



2007-03-01

Graphical and Bayesian Analysis of Unbalanced Patient Management Data

Emily Stewart Righter

Brigham Young University - Provo

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>

 Part of the [Statistics and Probability Commons](#)

BYU ScholarsArchive Citation

Righter, Emily Stewart, "Graphical and Bayesian Analysis of Unbalanced Patient Management Data" (2007). *All Theses and Dissertations*. 1102.

<https://scholarsarchive.byu.edu/etd/1102>

This Selected Project is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.

GRAPHICAL AND BAYESIAN ANALYSIS OF UNBALANCED PATIENT
MANAGEMENT DATA

by

Emily Stewart Righter

A Project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

April 2007

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a Project submitted by

Emily Stewart Righter

This Project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date

G. Bruce Schaalje, Chair

Date

David G. Whiting

Date

John S. Lawson

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the Project of Emily Stewart Righter in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

G. Bruce Schaalje
Chair, Graduate Committee

Accepted for the Department

Scott D. Grimshaw
Graduate Coordinator

Accepted for the College

Thomas W. Sederberg
Associate Dean, College of Physical and
Mathematical Sciences

ABSTRACT

GRAPHICAL AND BAYESIAN ANALYSIS OF UNBALANCED PATIENT MANAGEMENT DATA

Emily Stewart Righter

Department of Statistics

Master of Science

The International Normalizing Ratio (INR) measures the speed at which blood clots. Healthy people have an INR of about one. Some people are at greater risk of blood clots and their physician prescribes a target INR range, generally 2-3. The farther a patient is above or below their prescribed range, the more dangerous their situation. A variety of point-of-care (POC) devices has been developed to monitor patients.

The purpose of this research was to develop innovative graphics to help describe a highly unbalanced dataset and to carry out Bayesian analyses to determine which of five devices best manages patients. An initial Bayesian analysis compared a machine-identical beta-binomial model to a machine-specific beta-binomial model. The response variable was number of in-range visits. A second Bayesian analysis compared a machine-identical lognormal model, a machine-specific lognormal model, and a machine-specific lognormal model with lower therapeutic bound as a predictor. The response variable was INR. Machines were

compared using posterior predictive distributions of the absolute distance outside a patient's therapeutic range.

For the beta-binomial models, the machine-identical model had the lower DIC, meaning that POC device was not a strong predictor of success in keeping a patient in-range. The machine-specific lognormal model with a term for lower therapeutic bound had the lowest DIC of the three lognormal models, implying that the additional information of distance out of range revealed differences among the POC devices. Three of the machines had more favorable out-of-range distributions than the other two. Both Bayesian analyses provided useful information for medical practice in managing INR.

Acknowledgements

I am grateful to the Cleveland Clinic Foundation for providing the data for this project, particularly Jodie Fink and Ken Shermock, whose project this is, and Jason Connor, who was my mentor and oversaw all the early stages. Many thanks to the faculty at Brigham Young University for equipping me with the tools to do this project, and especially to Dr. Bruce Schaalje for agreeing to take it on in the first place and for working with me so diligently and patiently. Mostly I'm thankful to my dad, Walt Stewart, and my husband, AJ Righter, for pushing me to finish and to Joseph Ogden for keeping my engine running.

Contents

Chapter	
1	Introduction 1
2	Literature Review 4
2.1	INR 4
2.2	Exploratory Data Analysis 7
2.3	Bayesian Modeling and Inference 9
2.3.1	Gibbs sampling 11
2.3.2	Adaptive Rejection Sampling 11
2.3.3	Goodness-of-fit 12
2.3.4	Comparing Models: DIC 13
3	Methodology 14
3.1	Design of the Patient Management Study 14
3.2	Graphical Analysis 16
3.2.1	Visualizing Longitudinal Patterns: Spaghetti Plots 16
3.2.2	Visualizing Patient-Specific Badness: Bubble Plot 17

3.2.3	Crude Visualization of Autocorrelation: Correlation Plot . . .	18
3.2.4	Better Visualization of Autocorrelation: Semivariogram . . .	18
3.3	Hierarchical Bayesian Modeling and Analysis	19
3.3.1	Beta-Binomial Analysis of Out-of-Range Indicator	20
3.3.2	Bayesian Estimates of INR distributions by patient	21
4	Results	24
4.1	Graphical Analysis	24
4.2	Bayesian Analysis	28
4.2.1	Beta-Binomial Analysis	28
4.2.2	Lognormal Analysis	32
5	Recommendations and Contributions to Statistical Practice	39
5.1	The Study	39
5.2	Graphical Analysis	40
5.3	Statistical Analysis	40
5.4	Medical Practice Implications	41

Appendix

Bibliography	60
---------------------	-----------

Tables

Table

4.1	Alpha and Beta Summary Statistics	32
4.2	β , m , s , and τ Summary Statistics	36

Figures

Figure

- 4.1 INR vs. days for all follow-up times for all patients on all machines 25
- 4.2 Log INR values for each patient in each machine with overlaid typical therapeutic limits of (log)2 and (log)3. Each line connects a patient over time. 26
- 4.3 Log INR values for each patient in each machine with overlaid typical therapeutic limits of (log)2 and (log)3. Lines connect between patients graded the worst, best, median and 25th and 75th percentiles as determined using each patient's average badness. The hardest to manage patient is represented by the heavy red line, the 75th percentile is the light red line, the median patient is shown as the thin gray line, the light black line represents the 25th percentile patient, and the heavy black line represents the best, or most easily managed patient for each machine. 27

4.4	Patients vs. time, by machine. Out-of-range points are red, in-range black, with character size proportional to distance out-of-range. Patients are ordered by badness, with the worst patient (or hardest to manage) being at the top of the graphic, while the easiest to manage patients are at the bottom.	29
4.5	Autocorrelation is the solid line, 95% tolerance limit for zero autocorrelation is the dashed line	30
4.6	Semivariogram of days between measurements by patient.	31
4.7	Goodness-of-Fit Beta-Binomial Model: Overlaid density curves for the data (solid line) and the corresponding posterior predictive distribution (dashed line) of number of INR values in range per patient.	33
4.8	Goodness-of-Fit Lognormal Model: Overlaid density curves for the data (solid line) and the corresponding posterior predictive distribution (dashed line) of badness.	34
4.9	Zero-Added Posterior Predictive Distributions of Badness by Machine	38
5.1	Posterior Predictive Density of Probability In-Range with 95% PI .	42

Chapter 1

Introduction

The International Normalizing Ratio (INR) measures how quickly an individual's blood clots. Healthy people should have an INR of approximately one. However, if people are at higher risk for heart attacks, strokes, and other adverse events, their physician may prescribe a patient-specific therapeutic INR range to lower their risk, generally between two and three (Hirsch et al, 2001). The farther out of their therapeutic range the patient is in either direction, the more dangerous their situation is.

The lab is the gold standard as far as determining what a patient's INR is. However, it takes a long time to draw blood, send it to the lab, have it analyzed, and have the results returned to the doctor so they can make a dosing decision for the patient. Point-of-care (POC) devices have been developed to increase efficiency. In a matter of minutes, a doctor can take a fingerstick, have it analyzed by the POC device, and make a dosing decision. However, the accuracy of the readings is still unclear, as is the doctor's reaction to the reading as compared to the lab value.

The Cleveland Clinic Foundation carried out a study in 2002 to compare 5 POC devices for monitoring INR. The study was a large sample (over 300 patients) longitudinal study. The follow up times were highly irregular, as was the number of follow-ups per patient.

Previous research indicates that the POC devices may not all perform the same with respect to patient management (Sunderji et al. 2005, McBane et al. 2005, Bussey et al. 1997, Kaatz et al. 1995). The purpose of this project was to determine if the POC devices exhibit differences, and if so, which devices manage patients better. As opposed to the accuracy of the POC devices which has been studied elsewhere (Sunderji et al. 2005, McBane et al. 2005, Bussey et al. 1997, Kaatz et al. 1995), the concern of this research was the value of the devices in managing the patients' health.

One important aspect of this research is the patient-specific therapeutic range for INR. With such a unique feature for a highly unbalanced longitudinal dataset, exploratory data analysis through innovative graphical tools was essential. The results of the graphical analysis guided the development of further statistical analysis.

Each patient was nested within a machine, and there was inherent patient variability - some patients were simply easier to manage than others. So even though the randomization of patients to machines largely neutralized the between-patient variability, it was essential to model all sources of variability through a hierarchical model. Bayesian methods inherently deal well with hier-

archical situations.

The purpose of this study was to use a Bayesian hierarchical model to determine if the POC device by which a patient is managed makes a difference in how well that patient can be kept within their prescribed INR therapeutic range. If some devices manage patients better than others, physicians should exercise judgment in their choice of POC device.

Chapter 2

Literature Review

This chapter provides background on the subject matter. The first section deals with the medical background and implications. The later sections illuminate the statistical concepts.

2.1 INR

The INR is calculated by standardizing prothrombin time (PT) which is a measure of the clotting time of a blood sample. PT is measured by testing a blood sample with an enzyme that activates blood clotting factors, such as human placenta or rabbit brain. The formula for standardization is

$$INR = \left(\frac{PT_{patient}}{PT_{meannormal}} \right)^{ISI} \quad (2.1)$$

(Hirsch et al. 2001) where $PT_{meannormal}$ is the mean prothrombin time calculated from a set of normal, non-medicated patients at that particular lab (D'Angelo et al. 1997). The ratio of the patient's PT to this mean PT is raised to the power

of the international sensitivity index (ISI) of the thromboplastin. ISI reflects the responsiveness of the thromboplastin reagent (e.g. rabbit brain) to the World Health Organization standard. A healthy person would have a prothrombin time very near the $PT_{meannormal}$, making the ratio approximately one, regardless of the ISI.

Some people are at higher risk for blood clots. Issues such as venous thrombosis (blood clots in the vascular system), myocardial infarctions (heart attacks), complications due to irregular heart beats (atrial fibrillation), or complications due to major heart surgery (e.g. valve replacement or heart-lung bypass) may encourage a physician to prescribe a therapeutic INR range (Smith 2002). Prophylaxis of venous thrombosis, treatment of venous thrombosis, treatment of pulmonary embolism, prevention of systemic embolism, tissue heart valves, acute myocardial infarction, valvular heart disease, atrial fibrillation, and bileaflet mechanical valve in aortic position generally have a suggested therapeutic INR range of 2.0 - 3.0. Patients with mechanical prosthetic valves or needing to prevent recurrent myocardial infarction often have a higher suggested range such as 2.5 - 3.5.

INR values less than 2.0 for patients undergoing anticoagulation therapy renders the therapy less effective in most cases (Hirsch et al. 2001). Also, too high an INR (greater than 3) puts the patient at higher risk of bleeding complications, without increasing efficacy of the therapy, i.e., reducing clotting incidents (Hyers et al. 2001).

Maintaining a level of INR that reduces the adverse side effects of anti-

coagulation therapy while still enjoying the benefits of therapeutic effectiveness requires careful monitoring. Historically, a patient would visit their doctor and have blood drawn. The blood would then be sent to the lab for analysis. The lab would report back to the doctor, who would then make a dosing decision or other recommendation to the patient to keep their INR within the prescribed therapeutic range. While the gold standard of the lab analysis is accurate, there is a great loss of effectiveness in managing patients, as it takes several days to get lab results back and make dosing recommendations. This may be dangerous to the patient if adjustments need to be made to their treatment.

In response, several POC devices have been developed to monitor patient INR. These machines sit in the doctor's office and with a fingerstick can give an INR reading in a matter of minutes. Many studies have been done to measure and compare the accuracy of these types of machines (Sunderji et al. 2005, McBane et al. 2005, Bussey et al. 1997, Kaatz et al. 1995). All of these studies show varying degrees of disagreement between the POC device(s) and the lab. Collectively among the studies, there was no clear winner in accuracy. Shermock et al. (2002) carried out a study to determine dosing decision agreement between venous samples analyzed in a lab and two point-of-care devices. The bigger picture, however, is determining whether a patient can be safely managed solely using one of these machines, and consequently, if some machines are better than others.

2.2 Exploratory Data Analysis

Exploratory data analysis is essential to visualize patterns in the data (Tukey 1977). Particularly with longitudinal datasets, examining the data graphically is important in uncovering information on the correlation structure (Diggle et al. 1994). Innovation in graphics is essential because every dataset has unique details and attributes that must be illuminated.

The obvious first graph with longitudinal data is a simple scatter plot of the response variable vs. time. Jittering may be useful if the observation times tend to be common for many subjects. However, with large numbers of subjects, each with varying numbers of data points, the graph quickly would become cluttered and fail to properly display trends. Other important elements of the data, such as correlations within patients or differences between machines are impossible to pick out.

By connecting the dots by subject, the trends over time are better expressed. These are often called "spaghetti plots" (Wu and Tsang, 2004). Correlation structures and trends over time, as well as distributional aspects of the data then become slightly more evident. However, with so much data, the clutter often overwhelms the information.

Tufte (1983), emphasizes the data-ink ratio, and that one must not use much more ink than they have information to display. While it may be argued that occasionally a study has sufficient data to warrant lots of ink, often spaghetti

plots or collections of similar curves may be summarized and clarified by selecting a few representative lines or curves, which may be more effective from an information standpoint . Jones and Rice (1997) accomplish this by using principal components analysis to order the curves, and connecting the dots for subjects corresponding to the most extreme and the median of each major principal component. The use of principal components is only feasible with balanced data. Diggle et al. (2002) suggested ordering the curves based on other study-specific criteria.

On occasion, it is imperative to organize huge amounts of data for processing at a glance. Small multiples are series of graphics, each showing the same combination of variables, indexed by changes in another variable (Tufte, 1983). Small multiples invite comparison and are efficiently interpretable.

With longitudinal data, it is essential to explore the autocorrelation function. This is more easily done with balanced data where the data are collected at a fixed number of evenly spaced time points. Diggle et al. (2002) round their observation times to the nearest year, and that liberty may be taken with other slightly unbalanced datasets. The correlation may then be calculated for pairs of observations separated by all possible time intervals within subjects. Of course, the sample autocorrelation is subject to sampling variation, and a measure of variability is important to demonstrate the uncertainty in the estimate. A sample correlation coefficient has standard error of roughly $1/\sqrt{N}$ where N is the number of independent pairs of observations in the calculation under the assumption of zero correlation (Diggle et al. 1994). This permits plotting tolerance limits of zero

autocorrelation to determine if the sample autocorrelation is likely due to chance.

The semivariogram also characterizes the autocorrelation function (Curran 1988). The semivariogram is harder to compute and construct than the autocorrelation plot, but it is not distorted by rounding the observation times when the underlying stochastic process is observed at irregular time-points (Armitage and Colton 1999). Diggle popularized the semivariogram in the context of longitudinal data (Diggle et al. 1988). It is also possible to calculate tolerance limits for the semivariogram when the autocorrelation is zero, simply by computing the overall variance of the variable in question.

2.3 Bayesian Modeling and Inference

Bayesian inference is based, naturally, on Bayes Theorem (Gelman et al. 2003). In its simplest form, Bayes Theorem for two events A and B is

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \quad (2.2)$$

Thomas Bayes (1764) was apparently the first to use this theorem for parameter estimation, by substituting the parameter vector θ and data y for events:

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{f(y)} \quad (2.3)$$

or, more explicitly,

$$\pi(\theta|y) = \frac{f(y|\theta)\pi(\theta)}{\int f(y|\theta)\pi(\theta)d\theta}, \quad (2.4)$$

where $f(y|\theta)$ is the sampling distribution of the data given the parameter θ , $f(y)$ is the marginal distribution of the data, $\pi(\theta)$ is the prior distribution of the parameter

θ , and $\pi(\theta|y)$ is the posterior distribution of the parameter θ given the data y .

Modern Bayesian inference is also based on the posterior predictive distribution for future data,

$$\pi(y^*|y) = \int \pi(y^*|\theta)\pi(\theta|y)d\theta \quad (2.5)$$

where $\pi(y^*|\theta)$ is the sampling distribution for future data y^* given the parameter θ .

Bayesian inference focuses on using the features of the posterior or posterior predictive distribution to draw sensible conclusions. For example, the mean of the posterior distribution may be used as an estimate of the parameter. The highest density region of the predictive posterior distribution may be used as a prediction interval.

While Bayesian methodology was conceived in the 18th century, widespread use was not possible until recent years due to the difficulty in computation. For example, the denominator of equation 2.4, often referred to as the 'normalizing constant' is not easily integrable, even for simple situations, such as the normal distribution with unknown variance. Metropolis et al.(1953) were among the first to develop a practical approach using Markov Chain Monte Carlo (MCMC) simulation. Hastings (1970) generalized Metropolis' algorithm, and the resulting algorithm is commonly referred to as the Metropolis-Hastings algorithm.

2.3.1 Gibbs sampling

Occasionally, a prior distribution is conjugate, which means that by combining the data with the prior distribution, the posterior distribution is of the same family of distributions as the prior. Gibbs sampling involves knowing the full conditional distribution of each random variable, but not knowing the joint distribution (Geman and Geman 1984). If so, samples from complete conditional distributions can be alternately drawn, and the sequence converges to the posterior distribution of the parameters (Gelman et al. 2003). This is the easiest form of Metropolis-Hastings sampling. The WinBUGS software (WinBUGS 14) uses Gibbs sampling to draw from the posterior in these situations.

2.3.2 Adaptive Rejection Sampling

Rejection sampling involves defining an envelope function that can be sampled from rather than sampling from the posterior (Gelman et al. 2003). This method is used if it is too difficult to sample from the actual posterior distribution. To be an appropriate envelope function, the density must be greater than the posterior for all parameter values. A draw from the envelope distribution is accepted with probability equal to the ratio of the height of the posterior to the height of the envelope at that point.

The WinBUGS software uses a variation on rejection sampling when the prior is not conjugate (Gilks and Wild 1992). It is called adaptive rejection sampling because as sampling proceeds, a squeezing function (defined to always be

below the posterior distribution) (Lawson 2005) and the rejection envelope converge to the density function.

2.3.3 Goodness-of-fit

Bayesian modeling has progressed at a rate faster than its diagnostic tests have been able to maintain (Gelman et al. 1996). Several methods have been developed to assess goodness-of-fit of Bayesian models, with varying degrees of applicability and usefulness.

One relatively simple method is to draw samples corresponding to the design of the study from the posterior predictive distribution (equation 2.5) to investigate whether it resembles the data. If predictions are inconsistent with the observed data evidence is provided against model assumptions (Lynch and Western 2004). The posterior predictive distribution used in goodness-of-fit testing is commonly denoted

$$p(y^{rep}|y) = \int p(y^{rep}|\theta)p(\theta|y)d\theta, \quad (2.6)$$

where $p(y^{rep}|\theta)$ is the sampling distribution of the data, $p(\theta|y)$ is the posterior distribution of the parameters, and y^{rep} is the vector of predictions corresponding to the design of the study. A Q-Q plot (Baggerly 1998) or other graphical device is used to compare the distributions of y and y^{rep} . If the distributions are similar, it can be concluded that the model fits.

2.3.4 Comparing Models: DIC

In order to compare various Bayesian models, Spiegelhalter et al. (2002) derived a "Deviance Information Criterion" (DIC). If θ is the parameter vector in the model, let \bar{D} be the posterior mean of the deviance defined as negative twice the log likelihood of the data, i.e., $-2 * \log(p(y|\theta))$. \hat{D} is a point estimate of the deviance obtained by substituting in the posterior means $\bar{\theta}$ of θ , thus $\hat{D} = -2 * \log(p(y|\bar{\theta}))$. Then

$$DIC = \bar{D} + p_D = \hat{D} + 2 * p_D \quad (2.7)$$

where p_D is $\bar{D} - \hat{D}$, the number of effective parameters in the model. By penalizing for useless parameters, much like its sister criteria AIC and BIC, the DIC is useful both for model comparison and a measure of goodness-of-fit (Spiegelhalter et al. 2003).

Chapter 3

Methodology

This chapter discusses the methods used to collect and analyze the data.

3.1 Design of the Patient Management Study

In 2002, the Cleveland Clinic Foundation Anticoagulation Clinic in cooperation with the Department of Biostatistics and Epidemiology designed a double-blind randomized trial to compare patient management on five different point-of-care devices. Each of 371 patients receiving anticoagulation therapy was randomly assigned to one of five machines: (1) the ProTime Microcoagulation System (ProTime device; International Technidyne Corporation Limited, Edison NJ 08820), (2) Rapidpoint Coag (Rapidpoint machine; Bayer Healthcare, Newbury, Berkshire, RG14 1JA, UK), (3) the Hemochron Microcoagulation System (Hemochron device; International Technidyne Corporation Limited, Edison NJ 08820), (4) CoaguChek S System (CoaguChek; Roche Diagnostics, F. Hoffmann-La Roche Ltd, Basel, Switzerland), and (5) CoaguChek ProDM (CoaguChek; Roche Diagnostics,

F. Hoffmann-La Roche Ltd, Basel, Switzerland). Aside from the POC measurement, each patient had venous samples sent to a reference laboratory (the "gold standard"). The venous samples were analyzed at the Cleveland Clinic Hematology Laboratory using the Sysmex CA-6000 Coagulation Analyzer (Sysmex, Kobe, Japan). The thromboplastin reagent used to determine prothrombin time was Innovin (Dade Behring), which had an ISI of approximately 1.0. The gold standard lab values comprise the dataset for this project.

This was a longitudinal study. At irregular times over a period of time ranging from 14 to 152 days, patients were managed on the POC reading alone. The physicians were not informed of the lab results unless there was a huge discrepancy, in which case the patient was removed from the study.

A weakness of this dataset is that there is no record of dosing decisions. Also, there is no "doctor" variable, though there are pharmacist and lab tech variables, neither of which will be used in the current analysis.

Due to the irregular and unbalanced nature of the repeated measures, and to a patient-specific measure of "badness" (absolute distance out of the patient's therapeutic range), innovative analyses were required for this project. Initially, the data were examined graphically in several ways (section 3.2). The graphics helped to formulate useful statistical models (section 3.3) to compare the five POC devices.

3.2 Graphical Analysis

The dataset consisted of highly unbalanced and irregularly spaced longitudinal data. For each patient a specified central therapeutic range was "good", and the tails of the distribution represented a graded danger to the patient. Due to this, a careful and informative graphical analysis was a necessary component in order to better understand how to statistically model the data.

3.2.1 Visualizing Longitudinal Patterns: Spaghetti Plots

The initial scatter plot with horizontal jittering revealed very little about the pattern or the correlations. In the initial spaghetti plot, INR was plotted over time as small multiples by machine, with lines connecting values for each patient. Horizontal lines were overlaid on this plot at $\text{INR} = 2$ and $\text{INR} = 3$ because those are the most common INR therapeutic limits in our dataset (approx. 85%). In looking at that plot, it was clear that the data were right skewed. After log-transforming the INR, the distributions looked much more symmetric.

However, the spaghetti plots were still too crowded to be useful. We ranked subjects according to the average absolute distance out of therapeutic range. We then selected as representative curves (Jones and Rice 1992) the worst, the best, the median, and the 25th and 75th percentiles subjects per machine. We connected values only for those patients (appendix 1). While Jones and Rice (1992) utilized principal components to order their curves, this dataset called for the unique measure of badness by patient described above.

3.2.2 Visualizing Patient-Specific Badness: Bubble Plot

To get a useful depiction of how well each machine manages its patients, a plot showing the "badness" pattern of each machine was imperative. We used the graphical elements of color, symbol size, sorting, and small multiples to create an informative graphic.

Using the log of the INR, the distance outside of the log transformed therapeutic range was calculated for each data point. These were then summed up for each patient. Since the data were highly unbalanced, both in the spacing of observations and the number of observations per patient, it was necessary to divide the sum of the distance out-of-range for each patient by the number of data points for that patient as a summary measure of how poorly that patient was managed. Within machines, patients were then sorted by this summary measure.

The purpose of sorting patients by badness was to visualize how each machine performed. In plotting patients over time in order of badness, a pseudo-pareto longitudinal chart was created. This facilitated visual comparison among machines.

Color was used to distinguish data points within the prescribed therapeutic range (black) from those which were not (red). Lots of black points are good, lots of red points are bad. This was not sufficient, however, because as previously mentioned, how far a data point is out-of-range is of interest as well. Hence, the size of the graphical symbol was used to visualize the distance out of log therapeutic range of the log INR. The area of the out-of-range symbol was made

proportional to the distance out-of-range. Because the plotting symbol was a circle, we termed this a 'bubble plot' (appendix 1)

3.2.3 Crude Visualization of Autocorrelation: Correlation Plot

Exploring the autocorrelation function of longitudinal data is easiest with balanced data. In our first effort to visualize the autocorrelation, we artificially balanced the INR data by rounding the observation times to the nearest week. In the rare case where there was more than one observation in a week, the average INR of the two observations was used. Often there were no observations in a particular week, and these were treated as missing values.

Following that, all observations within patient were paired. Then pairs were grouped by their time lag, and correlations were calculated for each group. Correlations were then plotted versus time lag. As a reference, the upper 95% tolerance limit assuming zero autocorrelation was overlaid on the plot. If the sample autocorrelation was higher than the tolerance limit, the correlation was considered significant at the 5% level.

3.2.4 Better Visualization of Autocorrelation: Semivariogram

The semivariogram is a refinement of the autocorrelation plot described previously. It requires no rounding of the time variable, giving a less distorted visualization.

Within each patient, all possible pairings of INR values were made. The

time lag for these pairs was used as the horizontal axis on the graph. The vertical axis represented half the squared difference of the log INR values. Residuals from a crude model for the mean are often used as the vertical axis (Armitage and Colton 1999). However, the log INR values themselves were used in this case, under the assumption that machines and other effects were relatively small.

A smooth line using a spline routine (SAS Institute 2000) was fitted and overlaid on this plot to demonstrate trends in the autocorrelation over time. A 95% tolerance limit corresponding to zero autocorrelation was also calculated and overlaid on the plot.

3.3 Hierarchical Bayesian Modeling and Analysis

The overall strategy of the Bayesian modeling was to fit two models using MCMC, one allowing for differences among machines, the other assuming that machines were identical. We then carried out goodness-of-fit testing for the more complex model, the machine-specific model. If the fit was satisfactory, we then chose one of the two models using DIC. If the machine-specific model was chosen, we computed and compared posterior predictive distributions for the 5 POC devices. All of the calculations were done in WinBUGS 14, and goodness-of-fit plots were constructed using BRUGS (appendix 1).

3.3.1 Beta-Binomial Analysis of Out-of-Range Indicator

On a simple level, the more often a patient's INR is out of their prescribed therapeutic range, the worse the machine. This definition ignores distance out of range, which increases danger of adverse effects to patients, but it allows a valid initial analysis of the data.

Let y_{ij} be the number of in-range INR values for patient i managed on machine j . Let n_{ij} be the total number of INR measurements for patient i on machine j . Plots (figure 4.6) indicated the autocorrelation function was flat. Therefore, a reasonable machine-specific hierarchical Bayesian model for y_{ij} was

$$y_{ij} | \pi_{ij} \sim \text{binom}(\pi_{ij}, n_{ij})$$

$$\pi_{ij} \sim \text{beta}(\alpha_j, \beta_j)$$

$$\alpha_j \sim \text{gamma}(.001, .001)$$

$$\beta_j \sim \text{gamma}(.001, .001)$$

Note that vague prior distributions were used for all parameters.

The corresponding machine-identical model was:

$$y_{ij} \sim \text{Bin}(\pi_{ij}, n_{ij}),$$

$$\pi_{ij} \sim \text{Beta}(\alpha, \beta),$$

$$\alpha \sim \text{Gamma}(.001, .001),$$

$$\beta \sim \text{Gamma}(.001, .001).$$

Both models were fitted to the data using WinBUGS 14 (appendix 1). A burn-in of 1000 iterations was used. If the trace plots indicated convergence, 50,000 iterations were used to sample from the the posterior distributions of the parameters.

Goodness-of-fit was tested by drawing samples from the posterior predictive distribution in the same proportions among machines in the dataset. DIC was calculated for both models. Finally, 50,000 samples were drawn from appropriate posterior predictive distributions.

3.3.2 Bayesian Estimates of INR distributions by patient

While knowing if a patient was out-of-range was important, more can be learned from the dataset. The farther out-of-range a patient was, the more dangerous their situation. It became essential, then, to estimate the distribution of each patient's INR, and compare expected danger across machines.

Since INR is bounded below by zero and is strongly right skewed, it was reasonable to assume a lognormal distribution. The mean of the lognormal distribution was assumed to be normally distributed, and the precision parameter ($1/\textit{variance}$) was assumed to be gamma distributed. Parameters of each of these distributions were assigned hyperpriors that were diffuse enough to largely allow the data to determine the posterior distribution. If z_{ijk} is the INR of patient i on machine j at visit k and x_{ij} is the lower bound of the therapeutic range of patient

i on machine j , the machine specific model was:

$$z_{ijk} | \phi_{ij}, \beta, \tau_{ij} \sim \text{lognormal}(\phi_{ij} + \beta \log x_{ij}, \tau_{ij})$$

$$\phi_{ij} | m_j, s_j \sim \text{normal}(m_j, s_j)$$

$$m_j \sim \text{normal}(0, .0001)$$

$$s_j \sim \text{gamma}(.0001, .0001)$$

$$\beta \sim \text{normal}(0, .0001)$$

$$\tau_j \sim \text{gamma}(.0001, .0001)$$

.

A slightly simpler machine-specific model ignored the lower bound of the therapeutic range:

$$z_{ijk} | \mu_{ij}, \tau_j \sim \text{lognormal}(\mu_{ij}, \tau_j)$$

$$\mu_{ij} | m_j, s_j \sim \text{normal}(m_j, s_j)$$

$$m_j \sim \text{normal}(0, .0001)$$

$$s_j \sim \text{gamma}(.0001, .0001)$$

$$\tau_j \sim \text{gamma}(.0001, .0001)$$

The machine-identical model was very similar:

$$z_{ijk} | \mu_i, \tau \sim \text{lognormal}(\mu_i, \tau)$$

$$\mu_i | m, s \sim \text{normal}(m, s)$$

$$m \sim \text{normal}(0, .0001)$$

$$s \sim \text{gamma}(.0001, .0001)$$

$$\tau \sim \text{gamma}(.0001, .0001)$$

Goodness-of-fit was tested by drawing samples from the posterior predictive distribution of INR for the machine-specific model, and graphically comparing the draws to the data. DIC was calculated for all three models.

Devices were compared by estimating posterior predictive distributions of badness for each device. Badness was defined as the absolute distance out of therapeutic range for each visit.

Chapter 4

Results

This chapter offers the results of the graphical and Bayesian analyses.

4.1 Graphical Analysis

The initial scatterplot of INR over time (Figure 4.1) showed very little trend. Within patient correlations were not evident from the plot.

When the values of each figure were joined it became clear that there were consistent differences among patients (Figure 4.2). The irregular nature of the follow-ups was obvious. It was also clear that most of the data fell within the INR limits of 2-3. Positive deviations were more rare, but often of larger magnitude.

After ordering each patient by badness within machine, the minimum, maximum, median, first and third quartiles were selected. Those patients' INR values were plotted as lines in Figure 4.3, with points for all other INR values. These plots were not very informative.

The pseudo-pareto or bubble plots were plotted in Figure 4.4. After ordering patients by badness within machine, patients were plotted against time,

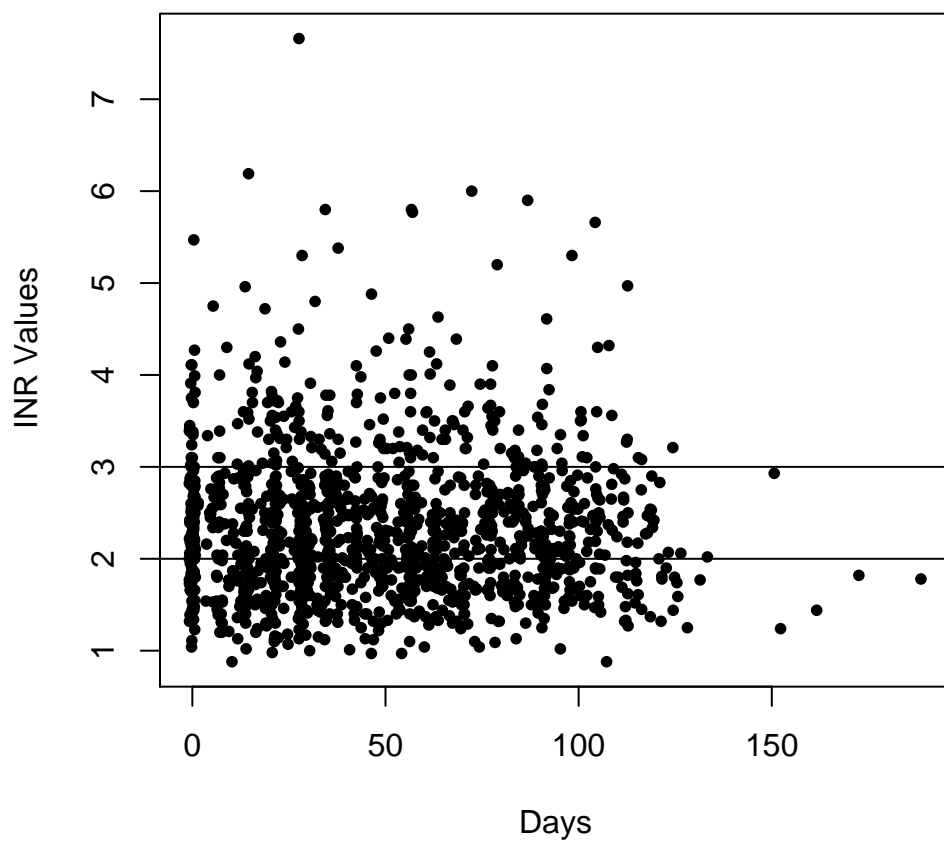


Figure 4.1: INR vs. days for all follow-up times for all patients on all machines

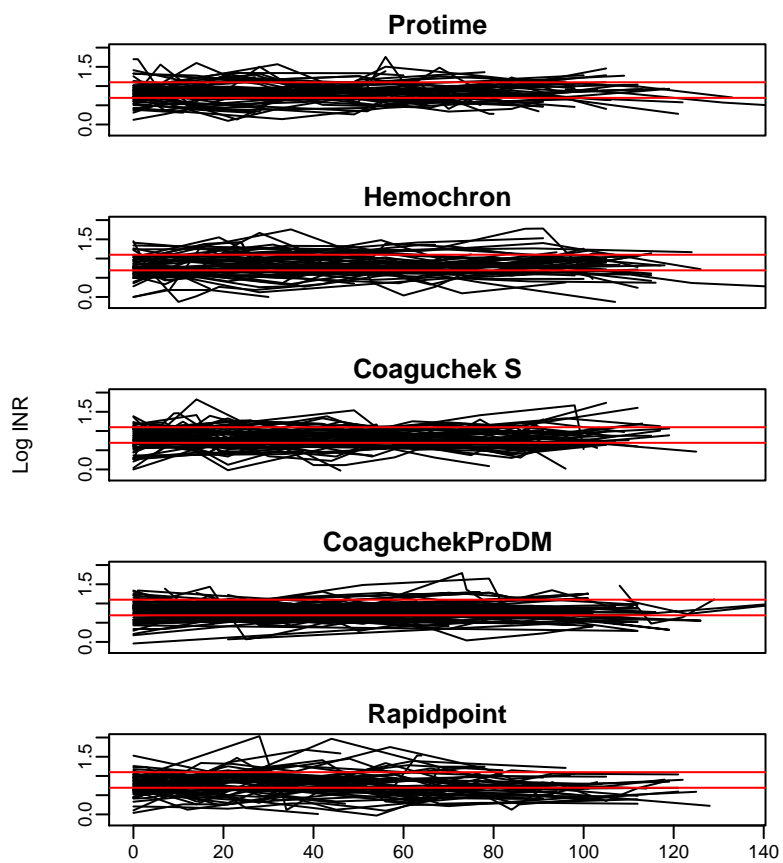


Figure 4.2: Log INR values for each patient in each machine with overlaid typical therapeutic limits of $(\log)2$ and $(\log)3$. Each line connects a patient over time.

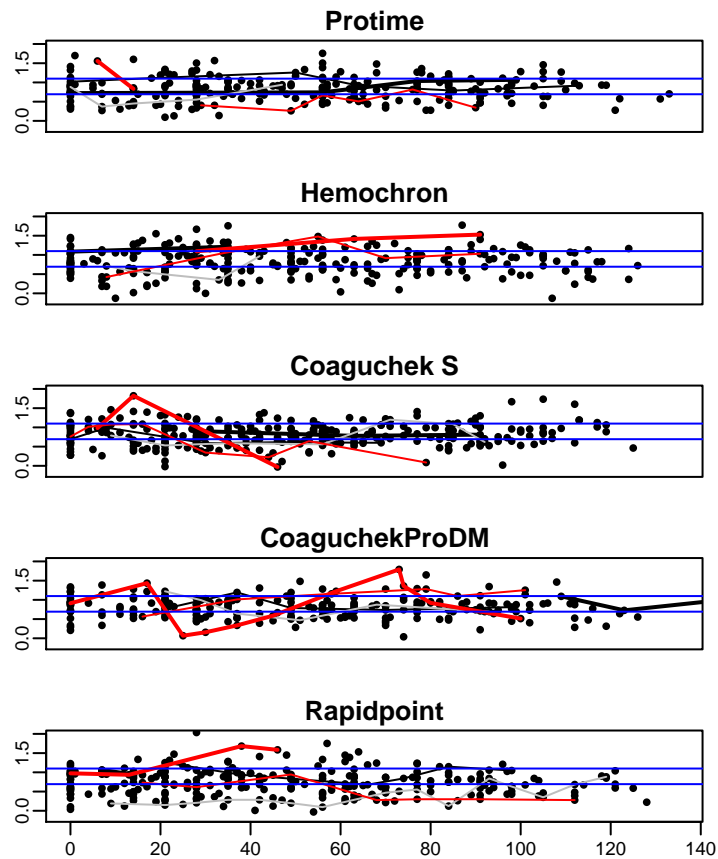


Figure 4.3: Log INR values for each patient in each machine with overlaid typical therapeutic limits of $(\log)2$ and $(\log)3$. Lines connect between patients graded the worst, best, median and 25th and 75th percentiles as determined using each patient's average badness. The hardest to manage patient is represented by the heavy red line, the 75th percentile is the light red line, the median patient is shown as the thin gray line, the light black line represents the 25th percentile patient, and the heavy black line represents the best, or most easily managed patient for each machine.

with the character color representing in-range (black) or out-of-range (red), and the character size being in proportion to the distance out-of-range. Bubble plots were very informative. There were no drastic differences among machines. There appear to be approximately the same proportions of harder-to-manage and easier-to-manage patients across machines, with a few slight exceptions. Rapidpoint seemed to have a higher proportion of large red bubbles. Coaguchek S seemed to have red bubbles extending further down the graph.

Figure 4.5 is the correlation plot. Each doctor visit was rounded to the nearest week, and the correlation between INR values within patient was calculated. There appeared to be constant autocorrelation for the first 13 weeks. Beyond that, we lack statistical power, since there were so few observations more than 13 weeks apart.

The semivariogram is the refined correlation plot - no rounding was necessary (Figure 4.6). There appears to be relatively constant autocorrelation. The observation of constant autocorrelation led to the inclusion of patient-specific parameters in the Bayesian models but no further modeling of the autocorrelation.

4.2 Bayesian Analysis

4.2.1 Beta-Binomial Analysis

After a burn-in period of 1000 iterations, 50,000 iterations were monitored for both of the beta-binomial models. Trace plots showed adequate convergence.

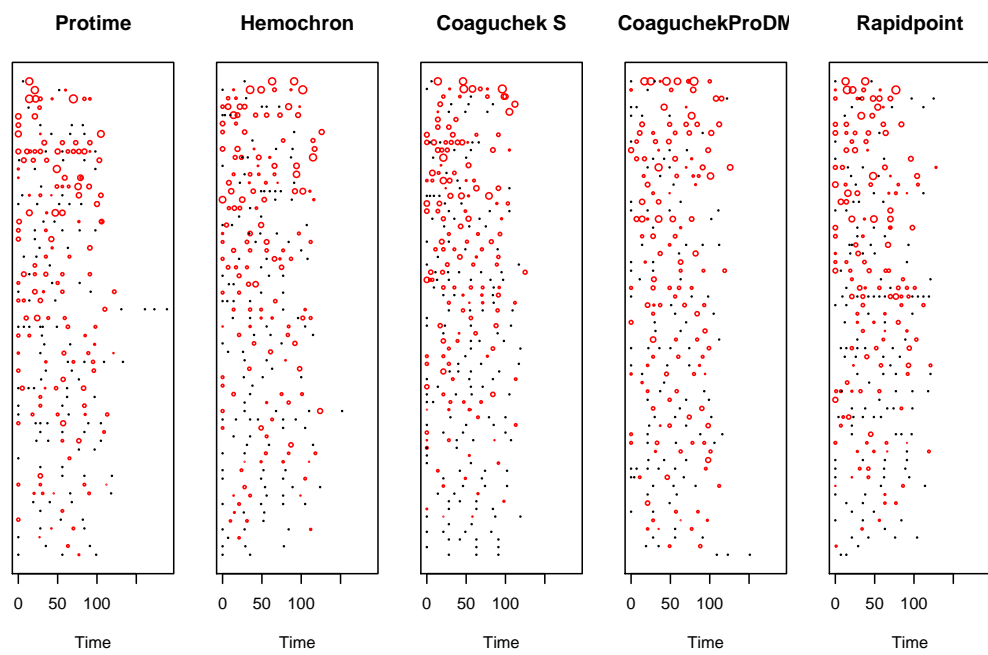


Figure 4.4: Patients vs. time, by machine. Out-of-range points are red, in-range black, with character size proportional to distance out-of-range. Patients are ordered by badness, with the worst patient (or hardest to manage) being at the top of the graphic, while the easiest to manage patients are at the bottom.

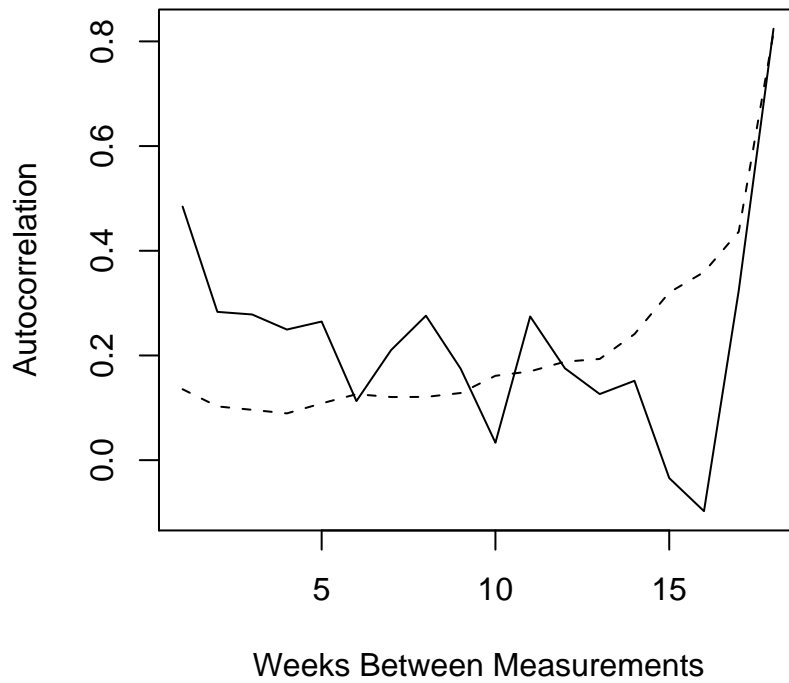


Figure 4.5: Autocorrelation is the solid line, 95% tolerance limit for zero autocorrelation is the dashed line

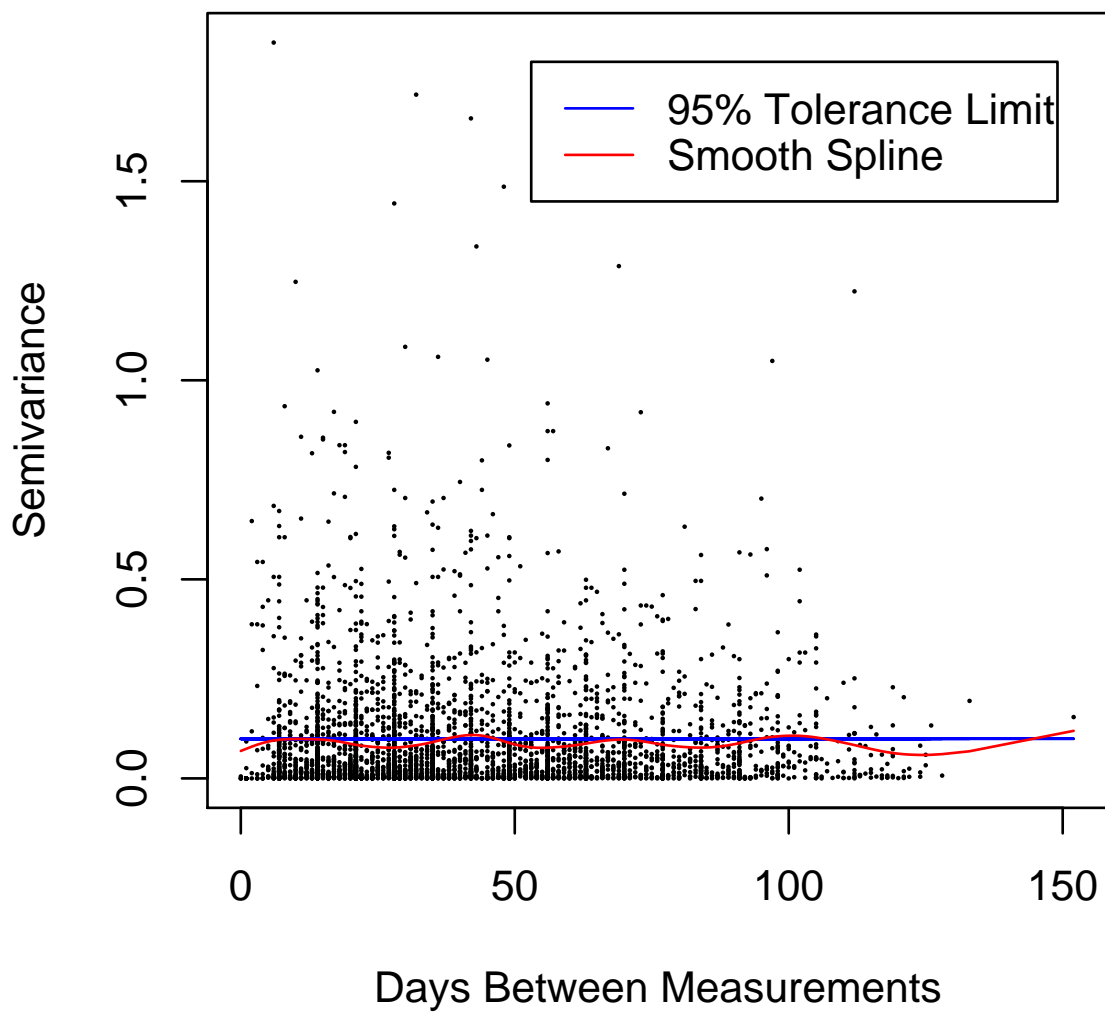


Figure 4.6: Semivariogram of days between measurements by patient.

Goodness-of-fit investigations (Figure 4.7) show that the data is sufficiently modeled by the machine-specific beta-binomial model, since the density of the INR agreed well with the density of the INR^{rep}.

The machine-identical model in the beta binomial analysis had a smaller DIC (882.4) than the DIC from the machine-specific model (893.0). Therefore, we conclude that the probability of a patient being within their therapeutic range is independent of POC device.

The estimate for the mean proportion of visits in range across individuals was 0.4634, with a 95% probability interval of (0.43110, 0.4960). The 95% limits of the distribution of proportion in-range for individuals ranged from 0.29166 to 0.63567 using this model. Table 4.1 contains the posterior estimates of α and β .

Statistics	Mean	SD	Median
Alpha	4.187	1.471	3.854
Beta	4.858	1.746	4.457

Table 4.1: Alpha and Beta Summary Statistics

4.2.2 Lognormal Analysis

Like the beta-binomial models, the lognormal models each had a burn-in of 1000 iterations followed by 50,000 iterations. The trace plots for each of the variables showed convergence (appendix 2). The goodness-of-fit of the most complex model (machine-specific including a term for the lower therapeutic bound) was satisfactory (Figure 4.8).

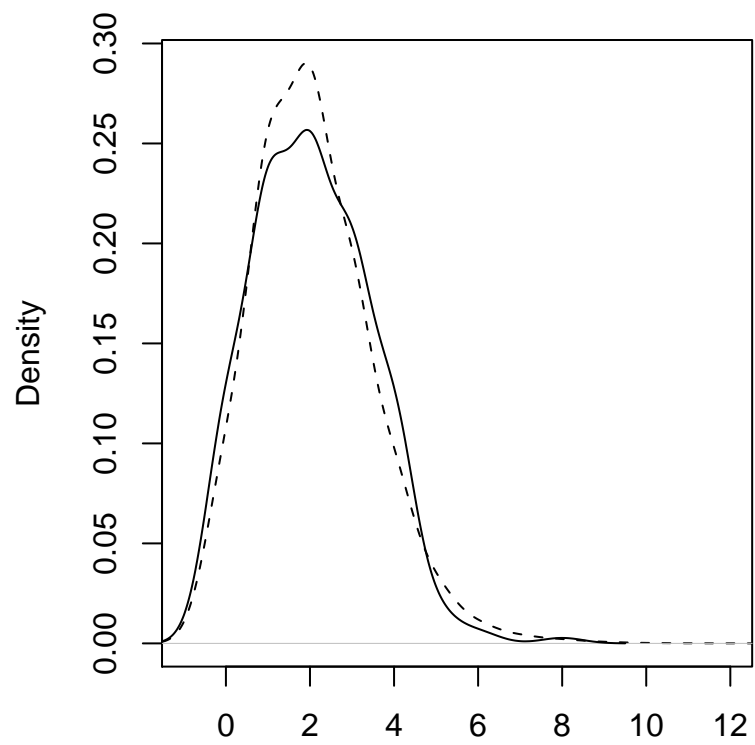


Figure 4.7: Goodness-of-Fit Beta-Binomial Model: Overlaid density curves for the data (solid line) and the corresponding posterior predictive distribution (dashed line) of number of INR values in range per patient.

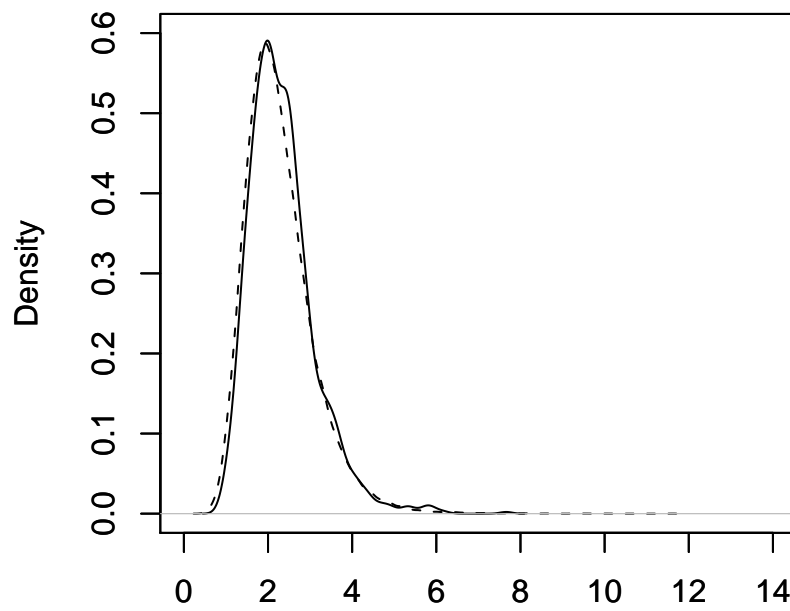


Figure 4.8: Goodness-of-Fit Lognormal Model: Overlaid density curves for the data (solid line) and the corresponding posterior predictive distribution (dashed line) of badness.

The DIC for the machine-identical model was 3389. The machine-specific model without a term for therapeutic lower bound had a DIC of 3232. The machine-specific model with a term for therapeutic lower bound had a DIC value of 3228. The most complex model had the lowest DIC making it the best fitting model of the three. Thus we conclude that machines were different and that the lower bound was predictive.

Of greatest interest were the posterior predictive distributions of badness, defined as the absolute distance of a patient's lab INR value outside the patient's therapeutic range. If a patient's INR is in-range, their badness measurement is zero for that visit.

Every draw from a posterior predictive distribution was a random draw from a patient on one of the five POC devices. For simplicity, each patient was assumed to have the prescribed therapeutic range of 2-3, since this was by far the most common in the actual dataset. Each POC device had 10,000 simulated patients with one draw of INR each. Table 2 contains the parameter estimates of β , the m 's, the s 's, and the τ 's.

The posterior predictive densities for the lognormal model are displayed as zero-inflated distributions (Figure 4.9) (Rodrigues 2003). The bars at zero on the badness plots reveal the differences between machines for the proportion of draws with zero badness. Trailing out to the right are the densities of positive badness. In comparing badness across machines, it seems that while there is no clear winner amongst machines although three of them, Protime, Hemochron,

Statistic	Mean	SD	Median
β	-0.03216	.03980	-0.03185
m [Protime]	0.8242	.02785	0.8242
m [Hemochron]	0.8224	.03114	0.8220
m [Coaguchek S]	0.7254	.05434	0.7251
m [CoaguchekProDM]	0.7728	.1042	0.7724
m [Rapidpoint]	0.7132	.1554	0.7146
s [Protime]	5681.0	6118.0	3600.0
s [Hemochron]	1866.0	3639.0	636.3
s [Coaguchek S]	1810.0	3758.0	362.4
s [CoaguchekProDM]	785.7	2514.0	26.6
s [Rapidpoint]	726.8	2528.0	36.2
τ [Protime]	10.50	.4776	10.49
τ [Hemochron]	10.110	.6373	10.09
τ [Coaguchek S]	7.68	1.481	7.53
τ [CoaguchekProDM]	761.40	2618.0	23.76
τ [Rapidpoint]	871.50	2807.0	40.86

Table 4.2: β , m , s , and τ Summary Statistics

and CoaguchekProDM, are somewhat superior to the other two, Rapidpoint and Coaguchek S. The bars for zero badness are higher for these three, and the positive badness densities are not only smaller (as would be necessary with higher zero bars), but also do not extend out as far as the other machines. Coaguchek S had a high density of lower-range positive badness. Rapidpoint also had a higher positive density, but the density had a thicker tail, implying that patients went farther out of range occasionally. This agrees with previous observations made from the bubble plot.

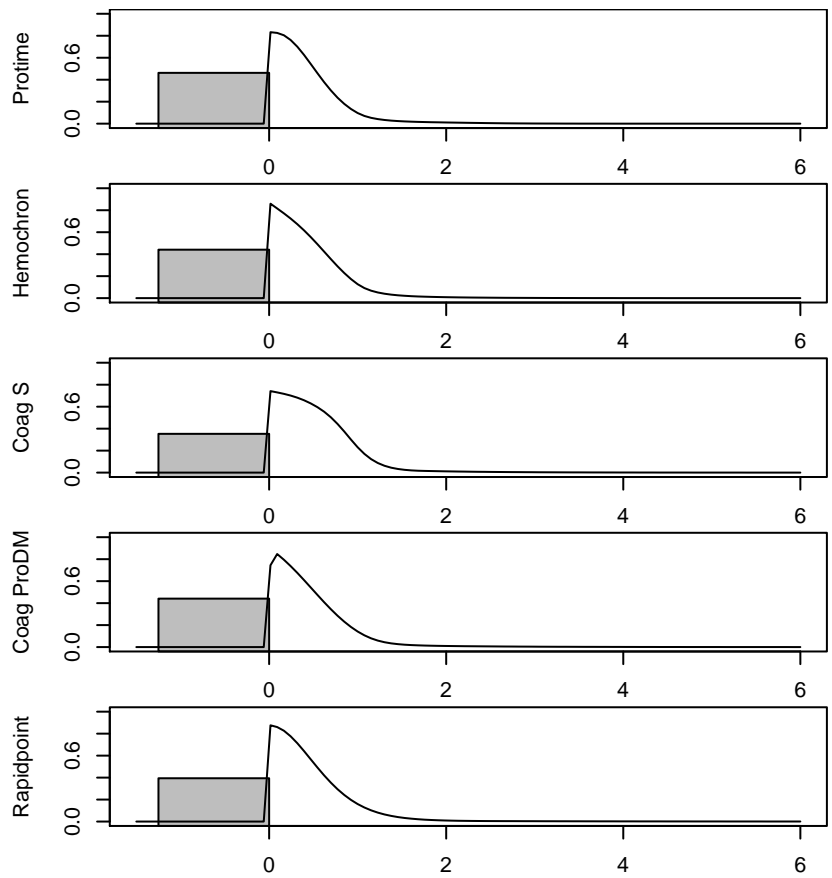


Figure 4.9: Zero-Added Posterior Predictive Distributions of Badness by Machine

Chapter 5

Recommendations and Contributions to Statistical Practice

This chapter discusses the results and their implications and suggests possible areas of future research. The first section examines some of the weaknesses of the dataset and offers remedies for future studies. The next sections outline further developments for statistical analysis of data of this nature.

5.1 The Study

While the study design was excellent, in a trial this large and long, simple, balanced data would be impossible to expect. However, some variables that may have been valuable in the analysis, like prescription adjustments, were not recorded.

By examining patient management, this project took a wider view on the outcome than the traditional approach of comparing accuracy of machine. To broaden this further, studies of life and health of patients over a very long period of time would yield valuable information.

5.2 Graphical Analysis

Innovative graphs were employed in an attempt to visually summarize this unbalanced longitudinal data set. The bubble plot was useful in summarizing the entire dataset and picking out trends. The bubble plot was more informative than the other plots, and dealt well with the unbalanced nature of the dataset. Therapeutic ranges were similar across patients, but the differences were accounted for in the bubble plot. Other patient management data with widely differing therapeutic ranges or other elements may benefit from the bubble plot. There may be other plots that could better display the trends and differences. The development of more graphical tools for patient management is suggested.

5.3 Statistical Analysis

The hierarchical Bayes models and WinBUGS dealt well with the large, complex dataset. It was a little surprising to learn that the lognormal model showed differences among machines whereas the beta-binomial model did not. The lognormal model contained additional information about the patient management and was therefore a more powerful model.

Improvements can be made through automation of the techniques we employed. Investigating other definitions of badness may be helpful.

5.4 Medical Practice Implications

Slightly fewer than half of the INR measurements can be expected to be within the prescribed therapeutic range for the average patient (Figure 5.1). A physician should not be surprised if a few patients are in range as little as 29% of the time. They should not expect patients to be in range more than 69% of the time. All else being equal, the better machines to manage a patient on are Protine, Hemochron, and CoaguchekProDM, but only because patients on those machines tended to be in range and closer to their prescribed range more often than the patients on the other machines.

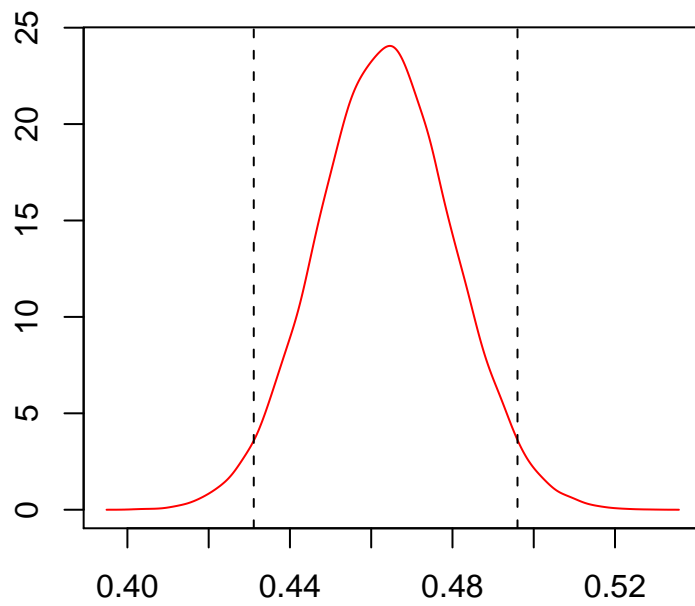


Figure 5.1: Posterior Predictive Density of Probability In-Range with 95% PI

Appendix 1

BUGS code:

Beta-Binomial Machine-Identical Model:

```
model for (i in 1:N) {  
in.range[i] ~ dbin(p[i], tot[i])  
p[i] ~ dbeta(alpha[i], beta[i])  
p.obs[i] < -in.range[i]/tot[i]  
alpha[i] < -A + 0 * Id[i] + 0 * Pro[i] + 0 * Hem[i] + 0 * CoS[i] + 0 * CoP[i] + 0 * Rap[i]  
beta[i] < -B + 0 * Id[i] + 0 * Pro[i] + 0 * Hem[i] + 0 * CoS[i] + 0 * CoP[i] + 0 * Rap[i]  
A ~ dgamma(.001, .001)  
B ~ dgamma(.001, .001)  
PP < -A/(A + B)}
```

Beta-Binomial Machine-Specific Model:

```
model for (i in 1:N) {  
in.range[i] ~ dbin(p[i], tot[i])  
p.obs[i] < -in.range[i]/tot[i]  
p[i] ~ dbeta(alpha[i], beta[i])  
alpha[i] < -A[1] * Pro[i] + A[2] * Hem[i] + A[3] * CoS[i] + A[4] * CoP[i] + A[5] * Rap[i] + 0 * Id[i]  
beta[i] < -B[1] * Pro[i] + B[2] * Hem[i] + B[3] * CoS[i] + B[4] * CoP[i] + B[5] * Rap[i] + 0 * Id[i]  
A[1] ~ dgamma(.001, .001)  
A[2] ~ dgamma(.001, .001)  
A[3] ~ dgamma(.001, .001)  
A[4] ~ dgamma(.001, .001)  
A[5] ~ dgamma(.001, .001)  
B[1] ~ dgamma(.001, .001)
```

$B[2] \sim \text{dgamma}(.001, .001)$
 $B[3] \sim \text{dgamma}(.001, .001)$
 $B[4] \sim \text{dgamma}(.001, .001)$
 $B[5] \sim \text{dgamma}(.001, .001)$
 $PP[1] < -A[1]/(A[1] + B[1])$
 $PP[2] < -A[2]/(A[2] + B[2])$
 $PP[3] < -A[3]/(A[3] + B[3])$
 $PP[4] < -A[4]/(A[4] + B[4])$
 $PP[5] < -A[5]/(A[5] + B[5])\}}\}$

Lognormal Machine-Identical Model:

```

model
{
  m ~ dnorm(1,.001)
  s ~ dgamma(.001,.001)
  tau ~ dgamma(.001,.001)
  for(iin1 : N){
    mu[i] ~ dnorm(m, s)
    for(jin1:17){
      inr[i, j] ~ dlnorm(mu[i], tau)
    }
  }
}

```

Lognormal Machine-Specific Model:

```

model
{
  beta ~ dnorm(0, .0001)
  tau ~ dgamma(.0001, .0001)
}

```

```

for(iin1 : mach){
m[i] ~ dnorm(1, .0001)
s[i] ~ dgamma(.0001, .0001)
med[i] < -exp(m[i] + beta * log(2))
mean[i] < -exp(m[i] + beta * log(2) + 1/2/s[i])
std[i] < -mean[i] * sqrt(exp(1/s[i]) - 1)
for(jin1 : sub){
mu[i, j] ~ dnorm(m[i], s[i])
meanie[i, j] < -(beta * log(lower[i, j])) + mu[i, j]
for(kin1 : vis){
inr[i, j, k] ~ dlnorm(meanie[i, j], tau)
}}}}

```

Lognormal Machine-Specific Model with Lower Bound Term:

```

model
{
beta ~ dnorm(0, .0001)
for(iin1 : mach){
m[i] ~ dnorm(1, .0001)
s[i] ~ dgamma(.0001, .0001)
tau[i] ~ dgamma(.0001, .0001)
for(jin1 : sub){
mu[i, j] ~ dnorm(m[i], s[i])
meanie[i, j] < -(beta * log(lower[i, j])) + mu[i, j]
for(kin1 : vis){
inr[i, j, k] ~ dlnorm(meanie[i, j], tau[i])
}}}}

```

R code:

Beta-Binomial model:

```
data1 <- read.csv("data1.csv")
```

```
attach(data1)
```

```
n1 <- sum(Pro)
```

```
n2 <- sum(Hem)
```

```
n3 <- sum(CoS)
```

```
n4 <- sum(CoP)
```

```
n5 <- sum(Rap)
```

```
N <- sum(n1,n2,n3,n4,n5)
```

```
# SEPARATE TOT BY MACHINE
```

```
tot1 <- rep(99,N)
```

```
tot2 <- rep(99,N)
```

```
tot3 <- rep(99,N)
```

```
tot4 <- rep(99,N)
```

```
tot5 <- rep(99,N)
```

```
for (i in 1:N)
```

```
{
```

```
if (Pro[i]==1) tot1[i] <- tot[i]
```

```
}
```

```
j <- 1
```

```
for (i in 1:N)
```

```
{
```

```

if (tot1[i] != 99)
{
tot1[j] < - tot1[i]
j i- j+1
}
}
tot1 < - tot1[1:n1]

```

(repeat for machines 2-5)

SEPARATE IN.RANGE BY MACHINE

```

in.range1 < - rep(99,N)
in.range2 < - rep(99,N)
in.range3 < - rep(99,N)
in.range4 < - rep(99,N)
in.range5 < - rep(99,N)

```

```

for (i in 1:293)
{if (Pro[i]==1) in.range1[i] < - in.range[i] else in.range1[i] < - 99}
for (i in 1:293)
{if (Hem[i]==1) in.range2[i] < - in.range[i] else in.range2[i] < - 99}
for (i in 1:293)
{if (CoS[i]==1) in.range3[i] < - in.range[i] else in.range3[i] < - 99}
for (i in 1:293)
{if (CoP[i]==1) in.range4[i] < - in.range[i] else in.range4[i] < - 99}
for (i in 1:293)

```



```
{if (Rap[i]==1) in.range5[i] < - in.range[i] else in.range5[i] < - 99}
```

```
    j < - 1
```

```
for (i in 1:N)
```

```
{
```

```
if (in.range1[i] != 99)
```

```
{
```

```
in.range1[j] < - in.range1[i]
```

```
j < - j+1
```

```
}}
```

```
in.range1 < - in.range1[1:n1]
```

```
(repeat for machines 2-5)
```

```
# START BRUGS
```

```
library(BRugs)
```

```
modelCheck('model2betabinom.txt')
```

```
modelData('data1.txt')
```

```
modelCompile(1)
```

```
modelInits( 'init2.txt' )
```

```
modelGenInits()
```

```
modelUpdate(1000)
```

```
dicSet()
```

```
samplesSet('PP')
```

```
samplesSet('p')
```

```

modelUpdate(50000)

dic2 <- dicStats()

PP1 <- data.frame(samplesSample('PP[1]'))
PP2 <- data.frame(samplesSample('PP[2]'))
PP3 <- data.frame(samplesSample('PP[3]'))
PP4 <- data.frame(samplesSample('PP[4]'))
PP5 <- data.frame(samplesSample('PP[5]'))

par(mfrow=c(1,1))
plot(density(PP1[,1]),ylim=c(0,12),xlim=c(.25,.75),xlab="",main="")
lines(density(PP2[,1]),col=2)
lines(density(PP3[,1]),col=3)
lines(density(PP4[,1]),col=4)
lines(density(PP5[,1]),col=5)
legend(x=.6,y=12,col=1:5,lty=1,legend=c("Pro", "Hem", "CoS", "CoP", "Rap"))

samplesDensity("PP")
samplesHistory("PP")
samplesStats("**") # the summarized results

# SAMPLE FROM POSTERIOR PREDICTIVE DISTRIBUTIONS

J <- rep(tot1, each = 100)
NJ <- length(J)

```

```

y.rep1 <- rep(999, NJ)

  for (i in 1:NJ)
  {
y.rep1[i] <- rbinom(1,J[i],PP1[i,1])
  }

  (repeat for machines 2-5)

par(mfrow=c(3,2),mar=c(2, 4, 1, 2) + 0.1)

plot(density(y.rep1,bw=.5),xlim=c(0,12),main="",xlab="")
lines(density(in.range1),lty=2)
plot(density(y.rep2,bw=.5),xlim=c(0,12),main="",xlab="")
lines(density(in.range2),lty=2)
plot(density(y.rep3,bw=.5),xlim=c(0,12),main="",xlab="")
lines(density(in.range3),lty=2)
plot(density(y.rep4,bw=.5),xlim=c(0,12),main="",xlab="")
lines(density(in.range4),lty=2)
plot(density(y.rep5,bw=.5),xlim=c(0,12),main="",xlab="")
lines(density(in.range5),lty=2)

par(mfrow=c(2,1))
alpha <- data.frame(samplesSample('alpha'))
beta <- data.frame(samplesSample('beta'))
plot(density(alpha))

```

```
plot(density(beta))

# GOF
par(mfrow=c(1,1))
plot(density(y.rep,bw=.5),main="",xlim=c(-1,12),xlab="",lty=2)
lines(density(in.range.all,bw=.5),xlim=c(-1,12))
```

Lognormal model:

```
# START BRUGS

library(BRugs)
modelCheck('model4new2.txt')
modelData('data4a.txt')
modelCompile(1)
modelInits('init4new2.txt' )
modelGenInits()
modelUpdate(1000)

dicSet()
samplesSet('m')
samplesSet('s')
samplesSet('tau')
samplesSet('beta')

modelUpdate(50000)
```

```

dic4b <- - dicStats()

samplesDensity("m",xlim=c(0,1))
par(new=T)
samplesDensity("beta")

samplesHistory("m",mfrow=c(3,2))
samplesHistory("s",mfrow=c(3,2))
samplesHistory("tau",mfrow=c(3,2))
par(new=T)
samplesHistory("beta",mfrow=c(3,2))
samplesStats("**") # the summarized results

# GOF
gof.data <- read.csv('gof4vis.csv')
attach(gof.data)
tot.mach1 <- 56
tot.mach2 <- 58
tot.mach3 <- 64
tot.mach4 <- 58
tot.mach5 <- 56
tot <- 292

beta <- samplesSample('beta')
m1 <- samplesSample('m[1]')
m2 <- samplesSample('m[2]')

```

```

m3 <- samplesSample('m[3]')
m4 <- samplesSample('m[4]')
m5 <- samplesSample('m[5]')
s1 <- samplesSample('s[1]')
s2 <- samplesSample('s[2]')
s3 <- samplesSample('s[3]')
s4 <- samplesSample('s[4]')
s5 <- samplesSample('s[5]')
tau1 <- samplesSample('tau[1]')
tau2 <- samplesSample('tau[2]')
tau3 <- samplesSample('tau[3]')
tau4 <- samplesSample('tau[4]')
tau5 <- samplesSample('tau[5]')

theta1 <- cbind(beta,m1,m2,m3,m4,m5,s1,s2,s3,s4,s5,tau1,tau2,tau3,tau4,tau5)
theta <- theta1[seq(from=1,to=50000, by=500),]
theta <- data.frame(theta)
theta1 <- data.frame(theta1)

mu.rep <- matrix(99,100,tot)
meanie <- matrix(99,100,tot)
inr.rep <- array(data=NA,dim=c(100,max(vis),tot))

for (i in 1:100)
{

```

```

    for (j in 1:tot.mach1)
    {
    if (mach[j]==1)
    {
    mu.rep[i,j] <- rnorm(1,mean=theta$m1[i],sd=sqrt(1/theta$s1[i]))
    meanie[i,j] <- mu.rep[i,j] + theta$beta[i]*log(inr0[j])
    for (k in 1:vis[j])
    {
    inr.rep[i,k,j] <- rlnorm(1,meanlog=meanie[i,j],sdlog=sqrt(1/theta$tau1[i]))
    }}}

```

```

    for (i in 1:100)
    {

    for (j in (tot.mach1+1):(tot.mach1+tot.mach2))
    {
    if (mach[j]==2)
    {
    mu.rep[i,j] <- rnorm(1,mean=theta$m2[i],sd=sqrt(1/theta$s2[i]))
    meanie[i,j] <- mu.rep[i,j] + theta$beta[i]*log(inr0[j])
    for (k in 1:vis[j])
    {
    inr.rep[i,k,j] <- rlnorm(1,meanlog=meanie[i,j],sdlog=sqrt(1/theta$tau2[i]))
    }}}

```

(repeat for machines 3-5)

```

#throw out NAs and plot vs. lab_inr

#overlapping density plots

data1 <- read.csv("pt.csv")

lab.inr <- data1$lab_inr

plotty <- inr.rep

a <- as.vector(plotty)

y.rep <- na.omit(a)

par(mfrow=c(1,1))

plot(density(y.rep ),xlim=c(0,14),ylim=c(0,.6),xlab="",main="",lty=2)

par(new=T)

plot(density(lab.inr),xlim=c(0,14),ylim=c(0,.6),main="",xlab="",ylab='Density')

# POSTERIOR PREDICTIVE DISTRIBUTION

mu.rep1 <- vector("numeric",10000)

meanie1 <- vector("numeric",10000)

inr.draw1 <- vector("numeric",10000)

badness1 <- vector("numeric",10000)

for (i in 1:10000)
{
mu.rep1[i] <- rnorm(1,mean=theta1$m1[i],sd=sqrt(1/theta1$s1[i]))
meanie1[i] <- mu.rep1[i] + theta1$beta[i]*log(2)
}

```



```

inr.draw1[i] <- rlnorm(1,meanlog=meanie1[i],sdlog=sqrt(1/theta1$tau1[i]))
badness1[i] <- max(0,(inr.draw1[i]-3),(2-inr.draw1[i]))

}

```

```

bad1 <- vector("numeric",10000)
for (i in 1:10000)
{if (badness1[i]==0) bad1[i] <- 1}
in1 <- sum(bad1)/length(badness1)

```

(repeat for machines 2-5)

Code for Plots:

Reduced Spaghetti Plots

```

badorder <- read.csv('c:/logbadorder.csv')
badorder2 <- badorder[,3:19,22]
badorder2 <- cbind(badorder2,logbadid)
attach(badorder2)

```

```

linr <- log(lab_inr)

```

FIND MIN, 25%, MED, 75%, MAX OF BADNESS TO HAVE LINES

```

select <- rep(0,length(id)) selectR <- rep(0,length(id[machine=='Rapidpoint']))
selectP <- rep(0,length(id[machine=='Protime']))
selectH <- rep(0,length(id[machine=='Hemochron']))

```

```

selectcp <- rep(0,length(id[machine=='Coaguchek S']))
selectCP <- rep(0,length(id[machine=='CoaguchekPr']))

# MAKE DIFFERENT LOGBADIDS FOR EACH MACHINE
for (i in 1:1312)

if (machine[i]==2) logbadid[i] <- logbadid[i]+100
if (machine[i]==3) logbadid[i] <- logbadid[i]+200
if (machine[i]==4) logbadid[i] <- logbadid[i]+300
if (machine[i]==5) logbadid[i] <- logbadid[i]+400

for (j in 1:5)
{
for (i in 1:length(logbadid[machine=='Rapidpoint']))
{
if (logbadid[machine=='Rapidpoint'][i]==fivenum(logbadid[machine=='Rapidpoint'])[j])
selectR[i] <- j
}
(repeat for machines 2-5)
}

select <- cbind(t(selectcp),t(selectCP),t(selectH),t(selectP),t(selectR))
select <- t(select)

detach(badorder2)

```

```

badorder2 <- cbind(badorder2,select)
attach(badorder2)
pro <- cbind(linr[machine=='Prottime'],cum.time[machine=='Prottime'],
logbadid[machine=='Prottime'],select[machine=='Prottime'])
rap <- cbind(linr[machine=='Rapidpoint'],cum.time[machine=='Rapidpoint'],
logbadid[machine=='Rapidpoint'],select[machine=='Rapidpoint'])
hem <- cbind(linr[machine=='Hemochron'],cum.time[machine=='Hemochron'],
logbadid[machine=='Hemochron'],select[machine=='Hemochron'])
cs <- cbind(linr[machine=='Coaguchek S'],cum.time[machine=='Coaguchek S'],
logbadid[machine=='Coaguchek S'],select[machine=='Coaguchek S'])
cp <- cbind(linr[machine=='CoaguchekPr'],cum.time[machine=='CoaguchekPr'],

logbadid[machine=='CoaguchekPr'],select[machine=='CoaguchekPr'])

detach(badorder2)

par(mfrow=c(5,1), mar=c(1.5,4,1.5,4)+.1, cex.axis=.9,mgp=c(3,.5,0))

#Prottime
pro <- data.frame(pro)

plot(pro$X2, pro$X1, type='p',pch=19,cex=.6, ylim=c(-.2,2),xlim=c(0,135), main='Prottime',
ylab="",xlab='Days', xaxt='n')

points(pro$X2[pro$X3==i], pro$X1[pro$X3==i], type='p',pch=19,cex=.6)
lines(pro$X2[pro$X4==5], pro$X1[pro$X4==5], type='l',lwd=2)

```

```

lines(pro$X2[pro$X4==4], pro$X1[pro$X4==4], type='l',lwd=1)
lines(pro$X2[pro$X4==3], pro$X1[pro$X4==3], type='l',col='gray')
lines(pro$X2[pro$X4==2], pro$X1[pro$X4==2], type='l',col='red',lwd=1)
lines(pro$X2[pro$X4==1], pro$X1[pro$X4==1], type='l',col='red',lwd=2)

```

```

  abline(h=log(2),col='blue')

```

```

abline(h=log(3),col='blue')

```

(repeat for machines 2-5)

Bubble plot

```

attach(badorder)

```

```

logbadid <- -logbadid

```

```

#red bubble plot

```

```

par(mfrow=c(1,5),mar=c(5, 2, 4, 0)+.1)

```

```

c <- 2*sqrt(logtail/pi)

```

```

plot(cum_time[machine=='Prottime'],logbadid[machine=='Prottime'], type='p', cex=c,
xlab='Time', ylab='Patient',main='Prottime', col='red',yaxt='n',xlim=c(0,190))
points(cum_time[machine=='Prottime'][is.na(tail)],logbadid[machine=='Prottime'][is.na(tail)], type='p',
pch=20, cex=.2)

```

(repeat for machines 2-5)

Bibliography

- Armitage, P. and Colton, T. (eds.) (1999), *Encyclopedia of Biostatistics*, John Wiley & Sons.
- Baggerly, K. A. (1998), "Empirical Likelihood as a Goodness-of-Fit Measure," *Biometrika*, 733–807.
- Bussey, H., Chiquette, E., Bianco, T., Lowder-Bender, K., Kraynak, M., Linn, W., Farnett, L., and Clark, G. (1997), "A statistical and clinical evaluation of fingerstick and routine laboratory prothrombin time measurements," *Pharmacotherapy. [Pharmacotherapy]*, 861–6.
- Curran, P. J. (1988), "The Semivariogram in Remote Sensing - An Introduction," *Remote Sensing of Environment*, 493–507.
- D'Angelo, A., Galli, L., and Lang, H. (1997), "Comparison of mean normal prothrombin time (PT) with PT of fresh normal pooled plasma or of a lyophilized control plasma (R82A) as denominator to express PT results: collaborative study of the International Federation of Clinical Chemistry," *Clinical Chemistry*, 2169–2174.
- Diggle, P. (1988), "An Approach to the analysis of repeated measurements," *Biometrics*.
- Diggle, P., Liang, K.-Y., and Zeger, S. L. (1994), *Analysis of longitudinal data*, vol. 1, Oxford University Press.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2003), *Bayesian Data Analysis*, vol. 2, Chapman & Hall/CRC.
- Gelman, A., Meng, X.-L., and Stern, H. (1996), "Posterior predictive assessment of model fitness via realized discrepancies," *Stat. Sin.*, 733–807.
- Geman, S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.

- Gilks, W. R. and Wild, P. (1992), “Adaptive rejection sampling for gibbs sampling,” *Applied Statistics*, 41, 337.
- Hirsch, J., Dalen, J. E., and Poller, L. (2001), “Oral Anticoagulants: Mechanism of action, Clinical Effectiveness, and Optimal Therapeutic Range,” *Chest*, 8S–21S.
- Hyers, T. M., Agnelli, G., Hull, R. D., and Morris, T. A. (2001), “Antithrombic therapy for venous thromboembolic disease,” *Chest*, 176S–193S.
- Jones, M. and Rice, J. (1992), “Displaying the Important Features of Large Collections of Similar Curves,” *The American Statistician*, 140–145.
- Kaatz, S., White, R., Hill, J., Mascha, E., Humphries, J., and Becker, D. (1995), “Accuracy of laboratory and portable monitor international normalized ratio determinations. Comparison with a criterion standard,” *Archives of internal medicine. [Arch Intern Med]*, 1861–7.
- Lawson, E. (2005), “Autologous Stem Cell Transplant: Factors Predicting the Yield of CD34+ Cells,” Master’s thesis, Brigham Young University.
- Lynch, S. M. and Western, B. (2004), “Bayesian Posterior Predictive Checks for Complex Models,” *Sociological Methods & Research*, 32, 301–335.
- McBane, R., Felty, C., Hartgers, M., Chaudhry, R., Beyer, L., and Santrach, P. (2005), “Importance of device evaluation for point-of-care prothrombin time international normalized ratio testing programs,” *Mayo Clinic proceedings. [Mayo Clin Proc]*, 181–6.
- R, S., K, G., K, S., and C, C. (2005), “Clinical impact of point-of-care vs laboratory measurement of anticoagulation,” *American journal of clinical pathology. [Am J Clin Pathol]*, 184–8.
- SAS Institute, I. (2000), “SAS online documentation, V8,” .
- Smith, N. T. (2004), “Validation of Criteria Used to Predict Warfarin Dosing Decisions,” Project, Brigham Young University.
- Spiegelhalter, Best, Carlin, and der Linde, V. (2003), “Bayesian Measures of Model Complexity and Fit (with Discussion),” *Journal of the Royal Statistical Society, Series B*, 64(4), 583–616.
- Tufte, E. (1983), *The visual display of quantitative information*, Graphics Press.
- Tukey, J. (1977), *Exploratory Data Analysis*, Addison-Wesley.
- Wu, F.-C. and Tsang, Y.-P. (2004), “Second-order Monte Carlo uncertainty/variability analysis using correlated model parameters: application to salmonid embryo survival risk assessment,” *Ecological Modelling*, 393–414.