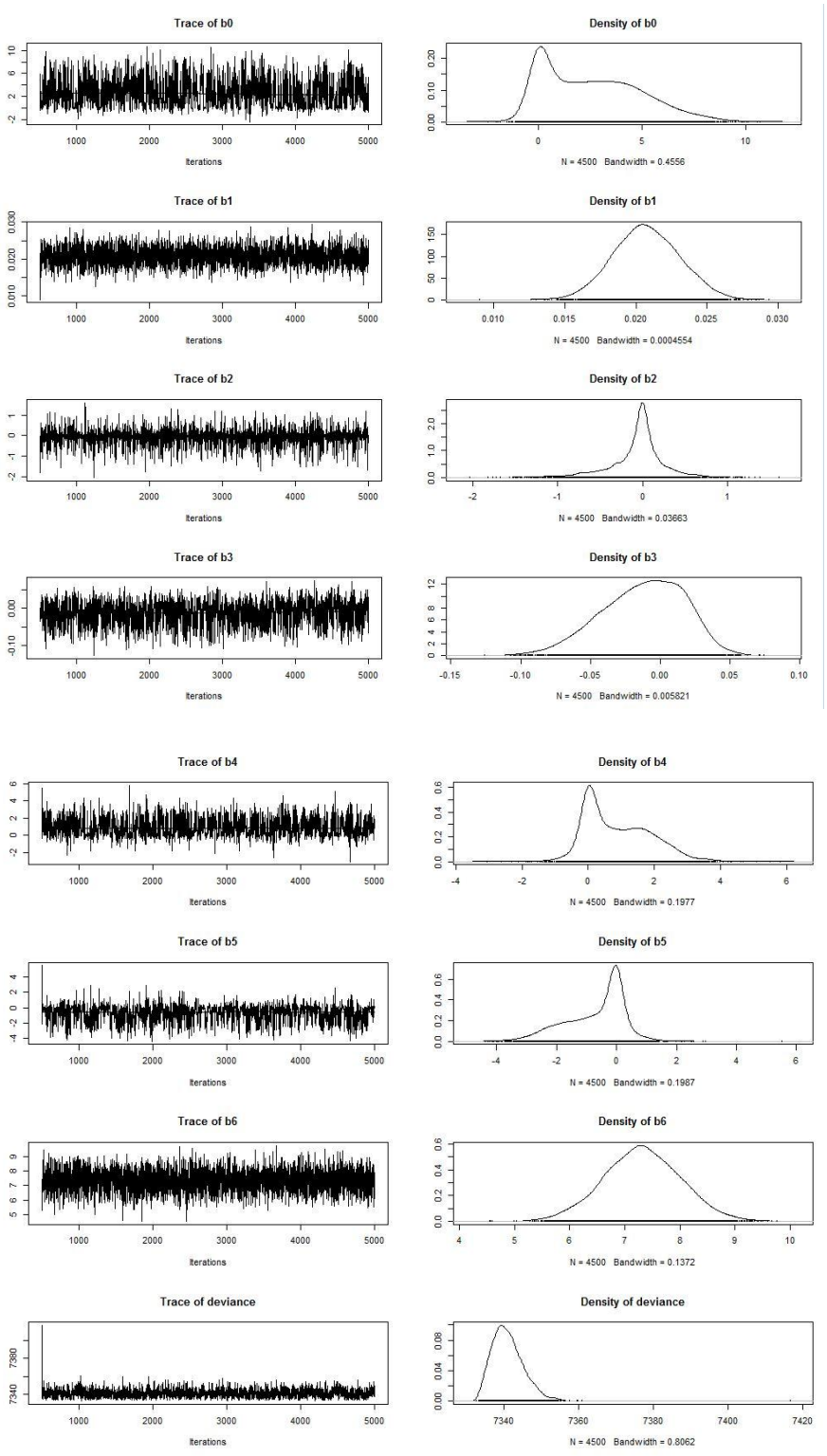

Figure 9a: Trace graphs of our parameters for Run 1 (continued on next page).

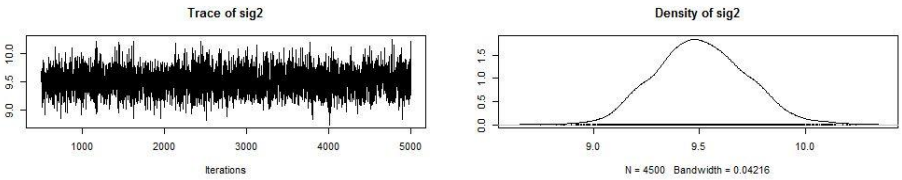

Figure 9a continued: Trace graphs of our parameters for Run 1.

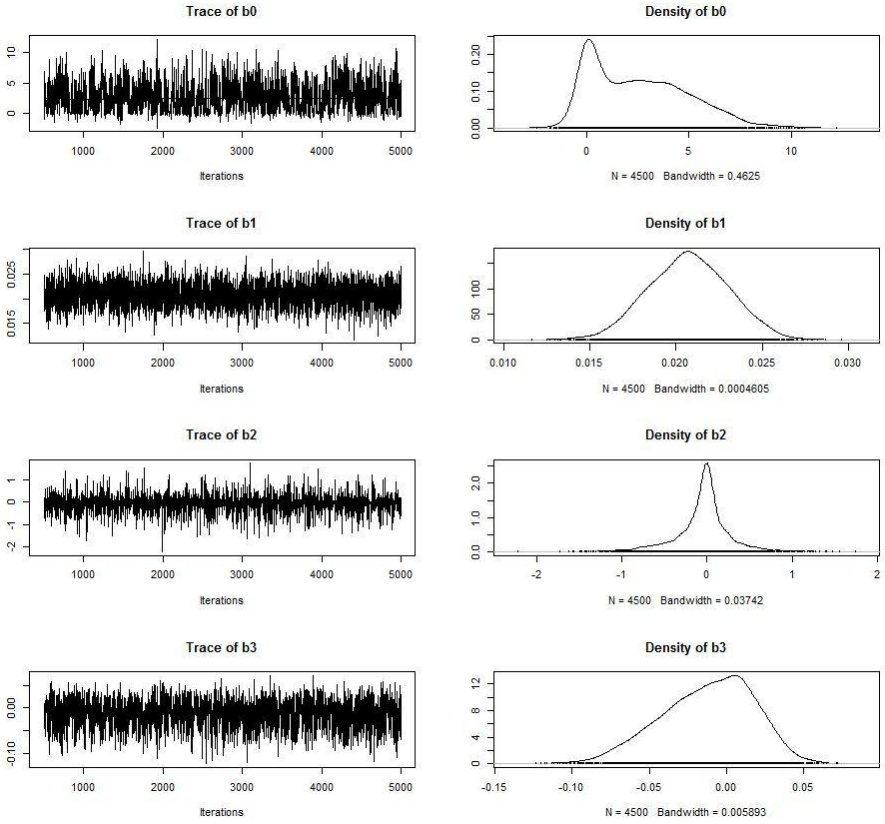

Figure 9b: Trace graphs of our parameters for Run 2 (continued on next page).

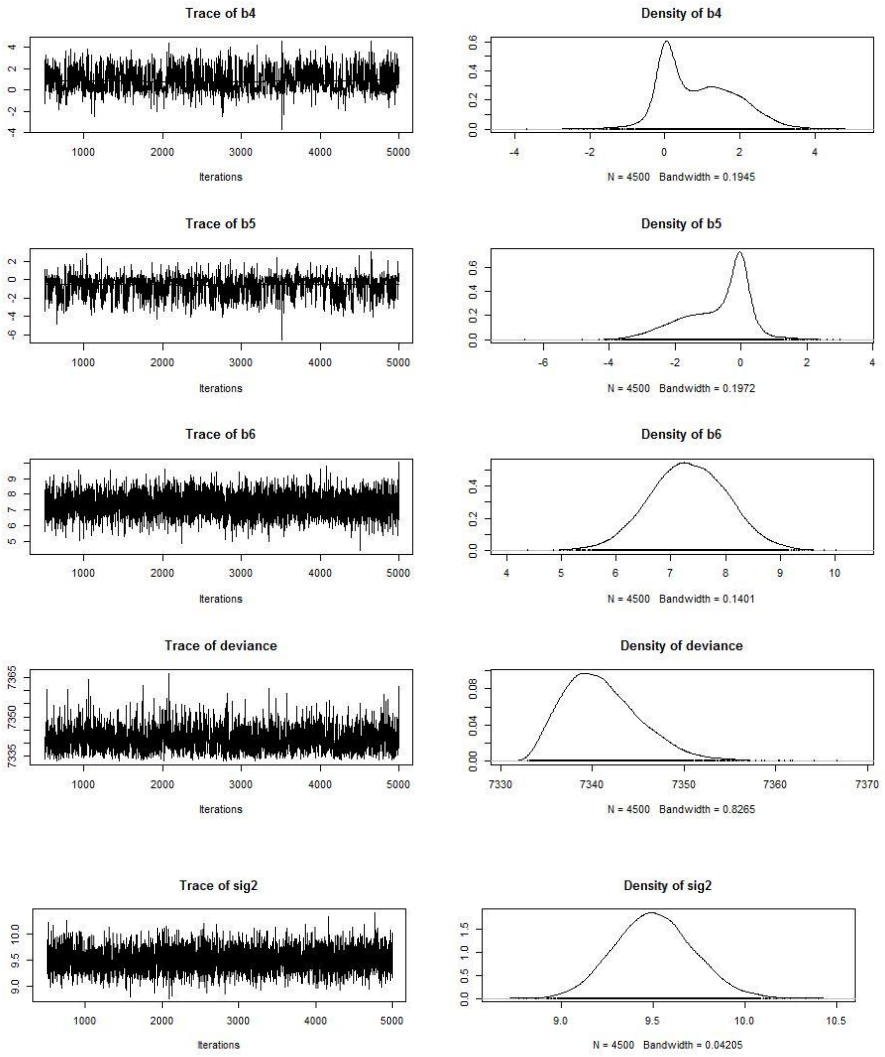

Figure 9b continued: Trace graphs of our parameters for Run 2.

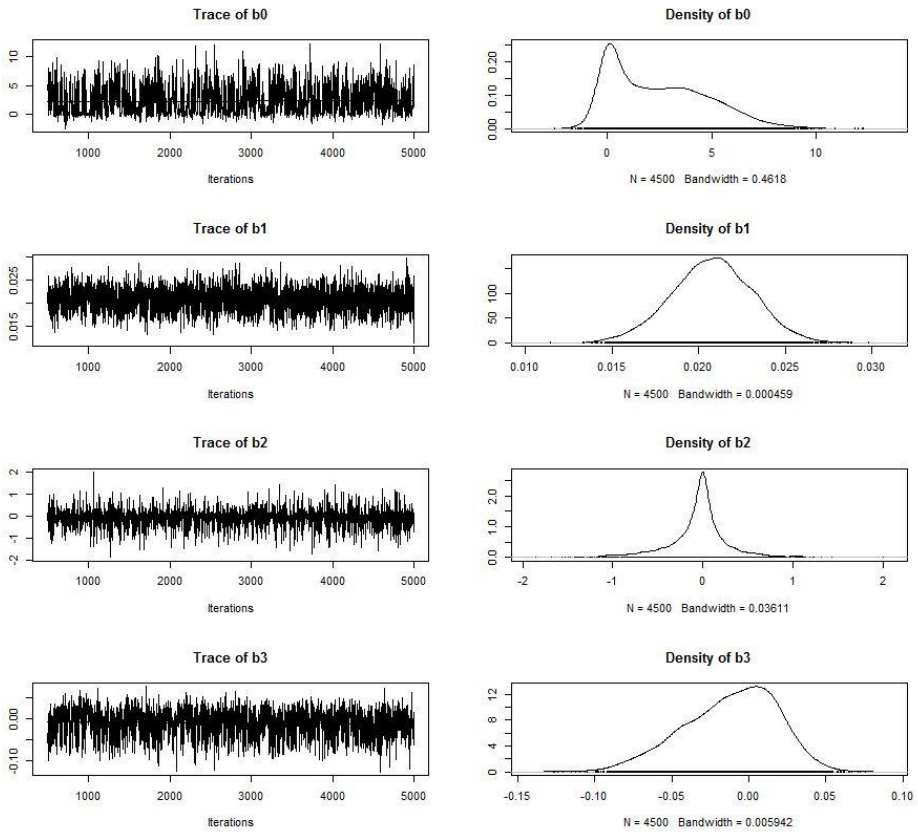

Figure 9c: Trace graphs of our parameters for Run 3 (continued on next page).

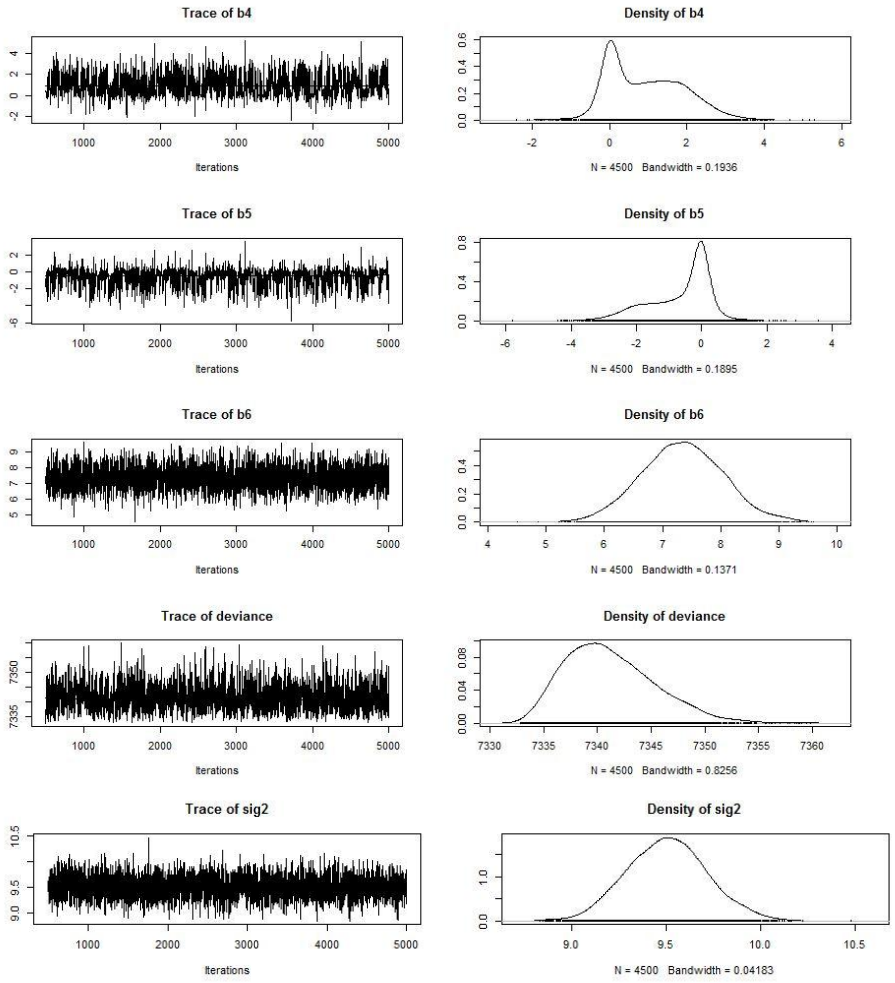

Figure 9c continued: Trace graphs of our parameters for Run 3.

BUGS Code

```
model{for (i in 1:N) {
        Y[i] ~ dnorm(mu[i], tau)
                x2[i] <- equals(sex[i],1)
                x4[i] <- equals(bo[i],1)
                x5[i] <- equals(lo[i],1)
                x6[i] <- equals(death[i],1)

mu[i] <- b0 + b1*t[i] + b2*x2[i] + b3*age[i] + b4*x4[i] + b5*x5[i] + b6*x6[i]
}
        b0 ~ dnorm(b0mu, b0tau)
        b1 ~ dnorm(b1mu, b1tau)
        b2 ~ dnorm(b2mu, b2tau)
        b3 ~ dnorm(b3mu, b3tau)
        b4 ~ dnorm(b4mu, b4tau)
        b5 ~ dnorm(b5mu, b5tau)
        b6 ~ dnorm(b6mu, b6tau)
b0mu ~ dnorm(0, 10000)
                b1mu ~ dnorm(0, 10000)
                b2mu ~ dnorm(0, 10000)
                b3mu ~ dnorm(0, 10000)
                b4mu ~ dnorm(0, 10000)
                b5mu ~ dnorm(0, 10000)
                b6mu ~ dnorm(0, 10000)
b0tau ~ dgamma(0.001, 0.001)
                        b1tau ~ dgamma(0.001, 0.001)
                        b2tau ~ dgamma(0.001, 0.001)
                        b3tau ~ dgamma(0.001, 0.001)
                        b4tau ~ dgamma(0.001, 0.001)
                        b5tau ~ dgamma(0.001, 0.001)
                        b6tau ~ dgamma(0.001, 0.001)
tau ~ dgamma(0.001, 0.001)
        sig2 <- 1/sqrt(tau)
                        b0sig2 <- 1/sqrt(b0tau)
                        b1sig2 <- 1/sqrt(b1tau)
                        b2sig2 <- 1/sqrt(b2tau)
                        b3sig2 <- 1/sqrt(b3tau)
                        b4sig2 <- 1/sqrt(b4tau)
                        b5sig2 <- 1/sqrt(b5tau)
                        b6sig2 <- 1/sqrt(b6tau)
}
list(N = 1753)

#load inits
###RUN #1
```

```
list(
        b0mu=3, b0tau=0.001,
        b1mu=0.019, b1tau=0.001,
        b2mu=-0.264, b2tau=0.001,
        b3mu=-0.024, b3tau=0.001,
        b4mu=1.256, b4tau=0.001,
        b5mu=-0.803, b5tau=0.001,
        b6mu=7.048, b6tau=0.001,
        tau=0.001
        )

###RUN #2
list(
        b0mu=60, b0tau=0.00001,
        b1mu=20, b1tau=0.00001,
        b2mu=-20, b2tau=0.00001,
        b3mu=-20, b3tau=0.00001,
        b4mu=30, b4tau=0.00001,
        b5mu=-20, b5tau=0.00001,
        b6mu=90, b6tau=0.00001,
        tau=0.1
        )

###RUN #3
list(
        b0mu=0, b0tau=10,
        b1mu=0, b1tau=10,
        b2mu=0, b2tau=10,
        b3mu=0, b3tau=10,
        b4mu=0, b4tau=10,
        b5mu=0, b5tau=10,
        b6mu=0, b6tau=10,
        tau=10
        )
#model data (head & tail of dataset)
sex[]   age[]   bo[]    lo[]    death[] t[]     Y[]
0       46      1       0       1       250     33
1       77      0       1       0       154     4
0       63      0       1       0       401     6
0       41      0       1       1       89      8
....
1       66      0       1       1       119     9
0       48      1       1       1       434     26
0       55      0       1       0       184     2
1       60      0       1       0       126     6
END
```

R Code

```
rm(list=ls())

setwd("C:/Users/katy/Copy/GradClasses/THESIS")

library(R2jags)
require(graphics)

scoredata<-read.table("scoredataR.txt", header = T)

sex<-as.vector(scoredata$sex)
age<-as.vector(scoredata$age)
bo<-as.vector(scoredata$bo)
lo<-as.vector(scoredata$lo)
death<-as.vector(scoredata$death)
score0<-as.vector(scoredata$score0)
scorefin<-as.vector(scoredata$scorefin)
time<-as.vector(scoredata$time)
Y<-as.vector(score0-scorefin)

summary(lm(Y~time))
summary(aov(Y~time))
Ytime.predlm<-lm(Y~time)
predict(lm(Y~time),interval="confidence",level=0.975)

scorechg.lm<-lm(Y~time+sex+age+bo+lo+death)
new<-data.frame(time=200,sex=0,age=55,bo=0,lo=1,death=0)
predict(scorechg.lm, new, interval="confidence",level=.975)
liz<-data.frame(time=730,sex=0,age=55,bo=0,lo=1,death=1)
predict(scorechg.lm, liz, interval="confidence",level=.975)




jagdata<-list(sex=sex, age=age, bo=bo, lo=lo,
                  death=death, time=time, Y=Y)
N<-1000
cat(
"model{
for (i in 1:1000) {
        Y[i] ~ dnorm(mu[i], tau)
        x2[i] <- equals(sex[i],1)
        x4[i] <- equals(bo[i],1)
        x5[i] <- equals(lo[i],1)
        x6[i] <- equals(death[i],1)
```

```
        mu[i] <- b0 + b1*time[i] + b2*x2[i] + b3*age[i] + b4*x4[i] + b5*x5[i] +
b6*x6[i]
                }
b0 ~ dnorm(b0mu, b0tau)
        b1 ~ dnorm(b1mu, b1tau)
        b2 ~ dnorm(b2mu, b2tau)
        b3 ~ dnorm(b3mu, b3tau)
        b4 ~ dnorm(b4mu, b4tau)
        b5 ~ dnorm(b5mu, b5tau)
        b6 ~ dnorm(b6mu, b6tau)
b0mu ~ dnorm(0, 10000)
                b1mu ~ dnorm(0, 10000)
                b2mu ~ dnorm(0, 10000)
                b3mu ~ dnorm(0, 10000)
                b4mu ~ dnorm(0, 10000)
                b5mu ~ dnorm(0, 10000)
                b6mu ~ dnorm(0, 10000)
b0tau ~ dgamma(0.001, 0.001)
                        b1tau ~ dgamma(0.001, 0.001)
                        b2tau ~ dgamma(0.001, 0.001)
                        b3tau ~ dgamma(0.001, 0.001)
                        b4tau ~ dgamma(0.001, 0.001)
                        b5tau ~ dgamma(0.001, 0.001)
                        b6tau ~ dgamma(0.001, 0.001)
tau ~ dgamma(0.001, 0.001)
        sig2 <- 1/sqrt(tau)
                        b0sig2 <- 1/sqrt(b0tau)
                        b1sig2 <- 1/sqrt(b1tau)
                        b2sig2 <- 1/sqrt(b2tau)
                        b3sig2 <- 1/sqrt(b3tau)
                        b4sig2 <- 1/sqrt(b4tau)
                        b5sig2 <- 1/sqrt(b5tau)
                        b6sig2 <- 1/sqrt(b6tau)
                }", file="jagmod")



############################################################
jaginits1<-list(list(
        "b0mu"=3, "b0tau"=0.001,
        "b1mu"=0.019, "b1tau"=0.001,
        "b2mu"=-0.264, "b2tau"=0.001,
        "b3mu"=-0.024, "b3tau"=0.001,
        "b4mu"=1.256, "b4tau"=0.001,
        "b5mu"=-0.803, "b5tau"=0.001,
        "b6mu"=7.048, "b6tau"=0.001,
```

```
        "tau"=0.001
        ))
parameters1<-c("b0mu","b0sig2","b0", "b1mu","b1sig2","b1",
"b2mu","b2sig2","b2",
                        "b3mu","b3sig2","b3", "b4mu","b4sig2","b4",
                        "b5mu","b5sig2","b5", "b6mu","b6sig2","b6",
                        "sig2", "mu")

#parameters1<-c("b0", "b1", "b2", "b3", "b4", "b5", "b6",
#                        "sig2")
jagmodel1<-jags(data = jagdata, inits = jaginits1,
                parameters.to.save = parameters1,
                model.file = "jagmod", n.chains = 1,
                n.iter = 5000, n.burnin = 500, n.thin = 1)
#jagmodel1
#plot(jagmodel1)
#traceplot(jagmodel1)
#plot(as.mcmc(jagmodel1),ask=TRUE)


############################################################
jaginits2<-list(list(
        "b0mu"=60, "b0tau"=0.00001,
        "b1mu"=20, "b1tau"=0.00001,
        "b2mu"=-20, "b2tau"=0.00001,
        "b3mu"=-20, "b3tau"=0.00001,
        "b4mu"=30, "b4tau"=0.00001,
        "b5mu"=-20, "b5tau"=0.00001,
        "b6mu"=90, "b6tau"=0.00001,
        "tau"=0.1
        ))
parameters2<-c("b0mu","b0sig2","b0", "b1mu","b1sig2","b1",
"b2mu","b2sig2","b2",
                        "b3mu","b3sig2","b3", "b4mu","b4sig2","b4",
                        "b5mu","b5sig2","b5", "b6mu","b6sig2","b6",
                        "sig2", "mu")
#parameters2<-c("b0", "b1", "b2", "b3", "b4", "b5", "b6",
#                        "sig2")
jagmodel2<-jags(data = jagdata, inits = jaginits2,
                parameters.to.save = parameters2,
                model.file = "jagmod", n.chains = 1,
                n.iter = 5000, n.burnin = 500, n.thin = 1)
#jagmodel2
#plot(jagmodel2)
#traceplot(jagmodel2)
#plot(as.mcmc(jagmodel2),ask=TRUE)
```

```
#############################################################
jaginits3<-list(list(
        "b0mu"=0, "b0tau"=10,
        "b1mu"=0, "b1tau"=10,
        "b2mu"=0, "b2tau"=10,
        "b3mu"=0, "b3tau"=10,
        "b4mu"=0, "b4tau"=10,
        "b5mu"=0, "b5tau"=10,
        "b6mu"=0, "b6tau"=10,
        "tau"=10
        ))
parameters3<-c("b0mu","b0sig2","b0", "b1mu","b1sig2","b1",
"b2mu","b2sig2","b2",
                        "b3mu","b3sig2","b3", "b4mu","b4sig2","b4",
                        "b5mu","b5sig2","b5", "b6mu","b6sig2","b6",
                        "sig2", "mu")
#parameters3<-c("b0", "b1", "b2", "b3", "b4", "b5", "b6",
#                        "sig2")
jagmodel3<-jags(data = jagdata, inits = jaginits3,
                parameters.to.save = parameters3,
                model.file = "jagmod", n.chains = 1,
                n.iter = 5000, n.burnin = 500, n.thin = 1)

#jagmodel3
#plot(jagmodel3)
#traceplot(jagmodel3)
#plot(as.mcmc(jagmodel3),ask=TRUE)


#############################################################
#############################################################
plot(time,Y,pch=20,col="gray")
X<-time+sex+age+bo+lo+death
regplot<-lm(Y~X)
abline(regplot,col="black",lwd=3)
reg<-lm(Y~time+sex+age+bo+lo+death)
summary(reg)

#############################################################
mujags1<-jagmodel1$BUGSoutput$mean$mu
points(time[1:N], mujags1, col="green", pch=22)
reg1<-lm(jagmodel1$BUGSoutput$mean$mu~time[1:N])
abline(reg1, col="green",lwd=4)
summary(reg1)
```

```
#############################################################
mujags2<-jagmodel2$BUGSoutput$mean$mu
points(time[1:N], mujags2, col="blue", pch=4)
reg2<-lm(jagmodel2$BUGSoutput$mean$mu~time[1:N])
abline(reg2, col="blue",lwd=3,lty="dashed")
summary(reg2)
#############################################################
mujags3<-jagmodel3$BUGSoutput$mean$mu
points(time[1:N], mujags3, col="orange", pch=20)
reg3<-lm(jagmodel3$BUGSoutput$mean$mu~time[1:N])
abline(reg3, col="red",lwd=4,lty="dotted")
summary(reg3)

legend('topright', c('reg', 'reg1', 'reg2', 'reg3'),
                   lty=c(1,4,6,8),
                   col=c('black','green','blue','red'),
                   bty='n', cex=.75)

#############################################################
#############################################################
plot(time,Y,pch=20,col="gray", xlim=c(100,300), ylim=c(0,15))
abline(regplot,col="black",lwd=3)
points(time[1:N], mujags1, col="green", pch=22)
abline(reg1, col="green",lwd=4)
points(time[1:N], mujags2, col="blue", pch=4)
abline(reg2, col="blue",lwd=3,lty="dashed")
points(time[1:N], mujags3, col="orange", pch=20)
abline(reg3, col="red",lwd=4,lty="dotted")

#############################################################
#############################################################
plot(time,Y,pch=20,col="gray", xlim=c(400,600), ylim=c(10,25))
abline(regplot,col="black",lwd=3)
points(time[1:N], mujags1, col="green", pch=22)
abline(reg1, col="green",lwd=4)
points(time[1:N], mujags2, col="blue", pch=4)
abline(reg2, col="blue",lwd=3,lty="dashed")
points(time[1:N], mujags3, col="orange", pch=20)
abline(reg3, col="red",lwd=4,lty="dotted")
```

Output

jagmodel1

Inference for Bugs model at "jagmod", fit using jags,
1 chains, each with 5000 iterations (first 500 discarded)
n.sims = 4500 iterations saved

	mu.vect	sd.vect	2.5%	25%	50%	75%	97.5%
b0	2.562	2.311	-0.312	0.313	2.271	4.230	7.527
b1	0.021	0.002	0.016	0.019	0.021	0.022	0.025
b2	-0.058	0.342	-0.896	-0.168	-0.016	0.081	0.626
b3	-0.012	0.030	-0.074	-0.032	-0.010	0.010	0.037
b4	0.884	1.003	-0.530	0.040	0.677	1.620	2.967
b5	-0.692	1.008	-2.942	-1.361	-0.337	0.004	0.803
b6	7.315	0.697	5.940	6.848	7.304	7.781	8.696
sig2	9.506	0.214	9.114	9.358	9.499	9.649	9.929

jagmodel2

Inference for Bugs model at "jagmod", fit using jags,
1 chains, each with 5000 iterations (first 500 discarded)
n.sims = 4500 iterations saved

	mu.vect	sd.vect	2.5%	25%	50%	75%	97.5%
b0	2.545	2.347	-0.353	0.283	2.246	4.205	7.413
b1	0.021	0.002	0.016	0.019	0.021	0.022	0.025
b2	-0.049	0.338	-0.851	-0.168	-0.019	0.086	0.642
b3	-0.012	0.030	-0.074	-0.032	-0.009	0.010	0.038
b4	0.879	0.987	-0.597	0.042	0.717	1.613	2.896
b5	-0.684	1.000	-2.969	-1.354	-0.323	0.007	0.755
b6	7.327	0.711	5.932	6.847	7.315	7.811	8.713
sig2	9.506	0.213	9.100	9.360	9.501	9.647	9.934

jagmodel3

Inference for Bugs model at "jagmod", fit using jags,
1 chains, each with 5000 iterations (first 500 discarded)
n.sims = 4500 iterations saved

	mu.vect	sd.vect	2.5%	25%	50%	75%	97.5%
b0	2.455	2.343	-0.410	0.235	2.058	4.122	7.490
b1	0.021	0.002	0.016	0.019	0.021	0.022	0.025
b2	-0.052	0.334	-0.866	-0.163	-0.017	0.082	0.631
b3	-0.012	0.030	-0.077	-0.032	-0.009	0.011	0.038
b4	0.933	0.982	-0.452	0.055	0.810	1.686	2.943
b5	-0.643	0.984	-2.881	-1.273	-0.251	0.015	0.715
b6	7.340	0.696	5.991	6.869	7.342	7.802	8.753
sig2	9.505	0.212	9.106	9.358	9.505	9.646	9.934