Characterization of the Mechanosensitivity of Tactile Receptors using Multivariate Logistical Regression

By
Sam Bradshaw

A Thesis
Submitted to the Faculty
of the
WORCESTER POLYTECHNIC INSTITUTE
In partial fulfillment of the requirements for the
Degree of Master of Science
In
Electrical Engineering
By

_____
May, 2001

Approved:

_____
Dr. Fred J. Looft
Thesis Advisor
Dept. of Electrical and Computer Engineering

_____
Dr. Sean S. Kohles
Committee Member
Dept. of Biomedical Engineering

_____
Dr. Peter Grigg
Committee Member
Dept. of Physiology
University of Mass. Medical Center

_____
Dr. Edward A. Clancy
Committee Member
Dept. of Electrical and Computer Engineering

Abstract

Tactile sensation is a complex manifestation of mechanical stimuli applied to the skin. At the most fundamental level of the somatosensory system is the cutaneous mechanoreceptor, making it the logical starting point in the bottom-up approach to understanding the somatosensory system and sensation, in general. Unfortunately, a consensus has not been reached in terms of the afferent behavior of mechanoreceptors subjected to compressive stimulation.

In this study, 10 afferent mechanoreceptors were isolated and mechanically stimulated with controlled compressive loads. Their responses were recorded and the sensitivities of the individual receptors to compressive stimulation were statistically evaluated by correlating the compressive state of the skin to the observed "all-or-nothing" responses. A host of linear techniques have been employed previously to describe this multiple-input, binary-output system; however, each of these techniques has associated shortcomings when employed in this context. In particular, two shortcomings are the assumption of linear system input-output and the inability of the model to assess individual input-output associations relative to concurrent input in a multivariate context with interacting input. Therefore, a non-linear regression technique called logistical regression was selected for characterizing the mechanoreceptor system. From this model, the relative contributions that each component of the stimulus has upon the neural response of the receptor can be quantitatively assessed and extrapolated to the greater population of cutaneous mechanoreceptors.

Since this study represents a novel approach to receptor characterization, a framework for the application of logistical regression to the time-series representation of the multiple-input, binary-output mechanoreceptor system was established and validated. Subsequently, in-vitro experiments were performed in which the afferent behavior of tactile receptors in rat hairy skin were recorded and the relative association between a number of biologically meaningful stimulus metrics and the observed neural response was evaluated for each receptor. Through the application of logistical regression, it was determined that cutaneous mechanoreceptors are preferentially sensitive to the rate of change of compressive stress when force-control stimulated and both stress and its rate of change when position-control stimulated.

Preface

This thesis represents the culmination of my studies at Worcester Polytechnic Institute. A number of people have helped make this possible and I would like to take a minute to recognize their contributions to this thesis and to my development as a student.

I am very grateful to Professor Fred Looft of the Electrical Engineering Department at WPI. It was only through his hard work and dedication in seeking funding that I was given the opportunity to work on such a fascination project. As my advisor, he challenged me to extend my capabilities by fostering an extremely stimulating, supportive working environment. More importantly, he is a good friend.

I would also like to thank Professor Sean Kohles of the Biomedical and Mechanical Engineering Departments at WPI, Professor Peter Grigg of the Physiology Department at the University of Massachusetts Medical School, Professor Ted Clancy of the Electrical Engineering Department at WPI, and Professor Len Polizzotto of the Electrical Engineering Department at WPI. Each has been instrumental in the success of this work. Professor Kohles helped me to understand the mechanics behind the skin model used in my experiments. Professor Grigg introduced me to logistical regression and provided the facilities in which the experiments were conducted. Professor Clancy provided insight into the delicacies of regression modeling and statistical analysis. Professor Polizzotto helped me to understand the psychosomatic issues inherent in sensation and perception.

Finally, I wish to acknowledge the National Science Foundation for the grant that was awarded to Professor Looft (IBN-9800116). It is through their munificence that this type of research continues.

Table of Contents

List of Figures and Tables

1.0 Introduction

Tactile sensation is a complex manifestation of mechanical stimuli applied to the skin. At the most fundamental level of the somatosensory system is the cutaneous mechanoreceptor. Thus, the bottom-up approach to understanding the somatosensory system and sensation, in general, begins with a characterization of tactile receptors. While tactile receptors have been studied extensively, a consensus has not been reached in terms of a robust model for describing and predicting the afferent behavior of tactile receptors subjected to compressive stimulation [47]. It is widely accepted that these receptors respond readily to pressure and low-frequency vibratory stimuli and studies involving populations of cutaneous receptors typically exploit this fact [17,18,21,29,36–39]. This study exploits the known sensitivities of tactile mechanoreceptors as well.

Tactile skin mechanoreceptors transduce externally-applied mechanical stimuli into a binary action potential event stream. The input stimuli can be described by multiple components, each quantified by a different metric. In order to determine the relationship between the response of the receptor and each component of the stimulus, researchers will systematically vary each component and look for changes in the response of the receptor [36-38,43,44,47,48]. In this type of experiment, it is difficult to evaluate the relationship between components of the stimulus to the observed nerve response for two reasons. First, the stimulus component controlled in the experiment may influence other components of the stimulus such that there is no clear single-input/single-output relationship. Second, the input-output behavior of the system is confounded when interactions of stimulus components are encoded by the receptor.

Multivariate logistical regression (MLR) is a mathematical technique in which multiple continuous explanatory variables, or covariates, are correlated to a dichotomous response variable. The technique was first employed in epidemiological studies to determine which habits and predispositions best predicted a particular disorder such as heart disease or kidney failure [5,17,51]. The explanatory variables could be, for example, age, weight, quantifiable eating habits, and disorder history. The response variable would be the presence or absence of the disorder, expressed as a binary output. MLR provides a quantification of the relative contribution that each explanatory variable has upon the binary response. The capabilities of MLR are realized without any predisposition towards known input/output correlations and MLR has the ability to account for confounded interactions among explanatory variables (such as height and weight in growing individuals).

Multivariate Logistical Regression Applied to Neural Systems

In this study, a multivariate logistical model was used to characterize the in-vitro behavior of isolated rapidly adapting mechanoreceptor afferents in rat hairy skin. The explanatory variables were controlled, time-sampled compressive stimulus states expressed as stress, strain, and their respective time-derivatives. The binary response variable was the presence or absence of a nerve action potential.

Since tactile receptors respond to a plurality of local stimuli, it is difficult to ascertain with certainty whether a mechanoreceptor responds more readily to internal strains or to external compressive stresses when stimulated via force or displacement applied orthogonally to the skin surface. The multivariate logistical model provides an elegant solution to this problem in that over a uniform distribution of stimulus states, the method can be used to quantify the relative spike output contributions from each of the components of the stimulus, including any stimulus state interactions. As a result, one can gain a more rigorous understanding of the incipient behavior of the system by analyzing the influence that interaction terms in the multivariate logistical model have upon the nerve response.

2.0 Research Summary

The goal of this research was to gain a more fundamental understanding of how tactile stimuli are transduced by a mechanoreceptor into a neural response. More specifically, the goal was to determine which stimulus metric or interaction of stimulus metrics is encoded by a mechanoreceptor afferent through the application of multivariate logistic regression.

This research involved stimulating an isolated, in-vitro patch of hairy skin from the inner leg of rat using indenting pseudorandom and non-repeating noise sequences as input, recording both the controlled input variables and corresponding nerve responses, and applying the multivariate logistical regression technique to the sampled data using commercial and custom designed software. Indentation stimuli were applied using an actuator to indent a tip of known dimensions into the surface of the skin directly above an isolated afferent receptor. Stimulus magnitude was varied from threshold to 8 times the receptor threshold with the actuator controlling either the indented position or the applied force. The position (displacement) and force data were then transformed offline into stress, strain, and their time derivatives.

After data were collected, MLR was performed on each data set and odds ratios calculated for each explanatory variable and interaction term in the regression model using a maximum likelihood estimation technique. The odds ratio is a quantitative measure of the degree to which the outcome variable changes for a single-unit change in a covariate. When the explanatory variables are normalized, the odds ratio can be interpreted as a dimensionless quantitative measure of the strength of association between the explanatory variable and the response relative, to all other explanatory variables included in the model. Therefore, the odds ratio was used to compare stimulus encoding characteristics among mechanoreceptors activated by dynamic compressive loads of varying magnitudes.

Two response characteristics, unique to neural encoders, were addressed during data processing. First, an evoked action potential may correspond to a stimulus state that occurred before the action potential was recorded. To address this problem, odds ratios were calculated for explanatory variable combinations that occurred up to 100 milliseconds before the observed action potential. This approach effectively decoupled the stimulus variables across samples, ensuring independence among explanatory variables that had been confounded in time by discrete sampling.

Second, nerve responses are followed by an absolute and relative refractory period, during which no magnitude of stimuli and elevated stimuli can elicit an action potential, respectively. In order to break this cross-sample dependency, post-spike sample points were assigned positive (binary 1) outcomes and added to the model as categorical (discrete) explanatory variables. An odds ratio was calculated for each and post-spike samples were discarded if they were calculated to be protective against spikes (i.e., an odds ratio less than 1). The regression coefficients and odds ratios were then recomputed for each explanatory variable.

3.0 Background

This *Background* section will serve to supplement the reader's knowledge of the afferent behavior of tactile receptors as well as introduce the concept of multivariate logistical regression.

3.1 Mechanoreceptors

Mechanoreceptors transduce static and dynamic mechanical deformations resulting from externally applied vibration, stretching, and pressure into action potential event streams. Mechanoreceptors found in the skin are known as tactile receptors and mediate cutaneous sensation. The nerve endings associated with cutaneous sensation are known as Meissner's corpuscles, Merkel cells, Pacinian corpuscles, and Ruffini corpuscles.

Mechanoreceptors found in muscles, joints, and tendons mediate the kinesthetic sensation by transmitting information on active muscle contraction, passive stretch of muscles, and actively or passively produced tension. These receptors are known as muscle spindles, Golgi tendon organs, and joint receptors.

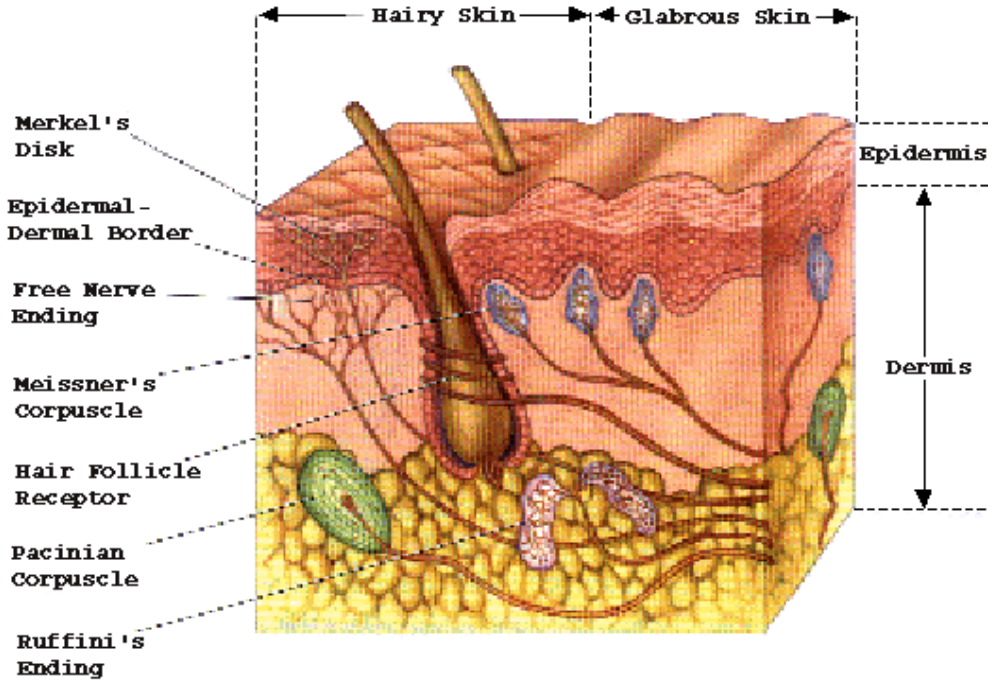


FIGURE 3.1: SKIN CROSS-SECTION WITH CUTANEOUS RECEPTORS

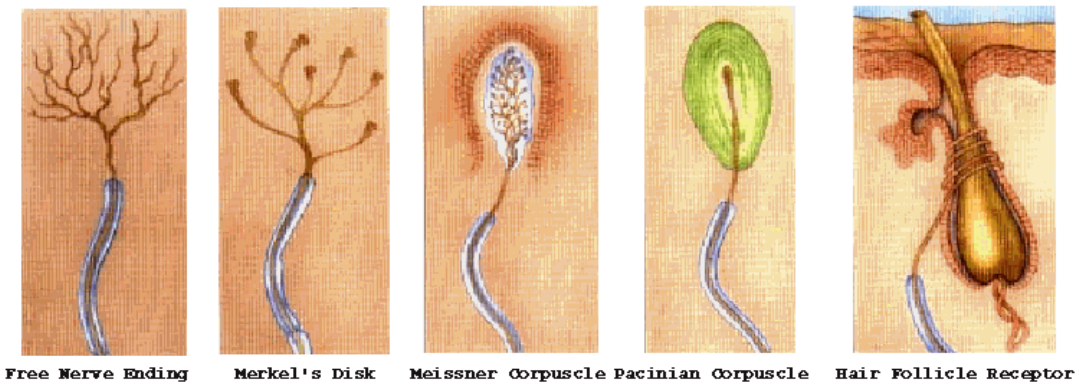


FIGURE 3.2: CUTANEOUS RECEPTOR MORPHOLOGY

3.1.1 Cutaneous Mechanoreceptor Behavior

Both the intensity and duration of the stimulation is encoded by the frequency of action potentials. However, mechanoreceptors exhibit adaptation when subjected to prolonged stimulation. This adaptation is either rapid or slow and each type of adaptation encodes different information [8,36,48].

Rapidly Adapting (RA) receptors encode rapid mechanical changes such as those produced by vibration. RA receptors fire at the onset and offset of a stimulus at or above the nerve's threshold and are generally quiet in between. RA afferents may also respond with a high frequency burst of action potentials at the onset of the stimulus where the burst rate is a function of the magnitude of the initial stimulation. The more intense or rapid the deformation of a single corpuscle, the higher the burst rate of nerve impulses generated in its neuron. If the stimulation is continuous but static in magnitude, an RA afferent will adapt, resulting in a decrease in the frequency of action potentials until none are observed. At this point, any change in the stimulus from this static state will again prompt a volley of action potentials, followed by the same pattern. Rapidly adapting afferents, therefore, tend to encode stimulus rate. An example of an RA tactile afferent is the Meissner corpuscle [8,36,48].

Pacinian corpuscles are rapidly adapting receptors that encode much higher frequency stimuli. Pacinian corpuscles are preferentially sensitive to vibrations up to approximately 400 Hz. They also have a much larger receptive field, resulting in fewer receptors per unit area in the skin, and are often found much deeper in the skin and subcutaneous tissues than other cutaneous receptors. Due to the large size of Pacinian corpuscles relative to other cutaneous receptors (~1 millimeter vs. 5 – 100 micrometers in diameter, respectively), the morphology of Pacinian corpuscles has been documented more thoroughly [8,36,48].

Slowly adapting (SA) receptors respond to slowly changing or static stimuli and can convey stimulus magnitude and duration more accurately than rapidly adapting receptors. To various degrees, SA afferents fire continuously for the duration of the above-threshold

stimulus. Rather than responding with a volley of impulses of decreasing frequency with time, slowly adapting afferents will respond to a constant compressive load with periodic impulses of low and relatively constant frequency. The adaptation time constant is much larger than that observed for RA afferents. Merkel cells and Ruffini corpuscles are slowly adapting and are referred to as slowly adapting type I (SAI) and type II (SAII), respectively [8,36,48].

Mechanoreceptor types can be further subdivided based on their receptive field size relative to one another, as shown in Table 3.1. Meissner corpuscles, Ruffini corpuscles, and Merkel cells typically resolve fine spatial differences, while the Pacinian corpuscles resolve coarse spatial differences. Pacinian corpuscles are sensitive to higher frequency stimulation than other cutaneous afferents and have the largest receptive field sizes.

	Adaptation	Receptive Field Size
Meissner's Corpuscle	Rapid	Small
Merkel Cells	Slow	Small
Pacinian Corpuscle	Rapid	Large
Ruffini's Corpuscle	Slow	Small

TABLE 3.1: RECEPTOR CHARACTERISTICS

3.2 The Linear Probability Model of Binary Data

When modeling a system with continuous output, common regression techniques attempt to provide a model that predicts the best estimate of the value of the continuous output. When dealing with binary output bounded by 0 and 1, on the other hand, regression techniques attempt to describe the conditional mean of a response variable, or the mean value of the response given the state of the input variables [2,34,40,41].

Linear regression is perhaps the most ubiquitous of data modeling techniques and involves fitting a line to a set of data points using the method of least. The equation for the first-order linear probability model for binary data with a single input variable takes the form:

$$Y = \beta_0 + \beta_1 x + \varepsilon_x \qquad \text{[Eq. 1]}$$

where:

- Y represents the dependent variable, in this case a binary (0 or 1) outcome
- β_1 represents the coefficient associated with the independent variable x
- x represents the value of a single independent variable or the value of a first order transformation of an independent variable
- β_0 represents the constant term coefficient
- ε_x represents the error in the model prediction

Throughout this document, the terms dependent variable, outcome, response, response variable, output, and risk factor will all refer to the dependent variable. Similarly, the terms independent variable, covariate, explanatory variable, input, and predictor will all refer to the independent variable(s).

The technique for determining the coefficients β_0 and β_1 is straightforward and resolves into selecting β_0 and β_1 to minimize ε_x where ε_x is the squared difference between the independent variable x and the value of Y predicted by the model. In other words, ε_x is the minimum chi-squared error of the estimate. When the coefficients are resolved, they can be directly interpreted as the probability of the given coefficient generating a binary 1 response [40,41].

While this model may correctly indicate the significance of the independent variable(s) based on the values of the coefficients in the fitted model, it violates several fundamental assumptions for statistical models because [2,7,12,13,22,28,30,34,41,45,53]:

- **The variance of the error term is dependent on the independent variable**. In other words, the error is not constant for every level of the independent variable. This is referred to as "heteroscedasticity."
- **The error is not normally distributed**. As a result, standard errors are inconsistent.
- **The dependent variable is binomially distributed**. Linear regression assumes a normally distributed dependent variable.
- **The predicted probability (expected value) of the dependent variable is not bounded by 0 and 1**. Mathematically, it can take on any value between -∞ and +∞. This result is inconsistent with the definition of probability.

For these and other reasons that will become apparent below, a logistic model is used instead of a linear model for predicting binary outcomes.

3.3 The Logistic Model of Binary Data

In the logistic model, the probability of the dependent variable, or the conditional mean of the system, is modeled by a logistic function. The logistic function, which will henceforth be denoted by $\pi(x)$, represents the probability of observing a positive outcome given the system inputs: Prob($Y = 1 \mid x$). It takes on the following form for the univariate case [13,22,25,30,31,33,34,41,45]:

$$\pi(x) = Prob(Y = 1 \mid x) = \frac{e^{\beta_0 + x\beta_1 + \varepsilon_x}}{1 + e^{\beta_0 + x\beta_1 + \varepsilon_x}} \qquad \text{[Eq. 2]}$$

where:

- Y represents the dependent variable, in this case a binary (0 or 1) outcome
- β_1 represents the coefficient associated with the independent variable x
- x represents the single independent variable, in this case a continuous variable
- β_0 represents the constant term coefficient
- ε_x represents the error associated with the independent variable x

The formula $\pi(x)$ is also described as the conditional mean of the dependent variable given x, or the expected value of the dependent variable, given x: ($E[Y|x]$). The assumptions inherent in the model will become apparent as this section progresses.

An example of a logistic function is shown in Figure 3.3. It is a sigmoid, or S-shaped, curve that looks similar to a *C*umulative *D*istribution *F*unction (CDF). In fact, the logistic function is often used as a CDF [3]. It should also be noted that the independent variable could be continuous or categorical (discrete) where an example of a categorical value is age or the year. The method for determining the coefficients β_0 and β_1 will be discussed in a subsequent section.

The logistic model is a non-linear transformation of the model used in linear regression. Because of this transform, coefficients can no longer be directly interpreted as probabilities as they were in linear regression [41]. This transform, called the "logit" or "log odds" transform, produces a linear combination of the independent variable(s) (Equation 3) and is significant because it represents the natural log of the probability that an event occurs ($P = \pi(x)$) divided by the probability that the event does not occur ($Q = 1 - P = 1 - \pi(x)$). The quantity ($P / 1 - P$) represents the "odds" of observing a particular outcome. The logit transform is the natural log of this value. Since the probability of an event occurring is modeled by a logistic function, the logit transform produces a linear combination of the explanatory variables [13,22,25,30,31,33,34,41,45]:

$$\text{logit}[\pi(x)] = \text{logit}[\text{Prob}(Y=1 \mid x)] = \ln\left[\frac{P}{Q}\right] = \ln\left[\frac{P}{1-P}\right] = \ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + x\beta_1 \quad \text{[Eq. 3]}$$

If, for example, the probability of a particular horse winning a given race were 20% (P = 0.20), that horse would be expected to win 1 race and lose 4 in every 5 races. In other words, the odds are 1 to 4, or 0.25, in favor of this horse winning a given race. Given this convention, it is clear that the lower bound on the odds is zero and the upper bound is infinite. The importance of these bounds on "odds" will become apparent shortly.

The logit transform is particularly useful as a basis for characterizing binary responses because a simple transformation of the probability of a positive response produces a linear function, making the task of regression and evaluating the "goodness of fit" considerably easier [25,26,30]. It is important to understand that the logit, the probability, and the odds are different methods for expressing the same function. The interpretation of probability and odds is more straightforward that the interpretation of the logit. However, the logit transformation removes the shortcomings associated with linear

regression for predicting probabilities and is a mathematically simpler form for expressing the relationship between a dichotomous value and continuous values. Since a logistic function is bounded by 0 and 1, it does not predict probabilities greater than 1 or less than 0 as shown in Figure 3.4 [7,12,41].

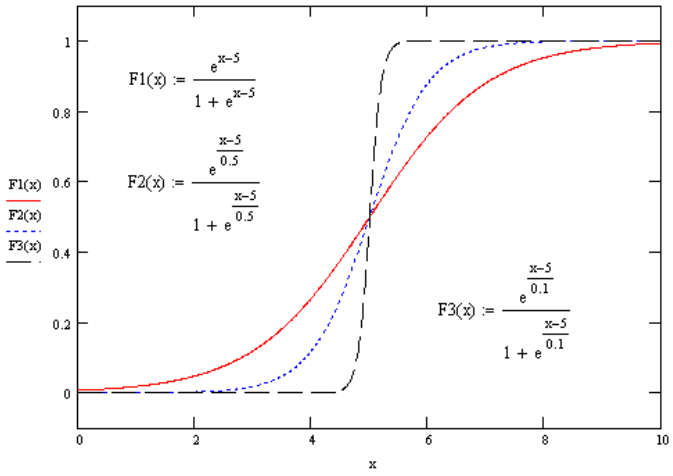


FIGURE 3.3: LOGISTIC FUNCTIONS

One can see from this graph that the S-shaped Logistic function can take on a variety of shapes ranging from continuous and predominantly linear to a step function.

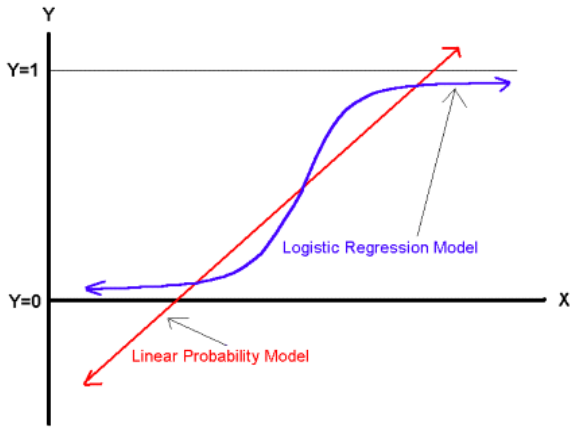


FIGURE 3.4: COMPARISON OF LINEAR PROBABILITY AND LOGISTIC PROBABILITY MODELS

The probability of the Logistic regression model is inherently bounded by 0 and 1; however, the linear probability model can predict response/event probabilities greater than 1 and less than 0.

To clarify, the odds of having a positive response are expressed as the ratio P/Q or $(Prob[Y = 1 \mid x])/(Prob[Y = 0 \mid x])$. In other words [13,22,25,30,31,33,34,41,45]:

$$\frac{\mathrm{Prob}\left[\mathrm{Y}=1|\mathrm{x}\right]}{\mathrm{Prob}\left[\mathrm{Y}=0|\mathrm{x}\right]} = \frac{\pi(\mathrm{x})}{1-\pi(\mathrm{x})} = \mathrm{e}^{\beta_0+\mathrm{x}\beta_1} \qquad \text{[Eq. 4]}$$

The logit transform of the conventional logistic function $\left[\frac{1}{1-\mathrm{e}^{z}}\right]$ would produce an undefined result, $ln(-e^{-z})$. Therefore, the model logistic function is modified such that the logit transform produces a linear result [17,25,31]. This will facilitate the task of model fitting.

3.3.1 The Multivariate Logistical Model

The logistical model can be extended into a multivariate case in which there are multiple independent, or explanatory, variables in the model. These variables are collectively referred to as "covariates." In the multivariate case, the natural log of Prob($Y = 1|x_n$) is a linear combination of "n" known independent variables ($x_{1\ldots n}$) multiplied by unknown parameters ($\beta_{1\ldots n}$), each representing an single covariate variable. In other words [2,17,22,25,31,41]:

$$logit(\pi(x_n)) = \ln\left[\frac{\pi(\mathrm{x}_n)}{1-\pi(\mathrm{x}_n)}\right] = \beta_0 + x_1\beta_1 + x_2\beta_2 + x_3\beta_3 + \ldots + x_n\beta_n \qquad \text{[Eq. 5]}$$

or

$$logit(\pi(x_n)) = ln[Prob(Y = 1|x_n)] = \sum_{i=1}^{n} x_i\beta_i = \boldsymbol{x\beta} \qquad \text{[Eq. 6]}$$

where:

- "n" is the number of independent variables included in the model
- x_n is a vector containing all of the input variables in the model
- $\boldsymbol{x\beta}$ is the linear combination of independent variables multiplied by their respective Beta coefficients

Since there are multiple independent variables in the multivariate model, additional covariates can be included that account for interactions among the independent variables. The interacting covariates are said to be "confounded". Confounded interactions occur when one or more covariates modify the affect that another covariate has upon the response variable [7,12,15,16,17,25,34,40,41]. Confounded variables are said to be associated both with the outcome variable and an independent variable.

The logistical model must be statistically adjusted to account for confounded interactions. This adjustment takes the form of an additional covariate that is included in the model. The value of this additional covariate is the normalized cross-product of the two

confounded covariates. Normalization is necessary because a covariate with a large standard deviation or a different mean may mask the effect of the other covariate in the interaction pair.

An example of a confounded interaction is a covariate and its derivative. If an explanatory variable follows a sinusoidal pattern, then its derivative will follow the pattern of a cosine, which is a sine wave with a 90° phase shift. If the response variable is correlated with either of these covariates, it is difficult to isolate which of the two has a more significant relationship with the sampled response [17,36,37,38].

Interaction terms can be included in the logistic model if they are *statistically significant* or have a *meaningful interpretation* [17,25,41]. Statistical significance implies that the model is more accurate if the interaction term is included and can be determined mathematically as will be discussed in subsequent sections. A meaningful interpretation implies that the interaction term has some significance in terms of the results of the study. In the previous example involving a sinusoid and its derivative, the second and even third derivatives can be calculated and proven mathematically to be confounded with the original sinusoid; however, such interaction terms should be eliminated if they cannot be interpreted in a meaningful way. Furthermore, certain biologically or physically confounded terms will not be statistically discernible, yet still warrant an interaction term in the model [25]. Therefore, it is necessary to use discretion when customizing the logistic function $\pi(x)$ such that statistically and experimentally meaningful confounded interactions are included. In the mechanoreceptor model discussed previously, interaction terms between *stress* and *strain* are meaningful because they describe mechanical states within the skin that cannot be described fully by either *stress* or *strain* alone.

Higher order (squared, cubed, etc.) interactions can be included. However, once again, it is important to be able to interpret the results of the model. Higher order interactions of *stress* and *strain* in the mechanoreceptor model do not yet have an interpretable physical or biological manifestation. In fact, over-fitting the model can lead to inaccuracies in the predicted probabilities and exaggerated independent variable standard errors because the model may be fitted to the random noise inherent in the system rather than the incipient behavior of the system [25,41].

A second consideration when fitting a model to sampled data is the issue of events per independent variable. More precisely, how many positive events are needed per covariate to obtain reliable estimates for the regression coefficients (Beta values)? A general heuristic is to include no more primary independent variables (disregarding interaction terms) than the number of events in the sample space divided by 10 [25]. This heuristic does not apply exclusively to logistical models; rather, it applies universally to data modeling techniques.

In the remaining sections of this *Background*, the terms logistical regression and logistical model will be assumed to refer to the multivariate model unless explicitly stated otherwise.

3.3.2 Building the Logistical Regression Model

There are two primary techniques for building a logistical model from sampled data. One method involves simply including all independent variables of interest in the model and evaluating the contribution that each has upon the accuracy of the fitted model [25,41].

The second, and more conventional technique, is "stepwise regression." In stepwise regression, each independent variable of interest is fitted with a univariate logistical model containing that independent variable only. Within this model, a strong correlation between the independent variable and the dependent variable indicates this independent should be included in the multivariate model. Next, independent variables that show associations with the dependent variable in their respective univariate model are included in the multivariate model. Each independent variable is first included then excluded and its significance tested by contrasting the "goodness-of-fit" of the model with and without the variable in question [25,41]. Two common statistics used to evaluate the significance of an independent variable within the logistical model are the *odds ratio* and the *likelihood ratio test*. The likelihood ratio test will be discussed in the next section and the odds ratio in a subsequent section.

Likelihood Ratio Test

In order to build a suitable logistical regression model, an indication of the relative significance of the marginal contribution of the $x_i\beta_i$ term to the entire logit transform of $\pi(x_n)$ is needed. A statistic called *deviance* (D) is useful for expressing this relationship and is defined as [3,16,25,28,31]:

$$D = -2 \ln\left[\frac{[\text{likelihood}_{\text{fittedmodel}}]}{[\text{likelihood}_{\text{saturatedmodel}}]}\right] \qquad \text{[Eq. 7]}$$

where:

- The saturated model is the model that contains all independent variables that showed significant association with the dependent variable in their respective univariate models
- The fitted model contains the independent variable under test
- The likelihood function indicates the probability of the observed outcome as a function of the unknown parameters and will be described below (Equation 8)
- The constant multiplication factor -2 forces the deviance function to be positive and to assume an approximately chi-squared distribution

Essentially, deviance represents the likelihood of the fitted model normalized by the saturated model. A large, positive value of the *i*th term of the deviance indicates the model does not provide a good fit for observation *i* [3,25,31]. If the deviance of a model calculated with a particular variable is significantly larger than that without the variable, then it would suffice to say that the variable should not be included in the model. Within

the brackets of Equation 7 is a ratio of likelihood functions, hence the qualification of deviance as the *likelihood ratio test*.

The deviance of the fitted logistical model is the equivalent of the residual sum of squares in linear regression. In fact, it produces the residual sum of squares value for a normal distribution [16,25,31,40]. Consequently, deviance will become smaller with each fitted parameter in the model [28,31].

Least Squares

In linear regression, the model is fitted to the data by a minimization of the least squares error in the model prediction for the system output relative to the observed system output [3,6,16,40]. If least squares were used for the logistical regression model, the unknown parameters β_0 and β_1 for the univariate case ($p = 1$) and all unknown parameters β_i for the multivariate case ($i = 1 \ldots p$) would be selected to minimize the chi-squared error of the estimation. This method, however, does not account for the variance in these estimated coefficients, particularly in the case where independent variables are categorical. Unless weighting is employed, this variance leads to an over-estimation of the standard error of the coefficients. Therefore, a more powerful technique called maximum likelihood estimation is used in logistical regression [2,6,7,16,19,24,25,34,49].

Maximum Likelihood

The principle of maximum likelihood estimation is to determine which Beta values maximize the probability of generating the observed outcome [2,6,7,16,19,24,25,34,41,49]. Maximum likelihood estimation involves first constructing a likelihood function, which indicates the probability of the observed outcome as a function of the unknown parameters (β_i). The likelihood function is [6,16,25,31,41]:

$$\text{Likelihood} = l(B) = \prod_{i=1}^{M} \pi(x)^{Y_i} [1 - \pi(x)]^{1-Y_i} \qquad \text{[Eq. 8]}$$

where:

- Y_i is the known outcome of the ith sample point
- M is the number of sample points

In essence, the likelihood function is the probability of the product of the observed dependent variables in the sample. In other words, the likelihood function is a measure of how likely it is that the observed outcome is predicted from the observed input. Maximum likelihood estimation involves estimating the parameters in the likelihood equation such that the likelihood of the observed outcome is maximized, given the recorded inputs to the model. For binary outcomes, the complexity of estimating the maximum likelihood is reduced by solving for the unknown variables that minimize $-2*L$ where L is the log of the likelihood function for the model. Minimizing the log of the likelihood function is mathematically equivalent to maximizing the likelihood function.

The shape of the minimized –2*log-likelihood function is an indication of the extent to which the values of the unknown parameters (β_i) of the model are the best fit.

For given values of "x", the log-likelihood function for the logistical regression model is [6,16,25,41]:

$$L(\beta) = \text{Log-Likelihood} = ln(l(\beta)) = \sum_{i=1}^{M}\left[Y_i \ln\left[\pi(x_n)\right] + \left[1 - Y_i\right]\left[\ln(1 - \pi(x_n))\right]\right] \quad \text{[Eq. 9]}$$

In a simplified form, Equation 9 becomes:

$$L(\beta) = \sum_{i=1}^{M} [Y_i\varphi - ln(1 + e^{\varphi})] \quad \text{[Eq. 10]}$$

where:

- φ is a vector representation of the linear combination of the explanatory variables, $\beta_0 + x_1\beta_1 + x_2\beta_2 + \ldots + x_n\beta_n$

The β values used to maximize this function (in other words, minimize the quantity –2*L) produce the best fit. To find these β values, the function in Equation 10 is differentiated with respect to β_i for $i = 0 \ldots n$ to produce [3,19,24,25,41]:

$$\sum_{i=1}^{M} [Y_i - \pi(x_i)] = 0; \text{ (for } \beta_0) \quad \text{[Eq. 11]}$$

and

$$\sum_{i=1}^{M} [x_i(Y_i - \pi(x_i))] = 0; \text{ (for } \beta_1\text{: } n = 1 = \text{univariate case)} \quad \text{[Eq. 12]}$$

For the multivariate case (n independent variables), Equation 12 becomes:

$$\sum_{i=1}^{M}\sum_{j=1}^{n} [x_{ij}(Y_i - \pi(x_{ij}))] = 0; \text{ (for } \beta_{1\ldots n}\text{: } j = 1 \ldots n = \text{multivariate case)} \quad \text{[Eq. 13]}$$

Equation 11 is used to solve for β_0 and Equation 12 is used to solve for β_1 in the univariate case. Equation 13 is used to solve for β_1 - β_n in the multivariate case. Equation 13 is inherently non-linear and its solution (β values) must be computed using iterative methods. The algorithm to compute the maximum likelihood starts with an initial arbitrary "guesstimate" for the logit coefficients. The algorithm then determines the direction and size change for the logit coefficients that will increase the likelihood function. After this initial function is estimated, the residuals are tested and another estimate is made with the improved function. This process is repeated typically about a half-dozen times until the logit function converges (i.e., the likelihood does not change significantly when the coefficients are re-estimated). For more information on this algorithm, the interested reader is encouraged to consult [4,19,24,25,41,46,49].

3.3.3 Interpreting the Logistical Regression Model

When interpreting the fitted model, one should ask, "What information do the estimated coefficients communicate regarding the questions motivating the study?" Techniques to help answer this question are discussed below.

The Odds Ratio

A significant and easy to interpret measure of association between explanatory variables and the dependent variable is the *odds ratio*. Essentially, the odds ratio is a quantitative measure of the degree to which the outcome variable changes for a single-unit change in the explanatory variable being tested [17,25,34,35,40,41,51,52]. In addition to being used to interpret the fitted logistical model, the odds ratio can be used in stepwise regression to help determine which independent variables are significant and, therefore, which independent variables should be included in the model. Mathematically, the odds ratio (denoted by Ψ) takes the form of the ratio of the odds for an input variable to be one over the odds for an input variable to be zero [16,24,25,40,41]:

$$\Psi = \left[\frac{\text{odds of } x = 1}{\text{odds of } x = 0}\right] = \frac{\left[\dfrac{\pi(1)}{1-\pi(1)}\right]}{\left[\dfrac{\pi(0)}{1-\pi(0)}\right]} \qquad \text{[Eq. 14]}$$

For the univariate case ($p = 1$), Equation 14 becomes:

$$\Psi = \frac{\left[\dfrac{e^{\beta_0+\beta_1}}{1+e^{\beta_0+\beta_1}}\right]\left[\dfrac{1}{1+e^{\beta_0}}\right]}{\left[\dfrac{e^{\beta_0}}{1+e^{\beta_0}}\right]\left[\dfrac{1}{1+e^{\beta_0+\beta_1}}\right]} = e^{\beta_1} \qquad \text{[Eq. 15]}$$

The logit difference is simply the natural log of Ψ. If, for example, an odds ratio was calculated to be 2, then the model including the particular explanatory variable corresponding to the β value used to determine the odds ratio would be twice as likely to predict a positive outcome than if the model did not include the explanatory variable. One could safely deduce that this explanatory variable should be included in the model due to its profound effect on the probability of a particular outcome (assuming the variable is of interest or can be interpreted in a meaningful fashion).

In the multivariate case, the independent variables can be normalized before determining the odds ratios for each variable [17,41]. The advantage of using standardized logit coefficients is that they indicate the relative significance of the independent variables with which they are associated, despite any disparity in the units of the independent variables [41,52]. When the independent variables have the same mean and standard

deviation, the Beta (β) values and odds ratios corresponding to each independent variable rank those independent variables in terms of their significance relative to one another within the fitted model. The odds ratios and relative significance of each independent variable can be compared by plotting them on the same axis measured in multiples of the standard deviation. Since normalization is a linear transform, the data can be converted back into original units in a straightforward manner, allowing for the comparison of sensitivity ranges among trials [17].

When odds ratios are calculated for normalized independent variables (zero mean, unit variance), the following categories can be used to contrast the influence that each independent variable has upon the response:

- **Odds ratio >> 1**: Very significant positive correlation between the independent and the response. When the level of the independent variable increases, the probability of generating a positive response *increases*.
- **Odds ratio ≈ 1**: No statistical correlation between the independent variable and the response. In other words, the independent variable is no more correlated with the response than random noise.
- **0 < Odds ratio << 1**: The independent variable is negatively correlated with, or protective against, the response. When the level of the independent increases, the probability of generating a positive response *decreases*.

For the situation in which the explanatory variables are continuous, the equation for the odds ratio becomes [25,41]:

$$\Psi(\mathrm{a},\mathrm{b}) = \frac{\left(\text{odds of } \mathrm{x}=\mathrm{a}\right)}{\left(\text{odds of } \mathrm{x}=\mathrm{b}\right)} = \frac{\left[\dfrac{\pi(\mathrm{x}=\mathrm{a})}{1-\pi(\mathrm{x}=\mathrm{a})}\right]}{\left[\dfrac{\pi(\mathrm{x}=\mathrm{b})}{1-\pi(x=b)}\right]} = \mathrm{e}^{\beta_1(\mathrm{a-b})} \qquad \text{[Eq. 16]}$$

This expression can also be used to evaluate pseudo-odds ratios that relate the change in the outcome variable for a change of ($a - b$) in the explanatory variable "x". It is important in this case to evaluate the logit at $x = a$ and $x = b$, rather than the situation above in which $a = 1$ and $b = 0$. It is conceivable that the odds ratio will be different depending on the levels of "a" and "b" used in the calculation; therefore, a useful quantity in this case is the average (or geometric mean) of the odds ratios for all "a" and "b" such that ($a - b$) = 1 [25,41]. In general, the relationship between the odds ratio and the range of the continuous explanatory variable should be selected so that it has a meaningful interpretation [17,25]. For example, for a model relating the outcome variable of heart disease to the continuous explanatory variable age, an odds ratio corresponding to a period of 1 month does not have a meaningful interpretation since the average person lives for approximately 70 years.

For a multivariate model in which there are only two covariates, "η" and "υ", and which are confounded, the logistic function becomes [16,25,40,41]:

$$\pi(\eta,\upsilon) = \frac{e^{\beta_0+\eta\beta_1+\upsilon\beta_2+\eta\upsilon\beta_3}}{1+e^{\beta_0+\eta\beta_1+\upsilon\beta_2+\eta\upsilon\beta_3}} \qquad \text{[Eq. 17]}$$

where

- $\eta\upsilon$ represents the normalized product of the two terms that are confounded

The linear outcome of the logit transform of $\pi(\eta,\upsilon)$ becomes:

$$\ln\left[\frac{\pi(\eta,\upsilon)}{1-\pi(\eta,\upsilon)}\right] = e^{\beta_0+\eta\beta_1+\upsilon\beta_2+\eta\upsilon\beta_3} \qquad \text{[Eq. 18]}$$

In order to calculate the odds ratio for models with interaction terms, all explanatory variables and interaction terms are held constant while the formula for the odds ratio of the variable in question is derived. The formula for the odds ratio of "η" is [25,40,41]:

$$\Psi(\eta|\upsilon) = \frac{(\text{odds of } \eta=\eta_1 \text{ given } \upsilon)}{(\text{odds of } \eta=\eta_2 \text{ given } \upsilon)} = \frac{\left[\frac{\pi(\eta=\eta_1)}{1-\pi(\eta=\eta_1)}\right]}{\left[\frac{\pi(\eta=\eta_1)}{1-\pi(\eta=\eta_2)}\right]} \qquad \text{[Eq. 19]}$$

Equation 19 translates into:

$$\Psi(\eta|\upsilon) = e^{(\eta_1-\eta_2)\beta_1} + e^{\tau(\eta_1-\eta_2)\beta_3} \qquad \text{[Eq. 20]}$$

For [$\eta_1 - \eta_2 = 1$], this value expresses the odds ratio of η for the given of υ. To calculate the true geometric mean of the odds ratio, one must now average the samples such that [$\eta_1 - \eta_2 = 1$] for all η and for all υ [25]. With multiple covariates and interaction terms, this task becomes increasingly compute intensive.

This method can easily be expanded for use with multivariate models comprising a larger set of covariates with multiple interaction terms.

The Wald Test

The Wald statistic is a computationally simple quantification of the significance of individual coefficients and their associated independent variables within the fitted model. There are two commonly employed forms of the Wald statistic. The first is the squared ratio of the unstandardized logit coefficient to the standard error of that coefficient [3,25,41]:

$$\text{Wald Statistic}_1 = \left[\frac{\beta_i}{SE(\beta_i)}\right]^2 \quad \text{[Eq. 21]}$$

where:

- β_i is the maximum likelihood fitted Beta value for the *i*th independent variable
- SE(β_i) is the standard error (the square root of the variance of β_i)

The first form is asymptotically distributed as a chi-square distribution.

The second form is the simply the ratio of the unstandardized logit coefficient to its standard error:

$$\text{Wald Statistic}_2 = \frac{\beta_i}{SE(\beta_i)} \quad \text{[Eq. 22]}$$

The second form is normally distributed.

A low Wald value relative to the Wald values for the other fitted parameters in the model indicates that the independent variable does not have a profound effect on the model, meaning it can be removed from the model without significantly changing the goodness-of-fit [3]. It should be noted that the Wald statistic should not be used exclusively to measure logit coefficient significance because significant coefficients with large standard errors may have deflated Wald values. Conversely, insignificant coefficients with small standard errors may be falsely identified as significant if the Wald test is used as the sole measure of significance.

Outliers

In any regression diagnosis, an analysis of outliers is necessary. Outliers are data points that fall outside the confidence interval of a fitted model. In other words, outliers fall beyond a reasonable deviation from the curve representing the fitted model. In logistical regression, the typical means for identifying and evaluating outliers are plots of Δ (delta) Beta vs. Probability, ΔDeviance vs. Probability, and ΔChi-square vs. Probability.

Before explaining how each of these plots is generated, two variables must be defined: the Pearson residual and the Deviance residual. A "residual" is an estimate of error that can be used not only to identify cases for which the model fits poorly, but also cases that have a disproportionately large influence on the estimated coefficients in the fitted model.

These residuals are used in the functions that generate ΔBeta, ΔDeviance, and ΔChi-square.

The Pearson residuals are elements of the Pearson chi-square that can be used to detect ill-fitted patterns in the independent variables. The Pearson residual, also known as the "standardized" or "chi" residual, defined for the *j*th independent variable pattern is given by [25,41]:

$$r_j = \frac{Y_j - m_j \overline{\pi_j}}{\sqrt{m_j \widehat{\pi_j}(1-\overline{\pi_j})}} \qquad \text{[Eq. 23]}$$

where:

- Y_j is the number of successes for the *j*th independent variable pattern
- m_j is the number of trials for the *j*th independent variable pattern
- $\overline{\pi_j}$ is the estimated probability for the *j*th independent variable pattern

The Pearson residual is simply the difference between the observed and estimated probabilities divided by the binomial standard deviation of the estimated probability. For large samples, this residual is normally distributed with a mean of 0 and a standard deviation of 1. A Pearson residual with a large absolute value for a particular observation indicates the model fits that observation poorly.

The deviance residual is similar to the Pearson. Deviance residuals are based on the model deviance and are also useful in identifying ill-fitted independent variable patterns. The model deviance is a goodness-of-fit statistic based on the log-likelihood function as discussed previously. The deviance residual defined for the *j*th independent variable pattern is given by [25, 41]:

$$d_j = \pm \left[2 \left(Y_j \log_e \left\{ \frac{Y_j}{m_j \overline{\pi_j}} \right\} + (m_j - Y_j) \log_e \left\{ \frac{m_j - Y_j}{m_j (1-\overline{\pi_j})} \right\} \right) \right]^{\frac{1}{2}} \qquad \text{[Eq. 24]}$$

Given these definitions, ΔBeta, ΔDeviance, and ΔChi-square are now discussed.

ΔBeta is useful for detecting independent variable patterns that have a strong influence on the estimates of the coefficients. It represents the standardized change in the regression coefficients if a particular case is deleted. The value itself is proportional to the Pearson residual. The change in the regression coefficients if case *j* is deleted is given by [25,41]:

$$\Delta\beta_j = \frac{r_j^{\,2} h_j}{1-h_j} \qquad \text{[Eq. 25]}$$

where:

- h_j is the leverage for the jth independent variable pattern
- r_j is the Pearson residual for the jth independent variable pattern (defined in Equation 23 above)

The leverage indicates whether an observation has unusual predictors. These observations have a large influence on the values of the estimated regression coefficients. It ranges from 0 (no influence) to 1 (completely determines the estimation of the coefficients in the fitted model). Since a thorough discussion of this value is beyond the scope of this section and the interpretation of this variable is straightforward, the interested reader is encouraged to consult [25,26,41,46] for a more thorough analysis. Most logistical regression software packages (SPSS, STATA, SAS, MINITAB) calculate this value automatically.

The Δchi-square is useful for detecting independent variable patterns that lie outside the confidence interval. Just as in the formula for ΔBeta, it is calculated by deleting the observations corresponding to a given pattern and contrasting the accuracy of the fit before and after the deletion. The change in chi-square if case j is deleted is given by [25,41]:

$$\Delta\chi_j^2 = \frac{r_j^2}{1-h_j} \qquad \text{[Eq. 26]}$$

Similarly, the Δdeviance statistic is a measure of how well a particular observation contributes to the fitted model. It is simply the deviance residual added to the ΔBeta and is given by [25,41]:

$$\Delta D_j = d_j^2 + \frac{r_j^2 h_j}{1-h_j} \qquad \text{[Eq. 27]}$$

where:

- d_j is the deviance residual for the jth independent variable pattern (defined in Equation 24 above)

In general, large values of ΔD_j and $\Delta\chi_j^2$ for a particular observation indicate that observation does not fit well within the model. A large value for $\Delta\beta_j$ indicates that the given observation exerts a particularly large influence on the estimated coefficients in the model. ΔD_j and $\Delta\chi_j^2$ are used to identify outliers while $\Delta\beta_j$ is used to assess the impact that the identified outliers.

The value of ΔD_j, $\Delta\chi_j^2$, or $\Delta\beta_j$ is not particularly useful in itself, however, when plotted versus the estimated logistic probability ($\overline{\pi_j}$) or leverage (h_j), one can visually

identify observations that lie outside the confidence interval. A plot of $\Delta\beta_j$ vs. $\overline{\pi_j}$ is shown in Figure 3.5. The RED dots indicate positive ($Y_j = 1$) observations. The BLUE dots represent binary 0 outcomes ($Y_j = 0$). The data points in the upper left hand corner of the graph are the observations that have the most influence on the estimated logit coefficients in the fitted model. In this graph, no observations are flagged as possible outliers since there are no data points that lie a significant distance away from the cluster.

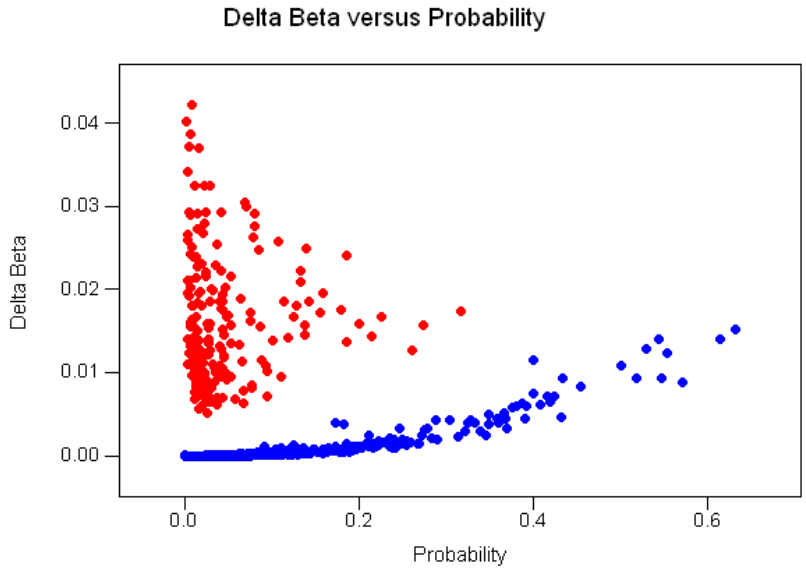


FIGURE 3.5: SAMPLE DELTA BETA VS. PROBABILITY
The RED dots indicate positive ($Y_j = 1$) observations. The BLUE dots represent binary 0 outcomes ($Y_j = 0$). No outliers are shown.

A plot of $\Delta\chi_j^2$ vs. $\overline{\pi_j}$ is shown in Figure 3.6. The RED dots indicate positive ($Y_j = 1$) observations. The BLUE dots represent binary 0 outcomes ($Y_j = 0$). In this plot, a single positive-outcome data point sits apart from the balance of the plot at [$\overline{\pi_j} \approx 0.0$, $\Delta\chi_j^2 \approx 380$]. This point, indicated by the GRAY circle, is flagged as an outlier.

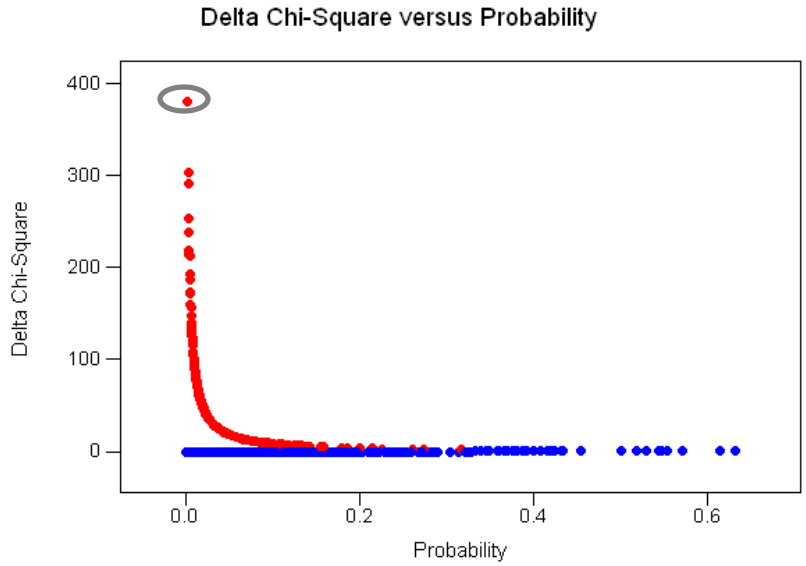


FIGURE 3.6: SAMPLE DELTA CHI-SQUARE VS. PROBABILITY

The RED dots indicate positive ($Y_j = 1$) observations. The BLUE dots represent binary 0 outcomes ($Y_j = 0$). The outlier is circled in GRAY.

A plot of ΔD_j vs. $\overline{\pi}_j$ is shown in Figure 3.7. The RED dots indicate positive ($Y_j = 1$) observations. The BLUE dots represent binary 0 outcomes ($Y_j = 0$). In this plot, 3 positive-outcome observations are flagged as outliers. These points exist at approximately [$\overline{\pi}_j \approx 0.55$, $\Delta D_j \approx 1$] and are circled in Figure 3.7 with a GRAY circle (the RED dots within the circle are the outliers). These data points fall slightly outside the balance of the slope representing positive outcomes.

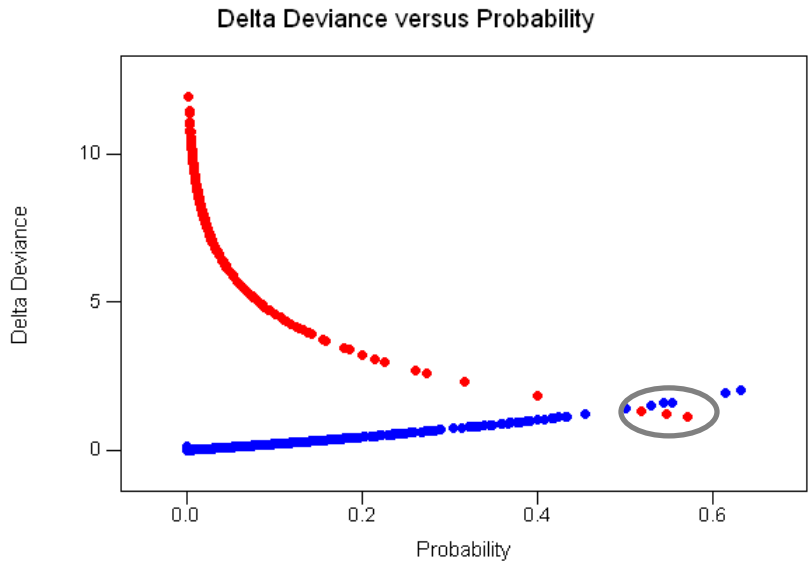


FIGURE 3.7: SAMPLE DELTA DEVIANCE VS. PROBABILITY

The RED dots indicate positive ($Y_j = 1$) observations. The BLUE dots represent binary 0 outcomes ($Y_j = 0$). The outliers are circled in GRAY (RED dots only within the circle).

Besides these purely numerical methods, several types of graphical methods can be employed to gain insight regarding the distribution of error terms associated with the fitted logistic regression model. These models, as discussed by Landwehr, et. al. [31], include local mean deviance plots and partial residual plots. It is, however, beyond the scope of this thesis to discuss these methods extensively; therefore, some general comments will be made regarding each of the methods.

Local mean deviance plots are generated from scatterplots of the system response on axes corresponding to the data producing the response. The responses are grouped by a proximity metric and the local mean deviance of each group is calculated. This value is plotted versus the degree of freedom used to generate the mean. It is useful for detecting the lack of a necessary interaction term in logistic function $\pi(x)$ [3,28,31].

Partial residual plots help to detect relationships between explanatory variables in the model and the response variables that do not adhere to the established model, signifying the need for the inclusion of non-linear terms (squared, absolute value, etc.). Such a method determines if improvements could be made to the model but does not indicate what types of improvements can be made [3,31].

3.3.4 Assessing the Fit of the Model

Assessments of the accuracy of the fitted model are a necessary component of any regression. The most common techniques for evaluating the goodness-of-fit are discussed below.

Classification Tables

Also known as contingency tables, classification tables are indices expressing the predictive efficiency of the fitted model. The tables organize observed and predicted data into a format that summarizes the predictive accuracy of the model. A 2 x 2 table is shown below in Table 3.2. In logistical regression, a 2 x 2 table is used because there are only two levels for the dependent variable. Each level can be correctly or incorrectly predicted, resulting in 4 values that can be organized in a 2 x 2 fashion.

	Observed		
Predicted	Dependent = 1	Dependent = 0	**Total**
Dependent = 1	*A*	*B*	*X*
Dependent = 0	*C*	*D*	*Y*
Total	*U*	*V*	*Z*

TABLE 3.2: *2* X *2* CONTINGENCY TABLE

In Table 3.2, the variables are defined as follows:

- A = # of responses that were both observed experimentally and predicted by the model as 1.
- B = # of responses that were observed to be 0 but predicted incorrectly as 1. These variables represent Type I statistical errors.
- C = # of responses that were observed to be 1 but predicted incorrectly as 0. These variables represent Type II statistical errors.
- D = # of responses that were both observed experimentally and predicted by the model as 0.
- X = Sum of $A + B$ = # of responses predicted by the fitted model to be binary 1.
- Y = Sum of $C + D$ = # of responses predicted by the fitted model to be binary 0.
- U = Sum of $A + C$ = # of observed binary 1 responses.
- V = Sum of $B + D$ = # of observed binary 0 responses.
- $Z = X + Y = U + V$ = total number of data points.

The predicted values of the fitted model are probabilities. If the predicted probability generated by the model for a particular observation is greater than a *cut value*, the observation is classified as a binary 1 in the table. Otherwise, it is classified as a binary 0. The value selected as the *cut value* strongly influences the efficiency of the model as defined in the context of a contingency table. A typical *cut value* is 0.5.

Three descriptive statistics are defined based on the values in Table 3.2, *sensitivity*, *specificity*, and *correct prediction %*. *Sensitivity* describes the % of the dependents in the binary 1 category that were correctly classified. It takes on the following form [25]:

$$Sensitivity = \frac{A}{U} \text{ x } 100 \qquad \text{[Eq. 28]}$$

The quantity [100 – *Sensitivity*] is a measure of the percent of Type II statistical errors in the prediction.

Specificity describes the % of the dependents in the binary 0 category that were correctly classified in the fitted model. It assumes the following form [25]:

$$Specificity = \frac{D}{V} \text{ x } 100 \qquad \text{[Eq. 29]}$$

The quantity [100 – *Specificity*] is a measure of the percent of Type I statistical errors in the prediction.

The *Correct Prediction %* defines the overall predictive accuracy of the fitted model. It is an average of the Specificity and Sensitivity that is weighted according to the number of observations in each category. It is defined as [25]:

$$\textit{Correct Prediction \%} = \frac{A+D}{Z} \text{ x } 100 \qquad \text{[Eq. 30]}$$

This value is sensitive to group sizes and favors the prediction % of the larger of the two groups.

R^2 Tests

In linear regression, a measurement called R^2 is often used to evaluate the error in the prediction of the fitted model. R^2 represents the proportion by which the regression equation reduces the error in the prediction relative to a simple prediction of the mean $\overline{Y}$. The linear regression R^2 value is composed of a number of simpler evaluations of the accuracy of the fitted model. These evaluations are the *total sum of squares*, the *error sum of squares*, and the *residual sum of squares* [41]:

$$\text{Total Sum of Squares} = SST = \sum_{i=1}^{N}\left(Y_i - \overline{Y}\right)^2 \qquad \text{[Eq. 31]}$$

where:

- Y_i is the *i*th value (0 or 1)
- $\overline{Y}$ is the mean of Y
- N is the number of samples

$$\text{Error Sum of Squares} = SSE = \sum_{i=1}^{N}\left(Y_i - \overline{\overline{Y}}_i\right)^2 \qquad \text{[Eq. 32]}$$

where:

- $\overline{\overline{Y}}_i$ is the *i*th predicted value for Y

$$\text{Residual Sum of Squares} = SSR = SST - SSE \qquad \text{[Eq. 33]}$$

Given these definitions, R^2, or the proportional reduction in error, is defined as:

$$R^2 = SSR/SST = 1 - (SSE/SST) \qquad \text{[Eq. 34]}$$

The value of R^2 ranges from 0 (the independent variables are no help at all) to 1 (the independent variables allow us to predict Y_i precisely).

While there are several functions that will produce psuedo-R^2 quantities in logistical regression diagnostics, none of the functions produce a measurement that is analogous to the R^2 value in linear regression, therefore, the psuedo-R^2 value is denoted R_L^2. R_L^2 is

based on the likelihood function and its interpretation becomes "*the proportional reduction in the absolute value of the log-likelihood*." It measures the degree to which including an independent variable in the model increases the "badness-of-fit" of the model and ranges from 0 (the independent variable is useless in predicting the dependent variable) to 1 (the model predicts the dependent variable perfectly incorporating the independent variable).

The formula for R_L^2 is given by [41]:

$$R_L^2 = 1 - \frac{L_P}{L_0} \qquad \text{[Eq. 35]}$$

where:

- L_P is the log-likelihood (Equation 8) of the model containing p covariates
- L_0 is the log-likelihood (Equation 8) of the model containing none of the covariates (the logistic function, therefore, contains only the intercept or B_0 coefficient)

While in linear regression, large R^2 indicate an accurate model and values very close to 1 are not uncommon, in logistical regression, R_L^2 values are considerably lower. For this reason, the R_L^2 values are typically used in model building rather than as a descriptive statistic.

Chi-Squared Goodness-of-Fit

A familiar statistic that is often used to determine the goodness of fit is the *chi-squared* (χ^2) statistic. It is a summation over all samples on which the model is based of the following empirical equation [16]:

$$\frac{\left(\text{observed \# of successes - fitted \# of successes}\right)^2}{\text{fitted \# of successes}} \qquad \text{[Eq. 36]}$$

In logistical regression, the formula for chi-squared error is [16,25,40]:

$$\chi^2 = \sum_{i=1}^{M} \frac{\left[Y_i - \pi(x)\right]^2}{\left[\pi(x)\right]\left[1-\pi(x)\right]} = \sum_{i=1}^{M} \frac{\left[Y_i - \left(\frac{e^{\beta_0 + x_1\beta_1 + \ldots + x_p\beta_p}}{1 + e^{\beta_0 + x_1\beta_1 + \ldots + x_p\beta_p}}\right)\right]^2}{\left[\frac{e^{\beta_0 + x_1\beta_1 + \ldots + x_p\beta_p}}{\left(1 + e^{\beta_0 + x_1\beta_1 + \ldots + x_p\beta_p}\right)^2}\right]} \qquad \text{[Eq. 37]}$$

where:

- M represents all of the sample points in the sample space
- Y_i represents the observed outcome (0 or 1) for the ith sample point
- $\pi(x)$ represents the familiar logistic function

The denominator in the equation, $[\pi(x)*(1 - \pi(x))]$, is the estimated variance of the given logistic distribution $\pi(x)$.

The chi-squared statistic does not give any indication of concentrations in the discrepancy between observed and fitted portions of the model or of the distribution of error beyond that immediately discernible from artificial grouping. Additionally, this statistic is unstable for fitted values near zero and one. Since logistical regression attempts to fit a logistic function to binary (0 or 1) data, this statistic, in itself, is not a definitive measure of the goodness of fit for the logistical regression model [6,25,31,41].

Hosmer & Lemeshow Tests

Hosmer & Lemeshow tests consist of first grouping observations based on their estimated probabilities and then evaluating the model fit for each of the groups. This method extends the notion of a "cut value" segregating groups into two categories into multiple cut values segregating groups into multiple categories. Strategies for group segregation can either be based on fixed values for the estimated probability or by dividing the probability range (0 to 1) into percentiles. Typical commercial software applications (namely SPSS, SAS, STATA) follow the latter technique, usually using 10 groups (percentiles of 10 %). The first group typically contains the smallest 10 % of the estimated probabilities while the last group contains the largest 10 % of the estimated probabilities. The Hosmer & Lemeshow goodness-of-fit statistic (denoted $\overline{C}$) is obtained by calculating the Pearson chi-square statistic from a contingency table of the number of groups (g) by the number of categories for the dependent variable (2 for binary) [25,26]:

$$\overline{C} = \sum_{k=1}^{g} \left[\frac{\left(\sum_{j=1}^{c_k} Y_j \right) - n_k' \sum_{j=1}^{c_k} \frac{m_j \overline{\pi}_j}{n_k'}}{n_k' \sum_{j=1}^{c_k} \frac{m_j \overline{\pi}_j}{n_k'} \left(1 - \sum_{j=1}^{c_k} \frac{m_j \overline{\pi}_j}{n_k'} \right)} \right] \qquad \text{[Eq. 38]}$$

where:

- g is the number of groups
- c_k is the number of levels of the covariates that were grouped into the kth group
- n_k' is the number of observations that fall into the kth group

- the quantity $\sum_{j=1}^{c_k} Y_j$ represents the number of responses (binary 1 values) among the c_k levels
- the quantity $\sum_{j=1}^{c_k} \frac{m_j \overline{\pi_j}}{n_k'}$ represents the average estimated probability

		Dependent = 1			Dependent = 0		
Group	**Prob**	*Observed*	*Expected*	*% Accuracy*	*Observed*	*Expected*	*% Accuracy*
1	A	K	U	EE	OO	YY	AJ
2	B	L	V	FF	PP	ZZ	AK
3	C	M	W	GG	QQ	AB	AL
4	D	N	X	HH	RR	AC	AM
5	E	O	Y	II	SS	AD	AN
6	F	P	Z	JJ	TT	AE	AO
7	G	Q	AA	KK	UU	AF	AP
8	H	R	BB	LL	VV	AG	AQ
9	I	S	CC	MM	WW	AH	AR
10	J	T	DD	NN	XX	AI	AS

TABLE 3.3: G X 2 HOSMER & LEMESHOW CONTINGENCY TABLE

Within Table 3.3, the variables A – AS are defined as:

- *A – J* = Average probability of the estimated probabilities of the observations that fall within group 1 – 10, respectively.
- *K – T* = Number of dependent variable observations that were binary 1 in the group 1 – 10, respectively.
- *U – DD* = Number of dependent variables estimated as binary 1 by the model for group 1 – 10, respectively.
- *EE – NN* = Percent accuracy within the indicated group, 1 – 10. Calculated for each group as:

$$\text{Percent Accuracy} = \frac{|Estimated - Observed|}{Observed}$$

- *OO – XX* = Number of dependent variable observations that were binary 0 in the group 1 – 10, respectively.
- *YY – AI* = Number of dependent variables estimated as binary 0 by the model for group 1 – 10, respectively.
- *AJ – AS* = Percent accuracy within the indicated group. Calculated the same as above.

Table 3.3 is the standard format for expressing the Hosmer & Lemeshow goodness-of-fit values.

3.4 Logistical Model Summary

Logistical regression has the benefits of linear regression but allows experimenters to overcome the limitations of linear regression models for certain systems. The following are modeling assumptions met by both logistical and linear regression [7,12,34,41]:

- **All relevant variables are included in the model**. If a relevant variable is omitted, the shared variability may be wholly attributed to the complementing variable, leading to inaccurate predictions or inflated error.
- **All irrelevant variables are excluded from the model**. If irrelevant variables are included in the model, the variance they share with relevant variables may be incorrectly attributed to the irrelevant variables, resulting in greater standard errors of the regression coefficients for these independent variables.
- **Interaction terms must be explicitly included**. The significance of interactions can be tested and included or not included, depending on their significance within the model.

The following are modeling assumptions unique to logistical regression [7,12,34,41]:

- **Logistic regression does not assume a linear relationship between the dependent variable (system output) and the independent variables (system input)**. It does, however, assume a linear relationship between the logit of the independents and the dependent variable. The model accounts for nonlinear effects even when exponential and polynomial terms are not explicitly added as additional independent variables; however, the model can be explicitly modified to account for pertinent interaction terms and higher order interactions.
- **Logistical regression does not assume a normally distributed dependent variable**. Linear regression assumes that the dependent variable is normally distributed. Since a binary dependent variable is not normally distributed, least squares estimates of coefficient significance tend to be incorrect and have an inflated standard error.
- **Logistical regression does not assume normally distributed error terms**. Logistical regression assumes error and the variance in the error to be independent of the independent variable(s).
- **Logistical regression does not make any assumptions regarding the distribution of the independent variable(s).** Normal, binomial, and uniform multivariate distributions are allowed.
- **Logistical regression does not require that the independent variables be continuous**. Continuous, categorical, and mixed continuous and categorical inputs are allowed.
- **Logistical regression does not require unbounded independent variables.** Independent variables can be discrete (categorical).
- **Logistical regression assumes a discrete dependent variable.** The dependent variable can only take on discrete values. In binary logistic regression, the dependent can only assume 2 different values. However, in multinomial logistic regression, the dependent variable can assume many discrete values.

4.0 Methodology

This *Methodology* establishes a framework for modeling the afferent behavior of mechanoreceptors under dynamic, transversely applied compressive loads using logistical regression and details an experiment in which the sensitivity of rapidly adapting afferents in rat hairy skin was quantitatively correlated to the stimulus metrics *stress*, *strain*, and their time derivatives. The results of this experiment are discussed in the subsequent *Results and Analysis* sections.

4.1 Experimental Design

This section details the experiments that were conducted. The experiments were not designed to correlate nerve response thresholds to absolute levels of induced *stress* and *strain* because the internal *stress* and *strain* components in the skin are a complex function of externally applied stimuli. Instead, this study examined the effect that external components of the stimuli have on the threshold of rapidly adapting afferents.

Briefly, stimuli applied externally to the skin were carefully quantified for both the absolute indented position of the stimulator and the force applied by that stimulator. These data were transformed into the cross-platform units *strain* and compressive *stress*. Finally, a multivariate logistical regression analysis was performed and the relative contributions of each recorded variable were assessed quantitatively.

4.1.1 Goals and Objectives

The purpose of this methodology is to explain the environment in which the following hypothesis was tested:

Rapidly adapting mechanoreceptors are preferentially sensitive to compressive stress and the rate of change of strain.

This hypothesis was conjectured from the findings of Looft [36-39], Del Prete and Grigg [17,18], and Grigg [21] in similar studies.

4.1.2 Experimental Setup

The experimental setup has been detailed previously [18,21,36-39] and, therefore, does not warrant an extensive discussion. The preparation can be described succinctly as:

1) Calibrate actuator (Model 300B Lever System, Aurora Scientific, Richmond Hills, CN).
2) Anesthetize an adult Sprague-Dawley rat using Pentobarbital Sodium.

3) Depilate and remove a skin patch from the inside of the rat's leg with an intact Saphenous nerve.
4) Transfer skin patch onto support substrate in plastic chamber bathed in artificial interstitial fluid (Hepes solution). See Figure 4.2 below.
5) Stretch skin to original dimensions (as determined by a circle drawn on the skin prior to excision) by attaching hooks to edges and corners of skin patch.
6) Measure skin thickness by recording the position voltage differential between the surface of the skin and the supporting substrate surface. Convert this value to an absolute thickness by multiplying by the calibration coefficient determined in Step (1).
7) Move nerve bundle through chamber wall into oil filled recording chamber. See Figure 4.2 below.
8) Micro-dissect nerve to isolate a fiber corresponding to a single, identifiable rapidly adapting afferent.
9) Wrap afferent fiber around electrode submersed in oil solution.
10) Position actuator tip above qualitatively-established most sensitive point for the nerve.
11) Visually verify receptor is not associated with hair shaft.

The actuator movements can be either *position*-controlled (± 2 millimeter range, ~1 micrometer resolution) or *force*-controlled (± 50 gram range, ~30 milligram resolution). Before data are recorded, the actuator tip is lowered to the surface of the skin and the *position* and *force* voltage feedback is zeroed. Then, the actuator tip is pre-indented approximately 300 – 600 micrometers for several seconds. The purpose of this pre-indentation is to prevent the actuator tip from losing contact with the surface of the skin during stimulation. This pre-indented state is referred to as "stimulus neutral." Both *position* and *force* voltage readings are recorded at "stimulus neutral" and are used in subsequent data manipulation. Indented *position* and *force* are recorded as positive values.

The waveforms used to control either the *position* or *applied force* of the actuator tip are either non-repeating noise with a 10 minute period or pseudorandom noise with a 0.5 second period. The non-repeating noise waveform is generated algorithmically using a noise generator (Hewlett Packard 8057A, 1 Kilohertz output bandwidth, band-pass filtered 0.5 – 80 Hz) and the pseudorandom noise sequences are 0.5-second noise records of non-repeating noise that can be looped under the control of software running on a personal computer.

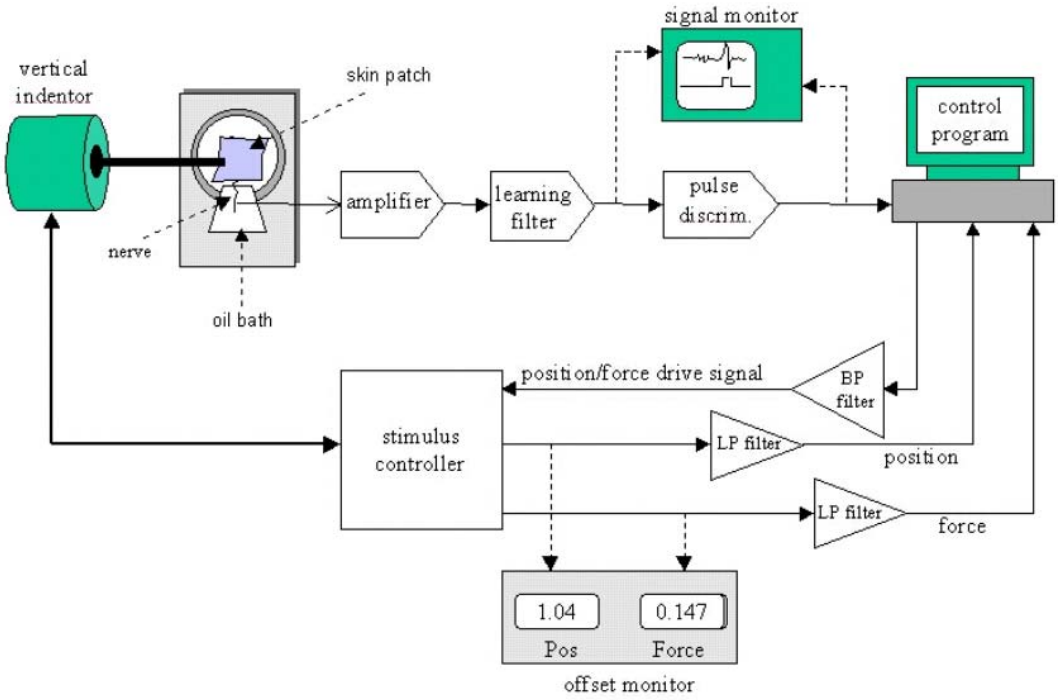


FIGURE 4.1: GENERAL SYSTEM SCHEMATIC
Image Courtesy of Looft [39]

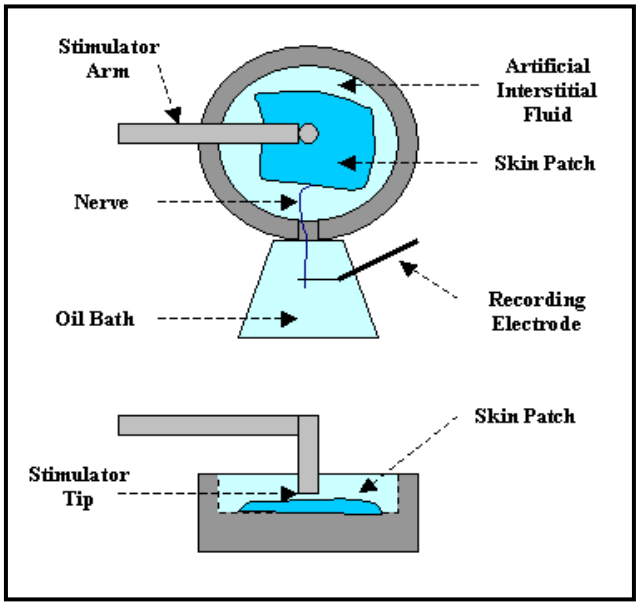


FIGURE 4.2: SKIN SUPPORT CONTAINER

4.1.3 Experimental Parameters

Parameters were varied in this experiment in order to ensure uniform coverage of the stimulus input space. These parameters include stimulus control, stimulus waveform, actuator tip area, and stimulus intensity.

Stimulus Control

Stimulus control refers to controlling either the applied compressive force or the indenting position of the actuator tip relative to the skin surface. This technique does not vary the distribution of inputs; rather, it varies the distribution of stimulus combinations. If position is the controlled variable, then the indented force will take on whatever value is necessary to ensure proper positioning of the actuator tip. Since skin is viscoelastic and exhibits non-linear *stress-strain* profiles, the force value needed to properly position the actuator tip may vary with time due to creep and stress-relaxation within the skin.

Stimulus Waveform

The stimulus waveform was either non-repeating noise (NRN) or pseudorandom noise (PRN). The distribution of inputs for non-repeating waveforms was purely Gaussian, while the distribution for pseudorandom inputs is approximately Gaussian. Non-repeating noise sequences have the advantage of covering all possible input combinations that occur naturally. With non-repeating noise input, rich waveform stimuli can be applied in a relatively short period of time. Pseudorandom inputs facilitate verification of the logistical regression technique and allow one to evaluate the effects that the time varying properties of skin have upon the receptor response. The analogy in epidemiology studies is to ask the same person the same set of questions repeatedly in order to compare the responses over time.

Sinusoidal or ramp inputs are not used because they suffer from the phenomenon known as phase-locking, as will be discussed subsequently.

Actuator Tip Area

The diameter of the stimulus tip contacting the skin was varied in order to change the physical manifestation of the *indented force*. A tip with a large, flat circular area creates non-localized compressive stress. A tip with a small circular area creates both localized compressive and shearing stress. Since the experimental setup does not allow the compressive, tensile, and shearing *stress* levels within the skin to be measured directly, it is conjectured that if varied sensitivity is detected for each stimulus tip, the receptors may respond preferentially to one type of induced *stress* or differently to each type of induced *stress*. The tip areas used in this experiment are summarized in Table 4.1. All tips were circular and planar where they contacted the skin surface.

Tip	Diameter (mm)	Area (mm^2)
T1	2.3	4.2
T2	4.6	16.6
T3	6.6	34.2

TABLE 4.1: STIMULUS TIP DIMENSIONS

Stimulus Intensity

The intensity of the stimulus was varied from the qualitatively established threshold to 8 times this threshold as established through voltage gain levels controlling the actuator. The higher intensity stimulation increases the number of action potentials that are elicited while the nerve is in its relative refractory period, allowing for a more concise determination of the length of the refractory period.

Other Parameters

Variables that could, but were not, varied directly during in this experiment include:

- Receptor type: Rapidly adapting only
- Substrate compliance: Non-compliant only
- Stimulus waveform bandwidth: Band-limited to 0.5 – 80 Hz

Rapidly adapting afferents were chosen due to the ease with which they can be identified and their high incidence rates within rat hairy skin. The number of identified and cleanly recorded slowly adapting receptors was small.

The substrate was selected to be non-compliant because the recorded *position* and *force* values could be linearly transformed into *strain* and induced *stress* values. If the substrate were compliant, this transform would no longer be linear. As will be indicated subsequently, the ability to convert normalized *stress* and *strain* back into their absolute levels is significant.

The stimulus waveforms were band-limited from 0.5 – 80 Hz. Rapidly adapting afferents are preferentially sensitive to frequencies at the higher end of this range.

4.1.4 Nerve Recordings

A coordinate plane must be established in order to understand the physical expression of the collected data. As shown in Figure 4.3, positive stimulus components are recorded as the actuator indents the stimulus tip into the surface of the skin. The sampling rate for

these recordings is fixed at 2 KiloHertz and a typical record lasts for 30 seconds (60000 sample points). The *position*, *force*, and nerve response signals are represented as 16-bit values corresponding to the voltage levels recorded as feedback from the actuator. These recordings were stored in binary datafiles on the personal computer. All sampling was completed with custom software and analog-to-digital converters. Additionally, the support substrate is non-compliant.

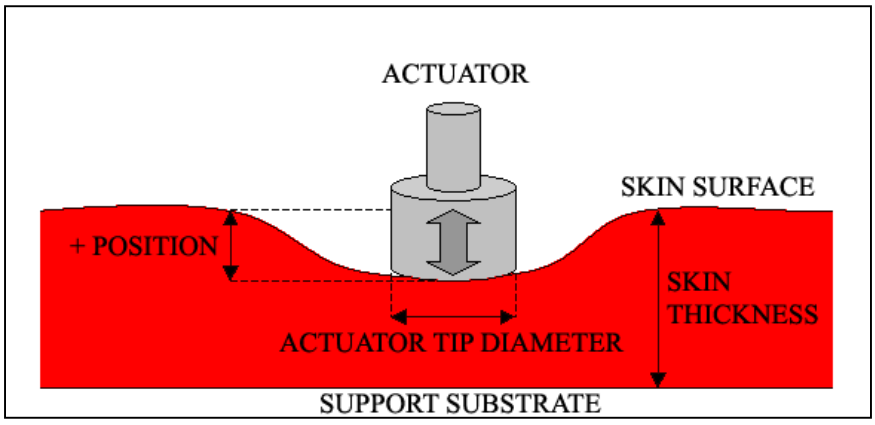


FIGURE 4.3: SYSTEM COORDINATE PLANE

In the datafiles, positive *position* is represented as the change in the actuator from its resting state on the surface of the skin to its instantaneous indented *position*. This value is converted numerically into *strain* by dividing the change in position by the measured skin thickness. Positive indented *force* is calculated as the change in the recorded *force* from the actuator's resting state on the surface of the skin to its instantaneous indented applied *force*. This value is converted into *stress* algorithmically by dividing it by the cross-sectional area of the actuator tip. The time derivatives of both *stress* and *strain* are calculated by averaging two rates of change. The first rate of change is calculated with the instantaneous recorded level of a variable and the level of that variable 1 sample prior. The second rate of change is calculated from the instantaneous recorded level of a variable and the level of that variable 2 samples prior. In pseudocode, the equations are:

$$DSTRESS/DT[I] := ((STRESS[I] - STRESS[I-1])/0.5 + (STRESS[I] - STRESS[I-2])/1)/2 \quad \text{[Eq. 39]}$$

DSTRESS/DT Units = KiloPascals/millisecond

$$DSTRAIN/DT[I] := ((STRAIN[I] - STRAIN[I-1])/0.5 + (STRAIN[I] - STRAIN[I-2])/1)/2 \quad \text{[Eq. 40]}$$

DSTRAIN/DT Units = milliseconds^{-1}

The calculated values of *stress*, *strain*, *dstress/dt*, and *dstrain/dt* are then normalized to a mean of 0 and a standard deviation of 1.

4.2 The Mechanoreceptor Model

The receptor population being studied in this experiment exhibits non-linear behavior such that the instantaneous system output is not purely a function of the instantaneous system input. This non-linear behavior results from a number of biological and mechanical mechanisms:

- Adaptation to prolonged stimuli
- Absolute and relative refractory periods
- Phase locking with periodic input
- Receptors suspended in a viscoelastic medium
- Variable action potential propagation velocity

This section will describe this behavior and the subsequent section will address these issues in the context of a logistical regression analysis.

4.2.1 Adaptation

Response adaptation is biologically significant because it prevents the nervous system from being saturated with information about relatively insignificant matters such as the touch and pressure of clothing.

Rapidly adapting (RA) afferents primarily encode stimulus rate by responding with a high frequency burst of action potentials at the onset of a stimulus. The magnitude of this stimulation is encoded in the frequency of the action potentials. If the load is continuous but static, the nerve adapts and the frequency of impulse generation decreases. However, any variation in the load will again prompt a volley of action potentials, followed by the same adaptation pattern [8,36,43,44,47,48].

Slowly adapting (SA) afferents are preferentially sensitive to static stimuli and respond to a constant compressive load with periodic impulses at a low, relatively constant rate roughly proportional to the load. The period of adaptation is longer relative to that observed in rapidly adapting afferents, hence the designation “slowly adapting” [8,36,43,44,47,48].

In this experiment, RA afferents were targeted because they respond more readily to dynamic stimuli, minimizing the effects of adaptation, and because of their ubiquitous presence in the skin of Sprague-Dawley rats.

4.2.2 Refractory Periods

When a nerve fires, an ionic depolarization gradient across the cell membrane propagates down the nerve axon from the point of transduction. This wave of depolarization is called the action potential. Immediately following this impulse, the receptor enters a repolarization phase in which sodium and potassium ions actively diffuse across the axon membrane in order for the nerve to return to its resting state. After repolarization, the nerve enters a hyperpolarized state in which the threshold is elevated. Following hyperpolarization, the stimulation threshold of the nerve returns to a steady-state.

The refractory period of a nerve exhibits both an absolute and a relative refractory period. During the absolute refractory period, the receptor will not respond to any level of stimulation. This period exists after an action potential has been generated up until repolarization occurs. During the relative refractory period, the receptor will respond to elevated levels of stimulation as compared to the steady-state of the receptor. This period exists after repolarization has begun and lasts until the resting cell potential is restored. The relative refractory period is characterized by the nerve, at first, responding to elevated threshold stimuli, followed by an exponentially decreasing threshold sensitivity until the steady-state threshold is reached at the end of the refractory phase.

Each receptor has a unique time constant that characterizes the transition from the beginning to the end of the relative refractory period. The mechanoreceptors targeted in this study have a refractory period (combined relative and absolute) of about 25 milliseconds. See Figure 4.4 below for an illustration of a nerve refractory period.

In a system in which elevated levels of post-spike stimulation elicit action potentials, the challenge is the treatment of data corresponding to this behavior. To include refractory response data could mean biasing the response of the model towards an elevated threshold. Various modeling approaches to address this issue are presented later in this chapter.

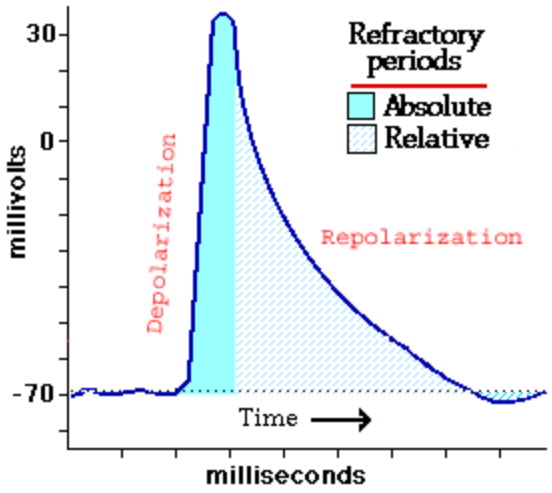


FIGURE 4.4: ABSOLUTE AND RELATIVE REFRACTORY STAGES OF A NERVE

This figure represents the voltage differential that is recorded across a nerve axon membrane when an action potential propagates through the region of the axon being recorded. The transition from the absolute refractory phase to the relative refractory phase occurs just after repolarization begins, or when active transport causes the ions that diffused across the axon membrane to cause the voltage differential to cross the membrane again back to their original locations. In a typical nerve at room temperature, each tick mark on the time axis represents approximately 5 milliseconds.

4.2.3 Phase Locking

Phase locking refers to an invariant, steady-state response to periodic stimulation. When mechanoreceptors are stimulated with a sinusoidal stimulus, impulses consistently occur at specific phases in the period of the stimulus. This is a result of stimulation occurring in the recovery phase of a receptor during which time the threshold of the receptor is elevated. When the stimulus reaches this elevated threshold, an impulse is generated followed by the same period of recovery. Over a multitude of repetitions, this threshold recovery causes a recurring steady-state response at a particular phase in the stimulus waveform. Phase locking should be avoided because it represents a decoupling of receptor response from absolute stimulus magnitude and may cause the receptor to appear to be sensitive to elevated stimulus levels. Additionally, since the mechanoreceptor system has a latency between stimulus at threshold and the observed nerve response when represented as time-series data, a waveform that is periodic also has a time derivative that is periodic, which makes it difficult to ascertain whether the receptor is sensitive to a stimulus metric or the time-derivative of that stimulus metric. See Figure 4.5 below for a diagram illustrating the phenomenon of phase locking.

In this experiment, non-deterministic (i.e., pseudo-random and non-repeating) noise sequences were used as stimulus input waveforms in order to mitigate the effects of phase locking behavior in mechanoreceptor afferents. Previous studies have shown that phase locking is alleviated by non-deterministic stimulus control [17,18,21,36-39]. In addition, non-deterministic waveforms represent a more natural stimulus as compared to deterministic waveforms because a continuum of stimulus states is applied.

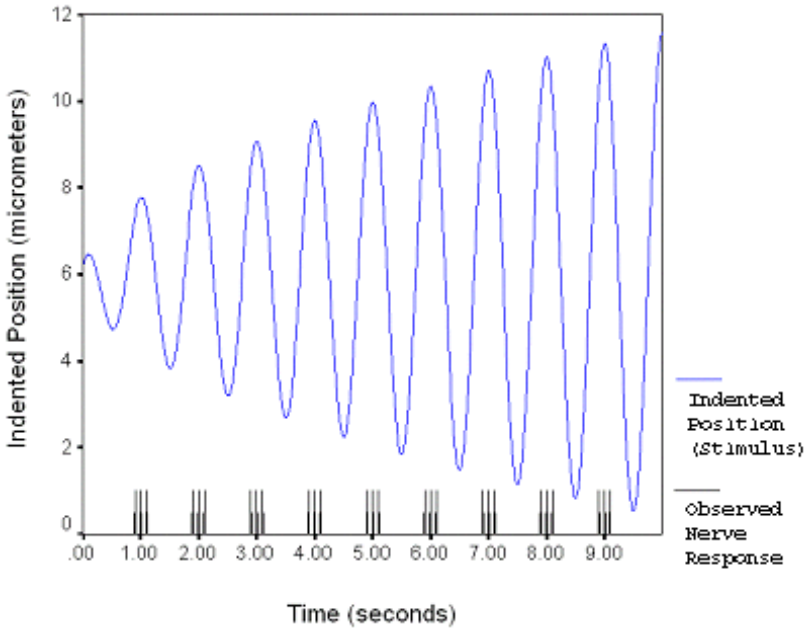


FIGURE 4.5: EXAMPLE OF A PHASE LOCKED NERVE RESPONSE TO SINUSOIDAL STIMULATION

This sample plot shows that when stimulated with periodic (in this case sinusoidal) input, a phase locked nerve will respond periodically in spite of increasing or decreasing stimulus intensity. Since it is widely accepted that tactile nerves respond differently with different stimulus intensity, phase locking represents a decoupling of the receptor response from the intensity level of the stimulus.

4.2.4 Viscoelastic Medium

The mechanoreceptors being studied in this experiment are suspended in the dermal and epidermal layers of the skin as shown in Figure 3.1. Skin is an inherently viscoelastic substrate. Viscoelastic materials exhibit both elastic and time-dependent deformation under loads, resulting in a non-linear *stress-strain* relationship and hysteresis as shown in Figure 4.6 below. Skin, and viscoelastic material, in general, exhibits the following characteristics:

- Stress relaxation – Skin exhibits decreasing levels of internal *stress* with time for a fixed initial indentation or expansion (i.e., constant *strain*).
- Creep - Skin exhibits continuing deformation with time under static loads (i.e., fixed *stress*).
- Creep recovery – Skin will recover from creep-induced deformation with time after the load is released.

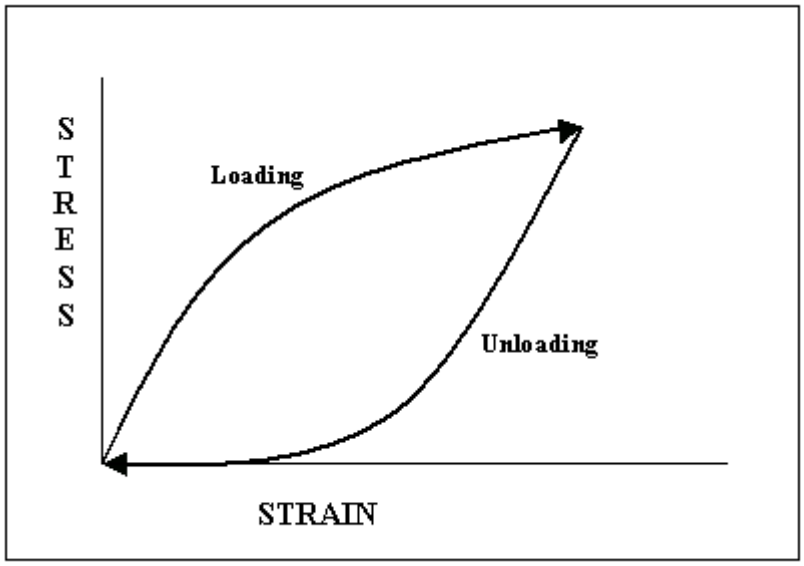


FIGURE 4.6: SAMPLE NON-LINEAR STRESS-STRAIN PROFILE (HYSTERESIS)
This chart exemplifies the stress-strain profile of viscoelastic skin when subjected to displacement controlled loading and unloading.

Creep and stress relaxation are relevant within the context of this study because with prolonged stimulation, these factors modify the physical manifestation of the load upon the receptor within the skin. This issue will be addressed by comparing the afferent behavior of the skin both before and after prolonged stimulation. It is assumed that the skin will have reached a steady-state deformation after prolonged stimulation.

During an experiment, creep recovery occurs after the stimulus is removed from the skin for an extended period of time. Recovery time is allowed between stimulation trials so that creep recovery has ample time to convert the skin back to its original dimensions and properties.

Figures 4.7, 4.8, and 4.9 show the interactions of indented-position/applied-force, applied-stress/indented-strain, and the time derivatives of applied-stress/indented-strain over time, respectively. The variable time difference between the peaks (example: *stress* vs. *strain* peaks in Figure 4.8) within each graph give an indication of the degree to which transversely compressed skin exhibits hysteresis.

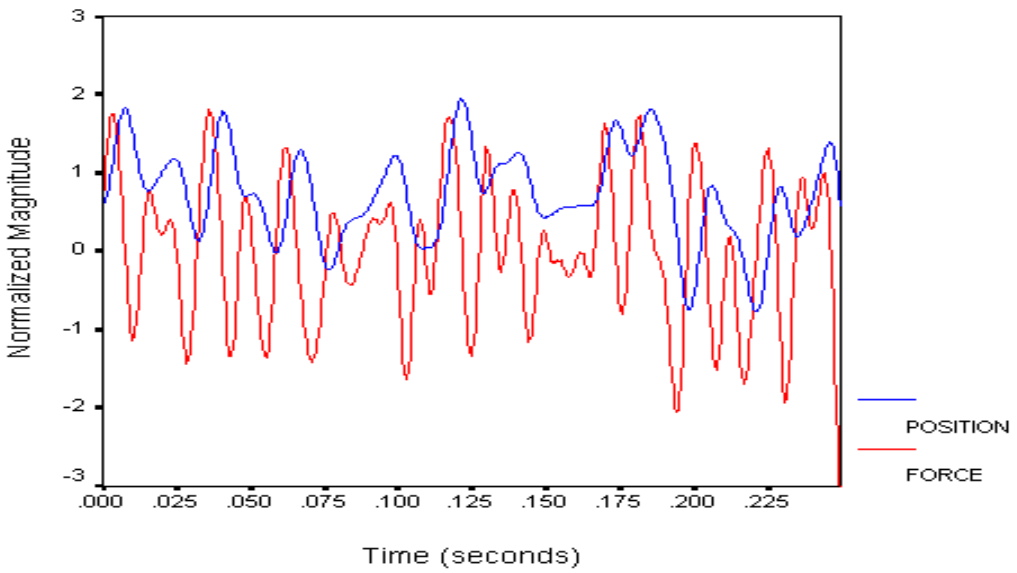


FIGURE 4.7: NORMALIZED POSITION AND FORCE PROFILES OVER TIME

This graphs shows that indented-position and force transversely applied to the skin exhibit a non-linear interaction. The values of position and force were normalized (zero mean, unit variance) before being plotted.

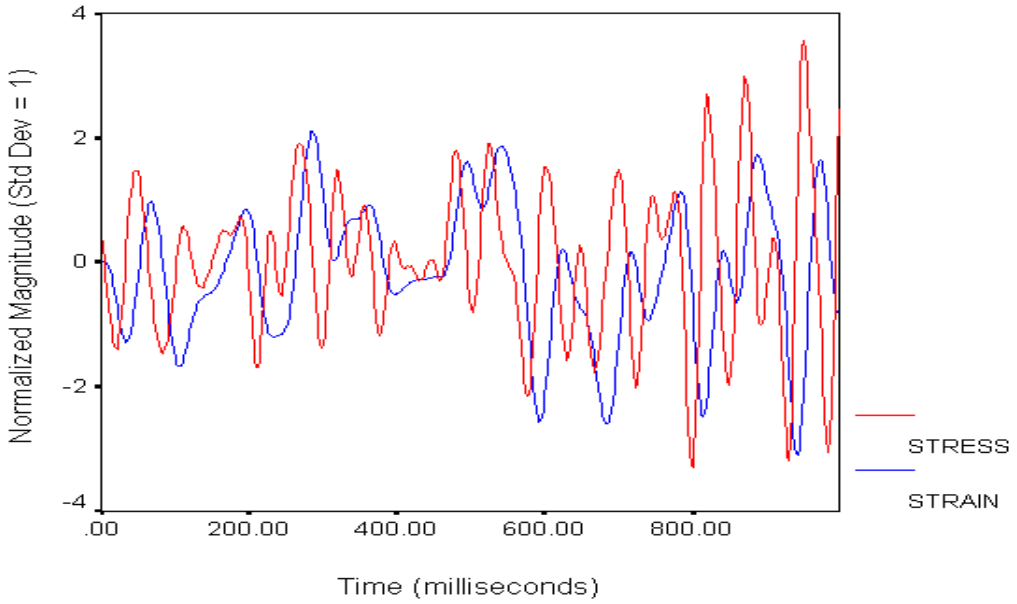


FIGURE 4.8: NORMALIZED STRESS AND STRAIN PROFILES OVER TIME

This graphs shows that indented-stress and indented-strain transversely applied to the skin exhibit a non-linear interaction. The values of stress and strain were normalized (zero mean, unit variance) before being plotted.

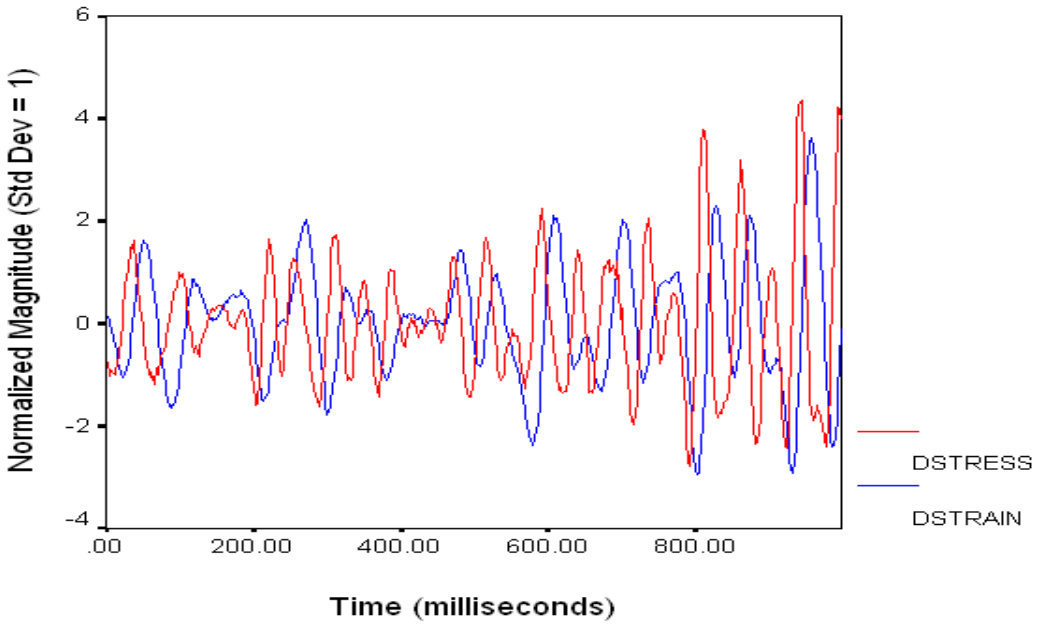


FIGURE 4.9: NORMALIZED DSTRESS/DT AND DSTRAIN/DT PROFILES OVER TIME
This graphs shows that the time derivative of indented-stress and the time-derivative of indented-strain transversely applied to the skin exhibit a non-linear interaction. The values of dstress/dt and dstrain/dt were normalized (zero mean, unit variance) before being plotted.

4.2.5 Propagation Delay

The speed at which an action potential propagates is a function of the diameter of the axon transmitting the signal, the ambient temperature, and the degree to which the axon in myelinated.

Propagation delays must be accounted for since they corrupt the timing relationship between the stimulus and the action potential resulting from that stimulus. The propagation latency can be estimated by using a step function to control indented position or force and measuring the time between the introduction of the stimulus and the recorded action potential. However, since it is not clear exactly which component of the stimulus causes the threshold to be exceeded and since components of compressive stimuli do not propagate through the skin uniformly due to the viscoelastic nature of skin, the latency introduced can only be estimated. This issue will be addressed in subsequent sections.

4.3 Logistical Regression Analysis

A multivariate logistical model of the system was chosen because the system contains continuous input variables and a singular binary output variable corresponding to an "all-or-nothing" nerve action potential. This section will examine the modeling assumptions of logistical regression and illustrate techniques for addressing these assumptions in the context of a model for the afferent behavior of mechanoreceptors subjected to dynamic transversely-applied compressive loads.

4.3.1 Addressing Modeling Assumptions

Logistical regression makes certain assumptions regarding the data under scrutiny in addition to conventional regression assumptions. These assumptions must be met in order to draw statistical inferences from the sample population to the population at large. These assumptions are:

1. The probability of generating a response given the system inputs follows a logistic distribution (in other words, the logit of $\pi(x)$ is linear as shown in Equations 3, 4, and 6)
2. The distribution of error is binomial
3. Each observation is independent

The first two assumptions are met by the mechanoreceptor system; however, the last assumption is not met directly. This assumption will be discussed below.

Observation Independency

Logistical regression assumes that every dependent variable and its corresponding independent variable(s) represent an independent observation. That is, a particular level of an independent variable or combination of independent variables will always produce the same result for the dependent variable no matter what other observations are included in the model or what order the observations were made.

The requirement of observational independency is not met by the time series data used in this experiment. In this experiment, an observed impulse may be the result of stimulus reaching threshold several sample periods prior to the sample period in which the impulse was observed. In contrast to epidemiology studies in which an observation corresponded to a single input combination, the input corresponding to a time series event is spread out in time across a multitude of consecutive, non-independent samples. Furthermore, if an observation falls within the refractory period of a nerve, elevated stimuli, which would elicit a response if observed outside a refractory period, may or may not evoke an action potential.

There are two techniques that can be employed to redress the issue of non-independence in time-quantized observations. The first involves generating and using post-impulse data points as categorical independent variables. The second involves calculating odds ratios using independent variable combinations that correspond to dependent variables prior to the observed dependent variable indicating an impulse. In both of these techniques, the independent variables must all be normalized to the same mean and standard deviation prior to fitting the model. The significance of normalization will become apparent as the methods are explained.

Lag Variable Correlations

Nerve responses are followed by an absolute and relative refractory period, during which time, no level of stimulation and elevated stimuli (exponentially decreasing to resting threshold) evokes an action potential, respectively. In order to mitigate the effects of this system feedback, post-spike sample points offset in time from the observed impulse were assigned positive (binary 1) outcomes and added to the model as categorical explanatory variables. These dummy variables are called "lag" variables since they represent a stimulus state that occurred some number of sample periods after the observed impulse.

When lag variables are included in the model, their odds ratios can be used to assess the correlation between their associated post-impulse sampling period and the impulse for all observed impulses in the data set. Subsequently, one can plot the odds ratios for each lag variable as a function of its offset from the observed impulse. This graph will look similar to the one shown in Figure 4.10. Post-impulse data samples with odds ratios below 1 are protective against impulses and can be discarded without significantly affecting the overall correlations predicted by the model. All sampling periods up until the sampling period with an odds ratio greater than 1 occur within the refractory period of the nerve, hence the protective correlation (inversely correlated: $0 < \text{odds ratio} < 1$) indicated by the odds ratio for the given sampling period. This technique is significant because it predicts statistically which post-impulse sample periods exist within the refractory period of the nerve [17].

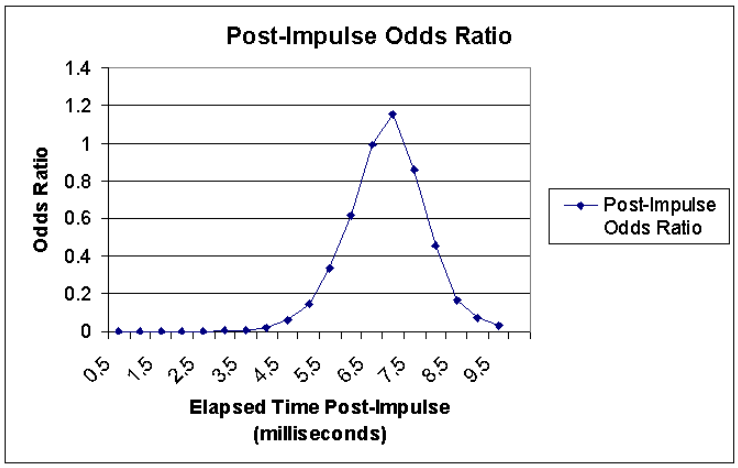


FIGURE 4.10: LAG VARIABLE CORRELATION

All data samples that fall between the observed spike (time = 0 in this graph) and point at which the odds ratio becomes larger than 1 (time = 7 milliseconds in this graph) fall within the nerve's refractory period. These odds ratios were calculated at a standard deviation of 0.

Memory Effects

Due to the non-linear nature of skin and the action potential propagation latency, an action potential may correspond to an above-threshold stimulus state that occurred several time samples prior to the sample during which the action potential was observed. In other words, the system has memory. To address this issue, pre-impulse independent variable combinations can be assigned positive (binary 1) dependent variable outcomes. These dummy dependent variables represent an impulse occurring some number of sample periods prior to the observed impulse. When a model is fitted to these data with the dummy variables as dependent variables, a plot of odds ratio versus sample period prior to the observed spike can be generated for each independent variable. When a particular component of the stimulus reaches threshold N samples before the observed impulse, the odds ratio for this stimulus component evaluated at a standard deviation of 0 will peak N samples prior to the sample during which the impulse was observed. This approach effectively decouples the components of the stimulus (independent variables) across samples [17].

In this experiment, odds ratios were systematically calculated for positive dependent variables offset in time up to 100 milliseconds prior to an observed impulse. The resulting graphs indicated the strength of the interaction between the stimulus component and the response as well as provided a quantification of the memory effects inherent in the system. A sample graph is shown in Figure 4.11.

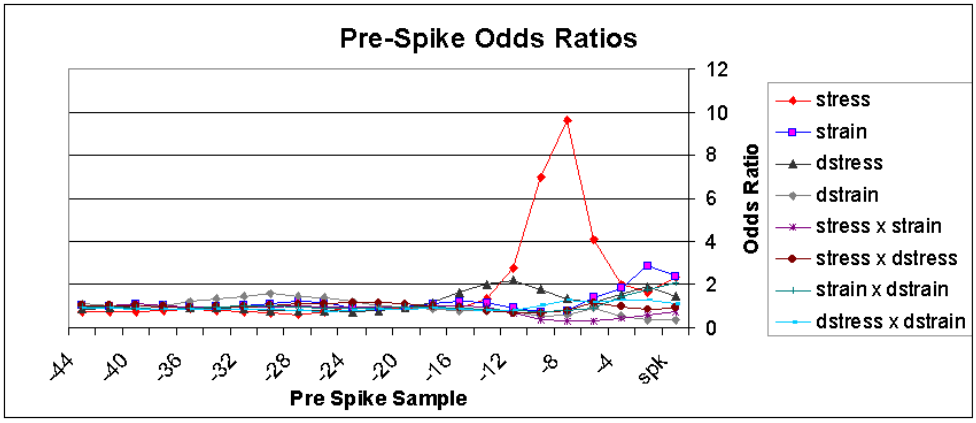


FIGURE 4.11: SAMPLE ODDS RATIOS VS. PRE-SPIKE TIME IN SAMPLES
1 Sample = 0.5 Milliseconds. This graph indicates that the nerve in question shows sensitivity to compressive stress. These odds ratios were calculated at a standard deviation of 0.

Interaction Term Effects

Since interaction terms are used in the logistical regression model to express the relationship between confounded independent variables, it is important to evaluate the effect of these interaction terms at varying levels of stimulation in order to answer the question: Are the effects of confounding manifested at high levels of stimulation?

The model being used in this experiment is as follows:

$$Y = \frac{e^{\alpha}}{1+e^{\alpha}} \qquad \text{[Eq. 41]}$$

where:

- $\alpha = \beta_0 + stress\,\beta_1 + strain\,\beta_2 + \frac{dstress}{dt}\beta_3 + \frac{dstrain}{dt}\beta_4 + stress*strain\,\beta_5 + stress*\frac{dstress}{dt}\beta_6 + strain*\frac{dstrain}{dt}\beta_7 + \frac{dstress}{dt}*\frac{dstrain}{dt}\beta_8$

This equation is condensed into:

$$Y = \frac{e^{\beta_0+\sigma\beta_1+\varepsilon\beta_2+\frac{d\sigma}{dt}\beta_3+\frac{d\varepsilon}{dt}\beta_4+\sigma\varepsilon\beta_5+\sigma\frac{d\sigma}{dt}\beta_6+\varepsilon\frac{d\varepsilon}{dt}\beta_7+\frac{d\sigma}{dt}\frac{d\varepsilon}{dt}\beta_8}}{1+e^{\beta_0+\sigma\beta_1+\varepsilon\beta_2+\frac{d\sigma}{dt}\beta_3+\frac{d\varepsilon}{dt}\beta_4+\sigma\varepsilon\beta_5+\sigma\frac{d\sigma}{dt}\beta_6+\varepsilon\frac{d\varepsilon}{dt}\beta_7+\frac{d\sigma}{dt}\frac{d\varepsilon}{dt}\beta_8}} \qquad \text{[Eq. 42]}$$

where:

- $\sigma = stress$
- $\varepsilon = strain$
- $\frac{d\sigma}{dt}$ = time derivative of *stress*

- $\frac{d\varepsilon}{dt}$ = time derivative of *strain*
- $\sigma\ \varepsilon$ = *stress*-*strain* interaction term
- $\sigma\frac{d\sigma}{dt}$ = *stress*-time derivative of *stress* interaction term
- $\varepsilon\frac{d\varepsilon}{dt}$ = *strain*-time derivative of *strain* interaction term
- $\frac{d\sigma}{dt}\frac{d\varepsilon}{dt}$ = time derivative of *stress*-time derivative of *strain* interaction term

Based on this model, the odds ratios for each independent variable are:

$$\psi_{\sigma} = e^{\beta_1+\varepsilon\beta_5+\frac{d\sigma}{dt}\beta_6} \qquad \text{[Eq. 43]}$$

$$\psi_{\varepsilon} = e^{\beta_2+\sigma\beta_5+\frac{d\varepsilon}{dt}\beta_7} \qquad \text{[Eq. 44]}$$

$$\psi_{d\sigma/dt} = e^{\beta_3+\sigma\beta_6+\frac{d\varepsilon}{dt}\beta_8} \qquad \text{[Eq. 45]}$$

$$\psi_{d\varepsilon/dt} = e^{\beta_4+\varepsilon\beta_7+\frac{d\sigma}{dt}\beta_8} \qquad \text{[Eq. 46]}$$

Each of these odds ratios is a function of the Beta (β) coefficient associated with the particular independent variable as well as the independent variables with which the particular independent interacts. Therefore, the odds ratios are not constant. Rather, they are a function of the levels of the interacting independents that are observed for the given level of the particular independent. In other words, due to the non-linear nature of skin, varying levels of *strain*, *dstress/dt*, and *dstrain/dt* can be observed for a particular level of *stress* and the effects of these varying combinations are a non-constant odds ratio with varying levels of *stress* because of the included interaction terms. This affect will become apparent in the next several sections as sample logistical regression analyses are performed.

4.3.2 Sample Logistical Regression Analysis

This section first describes the steps necessary for a logistical regression analysis of the data collected in the manner detailed in Section 4.1. After the steps in the analysis are explained, a sample analysis is performed using:

A) Artificial data that represents receptor behavior
B) Actual recordings from compressively stimulated receptors

The purpose of the first sample analysis is to validate the suitability of the logistical regression model by assessing the model's ability to predict known nerve sensitivities.

The purpose of the second sample analysis is to contrast the model of the artificial nerve data to the model fitted to actual experimental recordings and to prepare the reader for what follows in the *Results and Analysis* section.

The steps that were performed for the multivariate logistical regression analysis are as follows:

Step 1: Determination of Experimentally Pertinent Covariates and Interaction Terms for the Logistic Function

The first step in logistical regression is to determine which factors are to be included in the model. Both the applied compressive *stress* and the indented *strain* are included in the model, as well as their time derivatives, based on models used in similar studies [17].

As mentioned previously, the objective for selecting interaction terms to include interactions that are both significant and interpretable [17,25]. Under that premise, any variable and its derivative should be included since, in some cases, one is merely a phase-offset of the other and in other cases, the effect of the rate of change of a variable may only manifest itself above certain absolute thresholds [17]. Additionally, *stress* and *strain* should be included as an interaction term since they covary in this experiment. Furthermore, if two covariates warrant an interaction term, then their time derivatives also warrant an interaction term, meaning the term (*dstress/dt* x *dstrain/dt*) should be included. All of these terms are evaluated in the analysis to determine if their impact is significant enough to earn them a place in the final logistic model used in this experiment.

The logistic function, therefore, is:

$$Y = \frac{e^{\beta_0+\sigma\beta_1+\varepsilon\beta_2+d\sigma/dt\beta_3+d\varepsilon/dt\beta_4+\sigma X\varepsilon\beta_5+\sigma Xd\sigma/dt\beta_6+\varepsilon Xd\varepsilon/dt\beta_7+d\sigma/dtXd\varepsilon/dt\beta_8}}{1+e^{\beta_0+\sigma\beta_1+\varepsilon\beta_2+d\sigma/dt\beta_3+d\varepsilon/dt\beta_4+\sigma X\varepsilon\beta_5+\sigma Xd\sigma/dt\beta_6+\varepsilon Xd\varepsilon/dt\beta_7+d\sigma/dtXd\varepsilon/dt\beta_8}}$$

This equation is identical to Equation 42.

Step 2: Conditioning Data

The second step is to condition the data such that it can be imported into commercial logistical regression software. Each covariate is measured by a different metric and when comparing odds-ratios to determine whether a particular variable or interaction term should or should not be included in the model, it is necessary to compare data on the same scale [17]. Therefore, it is necessary to normalize all covariates and interaction terms to zero mean and unit variance before calculating the β values with the commercial regression software. Custom software was implemented to complete the task of conditioning and normalization.

Step 3: Derive Odds Ratios

Since the preliminary model includes interaction terms, it is necessary to derive the odds-ratios for each covariate. Following the explanation found in the *Background*, the odds ratios for the current model are:

$$\psi_{\sigma} = \mathrm{e}^{\beta_1 + \varepsilon\beta_5 + d\sigma/dt\,\beta_6}$$
$$\psi_{\varepsilon} = \mathrm{e}^{\beta_2 + \sigma\beta_5 + d\varepsilon/dt\,\beta_7}$$
$$\psi_{d\sigma/dt} = \mathrm{e}^{\beta_3 + \sigma\beta_6 + d\varepsilon/dt\,\beta_8}$$
$$\psi_{d\varepsilon/dt} = \mathrm{e}^{\beta_4 + \varepsilon\beta_7 + d\sigma/dt\,\beta_8}$$

Each of these odds-ratios is a function of two other covariates due to the inclusion of interaction terms in the model. These equations are identical to Equations 43 – 46.

Step 4: Qualitative Assessment of the Fit of the Data to a Logistical Model

The fourth step was to plot the logistic functions of each covariate. This can be completed by grouping the normalized magnitude of the covariate in question into bins and calculating the odds of a spike occurring in that bin. The plot of the odds of a nerve response occurring versus the average magnitude of the stimulus for each stimulus bin is the logistic function for that particular covariate. From this plot, one can determine graphically how well the data fit a logistic distribution and hypothesize as to which covariates will have a significant impact on the response based on how well the curve fits the idealized logistic curve.

Step 5: Calculate β Coefficients and Corresponding Odds Ratios

Step 5 is to determine the β coefficients for the experimentally established model and use these β values to calculate the odds ratios for each covariate. The β coefficients can be calculated by maximizing the likelihood function, as explained in the *Background* (or by minimizing the log-likelihood function) [16,25,31,40]. The interpretation of the β coefficients is quite straightforward in the absence of interaction terms. As the absolute value of the β coefficient increases, the covariate or interaction term correlates more strongly with the outcome; however, if the β value is becoming more negative, it indicates a covariate or interaction term that contributes negatively to the outcome. If the absolute value of the β coefficient is small, it indicates little to no correlation between the covariate or interaction term and the outcome. Since interaction terms are present, however, the interpretation is not quite as simple. Interaction terms often overcome the

predisposition of the β values because they are a function of other variables and as those other variables increase in magnitude, the effect of the interaction term is much more pronounced [25]. Such is the case with the model developed for this study; therefore, this study concentrated on a more meaningful quantitative measure of the significance of a covariate: the odds ratio.

The odds ratio is calculated using the maximum likelihood estimates of the β coefficients. As discussed in the *Background*, the odds ratio represents a quantitative measure of the degree to which the outcome variable changes for a single-unit change in the explanatory variable being tested. Since the data have been normalized to a constant standard deviation, the odds ratios can be graphed on the same axes and compared.

Step 6: Assessing the Accuracy of the Fitted Logistical Model versus Observed Data

The sixth and final step involves assessing the accuracy of the established model. In general, the assessment of the fit of the model is completed with a multitude of statistical techniques, some of which were discussed in the *Background* and the rest will be introduced in the following sample logistical regression analyses as well as the *Results and Analysis* section.

Sample Logistical Regression Analysis: Artificial Data

In order to validate the logistical regression model for tactile receptors, it is useful to manufacture and fit data sets that exemplify the behavior of a mechanoreceptor system under dynamic compressive loads. Neural discharges can be artificially correlated with any one of the independent variables at a given number of samples prior to the observed spike. Additionally, an artificial refractory period can be added. Assessing the accuracy of the fitted model resolves into comparing the artificially imposed variables to the fitted variables.

The first step in creating these data sets is to generate pseudorandom and non-repeating noise sequences similar to those used to control the voltage levels of the actuator when experimental data are collected. The non-repeating noise sequences used in the experiment are generated in real-time by a band-limiting noise generator. The pseudorandom noise sequences are simply 0.5-second recordings of the non-repeating noise that are looped and used to drive the actuator. Random noise was artificially superimposed on these values to mimic the conditions under which these values are recorded. Subsequently, these voltage levels were then linearly transformed into absolute *strain* and *stress* values.

After artificially selecting a component of the stimulus to which the impulse will be correlated with and a refractory period, the data are parsed into a form that is compatible with the logistical regression software packages being used for data analysis. Essentially, stimulus combinations are searched linearly in time for any supra-threshold levels. A

supra-threshold stimulus exists when the integral of an exponentiated linear sum of each stimulus component multiplied by its weighting factor crosses a fixed value. The weighting factors are determined by the chosen component of the stimulus. In pseudo-code, the algorithm is as follows (Algorithm 1):

```
EXP1 := e^(-1.0/50);                                                    // ~ 0.98
EXP2 := 1.0 – EXP1;                                                     // ~ 0.02
SUM := 0;
IMPULSE := FALSE;
FOR STIMULUS COMBINATIONS I := 1 TO N {
        VAR1 := STRESS_GAIN x STRESS[I] + STRAIN_GAIN x STRAIN[I] +
                DSTRESS_GAIN x DSTRESS[I] + DSTRAIN_GAIN x DSTRAIN[I] +
                STRESSXSTRAIN_GAIN x STRESS[I] x STRAIN[I] +
                STRESSXDSTRESS_GAIN x STRESS[I] x DSTRESS[I] +
                STRAINXDSTRAIN_GAIN x STRAIN[I] x DSTRAIN[I] +
                DSTRESSXDSTRAIN_GAIN x DSTRESS[I] x DSTRAIN[I];
        DIFF := VAR1 – SUM;
        SUM := DIFF x EXP2 + SUM x EXP1;
        IF (SUM < 0) {
                SUM := 0;
        } ENDIF;
        IF (SUM > FIXED_VALUE) {
                IMPULSE[I + PROPAGATION_DELAY_OFFSET] := TRUE;
                SUM := 0;
        } ENDIF;
} ENDFOR;
```

The fixed value in Algorithm 1 was experimentally derived such that approximately 5% of the stimulus component permutations produced an artificial impulse response when the weighting system was imposed.

Once supra-threshold stimulation is reached, a binary 1 value (corresponding to the observation of an impulse) is added to the spike category of the data file in place of a binary 0. The data value is written to the file some fixed number of samples in time after the observed above-threshold stimulus to account for the inherent propagation delays of the mechanoreceptor system as recorded experimentally. Next, the threshold is set to infinity for the duration of the absolute refractory period. This period is followed by a phase in which the threshold decreases exponentially from infinity down to the initial resting threshold. If, at any point during this period of elevated threshold, a supra-threshold stimulus level is encountered, a binary 1 data point is again added some fixed number of sample points following the sample point containing the supra-threshold stimulus. Following this observation, the threshold is again set to infinity followed by the same exponentially decreasing elevated threshold.

Trial	Actual/Predicted	Response Sensitivity	Refractory Length
1	*Actual*	Strain	5
	Predicted	Strain	4
2	*Actual*	Stress	10
	Predicted	Stress	8
3	*Actual*	Stress x Strain	15
	Predicted	Stress x Strain	14
4	*Actual*	dStrain/dt	10
	Predicted	dStrain/dt	8
5	*Actual*	dStress/dt	5
	Predicted	dStress/dt	4
6	*Actual*	dStress/dt x dStrain/dt	12
	Predicted	dStress/dt x dStrain/dt	10

TABLE 4.2: SAMPLE REGRESSION ANALYSIS OF ARTIFICIAL DATA SETS

Several trials were completed with the results summarized in Table 4.2. The results of Trial 1 are illustrated in Figure 4.12, 4.13, and 4.14. As can be seen from Table 4.2, the logistical regression model accurately detected the sensitivity of the system to a component of the stimulus. The logistical regression model predicted the refractory length with approximately 17% error.

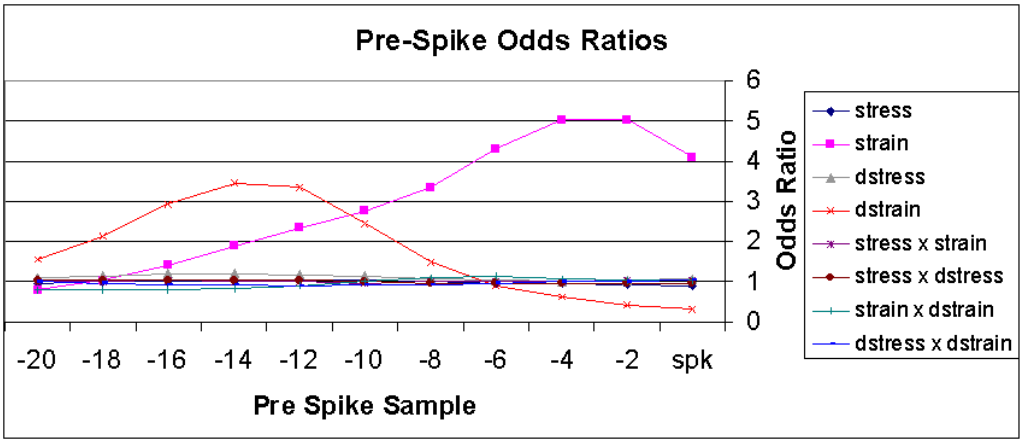


FIGURE 4.12: ANALYSIS OF ARTIFICIAL DATA - TRIAL 1 PRE-IMPULSE ODDS RATIOS
These odds ratios were calculated at a standard deviation of 0.

As is shown in Figure 4.12, the logistical regression model identified the *strain* sensitivity of the artificial receptor and predicted the strongest correlation at 4 samples (2

milliseconds) prior to the observed impulse. A *dstrain/dt* artifact was also detected at an offset of 14 samples (7 milliseconds). This is due to the inherent confounded relationship between *strain* and its derivative. Derivative artifacts of a strongly correlated independent will, in general, be evident up to 12.5 milliseconds before or after the peak correlation of that independent variable because the system inputs are band-limited up to 80 Hz.

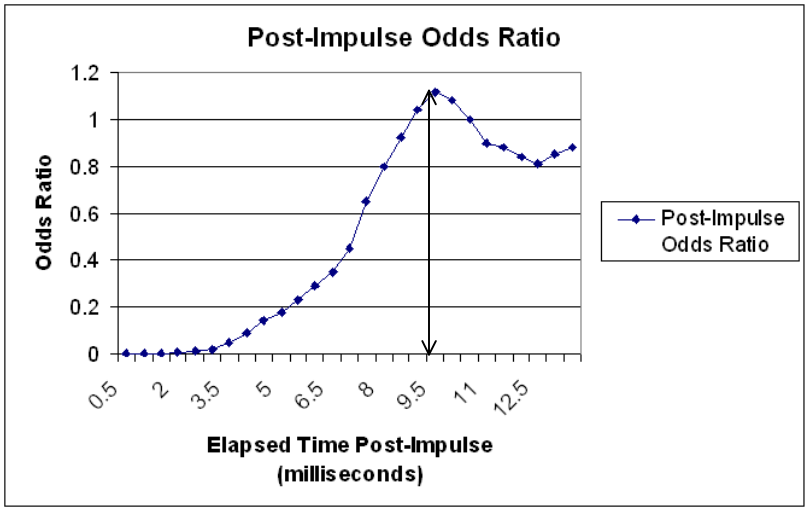


FIGURE 4.13: ANALYSIS OF ARTIFICIAL DATA - TRIAL 1 POST-IMPULSE ODDS RATIOS
These odds ratios were calculated at a standard deviation of 0.

In Figure 4.13, the post-impulse correlations are illustrated. At approximately 10 milliseconds following the observed impulse, the correlation between the dependent variable and the categorical dummy variable associated with that sampling period peaked, indicating a return of the nerve's threshold to a resting state.

In Figure 4.14, odds ratios are plotted versus the magnitude of the stimulus component in standard deviations. This plot was generated by converting the stimulus magnitude into bins and averaging the odds ratio of each stimulus component (calculated with Equations 43 - 46) for all samples of that stimulus component that fall within that particular bin. Since all components of the stimulus (*stress*, *strain*, *dstress/dt*, *dstrain/dt*) were normalized prior to being fitted, this graph indicates how the odds ratios of each component change with stimulus magnitude due to confounded interactions. Due to the fact that the odds ratios are relatively constant across all stimulus magnitudes, confounded interactions have a constant affect across all stimulus magnitudes.

From these experiments, one can see that logistical regression accurately assesses the sensitivity of a compressively stimulated tactile receptor system expressed as time-series input-output observations.

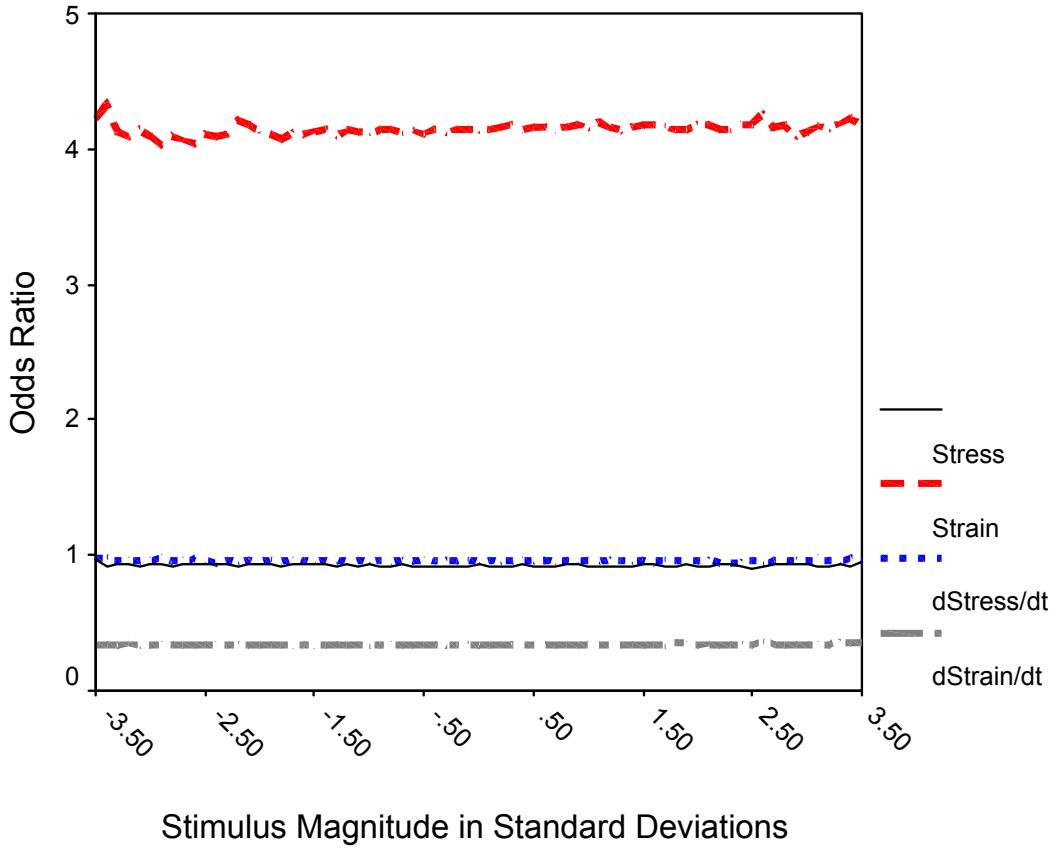


FIGURE 4.14: ANALYSIS OF ARTIFICIAL DATA - TRIAL 1 ODDS RATIOS VS. STIMULUS MAGNITUDE

Sample Logistical Regression Analysis: Collected Data

For the logistical regression analysis involving collected data, a dataset was chosen at random and the steps listed at the beginning of this section were performed. The extended model (described previously by Equations 41 and 42) was fitted using commercial software (SPSS Base and Systat, MINITAB) and the Beta values (β_0 - β_8) were estimated.

The fitted logistic curve is generated as shown in Figure 4.15. This curve represents the probability of an impulse given the stimulus state versus the magnitude of the stimulus. Similarly, the fitted curve can be plotted versus the normalized value of each independent, as shown in Figures 4.16 and 4.17.

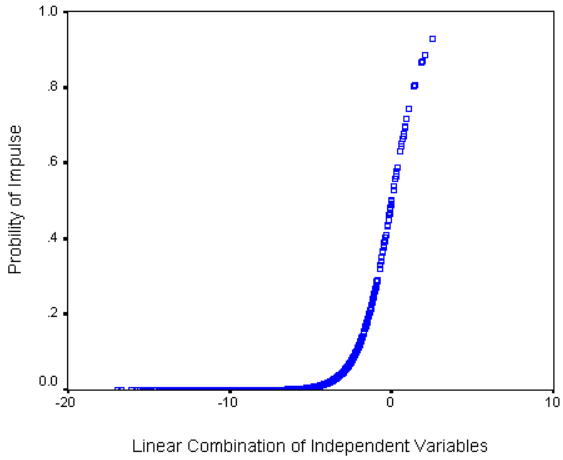


FIGURE 4.15: FITTED LOGISTIC REGRESSION CURVE
Dataset R1221911.DAT [Compressed 5:1]

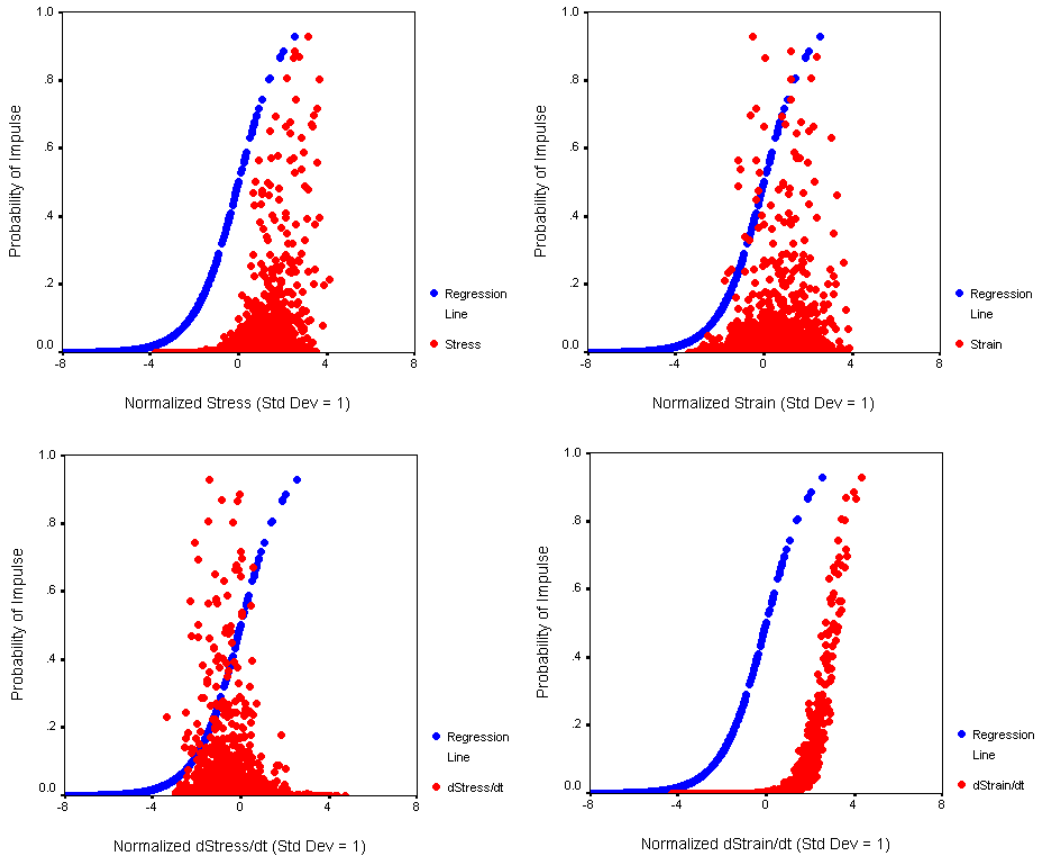


FIGURE 4.16: FITTED REGRESSION LINES
Dataset R1221911 [Compressed 5:1]
Due to limitations in available plotting software, these plots were generated using the first 3000 data samples of the compressed data file. The shift in the fitted line from the distribution indicated in the dStrain/dt data points and, to a lesser degree, in the Stress data points are due to the overall distribution of data points being shifted towards lower stimulus magnitudes (towards negative standard deviations).

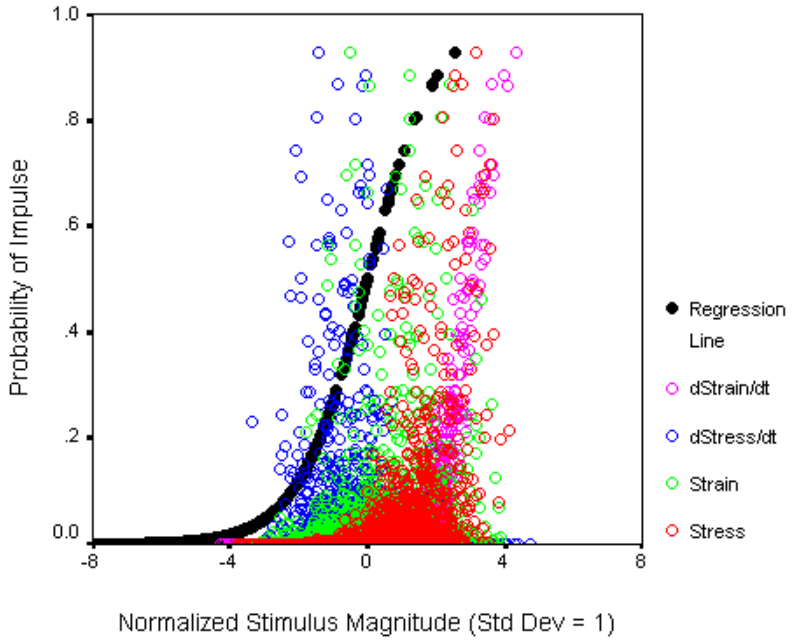


FIGURE 4.17: FITTED LINE
Dataset R1221911 [Compressed 5:1]

The accuracy of the fit of the interaction terms can be assessed visually by contrasting the predicted probability of the model as compared to the observed probability for each interaction term (*stress* x *strain*, *stress* x *dstress/dt*, *strain* x *dstrain/dt*, *dstress/dt* x *dstrain/dt*). These graphs are shown in Figures 4.18 through 4.21. In general, the *stress x dstress/dt* fit is more accurate than the other fits; however, the predicted probabilities of the other interaction terms indicate that the model predicts the spikes relatively accurately but falls victim to over-prediction in some regions outside those with high observed spike probabilities.

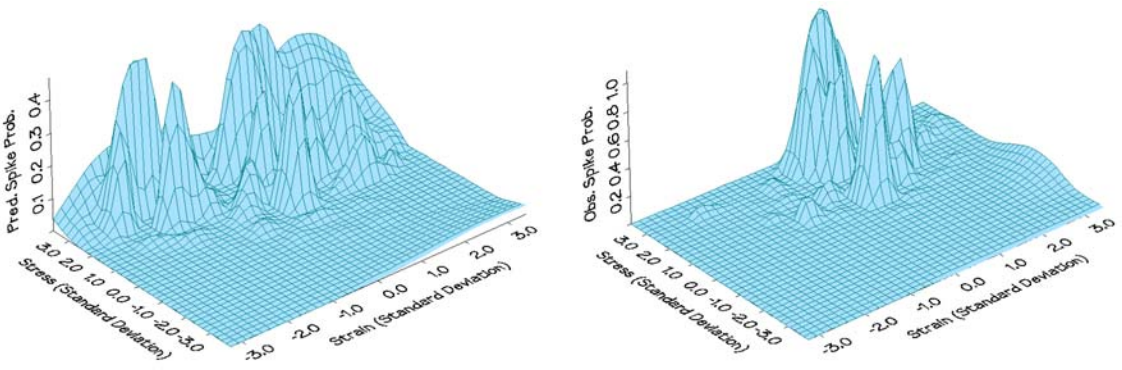


FIGURE 4.18: PREDICTED (LEFT) VS. OBSERVED (RIGHT) SPIKE PROBABILITIES FOR *STRESS-STRAIN* INTERACTION TERM
Dataset R1221911

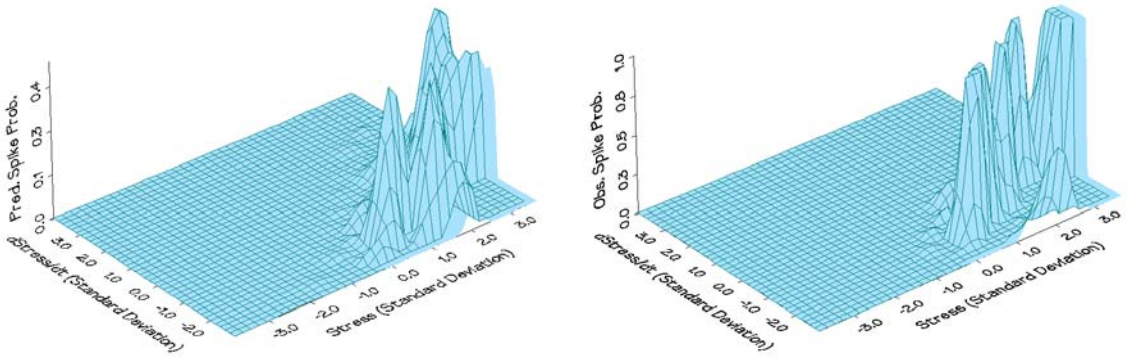


FIGURE 4.19: PREDICTED (LEFT) VS. OBSERVED (RIGHT) SPIKE PROBABILITIES FOR *STRESS-DSTRESS/DT* INTERACTION TERM
Dataset R1221911

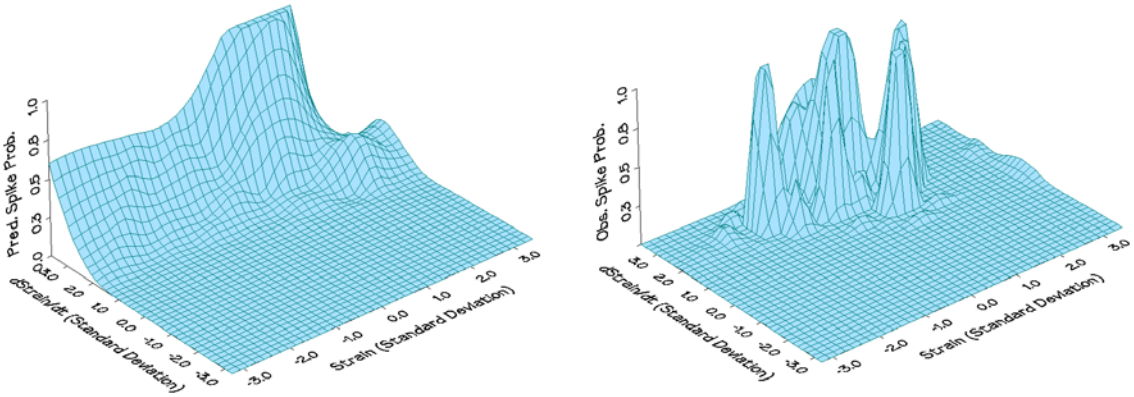


FIGURE 4.20: PREDICTED (LEFT) VS. OBSERVED (RIGHT) SPIKE PROBABILITIES FOR *STRAIN-DSTRAIN/DT* INTERACTION TERM
Dataset R1221911

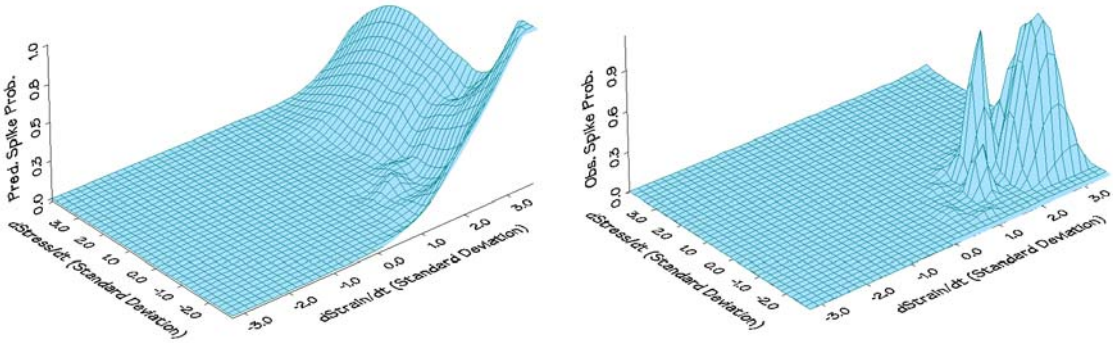


FIGURE 4.21: PREDICTED (LEFT) VS. OBSERVED (RIGHT) SPIKE PROBABILITIES FOR *DSTRESS/DT-DSTRAIN/DT* INTERACTION TERM
Dataset R1221911

The next step in the analysis is to perform a lag variable correlation. This correlation indicates which post-impulse data samples occur within the refractory period of the nerve. These samples will be removed from the data and a second logistical regression will be performed. The reason for the removal of these data points was explained previously: these data points represent non-independent observations and violate a fundamental assumption of logistical regression.

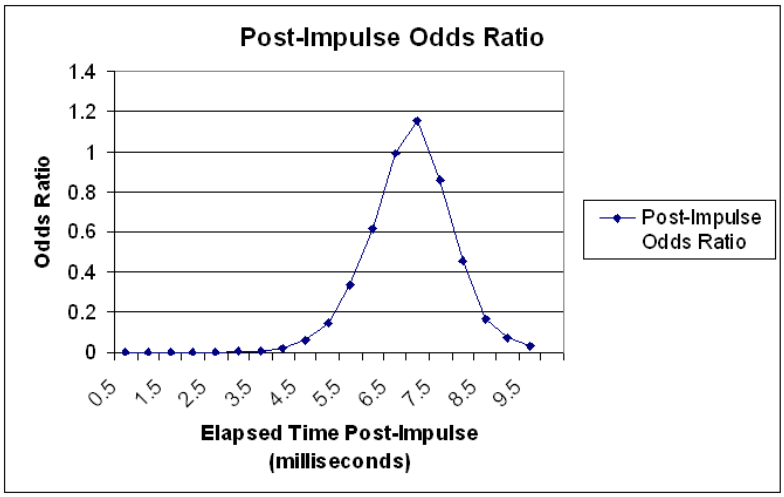


FIGURE 4.22: LAG VARIABLE CORRELATION
Dataset R1221911; 1 Millisecond = 2 Samples
These odds ratios were calculated at a standard deviation of 0.

The results from a lag variable correlation are illustrated in Figure 4.22. The peak of the curve exists at approximately 8 milliseconds (16 samples) post-impulse and indicates the statistical end of the nerve refractory period. The data file is then parsed such that 16 samples following every observed impulse are removed.

Following a lag variable correlation is a pre-impulse correlation for each independent variable and interaction term. This procedure was defined previously. The goal of this procedure is to determine which independent or interaction term has the strongest correlation to the impulse and at what pre-impulse offset. As shown in Figure 4.23, the 0-offset correlation indicates that the nerve has a strong sensitivity to *dstrain/dt*. However, at an offset of approximately 7 samples (3.5 milliseconds), the effects of the *dstrain/dt* sensitivity diminish and *dstress/dt* becomes the component of the stimulus to which the nerve is preferentially sensitive.

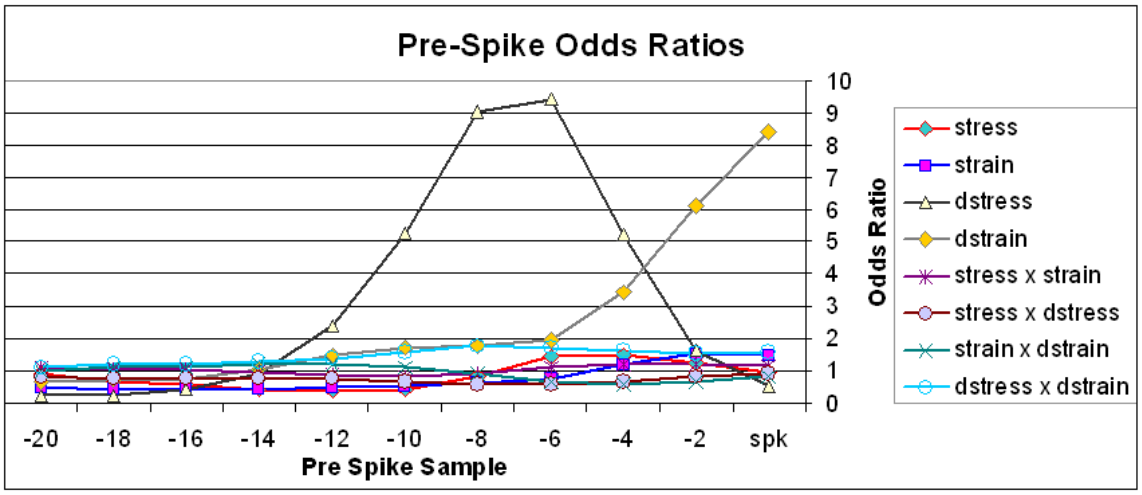


FIGURE 4.23: PRE-SPIKE ODDS RATIOS
Dataset R1221911; 1 Sample = 0.5 Milliseconds
These odds ratios were calculated at a standard deviation of 0.

The next step is to determine the effect that confounded interactions have upon the odds ratios of each stimulus component. The odds ratios for each stimulus magnitude bin were averaged and plotted versus the magnitude of the stimulus corresponding to that bin. The result is shown in Figure 4.24. The graph indicates strong *dstrain/dt* sensitivity across all magnitudes of the stimulus. The slope of the *dstrain/dt* curve is negative, indicating that the interaction between *dstrain/dt* and the other independent variables with which *dstrain/dt* interacts become weaker as the magnitude of the stimulus increases. Since the other independents show a constant odds ratio across all magnitudes of the stimulus, the interactions in which those independents participate do not affect the odds ratio of the independent significantly.

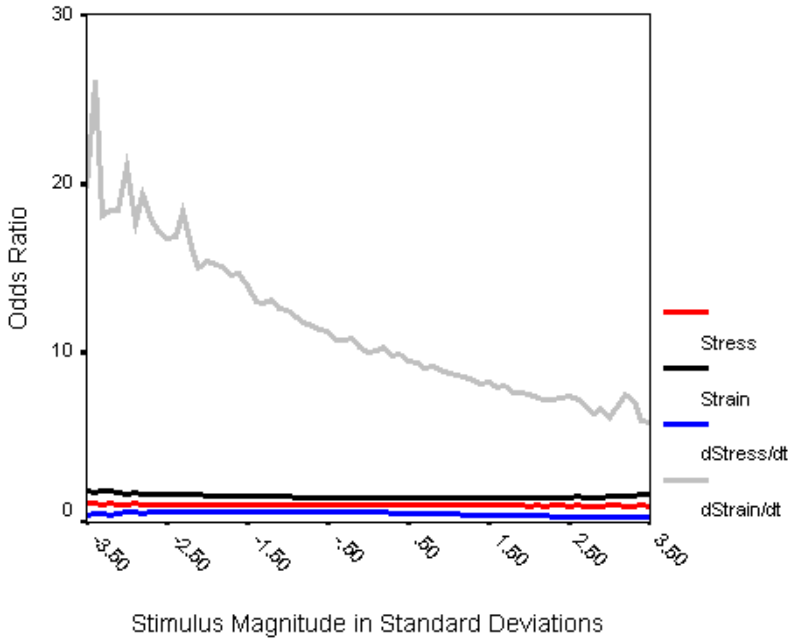


FIGURE 4.24: ODDS RATIOS VS. STIMULUS MAGNITUDE IN STANDARD DEVIATIONS
Dataset R1221911

4.4 Comparison of Data Analysis Techniques

Linear regression is the de facto standard for modeling the interactions between continuous independent variables and a continuous dependent. Linear regression models of dichotomous event probabilities, however, have some significant drawbacks, as noted in Section 3.2 of this report. Linear regression is an appealing technique for modeling dichotomous probabilities due to the direct interpretation of coefficients, but the model violates several fundamental regression modeling assumptions, leading researchers to use alternative models when the dependent variable is dichotomous. These alternative techniques include logit models using logistic or probit functions and discriminant analysis. Logistical regression using a logistic model has been previously contrasted to linear regression in this report for the sake of explanation; however, both probit analysis and discriminant analysis have yet to be addressed. This section provides a brief comparison of the practicality of these methods in a context in which the dependent variable is dichotomous.

Probit analysis is essentially identical to logistical regression modeling, save for the fact that a cumulative normal function is used in place of a logistic function and the coefficients are computationally more difficult to determine. The cumulative normal function has the familiar sigmoid shape synonymous with the logistic function. The only difference exists in the tails of the cumulative normal function. Just as in logistic regression, the estimated coefficients cannot be directly interpreted as the probability of

generating a positive response but can be converted into probabilities with a simple transform [2,12,34].

Discriminant analysis models dichotomous dependent variables by creating an equation to discriminate between two groups. Discriminant analysis predicts the probability of one group versus another by minimizing the cost of misclassification within the groups or by estimating the conditional probability of a particular event within the groups. The latter approach assumes that the multivariate distribution of the independent variables is normal [2,12,34]. The probability of an observation being in one group, given the observed independent variables [analogous to $\pi(x)$ in logistical regression] has the same functional form in discriminant analysis as it does in logistical regression. Computationally, discriminant analysis is more straightforward than logistical regression or probit analysis [2,12,34].

When the independent variables are distributed in a multivariate normal form, logistical regression, probit analysis, and discriminant analysis all produce approximately the same functional model. The most significant difference occurs between logistical regression and discriminant analysis when the predicted probabilities are very close to 0 or 1, in which case, the results of discriminant analysis may be misleading. Conversely, when the data are not distributed multivariate normal as is the case with categorical or discrete independent variables, logistical regression and probit analysis produce statistically consistent estimations for the modeling coefficients while discriminant analysis does not [2,12,34].

Of these methods, logistical regression was chosen for this application because:

- The dependent variable is dichotomous
- Some independent variables are categorical
- The multivariate distribution of the independent variables is not necessarily normal
- Confounded interactions exist within the independent variables
- Logistical regression is computationally simpler than probit analysis
- Commercial software for logistical regression modeling is readily available
- Logistical regression has been previously used for modeling mechanoreceptor systems [17]

4.5 Chapter Summary

In this chapter, a logistical regression framework for modeling afferent tactile mechanoreceptor behavior under dynamic compressive loads was developed. This model addressed the assumptions of logistical regression in an afferent tactile mechanoreceptor system. Subsequently, a systematic experiment was described based on this model. The purpose of the experiment was to determine if rapidly adapting mechanoreceptors are preferentially sensitive to compressive stress and the rate of change of strain in time. The next several sections describe the results of that experiment.

5.0 Results and Analysis

The recordings from 10 rapidly adapting mechanoreceptors from 7 different rats are discussed in this *Results and Analysis* section. One hundred ten trials were completed, 25 of which were force-controlled stimulation and the remainder of which were position-controlled stimulation.

In the first section, a single receptor is profiled with logistical regression techniques. This receptor is analyzed in the same fashion that all datasets were analyzed. After this detailed discussion, comments are made regarding general patterns that were observed in the analysis of all datasets with isolated examples given to illustrate the observations.

In the second section, more general patterns of sensitivity are assessed through a logistical regression analysis of all the nerves sampled in this study.

5.1 Single Neuron Analysis

The neuron profiled in this section is a rapidly adapting afferent. The component of the stimulus being controlled by the non-repeating noise sequences is indented position. Trials 1 and 2 used tip T2 (Area: 16.6 mm^2) while trials 4-7 used tip T1 (Area: 4.2 mm^2). All trials lasted for 30 seconds, resulting in 60000 samples at the fixed 2 KiloHertz sampling rate. Trial 3 was discarded due to a recording anomaly. The magnitude of the stimulus and number of observed impulses in these trials can be summarized as:

- Trial 1 (R1221911)
    - Threshold stimulation
    - 206 Impulses (0.34 % of samples)
- Trial 2 (R1221912)
    - 2x Threshold stimulation
    - 427 Impulses (0.71 % of samples)
- Trial 4 (R1221914)
    - Threshold stimulation
    - 154 Impulses (0.26 % of samples)
- Trial 5 (R1221915)
    - Threshold stimulation
    - 335 Impulses (0.56 % of samples)
- Trial 6 (R1221916)
    - 2x Threshold stimulation
    - 434 Impulses (0.72 % of samples)
- Trial 7 (R1221917)
    - 2x Threshold stimulation
    - 139 Impulses (0.23 % of samples)

Figures 5.1 and 5.2 indicate the correlations between the impulse and stimulus states that occurred up to 20 samples (10 milliseconds) prior to the observed impulse. Odds ratios were calculated up to 100 samples prior; however, the useful information in the graph is occurs in the 0 to 20 sample range. Trials 1 and 2 (Tip T2) indicate strong *dstress/dt* sensitivity at 4 milliseconds pre-impulse as well as *dstrain/dt* sensitivity at the instant the impulse was observed. Trials 4, 5, and 6 (Tip T1) also show a *dstress/dt* and *dstrain/dt* sensitivity at approximately 3 milliseconds pre-impulse, however, they also show both *stress* and *strain* sensitivity at approximately 2 milliseconds pre-impulse. Trial 7 shows purely *stress* and *strain* sensitivity. All of the trials indicate that the response of this receptor is statistically uncorrelated with any of the confounding interaction terms included in the model.

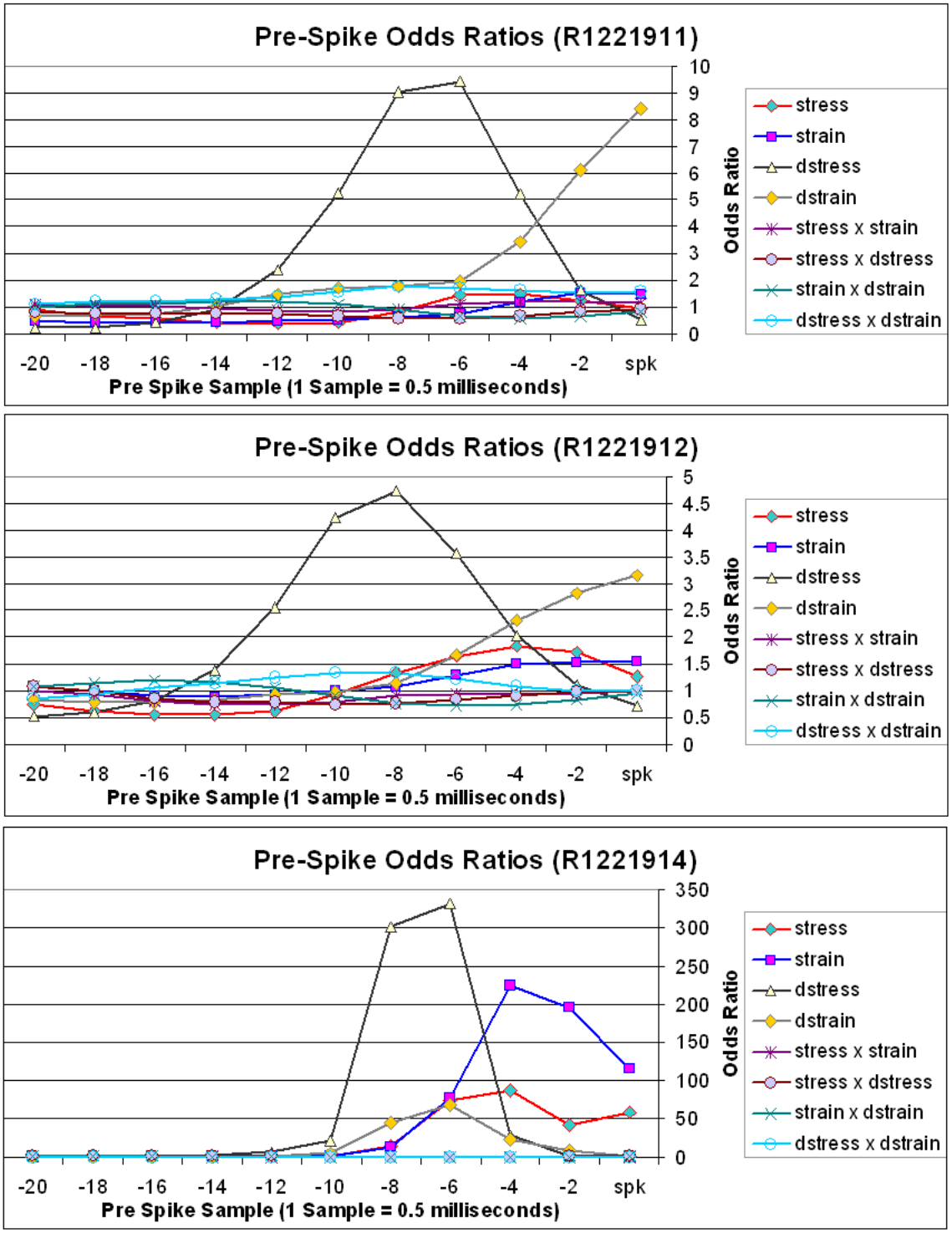


FIGURE 5.1: R122191X MEMORY EFFECTS [TRIAL 1, 2, 4]

Note the differing ranges for the abscissa. The interaction term odds ratios are not visible in trial 4 due to their low magnitude relative to the odds ratios corresponding to *stress*, *strain*, *dstress/dt*, and *dstrain/dt*.

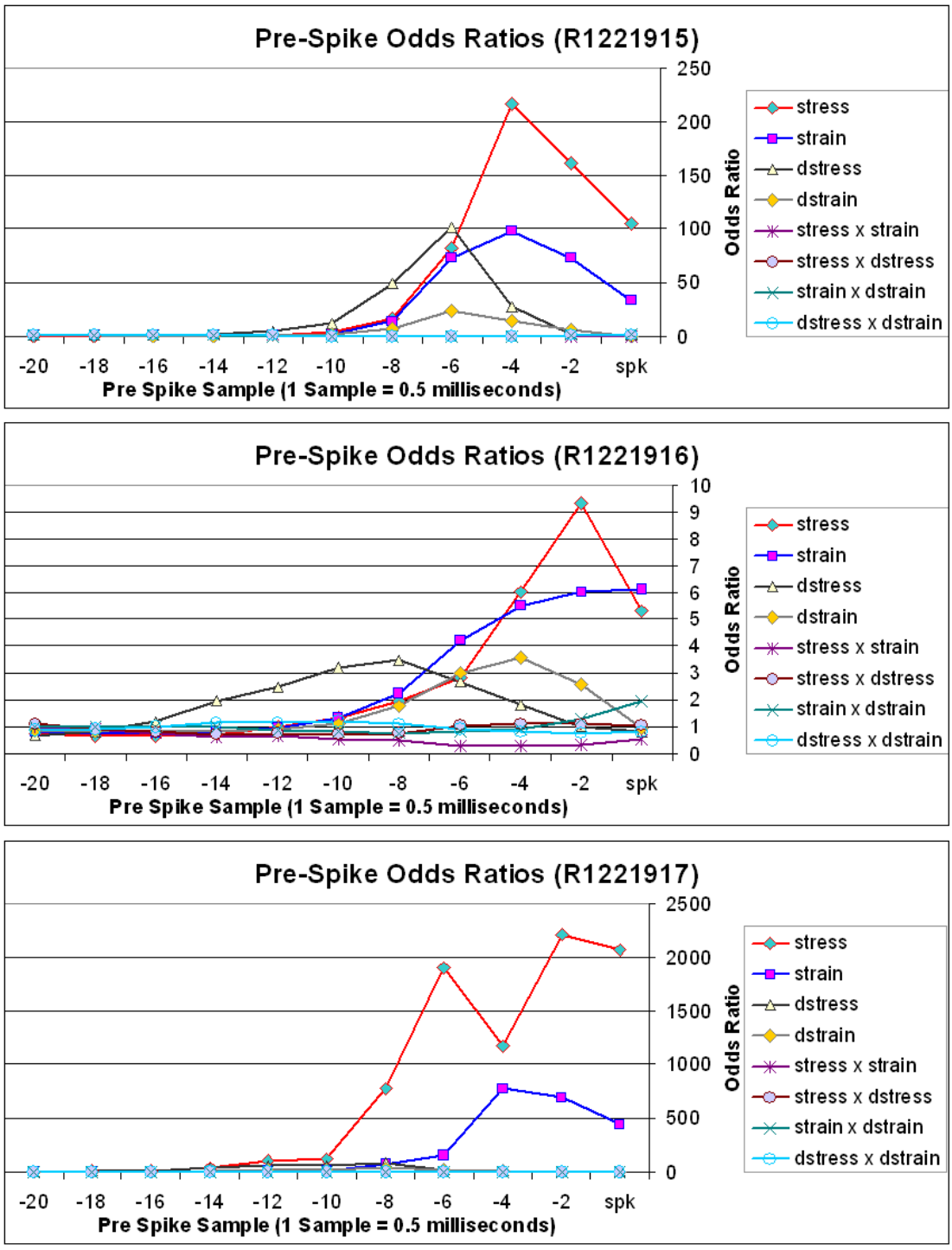


FIGURE 5.2: R122191X MEMORY EFFECTS [TRIAL 5, 6, 7]

Note the differing ranges for the abscissa. The interaction term odds ratios are not visible in trial 5 due to their low magnitude relative to the odds ratios corresponding to *stress*, *strain*, *dstress/dt*, and *dstrain/dt*. Similarly, the interaction term and time-derivative term odds ratios are not visible in trial 7 due to their low magnitude relative to the odds ratios corresponding to *stress* and *strain*.

Based on the odds ratio peaks shown in these graphs (Figures 5.1 and 5.2), this receptor is preferentially sensitive to the rate of change of *stress* and the rate of change of *strain* when stimulated with a tip of large area. When a tip of small area is used, the nerve is preferentially sensitive to the rate of change of *stress* and the rate of change of *strain* at low levels of stimulus magnitude and directly sensitive to *stress* and *strain* at all levels of stimulus magnitude.

5.1.1 Coefficient Significance

It is important to evaluate the significance of each coefficient within the fitted model in order to be able to interpret the odds ratios produced by the fitted model. Both the Wald statistic and the standard error indicate the significance of the associated independent variable with the standard error being the square root of the variance. An abnormally large standard error for a Beta coefficient (> 1) indicates that if the associated independent variable were removed, the fit of the model would be improved.

In Figures 5.3 and 5.4, the standard errors of the coefficients associated with each independent variable and interaction term are plotted up to 20 samples prior to the observed impulse. Trial 7 in Figure 5.4 shows the highest standard error of all the trials (*strain* peaks at 1, *stress* peaks at approximately 0.9). Trials 4 and 5 indicate elevated levels of standard error for *stress*, *strain*, and *dstress/dt*. These trials also have elevated odds ratios as shown in Figures 5.1 and 5.2. These elevated values are a direct result of the exponentiation operation performed on the coefficients to determine the odds ratio. With the exception of trial 1, the standard error curves follow a similar pattern for the independent variables that show correlations with the response. Any disparities are amortized over the range of standard errors, resulting in no radically erroneous coefficients. Such results lend credibility to the odds ratios predicted by the model.

The Wald statistic used in this example is calculated using Equation 22 from *Section 3.3.3*. It is effectively the value of the coefficient in the fitted model divided by the standard error of that coefficient. If the coefficient is large and has a small standard error, than the Wald statistic would indicate that the independent variable associated with the coefficient is both significant within the fitted model and strongly correlated with the outcome.

As shown in Figures 5.5 and 5.6, the Wald values are large for *dstress/dt* in trials 1, 2, 4, 5, and 6. Similarly, the Wald values for the *dstrain/dt* coefficient are large in trials 1 and 2. Therefore, in addition to these independent variables being correlated to the response, these independent variables have an elevated significance within the fitted model, thus indicating reliability both in the method and in the information embodied in the fitted model for these trials. The Wald values in trial 7 are similar and relatively constant for all coefficients, meaning that no one independent variable is particularly significant within the model. This result detracts from the credibility of the information predicted by the model fit in this trial.

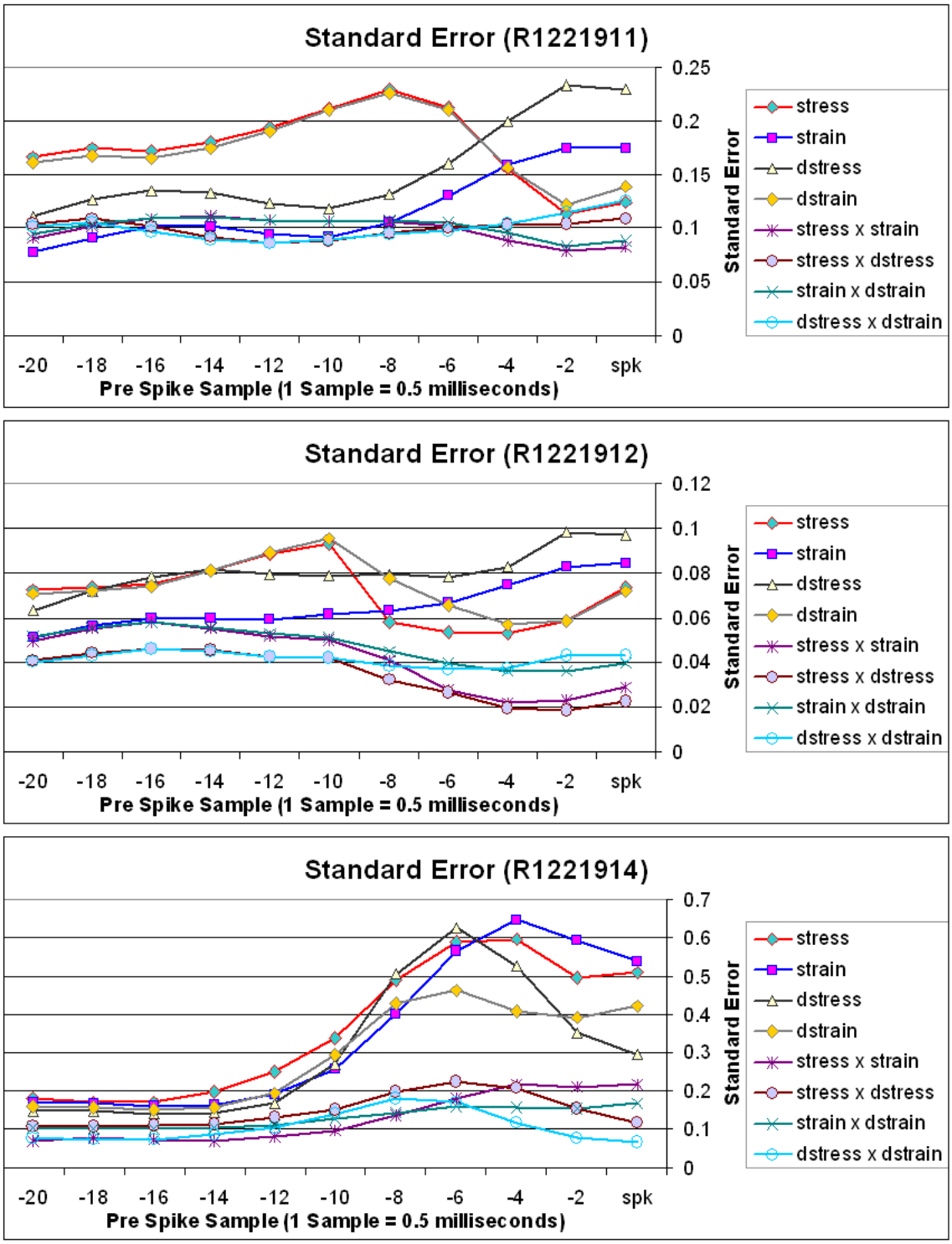


FIGURE 5.3: R122191X STANDARD ERROR [TRIALS 1, 2, 4]
The standard error in trial 4 is significantly higher than the standard error in trials 1 and 2.

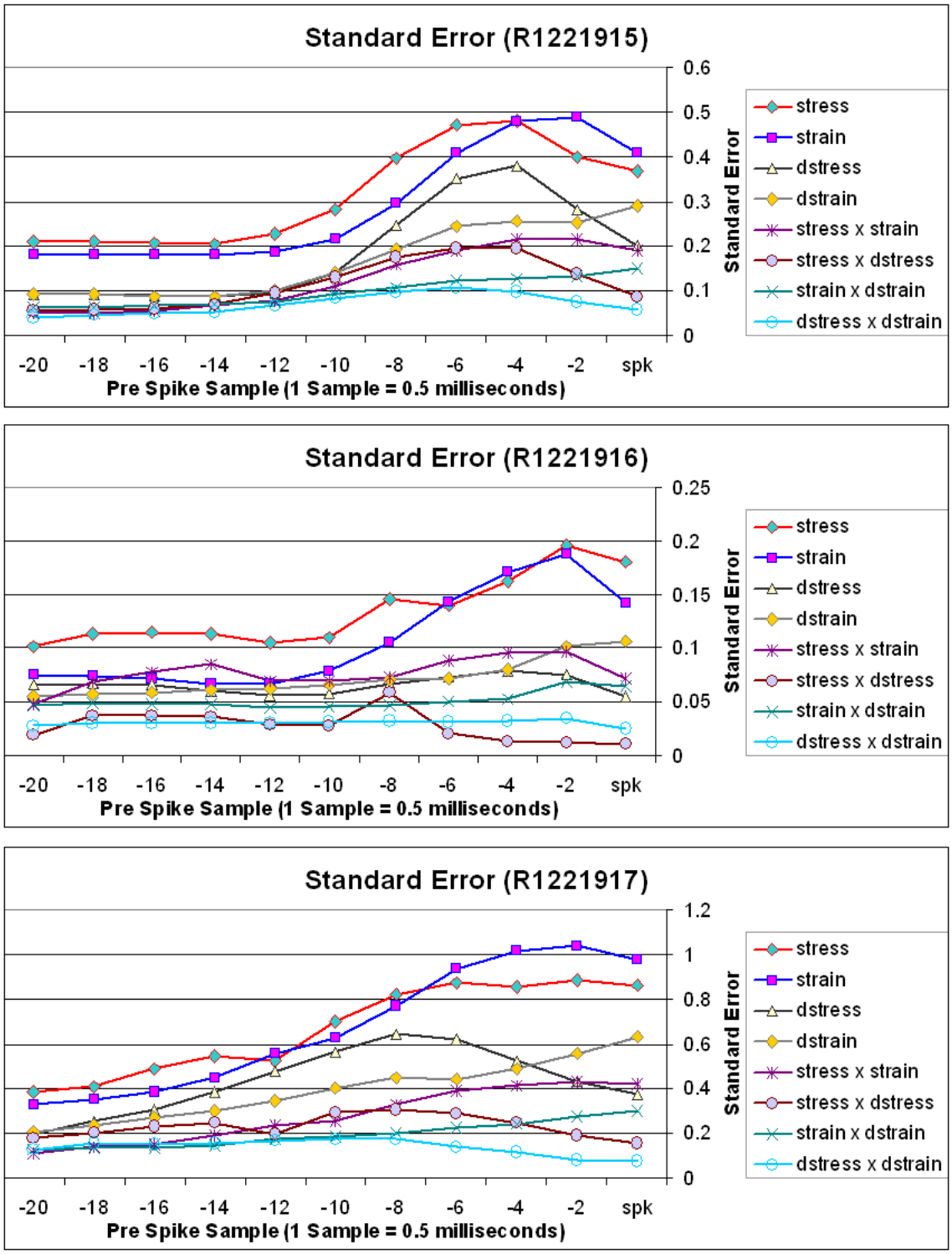


FIGURE 5.4: R122191X STANDARD ERROR [TRIALS 5, 6, 7]
The standard error in trial 7 is significantly higher than the standard error in trials 5 and 6.

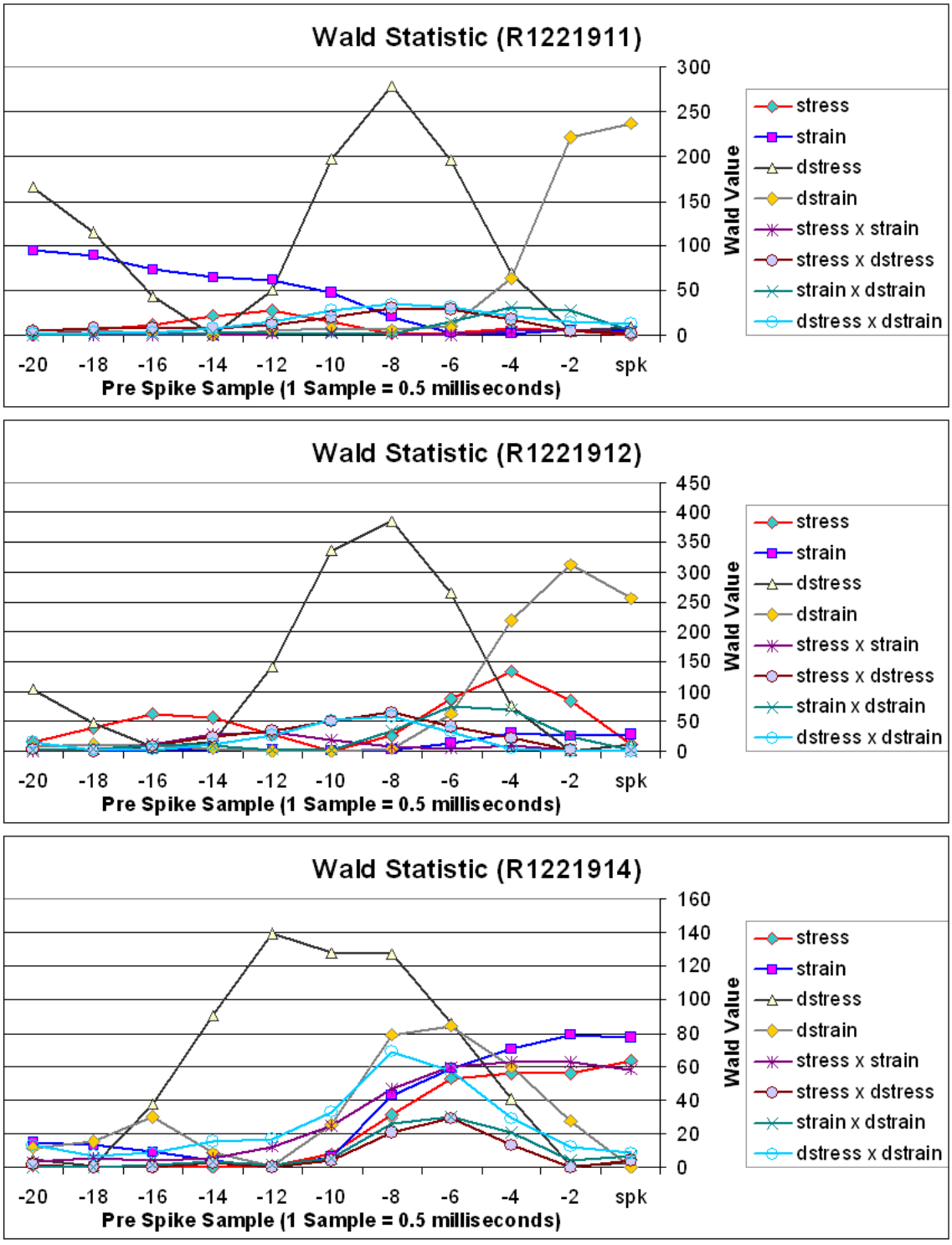


FIGURE 5.5: R122191X WALD STATISTICS [TRIALS 1, 2, AND 4]

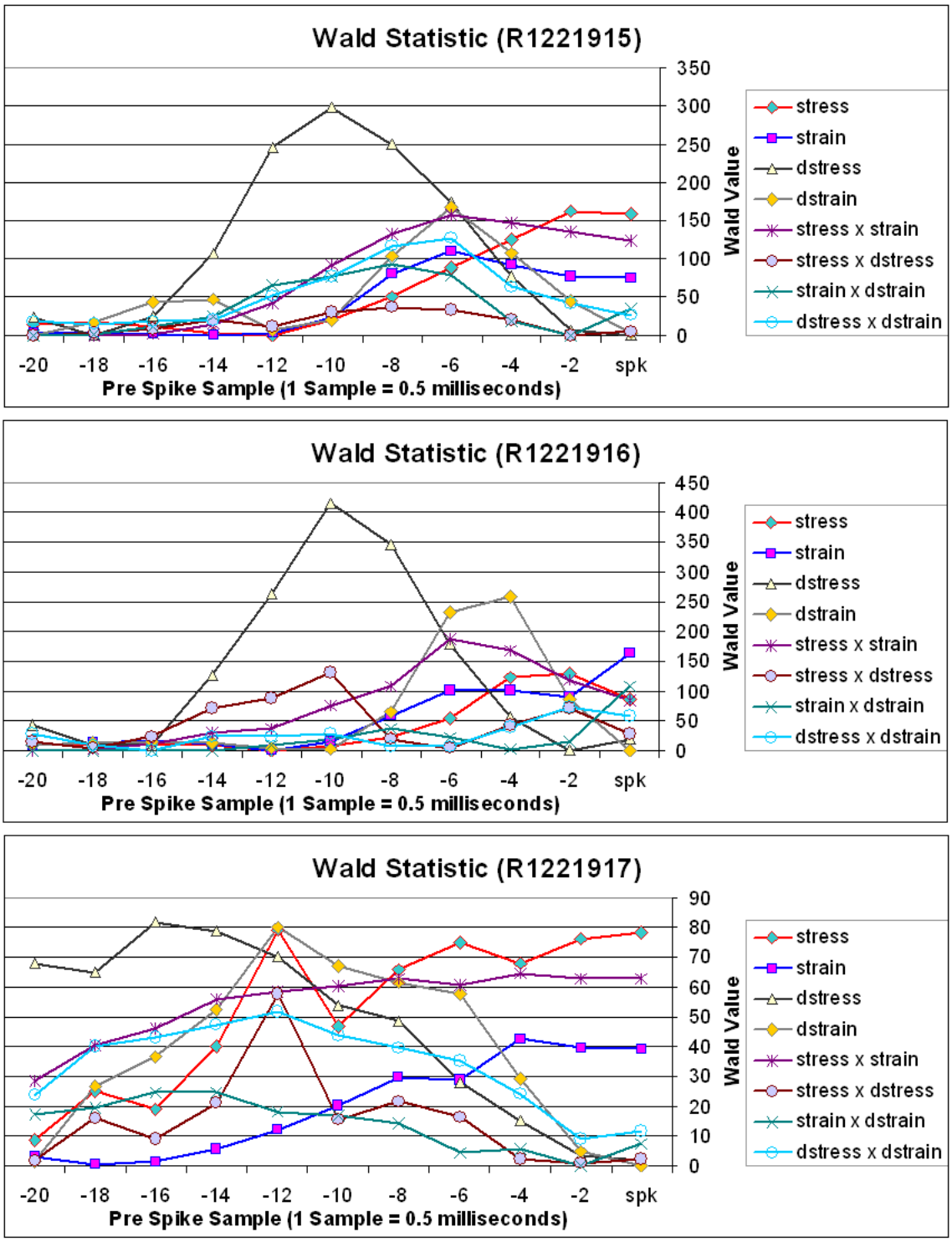


FIGURE 5.6: R122191X WALD STATISTICS [TRIALS 5, 6, AND 7]
Trial 7 indicates that no coefficient is particularly significant within the fitted model.

5.1.2 Lag Variable Correlation

The lag variable analysis performed for this receptor indicated no discernible difference between the pre-impulse odds ratios shown in Figures 5.1 and 5.2 and the same odds ratios calculated after statistically protective post-impulse data points were removed (as shown in Figures 5.7 and 5.8). This trend was observed in all datasets that were analyzed throughout the course of this experiment. Therefore, due to the fact that they do not provide any more information regarding the fit of the model, lag variable correlations will not be discussed further in this *Results and Analysis* section. It will, however, be addressed in the *Discussion* section that follows.

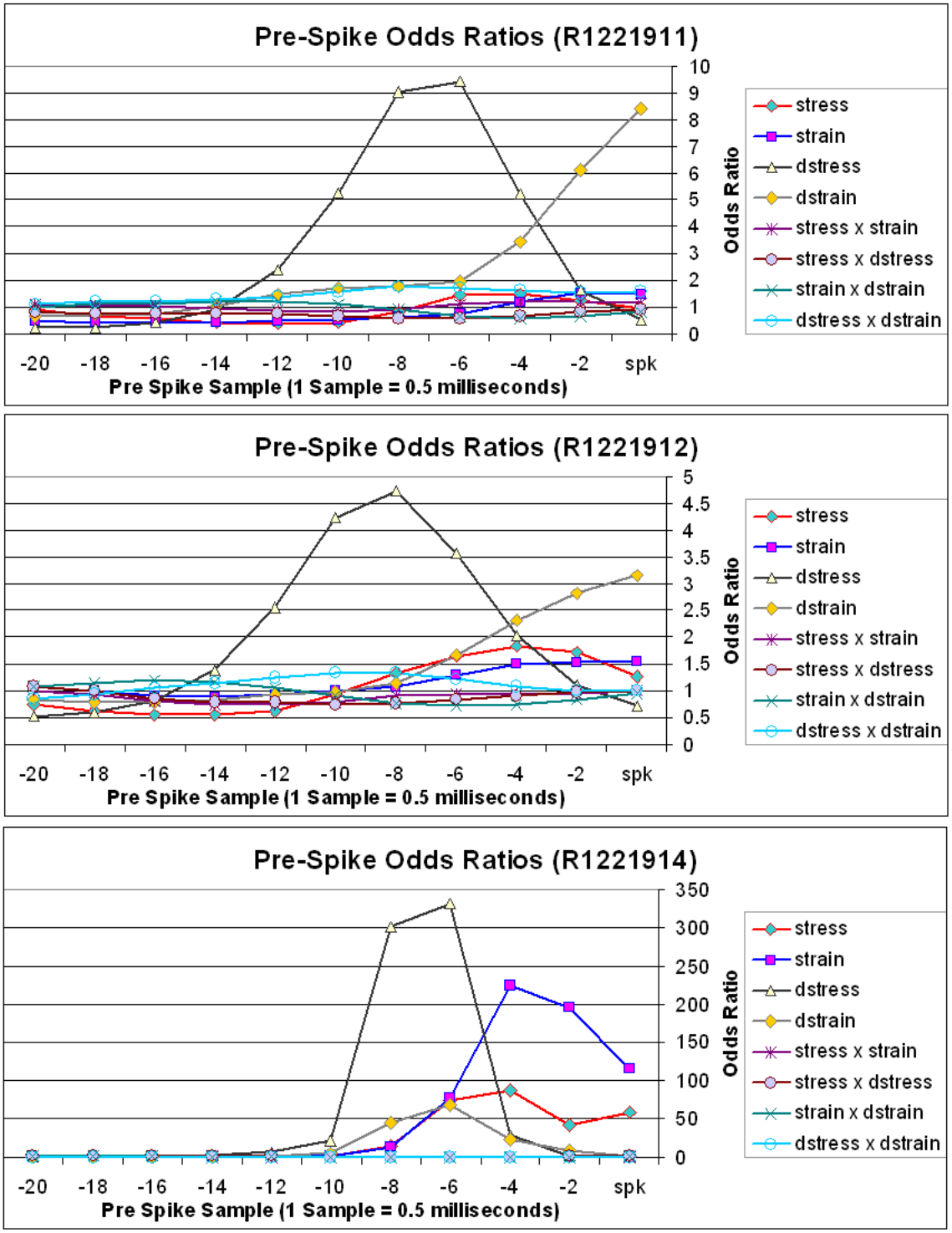


FIGURE 5.7: PRE-SPIKE ODDS RATIOS WITHOUT PROTECTIVE POST-SPIKE DATA SAMPLES
Contrast this graph to Figure 5.1. There are no noticeable differences.

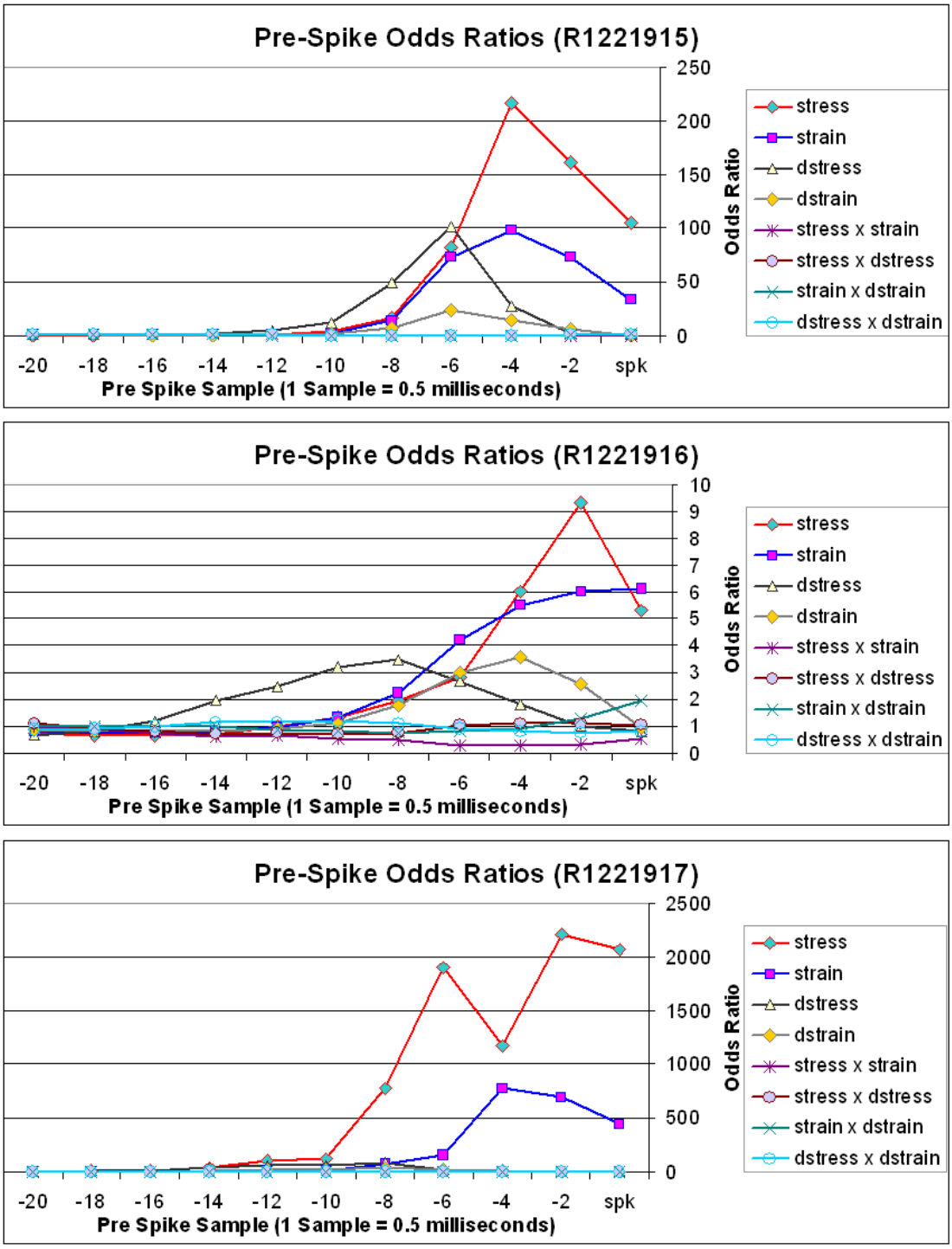


FIGURE 5.8: PRE-SPIKE ODDS RATIOS WITHOUT PROTECTIVE POST-SPIKE DATA SAMPLES
Contrast this graph to Figure 5.2. There are no noticeable differences.

5.1.3 Effects of Stress-Relaxation and Skin Creep

In order to assess the effect of stress-relaxation and creep within the skin on the sensitivity of the receptor, datasets were segmented in time into 6 equal-sized sets and pre-impulse odds ratios were calculated for each of the 6 sets. The purpose of this technique is to evaluate the correlations at the beginning of the trial compared to the correlations at the end of the trial after the skin has had ample time to reach a steady state. The results of this analysis for trial 1 are shown in Figures 5.9 and 5.10. These graphs indicate that adapting characteristics of the skin have a minimal effect on the regression coefficients within the fitted model, meaning the receptor in this trial responds relatively uniformly in spite of the varying properties of the medium in which the nerve is suspended. Trials 2, 4, 5, 6, and 7 exhibited similar behavior.

Figure 5.11 illustrates the Wald statistics for the first segment (first 10000 data samples) in the record and last segment (last 10000 data samples) in the record. It is clear from the maximum level of the peaks in the graphs that *dstress/dt* and *dstrain/dt* have more significance within the fitted model in the last segment of the dataset compared to the first segment. In addition to their improved significance over time, one can see from Figure 5.12 that the standard error of *dstress/dt* and *dstrain/dt* is nearly identical in the last segment whereas in the first segment the *dstrain/dt* standard error is elevated compared to the standard error of the other model coefficients. Based on these results, the predictive accuracy of the fitted model for this nerve increases as the skin reaches a steady state. The fitted model predicts elevated Wald values (coefficient divided by its standard error) for *dstress/dt* and *dstrain/dt* in the last segment of the dataset despite the elevated standard errors in the last segment compared to the first segment.

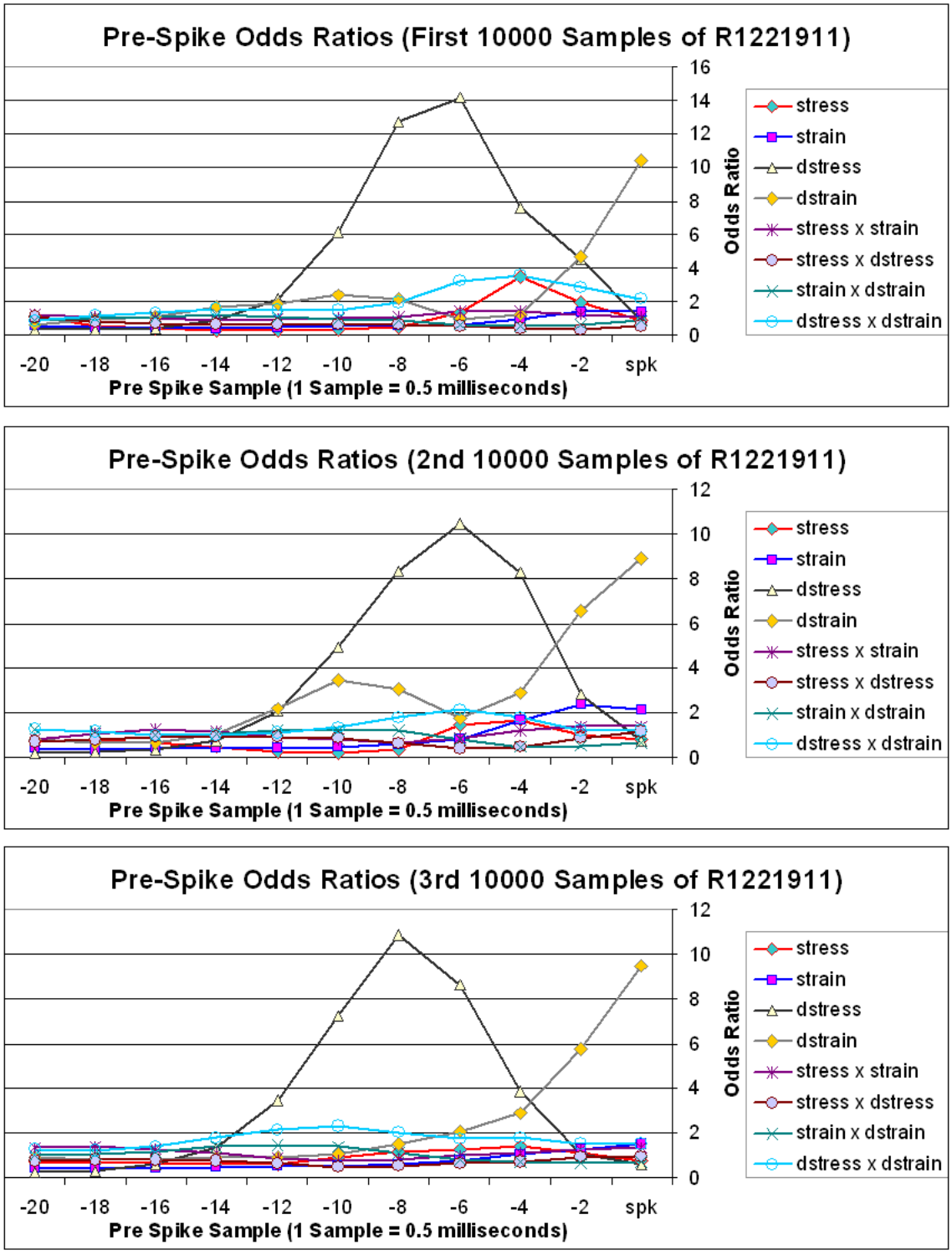


FIGURE 5.9: INFLUENCE OF SKIN BEHAVIOR ON ODDS RATIOS [R1221911 SEGMENTS 1, 2, AND 3]

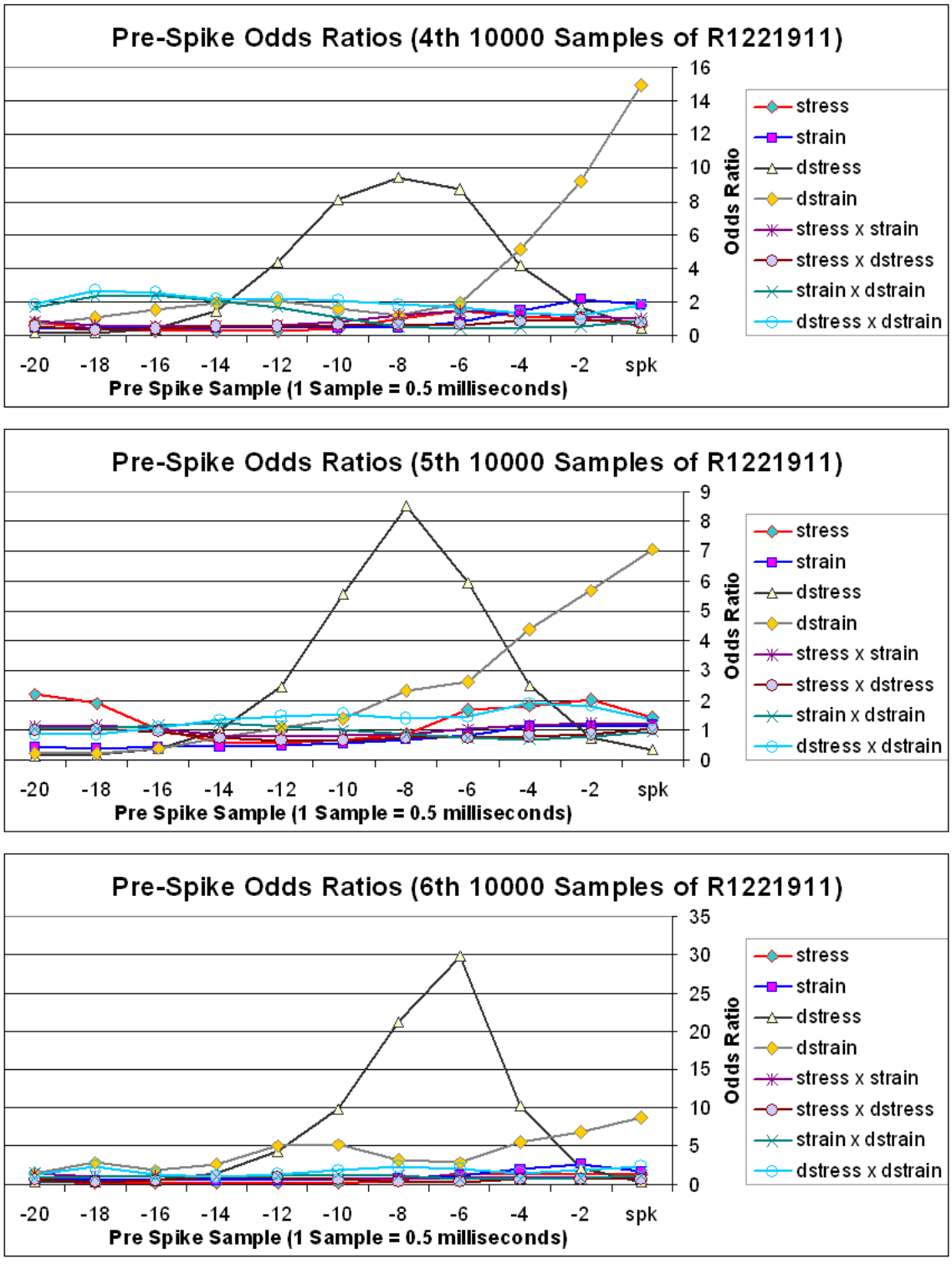


FIGURE 5.10: INFLUENCE OF SKIN BEHAVIOR ON ODDS RATIOS [R1221911 SEGMENTS 4, 5, AND 6]

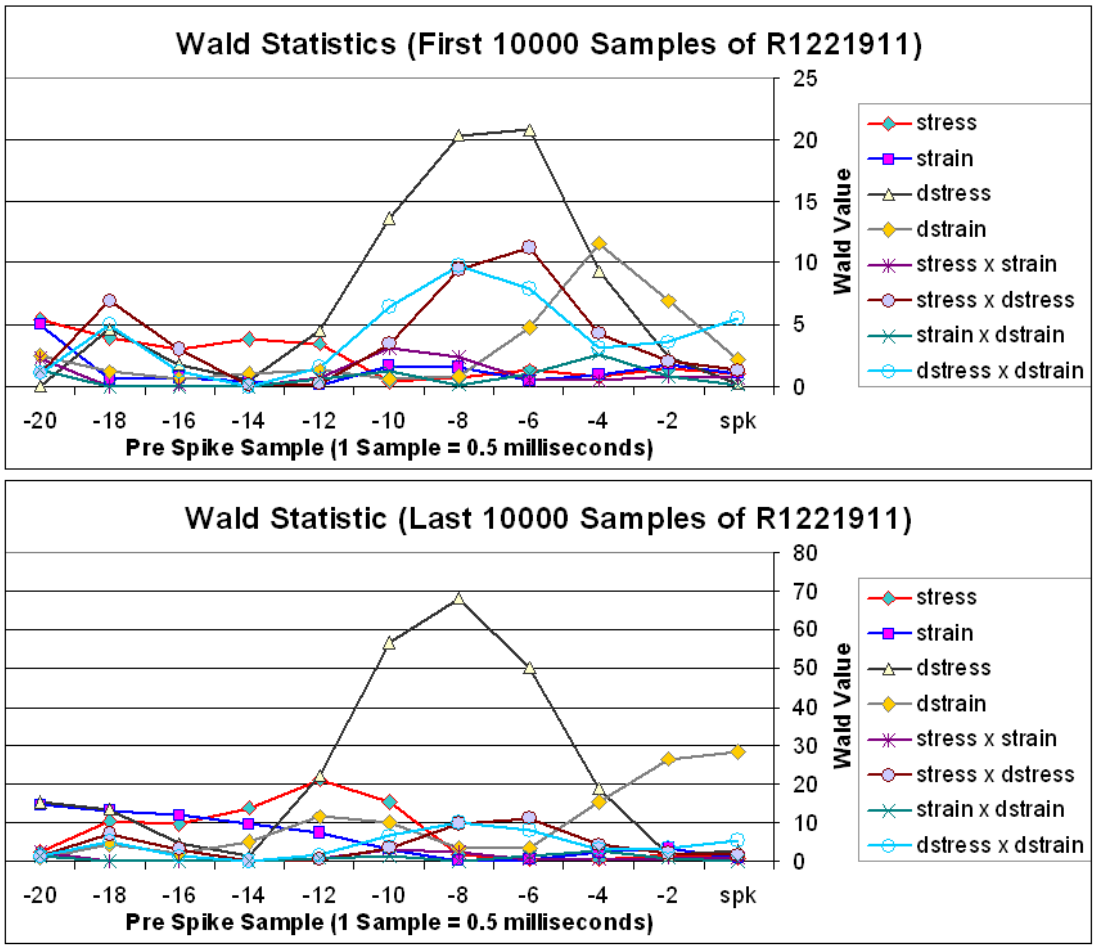


FIGURE 5.11: WALD STATISTICS – FIRST VS. LAST SEGMENTS OF DATASET [R1221911]

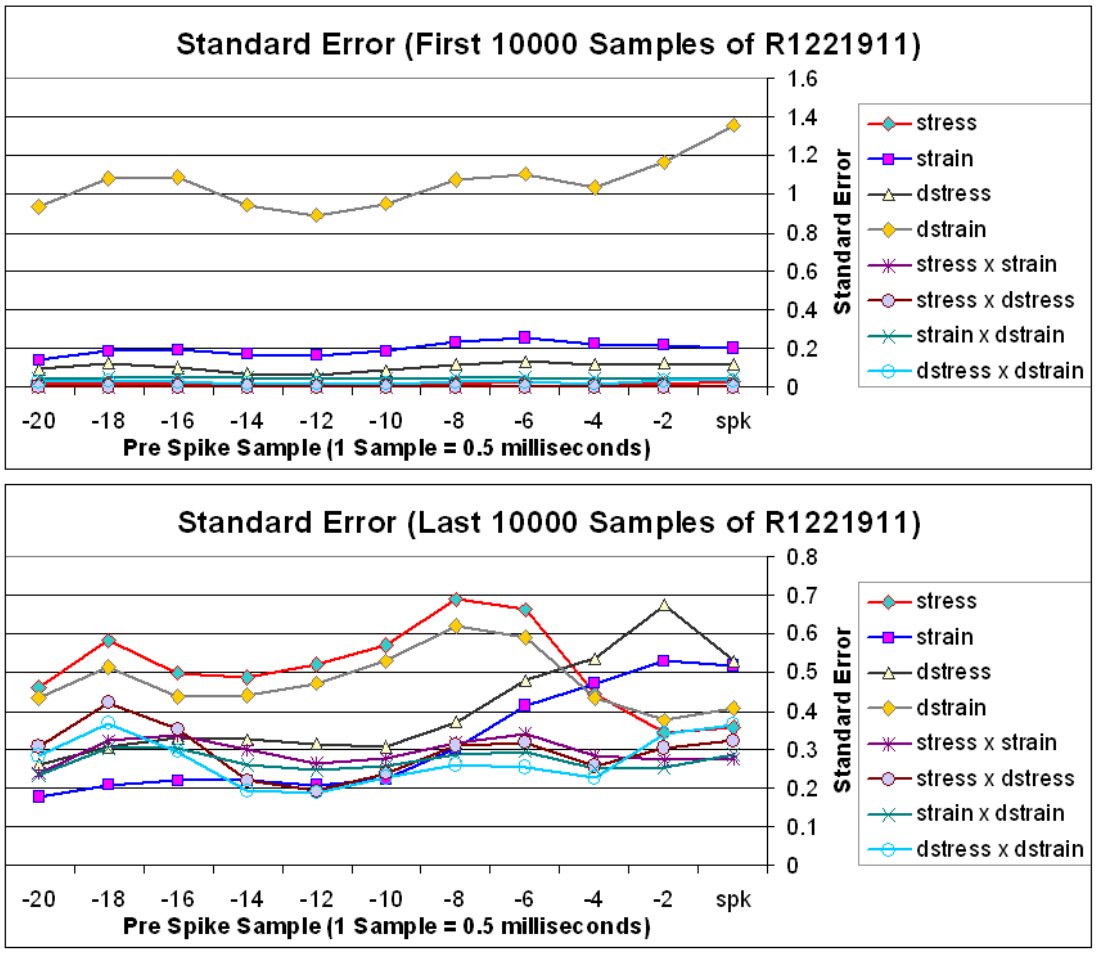


FIGURE 5.12: STANDARD ERROR – FIRST VS. LAST SEGMENTS OF DATASET [R1221911]

5.1.4 Interaction Term Effects

The inclusion or exclusion of interaction terms has a significant effect on the fitted model. In the fitted model for receptor R122191X, interactions among the independent variables were either statistically uncorrelated (odds ratio = 1 ±0.1), protective (0.5 < odds ratio < 1), or weakly correlated (1 < odds ratio < 1.25). More than 80 % of the odds ratios for interaction terms were determined to be approximately 0.9 with a standard deviation of 0.1. For this reason, pre-impulse odds ratios were calculated using the standard regression model (no interaction terms). The results are shown in Figures 5.13 and 5.14. Compared to the pre-impulse odds ratios from the extended model (shown in Figures 5.1 and 5.2), the results are similar in trials 1, 2, and 4, however, in trials 5, 6, and 7, the predominant *stress* sensitivity indicated by the extended model is not observed. Additionally, the *dstress/dt* sensitivity decreases as the stimulus intensity increases (trials 1, 4, and 5 at threshold vs. trials 2, 6, and 7 at 2x threshold).

The more important difference between the model with and without interaction terms is the fact that the fitted model without interaction terms produces odds ratios that fall within a reasonable range. In the extended model shown in Figure 5.1 and 5.2, the odds ratios for trials 4, 5, and 7 are an order of magnitude or more larger than those found in the other trials fit with the extended model (1, 2, and 6) and in comparison to the trials using the standard model, which have an upper limit of 10. This discrepancy is due to the fact that several statistically uncorrelated terms were included in the extended model. The influence of these interaction terms on the fitted parameters in the model increases with the number of samples in the experiment and since this model fit is based on approximately 60000 samples, the influence is significant enough that the extended model predicts odds ratios far beyond tangible levels. Despite this fact, the standard model validated the sensitivities predicted by the extended model.

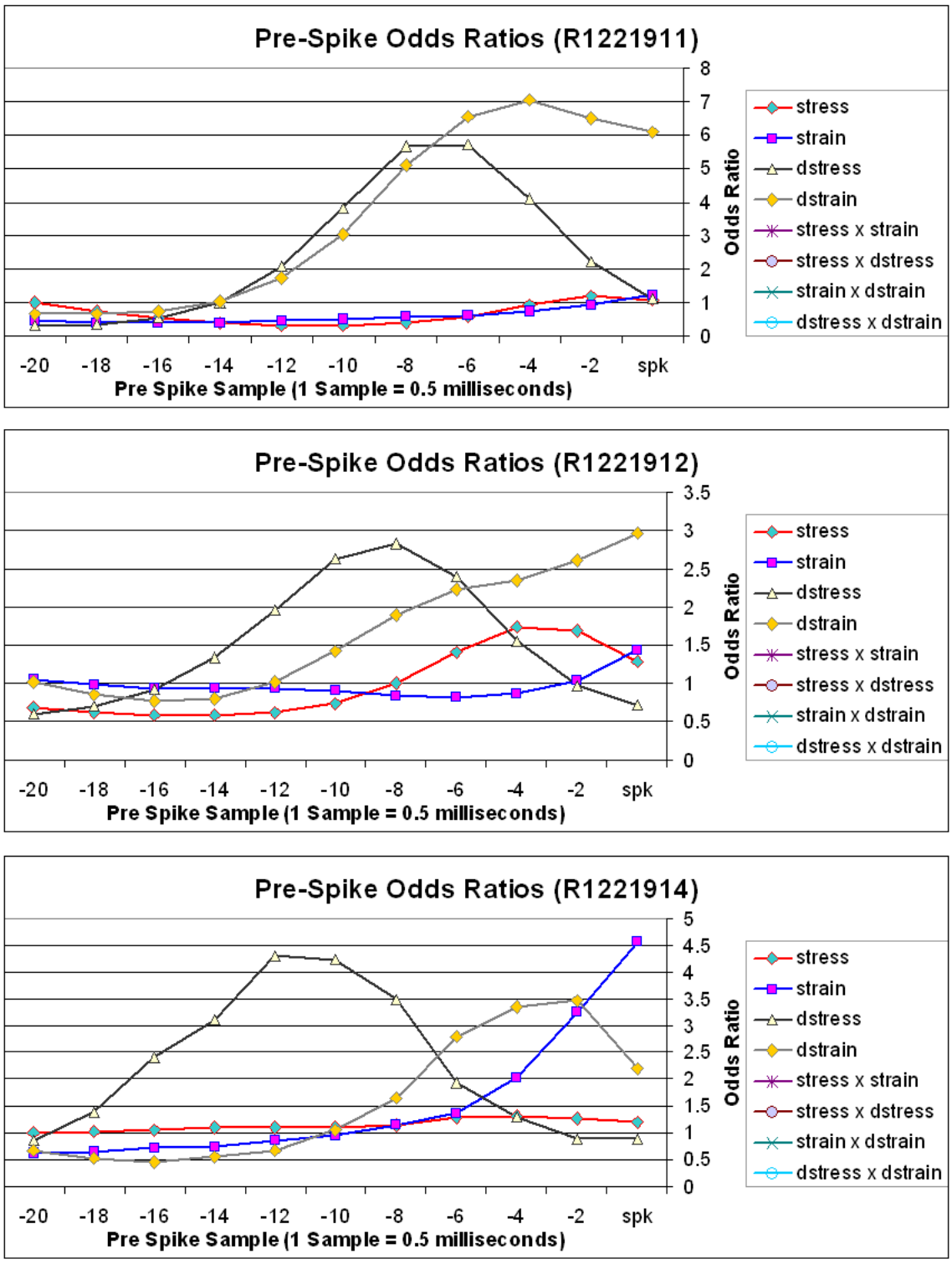


FIGURE 5.13: ODDS RATIOS USING STANDARD MODEL WITHOUT INTERACTION TERMS [TRIALS 1, 2, AND 4]

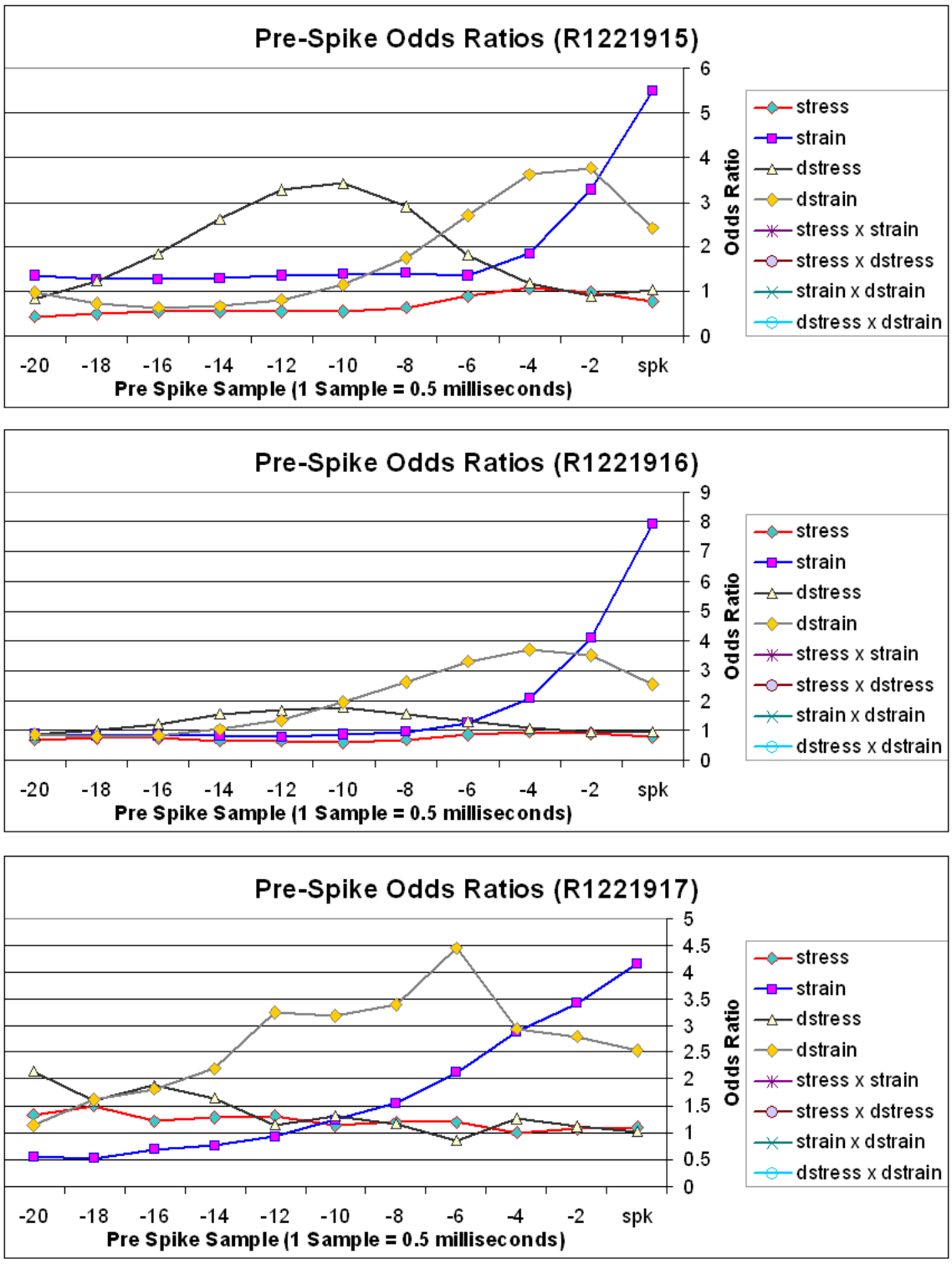


FIGURE 5.14: ODDS RATIOS USING STANDARD MODEL WITHOUT INTERACTION TERMS [TRIALS 5, 6, AND 7]

The effects of interaction terms at differing levels of stimulus intensity are assessed by plotting the odds ratio vs. stimulus magnitude (measured in standard deviations) for each independent variable. Odds ratios in Figures 5.1 – 5.14 are all calculated at a stimulus standard deviation of 0 and thus do not account for any interactions between the independent variables since the interaction effects are null at a standard deviation of 0. Figures 5.15 and 5.20 –5.24 indicate the significance of interaction terms on odds ratios. In Figure 5.15, the interactions of *dStrain/dt* and other independent variables observed at the instant the spike is observed cause the odds ratio to be highest approximately 3 standard deviations from the average towards the lower range of the stimulus intensity spectrum. Figure 5.16 shows the *Stress* interaction term effects (mean) from Figure 5.15 plotted vs. elapsed time pre-spike. At approximately 6 milliseconds pre-spike, the interaction term effects are shown to significantly affect the odds ratio for *Stress*, particularly at low intensity stimuli (standard deviation < -2). If one were to trace along the [*time* = 0] peak in Figure 5.16, one would generate the same jagged line shown for *dStrain/dt* in Figure 5.15. In Figure 5.17, similar patterns of elevated odds ratios exist for *Strain* at a pre-spike offset of 4 milliseconds, however, the odds ratios are relatively constant over the full range of stimulus intensities. At the same pre-spike offset of 6 milliseconds, the *dStress/dt*, and *dStrain/dt* odds ratios show considerable variance over the range of stimulus magnitudes (Figure 5.18 and Figure 5.19, respectively). Based on these results, interactions involving *stress* and *dStress/dt* are significant in the model whereas interactions involving *Strain* and *dStrain/dt* are not significant.

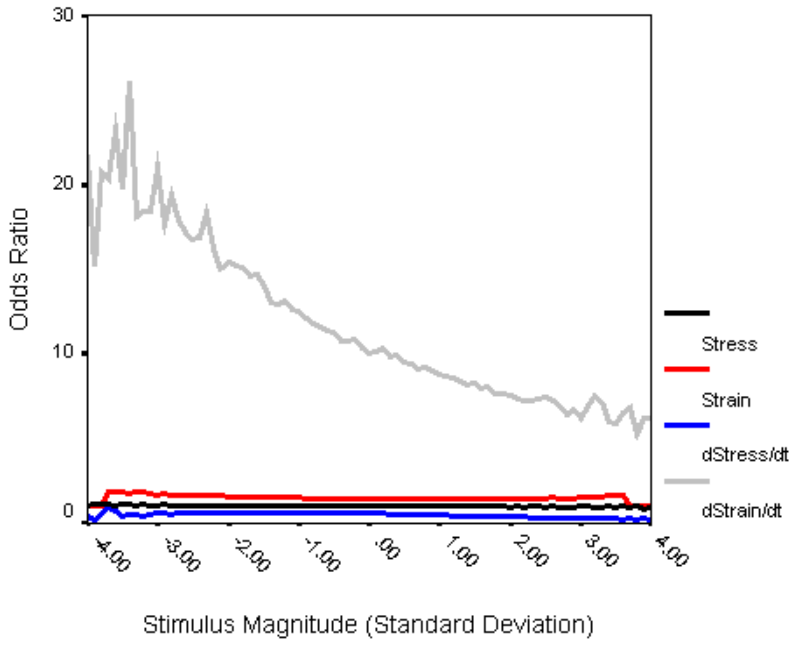


FIGURE 5.15: INDEPENDENT VARIABLE INTERACTION EFFECTS (TRIAL 1)

This graph shows that the odds ratio for *Stress*, *Strain*, and *dStress/dt* are relatively constant at all levels of the stimulus. On the other hand, *dStrain/dt* odds ratios decrease from a peak at the lowest of stimulus magnitudes to a low point at the highest of stimulus magnitudes. This trend indicates that interaction terms play a significant role in the calculated odds ratios, particularly as the stimulus magnitude approaches the low end of the intensity spectrum.

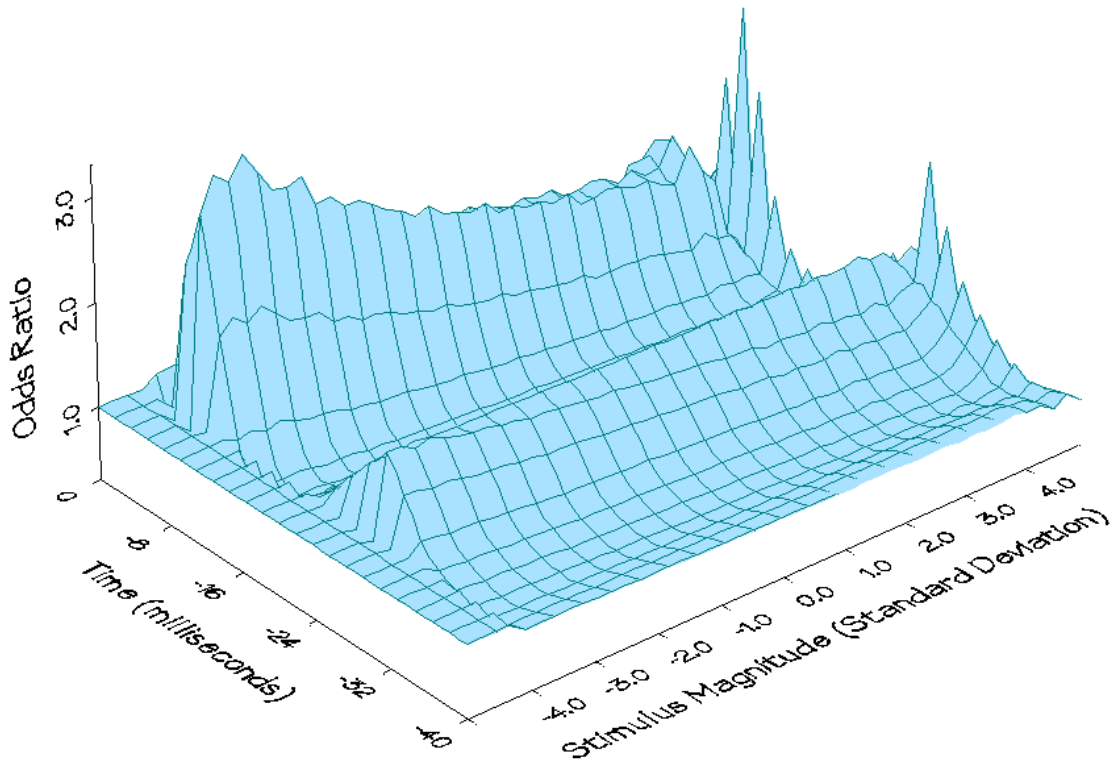


FIGURE 5.16: 3-D REPRESENTATION OF STRESS INTERACTION TERM EFFECTS VS. ELAPSED TIME PRE-SPIKE (TRIAL 1)

This graph indicates the *Stress* odds ratios vs. stimulus magnitudes for 40 samples leading up to an observed action potential (time = 0). One can see from the graph that at an offset of approximately 6 milliseconds pre-spike, the odds ratios peak. This peak is highest at low stimulus intensities and decreases as the stimulus intensity increases. This means that the interactions in which *Stress* takes part in the fitted model have a significant effect on the odds ratio for *Stress* at low stimulus intensities and this effect decreases as the stimulus intensities increase. The flattened region at the low end of the stimulus intensity spectrum indicates a region of no data. If one were to trace along the time = 0 line from low stimulus intensities to high stimulus intensities, one would see the same curve shown in Figure 5.15 for *Stress*.

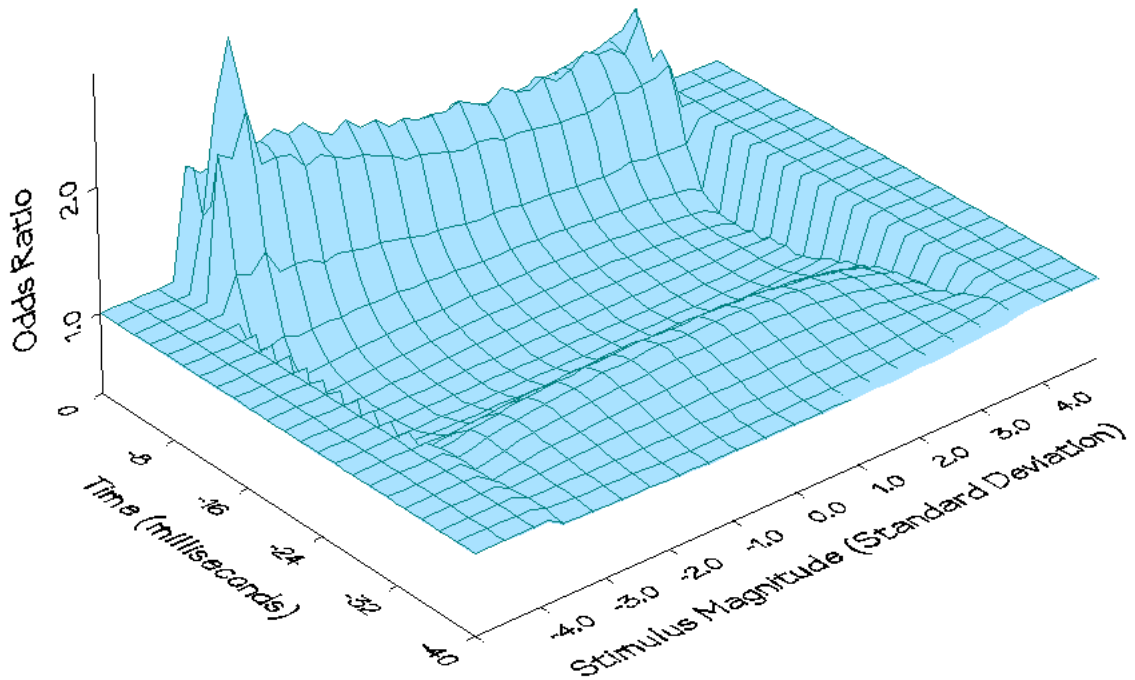


FIGURE 5.17: 3-D REPRESENTATION OF STRAIN INTERACTION TERM EFFECTS VS. ELAPSED TIME PRE-SPIKE (TRIAL 1)

This graph indicates the *Strain* odds ratios vs. stimulus magnitudes for 40 samples leading up to an observed action potential (time = 0). One can see from the graph that at an offset of approximately 4 milliseconds pre-spike, the odds ratios peak. This peak is relatively constant for all stimulus intensities, meaning the interaction terms that have *Strain* as a component do not have a significant effect on the *Strain* odds ratios. The flattened regions at the high and low ends of the stimulus intensity spectrum indicate regions of no data. If one were to trace along the time = 0 line from low stimulus intensities to high stimulus intensities, one would see the same curve shown in Figure 5.15 for *Strain*.

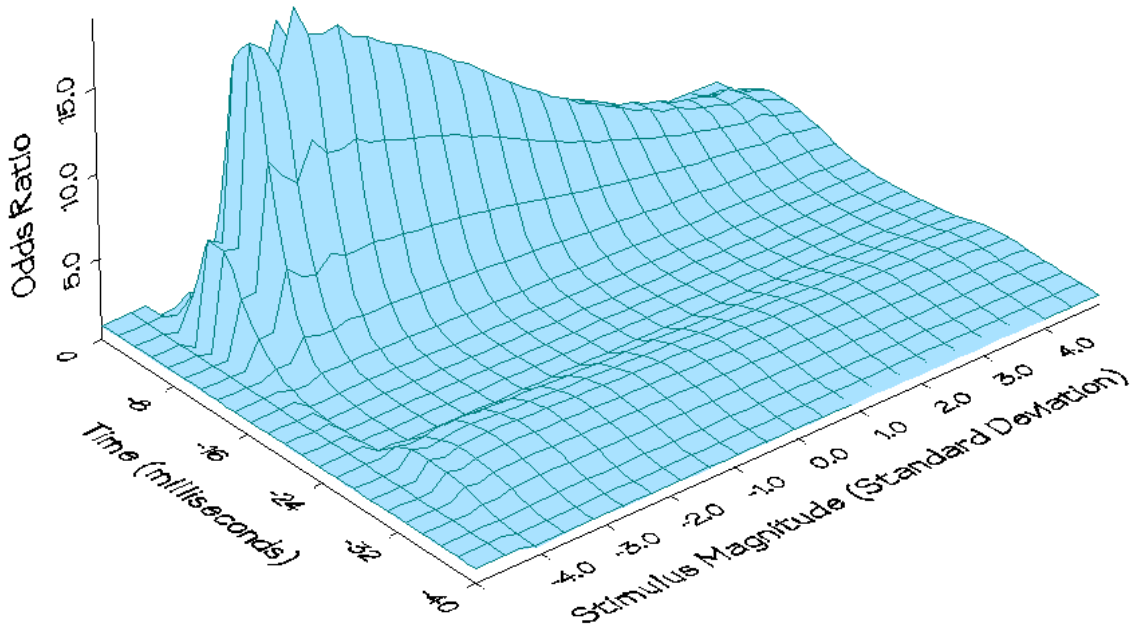


FIGURE 5.18: 3-D REPRESENTATION OF DSTRESS/DT INTERACTION TERM EFFECTS VS. ELAPSED TIME PRE-SPIKE (TRIAL 1)

This graph indicates the *dStress/dt* odds ratios vs. stimulus magnitudes for 40 samples leading up to an observed action potential (time = 0). One can see from the graph that at an offset of approximately 7 milliseconds pre-spike, the odds ratios peak. This peak is dramatically higher at low stimulus intensities than at high intensity stimuli. This means that the interactions in which *dStress/dt* takes part in the fitted model have a significant effect on the odds ratio for *dStress/dt* at low stimulus intensities. This effect decreases dramatically as the stimulus intensities increase. The flattened region at the low end of the stimulus intensity spectrum indicates no data for that region of the intensity spectrum. If one were to trace along the time = 0 line from low stimulus intensities to high stimulus intensities, one would see the same curve shown in Figure 5.15 for *dStress/dt*.

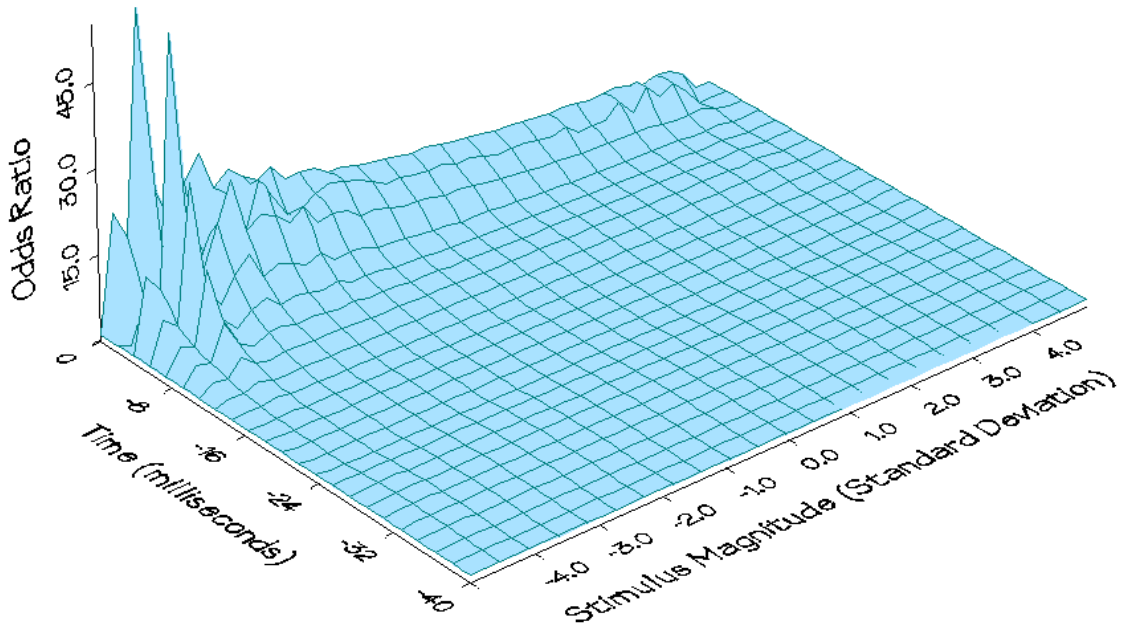


FIGURE 5.19: 3-D REPRESENTATION OF DSTRAIN/DT INTERACTION TERM EFFECTS VS. ELAPSED TIME PRE-SPIKE (TRIAL 1)

This graph indicates the *dStrain/dt* odds ratios vs. stimulus magnitudes for 40 samples leading up to an observed action potential (time = 0). One can see from the graph that between the time the spike is observed and 6 milliseconds prior to that point that the odds ratio peaks. This peak is very high relative to the overall distribution of odds ratios up to 40 milliseconds post spike. This result indicates that the interaction effects significantly affect the odds ratio of *dStrain/dt* at very low stimulus intensities. Since there are very few data points in the region showing extremely elevated odds ratios, one could conclude that there is not enough coverage in this region of the sample space to draw any conclusions from the odds ratio peaks. The flattened region at the high end of the stimulus intensity spectrum indicates a region of no data. If one were to trace along the time = 0 line from low stimulus intensities to high stimulus intensities, one would see the same curve shown in Figure 5.15 for *dStrain/dt*.

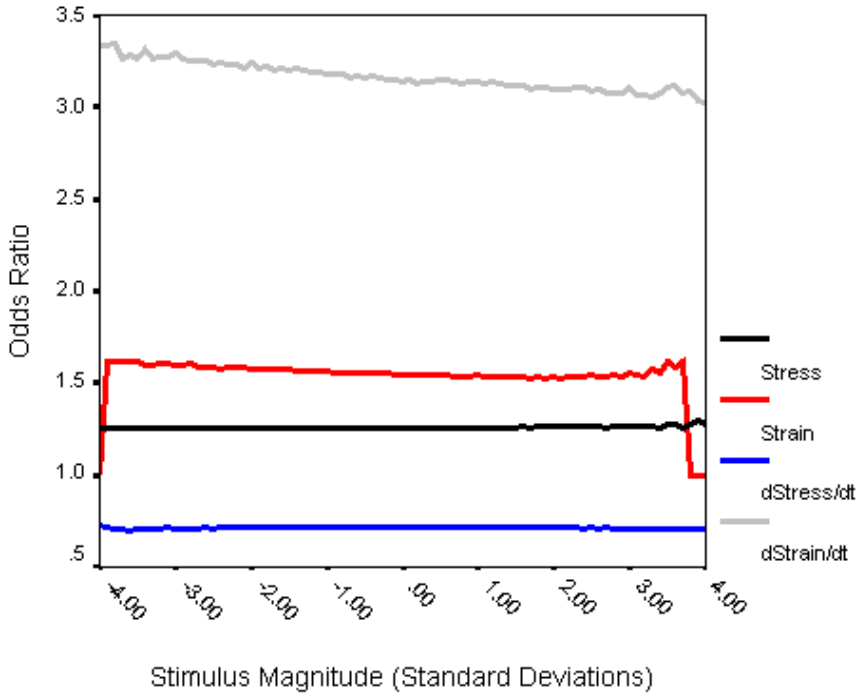


FIGURE 5.20: INDEPENDENT VARIABLE INTERACTION EFFECTS (TRIAL 2)

In this trial, the odds ratios are relatively constant for all levels of stimulus intensities.

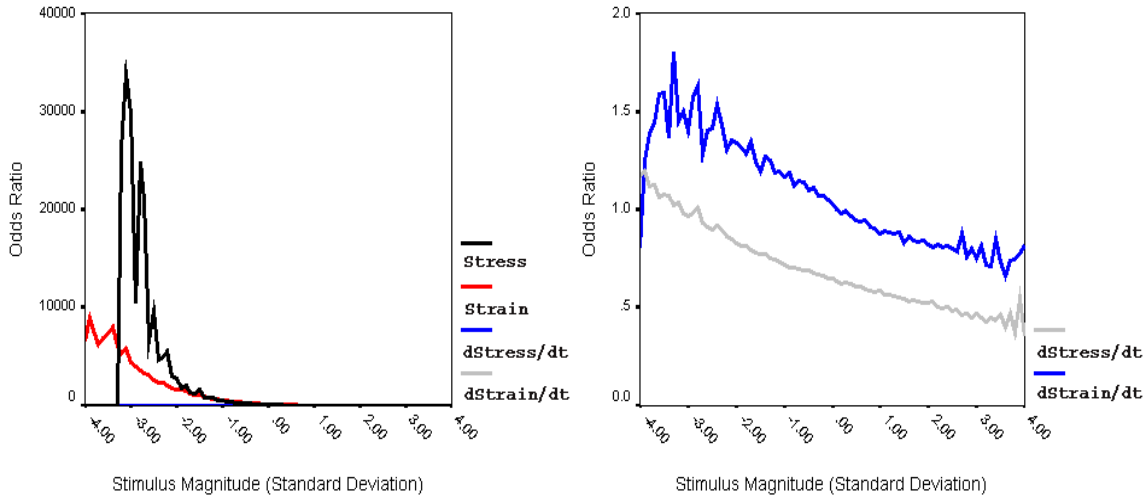


FIGURE 5.21: INDEPENDENT VARIABLE INTERACTION EFFECTS (TRIAL 4)

The results of this trial show extremely elevated odds ratios at the low end of the intensity spectrum. The extreme levels of these odds ratios immediately make the results questionable. Since they fall in a region in which very few samples exist, they are not considered to be indicative of the properties of the data, rather, they are simply statistical anomalies.

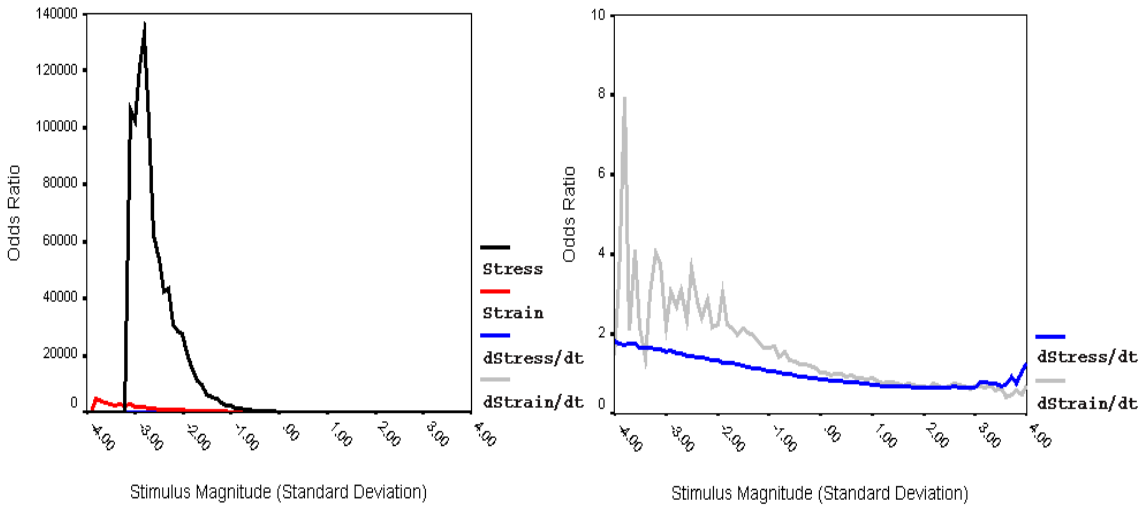


FIGURE 5.22: INDEPENDENT VARIABLE INTERACTION EFFECTS (TRIAL 5)
See caption under Figure 5.21.

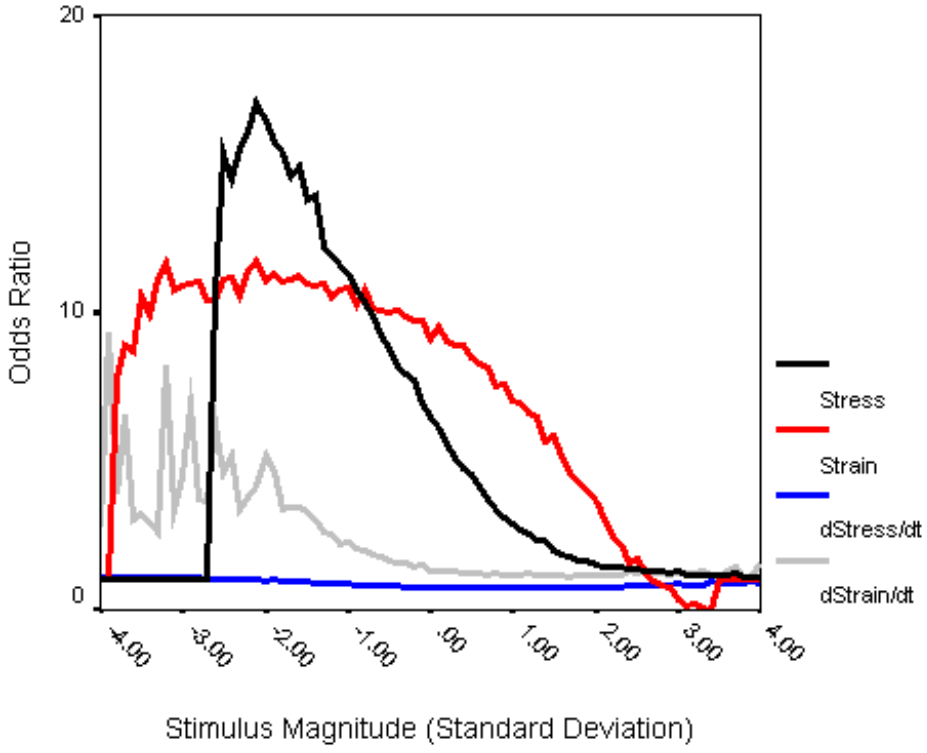


FIGURE 5.23: INDEPENDENT VARIABLE INTERACTION EFFECTS (TRIAL 6)
This graph shows clear trends: the interactions in which *Stress* and *Strain* participate have a significant effect on the odds ratios of those variables at low intensity stimuli. This effect mitigates as the stimulus intensity increases. The *dStrain/dt* odds ratios show similar behavior but to a lesser degree. The odds ratio of the *dStress/dt* variable is not affected by the interaction terms in which it participates.

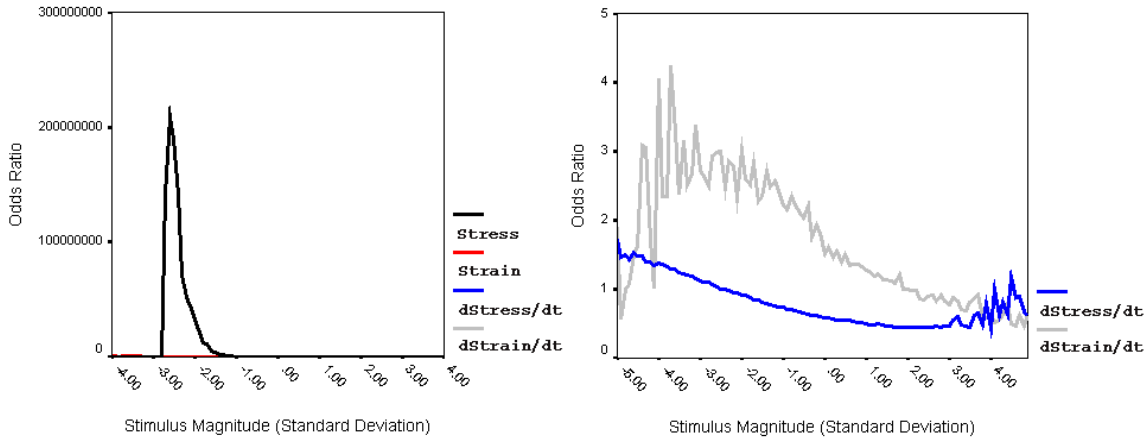


FIGURE 5.24: INDEPENDENT VARIABLE INTERACTION EFFECTS (TRIAL 7)
See caption under Figure 5.21.

5.1.5 Outliers

Outliers are data points that fall outside the confidence interval of the fitted model. In logistical regression, plots of Δ (delta) Beta vs. Probability, ΔDeviance vs. Probability, and ΔChi-square vs. Probability help identify outliers. The identification process is purely subjective as it is up to the researcher to identify data points that fall outside normal groupings or patterns, however, data points that fall outside 3 standard deviations from the distribution mean should be considered. Each of these plots contains 60000 data points, however, the binary 0 observations are tightly grouped such that they are not individually discernible.

Figures 5.25A through 5.25C show a plot of ΔBeta vs. Probability for nerve R122191X. Since positive outcomes corresponding to a nerve response are the minority (approximately 2 % of the total number of observations), positive outcome outliers are significant. In trial 2, positive outcomes (observed nerve impulses) are tightly clustered except for 7 data points. These 7 data points exert an above-average influence on the model. In trial 4, 2 data points corresponding to positive outcomes stand out as affecting the model more than the status quo and in trial 6, approximately 25 data points stand out apart from a tightly coupled cluster of positive outcome data points.

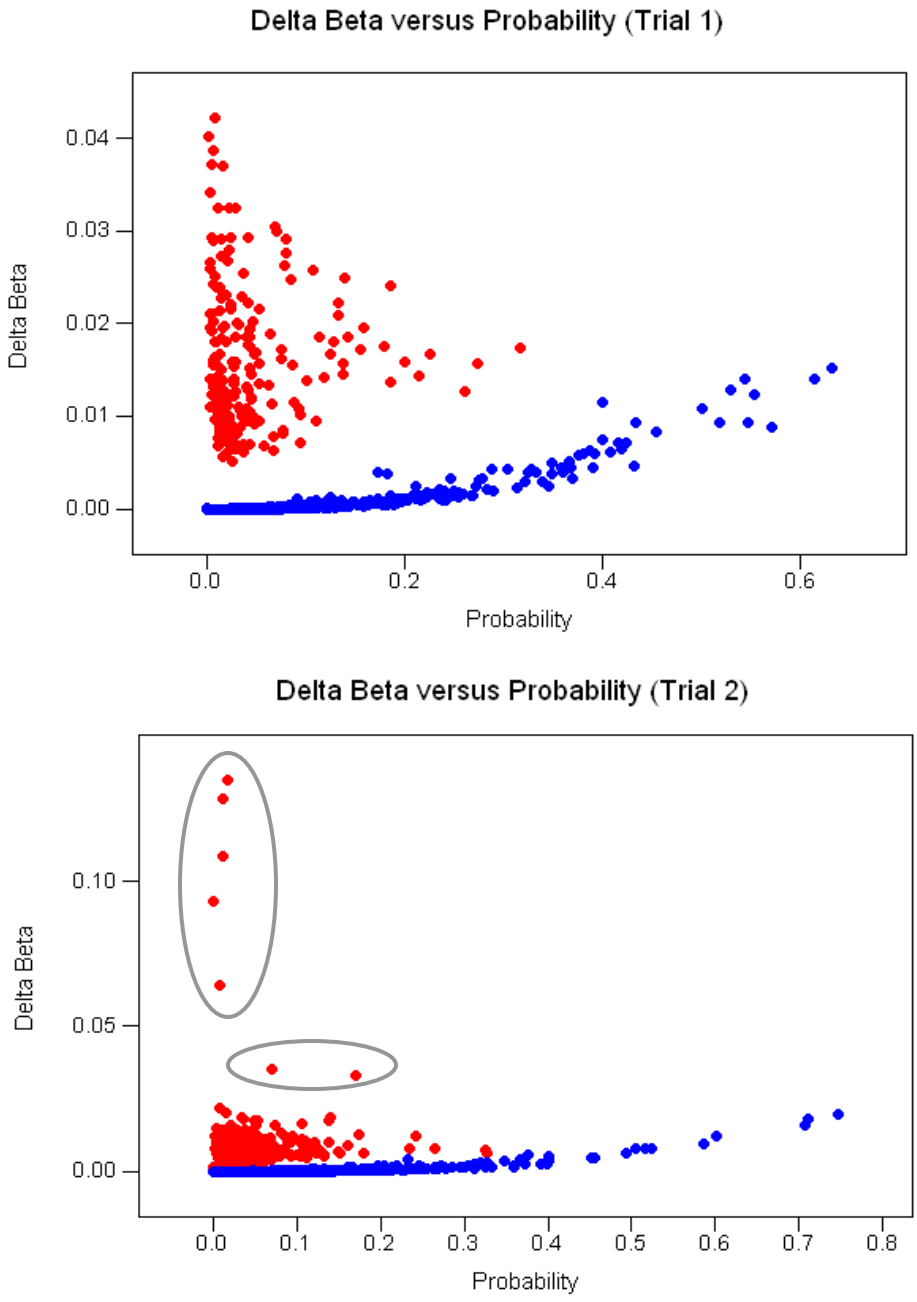


FIGURE 5.25A: DELTA BETA VS. PROBABILITY [TRIALS 1 AND 2]

The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

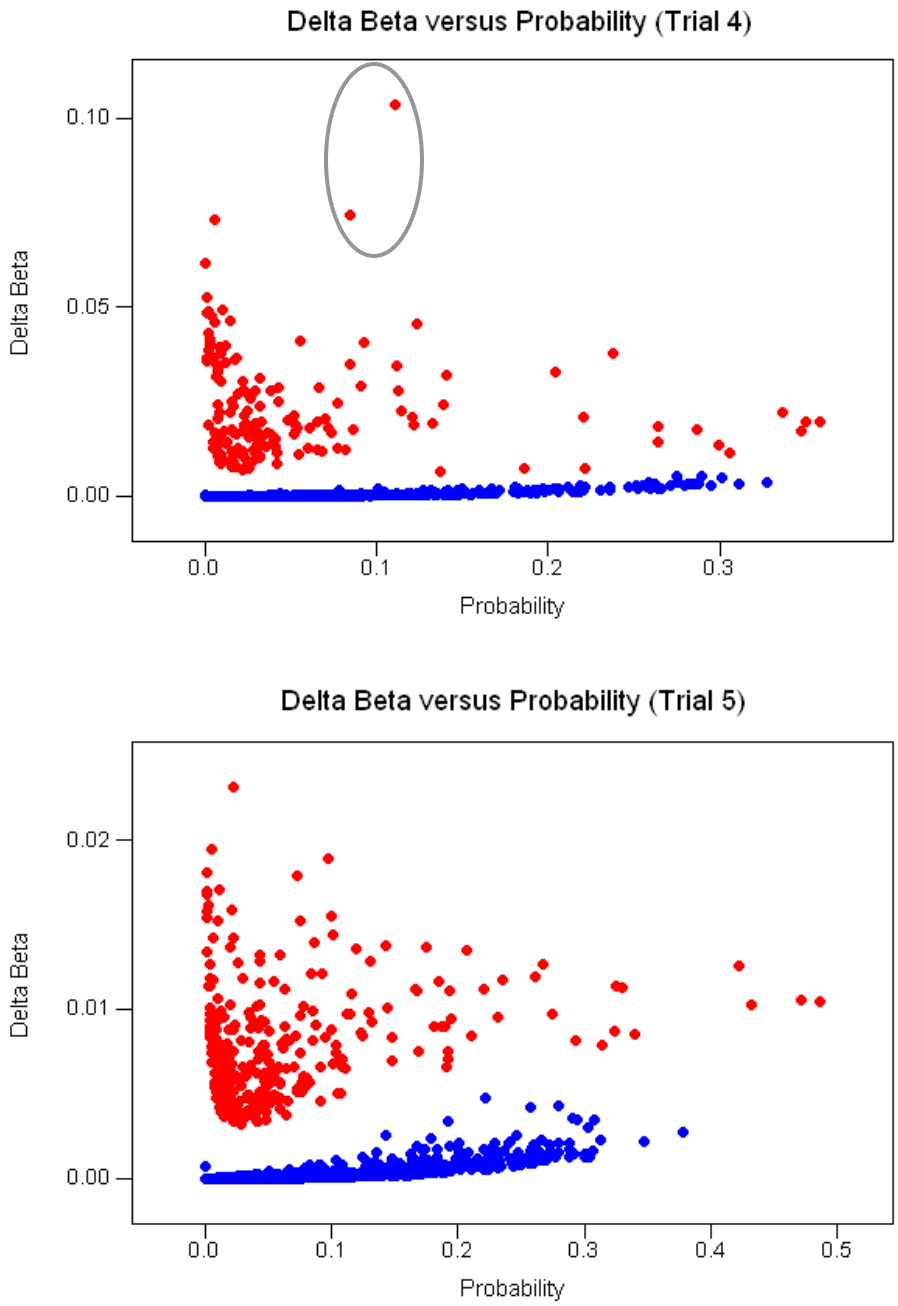


FIGURE 5.25B: DELTA BETA VS. PROBABILITY [TRIALS 4 AND 5]

The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

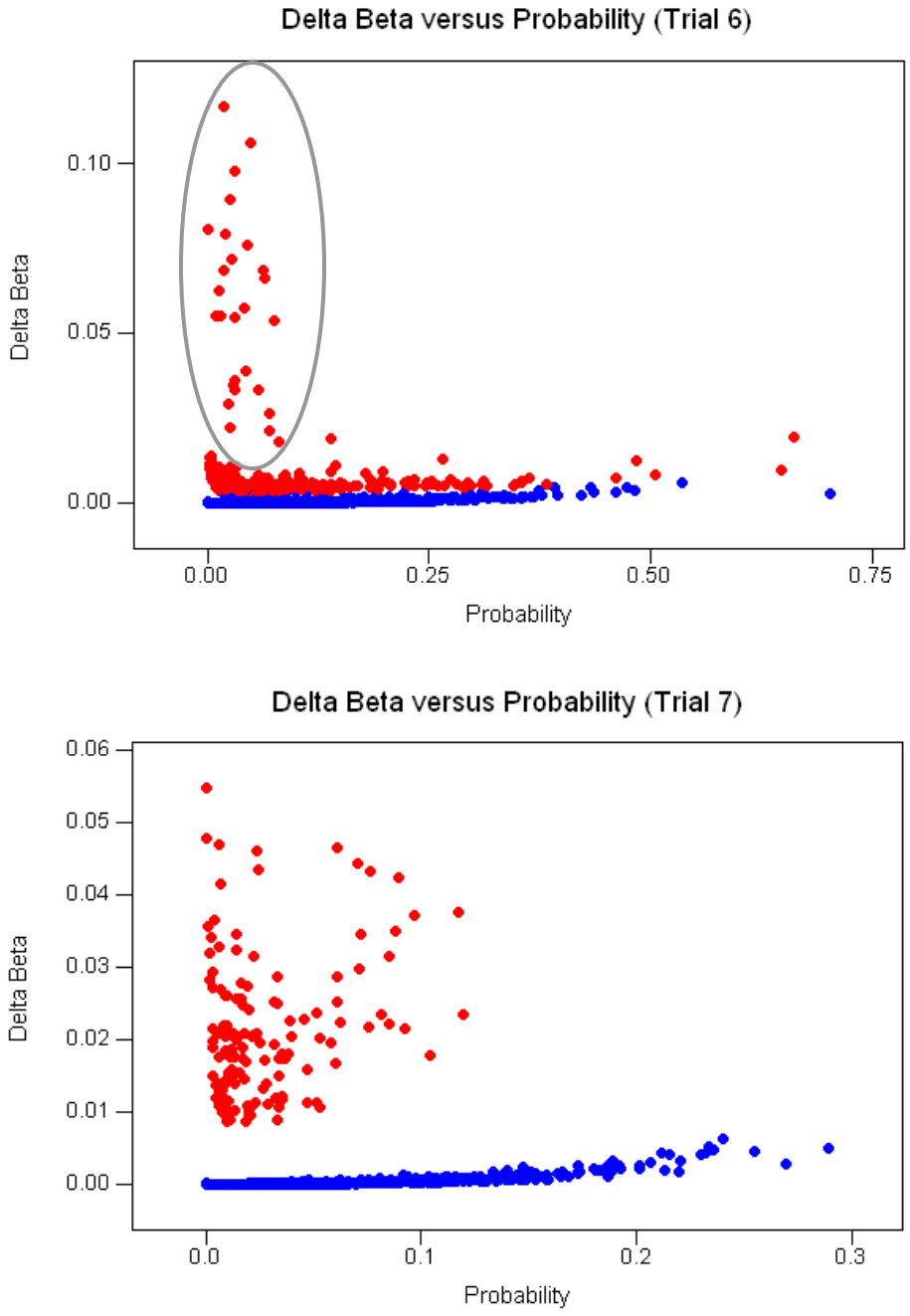


FIGURE 5.25C: DELTA BETA VS. PROBABILITY [TRIALS 6 AND 7]

The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

Figures 5.26A through 5.26C illustrate plots of ΔDeviance vs. Probability for nerve R122191X. Trials 2, 4, and 6 show outliers. Figure 5.27 shows the pre-impulse odds ratios for trial 2 after the outlier was removed and the odds ratios were re-calculated. The difference between these values and the odds ratios calculated with the outlier included are negligible, as shown in Figure 5.28. The maximum difference is 8 % but the average is less than 1 %. This is due primarily to the extremely large sample size (60000 observations).

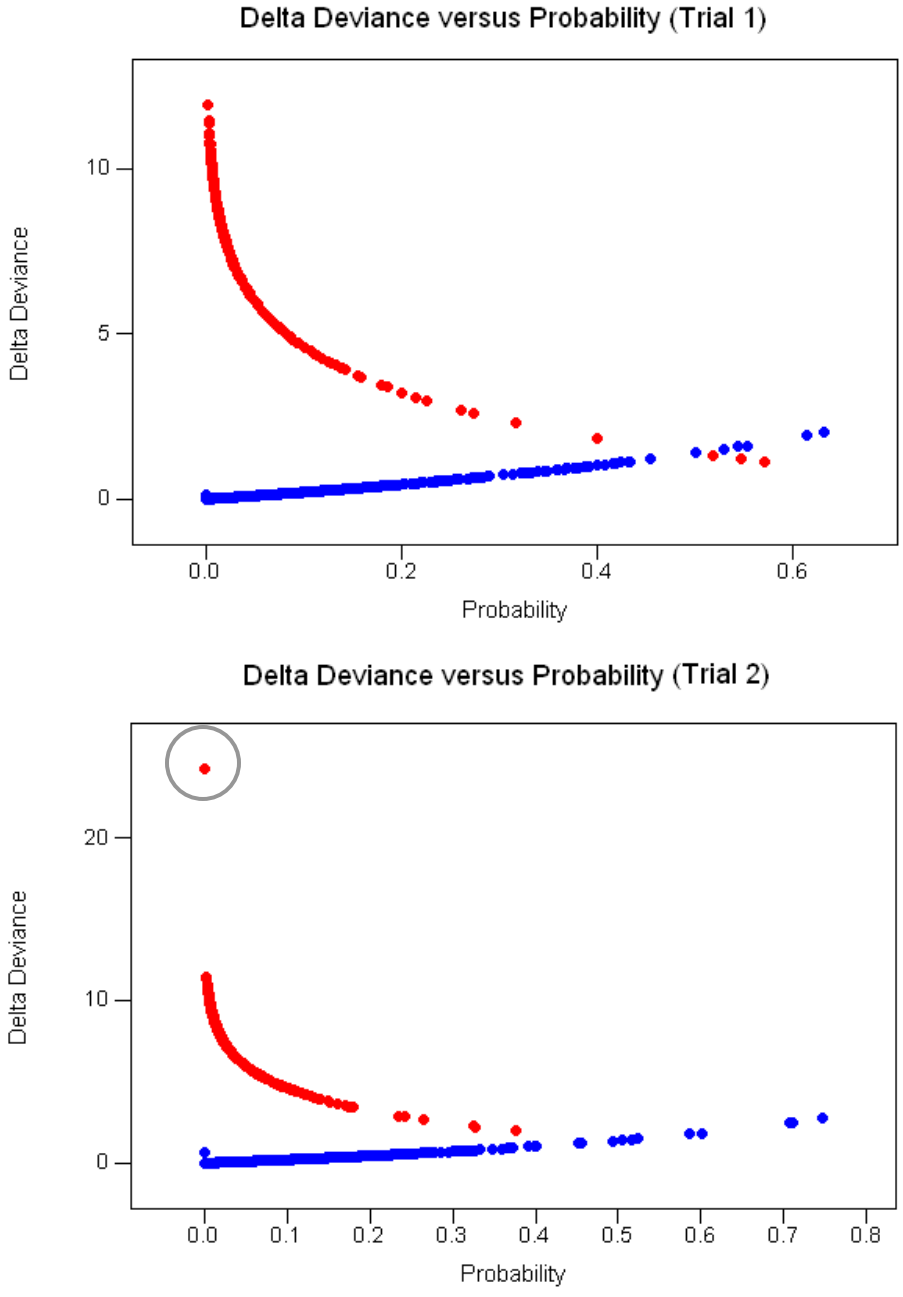


FIGURE 5.26A: DELTA DEVIANCE VS. PROBABILITY [TRIALS 1 AND 2]
The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

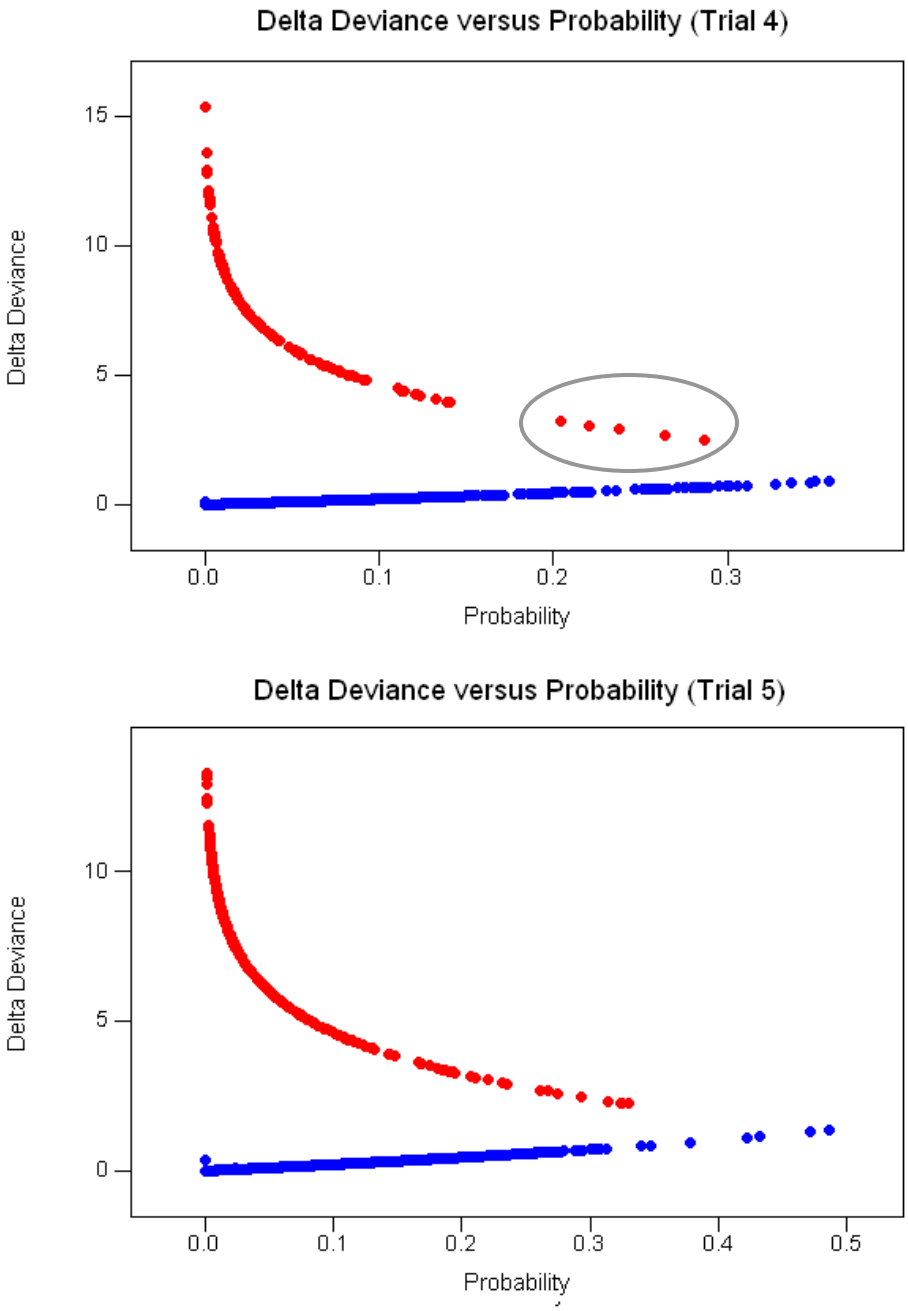


FIGURE 5.26B: DELTA DEVIANCE VS. PROBABILITY [TRIALS 4 AND 5]
The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

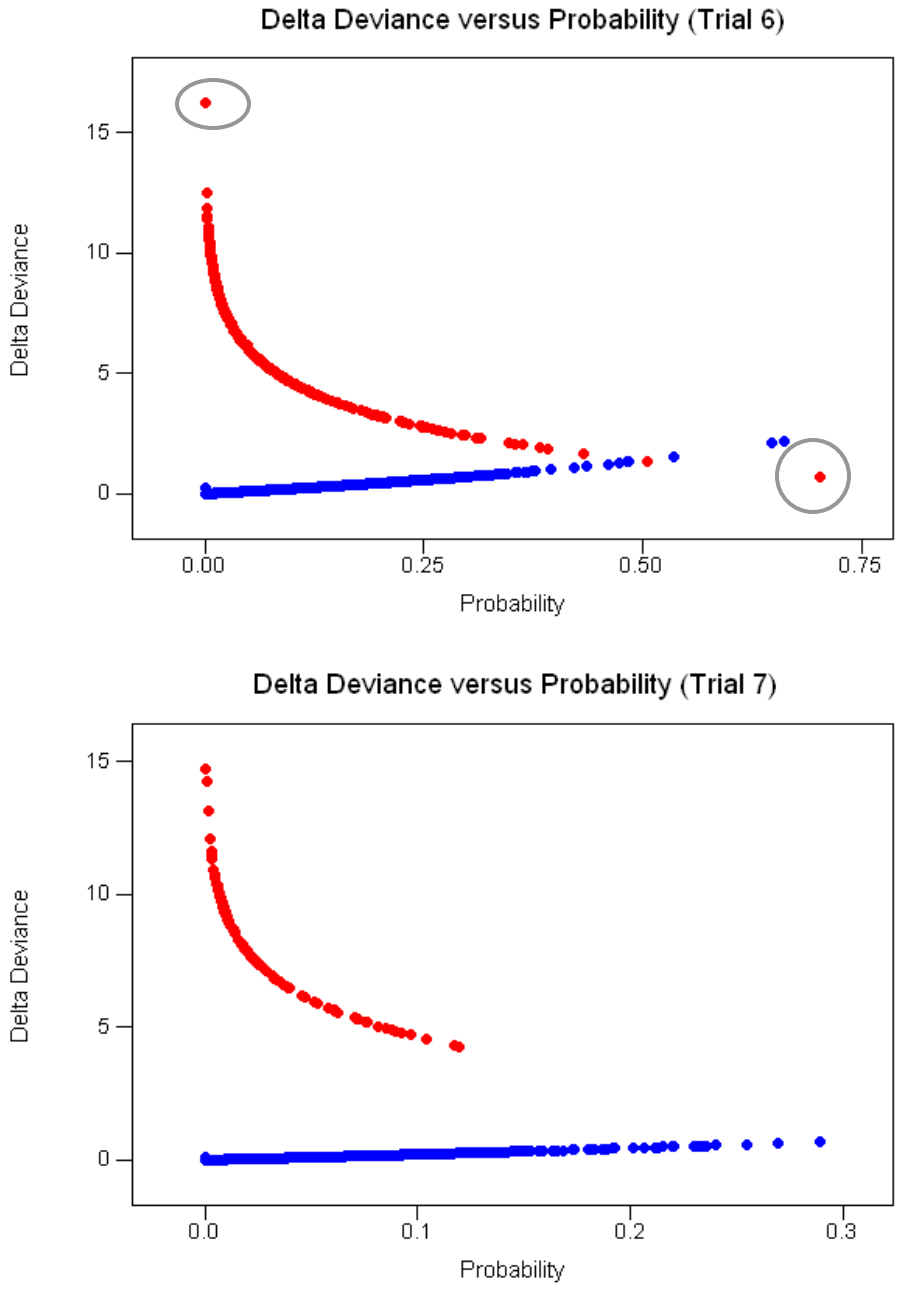


FIGURE 5.26C: DELTA DEVIANCE VS. PROBABILITY [TRIALS 6 AND 7]
The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

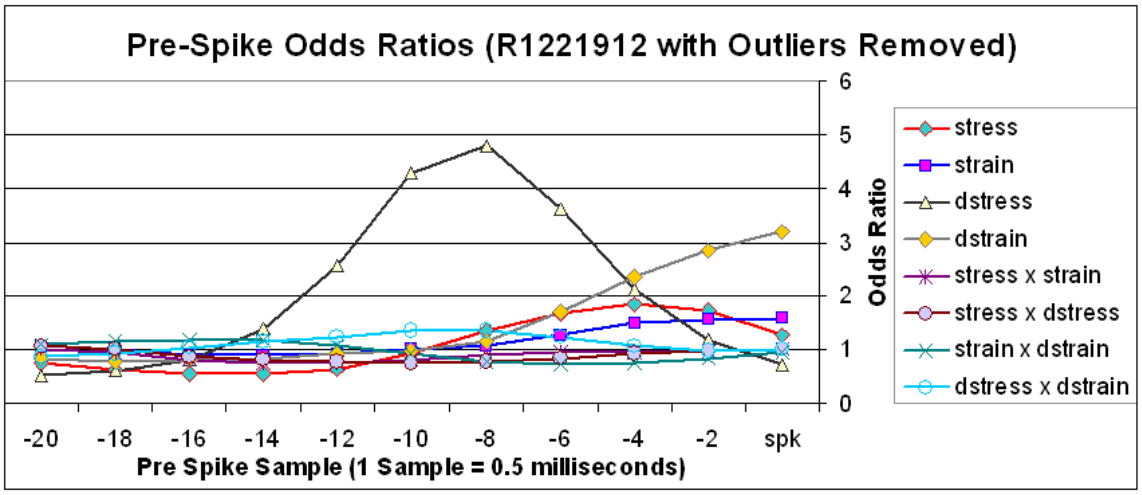


FIGURE 5.27: PRE-IMPULSE ODDS RATIOS WITH DEVIANCE OUTLIERS REMOVED [TRIAL 2]

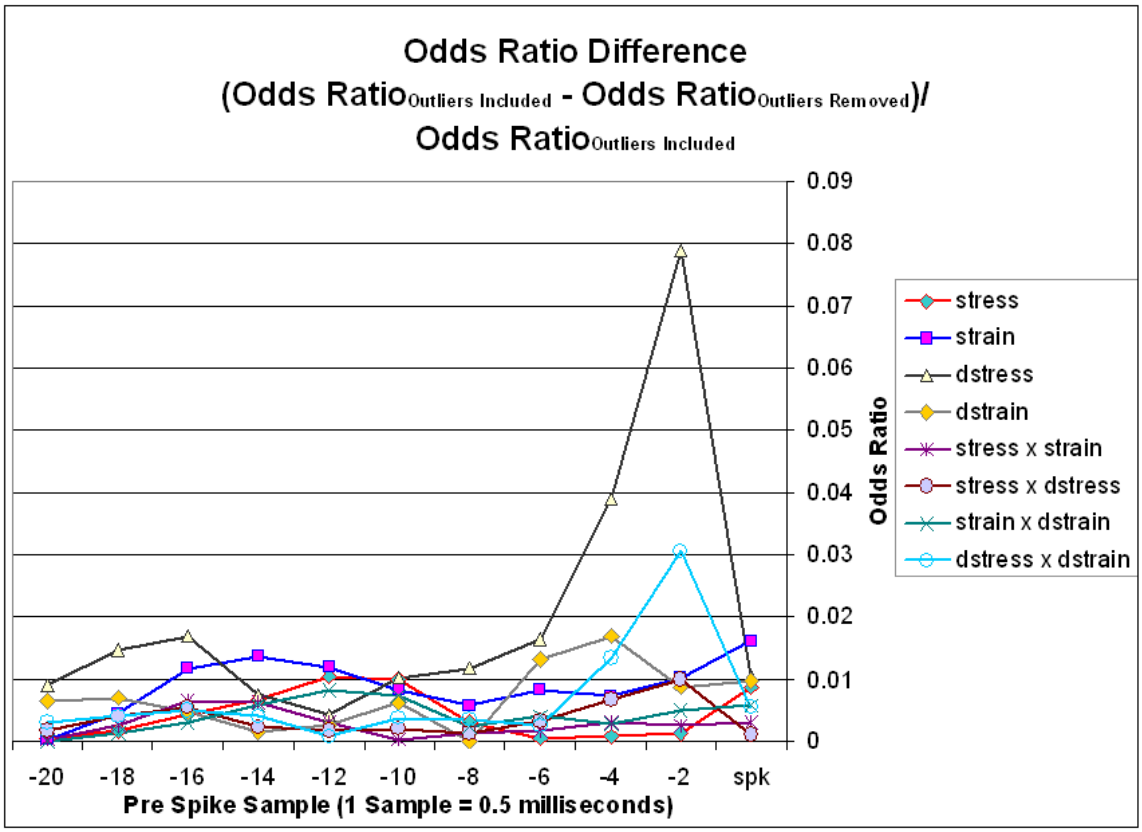


FIGURE 5.28: DIFFERENCE IN PRE-IMPULSE ODDS RATIOS WITH VS. WITHOUT DEVIANCE OUTLIERS [TRIAL 2]

Figures 5.29A through 5.29C show plots of ΔChi-square vs. Probability for nerve R122191X. All the trials have a small number of identifiable outliers, meaning the chi-square fit of the model shows a number of inconsistencies for this nerve. Figure 5.30 shows the pre-impulse odds ratios for trial 6 after the outlier was removed and the odds ratios were re-calculated. The difference between these values and the odds ratios calculated with the outlier included are significant, as shown in Figure 5.31. The maximum difference is approximately 25 % and the average is approximately 5 %. This plot indicates the sensitivity of the logistical regression estimation: a presence or absence of a single positive observation can change the calculated odds ratios by 25 %. Fortunately, each estimated coefficient is affected uniformly so the relative change is negligible. In other words, the incipient properties of the data are not lost if outliers are removed.

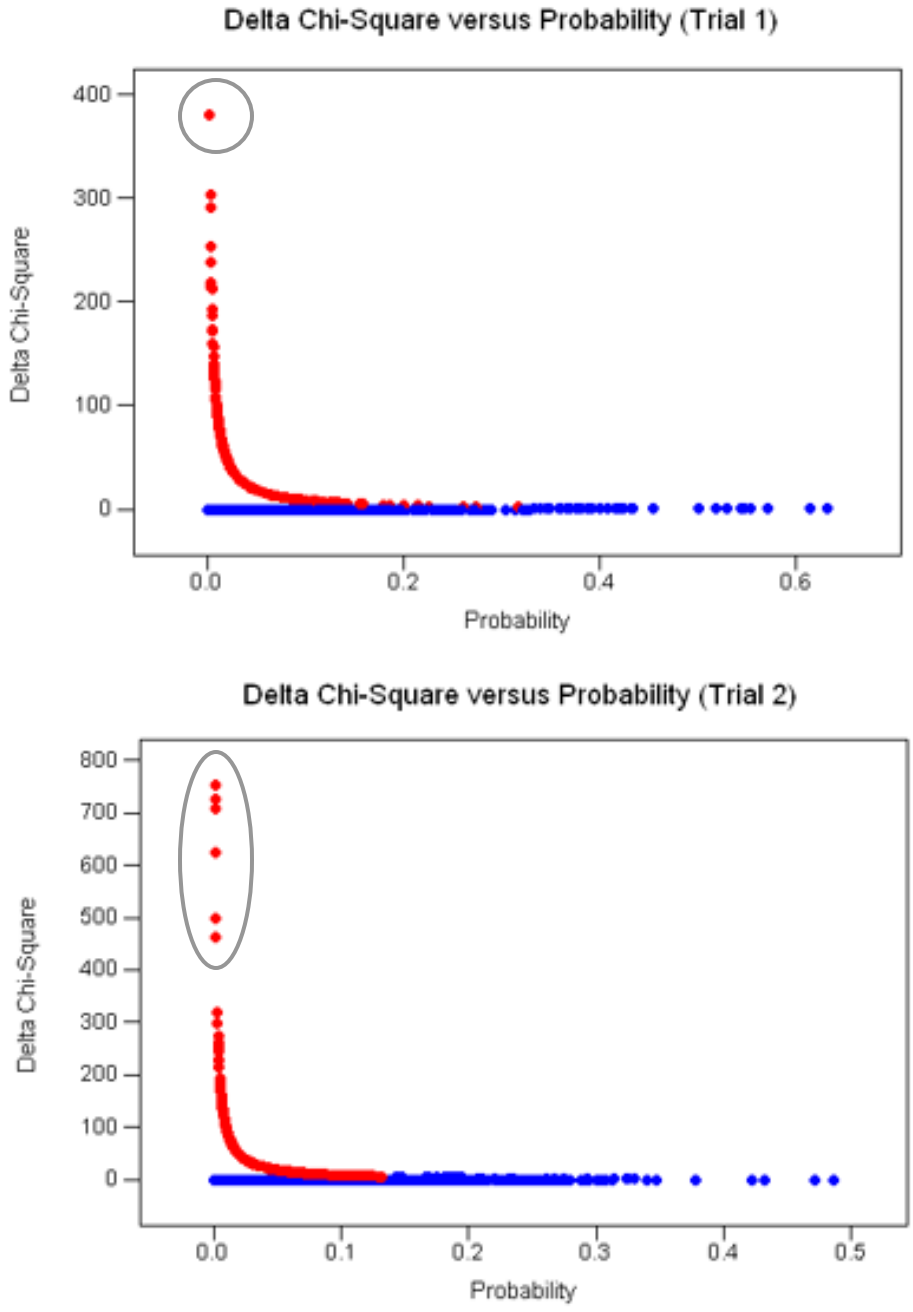


FIGURE 5.29A: DELTA CHI-SQUARE VS. PROBABILITY [TRIALS 1 AND 2]

The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

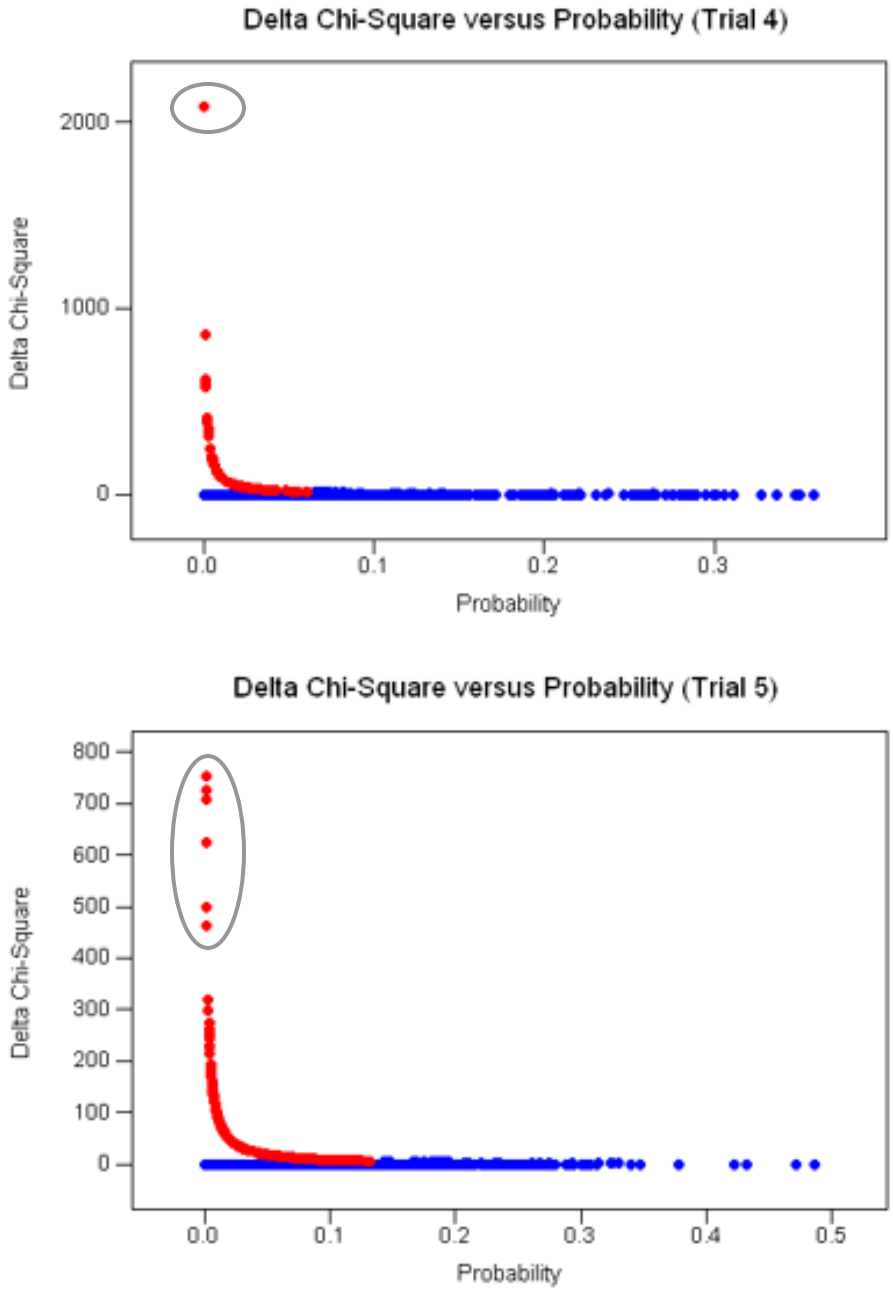


FIGURE 5.29B: DELTA CHI-SQUARE VS. PROBABILITY [TRIALS 4 AND 5]

The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

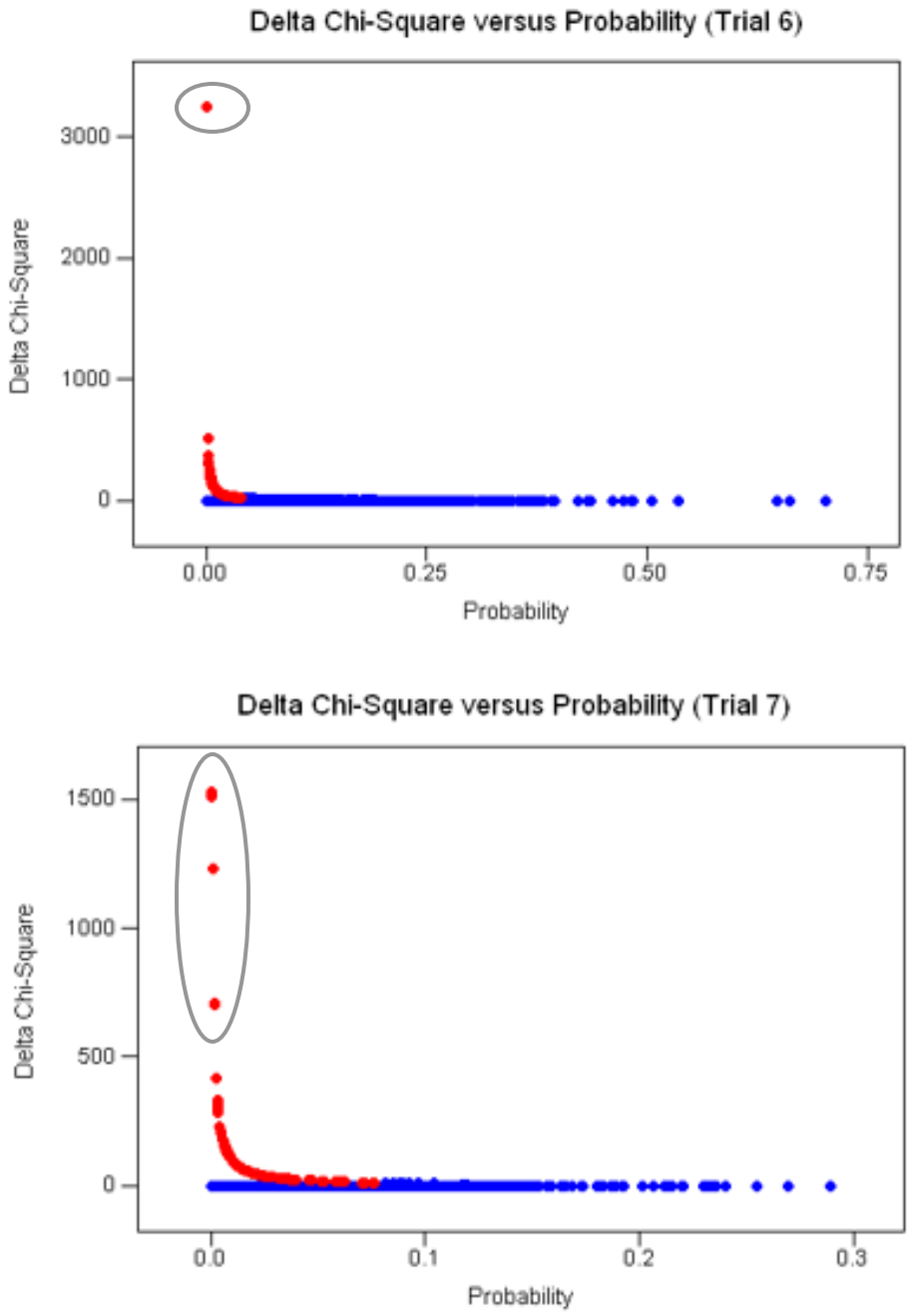


FIGURE 5.29C: DELTA CHI-SQUARE VS. PROBABILITY [TRIALS 6 AND 7]
The RED dots indicate positive (Y = 1) observations. The BLUE dots represent binary 0 outcomes (Y = 0). Possible outliers are circled in GRAY.

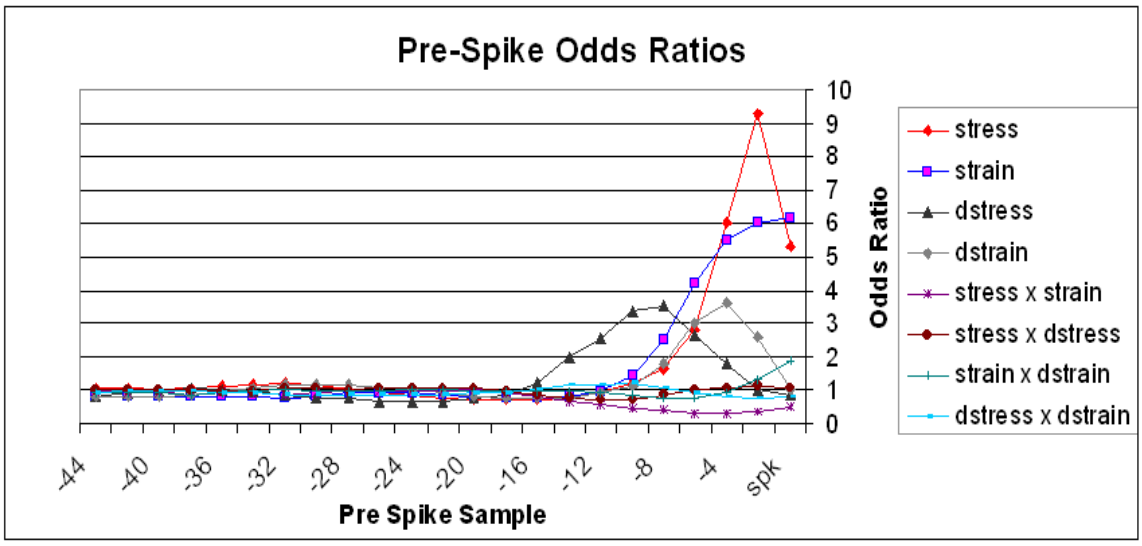


FIGURE 5.30: PRE-IMPULSE ODDS RATIOS WITH CHI-SQUARE OUTLIERS REMOVED [R1221912]

Contrast to Figure 5.1, Trial 2 [R1221912]

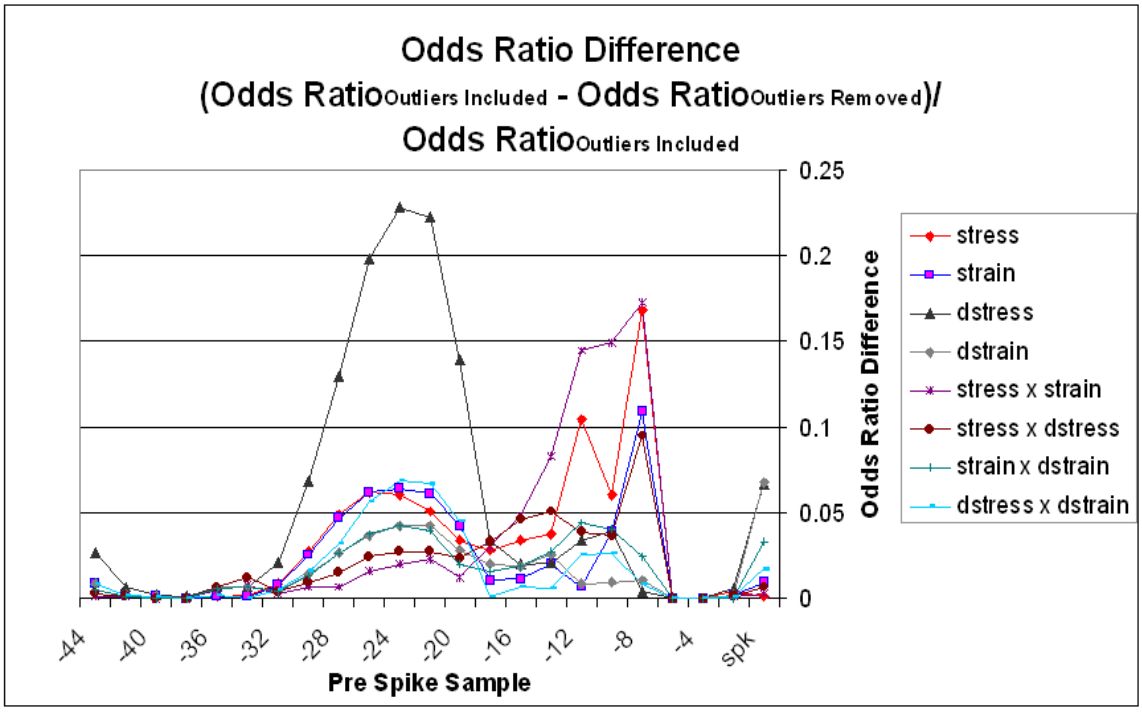


FIGURE 5.31: DIFFERENCE IN PRE-IMPULSE ODDS RATIOS WITH VS. WITHOUT CHI-SQUARE OUTLIERS [R1221912]

Peaks represent regions where there is a large discrepancy between the fitted model with outliers vs. without outliers.

5.1.6 Goodness of Fit

One can gain a visual impression of the accuracy of the logistical model by generating a plot of the probability of an impulse as determined by the logistic function with estimated coefficients vs. time and observed impulses vs. time. A 2.25-second sample of this plot is shown in Figure 5.32. This analysis is purely visual and subjective; however, it can point to regions of inaccurate predictions that warrant further examination. One can see that where the predicted probability peaks, a spike is consistently observed. This agreement between observed and predicted values means the model can predict spikes with relative accuracy. On the other hand, many spikes are observed where the fitted model predicts that no spike should be observed, given a cut value of 0.2, for example. This inaccuracy is described quantitatively in the contingency tables that were generated for this fitted model and are discussed below. Figure 5.33 illustrates the relative prediction error of the model vs. time. Large BLUE peaks indicate predictive inefficiency.

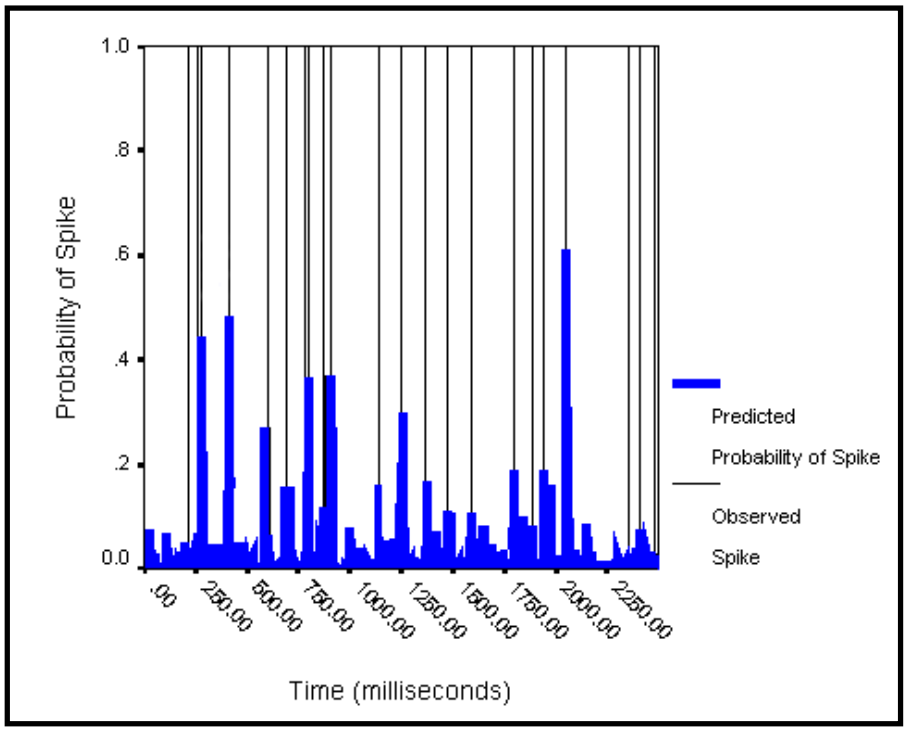


FIGURE 5.32: SPIKE PROBABILITY VS. TIME WITH OBSERVED SPIKES SUPERIMPOSED
Predicted probability peaks with observed spikes indicate regions where the model has elevated predictive efficiency. Peaks without observed spikes and valleys with observed spikes indicate regions of low predictive efficiency for the model.

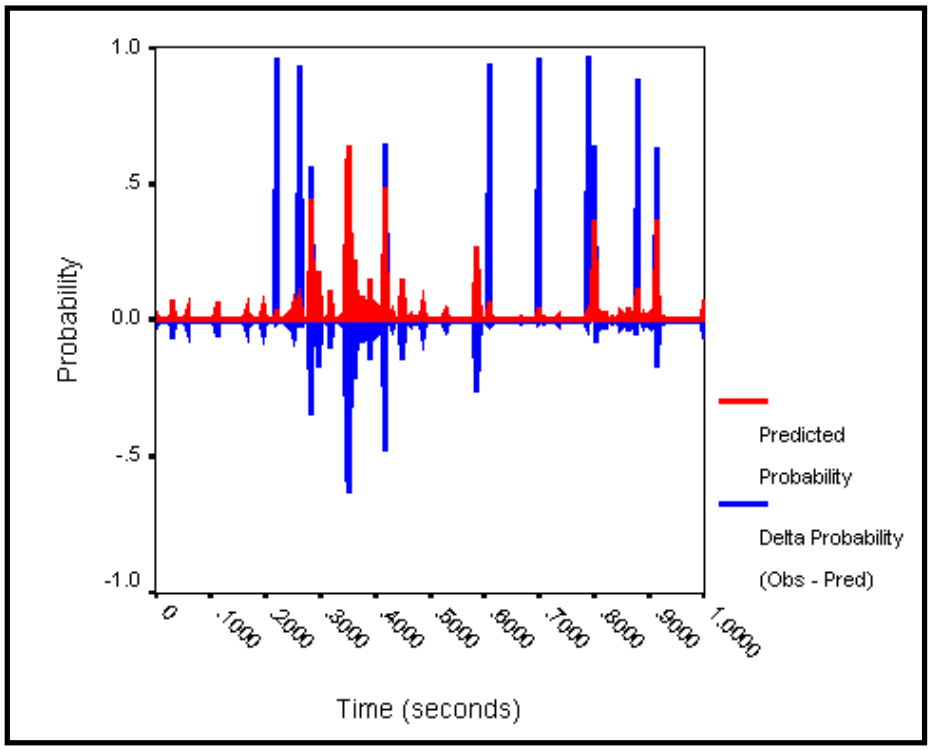


FIGURE 5.33: PREDICTED PROBABILITY AND SUPERIMPOSED PREDICTION ERROR VS. TIME

Blue peaks above 0 indicate regions where a spike was observed when predicted proabability was low. Blue peaks below zero indicate regions where the model predicted spikes but none were observed. The Red indicates the predicted spike probability.

		Trial 1	*Trial 2*	*Trial 4*	*Trial 5*	*Trial 6*	*Trial 7*
Chi-square[1]		897.254	1154.679	787.228	1488.922	1765.38	784.734
-2 log-likelihood[2]		1184.215	2535.886	859.103	1566.53	1964.62	729.905
Hos & Lem Chi-square[3]		15.028	54.858	1.671	1.822	13.163	0.207
***Contingency Table*[4]**							
spk = 0	**observed**	1010	913	1050	906	835	1062
	predicted	1026.138	924.143	1053.872	910.421	842.847	1062.55
	% error	1.5978218	1.220482	0.3687619	0.4879691	0.93976	0.05151
spk = 1	**observed**	187	284	147	291	362	135
	predicted	170.862	272.857	143.128	286.579	354.153	134.453
	% error	8.6299465	3.923592	2.6340136	1.519244	2.16768	0.40519

1. **Calculated as 2*[Log-likelihood$_{\text{final-model}}$ – Log-likelihood$_{\text{initial-model}}$], 0 represents a perfect fit.**
2. **Calculated using Equation 9 or 10, 0 represents a perfect fit.**
3. **Calculated using Equation 38, 0 represents a perfect fit.**
4. **Described in Section 3.3.4, 0 % error represents a perfect fit.**

TABLE 5.1: NERVE R122191X GOODNESS-OF-FIT SUMMARY STATISTICS

Other quantifications of the "goodness-of-fit" of the fitted model for each trial are shown in Table 5.1. The average % error for all trials for the [observed = 0, predicted = 0] category for group 10 of the Hosmer & Lemeshow *G x 2* contingency table is 0.78 %. The [observed = 0, predicted = 0] category showed higher average % error at 3.21 %. These values represent the predictive efficiency of only a subset of all the observed and predicted values, meaning these values are not representative of the overall predictive efficiency of the fitted model. The predictive efficiency of all the data samples is indicated in Figure 5.27 below. This plot shows *specificity* and *sensitivity* as a function of the cut value in a 2 x 2 contingency table. The most efficient prediction occurs at low cut values. This result makes sense given the small number of observed nerve responses. The fact that less than 1 % of all samples correspond to a nerve response would be an issue if stepwise regression techniques were employed, i.e., failure to reject the null model since it would be 99 % accurate. Additionally, trials 4 and 7 show lower chi-square and log-likelihood values than the other trials, indicating the fit of the model in those trials is more accurate than the fit of the model in the other trials.

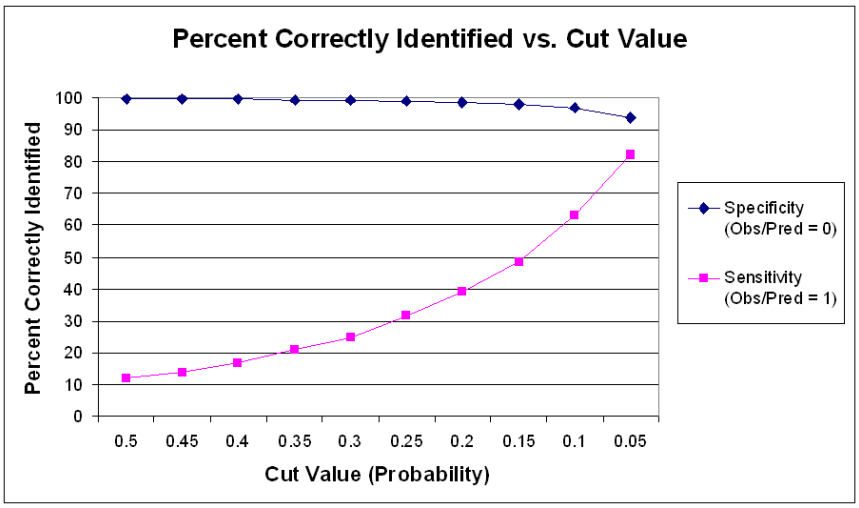


FIGURE 5.27: INFLUENCE OF 2 X 2 CONTINGENCY TABLE CUT VALUE ON PREDICTIVE EFFICIENCY

This graph shows that as the cut value that differentiates predicted responses as either 0 or 1 decreases, the efficiency in the model's prediction increases significantly in the [Observed = 1, Predicted = 1] group but decreases slightly in the [Observed = 0, Predicted = 0] group.

5.1.7 General Observations

Nerve R122191X is preferentially sensitive to *dstress/dt* and *dstrain/dt* at low levels of stimulus intensity. At higher levels of stimulus intensity and with a stimulus tip of smaller area, the sensitivity to *dstress/dt* diminishes and *stress*, *strain*, and *dstrain/dt* sensitivity predominates. High intensity stimuli produced estimated coefficients of higher-than-average standard error and inconclusive Wald quantifiers. Stress-relaxation and creep within the skin did not significantly affect the relative sensitivity of this nerve.

5.2 Multiple Neuron Analysis

In addition to the nerve profiled in the previous section, 9 other nerves were analyzed in the same fashion. This section will summarize the patterns observed in the regression analysis of these nerves.

5.2.1 Memory Effects

Tables 5.2 through 5.10 are tables of the nerve sensitivities predicted by the extended logistical regression model. The color-coding scheme in each chart is as follows:

	Tip T1
	Tip T2
	Tip T3
	Force-controlled
	No sensitivity
	Relatively Elevated

All trials are position-controlled unless specified as force-controlled by the above coding. The "*No Sensitivity*" rating means that no independent variables showed odds ratios above 2 in the span of samples up to 44 samples prior to the observed spike. The "*Relatively Elevated*" rating means the value in question is considerably higher than other values at the same pre-spike offset. This is a purely subjective, qualitative observation. The categories in the chart are described as follows:

- **Trial**: The number of the trial. The stimulus intensity, stimulus tip type, and stimulation type (force or position-controlled) are varied among different trials.

- **Tip**: The tip type used in the experiment, T1, T2, or T3.

- **Intensity**: Intensity is one of 4 types: *threshold, 2 x threshold, 4 x threshold, 6 x threshold, and 8 x threshold*. These values are qualitative assessments of the intensity of the stimulus as measured through voltage feedback from the indenter tip position or force.

- **Sensitivity**: Sensitivity refers to the independent variable that shows a strong correlation to the response. A strong correlation corresponds to an odds ratio peak of 2 or higher.

- **Offset**: This category represents the offset value (in pre-spike samples) corresponding to the peak of the odds ratio profile for the independent variable listed under *sensitivity*.

- **Odds Ratio**: The peak value that the odds ratio for the independent variable listed under *sensitivity* achieves up to 44 samples prior to the observed spike.

- **StdErr**: The standard error of the Beta coefficient associated with the independent variable listed under *sensitivity* at the given pre-spike sample offset. The standard error is the square root of the variance of the Beta coefficient, meaning a lower value (relative to the standard errors associated with the other fitted parameters in the model) represents a better fit.

- **Wald**: The Wald value of the Beta coefficient associated with the independent variable listed under *sensitivity* at the given pre-spike sample offset. The Wald values quantify the significance of the Beta values within the fitted model. A larger value (relative to the Wald values associated with other fitted parameters) indicates more significance within the model.

- **-2 Log-like**: The –2 Log-likelihood value of the fitted model corresponding to the given offset. A smaller value relative to the other –2 Log-likelihood values represents a better fit of the model to the data. Zero represents a perfect fit.

- **Chi-square**: The Chi-square value of the fitted model corresponding to the given offset. A smaller value relative to the other chi-square values represents a better fit of the model to the data. Zero represents a perfect fit.

- **Comments**: General comments about the recording or odds ratio profile.

Each nerve recording is detailed in the subsequent sections. Unless otherwise specified, each recording consists of 60000 samples with approximately 1 % of samples corresponding to nerve responses.

R010702X

Nerve R010702X was not consistently sensitive to any one or more of the independent variables in the first 4 trials as shown in Table 5.2. In trials 5, 6, and 7, however, the nerve shows sensitivity to *stress* and *dstress/dt* at a similar offset from the observed spike. The *stress* sensitivity occurs approximately 14 samples (7 milliseconds) pre-spike. The *dstress/dt* sensitivity occurs at the instant of the observed spike. The *dstress/dt* variable shows elevated significance within the fitted model without elevated levels of standard error as indicated by the Wald and standard error values for *dstress/dt*, meaning the estimated coefficient within the model is a good fit for the *dstress/dt* data. Trials 2 and 4 show *dstrain/dt* sensitivity at an offset of approximately 20 samples pre-spike as well as elevated Wald values without elevated standard errors.

R010702X										
Trial	Tip	Intensity	Sensitivity	Offset	Odds Ratio	StdErr	Wald	-2 Log-like	Chi-Square	Experiment Comments
1	T1	Thresh	*None*							
2	T1	2x	*dSTRAIN/dt*	-20	3.5	0.1	170	2243.7	470.5	*Long, Jagged peak*
	T1	2x	*dSTRESS/dt*	-42	3.5	0.1	0.5	1982.3	233.3	
3	T1	2x	*STRESS*	-32	23	0.2	128	4431.3	582.1	*Long, Jagged Peak*
4	T1	Thresh	*STRESS*	-20, spk	12	0.5, 0.5	30, 20	2138.7	552.7	*2 Long, Jagged Peaks*
	T1	Thresh	*dSTRAIN/dt*	-18	4	0.15	82	1767.5	459.9	*Smooth*
5	T1	2x	*STRESS*	-18	8.5	0.25	73	2717.7	909.9	*Smooth Peak, Sharp Rise*
	T1	2x	*STRAIN*	-4	4	0.4	9	2653.2	985.0	*Smooth*
	T1	2x	*dSTRESS/dt*	spk	7	0.1	200	2692.2	946.0	
6	T1	2x	*STRESS*	-14	4	0.2	57	4594.5	898.7	*Identical to 5*
	T1	2x	*STRAIN*	-2	2	0.2	9	4342.0	1218.8	
	T1	2x	*dSTRESS/dt*	spk	4.5	0.1	240	4287.5	1273.3	
7	T1	Thresh	*STRESS*	-12	4	0.25	32	2232.2	459.3	*Identical to 5*
	T1	Thresh	*dSTRESS/dt*	spk	5	0.1	183	2092.4	610.5	

TABLE 5.2: R010702X SENSITIVITY CHART

R010701X

The stimulated using force-controlled actuator movements, nerve R010701X responds readily to *dstress/dt* stimulation (trials 9, G, and H) as indicated in Table 5.3. In trials 9 and H, the Wald value of the fitted coefficient is elevated while the standard error is not, indicating a good model fit. When stimulated using position-controlled movements, the nerve responds to *stress* (trials 1-7, A, B, D), *dstress/dt* (trials 2, 3, 4, 6, 7, 8, A, B, D, E, F), and *dstrain/dt* (trials 1, 2, 3, 5, 6, 7, A, E). The *stress* sensitivity typically falls in the range between the observed spike and 6 samples (up to 3 milliseconds) prior to the observed spike. The *dstrain/dt* sensitivity occurs approximately 12 samples (6 milliseconds) pre-spike. The *dstress/dt* sensitivity typically occurs in the same sample as the observed spike. The *dstress/dt* odds ratio curve also shows dual odds ratio peaks in trials 6, 7, 9, E, and G ranging from 6 to 24 samples pre-spike. In several cases, the odds ratio quantified sensitivities of the nerve when stimulated with Tip T1 were often an order of magnitude lower than those when Tip T2 was used. Additionally, the Wald values for the *dstress/dt* sensitivity were consistently elevated when Tip T1 was used, emphasizing the significance of the nerve's sensitivity to *dstress/dt* when stimulated with a smaller diameter actuator tip.

R010701X										
Trial	Tip	Intensity	Sensitivity	Offset	Odds Ratio	StdErr	Wald	-2 Log-like	Chi-Square	Experiment Comments
1	T2	Thresh	*dSTRAIN/dt*	-8	100	0.31	224	2201.9	1189.0	
	T2		*STRESS*	-4	25	0.34	102	2129.9	1271.9	
2	T2	2x	*STRESS*	-2	20	0.27	120	3951.3	1830.3	
	T2		*dSTRESS/dt*	spk	16	0.18	226	3896.8	1884.8	
	T2		*dSTRAIN/dt*	-10	14	0.2	186	4639.6	1113.3	
3	T2	4x	*dSTRESS/dt*	spk	5.5	0.12	200	5322.2	1959.0	
	T2		*STRESS*	-2	3	0.2	34	5114.8	2166.5	
	T2		*STRAIN*	-2	3	0.14	0.3	5114.8	2166.5	
	T2		*dSTRAIN/dt*	spk	3	0.13	10	5322.2	1959.0	
4	T2	2x	*STRESS*	spk to -6	110	0.62	61	1184.8	681.7	
	T2		*dSTRESS/dt*	-6	32	0.54	41	1191.2	675.4	
5	T2	4x	*STRESS*	spk	110	0.44	112	2155.4	1202.7	
	T2		*dSTRAIN/dt*	-10	51	0.38	109	2323.5	1034.6	
	T2		*STRAIN*	-12	30	0.47	53	2378.3	979.8	
6	T2	2x	*dSTRESS/dt*	-24, spk	9, 8	0.27, 0.26	16, 66	3240.7	1643.7	*Dual Peak*
	T2		*dSTRAIN/dt*	-14	5.5	0.27	38	4141.5	722.9	
	T2		*STRESS*	-2	5.5	0.22	55	3225.8	1658.6	
7	T2	Thresh	*dSTRESS/dt*	-24, -6	55, 230	0.5, 1	13, 2	823.7	832.3	*Dual Peak, Strong Sensitivity*
	T2		*STRESS*	-4	293	0.95	36	823.7	832.3	*Strong Sensitivity (spk to -8)*
	T2		*STRAIN*	spk	145	1	25	833.4	822.6	*Strong Sensitivity (spk to -8)*
	T2		*dSTRAIN/dt*	-8	130	0.68	52	906.8	749.3	*Strong Sensitivity (spk to -8)*
8	T2	4x	*dSTRESS/dt*	spk	3	0.06	312	7758.6	2400.8	*Lifting, Discard*
9	T1	2x	*dSTRESS/dt*	-6, spk	8, 4	0.25, 0.2	65, 50	998.9	904.3	*Dual Peak*
	T1		*STRAIN*	spk	4.5	0.22	46	998.9	904.3	
a	T1	Thresh	*dSTRESS/dt*	spk	6	0.13	175	2785.3	1065.1	
	T1		*dSTRAIN/dt*	-12	3	0.14	75	3387.4	441.9	
	T1		*STRESS*	-2	3	0.19	35	2699.9	1150.5	
b	T1	2x	*dSTRESS/dt*	spk	7	0.09	542	4616.5	1861.4	*2 Units*
	T1		*STRESS*	spk to -8	2.5	0.21	23	4616.5	1861.4	
c	T1	Thresh	*None*							
d	T1	2x	*dSTRESS/dt*	spk	3	0.13	76	3045.3	804.3	
	T1		*STRESS*	-2	2.5	0.21	19	3097.6	752.1	
e	T1	Thresh	*dSTRESS/dt*	spk, -12	4, 2.5	0.12, 0.14	141, 36	3460.9	1293.4	
	T1		*dSTRAIN/dt*	-24	2.5	0.16	31	4067.7	209.4	
f	T1	2x	*dSTRESS/dt*	spk	3	0.05	573	6657.1	1610.9	*2 Units*
g	T1	Thresh	*dSTRESS/dt*	spk, -16	2.5, 2	0.15, 0.09	40, 55	2304.4	1086.2	
	T1		*dSTRAIN/dt*	-26	2.5	0.1	10	3091.5	146.7	
h	T1	2x	*dSTRESS/dt*	spk	2.5	0.05	422	4872.0	1404.3	

TABLE 5.3: R010701X SENSITIVITY CHART

R011901X

As shown in Table 5.4, under position-controlled stimulation, this nerve shows preferential sensitivity to *stress* (trials 1 – 6, B, C), *strain* (trials 1 – 6, A, B, D), *dstress/dt* (trials 1 – 6, A, B, C), and *dstrain/dt* (trials 1 – 6, B, D). The *stress* sensitivity has a low standard deviation and is consistently in the range between the observed spike and 8 samples prior. The *strain* sensitivity has a higher standard deviation and falls within a larger range from the observed spike up to 14 samples prior. The *dstress/dt* and *dstrain/dt* sensitivities fall between 4 and 14 pre-spike samples with equivalent standard deviations. Of these sensitivities, only *dstress/dt* showed consistently elevated significance (Wald values) in the fitted model. Both *stress* and *strain* showed elevated Wald values in approximately 60 % of the trials. Under force-controlled stimulation, this nerve indicates weak sensitivity (2 < odds ratio < 3.5) to both *dstress/dt* and *dstrain/dt* in a range from 12 to 30 pre-spike samples. The degree of the sensitivity predicted by the fitted model is influenced, in part, by the stimulus tip used. Trials in which Tip T2 was used show consistently and significantly higher odds ratios than those trials in which Tip

T3 was used. In trials 1 – 6 and B, the sensitivities are influenced partially by derivative effects. This phenomenon is apparent due to the consistent grouping of *dstress/dt* and *dstrain/dt* some number of samples from the terms from which they are derived, *stress* and *strain*.

R011901X										
Trial	**Tip**	**Intensity**	**Sensitivity**	**Offset**	**Odds Ratio**	**StdErr**	**Wald**	**-2 Log-like**	**Chi-Square**	**Experiment Comments**
1	T2	Thresh	*STRAIN*	spk to -6	50	0.33	130	2740.5	1402.4	*Low SNR*
	T2		*STRESS*	-2 to -4	35	0.34	125	2704.2	1438.7	
	T2		*dSTRESS/dt*	-10	22	0.28	125	2624.3	1518.6	
	T2		*dSTRAIN/dt*	-8	17	0.28	104	2563.9	1579.0	
2	T2	Thresh	*dSTRESS/dt*	-8	87	0.26	305	4184.7	2395.3	
	T2		*STRAIN*	spk to -6	17	0.22	220	4711.0	2080.5	
	T2		*STRESS*	spk to -8	18	0.22	180	4711.0	2080.5	
	T2		*dSTRAIN/dt*	-8	12	0.26	127	4184.7	2395.3	
3	T2	2x	*STRESS*	-2	9	0.14	255	6377.1	2713.8	*2 Units, Derivative Effects*
	T2		*STRAIN*	-6	8	0.18	136	5977.1	2629.1	
	T2		*dSTRESS/dt*	-10	6.5	0.09	418	6103.3	2339.6	
	T2		*dSTRAIN/dt*	-6	3	0.11	106	5977.1	2629.1	
4	T2	Thresh	*dSTRESS/dt*	-12	30	0.28	142	2913.1	1353.7	*Derivative Effects*
	T2		*dSTRAIN/dt*	-12	21	0.25	152	2913.1	1353.7	
	T2		*STRESS*	-8	18	0.38	60	2924.8	1372.9	
	T2		*STRAIN*	spk to -12	8	0.34	39	3000.6	1317.7	
5	T2	2x	*dSTRESS/dt*	-14	21	0.22	188	3977.1	1419.0	*2 Units, Derivative Effects*
	T2		*STRESS*	spk	20	0.24	156	4364.4	1758.7	
	T2		*dSTRAIN/dt*	-10	9	0.18	153	4044.3	1516.5	
	T2		*STRAIN*	-14	9	0.27	66	3977.1	1419.0	
6	T2	Thresh	*STRESS*	-6	5202	0.53	259	4068.7	3400.9	
	T2		*STRAIN*	-4	1570	0.38	379	4329.6	3388.8	
	T2		*dSTRESS/dt*	-8	407	0.31	384	4048.0	3403.7	
	T2		*dSTRAIN/dt*	-10	16	0.21	178	4260.2	3182.6	
7	T2	Thresh	*dSTRAIN/dt*	-12	3.5	0.15	73	3362.7	842.3	*1 Unit*
	T2		*dSTRESS/dt*	-24 to -30	2	0.15	10	3855.1	89.9	
8	T2	2x	*None*							
9	T2	2x	*None*							
a	T3	2x	*STRAIN*	-4	3.5	0.1	172	4028.8	1162.2	
	T3		*dSTRESS/dt*	-4	3	0.1	102	4028.8	1162.2	
b	T3	4x	*STRAIN*	-2	5	0.08	423	5931.6	1546.9	*2 Units, Derivative Effects*
	T3		*STRESS*	-2	2.5	0.12	61	5931.6	1546.9	
	T3		*dSTRAIN/dt*	-14	2.5	0.08	128	5360.0	791.3	
	T3		*dSTRESS/dt*	-12	2.5	0.06	205	5350.9	1053.0	
c	T3	2x	*STRESS*	spk to -4	25	0.24	185	2706.9	1153.9	
	T3		*dSTRESS/dt*	-20	17	0.23	160	2690.4	1001.1	
d	T3	4x	*dSTRAIN/dt*	-16	6	0.1	348	5796.7	764.7	
	T3		*STRAIN*	-4 to -8	3	0.06	380	5495.0	1424.5	
e	T3	2x	*dSTRESS/dt*	-10 to -16	2	0.09	64	6132.0	833.0	*Multiple Units*
	T3		*dSTRAIN/dt*	-26	2	0.12	10	5783.1	65.3	
f	T3	2x	*dSTRESS/dt*	-12	2	0.1	69	4991.3	598.4	
	T3		*dSTRAIN/dt*	-28	2	0.13	10	4904.9	68.9	
	T3		*STRAIN*	-2	2	0.1	52	5047.5	1132.0	

TABLE 5.4: R011901X SENSITIVITY CHART

R020902X

As indicated by Table 5.5, trials 4 – E of R020902X are either weakly sensitive to one of the independent variables (trials 7 and B) or not sensitive at all (trials 4 – 6, 8 – A, D, E). This lack of detectable sensitivity is due, in part, to weak signal-to-noise ratios in the recordings as well as multiple receptor action potentials detected on the same nerve axon. The other trials (trials 1 – 3) show inconsistent sensitivity strengths (trial 1 odds ratios:

stress = 16,826, *strain* = 3495, *dstrain/dt* = 755; trial 2 odds ratios: *stress* = 753). Based on these results, the logistical regression analysis of this nerve is inconclusive.

R020902X										
Trial	Tip	Intensity	Sensitivity	Offset	Odds Ratio	StdErr	Wald	-2 Log-like	Chi-Square	Experiment Comments
1	T2	Thresh	*STRESS*	-2	16826	0.85	131	1832.3	1569.5	
	T2		*STRAIN*	spk	3495	0.86	89	1891.1	1510.7	
	T2		*dSTRAIN/dt*	-8	755	0.56	139	1902.4	1499.4	
	T2		*dSTRESS/dt*	-8	55	0.56	49	1902.4	1499.4	
2	T2	2x	*STRESS*	-4	753	0.42	249	4071.6	2646.4	
	T2		*dSTRESS/dt*	-2	53	0.25	234	4098.1	2620.0	
	T2		*STRAIN*	-6	41	0.34	121	4117.4	2600.6	
	T2		*dSTRAIN/dt*	-6	51	0.25	248	4117.4	2600.6	
3	T2	4x	*STRAIN*	spk	9	0.17	159	5556.8	930.6	
	T2		*STRESS*	-10	3	0.23	26	5019.8	1458.3	
	T2		*dSTRAIN/dt*	-4	2.5	0.1	84	5051.5	1435.9	
4	T2	6x	*STRESS*	-8	3.5	0.12	111	4933.1	1554.3	
5	T2	8x	*None*							*2 Units*
6	T2	8x	*None*							*Multiple Units*
7	T2	4x	*dSTRESS/dt*	-4	2.5	0.06	221	3514.5	772.9	*Low SNR*
8	T2	2x	*None*							
9	T2	6x	*None*							*Multiple Units*
a	T2	8x	*None*							*Multiple Units + Noisy*
b	T2	6x	*STRAIN*	-16	2	0.18	25	2404.8	103.2	
	T2		*dSTRESS/dt*	-8	2	0.08	86	2314.6	456.3	
c	T2	6x	*STRAIN*	spk	2	0.07	99	9156.2	1641.9	
d	T2	8x	*None*							
e	T2	8x	*None*							*Multiple Units*

TABLE 5.5: R020902X SENSITIVITY CHART

R120891X

When force-control stimulated, nerve R120891X is weakly sensitive to *dstress/dt* (trials 4, 8, 9, C, D, E) as shown in Table 5.6. When position-control stimulated, the nerve exhibits preferentially sensitivity to *stress* and *dstress/dt* as well as a weaker sensitivity to *strain* and *dstrain/dt*. Of the position-controlled sensitivities, *stress* [mean = 5, standard deviation = 3.5], *strain* [mean = 12, standard deviation = 5.5], and *dstress/dt* [mean = 9, standard deviation = 3.5] show considerable disparity while *dstrain/dt* [mean = 10.5, standard deviation = 2] showed much less disparity in pre-spike offsets corresponding to the peak odds ratio. Although the *dstrain/dt* sensitivity was weaker than the *stress* or *dstress/dt* sensitivity, particularly at intensity levels above threshold (trials 3 and B), the consistency of the sensitivity lends credibility to this independent variable as being the primary sensitivity of this nerve. Additionally, the nerve's sensitivity to *dstress/dt* is substantiated by the consistently elevated Wald values, and, therefore, elevated significance within the fitted model. Trials 1, 2, 3, A, and B exhibit derivative effects, which means the model has mathematical difficulty determining to which covariate the nerve is sensitive: an independent variable or its time derivative.

R120891X										
Trial	Tip	Intensity	Sensitivity	Offset	Odds Ratio	StdErr	Wald	-2 Log-like	Chi-Square	Experiment Comments
1	T2	Thresh	*STRESS*	-2	107	0.55	72	816.5	481.3	
	T2		*STRAIN*	-6	45	0.85	20	817.6	480.2	
	T2		*dSTRESS/dt*	-12	28	0.45	54	793.6	504.2	
	T2		*dSTRAIN/dt*	-10 to -14	20	0.42	52	793.6	504.2	
2	T2	Thresh	*dSTRESS/dt*	-16	28	0.31	117	1601.1	542.5	*Derivative Effects*
	T2		*dSTRAIN/dt*	-12	15	0.3	84	1485.8	657.9	
	T2		*STRESS*	-8	10	0.29	64	1514.2	641.4	
	T2		*STRAIN*	-8	9	0.42	30	1514.2	641.4	
3	T2	2x	*STRESS*	-2, -12	5, 4	0.22, 0.28	60, 28	4035.4	1251.3	*2 Jagged Peaks*
	T2		*dSTRESS/dt*	-4	2.5	0.07	163	4116.8	1150.4	
	T2		*STRAIN*	-8	2	0.15	23	4334.4	893.7	
	T2		*dSTRAIN/dt*	-12	2	0.08	96	4171.7	849.7	
4	T2	2x	*STRESS*	spk	3	0.28	17	969.5	456.8	
	T2		*dSTRAIN/dt*	-2, -6	2	0.09, 0.15	63, 23	1142.0	449.7	
	T2		*dSTRESS/dt*	-4	2	0.14	28	1105.0	474.1	
5	T2	Thresh	*dSTRESS/dt*	-6, -44	3, 3.5	0.16	49	1418.4	374.4	*2 Long, Smooth Peaks*
	T2		*STRAIN*	-42	2.5	0.27	10	1598.7	119.7	*2 Long, Smooth Peaks*
6	T2	Thresh	*STRAIN*	-4	5	0.22	56	2930.6	654.3	
	T2		*dSTRESS/dt*	-2	3.5	0.11	148	2997.6	694	
7	T2	2x	*None*							
8	T2	2x	*dSTRESS/dt*	-32 to -38	2	0.13	10	2763.1	210.2	*Dual Peaks*
	T2		*STRAIN*	spk, -24	2	0.13	36, 17	3070.8	406.6	
9	T2	Thresh	*dSTRESS/dt*	spk, -40	4.5, 2.5	0.13, 0.15	137	5265.2	1017.3	*Dual Peaks*
	T2		*dSTRAIN/dt*	-24	2	0.17	10	5137.9	277.6	
a	T2	Thresh	*dSTRESS/dt*	-12	8	0.13	240	4253.4	1374.8	*Derivative Effects*
	T2		*STRESS*	spk to -10	6	0.14	177	4826.7	1632.9	
	T2		*dSTRAIN/dt*	-10 to -16	3.5	0.12	114	4319	1395.7	
	T2		*STRAIN*	-6 to -8	3.5	0.17	57	4666.4	1550.6	
b	T2	2x	*dSTRESS/dt*	-10	4	0.07	321	6078.4	1310.4	*Derivative Effects*
	T2		*STRESS*	spk	4	0.08	286	7189.6	2178	
	T2		*STRAIN*	-6	3.5	0.11	131	6355.4	1503.8	
	T2		*dSTRAIN/dt*	-4	2.5	0.08	118	6741.6	1769.9	
c	T2	Thresh	*STRESS*	spk	85	0.4	123	1745	797.6	
	T2		*dSTRESS/dt*	-18	22	0.26	143	1831.4	653.5	
	T2		*STRAIN*	spk to -10	6	0.36	26	1745	797.6	
d	T2	Thresh	*dSTRESS/dt*	-16	3.5	0.09	187	3855.1	514.3	*Derivative Effects*
	T2		*STRESS*	spk	2	0.13	28	3719.1	833.6	
e	T2	Thresh	*dSTRESS/dt*	-14	2	0.1	55	5685.2	277.3	

TABLE 5.6: R120891X SENSITIVITY CHART

R122391X

Nerve R122193X, as shown in Table 5.7, is consistently and preferentially sensitive to *strain* and *dstress/dt* when position-control stimulated. Of these two, only *dstress/dt* shows consistently elevated Wald values and low standard errors. When force-control stimulated, the sensitivity is primarily to *dstress/dt*. Two patterns emerge from this chart in terms of the position-controlled stimulation sensitivity: *dstress/dt* and *strain* odds ratios that peak very close to the observed nerve response (up to 4 samples pre-spike) and *dstress/dt* and *strain* odds ratios that peak a considerable amount of time prior to the observed impulse (20 to 40 samples prior). This observation leads one to believe that there are two receptors whose impulse responses are detected by the recording electrode attached to the nerve axon. In trial 2, the *dstress/dt* sensitivity is dual peaked with one peak lying very close to the observed impulse response and the other lying in the second range of observed peaks (20 to 40 samples pre-spike). When recording was taking place, this trial was visually interpreted as having multiple units attached to the recorded axon.

R122391X										
Trial	Tip	Intensity	Sensitivity	Offset	Odds Ratio	StdErr	Wald	-2 Log-like	Chi-Square	Experiment Comments
1	T2	Thresh	*STRESS*	-16	3	0.3	14	2707.9	420.6	*Very Sensitive*
	T2		*STRAIN*	-4	3	0.24	43	2296.5	832	
	T2		*dSTRESS/dt*	-4	2.5	0.12	67	2296.5	832	
2	T2	Thresh	*STRAIN*	-20	3	0.09	147	4164.4	428.8	*2 Units*
	T2		*dSTRESS/dt*	-24, spk	2.5	0.09	100, 111	3648.2	1216.2	*Dual Peaks*
3	T2	Thresh	*dSTRESS/dt*	spk	3	0.14	59	1617.1	551	*Clean*
	T2		*STRAIN*	spk	2.5	0.19	24	1617.1	551	
4	T2	2x	*STRAIN*	-32	4.5	0.1	35	2559.8	535.6	*Clean*
	T2		*dSTRESS/dt*	-40	4.5	0.1	110	2555.5	528.8	
6	T2	Thresh	*dSTRESS/dt*	-10	2.5	0.13	43	3752.3	647.8	*No Lifting*
7	T2	Thresh	*dSTRESS/dt*	-6	3	0.15	53	5901	1127.6	*No Lifting*
8	T2	Thresh	*None*							*No Lifting*
9	T2	Thresh	*dSTRESS/dt*	-20	2.5	0.12	61	3292.7	130.8	
	T2		*STRAIN*	-2	2.5	0.2	22	2612.4	940.3	
a	T2	2x	*STRAIN*	spk	2.5	0.13	48	4619.6	1360.7	*Good*
b	T2	2x	*STRAIN*	spk	3.5	0.21	30	1841.6	585.2	*Good*

TABLE 5.7: R122391X SENSITIVITY CHART

R122194X										
Trial	Tip	Intensity	Sensitivity	Offset	Odds Ratio	StdErr	Wald	-2 Log-like	Chi-Square	Experiment Comments
1	T2	Thresh	*dSTRESS/dt*	-6	188	0.51	104	1270.3	742.1	*Derivative Effects*
	T2		*STRESS*	-4	15	0.51	28	1268.7	743.7	
	T2		*STRAIN*	-2	10	0.44	26	1308	704.4	
	T2		*dSTRAIN/dt*	-6	8	0.49	18	1270.3	742.1	
2	T2	Thresh	*dSTRESS/dt*	-6	2928	0.91	76	790.3	637.3	
	T2		*STRESS*	-4	373	0.91	43	802.2	625.4	
	T2		*STRAIN*	spk	205	0.92	33	841.5	586	
	T2		*dSTRAIN/dt*	-6	74	0.65	44	790.3	637.3	
3	T2	2x	*STRESS*	-2	110	0.48	94	1589.3	999.3	
	T2		*STRAIN*	-2	100	0.57	65	1589.3	999.3	
	T2		*dSTRESS/dt*	-6	12	0.17	188	1711	877.6	
4	T2	4x	*STRESS*	spk	315	0.65	79	1411.6	684.2	
	T2		*STRAIN*	spk	100	0.55	70	1411.6	684.2	
5	T2	2x	*dSTRESS/dt*	-8	78	0.3	196	1946.9	845.4	*2 Units*
	T2		*STRAIN*	-2	51	0.44	81	1993.1	789.3	
	T2		*STRESS*	spk	36	0.45	63	1996.5	785.9	
6	T2	2x	*STRESS*	-2	6.5	0.21	75	2485.7	385	
	T2		*dSTRESS/dt*	-6	6	0.13	159	2223	658.9	
	T2		*STRAIN*	-2	3.5	0.23	26	2485.7	385	
	T2		*dSTRAIN/dt*	-8	3.5	0.12	113	2196.5	640.4	
7	T2	2x	*STRESS*	-2	31	0.33	108	1594.2	844.3	*Noisy*
	T2		*dSTRESS/dt*	-2	6.4	0.21	77	1594.2	844.3	
8	T2	2x	*STRESS*	spk	18	0.28	103	2598.1	1125.4	
	T2		*STRAIN*	spk to -8	5	0.23	37	2598.1	1125.4	

TABLE 5.8: R122194X SENSITIVITY CHART

R122194X

When either force or position-control stimulated, nerve R122194X shows preferential *stress*, *strain*, and *dstress/dt* sensitivity. See Table 5.8. The *dstress/dt* sensitivity profile consistently peaks at approximately 6 pre-spike samples and shows consistently elevated significance in the fitted model (Wald values). The *stress* profile peaks between the sample in which the impulse was observed and 4 samples prior to that point. The *strain* sensitivity is consistently highest at the sample in which the spike was observed. Neither *stress* nor *strain* shows consistent significance in the fitted model and both show elevated standard error on a number of occasions (trials 3 – 6).

R111791X										
Trial	Tip	Intensity	Sensitivity	Offset	Odds Ratio	StdErr	Wald	-2 Log-like	Chi-Square	Experiment Comments
1	T2	Thresh	*STRESS*	-8	10	0.2	131	5265.4	865.1	
	T2		*STRAIN*	-2	3	0.12	76	4773.1	1366.6	
2	T2	2x	*STRESS*	-8	14	0.16	276	8487.1	1190.5	
	T2		*STRAIN*	spk	5	0.1	249	7663.3	2086.7	
3	T2	Thresh	*STRESS*	-8	3.5	0.24	25	3878	528.6	*Clean*
	T2		*dSTRESS/dt*	spk	2	0.1	58	3783.5	782.5	
4	T2	Thresh	*None*							*Good*
5	T2	Thresh	*STRAIN*	-10	2.5	0.11	81	3643.9	409.5	*Messy*
7	T2	Thresh	*STRAIN*	-2	76	0.37	134	2943.1	2073.2	
	T2		*STRESS*	-4	25	0.24	174	2858	2158.4	
	T2		*dSTRAIN/dt*	-4	13.5	0.17	270	2858	2158.4	
	T2		*dSTRESS/dt*	-6	7.5	0.17	112	3019.8	1996.5	
8	T2	Thresh	*STRAIN*	-2	30	0.33	105	2231	1865.9	*Messy + Hysteresis*
	T2		*STRESS*	-2	9	0.23	88	2231	1865.9	
	T2		*dSTRAIN/dt*	-4	7	0.12	266	2274.2	1822.7	
9	T2	2x	*dSTRESS/dt*	-4	2	0.07	149	4435.4	248.7	*Messy + Hysteresis*
a	T2	4x	*STRESS*	-6	5.5	0.21	66	2701.4	319.7	*Messy + Hysteresis*
	T2		*STRAIN*	-14	2	0.15	30	2617.6	216.8	

TABLE 5.9: R111791X SENSITIVITY CHART

R111791X

In Table 5.9, nerve R111791X shows preferential sensitivity to *stress* and *strain* when subjected to position-controlled stimulation. The *stress* sensitivity peaks occurs between 2 and 8 samples prior to the observed spike while the *strain* sensitivity typically occur 2 samples prior. These variables also show frequently elevated variance (standard error) and Wald values, making the model prediction suspect. These inconsistencies are most likely due to poor experimental recordings. The force-controlled records of this nerve are inconclusive due to weak sensitivity and poor recordings.

R111793X

As shown in Table 5.10, the responses of nerve R111793X are weak and inconsistent. The results of this trial, therefore, are not considered indicative of mechanoreceptor sensitivities.

R111793X										
Trial	Tip	Intensity	Sensitivity	Offset	Odds Ratio	StdErr	Wald	-2 Log-like	Chi-Square	Experiment Comments
1	T2	Thresh	*None*							
2	T2	2x	*None*							*Hysteresis*
3	T2	4x	*None*							*Poor*
4	T2	2x	*dSTRAIN/dt*	-4 to spk	3	0.09	120	1765.3	607.6	
5	T2	4x	*dSTRAIN/dt*	-2	2	0.07	139	6778.4	1028	
	T2		*STRAIN*	spk	2	0.11	56	6366.4	1440	
7	T2	Thresh	*None*							
8	T2	Thresh	*STRAIN*	spk	2	0.14	35	5202.3	919.1	
9	T2	Thresh	*dSTRESS/dt*	-8	2	0.15	36	2608.8	236.8	
	T2		*dSTRAIN/dt*	-22	2	0.15	17	2649	85.7	
a	T2	2x	*None*							
c	T2	2x	*None*							

TABLE 5.10: R111793X SENSITIVITY CHART

Memory Effects: Comparison

The results of the comparison of the memory effects of the receptors in this study can be summarized in Table 5.11. In general, receptors subjected to force-controlled stimulation were preferentially sensitive to *dstress/dt*. This sensitivity was determined from 25 independent trials using 7 different nerves (2 of which produced inconclusive results due to poor recordings). Position-control stimulation of these receptors yielded different results: primary sensitivity to *stress* and *dstress/dt* as well as *strain* and *dstrain/dt* to a lesser degree. This sensitivity was determined from 79 independent trials using 9 different nerves (2 of which produced inconclusive results because of poor recordings).

	Position-controlled Sensitivity	***Force-controlled Sensitivity***
R010702X	*Stress, dStress/dt*	*NA*
R010701X	*Stress, dStress/dt, dStrain/dt*	*dStress/dt*
R011901X	*Stress, dStress/dt, dStrain/dt*	*dStress/dt, dStrain/dt*
R020902X	*Inconclusive*	*NA*
R120891X	*Stress, Strain, dStress/dt, dStrain/dt*	*dStress/dt*
R122391X	*Strain, dStress/dt*	*dStress/dt*
R122194X	*Stress, Strain, dStress/dt*	*Stress, Strain, dStress/dt*
R111791X	*Stress, Strain*	*Inconclusive*
R111793X	*Inconclusive*	*Inconclusive*

TABLE 5.11: NERVE SENSITIVITY SUMMARY

5.3 Results: Summary & Discussion

This section serves to put the results of the two previous sections into perspective. First, general trends that were observed during the execution the logistical regression analysis for each receptor considered in this study are summarized. Next, comments are made on the implications of experimental input/output coverage. Finally, the interpretation of the logistical regression results is discussed.

5.3.1 General Observations

There were several noticeable trends in the analysis that warrant discussion:

- The sensitivities of the nerves were consistent for the duration of the individual trials. In other words, the dynamic nature of skin did not have a significant effect on the estimated parameters of the fitted logistical regression model.

- Force-controlled stimulation produced markedly different patterns of sensitivity as compared to position-controlled stimulation. Force-control stimulation produced more consistent results when the experimental recordings were usable.

- The logistical model indicates that the receptors in this study are preferentially sensitive to *dstress/dt* when subjected either to position or force-controlled stimulation and *stress* when subjected exclusively to position-controlled stimulation.

- The size of the actuator tip used to compress the skin surface affected the level of the sensitivity predicted by the logistical regression model. Tip T2 (Area: 16.6 mm^2) got consistently higher, often an order of magnitude higher, sensitivity ratings than either Tip T1 (4.2 mm^2) or T3 (34.2 mm^2).

- High levels of stimulus intensity (6x and 8x) produced inconsistent results. The inconsistency is due to the actuator tip losing contact with the skin surface and creating non-linear input/output combinations.

- The sensitivities of the receptors in these trials were statistically uncorrelated to the interaction terms included in the extended model (*stress x strain*, *stress x dstress/dt*, *strain x dstrain/dt*, and *dstress/dt x dstrain/dt*), yet the interaction terms showed elevated Wald values in most of the trials, indicating that they are significant within the fitted model.

- Removing post-spike data points that correspond to observations within the refractory period of the nerve does not noticeably affect the estimated Beta coefficients, the significance of the estimated coefficients within the fitted model, or the accuracy of the model fit.

5.3.2 Input Space Coverage

Despite the fact that 110 trials were completed, the input space of this experiment was not covered thoroughly. Of the position-controlled trials, 75 % (64 trials) used Tip T2 and 78 % (66 trials) were low-level stimulation (*threshold* and *2 x threshold*). In the force-controlled trials, the disparity was greater with 80 % (20 trials) using Tip T2 and 96 % (24 trials) corresponding to low-level stimulation. These results are shown graphically in Figures 5.28 and 5.29.

The reason for the seemingly incomplete coverage is twofold. First, the receptors in question often became unresponsive or lifeless after only a few trials were completed. Second, the recording system was subject to transient noise, which often rendered trials inconclusive. These trials were identified manually and then discarded.

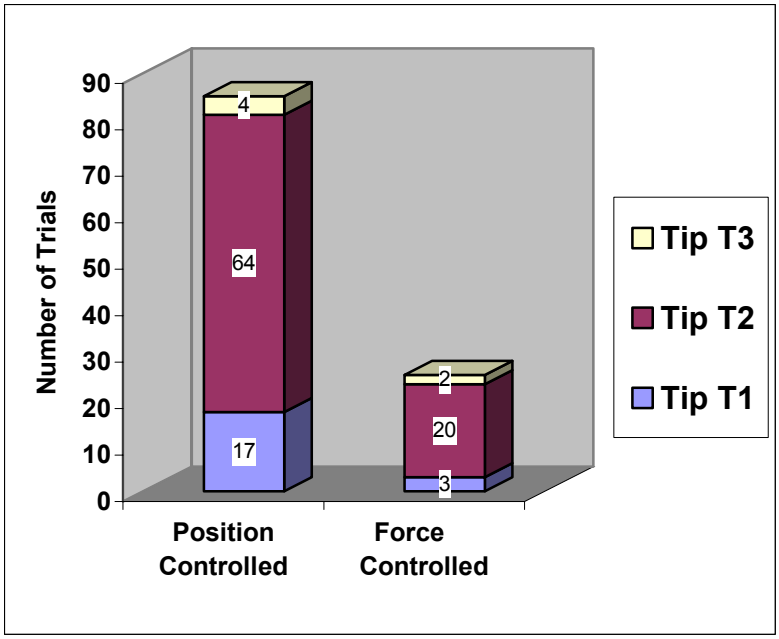


FIGURE 5.28: EXPERIMENTAL INPUT SPACE COVERAGE - ACTUATOR TIPS

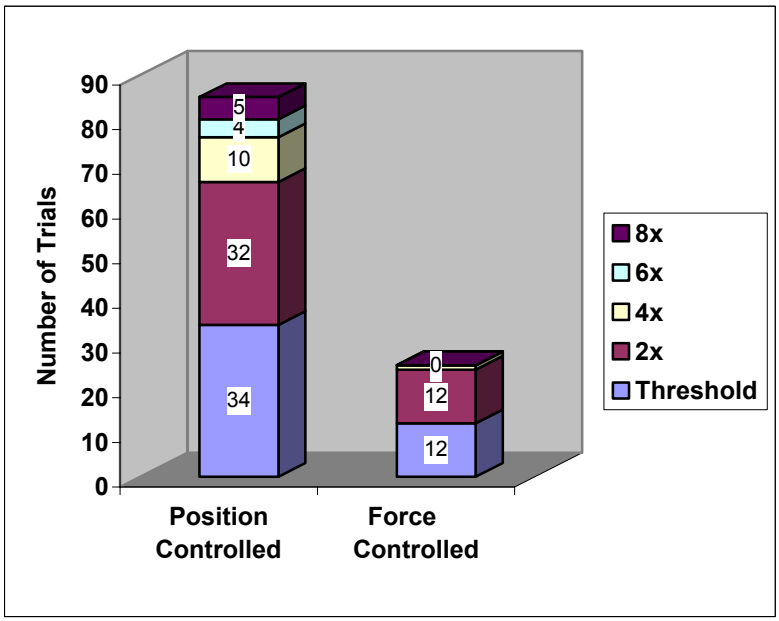


FIGURE 5.29: EXPERIMENTAL INPUT SPACE COVERAGE – STIMULUS INTENSITY

5.3.3 Logistical Regression Interpretation

In this analysis, several decisions were made that may or may not have influenced what the fitted model predicted as receptor sensitivities. These include:

1. Normalizing the independent variables prior to fitting and using the odds ratio as a measure of association
2. Assuming that the peaks in graphs of odds ratio vs. pre-spike sample for all coefficients represent receptor sensitivities
3. Using *stress*, *strain*, *dstress/dt*, and *dstrain/dt* as independent variables and *stress x strain*, *stress x dstress/dt*, *strain x dstrain/dt*, and *dstress/dt x dstrain/dt* as interaction terms in the model
4. Concentrating on correlated relationships as opposed to inverse or protective relationships

Each of these assumptions is now discussed.

1. **Normalizing the independent variables prior to fitting and using the odds ratio as a measure of association**

The odds ratio was selected as a quantitative measure of the input-output association for several reasons. First, it has an intuitively appealing interpretation. It quantifies the degree to which the probability of generating a positive response changes for a single unit change in a particular independent variable. Second, it is calculated in a straightforward manner and is listed in fitted-model summaries by commercial logistical regression software packages. Third, the odds ratio has been used as a measure of correlation in other studies [11,13,17,20,35,51].

Since the independent variables are normalized prior to being fit, the Beta coefficients themselves could be used directly as relative measures of association. When the coefficients are 0, the corresponding independent variable is statistically uncorrelated to the response. When the coefficients are less than 0, the corresponding independent variable has an inverse or protective relationship with the response. When the coefficients are greater than 0, the corresponding independent variable has a correlated relationship with the response. The disadvantage of comparing Beta coefficients directly is that they do not have the straightforward, easy-to-understand interpretation that the odds ratio has. Additionally, Beta coefficients do not account for the varied effects that interaction terms have upon the input-output correlation at varying levels of the independent variable.

The odds ratio does not give any more information on the relative input-output association than the Beta coefficients. In fact, both will order the input-output associations in the same fashion. However, the odds ratio is chosen over the Beta coefficient for its attractively simple interpretation.

The key to being able to make direct comparisons between Beta coefficients or odds ratios is the fact that the independent variables are normalized (standardized) prior to being fit. Normalization allows direct comparison of the coefficients associated with independent variables that inherently have different means and standard deviations as wall as units and ranges. Normalization, however, is not without experimental implications. During normalization, an independent variable undergoes a linear transform. The degree to which this variable is modified depends solely on the overall distribution of the variable and is not at all dependent on the distributions of the other independent variables. This implies that permutations comprising the independent variables that make up a particular sample or observation may be lost since one variable in that combination may undergo a dramatic change while others do not.

A potential solution to this dilemma is post-regression normalization as discussed by Menard [41]. In this method, the regression coefficients are estimated based on the absolute levels of the independent variables. Following estimation, the coefficients are normalized based on the levels of their associated independent variables. This transform allows direct comparison of the coefficients and odds ratios relative to one another. Due to the fact that post-regression normalization requires custom software (as it is not an option in commercial logistical regression software), this technique was not evaluated in this study.

2. **Assuming that the peaks in graphs of odds ratio vs. pre-spike sample for all coefficients represent receptor sensitivities**

The odds ratio peaks correspond to regions within the sample space in which a single unit change in a particular independent variable corresponds to an elevated increase in the probability of generating a response relative to the other independent variables. These peaks do not convey any information about average level of the odds. As an example, contrast the odds ratio for *stress* in Figures 5.30 and 5.31 below. The peak in Figure 5.30 has a much larger time variance than the one shown in Figure 5.31. This wider sensitivity range is not conveyed by peak value alone.

Alternative techniques to assess nerve sensitivity from odds ratio vs. pre-spike sample time could include integration of the curves over a finite interval no longer than the approximated conduction delay. The efficacy of alternative techniques has not been addressed in the literature on logistical regression or in studies in which logistical regression was employed. Most likely, the reason for this is the unique, unconventional context in which logistical regression is being applied in this study.

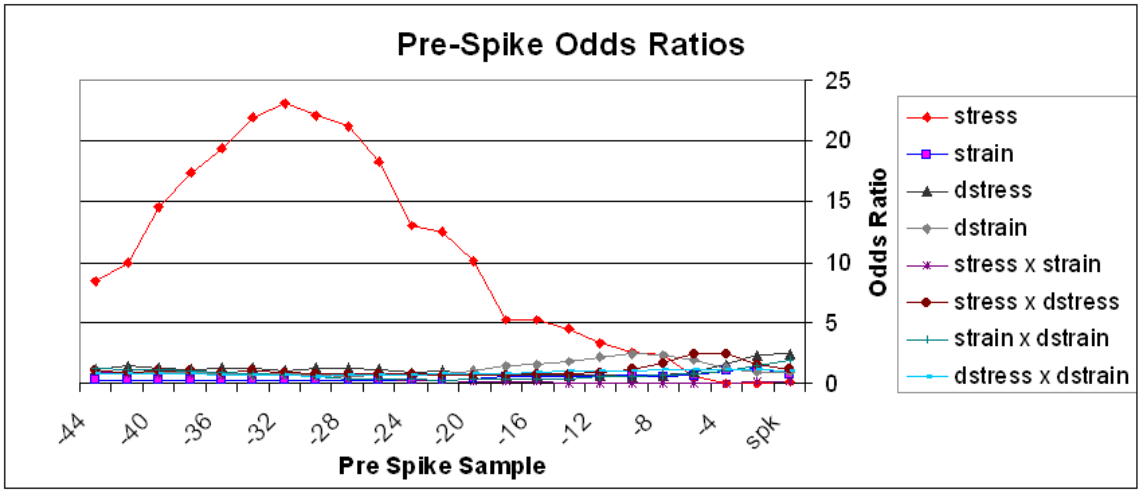


FIGURE 5.30: BROAD PEAK ODDS RATIO [NERVE R0107023]
Contrast the odds ratio peak shown in this figure to the one shown in Figure 5.31. The peak shown here is considerably wider, spanning approximately 20 samples.

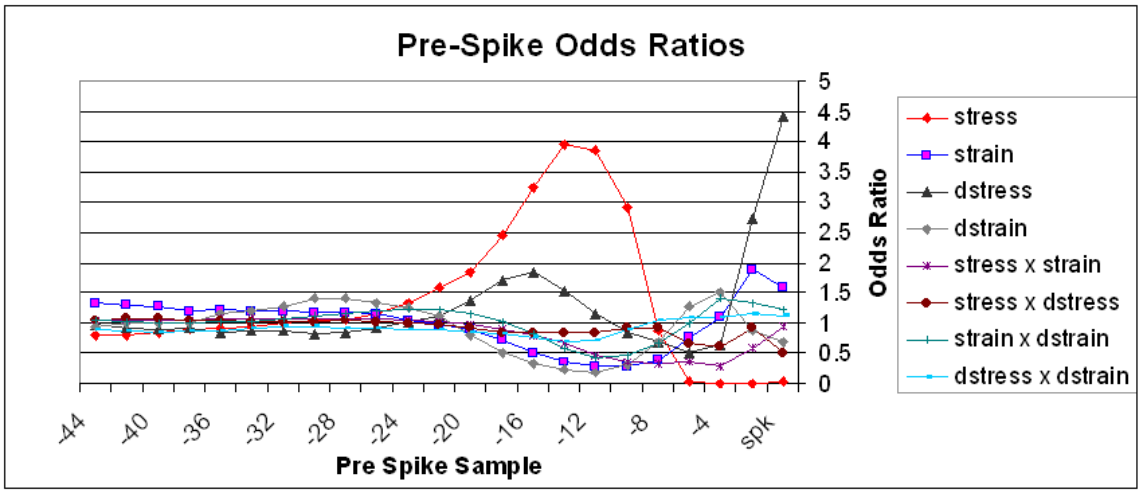


FIGURE 5.31: NARROW PEAK ODDS RATIO [NERVE R0107026]
Contrast the odds ratio peak shown in this figure to the one shown in Figure 5.30. The peak shown here is considerably narrower (less time variance), spanning only a few samples.

3. **Using *stress*, *strain*, *dstress/dt*, and *dstrain/dt* as independent variables and *stress x strain*, *stress x dstress/dt*, *strain x dstrain/dt*, and *dstress/dt x dstrain/dt* as interaction terms in the model**

The independent variables and interaction terms used in the model were selected based on a number of different reasons. First, the model was used in other studies in which logistical regression was used to model the responses of mechanoreceptors subjected to tensile loads [17].

Second, *stress* and *strain* are customary cross-platform metrics, meaning similar experiments could be performed in a different context and the results contrasted to this experiment. Although the *stress* and *strain* values used are linearly derived from the recorded indented position and applied force values, a cross-platform comparison would be complicated if indented position and applied force were used in place of *stress* and *strain*.

Third, the interaction terms were selected because they have experimental and physical significance in the system being modeled. The *stress x strain* interaction term, for example, is proportional to the energy applied to the skin in the elastic region of the *stress-strain* profile. Higher order interaction terms (2^{nd} derivatives, for example) do not have a meaningful interpretation in the context of the system being modeled. If interaction terms without biological or statistical significance were included in the model without due consideration, the danger is over-fitting, i.e., fitting the inherent noise in the system rather than the emergent properties of the system. Additionally, it demonstrates a lack of a thorough understanding of the system being modeled.

While not all of the interaction terms showed consistent significance within the fitted model across all the receptors included in this analysis, interaction terms did show consistent significance in individual receptors. Based on this observation, it was decided to uniformly use the extended model with all the interaction terms (*stress x strain*, *stress x dstress/dt*, *strain x dstrain/dt*, and *dstress/dt x dstrain/dt*). Although the standard model with no interaction terms may converge on nerve sensitivities more consistently, the disparity between the sensitivities predicted by the standard model and the extended model were not significant enough to warrant exclusive use of the standard model.

In a future analysis, it may be useful to perform stepwise inclusion or exclusion of interaction terms and independent variables based on statistical significance within the fitted model and compare the results of that analysis to the results predicted by an analysis using the standard or extended models. In stepwise logistical regression, however, the model is particularly sensitive to the number of binary 1 events relative to binary 0 events such that the method often fails to reject the null hypothesis when less than 10 % of the covariate combinations result in a binary 1 output. Since the average percentage of binary 1 observations in the time-series representation of the mechanoreceptors used in this study is considerably less than 10, stepwise regression may be ineffective without some type of pre-regression data parsing or compression.

4. **Concentrating on correlated relationships as opposed to inverse or protective relationships**

In this analysis, protective or inverse input-output relationships are not addressed. There is a wealth of information in these relationships. For example, a receptor may be preferentially sensitive to *strain* when the actuator tip is retracting from the skin (negative *dstrain/dt* values). This sensitivity may not be detected because the method correlates elevated stimulus levels to the response variable. This information is still present in the odds ratios embodied in the model fit; however, one must look for it statistically rather than qualitatively. In other words, the odds ratios map Beta coefficients from their inherent range of [$-\infty$ to $+\infty$] to a range of [0 to $+\infty$], meaning the quantitative assessments of the protective relationships in the fitted model ($-\infty <$ Beta coefficient < 0) are compressed into the region from $0 <$ odds ratio < 1. Therefore, minute changes in the odds ratios of protective independent variables correspond to dramatic changes in Beta coefficients. It is difficult to perceive such differences qualitatively. Instead, a more rigid statistical technique should be employed. Such an analysis is beyond the capabilities of the commercial software used in this study.

6.0 Discussion

Tactile sensation is a complex manifestation of mechanical stimuli applied to the skin. At the most fundamental or atomic level of the somatosensory system is the cutaneous mechanoreceptor. Therefore, the mechanoreceptor is a logical starting point in the bottom-up approach to understanding the somatosensory system and sensation, in general. In this study, we isolated several afferent mechanoreceptors, mechanically stimulated them, recorded their responses, and determined the sensitivities of each of the individual receptors by correlating the compressive state of the skin to the observed receptor responses.

What sets this study apart from other studies in which the input-output association of individual mechanoreceptors was evaluated is the fact that:

a) We proposed the novel application of a powerful statistical technique to characterize the response of a single mechanoreceptor subjected to controlled compressive loads.
b) We established and validated a framework for the application of this statistical method to a time-series representation of the multiple-input, binary-output system being modeled.
c) We performed in-vitro experiments to record the afferent behavior of tactile receptors in rat hairy skin.
d) We applied the technique to the recorded data to quantitatively evaluate the relative association between a number of biologically meaningful stimulus metrics and the observed neural response.

The statistical technique chosen for this study was multivariate logistical regression. The multivariate logistical model provides an elegant solution for modeling a system with confounded input combinations and a binary output. From this model, we can quantify the relative contributions that each of the components of the stimulus has upon the neural response of the receptor and extrapolate any input-output associations to the greater population of cutaneous mechanoreceptors.

In any regression method, there are fundamental assumptions concerning the data being modeled that must be met. We must meet these assumptions in order to draw statistical inferences from the sample population to the population at large. Violating some assumptions may carry little consequence. Violating others may quash much of the useful information embodied in the estimated parameters of the fitted model. Therefore, the decision to apply a statistical technique in a non-conventional manner carries the burden of heightened post-regression diagnostics. Nevertheless, we are dealing with a deterministic mathematical technique that will produce a given result regardless of the context in which the data were collected. In this study, we exploited that fact.

Multivariate logistical regression assumes that each observation or subject in a dataset is independent of the other observations or subjects included in the dataset. In a time-series representation of a system in which there is latency between input at a particular level

and the response of the system to that level, the requirement of observational independency is not met. Furthermore, when the response of the system is artificially depressed or elevated following a particular event, the requirement of observational independency is not met.

In this study we addressed these issues by manipulating the inputs and outputs of the system before performing the regression analysis such that the regression analysis produced results that could be interpreted as being derived from independent observations. In addition, we utilized a technique to statistically quantify the extent of the post-event observational dependencies such that these observations could be removed from the datasets. Subsequently, the non-independent observations were removed and the regression coefficients re-estimated. The result of the employment of both of these methods is a lack of model misspecification due to observational inter-dependency.

The application of multivariate logistical regression models to mechanoreceptor systems represents a significant deviation from conventional linear analysis techniques like stimulus-response histograms, raster plotting, cross-correlation, and transinformation analyses. Each of these linear techniques has associated shortcomings when employed in the context of mechanoreceptor systems, the most common being the assumption of linear system input-output and the inability to assess individual input-output associations relative to other input in a multivariate context with interactive input.

To conclude, multivariate logistical regression is a powerful technique for quantifying the input-output relationship between compressed skin states and the afferent response characteristics of mechanoreceptors. When applying this technique to sampled time-series representations of the input-output state of mechanoreceptors, special consideration is necessary to meet the modeling requirements of multivariate logistical regression.

7.0 Summary

Multivariate logistical regression is a mathematical technique that provides quantitative assessments of the contributions that each variable in a set of covariates has upon some binary response that is a complex, non-linear function of those inputs. The ability of logistical regression to quantify the relationship between inputs that covary and a response makes it the definitive statistical method for studies in which the input to a system corresponds to one or more measurable, continuous variables that elicit a dichotomous response from the system.

In this study, the system being modeled transduces mechanical stimuli applied to the skin into afferent nerve responses. The covariates are the position of the stimulator that indents into the skin, the force applied to the skin by that stimulator, and their respective time derivatives. The binary response is the presence or absence of a nerve response, given the state of the stimulus. A multivariate logistical model was developed for this system and the relative contributions of each covariate assessed in a quantitative fashion based on the multivariate logistical regression technique.

Appendix A: Logistical Regression Using SPSS

This *Appendix* contains a short tutorial on how to use SPSS to perform multivariate logistical regression.

Introduction

The purpose of this brief tutorial is to show the user how to use *SPSS Base with Regression Models v10* for performing multivariate logistical regression analyses. The SPSS product that is discussed in this tutorial is described on the SPSS website here:

http://www.spss.com/spss10/

A technical discussion of the regression component of the software can be found here:

http://www.spss.com/tech/stat/algorithms/regres.pdf

The SPSS product that is described here is Win32 compatible (Windows 95, 98, NT 4.0, and 2000). This tutorial references version 10 of the software but has also been verified to work with version 9 of the software. It is assumed that the user has the software installed and understands logistical regression before starting this tutorial.

Navigating the Software

When a particular subheading under one of the main categories (*File, Edit, View, Data, Transform, Analyze, Utilities, Graphs, Window,* or *Help*) is to be selected, it will be referenced as follows:

Main Heading ⇒ Selection ⇒ Subheading

For example, to select the *Binary Logistic Regression* subheading, one would select:

Analyze ⇒ Regression ⇒ Binary Logistic

This is shown in Figure 1. Once the *Binary Logistic* subheading is highlighted and clicked, a second window will pop up allowing the user to make selections and perform the regression analysis.

Importing Data

SPSS can import data from a number of different data sources and formats. Some of the more common formats are: **SPSS Systat (.sys), Microsoft Excel (.xls), Lotus (.w*), dBase (.dbf),** and **Text (.txt)**. The format that SPSS uses to store data once it has been imported is the **.sav** extension.

To import data in these formats, go to:

File ⇒ Open ⇒ Data

Select the format of the data from the pull down menu. Then highlight the file to be opened in the window and click *Open*. SPSS will then import the users data into columns in the main window as shown in Figure 2.

If the user chooses to import a file with a *.txt* extension, the *Text Import Wizard* will appear. At this point (Step 1 of the *Text Import Wizard*), one can select either to specify the format of the text or to import the text according to a pre-established format. Since no pre-established format exists yet, we will choose to specify the format of the text by selecting "no" where it asks, "Does your text file match a pre-defined format?" and then selecting *Next*.

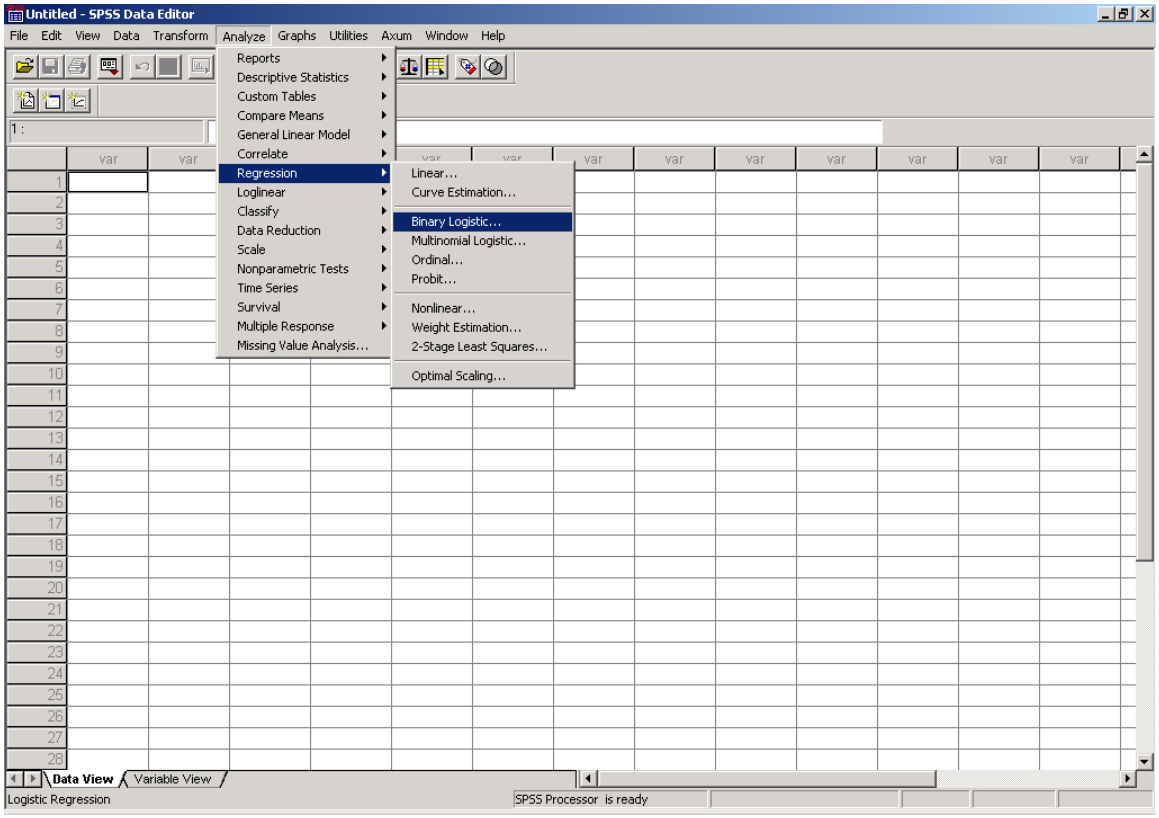
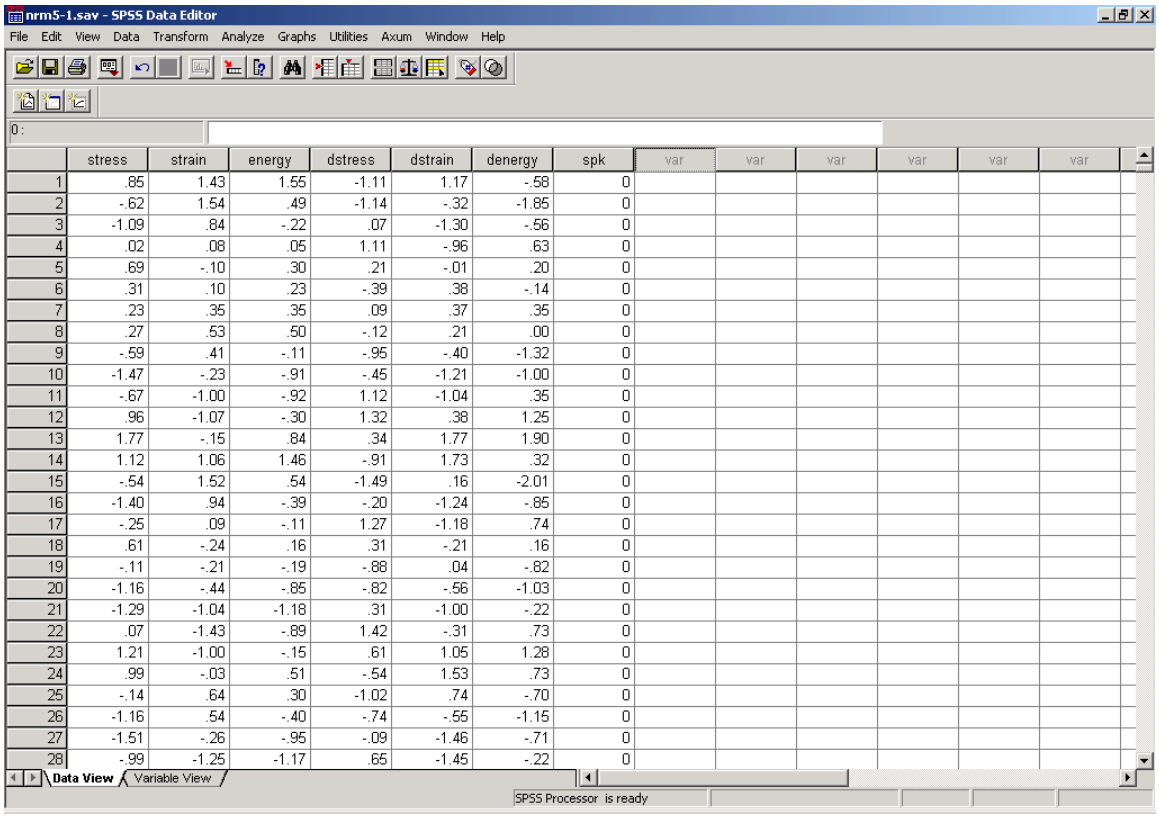


Figure 1: Subheading Selection (Binary Logistical Regression)

Step 2 of the *Text Import Wizard* asks questions about the format of the data file. After selecting the radio button for each question that describes the data file, click *Next*.

Step 3 of the *Text Import Wizard* asks questions regarding the number of cases you wish to import and how the cases are organized. Answer the questions and click *Next*.



	stress	strain	energy	dstress	dstrain	denergy	spk	var	var	var	var	var	var
1	.85	1.43	1.55	-1.11	1.17	-.58	0						
2	-.62	1.54	.49	-1.14	-.32	-1.85	0						
3	-1.09	.84	-.22	.07	-1.30	-.56	0						
4	.02	.08	.05	1.11	-.96	.63	0						
5	.69	-.10	.30	.21	-.01	.20	0						
6	.31	.10	.23	-.39	.38	-.14	0						
7	.23	.35	.35	.09	.37	.35	0						
8	.27	.53	.50	-.12	.21	.00	0						
9	-.59	.41	-.11	-.95	-.40	-1.32	0						
10	-1.47	-.23	-.91	-.45	-1.21	-1.00	0						
11	-.67	-1.00	-.92	1.12	-1.04	.35	0						
12	.96	-1.07	-.30	1.32	.38	1.25	0						
13	1.77	-.15	.84	.34	1.77	1.90	0						
14	1.12	1.06	1.46	-.91	1.73	.32	0						
15	-.54	1.52	.54	-1.49	.16	-2.01	0						
16	-1.40	.94	-.39	-.20	-1.24	-.85	0						
17	-.25	.09	-.11	1.27	-1.18	.74	0						
18	.61	-.24	.16	.31	-.21	.16	0						
19	-.11	-.21	-.19	-.88	.04	-.82	0						
20	-1.16	-.44	-.85	-.82	-.56	-1.03	0						
21	-1.29	-1.04	-1.18	.31	-1.00	-.22	0						
22	.07	-1.43	-.89	1.42	-.31	.73	0						
23	1.21	-1.00	-.15	.61	1.05	1.28	0						
24	.99	-.03	.51	-.54	1.53	.73	0						
25	-.14	.64	.30	-1.02	.74	-.70	0						
26	-1.16	.54	-.40	-.74	-.55	-1.15	0						
27	-1.51	-.26	-.95	-.09	-1.46	-.71	0						
28	-.99	-1.25	-1.17	.65	-1.45	-.22	0						

Figure 2: Imported Data

Step 4 of the *Text Import Wizard* asks what types of delimiters separate variables in the text file. Answer the question and click *Next*.

Step 5 of the *Text Import Wizard* allows you to name each of the columns in the text file. Use the mouse pointer to highlight a row in the *data preview* window as shown in Figure 3, then type the name for that column under *variable name*. You also have the ability to select how the data in that column are imported under *data format*. You can also choose not to import certain columns. Continue by clicking *Next*.

Step 6 of the *Text Import Wizard* allows you to save the current format so you don't have to go through each of these steps for the next file that you import that has this format.

Click *Finish* to complete the import and your data will appear in the columns of the primary SPSS window, as shown in Figure 2.

Note: You may need to go through these steps several times, making different selections each time, before the data is imported into SPSS in the desired format. Once the desired import format is determined, it is wise to save this import format as a template for future data imports.

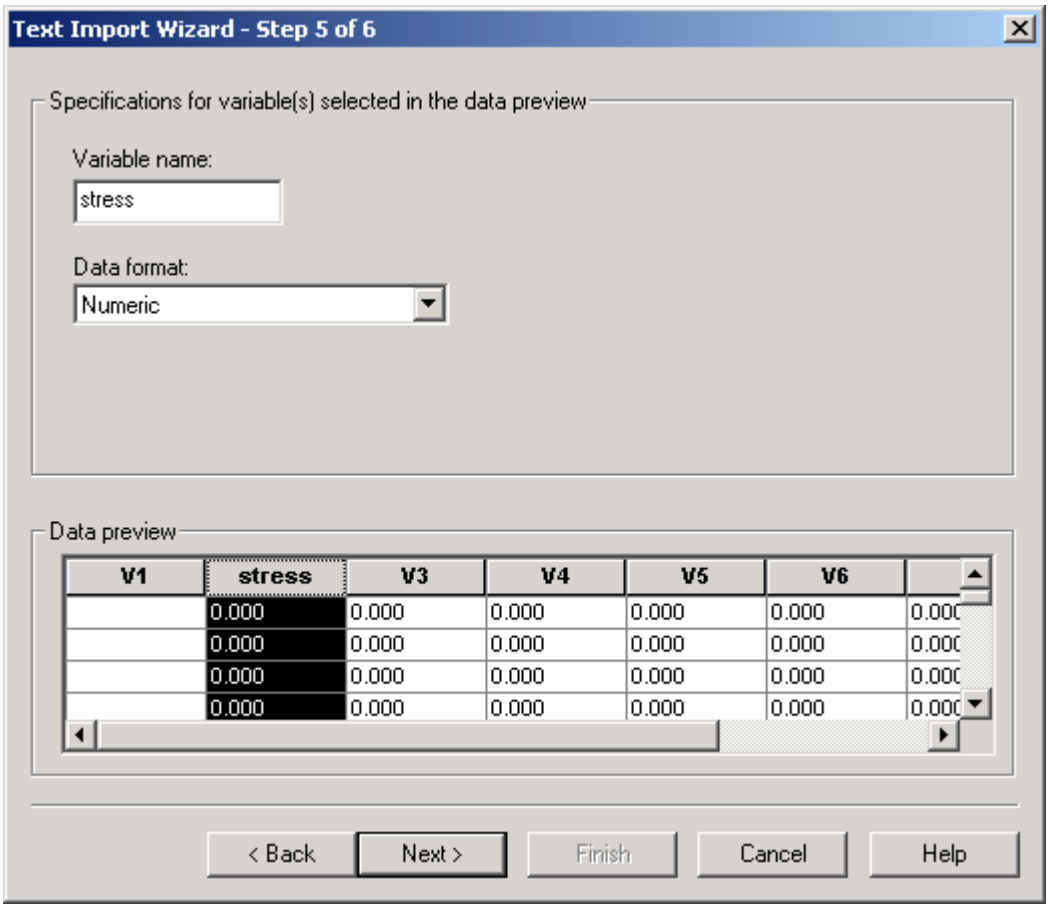
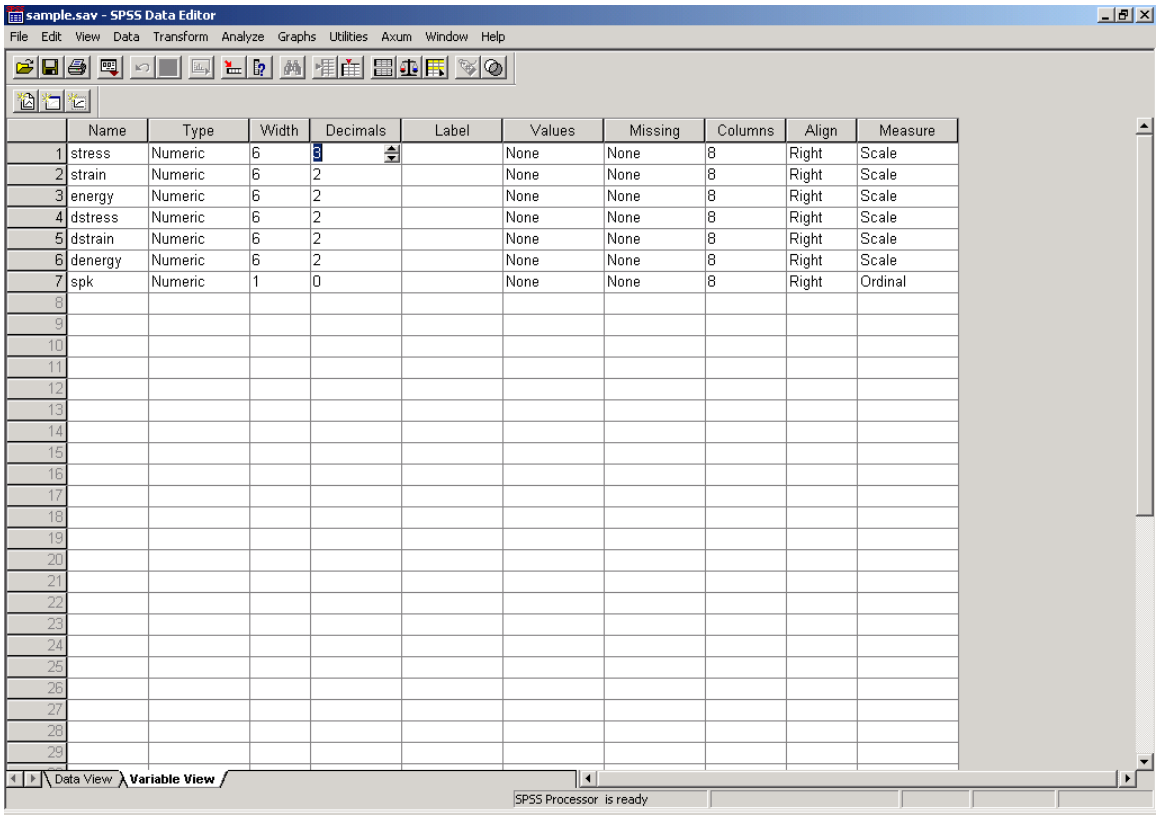


Figure 3: Text Import Wizard Step 5 – Column Naming

Once the data has been imported, changes may be made to column names, the precision of the displayed data, and type of the displayed data by clicking the *Variable View* tab in the lower left hand corner. Clicking this tab will change the main data window from a spreadsheet-type listing to a listing shown in Figure 4. You can then modify a number of different properties regarding the displayed data by clicking on the cell that contains the setting you wish to modify. To get back to the spreadsheet-type listing, simply click the *Data View* tab in the lower left-hand corner.

Plotting

A number of different graphs of the imported data can be generated. The complete listing is shown under the main heading *Graphs*. We will discuss a subset of these graphs here: *Line, Scatter, Histogram,* and *Interactive.*



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	stress	Numeric	6	3		None	None	8	Right	Scale
2	strain	Numeric	6	2		None	None	8	Right	Scale
3	energy	Numeric	6	2		None	None	8	Right	Scale
4	dstress	Numeric	6	2		None	None	8	Right	Scale
5	dstrain	Numeric	6	2		None	None	8	Right	Scale
6	denergy	Numeric	6	2		None	None	8	Right	Scale
7	spk	Numeric	1	0		None	None	8	Right	Ordinal

Figure 4: Variable View

Line Plots

SPSS is capable of generating a number of different line 2-dimensional line plots. Upon selecting *Line* from the *Graphs* heading, the window shown in Figure 5 will appear. If you wish to plot a single line (single column of Y-axis data points, single column of X-axis data points), highlight the *Simple* plot as shown in Figure 5. If you wish to plot multiple lines on the same axis (multiple columns of Y-axis data points, single column of X-axis data points), highlight *Multiple* instead of *Simple*.

Next, choose the type of line plot: Summaries for Groups of Cases, Summaries for Separate Variables, or Values of Individual Cases. After selecting the type of plot you wish to generate, click *Define*.

Summaries for groups of cases means you choose a variable/column to represent the X-axis and then choose a summary of that variable or another variable to represent the Y-axis, as shown in Figure 6. To select an X-axis variable, highlight the variable in the leftmost window then click on the arrow button just to the left of the *Category Axis* label. The variable you selected will now appear in this box. To select a Y-axis variable, you can either use a summary function of the X-axis variable or a summary function of another variable in the list in the left hand window. If you wish to select a different variable, select the radio button next to *Other Summary Function*. Next, highlight the variable you wish to use from the left-hand window and click the arrow next to *Variable*. The variable you selected will appear on this line with a *Mean* wrapper (**mean(variable)**). If you do not wish to use the mean of this variable, click *Change Summary* and you will have a number of different choices, as shown in the window labeled "*Summary Function*" on the right hand side of Figure 6. Upon making your selection, click *Continue* to return to the previous window, then click *OK* and the plot will be generated. To modify the appearance of the generated plot, go to the section of this tutorial labeled "*Plot Modification*."

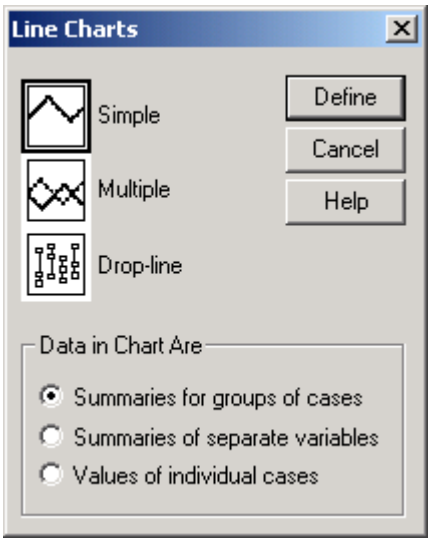


Figure 5: Line Plot Options

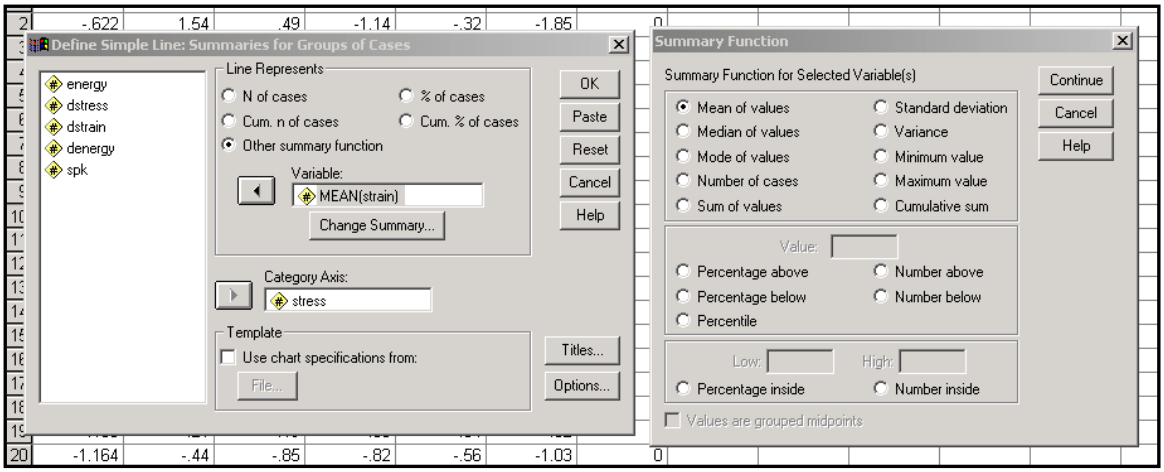


Figure 6: Summaries for Groups of Cases – Plotting Options

Summaries for separate variables allows you to choose a summary function to represent both the X-axis variable and the Y-axis variable. The summary functions are shown in Figure 7. The routine to select the variables you wish to plot consists of first highlighting the desired variable, clicking the arrow to the right of the window in which the variable was highlighted, then highlighting the same variable and clicking *Change Summary*. The *Summary Function* window shown as the right window in Figure 7 will appear. Make your selection and click *Continue* to return to the previous window. You can add a number of different variables in this way. When you have added all the variables you want, click *OK* and the plot will be generated. To modify the appearance of the plot, see the "*Plot Modification*" section of this tutorial.

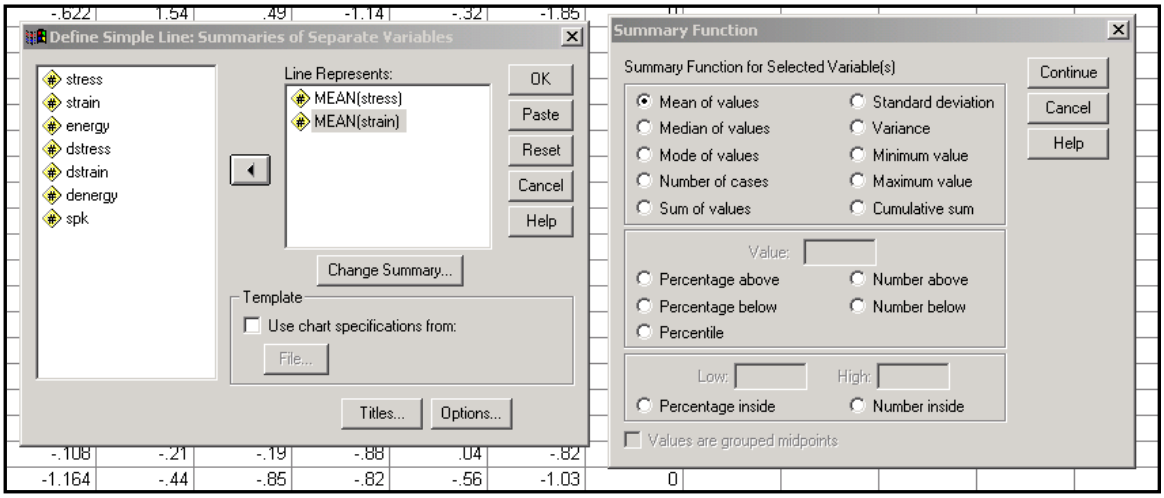


Figure 7: Summaries for Separate Variables – Plotting Options

Values for individual cases allows you to plot the actual values of the variables themselves, rather than a summary of the values. Select the variables that you wish to plot from the left hand window by highlighting them, then click the arrow to move the variable into the *Line Represents* space. This will be the Y-axis variable. Then choose the X-axis variable by first highlighting it, then clicking the *Variable* radio button, followed by the arrow under *Variable*. The variable you select will appear on the line under *Variable* and the window will look similar to Figure 8. To generate the plot, click *OK*. To modify the plot appearance, see "*Plot Modification*" in this tutorial.

To undo any variable selections, simply highlight the variable in question and click the arrow next to it to return it to the pool of unselected variables in the leftmost window.

If the *OK* button is invisible, it means the plot variables have not been completely specified, therefore, repeat the steps discussed above for the desired type of plot.

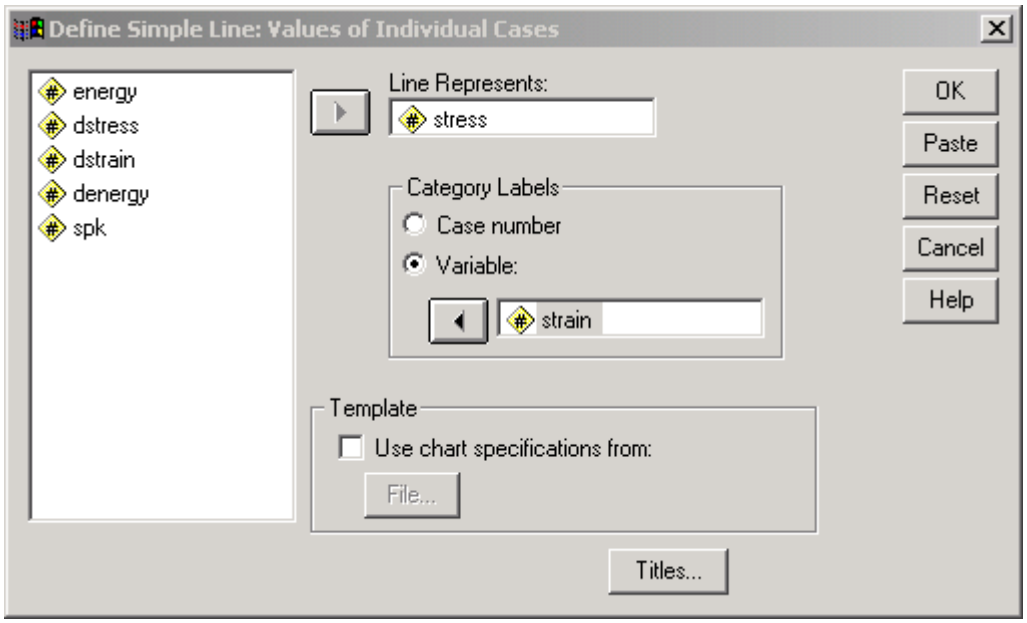


Figure 8: Values of Individual Cases – Plotting Options

The above descriptions pertain to *Simple Line* plots only. If you wish to generate *Multiple Line* plots, the steps are virtually identical, save for the fact that you have the option to select multiple Y-axis variables to plot versus one X-axis variable.

Scatter Plots

The scatter-plot option is available under **Graphs ⇒ Scatter**. The *Simple* scatter plot is described here. To select the *Simple* scatter plot, highlight *Simple* in the first window (Figure 9) and click *Define*. This will bring you to a window similar to the one shown in

Figure 10. Highlight the desired variables in the left-most window then click the arrow next to the X or Y-axis labels to move the variable into the label. Click *OK* to generate the plot. Information on modifying the appearance of this plot is found in the "*Plot Modification*" section of this tutorial.

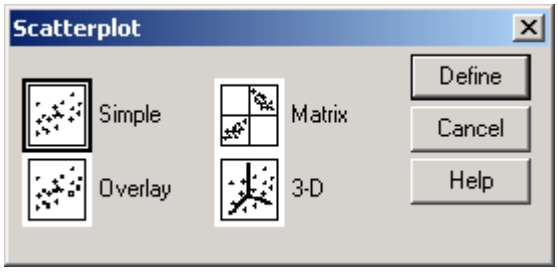


Figure 9: Scatter-Plot Types

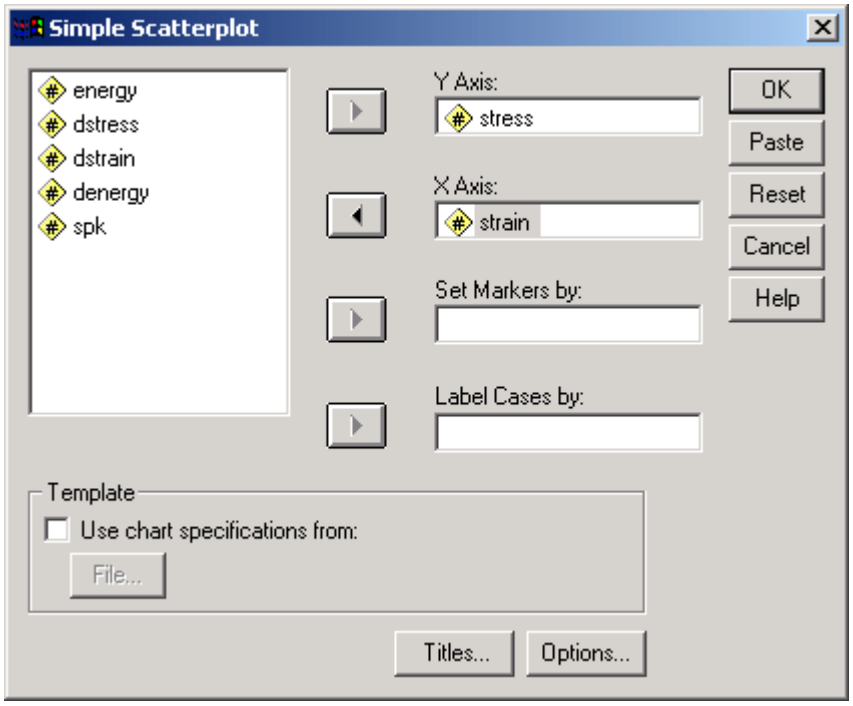


Figure 10: Scatter Plot Settings

Histogram Plots

Under **Graphs ⇒ Histogram**, one can generate a histogram of the data in any one of the variables/columns. In the window shown in Figure 11, highlight the variable to be plotted in the leftmost window, click the arrow to move the variable into the box labeled *Variable*, then click *OK* to plot the variable's histogram. If you want a normal regression curve fitted to the histogram, check the *Display Normal Curve* box before clicking *OK*.

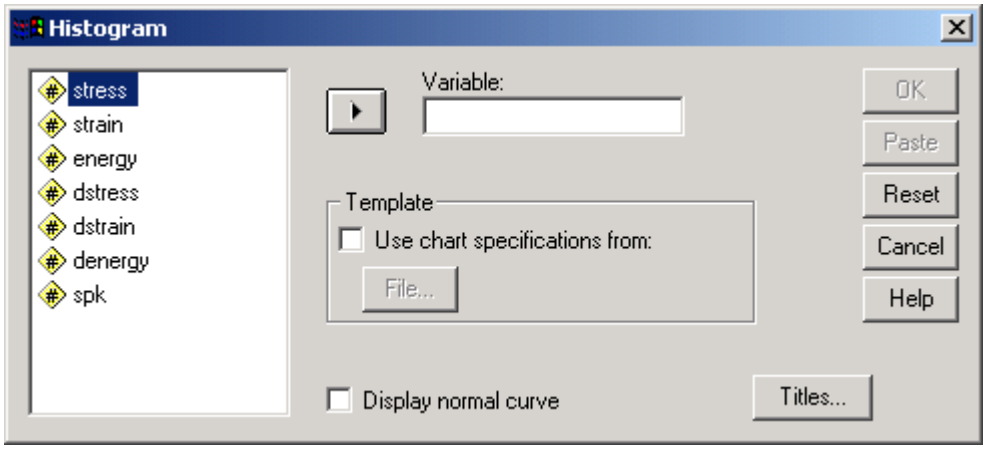


Figure 11: Histogram Plots

Interactive Plots

A number of different graphs can be custom-designed under the **Graphs ⇒ Interactive** heading. The steps for selecting the variables to be used in the graph is the same for all graphs, therefore, the steps will be detailed for the *Line* plot only.

Upon selecting **Graphs ⇒ Interactive ⇒ Line**, a window similar to the one shown in Figure 12 will appear. To select the X and Y-axis variables, highlight the variable in the leftmost window and drag it to the desired box to the right of the window. In Figure 12, the box with the variable [stress] in it corresponds to the Y-axis. The empty box corresponds to the X-axis. Upon dragging the variables to their appropriate boxes, you can either click *OK* at the bottom or choose from the various chart options available in the tabs at the top of the window (*Dots and Lines, Error Bars, Titles, and Options*).

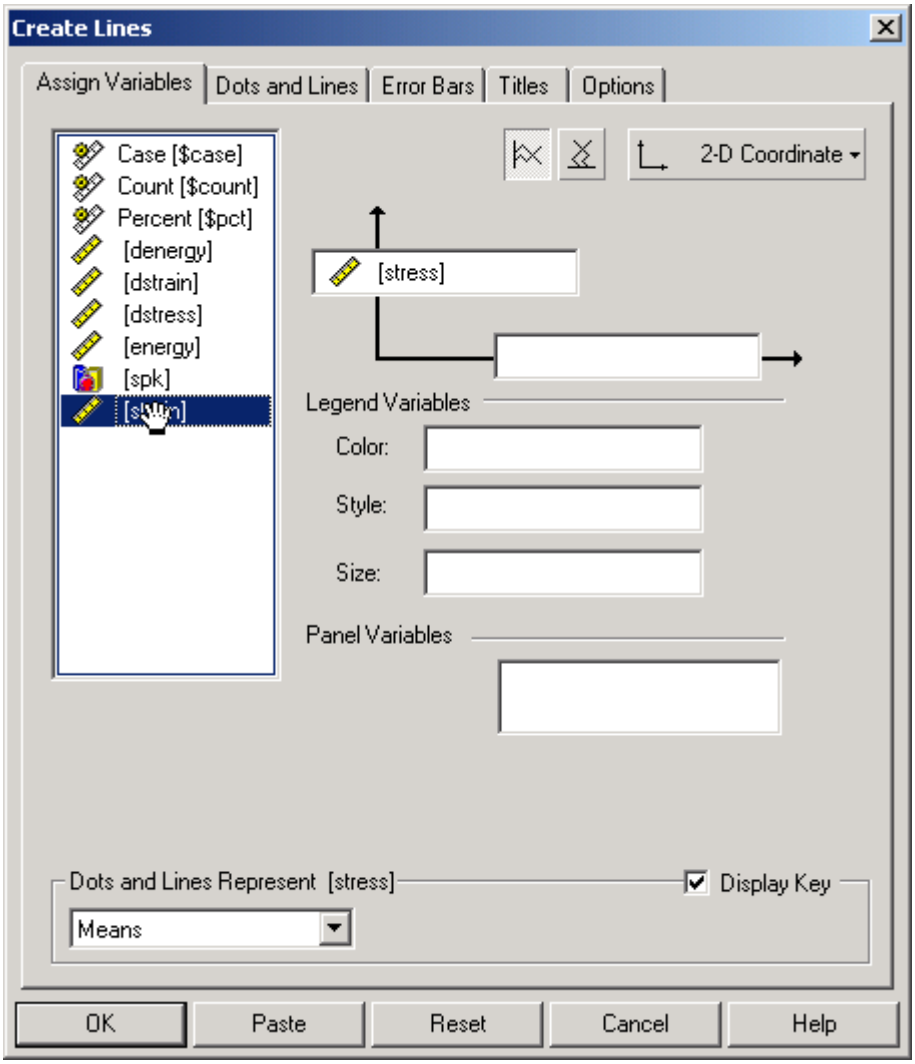


Figure 12: Interactive Line Plot Options

Plot Modification

Whenever a plot is generated using the techniques described above, a second window will pop up (assuming one is not already open) called *SPSS Viewer* containing the plot. If you wish to modify the appearance of the plot, double-click on the plot itself and another window called *SPSS Chart Editor* will appear as shown in Figure 13. The chart in *SPSS Viewer* will be darkened.

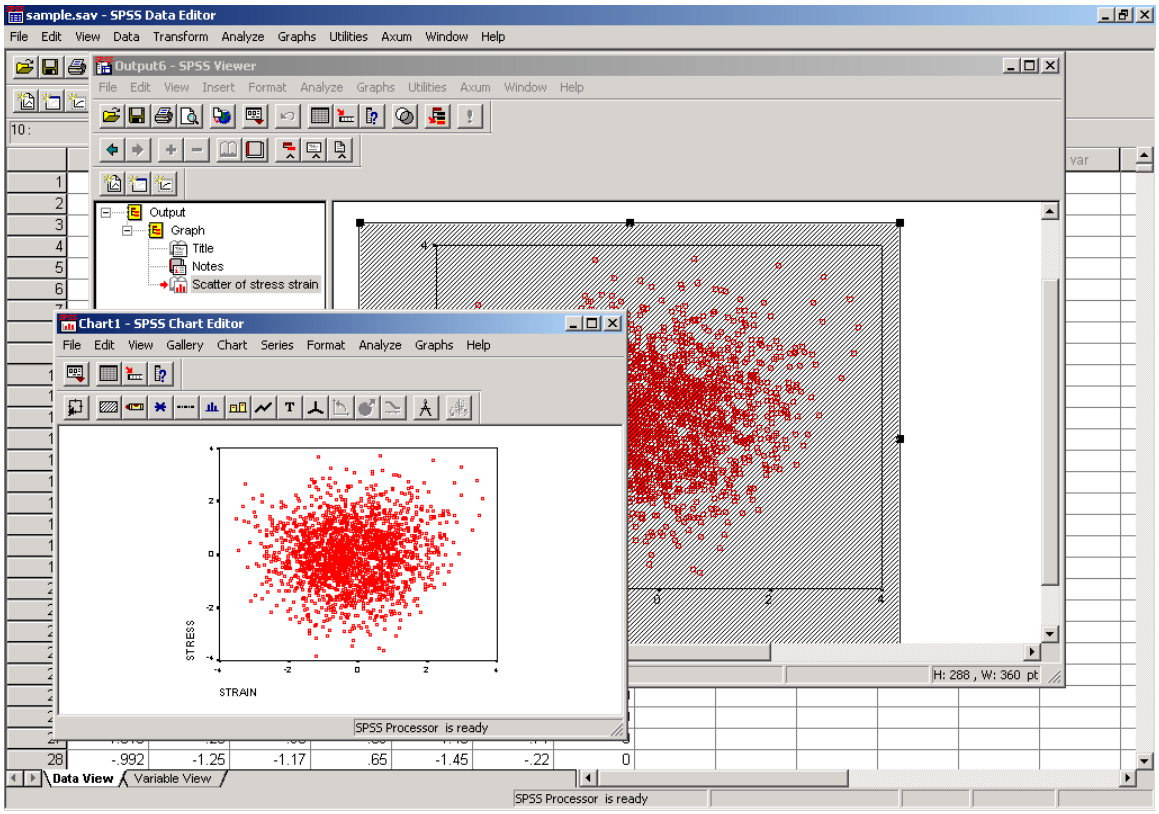


Figure 13: SPSS Chart Editor

In *SPSS Chart Editor*, one can change virtually all aspects of the plot. Under **Chart ⇒ Axis**, one can change the range, scale, titles and title justification, tick marks, and remove undesirable axis markers if needed. Under the **Format** menu, the color, line-type, fill-type, and marker for the data points can be changed as well as the font and font size for text. When changing any of these properties, you must first highlight the item or label in the chart that you wish to change before making the selection.

To change the color of the data points in the plot, first click on the data points themselves to highlight them. If there are many data points, only a few data points will be highlighted as shown in Figure 14. Then select **Format ⇒ Color** and a small window labeled "*Colors*" will appear. Click the desired color and the data points will be changed to that color, as shown in Figure 14.

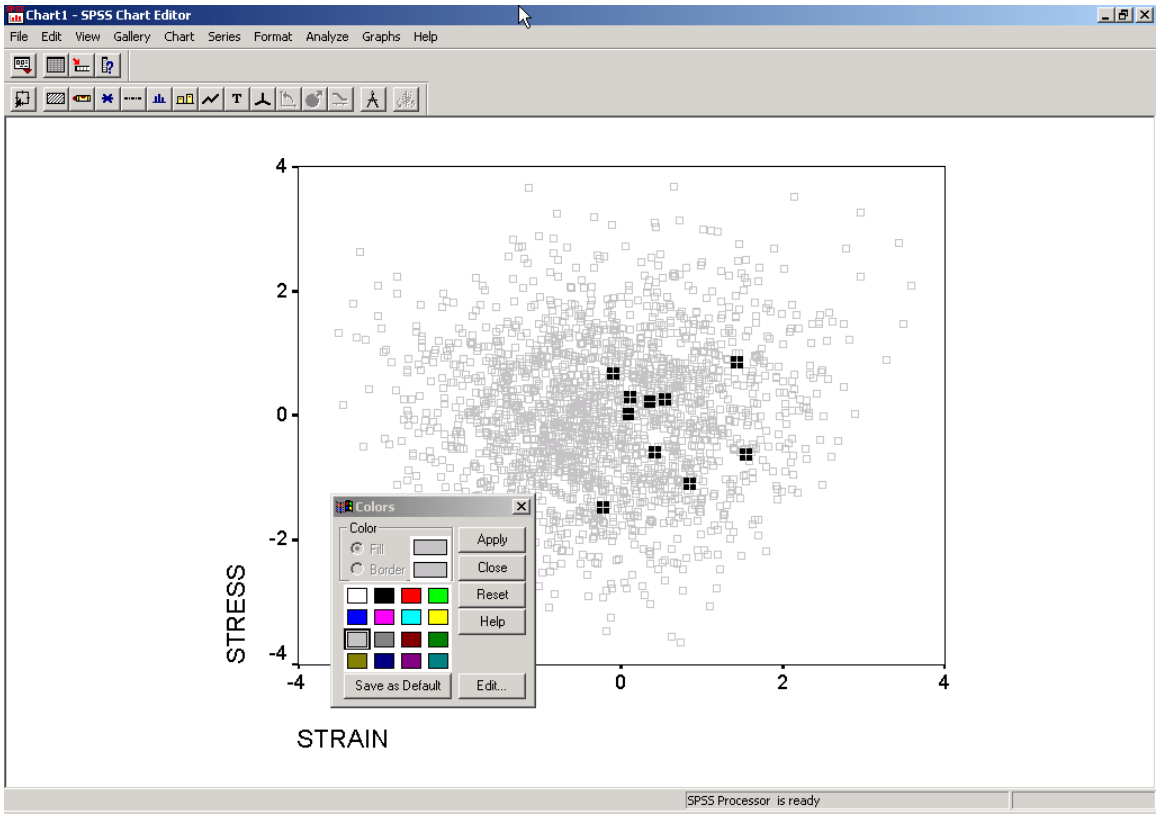


Figure 14: SPSS Chart Editor Use

To change the range of the Y-axis, for example, first click on the X-axis to highlight it (each tick mark will darken when the axis is highlighted). Then select Chart ⇒ Axis and a window will appear that is similar to the one labeled "*Y Scale Axis*" in Figure 15. From this window, one can chart a number of different properties including the title, title-justification, range, etc. To change the range, go to the *Range* portion of the window and adjust the minimum or maximum value to the desired value and click *OK*.

With a bit of experimentation, one can learn how to adjust all of the properties of the chart such that it has the desired appearance. Once one has made all of the adjustments, one can save the chart as a template so that the same chart can be generated for a different data set without having to change all the settings manually. To save the chart as a template, go to **File ⇒ Save Chart Template**, select a name, and save the template to the desired directory. When you wish to generate the same chart with different data, simply use select this chart template instead of selecting all the variables to generate the chart (for example, in Figure 11, instead of inputting the variables, check the "*Use chart specifications from*" box and point towards the template corresponding to the desired chart). When you are done editing the chart, close the *SPSS Chart Editor* to return to the *SPSS Viewer* and the main SPSS window containing the data. The SPSS Viewer contents can be saved or copied using the Windows clipboard, if desired.

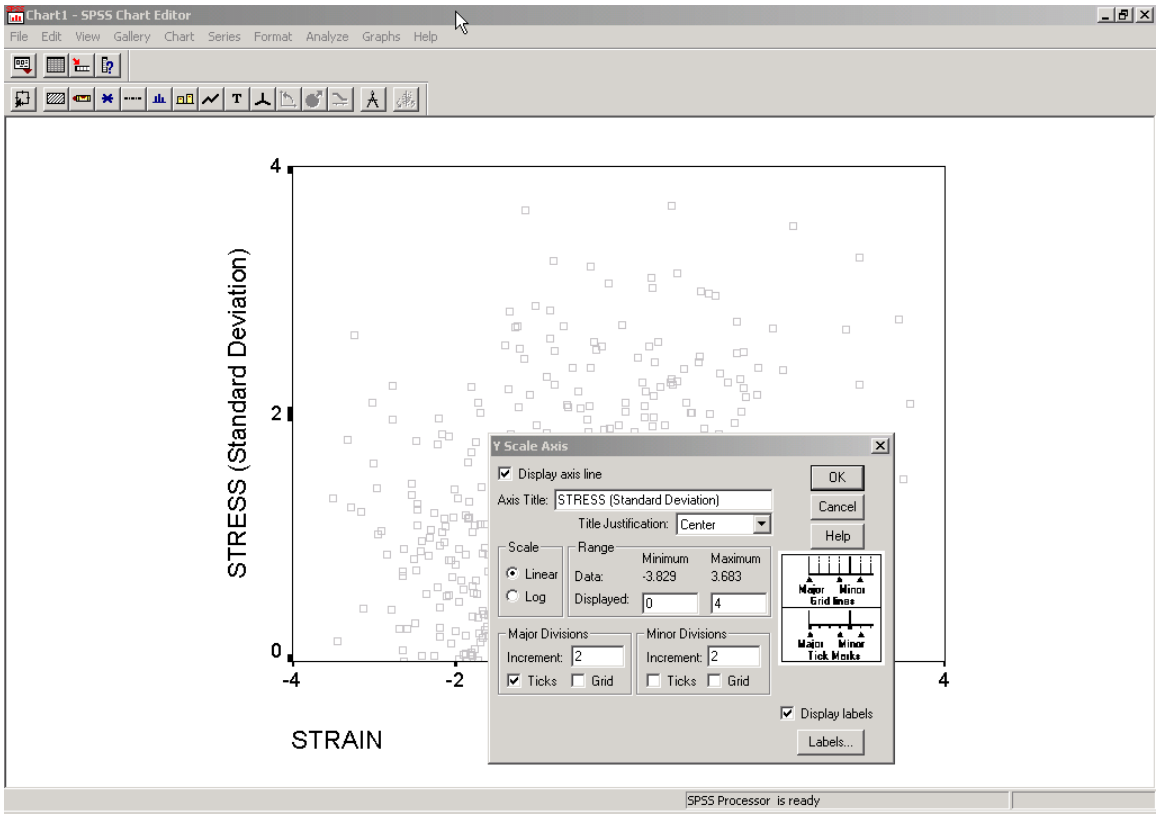


Figure 15: Chart Property Adjustment

Multivariate Logistical Regression

Before explaining how to use SPSS to perform logistical regression, let us discuss the format for the data that can be used in a logistical regression analysis.

- Binary logistical regression requires that the input(s) to the regression model (covariates or independent variables) be continuous or categorical. If the inputs are categorical, they must take on a finite number of values or strings. Any number of inputs can be used.

- The output (response or dependent variable) of the regression model must be binary. The output does not necessarily have to be coded as 0 or 1 values but it does have to be coded as either one value/string or another value/string (example: yes or no, lived or died). It can be determined from the post-regression results how the regression software treated the output variable (i.e., whether it encoded "lived" as 0 and "died" as 1 or "lived" as 1 and "died" as 0).

- An output variable must be present for every input combination included in the model.

- Each input variable must have its own column and the output variable must have its own column. Each row corresponds to an observation, i.e., a set of inputs and the binary output produced by those inputs. When imported into SPSS, the data should look similar to that in Figure 2. In Figure 2, there are 6 independent variables (*stress, strain, energy, dstress, dstrain, denergy*) and the dependent variable (*spk*).

With these requirements met, we can proceed to the *Logistical Regression* analysis. The first step is to select **Analyze ⇒ Regression ⇒ Binary Logistic**. A logistical regression window will appear as shown in Figure 16. To select the dependent variable, highlight the variable from the leftmost window and click the arrow next to the *Dependent* category. The variable will appear on the line just below the *Dependent* label. Similarly, to select the independent variables, highlight the desired variable and click the arrow next to *Covariates* to draw the variable from the leftmost window to the window below the label *Covariates*. For interaction terms, highlight both the variables that participate in the interaction and click the *>a*b>* button (as shown in Figure 16) to bring the two variables into the *Covariate* window as an interaction term. Select the type of regression you wish to perform from the *Method* pulldown menu (choices: *Enter, Forward (Conditional), Forward (LR), Forward (Wald), Backward (Conditional), Backward (LR), and Backward(Wald)*).

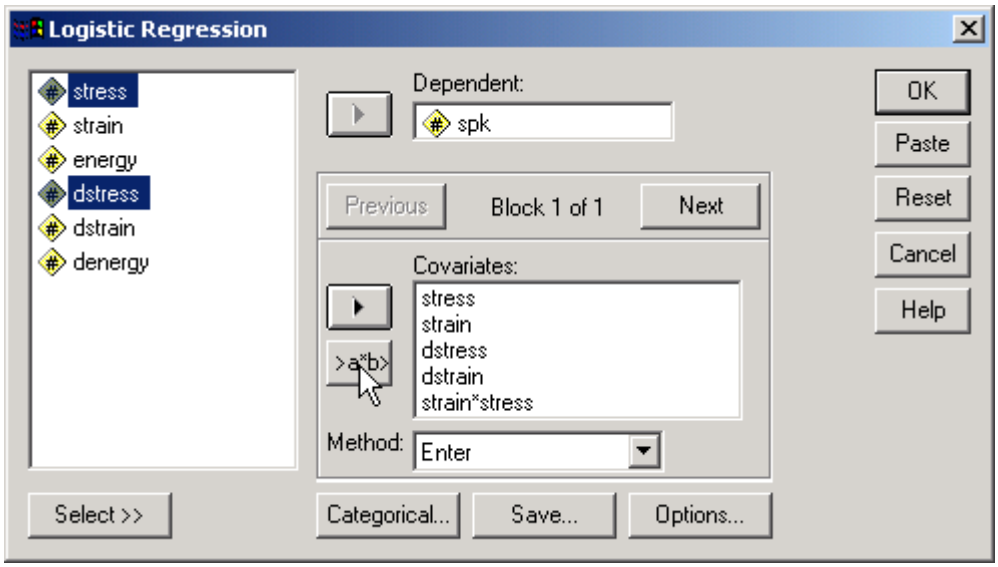


Figure 16: Logistical Regression Window

At this point, we have the option to specify if any of the covariates are categorical variables (the assumption when they are first selected is that they are continuous). To do this, click the *Categorical* label and another window will pop up. In this window, we can select from any of the non-interaction term covariates in the list of covariates. To choose

a covariate, highlight it and click the arrow to move it from the list in the leftmost window into the *Categorical Covariates* list, as shown in Figure 17. You then have the option to specify the contrast for the categorical variable as shown in Figure 17.

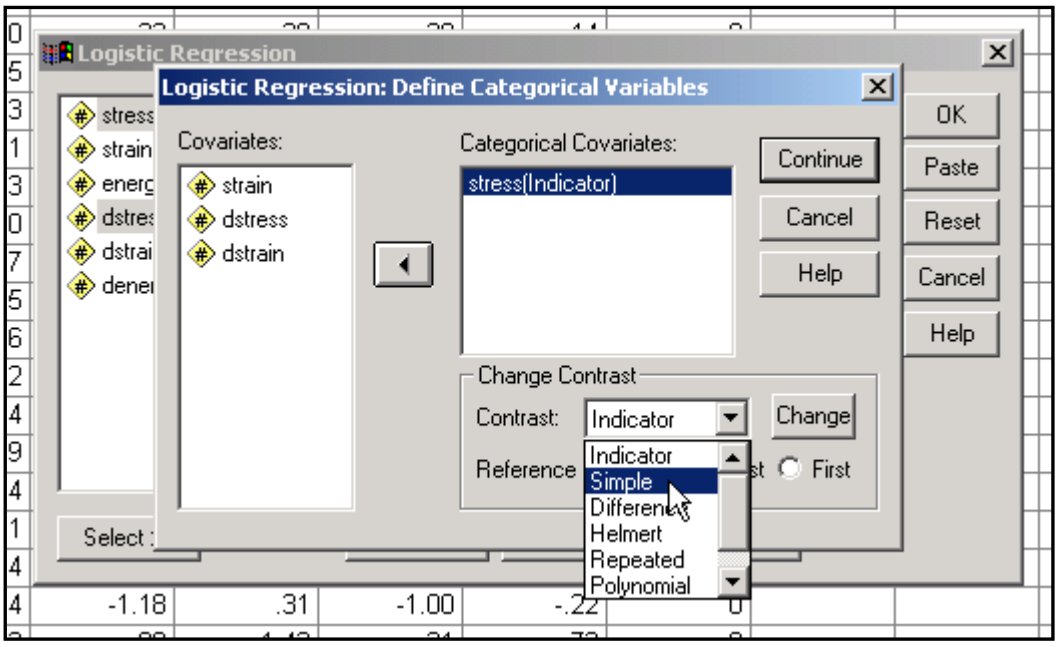


Figure 17: Categorical Variable Selection

Next, we have the option to save a number of different quantifications of the data once the regression has been performed. The variables that can be saved are: *Predicted Probabilities, Predicted Group Membership, Cook's Influence, Leverage Values, Delta Beta,* and a number of different *Residuals* (*Unstandardized, Logit, Studentized, Standardized,* and *Deviance*). To save any of these, first click *Save* from the logistical regression window. You will then be shown a window with a number of checkboxes similar to the one in Figure 18.

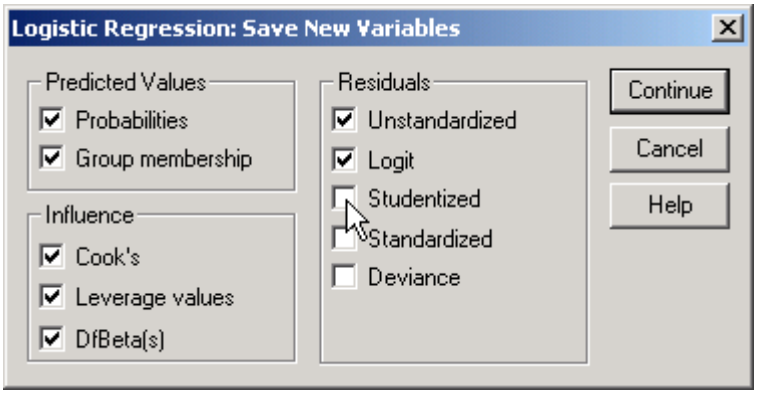


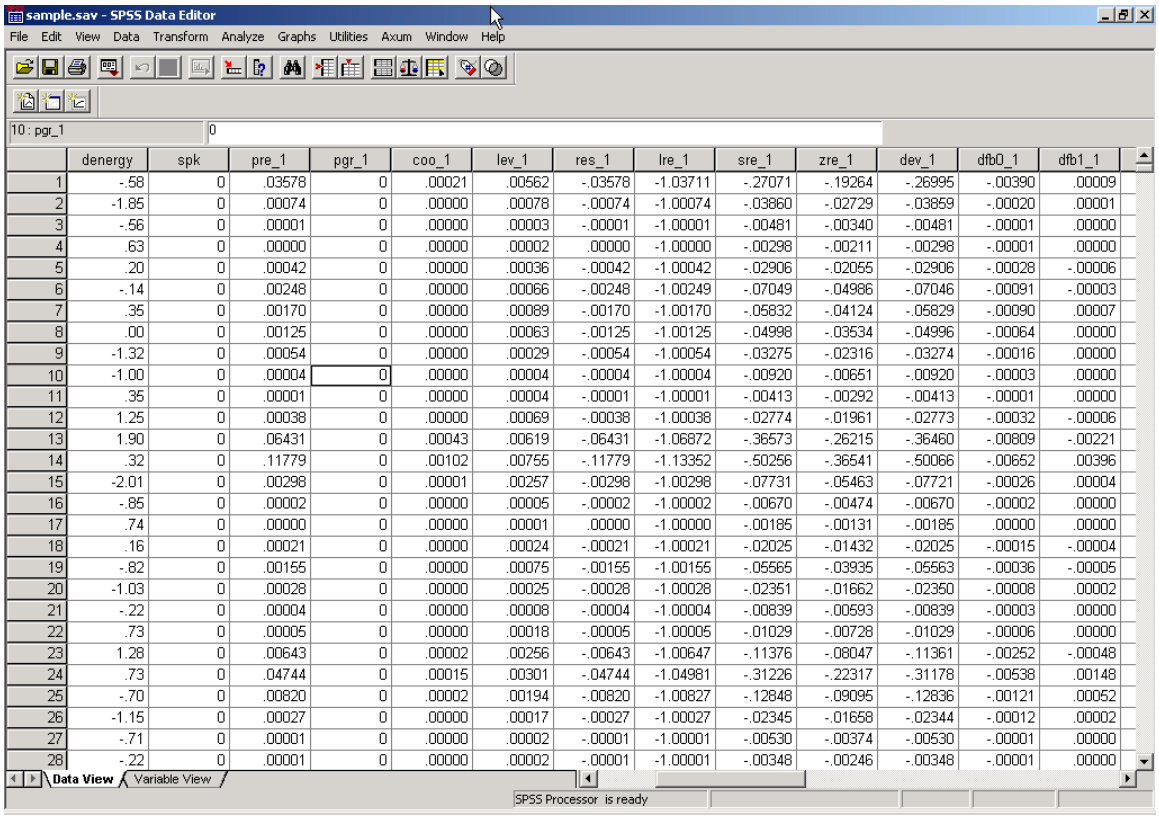
Figure 18: Saving Logistical Regression Variables

After the regression is performed, all of the values that were selected will appear in the main data window as shown in Figure 19. Each column will have a label corresponding to the regression variable. The encoding is shown in Table 1.

Regression Variable	Encoding	Type
Predicted Probability	pre_1	Continuous
Predicted Group	pgr_1	Binary
Cook's Influence	coo_1	Continuous
Leverage Value	lev_1	Continuous
Delta Beta	dfbX_1*	Continuous
Unstandardized Residual	res_1	Continuous
Logit Residual	lre_1	Continuous
Studentized Residual	sre_1	Continuous
Standardized Residual	zre_1	Continuous
Deviance Residual	dev_1	Continuous

Table 1: Regression Variable Encoding

*** X is a number corresponding to a particular covariate or interaction term that was included in the model. The encoding is based on the order of the variables in the *Covariates* list shown in Figure 16.**



	denergy	spk	pre_1	pgr_1	coo_1	lev_1	res_1	lre_1	sre_1	zre_1	dev_1	dfb0_1	dfb1_1
1	-.58	0	.03578	0	.00021	.00562	-.03578	-1.03711	-.27071	-.19264	-.26995	-.00390	.00009
2	-1.85	0	.00074	0	.00000	.00078	-.00074	-1.00074	-.03860	-.02729	-.03859	-.00020	.00001
3	-.56	0	.00001	0	.00000	.00003	-.00001	-1.00001	-.00481	-.00340	-.00481	-.00001	.00000
4	.63	0	.00000	0	.00000	.00002	.00000	-1.00000	-.00298	-.00211	-.00298	-.00001	.00000
5	.20	0	.00042	0	.00000	.00036	-.00042	-1.00042	-.02906	-.02055	-.02906	-.00028	-.00006
6	-.14	0	.00248	0	.00000	.00066	-.00248	-1.00249	-.07049	-.04986	-.07046	-.00091	-.00003
7	.35	0	.00170	0	.00000	.00089	-.00170	-1.00170	-.05832	-.04124	-.05829	-.00090	.00007
8	.00	0	.00125	0	.00000	.00063	-.00125	-1.00125	-.04998	-.03534	-.04996	-.00064	.00000
9	-1.32	0	.00054	0	.00000	.00029	-.00054	-1.00054	-.03275	-.02316	-.03274	-.00016	.00000
10	-1.00	0	.00004	0	.00000	.00004	-.00004	-1.00004	-.00920	-.00651	-.00920	-.00003	.00000
11	.35	0	.00001	0	.00000	.00004	-.00001	-1.00001	-.00413	-.00292	-.00413	-.00001	.00000
12	1.25	0	.00038	0	.00000	.00069	-.00038	-1.00038	-.02774	-.01961	-.02773	-.00032	-.00006
13	1.90	0	.06431	0	.00043	.00619	-.06431	-1.06872	-.36573	-.26215	-.36460	-.00809	-.00221
14	.32	0	.11779	0	.00102	.00755	-.11779	-1.13352	-.50256	-.36541	-.50066	-.00652	.00396
15	-2.01	0	.00298	0	.00001	.00257	-.00298	-1.00298	-.07731	-.05463	-.07721	-.00026	.00004
16	-.85	0	.00002	0	.00000	.00005	-.00002	-1.00002	-.00670	-.00474	-.00670	-.00002	.00000
17	.74	0	.00000	0	.00000	.00001	.00000	-1.00000	-.00185	-.00131	-.00185	.00000	.00000
18	.16	0	.00021	0	.00000	.00024	-.00021	-1.00021	-.02025	-.01432	-.02025	-.00015	-.00004
19	-.82	0	.00155	0	.00000	.00075	-.00155	-1.00155	-.05565	-.03935	-.05563	-.00036	-.00005
20	-1.03	0	.00028	0	.00000	.00025	-.00028	-1.00028	-.02351	-.01662	-.02350	-.00008	.00002
21	-.22	0	.00004	0	.00000	.00008	-.00004	-1.00004	-.00839	-.00593	-.00839	-.00003	.00000
22	.73	0	.00005	0	.00000	.00018	-.00005	-1.00005	-.01029	-.00728	-.01029	-.00006	.00000
23	1.28	0	.00643	0	.00002	.00256	-.00643	-1.00647	-.11376	-.08047	-.11361	-.00252	-.00048
24	.73	0	.04744	0	.00015	.00301	-.04744	-1.04981	-.31226	-.22317	-.31178	-.00538	.00148
25	-.70	0	.00820	0	.00002	.00194	-.00820	-1.00827	-.12848	-.09095	-.12836	-.00121	.00052
26	-1.15	0	.00027	0	.00000	.00017	-.00027	-1.00027	-.02345	-.01658	-.02344	-.00012	.00002
27	-.71	0	.00001	0	.00000	.00002	-.00001	-1.00001	-.00530	-.00374	-.00530	-.00001	.00000
28	-.22	0	.00001	0	.00000	.00002	-.00001	-1.00001	-.00348	-.00246	-.00348	-.00001	.00000

Figure 19: Regression Variable Lists

We also have the option of specifying several diagnostic procedures to be performed on the data during the regression. The statistics and plots that can be generated are: *Classification Plots, Hosmer & Lemeshow goodness-of-fit tests, Residual listing* (greater than a given standard deviation)*, Correlations of Estimates, Iteration History,* and *Confidence Intervals*. Additionally, one can specify the cutoff value used to categorize values in a contingency table and whether or not to include a constant in the model. To select any of these options, choose *Options* from the logistical regression window and check or enter the desired values or variables in the window that appears. The window is similar to the one shown in Figure 20. Click *Continue* when through.

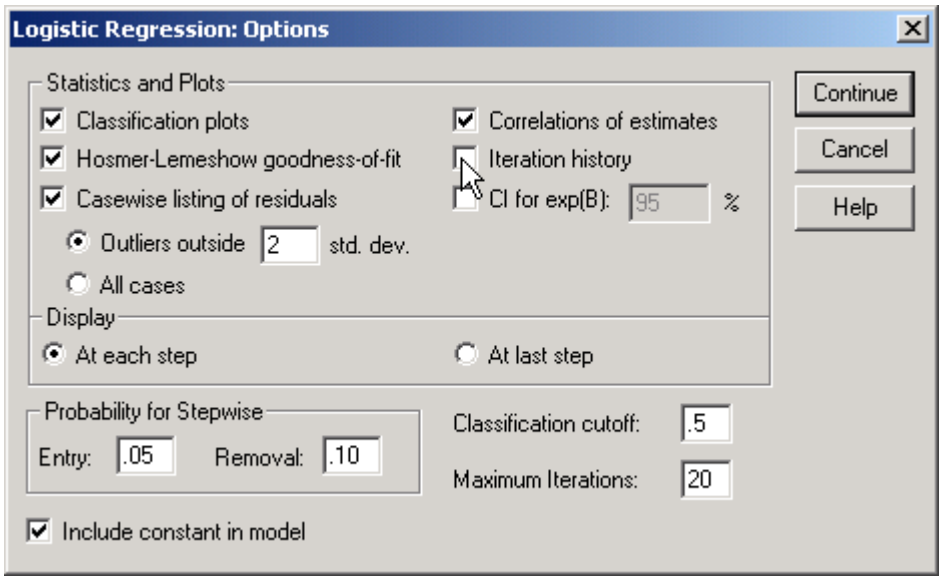


Figure 20: Logistical Regression Diagnostic Options

After you have completely specified the logistical regression variable, options, and diagnostics, click *OK* in the logistic regression window. After a brief delay during which time the regression is being performed, an *SPSS Viewer* window will appear with the results of the regression.

The contents of the *SPSS Viewer* window will now be discussed. Keep in mind that not all the charts that are discussed will appear in your *SPSS Viewer* window unless you select all the options available in Figure 20. Conversely, the charts describing the null model are not discussed here.

The case processing summary (Figure 21) details the number of cases included in the regression analysis and excluded from the analysis (missing cases).

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	2000	100.0
	Missing Cases	0	.0
	Total	2000	100.0
Unselected Cases		0	.0
Total		2000	100.0

a. If weight is in effect, see classification table for the total number of cases.

Figure 21: Case Processing Summary

The dependent variable encoding is shown in Figure 22. The encoding is the same as the original values due to the fact that our dependent variable was coded as either 0 or 1.

Dependent Variable Encoding

Original Value	Internal Value
0	0
1	1

Figure 22: Dependent Variable Encoding

Figure 23 shows the iteration history of the regression including the –2 Log-likelihood at each iteration as well as the Beta coefficients associated with each independent variable.

Iteration History[a,b,c,d]

Iteration		-2 Log likelihood	Coefficients								
			Constant	STRESS	STRAIN	DSTRESS	DSTRAIN	STRAIN by STRESS	DSTRESS by STRESS	DSTRAIN by STRAIN	DSTRAIN by DSTRESS
Step 1	1	684.025	-1.914	-.009	.028	-.006	.149	-.021	-.026	.060	-.009
	2	388.880	-2.921	-.036	.062	.008	.438	-.058	-.080	.153	-.021
	3	281.499	-3.833	-.087	.066	.099	.957	-.120	-.187	.262	-.008
	4	234.255	-4.812	-.096	.027	.233	1.558	-.194	-.346	.344	.085
	5	216.471	-5.803	-.083	-.014	.186	2.087	-.256	-.504	.422	.281
	6	210.920	-6.605	-.100	-.037	-.268	2.506	-.292	-.586	.470	.594
	7	209.517	-7.156	-.139	-.040	-.805	2.803	-.294	-.616	.477	.876
	8	209.426	-7.349	-.161	-.041	-.957	2.911	-.289	-.635	.475	.962
	9	209.426	-7.363	-.163	-.042	-.967	2.919	-.288	-.637	.475	.968

a. Method: Enter

b. Constant is included in the model.

c. Initial -2 Log Likelihood: 407.623

d. Estimation terminated at iteration number 9 because log-likelihood decreased by less than .010 percent.

Figure 23: Iteration History

The model summary is shown in Figure 24. It shows the goodness-of-fit measurements -2 Log-likelihood, Cox & Snell R^2, and the Nagelkerke R^2 values for the final model.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	209.426	.094	.512

Figure 24: Model Summary (Goodness of Fit)

Figure 25 indicates the Hosmer & Lemeshow Chi-square value for the final model.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	3.256	8	.917

Figure 25: Hosmer & Lemeshow Goodness of Fit

Figure 26 shows the Hosmer and Lemeshow contingency table. The distribution of response probabilities are divided into 10 equal groups and a contingency table is computed for each group. This is represented by this figure.

w Test

		SPK = 0		SPK = 1		
		Observed	Expected	Observed	Expected	Total
Step 1	1	200	200.000	0	.000	200
	2	200	199.998	0	.002	200
	3	200	199.989	0	.011	200
	4	200	199.966	0	.034	200
	5	200	199.915	0	.085	200
	6	200	199.819	0	.181	200
	7	200	199.580	0	.420	200
	8	200	198.894	0	1.106	200
	9	198	196.106	2	3.894	200
	10	160	163.734	40	36.266	200

Figure 26: Hosmer & Lemeshow Contingency Table

Figure 27 shows a simple contingency table with a cut value of 0.5. This is effectively the same as Figure 27 but instead of 10 groups, there are only 2 groups (response probability greater than 0.5 and less than 0.5).

Classification Table[a]

Observed			Predicted		
			SPK		Percentage Correct
			0	1	
Step 1	SPK	0	1949	9	99.5
		1	33	9	21.4
	Overall Percentage				97.9

a. The cut value is .500

Figure 27: Contingency Table

Figure 28 details the statistics associated with the fitted Beta coefficients in the model. The standard error (S.E.), Wald, and odds ratios (Exp(B)) are all listed in addition to the confidence interval for the coefficient.

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B) Lower	Upper
Step 1[a]	STRESS	-.163	.458	.127	1	.721	.849	.346	2.083
	STRAIN	-.042	.581	.005	1	.943	.959	.307	2.994
	DSTRESS	-.967	.820	1.391	1	.238	.380	.076	1.897
	DSTRAIN	2.919	.548	28.381	1	.000	18.519	6.328	54.199
	STRAIN by STRESS	-.288	.299	.928	1	.335	.750	.417	1.347
	DSTRESS by STRESS	-.637	.370	2.962	1	.085	.529	.256	1.092
	DSTRAIN by STRAIN	.475	.376	1.592	1	.207	1.608	.769	3.363
	DSTRAIN by DSTRESS	.968	.486	3.966	1	.046	2.632	1.015	6.820
	Constant	-7.363	.766	92.330	1	.000	.001		

a. Variable(s) entered on step 1: STRESS, STRAIN, DSTRESS, DSTRAIN, STRAIN * STRESS , DSTRESS * STRESS , DSTRAIN * STRAIN , DSTRAIN * DSTRESS .

Figure 28: Coefficient Statistics

Figure 29 shows the correlation matrix for the fitted coefficients. This gives an indication of the degree to which all the included independent variables interact with one another

Correlation Matrix

Step		Constant	STRESS	STRAIN	DSTRE	DSTRAIN	STRAIN by STRESS	DSTRESS by STRESS	DSTRAIN by STRAIN	DSTRAIN by DSTRESS
1	Constant	1.000	-.013	.125	.605	-.639	-.010	-.055	-.102	-.450
	STRESS	-.013	1.000	-.113	-.154	-.728	-.210	.374	.121	-.265
	STRAIN	.125	-.113	1.000	.496	-.018	-.009	.110	-.680	-.396
	DSTRESS	.605	-.154	.496	1.000	-.293	.047	.019	-.355	-.745
	DSTRAIN	-.639	-.728	-.018	-.293	1.000	.101	-.287	.028	.554
	STRAIN by STRESS	-.010	-.210	-.009	.047	.101	1.000	.208	-.676	-.191
	DSTRESS by STRESS	-.055	.374	.110	.019	-.287	.208	1.000	-.233	-.600
	DSTRAIN by STRAIN	-.102	.121	-.680	-.355	.028	-.676	-.233	1.000	.440
	DSTRAIN by DSTRESS	-.450	-.265	-.396	-.745	.554	-.191	-.600	.440	1.000

Figure 29: Correlation Matrix

Figure 30 is a casewise listing of residuals that fall outside 2 standard deviations from the mean of all the residuals. These cases are considered outliers since they lie outside the confidence interval.

Casewise List[b]

Case	Selected Status[a]	Observed SPK	Predicte	Predicted Group	Temporary Variable Resid	ZResid
89	S	1**	.074	0	.926	3.527
141	S	0**	.895	1	-.895	-2.923
244	S	1**	.091	0	.909	3.169
280	S	1**	.075	0	.925	3.519
316	S	1**	.042	0	.958	4.787
460	S	1**	.080	0	.920	3.380
550	S	1**	.104	0	.896	2.929
721	S	1**	.021	0	.979	6.907
823	S	0**	.870	1	-.870	-2.592
944	S	1**	.040	0	.960	4.897
964	S	1**	.131	0	.869	2.577
995	S	1**	.041	0	.959	4.865
1560	S	1**	.091	0	.909	3.167
1968	S	1**	.020	0	.980	6.962
1998	S	1**	.133	0	.867	2.553

a. S = Selected, U = Unselected cases, and ** = Misclassified cases.

b. Cases with studentized residuals greater than 2.000 are listed.

Figure 30: Residual Listing

Automation

If one had to complete all the steps detailed in the *Multivariate Logistical Regression* section for the hundreds of regressions that one must run for a typical analysis, it would be very time-consuming. Therefore, we now discuss the scripting capabilities of the software. The SPSS software comes with an excellent reference for scripting. It can be found in the **Help ⇒ Syntax Guide ⇒ Regression Models** help file under the heading *Logistic Regression* (page 608). This section completely explains how to write and execute scripts.

While scripts can automate the execution of regressions, they do not convert the results of the regression analysis into a format that can be easily plotted with SPSS or other software. For this reason, it is necessary to write additional scripts/macros (called "syntax" by SPSS) to convert the results of a regression into a format that is more conducive to plotting. The easiest way to complete this is to execute a macro that converts all the data from the *SPSS Viewer* into a Microsoft Excel format, which can then be parsed and the desired data extracted. The SPSS website has a section devoted to scripting (http://www.spss.com/tech/scptxchg/). There one can find a script called "Export_to_Excel_BIFF.SBS" (http://www.spss.com/tech/scptxchg/export.htm). This script will convert the data found in an *SPSS Viewer* window into Excel with one *SPSS Viewer* table per Excel sheet.

References

1. Agresti, A., Categorical Data Analysis, New York: Wiley, 1984

2. Aldrich, J.H., Nelson, F.D., Linear Probability, Logit, and Probit Models, *Series: Quantitative Applications in the Social Sciences*, Beverly Hills, CA: Sage Publications, 1984

3. Balakrishnan, N., Handbook of the Logistic Distribution, New York: Marcel Dekker, 1991

4. Bendel, R.B., Afifi, A.A., "Comparison of the Stopping Rules in Forward Regression," *Journal of the American Statistical Association*, 72: 46-53, 1977

5. Bender, R., Grouven, U., "Ordinal Logistic Regression in Medical Research," *Journal of the Royal College of Physicians of London*, 31: 546-551, 1997

6. Berkson, J., "Maximum Likelihood and Minimum Chi-squared Estimates of the Logistic Function," *Journal of the American Statistical Association*, 50: 130-162, 1955

7. Berry, W.D., Understanding Regression Assumptions, *Series: Quantitative Applications in the Social Sciences*, Newbury Park, CA: Sage Publications, 1993

8. Bi, Q., "Obtaining the Recovery Function from the Interval Histogram and the Poststimulus Time Histogram," *IEEE Transactions on Systems, Man, and Cybernetics*, 22: 580-585, 1992

9. Breslow, N., "Odds Ratio Estimators when the Data are Sparse," *Biometrika*, 68: 73-84, 1981

10. Carey, V., Zeger, S.L., Diggie, P., "Modeling Multivariate Binary Data with Alternating Logistic Regressions," *Biometrika*, 80: 517-526, 1993

11. Clayton, D.G., "Some Odds Ratio Statistics for the Analysis of Ordered Categorical Data," *Biometrika*, 61: 525-531, 1974

12. Cleary, P.D., Angel, R., "The Analysis of Relationships Involving Dichotomous Dependent Variables," *Journal of Health and Social Behavior*, 25: 334-348, 1984

13. Collica, R.S., Ramirez, J.G., Cantell, B.S., "Statistical Analysis of Particle/Defect Data Experiments Using Poisson and Logistic Regression," *The IEEE International Workshop on Defect and Fault Tolerance in VLSI Systems*, 127-135, 1994

14. Connolly, M.A., and Liang, K., "Conditional Logistic Regression Models for Correlated Binary Data," *Biometrika*, 75: 501-506, 1988

15. Costner, H.L., "Criteria for Measures of Association," *American Sociological Review*, 30: 341-353, 1965

16. Cox, D.R., Analysis of Binary Data, New York: Wiley & Sons, 1970

17. Del Prete, Z. and Grigg, P., "Identification of Encoded Variables in Sensory Neurons using Multivariate Logistic Regression Analysis," *Submitted to IEEE Transactions on Biomedical Engineering Jan. 2000*

18. Del Prete, Z. and Grigg, P., "Responses of Rapidly Adapting Afferent Neurons to Dynamic Stretch in Rat Hairy Skin," *Journal of Neurophysiology*, 80: 745-754, 1988

19. Eliason, S.R., Maximum Likelihood Estimation: Logic and Practice, *Series: Quantitative Applications in the Social Sciences*, Newbury Park, CA: Sage Publications, 1993

20. Fitzmaurice, G.M., "A Model for Binary Time Series Data with Serial Odds Ratio Patterns," *Applied Statistics*, 44: 51-61, 1995

21. Grigg, P., "Stretch Sensitivity of Mechanoreceptor Neurons in Rat Hairy Skin," *Journal of Neurophysiology*, 76: 2886-2895, 1996

22. Hall, G.H., Round, A.P., "Logistic Regression - Explanation and Use." *Journal of the Royal College of Physicians of London*, 28: 242-246, 1994

23. Hardy, M., Regression with Dummy Variables, *Series: Quantitative Applications in the Social Sciences*, Newbury Park, CA: Sage Publications, 1993

24. Hosmer, D.W., Wang, C.-Y., Lemeshow, S., "A Computer Program for Stepwise Logistic Regression using Maximum Likelihood Estimation," *Computer Programs in Biomedicine*, 8: 121-134, 1978

25. Hosmer, D.W., Lemeshow, S., Applied Logistic Regression Second Edition, New York: Wiley & Sons, 2000

26. Hosmer, D.W., Lemeshow, S., "Goodness-of-fit Testing for the Logistic Regression Model when Estimated Probabilities are Small," *Biometrical Journal*, 30: 911-924, 1988

27. Imrey, P.B., Koch, G.G., and Stokes, M.E., "Categorical Data Analysis: Some Reflections on the Log Linear Model and Logistic Regression. Part I: Historical

and Methodological Overview," *International Statistical Review*, 49: 265-283, 1981

28. Jennings, D.E., "Outliers and Residual Distributions in Logistic Regression," *Journal of the American Statistical Association*, 81: 987-990, 1986

29. Khalsa, P.S., LaMotte, R.H., and Grigg, P., "Tensile and Compressive Responses of Nociceptors in Rat Hairy Skin," *Journal of Neurophysiology*, 78: 492-505, 1997

30. Kleinbaum, D.G., Logistic regression: A Self-Learning Text. New York: Springer-Verlag, 1994

31. Landwehr, J.M., Pregibon, D., and Shoemaker, A.C., "Graphical Methods for Assessing Logistic Regression Models," *Journal of the American Statistical Association*, 79: 61-71, 1984

32. Le Cessie, S. and Van Houwelingen, J.C., "Logistic Regression for Correlated Binary Data," *Applied Statistics*, 43: 95-108, 1994

33. Liang, K., Zeger, S.L., and Qaqish, B., "Multivariate Regression Analysis for Categorical Data," *Journal of the Royal Statistical Society: Series B (Methodological)*, 54: 3-40, 1992

34. Liao, T.F., Interpreting Probability Models: Logit, Probit, and other Generalized Linear Models, *Series: Quantitative Applications in the Social Sciences*, Thousand Oaks, CA: Sage Publications, 1994

35. Lipsiz, S.R., Laird, N.M., and Harrington, D.P., "Generalized Estimating Equations for Correlated Binary Data: Using the Odds Ratio as a Measure of Association," *Biometrika*, 78: 153-160, 1991

36. Looft, F.J., "Response of Monkey Glabrous Skin Mechanoreceptors to Random Noise Sequences I: Temporal Response Characteristics," *Somatosensory and Motor Research*, 11: 327-344, 1994

37. Looft, F.J., "Response of Monkey Glabrous Skin Mechanoreceptors to Random Noise Sequences II: Dynamic Stimulus State," *Somatosensory and Motor Research*, 13: 11-28, 1996

38. Looft, F.J., "Response of Monkey Glabrous Skin Mechanoreceptors to Random Noise Sequences III: Spectral Analysis," *Somatosensory and Motor Research*, 13: 235-244, 1996

39. Looft, F.J., "Controlled In-Vitro Stimulation of Rat Tactile Mechanoreceptors: Response Variability," Grant Report, 2000

40. McCullagh, P., Nelder, J.A., Generalized Linear Models, Cambridge: University Press, 1983

41. Menard, S., Applied logistic regression analysis, *Series: Quantitative Applications in the Social Sciences*, Thousand Oaks, CA: Sage Publications, 1995

42. Pregibon, D., "Logistic Regression Diagnostics," *Annals of Statistics*, 9: 705-724, 1981

43. Pubols, B.H., "Factors Affecting Cutaneous Mechanoreceptor Response. I. Constant-Force versus Constant Displacement Stimulation," *Journal of Neurophysiology*, 47: 515-529, 1982

44. Pubols, B.H., "Factors Affecting Cutaneous Mechanoreceptor Response. II. Changes in the Mechanical Properties of Skin with Repeated Stimulation," *Journal of Neurophysiology*, 47: 530-542, 1982

45. Rice, J.C., "Logistic regression: An Introduction," In B. Thompson, ed., *Advances in Social Science Methodology*, 3: 191-245. Greenwich, CT: JAI Press, 1994

46. SPSS Statistical Algorithms, McGraw Hill: New York, NY, 1983

47. Srinivasan, M.A., "A Theoretical Framework to Represent and Predict the Responses of Cutaneous Mechanoreceptors," *Massachusetts Institute of Technology Technical Report*, 1986

48. Taniguchi, K., Ogawa, H., "Spectral Analysis of Instantaneous Frequency Responses to Sinusoidal Stimulation in Cutaneous Mechanoreceptor Afferent Units of Frogs," *Japanese Journal of Physiology*, 37: 1031-1049, 1987

49. Tritchler, D., "An Algorithm for Exact Logistic Regression," *Journal of the American Statistical Association*, 79: 709-711, 1984

50. Vermunt, J.K., Log-linear Models for Event Histories, *Series: Quantitative Applications in the Social Sciences*, London: Sage Publications, 1997

51. Want, M., Eddy, J.M., Fitzhugh, E.C., "Application of Odds Ratio and Logistic Models in Epidemiology and Health Research," *Health Values*, 19: 59-62, 1995

52. Whittemore, A.S., "Sample Size for Logistic Regression with Small Response Probability," *Journal of the American Statistical Association*, 76: 27-32, 1981

53. Wright, R.E., "Logistic Regression," In Grimm, L.G., & Yarnold, P.R., eds., *Reading and Understanding Multivariate Statistics*. Washington, DC: American Psychological Association, 1995