2012-03-09

Predicting Maximal Oxygen Consumption (VO2max) Levels in Adolescents

Brent A. Shepherd
Brigham Young University - Provo

Predicting Maximal Oxygen Consumption ($\dot{V}O_2max$) Levels in Adolescents

Brent Shepherd

A selected project submitted to the faculty of
Brigham Young University
in partial fulfillment of the requirements for the degree of

Master of Science

Gilbert W. Fellingham, Chair
Natalie J. Blades
Candace Berrett

Department of Statistics

Brigham Young University

April 2012

ABSTRACT

Predicting Maximal Oxygen Consumption ($\dot{V}O_2$max) Levels in Adolescents

Brent Shepherd
Department of Statistics, BYU
Master of Science

Maximal oxygen consumption ($\dot{V}O_2$max) is considered by many to be the best overall measure of an individual's cardiovascular health. Collecting the measurement, however, requires subjecting an individual to prolonged periods of intense exercise until their maximal level, the point at which their body uses no additional oxygen from the air despite increased exercise intensity, is reached. Collecting $\dot{V}O_2$max data also requires expensive equipment and great subject discomfort to get accurate results. Because of this inherent difficulty, it is often avoided despite its usefulness. In this research, we propose a set of Bayesian hierarchical models to predict $\dot{V}O_2$max levels in adolescents, ages 12 through 17, using less extreme measurements. Two models are developed separately, one that uses submaximal exercise data and one that uses physical fitness questionnaire data. The best submaximal model was found to include age, gender, BMI, heart rate, rate of perceived exertion, treadmill miles per hour, and an interaction between age and heart rate. The second model, designed for those with physical limitations, uses age, gender, BMI, and two separate questionnaire results measuring physical activity levels and functional ability levels, as well as an interaction between the physical activity level score and gender. Both models use separate model variances for males and females.

Keywords: $\dot{V}O_2$max, MCMC, Bayesian Hierarchical Models, Bayes Methods

ACKNOWLEDGMENTS

I'm especially grateful to Dr. Fellingham for his willingness to let me join him in this research. Since the day I received this dataset, he has answered my questions and helped me resolve my concerns. I am also grateful for his determination to help me thoroughly understand the problem I was working on. I have become a much better statistician by working with him these past few months.

In addition, I appreciate Dr. Evan Johnson for his encouragement to enter the Master's program here at BYU in the first place. Also, my appreciation goes out to all the faculty who I have had the privilege to take classes with. I've learned so much and really solidified my statistical knowledge and understanding while studying at BYU as a masters student.

I'd also like to thank my wonderful family and friends for their optimism, encouragement, and examples. They all remind me that I can set and achieve high goals and do hard things. Even more importantly, however, they also remind me to keep ambition in check and find balance in life.

CONTENTS

CHAPTER 1

INTRODUCTION

Man's ability to run is a specialized and unique trait. Bramble and Lieberman (2004) pointed out 26 traits of the human body that make it especially adapted to run long distances and even more traits are revealed in subsequent studies (see, for example, Rolian et al., 2008; Lieberman et al., 2006). Although we pale in comparison to other animals with respect to short bursts of speed, we are at the top of the line going "slow and steady." In fact, humans are more capable of running long distances than almost any other animal, and could most likely beat a horse in running a marathon's distance on a hot day (Lieberman and Bramble, 2007).

Perhaps it is because of our innate ability to run that it has long been popular with humans. Each year, millions compete in foot races designed to test their bodies' aerobic prowess, measured largely by the resulting race time. In addition to running, however, there are many ways people practice athletics and develop endurance. Examples include swimming or bicycling. When comparing athletes across disciplines, say, a long-distance swimmer versus a long-distance runner, race times and speeds mean little. Other measurements must be used in order to objectively compare fitness levels of these individuals.

Maximal oxygen consumption ($\dot{V}O_2max$) is the standard measurement in determining overall cardiovascular and aerobic health. It is defined as the maximum capacity of an individual's body to transport and use oxygen during exercise of increasing intensity. Although very useful in determining aerobic fitness, $\dot{V}O_2max$ is quite difficult to collect, as it requires subjecting the patient to maximal exercise.

In this project we are interested in developing a Bayesian hierarchical model that accurately predicts $\dot{V}O_2max$ without demanding maximal exercise from the individual. We

examine both submaximal exercise data and questionnaire data of male and female adolescents ages 12 to 17. The questionnaire data represent an individual's personal opinions on how physically fit and able he or she truly is. We not only determine the best-fitting Bayesian hierarchical model, but also consider the efficacy of a Bayesian model that completely avoids any exercise data whatsoever.

In order to find the best models, we use model selection methods including Bayes χ^2 tests, Bayes factors, deviance information criteria, and Bayesian predictive information criteria. We will compare all proposed models using these methods to find which models best fit the data. Determining these best-fitting models will allow prediction of $\dot{V}\mathrm{O_2max}$ levels for any individual, regardless of physical limitations.

CHAPTER 2

LITERATURE REVIEW

This chapter discusses the background and development of the necessary components in performing this analysis. The chapter is divided into four sections: first, maximal oxygen consumption is discussed. Bayes' theorem follows, after which Markov chain Monte Carlo methods are developed. Lastly, various Bayesian model selection criteria are defined and discussed.

2.1 Maximal oxygen consumption

Maximal oxygen consumption ($\dot{V}\mathrm{O}_2\mathrm{max}$, where $\dot{V}$ represents volume, O_2 represents oxygen, and max represents maximum) is a measure of the maximum capacity of an individual's body to transport and use oxygen during exercise of increasing intensity (Astrand and Saltin, 1961; Saltin and Astrand, 1967). More formally, it is defined by the equation

$$\dot{V}\mathrm{O}_2\mathrm{max} = Q(\mathrm{CaO}_2 - \mathrm{CvO}_2),$$

where Q is the cardiac output of the heart,CaO_2 is arterial oxygen content, and CvO_2 is the venous oxygen content, all at a maximal exercise level. This equation is called the Fick Principle, named after Adolf Eugen Fick who developed it (see Fick, 1870).

$\dot{V}\mathrm{O}_2\mathrm{max}$ is considered by many to be the best measurement of overall cardiovascular fitness and maximal aerobic power (Hyde and Gengenbach, 2009). It is collected by subjecting an individual to increasingly demanding aerobic exercise until their body's oxygen consumption remains the same despite increased workload. This is deemed the maximal level. At this maximal level, the concentrations of oxygen in the inhaled air versus the exhaled air are used to calculate a numeric value. $\dot{V}\mathrm{O}_2\mathrm{max}$ is expressed in either liters of

oxygen per minute (l/min) or milliliters of oxygen per kilogram of bodyweight per minute (ml/kg/min).

Various factors have been attributed to varying $\dot{V}\text{O}_2\text{max}$ levels. Bouchard et al. (1999) found that $\dot{V}\text{O}_2\text{max}$ levels are highly determined by genetics. In a standardized training program they implemented among various families, some families experienced significant changes in $\dot{V}\text{O}_2\text{max}$ while others did not. Other significant factors, as demonstrated in this and other studies, include age and gender, as well as exercise level and its correlated factors (heart rate, perceived exertion rate, and so forth).

Although useful, $\dot{V}\text{O}_2\text{max}$ is also extremely difficult to directly calculate. Gathering accurate $\dot{V}\text{O}_2\text{max}$ measurements is expensive and time consuming, requiring trained personnel and appropriate equipment (Astrand and Ryhming, 1954; Coleman, 1976; Macfarlane, 2001; Metz and Alexander, 1971; Siconolfi et al., 1982). In addition, it is taxing on the individuals being tested, who must exercise until reaching a point of exhaustion (Montoye et al., 1986). This can especially be dangerous to individuals with poor health. It also poses a significant risk to younger individuals, who are inherently more vulnerable to difficulties from studies involving highly strenuous exercise.

To account for the difficulty of collection, various methods have been proposed to predict $\dot{V}\text{O}_2\text{max}$. Many of these methods predict $\dot{V}\text{O}_2\text{max}$ using submaximal exercise tests. These tests are advantageous because they are less strenuous on the individual being tested. Also, many times submaximal exercise tests require less of the expensive and complex equipment that is necessary for direct $\dot{V}\text{O}_2\text{max}$ measurements.

The heart rate ratio method (Uth et al., 2004) simply calculates $\dot{V}\text{O}_2\text{max}$ via the equation

$$\dot{V}\text{O}_2\text{max} = 15\frac{\text{HR}_{\text{max}}}{\text{HR}_{\text{rest}}},$$

where HR_{max} is the heart rate of an individual at maximal exercise, and HR_{rest} is the heart rate of the individual while at rest. This equation approximates $\dot{V}\text{O}_2\text{max}$ in ml/kg/min. Al-

though this test requires the individuals to reach maximal exercise, it is sometimes preferred because less equipment is necessary to measure heart rate.

The multi-stage fitness test (Léger et al., 1988) approximates $\dot{V}\text{O}_2\text{max}$ by having an individual run 20-meter shuttle runs with a speed synchronized according to beeps playing from a sound track. Each additional shuttle run requires a faster pace as the beeps increase in frequency. Eventually, the individual being tested can no longer keep up with the pace set by the beeps in the soundtrack. The highest successful level attained by the individual before failing is recorded and converted, depending on factors such as age and gender, into an approximate $\dot{V}\text{O}_2\text{max}$ level.

Personal technologies have also been developed in this area. Cosmed (2006) recently introduced its FitMax which, in addition to measuring resting metabolic rate and oxygen consumption in general (Nieman et al., 2006), has also been shown as effective in predicting $\dot{V}\text{O}_2\text{max}$ (Lee et al., 2011) using these other measurements. The device, tested on young to middle-age adults, was shown to closely correlate to actual $\dot{V}\text{O}_2\text{max}$ levels although it uses submaximal exercise data gathered from the patient.

2.2 Bayes' Theorem

All Bayesian methodology revolves around Bayes' theorem. The theorem is named after Reverend Thomas Bayes, who used the theorem in his work with the binomial distribution (Bayes and Price, 1763). Although LaPlace also independently formulated the theorem himself shortly after (see Stigler, 1983), Bayes is traditionally credited as its discoverer.

Bayes' theorem comes from a simple but useful manipulation of conditional probabilities. As

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \qquad \text{and} \qquad P(A|B) = \frac{P(A \cap B)}{P(B)},$$

then

$$P(A \cap B) = P(B|A)P(A) = P(A|B)P(B).$$

Because

$$P(B|A)P(A) = P(A|B)P(B),$$

dividing both sides by $P(A)$ we have Bayes' theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

By the law of total probability, this can be rewritten as

$$P(B|A) = \frac{P(A|B)P(B)}{\sum_i P(A|B_i)P(B_i)}.$$

Bayes' theorem is often extended beyond discrete sets and into continuous probability distribution functions (pdfs). For example, given a vector of parameters $\boldsymbol{\theta}$ and a set of random variables $\mathbf{Y}$, Bayes theorem for the pdf, f, can be written as the combination of functions $\mathcal{L}, \pi$, and h as

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{h(\mathbf{y})}.$$

Again, using the law of total probabilities, this can be rewritten as

$$f(\boldsymbol{\theta}|\mathbf{y}) = \frac{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int \cdots \int_{\boldsymbol{\theta}} \mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}.$$

Because of this result, Bayes' Theorem is very useful for determining behaviors of unknown parameters explaining datasets. For any given data $\mathbf{y}$, one can derive a distribution for each parameter in $\boldsymbol{\theta}$. In this context, referring to the above equation the quantity $f(\boldsymbol{\theta}|\mathbf{y})$ is called the *posterior* distribution. It is the presumed distribution of the random variable $\mathbf{Y}$ after accounting for the data $\mathbf{y}$. $\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})$ is referred to as the *likelihood* function, and $\pi(\boldsymbol{\theta})$ is called the *prior* distribution. The prior distribution must be appropriately set before performing the analysis according to previous knowledge concerning the problem that hints toward a distribution for each $\boldsymbol{\theta}$. $h(\mathbf{y})$ is the marginal distribution of $\mathbf{y}$, and for a given value of $\mathbf{y}$, $h(\mathbf{y})$ is fixed and equivalent to a constant, c. It is often referred to as the *normalizing constant*.

2.3 Markov chain Monte Carlo methods

When finding a posterior distribution, calculating the normalizing constant necessary to make it a pdf can be extremely difficult, and the posterior function itself is usually quite messy. Using Markov chain Monte Carlo (MCMC) methods, Nicholas Metropolis and others formulated an algorithm for generating draws from this posterior distribution without having to know its exact form or normalizing constant (Metropolis et al., 1953).

The Metropolis algorithm uses Markov chain theory. The Markov chain in this context takes a user's guess regarding a value contained in the posterior distribution, $\boldsymbol{\theta}_0$, and then "jumps" to a different value, $\boldsymbol{\theta}^*$. The value $\boldsymbol{\theta}_1$ will assume either the value of $\boldsymbol{\theta}_0$ or $\boldsymbol{\theta}^*$, largely depending on which value has a greater likelihood of being within the posterior distribution. The process is then repeated: a new $\boldsymbol{\theta}^*$ is generated and compared against $\boldsymbol{\theta}_1$. This iteration process continues until a user-specified n iterations have been performed. The value of n need not be overwhelmingly large, but it must be sufficient to allow enough draws to be generated in order to capture the true parameter space. Note that because this is a Markov chain, only the previous value, $\boldsymbol{\theta}_{i-1}$, is used in determining the next drawn value, $\boldsymbol{\theta}_i$.

The jumps performed during the algorithm are randomly generated from a jumping distribution, J. The generated jump value is added to the previous value for $\boldsymbol{\theta}$ thus creating a newly proposed value, $\boldsymbol{\theta}^*$. Although the Metropolis algorithm demands that the jumping distribution be symmetric, Hastings generalized the algorithm and relaxed this constraint (see Hastings 1970).

In order to determine whether or not the newly drawn value, $\boldsymbol{\theta}^*$, is accepted, a ratio, r, must be calculated, where

$$r = \frac{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)}{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta}_{i-1})\pi(\boldsymbol{\theta}_{i-1})}.$$

When $r \geq 1$, the new draw, $\boldsymbol{\theta}^*$, is accepted and assigned as $\boldsymbol{\theta}_i$. When $r < 1$, $\boldsymbol{\theta}^*$ is accepted with probability r. If rejected, then $\boldsymbol{\theta}_i$ is set to equal $\boldsymbol{\theta}_{i-1}$. The iteration value i then increases by 1, and the process is repeated.

A synopsis of the algorithm is given in the following steps. For $i = 1, 2, \ldots, n$:

1. Draw $\boldsymbol{\theta}^*$ from $J_i(\boldsymbol{\theta}^*|\boldsymbol{\theta}_{i-1})$

2. Compute the ratio $r = g(\boldsymbol{\theta}^*)/g(\boldsymbol{\theta}_{i-1})$, where $g(\cdot) \equiv \mathcal{L}(\mathbf{y}|\cdot)\pi(\cdot)$.

3. Set $\boldsymbol{\theta}_i = \begin{cases} \boldsymbol{\theta}^*, & \text{with probability } \min(r, 1) \\ \boldsymbol{\theta}_{i-1}, & \text{otherwise} \end{cases}$

When choosing the scale parameter for a jumping distribution, one must be wary of choosing a variance that is too large or too small. If the variance is too large, the algorithm will rarely accept values, as it will constantly choose values outside the posterior distribution's space. On the other hand, if the variance is too small, then it will take a long time to capture the entire parameter space. Though any variance will work, both would require considerably more time to achieve the best results. Roberts et al. (1997) determined that for random walk metropolis algorithms, the optimal jumping distribution variances yielded a parameter acceptance rate of 0.234.

Depending on the starting values assigned by the user, the algorithm may require a substantial number of iterations before finding and converging on the appropriate distribution. For this reason, a user-designated number of initial iterations, M, are discarded from the Markov chain. These discarded iterations are known as the *burn-in*. Despite the removal of these values, however, there is no guarantee of the algorithm's convergence.

Various methods have been designed to diagnose whether or not a chain has converged to the proper distribution. The simplest method is to simply look at the resulting trace plot and assess the chain's jumps. If it takes a considerable number of draws to span the parameter space (i.e. the trace plot is more linear than "bushy"), then this indicates lack of convergence. Figure 2.1 demonstrates examples of both good and bad trace plots.

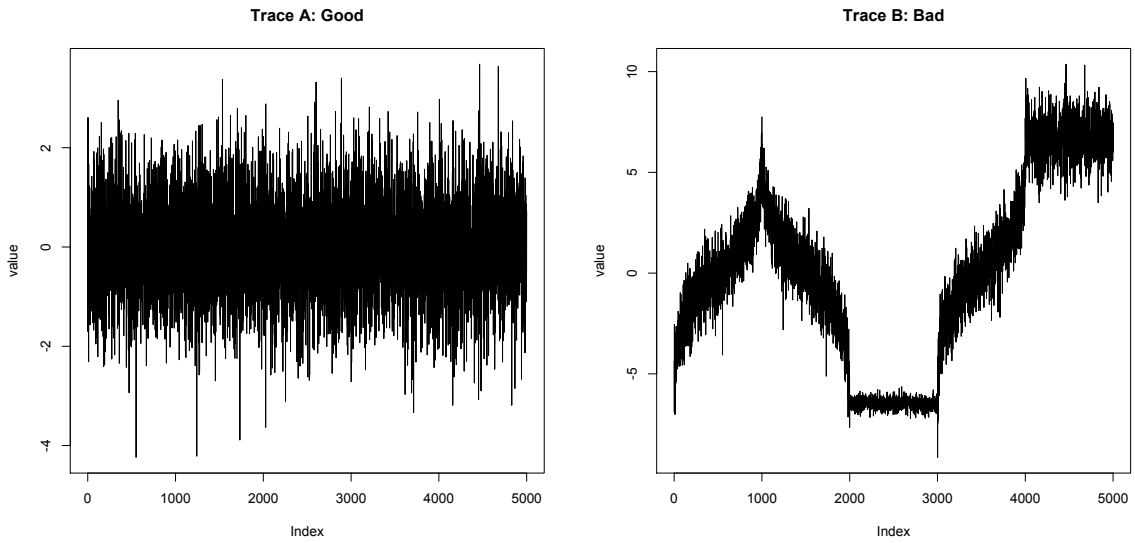

Figure 2.1: The contrast between good and bad trace plots. Trace plot A is jumping well, bushy, and seems to adequately capture the entire desired parameter space. Trace plot B, on the other hand, is struggling capture the parameter space, is not very bushy, and gets stuck at certain values.

The Gelman and Rubin diagnostic (Gelman and Rubin, 1992) compares the within-variance and between-variance of each individual parameter within two separate chains. A scale reduction factor, $\hat{R}$, is computed for each parameter by finding the ratio of a weighted combination of these two variances against the within-variance of a chain. A large value of $\hat{R}$ on any parameter indicates lack of convergence, and therefore one should continue iterating through the algorithm until the chains converge.

The Raftery and Lewis diagnostic (Raftery and Lewis, 1992) uses quantiles to assess chain convergence. The user inputs a posterior quantile, q, that he or she would like to estimate. In addition to this quantile, the user inputs a tolerance level, r, and a probability, s, associated with this quantile. The test returns a value, M, of the number of burn-in iterations necessary, as well as N, the number of iterations necessary after burn-in, required to attain the parameter value of $q \pm r$ with probability s. The Raftery and Lewis diagnostic also returns the dependence factor, denoted by I. When $I > 5$, this indicates either poor

initial values, high correlations, or poor jumping distribution selection. Ideally, a good Raftery and Lewis diagnostic will indicate a low I and a relatively low N.

2.4 Bayesian model selection

After proposing a model, it is almost always of interest to assess model fit, including proper prior specification. In addition, one might wish to compare two or more models and determine which single model has the best overall fit. Multiple methods have been proposed within the Bayesian paradigm to determine the adequacy of fit for a model. Some commonly used methods, along with their advantages and drawbacks, are presented here.

Bayesian χ^2 Goodness-of-Fit

Johnson (2004) proposed a Bayesian χ^2 goodness-of-fit test that implements MCMC draws in predicting model accuracy. Given a posterior distribution of parameters, $f(\boldsymbol{\theta}|\mathbf{y})$, let $\tilde{\boldsymbol{\theta}}$ represent a sample of draws from $f(\boldsymbol{\theta}|\mathbf{y})$. Also, set $a_0, a_1, a_2, \ldots, a_k$ such that $0 \equiv a_0 < a_1 < a_2 < \ldots < a_k \equiv 1$. Let $p_r = a_r - a_{r-1}$, where $r = 1, 2, \ldots, k$. Next, define

$$z_{jr}(\tilde{\boldsymbol{\theta}}) = \begin{cases} 1, & F(\mathbf{y}_j|\tilde{\boldsymbol{\theta}}) \in [a_{r-1}, a_r], \\ 0, & \text{otherwise} \end{cases},$$

where $F(\mathbf{y}_j|\tilde{\boldsymbol{\theta}})$ is the cumulative posterior distribution function for the random variables $\mathbf{Y}$ and n is the total number of observations, $\mathbf{y}$. Furthermore, let

$$m_r(\tilde{\boldsymbol{\theta}}) = \sum_{j=1}^{n} z_{jr}(\tilde{\boldsymbol{\theta}}),$$

or, in other words, the number of observations for which $F(\mathbf{y}_j|\tilde{\boldsymbol{\theta}})$ is in the rth bin. Now, let

$$R^B(\tilde{\boldsymbol{\theta}}) = \sum_{r=1}^{k} \left(\frac{m_r(\tilde{\boldsymbol{\theta}}) - np_r}{\sqrt{np_r}} \right)^2.$$

Then

$$R^B(\tilde{\boldsymbol{\theta}}) \sim \chi^2_{k-1}.$$

When choosing the number of bins, k, and each cell's respective probabilities, p_r, it is important to choose well. Too many bins results in a significant loss of power (Koehler and Gan, 1990), and false cell probabilities can adversely affect test results. Mann and Wald (1942) proposed a general method for choosing cell count while suggesting equally probable cells. This method, however, generally created too many bins. Koehler and Gan in their 1990 paper also addressed this issue and proposed the currently accepted method. They propose that all cells be equally probable, as Mann and Wald did, and suggest the number of bins to be approximately equal to $n^{0.4}$, which has provided near optimal results on various datasets.

This Bayesian χ^2 goodness-of-fit test is very useful in the sense that it is highly computationally efficient. The R^B values can be calculated quite simply once the MCMC posterior draws are generated.

One drawback for the Bayesian χ^2 analysis is that the distribution of $R^B(\tilde{\boldsymbol{\theta}})$ only provides a general idea of how well the model is performing. A more tedious and time-consuming method, as shown in Johnson's paper, is required when seeking a formal significance test value.

Bayes Factors

Bayes factors compare the ratio between the posterior and prior odds of the models. They are mainly attributed to Jeffreys (1935; 1961) and an in-depth analysis of Bayes factors is presented by Kass and Raftery (1995). The following derivation is quite similar to the explanation as given by Carlin and Louis (2009).

Let M_1 represent one possible model for a given dataset, $\mathbf{y}$, and let M_2 represent another possible model. Also, let $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$ represent the varying parameter values for M_1 and M_2, respectively. Then the marginal distributions of $\mathbf{Y}$ given either model are given by

$$g(\mathbf{y}|M_i) = \int \mathcal{L}(\mathbf{y}|\boldsymbol{\theta}_i, M_i)\pi_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i, \quad i = 1, 2.$$

By Bayes' theorem,

$$P(M_i|\mathbf{y}) = \frac{g(\mathbf{y}|M_i)P(M_i)}{P(\mathbf{y})}, \quad i = 1, 2.$$

The prior odds of M_1 against M_2 is equal to $P(M_1)/P(M_2)$. The posterior odds of model 1 against model 2 is $P(M_1|\mathbf{y})/P(M_2|\mathbf{y})$.

The Bayes factor, denoted BF, is defined as the ratio between these two odds. Therefore,

$$\begin{aligned} BF &= \frac{P(M_1|\mathbf{y})/P(M_2|\mathbf{y})}{P(M_1)/P(M_2)} \\ &= \frac{\left(\frac{g(\mathbf{y}|M_1)P(M_1)}{P(\mathbf{y})}\right) / \left(\frac{g(\mathbf{y}|M_2)P(M_2)}{P(\mathbf{y})}\right)}{P(M_1)/P(M_2)} \\ &= \frac{(g(\mathbf{y}|M_1)P(M_1))/(g(\mathbf{y}|M_2)P(M_2))}{P(M_1)/P(M_2)} \\ &= \frac{g(\mathbf{y}|M_1)}{g(\mathbf{y}|M_2)}. \end{aligned}$$

In determining model significance, the ratio presented above should be arranged such that the BF value is greater than 1 (in other words, it may be sometimes necessary to put $g(\mathbf{y}|M_2)$ in the numerator and $g(\mathbf{y}|M_1)$ in the denominator). The value of this ratio provides key evidence as to which model is better. Table 2.1 shows Jeffreys' criteria for determining the amount of evidence supporting the numerator model over the denominator model.

Table 2.1: The values of BF, the ratio between models, used for determining model significance. As BF increases, the stronger the evidence suggests that the numerator model is better than the denominator model.

BF	$\log_e(BF)$,	Evidence
1 to 3	0 to 1.1	Insignificant
3 to 10	1.1 to 2.3	Substantial
10 to 30	2.3 to 3.4	Strong
30 to 100	3.4 to 4.6	Very Strong
> 100	> 4.6	Decisive

Although this example presents only two models, it can easily be extended further. When comparing more than two models, one can simply take any two models and calculate

the better of these two using the methodology above. Once the better model has been determined, then this better model is then compared against the next model. The process is once again repeated, with the best model chosen in each case, until all proposed models have been evaluated.

While Bayes factors are useful, they also have their drawbacks. Deriving the posterior odds can be quite time-consuming. Furthermore, their calculations require a lot of computational time. Therefore, with numerous models this becomes quite impractical. Others suggest the use of the much simpler Schwartz criterion (also known as the Bayesian information criterion, or BIC) as a rough approximation to the logarithm of the Bayes factor (Kass and Raftery, 1995). Theoretically, Bayes factors also attach very little weight to the correct model as n gets large (Gelfand and Dey, 1994; see also Lindley, 1957, Bartlett, 1957), thus giving even more restrictions to their use. In addition, Bayes factors are uninterpretable with any improper prior distributions set on parameters because the marginal distribution of the data will also be improper.

Deviance Information Criterion

In addition to these two methods, a third method, the deviance information criterion (DIC) (Spiegelhalter, 2002), is commonly used. DIC takes both model fit and model complexity into account when comparing model. The fit is measured according to the model's *deviance*, defined as

$$D(\boldsymbol{\theta}) = -2\log(\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})).$$

$\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})$ represents the likelihood function with respect to the data. In the MCMC sense, $D(\boldsymbol{\theta})$ is calculated via the samples of $\boldsymbol{\theta}$ generated by simulation. The model's overall fit, $\bar{D}$, is then found, where

$$\bar{D} = E_{\boldsymbol{\theta}}[D(\boldsymbol{\theta})].$$

The lower the value of $\bar{D}$, the better the fit.

As with many information criteria, DIC also penalizes according to the number of effective parameters the model is using. The effective number of parameters, p_D, is found via the equation

$$p_D = \bar{D} - D(E[\boldsymbol{\theta}]).$$

From here, the DIC value is found, where

$$\text{DIC} = p_D + \bar{D}.$$

Smaller DIC values indicate better models. Thus, the DIC favors the best-fitting models with fewest parameters.

Although the DIC is widely used in assessing hierarchical Bayesian models, as with all other fitting methods discussed, it has its flaws. Mainly, the DIC parameter overfits the observed data, $\mathbf{y}$, in determining appropriate models, thus tending to select over-fitted models and biasing the results (Robert and Titterington, 2002).

Bayesian Predictive Information Criterion

The Bayesian predictive information criterion (BPIC) (Ando, 2007) is another method similar to DIC used to measure model fit. It was developed as a response to the inherent bias contained in DIC estimates and accounts for this issue. In order to define it, we first let

$$\hat{b}_{\boldsymbol{\theta}} \approx E_{\boldsymbol{\theta}|\mathbf{y}}[\log\{\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\}] - \log\{\mathcal{L}(\mathbf{y}|\hat{\boldsymbol{\theta}}_n)\pi(\hat{\boldsymbol{\theta}}_n)\} + \text{tr}\{J_n^{-1}(\hat{\boldsymbol{\theta}}_n)I_n(\hat{\boldsymbol{\theta}}_n)\} + p/2$$

where the approximation turns to an equality as n approaches ∞. Here, p is the dimension of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}_n$ is the posterior mode of $f(\boldsymbol{\theta}|\mathbf{y})$, and $I_n(\hat{\boldsymbol{\theta}}_n)$ and $J_n(\hat{\boldsymbol{\theta}}_n)$ are given by

$$I_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial\eta_n(\mathbf{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\frac{\partial\eta_n(\mathbf{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^T}\right\}, \quad J_n(\boldsymbol{\theta}) = -\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial^2\eta_n(\mathbf{y}_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^T}\right\}.$$

Here, $\eta_n(\mathbf{y}_i,\boldsymbol{\theta}) = \log\{\mathcal{L}(\mathbf{y}_i|\boldsymbol{\theta})\} + \log\{\pi(\boldsymbol{\theta})\}/n$. All other notation is preserved. Also, as with the DIC, let

$$D(\boldsymbol{\theta}) = -2\log(\mathcal{L}(\mathbf{y}|\boldsymbol{\theta}))$$

and

$$\bar{D} = E_{\boldsymbol{\theta}}[D(\boldsymbol{\theta})].$$

The BPIC is calculated as

$$\text{BPIC} = \bar{D} + 2n\hat{b}_{\boldsymbol{\theta}}.$$

The best fitting model minimizes the BPIC.

As mentioned previously, the BPIC does a good job at minimizing selection bias. In addition, unlike the DIC, it is not necessary to assume that the specified family of probability distributions contains the true model. However, as is apparent from the methodology leading to its calculation, it is much less straightforward than the DIC, and takes a significant amount of time to calculate. Also, as with other procedures, BPIC only works with independent, unimodal data.

CHAPTER 3

METHODS

$\dot{V}O_2$max levels were measured for 132 youth, male and female, ages 12-17, during a maximal exercise test. 20 of these individuals were tested again on a separate day for reliability purposes. In addition to measuring $\dot{V}O_2$max, these individuals also performed various submaximal tests at a difficulty they requested, and the results were recorded. The individuals also participated in two separate questionnaires which asked them to rate themselves on their own level of perceived functional abilities and physical activities. Other standard variables of interest were also recorded for each patient. A comprehensive list of these variables is given in Table 3.1.

Due to the difficulty of collecting maximal exercise data, it is of interest to find a reliable model to predict $\dot{V}O_2$max levels in individuals using data that is easier to obtain. An accurate model that uses only submaximal exercise data, along with a few other variables, would have great value. In addition, a comparable model that avoids exercise data altogether would be helpful for patients for whom exercise is difficult. In this work, we use MCMC methods to find the best-fitting overall Bayesian model using submaximal exercise data, as well as the best model that uses only questionnaire data.

To find the best models, we implemented a backward model selection procedure. We started with two full models, one using submaximal exercise data and the other questionnaire data. These full models contained all variables and interactions that we thought might be significant. In each of these models, the $\dot{V}O_2$max levels are assumed to be normally distributed, where

$$\dot{V}O_2\text{max} \sim N(\theta_{ij}, \sigma^2_{err_g}).$$

Table 3.1: The variables included in the dataset, used for determining the most appropriate model for determining $\dot{V}O_2$max levels. Different variables are present in different proposed models; some variables (including *Type* and all non-$\dot{V}O_2$max maximal exercise variables) are not present in any.

Variable	Description
ID	An identifier of each individual.
Gender	An indicator of gender (0 = female, 1 = male).
Age	The age of the individual.
Height	The height of the individual (in centimeters).
Weight	The weight of the individual (in kilograms).
BMI	The individual's body mass index.
SUMPFA	A numerical value from 0 to 26 indicating the individual's own perceived functional ability according to a questionnaire. 0 is the lowest, meaning very little ability, whereas 26 is the highest.
PAR	A numerical value from 0 to 10 indicating the individual's score on a questionnaire indicative of how physically active the individual is. 0 is the lowest, meaning no physical activity, whereas 10 is the highest.
Type	Indicates the type of exercise this individual does on a regular basis (W = walker, J = jogger, R = runner).
Day	Indicates whether the results were a result of the first test or the second test, performed on a different day (1 or 2).
MPH	Treadmill speed (in miles per hour) during submaximal exercise.
HR	Heart rate during submaximal exercise (in beats per minute).
RPE	Rating of perceived exertion (RPE) during submaximal exercise. The RPE scale is commonly used for determining exercise intensity levels.
MaxG	Treadmill grade during maximal exercise.
MaxHR	Heart rate during maximal exercise.
MaxRER	Ratio of volumes of expired carbon dioxide over consumed oxygen at maximal exercise.
MaxRPE	Rating of perceived exertion (RPE) during maximal exercise.
MaxVO2L	$\dot{V}O_2$max, measured in L/min.
MaxVO2ML	$\dot{V}O_2$max , measured in ml/kg/min.

Subscript i stands for the ith individual in the study, subscript j stands for the jth observation of the ith individual, subscript g stands for gender, and the $\dot{V}O_2$max levels are measured in ml/kg/min. For the submaximal model, θ_{ij} is defined as the linear combination

$$\begin{aligned}\theta_{ij} &= \beta_{0_i} + \beta_A \text{age}_{ij} + \beta_G \text{gender}_i + \beta_B \text{BMI}_{ij} + \beta_M \text{MPH}_{ij} + \beta_{HR} \text{HR}_{ij} + \beta_R \text{RPE} + \\ &\quad \beta_{AG} \text{age}_{ij} \cdot \text{gender}_i + \beta_{AB} \text{age}_{ij} \cdot \text{BMI}_{ij} + \beta_{AM} \text{age}_{ij} \cdot \text{MPH}_{ij} + \beta_{AH} \text{age}_{ij} \cdot \text{HR}_{ij} + \\ &\quad \beta_{AR} \text{age}_{ij} \cdot \text{RPE}_{ij} + \beta_{GB} \text{gender}_i \cdot \text{BMI}_{ij} + \beta_{GM} \text{gender}_i \cdot \text{MPH}_{ij} + \\ &\quad \beta_{GH} \text{gender}_i \cdot \text{HR}_{ij} + \beta_{GR} \text{gender}_i \cdot \text{RPE}_{ij}.\end{aligned}$$

For the questionnaire model, θ_{ij} is defined as the linear combination

$$\begin{aligned}\theta_{ij} &= \beta_{0_i} + \beta_A \text{age}_{ij} + \beta_G \text{gender}_i + \beta_B \text{BMI}_{ij} + \beta_S \text{SUMPFA}_{ij} + \beta_P \text{PAR}_{ij} + \\ &\quad \beta_{AG} \text{age}_{ij} \cdot \text{gender}_i + \beta_{AB} \text{age}_{ij} \cdot \text{BMI}_{ij} + \beta_{AS} \text{age}_{ij} \cdot \text{SUMPFA}_{ij} + \\ &\quad \beta_{AP} \text{age}_{ij} \cdot \text{PAR}_{ij} + \beta_{GB} \text{gender}_i \cdot \text{BMI}_{ij} + \beta_{GS} \text{gender}_i \cdot \text{SUMPFA}_{ij} + \\ &\quad \beta_{GP} \text{gender}_i \cdot \text{PAR}_{ij}.\end{aligned}$$

The priors (and, if applicable, hyperpriors) for each parameter, held constant in both models, are included in Table 3.2. Note that, within the table, all coefficients, with the exception of the intercept term, are represented by the symbol β_i. Also notice that both models include fairly non-informative priors.

After finding the best two models within each category using DIC, Bayes Factors and Bayesian χ^2 Goodness-of-Fit tests were calculated to further determine model adequacy and help decide the best-fitting model in each category.

Table 3.2: Priors and hyperpriors for the full models. For all normally distributed parameters, the two values represent the parameter mean and variance, respectively. Also, β_i represent all coefficients, with the exception of the intercept, as presented in either model.

Priors		Hyperpriors	
$\sigma^2_{err_g}$	$\sim Gam(1.5, \text{scale} = 40)$	(none)	
β_{0_i}	$\sim N\left(\beta_o, \sigma^2_{subj}\right)$	β_o σ^2_{subj}	$\sim N(0, 10000)$ $\sim Gam(1.5, \text{scale} = 20)$
β_i	$\sim N\left(0, 100\right)$	(none)	

CHAPTER 4

RESULTS

The submaximal model was fitted as specified in the previous chapter, and then the model parameters were thinned by removing insignificant parameters and taking into account the DIC. Also in determining adequate models, model simplicity was favored over including parameters with only slight gains. A log of this thinning procedure is given in Table 4.1. The results of these runs indicated that the majority of the interactions, with the exception of the age·HR and gender·HR interactions, didn't add to the model; however, all the main effects were significant. This thinning procedure motivated us to keep the following two models for further analysis: a model including main effects and an age·HR interaction, as well as a model including main effects, an age·HR interaction, and a gender·HR interaction.

The same model selection procedures used above were again performed on the proposed questionnaire models. Once again, in determining the models, model simplicity was of primary concern. A log of this thinning procedure is given in Table 4.2. Just as with the submaximal model, these results also suggested that all main effects were significant. Only one interaction, however, gender·PAR, hinted at being significant. As such, two models were selected as being the best: the main effects model and a model including the main effects and the gender·PAR interaction.

Table 4.1: Summary of backward model selection procedures to find the best two submaximal models. Models are shown in order they were formulated.

Model Parameters	DIC	Summary and Decision
full model	721.2	Age·RPE, gender·BMI, gender·MPH, and gender·RPE all centered around 0. Began by dropping gender·RPE.
main effects, age·gender, age·BMI, age·MPH, age·HR, age·RPE, gender·BMI, gender·MPH, gender·HR	714.7	Simpler model, lower DIC. Kept model. Age·RPE, gender·BMI, and gender·HR all centered around 0. Dropped age·RPE.
main effects, age·gender, age·BMI, age·MPH, age·HR, gender·BMI, gender·MPH, gender·HR	704.2	Simpler model, lower DIC. Kept model. Gender·BMI and gender·HR both centered around 0. Dropped gender·HR.
main effects, age·gender, age·BMI, age·MPH, age·HR, gender·BMI, gender·MPH	708.2	Slightly higher DIC, but simplicity favored. Kept model. Gender·BMI centered around 0, dropped from model.
main effects, age·gender, age·BMI, age·MPH, age·HR, gender·MPH	706.1	Simpler model, lower DIC. Kept model. Age·gender, age·BMI, age·MPH, and gender·MPH all contain 0 within probability intervals. Dropped age·gender.
main effects, age·BMI, age·MPH, age·HR, gender·MPH	705.3	Simpler model, lower DIC. Kept model. Age·MPH and gender·MPH both contain 0 within probability intervals. Dropped age·MPH.
main effects, age·BMI, age·HR, gender·MPH	706.7	Slightly higher DIC, but favored simplicity and kept model. Gender·MPH contains 0 within probability interval. Dropped in subsequent model.
main effects, age·BMI, age·HR	707.4	DIC slightly greater, but again favored simplicity and kept model. Dropped age·BMI which contains 0 within probability interval.
main effects, age·HR	699.2	Low DIC, quite simple. Kept model. Dropped final interaction to see how model does with just main effects.
main effects	734.1	Significant DIC increase. Kept previous model.
main effects, age·HR, gender·HR	702.2	Included as a check.
main effects, gender·HR	730.2	Significant DIC increase. Appears that age·HR is significant, whereas gender·HR is not. Keep age·HR interaction.

Table 4.2: Summary of backward model selection procedures to find the best two questionnaire models. Models are shown in order they were formulated.

Model Parameters	DIC	Summary and Decision
full model	749.3	Age·PFA, age·PAR, gender·BMI, and gender·PFA posterior distributions all centered around 0. Began by dropping gender·PFA.
main effects, age·gender, age·BMI, age·PFA, age·PAR, gender·BMI, gender·PAR	753.2	DIC only slightly higher, and model more simple. Kept this model. Gender·BMI centered around 0. Dropped this variable.
main effects, age·gender, age·BMI, age·PFA, age·PAR, gender·PAR	747.6	DIC better than previous two models. Kept this model. Age·PFA, age·PAR centered around 0. Dropped age·PFA.
main effects, age·gender, age·BMI, age·PAR, gender·PAR	747.6	Same DIC, simpler model. Kept this model. Age·PAR, gender·PAR contain 0 within the probability interval. Dropped gender·PAR.
main effects, age·gender, age·BMI, age·PAR	749.0	Again DIC is only slightly higher. Favored more simple model, kept this one. Age·PAR contains 0 within its probability interval. Dropped Age·PAR.
main effects, age·gender, age·BMI	750.7	DIC slightly higher, but not enough to outweigh desire for simplicity. Model kept. Age·gender contains 0 in probability interval and consequently dropped.
main effects, age·BMI	748.9	Lower DIC, simpler model. Model kept.Final interaction dropped to see if data can be adequately modeled with only main effects.
main effects	745.0	Lowest DIC yet–keep this model. Age was suggested as being insignificant, and thus tested next.
main effects minus age	754.6	High DIC. Keep age within the model.
main effects, gender·PAR	746.8	Tested to be sure that it wasn't significantly helping model. Higher DIC and more complicated. Keep main effects model.

Both models in each category were then assessed using a Bayesian χ^2 goodness-of-fit test. A plot of each model's density distribution, along with the actual χ^2 distribution, is given in Figure 4.1. The models from both the submaximal dataset as well as the questionnaire dataset performed quite well under this analysis. However, some models seem to fit better than others. Between the questionnaire models, the model that includes the interaction effect seems to more adequately fit the actual χ^2 distribution than the model with main effects only. This result contradicts the results obtained from the DIC. Among the submaximal models, it seems that the single interaction model performs better than the model with two interactions, supporting the DIC.

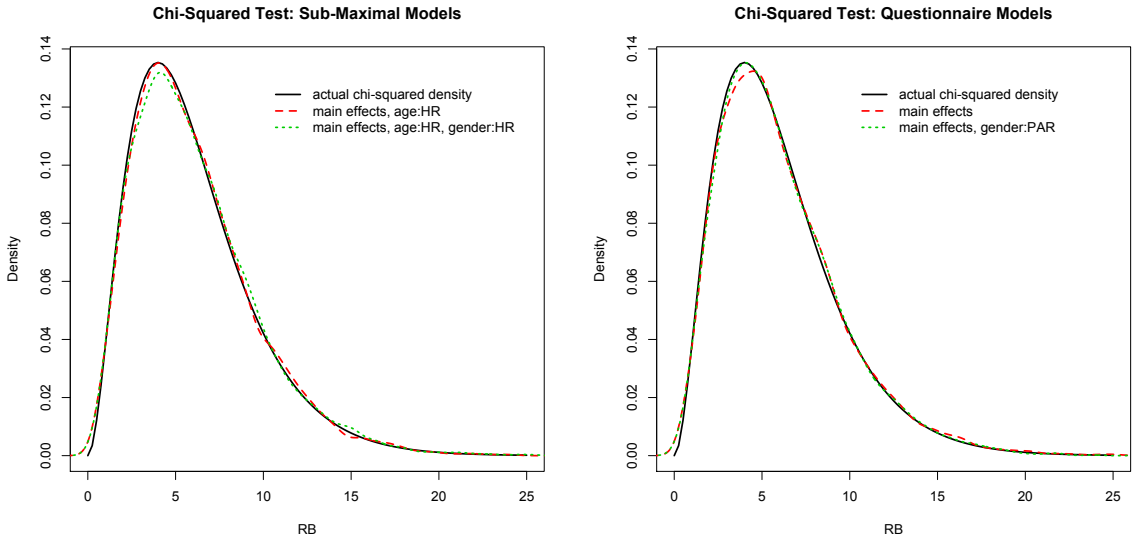

Figure 4.1: A comparison of model fits using the Bayesian χ^2 Goodness-of-Fit test. As is shown, all models seem to perform quite well under this analysis, indicating adequate model fit. However, the models that include a single interaction within each category seem to fit better than their counterparts in either case.

After testing χ^2 Goodness-of-Fit, Bayes factors were calculated between each model using the Laplace-Metropolis method (Lewis and Raftery, 1994). This method takes a second order Taylor series expansion of

$$g(\mathbf{y}|M_i) = \int \mathcal{L}(\mathbf{y}|\boldsymbol{\theta}_i, M_i)\pi_i(\boldsymbol{\theta}_i)d\boldsymbol{\theta}_i$$

and approximates it as

$$(2\pi)^{p_i/2}|\tilde{\Sigma}|_i^{1/2}\mathcal{L}(\mathbf{y}|\boldsymbol{\theta}_i, M_i)\pi_i(\boldsymbol{\theta}_i),$$

where p_i equals the number of parameters in the ith model. $|\tilde{\Sigma}|_i^{1/2}$ is approximately equal to the posterior variance matrix of $\boldsymbol{\theta}_i$. The log conditional probabilities of the data given each respective model are presented in Table 4.3, along with an overall ranking of model fit performance. These values suggest that, within the submaximal models, the model with only a single interaction term is best. Within the questionnaire models, the Bayes factor suggests using the model that has main effects only. Unlike the DIC, the Bayes factors also suggest that the submaximal models might actually be performing worse than the questionnaire models.

Table 4.3: A comparison of models according to Bayes factor analysis. The log conditional probabilities of each model are provided (higher is better). Also, recall that a difference of about 2.3 or more suggests strong evidence that the model with the higher log conditional probability is better than the other. These Bayes factors suggest that the questionnaire models perform better than the submaximal models. Also, the submaximal model with a single interaction performed best within the submaximal model category. Within the questionnaire model category, the main effects model performed best.

Model	$\log(g(\mathbf{y}\|M_i))$	Rank
Submax main effects, age:HR	-344.42	3
Submax main effects, age:HR, gender:HR	-349.19	4
Questionnaire main effects	-330.71	1
Questionnaire main effects, gender:PAR	-333.78	2

Although the questionnaire model choice is debatable, it seems the best models for each categories are the models that include the main effects and a single interaction term (age:HR for the submaximal model and gender:PAR for the questionnaire model). The estimated coefficients and their probability intervals for both the submaximal and questionnaire models, respectively, are given in Tables 4.4 and 4.5.

We collected posterior model estimates and prediction intervals from each model according to individual and also with respect to gender. These were all created assuming the mean values of each variable estimate. For females, age=14.5, BMI=21, PFA=14.5, PAR=6,

Table 4.4: Parameter Estimates for the coefficients of the submaximal model with 95% quantile and highest posterior density (HPD) intervals.

Parameter	Mean	95% PI (quantile)					95% PI (HPD)				
β_o	-71.102	-154.4	$\leq$	β_o	$\leq$	1.661	-157.4	$\leq$	β_o	$\leq$	-3.566
$\sigma^2_{err_F}$	2.687	0.558	$\leq$	$\sigma^2_{err_F}$	$\leq$	8.461	0.311	$\leq$	$\sigma^2_{err_F}$	$\leq$	6.849
$\sigma^2_{err_M}$	9.103	2.489	$\leq$	$\sigma^2_{err_M}$	$\leq$	22.662	1.758	$\leq$	$\sigma^2_{err_M}$	$\leq$	19.910
σ^2_{subj}	23.876	15.510	$\leq$	σ^2_{subj}	$\leq$	32.790	15.39	$\leq$	σ^2_{subj}	$\leq$	32.62
β_A	8.920	3.601	$\leq$	β_A	$\leq$	14.810	3.771	$\leq$	β_A	$\leq$	14.920
β_G	6.230	4.143	$\leq$	β_G	$\leq$	8.303	4.205	$\leq$	β_G	$\leq$	8.340
β_B	-0.873	-1.151	$\leq$	β_B	$\leq$	-0.586	-1.169	$\leq$	β_B	$\leq$	-0.607
β_M	4.984	3.421	$\leq$	β_M	$\leq$	6.579	3.382	$\leq$	β_M	$\leq$	6.512
β_H	0.658	0.195	$\leq$	β_H	$\leq$	1.174	0.215	$\leq$	β_H	$\leq$	1.184
β_R	-0.510	-0.998	$\leq$	β_R	$\leq$	-0.035	-0.996	$\leq$	β_R	$\leq$	-0.034
β_{AH}	-0.051	-0.088	$\leq$	β_{AH}	$\leq$	-0.019	-0.089	$\leq$	β_{AH}	$\leq$	-0.020

Table 4.5: Parameter Estimates for the coefficients of the questionnaire model with 95% quantile and highest posterior density (HPD) intervals.

Parameter	Mean	95% PI (quantile)					95% PI (HPD)				
β_o	37.834	27.850	$\leq$	β_o	$\leq$	46.540	28.5	$\leq$	β_o	$\leq$	47.02
$\sigma^2_{err_F}$	2.823	0.672	$\leq$	$\sigma^2_{err_F}$	$\leq$	8.464	0.435	$\leq$	$\sigma^2_{err_F}$	$\leq$	6.871
$\sigma^2_{err_M}$	15.593	5.094	$\leq$	$\sigma^2_{err_M}$	$\leq$	31.582	3.453	$\leq$	$\sigma^2_{err_M}$	$\leq$	28.800
σ^2_{subj}	19.270	11.15	$\leq$	σ^2_{subj}	$\leq$	28.53	11.12	$\leq$	σ^2_{subj}	$\leq$	28.41
β_P	.574	0.118	$\leq$	β_P	$\leq$	1.042	0.1242	$\leq$	β_P	$\leq$	1.045
β_G	4.660	-0.422	$\leq$	β_G	$\leq$	9.721	-0.438	$\leq$	β_G	$\leq$	9.699
β_B	-0.794	-1.062	$\leq$	β_B	$\leq$	-0.523	-1.057	$\leq$	β_B	$\leq$	-0.519
β_S	0.536	0.327	$\leq$	β_S	$\leq$	0.748	0.323	$\leq$	β_S	$\leq$	0.744
β_A	0.796	0.240	$\leq$	β_A	$\leq$	1.374	0.209	$\leq$	β_A	$\leq$	1.340
β_{GP}	0.293	-0.430	$\leq$	β_{GP}	$\leq$	1.008	-0.432	$\leq$	β_{GP}	$\leq$	1.005

MPH=5, HR=166, and RPE=13. For males, age=14, BMI=20.5, PFA=18, PAR=7, MPH=5.5, HR=160, and RPE=12. These plots are shown in Figures 4.2 and 4.3. Figure 4.2 groups the plots together according to model, and Figure 4.3 groups the plots together according to gender.

The model and prediction interval plots indicate that the chosen questionnaire and submaximal models are barely distinguishable. Both models perform quite accurately, and

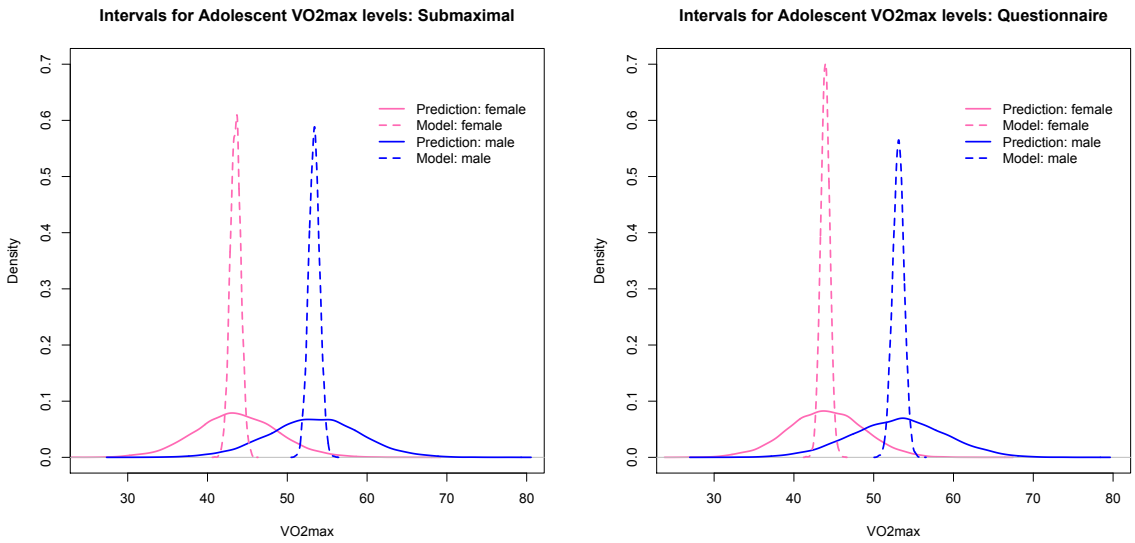

Figure 4.2: Model and prediction intervals for $\dot{V}O_2$max levels using final models, grouped by model. Shown in the plots are both the male and female model and prediction intervals, along with the overall prediction interval. These plots are created based off of the mean values for each variable for both male and female.

the questionnaire model interval for the females is actually narrower than the submaximal model interval for the females (see Table 4.6). This suggests that, for females, the questionnaire model provides more accurate results.

Table 4.6: 95% quantile and highest posterior density (HPD) model intervals. In this case, both intervals are practically identical for each model. The intervals for females were more narrow than the intervals for males, and the most narrow interval was the within the questionnaire model.

		95% PI (quantile)			95% PI (HPD)		
Model	Mean	Lower	Upper	Width	Lower	Upper	Width
Questionnaire: Female	43.93	42.79	45.07	2.28	42.77	45.04	2.27
Questionnaire: Male	53.09	51.70	54.48	2.78	51.71	54.49	2.78
Submaximal: Female	43.51	42.23	44.79	2.55	42.18	44.72	2.54
Submaximal: Male	53.37	51.99	54.72	2.73	52.00	54.72	2.73

We also assessed the residuals for both plots. These residuals, grouped according to gender as well as number of recorded observations for the individual, are given in Figure 4.4. In addition to the model diagnostics used previously, these residual plots confirm the separate

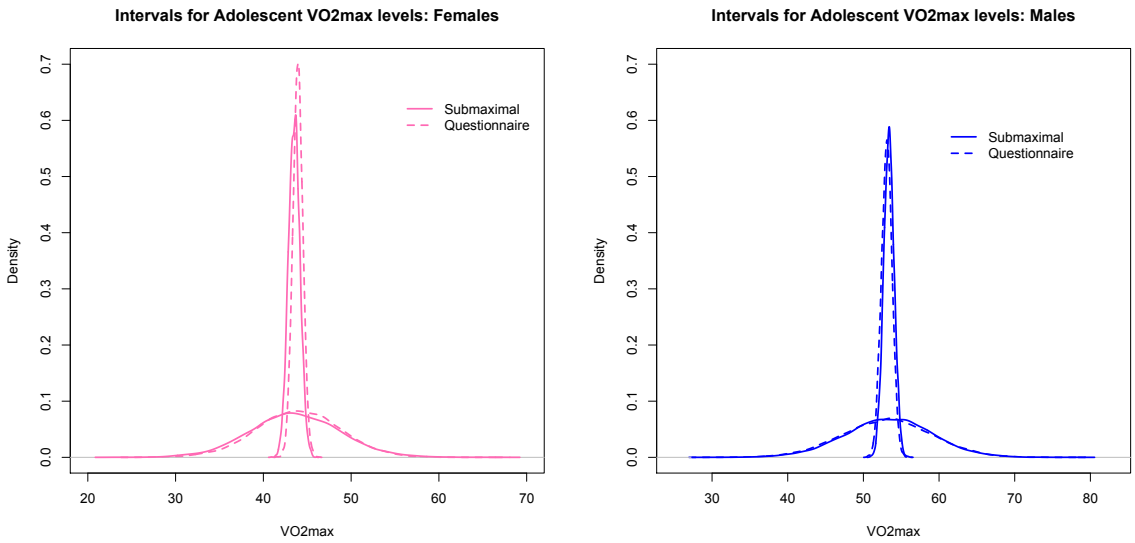

Figure 4.3: Model and prediction intervals for $\dot{V}O_2$max levels using final models, grouped by gender. According to these model and prediction intervals, the two models are practically interchangeable.

variances between male and female. Also, the individuals with multiple observations have tighter margins on their residual intervals, which is to be expected. It is also interesting to note from these residual plots that the width of the residual densities is smaller for males in the submaximal model than for the questionnaire model. This visual interpretation is verified by 95% highest posterior density interval calculations, presented in Table 4.7. From this information, in conjunction from what is apparent in Figure 4.3, the questionnaire model performs best for female adolescents, whereas the submaximal model performs best for male adolescents.

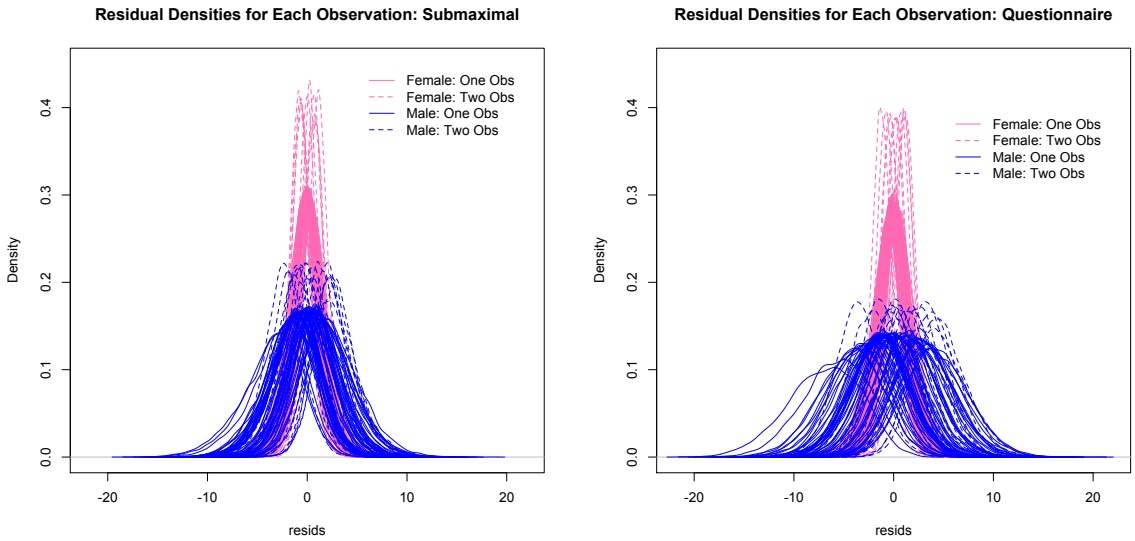

Figure 4.4: Density plots for each observational residual within each model. Through these plots it is apparent that two separate variances are required for males and females. The residuals created from individuals with more than one observation are also distinguished. As expected, the residual densities of the subjects with multiple observations are more narrow than the subjects with just a single observation.

Table 4.7: Average interval width for each type of residual, as shown in Figure 4.4, under both the submaximal and questionnaire models. These widths were created from 95% highest posterior density intervals.

Residual Type	Avg. Width
Questionnaire: Females, 2 obs	4.652
Questionnaire: Females, 1 obs	6.259
Questionnaire: Males, 2 obs	9.561
Questionnaire: Males, 1 obs	11.704
Submaximal: Females, 2 obs	4.623
Submaximal: Females, 1 obs	6.234
Submaximal: Males, 2 obs	8.245
Submaximal: Males, 1 obs	10.172

CHAPTER 5

DISCUSSION AND CONCLUSIONS

For youth whose $\dot{V}O_2max$ levels are either not possible or unsafe to collect, two separate models provide an alternative for adequately predicting this value. The first model is based on submaximal exercise data and uses age, gender, BMI, submaximal heart rate, submaximal treadmill MPH, and rate of perceived exertion during treadmill submaximal exercise. The main effects of these variables, as well as an interaction between age and heart rate, are all used to predict $\dot{V}O_2max$.

Where performing submaximal exercise tests is either not possible or inconvenient, another model using questionnaire data can help to adequately predict $\dot{V}O_2max$. In addition to age, gender, and BMI, this model also implements a physical activity score and a functional ability score as perceived by the subject. The main effects of these variables, as well as an interaction between gender and perceived physical activity, all contribute to predicting $\dot{V}O_2max$ within this model.

Both the submaximal and questionnaire models were implemented using two separate variances for male versus female individuals. In addition, both models include separate intercept priors for each individual and hyperpriors to determine an overall intercept and between subject variance.

It is interesting to note the similarity between these two models. Although one might assume that the submaximal model should perform better, the results show that the questionnaire model had similar predictive value (see Figure 4.3). In fact, the questionnaire model had an even tighter model interval for females than the submaximal model. The submaximal model, however, performed better in residual tests than the questionnaire model

for the males. Despite these differences, researchers can predict $\dot{V}O_2$max levels using either questionnaires or submaximal exercise tests with similar efficiency.

CHAPTER 6

APPENDIX

This appendix is included to provide all relevant code to the analysis. Highly repetitive code, such as different model fits within WinBUGS, was not all included. Rather, only samples from this code were inserted.

6.1 R and WinBUGS Code

Model Fitting: Sample Sub-Maximal Models

```
### STAT 595R class 26 January 2012

dat1 <- read.table(
'C:/Users/Brent Shepherd/Documents/MastersProj/ChildrensVO2Data.txt',
header=TRUE)

dat1[,12] <- as.numeric(as.character(dat1[,12]))

dat1 <- dat1[-34,]
for(i in 1:146){
  if(dat1[i,1] > 112){ dat1[i,1] <- dat1[i,1] - 1}
}

library(arm)

quadreg <- function(){
  for (i in 1:146){
    y[i] ~ dnorm(mu[i],prec[gender[i]+1]);
    mu[i] <- b0[id[i]] + beta[1]*age[i] + beta[2]*gender[i] +
      beta[3]*bmi[i] + beta[4]*mph[i] + beta[5]*hr[i] +
      beta[6]*rpe[i] + beta[7]*age[i]*hr[i];
  }

  for(i in 1:132){
    b0[i] ~ dnorm(bo,precsubj);
  }

  bo ~ dnorm(0,.0001);
  precsubj <- 1/varsubj;
  varsubj ~ dgamma(1.5,.05);
  for(i in 1:7){
```

```
    beta[i] ~ dnorm(0,.01);
  }

  for(i in 1:2){
    prec[i] <- 1/var[i];
    var[i] ~ dgamma(1.5,.025);
  }

}

filename <- file.path(tempdir(),'quadreg.bug')
write.model(quadreg,filename);

y <- dat1[,19]
age <- dat1[,3]
gender <- dat1[,2]
bmi <- dat1[,6]
mph <- dat1[,11]
hr <- dat1[,12]
rpe <- dat1[,13]
id <- dat1[,1]

data <- c('age','gender','bmi','mph','hr','rpe','id','y')
parameters <- c('b0','bo','beta','var','varsubj')
quadreg.sim <- bugs(data,inits=NULL,parameters,
    model.file='quadreg.bug',n.iter=2000000,n.burnin=1000000,
    n.chains=1,n.thin=100,bugs.seed=1234,debug=T)

attach(quadreg.sim$sims.list)
print(quadreg.sim)

### STAT 595R class 26 January 2012

dat1 <- read.table(
'C:/Users/Brent Shepherd/Documents/MastersProj/ChildrensVO2Data.txt',
header=TRUE)

dat1[,12] <- as.numeric(as.character(dat1[,12]))

dat1 <- dat1[-34,]
for(i in 1:146){
  if(dat1[i,1] > 112){ dat1[i,1] <- dat1[i,1] - 1}
}

library(arm)

quadreg <- function(){
  for (i in 1:146){
    y[i] ~ dnorm(mu[i],prec[gender[i]+1]);
    mu[i] <- b0[id[i]] + beta[1]*age[i] + beta[2]*gender[i] +
      beta[3]*bmi[i] + beta[4]*mph[i] + beta[5]*hr[i] +
      beta[6]*rpe[i] + beta[7]*age[i]*hr[i] +
      beta[8]*gender[i]*hr[i];
  }
```

```
  for(i in 1:132){
    b0[i] ~ dnorm(bo,precsubj);
  }

  bo ~ dnorm(0,.0001);
  precsubj <- 1/varsubj;
  varsubj ~ dgamma(1.5,.05);
  for(i in 1:8){
    beta[i] ~ dnorm(0,.01);
  }

  for(i in 1:2){
    prec[i] <- 1/var[i];
    var[i] ~ dgamma(1.5,.025);
  }

}

filename <- file.path(tempdir(),'quadreg.bug')
write.model(quadreg,filename);

y <- dat1[,19]
age <- dat1[,3]
gender <- dat1[,2]
bmi <- dat1[,6]
mph <- dat1[,11]
hr <- dat1[,12]
rpe <- dat1[,13]
id <- dat1[,1]

data <- c('age','gender','bmi','mph','hr','rpe','id','y')
parameters <- c('b0','bo','beta','var','varsubj')
quadreg.sim <- bugs(data,inits=NULL,parameters,
    model.file='quadreg.bug',n.iter=2000000,n.burnin=1000000,
    n.chains=1,n.thin=100,bugs.seed=1234,debug=T)

attach(quadreg.sim$sims.list)
print(quadreg.sim)
```

Model Fitting: Sample Questionnaire Models

```
### STAT 595R class 26 January 2012

dat1 <- read.table(
'C:/Users/Brent Shepherd/Documents/MastersProj/ChildrensVO2Data.txt',
header=TRUE)

dat1[,12] <- as.numeric(as.character(dat1[,12]))

dat1 <- dat1[-34,]
for(i in 1:146){
  if(dat1[i,1] > 112){ dat1[i,1] <- dat1[i,1] - 1}
}
```

```
library(arm)

quadreg <- function(){
  for (i in 1:146){
    y[i] ~ dnorm(mu[i],prec[gender[i]+1]);
    mu[i] <- b0[id[i]] + beta[1]*age[i] + beta[2]*gender[i] +
      beta[3]*bmi[i] + beta[4]*pfa[i] + beta[5]*par[i];
  }

  for(i in 1:132){
    b0[i] ~ dnorm(bo,precsubj);
  }

  bo ~ dnorm(0,.0001);
  precsubj <- 1/varsubj;
  varsubj ~ dgamma(1.5,.05);

  for(i in 1:5){
    beta[i] ~ dnorm(0,.01);
  }

  for(i in 1:2){
    prec[i] <- 1/var[i];
    var[i] ~ dgamma(1.5,.025);
  }

}

filename <- file.path(tempdir(),'quadreg.bug')
write.model(quadreg,filename);

y <- dat1[,19]
age <- dat1[,3]
gender <- dat1[,2]
bmi <- dat1[,6]
pfa <- dat1[,7]
par <- dat1[,8]
id <- dat1[,1]

data <- c('age','gender','bmi','pfa','par','id','y')
parameters <- c('b0','bo','beta','var','varsubj')
quadreg.sim <- bugs(data,inits=NULL,parameters,
    model.file='quadreg.bug',n.iter=2000000,n.burnin=1000000,
    n.chains=1,n.thin=100,bugs.seed=1234,debug=T)

attach(quadreg.sim$sims.list)
print(quadreg.sim)

### STAT 595R class 26 January 2012

dat1 <- read.table(
'C:/Users/Brent Shepherd/Documents/MastersProj/ChildrensVO2Data.txt',
header=TRUE)
```

```
dat1[,12] <- as.numeric(as.character(dat1[,12]))

dat1 <- dat1[-34,]
for(i in 1:146){
  if(dat1[i,1] > 112){ dat1[i,1] <- dat1[i,1] - 1}
}

library(arm)

quadreg <- function(){
  for (i in 1:146){
    y[i] ~ dnorm(mu[i],prec[gender[i]+1]);
    mu[i] <- b0[id[i]] + beta[1]*par[i] + beta[2]*gender[i] +
      beta[3]*bmi[i] + beta[4]*pfa[i] + beta[5]*age[i] +
      beta[6]*gender[i]*par[i];
  }

  for(i in 1:132){
    b0[i] ~ dnorm(bo,precsubj);
  }

  bo ~ dnorm(0,.0001);
  precsubj <- 1/varsubj;
  varsubj ~ dgamma(1.5,.05);

  for(i in 1:6){
    beta[i] ~ dnorm(0,.01);
  }

  for(i in 1:2){
    prec[i] <- 1/var[i];
    var[i] ~ dgamma(1.5,.025);
  }

}

filename <- file.path(tempdir(),'quadreg.bug')
write.model(quadreg,filename);

y <- dat1[,19]
gender <- dat1[,2]
bmi <- dat1[,6]
pfa <- dat1[,7]
par <- dat1[,8]
id <- dat1[,1]
age <- dat1[,3]

data <- c('gender','bmi','pfa','par','age','id','y')
parameters <- c('b0','bo','beta','var','varsubj')
quadreg.sim <- bugs(data,inits=NULL,parameters,
    model.file='quadreg.bug',n.iter=2000000,n.burnin=1000000,
    n.chains=1,n.thin=100,bugs.seed=1234,debug=T)
```

```
attach(quadreg.sim$sims.list)
print(quadreg.sim)
```

Further Model Selection

```
# Chi-sq analysis
n <- 146 # obs
K <- round(n^.4)+1
r <- K-1 # num bins
qs <- seq(0,1,length.out=K) # quantiles
k <- quantile(ypred,qs) # bin locs

p <- m <- c()
for(i in 1:r){
  p[i] <- qs[i+1] - qs[i]
  m[i] <- sum(y > k[i] & y <= k[i+1])
}

R <- sum(((m - n*p)/sqrt(n*p))^2)
1-pchisq(R,r-1)

# R q7 = .6157432
# R q9 = .6111673
# R sm8 = .6544418
# R sm10 = .7063084


# q7: age, gender, bmi, pfa, par
# q9: par, gender, bmi, pfa, age, gender*par
# sm8:  age, gender, bmi, mph, hr, rpe, age*hr
# sm10: age, gender, bmi, mph, hr, rpe, age*hr, gender*hr

detach(quadreg.sim$sims.list)
load('/Users/Brent/Desktop/WinMastersProj/Questionnaire/qtrim7Burn')
attach(quadreg.sim$sims.list)

n <- 146 # obs
K <- round(n^.4)+1
r <- K-1 # num bins
exp <- rep(n/r,r)
test <- rep(0,dim(beta)[1])
for(k in 1:dim(beta)[1]){
  cnt <- rep(0,r)
  for(i in 1:2){
    yG <- dat1[gender==(i-1),19]
    ageG <- dat1[gender==(i-1),3]
    genderG <- dat1[gender==(i-1),2]
    bmiG <- dat1[gender==(i-1),6]
    pfaG <- dat1[gender==(i-1),7]
    parG <- dat1[gender==(i-1),8]
    idG <- dat1[gender==(i-1),1]
    mphG <- dat1[gender==(i-1),11]
    hrG <- dat1[gender==(i-1),12]
    rpeG <- dat1[gender==(i-1),13]
```

```
    thetaq7 <- b0[k,idG] + ageG*beta[k,1] + genderG*beta[k,2] + bmiG*beta[
       k,3] + pfaG*beta[k,4] + parG*beta[k,5]; cnt <- cnt + hist(pnorm(yG,
       thetaq7,sqrt(var[k,i])),breaks=qs,plot=F)$counts
    #thetaq9 <- b0[k,idG] + parG*beta[k,1] + genderG*beta[k,2] + bmiG*beta
       [k,3] + pfaG*beta[k,4] + ageG*beta[k,5] + genderG*parG*beta[k,6];
       cnt <- cnt + hist(pnorm(yG,thetaq9,sqrt(var[k,i])),breaks=qs,plot=F
       )$counts
    #thetasm8 <- b0[k,idG] + ageG*beta[k,1] + genderG*beta[k,2] + bmiG*
       beta[k,3] + mphG*beta[k,4] + hrG*beta[k,5] + rpeG*beta[k,6] + ageG*
       hrG*beta[k,7]; cnt <- cnt + hist(pnorm(yG,thetasm8,sqrt(var[k,i])),
       breaks=qs,plot=F)$counts
    #thetasm10 <- b0[k,idG] + ageG*beta[k,1] + genderG*beta[k,2] + bmiG*
       beta[k,3] + mphG*beta[k,4] + hrG*beta[k,5] + rpeG*beta[k,6] + ageG*
       hrG*beta[k,7] + genderG*hrG*beta[k,8]; cnt <- cnt + hist(pnorm(yG,
       thetasm10,sqrt(var[k,i])),breaks=qs,plot=F)$counts
  }
  test[k] <- (1/(n/r))*sum((exp-cnt)^2)
}

crit <- qchisq(.95,r-1)
mean(test > crit)
testq7 <- test

detach(quadreg.sim$sims.list)
load('/Users/Brent/Desktop/WinMastersProj/Questionnaire/qtrim9Burn')
attach(quadreg.sim$sims.list)

n <- 146 # obs
K <- round(n^.4)+1
r <- K-1 # num bins
exp <- rep(n/r,r)
test <- rep(0,dim(beta)[1])
for(k in 1:dim(beta)[1]){
  cnt <- rep(0,r)
  for(i in 1:2){
    yG <- dat1[gender==(i-1),19]
    ageG <- dat1[gender==(i-1),3]
    genderG <- dat1[gender==(i-1),2]
    bmiG <- dat1[gender==(i-1),6]
    pfaG <- dat1[gender==(i-1),7]
    parG <- dat1[gender==(i-1),8]
    idG <- dat1[gender==(i-1),1]
    mphG <- dat1[gender==(i-1),11]
    hrG <- dat1[gender==(i-1),12]
    rpeG <- dat1[gender==(i-1),13]

    thetaq9 <- b0[k,idG] + parG*beta[k,1] + genderG*beta[k,2] + bmiG*beta[
       k,3] + pfaG*beta[k,4] + ageG*beta[k,5] + genderG*parG*beta[k,6];
       cnt <- cnt + hist(pnorm(yG,thetaq9,sqrt(var[k,i])),breaks=qs,plot=F
       )$counts

  }
  test[k] <- (1/(n/r))*sum((exp-cnt)^2)
```

```
}

crit <- qchisq(.95,r-1)
mean(test > crit)
testq9 <- test

detach(quadreg.sim$sims.list)
load('/Users/Brent/Desktop/WinMastersProj/subMax/smtrim8Burn')
attach(quadreg.sim$sims.list)

n <- 146 # obs
K <- round(n^.4)+1
r <- K-1 # num bins
exp <- rep(n/r,r)
test <- rep(0,dim(beta)[1])
for(k in 1:dim(beta)[1]){
  cnt <- rep(0,r)
  for(i in 1:2){
    yG <- dat1[gender==(i-1),19]
    ageG <- dat1[gender==(i-1),3]
    genderG <- dat1[gender==(i-1),2]
    bmiG <- dat1[gender==(i-1),6]
    pfaG <- dat1[gender==(i-1),7]
    parG <- dat1[gender==(i-1),8]
    idG <- dat1[gender==(i-1),1]
    mphG <- dat1[gender==(i-1),11]
    hrG <- dat1[gender==(i-1),12]
    rpeG <- dat1[gender==(i-1),13]

    thetasm8 <- b0[k,idG] + ageG*beta[k,1] + genderG*beta[k,2] + bmiG*beta
       [k,3] + mphG*beta[k,4] + hrG*beta[k,5] + rpeG*beta[k,6] + ageG*hrG*
       beta[k,7]; cnt <- cnt + hist(pnorm(yG,thetasm8,sqrt(var[k,i])),
       breaks=qs,plot=F)$counts
  }
  test[k] <- (1/(n/r))*sum((exp-cnt)^2)
}

crit <- qchisq(.95,r-1)
mean(test > crit)
testsm8 <- test

detach(quadreg.sim$sims.list)
load('/Users/Brent/Desktop/WinMastersProj/subMax/smtrim10Burn')
attach(quadreg.sim$sims.list)

n <- 146 # obs
K <- round(n^.4)+1
r <- K-1 # num bins
exp <- rep(n/r,r)
test <- rep(0,dim(beta)[1])
for(k in 1:dim(beta)[1]){
  cnt <- rep(0,r)
  for(i in 1:2){
    yG <- dat1[gender==(i-1),19]
```

```
    ageG <- dat1[gender==(i-1),3]
    genderG <- dat1[gender==(i-1),2]
    bmiG <- dat1[gender==(i-1),6]
    pfaG <- dat1[gender==(i-1),7]
    parG <- dat1[gender==(i-1),8]
    idG <- dat1[gender==(i-1),1]
    mphG <- dat1[gender==(i-1),11]
    hrG <- dat1[gender==(i-1),12]
    rpeG <- dat1[gender==(i-1),13]

    thetasm10 <- b0[k,idG] + ageG*beta[k,1] + genderG*beta[k,2] + bmiG*
       beta[k,3] + mphG*beta[k,4] + hrG*beta[k,5] + rpeG*beta[k,6] + ageG*
       hrG*beta[k,7] + genderG*hrG*beta[k,8]; cnt <- cnt + hist(pnorm(yG,
       thetasm10,sqrt(var[k,i])),breaks=qs,plot=F)$counts
  }
  test[k] <- (1/(n/r))*sum((exp-cnt)^2)
}

crit <- qchisq(.95,r-1)
mean(test > crit)
testsm10 <- test

curve(dchisq(x,r-1),xlim=c(0,25),lwd=2,main='Chi-Squared Test:
   Questionnaire Models',xlab='RB',ylab='Density')
lines(density(testq7),lty=2,col=2,lwd=2)
lines(density(testq9),lty=3,col=3,lwd=2)
legend(10,.13,c('actual chi-squared density','main effects','main effects,
    gender:PAR'),col=c(1,2,3),lty=c(1,2,3),lwd=c(2,2,2),bty='n')

curve(dchisq(x,r-1),xlim=c(0,25),lwd=2,main='Chi-Squared Test: Sub-Maximal
    Models',xlab='RB',ylab='Density')
lines(density(testsm8),lty=2,col=2,lwd=2)
lines(density(testsm10),lty=3,col=3,lwd=2)
legend(10,.13,c('actual chi-squared density','main effects, age:HR','main
   effects, age:HR, gender:HR'),col=c(1,2,3),lty=c(1,2,3),lwd=c(2,2,2),bty
   ='n')

# q7: 0.0546
# q9: 0.0519
# sm8: 0.0504
# sm10: 0.055

### Bayes Factor Masters Proj

# questionnaire -- trim 6, trim 8
# submax -- trim 8, trim 10
rm(list=ls())
dat1 <- read.table('~/Desktop/winMastersProj/ChildrensVO2Data.txt',header=
   TRUE)
dat1[,12] <- as.numeric(as.character(dat1[,12]))

dat1 <- dat1[-34,]
for(i in 1:146){
  if(dat1[i,1] > 112){ dat1[i,1] <- dat1[i,1] - 1}
```

```
}

ords <- order(dat1[,1])
dat1 <- dat1[ords,]

y <- dat1[,19]
age <- dat1[,3]
gender <- dat1[,2]
bmi <- dat1[,6]
pfa <- dat1[,7]
par <- dat1[,8]
id <- dat1[,1]
mph <- dat1[,11]
hr <- dat1[,12]
rpe <- dat1[,13]

datEval <- as.data.frame(cbind(y,age,gender,bmi,pfa,par,id,mph,hr,rpe))

##        full model (Q9)
##
##
##
##
##
##
detach(quadreg.sim$sims.list)
load('/Users/Brent/Desktop/WinMastersProj/Questionnaire/qtrim9Burn')
attach(quadreg.sim$sims.list)
##
##
##
##
##
##
##


# q9: par, gender, bmi, pfa, age, gender*par

loglik <- function(dat,b0,beta,sig2){
  NF <- sum(dat$gender==0)
  NM <- sum(dat$gender==1)
  ints <- rep(b0,as.vector(table(dat$id)))
  slps <- dat$par*beta[1] + dat$gender*beta[2] + dat$bmi*beta[3] + dat$pfa
     *beta[4] + dat$age*beta[5] + dat$gender*dat$par*beta[6]
  likF <- sum((dat$y[dat$gender==0] - ints[dat$gender==0] - slps[
     dat$gender==0])^2)
  likM <- sum((dat$y[dat$gender==1] - ints[dat$gender==1] - slps[
     dat$gender==1])^2)
  out <- -(NF/2)*log(sig2[1]) - (NF/2)*log(sig2[2]) - (1/(2*sig2[1]))*likF
      - (1/(2*sig2[2]))*likM
  return(out)
}
```

```
sigmahat <- var(cbind(bo,b0,beta,varsubj,var))
estsbo <- mean(bo)
estsbeta <- apply(beta,2,mean)
estsb0 <- apply(b0,2,mean)
estsvar <- c(mean(var[,1]),mean(var[,2]))
estsvarsubj <- mean(varsubj)

logpriors <- function(bo,beta,b0,var,varsubj){
  idNum <- 132
  betaNum <- 6
  Intpart <- 0
  for(i in 1:idNum){
    Intpart <- Intpart + log(dnorm(b0[i],bo,sqrt(varsubj)))
  }
  varpart <- 0
  for(i in 1:2){
    varpart <- varpart + log(dgamma(var[i],shape=1.5,scale=40))
  }
  subjpart <- log(dgamma(varsubj,shape=1.5,scale=20))
  bopart <- log(dnorm(bo,0,sqrt(10000)))
  betapart <- 0
  for(i in 1:betaNum){
    betapart <- betapart + log(dnorm(beta[i],0,sqrt(100)))
  }
  out <- Intpart + varpart + subjpart + bopart + betapart
  return(out)
}

logihat <- ((132+6+2+1+1)/2)*log(2*pi) + .5*log(det(sigmahat))
logihat <- logihat + loglik(datEval,estsb0,estsbeta,estsvar)
logihat <- logihat + logpriors(estsbo,estsbeta,estsb0,estsvar,estsvarsubj)

logihatQ9 <- logihat



##       reduced model (Q7)
##
##
##
##
##
##
detach(quadreg.sim$sims.list)
load('/Users/Brent/Desktop/WinMastersProj/Questionnaire/qtrim7Burn')
attach(quadreg.sim$sims.list)
##
##
##
##
##
##
##
```

```
# q7: age, gender, bmi, pfa, par

loglik <- function(dat,b0,beta,sig2){
  NF <- sum(dat$gender==0)
  NM <- sum(dat$gender==1)
  ints <- rep(b0,as.vector(table(dat$id)))
  slps <- dat$age*beta[1] + dat$gender*beta[2] + dat$bmi*beta[3] + dat$pfa
     *beta[4] + dat$par*beta[5]
  likF <- sum((dat$y[dat$gender==0] - ints[dat$gender==0] - slps[
     dat$gender==0])^2)
  likM <- sum((dat$y[dat$gender==1] - ints[dat$gender==1] - slps[
     dat$gender==1])^2)
  out <- -(NF/2)*log(sig2[1]) - (NF/2)*log(sig2[2]) - (1/(2*sig2[1]))*likF
      - (1/(2*sig2[2]))*likM
  return(out)
}

sigmahat <- var(cbind(bo,b0,beta,varsubj,var))
estsbo <- mean(bo)
estsbeta <- apply(beta,2,mean)
estsb0 <- apply(b0,2,mean)
estsvar <- c(mean(var[,1]),mean(var[,2]))
estsvarsubj <- mean(varsubj)

logpriors <- function(bo,beta,b0,var,varsubj){
  idNum <- 132
  betaNum <- 5
  Intpart <- 0
  for(i in 1:idNum){
    Intpart <- Intpart + log(dnorm(b0[i],bo,sqrt(varsubj)))
  }
  varpart <- 0
  for(i in 1:2){
    varpart <- varpart + log(dgamma(var[i],shape=1.5,scale=40))
  }
  subjpart <- log(dgamma(varsubj,shape=1.5,scale=20))
  bopart <- log(dnorm(bo,0,sqrt(10000)))
  betapart <- 0
  for(i in 1:betaNum){
    betapart <- betapart + log(dnorm(beta[i],0,sqrt(100)))
  }
  out <- Intpart + varpart + subjpart + bopart + betapart
  return(out)
}

logihat <- ((132+5+2+1+1)/2)*log(2*pi) + .5*log(det(sigmahat))
logihat <- logihat + loglik(datEval,estsb0,estsbeta,estsvar)
logihat <- logihat + logpriors(estsbo,estsbeta,estsb0,estsvar,estsvarsubj)

logihatQ7 <- logihat

##       full model (SM10)
##
##
```

```
##
##
##
##
detach(quadreg.sim$sims.list)
load('/Users/Brent/Desktop/WinMastersProj/subMax/smtrim10Burn')
attach(quadreg.sim$sims.list)
##
##
##
##
##
##
##

# sm10: age, gender, bmi, mph, hr, rpe, age*hr, gender*hr

loglik <- function(dat,b0,beta,sig2){
  NF <- sum(dat$gender==0)
  NM <- sum(dat$gender==1)
  ints <- rep(b0,as.vector(table(dat$id)))
  slps <- dat$age*beta[1] + dat$gender*beta[2] + dat$bmi*beta[3] + dat$mph
     *beta[4] + dat$hr*beta[5] + dat$rpe*beta[6] + dat$age*dat$hr*beta[7]
     + dat$gender*dat$hr*beta[8]
  likF <- sum((dat$y[dat$gender==0] - ints[dat$gender==0] - slps[
     dat$gender==0])^2)
  likM <- sum((dat$y[dat$gender==1] - ints[dat$gender==1] - slps[
     dat$gender==1])^2)
  out <- -(NF/2)*log(sig2[1]) - (NF/2)*log(sig2[2]) - (1/(2*sig2[1]))*likF
      - (1/(2*sig2[2]))*likM
  return(out)
}

sigmahat <- var(cbind(bo,b0,beta,varsubj,var))
estsbo <- mean(bo)
estsbeta <- apply(beta,2,mean)
estsb0 <- apply(b0,2,mean)
estsvar <- c(mean(var[,1]),mean(var[,2]))
estsvarsubj <- mean(varsubj)

logpriors <- function(bo,beta,b0,var,varsubj){
  idNum <- 132
  betaNum <- 8
  Intpart <- 0
  for(i in 1:idNum){
    Intpart <- Intpart + log(dnorm(b0[i],bo,sqrt(varsubj)))
  }
  varpart <- 0
  for(i in 1:2){
    varpart <- varpart + log(dgamma(var[i],shape=1.5,scale=40))
  }
  subjpart <- log(dgamma(varsubj,shape=1.5,scale=20))
  bopart <- log(dnorm(bo,0,sqrt(10000)))
  betapart <- 0
```

```
  for(i in 1:betaNum){
    betapart <- betapart + log(dnorm(beta[i],0,sqrt(100)))
  }
  out <- Intpart + varpart + subjpart + bopart + betapart
  return(out)
}

logihat <- ((132+8+2+1+1)/2)*log(2*pi) + .5*log(det(sigmahat))
logihat <- logihat + loglik(datEval,estsb0,estsbeta,estsvar)
logihat <- logihat + logpriors(estsbo,estsbeta,estsb0,estsvar,estsvarsubj)

logihatSM10 <- logihat

##        full model (SM8)
##
##
##
##
##
##
detach(quadreg.sim$sims.list)
load('/Users/Brent/Desktop/WinMastersProj/subMax/smtrim8Burn')
attach(quadreg.sim$sims.list)
##
##
##
##
##
##
##
##

# sm8: age, gender, bmi, mph, hr, rpe, age*hr

loglik <- function(dat,b0,beta,sig2){
  NF <- sum(dat$gender==0)
  NM <- sum(dat$gender==1)
  ints <- rep(b0,as.vector(table(dat$id)))
  slps <- dat$age*beta[1] + dat$gender*beta[2] + dat$bmi*beta[3] + dat$mph
     *beta[4] + dat$hr*beta[5] + dat$rpe*beta[6] + dat$age*dat$hr*beta[7]
  likF <- sum((dat$y[dat$gender==0] - ints[dat$gender==0] - slps[
     dat$gender==0])^2)
  likM <- sum((dat$y[dat$gender==1] - ints[dat$gender==1] - slps[
     dat$gender==1])^2)
  out <- -(NF/2)*log(sig2[1]) - (NF/2)*log(sig2[2]) - (1/(2*sig2[1]))*likF
      - (1/(2*sig2[2]))*likM
  return(out)
}

sigmahat <- var(cbind(bo,b0,beta,varsubj,var))
estsbo <- mean(bo)
estsbeta <- apply(beta,2,mean)
estsb0 <- apply(b0,2,mean)
estsvar <- c(mean(var[,1]),mean(var[,2]))
estsvarsubj <- mean(varsubj)
```

```
logpriors <- function(bo,beta,b0,var,varsubj){
  idNum <- 132
  betaNum <- 7
  Intpart <- 0
  for(i in 1:idNum){
    Intpart <- Intpart + log(dnorm(b0[i],bo,sqrt(varsubj)))
  }
  varpart <- 0
  for(i in 1:2){
    varpart <- varpart + log(dgamma(var[i],shape=1.5,scale=40))
  }
  subjpart <- log(dgamma(varsubj,shape=1.5,scale=20))
  bopart <- log(dnorm(bo,0,sqrt(10000)))
  betapart <- 0
  for(i in 1:betaNum){
    betapart <- betapart + log(dnorm(beta[i],0,sqrt(100)))
  }
  out <- Intpart + varpart + subjpart + bopart + betapart
  return(out)
}

logihat <- ((132+7+2+1+1)/2)*log(2*pi) + .5*log(det(sigmahat))
logihat <- logihat + loglik(datEval,estsb0,estsbeta,estsvar)
logihat <- logihat + logpriors(estsbo,estsbeta,estsb0,estsvar,estsvarsubj)

logihatSM8 <- logihat

logihatQ9
logihatQ7
logihatSM10
logihatSM8
```

Prediction and Residual Analysis

```
## prediction

##### Bayes Chi-Squared Masters Project

# sub-max -- 8, 10; questionnaire -- 7, 9

dat1 <- read.table('~/Desktop/winMastersProj/ChildrensVO2Data.txt',header=
    TRUE)
dat1[,12] <- as.numeric(as.character(dat1[,12]))

dat1 <- dat1[-34,]
for(i in 1:146){
  if(dat1[i,1] > 112){ dat1[i,1] <- dat1[i,1] - 1}
}

y <- dat1[,19]
age <- dat1[,3]
gender <- dat1[,2]
bmi <- dat1[,6]
```

```
pfa <- dat1[,7]
par <- dat1[,8]
id <- dat1[,1]
mph <- dat1[,11]
hr <- dat1[,12]
rpe <- dat1[,13]
day <- dat1[,9]

library(coda)

load('/Users/Brent/Desktop/WinMastersProj/Questionnaire/qtrim9Burn')
betaQ <- quadreg.sim$sims.list$beta
b0Q <- quadreg.sim$sims.list$b0
boQ <- quadreg.sim$sims.list$bo
varQ <- quadreg.sim$sims.list$var
varsubjQ <- quadreg.sim$sims.list$varsubj
load('/Users/Brent/Desktop/WinMastersProj/subMax/smtrim8Burn')
betaSM <- quadreg.sim$sims.list$beta
b0SM <- quadreg.sim$sims.list$b0
boSM <- quadreg.sim$sims.list$bo
varSM <- quadreg.sim$sims.list$var
varsubjSM <- quadreg.sim$sims.list$varsubj

xfQ <- c(6,0,21,14.5,14.5,0)
xmQ <- c(7,1,20.5,18,14,7)
xfSM <- c(14.5,0,21,5,166,13,14.5*166)
xmSM <- c(14,1,20.5,5.5,160,12,14*160)

# females
dat1 <- dat1[,-c(9)]
apply(dat1[dat1$Gender == 0, ],2,mean)
# age = 14.5, BMI = 21, PFA = 14.5, PAR = 6, MPH = 5, HR = 166, RPE = 13,
   vo2 = 44

# males
apply(dat1[dat1$Gender == 1, ],2,mean)
# age = 14, BMI = 20.5, PFA = 18 PAR = 7 MPH = 5.5, HR = 160, RPE = 12,
   vo2 = 54

#femalehat <- (cbind(b0,bgen,bhr,bmph,bbmi,brpe)%*%xx)
#ppfem <- femalehat + rnorm(10000,0,sqrt(var)) # this gives prediction
   interval
#plot(density(ppfem),col='red',ylim=c(0,.7))
#lines(density(femalehat))

thetafQ <- betaQ%*%xfQ
thetamQ <- betaQ%*%xmQ
thetafSM <- betaSM%*%xfSM
thetamSM <- betaSM%*%xmSM

# prediction intervals
ypredfQ <- rnorm(10000,thetafQ,sqrt(varQ[,1])) + rnorm(10000,boQ,sqrt(
   varsubjQ))
```

```
ypredmQ <- rnorm(10000,thetamQ,sqrt(varQ[,2])) + rnorm(10000,boQ,sqrt(
   varsubjQ))
ypredfSM <- rnorm(10000,thetafSM,sqrt(varSM[,1])) + rnorm(10000,boSM,sqrt(
   varsubjSM))
ypredmSM <- rnorm(10000,thetamSM,sqrt(varSM[,2])) + rnorm(10000,boSM,sqrt(
   varsubjSM))

ypredSM <- c(ypredfSM,ypredmSM)
ymodFSM <- thetafSM + boSM
ymodMSM <- thetamSM + boSM
ypredQ <- c(ypredfQ,ypredmQ)
ymodFQ <- thetafQ + boQ
ymodMQ <- thetamQ + boQ

# plot Questionnaire
plot(density(ypredfQ),col='hotpink',xlim=c(25,80),lwd=2,main='Intervals
   for Adolescent VO2max levels: Questionnaire',xlab='VO2max',ylim=c(0,.7)
   )
lines(density(ypredmQ),col='blue',lwd=2)
lines(density(ymodFQ),lty=2,lwd=2,col='hotpink')
lines(density(ymodMQ),lty=2,lwd=2,col='blue')
legend(60,.65,c('Prediction: female','Model: female','Prediction: male','
   Model: male'),col=c('hotpink','hotpink','blue','blue'),lwd=c(2,2,2,2),
   bty='n',lty=c(1,2,1,2))

# plot SM
plot(density(ypredfSM),col='hotpink',xlim=c(25,80),lwd=2,main='Intervals
   for Adolescent VO2max levels: Submaximal',xlab='VO2max',ylim=c(0,.7))
lines(density(ypredmSM),col='blue',lwd=2)
lines(density(ymodFSM),lty=2,lwd=2,col='hotpink')
lines(density(ymodMSM),lty=2,lwd=2,col='blue')
legend(60,.65,c('Prediction: female','Model: female','Prediction: male','
   Model: male'),col=c('hotpink','hotpink','blue','blue'),lwd=c(2,2,2,2),
   bty='n',lty=c(1,2,1,2))

# plot Female Q and SM
plot(density(ypredfQ),col='hotpink',xlim=c(20,70),lwd=2,main='Intervals
   for Adolescent VO2max levels: Females',xlab='VO2max',ylim=c(0,.7),lty
   =2)
lines(density(ypredfSM),col='hotpink',lwd=2)
lines(density(ymodFQ),col='hotpink',lwd=2,lty=2)
lines(density(ymodFSM),col='hotpink',lwd=2)
legend(55,.65,c('Submaximal','Questionnaire'),col=c('hotpink','hotpink'),
   lwd=c(2,2),lty=c(1,2),bty='n')

# plot Male Q and SM
plot(density(ypredmQ),col='blue',xlim=c(25,83),lwd=2,main='Intervals for
   Adolescent VO2max levels: Males',xlab='VO2max',ylim=c(0,.7),lty=2)
lines(density(ypredmSM),col='blue',lwd=2)
lines(density(ymodMQ),col='blue',lwd=2,lty=2)
lines(density(ymodMSM),col='blue',lwd=2)
legend(60,.6,c('Submaximal','Questionnaire'),col=c('blue','blue'),lwd=c
   (2,2),lty=c(1,2),bty='n')
```

```
# residuals

# resids at every iteration -- Q
yhatsQ <- resQ <- matrix(NA,nrow=10000,ncol=146)
for(i in 1:146){
  yhatsQ[,i] <- b0Q[,id[i]] + betaQ[,1]*par[i] + betaQ[,2]*gender[i] +
      betaQ[,3]*bmi[i] + betaQ[,4]*pfa[i] + betaQ[,5]*age[i] +
      betaQ[,6]*gender[i]*par[i]
      resQ[,i] <- y[i] - yhatsQ[,i]
}

# resids at every iteration -- SM
yhatsSM <- resSM <- matrix(NA,nrow=10000,ncol=146)
for(i in 1:146){
  yhatsSM[,i] <- b0SM[,id[i]] + betaSM[,1]*age[i] + betaSM[,2]*gender[i] +
      betaSM[,3]*bmi[i] + betaSM[,4]*mph[i] + betaSM[,5]*hr[i] +
      betaSM[,6]*rpe[i] + betaSM[,7]*age[i]*hr[i]
      resSM[,i] <- y[i] - yhatsSM[,i]
}

# plot the residual densities: Q

plot(density(resQ[,1]),xlab='resids',xlim=c(-22,22),ylim=c(0,.4),main='
   Residual Densities for Each Observation: Questionnaire')
for(i in 2:146){
  lines(density(resQ[,i]),col=i)
}

plot(density(resQ[,1]),xlab='resids',xlim=c(-22,22),ylim=c(0,.4),col='
   hotpink',main='Residual Densities for Each Observation: Questionnaire')
for(i in 2:146){
  if(gender[i] == 0){
    lines(density(resQ[,i]),col='hotpink')
  } else lines(density(resQ[,i]),col='blue')
}
legend(7,.37,c('Female','Male'),col=c('hotpink','blue'),lty=c(1,1),bty='n
   ')


day2 <- day == 2
for(i in 1:146){
  if(day2[i] == TRUE){day2[i-1] = TRUE}

}

plot(density(resQ[,1]),xlab='resids',xlim=c(-22,22),ylim=c(0,.4),col='
   hotpink',main='Residual Densities for Each Observation: Questionnaire')
for(i in 2:146){
  if(gender[i] == 0){
    if(day2[i] == TRUE){
      lines(density(resQ[,i]),col='hotpink',lty=2)
    } else lines(density(resQ[,i]),col='hotpink')
  } else if(day2[i] == TRUE){
    lines(density(resQ[,i]),col='blue',lty=2)
```

```
  } else lines(density(resQ[,i]),col='blue')
}
legend(5,.4,c('Female: One Obs','Female: Two Obs','Male: One Obs','Male:
    Two Obs'),col=c('hotpink','hotpink','blue','blue'),lty=c(1,2,1,2),bty='
    n')

# plot the residual densities: SM

plot(density(resSM[,1]),xlab='resids',xlim=c(-22,22),ylim=c(0,.45),main='
    Residual Densities for Each Observation: Submaximal')
for(i in 2:146){
  lines(density(resSM[,i]),col=i)
}

plot(density(resSM[,1]),xlab='resids',xlim=c(-22,22),ylim=c(0,.45),col='
    hotpink',main='Residual Densities for Each Observation: Submaximal')
for(i in 2:146){
  if(gender[i] == 0){
    lines(density(resSM[,i]),col='hotpink')
  } else lines(density(resSM[,i]),col='blue')
}
legend(7,.4,c('Female','Male'),col=c('hotpink','blue'),lty=c(1,1),bty='n')

plot(density(resSM[,1]),xlab='resids',xlim=c(-22,22),ylim=c(0,.45),col='
    hotpink',main='Residual Densities for Each Observation: Submaximal')
for(i in 2:146){
  if(gender[i] == 0){
    if(day2[i] == TRUE){
      lines(density(resSM[,i]),col='hotpink',lty=2)
    } else lines(density(resSM[,i]),col='hotpink')
  } else if(day2[i] == TRUE){
    lines(density(resSM[,i]),col='blue',lty=2)
  } else lines(density(resSM[,i]),col='blue')
}
legend(5,.45,c('Female: One Obs','Female: Two Obs','Male: One Obs','Male:
    Two Obs'),col=c('hotpink','hotpink','blue','blue'),lty=c(1,2,1,2),bty='
    n')


# density plotted of all possible resids: female
# Q
plot(density(resQ[,1:70]),lwd=2)
curve(dnorm(x,0,sqrt(mean(varQ[,1]))),col='hotpink',add=T,lwd=2) # what it
     should look like
# SM
plot(density(resSM[,1:70]),lwd=2)
curve(dnorm(x,0,sqrt(mean(varSM[,1]))),col='hotpink',add=T,lwd=2) # what
    it should look like

# density plotted of all possible resids: male
# Q
plot(density(resQ[,71:146]),lwd=2)
curve(dnorm(x,0,sqrt(mean(varQ[,2]))),col='blue',add=T,lwd=2) # what it
    should look like
```

```
# SM
plot(density(resSM[,71:146]),lwd=2)
curve(dnorm(x,0,sqrt(mean(varSM[,2]))),col='blue',add=T,lwd=2) # what it
    should look like


##### SM8

#
#
#
#
# This code was used to discover the source of the smaller variance within
    gender resids
#
#
#
#
weirds <- c(3,4,17,18,19,20,56,57,64,65)
notweirds <- 1:70; notweirds <- notweirds[-weirds]
plot(density(resSM[,1]),xlab='resids',ylim=c(0,.5))
for(i in notweirds){
  lines(density(resSM[,i]))
}

for(i in weirds){
  lines(density(resSM[,i]),col='blue')
}

wdat <- dat1[weirds,c(3,4,5,6,10)]
nwdat <- dat1[notweirds,c(3,4,5,6,10)]
apply(as.matrix(wdat),2,mean)
apply(as.matrix(nwdat),2,mean)
# it's the number of obs-- those tested twice have less variance in resids
    (this makes sense)
```

BIBLIOGRAPHY

Ando, T. (2007), "Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models," *Biometrika*, 94, 443–458.

Astrand, P., and Ryhming, I. (1954), "A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work," *Journal of Applied Physiology*, 7, 218–221.

Astrand, P., and Saltin, B. (1961), "Maximal oxygen uptake and heart rate in various types of muscular activity," *Journal of Applied Physiology*, 16, 917–976.

Bartlett, M. (1957), "A comment on D.V. Lindley's Statistical Paradox," *Biometrika*, 44, 533–534.

Bayes, T., and Price, R. (1763), "An Essay towards solving a Problem in the Doctrine of Chance. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S." *Philosophical Transactions of the Royal Society of London*, 53, 370–418.

Bouchard, C., An, P., Rice, T., Skinner, J. S., Wilmore, J. H., Gagnon, J., Perusse, L., Leon, A. S., and Rao, D. (1999), "Familial aggregation of $\dot{V}O_2$max response to exercise training: results from the HERITAGE Family Study," *Journal of Applied Physiology*, 87, 1003–1008.

Bramble, D. M., and Lieberman, D. E. (2004), "Endurance running and the evolution of *Homo*," *Nature*, 432, 345–352.

Carlin, B. P., and Louis, T. A. (2009), *Bayesian Methods for Data Analysis* (3rd ed.), Chapman & Hall/CRC.

Coleman, E. (1976), "Validation of a submaximal test of maximal oxygen uptake," *Journal of Sports Medicine and Physical Fitness*, 16, 106–111.

Cosmed (2006), "Cosmed FitMax," http://www.cosmed.it/.

Fick, A. E. (1870), "ber die blutquantums in den herzventrikeln," *Sitzung Phys med Gesell Würzburg*, July, 16.

Gelfand, A., and Dey, D. (1994), "Bayesian Model Choice: Asymptotics and Exact Calculations," *Journal of the Royal Statistical Society, Series B*, 56, 501–514.

Gelman, A., and Rubin, D. B. (1992), "Inference from Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472.

Hastings, W. K. (1970), "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika*, 57, 97–109.

Hyde, T. E., and Gengenbach, M. S. (2009), *Conservative Management of Sports Injuries* (2nd ed.), Jones & Bartlett.

Jeffreys, H. (1935), "Some tests of significance, treated by the theory of probability," *Proceedings of the Cambridge Philosophy Society*, 203–222.

—— (1961), *Theory of Probability* (3rd ed.), Oxford University Press.

Johnson, V. E. (2004), "A Bayesian χ^2 test for goodness-of-fit," *The Annals of Statistics*, 32, 2361–2384.

Kass, R. E., and Raftery, A. E. (1995), "Bayes Factors," *Journal of the American Statistical Association*, 90, 773–795.

Koehler, K., and Gan, F. (1990), "Chi-squared goodness-of-fit tests: Cell selection and power," *Communications in Statistics–Simulation and Computation*, 19, 1265–1278.

Lee, J.-M., Bassett, D. R. J., Thompson, D. L., and Fitzhugh, E. C. (2011), "Validation of the cosmed fitmate for prediction of maximal oxygen consumption," *Journal of Strength and Conditioning Research*, 25, 2573–2579.

Léger, L., Mercier, D., Gadoury, C., and Lambert, J. (1988), "The multistage 20 metro shuttle run test for aerobic fitness," *Journal of Sports Sciences*, 6, 93–101.

Lewis, S. M., and Raftery, A. E. (1994), "Estimating Bayes Factors via Posterior Simulation with the Laplace-Metropolis Estimator," *Journal of the American Statistical Association*, 92, 648–655.

Lieberman, D. E., and Bramble, D. M. (2007), "The evolution of marathon running capabilities in humans," *Sports Medicine*, 37, 288–290.

Lieberman, D. E., Raichlen, D. A., Pontzer, H., Bramble, D. M., and Cutright-Smith, E. (2006), "The human gluteus maximus and its role in running," *The Journal of Experimental Biology*, 209, 2143–2155.

Lindley, D. (1957), "A Statistical Paradox," *Biometrika*, 44, 187–192.

Macfarlane, D. (2001), "Automated metabolic gas analysis systems: A review," *Sports Medicine*, 31, 841–861.

Mann, H., and Wald, A. (1942), "On the choice of the number of class intervals in the application of the chi-square test," *The Annals of Mathematical Statistics*, 13, 306–317.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953), "Equations of state calculations by fast computing machines," *Journal of Chemical Physics*, 21, 1087–1092.

Metz, K., and Alexander, J. (1971), "Estimation of maximal oxygen intake from submaximal work parameters," *Research Quarterly of the AAHPER*, 42, 187–193.

Montoye, H., Ayen, T., and Washburn, R. (1986), "The estimation of $\dot{V}O_2$max from maximal and sub-maximal measurements in males, age 10-39," *Research Quarterly for Exercise & Sport*, 57, 25–253.

Nieman, D., Austin, M., Benezra, L., Pearce, S., McInnis, T., Unick, J., and Gross, S. (2006), "Validation of Cosmed's FitMate in measuring oxygen consumption and estimating resting metabolic rate," *Research Sports Med*, 14, 89–96.

Raftery, A. E., and Lewis, S. M. (1992), "One long run with diagnostics: Implementation strategies for Markov chain Monte Carlo," *Statistical Science*, 7, 493–497.

Robert, C. P., and Titterington, D. M. (2002), "Discussion of paper by D. J. Spiegelhalter et al," *Journal of the Royal Statistical Society*, B, 621–622.

Roberts, G., Gelman, A., and Gilks, W. (1997), "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms," *The Annals of Applied Probability*, 7, 110–120.

Rolian, C., Lieberman, D. E., Hamill, J., Scott, J. W., and Werbel, W. (2008), "Walking, running and the evolution of short toes in humans," *The Journal of Experimental Biology*, 212, 713–721.

Saltin, B., and Astrand, P. (1967), "Maximal oxygen uptake in athletes," *Journal of Applied Physiology*, 23, 353–358.

Siconolfi, S., Cullinane, E., Carleton, R., and Thompson, P. (1982), "Assessing $\dot{V}O_2$max in epidemiological studies: modification of the Astrand-Ryhming test," *Medicine and Science in Sports and Exercise*, 14, 335–338.

Spiegelhalter, D. J. (2002), "Bayesian measures of model complexity and fit (with discussion)," *Journal of the Royal Statistical Society*, 64, 583–639.

Stigler, S. M. (1983), "Who Discovered Bayes' Theorem?" *The American Statistician*, 37, 290–296.

Uth, N., Sørensen, H., Overgaard, K., and Pedersen, P. K. (2004), "Estimation of VO2max from the ratio between HRmax and HRrest–the Heart Rate Ratio Method," *European Journal of Applied Physiology*, 91, 111–115.