SUMMARIZING FLARE ASSAY IMAGES IN COLON CARCINOGENESIS

A Dissertation

by

MAŁGORZATA LEYK WILLIAMS

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

December 2004

Major Subject: Statistics

SUMMARIZING FLARE ASSAY IMAGES IN COLON CARCINOGENESIS

A Dissertation

by

MAŁGORZATA LEYK WILLIAMS

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Raymond J. Carroll
(Chair of Committee)

Naisyin Wang
(Member)

Robert S. Chapkin
(Member)

Tailen Hsing
(Member)

Michael Longnecker
(Interim Head of Department)

December 2004

Major Subject: Statistics

ABSTRACT

Summarizing FLARE Assay Images in
Colon Carcinogenesis. (December 2004)
Małgorzata Leyk Williams, B.S.; M.S., Texas A&M University
Chair of Advisory Committee: Dr. Raymond J. Carroll

Intestinal tract cancer is one of the more common cancers in the United States. While in some individuals a genetic component causes the cancer, the rate of cancer in the remainder of the population is believed to be affected by diet. Since cancer usually develops slowly, the amount of oxidative damage to DNA can be used as a cancer biomarker. This dissertation examines effective ways of analyzing FLARE assay data, which quantifies oxidative damage. The statistical methods will be implemented on data from a FLARE assay experiment, which examines cells from the duodenum and the colon to see if there is a difference in the risk of cancer due to corn or fish oil diets. Treatments of the oxidizing agent dextran sodium sulfate (DSS), DSS with a recovery period, as well as a control will also be used.

Previous methods presented in the literature examined the FLARE data by summarizing the DNA damage of each cell with a single number, such as the relative tail moment (RTM). Variable skewness is proposed as an alternative measure, and shown to be as effective as the RTM in detecting diet and treatment differences in the standard analysis. The RTM and skewness data is then analyzed using a hierarchical model, with both the skewness and RTM showing diet/treatment differences. Simulated data for this model is also considered, and shows that a Bayes Factor (BF) for

higher dimensional models does not follow guidelines presented by Kass and Raftery (1995).

It is hypothesized that more information is obtained by describing the DNA damage functions, instead of summarizing them with a single number. From each function, seven points are picked. First, they are modeled independently, and only diet effects are found. However, when the correlation between points at the cell and rat level is modeled, much stronger diet and treatment differences are shown both in the colon and the duodenum than for any of the previous methods. These results are also easier to interpret and represent graphically, showing that the latter is an effective method of analyzing the FLARE data.

TABLE OF CONTENTS

Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

CHAPTER I

BIOLOGICAL BASIS OF THE FLARE ASSAY EXPERIMENT: INTRODUCTION

This dissertation studies methods of analyzing FLARE assay data. Before these methods are discussed, the biological basis of the experiment will be given. This background information gives a better understanding of the experiment, as well as the expected results.

1.1 Introduction

The FLARE assay data measures the amount of oxidative DNA damage, which is used as a biomarker for cancer. Dependencies of this damage in the intestine tract upon diet are of interest. In this chapter, a description of how cancer develops will be presented. A duodenum and colon specific description of the cancer will then follow. In the experiment, oxidative damage itself was measured by the FLARE assay. Explanations will be given of why oxidation can be used as a cancer biomarker, how oxidation occurs, how use of dextran sodium sulfate (DSS) is expected to affect it, and how FPG enzyme can be used to measure it. Also, corn and fish oil diet will be discussed, and how they are believed to effect intestine tract cancer. Then the FLARE assay, which is a modification of the comet assay and single-cell electrophoresis, will be described, along with summary statistics previously used to quantify this type of data.

This dissertation follows the style and format of the *Journal of the American Statistical Association*.

1.2 Cancer

Cancer is one of the leading causes of death in the United States (Jemal et al. 2002). All forms of cancer occur in part due to uncontrollable cell divisions. In a normal cell, this division is carefully controlled by various mechanisms. However, tumors usually occur when these mechanisms fail to work correctly. For example, when regulatory pathways are inhibited or oncogenes are activated. The change from a normal cell to a cancerous cell usually begins with mutations in the genomic DNA. The transformation of a normal cell to a cancerous cell requires more than one mutation (Alberts et al. 1994).

There are many regulatory genes in a cell, and when these genes are altered, the cell can loose control of repair and proliferation functions. Tumor development can begin with a mutation that may occur only in one cell. This has been tested by noting similar mutation(s) in the tumor cells. Another characteristic of tumor cells is their quick, unrestrained replication that may incur more mutations. While normal cells carefully control their replication, the mutated cells tend to be less stable. The rate of mutations in a cancerous cell is higher than what would be expected in a normal cell (Jackson and Loeb 2001).

1.3 Tissue Differences in Cancer Incidence

The analyzed data comes from the gastrointestinal tract, and principally focuses on cancer of the duodenum and colon. Colon cancer accounts for 36% of all deaths due to cancers in the digestive system (Jemal et al. 2002).

The progression of colon cancer has been identified by a series of mutations. One mutation that occurs in more than 70% of the tumors is an inhibition of the tumor suppressor APC (*Adenomatous Polyposis Coli*), see Marx (1992). The change

is believed to occur in the early stages of cancer progression (Kinzler and Vogelstein 1997). The next gene believed to be activated by DNA mutation is the *ras* oncogene, a member of a MAP kinase cell proliferation pathway, and is found in 50% of tumors. It is followed by a mutation of chromosome 18q in more than 70% of colon cancers, which causes a deletion of the tumor suppressor DCC (*Deleted in Colon Carcinoma*). Loss of p53 usually occurs in more developed tumors, and it has been detected in more than 70% of tumors (Kinzler and Vogelstein 1997, Marx 1992). The tumor suppressor gene p53 is located on chromosome 17p, and acts as a transcriptional promoter of several genes, such as p21, Fas, Bax, 14-3-3, and mdm2. The genes p21 and 14-3-3 arrest the cell cycle in the G1 and G2 phase, respectively. Bax would move to the mitochondria for cytochrome c release. Cytochrome c would then form the apoptosome in the cytosol to initiate apoptosis.

Small intestine cancer is rare, accounting for less than 1% of all deaths due to cancers in the digestive system (Jemal et al. 2002). Even though the small intestine has a larger surface area than the colon, the reasons for the rarity of this cancer are not known. However, since it is rare, it is not as widely studied as colon cancer. It has been observed that small intestine tumors do not usually exhibit mutations in the tumor suppressors APC or DCC (Arber et al. 1999, Wheeler et al. 2002). Small intestine tumors have been found to contain p53 mutations in approximately 50% of the benign tumors, and in over 60% of malignant tumors. About 50% of both types of tumors also had p21 mutations. In two other tumor suppressor genes, p16 and p27, mutations were found in over 75% of both types of tumors (Arber et al. 1999).

1.4 Causes of Cancer

Cancer usually develops slowly; hence, a biomarker of cancer to measure if a substance is carcinogenic would be beneficial. Studies have shown that diets high in fruits and

vegetables decrease the rate of cancer (Halliwell 2002). Individuals on this type of diet also incurred less oxidative DNA damage. However, cigarettes, diets high in fat, and chronic inflammation increase the rate of cancer. Individuals in these situations incur more oxidative damage to DNA. This would suggest that oxidative damage can be used as a biomarker (Halliwell 2002).

While oxidative damage naturally occurs in the body, the body is able to either repair the cell or induce the cell's death (apoptosis). However, an excess of reactive oxygen species (ROS) can interfere with cell's processes, including cell replication, and it can cause damage to the DNA (Loft and Poulsen 1996). One example of a common adduct caused by ROS is the creation of higher levels of the altered guanine base 8-hydroxy-2'-deoxyguanosine (8OHdG), which is highly mutagenic (Halliwell 2002, Jackson and Loeb 2001, Loft and Poulsen 1996). The guanine in the DNA is oxidized, which in turn causes GC→GA change during DNA replication. Upon the second replication, the adenine is matched with thymine, resulting in an overall GC→TA mutation in the DNA (Boiteux and Radicella 1999).

To increase the amount of 8OHdG present, dextran sodium sulfate (DSS) can be administered. When DSS is administered at 3-6% in the drinking water for 2 days, it causes inflammation of the colon (Tardieu et al. 1998, 2000). It has been documented that chronic inflammation of the bowel leads to increased rate of colon cancer (Jackson and Loeb 2001). This increase is believed to be linked to increased creation of ROS (Jackson and Loeb 2001).

An enzyme capable of repairing 8OHdG adducts is OGG1. The procaryotic xenolog of OGG1 is formamidopyrimidine-DNA glycosylase enzyme (FPG), found in *Escherichia coli*. FPG is able to repair damage caused by 8OHdG by removing the oxidized guanine. Another repair enzyme in *Escherichia coli* is MutY, which removes the adenosine when it is paired with an oxidized guanine. Together, FPG

and MutY can prevent GC→TA mutations (Boiteux and Radicella 1999). Without timely repair of an oxidized base, the GC→TA mutation will become permanent. Although GC→TA mutations are not the only genetic alterations produced by ROS, they do occur with high frequency (Jackson and Loeb 2001).

1.5 Dietary Lipid Sources

The effect of corn oil, fish oil, and DSS on oxidative damage will be analyzed. It is expected that the cell's nucleus exposed to fish oil will have less oxidative damage than the cell's nuclei exposed to corn oil. The basis for this claim can be seen by observing populations that have a fish-based diet. Both Eskimos and Japanese fishermen have a high intake of fish and low rates of colon cancer (Bartsch et al. 1999).

While this is only an observation, and not conclusive evidence, experiments performed on animal models provide similar results. Diets which include n-6 polyunsaturated fatty acid (PUFA), which is in corn oil as well as other vegetable oils, were found to promote cancer. On the other hand, diets containing n-3 PUFA, which is present in fish oil, were found not to promote cancer, and possibly even have a cancer preventing (chemopreventive) effect (Bartsch et al. 1999, Diggle 2002, Hong et al. 2002, Sugimura 2000).

1.6 FLARE Assay

The FLARE assay, also known as the FPG-comet assay, is able to quantify the amount of oxidative damage. The assay is performed on a cell level. In this case, it is performed on cells from the duodenum and colon. Half the cell are exposed to FPG, which cuts the cell's DNA at oxidized guanine. A gel electrophoresis is then performed on DNA from the isolated cells. Migration of the DNA is dependent on its size, with smaller pieces migrating farther in the gel than larger pieces, resulting in

an image that gives the appearance of a comet. Figure 1 contains examples of comets with varying levels of DNA damage. The comet in the top left corner has the least amount of damage, while the comet in the bottom right corner has the most damage. The quantified results for cells with FPG and without FPG can then be compared to determine amount of damage due to oxidation.

The comet is composed of two parts. The circular part to the left is called the head (or nucleus). When there is no damage, there will only be a head; however, when there is damage, the comet will also have a tail. Everything to the right of the head is called the tail. This method is able to detect 1 single-strand DNA break per 3×10^9 Daltons of DNA (Singh 2000).

1.7 Previous Statistical Methods

One of the first methods for analyzing the comet data was the distance DNA migrated (Singh et al. 1988). A similar measure is the tail length, especially when measured from the center of the head. Kent et al. (1995) pointed out that initially the tail length increases linearly with increase in damage of the cell's nucleus. However, there occurs a point where tail length does not increase, but the proportion of damage does. A measure that takes this fact into account is the tail moment. Olive et al. (2001) introduced this measure, which is defined as the percentage of DNA in the comet tail times the distance between the moments of the head and tail DNA distributions. Symmetry of the head was assumed, and hence the tail area could be identified.

Another measure used in comet analysis is the relative tail moment (RTM) (Riso et al. 1999, Morris et al. 1999, Hellman et al. 1995), defined as

$$\text{RTM} = 100^*(\text{tail moment})/(\text{tail moment} + \text{head moment})$$

where head or tail moment is defined as the sum of the distance away from the center of the head times the amount of DNA in head or tail at that distance, respectively.

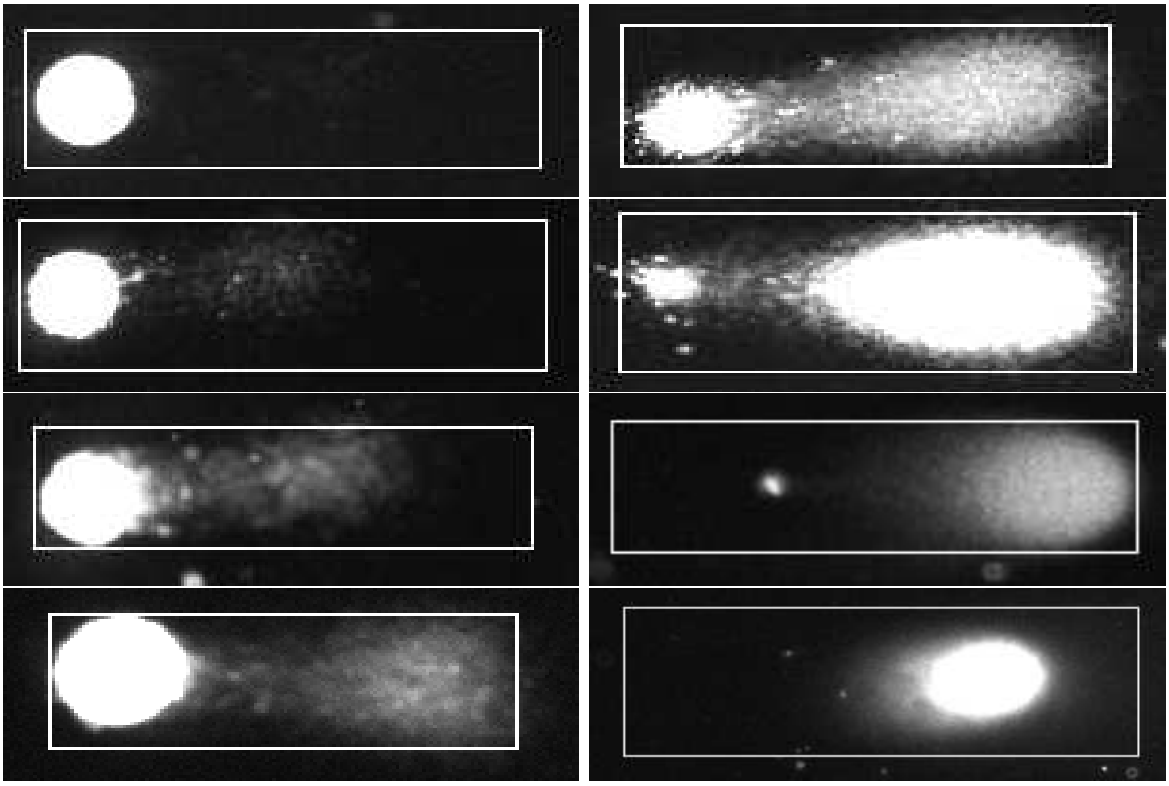

Figure 1. Different levels of damage in comets, with damage level increasing for comets moving from top to bottom, and from left to right.

The advantage of this variable is that it does not depend on the scale used to measure the distance (as long as the same scale is used for both the head and the tail moment), or whether the intensity of the comet is standardized.

1.8 Summary of the Experiment

The aim of the experiment which will be analyzed in this dissertation is to determine if there is a difference in the rates of intestine tract cancer as related to the consumption of corn or fish oil in a diet. From previously published results, it is believed that a fish oil diet is more beneficial than a corn oil diet. The experiment will use the FLARE assay to test this. The assay will produce images of amount of damage to cell's DNA. Half the cells for each diet will be treated with FPG enzyme. Consequently, the images from cells not treated with FPG will show "naturally" occurring damage. The images for cells treated with FPG will show both the oxidative and "naturally" occurring damage. Analysis of these images, through the use of summary statistics, will enable estimation of amount of oxidative damage, and consequently the determination of oxidative differences due to diet.

CHAPTER II

COMPARING AUTOMATIC AND MANUAL IMAGE PROCESSING IN FLARE ASSAY ANALYSIS FOR COLON CARCINOGENESIS

Measurement of the amount of oxidative damage to DNA is one tool that can be used to estimate the beneficial effect of diet on the prevention of colon carcinogenesis. The FLARE assay is a modification of the single-cell gel electrophoresis (Comet) assay, and provides a measure of the 8OHdG adduct in the cells. In this chapter, two innovations to the existing methods of analysis are presented. The first one is related to the FLARE assay itself. An automated image analysis technique will be described that can be expected to measure oxidative damage faster, reproducibly, with less noise, and hence achieve greater statistical power. The proposed technique is compared to an existing technique, which was more manual and thus slower. The second innovation is the statistical analysis: the shape of FLARE intensity histograms is exploited, and statistically significant diet effects in the duodenum are shown. Previous analysis of these data concentrated on simple summary statistics, and found only marginally statistically significant diet effects. With the new imaging method and measure of oxidative damage, cells in the duodenum exposed to fish oil are shown as having more oxidative damage than cells exposed to corn oil.

2.1 Introduction

Cancer usually develops slowly; hence, a reliable biomarker of oxidative damage would be beneficial. Altered guanine base 8-hydroxy-2'-deoxyguanosine (8OHdG) can be considered as one such biomarker (Boiteux and Radicella 1999, Halliwell 2002, Jackson and Loeb 2001, Loft and Poulsen 1996). 8OHdG has been found to be highly

mutagenic (Boiteux and Radicella 1999). Hence, an increase in the amount of 8OHdG damage to DNA should be correlated with increased rates of cancer (Halliwell 2002).

To be able to measure increased levels of 8OHdG, the FLARE assay can be utilized. The FLARE assay uses the single-cell gel electrophoresis assay, also known as Comet assay, to produce images of the amount of damage to DNA of a cell's nucleus (Collins et al. 1995, Singh et al. 1988). Moreover, half of the cell's nuclei are exposed to the repair enzyme formamidopyrimidine-DNA glycosylase enzyme (FPG). FPG is found in *Escherichia coli*, and contributes to the repair process by removing the oxidized guanine (Boiteux and Radicella 1999). DNA of cells that are exposed to FPG should contain both naturally occurring breaks in DNA, as well as breaks caused by FPG at places containing oxidized guanine base. DNA of cells not exposed to FPG can serve as a baseline for naturally occurring DNA breaks.

Existing methods of image analysis involve manually identifying each comet on an image, and utilizing a macro on each identified comet to compute a measure of DNA damage (Bancroft et al. 2003). Doing this for thousands of comet images is slow, and prone to possible error. We propose an automated image analysis that will automatically identify comets in an image, as well as compute a desired measure of damage for each comet, which would be faster and more consistent than the current image analysis method.

In the chapter, the two techniques are compared to determine how they differ by analyzing the amount of 8OHdG DNA damage in colonocytes from rats consuming diets containing either fish oil or corn oil (Bartsch et al. 1999, Diggle 2002, Hong et al. 2002, Sugimura 2000), and with or without dextran sodium sulfate (DSS) treatment (Tardieu et al. 1998, 2000), which is an oxidizing agent. Analysis of the data using existing methodologies yielded marginal results. The analysis using the proposed automated imaging algorithm produced more significant results.

It is also hypothesized that an important component of the collected data is the shape of the image. This chapter reports a new statistical variable that describes the shape of a comet, along with the automated imaging algorithm, and statistically significant diet effects.

2.2 Materials and Methods

Two groups of Sprague-Dawley male rats were established. One group consumed a diet containing corn oil at 15% by weight of the diet, while the other group of rats consumed a diet containing fish oil at 15% by weight of the diet (Bancroft et al. 2003). Three further subgroups were made in each of the diet groups. The first subgroup was treated with 3% DSS (ICN Biomedicals; Aurora, OH) in the drinking water for 48 hours. Another subgroup was also treated with DSS for 48 hours, but was then given 48 hours before euthanasia, to allow for recovery and DNA repair. The final subgroup was the control which received no DSS treatment. Each of the subgroups contained ten rats. Results from one rat fed corn oil with DSS treatment was excluded due to a sample preparation problem.

For each rat, cells were obtained from both the colon (large intestine) and the duodenum (small intestine). There were usually more than 200 cells per rat in each location available for analysis. Half of the cells obtained from each location were treated with FPG, while the other half remained untreated. The cells were then processed using the FLARE assay procedure described by Bancroft et al. (2003).

After the preparation, images were obtained from the processed cells. An image algorithm was then used, which computed a measure of the amount of damage to DNA in the cells. The amount of 8OHdG adducts for each rat-location used in the analysis was the difference between the FPG measure and the no-FPG measure.

2.3 Existing Image Analysis and Calculations

The existing analysis used the Metamorph Imaging System (Version 4.6r3, Universal Imaging Corporation, Downingtown, PA) for capture and analysis of the image, as follows. Images of randomly selected comets were captured and stored on the computer. Each comet image was then identified manually on the slide. For each comet, the head of the comet was identified by selecting a threshold. Pixels with intensity higher or equal to the threshold were selected. If pixels that were not part of the comet were selected, then a freeform tool was used to define the area of the image where the comet was located. If pixels in the tail were included in the head after thresholding, the user could create a two pixel gap between the head and the tail to help the program identify the end of the head. Any pixels still identified by thresholding, but not part of the head, could be removed from calculations by manually identifying them.

A box was then drawn that contains the comet, with the right side of the box extending 10 to 20 pixels beyond the end of the comet tail. A macro was executed that circled the thresholded head and found its center. Average intensity of the pixels to the right of the tail was used to estimate the background intensity, which was subtracted from the result. A measure of the amount of damage was then computed by the program for each comet.

The standard output from FLARE analysis is the relative tail moment (RTM) (Hellman et al. 1995, Morris et al. 1999, Riso et al. 1999), defined as

$$\text{RTM} = 100^{*}(\text{tail moment})/(\text{tail moment} + \text{head moment})$$

where head or tail moment is defined as the sum of the distance away from the center of the head times the amount of DNA in head or tail at that distance, respectively.

2.4 Proposed New Image Analysis

The existing image analysis method did not have the benefit of a standardized routine of data generation, making it less objective than desired. Therefore, the possibility exists for inconsistent comet identification and processing. The new method of image analysis uses a set algorithm for identifying a comet and its components, thus reducing the variability introduced by human error in the existing image analysis system. It provides a methodical approach that is reproducible, and concentrates on eliminating as much of the background noise as possible before processing.

The processing of the FLARE images was done in the following manner. First, a grayscale FLARE image was converted to a zero-minimum-intensity image by subtracting the minimum intensity from all the pixels. The image was then enhanced by a nonlinear scaling of the pixel intensities: each pixel intensity was raised to the power 1.05. The factor of 1.05, as well as many of the other numbers in this description, is application specific, and was determined by iterative analysis of the procedure and results. The enhanced image was converted to a binary image by thresholding at a grayscale level of 50. This binary image had legitimate comets together with abnormal non-comet areas of staining on a black background with speckled noise. From observation it was found that comets had an area of less than 11000 pixels. Anything larger than that was considered non-specific, and was removed by using a sequence of morphological operations (Dougherty 1992, Dougherty and Astola 1999, Giardina and Dougherty 1988, Serra 1982).

Incomplete comet images (those partially obstructed by the image boundary) were eliminated using morphological operators for removing boundary regions (Dougherty 1992). The speckled noise in the resulting binary image was removed by applying morphological opening on the image using a disk of radius 1. A grayscale image cor-

responding to the original enhanced image was created using the morphological operations of thinning and reconstruction and using the above binary image as the mask. Morphological methods for reconstruction and thinning may be found in Dougherty (1992) and the SDC morphology toolbox for Matlab.

The resulting image, J, was then used to construct markers for the potential comet regions, which could be head, tail, head with tail, or abnormal shaped regions. Markers for heads were constructed through a sequence of morphological operations. First the image J was negated and basins with contrast greater than 170 were found using the SDC morphology toolbox for Matlab. These regions were labeled as the heads since it was found that the head region had a very high contrast (greater than 170) compared to tails. To remove noise, shapes with a disk radius of less than 3 were removed by morphological opening. This image was converted into a binary image via thresholding as discussed earlier. A thinning operation on the binary image provided the markers for the heads. The head was reconstructed from the original enhanced image via reconstruction using morphological operators and the markers for the head. The process of thinning and reconstruction removed some of the noise not removed by the morphological opening used earlier.

A similar operation for reconstructing the tails was used. Regions corresponding to the head were subtracted from the original image and markers for the tail were generated as in the case of the head after cleaning the image. The tails were then reconstructed. It should be noted that the regions detected may not necessarily be tails; they could be comets complete with a head and tail or heads which are weak in signal intensity. The two binary images of the head and the tail regions were combined and each region was labeled. Thus the outline of all clearly visible heads and tails of comets in an image were obtained. Tails with weak signal intensities and low density may not be detected and were specially processed later on.

Before further processing of the head and tail regions, regions with abnormal shapes should be removed. To do this, an intensity vector was constructed for each region by summing along the vertical length of the region for each pixel along the horizontal length of the image. The shape of the intensity vector was used to determine if the region is a legitimate head/tail region. The intensity vector was smoothed using an FIR filter of order 25 and cut-off frequency 0.1. A measure, defined as the square root of the weighted sum of the squares of the intensity values and the respective index values (of the intensity vector), was then computed. This measure was indicative of the amount of matter around a peak: the region was declared as abnormal if either this measure was too low or if the measure was too large and the number of peaks was very high. This method was not 100% effective; however, it removed most of the abnormal regions.

Classification of the labeled regions as head, heavy tail or head with tail was necessary to identify the comets in the given image. For this purpose, four measures were computed for each labeled region. Two of these measures were the area and the width-to-height ratio of the bounding box containing the labeled region. Two other measures, r_o and r_i, were computed. The measure r_o was computed by flipping the left half of the region onto the right and finding the excess of the original region on the right half with respect to the flipped one. The measure r_i, on the other hand, was computed by flipping the left half of the region onto the right and finding the excess of the flipped half with respect to the original right half. The four measures form a parameter space and heuristics/conditions based on intuition. Trial and error were used to split the space into the following six classes: head, small head, significant tail, heavy tail, head and tail, or anomalous.

Once the labeled regions were classified, the comets were reconstructed. First solitary heads and tails were considered and matched to see if they form a head-tail

pair. The matching conditions were straightforward: the tail should be to the right of the unpaired head (in the direction of the current flow) and the head to the left of the unpaired tail. The region to the right of the unpaired head was masked: 130 pixels to the right of the head were considered. This region was enhanced by thresholding the particular region in the original image with a low threshold level of 30. Post processing on the region was done to extract the signal from the noisy background.

For the unpaired tail, the region to the left was masked. Here again, re-thresholding at a lower threshold value was done and the density of signal (the sum of the signal intensities, in the bounding box placed left of the tail with a length of at most 65 pixels and a height equal to that of the tail, divided by the area of the bounding box) to the left of the tail was computed. If this density was greater than 1, then further checking was done to ensure that the region does have a head. This involved identifying the peak of the head from the intensity vector. If the region about the peak was symmetric, then the region was classified as a head, and a mask, i.e. a new labeled region, which contains both the head and the tail was drawn. The labeled regions were thus classified as heads with tails, head only or tail only. The anomalous regions were neglected and masked out.

Following identification of the comets, the intensities for each comet were computed. Before the actual raw intensities were computed, the background intensities were computed locally for each region in the following manner. The mean intensities on the four sides of the labeled region viz., the top, bottom, right and left were computed. The median of these four background estimates was then taken as the background estimate for the labeled region.

The intensity estimation was done for the head and the tail separately. For the new labeled regions, which have been classified as head with tail or just head, the head intensity was calculated by creating a mask symmetric about the vertical axis passing

through the center of the head. This mask was created by replacing the right half of the head mask by the left half. The original image under the mask was reconstructed and the pixel intensities along the vertical axis were summed up for each pixel along the length of the head. The rest of the region in the bounding box was considered as tail, and the intensities were integrated as for the head.

For the regions classified as not having a head or a very small head, the processing was different. The head mask was identified by defining a box around it and the intensities were integrated. The rest of the region was considered as tail, and the intensities for the tail were integrated. Finally, the background intensities computed locally for each region were subtracted from the intensities computed above. Various measures could be computed from these raw intensities.

Figure 1 shows FLARE images isolated by the program. After the head and tail intensities were summed over in the vertical direction, an intensity histogram for the FLARE was created. An example of such a histogram is shown in Figure 2. This histogram corresponds to the first comet in the second column of Figure 1. To create intensity histograms, the head and tail intensities were summed over in the vertical direction. Summing intensities in the vertical direction is feasible since the electric current only has a left-right effect, so the vertical position of a stained particle does not provide more information about the damage in a comet. The vertical axis of the intensity histogram, the example of which is shown in Figure 2, corresponds to the intensity of the comet for a given distance in the horizontal direction. In Figure 2, the solid line is the tail, while the dashed line corresponds to the head of the comet.

2.5 New Statistical Methods

The statistical analysis is based upon two steps: (a) the measure of oxidative DNA damage computed for each cell, and (b) the aggregation of cell-level summaries to

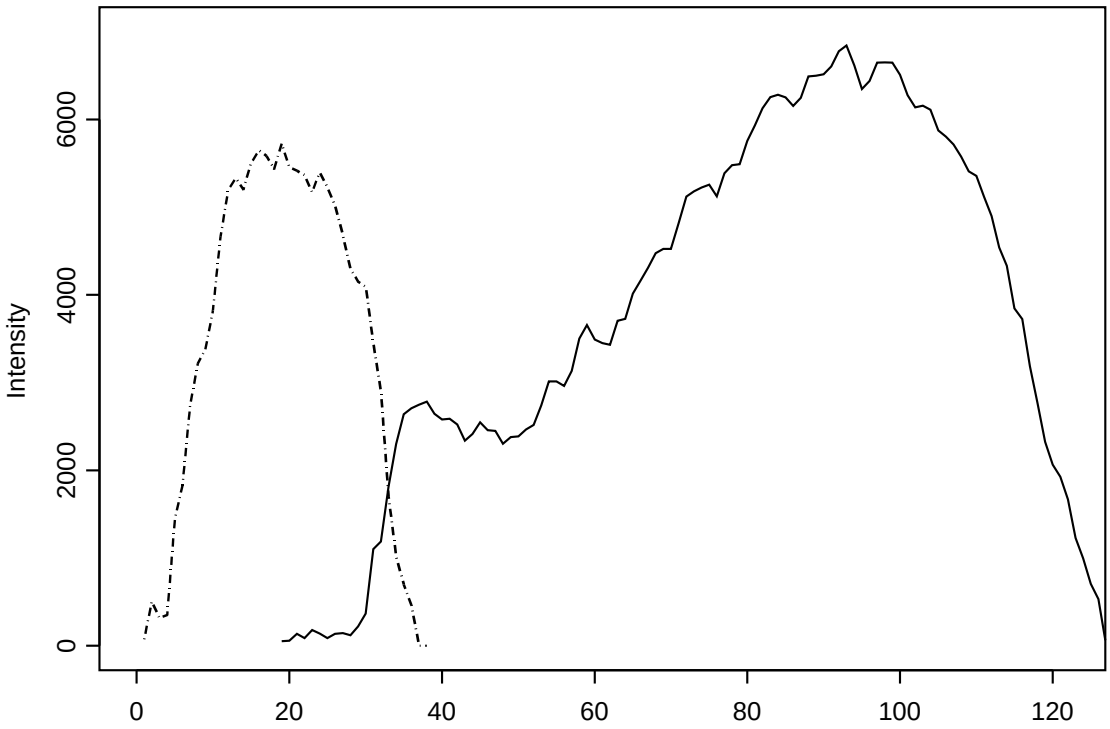

Figure 2. Example of intensity histograms for a comet, with head (dashed) and tail (solid) intensities.

rat-level summary statistic. Since the quantity of 8OHdG is the variable of interest, the difference between the results with and without FPG needs to be calculated at the rat-location level. A graphical representation of the statistical analysis is given in Figure 3.

First consider the measure of DNA damage. While RTM, used in previous statistical analysis, does contain information about the amount of oxidative damage, additional information about the level of damage is contained in the shape of the comet. When presented with an image of the comet, severity of damage can be easily assessed just by looking at the shape of the comet. By describing the shape, and not just means, the differences between comets should be more apparent than what is provided by some of the previous estimators. Consequently, in addition to RTM, skewness was used to test for differences in comet shape, which is denoted by $\tilde{u}_3$. Formally, this is defined as

$$\text{skewness} = \tilde{u}_3 = u_3/(u_2)^{3/2}$$

where

$$u_i = \sum_j (x_j)^i \times (\text{proportion of intensity at point } x_j).$$

In the above equation, x_j is the value on the horizontal axis of the intensity histogram, with the center of the head of the comet being 0 on the horizontal axis. Note that to compute $\tilde{u}_3$, the comet does not need to be separated into a head and a tail part.

The second part of the statistical analysis is based on aggregating the cell-level measurement (RTM or skewness) to form a summary measure for each rat. In addition to the mean and the median, the first and third quartiles (25^{th} and 75^{th} percentiles, respectively) were also used. The mean and median are two ways of describing the distributions of the RTM or skewness. The median describes only what has happened

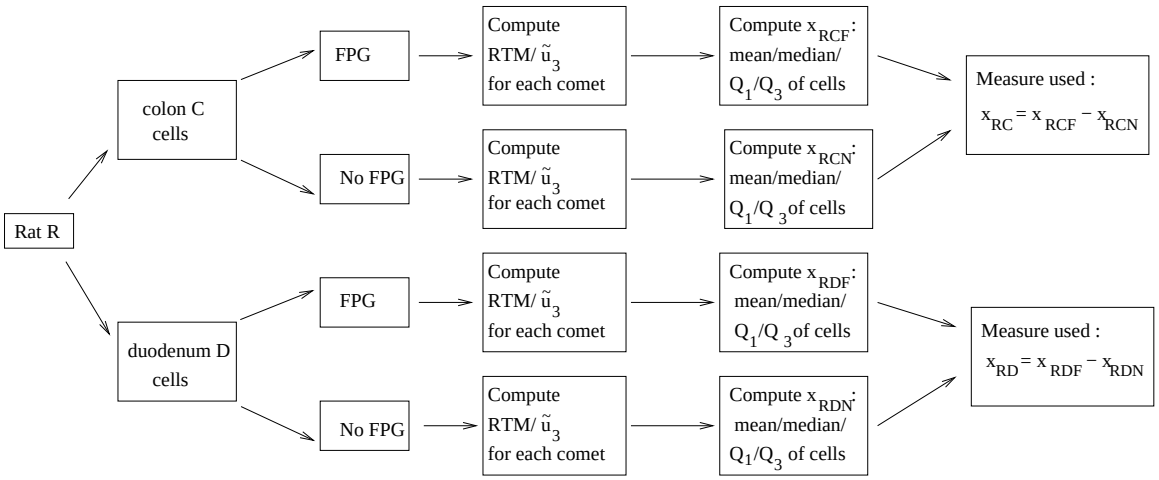

Figure 3. Representation of how the final measure of oxidative damage for each rat is computed in the new analysis.

in the middle of the distribution. However, the first and third quartiles describe what has happened to the left and the right of the center. These three are not necessarily dependent on each other. In particular, for the same median, there can be many possible values for the two quartiles. Therefore, additional information can be obtained about the distributions by examining the first and third quartiles.

It should be noted that the mean will depend on the median, first and third quartile values. If the distribution of RTM or skewness is skewed to the right, the location of the mean will be to the right of the median. Therefore, any significant results seen when using the mean may be explained by using the quartiles results.

2.6 Results

As described above, the previous technique used a manual image analysis, computed the RTM for each comet, and aggregated it to the rat-level by computing the mean, median, first and third quartile RTM over all cells in a rat. The level of 8OHdG was then assessed as the difference between the aggregated value for FPG and no-FPG cells. A graphical representation of this process is similar to the process given in Figure 3, except only RTM is computed for each comet.

With previous imaging and summary statistics, the only significant diet effect was found in the colon when aggregation was performed using the first quartile RTM (p-value of 0.044, see Table 1). Here, the corn oil diet estimate was higher than the fish oil diet estimate, indicating more oxidative damage to cells exposed to corn oil (Table 2). The only other significant effect was in the colon when aggregation was done using the third quartile (p-value of 0.030, Table 1). It indicated that administration of DSS resulted in a statistically significant increase in oxidative damage over control or DSS with a recovery period. The estimates for the statistically significant treatment and diet effects are given in Table 2.

Table 1. Comparing p-values for RTM and skewness when the existing or new imaging method is used.

Location	Effect	Mean p-value	1^{st} Quartile p-value	Median p-value	3^{rd} Quartile p-value
		Existing	*Method*	*(RTM)*	
Colon	Treatment	0.266	0.991	0.613	0.030
Colon	Diet	0.119	0.044	0.151	0.690
Colon	Diet*Treatment	0.650	0.664	0.417	0.766
Duodenum	Treatment	0.337	0.383	0.470	0.555
Duodenum	Diet	0.805	0.869	0.059	0.806
Duodenum	Diet*Treatment	0.858	0.894	0.696	0.566
		New	*Method*	*(RTM)*	
Colon	Treatment	0.608	0.855	0.626	0.090
Colon	Diet	0.961	0.583	0.587	0.778
Colon	Diet*Treatment	0.435	0.526	0.175	0.858
Duodenum	Treatment	0.474	0.159	0.919	0.724
Duodenum	Diet	0.275	0.396	0.721	0.007
Duodenum	Diet*Treatment	0.417	0.463	0.557	0.180
		New	*Method*	*(skewness)*	
Colon	Treatment	0.441	0.928	0.852	0.254
Colon	Diet	0.385	0.205	0.202	0.915
Colon	Diet*Treatment	0.223	0.412	0.226	0.347
Duodenum	Treatment	0.412	0.244	0.671	0.519
Duodenum	Diet	0.024	0.482	0.159	0.027
Duodenum	Diet*Treatment	0.847	0.700	0.792	0.811

Table 2. Estimates of significant treatment effects for RTM and skewness when the existing or new imaging method is used.

Location	Diet/Treatment Effect	FPG Difference Type	Estimate ± Standard Error
	Existing Method	*(RTM)*	
Colon	Control	Q_3	2.44 ± 3.14 [a1]
Colon	DSS	Q_3	12.64 ± 3.23 [a2]
Colon	Recovery	Q_3	1.42 ± 3.14 [a1]
Colon	Corn	Q_1	8.79 ± 2.48 [b1]
Colon	Fish	Q_1	1.62 ± 2.43 [b2]
	New Method	*(RTM)*	
Duodenum	Corn	Q_3	-1.03 ± 1.06 [c1]
Duodenum	Fish	Q_3	3.13 ± 1.05 [c2]
	New Method	*(skewness)*	
Duodenum	Corn	Mean	-0.05 ± 0.05 [d1]
Duodenum	Fish	Mean	-0.21 ± 0.05 [d2]
Duodenum	Corn	Q_3	-0.13 ± 0.09 [e1]
Duodenum	Fish	Q_3	-0.41 ± 0.09 [e2]

NOTE: Effects with the same superscript letter (a, b, c, d, or e) are compared for statistical significance. Different superscript number (1, 2) following the same letter denotes statistically significant difference.

In contrast with the previous imaging and statistical analysis, the new imaging and new statistical analysis showed stronger diet effects, especially in the duodenum.

In the new image analysis, first RTM was considered. Here we see a statistically significant result for diet in the duodenum when the summary measure aggregated for the rat is the third quartile of the RTM (p-value of 0.007, Table 1). The fish oil caused more oxidative damage than the corn oil (see Table 2). This finding suggests that diet influences the distribution of DNA damage. This is the only statistically significant effect found when RTM was examined.

With this background, now skewness ($\tilde{u}_3$) is considered in the place of the RTM. Because shape was conjectured to be critical in elucidating diet effects, we would expect to see diet effects to become more obvious. This happens in the duodenum since both the mean and the third quartile of the shapes (skewness) show highly statistically significant diet effects (Tables 1 and 2).

2.7 Discussion

The main advantage of automated image analysis is that it is faster and more consistent than processing 35,000 comets manually. While the current image analysis is not perfect, it identifies parts of the head and tail the same way for every comet. There is no variability or bias due to a technician's error. It is also reproducible, meaning that if the algorithm was performed again on the same FLARE images, it would produce essentially the same results. This is not true with an analysis that involves extensive human intervention.

While the new image analysis approach is beneficial in processing large amounts of data, it also has a few drawbacks. The program processing the images finds it difficult to distinguish a comet with a small head from a clustering of the fluorescent material that is not a comet. This type of clustering seems to occur naturally, ap-

pearing as little specs on the image. One solution to this problem is not to consider "comets" that are smaller in height than a certain threshold. In this analysis, the threshold of 12 pixels was established by comparing images and their corresponding intensity histograms. Specs that were too small to be nuclei of cells were identified. The height of these specs was considered, based on which threshold was established. The comets below that threshold (2.4%) were not considered in the analysis.

Another problem that occurred in 4.4% of comets was that specs were identified as small heads, and a comet with a head right behind the spec was identified as all tail. While this type of problem could be fixed by adjusting the threshold for the smallest amount of intensity in an area that could be considered as a head, doing this would also eliminate from the analysis some comets with a substantial amount of damage. The last major problem with the analysis was when two comets were too close together. In 3.5% of the cases, the comets were overlapping enough to be identified as one comet.

Some of the problems mentioned above might be corrected by adjusting the thresholds on the comet specification. However, this will also cause a certain percentage of good comets not to be processed because they did not meet the specifications. This will most likely hold true for comets that had a substantial amount of damage. Eliminating these types of comets would be counterproductive, so the thresholds on the comet specification were not adjusted further. As is typically the case with a prototype imaging algorithm, further refinement to address problematic issues is certainly possible. It is not believed that extra time and resources required is called for at this time, since the current imaging algorithm has correctly processed 90% of the images, while at the same time yielding gains in processing speed, precision, and reproducibility.

Using the new image analysis, diet effects turned out to be stronger than in

the previous image analysis (Bancroft et al. 2003). The two imaging techniques can be compared based on the results obtained for RTM values. The RTM p-value was more significant when the new imaging method was used rather than when the previous imaging method was used. However, the location of the diet differences was not the same. In the previous imaging method, the marginally statistically significant diet effect was only seen in the colon. Using the new imaging method, the statistically significant diet effect was seen only in the duodenum, but not in the colon. Different results in the two methods must be due to difference in image processing since the same set of images was used for both procedures. The new imaging method is preferred since it is reproducible, and puts more emphasis on eliminating noise.

From this new data analysis it can be concluded that fish oil causes more oxidative damage than corn oil in the duodenum. No significant diet effects were found in the colon. While there is also a strong diet effect for the skewness ($\tilde{u}_3$) in the duodenum, the result is not as easily interpretable. The difference between corn and fish oil estimates is positive, but each of the individual diet estimates is negative (see Table 2). The corn oil diet estimate showed no statistical difference between data with and without FPG incubation, indicating that there is no oxidation effect for corn oil results (p-values 0.3576 and 0.1421 for mean and third quartile, respectively). However, for fish oil the difference in FPG types was negative (p-values less than 0.0001 for both mean and third quartile).

In order to explain the negative values obtained, the following was done. A figure was created using the intensity histograms for all comets. The center of the head was set as zero on the horizontal axis so that they would be aligned. The intensities for all the comets were standardized so that the area under the intensity histogram was one. Let the intensity histograms of these comets be denoted by $f_{drlf}(x_i)$, where d

denotes the diet type, r the rat given diet d, l the location in the rat, f the FPG type, and c a comet obtained from cell in location l from rat r. Then for all the comets for rat r, location l, and FPG type f, the mean intensity for each horizontal axis location was found. In the above notation, this means that:

$$f_{drlf}(x_i) \quad = \quad \sum_{c \subset rlf} f_{drlfc}(x_i)/n_{rlf}$$

for each value of x_i on horizontal axis. Above, n_{rlf} is the number of comets for rat r and location l. The average is then taken for all the rats with the same location and FPG type for each horizontal axis location:

$$f_{dlf}(x_i) \quad = \quad \sum_{r \subset dl} f_{drlf}(x_i)/n_{dl}$$

for each value of x_i on horizontal axis. Here, n_{dl} is the number of rats with diet d and location l. From the resulting function, Figure 4 was created. Figure 4 shows the FPG (solid) and no-FPG (dashed) for fish and corn oil diet in the duodenum and colon. The functions can be thought of as representative standardized intensity functions for the comets in the given location, diet, and FPG type groups. If $\tilde{u}_3$ is computed for both the fish FPG functions for duodenum in Figure 4, and then the difference is taken, this difference is negative. From this figure, it can be concluded that there is more damage when FPG is present since the level of intensity (vertical axis) in the head is lower for FPG, and thus higher intensity must be in the tail than when FPG is not present. Similar behavior can be seen for the two fish FPG functions in the colon, but to a lesser extent. For the corn oil functions, there is almost no difference between the FPG functions. To summarize, differences in FPG functions for fish oil are more pronounced than for corn oil. Also, the negative difference between skewness of the FPG function and the no-FPG function for fish oil implies higher level of damage for the FPG function than the no-FPG function, as is expected.

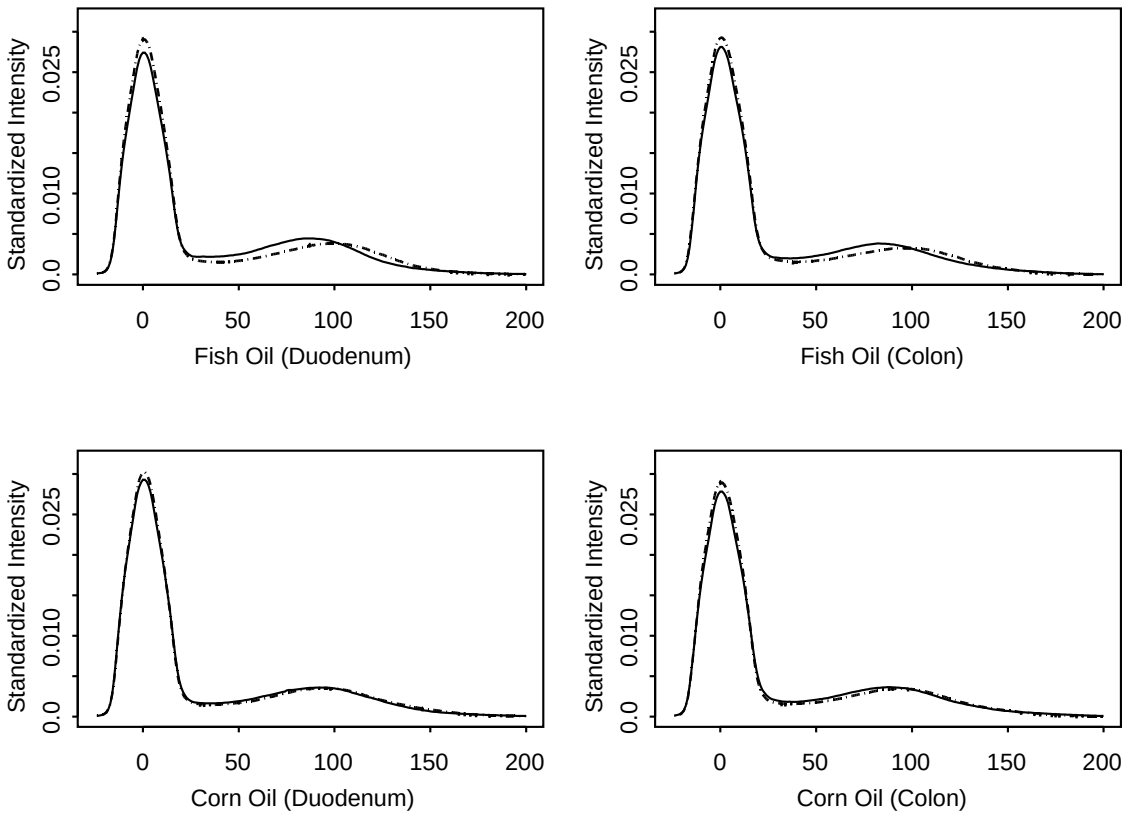

Figure 4. Representative functions for FPG (solid) and no-FPG (dashed) in the duodenum and colon for rats fed a fish or corn oil diet.

Consequently, the data suggests that fish oil causes more oxidative damage than corn oil. However, further studies of the colon from the same rats found that fish oil enhances apoptosis (Bancroft et al. 2003). From the FLARE assay, it could not be determined which cells were in the process of apoptosis, and would consequently have more DNA fragmentation. However, if the cells exposed to fish oil and an oxidizing agent were more likely to undergo apoptosis, then this would imply that the larger amount of damage seen when fish oil is in the diet is not necessarily harmful. The cells that were in the process of apoptosis would show more damage, but they would not pose a risk of increased digestive tract cancer since they would be eliminated.

CHAPTER III

SINGLE VARIABLE BAYESIAN MODELS FOR FLARE FUNCTIONS

The FLARE assay, a modification of the comet assay, is used to determine the amount of oxidative damage to DNA in a cell. The results of these assays are intensity histograms denoting the amount of damage to DNA in a cell. Previously this data was examined by using frequentist analysis on the function summarizing variables of the relative tail moment (RTM) and skewness. In this chapter, single variable Bayesian models will be examined using function summarizing variables.

First, simulations will be performed on Bayesian models to examine how the Bayes Factor (BF) is influenced by prior information and estimation of posterior variables. The values of the Likelihood Ratio Test (LRT) in each model comparison will also be computed. The BF and LRT will be computed for 300 simulated data sets to show that while LRT follows a χ^2 distribution, the distribution of BF is harder to describe.

We will then examine a hierarchical model for analyzing the RTM and skewness data. The models will be implemented on the FLARE data, which aims to test differences between oxidative levels in corn and fish oil diets in the intestine tract. The results show that diet and treatment covariance differences were detected for the RTM in the duodenum. For skewness, diet and treatment differences were detected for the mean vector in the colon, and only diet differences were detected in the duodenum.

3.1 Introduction

The use of a FLARE assay enables the estimation of the amount of damage to a cell's DNA that is caused by oxidation. The estimation can be attributed to the use of

the enzyme FPG. Without FPG, the FLARE assay produces images of "naturally" occurring damage in a cell. However, with FPG, these images show the "naturally" occurring damage, as well as the damage due to oxidation. The images can be summarized by intensity histograms. In this chapter, the intensity histograms will be used in computing values for the single variable model.

Previous results for this data looked at the relative tail moment (RTM) and skewness using standard analysis. While there appeared to be differences due to diet when skewness was considered, the exact nature of the differences was difficult to describe. In part, this was due to the fact that each function was described by a single number, and differences in functions due to diets seemed to be non-trivial. Since there were no significant diet or treatment differences with the no-FPG data, the difference between FPG and no-FPG values at the rat level was modeled, making it harder to interpret the results, especially when differences in skewness values were considered.

In this chapter, we will present Bayesian models for FLARE data analysis. The Likelihood Ratio Test (LRT) and Bayes Factor (BF) will be used for model selection. While use and interpretation of LRT is well established, that is not the case with BF. For BF, there are only a few suggested guidelines for cut-off points to use in determining a significant difference between models (Kass and Raftery 1995). We will present simulations from the single variable hierarchical model that aims to compare the LRT and BF variable distributions. The effect of prior and estimation of posterior variables will also be examined.

We will also present a Bayesian model for a single variable. The estimation was done for both FPG types. This enabled a better interpretation of the data than modeling differences between FPG types at the rat level. Both the FPG types were modeled simultaneously, which allowed modeling of the correlation between the FPG

types at the rat level. The estimation of each of the FPG values also allows for a better understanding how the values change within a diet or treatment, and how they change between diets or treatments. The single variable analysis will be done for both the RTM and the skewness variables.

3.2 The Bayesian Model for a Single Variable

First define the following notation. Let

- d=1,..., D=2 denote corn and fish oil diet, respectively,
- t=1,..., T=3 denote control, DSS, and DSS and recovery treatments, respectively
- ℓ=1, ..., L=2 denote colon and duodenum location, respectively,
- f=1, ..., F=2 denote no-FPG and FPG type, respectively,
- r=1, ..., n_{dtl} denote rat with location ℓ, treatment t, and diet d, where n_{dtl} is the number of rats with treatment t and diet d, and
- c=1, ..., n_{rlf} denote cell from location ℓ in a rat r with FPG type f, where n_{rlf} is the number of observations for rat r in location ℓ with FPG type f.

The following model will use a single variable λ_{rlfc}, which will denote RTM or skewness. For a specific diet d, treatment t, and location ℓ:

$$\begin{aligned} \lambda_{rlfc} &= \lambda_{rlf} + \epsilon_{rlfc}\ , \quad \text{where } \epsilon_{rlfc} \sim \text{Normal}(0\ ,\ \sigma^2_{elf}); \\ \Lambda_{rl} &= \Lambda_{dtl} + \xi_{rl}\ , \quad \text{where } \xi_{rl} \sim \text{Normal}(\mathbf{0}\ ,\ \Sigma_{dtl}), \end{aligned} \tag{3.1}$$

where λ_{rlf} gives the mean for rat r at location ℓ for FPG type f, while ϵ_{rlfc} is the cell-level error term. The no-FPG term $\lambda_{rl(f=1)}$ and the FPG term $\lambda_{rl(f=2)}$ are combined

to form the 2×1 vector Λ_{rl}, respectively. The rat-level mean vector Λ_{rl} is modeled as a 2×1 diet-treatment mean vector Λ_{dtl} plus the rat-level error term ξ_{rl}.

The following priors are assumed:

$$\begin{aligned} \Lambda_{dtl} &\sim \text{Normal}(\Lambda_0 \ , \ \Sigma_0); \\ \Sigma_{dtl} &\sim \text{InverseWishart}(\upsilon_{r0} \ , \ \Sigma_{r0}); \\ \sigma^2_{elf} &\sim \text{InverseGamma}(a_{f0} \ , \ b_{f0}), \end{aligned} \tag{3.2}$$

where f is 1 for no-FPG and 2 for FPG. The InverseWishart(υ, Σ) and the InverseGamma(a, b) are defined in Appendix A, along with the variables' mean and variance or second moment.

To compute the likelihood, first note that the mean vectors Λ_{rl} at the rat level can be stated as follows:

$$\overline{\Lambda}_{rl} \sim \text{Normal}(\Lambda_{dtl} \ , \ \Sigma_{dtl} + \Sigma_{erl}),$$

where Σ_{erl} is a 2×2 matrix, with $\sigma^2_{el(f=1)}/n_{rl(f=1)}$ and $\sigma^2_{el(f=2)}/n_{rl(f=2)}$ on the diagonal, respectively, and zeros off the diagonal. The $\overline{\Lambda}_{rl}$ is the no-FPG and FPG mean of observations λ_{rlfc} for rat r and location ℓ. Then the likelihood for each location ℓ can be written as:

$$\begin{aligned} \prod_{\text{All } r} f(\overline{\Lambda}_{rl} | \cdots) &= \prod_{\text{All } r} |2\pi(\Sigma_{dtl} + \Sigma_{erl})|^{-1/2} \\ &\quad \times \exp\{-(\overline{\Lambda}_{rl} - \Lambda_{dtl})^{\mathrm{T}} (\Sigma_{dtl} + \Sigma_{erl})^{-1} (\overline{\Lambda}_{rl} - \Lambda_{dtl})/2\}. \end{aligned} \tag{3.3}$$

The following are the resulting posteriors for the diet-treatment mean vector Λ_{dtl}, rat-level mean λ_{rlf}, rat-level covariance matrix Σ_{dtl}, and cell-level variance σ^2_{elf}:

$$\begin{aligned}
\pi(\Lambda_{dtl}|\cdots) &= \text{Normal}\{(n_{dtl}\Sigma_{dtl}^{-1}+\Sigma_0^{-1})^{-1}(n_{dtl}\Sigma_{dtl}^{-1}\overline{\Lambda}_{dtl}+\Sigma_0^{-1}\Lambda_0), \\
&\qquad\qquad (n_{dtl}\Sigma_{dtl}^{-1}+\Sigma_0^{-1})^{-1}\}; \qquad (3.4)\\
\pi(\lambda_{rlf}|\cdots) &= \text{Normal}\left(\frac{n_{rlf}\sigma^2_{dtlf}\overline{\lambda}_{rlf}+\sigma^2_{elf}\lambda_{dtlf}}{n_{rlf}\sigma^2_{dtlf}+\sigma^2_{elf}},\ \frac{\sigma^2_{dtlf}\sigma^2_{elf}}{n_{rlf}\sigma^2_{dtlf}+\sigma^2_{elf}}\right); \\
\pi(\Sigma_{dtl}|\cdots) &= \text{InverseWishart}[n_{dtl}+\upsilon_{r0},\ \{\Sigma_{r0}^{-1}+\sum_{r=1}^{n_{dtl}}(\Lambda_{rl}-\Lambda_{dtl})(\Lambda_{rl}-\Lambda_{dtl})^{\mathrm{T}}\}^{-1}]; \\
\pi(\sigma^2_{elf}|\cdots) &= \text{InverseGamma}[a_{f0}+\frac{1}{2}n_{rlf},\ \{1/b_{f0}+0.5\sum_{r=1}^{n_{dtl}}\sum_{c=1}^{n_{rlf}}(\lambda_{rlfc}-\lambda_{rlf})^2\}^{-1}]
\end{aligned}$$

where σ^2_{dtlf} is the value of the appropriate diagonal entry of Σ_{dtl}, λ_{rlf} is the appropriate vector entry of Λ_{rl}, $\overline{\Lambda}_{dtl}$ is the average of Λ_{rl} for all rats r with diet d and treatment t, and $\overline{\lambda}_{rlf}$ is the average over all the values λ_{rlfc} for the same rat r, location ℓ, and FPG type f. The calculations are given in Appendix B.

3.3 Computing the Bayes Factor

The differences due to diet or treatment between the covariance matrices Σ as well as the mean vectors Λ will be tested. This will be done by comparing the results from the models using the covariance and means vector which depend on diet and treatment to models in which the covariance or mean vector depends on diet only, on treatment only, or on neither. To compare the different models, the Bayes Factor (BF) and the Likelihood Ratio Test (LRT) will be used.

For the LRT, each of the two likelihoods is computed using usual m.l.e. estimates, given in Appendix C.

The computation of the BF will follow the methodology presented by Chib (1995). The BF for comparison of models M_i and M_j can be written as given in

Section 2 of Chib (1995):

$$BF_{ij} \quad = \quad \exp[\log\{m(y|M_i)\} - \log\{m(y|M_j)\}].$$

The marginal likelihood $m(y|M)$ can be estimated by the log-likelihood plus the log-prior minus the log-posterior:

$$\log \widehat{m}(y|M) \quad = \quad \log\{f(y|\theta^*|M)\} + \log\{\pi(\theta^*, M)\} - \log\{\widehat{\pi}(\theta^*|y, M)\}, \quad (3.5)$$

where $f(y|\theta, M)$ is the density function of the data under model M, and θ^* is the estimated model parameters at their posterior mean or median.

For the FLARE model, each observation is represented by λ_{rlfc}, which is the data at the cell level, and θ is all the parameters of the model. The estimation of (3.5) consists of several steps. The first one of these is the selection of θ^*. This can be done by first running the usual Gibbs sampler given in (3.4). Either the resulting posterior mean or median can be chosen as to estimate parameters (θ^*). The log-likelihood $\log\{f(y|\theta^*)\}$ can then easily be computed by substituting θ^* into (3.3), and computing the resulting value. Similarly, the prior $\log\{\pi(\theta^*)\}$ can be computed by evaluating distributions given in (3.2) at θ^*.

However, estimating $\log\{\pi(\theta^*|y)\}$ is a little more challenging. Chib (1995) presents a method for estimating this value. In this evaluation, the rat-level mean vector Λ_{rlf} will be considered as a latent variable. It should be noted that the posterior of variance σ^2_{elf} at the rat-level depends only on Λ_{rlf}. Therefore, the estimate of $\widehat{\pi}(\sigma^{2\,*}_{elf}|y)$ can be computed using G iterations of the Gibbs sampler:

$$\widehat{\pi}(\sigma^{2\,*}_{elf}|\lambda_{rlfc}) \quad = \quad G^{-1}\sum_{g=1}^{G}\pi(\sigma^{2\,*}_{elf}|\lambda_{rlfc}, \Lambda^{(g)}_{rlf}).$$

In executing the above G iterations of the Gibbs sampler, all the variables are updated, including σ^2_{elf}. However, the posterior of σ^2_{elf} at each iteration is evaluated at $\sigma^{2\,*}_{elf}$.

Refer to Sections 2.1.2 and 2.1.3 in Chib (1995) for a general description of this algorithm.

The posterior of covariance Σ_{dtl} at the rat-level depends on the diet-treatment mean vectors Λ_{dtl} and the rat-level mean vectors Λ_{rlf}. Its posterior can be estimated in the following manner:

$$\widehat{\pi}(\Sigma^*_{dtl}|\lambda_{rlfc}) \;=\; G^{-1}\sum_{g=1}^{G}\pi(\Sigma^*_{dtl}|\lambda_{rlfc},\Lambda^{(g)}_{dtl},\Lambda^{(g)}_{rlf}).$$

Since Σ_{dtl} and σ^2_{elf} are independent, then both of their posteriors can be estimated in the same G iterations.

The posterior of diet-treatment mean vector Λ_{dtl} depends on the rat-level covariance Σ_{dtl} and the average Λ_{rlf} of cell-level data with same rat r, location ℓ, and FPG type f. Consequently, the posterior estimation of Λ_{dtl} requires an additional G iterations. The following estimates this posterior:

$$\widehat{\pi}(\Lambda^*_{dtl}|\lambda_{rlfc},\Sigma^*_{dtl}) \;=\; G^{-1}\sum_{g=1}^{G}\pi(\Lambda^*_{dtl}|\lambda_{rlfc},\Sigma^*_{dtl},\Lambda^{(g)}_{rlf}).$$

The additional iterations are performed by updating diet-treatment mean vectors Λ_{dtl} and rat-level mean vectors Λ_{rlf}. However, the values of rat-level covariance Σ_{dtl} and cell-level variance σ^2_{elf} are fixed at Σ^*_{dtl} and $\sigma^{2\,*}_{elf}$, respectively. A general example of this implementation is given at the end of Section 2.1.3 in Chib (1995).

Note that for all the models that use one variable, 10,000 iterations will be performed. The last 8,000 iterations will be used to find the posterior mean, and the last 2,000 iterations will be used to compute the posterior median.

3.4 Determining BF and LRT Distributions using Simulated Data

The simulated data will be used to see how using the BF to test model differences compares to using the LRT. The effect of using the posterior mean and median, as

well as different priors, in detection of the model differences will be considered. The model given in (3.1) will be fit to the simulated data. The simulated data will have similar structure to the actual colon FLARE data. The number of rats within each group will remain the same, but there will be 100 observations per rat. The function estimating values will be simulated.

Comparing the models will be done using the BF and the LRT. While the LRT has a known asymptotic distribution, the BF does not. However, Kass and Raftery (1995) discussed the cut-off points that can be used for the BF. They state that if the value of 2log(BF) is less than 2, it does not give much evidence against H_o. If it is between 2 and 6, it gives some positive evidence against H_o. Between 6 and 10, this is strong evidence, and above 10 it is very strong evidence. The above guidelines will be compared against the simulation results.

The data structure used to simulate the data is one where all mean vectors Λ_{rl} at rat level were sampled from multivariate normal with the mean of $[2.0, 2.2]^{\mathrm{T}}$, variance of 0.30, and covariance of 0.06 between the FPG types. The simulated variance at the cell level, σ^2_{elf}, was 0.0025. From the data structure, 300 data sets were simulated. The LRT and the BF for each data set were computed, and the distributions of the LRT and the BF were obtained.

In the data structure neither the covariance Σ_l nor the mean vector Λ_l at the rat level is dependent on diet or treatment. Prior 1 values from Table 3 were used. Model with covariance Σ_{dtl} depending on diet and treatment was compared to models with covariance Σ_{dl} depending on diet, covariance Σ_{tl} depending on treatment, and covariance Σ_l depending on just location, while mean vector Λ_{dtl} depending on diet and treatment was used in each case. Then the model with the mean vector Λ_{dtl} depending on diet and treatment was compared to the models with the mean vector Λ_{dl} depending on diet, mean vector Λ_{tl} depending on treatment, and mean vector Λ_l

Table 3. Single variable model priors.

Parameter	Prior1	Prior2	Prior3	Prior4	Prior5	Prior6
$(\Lambda_0)_1$	2.0	1.9	63.5	60.0	2.5	2.8
$(\Lambda_0)_2$	2.2	2.3	66.5	70.0	2.4	2.2
$(\Sigma_0)_{11}$, $(\Sigma_0)_{22}$	0.35	1.0	150.0	250.0	0.20	0.5
$(\Sigma_0)_{12}$, $(\Sigma_0)_{21}$	0.07	0.0	100.0	60.0	0.05	0.0
υ_{r0}	13	7	13	7	13	7
$(\Sigma_{r0}^{-1})_{11}$, $(\Sigma_{r0}^{-1})_{22}$	0.30	0.5	160.0	200.0	0.125	0.25
$(\Sigma_{r0}^{-1})_{12}$, $(\Sigma_{r0}^{-1})_{21}$	0.06	0.0	90.0	60.0	0.05	0.005
$a_{(f=1)0}$	5.0	5.0	5.0	5.0	5.0	5.0
$b_{(f=1)0}$	100.0	75.0	0.00025	0.00017	0.139	0.063
$a_{(f=2)0}$	5.0	5.0	5.0	5.0	5.0	5.0
$b_{(f=2)0}$	100.0	75.0	0.00025	0.00017	0.139	0.063

depending just on location, while the covariance Σ_l depending on just location was used in each case. Results were repeated using Prior 2 from Table 3. In each case, the H_o hypothesis of no difference due to diet or treatment holds.

Figure 5 displays the distribution of the LRT and the BF using posterior mean and posterior median when the model with covariance Σ_{dtl} depending on diet and treatment was compared to the models with covariance Σ_l depending on just location, Σ_{dl} depending on diet, and covariance Σ_{tl} depending on treatment, while the mean vector Λ_{dtl} depending on diet and treatment was used in each case. The LRT distribution had a χ^2 distribution superimposed in red. As can be seen from Figure 5, the LRT distributions follows a χ^2 distribution. However, the BF distributions are different from the LRT distributions. They also do not seem to follow the recommendations set by Kass and Raftery (1995). From the 300 simulations, all the values of $2log(BF)$ were greater than 10 for distributions testing for diet/treatment and diet differences, even though the H_o hypothesis of no difference due to diet or treatment holds. For the BF distribution testing for the treatment difference, more than 75% of values were greater than 10. The resulting distribution is similar whether the posterior mean or posterior median is used to compute BF values.

There are also differences between the BF distributions. The distribution testing for the diet/treatment difference is centered around 30, the one testing for diet difference is centered around 20, and the distribution testing for treatment difference is centered around 12. From this it can be concluded that as degrees of freedom become higher, the center of the distribution is also higher. Therefore the cut-off points specified by Kass and Raftery (1995) may not hold because these are higher dimensional models. The mean, along with the 50%, 75%, 90%, and 95% perctentiles for each distribution in Figure 5 are given in Table 4.

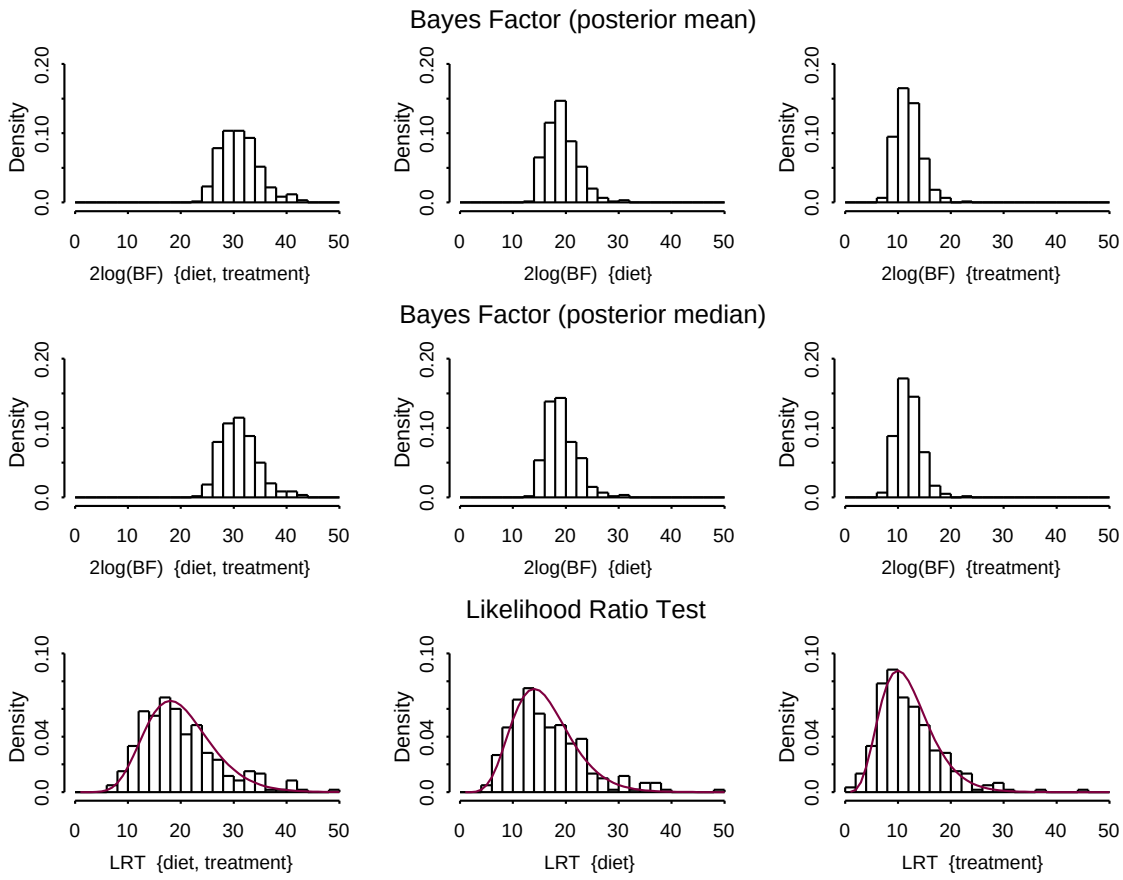

Figure 5. Distribution of 2log(BF) and LRT testing for diet and/or treatment effect in the covariance matrix when there is not a difference in the covariance matrix due to diet or treatment. For BF, prior 1 is used. For LRT, χ_{20}, χ_{16}, and χ_{12} densities are superimposed in red, respectively.

Table 4. Description of simulated distributions.

Variable	Prior	Posterior Type	Mean	50% Perc.	75% Perc.	90% Perc.	95% Perc.
		Σ_l *vs* Σ_{dtl}					
2log(BF)	1	mean	31.13	30.84	33.34	35.52	37.18
2log(BF)	2	mean	33.77	33.28	35.39	37.31	40.16
2log(BF)	1	median	31.08	30.86	33.24	35.25	37.10
2log(BF)	2	median	34.14	33.64	35.75	37.72	40.19
LRT			19.88	18.39	23.50	30.07	34.25
		Σ_{dl} *vs* Σ_{dtl}					
2log(BF)	1	mean	19.24	18.77	21.19	23.38	24.41
2log(BF)	2	mean	20.71	20.30	22.23	24.57	26.12
2log(BF)	1	median	19.21	18.80	20.96	22.98	24.03
2log(BF)	2	median	21.36	20.95	22.81	25.06	26.49
LRT			16.38	14.93	20.03	24.74	30.49
		Σ_{tl} *vs* Σ_{dtl}					
2log(BF)	1	mean	11.95	11.79	13.19	15.17	16.07
2log(BF)	2	mean	12.72	12.41	14.03	15.76	16.96
2log(BF)	1	median	11.99	11.78	13.22	14.92	15.72
2log(BF)	2	median	13.39	13.03	14.55	16.19	17.76
LRT			12.13	11.05	15.06	19.55	22.61
		Λ_l *vs* Λ_{dtl}					
2log(BF)	1	mean	31.08	30.30	34.26	37.73	40.50
2log(BF)	2	mean	31.67	30.91	34.16	37.09	38.92
2log(BF)	1	median	31.30	30.47	34.42	37.81	40.53
2log(BF)	2	median	31.84	31.13	34.24	37.31	39.01
LRT			10.96	10.20	13.83	17.21	19.88
		Λ_{dl} *vs* Λ_{dtl}					
2log(BF)	1	mean	20.74	19.65	23.11	26.74	30.11
2log(BF)	2	mean	21.16	20.33	23.34	26.30	28.88
2log(BF)	1	median	20.86	19.81	23.19	26.81	30.27
2log(BF)	2	median	21.41	20.57	23.60	26.64	29.15
LRT			8.64	7.65	11.08	14.22	18.25
		Λ_{tl} *vs* Λ_{dtl}					
2log(BF)	1	mean	14.14	13.44	16.22	19.43	21.73
2log(BF)	2	mean	14.39	13.84	16.23	18.97	20.93
2log(BF)	1	median	14.18	13.43	16.26	19.36	21.66
2log(BF)	2	median	14.53	13.98	16.29	19.17	20.89
LRT			6.72	5.88	9.00	12.30	14.34

The models were recalculated using Prior 2 from Table 3. The LRT and the BF distributions are shown in Figure 6, and the percentiles of these distribution are given in Table 4. Again, the results using posterior mean or posterior median are fairly close. Bigger differences are seen between the distributions when results from Prior 1 and 2 are compared. Note that the LRT results do not depend on a prior.

Models testing for changes in the mean vectors were fitted next. The model with mean vector Λ_{dtl} depending on diet and treatment was compared to the models with mean vector Λ_l depending just on location, Λ_{dl} depending on diet, and mean vector Λ_{tl} depending on diet, while the covariance Σ_l depending on just location was used in each case. Resulting distributions are shown in Figure 7 when Prior 1 was used, and in Figure 8 when Prior 2 was used. Here again we see that LRT follows a χ^2 distribution, but the BF does not follow a χ^2 distribution. There again is not much difference between the distributions of the BF using posterior mean or posterior median. There is more of a difference between the BF distributions when Prior 2 is used instead of Prior 1. The mean and the 50%, 75%, 90%, and 95% perctentiles for these distributions are given in Table 4.

From the above simulation results, while the LRT followed a χ^2 distribution as was expected, the BF did not. The distribution of the BF appeared to be influenced by the type of prior that was used, but not as much whether the posterior mean or the posterior median was used. As the number of degrees of freedom increased, the listed percentile values also increased. However, these values were higher than what was recommended by Kass and Raftery (1995). Consequently, the value of the 95% percentile will be used as the new cut-off points in determining differences between models. However, this will not be a strict guideline since it was found that these values are influenced by the prior used. Lastly, it should be noted that there does not always seem to be a direct correspondence between the degrees of freedom and the

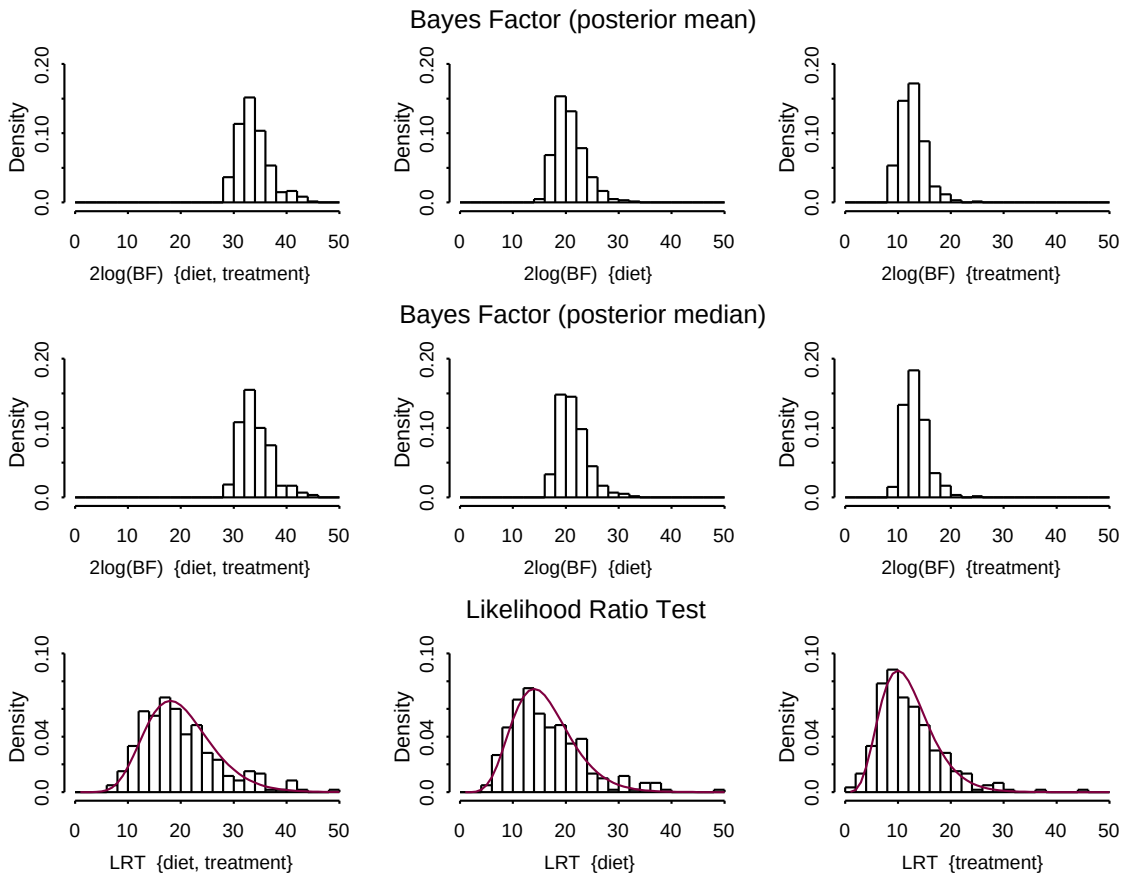

Figure 6. Distribution of 2log(BF) and LRT testing for diet and/or treatment effect in the covariance matrix when there is not a difference in the covariance matrix due to diet or treatment. For BF, prior 2 is used. For LRT, χ_{20}*,* χ_{16}*, and* χ_{12} *densities are superimposed in red, respectively.*

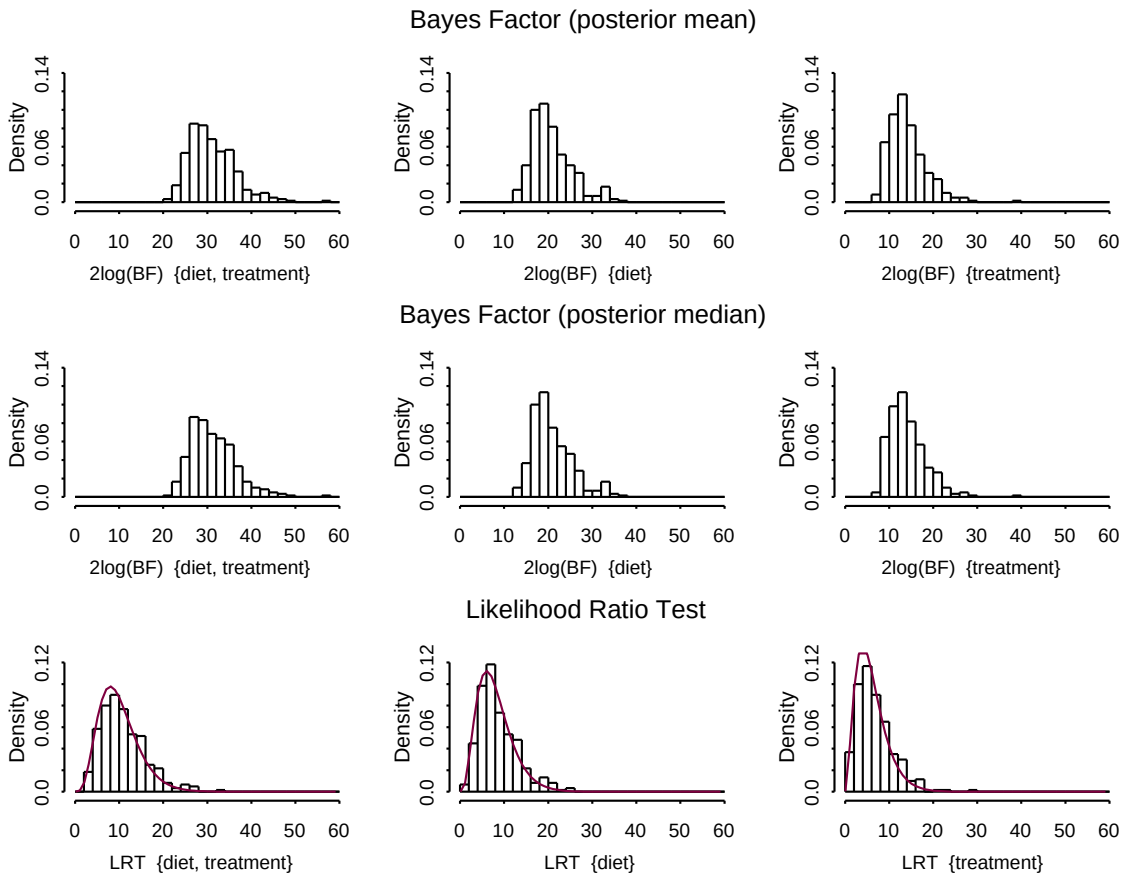

Figure 7. Distribution of 2log(BF) and LRT testing for diet and/or treatment effect in the mean vector when there is not a difference in the mean vector due to diet or treatment. For BF, prior 1 is used. For LRT χ_{10}, χ_8, and χ_6 densities are superimposed in red, respectively.

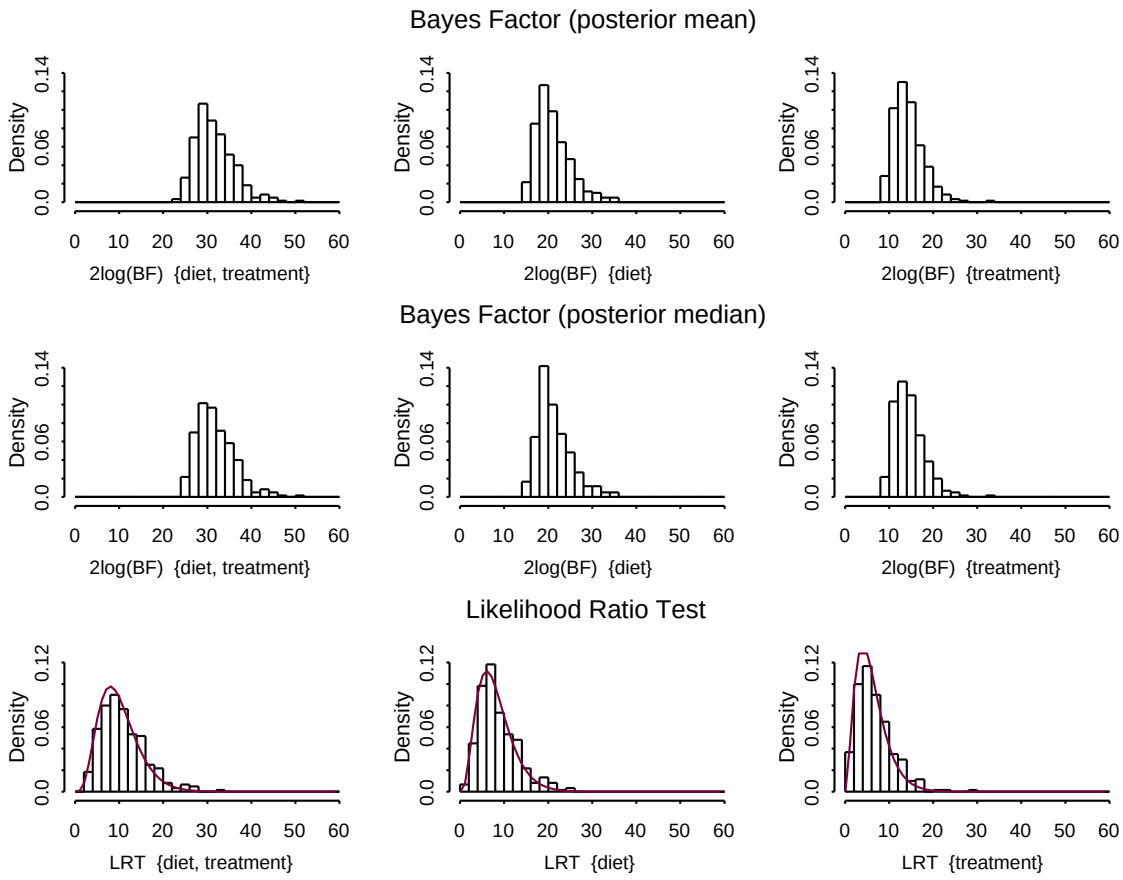

Figure 8. Distribution of 2log(BF) and LRT testing for diet and/or treatment effect in the mean vector when there is not a difference in the mean vector due to diet or treatment. For BF, prior 2 is used. For LRT, χ_{10}, χ_8, and χ_6 densities are superimposed in red, respectively.

value of the BF cut-off points. When testing for the diet and treatment differences in the mean vector, the cut-off points appear to be similar to the cut-off points for testing diet and treatment differences for the covariance matrix.

3.5 Modeling RTM Data

One of the variables previously used to summarize each comet function was the relative tail moment (RTM). To test the diet and treatment effects in colon and duodenum, the model similar in structure to the one given in (3.1) was used. Models will be fit to the data to determine whether the covariance Σ and the mean vector Λ at the rat level depend on diet and/or treatment.

The models will be compared using the LRT and the BF. For the LRT, a χ^2 distribution is assumed. The 95% percentile of the BF simulated distribution will be used as a cut-off point for determining a difference in the models being compared. Again, these will be only used as guidelines since the cut-off points were found to vary depending on the prior used, and the priors and simulated data differed from the priors and data for the RTM. For the comparison of the two models, both the posterior mean and posterior median were used as variable estimates. Also, Priors 3 and 4 were used, for which values are given in Table 3. The prior values were set relative to the m.l.e. values, given in Table 5.

The results for the RTM in the colon were first considered. To determine if the covariance Σ and the mean vector Λ are dependent on diet or treatment, different models were compared. The models with the mean vector dependent on diet and treatment (Λ_{dtl}), dependent on diet (Λ_{dl}),dependent on treatment (Λ_{tl}), and not dependent on them (Λ_l) were compared, while the covariance Σ_{dtl} depending on diet and treatment was used in all of the models. There were no differences based on diet or treatment in the means vector since the LRT p-value was greater than 0.05 and

Table 5. RTM estimated parameters.

Diet	Treatment	Prior	Estimator	$(\widehat{\Lambda})_1$	$(\widehat{\Lambda})_2$	$(\widehat{\Sigma})_{11}$	$(\widehat{\Sigma})_{22}$	$(\widehat{\Sigma})_{12}$
			Colon					
			m.l.e.	61.46	65.36	131.01	111.33	90.50
		3	mean	61.54	65.42	130.95	115.12	83.55
		3	median	61.53	65.40	127.92	112.72	81.42
		4	mean	61.49	65.39	130.49	112.95	80.75
		4	median	61.53	65.39	126.82	108.79	77.78
			Duodenum					
corn	cont		m.l.e.	65.72	68.03	26.00	116.60	26.51
corn	cont	5	mean	65.70	68.04	95.67	137.20	57.58
corn	cont	5	median	65.69	68.02	88.88	128.54	52.42
corn	cont	6	mean	65.70	68.08	77.98	139.16	34.37
corn	cont	6	median	65.74	68.11	71.98	127.73	31.04
corn	DSS		m.l.e.	65.72	68.03	216.80	159.46	159.26
corn	DSS	5	mean	65.70	68.04	182.88	157.03	116.94
corn	DSS	5	median	65.69	68.02	171.01	145.71	107.66
corn	DSS	6	mean	65.70	68.08	206.88	168.20	120.98
corn	DSS	6	median	65.74	68.11	183.74	150.57	104.95
corn	rec		m.l.e.	65.72	68.03	40.76	99.91	45.66
corn	rec	5	mean	65.70	68.04	101.85	129.27	66.22
corn	rec	5	median	65.69	68.02	94.90	120.32	60.67
corn	rec	6	mean	65.70	68.08	87.19	127.80	46.88
corn	rec	6	median	65.74	68.11	80.49	117.14	41.26
fish	cont		m.l.e.	65.72	68.03	156.97	186.62	151.91
fish	cont	5	mean	65.70	68.04	157.15	171.86	116.72
fish	cont	5	median	65.69	68.02	147.49	162.22	107.79
fish	cont	6	mean	65.70	68.08	165.27	187.56	117.91
fish	cont	6	median	65.74	68.11	147.91	169.64	104.75
fish	DSS		m.l.e.	65.72	68.03	129.12	90.25	85.45
fish	DSS	5	mean	65.70	68.04	140.33	125.37	84.31
fish	DSS	5	median	65.69	68.02	129.60	116.71	78.01
fish	DSS	6	mean	65.70	68.08	143.39	119.74	71.62
fish	DSS	6	median	65.74	68.11	128.87	107.43	62.99
fish	rec		m.l.e.	65.72	68.03	236.95	295.06	183.01
fish	rec	5	mean	65.70	68.04	194.50	223.48	132.16
fish	rec	5	median	65.69	68.02	181.40	211.41	123.22
fish	rec	6	mean	65.70	68.08	219.49	258.85	139.49
fish	rec	6	median	65.74	68.11	202.39	230.95	125.09

NOTE: For the estimator, "mean" denotes posterior mean, and "median" denotes posterior median; for the treatment, "cont" denotes control treatment, and "rec" denotes the recovery treatment.

the $2log(BF)$ values were less than 38.92. Please refer to Table 6 for the BF and the LRT values, and to Table 4 for the BF cut-off values. It was also tested whether the covariance should depend on diet and treatment (Σ_{dtl}), on diet (Σ_{dl}), on treatment (Σ_{tl}), or not (Σ_l), while the mean vector Λ_{dtl} was considered to be diet and treatment dependent. Both the BF and the LRT indicated that the covariance Σ_l depending only on location could be used for the colon model since all of the LRT p-values were greater than 0.05 and the 2log(BF) values were less than 37.10 (refer to Table 4 for cut-off values). Consequently, a model with Λ_l and Σ_l not depending on diet or treatment sufficiently describes the RTM colon data.

The variable estimates using the posterior mean, posterior median, Prior 3, and Prior 4 are given in Table 5. It can also be seen that the estimates of the mean vector Λ_l do not differ much across the two priors or based on the variable estimate used. However, the estimate of covariance Σ_l differs more. This is partly due to the fact that the posterior distribution of means vector Λ_l is symmetric, and the posterior distribution of covariance Σ_l is skewed. The estimates associated with the posterior median are smaller for the covariance Σ_l than the estimates associated with the posterior mean. Note that for the above results, no-FPG estimate is denoted by $\widehat{\Lambda}_1$, and the FPG estimate is denoted by $\widehat{\Lambda}_2$.

Similarly, the results for the RTM in the duodenum were analyzed. Using the LRT, while the mean vector Λ_l was not found to depend on the diet or treatment, the covariance Σ_{dtl} was found to be dependent on both diet and treatment. For the BF values, no diet or treatment effect was found for the mean vector Λ_l. In testing for diet and/or treatment differences for the covariance matrix, the values using Prior 4 are close to the cut-off points given in Table 4, indicating that there may be a possible diet or treatment effect. The LRT and the BF values are given in Table 6. The estimates using the m.l.e., the posterior mean and median for variable estimates,

Table 6. Model comparison for RTM data using priors 3 and 4.

Location	Prior Used	Models Compared	2log(BF) (p. mean)	2log(BF) (p. median)	LRT	df	LRT p-value
Colon	3	Λ_l vs. Λ_{dtl}	23.18	23.82	4.69	10	0.9109
Colon	3	Λ_{dl} vs. Λ_{dtl}	15.03	15.52	4.80	8	0.7787
Colon	3	Λ_{tl} vs. Λ_{dtl}	8.76	9.18	2.38	6	0.8821
Colon	3	Σ_l vs. Σ_{dtl}	24.84	25.42	19.88	20	0.4652
Colon	3	Σ_{dl} vs. Σ_{dtl}	14.21	14.51	19.66	16	0.2361
Colon	3	Σ_{tl} vs. Σ_{dtl}	8.34	8.93	12.88	12	0.3775
Colon	4	Λ_l vs. Λ_{dtl}	23.47	24.26	4.69	10	0.9109
Colon	4	Λ_{dl} vs. Λ_{dtl}	15.33	15.90	4.80	8	0.7787
Colon	4	Λ_{tl} vs. Λ_{dtl}	8.86	9.26	2.38	6	0.8821
Colon	4	Σ_l vs. Σ_{dtl}	34.42	35.10	19.88	20	0.4652
Colon	4	Σ_{dl} vs. Σ_{dtl}	20.61	21.03	19.66	16	0.2361
Colon	4	Σ_{tl} vs. Σ_{dtl}	12.33	13.17	12.88	12	0.3775
Duodenum	3	Λ_l vs. Λ_{dtl}	25.33	25.91	9.93	10	0.4468
Duodenum	3	Λ_{dl} vs. Λ_{dtl}	15.01	15.42	5.95	8	0.6526
Duodenum	3	Λ_{tl} vs. Λ_{dtl}	10.71	10.97	4.34	6	0.6308
Duodenum	3	Σ_l vs. Σ_{dtl}	27.73	27.80	30.42	20	0.0634
Duodenum	3	Σ_{dl} vs. Σ_{dtl}	16.51	16.78	27.27	16	0.0386
Duodenum	3	Σ_{tl} vs. Σ_{dtl}	11.33	11.59	23.10	12	0.0269
Duodenum	4	Λ_l vs. Λ_{dtl}	25.80	26.60	9.93	10	0.4468
Duodenum	4	Λ_{dl} vs. Λ_{dtl}	15.00	15.64	5.95	8	0.6526
Duodenum	4	Λ_{tl} vs. Λ_{dtl}	10.80	11.22	4.34	6	0.6308
Duodenum	4	Σ_l vs. Σ_{dtl}	36.69	37.27	30.42	20	0.0634
Duodenum	4	Σ_{dl} vs. Σ_{dtl}	22.10	22.45	27.27	16	0.0386
Duodenum	4	Σ_{tl} vs. Σ_{dtl}	15.77	16.12	23.10	12	0.0269

NOTE: It is tested whether covariance structure is diet and treatment dependent (Σ_{dtl}), diet dependent (Σ_{dl}), treatment dependent (Σ_{tl}), or not (Σ_l), and whether the means structure is diet and treatment dependent (Λ_{dtl}), diet dependent (Λ_{dl}), treatment dependent (Λ_{tl}), or only location dependent (Λ_l). Computed is the likelihood ratio test (LRT) and the corresponding p-value, as well as 2log(Bayes Factor). Results are given when variables are evaluated at the posterior mean and median.

and Priors 3 and 4 are given in Table 5. The results for the mean vector estimates are similar to the m.l.e results, also given in the same table. However, there is some distinction between the m.l.e. and the Bayesian covariate estimates.

Consequently, when using the RTM to summarize FLARE functions, there are no diet or treatment effects found in the colon, but there are covariance differences in the duodenum due to diet and treatment.

3.6 Modeling Skewness Data

The variable skewness will be considered since it measures the shape of the comet. It is hypothesized that the quantified shape of the comet will be a better measure of damage than the relative tail moment RTM. When shown an intensity histogram of a comet, it is easy to determine how much damage is present just by looking at the shape of the histogram. Here the shape of a comet function will be quantified by skewness. The skewness data will be analyzed separately for the colon and the duodenum. The model used will be similar to the model presented in (3.1). As with the RTM data, first a proper model will be chosen. Once the model is chosen, the estimates will be considered to see how diet or treatment estimates differ.

The models will be implemented using posterior median and posterior mean to estimate the variables. Also, Priors 5 and 6 will be considered, the values for which are given in Table 3. The prior values were set relative to the m.l.e. values, given in Tables 7 and 8.

A diet and treatment effect was found for the mean vector Λ_{dtl} in the colon, and a diet effect for the mean vector Λ_{dl} in the duodenum since the LRT p-values were less than 0.05. Refer to Table 9 for the results. While the colon $2log(BF)$ values are less than 38.92 (refer to Table 4 for cut-off values), values in the duodenum are close to that cut-off point. Comparing the model where the covariance matrix Σ_{dtl} was

Table 7. Skewness estimated parameters for colon.

Diet	Treatment	Prior	Estimator	$(\widehat{\Lambda})_1$	$(\widehat{\Lambda})_2$	$(\widehat{\Sigma})_{11}$	$(\widehat{\Sigma})_{22}$	$(\widehat{\Sigma})_{12}$
corn	cont		m.l.e.	2.46	2.47	0.063	0.063	0.034
corn	cont	5	mean	2.46	2.46	0.070	0.071	0.030
corn	cont	5	median	2.46	2.46	0.069	0.070	0.029
corn	cont	6	mean	2.46	2.47	0.074	0.075	0.025
corn	cont	6	median	2.47	2.47	0.072	0.073	0.024
corn	DSS		m.l.e.	2.40	2.32	0.063	0.063	0.034
corn	DSS	5	mean	2.41	2.32	0.070	0.071	0.030
corn	DSS	5	median	2.41	2.33	0.069	0.070	0.029
corn	DSS	6	mean	2.41	2.32	0.074	0.075	0.025
corn	DSS	6	median	2.41	2.32	0.072	0.073	0.024
corn	rec		m.l.e.	2.51	2.29	0.063	0.063	0.034
corn	rec	5	mean	2.51	2.29	0.070	0.071	0.030
corn	rec	5	median	2.50	2.29	0.069	0.070	0.029
corn	rec	6	mean	2.51	2.29	0.074	0.075	0.025
corn	rec	6	median	2.51	2.29	0.072	0.073	0.024
fish	cont		m.l.e.	2.63	2.43	0.063	0.063	0.034
fish	cont	5	mean	2.62	2.43	0.070	0.071	0.030
fish	cont	5	median	2.63	2.43	0.069	0.070	0.029
fish	cont	6	mean	2.63	2.43	0.074	0.075	0.025
fish	cont	6	median	2.63	2.43	0.072	0.073	0.024
fish	DSS		m.l.e.	2.60	2.49	0.063	0.063	0.034
fish	DSS	5	mean	2.60	2.50	0.070	0.071	0.030
fish	DSS	5	median	2.60	2.50	0.069	0.070	0.029
fish	DSS	6	mean	2.61	2.50	0.074	0.075	0.025
fish	DSS	6	median	2.60	2.50	0.072	0.073	0.024
fish	rec		m.l.e.	2.34	2.19	0.063	0.063	0.034
fish	rec	5	mean	2.35	2.20	0.070	0.071	0.030
fish	rec	5	median	2.36	2.20	0.069	0.070	0.029
fish	rec	6	mean	2.35	2.19	0.074	0.075	0.025
fish	rec	6	median	2.35	2.19	0.072	0.073	0.024

NOTE: For the estimator, "mean" denotes posterior mean, and "median" denotes posterior median; for the treatment, "cont" denotes control treatment, and "rec" denotes the recovery treatment.

Table 8. Skewness estimated parameters for duodenum.

Diet	Prior	Estimator	$(\widehat{\Lambda})_1$	$(\widehat{\Lambda})_2$	$(\widehat{\Sigma})_{11}$	$(\widehat{\Sigma})_{22}$	$(\widehat{\Sigma})_{12}$
corn		m.l.e.	2.53	2.48	0.120	0.151	0.101
corn	5	mean	2.53	2.48	0.110	0.138	0.079
corn	5	median	2.53	2.48	0.108	0.135	0.076
corn	6	mean	2.53	2.48	0.118	0.150	0.079
corn	6	median	2.53	2.48	0.117	0.145	0.077
fish		m.l.e.	2.47	2.26	0.120	0.151	0.101
fish	5	mean	2.48	2.26	0.110	0.138	0.079
fish	5	median	2.48	2.27	0.108	0.135	0.076
fish	6	mean	2.47	2.26	0.118	0.150	0.079
fish	6	median	2.48	2.26	0.117	0.145	0.077

NOTE: For the estimator, "mean" denotes posterior mean, and "median" denotes posterior median.

Table 9. Model comparison for skewness data using priors 5 and 6.

Location	Prior Used	Models Compared	2log(BF) (p. mean)	2log(BF) (p. median)	LRT	df	LRT p-value
Colon	5	Λ_l vs. Λ_{dtl}	30.56	31.07	19.12	10	0.0388
Colon	5	Λ_{dl} vs. Λ_{dtl}	21.43	21.98	16.79	8	0.0324
Colon	5	Λ_{tl} vs. Λ_{dtl}	13.46	13.77	14.91	6	0.0210
Colon	5	Σ_l vs. Σ_{dtl}	27.99	28.80	23.84	20	0.2496
Colon	5	Σ_{dl} vs. Σ_{dtl}	14.79	15.89	10.64	16	0.8314
Colon	5	Σ_{tl} vs. Σ_{dtl}	8.40	9.18	17.68	12	0.1257
Colon	6	Λ_l vs. Λ_{dtl}	29.23	30.18	19.12	10	0.0388
Colon	6	Λ_{dl} vs. Λ_{dtl}	20.33	20.96	16.79	8	0.0324
Colon	6	Λ_{tl} vs. Λ_{dtl}	12.96	13.42	14.91	6	0.0210
Colon	6	Σ_l vs. Σ_{dtl}	47.61	49.65	23.84	20	0.2496
Colon	6	Σ_{dl} vs. Σ_{dtl}	22.62	25.09	10.64	16	0.8314
Colon	6	Σ_{tl} vs. Σ_{dtl}	12.96	14.36	17.68	12	0.1257
Duodenum	5	Λ_l vs. Λ_{dtl}	36.77	37.46	21.39	10	0.0185
Duodenum	5	Λ_{dl} vs. Λ_{dtl}	22.74	23.10	13.47	8	0.0966
Duodenum	5	Λ_{tl} vs. Λ_{dtl}	18.55	18.81	15.44	6	0.0171
Duodenum	5	Σ_l vs. Σ_{dtl}	25.40	24.85	9.27	20	0.9796
Duodenum	5	Σ_{dl} vs. Σ_{dtl}	12.76	12.34	6.78	16	0.9773
Duodenum	5	Σ_{tl} vs. Σ_{dtl}	8.50	8.17	8.07	12	0.7797
Duodenum	6	Λ_l vs. Λ_{dtl}	34.68	35.55	21.39	10	0.0185
Duodenum	6	Λ_{dl} vs. Λ_{dtl}	21.66	22.39	13.47	8	0.0966
Duodenum	6	Λ_{tl} vs. Λ_{dtl}	16.93	17.36	15.44	6	0.0171
Duodenum	6	Σ_l vs. Σ_{dtl}	38.84	39.26	9.27	20	0.9796
Duodenum	6	Σ_{dl} vs. Σ_{dtl}	18.63	19.67	6.78	16	0.9773
Duodenum	6	Σ_{tl} vs. Σ_{dtl}	11.18	11.51	8.07	12	0.7797

NOTE: It is tested whether covariance structure is diet and treatment dependent (Σ_{dtl}), diet dependent (Σ_{dl}), treatment dependent (Σ_{tl}), or not (Σ_l), and whether the means structure is diet and treatment dependent (Λ_{dtl}), diet dependent (Λ_{dl}), treatment dependent (Λ_{tl}), or only location dependent (Λ_l). Computed is the likelihood ratio test (LRT) and the corresponding p-value, as well as 2log(Bayes Factor). Results are given when variables are evaluated at the posterior mean and median.

set to be diet and treatment dependent to model with covariance matrix Σ_l only location dependent, no difference was found between the two models using LRT in both colon and the duodenum. Refer to Table 9 for the LRT values. Using the BF and cut-off values in Table 4, there may be a difference in the covariance due to diet and/or treatment when Prior 6 is used. However, since the cut-off values are not confirmed to be reliable, the LRT results will be used instead. Considering these results, the model with the mean vector Λ_{dtl} depending on diet and treatment and the covariance Σ_l at the rat level depending only on location was chosen as sufficiently describing the skewness data for the colon. The model with the mean vector Λ_{dl} depending on diet and the covariance Σ_l at the rat level depending only on location was chosen for the duodenum.

The resulting estimates for the m.l.e., posterior medians and posterior means, as well as the Prior 5 and Prior 6 are given in Table 7 for colon and Table 8 for duodenum. The Bayesian estimates are fairly close together. While the m.l.e. mean vector estimates are also similar to the Bayesian estimates, they differ slightly for the covariance estimates. In the results, the $\widehat{\Lambda}_1$ estimate corresponds to the no-FPG skewness estimate, and the $\widehat{\Lambda}_2$ estimate to the FPG skewness estimate. We can see that the mean vector for FPG skewness is smaller than for the no-FPG skewness. This was something that was observed in Chapter II analysis of the skewness data as well.

3.7 Discussion

The simulated data was used to examine effectiveness of LRT and BF in choosing the correct model, and how the BF is effected by priors and posterior means and medians. While it was shown that the LRT follows a χ^2 distribution, the BF distribution does not. Initially, the cut-off points suggested by Kass and Raftery (1995) were intended

to be used. However, these were not found to be effective for these higher dimensional models. While their paper suggested that a difference of 10 gave a very strong evidence again H_o, in the simulated data results most of the $2log(BF)$ values were greater than 10, even though H_o hypothesis was valid.

From the simulated results, it was also found that the type of prior chosen will influence the cut-off point (the 95% percentile of the distribution). The effect of using the posterior mean or posterior median was not as influential on the cut-off point as was the prior. Because of the variability in the cut-off point due to the prior, for analyzing diet and treatment differences for the variables RTM and skewness, the BF results were not given as much weight as the LRT results.

For both RTM and skewness differences were detected. For the RTM colon, no diet or treatment differences were found in the mean vector or the covariance matrix. For the RTM duodenum results, while there were no differences in the mean vector due to diet or treatment, differences in the covariance matrix were detected. The fish no-FPG and FPG variance estimates were more similar than the corn estimates. To test where the differences occur within each FPG type, the LRT was performed for the FPG and the no-FPG results separately using the m.l.e. values. While no effect due to diet or treatment was detected in the FPG variance, a diet and treatment effect was detected in the no-FPG variance. The LRT p-value is 0.0006 for testing a diet effect in the variance, and 0.0027 for testing a treatment effect in the variance. For the control and recovery treatments, the no-FPG fish variance estimate is larger than the no-FPG corn variance estimate (refer to Table 5). For the DSS, the reverse holds.

The differences found using skewness occurred in both the colon and the duodenum. In the colon, differences due to diet and treatment were detected for the means vector, while in the duodenum only diet differences were detected for the means vec-

tor. In the colon, the largest difference between the FPG estimates for corn oil occurs for the recovery treatment, while the biggest difference for the fish oil occurs for the control treatment. However, estimates for no-FPG and FPG for the fish oil and recovery treatment in the colon are lower the remaining corresponding estimates. For the duodenum, the difference between the FPG estimates is seen mostly for the fish oil. The LRT was then performed for the FPG and the no-FPG results separately, using the m.l.e. values. For the no-FPG colon, diet differences were detected (p-value 0.0274). For the FPG colon, the treatment differences appeared (p-value 0.0208), with the recovery treatment value being lower than the control or the DSS values (refer to Table 7). In the duodenum, when the separate analysis was performed, only the FPG diet differences were detected (p-value 0.0188), with the fish estimate being lower than the corn estimate (refer to Table 8).

CHAPTER IV

MULTI VARIABLE BAYESIAN MODEL FOR FLARE FUNCTIONS

Previous studies examined FLARE data by summarizing each function using the relative tail moment (RTM) or the skewness. In this chapter, we will examine a model for the FLARE function, as opposed to a single summary measure of it.

To obtain a basic idea of the shape of the FLARE functions, first a non-model based representative function at the diet and treatment level will be shown. Then a seven point model will be presented, where each of the seven points is modeled independently. Finally, a hierarchical model, which account for correlation of the points at the rat and comet function level, will be fitted using seven points from each comet function. Models using the Bayesian and the m.l.e. parameter estimates will be fitted.

Diet and treatment differences will be shown in the colon and the duodenum when the 7-point model is used to represent the data. The latter type of analysis is more informative than when functions are summarized by a single number since the results are functions that can be easily compared for the diet or treatment differences.

4.1 Introduction

While using the RTM or skewness to summarize each function shows if there are any diet or treatment differences, it is hypothesized that a more powerful analysis will be modeling the function itself. The nature of the comet function will be first described by computing representative functions. These will describe the expected comet functions for different diet and treatments, without formally testing if there are significant differences between them. Then each function will be summarized

by seven equally-spaced points. The seven points will first be analyzed separately, then by using a Bayesian approach that involves modeling the correlation between the points at the comet as well as the rat level. The FPG and the no-FPG 7-point functions will be estimated individually and modeled simultaneously.

Diet and treatment differences will be more apparent when a 7-point model is considered rather than when a single variable model is used. The results from the 7-point analysis are also easier to represent and interpret visually than results from analysis where each function is summarized by a single number.

4.2 Representative Functions

This section will describe the basic shape of the comet functions using representative functions, without performing actual tests for differences due to diet or treatment. Previously, each function was summarized by a single number. By considering the function itself, and not just a function summary, changes due to diet or treatment should be more apparent.

The representative functions show the expected shape of a comet at a diet or treatment level. To obtain them, the mean is taken of all the comet functions for each rat, then the mean is taken over the obtained rat comet functions for all the rats with the same diet or treatment. Even though the means are taken over each location on the horizontal axis separately, the area under each of the resulting functions is one.

The representative functions are constructed in the following manner. First, the intensity histograms for all comets had the center of the head set as zero on the horizontal axis so that they would be aligned. The intensities for all the comets were standardized so that the area under the intensity histogram was one. Let the intensity histograms of these comets be denoted by $f_{dtrlfc}(x_i)$, where d denotes the diet type, t denotes the treatment, r the rat given diet d and treatment t, ℓ the location in the

rat, f the FPG type, and c a comet obtained from rat r in location ℓ. Then for all the comets for rat r, location ℓ, and FPG type f, the mean intensity for each horizontal axis location is found. In the above notation, this means that

$$f_{dtrlf}(x_i) = \sum_{c \subset rlf} f_{dtrlfc}(x_i)/n_{rlf}$$

for each value of x_i. Above, n_{rlf} is the number of comets for rat r, location ℓ, and FPG type f. The average is then taken for all the rats with the same location and FPG type

$$f_{dtlf}(x_i) = \sum_{r \subset dt} f_{dtrlf}(x_i)/n_{dtl}$$

for each horizontal axis location x_i. Here, n_{dtl} is the number of rats on diet d and treatment t. From the resulting function $f_{dtlf}(x_i)$, Figure 9 was created. The resulting functions are called representative functions since each function is the average function for a specific diet, treatment, and FPG type. The area under all the functions shown in Figure 9 is one. Each box shows the combination of diet and FPG type functions for each location and treatment. By considering Figure 9, there appears to be more oxidative damage for results from the fish oil diet rather than from the corn oil diet for the representative functions. This difference appears to be more noticeable in the duodenum rather than the colon, especially when the treatment DSS with recovery was used.

While the representative functions are informative in where we expect to see diet or treatment differences, they are not conclusive since they do not describe the amount of variability around them. Therefore, it is not possible to describe if any of the differences are significant. The next three sections on the 7-point function analysis will statistically evaluate these differences.

Additional figures for the representative functions are available in Appendix D.

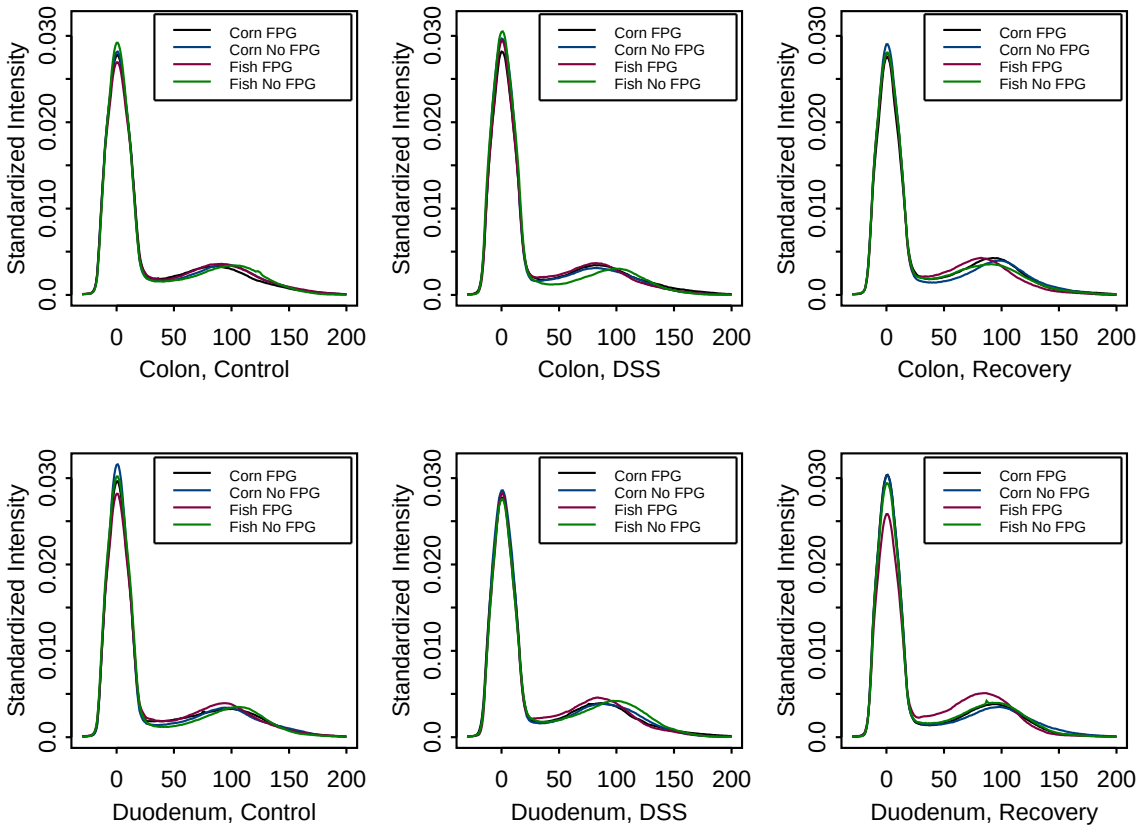

Figure 9. Representative functions for diet and FPG types for each location and treatment. Black is corn FPG, blue is corn no-FPG, red is fish FPG, and green is fish no-FPG.

4.3 Model for Seven Points: Independent Analysis

The representative functions do not include a measurement of variability. Therefore, to test any significant diet differences, seven evenly-spaced points were picked out along the length of each of the functions, where the length of the comet was measured in number of horizontal pixels in the image. The standardized intensity at the chosen length, along with the standardized intensities for two pixels on each side of the chosen length, were averaged. The average was used as the measure of the intensity of the function at each chosen location. Like in the representative functions, the considered intensities were standardized by the total intensity of the comet. For each location separately, it was tested whether there were any significant diet or treatment differences.

The seven points were first picked only from the FPG functions. A standard mixed model analysis was performed on each set of points, with rats within diet-treatment being defined as random. The functions are shown in Figure 10. From the figure, it can be seen that there is a bigger departure between the diets in the duodenum rather than in the colon. The fish oil function has a lower head intensity and a higher tail intensity, which corresponds to a greater proportion of damage. However, there was not a statistically significant diet difference in any of the seven locations.

Similarly, this was done for the no-FPG functions only. The results are also shown in Figure 10. There is less of a difference between the two diets in duodenum for the no-FPG functions than there was for the FPG functions. Again, none of the diet or treatment effects are significant.

Lastly, similar analysis was performed for the points where the FPG difference was taken at the rat level. Here, the only significant difference was between diets at

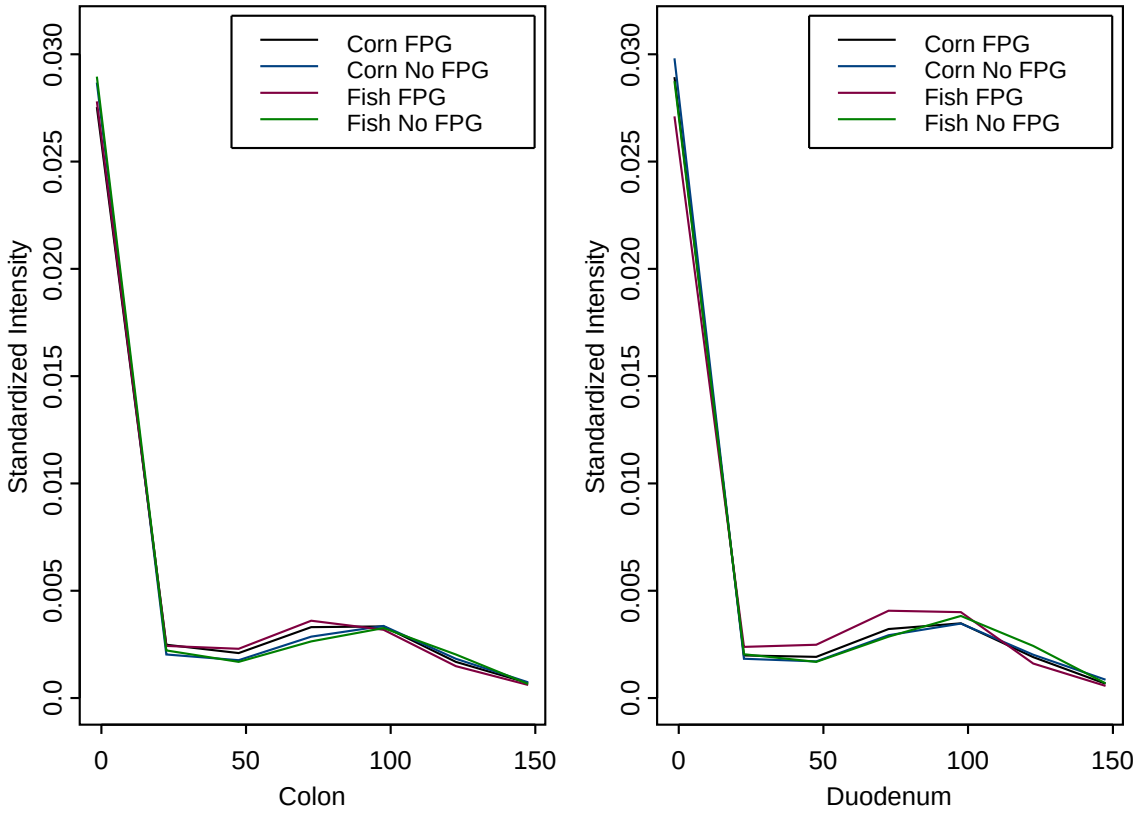

Figure 10. FPG and No-FPG 7-point functions for diet at each location, where the seven points were modeled independently. Black is corn FPG, blue is corn no-FPG, red is fish FPG, and green is fish no-FPG.

lengths 50 and 75 for the duodenum, with fish oil diet showing more oxidative damage than corn oil diet (p-values of 0.0497 and 0.0065, respectively).

This analysis shows that the diet differences are not very big. Any differences that may be found would more likely be in the duodenum, with fish oil diet showing more oxidative damage.

Additional table with p-values and figures for this analysis are in Appendix D.

4.4 Model for Seven Points: Bayesian Analysis with Correlation

In Section 4.3, each of the seven points were modeled independently. In this section, the seven points will be modeled, along with the covariance between them at the comet and the rat level. This is an improved model because the seven points from a single comet should be correlated. If the estimate at point 0, which is the center of the head, is lower, it is expected that at least one of the remaining points is higher since the area under the function is one. The model used for the seven points will be an expansion of the model presented in (3.1).

The following model will use seven points from each comet function. The points will be denoted by Λ_{rlfc}. Refer to Section 3.2 for the definition of notation. For a specific diet d, treatment t, and location ℓ:

$$\begin{aligned} \Lambda_{rlfc} &= \Lambda_{rlf} + \xi_{rlfc}\ , \quad \text{where } \ \xi_{rlfc} \sim \text{Normal}(\mathbf{0}\ ,\ \Sigma_{elf}); \\ \Lambda_{rl} &= \Lambda_{dtl} + \xi_{rl}\ , \quad \text{where } \ \xi_{rl} \sim \text{Normal}(\mathbf{0}\ ,\ \Sigma_{dtl}), \end{aligned} \tag{4.1}$$

where the mean vector Λ_{rlf} at the rat level and error term ξ_{rlf} are vectors of a length of 7. Mean vector Λ_{rl} at the rat-level, mean vector Λ_{dtl} at the diet-treatment level, and error term ξ_{rl} at the rat level have a length of 14, with parameters for no-FPG

and FPG, respectively. The following priors are assumed:

$$\begin{aligned}
\Lambda_{dtl} &\sim \text{Normal}(\Lambda_0\ ,\ \Sigma_0);\\
\Sigma_{dtl} &\sim \text{InverseWishart}(\upsilon_{r0}\ ,\ \Sigma_{r0});\\
\Sigma_{elf} &\sim \text{InverseWishart}(\upsilon_{ef0}\ ,\ \Sigma_{ef0}).
\end{aligned}$$

To compute the likelihood, first note that the sample mean vector at the rat level $\overline{\Lambda}_{rl} \sim \text{Normal}(\Lambda\ ,\ \Sigma_{dtl}+\Sigma_{el})$, where covariance at the cell level Σ_{el} is a 14×14 matrix, with the no-FPG covariance $\Sigma_{el(f=1)}/n_{rl(f=1)}$ and FPG covariance $\Sigma_{el(f=2)}/n_{rl(f=2)}$ in the top-left 7×7 corner and bottom-right 7×7 corner, respectively, and zeros everywhere else. The value of n_{rlf} is the number of observations for rat r in location ℓ with FPG type f. Then the likelihood can be written as:

$$\begin{aligned}
\prod_{\text{All } r} f(\overline{\Lambda}_{rl}|\cdots) &= \prod_{\text{All } r} |2\pi(\Sigma_{dtl}+\Sigma_{el})|^{-1/2}\\
&\qquad \times\exp\{(\overline{\Lambda}_{rl}-\Lambda_{dtl})^{\mathrm{T}}(\Sigma_{dtl}+\Sigma_{el})(\overline{\Lambda}_{rl}-\Lambda_{dtl})/2\}.
\end{aligned}$$

The following are the resulting posteriors:

$$\begin{aligned}
\pi(\Lambda_{dtl}|\cdots) &= \text{Normal}\{(n_{dtl}\Sigma_{dtl}^{-1}+\Sigma_0^{-1})^{-1}(n_{dtl}\Sigma_{dtl}^{-1}\overline{\Lambda}_{dtl}+\Sigma_0^{-1}\Lambda_0)\ ,\\
&\qquad (n_{dtl}\Sigma_{dtl}^{-1}+\Sigma_0^{-1})^{-1}\};\\
\pi(\Lambda_{rlf}|\cdots) &= \text{Normal}\{(n_{rlf}\Sigma_{elf}^{-1}+\Sigma_{dtlf}^{-1})^{-1}(n_{rlf}\Sigma_{elf}^{-1}\overline{\Lambda}_{rlf}+\Sigma_{dtlf}^{-1}\Lambda_{dtlf})\ ,\\
&\qquad (n_{rlf}\Sigma_{elf}^{-1}+\Sigma_{dtlf}^{-1})^{-1}\};
\end{aligned}$$

$$\begin{aligned}
\pi(\Sigma_{dtl}|\cdots) &= \text{InverseWishart}[n_{dtl}+\upsilon_{r0}\ ,\ \{\Sigma_{r0}^{-1}+\sum_{rl\subset dtl}(\Lambda_{rl}-\Lambda_{dtl})(\Lambda_{rl}-\Lambda_{dtl})^{\mathrm{T}}\}^{-1}];\\
\pi(\Sigma_{elf}|\cdots) &= \text{InverseWishart}[n_{rlf}+\upsilon_{ef0}\ ,\\
&\qquad \{\Sigma_{ef0}^{-1}+\sum_{rl\subset dtl}\sum_{c\subset rlf}(\Lambda_{rlfc}-\Lambda_{rlf})(\Lambda_{rlfc}-\Lambda_{rlf})^{\mathrm{T}}\}^{-1}],
\end{aligned}$$

where covariance matrix Σ_{dtlf} is the appropriate upper-left or lower-right 7×7 submatrix of covariance matrix Σ_{dtl} at the rat level, mean vector Λ_{rlf} is the appropriate

upper or lower 7-entry sub-vector of mean vector Λ_{rl} for rat r and location ℓ, mean vector $\overline{\Lambda}_{dtl}$ is the average of all mean vectors Λ_{rl} for all rats r with diet d and treatment t, and mean vector $\overline{\Lambda}_{rlf}$ is the average over all the values Λ_{rlfc} for comet-level results from the same rat r, location ℓ, and FPG type f.

4.5 Modeling the 7-point Correlated Data

The 7-point data will be modeled using (4.1). The LRT and the BF will be used to determine whether rat-level covariance Σ and mean vector Λ depends on the diet or treatment. For the Bayesian analysis, the considered models will be implemented using the posterior mean and median as the variable estimating points. The m.l.e. estimates will be used in computing LRT. All the data points were scaled by 10000.

The prior used for the Bayesian analysis was prior 7: the values for the upper and lower 7 elements of Λ_0 are 250, 25, 25, 25, 25, 25, 25, respectively. The values for Σ_0 are 500 on the diagonal and 0.0 off the diagonal. The value of υ_{r0} is 18, and Σ_{r0}^{-1} has values of 1000, 100, 100, 100, 100, 100, 100, 1000, 100, 100, 100, 100, 100, 100 on the diagonal, while the off-diagonal values were 0.0. The value of υ_{ef0} is 11. Finally, Σ_{ef0} has values 10000, 1000, 1000, 1000, 1000, 1000, 500 on the diagonal, respectively, and 0.0 off the diagonal.

The results are based on 10,000 simulations. Only the last 5,000 were used to compute the posterior mean. Since the posterior median requires sorting, only the last 190 iterations were used to compute the median, which was the largest number that allowed for the execution of the program.

4.5.1 Correlated 7-point Model Results

Results from the simulations for the colon and duodenum are available in Table 10. Models with the mean vector dependent on only diet (Λ_{dl}), dependent on only treat-

Table 10. Model comparison for colon and duodenum 7-point data using prior 7.

Location	Prior Used	Models Compared	2log(BF) (mean)	2log(BF) (median)	LRT	df	LRT p-value
Colon	7	Λ_{dl} vs. Λ_{dtl}	276.47	130.92	1970.00	56	<0.0001
Colon	7	Λ_{tl} vs. Λ_{dtl}	315.93	266.40	2101.74	42	<0.0001
Colon	7	Σ_{dl} vs. Σ_{dtl}	598.28	587.62	10539.10	784	<0.0001
Colon	7	Σ_{tl} vs. Σ_{dtl}	405.05	402.46	10318.88	588	<0.0001
Duo.	7	Λ_{dl} vs. Λ_{dtl}	332.86	192.18	1895.06	56	<0.0001
Duo.	7	Λ_{tl} vs. Λ_{dtl}	330.68	281.45	1941.28	42	<0.0001
Duo.	7	Σ_{dl} vs. Σ_{dtl}	719.66	681.64	10390.48	784	<0.0001
Duo.	7	Σ_{tl} vs. Σ_{dtl}	451.68	420.36	10190.10	588	<0.0001

NOTE: It is tested whether covariance structure is diet and treatment dependent (Σ_{dtl}), diet dependent (Σ_{dl}), or treatment dependent (Σ_{tl}), and whether the means structure is diet and treatment dependent (Λ_{dtl}), diet dependent (Λ_{dl}), or treatment dependent (Λ_{tl}). Computed is the likelihood ratio test (LRT) and the corresponding p-value, as well as 2log(Bayes Factor). Results are given when variables are evaluated at the posterior mean and median.

ment (Λ_{tl}), and diet and treatment (Λ_{dtl}) were fitted, while the covariance matrix Σ_{dtl} used in the model was diet and treatment dependent. The values for the BF using posterior mean and posterior median are of the same magnitude. Determination whether these values are significant is hard to determine since, as was discussed in Chapter III, cut-off points given by Kass and Raftery (1995) do not hold. However, LRT using the m.l.e. estimates has a know asymptotic distribution. The LRT p-values, given in Table 10, indicate that there is a diet and treatment difference in the mean vector. Similarly, models with the covariance matrix dependent on only diet (Σ_{dl}), dependent on only treatment (Σ_{tl}), and diet and treatment (Σ_{dtl}) were fitted, while the mean vector Λ_{dtl} used in the model was diet and treatment dependent. Again, the LRT indicates that there is a diet and treatment differences in the covariance matrix.

In the model, the FPG and no-FPG results were estimated simultaneously. In both the colon and duodenum, significant diet and treatment results were seen for the mean vector (Λ_{dtl}) and the covariance matrix (Σ_{dtl}). In the results presented above, differences between the models were tested for both FPG types simultaneously, but now the diet and treatment differences will be tested separately for each FPG type. This will be done using the LRT only since the cut-off points for BF are not known. Table 11 shows the LRT results, along with the corresponding p-value, for both FPG types. In no-FPG colon, no-FPG duodenum, and FPG duodenum, both the mean vector and the covariance were found to be diet and treatment dependent. In the FPG colon, only the mean vector was diet and treatment dependent. The covariance vector was dependent only on location.

Table 11. Model comparison by FPG type for colon and duodenum 7-point data.

Location	Models Compared	df	LRT (no-FPG)	p-value (no-FPG)	LRT (FPG)	p-value (FPG)
Colon	Λ_l vs. Λ_{dtl}	35	119.78	<0.0001	75.29	0.0001
Colon	Λ_{dl} vs. Λ_{dtl}	28	112.65	<0.0001	69.42	<0.0001
Colon	Λ_{tl} vs. Λ_{dtl}	21	116.47	<0.0001	51.79	0.0002
Colon	Σ_l vs. Σ_{dtl}	245	330.29	0.0002	231.06	0.7296
Colon	Σ_{dl} vs. Σ_{dtl}	196	282.98	<0.0001	198.37	0.4392
Colon	Σ_{tl} vs. Σ_{dtl}	147	237.81	<0.0001	149.63	0.4242
Duodenum	Λ_l vs. Λ_{dtl}	35	71.01	0.0003	83.42	<0.0001
Duodenum	Λ_{dl} vs. Λ_{dtl}	28	66.57	0.0001	68.14	<0.0001
Duodenum	Λ_{tl} vs. Λ_{dtl}	21	58.99	<0.0001	41.43	0.0050
Duodenum	Σ_l vs. Σ_{dtl}	245	289.67	0.0264	299.43	0.0100
Duodenum	Σ_{dl} vs. Σ_{dtl}	196	258.01	0.0020	245.58	0.0093
Duodenum	Σ_{tl} vs. Σ_{dtl}	147	223.93	<0.0001	206.85	0.0008

NOTE: It is tested whether covariance structure is diet and treatment dependent (Σ_{dtl}), diet dependent (Σ_{dl}), treatment dependent (Σ_{tl}), or not (Σ_l), and whether the means structure is diet and treatment dependent (Λ_{dtl}), diet dependent (Λ_{dl}), treatment dependent (Λ_{tl}), or only location dependent (Λ_l). Computed is the likelihood ratio test (LRT) and the corresponding p-value.

4.6 Discussion

Presented in this chapter were additional ways of analyzing FLARE data. A method present in the literature for analyzing this type of data was to summarize each function by the RTM. In the previous chapter, it was shown that when using single values to summarize each function, the diet and treatment differences appeared. However, using seven points to summarize each function, and then modeling that data, proved to be more effective. Diet and treatment differences were found to be much more apparent in the colon and duodenum. Also, the results from this analysis are more easily interpretable when plotted.

In the standard analysis in Chapter II and Section 4.3 on independently modeling the seven points, we only saw diet differences, and only in the duodenum. In the analysis where the points are correlated we see diet and treatment differences both in the colon and duodenum. This may be because modeling the correlation between the seven points is a more powerful approach to modeling the data.

The difference between FPG and no-FPG functions for the fish oil diet are larger than for the corn oil diet. This can be seen from Figures 11 and 12. The Bayesian estimates for the mean vectors Λ_{dtl} that depend on diet and treatment are plotted in Figure 11 (using posterior mean). The m.l.e. estimates are plotted in Figure 12. The differences also appear to be larger in the duodenum than in the colon. The biggest difference between FPG types occurs in the duodenum for the recovery treatment. There, the function for the fish oil diet comets exposed to FPG has a lower peak at zero (center of the head) and an elevated level of damage in the first part of the tail. The lower peak at zero indicates more damage to DNA since when there is less intensity in the head, then there must be more intensity in the tail because the area under the function is one. Therefore, the fish oil diet function in the duodenum and

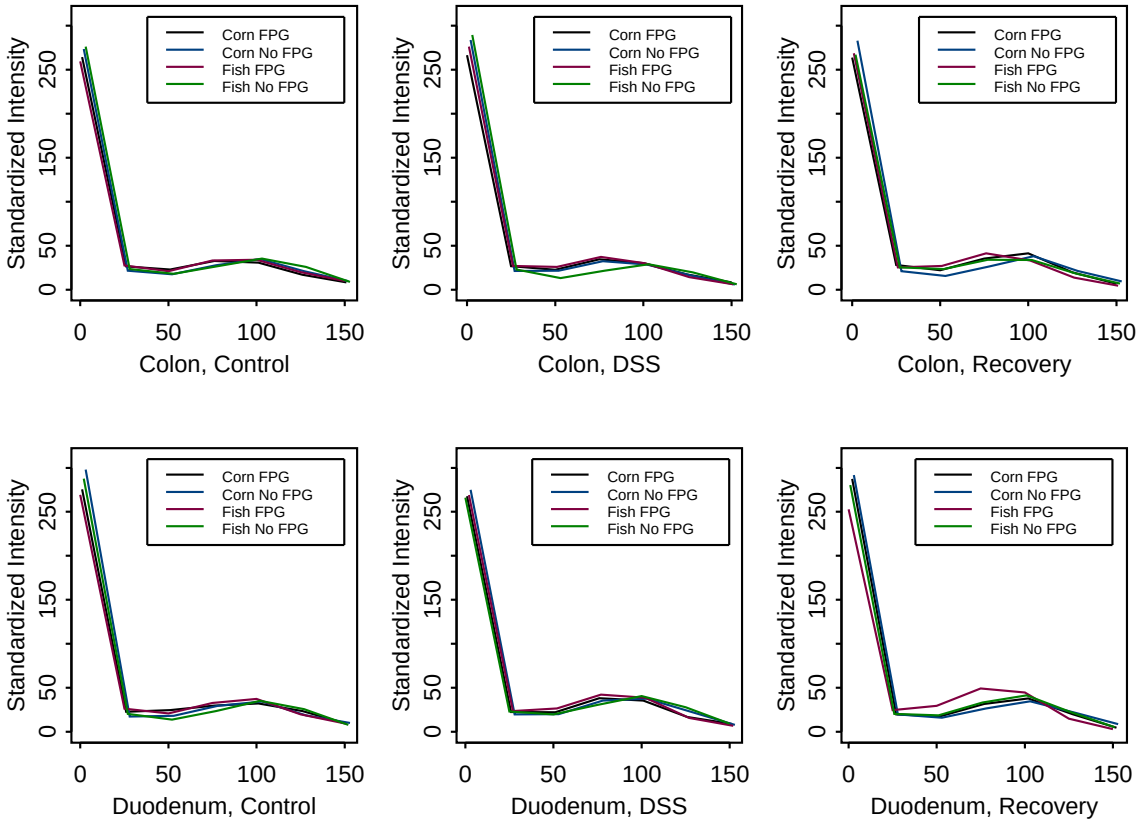

Figure 11. Bayesian FPG and No-FPG 7-point functions for diet at each location and treatment, where functions are modeled accounting for correlation between the seven points. Black is corn FPG, blue is corn no-FPG, red is fish FPG, and green is fish no-FPG.

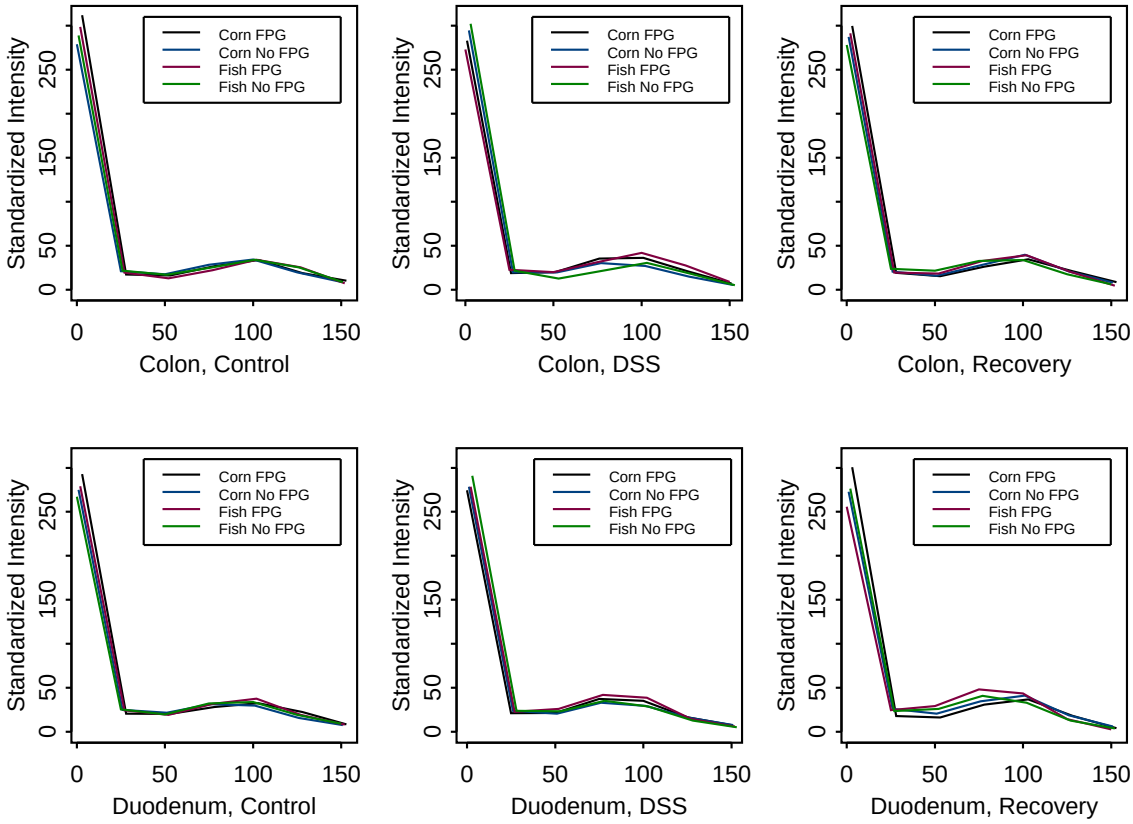

Figure 12. MLE FPG and No-FPG 7-point functions for diet at each location and treatment, where functions are modeled accounting for correlation between the seven points. Black is corn FPG, blue is corn no-FPG, red is fish FPG, and green is fish no-FPG.

for recovery treatment experiences the largest change in shape due to oxidation.

Note that these figures are similar to Figure 9 for the representative functions, where only the mean of the functions was taken without fitting a model. In the figures, the biggest departure from the remaining plotted functions is for the FPG fish oil function. The other function that differs from the rest, but to a lesser extent, is the fish function with no FPG present. This indicates that for all of the figures the corresponding estimated mean vectors are fairly similar.

CHAPTER V

CONCLUSION

This dissertation examined ways of analyzing FLARE assay data. This was first done by summarizing each FLARE function by a single number, and then by using seven points from each function. While diet and treatment differences were detected using RTM and skewness, the differences were more profound when the correlated 7-point model was used. The results from this last analysis were also more easily to interpret when plotted.

For both RTM and skewness differences were detected. For the RTM colon, no diet or treatment differences were found in the mean vector or the covariance matrix. For the RTM duodenum results, while there were no differences in the mean vector due to diet or treatment, differences in the covariance matrix were detected. When skewness was used, differences were found in both the colon and the duodenum. In the colon, differences due to diet and treatment were detected for the means vector, while in the duodenum only diet differences were detected for the means vector.

In the analysis where the points are correlated, we see diet and treatment differences in both the colon and duodenum. These differences are stronger than in the previous analyses, including the analysis where the seven points were modeled independently. This may be because modeling the correlation between the seven points is a more powerful approach to modeling the data. For this analysis, the difference between FPG and no-FPG functions for the fish oil diet is larger than for the corn oil diet. The differences also appear to be larger in the duodenum than in the colon. The biggest difference between FPG types occurs in the duodenum for DSS and recovery treatment. Here, the fish oil diet function experiences the largest change in shape

due to oxidation.

The appropriate model for the data was chosen using the LRT and the BF. To examine distributions of the LRT and the BF, simulated data was used. While it was shown that the LRT follows a χ^2 distribution, the BF distribution does not. Initially, the cut-off points suggested by Kass and Raftery (1995) were intended to be used. However, these were not found to be effective for these higher dimensional models. While their paper suggested that a difference of 10 gave very strong evidence again H_o, in the simulated data results most of the $2log(BF)$ values were greater than 10, even though H_o hypothesis was valid.

From the simulated results, it was also found that the type of prior chosen will influence the cut-off point (the 95% percentile of the distribution). The effect of using the posterior mean or the posterior median was not as influential on the cut-off point as was the prior. Because of the variability in the cut-off point due to the prior for analyzing diet and treatment differences, the BF results were not given as much weight as the LRT results.

REFERENCES

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1994), *Molecular Biology of the Cell* (Third edition), New York: Garland Publishing.

Arber, N., Hibshoosh, H., Yasui, W., Neugut, A. I., Hibshoosh, A., Yao, Y., Sgambato, A., Yamamoto, H., Shapira, I., Rosenman, D., Fabian, I., Weinstein, I. B., Tahara, E., and Holt, P. R. (1999), "Abnormalities in the expression of cell cycle-related proteins in tumors of the small bowel," *Cancer Epidemiology, Biomarkers & Prevention*, 8, 1101-1105.

Bancroft, L. K., Lupton, J. R., Davidson, L. A., Taddeo, S. S., Murphy, M. E., Carroll, R. J., and Chapkin, R. S. (2003), "Dietary fish oil reduces oxidative DNA damage in rat colonocytes," *Free Radical Biology and Medicine*, 35, 149-159.

Bartsch, H., Nair, J., and Owen, R. W. (1999), "Dietary polyunsaturated fatty acids and cancers of the breast and colorectum: emerging evidence for their role as risk modifiers," *Carcinogenesis*, 20, 2209-2218.

Boiteux, S., and Radicella, J. P. (1999), "Base excision repair of 8-hydroxyguanine protects DNA from endogenous oxidative stress," *Biochimie*, 81, 59-67.

Chib, S. (1995), "Marginal likelihood from the Gibbs output," *Journal of the American Statistical Association* , 432, 1313-1321.

Collins, A. R., Ma, A. G., and Duthie, S. J. (1995), "The kinetics of repair of oxidative DNA damage (strand breaks and oxidised pyrimidines) in human cells," *Mutation Research*, 336, 69-77.

Diggle, C. P. (2002), "In vitro studies on the relationship between polyunsaturated fatty acids and cancer: tumor or tissue specific effects?," *Progress in Lipid Research*, 41, 240-253.

Dougherty, E. R. (1992), *An Introduction to Morphological Image Processing*, Bellingham, WA: SPIE Press.

Dougherty, E. R., and Astola, J. (1999), *Nonlinear Image Filtering*, Bellingham, WA: SPIE and IEEE Presses.

Giardina, C. R., and Dougherty, E. R. (1988), *Morphological Methods in Image and Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall.

Halliwell, B. (2002), "Effect of diet on cancer development: is oxidative DNA damage a biomarker?," *Free Radical Biology & Medicine*, 32, 968-974.

Hellman, B., Vaghef, H., and Bostrom, B. (1995), "The concepts of tail moment and tail inertia in the single cell gel electrophoresis assay," *Mutation Research*, 336, 123-31.

Hong, M. Y., Chapkin, R. S., Barhoumi, R., Burghardt, R. C., Turner, N. D., Henderson, C. E., Sanders, L. M., Fan, Y. Y., Davidson, L. A., Murphy, M. E., Spinka, C. M., Carroll, R. J., and Lupton, J. R. (2002), "Fish oil increases mitochondrial phospholipid unsaturation, upregulating reactive oxygen species and apoptosis in rat colonocytes," *Carcinogenesis*, 23, 1919-25.

Jackson, A. L., and Loeb, L. A. (2001), "The contribution of endogenous sources of DNA damage to multiple mutations in cancer," *Mutation Research*, 477, 7-21.

Jemal, A., Thomas, A., Murray, T., and Thun, M. (2002), "Cancer statistics, 2002," *CA A Cancer Journal for Clinicians*, 52, 23-47.

Kass, R. E. and Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association* , 430, 773-795.

Kent, C. R., Eady, J. J., Ross, G. M., and Steel, G.G. (1995), "The comet moment as a measure of DNA damage in the comet assay," *International Journal of Radiation Biology*, 67, 655-660.

Kinzler, K. W., and Vogelstein, B. (1997), *The Genetic Basis of Human Cancer*, New

York: McGraw-Hill Professional.

Loft, S., and Poulsen, H. E. (1996), "Cancer risk and oxidative DNA damage in man," *Journal of Molecular Medicine*, 74, 297-312.

Marx, J. L. (1992), "New colon cancer gene discovered," *Science*, 260, 751-752.

Morris, E. J., Dreixler, J. C., Cheng, K. Y., Wilson, P. M., Gin, R. M., and Geller, H. M. (1999), "Optimization of single-cell gel electrophoresis (SCGE) for quantitative analysis of neuronal DNA damage," *Biotechniques*, 26, 282-9.

Olive, P. L., Durand, R. E., Banath, J. P., and Johnston, P. J. (2001), "Analysis of DNA damage in individual cells" in *Methods in Cell Biology: Flow Cytometry (Third Edition)*, 235-249. eds. Z. Darzinkiewvicz, H.A. Chrissman, and J.P. Robinson, San Diego: Academic Press.

Riso, P., Santangelo, A., and Porrini, M. (1999), "The comet assay for the evaluation of the cell resistance to oxidative stress," *Nutrition Research*, 19, 325-333.

Serra, J. (1982), *Image Analysis and Mathematical Morphology*, New York: Academic Press.

Singh, N. P., McCoy, M. T., Tice, R. R., and Schneider, E. L. (1988), "A simple technique for quantitation of low levels of DNA damage in individual cells," *Experimental Cell Research*, 175, 184-191.

Singh, N. P. (2000), "Microgels for estimation of DNA strand breaks, DNA protein crossslinks and apoptosis," *Mutation Research*, 455, 111-127.

Siskind, V. (1972), "Second moments of inverse Wishart-matrix elements," *Biometrika*, 59, 690-691.

Sugimura, T. (2000), "Nutrition and dietary carcinogens," *Carcinogenesis*, 21, 387-395.

Tardieu, D., Jaeg, J. P., Cadet, J., Embvani, E., Corpet, D. E., and Petit, C. (1998), "Dextran sulfate enhances the level of an oxidative DNA damage biomarker, 8-

oxo-7,8- dihydro-2'-deoxyguanosine, in rat colonic mucosa," *Cancer Letters*, 134, 1-5.

Tardieu, D., Jaeg, J. P., Deloly, A., Corpet, D. E., Cadet, J., and Petit, C. R. (2000), "The COX-2 inhibitor nimesulide suppresses superoxide and 8-hydroxy-deoxyguanosine formation, and stimulates apoptosis in mucosa during early colonic inflammation in rats," *Carcinogenesis*, 21, 973-976.

Wheeler, J. M., Warren, B. F., Mortensen, N. J., Kim, H. C., Boddolph, S. C., Elia, G., Beck, N. E., Williams, G. T., Shepherd, N. A., Bateman, A. C., and Bodmer, W. F. (2002), "An insight into the genetic pathway of adenocarcinoma of the small intestine," *Gut*, 50, 218-223.

APPENDIX A

DEFINITIONS OF DISTRIBUTIONS

The density function for InverseGamma(a, b) used in the dissertation is the following:

$$\mathrm{p}(x) = \frac{1}{b^a \Gamma(a)} x^{-(a+1)} e^{-1/(bx)},\ x > 0.$$

The expected value and the variance of x are

$$\begin{aligned} \mathrm{E}(x) &= \frac{1}{b(a-1)}, \quad \text{for } a > 1; \\ \mathrm{V}(x) &= \frac{1}{b^2(a-1)^2(a-2)}, \quad \text{for } a > 2. \end{aligned}$$

The density function for InverseWishart(υ, Σ), where W is positive definite and has dimension of $k \times k$, used in the dissertation is the following:

$$\begin{aligned} \mathrm{p}(W) = {} & [2^{\upsilon k/2} \pi^{k(k-1)/4} \prod_{i=1}^{k} \Gamma\{(\upsilon + 1 - i)/2\}]^{-1} \\ & \times |\Sigma|^{-\upsilon/2} |W|^{-(\upsilon+k+1)/2} \exp\{-0.5 \mathrm{tr}(\Sigma^{-1} W^{-1})\}. \end{aligned}$$

The expected value of W is

$$\mathrm{E}(W) = (\upsilon - k - 1)^{-1} \Sigma^{-1}.$$

The second moments of inverse Wishart matrix elements, but using the above notation, are (Siskind 1972):

Let t be a constant $p \times 1$ vector and W a $k \times k$ inverse Wishart matrix with $\upsilon > k+3$ degrees of freedom and expectation $(\upsilon - k - 1)^{-1}\Sigma^{-1}$. Then

$$(\upsilon - k)(\upsilon - k - 3)\mathrm{E}(W t t^T W) = \Sigma^{-1} t t^T \Sigma^{-1} + \Sigma^{-1}(t^T \Sigma^{-1} t)/(\upsilon - k - 1).$$

From the above equation, the equation below follows (Siskind 1972):

$$(\upsilon - k)(\upsilon - k - 1)(\upsilon - k - 3)\mathrm{E}(w_{ij} w_{rs}) = (\upsilon - k - 2)\sigma^{ij}\sigma^{rs} + \sigma^{ir}\sigma^{js} + \sigma^{is}\sigma^{jr},$$

with superscripts denoting elements of an inverse matrix.

APPENDIX B

CALCULATION OF POSTERIOR DISTRIBUTIONS

The posteriors given in (3.4) are calculated below for the model presented (3.1), with the priors given in (3.2).

The posterior for mean vector for diet d, treatment t, and location ℓ is

$$\begin{aligned}
\pi(\Lambda_{dtl}|\cdots) &\propto \text{Normal}(\Lambda_{dtl}, \Sigma_{dtl}) \times \text{Normal}(\Lambda_0, \Sigma_0) \\
&\propto \exp\{-0.5\ n_{dtl}(\overline{\Lambda}_{rl} - \Lambda_{dtl})^T \Sigma_{dtl}^{-1}(\overline{\Lambda}_{rl} - \Lambda_{dtl})\} \\
&\qquad \times\exp\{-0.5(\Lambda_{dtl} - \Lambda_0)^T \Sigma_0^{-1}(\Lambda_{dtl} - \Lambda_0)\} \\
&\propto \exp[-0.5\{\ n_{dtl}(\Lambda_{dtl}^T \Sigma_{dtl}^{-1}\Lambda_{dtl} - 2\Lambda_{dtl}^T \Sigma_{dtl}^{-1}\overline{\Lambda}_{rl}) \\
&\qquad +(\Lambda_{dtl}^T \Sigma_0^{-1}\Lambda_{dtl} - 2\Lambda_{dtl}^T \Sigma_0^{-1}\Lambda_0)\}] \\
&\propto \exp[-0.5\{\Lambda_{dtl}^T(n_{dtl}\Sigma_{dtl}^{-1} + \Sigma_0^{-1})\Lambda_{dtl} - 2\Lambda_{dtl}^T(n_{dtl}\Sigma_{dtl}^{-1}\overline{\Lambda}_{rl} + \Sigma_0^{-1}\Lambda_0)\}] \\
&= \text{Normal}\{(n_{dtl}\Sigma_{dtl}^{-1} + \Sigma_0^{-1})^{-1}(n_{dtl}\Sigma_{dtl}^{-1}\overline{\Lambda}_{dtl} + \Sigma_0^{-1}\Lambda_0), \\
&\qquad (n_{dtl}\Sigma_{dtl}^{-1} + \Sigma_0^{-1})^{-1}\}.
\end{aligned}$$

The posterior for mean for rat r, FPG type f, and location ℓ is

$$\begin{aligned}
\pi(\lambda_{rlf}|\cdots) &\propto \text{Normal}(\lambda_{rlf}, \sigma^2_{elf}) \times \text{Normal}(\lambda_{dtlf}, \sigma^2_{dtlf}) \\
&\propto \exp\{-0.5\ n_{rlf}(\overline{\lambda}_{rlf} - \lambda_{rlf})^2/\sigma^2_{elf}\} \times \exp\{-0.5(\lambda_{rlf} - \lambda_{dtlf})^2/\sigma^2_{dtlf}\} \\
&\propto \exp\{-0.5\ n_{rlf}(\lambda^2_{rlf}/\sigma^2_{elf} - 2\lambda_{rlf}\overline{\lambda}_{rlf}/\sigma^2_{elf} \\
&\qquad -0.5(\lambda^2_{rlf}/\sigma^2_{dtlf} - 2\lambda_{rlf}\lambda_{dtlf}/\sigma^2_{dtlf})\} \\
&\propto \exp[-0.5\{\lambda^2_{rlf}(n_{rlf}\sigma^{-1}_{elf} + \sigma^{-1}_{dtlf}) - 2\lambda_{rlf}(n_{rlf}\sigma^{-1}_{elf}\overline{\lambda}_{rlf} + \sigma^{-1}_{dtlf}\Lambda_{dtlf})\}] \\
&\propto \exp\left\{-0.5\left(\lambda^2_{rlf}\frac{\sigma^2_{elf} + n_{rlf}\sigma^2_{dtlf}}{\sigma^2_{elf}\sigma^2_{dtlf}} - 2\lambda_{rlf}\frac{n_{rlf}\overline{\lambda}_{rlf}\sigma^2_{dtlf} + \lambda_{dtlf}\sigma^2_{elf}}{\sigma^2_{elf}\sigma^2_{dtlf}}\right)\right\} \\
&= \text{Normal}\left(\frac{n_{rlf}\sigma^2_{dtlf}\overline{\lambda}_{rlf} + \sigma^2_{elf}\lambda_{dtlf}}{n_{rlf}\sigma^2_{dtlf} + \sigma^2_{elf}},\ \frac{\sigma^2_{dtlf}\sigma^2_{elf}}{n_{rlf}\sigma^2_{dtlf} + \sigma^2_{elf}}\right).
\end{aligned}$$

The posterior for covariance at the rat level is

$$\begin{aligned}
\pi(\Sigma_{dtl}|\cdots) &\propto \text{Normal}(\Lambda_{dtl}, \Sigma_{dtl}) \times \text{InverseWishart}(\upsilon_{r0}, \Sigma_{r0}) \\
&\propto |\Sigma_{dtl}|^{-n_{dtl}/2} \exp\{-0.5 \sum_{r=1}^{n_{dtl}} (\Lambda_{rl} - \Lambda_{dtl})^T \Sigma^{-1}_{dtl}(\Lambda_{rl} - \Lambda_{dtl})\} \\
&\qquad \times |\Sigma_{dtl}|^{-(\upsilon_{r0}+k+1)/2} \exp\{-0.5\ \text{tr}(\Sigma^{-1}_{dtl}\Sigma^{-1}_{r0})\} \\
&\propto |\Sigma_{dtl}|^{-(n_{dtl}+\upsilon_{r0}+k+1)/2} \\
&\qquad \times \exp[-0.5\ \text{tr}\{\Sigma^{-1}_{dtl}\sum_{r=1}^{n_{dtl}}(\Lambda_{rl} - \Lambda_{dtl})(\Lambda_{rl} - \Lambda_{dtl})^T\} - 0.5\ \text{tr}(\Sigma^{-1}_{dtl}\Sigma^{-1}_{r0})] \\
&\propto |\Sigma_{dtl}|^{-(n_{dtl}+\upsilon_{r0}+k+1)/2} \\
&\qquad \times \exp(-0.5\ \text{tr}[\Sigma^{-1}_{dtl}\{\sum_{r=1}^{n_{dtl}}(\Lambda_{rl} - \Lambda_{dtl})(\Lambda_{rl} - \Lambda_{dtl})^T + \Sigma^{-1}_{r0}\}]) \\
&= \text{InverseWishart}[n_{dtl} + \upsilon_{r0},\ \{\Sigma^{-1}_{r0} + \sum_{r=1}^{n_{dtl}}(\Lambda_{rl} - \Lambda_{dtl})(\Lambda_{rl} - \Lambda_{dtl})^{\mathrm{T}}\}^{-1}].
\end{aligned}$$

The posterior for variance at the cell level is

$$\begin{aligned}
\pi(\sigma^2_{elf}|\cdots) &\propto \text{Normal}(\lambda_{rlf}, \sigma^2_{elf}) \times \text{InverseGamma}(a_{f0}, b_{f0}) \\
&\propto (\sigma^2_{elf})^{-n_{rlf}/2} \exp\{-0.5 \sum_{r=1}^{n_{dtl}} \sum_{c=1}^{n_{rlf}} (\lambda_{rlfc} - \lambda_{rlf})^2/(\sigma^2_{elf})\} \\
&\qquad \times (\sigma^2_{elf})^{-(a_{f0}+1)} \exp\{-1/(b_{f0}\sigma^2_{elf})\} \\
&\propto (\sigma^2_{elf})^{-(a_{f0}+n_{rlf}/2+1)} \exp[-\{1/b_{f0} + 0.5 \sum_{r=1}^{n_{dtl}} \sum_{c=1}^{n_{rlf}} (\lambda_{rlfc} - \lambda_{rlf})^2\}/\sigma^2_{elf}] \\
&= \text{InverseGamma}[a_{f0} + \frac{1}{2} n_{rlf},\ \{1/b_{f0} + 0.5 \sum_{r=1}^{n_{dtl}} \sum_{c=1}^{n_{rlf}} (\lambda_{rlfc} - \lambda_{rlf})^2\}^{-1}].
\end{aligned}$$

APPENDIX C

CALCULATION OF MAXIMUM LIKELIHOOD ESTIMATORS

The following two log-likelihoods need to be maximized for the variance σ_{elf} at the cell level, the covariance Σ_{dtl} at the rat level, the diet-treatment mean vector Λ_{dtl}, and the rat-level mean Λ_{rl}. The two likelihoods are:

$$\begin{aligned}
\log \mathrm{L}_1 &= -n_{rc}\log(2\pi\sigma^2_{elf})/2 - \sum_{r=1}^{n_{dtl}}\sum_{c=1}^{n_{rlf}}(\lambda_{rlfc}-\lambda_{rlf})^2/(2\sigma^2_{elf});\\
\log \mathrm{L}_2 &= -\frac{1}{2}\sum_{r=1}^{n_{dtl}}\log|2\pi(\Sigma_{dtl}+\Sigma_{erl})| - \frac{1}{2}\sum_{r=1}^{n_{dtl}}(\overline{\Lambda}_{rl}-\Lambda_{dtl})^{\mathrm{T}}(\Sigma_{dtl}+\Sigma_{erl})^{-1}(\overline{\Lambda}_{rl}-\Lambda_{dtl}),
\end{aligned}$$

where n_{rc} is the total number of observations λ_{rlfc} (for all the rats), and $\overline{\Lambda}_{rl}$ is the no-FPG and FPG mean value of the observations λ_{rlfc} for rat r and location ℓ, respectively. As was defined in Section 3.2, Σ_{erl} is a 2×2 matrix, with $\sigma^2_{el(f=1)}/n_{rl(f=1)}$ and $\sigma^2_{el(f=2)}/n_{rl(f=2)}$ on the diagonal, respectively, and zeros off the diagonal.

The maximization will be done with respect to both of the equations log L_1 and log L_2. First, the m.l.e. estimate is obtained for the rat-level mean value λ_{rlf} for rat r, with location ℓ and FPG type f:

$$\begin{aligned}
\frac{\partial \log \mathrm{L}_1}{\partial \lambda_{rlf}} &= \sum_{c=1}^{n_{rlf}}(\lambda_{rlfc}-\lambda_{rlf})/(\sigma^2_{elf}) = 0;\\
\Rightarrow \widehat{\lambda}_{rlf} &= \frac{1}{n_{rlf}}\sum_{c=1}^{n_{rlf}}\lambda_{rlfc}.
\end{aligned}$$

Next, the m.l.e. estimate is obtained for the variance σ^2_{elf} at the cell level for location ℓ and FPG f:

$$\begin{aligned}
\frac{\partial \log \mathrm{L}_1}{\partial \sigma^2_{elf}} &= -n_{rc}/(2\sigma^2_{elf}) + \sum_{r=1}^{n_{dtl}}\sum_{c=1}^{n_{rlf}}(\lambda_{rlfc}-\lambda_{rlf})^2/(2\sigma^4_{elf}) = 0;\\
\Rightarrow \widehat{\sigma}^2_{elf} &= \frac{1}{n_{rc}}\sum_{r=1}^{n_{dtl}}\sum_{c=1}^{n_{rlf}}(\lambda_{rlfc}-\lambda_{rlf})^2.
\end{aligned}$$

To differentiate log L_2 with respect to the covariance, the following three equations will be used:

$$\begin{aligned}
\sum_i x_i^{\mathrm{T}}\Sigma^{-1}x_i &= \mathrm{tr}(\Sigma^{-1}\sum_i x_i x_i^{\mathrm{T}});\\
\frac{\partial \log|\Sigma^{-1}|}{\partial\Sigma^{-1}} &= 2\Sigma - \mathrm{Diag}\Sigma;\\
\frac{\partial \mathrm{tr}(\Sigma^{-1}S)}{\partial\Sigma^{-1}} &= 2S - \mathrm{Diag}(S).
\end{aligned}$$

Using the first of the three equations, log L_2 can be approximated by:

$$\log\ \mathrm{L}_2 = -\frac{n_{dtl}}{2}\log|2\pi(\Sigma_{dtl}+\widetilde{\Sigma}_{el})| - \frac{1}{2}\mathrm{tr}\{(\Sigma_{dtl}+\widetilde{\Sigma}_{el})^{-1}n_{dtl}S_{dtl}\},$$

where $n_{dtl}S_{dtl}$ is $\sum_{r=1}^{n_{dtl}}(\overline{\Lambda}_{rl}-\Lambda_{dtl})(\overline{\Lambda}_{rl}-\Lambda_{dtl})^{\mathrm{T}}$. The $\widetilde{\Sigma}_{el}$ is 2×2 matrix, with $\sigma^2_{el(f=1)}\sum_{r=1}^{n_{dtl}} n^{-1}_{rl(f=1)}/n_{dtl}$ and $\sigma^2_{el(f=2)}\sum_{r=1}^{n_{dtl}} n^{-1}_{rl(f=1)}/n_{dtl}$ on the diagonal, respectively, and zeros off the diagonal. Note that if the number of observations in each rat was the same, then $\widetilde{\Sigma}_{el}$ would be the same as Σ_{erl}. Then, using the remaining two equations:

$$\begin{aligned}
\frac{\partial \log\ \mathrm{L}_2}{\partial(\Sigma_{dtl}+\widetilde{\Sigma}_{el})^{-1}} &= \frac{n_{dtl}}{2}\{2(\Sigma_{dtl}+\widetilde{\Sigma}_{el}) - \mathrm{Diag}(\Sigma_{dtl}+\widetilde{\Sigma}_{el}) - 2S_{dtl} - \mathrm{Diag}(S_{dtl})\} = 0;\\
\Rightarrow 0 &= 2(\Sigma_{dtl}+\widetilde{\Sigma}_{el}-S_{dtl}) - \mathrm{Diag}(\Sigma_{dtl}+\widetilde{\Sigma}_{el}-S_{dtl});\\
\Rightarrow 0 &= \Sigma_{dtl}+\widetilde{\Sigma}_{el}-S_{dtl};\\
\Rightarrow \widehat{\Sigma}_{dtl} &= S_{dtl}-\widetilde{\Sigma}_{el}.
\end{aligned}$$

The m.l.e. estimate for diet-treatment mean Λ_{dtl} is:

$$\begin{aligned}
\frac{\partial \log\ \mathrm{L}_2}{\partial\Lambda_{dtl}} &= \sum_{r=1}^{n_{dtl}}(\Sigma_{dtl}+\widetilde{\Sigma}_{el})^{-1}(\overline{\Lambda}_{rl}-\Lambda_{dtl}) = 0;\\
\Rightarrow \widehat{\Lambda}_{dtl} &= \frac{1}{n_{dtl}}\sum_{r=1}^{n_{dtl}}\overline{\Lambda}_{rl}.
\end{aligned}$$

APPENDIX D

MULTI VARIABLE MODEL RESULTS

- Table 12 shows the diet effect p-values from independent analysis of each of the seven points. When FPG difference is taken at the rat level, diet differences appear at points 50 and 75. No diet differences occur when only the FPG or the no-FPG values are analyzed separately. No treatment effect was found for FPG, no-FPG, or FPG difference.

- Figure 13 shows the representative functions for each location. Figure 13 was created by averaging over treatment t:

$$f_{dlf}(x_i) = \sum_{r \subset d} f_{dtrlf}(x_i)/n_{dtl}.$$

  In this figure, we again see that there is more departure in the representative functions for comets exposed to the fish oil diet than to the corn oil diet, implying more change due to oxidation for fish oil diet than for the corn oil diet. Biggest departure between the FPG types is seen for the fish oil diet in the duodenum.

- Figure 14 shows the FPG difference representative functions for each location and treatment. The difference between the FPG types was taken at the rat level. This means that once $f_{dtrlf}(x_i)$ was obtained, the difference between $f_{dtrl(f=1)}(x_i)$ and $f_{dtrl(f=2)}(x_i)$ was taken at each x_i to obtain $f_{dtrl}(x_i)$. Then the function was averaged over the rats with the same diet and treatment:

$$f_{dtl}(x_i) = \sum_{r \subset dt} f_{dtrl}(x_i)/n_{dtl}.$$

Here, the sum of all the values in each function is zero. Each representative function is the average difference between the FPG and the no-FPG function. Each box shows the two diet functions for the same location and treatment. Again, we are looking at which of the two functions deviates more from the x-axis. Although it is not as clear to see as in Figures 9 and 13, difference for each diet is usually negative around zero, indicating that there is more damage for the FPG function than the no-FPG function. Also, usually this difference is bigger for the fish oil diet than the corn oil diet. The biggest difference is seen again for comets from the duodenum with the DSS and recovery treatment. There, initially the fish function (red) has a big dip, indicating that the head part of the FPG function has a lower intensity than the no-FPG function. From length 25 to 100, there is an increase in the fish function, indicating that for this part the FPG function has a higher value than the no-FPG function. Since for the fish oil diet the dip around zero is lower than for the corn oil diet, the conclusion here would be the same: fish oil shows more oxidative damage for the duodenum and DSS with recovery.

- Figure 15 shows the FPG difference representative functions for each location. In Figure 15, the functions were averaged over the rats with the same diet:

$$f_{dtl}(x_i) = \sum_{r \subset d} f_{dtrl}(x_i)/n_{dtl}.$$

The fish function has bigger departures from the x-axis than the corn function. This is more apparent in duodenum than the colon.

- Figure 16 shows the results from the independent 7-point analysis for the FPG data only. The functions have 95% simultaneous Bonferroni confidence intervals for each of the seven points. The blue fish function was offset from the black

corn function by 3 units to better show confidence intervals. Bigger difference is seen between the fish and the corn oil diet in the duodenum rather than the colon.

- Figure 17 shows the results from the independent 7-point analysis for the no-FPG data only. The functions have 95% simultaneous Bonferroni confidence intervals for each of the seven points. The blue fish function was offset from the black corn function by 3 units to better show confidence intervals. The fish and the corn oil diet estimates are fairly similar.

- Figure 18 shows the results from the independent 7-point analysis for the FPG difference data. The functions have 95% simultaneous Bonferroni confidence intervals for each of the seven points. The blue fish function was offset from the black corn function by 3 units to better show confidence intervals. Bigger difference between the FPG types is seen for the fish rather than the corn oil diet. These differences in the fish oil diet are more profound in the duodenum rather than the colon.

Table 12. Diet p-values for independent models computed for the 7-point data.

Location	Point	FPG p-value	No-FPG p-value	FPG Difference p-value
Colon	Point 0	0.8049	0.7984	0.9906
Colon	Point 25	0.8443	0.4438	0.3917
Colon	Point 50	0.5298	0.7807	0.2372
Colon	Point 75	0.4684	0.4959	0.0692
Colon	Point 100	0.6290	0.7980	0.8213
Colon	Point 125	0.4248	0.4084	0.1824
Colon	Point 150	0.6002	0.6966	0.9102
Duodenum	Point 0	0.1827	0.3615	0.5085
Duodenum	Point 25	0.1065	0.2580	0.3908
Duodenum	Point 50	0.1773	0.9418	0.0497
Duodenum	Point 75	0.1018	0.8738	0.0065
Duodenum	Point 100	0.2265	0.4458	0.6629
Duodenum	Point 125	0.2955	0.2075	0.0776
Duodenum	Point 150	0.6035	0.4649	0.7711

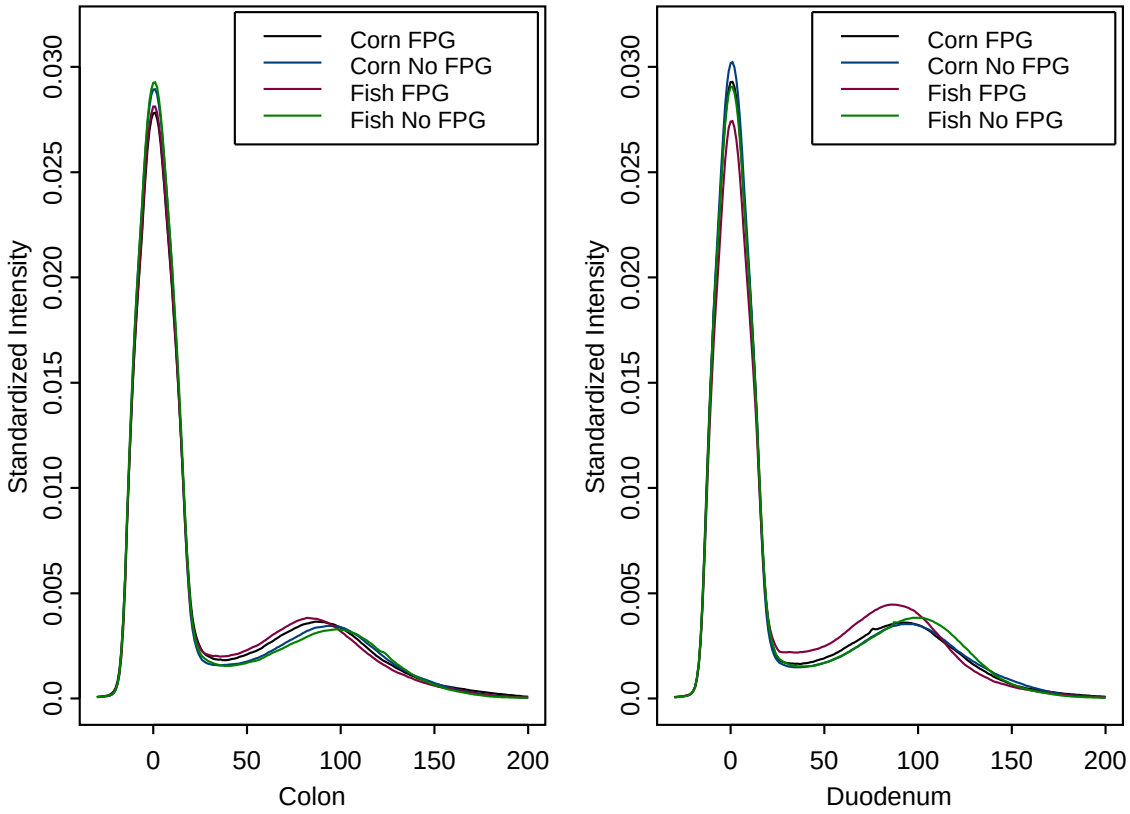

Figure 13. Representative functions for diet and FPG types for each location. Black is corn FPG, blue is corn no-FPG, red is fish FPG, and green is fish no-FPG.

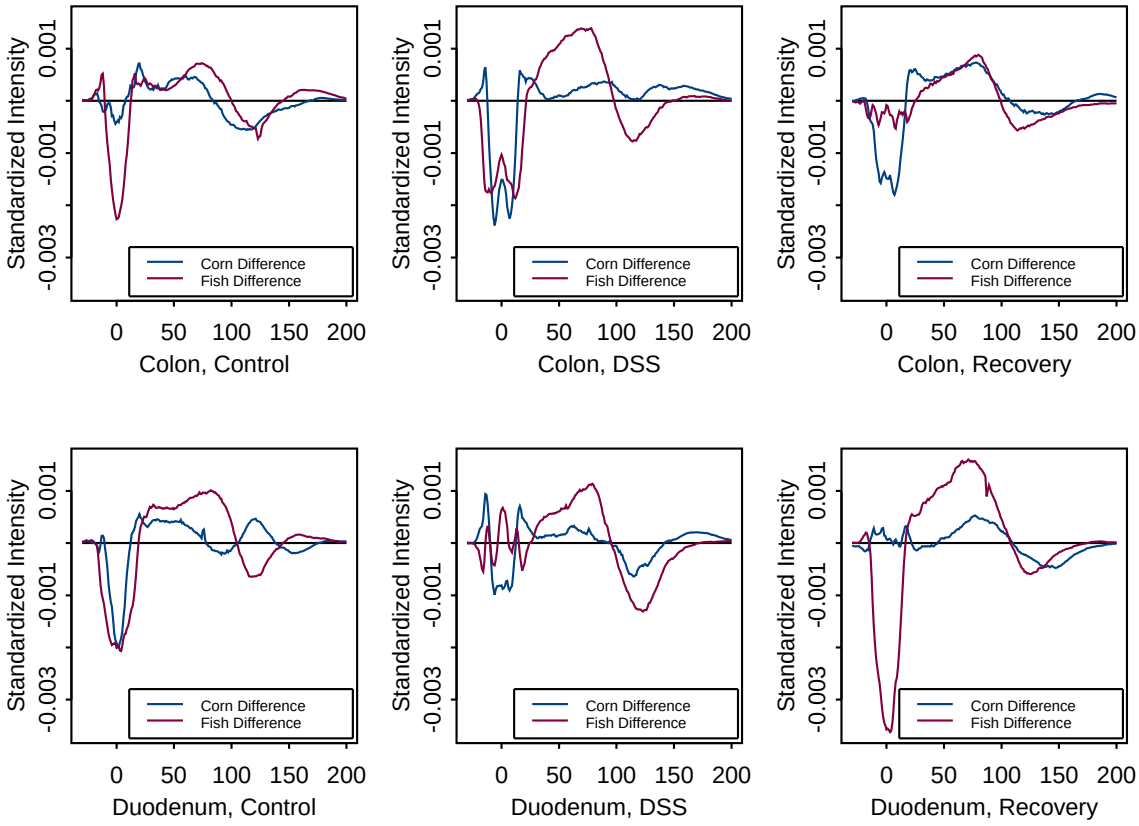

Figure 14. FPG difference representative functions for diet for each location and treatment. Blue is corn difference, and red is fish difference.

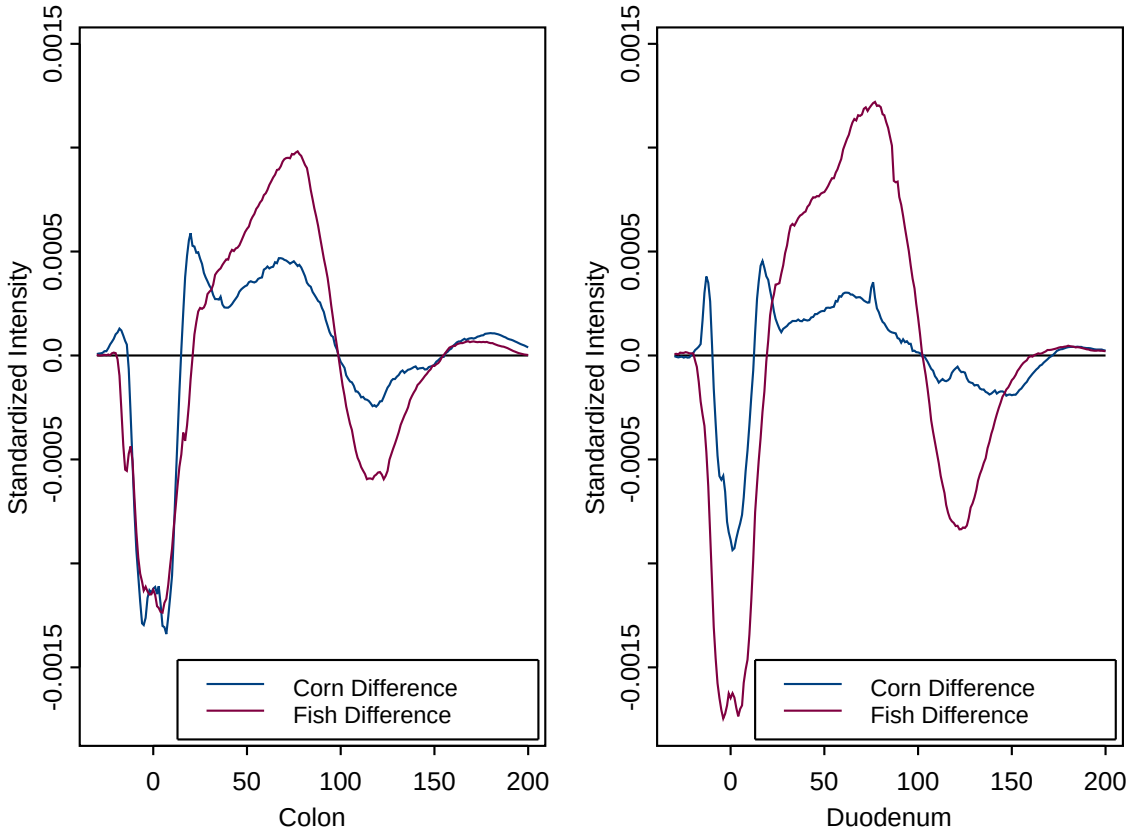

Figure 15. FPG difference representative functions for diet at each location. Blue is corn difference, and red is fish difference.

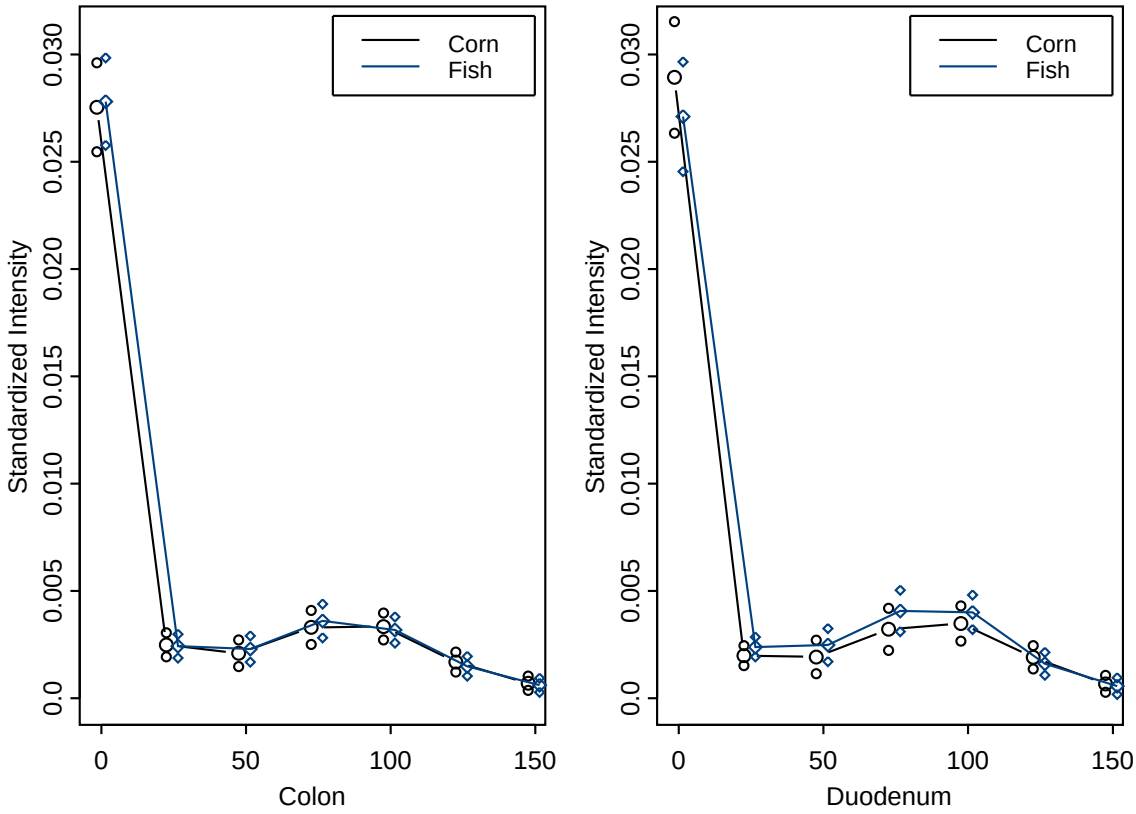

Figure 16. FPG 7-point functions for diet at each location, where the seven points were modeled independently. Black is corn, and blue is fish.

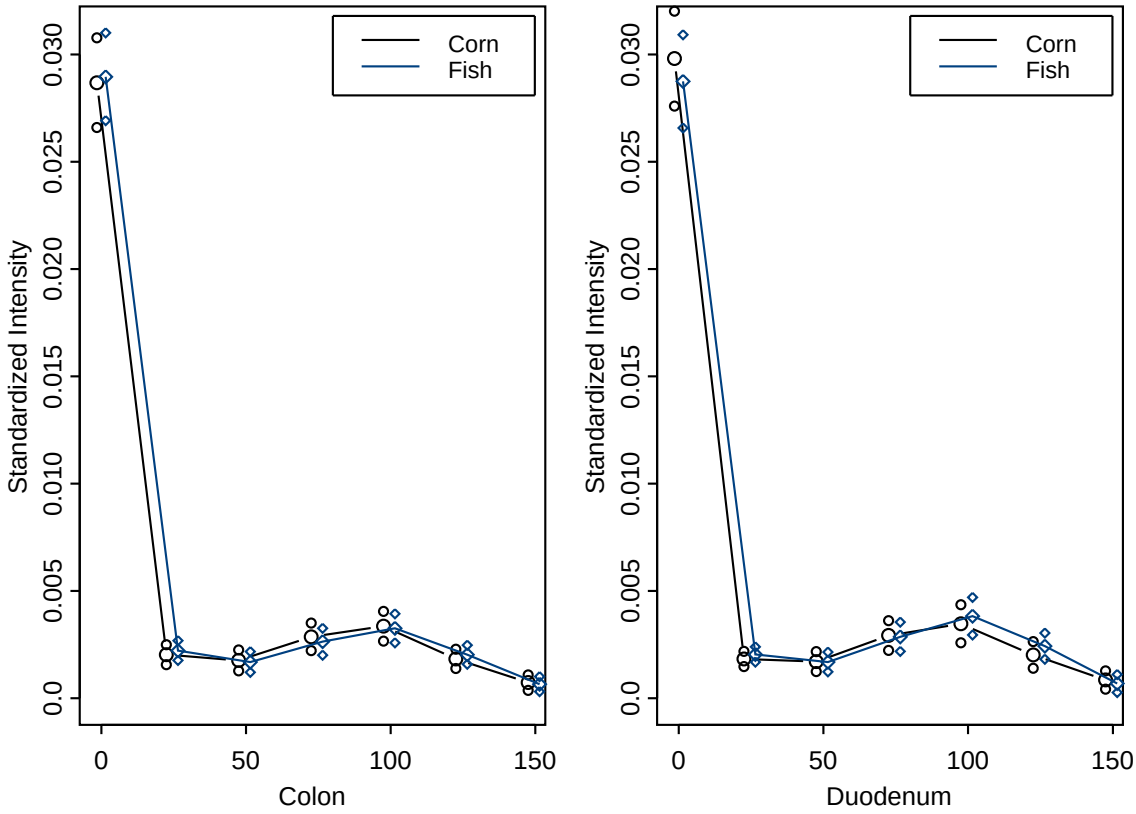

Figure 17. No-FPG 7-point functions for diet at each location, where the seven points were modeled independently. Black is corn, and blue is fish.

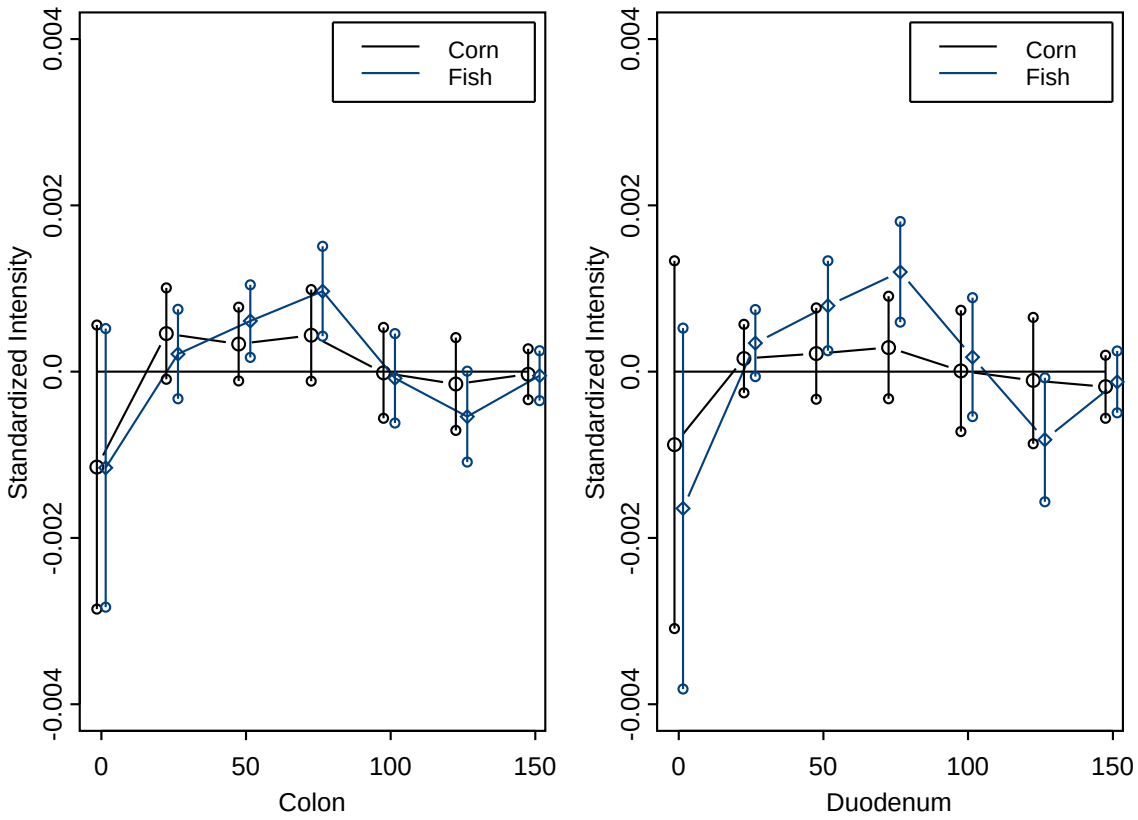

Figure 18. FPG difference 7-point functions for diet at each location, where the seven points were modeled independently. Black is corn, and blue is fish.

VITA

Małgorzata Leyk Williams
Department of Statistics, Texas A&M University
College Station, TX 77843-3143

Degrees

Ph.D. in Statistics, December 2004
Texas A&M University, College Station, Texas

M.S. in Statistics, May 2002
Texas A&M University, College Station, Texas

B.S. in Applied Mathematics, December 1999
Texas A&M University, College Station, Texas

Publications

Leyk, M., Nguyes, D. V., Attoor, S. N., Dougherty, E. R., Turner, N. D., Bancroft, L. K., Chapkin, R. S., Lupton, J. R., and Carroll, R. J. "Comparing Automatic and Manual Image Processing in FLARE Assay Analysis for Colon Carcinogenesis," *Journal of Biomedicine and Biotechnology*, manuscript under review.

Williams, E.T., Leyk, M., Wrighton, S. A., Davies, P. J. A., Loose-Mitchell, D. S., Shipley, G. L., and Strobel, H. W. "Estrogen Regulation of the Cytochrome P450 3A Subfamily in Humans," *Journal of Pharmacology and Experimental Therapeutics*, manuscript in press.

Phillips, C.D., Hawes, C., and Leyk Williams, M. (2003), *Nursing Homes in Rural and Urban Areas, 2000,* College Station, TX: Texas A&M University System Health Science Center, School of Rural Public Health, Southwest Rural Health Research Center.

Experience

Research Assistant, Texas A&M University — Spring 2002 - Spring 2004
Assessed quality of life and care in nursing homes for rural and urban areas; Performed data preparation and data analysis using SAS and SUDAAN.

Teaching, Texas A&M University — Fall 2000 - Fall 2001
Taught statistics classes aimed at students in biological, agricultural, and social sciences.

Teaching Assistant, Texas A&M University — Spring 1998 - Fall 1999
Aided students from calculus classes understand the course work.