ALTERNATIVE STATISTICAL MODELS THAT ACCOUNT FOR CLUSTERING IN DENTAL IMPLANT FAILURE DATA

by

Heidi M. Huber

BS, University of Pittsburgh, 1983

D.M.D., University of Pittsburgh School of Dental Medicine, 1987

Submitted to the Graduate Faculty of

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2004

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Heidi M. Huber

It was defended on

August 26, 2004

and approved by

Robert Weyant, D.M.D., DrPH, Chair, Department of Dental Public Health, School of Dental Medicine, University of Pittsburgh

Joseph P. Costantino, DrPH, Professor and Director NSABP Biostistical Center, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Thesis Director: Roslyn A. Stone, Ph.D., Associate Professor, Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh

Roslyn A. Stone, Ph.D.

ALTERNATIVE STATISTICAL MODELS THAT ACCOUNT FOR CLUSTERING IN DENTAL IMPLANT FAILURE DATA

Heidi M. Huber, M.S.

University of Pittsburgh, 2004

ABSTRACT

Longitudinal data analysis is a major component of public health care assessment. It is important to know how treatments compare over time, how diseases occurr and recurr, and how environmental or other exposures influence to a disease processes over time. Investigations of such topics involve the statistical analysis of time-to-event data in various areas of health care.

Long term dental assessment of dental restorations have typically employed statistical analyses that assume independence of the restorations within the patient. Dental data naturally occur in the form of clusters. The patient is a cluster of correlated dental units (teeth) to be evaluated. Statistical analysis of the dental units without acknowledgement of within-cluster correlation can underestimate standard errors, which can erroneously inflate the significance level of between-cluster predictors in a model.

The purpose of this thesis is to 1) review the statistical literature on the analysis of dental implant data, 2) create a suitable longitudinal data file of dental implant failure, 3) describe the data management and statistical methods used, 4) compare alternative statistical models to analyze clustered survival data, and 5) show how these models can be used to identify some patient-level and implant site-level predictors of implant failure. We consider logistic regression, discrete survival, generalized estimating equations and the Cox model with and without frailty, and examine the associations between implant failure and patient race, implant type, and oral location of implant. Models that ignore the clustering consistently overestimate the significance of patient race.

ACKNOWLEDGMENTS

I am truly grateful to my Lord for blessing me with the people who helped me with this thesis. Roslyn Stone is my director and inspiration for this project. I thank you Dr. Stone for your patience, direction, wisdom and support. I thank Dr. Robert Weyant for permitting me to use the data for this thesis and for introducing me to Biostatistics. I thank Dr. Joseph Costantino for his support and advice during my degree process. I thank my family (especially my daughter Sarah Richards) for giving me the needed encouragement to complete this thesis.

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction

Root form or endosseous dental implants were introduced to the dental community by Dr. Per Ingmar Branemark, an orthopedic surgeon, in 1969 (Branemark. P.I., 1969, and 1983). Dr. Branemark (Branemark. P.I., 1981, 1983, 1984) has been a major contributor to the literature on scientific studies of root form dental implants as well as their current clinical application. Clinically, dental implants have provided patients with the ability to wear prostheses with considerable comfort, function, and esthetic advantages. However, the prosthetic advantages depend on the survival of the actual implant fixtures.

A dental implant is a surgically placed element that can support a prosthetic replacement of an edentulous region. The word “edentulous” refers to a “patient who is without teeth” or “region of the mouth that is without teeth”. Implants are also a useful, and often necessary, means of retaining extraoral prostheses for patients who have facial defects due to surgery for tumor removal, trauma, or congenitally missing tissues. These implants do not replace teeth, but serve to retain an often large prosthesis or a prosthesis that possibly will lie on a tissue bed that is either mobile or not conducive to holding a prosthesis by conventional means (e.g. retention via either mechanical or adhesive modes). Dental implants can be of various materials (e.g. titanium, ceramic, glass, or other metals), forms (e.g. root form, staple, blade), and coatings and surfaces (e.g. hydroxyappatite, titanium, serrated, or acid etched).

This thesis will focus on only the endosseous root form dental implants that are involved with intraoral treatment. Implants of this type are placed under several circumstances. In scenario (1), patients require the replacement of only one tooth (single-tooth replacement). Typically these are patients who have an anterior (front tooth region) missing and do not want the teeth adjacent to the edentulous space to be altered, defaced, or used as abutments for a bridge or partial denture as replacement for this missing tooth. Also, patients can have single posterior (back tooth region) implant placed. In scenario (2) patients are partially edentulous in either arch (maxilla (upper arch) or mandible (lower arch)). These patients can have two or more implants, depending on the edentulous span and the treatment plan of the prosthodontist and surgeon. Sometimes a single implant that is linked to a natural tooth in a bridge is placed to decrease the number of implants required, due to either the finances of the patient and/or the compromised and questionable health status of the natural tooth. The bridge usually has a precision attachment to allow for the provision of loss of the tooth without total destruction of all bridge units, allowing placement of an implant in the region where the tooth is lost. The inequitable force distribution in this restorative design has been implicated as a cause of the ultimate failure of the restoration. The implant is anklosed (the implant is fused to the bone) whereas the tooth is joined to the bone by a periodontal ligament. The implant will not move under function, but a natural tooth would. This situation does not occur very frequently. In scenario (3) patients are totally edentulous in either arch (maxilla or mandible).

In the first scenario, the restoration is in the form of a crown that is cement or screw retained. In the second scenario the form of the restoration is usually a fixed bridge or independent units and may or may not be removable. In the third scenario, a fixed screw-retained bar secures a denture that can be removed by the patient. Another option for some patients is a “hybrid prosthesis”, which is a metal substructure with an acrylic super-structure that is screw-retained. This prosthesis is only removed by the prosthodontist for annual evaluations, cleanings and maintenance. Scenarios two and three present opportunities for multiple implants and multiple implant failures per patient. All three scenarios present the opportunity for multiple failures and subsequent replacement of the failed implants.

Statistical analysis of implant failure times can be based on survival analysis, or time-to-event analysis, which incorporates placement dates and evaluations over time, and a censoring indicator to denote whether an implant has in fact failed by the end of follow-up. Standard survival methods assume that observations are independent. However, dental implant data are often clustered. Each individual has potentially 32 teeth to be evaluated. Spatial clustering can also occur, such as teeth within a quadrant or other region in the mouth. For example anterior or posterior teeth or teeth that are alike contralaterally in the same arch may be more similar to each other than to teeth in other regions. In this thesis we focus on within person clustering and investigate some aspects of spatial characteristics with respect to implant survival.

Another characteristic in survival data is that of variable time at risk. Dental implants may be placed in a patient at different times, or a patient may have several implants placed at one time. It is clinically practical to place several implants at one time. Also, implants may be replaced after failing. The survival analysis of dental implant failure presents the complexities of varying time at risk, repeated failures, and clustered observations.

This thesis involves the secondary analysis of an existing dataset from the dissertation of Robert Weyant, D.M.D., Dr. P.H. (1991), which includes placement dates and follow-up for dental implants over almost 7 years, for 1,246 patients in five participating sites. These data were obtained from the Department of Veterans Affairs' Dental Implant Registry, which was created in 1987. In his initial investigation of these data, Weyant described a "quasi-experimental study design" whose purpose was to evaluate the association of patient and treatment facility characteristics on dental implant performance and to estimate the survival probabilities of various dental implants. He uses a correlated binomial model to determine the degree of intraclass (within-patient) correlation and to adjust the binomial probability of several dependent variables (surgical complications, implant health status, and implant failures). In his survival analysis of dental implant data, Weyant used the Kaplan-Meier (1959) or life-table methods, ignoring the dependency of the implants within each patient. In his dissertation, Dr. Weyant acknowledges the need to account for intra-patient correlation and suggests but does not implement two procedures to address this issue statistically, bootstrapping and ordinary least squares (linear) regression,.

Statistical analysis of clustered data should take into account the dependence of the units within the cluster. The purpose of this thesis is to 1) review the statistical literature on the analysis of dental implant data, 2) create a suitable longitudinal data file of dental implant failure, 3) describe the data management and statistical analysis methods used on the Weyant data, 4) compare alternative statistical models to analyze clustered survival data, and 5) Show how these models can be used to identify some patient-level and implant site-level predictors of implant failure.

CHAPTER 2

Review of the Literature

Statistical Approaches Taken in the Dental Literature on Implant Failure

The majority of articles discussing the survival analysis of dental implants utilize statistical methods that ignore the correlation structure or innate clustering of such dental data. Such analyses treat multiple dental implants within each patient as independent units. The implicit assumption is that the failure of each implant does not depend on either the status of any other implant unit within the same patient or patient characteristics shared by all implants within a patient. This assumption is not justified clinically. Often the failure of an implant in one region of the mouth may coincide with bone loss around the failing implant. Implants near the failing implant may also fail due to this bone loss and subsequent lack of bone. Local or disperse periodontal problems, infection, patient habits, treatments or medications, and other anatomic or systemic problems are patient-specific factors that can contribute to implant failure. Positional and systemic variables potentially influence the failure of all implants placed in each patient. These effects contribute to within-patient correlation.

Some of the more recent articles on the analysis of implant survival have acknowledged the issue of intracluster dependence, and have used statistical techniques to account for this. We now review some statistical methods used to analyze dental implant failure data.

Kaplan-Meier Estimator

The Kaplan-Meier, or Product-Limit, estimator is typically presented as an overall measure of implant survival (Kaplan-Meier 1959). The majority of studies reporting this type of analysis implicitly assume that implants are independent of each other (Wheeler, 1996; Buser et. al., 1997; Brocard et. al., 2000). The Kaplan-Meier estimator is a step function that jumps at the event times and can accommodate more than one failure at each event time. The survival estimator at a given time t_k is the product of conditional probabilities of survival at the previous event times. The conditional probability of survival beyond t_k is given by $[1-d_k/n_k]$ where k is the time interval of interest, d_k denotes the number of implant failures at t_k and n_k is the total number of implants at risk for failure just prior to t_k (e.g. not yet failed and still under observation). No assumption is made about the functional form of the survival function. The variance estimator does not account for the clustering of implants within patients.

Logistic Regression

Logistic regression is used frequently to describe biological relationships between predictor variables and dichotomous or binary outcome variables (Hosmer and Lemeshow (1989)). The logistic distribution is a flexible and convenient function from a mathematical perspective, and provides a suitable model for many biological mechanisms. Logistic regression is a generalized linear model with the logit link function (McCullough and Nelder), where logit p = $\log[(p)/(1-p)]$ and p is the probability of an event in a fixed interval of time.

In logistic regression, the likelihood function is constructed under the assumption that each observation is independent. For the logistic model, $E(y_{ij}) = \pi_{ij} = (1 + e - X'_{ij}T\beta)^{-1}$ where X_{ij} is the column vector of covariates x_{ij}, i indexes the cluster (patient) and j indexes the observations within the cluster. Parameters in a logistic model are estimated by maximizing the binomial likelihood. The log likelihood equations are differentiated with respect to the parameters, and the resulting score equations, $X^T AV^{-1}(y-\pi)=0$, are set equal to zero to obtain maximum likelihood estimates of the parameters. Because these likelihood equations are not linear in the parameters, iterative methods are required for their solution using generalized weighted least squares (McCullough and Nelder). The score equation for the logistic regression model is modified as follows to define generalized estimating equations:

One problem in using a simple logistic model for time-to-event outcomes is that every patient is assumed to be at risk for the entire time interval. This assumption may not be valid for studies with long follow-up or other situations where patients have variable time at risk.

In several studies (Albrektsson et. al., 1996; Jemt et. al., 1996; Lazzara et. al., 1996; Rosenquist and Grenthe, 1996; Hising et. al., 2001) oral implant survival data are analyzed using survival as a binary outcome over a fixed interval of time. This approach often overestimates survival because long-term failures are mixed with the early success of recently placed implants (Eckert and Wollan, 1998).

The Discrete Proportional Odds and Discrete Proportional Hazards Model

The discrete-time proportional odds model (Cox, 1972) is an extension of logistic regression that accounts for time at risk. In this model, the conditional probability of an event (e.g. implant failure) during time interval m, $(m=1,....,M)$, is p_m, and logit $p_m(x)=\alpha_m+x^t\beta$. The α_m parameters represent time-interval specific intercepts for a patient with a reference vector of regression variables (x=0), the log-odds of failure in interval m conditional on not failing prior to m. The $x^t\beta$ is the linear predictor, which is interpreted as the logarithm of the relative risk of failure at time t_m for an individual with covariates $x \neq 0$ relative to an individual with x=0.

Under this model, the odds ratio of an event at time m for two individuals with covariates x_1 and x_2 respectively, does not depend on the time interval m, which is the *proportional-odds assumption*:

$$p_m(x_1)[1-p_m(x_2)]/[1-p_m(x_1)p_m(x_2)] = \exp\{(x_1-x_2)^t \beta\} \qquad \text{(1) (Breslow, Nelson, 1992)}$$

The odds ratio approximates the hazard ratio (relative risk) when the probability of an event in the time interval is small.

Prentice and Gloeckler (1978) showed that the discrete-time analog of the continuous time proportional hazards model is:

$$\log(-\log(1-p_m(x))) = \alpha_m + x^t\beta \qquad (2)$$

Here $\log[-\log(1-p)]$ is the complementary log-log (c-log-log) transform. The conditional probability of an event occurring in each time interval is assumed to be binomial with the denominator equal to the number at risk at time m. If two time intervals are of interest, the conditional probability of survival over the two periods is $c-\log(p) = [1-p_1(x)][1-p_2(x)] = \exp\left[-\left(e_1^{\alpha}+e_2^{\alpha}\right)e^{x^t\beta}\right]$ which is linear in x. When p is small, there is not a substantial difference numerically between the discrete proportional odds and discrete proportional hazards models. However, the interpretations of the parameter estimates are different. The β_m represent log hazard ratios in the discrete proportional hazards model.

For both models, patients contribute $1-p_m(x)$ to the likelihood function for each interval m in which they have not yet failed. Patients experiencing a failure contribute $p_m(x)$. Computationally, a separate record for each time interval for each patient is created; this data set-up also accommodates time-dependent predictors. In longitudinal data, time-dependent predictors can be key to understanding a history and mechanism of a potential disease process.

These variables change in value over the time period of study and can include history of previous failures as predictors. The proportional odds (or hazard) assumption can be tested by including interactions of time and x in the model.

Robust Variance Estimation

Four assumptions are made with a logistic regression model: (1) the link function is specified correctly, (2) the error structure is specified correctly, and (3) the form of the linear predictor $\left(x^t\beta\right)$ is correct, and (4) the observations are independent. The score equations (or likelihood equations) are:

$$U\left(\beta_k\right)=\frac{\partial l}{\partial\beta_k}=\sum_i\sum_k x_{ik}\left(y_i-\pi_i\right)=0 \quad \text{(3) (Carlin, J.B., et.al.,1999)}$$

Here k indexes the β parameters and i indexes the patients. A vector form of the score equations is presented as:

$$\mathbf{U}(\beta)=\boldsymbol{X}^{\mathbf{T}}(\mathbf{y}-\pi)=0 \text{ (4)} \qquad \text{(4)} \qquad \text{(Carlin, J.B., et.al.,1999)}$$

where **y** and π are vectors of the data and parameters respectively and ***X*** is a design matrix with the number of rows equal to the length of the **y** vector (***n***), and the number of columns equal to the number of estimated parameters. The corresponding information matrix is:

$$\boldsymbol{COV(}\hat{\beta}^{ML}\boldsymbol{)}=(\boldsymbol{X}^{\mathbf{t}}\ \hat{A}\,\boldsymbol{X})^{-1} \qquad \text{(5)} \qquad \text{(Carlin, J.B., et.al.,1999)}$$

where $\hat{A}$ =diag$\left(\hat{\pi}_i\left(1-\hat{\pi}_i\right)\right)$, a diagonal matrix of the binomial variances calculated at the values of π as the solution to the maximum likelihood (ML) equations. This is the model-based variance.

When responses are potentially correlated, consistent estimates of $\hat{\beta}$ can be obtained using ML as long as the first-order specification is correct (this means that the model for the mean of ***y*** is correct). Consistency means that point estimates become close to the true population values as the sample size increases. However, the standard errors of between-cluster predictors generally will tend to be underestimated, because the covariance matrix will be estimated based on the assumption that the observations are independent. Some of the methods proposed to account for this dependence are the Jack-knife and Bootstrap, which involve resampling with replacement (for the Bootstrap) and without replacement (for the Jack-knife) (J.B. Carlin et. al., 1999). Another general approach is to use the information-sandwich variance estimator variance proposed by Huber and White (1967). This approach incorporates a "”robust” variance estimator is consistent even when the covariance structure is not correctly specified. The robust estimator is:

$$\boldsymbol{COV}_{\mathbf{R}}(\hat{\beta}^{ML}) = (X^T \hat{A} X)^{-1} \sum_{i=1}^{n} \{X^T (y_i - \pi_i)(y_i - \pi_i)^T X_j (X^T \hat{A} X)^{-1}\} \qquad (6)$$

The robust estimator is often called the sandwich estimator because the “bread” is the $\boldsymbol{COV}_{\mathbf{R}}(\hat{\beta})^{ML}$ and the empirical estimator of the variance is the filling.

This empirical correction can be summed over independent observations (***i***=1,….,***n***) or over clusters (***i***=1,….,***C***). The “poor man’s GEE” approach is to fit a logistic regression model ignoring the clustering and use a robust variance estimator calculated at the cluster level.

Marginal Models

(Generalized Estimating Equations) GEE

GEE is an extension of generalized linear models that relaxes the independence assumption. In this quasi-likelihood approach, parameters are estimated by solving the quasi-score equations:

$$U^q(\beta) = D^T V^{-1}(y - \pi) = 0 \qquad (7)$$

where D is an $(n \times k)$ matrix of the derivatives of the expectation of the response variable with respect to β. The covariance matrix, $V = Cov(y)$, may not correspond to a likelihood

In **GEE,** the variance matrix in the score equation is a block diagonal with n submatrices, V_i where:

$$V_i = A_i^{1/2} R(\alpha) A_i^{1/2} \qquad (8)$$

and $R(\alpha)$ is the "working" correlation matrix. This $R(\alpha)$ may contain unknown parameters α that specify the correlation structure. Provided that the model for the mean is correctly specified, the standard error estimates obtained using **GEE** are consistent, even if $R(\alpha)$ is misspecified. However, the efficiency of estimating (β) increases when the correlation structure is more accurately specified.

The commonly specified working correlation structures include: (1) exchangeable; where the observations are equally correlated within a cluster, (2) autoregressive; where the correlation between two observations decreases exponentially over time, (3) stationary; where the correlation between observations depends on how far apart they are in time but not on the specific time points, or (4) unstructured, where the α_{st} allows for arbitrary correlation between observations at times s and t.

The **GEE** model can be fit using either a "model-based" or a "robust" variance estimator. The "robust" information-sandwich matrix in GEE is:

$$COV_R(\hat{\beta}^{GEE}) = (D^T V^{-1} D)^{-1} \sum_{i=1}^{n} \{D^T V_i^{-1} (y_i - \hat{\pi}_i)(y_i - \hat{\pi}_i)^T V_i^{-1} D_i (D^T V^{-1} D)^{-1}\} \qquad (9)$$

Although GEE for longitudinal data with time-dependent and time-independent predictors was proposed in 1986, these methods have only recently appeared in the dental literature. For example, Lambert PM et. al. (2000), use (GEE) to analyze the survival of dental implants. Morris et. al. (2000) evaluated implant survival in patients with type 2 diabetes over a period of 36 months and report that diabetic patients had more failures than non-diabetic patients. They compare models assuming independence vs. those considering correlation.

Ochi (2000) elaborates on the evaluation of clustered dental implant data. The authors explain that the implants are highly clustered in several hierarchical levels (i.e. implants within cases, implants within patients and implants within hospitals). The statistical methods used for this study involved a logistic regression analysis of the effects of predictors on survival to given stages. The authors used GEE as implemented in **SUDAAN** (*Research Triangle Institute, Research Triangle Park, NC.*) where the patient was the primary cluster. The primary clusters in some analyses were the participating institutions. Exchangeable and independent working correlations were assumed and statistical results were compared with the logistic regression analyses. Jacknifing was attempted and required very long computational times especially with large data sets. Kaplan-Meier survival analysis was done, and the authors state that the Cox regression plots and analyses were not routinely performed because of uncertainty of assessing survival status using a scheduled uncovering surgery date. In their paper, logistic regression was used to model the probability of failure by a specific timepoint. Despite reported difficulties with availability of software to handle the analysis of clustered data, this group acknowledges the need to account for this clustering statistically.

Survival Models

Cox Model

The Cox Proportional Hazards model is a semiparametric approach to survival analysis where failures are assessed in continuous time (Cox, 1972, 1975). In the Cox proportional hazards model the hazard of an event at time *t* in a patient with covariates x is:

$$h(t) = h_0(t)e^{x\beta} \qquad (10)$$

where, $h_0(t)$ is the baseline hazard and the covariates multiply the baseline hazard. The baseline hazard is not parameterized and the hazard shape over time is not specified. The partial likelihood function (Cox, 1975) is:

$$\prod_{i=1}^{I} \frac{\exp(x\beta)}{\sum_{l\varepsilon R(t_i)} \exp(x\beta)} \tag{11}$$

The value $R(t_i)$ represents a risk set at t_i, and includes those patients who have not yet failed and are under observation just before time t_i, the failure time for the i^{th} failure. The covariates of the patient who experienced the failure at t_i appear in the x term in the numerator. The parameter β is estimated by maximizing the partial likelihood function. The reference cumulative incidence function is:

$$H_0(t) = \int_0^t h_0(s)ds \tag{12}$$

The Breslow estimator of this function, which accommodates covariates, is:

$$\hat{H}_0(t) = \sum_{t_i \le t} \frac{d_i}{\sum_{l\varepsilon R(t_i)} \exp\left(x_l{}^t \beta\right)} \tag{13}$$

The t_j term indicates times that patients have failed, d_i indicates the number of cases at the i^{th} failure time ($d_i \ge 1$) (Breslow, 1974). When $\hat{\beta} = 0$ the denominator sum equals the total number at risk at t_i in equation (13).

The simplest version of this model assumes that the relative risk of an event for two groups of individuals with different covariate values is constant across the time interval studied. This is the “proportional hazards” assumption. However, the underlying incidence rate for the two groups is permitted to be different in a structured manner. The hazard of an event at thime t

for a person with predictors x_i compared to a person with predictors x_j, under the Cox proportional hazards model is:

$$h_0\, e^{x_i\beta} / h_0 e^{x_j\beta} \qquad (14)$$

If the covariates x_i and x_j are constant over time, then the above ratio is constant. In fact the baseline hazard cancels out of the calculation. The $e^{(x\beta)} = e^{(x_1\beta_1 + x_2\beta_2 + \ldots + x_j\beta_j)}$ part of the Cox model represents the hazard relative to a patient with $x = 0$, and $x\beta$ is the log-relative hazard. A parametric form is assumed for the covariates involved with the model but not for the baseline hazard.

The *ordering* of the failure times is the essential information used in the Cox model, not the actual failure time values. The Cox likelihood is a partial likelihood because the estimate of β obtained by maximizing this partial likelihood produces an asymptotic normal distribution with a mean equal to β and a variance-covariance matrix equal to the matrix of second derivatives of the partial likelihood with respect to β (Kalbfleisch and Prentice, 1980).

Tied Failure Times

The Cox model assumes that failure times are distinct, although ties do occur in practice. One way of dealing with tied failure times in the Cox model is by a marginal or continuous-time calculation. We do not know the exact ordering of the failures and can consider the possibility that implant a failed slightly before b. As implants are considered to fail in various orders the risk set will change to exclude the implants that failed. Since we are unsure of the order of implant failures, the marginal calculation uses both probabilities in the calculation $(P_{ab} + P_{ba})$. The term continuous-time arises because there is no assumption that the implants failed at the exact same time.

Another method of calculating the probability of tied failures is the partial, conditional logistic or discrete-time calculation (Peto, R., 1972). There is an assumption that the implants failed at the same time and the computation becomes a multinomial calculation where all the possibilities of implant failures is considered.

Another method of calculating the probability of tied failures is the Breslow (Breslow, 1974) approximation which is a less computationally intensive method. The calculation is an approximation of the exact marginal probability of tied failures. The risk set is not adjusted for prior failures. This approximation is adequate when failures are a small fraction of the risk set.

Another approximation that handles tied failures is the Efron approximation (Efron, B 1977), which adjusts the risk sets using probability weights and averages the risk sets. This approach is more accurate than the Breslow approximation, although computation time is higher.

Stratified Cox Model

A stratified Cox model allows a separate baseline hazard for each group. The proportionality assumption between groups is dropped. However the estimates β are constrained to be the same. The stratified model is:

$h_x(t) = h_{0S} e^{(x\beta)}$ where S denotes the stratum. (15)

The multiplicative effect covariate x is $e^{(x\beta)}$ in each stratum.

Manz M (2000) utilized a stratified analysis that indicated different bone loss patterns where the stata analyzed were, arch (maxillary vs. mandibular), jaw region (anterior vs. posterior), bone quality surface type (texture status), implant design (endosseous vs. other), smoking status (smoker vs. non-smoker), and postoperative antibiotic treatment (treatment vs. no-treatment). Manz (2000) points out the importance of controlling for confounding and accounting for correlation of data over time within patient.

Marginal Model

The marginal model, introduced by Lee et al. (1992), assumes proportional hazards for each implant given the patient's covariates. The model is:

$$h_{ij}\left(t \middle| X_{ij}\right) = h_0(t)e^{(X_{ij}\beta)},\ i = 1,\dots,n \quad j = 1,\dots\dots, J_i \qquad (16)$$

The estimation of β is approached with an independence working model for the data. This assumes that the observations within a patient are independent and the estimation is based on partial likelihood. According to Lee et. al. (1992) the estimator for β is consistent if the marginal model is specified correctly. The variance-covariance matrix of the estimator, β, is not valid when obtained from the corresponding information matrix.

The robust or "sandwich" estimator adjusts the covariance matrix for correlations between implants within patients. Based on the independence working model, the estimate of the variance correction matrix utilizes the following definitions:

T_{ij} is the time of evaluation of implant j within patient i, δ_{ij} is the failure indicator, and X_{ij} is a covariate vector for the jth implant in the ith patient. $Y_{ij}(t)$ is an indicator that implant j in patient i is at risk at time t. The survival functions are:

$$S_0(t) = \sum_{i=}^{n}\sum_{j=}^{Ji}\left[Y_{ij}(t)X_{ijk}\exp\left(\beta t X_{ij}\right)\right], \quad \text{and} \quad S_{1k}(t) = \sum_{i=}^{n}\sum_{j=}^{Ji}\left[Y_{ij}(t)X_{ijk}\exp\left(\beta t X_{ij}\right)\right], \quad k = 1,\dots, p \qquad (17)$$

where k is the number of covariates. The adjusted estimator of the variance of β presents is:

$$V = \hat{V}C\hat{V} \qquad (18)$$

The β estimator follows a large sample p-variate normal distribution with a mean of β and variance estimate obtained from V. A Wald test can be employed locally and globally. This model provides no estimate of the correlation between observations within a person.

Lin and Wei (1993) also consider the situation where the baseline hazard rate is different for each group and a common β represents covariate effects. They use the independence working model with a sandwich estimator for the variance.

Spiekerman and Lin (1998) evaluated the survival of teeth that are in different positions (anterior vs. posterior) relative to each other. Their analyses indicate that teeth in similar positions contralaterally tend to have similar survival distributions. The authors extend the concept of Lee, Wei, and Amato (1992) and Wei, Lin, and Weissfeld (1989) using the "quasi-likelihood" estimating equations with an independence working assumption and relate this to a stratified Cox model for univariate failure data (Kalbfleisch and Prentice (1980) where the strata (anterior vs. posterior regions of the mouth) are correlated and there is clustering of failure times within each stratum.

Frailty Models

Vaupel et. al. (1979) first presented the term "frailty" for the analysis of univariate data. Clayton (1979) considered frailty for the analysis of multivariate survival data. The frailty model is often described as a "random effects" model for time-to-event data. However, this model can be further categorized into two types, *shared* (random-effects) and *unshared* (overdispersion and heterogeneity).

Shared Frailty

The hazard calculated by averaging over the surviving population is termed the population hazard, which can differ from that displayed by individuals. If the study population encompasses significant heterogeneity, the population hazard can decrease with time as the risk set becomes more dense with patients who are less frail and less likely to experience the event. This phenomenon is known as the "frailty effect". In the present study each patient would have a frailty that would be shared by all the implants that he or she had placed. In the framework of the Cox model, a frailty is a latent random effect that multiplies the hazard. For the *j*th implant in the *i*th patient, the frailty model is:

$h_{ij}(t) = h_0(t)\alpha_i \exp(x_{ij}\beta)$ (19) Cleves, M.A., Gould, W.W. and Gutierrez, R.G. 2004, Revised Edition.

For $\upsilon_i = \log \alpha_i$, this model can be rewritten as $h_{ij}(t) = h_0(t)\exp(x_{ij}\beta t \upsilon_i)$, where the log frailties, υ_i are analogous to random effects in standard linear models. The estimated variance of the frailty parameters is compared to a 50:50 mixture of $\chi^2(0)$ and $\chi^2(1)$ distributions.

Andersen and Commenges (1995) derived a score test to assess association between groups of patients, after adjustment for covariate effects in a Cox proportional hazards model. This test may also be utilized for the assessment of overdispersion in (stratified and unstratified) Cox proportional hazards models.

The hazard rate for the frailty model can be written as:

$$h_{ij}(t) = h_0(t)\exp((\theta)w_i + \beta_i X_{ij}),\ i = 1 \text{ to } n \text{ and } j = 1 \text{ to } J_i \qquad (20)$$

where $h_0(t)$ is the baseline hazard for the jth implant in the ith patient, X_{ij} is the covariate vector, β is the regression coefficient vector, w_i represents the frailties, and θ is the variance of the frailty. When θ equals zero, this model becomes the proportional hazards model.

Likelihood Derivation

Therneau and Gramsch (2002) describe the estimation of θ as a maximum profile log-likelihood. The value for θ is fixed as β and r_i are estimated by maximizing the likelihood as follows:

$$L(\theta) = L_c(\beta, r_i) + \sum_{i=1}^{N} \left[\frac{1}{\theta}\{r_i - \exp(r_i)\} + \left(\frac{1}{\theta} + d_i\right)\left\{1 - \ln\left(\frac{1}{\theta} + d_i\right)\right\} - \frac{\ln\theta}{\theta} + \ln\Gamma\left(\frac{1}{\theta} + d_i\right) - \ln\Gamma\left(\frac{1}{\theta}\right)\right]$$

(21)

(Survival Analysis and Epidemiological Tables, STATA Manual release 8, 2003)

Where $L_c(\beta, r_i)$ is the traditional Cox partial likelihood, the r_i represent the coefficients of indicator variables for the patitents and d_i indicates the number of implant failures for patient i, which ranges from 1 to J_i. In this calculation each observation for the *i*th patient has a log-relative hazard represented by:

$$x_{ij}\beta + r_i \qquad (22)$$

The estimates of θ, β and r_i are those that maximize $L(\theta)$.

A variance-covariance matrix of $(\hat{\beta}, \hat{r}_i)$ is obtained from the inverse of the negative Hessian Matrix of $L(\hat{\theta})$. The variance-covariance matrix of $\hat{\beta}$ can be found as a submatrix of the variance-covariance matrix of $(\hat{\beta}, \hat{r}_i)$. Any inference based on the estimation of β is conditional on the estimation of θ.

Recent Articles of Analysis of Dental Implant Survival

Herrmann I et. al., (1999) discuss the risk of failure of implants in each patient after any one failure in the same patient. If one implant fails, will the risk of subsequent failures increase? The hypothesis evaluated was whether dependency exists among implants in the same patient/jaw. This article identified a dependency among implants that existed prior to functional loading, i.e. the risk for failure among remaining implants in the same patient/jaw increased after the first failure. The authors state that study design and statistical analysis are important when comparing success rates from various investigations, since dependency among implants in the same patient/jaw may influence success rates. Chuang et. al. (2001) compare three statistical models for survival estimation. The first model involved randomly selecting one implant per patient. The second statistical model evaluated utilized all implants, assuming independence among implants from the same subject and the third model used all implants, assuming dependence among implants from the same subject (The GEE approach was employed). These authors of this study state that the point estimates for five-year survival were similar for all three approaches. The differences in the standard error estimates were small as well. However, the authors state also that the assumption of independent observations produces statistically invalid results.A few articles address the interdependency of implants with respect to survival analysis (Mau (1993) and Haas et. al. (1996)). These authors state that independence of implants cannot be assumed in patients with multiple implants (especially when multiple implants exist in one arch) and that the total number of implants should not be used to obtain statistical results for survival analysis. The statistical method of handling dependent observations is discussed further by Haas et. al. (1996), Ivanoff et. al. (1999), Lekholm et. al. (1999), and Herman et. al. (1999). These authors recommend the random selection of one implant per patient, where the sample procedure is repeated several times, to guarantee representative results. This method is inefficient with respect to estimation because not all observations are used at the same time during sampling.

To our knowledge, no articles address the concept of frailty in the analysis of survival of dental implants. This will be the focus of this thesis.

CHAPTER 3

Methods

Creation of an Analytical Dataset

The data that were provided to me by Dr. Weyant included survival information with almost 7 years of follow-up (maximum follow-up time=2,520 days). However, these data were not in longitudinal follow-up form, and considerable data management was required to create a suitable analytic dataset for this thesis. We describe the creation of an analytic data file, data cleaning and formatting for analysis. This process is summarized in Figure 1.

Data Forms and Corresponding Files

Clinicians involved with the study were required to fill out six data forms (**Form A, Form B1, Form C, Form D, Form U, and Form X**) during their clinical evaluations of study patients.

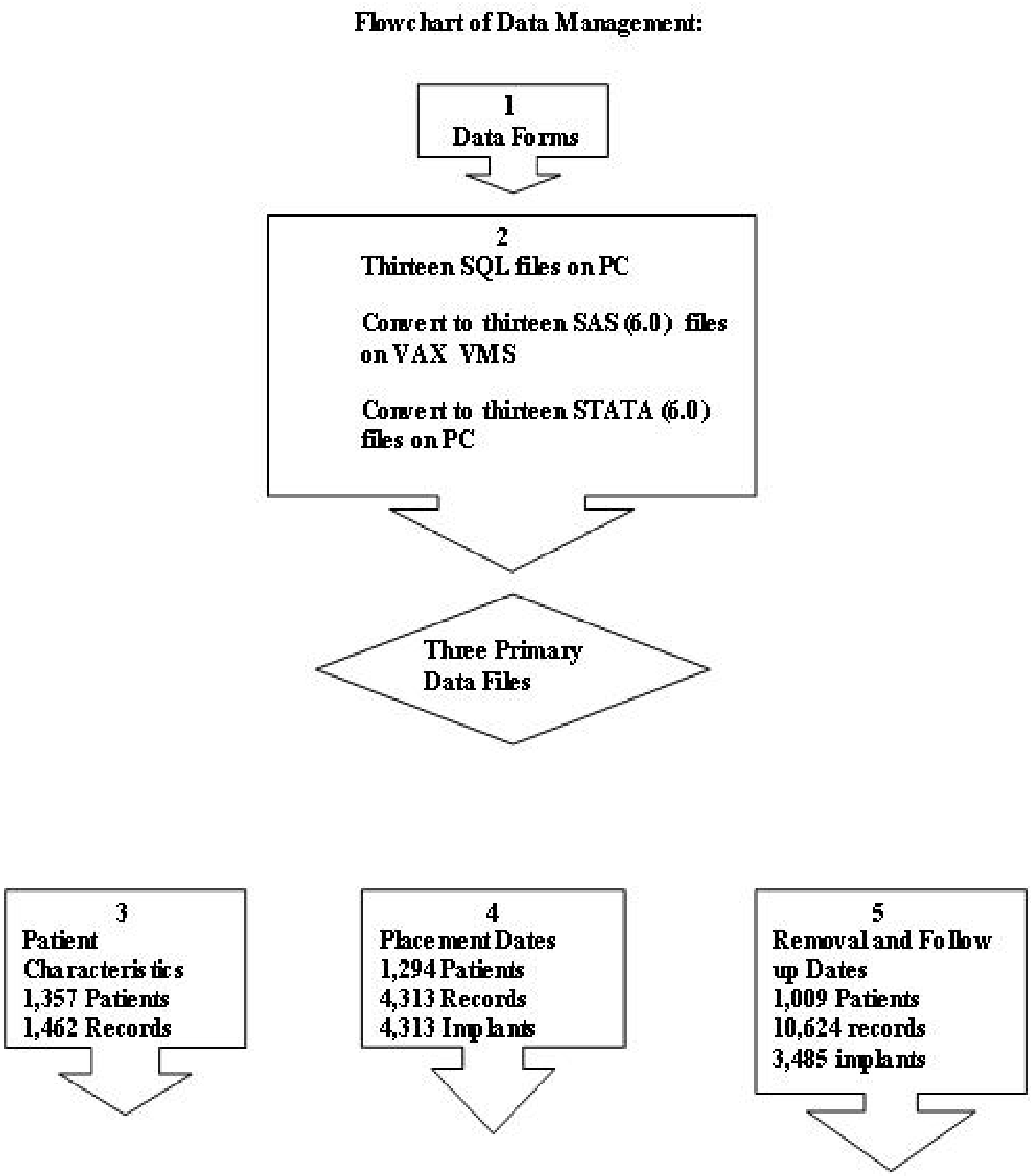

Figure 1 Flowchart of Data Management

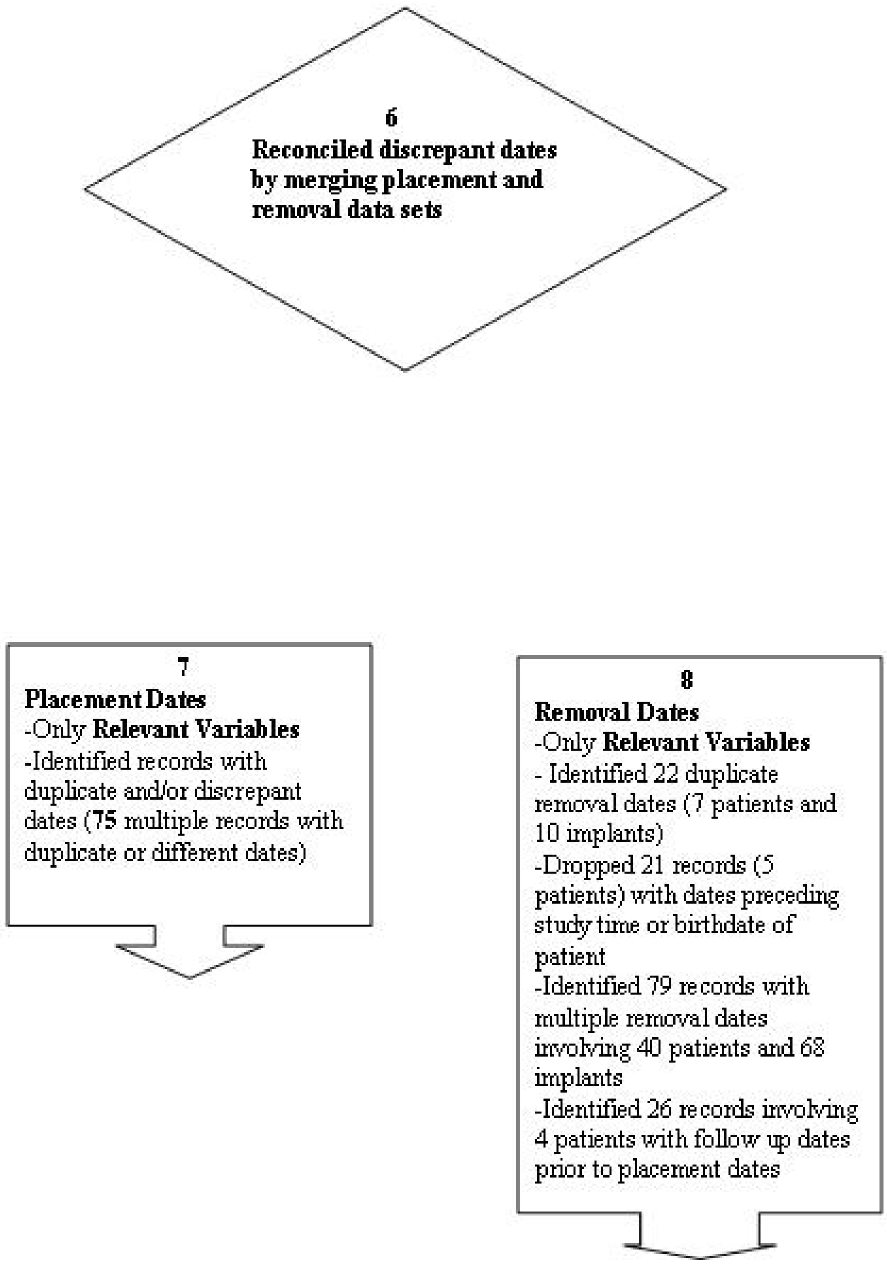

Figure 1 continued

9
Compared multiple placement and removal dates to assess clinical feasibility
-Dropped 1929 observations with placement dates and no evaluation dates (511 patients)
-Identified 15 patients with two placement dates and one removal date
-Dropped 1 patient with three duplicate placement dates and one removal date
-Dropped 4 patients with duplicate follow up dates prior to placement (26 observations)

10
Merge of edited Placement and Removal datasets with Patient Characteristics Datasets

This merge resulted in dropping 653 records that were not associated with follow up times or identifiers from the datasets that were in the merged **Placement Dates** and **Removal Dates** dataset. There was maintenance of 781 patients, 8,129 observations and 103 failures at this point.
-26 records involving 4 patients with follow up time after failure were dropped.
-117 records were dropped via manual edition due to date and record discrepancies revealed after the STSET procedure in STATA.

11
The **Analytic Dataset** had 777 patients, 7,986 records, 2,305 implants and 103 failures

Figure 1 continued

These six forms are shown in **Appendix A**. The information from these forms was entered into thirteen separate SQL files on a Personal Computer. The data were available on a VAX VMS system in the form of SAS (version 6.0) files. The data were transferred from SAS (version 6.0) to STATA (version 6.0) system files, utilizing STAT Transfer (version 5.0). The files in both systems were compared for data transfer errors. This process is summarized in the first two steps of Figure 1.

A data dictionary was not available, and the coding of some variables in the datasets did not agree with the forms. These variables were not considered further. Also, there were no variables to represent natural dentition, although it was mentioned in Dr. Weyant's thesis. I do not have the same core data that were used in Dr. Weyant's dissertation, because his data are reported to span a three year time period and the present data span 6.9 years and include a larger number of patients. The relevant information for the present analysis was obtained by merging the three datasets corresponding to Forms A, B, and C.

Three Primary Data Files

The primary data required for a survival analysis include a unique patient identifier, an implant identifier, a starting time (placement date of an implant) and follow-up time(s) (evaluation date(s) of implants), and a variable to indicate censorship or failure. The component datasets are described in **Appendix B,** and the number of patients, records, and implants in each data set is summarized in steps 3-5 of Figure 1. The number of implants placed per patient for each dataset is shown in Figure 2, which shows that many patients have more than one implant placed.

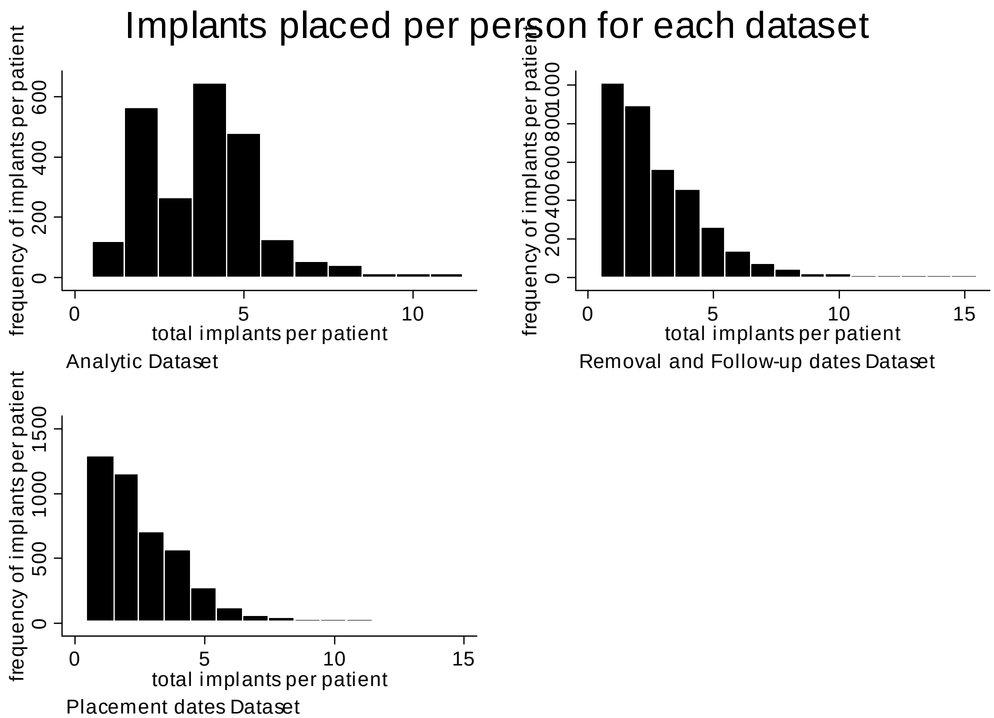

Figure 2 Implants placed in the Component Datasets

Patient Characteristics Dataset

The **Patient Characteristics** dataset contains basic patient characteristics such as each patient's unique identifier, birth date, race, and gender. **Patient Characteristics** contains no data on implants. There are more records (1,462) than patients (1,357) in the **Patient Characteristics** dataset, indicating duplicate records (**Figure 1, Step 3**). These were ultimately deleted.

Placement Dates Datasets

The **Placement Dates** dataset contains the unique patient identifier and the placement date of each implant. Other variables in this dataset include description of implants (i.e. brand, coating, length, width, etc.), bone width and height, gingival attachment measurements, and a description of opposing occlusion. There are 1,294 reported patients and 4,313 reported implants (**Figure 1, Step 4**). The greater number of implants than patients reflects the multiple placements of implants per patient as well as duplicate records per implant and multiple placements per implant-site. Indicators for records with duplicate placement dates and different placement dates were created. Records for missing placement dates were dropped.

The **Placement Dates** dataset required substantial editing. The key variables for patient identifier, implant placement date, and site of placement had to be converted from string to numeric. In STATA version 7 the "destring" command was used. Because the **site** variable was differently named than in the other datasets, a common variable name was coded. The patient identifier and placement date variables were retyped to numeric form and renamed.

Multiple placement dates at an implant site

The next step was to evaluate the multiple placement dates at some implant sites for the same patient. There appeared to be 75 records with multiple placements per implant. In the one case where a placement date was duplicated, the duplicate record was dropped. One implant had three different placement dates and the remaining implants had two different placement dates

Removal and Follow up Dates Dataset

The **Removal and Follow up Dates** dataset contains the unique patient identifier, the dates of follow-up visits and possibly removal of each implant, and other descriptive variables (**Figure 1, Step 5**). This dataset included 1009 patients, 10,624 records, and 3,485 implants. Each patient could have multiple follow-up visits for each implant. This dataset also contains duplicate records and implants with multiple removal dates. Indicators for duplicate and different placement dates were created.

The Primary issues to address with the Removal and Follow up Dates dataset:

(1) How many implant removal dates existed?

(2) Were there multiple removal dates per implant?

(3) If multiple removals existed, were these replacement dates or typographical errors?

(4) Were there implant removal dates before evaluation dates?

Indeed, some implant removal dates were found to precede some first evaluation dates. Implants without placement dates were dropped from the dataset. A variable, **ctr1**, was created to count the multiple removal dates. Another variable, **dup**, was created to count the duplicates among the multiple removal dates. A censoring variable, **rem**, was created with a dichotomous coding of 0 or 1, where 1 indicated failure and evidence of an implant removal date, and 0 otherwise. In the **Removal and Follow up Dates** dataset, a new variable, **followup**, was created for follow-up times for those implants that were either censored or failed. This was the time variable used for statistical analysis. The following STATA code was used: **g followup=nevldate, replace followup=nimrdate if rem=1**. These commands identified 43 records as containing an evaluation date and an implant removal date that differed and therefore, indexed those implants as failures. The data were sorted by patient identifier, site and followup. Date values preceding the study time or possibly the birth date of the patient were deleted; 21 observations and 5 patients were dropped from the dataset. This did not affect removal dates, of which there were 158 in this dataset.

The Number of Implant Removal Dates

An evaluation of implant removal revealed a total of 3,434 implants that were ever evaluated and not removed, and 124 implants that were “ever” removed, giving a total of 3558 implants with at least one evaluation or follow-up date. There were 1009 patients in this file.

Multiple Removal Dates

There were 10,624 records for these 1009 patients, including 22 records in which there are duplicate removal dates on the same implant. This involved 7 patients and 10 implants. There are 79 records with multiple removal dates that are different. This involved 40 patients and 68 implants.

The multiple different removal dates were compared with the placement dates to assess: (1) the potential for multiple corresponding placement dates (2) the clinical feasibility of the placement, removal and evaluation dates and (3) the potential for analysis of repeated failures per implant site in the resulting dataset.

The **Placement Dates** and **Removal and Follow up Dates** datasets were first subsetted into separate files that included only the relevant variables (patient identifier, implant site, placement date, removal date and evaluation dates) and the corresponding indicators of problematic records that were created in **Steps 6** and **7** of **Figure 1**. The abbreviated files were then merged and evaluated for date and record discrepancies (**Figure 1, Step 8**). Duplicate observations were removed and the ordering of multiple placements and removals were assessed for clinical feasibility. This cleaned dataset was remerged with the two separate source datasets.

The Process of Evaluation

The **Placement Dates** and **Removal and Follow up Dates** datasets were merged by employing the "joinby" command in STATA **(Step 9 of Figure 1)**. The "master" dataset was the abbreviated **Placement Dates** dataset, which had 4,297 records, and the "using" dataset was the abbreviated **Removal and Follow up Dates** dataset, which had 10,308 records. STATA's joinby procedure enables one to track the source of the records in the combined dataset and discern discrepancies in merging with a **_merge** variable. A tabulation of the **_merge** variable showed that 1,929 records were only in the "master" (**Placement Dates**) dataset and did not merge with the "using" (**Removal and Follow up Dates**) dataset, giving 8,379 records in the combined dataset. These 1,929 dropped records (511 patients) have only placement dates and no evaluation dates or implant removal dates. The total number of unique patient identifiers in the combined dataset is now 781.

A new variable, **place**, represents the implant placement date for each implant. A variable named **failure** was created to denote censoring or failure of each implant at any specific follow-up time. There were 15 verified patients who had two separate implant placement dates and only one removal date. The decision was made to choose the first, or earliest, implant placement date for the analysis. This decision was based on the clinical feasibility of the timing of the placement dates. There was one case that had three duplicate placement dates and only one removal date. We decided to manually edit the data for circumstances involving:

(1) Multiple placement dates and no removal dates,

(2) Multiple placement dates and one removal date

(3) Multiple placement dates and multiple removal dates,

(4) One unique placement date and multiple removal dates, and

(5) Duplicate placement dates and/or duplicate removal dates.

The **followup** variable was evaluated and the records were sorted to have the last possible evaluation date listed with the **failure** variable changed to a 0 or 1 to denote censoring or removal, respectively. Duplicate follow-up, placement and removal times were deleted. There were four patients who had evaluation dates or follow up visits prior to the date of placement. If this occurred on the first visit, perhaps as an evaluation before placement of an implant, this could be somewhat reasonable. However, there are several visits (26) for evaluations of implants with no placement records; there were no removals for such observations. These records were removed from the analytic file.

There were situations where an implant was removed and follow up times were present after removal, for the same implant. One question that clinicians may pose is; “Why was a site evaluated several times after removal”? This may have been an oversight on the part of anyone evaluating the patient, he or she may have been evaluating other sites, or a subsequent placement date could be missing. However, in evaluating sites we are assuming that there is an implant to evaluate. If an implant is not present, implant failure cannot be assessed. These observations were left until the demographic variables in the **Patient Characteristics** dataset were merged with this dataset and then dropped.

Creating the Analytic Dataset

The analytic dataset was formed by using the **joinby** command with the merged abbreviated dataset and the three original datasets separately in order to collect all the variables (**Figure 1, Step 10**). In merging the subsetted datasets, it became apparent that there were site indicators that differed in both datasets and had to be renamed in one. Careful inspection of all the variables with respect to type is required with all data merging, procedures and especially so here. Merging can be unsuccessful if variables are typed or named differently in the component datasets.

The **Patient Characteristics** dataset was merged with the combined cleaned dataset (**Placement Dates/ Removal and Follow up Dates**). After merging there were 95 records that did not have follow-up times corresponding with the **Removal and Follow up Dates** dataset and did not have placement date information in the **Placement Dates** dataset. There were more observations in the **Patient Characteristics** dataset than in the combined dataset. This could be attributed to unmatching patient identifiers. A total of 653 records were lost in the merge because of matching problems. Records in which there were no demographic data were kept because survival data exist; at this point there were 781 patients, 8,129 records and 103 records with failures. The STSET procedure sets the data for survival analysis in STATA with the appropriate unique identifier and time variables. This procedure also identifies date and record errors pertinent to survival analysis. Twenty-six records involving 4 patients were identified with follow up times occurring after a failure. These observations were dropped with manual editing. After manually editing for date and record discrepancies that were revealed in the STSET procedure in STATA, a total of 777 patients, 7,986 records, 2,305 implants and 103 failures were maintained in the final analytic dataset (**Step 11, Figure 1**)

Statistical Models

We will illustrate the following models using implant-level predictors (type and location) and patient-level predictors (race/ethnicity).

1. Logistic regression of first implant per patient with first year follow-up
2. Logistic regression of multiple implants per patient with first year follow-up
3. Logistic regression of multiple implants per patient first year followup using Generalized Estimating Equations (GEE)
4. Discrete Proportional Odds of first implant per patient with multiple time intervals
5. Discrete Proportional Hazards of first implant per patient with multiple time intervals
6. Discrete Proportional Odds of multiple implants per patient with multiple time intervals
7. Discrete Proportional Odds of multiple implants per patient with multiple time intervals using (GEE)
8. Discrete Proportional Hazards of multiple implants per patient with multiple time intervals using (GEE)
9. Continuous-time Cox Model of single implant per patient over time
10. Continuous-time Cox Model of multiple implants per patient over time
11. Continuous-time Shared Frailty Model of multiple implants per patient over time

Table 1 describes the statistical models evaluated for the various data situations involved which corresponds with the preceding list.

Table 1 Statistical Models Evaluated

1	Logistic regression	$\log it(p_i) = \beta_0 + \beta_1(loc2)_i + \beta_2(loc3)_i + \beta_3(loc4)_i + \beta_4(type2)_i + \beta_5(type2)_i$	Single site per person and single time interval
2,3	Logistic regression, Generalized Estimating Equations (GEE)	$\log it(p_{ijt}) = \beta_0 + \alpha_t + \beta_1(loc2)_{ij} + \beta_2(loc3)_{ij} + \beta_3(loc4)_{ij} + \beta_4(race2)_i + \beta_5(type2)_j$	Multiple sites per person and single time interval
4,5	Logistic regression with Discrete proportional odds and hazards	$\log it(p_{it}) = \beta_0 + \alpha_t + \beta_1(loc2) + \beta_2(loc3) + \beta_3(loc4) + \beta_4(race2) + \beta_5(type2)$ $c - \log - \log(p_{it}) = \beta_0 + \alpha_t + \beta_1(loc2) + \beta_2(loc3) + \beta_3(loc4) + \beta_4(race2) + \beta_5(type2)$	Single site per person with multiple time intervals
6,7, 8	Discrete Proportional Odds and hazards and (GEE) model	$\log it(p_{ijt}) = \beta_0 + \alpha_t + \beta_1(loc2)_{jt} + \beta_2(loc3)_{jt} + \beta_3(loc4)_{jt} + \beta_4(race2)_i + \beta_5(type2)_{jt}$ $c - \log - \log(p_{ijt}) = \beta_0 + \alpha_t + \beta_1(loc\ 2)_{jt} + \beta_2(loc\ 3)_{jt} + \beta_3(loc\ 4)_{jt} + \beta_4(race\ 2)_i + \beta_5(type\ 2)_{jt}$	Multiple sites per person and multiple time intervals
9	Continuous-time Survival (Cox model)	$h(t; x_i) = h_0(t)e^{(\beta_1(loc2)_{ij} + \beta_2(loc3)_{ij} + \beta_3(loc4)_{ij} + \beta_4(race2)_i + \beta_5(type2)_{ij})}$	Single site per person and continuous time
10	Continuous-time Survival (Cox model)	$\lambda(t; x_{ij}) = \lambda_0(t)e^{(\beta_1(loc2)_{ij} + \beta_2(loc3)_{ij} + \beta_3(loc4)_{ij} + \beta_4(race2)_j + \beta_5(type2)_{ij})}$	Multiple sites per person and continuous time
11	Shared Frailty Model	$\lambda(t; x_{ij}) = \lambda_0(t)e^{(\beta_1(loc\ 2)_{ij} + \beta_2(loc\ 3)_{ij} + \beta_3(loc\ 4)_{ij} + \beta_4(race\ 2)_j + \beta_5(type\ 2)_{ij})}$	Multiple sites per person and continuous time

Coding of predictor variables

The models discussed in this thesis incorporate both between and within patient variables (oral location, implant type and race of patient). We also considered differences between model-based and robust variance estimates.

The variable for oral location was created by the site variable and is categorized as follows: Iloc_1=mandibular anterior, Iloc_2=maxillary anterior, Iloc_3=mandibular posterior, and Iloc_4=maxillary posterior. Therefore, the mandibular anterior region was considered the baseline value.

The variables of implant type and race were categorized as well. Implant type started out with 7 unique values. Due to the small numbers in all but type 19 and 4, the variable was collapsed to two categories. Type 19 was the baseline or (Itype_1) and all other types were grouped into Itype_2. The same was done for race where the baseline race was white (Irace_1) and non-white was Irace_2.

CHAPTER 4

Descriptive Analysis

Patient Demographic Characteristics

Patient demographic characteristics are summarized in Table 2. The high percentage of males reflects a typical Veterans Administration population. Both gender and race/ethnicity were missing for 36 patients (278 records) and ethnicity was unknown for 9 patients. The mean age of the patient population is 62 years, with a minimum and maximum of 25 and 82 years, respectively. A variable was created to assess the possibility of multiple recordings of gender and race/ethnicity across visits. There was one such patient, who had records specifying Hispanic and White, and this patient was counted as Hispanic. The majority of these patients were white (80.8%).

Table 2 Patient Demographic Characteristics

(n=777)

	Characteristic	Frequency	Percent
Gender	Male	710	95.8
	Female	20	2.7
	Unknown/ Missing	47	3.5
Race/ Ethnicity	White	628	80.8
	Black	77	9.9
	Asian	1	0.1
	Native Am.	2	0.3
	Hispanic	22	2.8
	Other	2	0.3
	Missing	45	5.8
Those not recording any ethnic value (i.e. only 0's)			
Total number of Patients		777	100.0

Implant Characteristics

As can be seen in this Table 3, 36.3% of the patients have two implants and 20.7% have four implants. A single implant is the third most frequent situation, occurring in 14.9% of patients. In this dataset, 85.1% of patients have multiple implants.

Table 3 Distribution of Number of Implants: Overall and By Patient Frequencies and Percents

Number Of Implants	Implants (k=2305)		Patients (n=777)	
	Freq	Percent	Freq	Percent
1	116	5.03	116	14.9
2	564	24.47	282	36.3
3	261	11.32	87	11.2
4	644	27.94	161	20.7
5	475	20.61	95	12.2
6	126	5.47	21	2.7
7	49	2.13	7	0.9
8	40	1.87	5	0.6
9	9	0.39	1	0.1
10	10	0.43	1	0.1
11	11	0.48	1	0.1

Although seven implant types are reported in this dataset, only one, Type 19, was used for the vast majority (725) of patients (Table 4). A total of 45 (5.8%) of patients received implants of Type 4, and very few patients received the other implant types. The six patients who received more than one type of implant are listed more than once in column 2 of Table 3. A total of 94% of the implants were of Type 19 and 4% were of Type 4. Sixty-four patients experienced at least one implant failure. A total of 103 failures were observed out of the 2,305 implants placed. The "Within" Percent value indicates that the patients who have received implant Type 19 received this implant type 99.5% of the time, while patients who received Type 4 received this type 85.6% of the time.

Table 4 Numbers of Patients, Implants by Type of Implant and Implant Failures

Type of Implant	Number of Patients	Number of implants	Number of failures	Failure rate	Number of patients with failures	Within Percent
2	5	13	0	0.00	0	100.0
4	45	95	6	0.06	3	85.6
5	1	4	0	0.00	0	100.0
6	2	6	4	0.70	1	60.0
10	1	2	0	0.00	0	33.3
18	4	16	0	0.00	0	94.1
19	725	2169	93	0.04	60	99.5
Total	783*	2305	103	0.04	64	98.4

*Note: Six (6) patients had more than one type of implant

Figure 3 displays the frequency of implants placed per patient in the analytic dataset. As shown, there are many opportunities to evaluate multiple failures per patient with most patients having more than one implant.

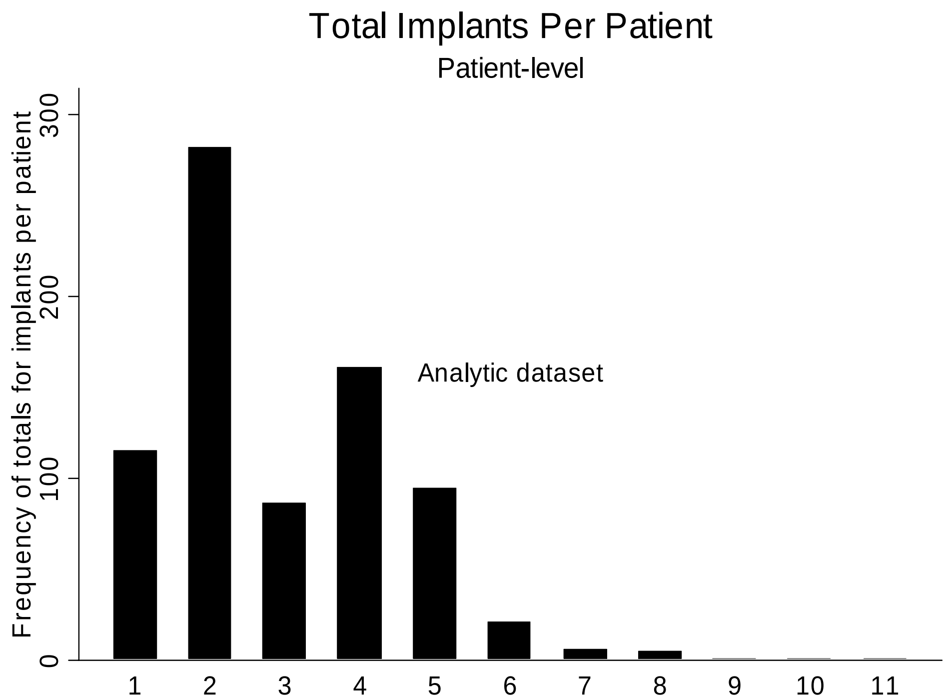

Figure 3 Total Implants per Patient in the Analytic Dataset

Figure 4 displays the frequency of implants placed by site per patient. The two histograms separate the frequencies by dental arch ((maxillary-upper jaw) vs. (mandibular-lower jaw)). For both arches the higher frequencies occur in the canine regions, where the bone density may be greater.

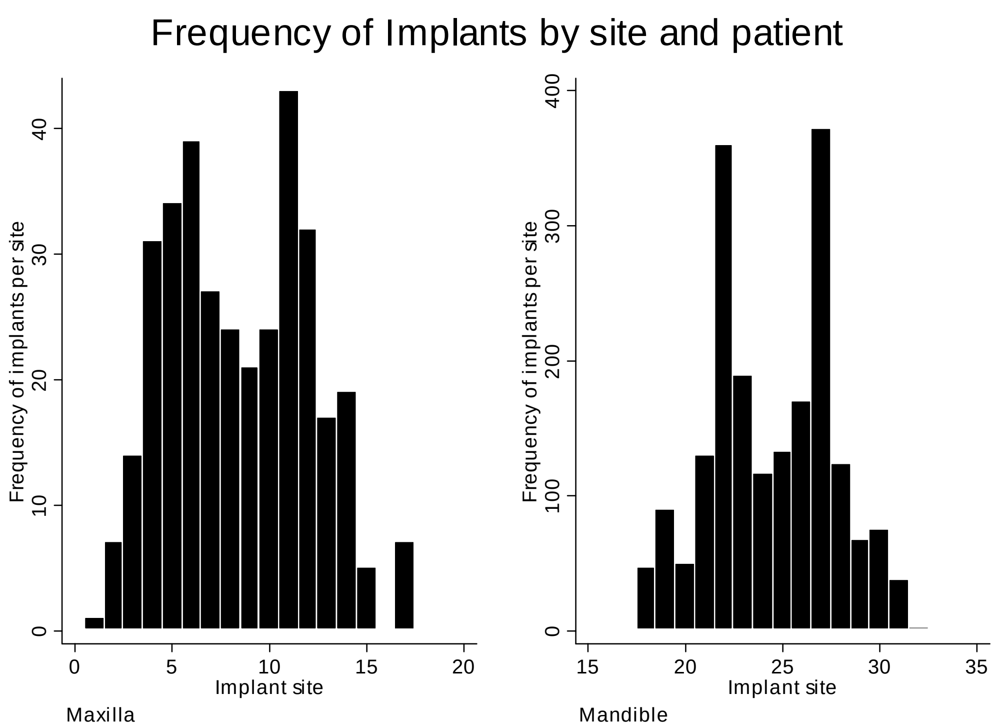

Figure 4 Frequency of Implants placed by site and patient

Table 5 shows the frequencies of first, second and subsequent visits. One patient had 25 visits.

Table 5 Distribution of follow-up visits

Number of Visits	Frequency of Visits	Percent of Visits
1	777	100.0
2	548	70.5
3	394	50.7
4	279	35.9
5	196	25.2
6	146	18.8
7	105	13.5
8	80	10.3
9	63	8.1
10	44	5.7
11	37	4.8
12	25	3.2
13	19	2.5
14	12	1.5
15	11	1.4
16	8	1.0
17	8	1.0
18	5	0.6
19	5	0.6
20	5	0.6
21	4	0.5
22	4	0.5
23	4	0.5
24	1	0.1
25	1	0.1
Total	2,781	357.9

It should be noted that this table presents the frequencies and percents at the patient-level where each patient could be included in several visit categories. A patient who had seventeen visits also was included in the count for one through sixteen visits. Therefore, although there are 777 patients in the study, each patient may be counted more than once for each visit that they participated in. This follows for the "percent" values. The visit frequencies and percents are evaluated at each visit.

CHAPTER 5

Modeling Results

Table 6 summarizes Logistic regression analysis of the first implant per patient with one year of follow-up (Model 1). There is a nonsignificant elevation of the odds ratios for the maxillary anterior and posterior regions relative to the mandibular anterior region. Implant types other than type 19 have a nonsignificantly decreased odds of failure. Non-whites have nonsignificantly increased odds of failure relative to whites. The model-based and robust standard errors were virtually identical.

Table 6 Model 1 Results: Logistic regression of first implant with one year follow-up

Number of observations: 872

Predictor	Estimated Odds Ratio	Model-based Standard error	P-value	Robust Standard error	P-value
Oral Location					
Mandibular anterior	1.0	—		—	
Maxillary anterior	1.52	1.21	0.60	1.20	0.60
Mandibular posterior	0.64	0.34	0.40	0.33	0.40
Maxillary posterior	2.20	1.50	0.26	1.50	0.25
Type					
Type 19	1.0	-		-	
Others	0.71	0.55	0.66	0.55	0.66
Race					
White	1.0	-		-	
Others	1.88	1.00	0.24	1.01	0.24

Table 7 summarizes the Logistic regression analysis of multiple implants per patient with one year of follow-up (Model 2). There is a nonsignificant increase in the odds of failure for the maxillary anterior and mandibular posterior regions relative to the mandibular anterior region. The maxillary posterior region had a significantly elevated odds of implant failure based on a model-based standard error (p=0.04), but this odds ratio of 2.86 is not statistically significantly elevated when a robust standard error was used. The robust standard error accounts for the multiple implants per patient.

Table 7 Model 2 Results: Logistic regression of multiple implants with one year follow up
Number of observations: 2610

Predictor	**Estimated Odds Ratio**	**Model-based Standard error**	**P-value**	**Robust Standard error**	**P-value**
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	2.15	0.99	0.10	1.13	0.15
Mandibular posterior	1.12	0.37	0.74	0.37	0.73
Maxillary posterior	2.86	1.43	0.04	1.89	0.11
Type					
Type 19	1.0	-			
Others	0.85	0.46	0.77	0.64	0.84
Race					
White	1.0	-			
Others	1.97	0.67	0.05	0.95	0.16

Table 8 summarizes the a GEE (Logistic regression) analysis of multiple implants per patient over the first year of follow-up (Model 3) and an assumed exchangeable correlation structure. The highest odds of failure is observed for the maxillary anterior and posterior regions. The odds ratio was somewhat elevated for the mandibular posterior region. However, no region was statistically significant.

Non-whites have a significantly elevated odds of failure based on the model-based standard error (p=0.04), but this odds ratio of 2.18 is not statistically significant when a robust standard error was used.

Table 8 Model 3 Results: GEE (Logistic Regression), Multiple Implants, First year followup

Number of observations: 2610

Predictor	**Estimated Odds Ratio**	**Model-based Standard error**	**P-value**	**Robust Standard error**	**P-value**
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	1.83	0.99	0.26	1.08	0.31
Mandibular posterior	1.26	0.41	0.48	0.34	0.40
Maxillary posterior	2.62	1.43	0.08	1.69	0.14
Type					
Type 19	1.0	-			
Others	0.80	0.51	0.72	0.64	0.78
Race					
White	1.0	-			
Others	2.18	0.84	0.04	1.06	0.11

Table 9 summarizes the discrete proportional odds model analysis for the first implant per patient with multiple time intervals of follow-up (Model 4). We see a significantly decreased odds of failure in year 2 relative to year 1, with a nonsignificant decrease in the odds of implant failure in subsequent years until year 8. In year 8, four implants failed among the 16 patients still at risk.

The maxillary anterior and posterior regions had elevated odds ratios (p=0.001 and 0.07 respectively) relative to the mandibular anterior region. The odds of failure for the two mandibular regions were similar. The model-based and robust standard errors are virtually identical.

Table 9 Model 4 Results for Discrete Proportional Odds. First implant per patient with multiple time intervals

Number of Observations: 3651

Predictor	**Estimated Odds Ratio**	**Model-based Standard error**	**P-value**	**Robust Standard error**	**P-value**
Year					
Year 1	1.0	-		-	
Year 2	0.37	0.17	0.03	0.17	0.03
Year 3	0.57	0.28	0.26	0.28	0.25
Year 4	0.20	0.20	0.11	0.20	0.11
Year 5	0.37	0.38	0.33	0.38	0.33
Year 6	0.82	0.85	0.85	0.85	0.85
Year 7	-	-	-	-	-
Year 8	12.7	14.4	0.02	13.7	0.02
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	5.1	2.46	0.001	2.51	0.001
Mandibular posterior	0.95	0.41	0.90	0.40	0.90
Maxillary posterior	2.6	1.41	0.07	1.40	0.07
Type					
Type 19	1.0	-		-	
Others	1.4	0.78	0.60	0.71	0.55
Race					
White	1.0	-		-	
Others	1.3	0.58	0.60	0.60	0.61

Table 10 summarizes the discrete proportional hazards model analysis for the first implant per patient with multiple time intervals for follow-up using the Clog-log function (Model 5). Numerically, these estimates are very similar to the comparable discrete proportional odds model shown in Table 9 (Model 4). However, these parameter estimates are hazard ratios rather than odds ratios.

Table 10 Model 5 Results for Discrete Proportional Hazards using the Cloglog function First implant per patient with multiple time intervals

Number of Observations: 3651

Predictor	**Estimated Hazard Ratio**	**Model-based Standard error**	**P-value**	**Robust Standard error**	**P-value**
Year					
Year 1	1.0	-		-	
Year 2	0.37	0.17	0.03	0.18	0.04
Year 3	0.57	0.28	0.26	0.28	0.26
Year 4	0.19	0.20	0.11	0.20	0.11
Year 5	0.37	0.38	0.33	0.38	0.33
Year 6	0.87	0.85	0.85	0.86	0.86
Year 7	-	-	-	-	-
Year 8	11.89	12.46	0.02	11.91	0.01
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	4.98	2.38	0.001	2.44	0.001
Mandibular posterior	0.96	0.41	0.92	0.40	0.92
Maxillary posterior	2.62	1.40	0.07	1.40	0.07
Type					
Type 19	1.0	-		-	
Others	1.38	0.78	0.58	0.71	0.54
Race					
White	1.0	-		-	
Others	1.27	0.57	0.60	0.59	0.61

Table 11 summarizes the discrete proportional odds model analysis for multiple implants per patient with multiple time intervals of follow-up (Model 6). The robust standard errors are consistently larger than the corresponding model-based values. In several instances (year 2, year 4, mandibular posterior, and non-white race) the parameter estimates change from significant to nonsignificant when the robust standard errors are used.

Table 11 Model 6 Results for Discrete Proportional odds
Multiple implants per patient over time
Number of observations: 11,217

Predictor	Estimated Odds Ratio	Model-based Standard error	P-value	Robust Standard error	P-value
Year					
Year 1	1.0	-			
Year 2	0.55	0.14	0.02	0.21	0.12
Year 3	0.42	0.15	0.02	0.17	0.03
Year 4	0.33	0.17	0.03	0.23	0.09
Year 5	0.59	0.31	0.32	0.48	0.52
Year 6	0.94	0.57	0.93	0.73	0.95
Year 7	0.72	0.73	0.75	0.75	0.75
Year 8	27.73	17.13	0.000	29.74	0.002
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	4.70	1.42	0.000	1.89	0.000
Mandibular posterior	1.40	0.35	0.20	0.35	0.17
Maxillary posterior	3.39	1.13	0.000	1.35	0.002
Type					
Type 19	1.0	-		-	
Others	1.42	0.52	0.33	0.64	0.42
Race					
White	1.0	-		-	
Others	1.79	0.45	0.02	0.72	0.15

Table 12 summarizes the discrete proportional odds model for multiple implants per patient with multiple time intervals for follow-up using a GEE analysis (Model 7) and an assumed exchangeable correlation structure. After the first year the odds ratios appear to be less than one until year six. The odds ratios for the maxillary anterior and posterior regions are significantly elevated relative to the mandibular anterior region when the model-based or robust standard errors are employed. The type of implant does not appear to be significant in this model. Non-white race appears to be significantly associated with increased failure using the model-based standard error but not the robust standard error.

Table 12 Model 7 Results for Discrete Proportional odds

Multiple implants per patient over time with GEE analysis

Number of observations: 11,217

Predictor	**Estimated Odds Ratio**	**Model-based Standard Error**	**P-value**	**Robust Standard error**	**P-value**
Year					
Year 1	1.0	-		-	
Year 2	0.70	0.16	0.12	0.22	0.25
Year 3	0.63	0.19	0.13	0.19	0.12
Year 4	0.55	0.22	0.15	0.23	0.16
Year 5	0.81	0.37	0.64	0.42	0.69
Year 6	1.36	0.69	0.54	0.76	0.58
Year 7	1.20	0.95	0.81	0.73	0.75
Year 8	28.67	17.4	0.000	29.89	0.001
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	4.06	1.35	0.000	1.41	0.000
Mandibular posterior	1.38	0.34	0.19	0.28	0.11
Maxillary posterior	3.08	1.12	0.002	1.17	0.003
Type					
Type 19	1.0	-		-	
Others	1.35	0.57	0.48	0.63	0.53
Race					
White	1.0	-		-	
Others	1.93	0.56	0.02	0.78	0.10

Table 13 summarizes the a discrete proportional hazards model for multiple implants per patient with multiple time intervals of follow-up with the Clog-log link using GEE (Model 8) and an assumed exchangeable correlation structure. The estimates are similar numerically to those in the comparable discrete proportional odds analysis using GEE in Table 12.

Table 13 Model 8 Results for Discrete Proportional hazards using C-log-log and GEE analysis

Number of observations: 11,217

Predictor	Estimated Hazards Ratio	Model-based Standard Error	P-value	Robust Standard error	P-value
Year					
Year 1	1.0	-		-	
Year 2	0.70	0.16	0.13	0.22	0.26
Year 3	0.63	0.19	0.13	0.19	0.12
Year 4	0.55	0.23	0.15	0.23	0.16
Year 5	0.81	0.37	0.64	0.42	0.68
Year 6	1.36	0.68	0.54	0.75	0.58
Year 7	1.20	0.93	0.82	0.72	0.76
Year 8	24.07	12.74	0.000	21.38	0.000
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	4.03	1.32	0.000	1.39	0.000
Mandibular posterior	1.40	0.34	0.17	0.28	0.10
Maxillary posterior	3.08	1.11	0.002	1.16	0.003
Type					
Type 19	1.0	-		-	
Others	1.43	0.58	0.38	0.63	0.42
Race					
White	1.0	-		-	
Others	1.87	0.54	0.03	0.75	0.12

Table 14 summarizes the continuous-time proportional Cox model analysis for the first implant per patient over time (Model 9). The hazard ratios are significantly elevated for the maxillary anterior region and nonsignificantly elevated for the maxillary posterior region, both relative to the mandibular anterior region. The model-based and robust standard errors are virtually identical.

Table 14 Model 9 Results for Continuous-time Cox Model
Single implant per patient over time

Number of observations: 2,483

Predictor	Estimated Hazard Ratio	Model-based Standard error	P-value	Robust Standard error	P-value
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	4.04	1.92	0.003	1.92	0.003
Mandibular posterior	0.88	0.37	0.76	0.36	0.75
Maxillary posterior	2.05	1.09	0.18	1.09	0.18
Type					
Type 19	1.0	-		-	
Others	1.58	0.92	0.43	0.80	0.36
Race					
White	1.0	-		-	
Others	1.31	0.59	0.55	0.60	0.55

Table 15 summarizes the continuous-time proportional Cox model analysis for the multiple implants per patient over time. The hazard ratios for the maxillary regions are significantly elevated when either the model-based or robust standard errors are used. The odds ratio of 1.80 associated with non-white race is significant using the model-based standard error and not significant using the robust standard error estimator.

Table 15 Model 10 Results for Continuous-time Cox Model
Multiple implants per patient over time

Number of observations: 7,633

Predictor	**Estimated Hazard Ratio**	**Model-based Standard error**	**P-value**	**Robust Standard error**	**P-value**
Oral Location					
Mandibular anterior	1.0	-		-	
Maxillary anterior	3.85	1.14	0.000	1.55	0.001
Mandibular posterior	1.29	0.32	0.30	0.31	0.28
Maxillary posterior	2.64	0.87	0.003	1.03	0.01
Type					
Type 19	1.0	-		-	
Others	1.54	0.56	0.23	0.67	0.32
Race					
White	1.0	-		-	
Others	1.80	0.45	0.02	0.70	0.13

Table 16 summarizes the continuous-time shared frailty model for multiple implants per patient over time. The maxillary arch in both regions shows significantly elevated hazard ratios relative to the mandibular anterior region. The non-white race level presents a highly elevated hazard ratio, which is significant (p=0.002). The frailty estimate for this model is highly significant (Chibar=0.000) which means that there is significant unobserved patient-level frailty.

Table 16 Model 11 Results: Continuous time shared frailty model for multiple implants per patient over time

Number of observations: 7,633
Number of groups: 732

Predictor	**Estimated Hazard Ratio**	**Standard error**	**P-value**
Oral Location			
Mandibular anterior	1.0	—	
Maxillary anterior	5.76	4.23	0.02
Mandibular posterior	1.51	0.54	0.24
Maxillary posterior	5.82	4.59	0.03
Type			
Type 19	1.0	-	
Others	1.16	1.12	0.88
Race			
White	1.0	-	
Others	17.74	16.09	0.002

Likelihood-ratio test of θ=0: $\chi^2(0.1) = 225.1$ **and p=0.000**

The year-specific numbers of implant failures, implants, and proportion of failures are summarized in Table 17 by Intraoral region and in Table 18 by Type of Implant. The year- and race- specific distributions are shown in Table 19.

Table 17 Implant Failure rates by Intraoral Region and Year

Intraoral location	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total failures
Mandibular Anterior region									
r	28	5	6	1	3	2	1	2	48
n	1,339	974	623	415	264	143	54	11	
p	0.021	0.0051	0.010	0.0024	0.011	0.014	0.019	0.18	
Maxillary Anterior region									
r	6	8	1	1	0	0	0	0	16
n	178	109	61	23	9	0	0	0	
p	0.004	0.07	0.020	0.043	0.00	0.00	0.00	0.00	
Mandibular Posterior region									
r	15	9	3	0	1	0	0	2	27
n	628	424	74	122	63	33	15	4	
p	0.024	0.021	0.041	0.00	0.016	0.00	0.00	0.50	
Maxillary Posterior region									
r	5	1	3	2	0	1	0	0	12
n	160	123	74	26	9	5	0	0	
p	0.031	0.0081	0.041	0.077	0.00	0.20	0.00	0.00	
Total failures									
r	54	23	10	4	4	3	1	4	103
n	2,305	1630	982	586	345	181	69	15	
p	0.023	0.014	0.010	0.0068	0.012	0.017	0.014	0.27	

r=number of implant failures, n=number of implants
p=proportion of failures=number of failures/n

Table 18 Implant Failure Rates by Type and Year

Implant Type	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total failures
Type 19									
r	50	23	10	4	4	1	1	0	93
n	2,169	154	928	547	314	154	55	7	
p	0.023	0.15	0.011	0.0073	0.013	0.013	0.018	0.00	
Other than Type 19									
r	4	0	0	0	0	2	0	4	10
n	136	89	53	39	31	27	14	8	
p	0.029	0.00	0.00	0.00	0.00	0.074	0.00	0.50	
Total failures									
r	54	23	10	4	4	3	1	4	103
n	2,305	1,630	982	586	345	181	69	15	
p	0.023	0.014	0.010	0.0068	0.012	0.017	0.014	0.27	

r=number of implant failures, n=number of implants
p=proportion of failures=number of failures/n

Table 19 Implant Failure Rates by Race and Year

Race	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7	Year 8	Total failures
White									
r	42	15	8	4	4	3	1	4	81
n	1,891	1,332	811	489	291	148	48	12	
p	0.022	0.011	0.010	0.0082	0.014	0.020	0.021	0.33	
Non-White									
r	12	8	1	0	0	0	0	0	21
n	283	201	98	67	40	25	19	3	
p	0.042	0.040	0.010	0.00	0.00	0.00	0.00	0.00	
Total failures									
r	54	23	9	4	4	3	1	4	* 102
n	2,174	1,533	909	556	331	173	67	15	
p	0.025	0.015	0.01	0.0072	0.012	0.017	0.015	0.27	

r=number of implant failures, n=number of implants
p=proportion of failures=number of failures/n
***Note: The discrepancy in one failure was due to the Hispanic ethnicity variable being labeled as white and non-white.**

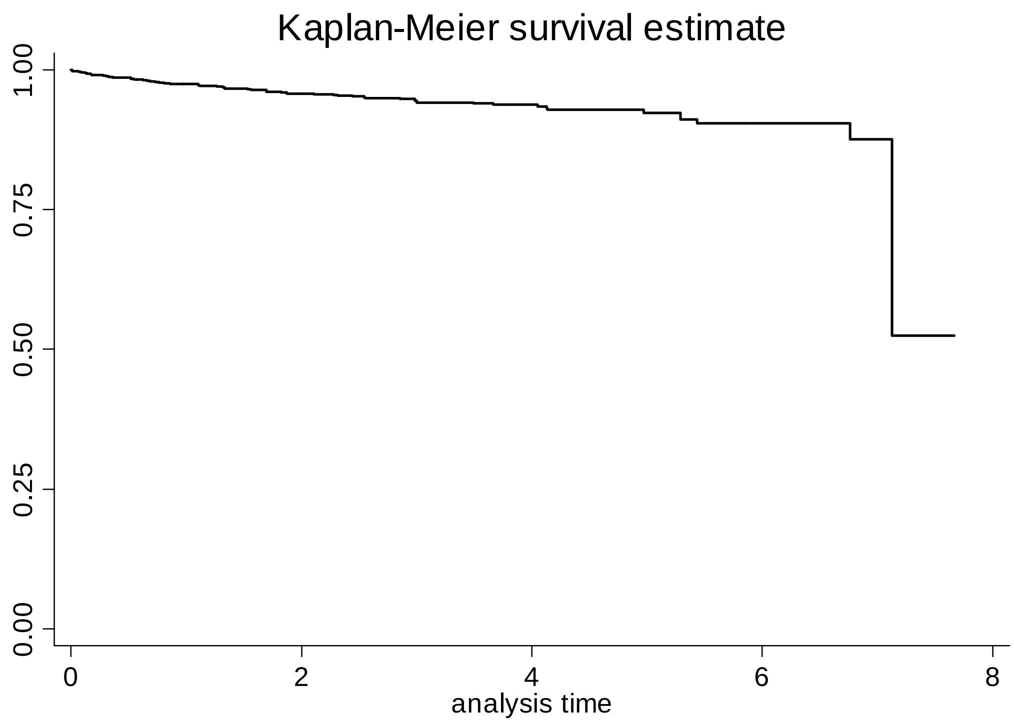

Figure 5 Kaplan-Meier Estimate

Figure 5 displays the Kaplan-Meier estimate of implant survival. The survival estimates slowly but steadily decrease over the first four to six years. The extreme drop occurs at year 7, when the risk set is very small. This estimate does not account for clustering of the data.

CHAPTER 6

Discussion

Longitudinal data can be analyzed by various methods. As demonstrated in this thesis, each method requires distinct formatting and therefore knowledge of the data. The Weyant data came in several independent datasets that required thorough evaluation and cleaning prior to linking. Our analysis occurred years after data collection. Ideally, planning of the study would consider the data analysis prior to data collection, so that data collection could be managed more efficiently. These data include many variables that would have been interesting to analyze with respect to implant failure. However, a variety of data discrepancies precluded such analyses. Typically age would be included as a basic univariate descriptive of the population being studied. However, there are 5,999 missing values for age out of 7,986. Other variables, such as descriptives of implants or the surrounding periodontium, were also often missing. However, the basic variables required for survival analysis exist, and the strengths of the analysis include a large number of patients with long follow-up time. The statistical assumption of random censorship was made throughout, i.e, that the probability of loss to follow-up is unrelated to the probability of failure. This assumption may be suspect if sicker patients may not return for follow-up visits.

The logistic regression analyses have the limitation of only allowing a view of survival probability over the entire study period as a single time interval. There is an assumption that patients are at risk over the entire study period. This may not be true for all patients and variable time at risk is not addressed. **Models 1** and **2** indicate that the failure risk is not significantly influenced by any of the variables in the model. However, the discrete analyses, **Models 4, 5, 6,**

7, and 8, permit a view of failure risk over time. The decreased odds ratio during the third through fifth years may be an indicator for the osseointegration process within the first two years of placement. If the implant does not fail within the first two years, the expectation of survival thereafter may be higher. The higher odds ratios in the last year are to be considered with caution because of the low number of patients remaining at that time. The plateau of failures during the three-five year period can be likened to a "frailty effect" where the implants (or clusters of implants) with higher frailty will not be in the risk set after the first two years. However, the more robust or stable implants still will be at risk after the first two years. The shared frailty analysis is computationally intensive because this is an iterative process involving several iterations per cluster. Each of the 777 patients is a cluster with a frailty value that is shared by all implants within a cluster. A STATA statistician suggested the robust variance option and clustering on the patient as an alternative to the shared frailty procedure. This option is much less computer-intensive. The shared frailty model (**Model 11**) displayed a significant frailty effect. Also hazard ratios for the maxillary anterior regions and non-white race were significantly elevated. This was consistent with the analysis for **Model 10** (the continuous-time Cox model for multiple implants per patient). This frailty model (which is comparable to a random-effects model) indicates that there is an unobserved patient-level effect that influences the hazard ratio.

An interesting finding with all models is that an implant placed in the maxillary arch is at greater risk of failure than an implant placed in the mandible. This is witnessed clinically. Some attributing factors involved may include the difference in bone integrity and vascularity between the arches. The proximity of the maxillary sinuses in the posterior regions can present more infection, which is a potential influence on implant failure.

Implants placed in patients of non-white race appear to be at greater risk of failure than those placed in white patients. However, when the robust variance is used, the significance of the difference between the race levels disappears. It is important to address the difference between patient-level and implant-level variables with respect to the different results obtained in our models. Race is a patient-level variable. The robust variance calculated at the patient-level accounts for the repeated observations per patient. The model-based variance assuming independence presumes an inappropriately larger number of independent observations, and generally underestimates the variance of the cluster-level predictors. However, with an implant-level variable (i.e. intraoral location), variances can be over or underestimated when the clustering is ignored. In this study, variances of such variables were generally underestimated when clustering was ignored.

It is clear that the correlation structure of dental implant data must be considered with a time to failure analysis. Each patient represents a cluster of implants which are correlated with respect to failure. Our analyses show that although predictor variables are significant influences on the risk to failure when clustering is ignored, introduction of the cluster-level robust variance often deflates this significance. When we utilize the robust variance analysis at the observation (implant) level in the logistic regression model, the model-based binomial variance structure is relaxed. However, the robust variance at the cluster level relaxes the model-based variance structure and calculates each cluster's independent contribution to the variance.

The need to adjust for correlation between observations becomes apparent in most dental data. Most studies of implant failure and certainly implant companies have employed Kaplan-Meir and Cox models without regarding the correlation between observations. Only more recent dental implant studies have executed GEE or robust variance analyses. The need for adjusting for correlated observations has been acknowledged in earlier studies (within the recent decade).

Some issues of interest to address in future studies of time-to-failure of dental implant would include an assessment of repeated failures. This could not be addressed with this data due to confusion with respect to placement and follow up dates per implant, as discussed in Chapter 3. Also, the influence of natural teeth approximating implants and their risk of failure is a clinical topic not yet evaluated. Clinical issues involving smoking, medication use, the patient's current prosthetic or restorative status and periodontal status could be investigated in other studies employing some of the methods described in this thesis. Another statistical issue to investigate would be the risk of failure of implants that approximate a site of an implant that has failed. This can be done with these data but requires considerably more time for programming with respect to evaluating each implant site conceptually and determining whether or not an implant was adjacent to it and if so, whether the implant failed. Other possible approaches not considered here are spatial analysis and Bayesian techniques.

Researchers in clinical dentistry need to be informed of the unique clustering of observations involved with this health specialty. Planning for dental studies requires acknowledgement of such clustering and subsequent planning for proper data collection, formatting and analysis.

APPENDIX A
DATAFORMS

FORM A Patient History Implant Rationale

VA DENTAL IMPLANT REGISTRY

PATIENT SSN | | | | | | | | | |

Provider ID | | | | |

Station No. | | | |

1. Patient date of BIRTH ? M M D D Y Y

2. Patient ethnic identification ? (check all that apply)
- ○ White ○ Asian ○ Hispanic
- ○ Black ○ Native Am ○ Other

3. Patient Gender ○ Male ○ Female

4. How will this case be paid for ? (check all that apply)
- ○ Patient ○ Insurance ○ Research
- ○ Teaching ○ Medicaid ○ No fee ○ OTHER

5. What was the primary source of demand for implants in this case ? (Check all that apply)
- ○ Patient initially requested implants
- ○ Dentist initially suggested implants
- ○ Patient was referred by dentist
- ○ Patient was referred by physician
- ○ Unknown ○ OTHER

6. What factors motivated the use of implants in this case ? (check all that apply)

Max	Mand (indicate to which arch this applies)
○	○ Previous prosthetic failure
○	○ Inadequate alveolus
○	○ Post maxillofacial trauma
○	○ Post maxillofacial pathology
○	○ Gagging
○	○ Anatomic anomaly
○	○ Previous implants
○	○ Single tooth replacement
○	○ Poor denture retention
○	○ Patient dissatisfaction
○	○ Provider preference/Tx of choice
○	○ OTHER ____________________

7. What PROSTHESES is the patient currently wearing ? (check all that apply)

Max	Mand	
○	○ CD	○ No Max Dentures
○	○ RPD	○ No Mand Dentures
○	○ FPD	

8. Does patient have Oral Habits ? (check all that apply)
- ○ Bruxism/clenching
- ○ Tongue thrust
- ○ Foreign objects (e.g. pipe)
- ○ Mouth breather
- ○ OTHER
- ○ NONE

9. What is the date of this EXAM? M M D D Y Y

10. Medical History (check all that apply)

○ Bleeding Disorders	○ Hypersensitivities (metals)
○ Other Blood Disorders	○ Allergies
○ High Blood Pressure	○ TMJ Disorders
○ Cardiovascular	○ Hx Rad Therapy (head & neck)
○ Cancer	○ Chemotherapy
○ Diabetes	○ Hx Tobacco Use
○ Endocrine (non diabetes)	○ Hx Alcoholism
○ Genitourinary	○ Hx Other Substance Abuse
○ Neuromuscular	○ Xerostomia
○ Osteoporosis	○ Ectodermal Dysplasia
○ Pulmonary	○ OTHER
○ Psychiatric	
○ Immune Disorder	○ NONE NOTED

11. What medications is the patient currently taking (check all that apply)

○ Analgesics	○ Cardiovascular
○ NSAID Analgesics	○ Corticosteroids
○ Anticoagulants	○ Digoxin
○ Anticonvulsants	○ G.I Preps
○ Antidepressant	○ Hormone
○ Antihypertensive	○ Oral Hypoglycemics
○ Antihistamine/decongst	○ Injectable Insulin
○ Antimicrobial	○ Minor Antianxiety
○ Antineoplastic	○ Sedative/Hypnotics
○ Antipsychotic	○ OTHER
○ Bronchodilators	○ NONE NOTED

12. What is the patient's ASA rating (1-5) [ ]

ASA CODES
- 1= normal healthy patient
- 2= mild to moderate systemic disease
- 3= severe but not incapacitating systemic disease
- 4= severe disease that limits activity and is a constant threat to life
- 5= moribund

13. What is the patient's JAW RELATIONS
- ○ Class I
- ○ Class II
- ○ Class III

14. Is patient COMPLETELY edentulous ?

○ YES ○ NO: (If patient has ANY remaining natural teeth complete BACK of this form)

- OVER -

6/92

1. EXISTING DENTITION
Place an X over the number of all MISSING NATURAL TEETH
(large spans of missing teeth can be indicated with two Xs connected by a line)

UR					UA						UL				
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
32	31	30	29	28	27	26	25	24	23	22	21	20	19	18	17
LR					LA						LL				

The following is an assessment of the **Caries** and **Periodontal Status** of the patient. The assessment is done by sextant. Please answer each question by marking ALL sextants that have the characteristic indicated.
(KEY: UR=upper right, UA=upper anterior, UL=upper left, LR=lower right, LA=lower anterior, LL=lower left)
please indicate NONE if condition is absent in all sextants.

Please mark for each sextant where there is ANY occurance of the following:

2.	NONE	UR	UA	UL	LL	LA	LR
Periodontal bleeding on gentle probing	○	○	○	○	○	○	○
Supragingival calculus	○	○	○	○	○	○	○
Subgingival calculus	○	○	○	○	○	○	○
Periodontal pockets between 3-5 mm	○	○	○	○	○	○	○
Periodontal pockets greater than > 5 mm	○	○	○	○	○	○	○
Unfilled coronal caries	○	○	○	○	○	○	○
Unfilled root caries	○	○	○	○	○	○	○
+1 Tooth mobility	○	○	○	○	○	○	○
+2 Tooth mobility	○	○	○	○	○	○	○
+3 Tooth mobility	○	○	○	○	○	○	○
Received scaling/root planing within last 6 mo	○	○	○	○	○	○	○
Will be scaled/root planed before implant therapy	○	○	○	○	○	○	○

3. Please evaluate patient's **oral hygiene**

- ○ Excellent (no debris, plaque, or calculus detectable)
- ○ Good (no debris, minor isolated areeas of plaque/calculus)
- ○ Fair (most teeth affected by some plaque/calculus)
- ○ Poor (widespread oral debris, plaque/calculus affecting most teeth)

Comments (note: comments will not be stored in registry and cannot be retreived in reports.):

SUBMIT ORIGINAL FORMS ONLY TO REGISTRY

VA FORM A 6/92

FORM B1 Cylinder Implant Surgery

VA DENTAL IMPLANT REGISTRY

PATIENT SSN [ ][ ][ ]-[ ][ ]-[ ][ ][ ][ ] Provider ID [ ][ ][ ][ ] Station No. [ ][ ][ ]

USE A SEPARATE SHEET FOR EACH ARCH

1. This form is for implants in which arch ?

○ Maxilla ○ Mandible

2. TOTAL number of implants in this ARCH? [ ]

3. Date of implant placement ? [ ][ ][ ][ ][ ][ ]
M M D D Y Y

4. Location of implant surgery ?

○ Dental operatory ○ Dental Office surgical suite
○ Hospital OR ○ OTHER

5. Patient's status for implant surgery

○ Outpatient ○ Inpatient ○ OTHER

6. What anesthesia was used for implant surgery (mark all that apply)

○ General ○ Local (block)
○ I.V. Sedation ○ Local (infiltration)
○ Nitrous Oxide ○ OTHER

7. How long did implant surgery take (incision to close) **?**

○ Less than 1/2 hr. ○ 1 to 2 hrs
○ 1/2 to 1 hr ○ more than 2 hrs

8. Please rate the complexity of the surgery ?

○ Highly complex ○ Complex ○ Routine

9. If management of case is complex, is it due to: (check all that apply)

○ Patient health complications
○ Implant surgery requirements
○ Prosthetic requirements
○ OTHER

10. Which medications were given in association with the implant surgery ? (check all that apply and note duration)

Antibiotics/Antiseptics: duration (days)

○ Penicillin ______
○ Ampicillin ______
○ Amoxycillin ______
○ Erythromycin ______
○ Tetracycline ______
○ Cephalosporin ______
○ Doxycycline ______
○ Clindamycin ______
○ Chlorhexidine ______
○ OTHER______________ ______
○ NONE

Analgesics/Sedative:

○ ASA ______
○ Acetaminophen ______
○ Narcotic ______
○ ASA+Narcotic ______
○ Acetaminophen+Narcotic ______
○ NSAID ______
○ Barbiturate ______
○ Diazepam ______
○ Other ______________ ______
○ Steroid:______________ ______
○ NONE

11. Will the uncovering procedure be done by another dentist ?

○ No ○ Yes ○ Not applicable

12. Will the prosthesis be done by another dentist ?

○ No ○ Yes

Comments:

-OVER- 6/92

IMPLANT DATA: Please use the following codes to decribe EACH cylinder implant used. Place the appropriate number in the blocks under the implant site numbers. (USE FORM B2 FOR NON-CYLINDER IMPLANTS)

1. Manufacturer Codes		2. Implant Material Codes	3. Implant Coating Codes
1 Branemark	10 OTC	1 Titanium - commercially pure	1 NONE
2 Calcitek	11 Osteo	2 Titanium - alloy	2 Hydroxylapatite
3 Collagen	12 Park Dental	3 Vitallium - surgical grade	3 Titanium plasma spray
4 Core Vent	13 Stryker	4 Vitreous carbon	4 Ceramic plasma spray
5 Denar	14 Synthes	5 Alluminum oxide or crystal	5 Carbon
6 IMZ	15 Titanodont	6 Ceramic	6 Titanium /HA
7 ITI	16 Ultimatics	7 Stainless steel	7 Titanium macropore
8 Miter	17 Zimmer (staple)	8 Tantalum	8 OTHER
9 Oratronic	18 OTHER(describe)	9 OTHER	

4. Stage Codes (cylinder/blade only)

1 One stage
2 Two stage (submerged healing)
3 Not applicable

5. Implant Morphology Codes

1 Basket
2 Bullet
3 Screw
4 OTHER
5 Not Applicable

The questions below apply to CYLINDER IMPLANTS ONLY. Up to 8 implants can be described in the columns below. If more than 8 implants are used in this arch - use additional sheets. Please use the other specified forms for - blades, staples, and subperiosteal implants. Refer to each cylinder by reference to the number of the tooth it replaces (use standard 1-32 tooth numbering)

IMPLANT SITE NUMBER................	32	26						
1. Implant Manufacturer Code....................								
2. Implant Material Code....................								
3. Implant Coating Code....................								
4. Stage Code....................								
5. Implant Morphology Code....................								
6. Implant height (mm)(top to bottom).........								
7. Implant width/diameter (mm)....................								
8. Height of available bone (mm)....................								
9. Width of available bone (mm)....................								
10. Attached Gingiva (bucc/ling width in mm)....								
11. Branemark bone classification (1-4)..........								

12. SURGICAL DETAILS
For each implant note any occurence of the following during the implant surgery (mark all that apply)

Implant altered (describe below)....	O	O	O	O	O	O	O	O
Alveoler ridge perforation..........	O	O	O	O	O	O	O	O
Jaw Fracture..........	O	O	O	O	O	O	O	O
Neurological damage..........	O	O	O	O	O	O	O	O
Inf.Mandibular Border Perf..........	O	O	O	O	O	O	O	O
Sinus Lift..........	O	O	O	O	O	O	O	O
Perf sinus/nasal cavity..........	O	O	O	O	O	O	O	O
Equipment complications..........	O	O	O	O	O	O	O	O
Unable to seat implant..........	O	O	O	O	O	O	O	O
Implant not well adapted to site..	O	O	O	O	O	O	O	O
Ridge augmentation used..........	O	O	O	O	O	O	O	O
Periodontal tissue damage..........	O	O	O	O	O	O	O	O
Patient experienced pain..........	O	O	O	O	O	O	O	O
Excessive bleeding..........	O	O	O	O	O	O	O	O
Guided tissue regeneration.......... (Membrane e.g. GoreTex).	O	O	O	O	O	O	O	O
OTHER	O	O	O	O	O	O	O	O

VA Form B1 6/92

FORM C Prosthetic Restoration

VA DENTAL IMPLANT REGISTRY

PATIENT SSN [ ][ ][ ]-[ ][ ]-[ ][ ][ ][ ]   Provider ID [ ][ ][ ][ ]   Station No. [ ][ ][ ]

Complete one FORM C for EACH PROSTHESIS

1.This form is for a prosthesis in which arch ?

○ Maxilla ○ Mandible

2. Is the prosthesis :

○ Anterior ○ Posterior ○ Full Arch

3. Date implants **UNCOVERED** ?

○ Not applicable
○ Unknown

M M D D Y Y

4. Date **TEMPORARY PROSTHESIS** inserted ?

○ Not applicable
○ Unknown

M M D D Y Y

5. Date **FINAL PROSTHESIS** inserted ?

M M D D Y Y

6. Was this prosthesis done by the same dentist who surgically placed the implants ? my interest

○ Yes ○ No

7. Has the patient's medical status changed significantly since the implant surgery ? (if yes please describe below)

○ Yes ○ No

8. What type of dentition opposes this prosthesis ? (check all that apply)

○ Natural dentition only
○ RPD
○ Complete denture
○ OTHER
○ Edentulous
○ FPD
○ Implants

9. Are there implants in the opposing arch ?

○ Yes ○ No

10. What type of prosthesis was placed in this arch ?

Partial Arch:

○ Single tooth (free standing)
○ Sinlge tooth (rests on adjacent teeth)
○ Implant abutment for RPD
○ Implant abutment rigidly fixed to natural teeth
○ Implant abutment with nonrigid attachment to nat. teeth
○ Totally implant borne FPD
○ OTHER

Complete Arch:

○ Overdenture without bar
○ Fixed detachable prosthesis
○ Cast metal bar with overdenture
○ Fixed bridge
○ OTHER

11. Was implant position/alignment acceptable for planned prosthesis ?

○ Yes ○ No

12. Was surgical guide/stent used at time of surgery ?

○ Yes ○ No

13. The occlusal surface(s) of the prosthesis was constructed of (check all that apply)

○ Porcelain ○ Metal
○ Resin ○ OTHER

14. What is your assessment of the patient's oral hygiene ?

○ Excellent (no debris, plaque, or calculus detectable)
○ Good (no debris, minor isolated areeas of plaque/calculus)
○ Fair (most teeth affected by small amounts of plaque/calculus)
○ Poor (widespread oral debris, plaque/calculus affecting most teeth)

15. How is abutment attached to implant ? (check all that apply)

○ Screw ○ Cement
○ Friction ○ OTHER

16. How is substructure attached to abutment? (check all that apply)

○ Screw ○ Cement
○ Friction ○ No Substructure
○ OTHER

17. If substructure is used, how is prosthesis retained ?

○ Not applicable
○ Screw
○ Magnet
○ Clips (all kinds)
○ Cement
○ Rubber O-ring
○ Resilient liner
○ OTHER

Comments:

SUBMIT ORIGINALS ONLY TO REGISTRY 6/92

IMPLANT DATA at each abutment: Use standard 1-32 tooth numbering to identify each implant.
Be sure implant numbers are consistant with FORM B

IMPLANT SITE NUMBER —— 27 | 20

IMPLANT HEALTH STATUS

1. PERIODONTAL (Associated with implant)

- Bleeding on probing
- Gingival recession >3 mm
- Probing depth > 3mm
- Dehiscence
- Hyperplasia (peri-implant)
- Mobility

Peri-implant inflamation
(0=none, 1=slight, 2=diffuse, 3=general)

Width of keratinized mucosa (mm): Mandible buccal:
Mandible: lingual :
Maxilla: buccal:

2. RADIOGRAPHIC

- Bone not well adapted to implant
- Cervical cupping/cratering
- Bone loss <1/4 of implant lengh
- Bone loss 1/4 - 1/2 implant length
- Bone loss 1/2- - 3/4 implant length
- Peri-implant radiolucency
- Radiographically within normal limits
- Not assessed/not reported
- OTHER

3. IMPLANT HEALTH CATEGORY (Loma Linda)

1= no clinical pathology at implant site
2= minor clinical path., no intervention, prognosis good
3= clinical path., surgical intervention needed, prognosis good
4= clinical path., intervention need, prognosis poor
5= NOT treatable, prognosis poor, implant still in place
6= Implant removed (note date below)

DATE OF IMPLANT REMOVAL (IF APPLICABLE) M M D D Y Y

4. FUNCTIONALITY

- Implant functioning well
- Implant less than optimal but serviceable
- Implant SUBMERGED (sleeper)
- Implant TO BE REMOVED
- OTHER

5. PATIENT COMPLAINTS (implant related)

- Pain associated with implant
- Pain elsewhere due to implant
- Loss of sensation near implant
- Esthetics due to implant
- Mastication problems due to implant
- Speech problems due to implant
- OTHER

6. Check here if a COMPROMISED IMPLANT INTERVENTION WAS USED
(e.g. bone graft, bone substitute, guided tissue regeneration)

VA FORM C/D 6/92

FORM D Recall VA DENTAL IMPLANT REGISTRY

PATIENT SSN [| |]-[|]-[| | |] Provider ID [| | |] Station No. [| |]

Use a separate form for EACH PROSTHESIS

1. This form is for the prosthesis in which arch ?
   - ○ Maxilla ○ Mandible

2. Date of this appointment ? [| | | | |] M M D D Y Y

3. The dentist completing this form provided which service(s) ? (check all that apply)
   - ○ Implant surgery ○ OTHER
   - ○ Prosthodontics

4. Is this appointment ?
   - ○ Routine recall (scheduled by dentist)
   - ○ Emergency appointment
   - ○ OTHER

5. Who initiated this appointment ? (check all that apply)
   - ○ Implant surgeon ○ Patient (or family)
   - ○ Prosthesis provider ○ Physician
   - ○ Other dentist ○ OTHER

6. Has the patient's medical status changed significantly since last visit ?
   - ○ Yes ○ No ○ First visit with patient

7. Does patient have any new oral habits which affect this case?
   - ○ Bruxism/clenching ○ Mouth breathing
   - ○ Tongue thrust ○ OTHER
   - ○ Foreign objects ○ NONE

8. What is your assessment of the patients oral hygiene ?
   - ○ Excellent (no debris, plaque, or calculus detectable)
   - ○ Good (no debris, minor isolated areeas of plaque/calculus)
   - ○ Fair (most teeth affected by small amounts of plaque/calculus)
   - ○ Poor (widespread oral debris, plaque/calculus affecting most teeth)

9. What services were provided at this appointment? (check all that apply)
   - ○ Examination ○ Radiographs
   - ○ Prophylaxis ○ Implant adjustment
   - ○ Denture adjustment ○ Implant removed
   - ○ Occlusal adjustment ○ OTHER
   - ○ Prosthesis tightened

10. What was the patient's level of satisfaction with the prosthesis at this appointment ?
    - ○ Very Satisfied ○ Satisfied ○ Unsatisfied ○ Very Unsatisfied

11. If patient was dissatisfied, why ?(check all that apply)
    - ○ Not applicable (patient satisfied)
    - ○ Pain with prosthesis
    - ○ Speech/phonetics difficulties
    - ○ Esthetics
    - ○ Mastication
    - ○ Retention problems
    - ○ Stability problems
    - ○ Poor cleaning access
    - ○ No improvement over previous prosthesis
    - ○ OTHER

12. How was patient satisfaction determined:
    - ○ Clinical impression of dentist
    - ○ Patient volunteered information
    - ○ Dentist asked patient directly
    - ○ OTHER

13. What is your evaluation of this prosthesis ? (check all that apply)
    - ○ Prosthesis function is acceptable
    - ○ Prosthesis esthetics is acceptable
    - ○ Prosthesis requires remake/major revison
    - ○ Implant attachment breakage
    - ○ Prosthesis fracture
    - ○ Restorative material freacture
    - ○ Loose screw
    - ○ Occlusal difficulties
    - ○ Stability difficulties
    - ○ Retention difficulties
    - ○ Esthetics unacceptable
    - ○ OTHER
    - ○ Prothesis NOT EVALUATED

14. Do you believe the implant(s) were at least in part responsible for any prosthesis difficulties ?
    - ○ Yes ○ No ○ Not applicable

EVALUATE IMPLANTS ON BACK

6/92

IMPLANT DATA at each abutment: Use standard 1-32 tooth numbering to identify each implant. Be sure implant numbers are consistant with FORM B

IMPLANT SITE NUMBER —

IMPLANT HEALTH STATUS

1. PERIODONTAL (Associated with implant)

- Bleeding on probing
- Gingival recession >3 mm
- Probing depth > 3mm
- Dehiscence
- Hyperplasia (peri-implant)
- Mobility

Peri-implant inflamation (0=none, 1=slight, 2=diffuse, 3=general)

Width of keratinized mucosa (mm): Mandible buccal:

Mandible: lingual:

Maxilla: buccal:

2. RADIOGRAPHIC

- Bone not well adapted to implant
- Cervical cupping/cratering
- Bone loss <1/4 of implant length
- Bone loss 1/4 - 1/2 implant length
- Bone loss 1/2- - 3/4 implant length
- Peri-implant radiolucency
- Radiographically within normal limits
- Not assessed/not reported
- OTHER

3. IMPLANT HEALTH CATEGORY (Loma Linda)

1= no clinical pathology at implant site
2= minor clinical path., no intervention, prognosis good
3= clinical path., surgical intervention needed, prognosis good
4= clinical path., intervention need, prognosis poor
5= NOT treatable, prognosis poor, implant still in place
6= Implant removed (note date below)

DATE OF IMPLANT REMOVAL (IF APPLICABLE) M M D D Y Y

4. FUNCTIONALITY

- Implant functioning well
- Implant less than optimal but serviceable
- Implant SUBMERGED (sleeper)
- Implant TO BE REMOVED
- OTHER

5. PATIENT COMPLAINTS (implant related)

- Pain associated with implant
- Pain elsewhere due to implant
- Loss of sensation near implant
- Esthetics due to implant
- Mastication problems due to implant
- Speech problems due to implant
- OTHER

6. Check here if a COMPROMISED IMPLANT INTERVENTION WAS USED (e.g. bone graft, bone substitute, guided tissue regeneration)

VA FORM C/D 6/92

FORM U Implant Uncovering

VA DENTAL IMPLANT REGISTRY

PATIENT SSN [ ][ ][ ]-[ ][ ]-[ ][ ][ ][ ] Provider ID [ ][ ][ ][ ] Station No [ ][ ][ ]

Please Complete one FORM U for **EACH ARCH**

1.This form is for a prosthesis in which arch ?
○ Maxilla ○ Mandible

3. Date implants **UNCOVERED** ?
[ ][ ] [ ][ ] [ ][ ]
M D Y

IMPLANT DATA: AT UNCOVERING

Place the tooth number for EACH cyclinder in the space provided. Use standard 1-32 tooth numbering to identify each implant

IMPLANT SITE NUMBER —— [ ] [ ] [ ] [ ] [ ] [ ] [ ] [ ]

IMPLANT HEALTH STATUS

1. RADIOGRAPHIC

Bone not well adapted to implant	○	○	○	○	○	○	○	○
Cervical cupping/cratering	○	○	○	○	○	○	○	○
Bone loss <1/4 of implant lengh	○	○	○	○	○	○	○	○
Bone loss 1/4 - 1/2 implant length	○	○	○	○	○	○	○	○
Bone loss 1/2- - 3/4 implant length	○	○	○	○	○	○	○	○
Peri-implant readiolucency	○	○	○	○	○	○	○	○
Radiographically within normal limits	○	○	○	○	○	○	○	○
Not assessed/not reported	○	○	○	○	○	○	○	○
OTHER	○	○	○	○	○	○	○	○

2. IMPLANT MOBILITY

YES	○	○	○	○	○	○	○	○
NO	○	○	○	○	○	○	○	○

Width of keratinized mucosa (mm):

Mandible buccal:	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Mandible: lingual :	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Maxilla: buccal:	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

3. IMPLANT UNCOVERED BY:

Biopsy Punch	○	○	○	○	○	○	○	○
Crestal Incision	○	○	○	○	○	○	○	○
Already Exposed at this Appointment	○	○	○	○	○	○	○	○
OTHER	○	○	○	○	○	○	○	○

4. Check here if IMPLANT WAS REMOVED at THIS APPOINTMENT	○	○	○	○	○	○	○	○
5. Implant NOT UNCOVERED (Sleeper)	○	○	○	○	○	○	○	○

Comments:

6/92

FORM **X** Patient Inactivation

VA DENTAL IMPLANT REGISTRY

PATIENT SSN [ ][ ][ ]-[ ][ ]-[ ][ ][ ][ ] Provider ID [ ][ ][ ][ ] Station No. [ ][ ][ ]

2. Date of this report ? [ ][ ][ ][ ][ ][ ]
M M D D Y Y

1. The patient noted above will be flagged in the registry for the following reason:

- ○ Patient lost to follow-up reason unknown

- ○ Patient moved
- ○ Patient deceased
- ○ Patient seeking care elsewhere
- ○ All implants removed or buried
- ○ Eligibility for VA care expired
- ○ OTHER (please note below)

2. Please indicate if to the best of your knowledge any of this patients implants were removed.

- ○ No implants removed
- ○ Status of implants unknown
- ○ All implants removed
- ○ Some implants removed: (please indicate below the abutment number of removed implants)

Implant site number of REMOVED implants	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
DATE of IMPLANT REMOVAL (if known) mm/dd/yy	/ /	/ /	/ /	/ /	/ /	/ /

Comments

6/92

APPENDIX B
CODEBOOK LISTING AND VARIABLES OF DATASET

Description of Premerged Datasets

Form A Dataset: (See Table 1)
There are 1,462 records in this dataset.
There are 1,357 unique identifiers (representing patients) indicating duplicate records.

Variable names:
ssn   Social Security Number (scrambled)
xdate   Date of initial examination
provid Provider
station Location of treatment
bdate   Birthdate

Ethnicity:
ethw   White
ethb   Black
etha   Asian
ethnam Native American
ethhis   Hispanic
ethoth   Other

sex   Sex

Diagnostic variables affiliated with the FormA dataset:
These are dichotomous variables that can not be validated (as one can see from a comparison of the hard copy form and the order of the suffix numbers attached to the prefix dia_). Much information can be obtained from such variables if (1) more descript names were provided and/or (2) a data dictionary or labeling was provided with the dataset(s). Because validation of these variables is impossible, these variables were not included in the final dataset.

diag1-diag68
sedhyp Sedative/Hypnotic medications used by the patient
othmed Other medications used by the patient

mednone No medications noted
asarate   ASA rating (1-5):
1=normal healthy patient

2=mild to moderate systemic disease
3=severe but not incapacitating systemic disease
4=severe disease that limits activity and is a constant threat to life
5=moribund
edenttot   Is the patient completely edentulous? (yes/no).
oralhyg   Oral Hygiene:
A-Excellent
B-Good
C-Fair
D-Poor

The following variables were presented initially in string format and were converted to numeric form so that they may be used for analysis.
newssn Numeric Social security numbers
nxdate   Numeric initial examination date
nbdate   Numeric birthdate
gender   Numeric Sex

There are no variables in the dataset with names that would correspond with information on existing teeth, periodontal status, prosthesis type that the patient is currently wearing, jaw relation, primary source of demand for implants, or how the patient paid for the treatment. However, these variables are listed on the questionnaire for the Form A dataset. It was also stated, via personal communication with Robert Weyant, that the order of the dataset variables were to follow the order of the questions in the questionnaire. As mentioned above, this ordering was not followed.

Placement Dataset: (See figure 1)
This is the dataset that incorporates the placement dates for the implants.

There are 4,313 observations or records.
There are 1,294 unique identifiers.

There appears to be 1294 patients (all patients would be expected to have at least one visit) with the first placement date, 46 patients who have ever had a second visit or placement date, and only one person who has ever experienced a third placement date for an implant in the same implant site.
The range of implants placed per person is from 1-14 implants. All patients have at least one (1) implant placed

Variable names:
ssn - Social Security (string format)

isdate - Placement date
imparch - Arch (Maxilla or Mandible)
imp1-Implant site locator (1-32)

imp2 – Implant manufacturing code
imp3 – Implant material code
imp4 – Implant Coating code
imp5 – Stage code
imp6 – Implant Morphology code
imp7 – Implant Height (mm)(top to bottom)
imp8 - Implant Width/diameter (mm)
imp9 – Height of available bone (mm)
imp10 – Width of available bone (mm)
avboneht – Average bone height (mm)
avbonewi – Average bone width (mm)
attginwi – Attached gingival width
bonclass – Bone classification (I assume Branemark Classification)

Surgical Details:
surocc1 – Implant altered
surocc2 – Alveolar ridge perforation
surocc3 – Jaw Fracture
surocc4 – Neurological damage
surocc5 – Inferior Mandibular Border Perforation
surocc6 – Sinus Lift
surocc7 – Perforated Sinus/Nasal Cavity
surocc8 – Equipment complications
surocc9 – Unable to seat implant
surocc10 – Implant not well adapted to site
surocc11 - Ridge augmentation used
surocc12 – Periodontal tissue damage
surocc13 – Patient experienced pain
surocc14 – Excessive bleeding
surocc15 – Guided tissue regeneration (Membrane e.g. Gore Tex)
surocc16 - Other
newssn – Numeric Social Security Number
nisdate – Numeric placement date
newid – Numeric id used which incorporates the site with an individual social security number.

Removal Dataset: (See figure 1)
This is the dataset that incorporates the removal and evaluation or follow-up dates for the implants.

There are 10,624 observations in the dataset.
There are 1009 unique values in the dataset.

Newid is a variable which indicates the number of patient-sites in the dataset which equals 3485 in this dataset.
The variable sitetot2 indicates the total number of implants per individual. A summary of sitetot2 presents that the mean implants per patient was 2-3 per patient and that there is the possibility of having 15 implants ever placed.

The visit2 variable indicates the number of visits the patients had. It appears that the range of visits was from 1-25. There are 1009 patients with at least one visit and one patient who ever had 25 visits. There is a mean of 3 visits per individual.

Variable names:
Ssn - Social Security number (string format)
Site – Implant site indicator (1-32)
Evaldate – Evaluation date (string format)
Mobil - Mobility
Periminf – Peri-implant inflammation
Imphcat – Implant Health Category
Imprdate – Implant removal date
Impfunc - Functionality
Impltopt – Implant less than optimal but serviceable
Impnonsp – Can't be verified (Could refer to whether or not the implant is submerged)
imp2brmv – Implant to be removed
funother – Can not be verified
painlswr – Can not be verified(Could refer to pain associated with implant or elsewhere)
esthetic – Esthetics due to implant
mastprob – Mastication problems due to implant
speechpr – Speech problems due to implant
cmplnoth – Implant related complaints-other
compimpi – If compromised implant intervention was used (bone graft, bone substitute, guided tissue regeneration)
newssn – Numeric Social Security
nevldate – Numeric evaluation date
nimrdate – Numeric implant removal date
newid - Numeric id used which incorporates the site with an individual social security number.
Imph – A variable created to be an index for the 6th level of implant health category which indicated removal of an implant.
Rem – A variable used to index removal of an implant and used in the creation of other variables
Post – A variable used as an index for times after removal of an implant
Ps – A variable used as an index to remove implant removal dates after an implant was removed

Analytical Dataset

Commands for created variables

A variable "freq" was created to count the first records for each patient and each site. The xttab command for the overall frequency and percent calculations accounts for all records and this includes all follow-up records. This inflates the value for implants.

```
. by id site:gen freq=followup[1]

. by id site:replace freq=1 if followup==freq
(2305 real changes made)
 replace freq=0 if freq~=1
(5681 real changes made)
. tabulate freq
```

freq	Freq.	Percent	Cum.
0	5681	71.14	71.14
1	2305	28.86	100.00
Total	7986	100.00	

It appears that there are 2305 total implants in 777 patients.
An xttab procedure on the records representing the first placement date produces the following:

```
. iis id
. tis followup
. xttab  imptype if freq==1
```

	Overall		Between		Within
imptype	Freq.	Percent	Freq.	Percent	Percent
2	13	0.56	5	0.64	100.00
4	95	4.12	45	5.79	85.59
5	4	0.17	1	0.13	100.00
6	6	0.26	2	0.26	60.00
10	2	0.09	1	0.13	33.33
18	16	0.69	4	0.51	94.12
19	2169	94.10	725	93.31	99.45
Total	2305	100.00	783	100.77	98.44
			(n = 777)		

Here the "Between frequency" and "percent" have not changed. However, the "Overall Frequency" and "Percent" changed because the records of followup were not counted.

Codebook for Analytical Dataset:
. codebook

```
surocc9 ---------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [0,1]                    units:  1
      unique values:  2                coded missing:  5683 / 7986

         tabulation:  Freq.  Value
                      2277  0
                        26  1


surocc10 --------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [0,1]                    units:  1
      unique values:  2                coded missing:  5701 / 7986

         tabulation:  Freq.  Value
                      2222  0
                        63  1

surocc11 --------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [0,1]                    units:  1
      unique values:  2                coded missing:  5733 / 7986

         tabulation:  Freq.  Value
                      2242  0
                        11  1
surocc12 --------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [0,1]                    units:  1
      unique values:  2                coded missing:  5793 / 7986

         tabulation:  Freq.  Value
                      2164  0
                        29  1
```

surocc13 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 5837 / 7986

tabulation: Freq. Value
2134 0
15 1

surocc14 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 5934 / 7986

tabulation: Freq. Value
1996 0
56 1

surocc15 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 6391 / 7986

tabulation: Freq. Value
1556 0
39 1

surocc16 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 7729 / 7986

tabulation: Freq. Value
256 0
1 1

```
_merge ------------------------------------------------------------ (unlabeled)
                type:  numeric (byte)

               range:  [2,3]                     units:  1
       unique values:  2                 coded missing:  0 / 7986

          tabulation:  Freq.  Value
                         278  2
                        7708  3


age1 -------------------------------------------------------------- (unlabeled)
                type:  numeric (float)

               range:  [.40273973,82.331505]     units:  1.000e-08
       unique values:  306               coded missing:  5999 / 7986

                mean:   62.6349
            std. dev:   10.4379

         percentiles:        10%       25%       50%       75%       90%
                         45.3589   58.8411   65.8192   69.1671   72.6219


newid2 ------------------------------------------------------ group(id site)
                type:  numeric (float)

               range:  [1,2305]                  units:  1
       unique values:  2305              coded missing:  0 / 7986

                mean:   1130.18
            std. dev:   648.061

         percentiles:        10%       25%       50%       75%       90%
                             199       600      1098      1659      2007
y ----------------------------------------------------------------- (unlabeled)
                type:  numeric (float)

               range:  [1,1]                     units:  1
       unique values:  1                 coded missing:  7961 / 7986

          tabulation:  Freq.  Value
                          25  1
```

```
id ------------------------------------------------------------------------ id
                type:  numeric (double)

               range:  [1,1492]                    units:  1
       unique values:  777                 coded missing:  0 / 7986

                mean:   753.539
            std. dev:   406.257

         percentiles:        10%       25%       50%       75%       90%
                             129       461       738      1071      1296


nisdate ----------------------------------------------------------------- isdate
                type:  numeric daily date (double)

               range:  [7748,12386]                units:  1
     or equivalently:  [19mar1981,29nov1993]       units:  days
       unique values:  638                 coded missing:  0 / 7986

                mean:   10689.4 = 07apr1989 (+ 9 hours)
            std. dev:   710.906

         percentiles:        10%       25%       50%       75%       90%
                            9799     10267     10713     11170     11562
                       30oct1986 10feb1988 01may1989 01aug1990 28aug1991


place ----------------------------------------------------------- (unlabeled)
                type:  numeric daily date (float)

               range:  [7748,12386]                units:  1
     or equivalently:  [19mar1981,29nov1993]       units:  days
       unique values:  638                 coded missing:  0 / 7986

                mean:   10689.3 = 07apr1989 (+ 8 hours)
            std. dev:   710.893

         percentiles:        10%       25%       50%       75%       90%
                            9799     10267     10713     11170     11562
                       30oct1986 10feb1988 01may1989 01aug1990 28aug1991
```

```
nevldate ----------------------------------------------------------- evaldate
                type:  numeric daily date (double)

               range:  [7906,12745]              units:  1
     or equivalently:  [24aug1981,23nov1994]     units:  days
       unique values:  1462                  coded missing:  0 / 7986

                mean:     11375 = 22feb1991 (+ 1 hour)
            std. dev:   768.548

         percentiles:        10%       25%       50%       75%       90%
                           10307     10898   11442.5     11995     12331
                       21mar1988 02nov1989 30apr1991 03nov1992 05oct1993

followup --------------------------------------------------------- (unlabeled)
                type:  numeric daily date (float)

               range:  [7906,12745]              units:  1
     or equivalently:  [24aug1981,23nov1994]     units:  days
       unique values:  1465                  coded missing:  0 / 7986
                mean:   11374.6 = 21feb1991 (+ 16 hours)
            std. dev:   768.485

         percentiles:        10%       25%       50%       75%       90%
                           10307     10898     11442     11994     12331
                       21mar1988 02nov1989 30apr1991 02nov1992 05oct1993

nimrdate ----------------------------------------------------------- imprdate
                type:  numeric daily date (double)

               range:  [9735,12589]              units:  1
     or equivalently:  [27aug1986,20jun1994]     units:  days
       unique values:  75                    coded missing:  7872 / 7986

                mean:   11068.7 = 21apr1990 (+ 17 hours)
            std. dev:   748.244

         percentiles:        10%       25%       50%       75%       90%
                           10074     10455   10942.5     11648     12087
                       01aug1987 16aug1988 16dec1989 22nov1991 03feb1993
```

```
site ------------------------------------------------------------------- site
             type:  numeric (double)

            range:  [1,32]                   units:  1
      unique values:  31                 coded missing:  0 / 7986

             mean:   22.8123
          std. dev:   5.72435

        percentiles:       10%       25%       50%       75%       90%
                            14        22        23        27        28


failure --------------------------------------------------------- (unlabeled)
             type:  numeric (float)

            range:  [0,1]                   units:  1
      unique values:  2                 coded missing:  0 / 7986

         tabulation:  Freq.  Value
                      7883  0
                       103  1


rownames -------------------------------------------------------- (unlabeled)
             type:  string (str5)

      unique values:  7984               coded missing:  2 / 7986

          examples:  "240"
                     "4342"
                     "6116"
                     "82"
evaldate -------------------------------------------------------- (unlabeled)
             type:  string (str11)

      unique values:  1462               coded missing:  2 / 7986

          examples:  "07-APR-1993"
                     "13-APR-1993"
                     "18-SEP-1986"
                     "24-MAY-1988"
```

mobil -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
7885 0
99 1

periminf -- (unlabeled)
type: string (str1)

unique values: 4 coded missing: 3562 / 7986

tabulation: Freq. Value
3426 "0"
801 "1"
126 "2"
71 "3"

imphcat --- (unlabeled)
type: numeric (float)

range: [1,6] units: 1
unique values: 6 coded missing: 186 / 7986

tabulation: Freq. Value
6556 1
771 2
213 3
35 4
91 5
134 6

imprdate -- (unlabeled)
type: string (str11)
unique values: 75 coded missing: 7872 / 7986

examples: ""
""
""
""

impfunc -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
3663 0
4321 1

impltopt --- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
7812 0
172 1

impnonsp -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
7912 0
72 1

imp2brmv -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
7926 0
58 1

funother -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
7934 0
50 1

painlswr -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
7960 0
24 1

esthetic -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
7970 0
14 1

mastprob -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 2 / 7986

tabulation: Freq. Value
7932 0
52 1

```
speechpr ------------------------------------------------------------ (unlabeled)
                  type:  numeric (float)

                 range:  [0,1]                    units:  1
         unique values:  2                coded missing:  2 / 7986

            tabulation:  Freq.  Value
                          7979  0
                             5  1
cmplnoth ------------------------------------------------------------ (unlabeled)
                  type:  numeric (float)

                 range:  [0,1]                    units:  1
         unique values:  2                coded missing:  2 / 7986

            tabulation:  Freq.  Value
                          7867  0
                           117  1
compimpi ------------------------------------------------------------ (unlabeled)
                  type:  numeric (float)

                 range:  [0,1]                    units:  1
         unique values:  2                coded missing:  2 / 7986

            tabulation:  Freq.  Value
                          7964  0
                            20  1


newid --------------------------------------------------------------- group(newssn)
                  type:  numeric (float)

                 range:  [1.22,1005.19]           units:  .01
         unique values:  2305             coded missing:  2 / 7986

                  mean:   496.084
              std. dev:    282.62

           percentiles:        10%       25%       50%       75%       90%
                             77.08    274.27    485.73   724.245    877.23
```

```
isdate ---------------------------------------------------------- (unlabeled)
              type:  string (str11)

       unique values:  638               coded missing:  0 / 7986

           examples:  "06-OCT-1988"
                      "12-OCT-1989"
                      "19-MAR-1981"
                      "25-JUL-1991"


imparch ---------------------------------------------------------- (unlabeled)
              type:  string (str1)

       unique values:  2                 coded missing:  0 / 7986

         tabulation:  Freq.  Value
                       7155  "N"
                        831  "X"


imp1 ------------------------------------------------------------ (unlabeled)
              type:  numeric (float)

              range:  [1,32]                   units:  1
       unique values:  31                coded missing:  0 / 7986

               mean:   22.8123
           std. dev:   5.72435

        percentiles:        10%       25%       50%       75%       90%
                             14        22        23        27        28
```

imptype -- Implant Type
type: numeric (float)

range: [2,19] units: 1
unique values: 7 coded missing: 0 / 7986

tabulation: Freq. Value

Freq.	Value
30	2
466	4
16	5
46	6
2	10
28	18
7398	19

matcode -- Material Code
type: numeric (float)

range: [1,19] units: 1
unique values: 14 coded missing: 0 / 7986

mean: 6.05672
std. dev: 5.16541

percentiles:	10%	25%	50%	75%	90%
	1	1	4	12	12

coatcode -- Coating Code
type: numeric (float)

range: [1,9] units: 1
unique values: 6 coded missing: 0 / 7986

tabulation: Freq. Value

Freq.	Value
3643	1
4222	2
71	3
13	6
20	7
17	9

```
stagecode ------------------------------------------------------- Stage Code
              type:  numeric (float)

             range:  [0,25]                    units:  1
     unique values:  8                 coded missing:  0 / 7986

        tabulation:  Freq.  Value
                       484  0
                      1349  1
                      3134  2
                      2617  3
                       376  4
                        17  8
                         5  13
                         4  25


morphcode -------------------------------------------------- Morphology Code
              type:  numeric (float)

             range:  [0,24]                    units:  .01
     unique values:  15                coded missing:  0 / 7986

              mean:   1.09504
          std. dev:   1.30029

       percentiles:        10%       25%       50%       75%       90%
                             0         0         1         2         2


implantheight ----------------------------------------- Implant Height (mm)
              type:  numeric (float)

             range:  [0,37]                    units:  .1
     unique values:  23                coded missing:  0 / 7986

              mean:   1.03418
          std. dev:   2.62501

       percentiles:        10%       25%       50%       75%       90%
                             0         0         0         2         3
```

```
implantwidth -------------------------------------- Implant Width/Diameter (mm)
              type:  numeric (float)

             range:  [0,45]                   units:  .01
     unique values:  28                coded missing:  35 / 7986

              mean:   3.55757
          std. dev:   6.05446

       percentiles:        10%       25%       50%       75%       90%
                             0         0         0         8        13
availboneheight --------------------------------- Height of Available Bone (mm)
              type:  numeric (float)

             range:  [0,40]                   units:  .01
     unique values:  23                coded missing:  262 / 7986

              mean:   1.18031
          std. dev:    2.8932

       percentiles:        10%       25%       50%       75%       90%
                             0         0         0     3.125      3.75

availbonewidth ----------------------------------- Width of Available Bone (mm)
              type:  numeric (float)

             range:  [0,35]                   units:  .01
     unique values:  35                coded missing:  1 / 7986

              mean:    4.0065
          std. dev:   7.35246

       percentiles:        10%       25%       50%       75%       90%
                             0         0         0       3.8        17

avboneht ---------------------------------------------------------- (unlabeled)
              type:  numeric (float)
             range:  [0,30]                   units:  .1
     unique values:  29                coded missing:  8 / 7986
              mean:   2.10834
          std. dev:   4.20306
       percentiles:        10%       25%       50%       75%       90%
                             0         0         0         4         7
```

avbonewi -- (unlabeled)
type: numeric (float)

range: [0,15] units: .1
unique values: 17 coded missing: 29 / 7986

mean: 1.17906
std. dev: 2.32455

percentiles:	10%	25%	50%	75%	90%
	0	0	0	1	5

attginwi -- (unlabeled)
type: numeric (float)

range: [0,25] units: .1
unique values: 12 coded missing: 7 / 7986

mean: .477503
std. dev: 1.15463

percentiles:	10%	25%	50%	75%	90%
	0	0	0	0	2

bonclass -- (unlabeled)
type: numeric (float)

range: [0,4] units: 1
unique values: 5 coded missing: 7 / 7986

tabulation: Freq.	Value
7716	0
30	1
163	2
52	3
18	4

surocc1 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 0 / 7986

tabulation: Freq. Value
7904 0
82 1

surocc2 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 0 / 7986

tabulation: Freq. Value
7887 0
99 1

surocc3 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 0 / 7986

tabulation: Freq. Value
7932 0
54 1

surocc4 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 0 / 7986

tabulation: Freq. Value
7915 0
71 1

surocc5 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 0 / 7986

tabulation: Freq. Value
7970 0
16 1

surocc6 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 0 / 7986

tabulation: Freq. Value
7845 0
141 1

surocc7 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 484 / 7986

tabulation: Freq. Value
7137 0
365 1

surocc8 -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 3072 / 7986

tabulation: Freq. Value
4574 0
340 1

provid -- (unlabeled)
type: string (str4)

unique values: 52   coded missing: 278 / 7986

examples: "2301"
"2301"
"2301"
"2307"

warning: variable has leading blanks

station -- (unlabeled)
type: numeric (float)

range: [.,.]   units: .
unique values: 0   coded missing: 7986 / 7986

tabulation: Freq. Value

ethw -- (unlabeled)
type: numeric (float)

range: [0,1]   units: 1
unique values: 2   coded missing: 278 / 7986

tabulation: Freq. Value
1037 0
6671 1

ethb -- (unlabeled)
type: numeric (float)

range: [0,1]   units: 1
unique values: 2   coded missing: 278 / 7986

tabulation: Freq. Value
6884 0
824 1

etha -- (unlabeled)
type: numeric (float)

range: [0,1]   units: 1
unique values: 2   coded missing: 278 / 7986

tabulation: Freq. Value
7702 0
6 1

ethnam -- (unlabeled)
type: numeric (float)

range: [0,1]   units: 1
unique values: 2   coded missing: 278 / 7986

tabulation: Freq. Value
7691 0
17 1

ethhis --- (unlabeled)
type: numeric (float)

range: [0,1]   units: 1
unique values: 2   coded missing: 278 / 7986

tabulation: Freq. Value
7599 0
109 1

ethoth -- (unlabeled)
type: numeric (float)

range: [0,1]   units: 1
unique values: 2   coded missing: 278 / 7986

tabulation: Freq. Value
7700 0
8 1

```
sex --------------------------------------------------------------- (unlabeled)
              type:  string (str1)

       unique values:  4                coded missing:  278 / 7986

         tabulation:  Freq.  Value
                        63  "0"
                       137  "F"
                      7501  "M"
                         7  "X"


asarate ----------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [1,4]                    units:  1
       unique values:  4                coded missing:  7581 / 7986

         tabulation:  Freq.  Value
                        16  1
                       268  2
                       116  3
                         5  4


edenttot ---------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [1,2]                    units:  1
       unique values:  2                coded missing:  7579 / 7986

         tabulation:  Freq.  Value
                       319  1
                        88  2
```

```
nxdate -------------------------------------------------------- numeric xdate
              type:  numeric daily date (float)

             range:  [-17583,12610]             units:  1
     or equivalently:  [11nov1911,11jul1994]       units:  days
       unique values:  634                  coded missing:  278 / 7986

              mean:   10631.1 = 08feb1989 (+ 3 hours)
          std. dev:   1180.38

         percentiles:        10%       25%       50%       75%       90%
                           9799     10275     10685     11140     11521
                      30oct1986 18feb1988 03apr1989 02jul1990 18jul1991


nbdate -------------------------------------------------------- numeric bdate
              type:  numeric daily date (float)

             range:  [-18323,11846]             units:  1
     or equivalently:  [01nov1909,07jun1992]       units:  days
       unique values:  288                  coded missing:  5999 / 7986

              mean:  -11443.5 = 02sep1928 (+ -13 hours)
          std. dev:   3871.09

         percentiles:        10%       25%       50%       75%       90%
                         -14914    -13887    -12473     -9801     -4903
                      03mar1919 24dec1921 07nov1925 02mar1933 30jul1946


gender --------------------------------------------------------- numeric sex
              type:  numeric (long)
             label:  gender

             range:  [1,6]                      units:  1
       unique values:  4                    coded missing:  278 / 7986

          tabulation:  Freq.   Numeric  Label
                          63         1  0
                         137         3  F
                        7501         4  M
                           7         6  X
```

rem -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 0 / 7986

tabulation: Freq. Value
7872 0
114 1

index -- (unlabeled)
type: numeric (float)

range: [2,2] units: 1
unique values: 1 coded missing: 7985 / 7986

tabulation: Freq. Value
1 2

ind -- (unlabeled)
type: numeric (float)

range: [1,2] units: 1
unique values: 2 coded missing: 7802 / 7986

tabulation: Freq. Value
183 1
1 2

ctr -- (unlabeled)
type: numeric (float)

range: [0,1] units: 1
unique values: 2 coded missing: 0 / 7986

tabulation: Freq. Value
7872 0
114 1

```
ctr1 ------------------------------------------------------------- (unlabeled)
               type:  numeric (float)

              range:  [1,3]                    units:  1
      unique values:  3                coded missing:  7880 / 7986

         tabulation:  Freq.  Value
                         97  1
                          7  2
                          2  3


dupimp ---------------------------------------------------------- (unlabeled)
               type:  numeric (float)

              range:  [1,1]                    units:  1
      unique values:  1                coded missing:  7982 / 7986

         tabulation:  Freq.  Value
                          4  1


y2 --------------------------------------------------------------- (unlabeled)
               type:  numeric (float)

              range:  [.,.]                    units:  .
      unique values:  0                coded missing:  7986 / 7986

         tabulation:  Freq.  Value


visit ------------------------------------------------------------ (unlabeled)
               type:  numeric (float)

              range:  [1,25]                   units:  1
      unique values:  25               coded missing:  0 / 7986

               mean:   3.73616
           std. dev:   3.36536

        percentiles:        10%       25%       50%       75%       90%
                              1         1         3         5         8
```

```
vistot ---------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [1,25]                    units:  1
     unique values:  19                coded missing:  0 / 7986

              mean:   6.47245
          std. dev:   4.87356

       percentiles:        10%       25%       50%       75%       90%
                             2         3         5         9        13


sittot ---------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [1,25]                    units:  1
     unique values:  25                coded missing:  0 / 7986

              mean:   3.73616
          std. dev:   3.36536

       percentiles:        10%       25%       50%       75%       90%
                             1         1         3         5         8


sit ------------------------------------------------------------- (unlabeled)
              type:  numeric (float)

             range:  [0,1]                     units:  1
     unique values:  2                 coded missing:  0 / 7986

        tabulation:  Freq.  Value
                      5681  0
                      2305  1
```

```
sitetot ---------------------------------------------------------- (unlabeled)
                  type:  numeric (float)

                 range:  [1,11]                    units:  1
         unique values:  11                coded missing:  0 / 7986

                  mean:   2.40947
              std. dev:   1.41142

           percentiles:        10%       25%       50%       75%       90%
                                 1         1         2         3         4


sittotal --------------------------------------------------------- (unlabeled)
                  type:  numeric (float)

                 range:  [1,11]                    units:  1
         unique values:  11                coded missing:  0 / 7986

                  mean:   2.40947
              std. dev:   1.41142

           percentiles:        10%       25%       50%       75%       90%
                                 1         1         2         3         4


sittotal2 -------------------------------------------------------- (unlabeled)
                  type:  numeric (float)

                 range:  [1,11]                    units:  1
         unique values:  11                coded missing:  0 / 7986

                  mean:   3.82795
              std. dev:   1.58522

           percentiles:        10%       25%       50%       75%       90%
                                 2         2         4         5         5
```

```
_st -------------------------------------------------------------- (unlabeled)
              type:  numeric (byte)

             range:  [1,1]                    units:  1
     unique values:  1                coded missing:  0 / 7986

        tabulation:  Freq.  Value
                      7986  1
_d --------------------------------------------------------------- (unlabeled)
              type:  numeric (byte)

             range:  [0,1]                    units:  1
     unique values:  2                coded missing:  0 / 7986

        tabulation:  Freq.  Value
                      7883  0
                       103  1

_origin ---------------------------------------------------------- (unlabeled)
              type:  numeric (int)

             range:  [7748,12386]             units:  1
     unique values:  637              coded missing:  0 / 7986

              mean:   10689.3
          std. dev:   710.895

       percentiles:        10%       25%       50%       75%       90%
                          9799     10267     10713     11170     11562

_t --------------------------------------------------------------- (unlabeled)
              type:  numeric (double)

             range:  [.00273785,7.6687201]    units:  1.000e-10
     unique values:  1327             coded missing:  0 / 7986

              mean:   1.87636
          std. dev:   1.43829

       percentiles:        10%       25%       50%       75%       90%
                       .503765   .799452   1.41684   2.64476   4.01095
```

```
_t0 -------------------------------------------------------------- (unlabeled)
              type:  numeric (double)

             range:  [0,7.2443532]              units:  1.000e-10
     unique values:  1074                coded missing:  0 / 7986

              mean:   1.25179
          std. dev:   1.38378

       percentiles:        10%       25%       50%       75%       90%
                             0         0   .859685   1.86174   3.32923


freq --------------------------------- The counter for all first records by id
                              and site
              type:  numeric (float)

             range:  [0,1]                      units:  1
     unique values:  2                   coded missing:  0 / 7986

        tabulation:  Freq.  Value
                      5681  0
                      2305  1


AVBHPlus1 ------------------------------------  Availboneheight+1 for log scale
              type:  numeric (float)

             range:  [1,41]                     units:  .01
     unique values:  23                  coded missing:  262 / 7986

              mean:   2.18031
          std. dev:    2.8932

       percentiles:        10%       25%       50%       75%       90%
                             1         1         1     4.125      4.75
```

AVBWPlus1 -------------------------------------- Availbonewidth+1 for log scale
type: numeric (float)

range: [1,36] units: .01
unique values: 35 coded missing: 1 / 7986

mean: 5.0065
std. dev: 7.35246

percentiles:	10%	25%	50%	75%	90%
	1	1	1	4.8	18

implhtPlus1 ------------------------------------- implantheight+1 for log scale
type: numeric (float)

range: [1,38] units: .1
unique values: 23 coded missing: 0 / 7986

mean: 2.03418
std. dev: 2.62501

percentiles:	10%	25%	50%	75%	90%
	1	1	1	3	4

implwdthPlus1 ----------------------------------- implantwidth+1 for log scale
type: numeric (float)

range: [1,46] units: .01
unique values: 28 coded missing: 35 / 7986

mean: 4.55757
std. dev: 6.05446

percentiles:	10%	25%	50%	75%	90%
	1	1	1	9	14

APPENDIX C
ANNOTATIONS FOR FIGURES AND TABLES

Annotations for Figure 1:

Patient Characteristics dataset (otherwise presented as the forma.dta dataset).

desc

Contains data from C:\DATA\forma.dta
obs: 1,462

There are 1,462 records in this dataset which is contains information at the patient level. Each patient entering the study to receive implant(s) is expected to be in this dataset. Therefore, there should be one record per patient. The fact that there are 1,357 patients and 1,462 records indicates that there are multiple records that were ultimately removed from the final analytic dataset.

Placement Dates dataset (otherwise presented as the surgsite.dta dataset)

There are 1,294 patients with 4,313 records and it is and assumed that there is one placement date per implant.

. desc

Contains data from C:\DATA\surgsite.dta
obs: 4,313

A variable freqs1 was temporarily generated to evaluate the number of multiple placement dates (in this dataset the placement date was called nisdate) and hence multiple records that are initially in this dataset.

. gen freqs1=1 if newssn[_n]==newssn[_n+1] & imp1[_n]==imp1[_n+1] & nisdate[_n]~= nisdate[_n+1]
(4234 missing values generated)

Since there are 4,313 records and from this STATA command, there appears to be a total of (4,313-4,234=79) 79 records with multiple placement dates that are different.

. replace freqs=1 if newssn[_n]==newssn[_n+1] & imp1[_n]==imp1[_n+1] & nisdate[_n]==nisdate[_n+1]
(2 real changes made)
The two changes made reflect the number of multiple records that have the same placement date. Therefore, there are 81 total multiple records that ultimately were removed from the final analytic dataset.

Removal Dates dataset (otherwise presented as the evalx.dta dataset)
. desc

Contains data from C:\DATA\evalx.dta
obs: 10,624

This is the original dataset with removal dates. Note that this dataset also has multiple removal dates (in this dataset nevldate was the variable name for evaluation dates) that must be accounted for.

Sorted by: newssn site nevldate nimrdate

A temporary variable, freqsurg, was created to evaluate the number of implants placed in all patients.

. by newssn site:gen freqsurg=nevldate[1]

. by newssn site:replace freqsurg=1 if nevldate==freqsurg
(3485 real changes made)

This variable is accounting for only one placement of and implant. That is to say that if an implant were removed more than once (i.e. multiple removal), it should be verified that it is placed more than once. If there is only one placement date, then subsequent removal dates were removed. The first evaluation date was used for the purpose of establishing "one" implant placement, since there is the possibility of any implant having multiple evaluation dates and hence multiple records. Otherwise the number of implants will be evaluated wrongly as the number of records.

```
. replace freqsurg=0 if freqsurg~=1
(7139 real changes made)

. tabulate freqsurg

   freqsurg |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |       7139       67.20       67.20
          1 |       3485       32.80      100.00
------------+-----------------------------------
      Total |      10624      100.00
```

There are 3485 implants in 1009 patients, not accounting for multiple records for implant removal.

```
. iis newssn

. tis nevldate

. xttab freqsurg

                  Overall             Between            Within
freqsurg|    Freq.  Percent      Freq.  Percent         Percent
----------+----------------------------------------------------
       0   |    7139     67.20        648      64.22         75.13
       1   |    3485     32.80       1009    100.00         32.80
----------+----------------------------------------------------
    Total  | 10624     100.00      1657    164.22         49.36
                                 (n = 1009)
```

The xttab presents the number of first records (3,485 overall) and the remainder (7,139) or those records which are follow-up evaluation dates. The between frequencies and percents reflect the number of implants at the patient level. Therefore, there are 1009 patients with implants placed. There are 648 patients that have ever had freqsurg=0 or with follow-up records.

Analysis dataset (otherwise presented as the final9.dta dataset)

. desc

Contains data from C:\Stata\final9.dta
obs: 7,986

The variable freq was temporarily created to indicate all first records by patient id and site of implant. The tabulation command shows that there are 2,305 total implants. The number of patients with implants is 777 and there are 7,986 total records. This dataset does not contain duplicate records or multiple placement or removal dates. There are 5,681 follow-up records represented in this dataset.

. tabulate freq

```
The counter |
    for all |
      first |
 records by |
id and site |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |       5681       71.14       71.14
          1 |       2305       28.86      100.00
------------+-----------------------------------
      Total |       7986      100.00
```

Percentile results from surgsite "placement" dataset for fig 1.

Contains data from C:\DATA\surgsite.dta
. summarize sitetot1,detail

```
                   sitetot1
-----------------------------------------------------------
     Percentiles      Smallest
 1%           1            1
 5%           1            1
10%           1            1       Obs              4313
25%           1            1       Sum of Wgt.        4313

50%           2                    Mean          2.664503
                   Largest        Std. Dev.      1.809679
75%           4           13
90%           5           14       Variance       3.274939
95%           6           14       Skewness       1.78664
99%          10           14       Kurtosis       7.843539
```

. desc

Contains data from C:\DATA\evalx.dta

Percentile results from evalx "removal" dataset for figure 1.

. summarize sitetot2,detail

```
                   sitetot2
-----------------------------------------------------------
     Percentiles      Smallest
 1%           1            1
 5%           1            1
10%           1            1       Obs             10624
25%           1            1       Sum of Wgt.       10624

50%           2                    Mean          2.817771
                   Largest        Std. Dev.      1.820817
75%           4           14
90%           5           14       Variance       3.315373
95%           6           15       Skewness       1.288991
99%           8           15       Kurtosis       5.160435
```

Contains data from C:\Stata\final9.dta

Percentile results from final9 "analysis" dataset for figure 1.

```
. summarize sittotal2,detail

                        sittotal2
-------------------------------------------------------------
      Percentiles      Smallest
 1%            1             1
 5%            2             1
10%            2             1       Obs                7986
25%            2             1       Sum of Wgt.        7986

50%            4                     Mean           3.827949
                        Largest      Std. Dev.      1.585217
75%            5            11
90%            5            11       Variance       2.512912
95%            6            11       Skewness       .6826816
99%            9            11       Kurtosis       4.263949
```

Annotations for Table 2:
A univariate evaluation of race or ethnicity looking at the between values primarily.

```
. sort id site place followup
iis newid2

. iis id

. tis followup

. xttab  ethw

                  Overall             Between            Within
    ethw |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
       0 |     1037    13.45        112    15.11         100.00
       1 |     6671    86.55        629    84.89         100.00
----------+-----------------------------------------------------
   Total |     7708   100.00        741   100.00         100.00
                                 (n = 741)
```

. xttab ethb

ethb	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
0	6884	89.31	664	89.61	100.00
1	824	10.69	77	10.39	100.00
Total	7708	100.00	741	100.00	100.00
			(n = 741)		

. xttab etha

etha	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
0	7702	99.92	740	99.87	100.00
1	6	0.08	1	0.13	100.00
Total	7708	100.00	741	100.00	100.00
			(n = 741)		

. xttab ethnam

ethnam	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
0	7691	99.78	739	99.73	100.00
1	17	0.22	2	0.27	100.00
Total	7708	100.00	741	100.00	100.00
			(n = 741)		

. xttab ethhis

ethhis	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
0	7599	98.59	719	97.03	100.00
1	109	1.41	22	2.97	100.00
Total	7708	100.00	741	100.00	100.00
			(n = 741)		

```
. xttab  ethoth

                  Overall             Between            Within
ethoth |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        0 |    7700     99.90       739     99.73         100.00
        1 |         8       0.10          2       0.27         100.00
----------+-----------------------------------------------------
    Total |    7708    100.00       741    100.00         100.00
                              (n = 741)

. xttab gender

                   Overall              Between             Within
   gender |    Freq.  Percent      Freq.  Percent         Percent
----------+-----------------------------------------------------
           0 |      63      0.82        10      1.35         100.00
           F |    137     1.78         20      2.70         100.00
          M |  7501   97.31       710    95.82         100.00
           X |       7      0.09          1      0.13         100.00
----------+-----------------------------------------------------
    Total |   7708  100.00      741   100.00         100.00
                              (n = 741)
```

*This generates a minority variable to determine potentially multiple recordings of ethnicity.

```
. gen minor=0 if ethw==1
(1315 missing values generated)

. replace minor=1 if ethb+etha+ethnam+ethhis+ethoth>0
(1242 real changes made)
```

```
. codebook minor
--------------------------------------------------------------------------------
minor                                                                (unlabeled)
--------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,1]                    units:  1
         unique values:  2                    missing .:  75/7986

            tabulation:  Freq.  Value
                          6669  0
                          1242  1
                            75  .
. *There are 75 records involved with a missing minor variable.
. list id if minor==.
      |   id |
1218. |  241 |
1219. |  241 |
1220. |  241 |
1221. |  241 |
1222. |  245 |
      |------|
1223. |  245 |
1224. |  245 |
1225. |  245 |
1346. |  299 |
1347. |  299 |
      |------|
1348. |  299 |
1349. |  299 |
1471. |  344 |
1472. |  344 |
1473. |  344 |
      |------|
1474. |  344 |
1475. |  344 |
1476. |  344 |
1477. |  344 |
1478. |  344 |
```

1479.	344
1480.	344
1481.	344
1482.	344
1483.	344
1484.	344
1966.	455
1967.	455
1968.	455
1969.	455
1970.	455
1971.	455
3280.	654
3281.	654
3282.	654
3283.	654
3284.	654
3285.	654
3286.	654
3287.	654
3288.	654
3289.	654
3290.	654
3291.	654
3292.	654
3293.	654
3294.	654
3295.	654
3296.	654
3297.	654
3298.	654
3299.	654
3300.	654
3301.	654
3302.	654

```
      |------|
3303. |  654 |
3304. |  654 |
3305. |  654 |
3306. |  654 |
3307. |  654 |
      |------|
6045. | 1089 |
6236. | 1157 |
6237. | 1157 |
6238. | 1157 |
6239. | 1157 |
      |------|
6240. | 1157 |
6241. | 1157 |
6242. | 1157 |
6243. | 1157 |
6244. | 1157 |
      |------|
6245. | 1157 |
7916. | 1466 |
7917. | 1466 |
7918. | 1466 |
7919. | 1466 |
      +------+

. * this involves nine patients.
. list eth* if id==241
      +---------------------------------------------+
      | ethw   ethb   etha   ethnam   ethhis   ethoth |
      |---------------------------------------------|
1218. |    0      0      0        0        0        0 |
1219. |    0      0      0        0        0        0 |
1220. |    0      0      0        0        0        0 |
1221. |    0      0      0        0        0        0 |
      +---------------------------------------------+

. list eth* if id==245
      +---------------------------------------------+
      | ethw   ethb   etha   ethnam   ethhis   ethoth |
      |---------------------------------------------|
1222. |    0      0      0        0        0        0 |
1223. |    0      0      0        0        0        0 |
1224. |    0      0      0        0        0        0 |
1225. |    0      0      0        0        0        0 |
```

```
     +-----------------------------------------------+


. list eth* if id==299

     +-----------------------------------------------+
     | ethw   ethb   etha   ethnam   ethhis   ethoth |
     |-----------------------------------------------|
1346. |    0      0      0        0        0        0 |
1347. |    0      0      0        0        0        0 |
1348. |    0      0      0        0        0        0 |
1349. |    0      0      0        0        0        0 |
     +-----------------------------------------------+

. list eth* if id==344

     +-----------------------------------------------+
     | ethw   ethb   etha   ethnam   ethhis   ethoth |
     |-----------------------------------------------|
1471. |    0      0      0        0        0        0 |
1472. |    0      0      0        0        0        0 |
1473. |    0      0      0        0        0        0 |
1474. |    0      0      0        0        0        0 |
1475. |    0      0      0        0        0        0 |
     |-----------------------------------------------|
1476. |    0      0      0        0        0        0 |
1477. |    0      0      0        0        0        0 |
1478. |    0      0      0        0        0        0 |
1479. |    0      0      0        0        0        0 |
1480. |    0      0      0        0        0        0 |
     |-----------------------------------------------|
1481. |    0      0      0        0        0        0 |
1482. |    0      0      0        0        0        0 |
1483. |    0      0      0        0        0        0 |
1484. |    0      0      0        0        0        0 |
     +-----------------------------------------------+

. list eth* if id==455

     +-----------------------------------------------+
     | ethw   ethb   etha   ethnam   ethhis   ethoth |
     |-----------------------------------------------|
1966. |    0      0      0        0        0        0 |
1967. |    0      0      0        0        0        0 |
1968. |    0      0      0        0        0        0 |
1969. |    0      0      0        0        0        0 |
1970. |    0      0      0        0        0        0 |
     |-----------------------------------------------|
1971. |    0      0      0        0        0        0 |
```

```
     +--------------------------------------------+

. list eth* if id==654

     +--------------------------------------------+
     | ethw   ethb   etha   ethnam   ethhis   ethoth |
     |--------------------------------------------|
3280. |   0      0      0       0       0       0 |
3281. |   0      0      0       0       0       0 |
3282. |   0      0      0       0       0       0 |
3283. |   0      0      0       0       0       0 |
3284. |   0      0      0       0       0       0 |
     |--------------------------------------------|
3285. |   0      0      0       0       0       0 |
3286. |   0      0      0       0       0       0 |
3287. |   0      0      0       0       0       0 |
3288. |   0      0      0       0       0       0 |
3289. |   0      0      0       0       0       0 |
     |--------------------------------------------|
3290. |   0      0      0       0       0       0 |
3291. |   0      0      0       0       0       0 |
3292. |   0      0      0       0       0       0 |
3293. |   0      0      0       0       0       0 |
3294. |   0      0      0       0       0       0 |
     |--------------------------------------------|
3295. |   0      0      0       0       0       0 |
3296. |   0      0      0       0       0       0 |
3297. |   0      0      0       0       0       0 |
3298. |   0      0      0       0       0       0 |
3299. |   0      0      0       0       0       0 |
     |--------------------------------------------|
3300. |   0      0      0       0       0       0 |
3301. |   0      0      0       0       0       0 |
3302. |   0      0      0       0       0       0 |
3303. |   0      0      0       0       0       0 |
3304. |   0      0      0       0       0       0 |
     |--------------------------------------------|
3305. |   0      0      0       0       0       0 |
3306. |   0      0      0       0       0       0 |
3307. |   0      0      0       0       0       0 |
     +--------------------------------------------+
```

```
. list eth* if id==1089

      +---------------------------------------------+
      | ethw   ethb   etha   ethnam   ethhis   ethoth |
      |---------------------------------------------|
6045. |    0      0      0        0        0        0 |
      +---------------------------------------------+

. list eth* if id==1157

      +---------------------------------------------+
      | ethw   ethb   etha   ethnam   ethhis   ethoth |
      |---------------------------------------------|
6236. |    0      0      0        0        0        0 |
6237. |    0      0      0        0        0        0 |
6238. |    0      0      0        0        0        0 |
6239. |    0      0      0        0        0        0 |
6240. |    0      0      0        0        0        0 |
      |---------------------------------------------|
6241. |    0      0      0        0        0        0 |
6242. |    0      0      0        0        0        0 |
6243. |    0      0      0        0        0        0 |
6244. |    0      0      0        0        0        0 |
6245. |    0      0      0        0        0        0 |
      +---------------------------------------------+

. list eth* if id==1466

      +---------------------------------------------+
      | ethw   ethb   etha   ethnam   ethhis   ethoth |
      |---------------------------------------------|
7916. |    0      0      0        0        0        0 |
7917. |    0      0      0        0        0        0 |
7918. |    0      0      0        0        0        0 |
7919. |    0      0      0        0        0        0 |
      +---------------------------------------------+

. *All ethnicity values reveal 0's and do not report a missing value and do
not report any ethnicity.
. count if ethhis==1 & ethw==1
   2
```

```
. *There are two records having multiple recordings of ethw and ethhis.
. list id if ethhis==1 & ethw==1

     +-----+
     |  id |
     |-----|
953. | 183 |
954. | 183 |
     +-----+

. list eth* if id==183

     +----------------------------------------------+
     | ethw   ethb   etha   ethnam   ethhis   ethoth |
     |----------------------------------------------|
953. |    1      0      0        0        1        0 |
954. |    1      0      0        0        1        0 |
     +----------------------------------------------+

--------------------------------------------------------------------------------
gender                                                              numeric sex
--------------------------------------------------------------------------------

                  type:  numeric (long)
                 label:  gender

                 range:  [1,6]                        units:  1
         unique values:  4                        missing .:  278/7986

            tabulation:  Freq.   Numeric  Label
                            63         1  0
                           137         3  F
                          7501         4  M
                             7         6  X
                           278         .

. count if gender==3 & gender==4
    0

. *There are no patients listed in both categories of gender.
. codebook sitetot
```

Annotations for Table 3 and Table 4:

A variable "freq" was created to count the first records for each patient and each site. The xttab command for the overall frequency and percent calculations accounts for all records and this includes all follow-up records. This inflates the value for implants.

```
. by id site:gen freq=followup[1]
. by id site:replace freq=1 if followup==freq
(2305 real changes made)
. replace freq=0 if freq~=1
(5681 real changes made)
. tabulate freq
       freq |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |       5681       71.14       71.14
          1 |       2305       28.86      100.00
------------+-----------------------------------
      Total |       7986      100.00
```

It appears that there are 2305 total implants in 777 patients.
An xttab procedure on the records representing the first placement date produces the following:

```
. iis id
. tis followup
. xttab  imptype if freq==1

                  Overall             Between            Within
  imptype |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        2 |      13      0.56         5      0.64         100.00
        4 |      95      4.12        45      5.79          85.59
        5 |       4      0.17         1      0.13         100.00
        6 |       6      0.26         2      0.26          60.00
       10 |       2      0.09         1      0.13          33.33
       18 |      16      0.69         4      0.51          94.12
       19 |    2169     94.10       725     93.31          99.45
----------+-----------------------------------------------------
    Total |    2305    100.00       783    100.77          98.44
                                 (n = 777)
```

Here the "Between frequency" and "percent" have not changed. However, the "Overall Frequency" and "Percent" changed because the records of followup were not counted.

. by imptype:xttab failure if freq==1

_
```
-> imptype = 2

                 Overall             Between            Within
  failure |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        0 |       13    100.00          5    100.00         100.00
----------+-----------------------------------------------------
    Total |       13    100.00          5    100.00         100.00
                                  (n = 5)
```

```
imptype = 4

                 Overall             Between            Within
  failure |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        0 |       93     97.89         45    100.00          97.89
        1 |        2      2.11          1      2.22          50.00
----------+-----------------------------------------------------
    Total |       95    100.00         46    102.22          96.85
                                  (n = 45)
```

```
imptype = 5

                 Overall             Between            Within
  failure |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        0 |        4    100.00          1    100.00         100.00
----------+-----------------------------------------------------
    Total |        4    100.00          1    100.00         100.00
                                  (n = 1)
```

imptype = 6

failure	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
0	6	100.00	2	100.00	100.00
Total	6	100.00	2	100.00	100.00
			(n = 2)		

imptype = 10

failure	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
0	2	100.00	1	100.00	100.00
Total	2	100.00	1	100.00	100.00
			(n = 1)		

imptype = 18

failure	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
0	16	100.00	4	100.00	100.00
Total	16	100.00	4	100.00	100.00
			(n = 4)		

imptype = 19

failure	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
0	2130	98.20	717	98.90	98.84
1	39	1.80	28	3.86	41.94
Total	2169	100.00	745	102.76	96.70
			(n = 725)		

This was the xttab procedure using only the first records, (i.e. freq=1) subset of patients.
The “Between Frequencies” are reported in the table as the “Number of Failures” by type. The “Overall Frequencies” are reported in the table as the “Number of Patients With Failures”.

Annotations for Figure 4

```
. iis newid2
. tis followup
. xttab  sittotal2
```

sittotal2	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
1	288	3.61	116	5.03	100.00
2	1815	22.73	564	24.47	100.00
3	779	9.75	261	11.32	100.00
4	2720	34.06	644	27.94	100.00
5	1654	20.71	475	20.61	100.00
6	346	4.33	126	5.47	100.00
7	125	1.57	49	2.13	100.00
8	149	1.87	40	1.74	100.00
9	89	1.11	9	0.39	100.00
10	10	0.13	10	0.43	100.00
11	11	0.14	11	0.48	100.00
Total	7986	100.00	2305	100.00	100.00
			(n = 2305)		

```
. iis id

. tis site

. xttab  sittotal2
```

sittotal2	Overall Freq.	Overall Percent	Between Freq.	Between Percent	Within Percent
1	288	3.61	116	14.93	100.00
2	1815	22.73	282	36.29	100.00
3	779	9.75	87	11.20	100.00
4	2720	34.06	161	20.72	100.00
5	1654	20.71	95	12.23	100.00
6	346	4.33	21	2.70	100.00
7	125	1.57	7	0.90	100.00
8	149	1.87	5	0.64	100.00
9	89	1.11	1	0.13	100.00
10	10	0.13	1	0.13	100.00
11	11	0.14	1	0.13	100.00
Total	7986	100.00	777	100.00	100.00
			(n = 777)		

The between freq/percent using id as an iis variable is appropriate for assessing the number of "patients" with implants of the site total indexed. That is to say that there are 116 patients with one implant and there are 95 patients with 5 implants. However, the overall frequency and percent assess the follow-up times or visits for each implant and is not useful information to present. When iis is newid2, the between frequency and percent are a true assessment of the number of implants for patients at the various site total levels.

Annotations for Table 5

```
. iis id
. tis followup
. xttab visit
                  Overall              Between            Within
    visit |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     2305     28.86        777    100.00          28.86
        2 |     1544     19.33        548     70.53          21.07
        3 |     1102     13.80        394     50.71          16.87
        4 |      777      9.73        279     35.91          13.73
        5 |      563      7.05        196     25.23          11.82
        6 |      430      5.38        146     18.79          10.47
        7 |      316      3.96        105     13.51           9.27
        8 |      244      3.06         80     10.30           8.52
        9 |      181      2.27         63      8.11           7.54
       10 |      133      1.67         44      5.66           6.98
       11 |      109      1.36         37      4.76           6.48
       12 |       69      0.86         25      3.22           5.96
       13 |       51      0.64         19      2.45           5.43
       14 |       31      0.39         12      1.54           5.00
       15 |       30      0.38         11      1.42           4.95
       16 |       19      0.24          8      1.03           4.17
       17 |       19      0.24          8      1.03           4.17
       18 |       10      0.13          5      0.64           3.47
       19 |       10      0.13          5      0.64           3.47
       20 |       10      0.13          5      0.64           3.47
       21 |        9      0.11          4      0.51           3.60
       22 |        9      0.11          4      0.51           3.60
       23 |        9      0.11          4      0.51           3.60
       24 |        3      0.04          1      0.13           3.16
       25 |        3      0.04          1      0.13           3.16
----------+-----------------------------------------------------
    Total |     7986    100.00       2781    357.92          18.52
                                  (n = 777)
```

. *This attempts to evalutate visits per patient. It appears that naturally
there is at least one visit and therefore this would lend well to being the highest value.
Also the greater the visit frequency the less patients involved. The
minimum value is one and the maximum is 25.
The overall freq/percent reveal values for those who have had visits and
therefore is cumulative. This could be presented as "There are 146 patients who
have ever had six (6) visits." There are patients who are in this category that have also been
in the category of those who have had seven (7) implants. Therefore, the same patients who have been counted for the six (6) visits are also in the seven visit category.

APPENDIX D
ANNOTATIONS AND PROGRAMS FOR ANALYSIS

Program for Analyses producing Tables for Models 1 through 11 and Tables 6-19

```
use "C:\unzipped\final9folder\final9.dta", clear
sort newid2 site place followup
*The final9 data set that has been stset for continuous time survival
analysis.
desc
*note the number of observations being 7,986
codebook race race2 type loc
*The race variable was created with the following value labels:
*1-white, 2-black, 3-asian, 4-native american, 5-hispanic, 6-other.
*I need to account for the missing values.
*There are 353 missing values that are maintained
*The race variable was then changed to collapse the cells to the following
value labels for the variable race2:
*1-white, 2-black+asian+native american+hispanic, 3-other, missing.
*In my analysis for discrete survival and continuous survival I found that
the number of failures for each category was sparse and therefore further
collapsed the cells and created a race3 variable.
/*gen race3=1 if race2==1 & race2~=.
codebook race2 race3
replace race3=2 if race2==2|race2==3 & race2~=.*/
codebook race2 race3
*This will present the value labels for race3 as 1-white and 2-other and
missing.
*Now I will also collapse the cells for the loc (location variable) into four
instead of 6 cells.
/*gen loc2=1 if loc==5 & loc~=.
codebook loc loc2
replace loc2=2 if loc==2 & loc~=.
codebook loc loc2
replace loc2=3 if loc==4|loc==6 & loc~=.
codebook loc loc2
replace loc2=4 if loc==1|loc==3 & loc~=.
codebook loc loc2*/
tab failure loc
*Clearly the higher failure frequencies occur in the loc==5 and 2 regions
which influenced the value labels in the loc2 variable which designates the 1
and 2 values as these to regions.
*The value labels for loc2 are as follows:
*1-mandibular anterior region, 2-maxillary anterior region, 3-mandibular
posterior region, 4-maxillary posterior region.
```

```
/*save "C:\unzipped\final9Folder\final9.dta", replace*/
sort failure
tab failure loc2
by failure:xttab loc2
by failure:xttab type2
by failure:xttab race3
sort id site place followup
*Here in the loc2 variable the failures do not increase but the frequency is
higher in the four cells which may present an analysis with less of an issue
regarding "perfect predictors" due to sparse failure counts.
*Now I will attempt to analyze the data using the six models discussed in my
thesis.
*(A) Single site per person and single time interval
*I need to limit my evaluation to the first year in the study and that means
that [first year of follow up-place (placement date of implant)] needs to be
indicated.
*I also must assure that the censoring variable is maintained.
sort id site place followup
/*by id:gen year=1 if followup-place<=365.25
replace year=2 if followup-place<=2*365+0.25 & year~=1
replace year=3 if followup-place<=3*365+0.25 & year~=2
replace year=4 if followup-place<=4*365+0.25 & year~=3
replace year=5 if followup-place<=5*365+0.25 & year~=4
replace year=6 if followup-place<=6*365+0.25 & year~=5
replace year=7 if followup-place<=7*365+0.25 & year~=6
replace year=8 if followup-place<=8*365+0.25 & year~=7
codebook year*/
codebook year
*Now I will attempt a logistic regression for a single site per person and
single time.  I will also incorporate in the code a variable called firstimp
to indicate the first record and only include this implant for evaluation.
*The code used to generate such a variable follows.
/*firstimp=0
by id:gen firstimp=1 if site==site[1]*/
codebook firstimp
*firstimp indicates implant first records at an implant level
list newid2 id site place followup firstimp failure in 1/72
*(A) Single site per person and single time interval (first year of study).
count if firstimp==1 & year==1
xi:logistic failure i.loc i.type i.race if firstimp==1 & year==1
*I will use the other variables that were created to evaluate this model.
xi:logistic failure i.loc2 i.type i.race3 if firstimp==1 & year==1
testparm _Iloc*
*Type is still an issue.
codebook type
/*gen type2=1 if type=1 and type~=.
gen type2=1 if type==1 and type~=.
gen type2=1 if type==1 & type~=.
codebook type type2
replace type2=2 if type==2|type==3 & type~=.
codebook type type2
save "C:\STATA\final9.dta", replace*/
xi:logistic failure i.loc2 i.type2 i.race3 if firstimp==1 & year==1
```

```
*The type variable no longer is presented as being problematic.
testparm _Iloc*
*I will now incorporate the robust variance into the analysis.
xi:logistic failure i.loc2 i.type2 i.race3 if firstimp==1 & year==1, robust
cluster(id)
testparm _Iloc*
*(B) The next model evaluates the situation for Multiple sites per person and
a single time interval.
count if year==1
xi:logistic failure i.loc2 i.type2 i.race3 if year==1
testparm _Iloc*
*This does not incorporate the robust variance
xi:logistic failure i.loc2 i.type2 i.race3 if year==1,robust cluster(id)
testparm _Iloc*
*Now using the xt command structure in Stata to evaluate GEE
set matsize 80
xi:xtgee failure i.loc2 i.type2 i.race3 if year==1, family(bin) link(logit)
corr(exc) i(id) eform
testparm _Iloc*
*I will now analyze the data using the robust variance analysis.
xi:xtgee failure i.loc2 i.type2 i.race3 if year==1, family(bin) link(logit)
corr(exc) i(id) eform
testparm _Iloc*
*These are the population averaged models.  The second model incorporating
the robust variance procedure.
*The next situation to evaluate is the single site per person with multiple
time intervals.  In this situation I must reorganize the data to accommodate
one record per person per time interval.  I will use the stsplit command in
Stata where time intervals will be established and the observation level will
change according to this.  I do not want to change this dataset and will use
another dataset that has been established for this analysis (final9b.dta).
clear
use "C:\unzipped\final9Folder\final9b.dta", clear
desc
sort id site place followup
*This data was modified using the stsplit command to create a categorical
variable called annualt which separates the data into yearly time intervals
and allows for discrete survival analysis.  Also the analysis requires that
other variables be created: (A) one to index each patient (B) a binary
dependent variable to indicate censorship within the time intervals, and (C)
a variable to summarize the pattern of duration dependence.
*The data in it's current form has more observations than the final9.dta.
Also, the race loc and type variables need to be collapsed as well.
codebook   race race2 type loc
/*Code for generating race3, loc2 and type2 variables
gen race3=1 if race2==1 &  race2~=.
codebook   race2 race3
replace race3=2 if race2==2|race2==3 &  race2~=.
codebook   race2 race3
save "C:\unzipped\final9Folder\final9b.dta", replace
gen  type2=1 if type==1 & type~=.
codebook type type2
replace type2=2 if type==2|type==3 & type~=.
codebook type type2
```

```
save "C:\unzipped\final9Folder\final9b.dta", replace
gen loc2=1 if loc==5 & loc~=.
codebook loc loc2
replace loc2=2 if loc==2 & loc~=.
codebook loc loc2
replace loc2=3 if loc==4|loc==6 & loc~=.
codebook loc loc2
replace loc2=4 if loc==1|loc==3 & loc~=.
codebook loc loc2*/
codebook annualt
*I need to create another censorship indicator
codebook _d
/*gen dfail=_d*/
codebook dfail
*I will also need to code a variable to indicate the first record in this
dataset "firstimp" and only include this implant for evaluation.
*The code used to generate such a variable follows.
/*gen firstimp=0
by id:gen firstimp=1 if site==site[1]*/
list id site place followup annualt firstimp in 1/72
*firstimp indicates implant first records at an implant level.
tabulate failure annualt
*(C)Discrete Proportional Odds model.
*This is the situation of the single site per patient with multiple time
intervals.
count if firstimp==1
xi:logit dfail i.annualt i.loc2 i.type2 i.race3 if firstimp==1
logit, or
testparm _Iloc*
testparm _Iannualt*
xi:logit dfail i.annualt i.loc2 i.type2 i.race3 if firstimp==1, robust
cluster(id)
logit, or
testparm _Iloc*
testparm _Iannualt*
*Now to evaluate the Discrete proportional hazards using the cloglog function
xi:cloglog dfail i.annualt i.loc2 i.type2 i.race3 if firstimp==1
matrix b=e(b)
matrix v=e(V)
ereturn post b v
ereturn display, eform(exp_b)
testparm _Iloc*
testparm _Iannualt*
xi:cloglog dfail i.annualt i.loc2 i.type2 i.race3 if firstimp==1,robust
cluster(id)
matrix b=e(b)
matrix v=e(V)
ereturn post b v
ereturn display, eform(exp_b)
testparm _Iloc*
testparm _Iannualt*
*Next I will evaluate the situation of multiple sites per patient and
multiple time intervals.  This will be a Discrete Proportional Odds model.
```

```
*(D) Multiple sites per person and multiple time intervals
*Discrete Proportional Odds.
xi:logit dfail i.annualt i.loc2 i.type2 i.race3
logit,or
testparm _Iloc*
testparm _Iannualt*
*Next I will incorporate the robust variance
xi:logit dfail i.annualt i.loc2 i.type2 i.race3, robust cluster(id)
logit, or
testparm _Iloc*
testparm _Iannualt*
*Now we will analyze the data using GEE for the Discrete Proportional Odds
and Cloglog Models.
set matsize 110
xi:xtgee dfail i.annualt i.loc2 i.type2 i.race3, family(bin) link(logit)
corr(exc) i(id) eform
testparm _Iloc*
testparm _Iannualt*
xi:xtgee dfail i.annualt i.loc2 i.type2 i.race3, family(bin) link(logit)
corr(exc) i(id) eform robust
testparm _Iloc*
testparm _Iannualt*
*Now for the Cloglog evaluation of GEE
xi:xtcloglog dfail i.annualt i.loc2 i.type2 i.race3, pa i(id)
testparm _Iloc*
testparm _Iannualt*
matrix b=e(b)
matrix v=e(V)
ereturn post b v
ereturn display, eform(exp_b)
xi:xtcloglog dfail i.annualt i.loc2 i.type2 i.race3, pa robust i(id)
testparm _Iloc*
testparm _Iannualt*
matrix b=e(b)
matrix v=e(V)
ereturn post b v
ereturn display, eform(exp_b)
*The next situation to evaluate involves the single site per patient with
continuous time.  This would involve the Cox model and I will now evaluate
this on the final9.dta dataset.
/*save "C:\unzipped\final9Folder\final9b.dta", replace
clear*/
use "C:\unzipped\final9Folder\final9.dta", clear
desc
*(E)Single site per patient and continuous time Cox Proportional Hazards
Model
count if firstimp==1
xi: stcox  i.loc2 i.type2 i.race3 if firstimp==1
testparm _Iloc*
xi: stcox  i.loc2 i.type2 i.race3 if firstimp==1, robust cluster(id)
testparm _Iloc*
*The next situation to evaluate involves multiple sites per patient with
continuous time.  This also would involve the Cox model.
```

```
*(F) Multiple sites per patient and continuous time Cox Proportional Hazards
Model
xi: stcox  i.loc2 i.type2 i.race3
testparm _Iloc*
xi: stcox  i.loc2 i.type2 i.race3 , robust cluster(id)
testparm _Iloc*
xi: stcox  i.loc2 i.type2 i.race3 , shared(id)
end of log
```

Log file from Analysis Program:

```
----------------------------------------------------------------------------
       log:  C:\DATA\aug1b2004.smcl
  log type:  smcl

. do c:\stata\aug1ed2004.txt

. use "C:\unzipped\final9folder\final9.dta", clear

. sort newid2 site place followup

. *The final9 data set that has been stset for continuous time survival
analysis.
. desc

Contains data from C:\unzipped\final9folder\final9.dta
  obs:         7,986
 vars:           121
 size:     4,016,958 (68.1% of memory free)
----------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
----------------------------------------------------------------------------
implwdthPlus1   float  %9.0g                   implantwidth+1 for log scale
nxdate          float  %d                     numeric xdate
nbdate          float  %d                     numeric bdate
surocc9         float  %9.0g                  Unable to Seat Implant
surocc10        float  %9.0g                  Implant Not Well Adapted to
Site
surocc11        float  %9.0g                  Ridge Augmentation Used
surocc12        float  %9.0g                  Periodontal Tissue Damage
surocc13        float  %9.0g                  Patient Experienced Pain
surocc14        float  %9.0g                  Excessive Bleeding
surocc15        float  %9.0g                  Guided Tissue Regeneration
surocc16        float  %9.0g
_merge          byte   %8.0g
age1            float  %9.0g
newid2          float  %9.0g                  group(id site)
y               float  %9.0g
id              double %9.0g                  id
nisdate         double %d                     isdate
place           float  %d
nevldate        double %d                     evaldate
followup        float  %d
nimrdate        double %d                     imprdate
```

```
site            double %9.0g                 site
failure         float  %9.0g
rownames        str5   %5s
evaldate        str11  %11s
mobil           float  %9.0g
periminf        str1   %1s                   Peri-implant Inflammation
imphcat         float  %9.0g                 Implant Health Category
imprdate        str11  %11s
impfunc         float  %9.0g                 Implant Functionality
impltopt        float  %9.0g                 Implant Less Than Optimal but
                                               Functional
impnonsp        float  %9.0g
imp2brmv        float  %9.0g
funother        float  %9.0g
painlswr        float  %9.0g
esthetic        float  %9.0g
mastprob        float  %9.0g                 Mastication Problems Due to
                                               Implant
speechpr        float  %9.0g                 Speech Problems Due to Implant
cmplnoth        float  %9.0g
compimpi        float  %9.0g                 If Compromised Implant
                                               Intervention Was Used
newid           float  %9.0g                 group(newssn)
isdate          str11  %11s
imparch         str1   %1s                   Arch Location
imp1            float  %9.0g
imptype         float  %9.0g                 Implant Type
matcode         float  %9.0g                 Material Code
coatcode        float  %9.0g                 Coating Code
stagecode       float  %9.0g                 Stage Code
morphcode       float  %9.0g                 Morphology Code
implantheight   float  %9.0g                 Implant Height (mm)
implantwidth    float  %9.0g                 Implant Width/Diameter (mm)
availboneheight float  %9.0g                 Height of Available Bone (mm)
availbonewidth  float  %9.0g                 Width of Available Bone (mm)
avboneht        float  %9.0g
avbonewi        float  %9.0g
attginwi        float  %9.0g
bonclass        float  %9.0g                 Bone Classification
surocc1         float  %9.0g                 Implant Altered
surocc2         float  %9.0g                 Alveolar Ridge Perforation
surocc3         float  %9.0g                 Jaw Fracture
surocc4         float  %9.0g                 Neurological Damage
surocc5         float  %9.0g                 Inferior Mandibular Border
                                               Perforation
surocc6         float  %9.0g                 Sinus Lift
surocc7         float  %9.0g                 Perforated Sinus/Nasal Cavity
surocc8         float  %9.0g                 Equipment Complications
provid          str4   %4s
station         float  %9.0g
ethw            float  %9.0g
ethb            float  %9.0g
etha            float  %9.0g
ethnam          float  %9.0g
```

```
ethhis          float  %9.0g
ethoth          float  %9.0g
sex             str1   %1s
asarate         float  %9.0g
edenttot        float  %9.0g
gender          long   %8.0g      gender      numeric sex
rem             float  %9.0g
ind             float  %9.0g
ctr             float  %9.0g
ctr1            float  %9.0g
dupimp          float  %9.0g
y2              float  %9.0g
visit           float  %9.0g
vistot          float  %9.0g                  Total number of Visits
sittot          float  %9.0g
sit             float  %9.0g
sitetot         float  %9.0g
sittotal        float  %9.0g
sittotal2       float  %9.0g                  Total number of sites per
                                                patient
freq            float  %9.0g                  The counter for all first
                                                records by id and site
AVBHPlus1       float  %9.0g                   Availboneheight+1 for log
scale
AVBWPlus1       float  %9.0g                  Availbonewidth+1 for log scale
implhtPlus1     float  %9.0g                  implantheight+1 for log scale
age2            float  %9.0g
agecat          float  %9.0g
arch            float  %9.0g
_st             byte   %8.0g
_d              byte   %8.0g
_origin         int    %10.0g
_t              double %10.0g
_t0             double %10.0g
failind         byte   %8.0g
seq             float  %9.0g
firstrec        float  %9.0g
race            float  %9.0g
race2           float  %9.0g
type            float  %9.0g
loc             float  %9.0g
race3           float  %9.0g
loc2            float  %9.0g
type2           float  %9.0g
index           float  %9.0g
folupind        float  %9.0g
maxfolup        byte   %10.0g
maxind          float  %9.0g
_Iloc2_2        byte   %8.0g                  loc2==2
_Iloc2_3        byte   %8.0g                  loc2==3
_Iloc2_4        byte   %8.0g                  loc2==4
year            float  %9.0g
firstimp        float  %9.0g
. *note the number of observations being 7,986
. codebook race race2 type loc
```

```
race
(unlabeled)
-----------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,6]                        units:  1
         unique values:  6                        missing .:  353/7986


           tabulation:  Freq.  Value
                         6669  1
                          824  2
                            6  3
                           17  4
                          109  5
                            8  6
                          353  .

-----------------------------------------------------------------------------
race2
(unlabeled)
-----------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,3]                        units:  1
         unique values:  3                        missing .:  353/7986

           tabulation:  Freq.  Value
                         6669  1
                          956  2
                            8  3
                          353  .

-----------------------------------------------------------------------------
type
(unlabeled)
-----------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,3]                        units:  1
         unique values:  3                        missing .:  0/7986

           tabulation:  Freq.  Value
                         7398  1
                          466  2
                          122  3
```

```
------------------------------------------------------------------------------
loc
(unlabeled)
------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,6]                        units:  1
         unique values:  6                        missing .:  0/7986

            tabulation:  Freq.  Value
                           208  1
                           415  2
                           209  3
                           986  4
                          5238  5
                           930  6

. *The race variable was created with the following value labels:
. *1-white, 2-black, 3-asian, 4-native american, 5-hispanic, 6-other.
. *I need to account for the missing values.
. *There are 353 missing values that are maintained
. *The race variable was then changed to collapse the cells to the following
value labels for the variable race2:
. *1-white, 2-black+asian+native american+hispanic, 3-other, missing.
. *In my analysis for discrete survival and continuous survival I found that
the number of failures for each category was sparse and therefore further
collapsed the cells and created a race3 variable.
. /*gen race3=1 if race2==1 & race2~=.
> codebook race2 race3
> replace race3=2 if race2==2|race2==3 & race2~=.*/
. codebook race2 race3

------------------------------------------------------------------------------
race2
(unlabeled)
------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,3]                        units:  1
         unique values:  3                        missing .:  353/7986

            tabulation:  Freq.  Value
                          6669  1
                           956  2
                             8  3
                           353  .
```

```
------------------------------------------------------------------------------
race3
(unlabeled)
------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,2]                        units:  1
         unique values:  2                        missing .:  353/7986

            tabulation:  Freq.  Value
                          6669  1
                           964  2
                           353  .

. *This will present the value labels for race3 as 1-white and 2-other and
missing.
. *Now I will also collapse the cells for the loc (location variable) into
four instead of 6 cells.
. /*gen loc2=1 if loc==5 & loc~=.
> codebook loc loc2
> replace loc2=2 if loc==2 & loc~=.
> codebook loc loc2
> replace loc2=3 if loc==4|loc==6 & loc~=.
> codebook loc loc2
> replace loc2=4 if loc==1|loc==3 & loc~=.
> codebook loc loc2*/

. tab failure loc

           |                     loc

   failure |         1          2          3          4 |     Total
-----------+--------------------------------------------+----------
         0 |       200        399        205        975 |     7,883
         1 |         8         16          4         11 |       103
-----------+--------------------------------------------+----------
     Total |       208        415        209        986 |     7,986


           |          loc
   failure |         5          6 |     Total
-----------+----------------------+----------
         0 |     5,190        914 |     7,883
         1 |        48         16 |       103
-----------+----------------------+----------
     Total |     5,238        930 |     7,986

. *Clearly the higher failure frequencies occur in the loc==5 and 2 regions
which influenced the value labels in the loc2 variable which designates the 1
and 2 values as these to regions.
. *The value labels for loc2 are as follows:
. *1-mandibular anterior region, 2-maxillary anterior region, 3-mandibular
posterior region, 4-maxillary posterior region.
```

```
. /*save "C:\unzipped\final9Folder\final9.dta", replace*/
. sort failure

. tab failure loc2

           |                        loc2
   failure |         1          2          3          4 |     Total
-----------+--------------------------------------------+----------
         0 |     5,190        399      1,889        405 |     7,883
         1 |        48         16         27         12 |       103
-----------+--------------------------------------------+----------
     Total |     5,238        415      1,916        417 |     7,986

. by failure:xttab loc2
```

```
-> failure = 0

                Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     5190     65.84      1317     58.17        100.00
        2 |      399      5.06       174      7.69        100.00
        3 |     1889     23.96       616     27.21        100.00
        4 |      405      5.14       157      6.93        100.00
----------+-----------------------------------------------------
    Total |     7883    100.00      2264    100.00        100.00
                               (n = 2264)

-----------------------------------------------------------------------------
-> failure = 1

                Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       48     46.60        48     46.60        100.00
        2 |       16     15.53        16     15.53        100.00
        3 |       27     26.21        27     26.21        100.00
        4 |       12     11.65        12     11.65        100.00
----------+-----------------------------------------------------
    Total |      103    100.00       103    100.00        100.00
                                (n = 103)
. by failure:xttab type2

-----------------------------------------------------------------------------
-> failure = 0

                Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     7305     92.67      2130     94.08        100.00
        2 |      578      7.33       134      5.92        100.00
----------+-----------------------------------------------------
    Total |     7883    100.00      2264    100.00        100.00
                               (n = 2264)
```

```
------------------------------------------------------------------------------
-> failure = 1

                 Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       93     90.29         93     90.29         100.00
        2 |       10      9.71         10      9.71         100.00
----------+-----------------------------------------------------
    Total |      103    100.00        103    100.00         100.00
                               (n = 103)

. by failure:xttab race3

------------------------------------------------------------------------------
-> failure = 0

                 Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     6588     87.48       1859     87.15         100.00
        2 |      943     12.52        274     12.85         100.00
----------+-----------------------------------------------------
    Total |     7531    100.00       2133    100.00         100.00
                               (n = 2133)

------------------------------------------------------------------------------
-> failure = 1

                 Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       81     79.41         81     79.41         100.00
        2 |       21     20.59         21     20.59         100.00
----------+-----------------------------------------------------
    Total |      102    100.00        102    100.00         100.00
                               (n = 102)

. sort id site place followup

. *Here in the loc2 variable the failures do not increase but the frequency
is higher in the four cells which may present an analysis with less of an
issue regarding "perfect predictors" due to sparse failure counts.
. *Now I will attempt to analyze the data using the six models discussed in
my thesis.
. *(A) Single site per person and single time interval
. *I need to limit my evaluation to the first year in the study and that
means that [first year of follow up-place(placement date of implant)] needs
to be indicated.
. *I also must assure that the censoring variable is maintained.
```

```
. sort id site place followup
. /*by id:gen year=1 if followup-place<=365.25
> replace year=2 if followup-place<=2*365+0.25 & year~=1
> replace year=3 if followup-place<=3*365+0.25 & year~=2
> replace year=4 if followup-place<=4*365+0.25 & year~=3
> replace year=5 if followup-place<=5*365+0.25 & year~=4
> replace year=6 if followup-place<=6*365+0.25 & year~=5
> replace year=7 if followup-place<=7*365+0.25 & year~=6
> replace year=8 if followup-place<=8*365+0.25 & year~=7
> codebook year*/
. codebook year

-----------------------------------------------------------------------------
year
(unlabeled)
-----------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,8]                        units:  1
         unique values:  8                        missing .:  0/7986

            tabulation:  Freq.  Value
                          2738  1
                          2442  2
                          1218  3
                           774  4
                           457  5
                           231  6
                           110  7
                            16  8

. *Now I will attempt a logistic regression for  a single site per person and
single time.  I will also incorporate in the code a variable called firstimp
to indicate the first record and only include this implant for evaluation.
. *The code used to generate such a variable follows.
. /*firstimp=0
> by id:gen firstimp=1 if site==site[1]*/
. codebook firstimp

-----------------------------------------------------------------------------
firstimp
(unlabeled)
----------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/7986

            tabulation:  Freq.  Value
                          5377  0
                          2609  1
```

```
. *firstimp indicates implant first records at an implant level
. list newid2 id site place followup firstimp failure in 1/72

     +-------------------------------------------------------------------+
     | newid2   id   site       place    followup   firstimp   failure |
     |-------------------------------------------------------------------|
  1. |      1    1     22   22jun1993   09jun1994          1         0 |
  2. |      2    1     23   22jun1993   09jun1994          0         0 |
  3. |      3    1     25   22jun1993   09jun1994          0         0 |
  4. |      4    1     26   22jun1993   09jun1994          0         0 |
  5. |      5    1     27   22jun1993   09jun1994          0         0 |
     |-------------------------------------------------------------------|
  6. |      6    9     22   16jan1984   25may1984          1         0 |
  7. |      6    9     22   16jan1984   09aug1984          1         0 |
  8. |      6    9     22   16jan1984   24sep1984          1         0 |
  9. |      6    9     22   16jan1984   15nov1984          1         0 |
 10. |      6    9     22   16jan1984   01may1985          1         0 |
     |-------------------------------------------------------------------|
 11. |      6    9     22   16jan1984   16aug1985          1         0 |
 12. |      6    9     22   16jan1984   13dec1985          1         0 |
 13. |      6    9     22   16jan1984   02jun1986          1         0 |
 14. |      6    9     22   16jan1984   05dec1986          1         0 |
 15. |      7    9     27   16jan1984   25may1984          0         0 |
     |-------------------------------------------------------------------|
 16. |      7    9     27   16jan1984   09aug1984          0         0 |
 17. |      7    9     27   16jan1984   24sep1984          0         0 |
 18. |      7    9     27   16jan1984   15nov1984          0         0 |
 19. |      7    9     27   16jan1984   01may1985          0         0 |
 20. |      7    9     27   16jan1984   16aug1985          0         0 |
     |-------------------------------------------------------------------|
 21. |      7    9     27   16jan1984   13dec1985          0         0 |
 22. |      7    9     27   16jan1984   02jun1986          0         0 |
 23. |      7    9     27   16jan1984   05dec1986          0         0 |
 24. |      8   10     22   16may1990   27mar1991          1         0 |
 25. |      8   10     22   16may1990   25apr1991          1         0 |
     |-------------------------------------------------------------------|
 26. |      8   10     22   16may1990   29may1991          1         0 |
 27. |      8   10     22   16may1990   25sep1991          1         0 |
 28. |      8   10     22   16may1990   08mar1992          1         0 |
 29. |      8   10     22   16may1990   06apr1992          1         0 |
 30. |      9   10     27   16may1990   27mar1991          0         0 |
     |-------------------------------------------------------------------|
 31. |      9   10     27   16may1990   25apr1991          0         0 |
 32. |      9   10     27   16may1990   29may1991          0         0 |
 33. |      9   10     27   16may1990   25sep1991          0         0 |
 34. |      9   10     27   16may1990   08mar1992          0         0 |
 35. |      9   10     27   16may1990   06apr1992          0         0 |
     |-------------------------------------------------------------------|
 36. |     10   14     30   29apr1987   13jan1988          1         0 |
 37. |     11   16     22   01jul1992   10may1994          1         0 |
 38. |     12   16     23   01jul1992   10may1994          0         0 |
 39. |     13   16     25   01jul1992   10may1994          0         0 |
 40. |     14   16     26   01jul1992   10may1994          0         0 |
```

```
     |-----------------------------------------------------------------|
 41. |     15   16     27   01jul1992   10may1994         0         0 |
 42. |     16   17     22   12jun1992   10dec1992         1         0 |
 43. |     16   17     22   12jun1992   24dec1992         1         0 |
 44. |     17   17     24   12jun1992   10dec1992         0         0 |
 45. |     17   17     24   12jun1992   24dec1992         0         0 |
     |-----------------------------------------------------------------|
 46. |     18   17     25   12jun1992   10dec1992         0         0 |
 47. |     18   17     25   12jun1992   24dec1992         0         0 |
 48. |     19   17     27   12jun1992   10dec1992         0         0 |
 49. |     19   17     27   12jun1992   24dec1992         0         0 |
 50. |     20   18      6   10jan1991   29jun1994         1         0 |
     |-----------------------------------------------------------------|
 51. |     21   20     23   08feb1990   18jun1990         1         0 |
 52. |     21   20     23   08feb1990   18sep1990         1         0 |
 53. |     21   20     23   08feb1990   21nov1990         1         0 |
 54. |     21   20     23   08feb1990   03may1991         1         0 |
 55. |     21   20     23   08feb1990   25oct1991         1         0 |
     |-----------------------------------------------------------------|
 56. |     21   20     23   08feb1990   06apr1992         1         0 |
 57. |     22   20     26   08feb1990   18jun1990         0         0 |
 58. |     22   20     26   08feb1990   18sep1990         0         0 |
 59. |     22   20     26   08feb1990   21nov1990         0         0 |
 60. |     22   20     26   08feb1990   03may1991         0         0 |
     |-----------------------------------------------------------------|
 61. |     22   20     26   08feb1990   25oct1991         0         0 |
 62. |     22   20     26   08feb1990   06apr1992         0         0 |
 63. |     23   20     28   08feb1990   18jun1990         0         0 |
 64. |     23   20     28   08feb1990   18sep1990         0         0 |
 65. |     23   20     28   08feb1990   21nov1990         0         0 |
     |-----------------------------------------------------------------|
 66. |     23   20     28   08feb1990   03may1991         0         0 |
 67. |     23   20     28   08feb1990   25oct1991         0         0 |
 68. |     23   20     28   08feb1990   06apr1992         0         0 |
 69. |     24   21     21   11aug1987   16feb1988         1         0 |
 70. |     24   21     21   11aug1987   20oct1988         1         0 |
     |-----------------------------------------------------------------|
 71. |     24   21     21   11aug1987   20apr1989         1         1 |
 72. |     25   21     23   11aug1987   16feb1988         0         0 |
     +-----------------------------------------------------------------+

. *(A) Single site per person and single time interval (first year of study).
. count if firstimp==1 & year==1
  920

. xi:logistic failure i.loc i.type i.race if firstimp==1 & year==1
i.loc             _Iloc_1-6           (naturally coded; _Iloc_1 omitted)
i.type            _Itype_1-3          (naturally coded; _Itype_1 omitted)
i.race            _Irace_1-6          (naturally coded; _Irace_1 omitted)

note: _Itype_3 != 0 predicts failure perfectly
      _Itype_3 dropped and 14 obs not used
```

```
note: _Irace_3 != 0 predicts failure perfectly
      _Irace_3 dropped and 1 obs not used

note: _Irace_4 != 0 predicts failure perfectly
      _Irace_4 dropped and 1 obs not used

note: _Irace_5 != 0 predicts failure perfectly
      _Irace_5 dropped and 19 obs not used


Logistic regression                          Number of obs = 837
                                             LR chi2(8)= 9.83
                                             Prob > chi2 = 0.2771
Log likelihood = -96.846324                  Pseudo R2 = 0.0483

-----------------------------------------------------------------------------
     failure | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
     _Iloc_2 |   .7782141    .797963    -0.24   0.807     .1043029
5.80633
     _Iloc_3 |   2.278973    2.93567     0.64   0.523     .1824986
28.45894
     _Iloc_4 |   .2561431   .2311472    -1.51   0.131     .0436864
1.501825
     _Iloc_5 |   .4581945   .3684603    -0.97   0.332     .0947437
2.215897
     _Iloc_6 |   .6365915   .6576019    -0.44   0.662     .0840554
4.821209
    _Itype_2 |   .8837802   .7070266    -0.15   0.877     .1842386
4.239434
    _Irace_2 |   1.790739   1.043359     1.00   0.317     .5715933
5.610191
    _Irace_6 |   57.57701   84.56057     2.76   0.006     3.236909
1024.16
-----------------------------------------------------------------------------
```

```
. *I will use the other variables that were created to evaluate this model.
. xi:logistic failure i.loc2 i.type i.race3 if firstimp==1 & year==1
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type            _Itype_1-3          (naturally coded; _Itype_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

note: _Itype_3 != 0 predicts failure perfectly
      _Itype_3 dropped and 14 obs not used

Logistic regression                                 Number of obs = 858
                                                    LR chi2(5) = 3.70
                                                    Prob > chi2 = 0.5938
Log likelihood = -100.46537                         Pseudo R2 = 0.0181

-----------------------------------------------------------------------------
     failure | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
    _Iloc2_2 |   1.546426   1.233396     0.55   0.585     .3239135
7.382939
    _Iloc2_3 |   .6706245   .3635448    -0.74   0.461      .231763
1.940505
    _Iloc2_4 |   2.209078   1.513042     1.16   0.247     .5770404
8.456994
    _Itype_2 |   .7968733   .6342417    -0.29   0.775     .1674584
3.792028
   _Irace3_2 |   1.846023    .983033     1.15   0.250     .6500705
5.242202
-----------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =    3.02
         Prob > chi2 =    0.3892

. *Type is still an issue.
```

```
. codebook type

--------------------------------------------------------------------------------
type
(unlabeled)
--------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,3]                        units:  1
         unique values:  3                        missing .:  0/7986

            tabulation:  Freq.  Value
                          7398  1
                           466  2
                           122  3

. /*gen type2=1 if type=1 and type~=.
> gen type2=1 if type==1 and type~=.
> gen type2=1 if type==1 & type~=.
> codebook type type2
> replace type2=2 if type==2|type==3 & type~=.
> codebook type type2
> save "C:\STATA\final9.dta", replace*/
. xi:logistic failure i.loc2 i.type2 i.race3 if firstimp==1 & year==1
i.loc2           _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2          _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3          _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Logistic regression                                   Number of obs = 872
                                                      LR chi2(5) = 4.04
                                                      Prob > chi2 =0.5437
Log likelihood = -100.65474                           Pseudo R2 = 0.0197

--------------------------------------------------------------------------------
     failure | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+------------------------------------------------------------------
    _Iloc2_2 |   1.515704   1.205282     0.52   0.601     .3189646
7.202555
    _Iloc2_3 |    .636673   .3393099    -0.85   0.397      .224014
1.809496
    _Iloc2_4 |   2.164733   1.476601     1.13   0.258     .5685721
8.241823
   _Itype2_2 |   .7100469   .5546303    -0.44   0.661     .1536025
3.282281
   _Irace3_2 |   1.875067   .9965234     1.18   0.237     .6616631
5.313694
--------------------------------------------------------------------------------
```

```
. *The type variable no loger is presented as being problematic.
. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =    3.16
         Prob > chi2 =    0.3670

. *I will now incorporate the robust variance into the analysis.
. xi:logistic failure i.loc2 i.type2 i.race3 if firstimp==1 & year==1, robust
cluster(id)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Logistic regression                      Number of obs = 872
                                         Wald chi2(5)= 3.78
                                         Prob > chi2 = 0.5817
Log pseudo-likelihood = -100.65474       Pseudo R2 = 0.0197

                               (standard errors adjusted for clustering on
id)
------------------------------------------------------------------------------
             |               Robust
     failure | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
    _Iloc2_2 |   1.515704   1.195212     0.53   0.598      .323145
7.109377
    _Iloc2_3 |    .636673    .335304    -0.86   0.391     .2267936
1.787319
    _Iloc2_4 |   2.164733   1.453811     1.15   0.250     .5804258
8.073504
   _Itype2_2 |   .7100469   .5712074    -0.43   0.670     .1467323
3.435961
   _Irace3_2 |   1.875067   1.014503     1.16   0.245     .6493439
5.414504
------------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =    3.05
         Prob > chi2 =    0.3846
```

```
. *(B) The next model evaluates the situation for Multiple sites per person
and
>  a single time interval.
. count if year==1
 2738

. xi:logistic failure i.loc2 i.type2 i.race3 if year==1
i.loc2            _Iloc2_1-4         (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2        (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2        (naturally coded; _Irace3_1 omitted)

Logistic regression                   Number of obs = 2610
                                      LR chi2(5)= 8.89
                                      Prob > chi2 =  0.1136
Log likelihood = -258.41238           Pseudo R2 = 0.0169

------------------------------------------------------------------------------
     failure | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
    _Iloc2_2 |   2.147211   .9929313     1.65   0.098     .8674711
5.314892
    _Iloc2_3 |   1.119069   .3712169     0.34   0.735     .5841129
2.143961
    _Iloc2_4 |   2.861581   1.433786     2.10   0.036     1.071801
7.640077
   _Itype2_2 |   .8545112   .4567328    -0.29   0.769     .2997464
2.436024
   _Irace3_2 |   1.966096   .6707841     1.98   0.048     1.007385
3.837196
------------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =    6.30
         Prob > chi2 =    0.0981

. *This does not incorporate the robust variance
```

```
. xi:logistic failure i.loc2 i.type2 i.race3 if year==1,robust cluster(id)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Logistic regression                            Number of obs = 2610
                                               Wald chi2(5)= 5.00
                                              Prob > chi2 = 0.4161
Log pseudo-likelihood = -258.41238            Pseudo R2 = 0.0169
                              (standard errors adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
     failure | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
    _Iloc2_2 |   2.147211   1.134491     1.45   0.148     .7623208
6.047999
    _Iloc2_3 |   1.119069   .3654506     0.34   0.730     .5900419
2.122418
    _Iloc2_4 |   2.861581   1.885753     1.60   0.111     .7864529
10.41213
   _Itype2_2 |   .8545112   .6437418    -0.21   0.835     .1951953
3.740815
   _Irace3_2 |   1.966096   .9476919     1.40   0.161     .7643828
5.057064
------------------------------------------------------------------------------
. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =    2.86
         Prob > chi2 =    0.4131

. *Now using the xt command structure in Stata to evaluate GEE
. set matsize 80
```

```
. xi:xtgee failure i.loc2 i.type2 i.race3 if year==1, family(bin) link(logit)
corr(exc) i(id) eform
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Iteration 1: tolerance = .10700382
Iteration 2: tolerance = .00799128
Iteration 3: tolerance = .00033571
Iteration 4: tolerance = .00004652
Iteration 5: tolerance = 7.987e-06
Iteration 6: tolerance = 1.144e-06
Iteration 7: tolerance = 1.803e-07

GEE population-averaged model                   Number of obs=2610
Group variable: id                              Number of groups=557
Link: logit      Obs per group: min =1
Family: binomial                                avg =4.7
Correlation: exchangeable                        max =74
                                                Wald chi2(5) =8.21
Scale parameter: 1                              Prob > chi2 =0.1451

------------------------------------------------------------------------------
     failure | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
    _Iloc2_2 |   1.828297   .9883837     1.12   0.264     .6337011
5.274839
    _Iloc2_3 |   1.255482   .4067927     0.70   0.483     .6652885
2.369251
    _Iloc2_4 |   2.619865     1.4343     1.76   0.079      .895923
7.661029
   _Itype2_2 |   .7972603   .5085502    -0.36   0.722     .2283717
2.783287
   _Irace3_2 |   2.177236   .8389304     2.02   0.043     1.023108
4.633292
------------------------------------------------------------------------------
. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =    3.63
         Prob > chi2 =    0.3040
```

```
. *I will now analyze the data using the robust variance analysis.
. xi:xtgee failure i.loc2 i.type2 i.race3 if year==1, family(bin) link(logit)
corr(exc) i(id) eform
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)
Iteration 1: tolerance = .10700382
Iteration 2: tolerance = .00799128
Iteration 3: tolerance = .00033571
Iteration 4: tolerance = .00004652
Iteration 5: tolerance = 7.987e-06
Iteration 6: tolerance = 1.144e-06
Iteration 7: tolerance = 1.803e-07

GEE population-averaged model                    Number of obs = 2610
Group variable: id                               Number of groups = 557
Link: logit                                      Obs per group: min = 1
Family: binomial                                             avg = 4.7
Correlation: exchangeable                                    max =74
                                                 Wald chi2(5) = 8.21
Scale parameter:1                              Prob > chi2 = 0.1451
------------------------------------------------------------------------------
     failure | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
    _Iloc2_2 |   1.828297   .9883837     1.12   0.264     .6337011
5.274839
    _Iloc2_3 |   1.255482   .4067927     0.70   0.483     .6652885
2.369251
    _Iloc2_4 |   2.619865     1.4343     1.76   0.079      .895923
7.661029
   _Itype2_2 |   .7972603   .5085502    -0.36   0.722     .2283717
2.783287
   _Irace3_2 |   2.177236   .8389304     2.02   0.043     1.023108
4.633292
------------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =    3.63
         Prob > chi2 =    0.3040

. *These are the population averaged models.  The second model incorporating
the robust variance procedure.
```

```
. *The next situation to evaluate is the single site per person with multiple
time intervals.  In this situation I must reorganize the data to accomodate
one record per person per time interval.  I will use the stsplit command in
Stata where time intervals will be established and the observation level will
change according to this.  I do not want to change this dataset and will use
another dataset that has been established for this analysis (final9b.dta).
. clear

. use "C:\unzipped\final9Folder\final9b.dta", clear

. desc

Contains data from C:\unzipped\final9Folder\final9b.dta
  obs:        11,794
 vars:           136
 size:     6,038,528 (52.0% of memory free)
-----------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
-----------------------------------------------------------------------------
implwdthPlus1   float  %9.0g                   implantwidth+1 for log scale
nxdate          float  %d                     numeric xdate
nbdate          float  %d                     numeric bdate
surocc9         float  %9.0g                  Unable to Seat Implant
surocc10        float  %9.0g                  Implant Not Well Adapted to
Site
surocc11        float  %9.0g                  Ridge Augmentation Used
surocc12        float  %9.0g                  Periodontal Tissue Damage
surocc13        float  %9.0g                  Patient Experienced Pain
surocc14        float  %9.0g                  Excessive Bleeding
surocc15        float  %9.0g                  Guided Tissue Regeneration
surocc16        float  %9.0g
_merge          byte   %8.0g
age1            float  %9.0g
newid2          float  %9.0g                  group(id site)
y               float  %9.0g
id              double %9.0g                  id
nisdate         double %d                     isdate
place           float  %d
nevldate        double %d                     evaldate
followup        float  %d
nimrdate        double %d                     imprdate
site            double %9.0g                  site
failure         float  %9.0g
rownames        str5   %5s
evaldate        str11  %11s
mobil           float  %9.0g
periminf        str1   %1s                    Peri-implant Inflammation
imphcat         float  %9.0g                  Implant Health Category
```

imprdate	str11	%11s	
impfunc	float	%9.0g	Implant Functionality
impltopt	float	%9.0g	Implant Less Than Optimal but Functional
impnonsp	float	%9.0g	
imp2brmv	float	%9.0g	
funother	float	%9.0g	
painlswr	float	%9.0g	
esthetic	float	%9.0g	
mastprob	float	%9.0g	Mastication Problems Due to Implant
speechpr	float	%9.0g	Speech Problems Due to Implant
cmplnoth	float	%9.0g	
compimpi	float	%9.0g	If Compromised Implant Intervention Was Used
newid	float	%9.0g	group(newssn)
isdate	str11	%11s	
imparch	str1	%1s	Arch Location
imp1	float	%9.0g	
imptype	float	%9.0g	Implant Type
matcode	float	%9.0g	Material Code
coatcode	float	%9.0g	Coating Code
stagecode	float	%9.0g	Stage Code
morphcode	float	%9.0g	Morphology Code
implantheight	float	%9.0g	Implant Height (mm)
implantwidth	float	%9.0g	Implant Width/Diameter (mm)
availboneheight	float	%9.0g	Height of Available Bone (mm)
availbonewidth	float	%9.0g	Width of Available Bone (mm)
avboneht	float	%9.0g	
avbonewi	float	%9.0g	
attginwi	float	%9.0g	
bonclass	float	%9.0g	Bone Classification
surocc1	float	%9.0g	Implant Altered
surocc2	float	%9.0g	Alveolar Ridge Perforation
surocc3	float	%9.0g	Jaw Fracture
surocc4	float	%9.0g	Neurological Damage
surocc5	float	%9.0g	Inferior Mandibular Border Perforation
surocc6	float	%9.0g	Sinus Lift
surocc7	float	%9.0g	Perforated Sinus/Nasal Cavity
surocc8	float	%9.0g	Equipment Complications
provid	str4	%4s	
station	float	%9.0g	
ethw	float	%9.0g	
ethb	float	%9.0g	
etha	float	%9.0g	
ethnam	float	%9.0g	
ethhis	float	%9.0g	
ethoth	float	%9.0g	
sex	str1	%1s	
asarate	float	%9.0g	
edenttot	float	%9.0g	

```
gender          long   %8.0g       gender     numeric sex
rem             float  %9.0g
index           float  %9.0g
ind             float  %9.0g
ctr             float  %9.0g
ctr1            float  %9.0g
dupimp          float  %9.0g
y2              float  %9.0g
visit           float  %9.0g
vistot          float  %9.0g                  Total number of Visits
sittot          float  %9.0g
sit             float  %9.0g
sitetot         float  %9.0g
sittotal        float  %9.0g
sittotal2       float  %9.0g                  Total number of sites per
                                                patient
freq            float  %9.0g                  The counter for all first
                                                records by id and site
AVBHPlus1       float  %9.0g                   Availboneheight+1 for log
scale
AVBWPlus1       float  %9.0g                  Availbonewidth+1 for log scale
implhtPlus1     float  %9.0g                  implantheight+1 for log scale
age2            float  %9.0g
agecat          float  %9.0g
arch            float  %9.0g
failind         byte   %8.0g
seq             float  %9.0g
firstrec        float  %9.0g
race            float  %9.0g
race2           float  %9.0g
type            float  %9.0g
loc             float  %9.0g
_st             byte   %8.0g
_d              byte   %8.0g
_origin         int    %10.0g
_t              double %10.0g
_t0             double %10.0g
annualt         byte   %9.0g
race3           float  %9.0g
type2           float  %9.0g
loc2            float  %9.0g
durat1          byte   %8.0g                  annualt==     0.0000
durat2          byte   %8.0g                  annualt==     1.0000
durat3          byte   %8.0g                  annualt==     2.0000
durat4          byte   %8.0g                  annualt==     3.0000
durat5          byte   %8.0g                  annualt==     4.0000
durat6          byte   %8.0g                  annualt==     5.0000
durat7          byte   %8.0g                  annualt==     6.0000
durat8          byte   %8.0g                  annualt==     7.0000
dfail           float  %9.0g
_Iannualt_1     byte   %8.0g                  annualt==1
_Iannualt_2     byte   %8.0g                  annualt==2
_Iannualt_3     byte   %8.0g                  annualt==3
```

```
_Iannualt_4     byte   %8.0g                  annualt==4
_Iannualt_5     byte   %8.0g                  annualt==5
_Iannualt_6     byte   %8.0g                  annualt==6
_Iannualt_7     byte   %8.0g                  annualt==7
_Iloc2_2        byte   %8.0g                  loc2==2
_Iloc2_3        byte   %8.0g                  loc2==3
_Iloc2_4        byte   %8.0g                  loc2==4
_Itype2_2       byte   %8.0g                  type2==2
_Irace3_2       byte   %8.0g                  race3==2
firstimp        float  %9.0g
-----------------------------------------------------------------------------
Sorted by:  id  site  place  followup

. *This data was modified using the stsplit command to create a categorical
variable called annualt which separates the data into yearly time intervals
and allows for dicrete survival analysis.  Also the analysis requires that
other variables be created: (A) one to index each patient (B) a binary
dependent variable to indicate censorship within the time intervals, and (C)
a variable to summarize the pattern of duration dependence.
. *The data in it's current form has more observations than the final9.dta
dataset.  Also, the race loc and type variables need to be collapsed as well.
. codebook   race race2 type loc

-----------------------------------------------------------------------------
race
(unlabeled)
-----------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,6]                        units:  1
         unique values:  6                        missing .:  577/11794

           tabulation:  Freq.  Value
                         9800  1
                         1207  2
                            9  3
                           23  4
                          164  5
                           14  6
                          577  .
```

```
--------------------------------------------------------------------------------
race2
(unlabeled)
--------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,3]                        units:  1
         unique values:  3                        missing .:  577/11794

            tabulation:  Freq.  Value
                          9800  1
                          1403  2
                            14  3
                           577  .

--------------------------------------------------------------------------------
type
(unlabeled)
--------------------------------------------------------------------------------
                  type:  numeric (float)

                 range:  [1,3]                        units:  1
         unique values:  3                        missing .:  0/11794

            tabulation:  Freq.  Value
                         10945  1
                           660  2
                           189  3

--------------------------------------------------------------------------------
loc
(unlabeled)
--------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [1,6]                        units:  1
         unique values:  6                        missing .:  0/11794

            tabulation:  Freq.  Value
                           326  1
                           617  2
                           328  3
                          1475  4
                          7722  5
                          1326  6

. /*Code for generating race3, loc2 and type2 variables
> gen race3=1 if race2==1 &  race2~=.
> codebook   race2 race3
```

```
> replace race3=2 if race2==2|race2==3 &  race2~=.
> codebook   race2 race3
> save "C:\unzipped\final9Folder\final9b.dta", replace
> gen  type2=1 if type==1 & type~=.
> codebook type type2
> replace type2=2 if type==2|type==3 & type~=.
> codebook type type2
> save "C:\unzipped\final9Folder\final9b.dta", replace
> gen loc2=1 if loc==5 & loc~=.
> codebook loc loc2
> replace loc2=2 if loc==2 & loc~=.
> codebook loc loc2
> replace loc2=3 if loc==4|loc==6 & loc~=.
> codebook loc loc2
> replace loc2=4 if loc==1|loc==3 & loc~=.
> codebook loc loc2*/

. codebook annualt

-----------------------------------------------------------------------------
annualt
(unlabeled)
-----------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,7]                        units:  1
         unique values:  8                        missing .:  0/11794

           tabulation:  Freq.  Value
                         4368  0
                         3424  1
                         1804  2
                         1119  3
                          638  4
                          304  5
                          121  6
                           16  7

. *I need to create another censorship indicator
```

```
. codebook _d

-------------------------------------------------------------------------------
_d
(unlabeled)
-------------------------------------------------------------------------------

                  type:  numeric (byte)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/11794

            tabulation:  Freq.  Value
                         11691  0
                           103  1

. /*gen dfail=_d*/
. codebook dfail

-------------------------------------------------------------------------------
dfail
(unlabeled)
-------------------------------------------------------------------------------

                  type:  numeric (float)

                 range:  [0,1]                        units:  1
         unique values:  2                        missing .:  0/11794

            tabulation:  Freq.  Value
                         11691  0
                           103  1

. *I will also need to code a variable to indicate the first record in this
dataset "firstimp" and only include this implant for evaluation.
. *The code used to generate such a variable follows.
. /*gen firstimp=0
> by id:gen firstimp=1 if site==site[1]*/
. list id site place followup annualt firstimp in 1/72

     +-----------------------------------------------------------+
     | id   site       place    followup   annualt   firstimp |
     |-----------------------------------------------------------|
  1. |  1     22   22jun1993   09jun1994         0          1 |
  2. |  1     23   22jun1993   09jun1994         0          0 |
  3. |  1     25   22jun1993   09jun1994         0          0 |
  4. |  1     26   22jun1993   09jun1994         0          0 |
  5. |  1     27   22jun1993   09jun1994         0          0 |
     |-----------------------------------------------------------|
  6. |  9     22   16jan1984   25may1984         0          1 |
```

7.	9	22	16jan1984	09aug1984	0	1
8.	9	22	16jan1984	24sep1984	0	1
9.	9	22	16jan1984	15nov1984	0	1
10.	9	22	16jan1984	15jan1985	0	1
11.	9	22	16jan1984	01may1985	1	1
12.	9	22	16jan1984	16aug1985	1	1
13.	9	22	16jan1984	13dec1985	1	1
14.	9	22	16jan1984	15jan1986	1	1
15.	9	22	16jan1984	02jun1986	2	1
16.	9	22	16jan1984	05dec1986	2	1
17.	9	27	16jan1984	25may1984	0	0
18.	9	27	16jan1984	09aug1984	0	0
19.	9	27	16jan1984	24sep1984	0	0
20.	9	27	16jan1984	15nov1984	0	0
21.	9	27	16jan1984	15jan1985	0	0
22.	9	27	16jan1984	01may1985	1	0
23.	9	27	16jan1984	16aug1985	1	0
24.	9	27	16jan1984	13dec1985	1	0
25.	9	27	16jan1984	15jan1986	1	0
26.	9	27	16jan1984	02jun1986	2	0
27.	9	27	16jan1984	05dec1986	2	0
28.	10	22	16may1990	27mar1991	0	1
29.	10	22	16may1990	25apr1991	0	1
30.	10	22	16may1990	16may1991	0	1
31.	10	22	16may1990	29may1991	1	1
32.	10	22	16may1990	25sep1991	1	1
33.	10	22	16may1990	08mar1992	1	1
34.	10	22	16may1990	06apr1992	1	1
35.	10	27	16may1990	27mar1991	0	0
36.	10	27	16may1990	25apr1991	0	0
37.	10	27	16may1990	16may1991	0	0
38.	10	27	16may1990	29may1991	1	0
39.	10	27	16may1990	25sep1991	1	0
40.	10	27	16may1990	08mar1992	1	0
41.	10	27	16may1990	06apr1992	1	0
42.	14	30	29apr1987	13jan1988	0	1
43.	16	22	01jul1992	01jul1993	0	1
44.	16	22	01jul1992	10may1994	1	1
45.	16	23	01jul1992	01jul1993	0	0
46.	16	23	01jul1992	10may1994	1	0
47.	16	25	01jul1992	01jul1993	0	0
48.	16	25	01jul1992	10may1994	1	0
49.	16	26	01jul1992	01jul1993	0	0
50.	16	26	01jul1992	10may1994	1	0

```
 51. | 16     27   01jul1992   01jul1993         0          0 |
 52. | 16     27   01jul1992   10may1994         1          0 |
 53. | 17     22   12jun1992   10dec1992         0          1 |
 54. | 17     22   12jun1992   24dec1992         0          1 |
 55. | 17     24   12jun1992   10dec1992         0          0 |
     |---------------------------------------------------------|
 56. | 17     24   12jun1992   24dec1992         0          0 |
 57. | 17     25   12jun1992   10dec1992         0          0 |
 58. | 17     25   12jun1992   24dec1992         0          0 |
 59. | 17     27   12jun1992   10dec1992         0          0 |
 60. | 17     27   12jun1992   24dec1992         0          0 |
     |---------------------------------------------------------|
 61. | 18      6   10jan1991   10jan1992         0          1 |
 62. | 18      6   10jan1991   09jan1993         1          1 |
 63. | 18      6   10jan1991   09jan1994         2          1 |
 64. | 18      6   10jan1991   29jun1994         3          1 |
 65. | 20     23   08feb1990   18jun1990         0          1 |
     |---------------------------------------------------------|
 66. | 20     23   08feb1990   18sep1990         0          1 |
 67. | 20     23   08feb1990   21nov1990         0          1 |
 68. | 20     23   08feb1990   08feb1991         0          1 |
 69. | 20     23   08feb1990   03may1991         1          1 |
 70. | 20     23   08feb1990   25oct1991         1          1 |
     |---------------------------------------------------------|
 71. | 20     23   08feb1990   08feb1992         1          1 |
 72. | 20     23   08feb1990   06apr1992         2          1 |
     +---------------------------------------------------------+

. *firstimp indicates implant first records at an implant level.

. tabulate failure annualt

           |                      annualt
   failure |         0          1          2          3 |     Total
-----------+--------------------------------------------+----------
         0 |     2,684      2,419      1,208        770 |     7,883
         1 |        54         23         10          4 |       103
-----------+--------------------------------------------+----------
     Total |     2,738      2,442      1,218        774 |     7,986


           |                      annualt
   failure |         4          5          6          7 |     Total
-----------+--------------------------------------------+----------
         0 |       453        232        105         12 |     7,883
         1 |         4          3          1          4 |       103
-----------+--------------------------------------------+----------
     Total |       457        235        106         16 |     7,986
```

```
. *(C)Discrete Proportional Odds model.
. *This is the situation of the single site per patient with multiple time
intervals.
. count if firstimp==1
 3896
. xi:logit dfail i.annualt i.loc2 i.type2 i.race3 if firstimp==1
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

note: _Iannualt_6 != 0 predicts failure perfectly
      _Iannualt_6 dropped and 43 obs not used
Iteration 0:   log likelihood =  -206.7099
Iteration 1:   log likelihood =  -206.1689
Iteration 2:   log likelihood =  -196.4683
Iteration 3:   log likelihood = -194.11401
Iteration 4:   log likelihood = -193.99376
Iteration 5:   log likelihood = -193.99212
Iteration 6:   log likelihood = -193.99212

Logit estimates                                   Number of obs =3651
                                                  LR chi2(11) =25.44
                                                  Prob > chi2=0.0079
Log likelihood = -193.99212                       Pseudo R2 =0.0615
-----------------------------------------------------------------------------
       dfail |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
 _Iannualt_1 |  -1.004123   .4644104    -2.16   0.031     -1.91435    -
.093895
 _Iannualt_2 |  -.5688591   .4999297    -1.14   0.255    -1.548703
.4109851
 _Iannualt_3 |  -1.633965   1.026056    -1.59   0.111    -3.644999
.377068
 _Iannualt_4 |  -.9991932   1.028893    -0.97   0.331    -3.015787
1.017401
 _Iannualt_5 |  -.1925976     1.0361    -0.19   0.853    -2.223315
1.83812
 _Iannualt_7 |   2.543793   1.130133     2.25   0.024     .3287724
4.758813
    _Iloc2_2 |   1.624892   .4848263     3.35   0.001     .6746496
2.575134
    _Iloc2_3 |  -.0523555   .4280398    -0.12   0.903    -.8912981
.7865871
    _Iloc2_4 |    .964416   .5385706     1.79   0.073    -.0911628
2.019995
   _Itype2_2 |   .3072857    .577294     0.53   0.595    -.8241898
1.438761
   _Irace3_2 |   .2383802   .4580788     0.52   0.603    -.6594378
1.136198
       _cons |  -4.479754   .3276953   -13.67   0.000    -5.122025   -
3.837483
-----------------------------------------------------------------------------
```

```
. logit, or

note: _Iannualt_6 != 0 predicts failure perfectly
      _Iannualt_6 dropped and 43 obs not used


Logit estimates                                   Number of obs =3651
                                                  LR chi2(11) = 25.44
                                                  Prob > chi2=0.0079
Log likelihood = -193.99212                       Pseudo R2 =0.0615

------------------------------------------------------------------------------
       dfail | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
 _Iannualt_1 |    .366366   .1701441    -2.16   0.031     .1474376
.9103783
 _Iannualt_2 |    .566171   .2830457    -1.14   0.255     .2125234
1.508303
 _Iannualt_3 |   .1951542   .2002392    -1.59   0.111     .0261214
1.458003
 _Iannualt_4 |   .3681764   .3788142    -0.97   0.331     .0490072
2.765996
 _Iannualt_5 |   .8248138   .8545892    -0.19   0.853     .1082496
6.284713
 _Iannualt_7 |   12.72785   14.38417     2.25   0.024     1.389262
116.6074
    _Iloc2_2 |   5.077869   2.461885     3.35   0.001     1.963345
13.13308
    _Iloc2_3 |   .9489914   .4062061    -0.12   0.903      .410123
2.195889
    _Iloc2_4 |   2.623255   1.412808     1.79   0.073      .912869
7.538287
   _Itype2_2 |   1.359729   .7849637     0.53   0.595     .4385902
4.215471
   _Irace3_2 |   1.269192   .5813897     0.52   0.603      .517142
3.114903
------------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   14.90
         Prob > chi2 =    0.0019
```

```
. testparm _Iannualt*

 ( 1)  _Iannualt_1 = 0
 ( 2)  _Iannualt_2 = 0
 ( 3)  _Iannualt_3 = 0
 ( 4)  _Iannualt_4 = 0
 ( 5)  _Iannualt_5 = 0
 ( 6)  _Iannualt_7 = 0

           chi2(  6) =   13.93
         Prob > chi2 =    0.0305

. xi:logit dfail i.annualt i.loc2 i.type2 i.race3 if firstimp==1, robust
cluster(id)
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

note: _Iannualt_6 != 0 predicts failure perfectly
      _Iannualt_6 dropped and 43 obs not used

Iteration 0:   log pseudo-likelihood =  -206.7099
Iteration 1:   log pseudo-likelihood =  -206.1689
Iteration 2:   log pseudo-likelihood =  -196.4683
Iteration 3:   log pseudo-likelihood = -194.11401
Iteration 4:   log pseudo-likelihood = -193.99376
Iteration 5:   log pseudo-likelihood = -193.99212
Iteration 6:   log pseudo-likelihood = -193.99212
```

```
Logit estimates                                    Number of obs = 3651
                                                   Wald chi2(11) = 41.31
                                                   Prob > chi2 = 0.0000
Log pseudo-likelihood = -193.99212                 Pseudo R2 =0.0615

                              (standard errors adjusted for clustering on id)
-----------------------------------------------------------------------------
             |               Robust
       dfail |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
 _Iannualt_1 |  -1.004123   .4743023    -2.12   0.034    -1.933738   -
.0745071
 _Iannualt_2 |  -.5688591   .4991743    -1.14   0.254    -1.547223
.4095044
 _Iannualt_3 |  -1.633965   1.023105    -1.60   0.110    -3.639214
.3712827
 _Iannualt_4 |  -.9991932   1.029963    -0.97   0.332    -3.017884
1.019498
 _Iannualt_5 |  -.1925976    1.03659    -0.19   0.853    -2.224277
1.839082
 _Iannualt_7 |   2.543793    1.07903     2.36   0.018     .4289335
4.658652
    _Iloc2_2 |   1.624892   .4946605     3.28   0.001     .6553749
2.594409
    _Iloc2_3 |  -.0523555   .4169062    -0.13   0.900    -.8694767
.7647657
    _Iloc2_4 |    .964416   .5341816     1.81   0.071    -.0825607
2.011393
   _Itype2_2 |   .3072857    .519319     0.59   0.554    -.7105609
1.325132
   _Irace3_2 |   .2383802   .4668716     0.51   0.610    -.6766713
1.153432
       _cons |  -4.479754   .3592683   -12.47   0.000    -5.183907   -
3.775601
-----------------------------------------------------------------------------
```

```
. logit, or

note: _Iannualt_6 != 0 predicts failure perfectly
      _Iannualt_6 dropped and 43 obs not used

Logit estimates                                   Number of obs = 3651
                                                  Wald chi2(11) = 41.31
                                                  Prob > chi2 = 0.0000
Log pseudo-likelihood = -193.99212                Pseudo R2 =0.0615

                              (standard errors adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
       dfail | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
 _Iannualt_1 |    .366366   .1737682    -2.12   0.034     .1446066
.9282009
 _Iannualt_2 |    .566171    .282618    -1.14   0.254     .2128383
1.506071
 _Iannualt_3 |   .1951542   .1996631    -1.60   0.110      .026273
1.449593
 _Iannualt_4 |   .3681764   .3792082    -0.97   0.332     .0489046
2.771803
 _Iannualt_5 |   .8248138   .8549939    -0.19   0.853     .1081456
6.290759
 _Iannualt_7 |   12.72785   13.73373     2.36   0.018     1.535619
105.4938
    _Iloc2_2 |   5.077869   2.511822     3.28   0.001     1.925864
13.38867
    _Iloc2_3 |   .9489914   .3956404    -0.13   0.900     .4191709
2.148491
    _Iloc2_4 |   2.623255   1.401295     1.81   0.071     .9207555
7.473719
   _Itype2_2 |   1.359729   .7061333     0.59   0.554     .4913685
3.762683
   _Irace3_2 |   1.269192   .5925495     0.51   0.610     .5083062
3.169049
------------------------------------------------------------------------------

. testparm _Iloc*
 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0
           chi2(  3) =   14.60
         Prob > chi2 =    0.0022
```

```
. testparm _Iannualt*

 ( 1)  _Iannualt_1 = 0
 ( 2)  _Iannualt_2 = 0
 ( 3)  _Iannualt_3 = 0
 ( 4)  _Iannualt_4 = 0
 ( 5)  _Iannualt_5 = 0
 ( 6)  _Iannualt_7 = 0

           chi2(  6) =   14.93
         Prob > chi2 =    0.0208

. *Now to evaluate the cloglog function
. xi:cloglog dfail i.annualt i.loc2 i.type2 i.race3 if firstimp==1
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)
note: _Iannualt_6 != 0 predicts failure perfectly
      _Iannualt_6 dropped and 43 obs not used


Iteration 0:   log likelihood = -194.13645
Iteration 1:   log likelihood = -194.01783
Iteration 2:   log likelihood = -194.01629
Iteration 3:   log likelihood = -194.01629

Complementary log-log regression                Number of obs = 3651
                                                Zero outcomes = 3614
                                               Nonzero outcomes = 37

                                                LR chi2(11)=25.39
Log likelihood = -194.01629                     Prob > chi2=0.0080
-----------------------------------------------------------------------------
       dfail |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
 _Iannualt_1 |  -.9904265   .4617544    -2.14   0.032    -1.895448   -
.0854046
 _Iannualt_2 |  -.5596892   .4960815    -1.13   0.259    -1.531991
.4126126
 _Iannualt_3 |  -1.623483   1.023573    -1.59   0.113     -3.62965
.382684
 _Iannualt_4 |  -.9926252   1.025086    -0.97   0.333    -3.001757
1.016506
 _Iannualt_5 |   -.189057   1.030384    -0.18   0.854    -2.208573
1.830459
 _Iannualt_7 |   2.475689   1.047728     2.36   0.018     .4221803
4.529197
    _Iloc2_2 |   1.605121   .4777364     3.36   0.001      .668775
2.541467
    _Iloc2_3 |  -.0427614   .4255752    -0.10   0.920    -.8768734
.7913505
```

```
    _Iloc2_4 |   .9630951   .5338906     1.80   0.071    -.0833113
2.009502
   _Itype2_2 |   .3194099   .5702439     0.56   0.575    -.7982476
1.437067
   _Irace3_2 |   .2353798   .4526449     0.52   0.603    -.6517879
1.122548
       _cons |  -4.493444    .326922   -13.74   0.000      -5.1342   -
3.852689
------------------------------------------------------------------------------

. matrix b=e(b)

. matrix v=e(V)

. ereturn post b v

. ereturn display, eform(exp_b)
------------------------------------------------------------------------------
             |      exp_b   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
dfail        |
 _Iannualt_1 |   .3714182    .171504    -2.14   0.032     .1502509
.9181408
 _Iannualt_2 |   .5713866   .2834543    -1.13   0.259      .216105
1.51076
 _Iannualt_3 |   .1972107   .2018596    -1.59   0.113     .0265255
1.466215
 _Iannualt_4 |   .3706025   .3798994    -0.97   0.333     .0496997
2.763523
 _Iannualt_5 |   .8277393   .8528894    -0.18   0.854     .1098573
6.236746
 _Iannualt_7 |   11.88989   12.45737     2.36   0.018     1.525283
92.68411
    _Iloc2_2 |   4.978462   2.378393     3.36   0.001     1.951845
12.69829
    _Iloc2_3 |   .9581399   .4077606    -0.10   0.920     .4160818
2.206374
    _Iloc2_4 |   2.619792   1.398683     1.80   0.071     .9200647
7.459598
   _Itype2_2 |   1.376315   .7848354     0.56   0.575     .4501171
4.208336
   _Irace3_2 |   1.265389    .572772     0.52   0.603     .5211132
3.072672
------------------------------------------------------------------------------
. testparm _Iloc*

 ( 1)  [dfail]_Iloc2_2 = 0
 ( 2)  [dfail]_Iloc2_3 = 0
 ( 3)  [dfail]_Iloc2_4 = 0

           chi2(  3) =   14.97
         Prob > chi2 =    0.0018
```

```
. testparm _Iannualt*

 ( 1)  [dfail]_Iannualt_1 = 0
 ( 2)  [dfail]_Iannualt_2 = 0
 ( 3)  [dfail]_Iannualt_3 = 0
 ( 4)  [dfail]_Iannualt_4 = 0
 ( 5)  [dfail]_Iannualt_5 = 0
 ( 6)  [dfail]_Iannualt_7 = 0

           chi2(  6) =   14.56
         Prob > chi2 =    0.0240

. xi:cloglog dfail i.annualt i.loc2 i.type2 i.race3 if firstimp==1,robust
clust
> er(id)
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)
note: _Iannualt_6 != 0 predicts failure perfectly
      _Iannualt_6 dropped and 43 obs not used


Iteration 0:   log pseudo-likelihood = -194.13645
Iteration 1:   log pseudo-likelihood = -194.01783
Iteration 2:   log pseudo-likelihood = -194.01629
Iteration 3:   log pseudo-likelihood = -194.01629

Complementary log-log regression                Number of obs  = 3651
                                                Zero outcomes =3614
                                                Nonzero outcomes = 37

                                                Wald chi2(11) =42.50
Log pseudo-likelihood = -194.01629              Prob > chi2=0.0000

                               (standard errors adjusted for clustering on
id)
------------------------------------------------------------------------------
             |               Robust
       dfail |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
 _Iannualt_1 |  -.9904265   .4754945    -2.08   0.037    -1.922379   -
.0584745
 _Iannualt_2 |  -.5596892   .4964257    -1.13   0.260    -1.532666
.4132874
 _Iannualt_3 |  -1.623483   1.019574    -1.59   0.111    -3.621812
.3748463
 _Iannualt_4 |  -.9926252   1.024684    -0.97   0.333    -3.000968
1.015718
 _Iannualt_5 |   -.189057   1.033048    -0.18   0.855    -2.213793
1.835679
```

```
 _Iannualt_7 |   2.475689   1.001414     2.47   0.013     .5129531
4.438424
    _Iloc2_2 |   1.605121   .4904733     3.27   0.001      .643811
2.566431
    _Iloc2_3 |  -.0427614   .4182068    -0.10   0.919    -.8624317
.7769088
    _Iloc2_4 |   .9630951   .5319986     1.81   0.070     -.079603
2.005793
   _Itype2_2 |   .3194099   .5182746     0.62   0.538    -.6963897
1.33521
   _Irace3_2 |   .2353798   .4624613     0.51   0.611    -.6710277
1.141787
       _cons |  -4.493444   .3612863   -12.44   0.000    -5.201552   -
3.785336
------------------------------------------------------------------------------

. matrix b=e(b)

. matrix v=e(V)

. ereturn post b v

. ereturn display, eform(exp_b)
------------------------------------------------------------------------------
             |      exp_b   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
dfail        |
 _Iannualt_1 |   .3714182   .1766073    -2.08   0.037     .1462587
.9432023
 _Iannualt_2 |   .5713866    .283651    -1.13   0.260     .2159592
1.511779
 _Iannualt_3 |   .1972107   .2010709    -1.59   0.111     .0267342
1.454768
 _Iannualt_4 |   .3706025   .3797503    -0.97   0.333     .0497389
2.761344
 _Iannualt_5 |   .8277393   .8550942    -0.18   0.855     .1092853
6.269392
 _Iannualt_7 |   11.88989   11.90671     2.47   0.013     1.670216
84.64146
    _Iloc2_2 |   4.978462   2.441803     3.27   0.001     1.903722
13.01928
    _Iloc2_3 |   .9581399   .4007006    -0.10   0.919     .4221343
2.174739
    _Iloc2_4 |   2.619792   1.393726     1.81   0.070     .9234829
7.431987
   _Itype2_2 |   1.376315   .7133093     0.62   0.538     .4983814
3.800792
   _Irace3_2 |   1.265389   .5851935     0.51   0.611      .511183
3.132362
------------------------------------------------------------------------------
```

```
. testparm _Iloc*

 ( 1)  [dfail]_Iloc2_2 = 0
 ( 2)  [dfail]_Iloc2_3 = 0
 ( 3)  [dfail]_Iloc2_4 = 0

           chi2(  3) =   14.61
         Prob > chi2 =    0.0022

. testparm _Iannualt*

 ( 1)  [dfail]_Iannualt_1 = 0
 ( 2)  [dfail]_Iannualt_2 = 0
 ( 3)  [dfail]_Iannualt_3 = 0
 ( 4)  [dfail]_Iannualt_4 = 0
 ( 5)  [dfail]_Iannualt_5 = 0
 ( 6)  [dfail]_Iannualt_7 = 0

           chi2(  6) =   15.71
         Prob > chi2 =    0.0154

. *Next I will evaluate the situation of multiple sites per patient and
multiple time intervals.  This will be a Discrete Proportional Odds model.
. *(D) Multiple sites per person and multiple time intervals
. *Discrete Proportional Odds.
. xi:logit dfail i.annualt i.loc2 i.type2 i.race3
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Iteration 0:   log likelihood = -580.95655
Iteration 1:   log likelihood = -575.74007
Iteration 2:   log likelihood = -555.16986
Iteration 3:   log likelihood = -547.71192
Iteration 4:   log likelihood = -546.98009
Iteration 5:   log likelihood = -546.97555
Iteration 6:   log likelihood = -546.97555

Logit estimates                                   Number of obs =11217
                                                  LR chi2(12) = 67.96
                                                  Prob > chi2] = 0.0000
Log likelihood = -546.97555                       Pseudo R2 = 0.0585

-----------------------------------------------------------------------------
       dfail |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
 _Iannualt_1 |  -.5987292    .251468    -2.38   0.017    -1.091597    -
.105861
 _Iannualt_2 |   -.870447   .3622894    -2.40   0.016    -1.580521   -
.1603728
```

```
 _Iannualt_3 |  -1.104999   .5209809    -2.12   0.034    -2.126103   -
.0838952
 _Iannualt_4 |  -.5233748   .5228846    -1.00   0.317     -1.54821
.5014603
 _Iannualt_5 |  -.0529839   .6011675    -0.09   0.930     -1.23125
1.125283
 _Iannualt_6 |  -.3296137   1.020678    -0.32   0.747    -2.330105
1.670878
 _Iannualt_7 |   3.322425   .6178347     5.38   0.000     2.111491
4.533359
    _Iloc2_2 |   1.547903   .3022753     5.12   0.000     .9554539
2.140351
    _Iloc2_3 |   .3409464   .2504324     1.36   0.173    -.1498922
.8317849
    _Iloc2_4 |   1.221204   .3344616     3.65   0.000     .5656716
1.876737
   _Itype2_2 |   .3562401   .3649522     0.98   0.329    -.3590532
1.071533
   _Irace3_2 |   .5815332   .2540745     2.29   0.022     .0835563
1.07951
       _cons |  -4.830381   .1933834   -24.98   0.000    -5.209406   -
4.451357
------------------------------------------------------------------------------

. logit,or

Logit estimates                                   Number of obs =11217
                                                  LR chi2(12) =67.96
                                                  Prob > chi2 = 0.0000
Log likelihood = -546.97555                       Pseudo R2 =0.0585

------------------------------------------------------------------------------
       dfail | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
 _Iannualt_1 |   .5495095    .138184    -2.38   0.017     .3356798
.8995496
 _Iannualt_2 |   .4187643   .1517139    -2.40   0.016     .2058678
.8518262
 _Iannualt_3 |   .3312112   .1725547    -2.12   0.034     .1193013
.9195276
 _Iannualt_4 |   .5925176   .3098183    -1.00   0.317     .2126283
1.651131
 _Iannualt_5 |   .9483953   .5701444    -0.09   0.930     .2919273
3.081088
 _Iannualt_6 |   .7192015    .734073    -0.32   0.747     .0972855
5.316834
 _Iannualt_7 |    27.7275   17.13101     5.38   0.000     8.260548
93.07063
    _Iloc2_2 |   4.701599   1.421177     5.12   0.000      2.59985
8.502425
    _Iloc2_3 |   1.406278   .3521775     1.36   0.173     .8608008
2.297416
```

```
    _Iloc2_4 |    3.39127    1.13425     3.65   0.000      1.76063
6.532157
   _Itype2_2 |    1.42795   .5211337     0.98   0.329     .6983372
2.919853
   _Irace3_2 |   1.788779   .4544831     2.29   0.022     1.087146
2.943237
------------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   32.63
         Prob > chi2 =    0.0000

. testparm _Iannualt*

 ( 1)  _Iannualt_1 = 0
 ( 2)  _Iannualt_2 = 0
 ( 3)  _Iannualt_3 = 0
 ( 4)  _Iannualt_4 = 0
 ( 5)  _Iannualt_5 = 0
 ( 6)  _Iannualt_6 = 0
 ( 7)  _Iannualt_7 = 0

           chi2(  7) =   47.32
         Prob > chi2 =    0.0000

. *Next I will incorporate the robust variance
. xi:logit dfail i.annualt i.loc2 i.type2 i.race3, robust cluster(id)
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Iteration 0:   log pseudo-likelihood = -580.95655
Iteration 1:   log pseudo-likelihood = -575.74007
Iteration 2:   log pseudo-likelihood = -555.16986
Iteration 3:   log pseudo-likelihood = -547.71192
Iteration 4:   log pseudo-likelihood = -546.98009
Iteration 5:   log pseudo-likelihood = -546.97555
Iteration 6:   log pseudo-likelihood = -546.97555
```

Logit estimates Number of obs = 11217
Wald chi2(12)=45.93
Prob > chi2 =0.0000
Log pseudo-likelihood = -546.97555 Pseudo R2 =0.0585

(standard errors adjusted for clustering on id)

dfail	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Iannualt_1	-.5987292	.3819597	-1.57	0.117	-1.347356	.149898
_Iannualt_2	-.870447	.4105935	-2.12	0.034	-1.675195	-.0656985
_Iannualt_3	-1.104999	.6580218	-1.68	0.093	-2.394698	.1847002
_Iannualt_4	-.5233748	.8091448	-0.65	0.518	-2.109269	1.06252
_Iannualt_5	-.0529839	.7742164	-0.07	0.945	-1.57042	1.464452
_Iannualt_6	-.3296137	1.03622	-0.32	0.750	-2.360568	1.70134
_Iannualt_7	3.322425	1.072642	3.10	0.002	1.220086	5.424764
_Iloc2_2	1.547903	.4025621	3.85	0.000	.7588954	2.33691
_Iloc2_3	.3409464	.247351	1.38	0.168	-.1438527	.8257454
_Iloc2_4	1.221204	.3990595	3.06	0.002	.4390621	2.003347
_Itype2_2	.3562401	.4451742	0.80	0.424	-.5162852	1.228765
_Irace3_2	.5815332	.4001928	1.45	0.146	-.2028302	1.365897
_cons	-4.830381	.2521104	-19.16	0.000	-5.324508	-4.336254

```
. logit, or
Logit estimates                                   Number of obs =11217
                                                  Wald chi2(12) = 45.93
                                                  Prob > chi2 = 0.0000
Log pseudo-likelihood = -546.97555                Pseudo R2 =0.0585

                             (standard errors adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
       dfail | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
 _Iannualt_1 |   .5495095   .2098905    -1.57   0.117     .2599265
1.161716
 _Iannualt_2 |   .4187643   .1719419    -2.12   0.034     .1872716
.9364131
 _Iannualt_3 |   .3312112   .2179442    -1.68   0.093     .0912002
1.202858
 _Iannualt_4 |   .5925176   .4794325    -0.65   0.518     .1213266
2.893653
 _Iannualt_5 |   .9483953   .7342632    -0.07   0.945     .2079578
4.325174
 _Iannualt_6 |   .7192015    .745251    -0.32   0.750     .0943666
5.481289
 _Iannualt_7 |    27.7275   29.74168     3.10   0.002     3.387478
226.9577
    _Iloc2_2 |   4.701599   1.892686     3.85   0.000     2.135916
10.34921
    _Iloc2_3 |   1.406278   .3478442     1.38   0.168     .8660153
2.283582
    _Iloc2_4 |    3.39127   1.353318     3.06   0.002     1.551252
7.413826
   _Itype2_2 |    1.42795   .6356866     0.80   0.424     .5967332
3.417008
   _Irace3_2 |   1.788779   .7158564     1.45   0.146     .8164168
3.919236
------------------------------------------------------------------------------

. testparm _Iloc*
 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0
           chi2(  3) =   15.42
         Prob > chi2 =    0.0015
```

```
. testparm _Iannualt*

 ( 1)  _Iannualt_1 = 0
 ( 2)  _Iannualt_2 = 0
 ( 3)  _Iannualt_3 = 0
 ( 4)  _Iannualt_4 = 0
 ( 5)  _Iannualt_5 = 0
 ( 6)  _Iannualt_6 = 0
 ( 7)  _Iannualt_7 = 0

           chi2(  7) =   19.39
         Prob > chi2 =    0.0070

. *Now we will analyze the data using GEE for the Discrete Proportional Odds
an
> d Cloglog Models.
. set matsize 110

. xi:xtgee dfail i.annualt i.loc2 i.type2 i.race3, family(bin) link(logit)
corr
> (exc) i(id) eform
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Iteration 1: tolerance = .43971565
Iteration 2: tolerance = .03868157
Iteration 3: tolerance = .01028691
Iteration 4: tolerance = .0022018
Iteration 5: tolerance = .00046132
Iteration 6: tolerance = .00009796
Iteration 7: tolerance = .00002114
Iteration 8: tolerance = 4.619e-06
Iteration 9: tolerance = 1.018e-06
Iteration 10: tolerance = 2.257e-07

GEE population-averaged model                   Number of obs = 11217
Group variable: id                              Number of groups = 732
Link: logit                                     Obs per group: min = 1
Family: binomial                                            avg =15.3
Correlation: exchangeable                                   max =106
                                                Wald chi2(12)=64.41
Scale parameter:1                                Prob > chi2 = 0.0000

-----------------------------------------------------------------------------
       dfail | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
 _Iannualt_1 |    .700332     .16206    -1.54   0.124     .4449714
1.102239
 _Iannualt_2 |   .6281056   .1936827    -1.51   0.132     .3432069
1.149501
```

```
 _Iannualt_3 |   .5517458   .2289641    -1.43   0.152     .2446282
1.244433
 _Iannualt_4 |   .8111113   .3673954    -0.46   0.644     .3338306
1.970765
 _Iannualt_5 |   1.361996   .6852592     0.61   0.539     .5080561
3.651234
 _Iannualt_6 |    1.21016   .9466859     0.24   0.807     .2611943
5.60689
 _Iannualt_7 |   28.66883    17.4287     5.52   0.000     8.708378
94.38058
    _Iloc2_2 |   4.055249   1.348378     4.21   0.000     2.113448
7.781146
    _Iloc2_3 |   1.382045   .3409234     1.31   0.190     .8522114
2.241284
    _Iloc2_4 |   3.077276   1.122668     3.08   0.002     1.505313
6.290805
   _Itype2_2 |   1.346909   .5703833     0.70   0.482     .5873202
3.088883
   _Irace3_2 |   1.930223   .5635115     2.25   0.024     1.089198
3.420647
-----------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   21.17
         Prob > chi2 =    0.0001

. testparm _Iannualt*

 ( 1)  _Iannualt_1 = 0
 ( 2)  _Iannualt_2 = 0
 ( 3)  _Iannualt_3 = 0
 ( 4)  _Iannualt_4 = 0
 ( 5)  _Iannualt_5 = 0
 ( 6)  _Iannualt_6 = 0
 ( 7)  _Iannualt_7 = 0

           chi2(  7) =   39.58
         Prob > chi2 =    0.0000

. xi:xtgee dfail i.annualt i.loc2 i.type2 i.race3, family(bin) link(logit)
corr
> (exc) i(id) eform robust
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)
```

```
Iteration 1: tolerance = .43971565
Iteration 2: tolerance = .03868157
Iteration 3: tolerance = .01028691
Iteration 4: tolerance = .0022018
Iteration 5: tolerance = .00046132
Iteration 6: tolerance = .00009796
Iteration 7: tolerance = .00002114
Iteration 8: tolerance = 4.619e-06
Iteration 9: tolerance = 1.018e-06
Iteration 10: tolerance = 2.257e-07

GEE population-averaged model                 Number of obs =11217
Group variable: id                            Number of groups = 732
Link: logit                                   Obs per group: min = 1
Family: binomial                              avg = 15.3
Correlation: exchangeable                     max =106
                                              Wald chi2(12)=42.92
Scale parameter: 1                            Prob > chi2 = 0.0000

                             (standard errors adjusted for clustering on id)
```

dfail	Odds Ratio	Semi-robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
_Iannualt_1	.700332	.2164383	-1.15	0.249	.3821548	1.28342
_Iannualt_2	.6281056	.1885218	-1.55	0.121	.3487788	1.131137
_Iannualt_3	.5517458	.2339718	-1.40	0.161	.240315	1.266768
_Iannualt_4	.8111113	.4212591	-0.40	0.687	.2930894	2.244713
_Iannualt_5	1.361996	.7599633	0.55	0.580	.4562723	4.065625
_Iannualt_6	1.21016	.7288949	0.32	0.751	.3716665	3.940328
_Iannualt_7	28.66883	29.88974	3.22	0.001	3.714998	221.2388
_Iloc2_2	4.055249	1.408705	4.03	0.000	2.052716	8.01136
_Iloc2_3	1.382045	.2795556	1.60	0.110	.929702	2.054473
_Iloc2_4	3.077276	1.167701	2.96	0.003	1.46275	6.473853
_Itype2_2	1.346909	.6323171	0.63	0.526	.536704	3.380193
_Irace3_2	1.930223	.7799835	1.63	0.104	.8742698	4.261568

```
. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   16.60
         Prob > chi2 =    0.0009

. testparm _Iannualt*

 ( 1)  _Iannualt_1 = 0
 ( 2)  _Iannualt_2 = 0
 ( 3)  _Iannualt_3 = 0
 ( 4)  _Iannualt_4 = 0
 ( 5)  _Iannualt_5 = 0
 ( 6)  _Iannualt_6 = 0
 ( 7)  _Iannualt_7 = 0

           chi2(  7) =   16.46
         Prob > chi2 =    0.0212

. *Now for the Cloglog evaluation of GEE
. xi:xtcloglog dfail i.annualt i.loc2 i.type2 i.race3, pa i(id)
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Iteration 1: tolerance = .43576416
Iteration 2: tolerance = .04073009
Iteration 3: tolerance = .01053104
Iteration 4: tolerance = .0022568
Iteration 5: tolerance = .00047061
Iteration 6: tolerance = .0000994
Iteration 7: tolerance = .00002126
Iteration 8: tolerance = 4.598e-06
Iteration 9: tolerance = 1.002e-06
Iteration 10: tolerance = 2.196e-07
```

```
GEE population-averaged model                   Number of obs =11217
Group variable: id                              Number of groups = 732
Link: cloglog                                   Obs per group: min = 1
Family: binomial                                avg =15.3
Correlation: exchangeable                       max = 106
                                                Wald chi2(12) =72.47
Scale parameter: 1                              Prob > chi2=0.0000

-----------------------------------------------------------------------------
       dfail |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
 _Iannualt_1 |  -.3497581    .229746    -1.52   0.128     -.800052
.1005358
 _Iannualt_2 |  -.4604452    .306653    -1.50   0.133    -1.061474
.1405838
 _Iannualt_3 |  -.5895319   .4131385    -1.43   0.154    -1.399268
.2202047
 _Iannualt_4 |  -.2111567   .4509183    -0.47   0.640     -1.09494
.6726269
 _Iannualt_5 |   .3063307   .4988064     0.61   0.539    -.6713119
1.283973
 _Iannualt_6 |   .1820401   .7774632     0.23   0.815     -1.34176
1.70584
 _Iannualt_7 |   3.181051   .5293436     6.01   0.000     2.143557
4.218545
    _Iloc2_2 |   1.394244   .3283655     4.25   0.000     .7506597
2.037829
    _Iloc2_3 |   .3333388   .2432716     1.37   0.171    -.1434647
.8101424
    _Iloc2_4 |   1.124071   .3609903     3.11   0.002      .416543
1.831599
   _Itype2_2 |   .3553732   .4048151     0.88   0.380    -.4380498
1.148796
   _Irace3_2 |   .6282387   .2877997     2.18   0.029     .0641617
1.192316
       _cons |  -4.772188   .2105216   -22.67   0.000    -5.184803   -
4.359573
-----------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   21.48
         Prob > chi2 =    0.0001
```

```
. testparm _Iannualt*

 ( 1)  _Iannualt_1 = 0
 ( 2)  _Iannualt_2 = 0
 ( 3)  _Iannualt_3 = 0
 ( 4)  _Iannualt_4 = 0
 ( 5)  _Iannualt_5 = 0
 ( 6)  _Iannualt_6 = 0
 ( 7)  _Iannualt_7 = 0

           chi2(  7) =   46.15
         Prob > chi2 =    0.0000

. matrix b=e(b)

. matrix v=e(V)

. ereturn post b v

. ereturn display, eform(exp_b)
------------------------------------------------------------------------------
             |      exp_b   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
 _Iannualt_1 |   .7048586   .1619384    -1.52   0.128     .4493056
1.105763
 _Iannualt_2 |   .6310027   .1934989    -1.50   0.133     .3459455
1.150945
 _Iannualt_3 |   .5545868   .2291212    -1.43   0.154     .2467774
1.246332
 _Iannualt_4 |   .8096472   .3650847    -0.47   0.640     .3345596
1.959378
 _Iannualt_5 |   1.358432   .6775944     0.61   0.539     .5110377
3.610959
 _Iannualt_6 |   1.199662   .9326933     0.23   0.815     .2613853
5.506009
 _Iannualt_7 |   24.07204   12.74238     6.01   0.000      8.52972
67.9346
    _Iloc2_2 |   4.031927   1.323946     4.25   0.000     2.118397
7.67393
    _Iloc2_3 |    1.39562   .3395147     1.37   0.171     .8663513
2.248228
    _Iloc2_4 |   3.077356   1.110896     3.11   0.002     1.516709
6.243862
   _Itype2_2 |   1.426713    .577555     0.88   0.380     .6452937
3.154393
   _Irace3_2 |   1.874306   .5394248     2.18   0.029     1.066265
3.294702
------------------------------------------------------------------------------
```

```
. xi:xtcloglog dfail i.annualt i.loc2 i.type2 i.race3, pa robust i(id)
i.annualt         _Iannualt_0-7       (naturally coded; _Iannualt_0 omitted)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

Iteration 1: tolerance = .43576416
Iteration 2: tolerance = .04073009
Iteration 3: tolerance = .01053104
Iteration 4: tolerance = .0022568
Iteration 5: tolerance = .00047061
Iteration 6: tolerance = .0000994
Iteration 7: tolerance = .00002126
Iteration 8: tolerance = 4.598e-06
Iteration 9: tolerance = 1.002e-06
Iteration 10: tolerance = 2.196e-07

GEE population-averaged model                    Number of obs =11217
Group variable: id                               Number of groups =732
Link: cloglog                                    Obs per group: min =1
Family: binomial                                 avg =15.3
Correlation: exchangeable                        max =106
                                                 Wald chi2(12) =45.89
Scale parameter: 1                               Prob > chi2=0.0000

                             (standard errors adjusted for clustering on id)
-----------------------------------------------------------------------------
             |             Semi-robust
       dfail |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
 _Iannualt_1 |  -.3497581   .3071623    -1.14   0.255    -.9517851
.2522689
 _Iannualt_2 |  -.4604452   .2987017    -1.54   0.123     -1.04589
.1249993
 _Iannualt_3 |  -.5895319     .42337    -1.39   0.164    -1.419322
.240258
 _Iannualt_4 |  -.2111567   .5168515    -0.41   0.683    -1.224167
.8018536
 _Iannualt_5 |   .3063307    .554254     0.55   0.580    -.7799872
1.392649
 _Iannualt_6 |   .1820401   .5986252     0.30   0.761    -.9912436
1.355324
 _Iannualt_7 |   3.181051   .8879963     3.58   0.000      1.44061
4.921492
    _Iloc2_2 |   1.394244     .34364     4.06   0.000     .7207223
2.067766
    _Iloc2_3 |   .3333388   .1997842     1.67   0.095     -.058231
.7249087
    _Iloc2_4 |   1.124071   .3761522     2.99   0.003     .3868262
1.861316
```

```
   _Itype2_2 |   .3553732   .4426169     0.80   0.422      -.51214
1.222886
   _Irace3_2 |   .6282387   .4001439     1.57   0.116    -.1560289
1.412506
       _cons |  -4.772188   .2395996   -19.92   0.000    -5.241795   -
4.302581
------------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   16.85
         Prob > chi2 =    0.0008

. testparm _Iannualt*

 ( 1)  _Iannualt_1 = 0
 ( 2)  _Iannualt_2 = 0
 ( 3)  _Iannualt_3 = 0
 ( 4)  _Iannualt_4 = 0
 ( 5)  _Iannualt_5 = 0
 ( 6)  _Iannualt_6 = 0
 ( 7)  _Iannualt_7 = 0

           chi2(  7) =   19.32
         Prob > chi2 =    0.0072

. matrix b=e(b)

. matrix v=e(V)

. ereturn post b v

. ereturn display, eform(exp_b)
------------------------------------------------------------------------------
             |      exp_b   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
 _Iannualt_1 |   .7048586    .216506    -1.14   0.255     .3860513
1.286942
 _Iannualt_2 |   .6310027   .1884816    -1.54   0.123     .3513791
1.133148
 _Iannualt_3 |   .5545868   .2347954    -1.39   0.164      .241878
1.271577
 _Iannualt_4 |   .8096472   .4184674    -0.41   0.683     .2940025
2.22967
 _Iannualt_5 |   1.358432   .7529162     0.55   0.580     .4584119
4.025498
 _Iannualt_6 |   1.199662    .718148     0.30   0.761     .3711149
3.878017
```

```
 _Iannualt_7 |   24.07204   21.37588     3.58   0.000     4.223272
137.2071
    _Iloc2_2 |   4.031927   1.385531     4.06   0.000     2.055918
7.907141
    _Iloc2_3 |    1.39562   .2788229     1.67   0.095      .943432
2.064543
    _Iloc2_4 |   3.077356   1.157554     2.99   0.003     1.472301
6.432194
   _Itype2_2 |   1.426713   .6314873     0.80   0.422     .5992119
3.396979
   _Irace3_2 |   1.874306   .7499923     1.57   0.116     .8555345
4.106234
-----------------------------------------------------------------------------

. *The next situation to evaluate involves the single site per patient with
continuous time.  This would involve the Cox model and I will now evaluate
this on the final9.dta dataset.
. /*save "C:\unzipped\final9Folder\final9b.dta", replace
> clear*/
. use "C:\unzipped\final9Folder\final9.dta", clear

. desc

Contains data from C:\unzipped\final9Folder\final9.dta
  obs:         7,986
 vars:           121
 size:     4,016,958 (68.1% of memory free)

. *(E)Single site per patient and continuous time Cox Proportional Hazards
Model
. count if firstimp==1
 2609

. xi: stcox  i.loc2 i.type2 i.race3 if firstimp==1
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

         failure _d:  failure
   analysis time _t:  (followup-origin)/365.25
             origin:  time place
                 id:  newid2

Iteration 0:   log likelihood = -222.01536
Iteration 1:   log likelihood = -219.82572
Iteration 2:   log likelihood = -217.08173
Iteration 3:   log likelihood = -217.02842
Iteration 4:   log likelihood = -217.02835
Refining estimates:
Iteration 0:   log likelihood = -217.02835

Cox regression -- Breslow method for ties
```

```
No. of subjects = 732                              Number of obs=2483
No. of failures = 37
Time at risk =1582.020534
                                                      LR chi2(5) =9.97
Log likelihood =-217.02835                       Prob > chi2=0.0760

-----------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
    _Iloc2_2 |   4.035504   1.915437     2.94   0.003     1.591762
10.23098
    _Iloc2_3 |   .8784375   .3730693    -0.31   0.760     .3821278
2.019357
    _Iloc2_4 |   2.045869   1.087823     1.35   0.178     .7215719
5.800642
   _Itype2_2 |   1.583092   .9157907     0.79   0.427     .5094496
4.919389
   _Irace3_2 |   1.312117   .5929124     0.60   0.548     .5411727
3.181332
-----------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   11.66
         Prob > chi2 =    0.0086

. xi: stcox  i.loc2 i.type2 i.race3 if firstimp==1, robust cluster(id)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

         failure _d:  failure
   analysis time _t:  (followup-origin)/365.25
             origin:  time place
                 id:  newid2

Iteration 0:   log pseudo-likelihood = -222.01536
Iteration 1:   log pseudo-likelihood = -219.82572
Iteration 2:   log pseudo-likelihood = -217.08173
Iteration 3:   log pseudo-likelihood = -217.02842
Iteration 4:   log pseudo-likelihood = -217.02835
Refining estimates:
Iteration 0:   log pseudo-likelihood = -217.02835

Cox regression -- Breslow method for ties
```

```
No. of subjects = 732                              Number of obs = 2483
No. of failures = 37
Time at risk = 1582.020534
                                                   Wald chi2(5)= 11.81
Log pseudo-likelihood = -217.02835                 Prob > chi2 = 0.0376

                              (standard errors adjusted for clustering on id)
-----------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
    _Iloc2_2 |   4.035504    1.92232     2.93   0.003      1.58645
10.26524
    _Iloc2_3 |   .8784375   .3571286    -0.32   0.750     .3959634
1.948797
    _Iloc2_4 |   2.045869   1.085041     1.35   0.177     .7234974
5.785204
   _Itype2_2 |   1.583092   .8014075     0.91   0.364     .5869528
4.269817
   _Irace3_2 |   1.312117    .600089     0.59   0.553     .5354023
3.215619
-----------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   11.58
         Prob > chi2 =    0.0090

. *The next situation to evaluate involves multiple sites per patient with
continuous time.  This also would involve the Cox model.
. *(F) Multiple sites per patient and continuous time Cox Proportional
Hazards Model
. xi: stcox  i.loc2 i.type2 i.race3
i.loc2            _Iloc2_1-4         (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2        (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2        (naturally coded; _Irace3_1 omitted)

         failure _d:  failure
   analysis time _t:  (followup-origin)/365.25
             origin:  time place
                 id:  newid2

Iteration 0:   log likelihood = -707.69229
Iteration 1:   log likelihood = -697.89642
Iteration 2:   log likelihood = -694.96837
Iteration 3:   log likelihood = -694.92372
Iteration 4:   log likelihood = -694.92369
```

```
Refining estimates:
Iteration 0:   log likelihood = -694.92369

Cox regression -- Breslow method for ties

No. of subjects = 2174                          Number of obs = 7633
No. of failures =102
Time at risk = 4694.20397
                                                   LR chi2(5)= 25.54
Log likelihood  = -694.92369                     Prob > chi2 = 0.0001

-----------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
    _Iloc2_2 |   3.847636   1.144967     4.53   0.000     2.147318
6.894324
    _Iloc2_3 |   1.294654   .3192834     1.05   0.295     .7984228
2.099299
    _Iloc2_4 |   2.635487   .8691616     2.94   0.003     1.380834
5.030139
   _Itype2_2 |   1.542095   .5613253     1.19   0.234     .7555655
3.147386
   _Irace3_2 |   1.796049   .4503715     2.34   0.020     1.098686
2.936045
-----------------------------------------------------------------------------
. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   24.44
         Prob > chi2 =    0.0000

. xi: stcox  i.loc2 i.type2 i.race3 , robust cluster(id)
i.loc2            _Iloc2_1-4          (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2         (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2         (naturally coded; _Irace3_1 omitted)

         failure _d:  failure
   analysis time _t:  (followup-origin)/365.25
             origin:  time place
                 id:  newid2

Iteration 0:   log pseudo-likelihood = -707.69229
Iteration 1:   log pseudo-likelihood = -697.89642
Iteration 2:   log pseudo-likelihood = -694.96837
Iteration 3:   log pseudo-likelihood = -694.92372
Iteration 4:   log pseudo-likelihood = -694.92369
```

```
Refining estimates:
Iteration 0:   log pseudo-likelihood = -694.92369

Cox regression -- Breslow method for ties

No. of subjects = 2174                         Number of obs =7633
No. of failures = 102
Time at risk = 4694.20397
                                               Wald chi2(5)= 16.40
Log pseudo-likelihood = -694.92369             Prob > chi2 = 0.0058

                              (standard errors adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
    _Iloc2_2 |   3.847636    1.55498     3.33   0.001      1.74257
8.495672
    _Iloc2_3 |   1.294654   .3076313     1.09   0.277      .812632
2.062592
    _Iloc2_4 |   2.635487   1.031562     2.48   0.013     1.223743
5.675859
   _Itype2_2 |   1.542095   .6704176     1.00   0.319     .6577421
3.615484
   _Irace3_2 |   1.796049    .697502     1.51   0.132     .8389788
3.844902
------------------------------------------------------------------------------

. testparm _Iloc*

 ( 1)  _Iloc2_2 = 0
 ( 2)  _Iloc2_3 = 0
 ( 3)  _Iloc2_4 = 0

           chi2(  3) =   11.34
         Prob > chi2 =    0.0100

end of do-file
```

Log file for Frailty Model 11 and data in Table 16

```
. xi:stcox i.loc2 i.type2 i.race3, shared(id)
i.loc2            _Iloc2_1-4         (naturally coded; _Iloc2_1 omitted)
i.type2           _Itype2_1-2        (naturally coded; _Itype2_1 omitted)
i.race3           _Irace3_1-2        (naturally coded; _Irace3_1 omitted)
         failure _d:  failure
   analysis time _t:  (followup-origin)/365.25
             origin:  time place
Fitting comparison Cox model:
Estimating frailty variance:
Iteration 0:   log profile likelihood = -690.08355
Iteration 1:   log profile likelihood = -685.99749
Iteration 2:   log profile likelihood = -685.99734
Iteration 3:   log profile likelihood = -685.99734

Fitting final Cox model:
Iteration 0:   log likelihood = -988.28436
Iteration 1:   log likelihood = -873.39136
Iteration 2:   log likelihood = -742.52786
Iteration 3:   log likelihood =    -696.17
Iteration 4:   log likelihood = -687.09208
Iteration 5:   log likelihood = -686.04859
Iteration 6:   log likelihood = -685.99776
Iteration 7:   log likelihood = -685.99734
Iteration 8:   log likelihood = -685.99734
Refining estimates:
Iteration 0:   log likelihood = -685.99734

Cox regression -- Breslow method for ties       Number of obs = 7633
Gamma shared frailty                            Number of groups = 732
Group variable: id
No. of subjects = 7633                          Obs per group: min = 1
No. of failures = 102                           avg =10.4276
Time at risk = 14366.74606                      max = 95
                                                Wald chi2(5)= 18.75
Log likelihood = -685.99734                     Prob > chi2 = 0.0021
-----------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+---------------------------------------------------------------
    _Iloc2_2 |   5.759403   4.227933     2.39   0.017     1.366209
24.27939
    _Iloc2_3 |   1.511322   .5351903     1.17   0.244     .7549687
3.025416
    _Iloc2_4 |   5.823392   4.594466     2.23   0.026     1.240526
27.33671
   _Itype2_2 |   1.158311   1.124599     0.15   0.880     .1727419
7.766989
   _Irace3_2 |   17.74163   16.08504     3.17   0.002     3.001038
104.8856
-------------+---------------------------------------------------------------
       theta |   29.63646   6.384354
-----------------------------------------------------------------------------
Likelihood-ratio test of theta=0: chibar2(01) = 225.07 Prob>=chibar2 = 0.000

Note: Standard errors of hazard ratios are conditional on theta.
```

Log file for data in Table 17 Table 18 and Table 19

```
----------------------------------------------------------------------------
. sort failure

. iis id
. by failure:xttab loc if year==1
----------------------------------------------------------------------------
-> failure = 0
                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      54     2.01         33     5.66           42.52
        2 |     166     6.18         54     9.26           69.17
        3 |      50     1.86         28     4.80           38.76
        4 |     342    12.74        164    28.13           34.76
        5 |    1702    63.41        402    68.95           79.02
        6 |     370    13.79        177    30.36           34.87
----------+-----------------------------------------------------
    Total |    2684   100.00        858   147.17           58.11
                               (n = 583)
----------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       2     3.70          2     5.56           28.57
        2 |       6    11.11          5    13.89           60.00
        3 |       3     5.56          3     8.33           37.50
        4 |       6    11.11          6    16.67           54.55
        5 |      28    51.85         20    55.56           90.32
        6 |       9    16.67          9    25.00           56.25
----------+-----------------------------------------------------
    Total |      54   100.00         45   125.00           69.10
                                (n = 36)

. by failure:xttab loc if year==2
----------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      89     3.68         34     7.17           38.20
        2 |     122     5.04         37     7.81           46.21
        3 |      82     3.39         27     5.70           38.14
        4 |     329    13.60        142    29.96           33.99
        5 |    1512    62.51        327    68.99           79.04
        6 |     285    11.78        135    28.48           33.33
----------+-----------------------------------------------------
    Total |    2419   100.00        702   148.10           55.85
                               (n = 474)
```

```
-----------------------------------------------------------------------------

-> failure = 1

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1     4.35          1     8.33         100.00
        2 |        8    34.78          5    41.67         100.00
        4 |        3    13.04          3    25.00          27.27
        5 |        5    21.74          3    25.00          55.56
        6 |        6    26.09          5    41.67          46.15
----------+-----------------------------------------------------
    Total |       23   100.00         17   141.67          63.49
                                 (n = 12)

. by failure:xttab loc if year==3

-----------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       46     3.81         21     7.34          40.71
        2 |       73     6.04         21     7.34          50.00
        3 |       52     4.30         20     6.99          43.33
        4 |      142    11.75         77    26.92          34.89
        5 |      764    63.25        208    72.73          81.62
        6 |      131    10.84         68    23.78          36.49
----------+-----------------------------------------------------
    Total |     1208   100.00        415   145.10          60.04
                                (n = 286)

-----------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        3    30.00          2    22.22         100.00
        2 |        1    10.00          1    11.11         100.00
        5 |        6    60.00          6    66.67         100.00
----------+-----------------------------------------------------
    Total |       10   100.00          9   100.00         100.00
                                  (n = 9)
```

```
. by failure:xttab loc if year==4

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        8     1.04          7     3.59          57.14
        2 |       26     3.38         11     5.64          76.47
        3 |       11     1.43          6     3.08          55.00
        4 |       82    10.65         46    23.59          34.75
        5 |      578    75.06        154    78.97          85.88
        6 |       65     8.44         38    19.49          32.66
----------+-----------------------------------------------------
    Total |      770   100.00        262   134.36          67.32
                               (n = 195)

-------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1    25.00          1    33.33         100.00
        2 |        1    25.00          1    33.33          50.00
        3 |        1    25.00          1    33.33          50.00
        5 |        1    25.00          1    33.33         100.00
----------+-----------------------------------------------------
    Total |        4   100.00          4   133.33          75.00
                                 (n = 3)

. by failure:xttab loc if year==5

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent

----------+-----------------------------------------------------
        1 |        2     0.44          2     1.87          50.00
        2 |       12     2.65          5     4.67         100.00
        3 |        7     1.55          3     2.80          87.50
        4 |       41     9.05         29    27.10          29.29
        5 |      353    77.92         89    83.18          84.86
        6 |       38     8.39         21    19.63          30.65
----------+-----------------------------------------------------
    Total |      453   100.00        149   139.25          66.49
                               (n = 107)
```

```
-------------------------------------------------------------------------------
-> failure = 1
                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        4 |        1    25.00          1    50.00          33.33
        5 |        3    75.00          2   100.00          75.00
----------+-----------------------------------------------------
    Total |        4   100.00          3   150.00          61.11
                                 (n = 2)

. by failure:xttab loc if year==6

-------------------------------------------------------------------------------
-> failure = 0
                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1     0.44          1     1.75          25.00
        3 |        3     1.32          1     1.75          75.00
        4 |       26    11.40         16    28.07          38.24
        5 |      182    79.82         50    87.72          85.45
        6 |       16     7.02         11    19.30          30.19
----------+-----------------------------------------------------
    Total |      228   100.00         79   138.60          67.29
                                (n = 57)

-------------------------------------------------------------------------------
-> failure = 1
                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1    33.33          1    50.00         100.00
        5 |        2    66.67          1    50.00         100.00
----------+-----------------------------------------------------
    Total |        3   100.00          2   100.00         100.00
                                 (n = 2)

. by failure:xttab loc if year==7

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        4 |       12    11.01          8    33.33          35.29
        5 |       90    82.57         21    87.50          86.54
        6 |        7     6.42          5    20.83          31.82
----------+-----------------------------------------------------
    Total |      109   100.00         34   141.67          66.43
                                (n = 24)
```

```
------------------------------------------------------------------------------

-> failure = 1

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        5 |        1   100.00          1   100.00         100.00
----------+-----------------------------------------------------
    Total |        1   100.00          1   100.00         100.00
                                 (n = 1)

. by failure:xttab loc if year==8

------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        4 |        1     8.33          1    20.00          33.33
        5 |        9    75.00          4    80.00          90.00
        6 |        2    16.67          1    20.00         100.00
----------+-----------------------------------------------------
    Total |       12   100.00          6   120.00          82.22
                                 (n = 5)

------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
      loc |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        4 |        1    25.00          1   100.00          25.00
        5 |        2    50.00          1   100.00          50.00
        6 |        1    25.00          1   100.00          25.00
----------+-----------------------------------------------------
    Total |        4   100.00          3   300.00          33.33
                                 (n = 1)

. by failure:xttab loc2 if year==1

------------------------------------------------------------------------------
-> failure = 0
                  Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     1702    63.41        402    68.95          79.02
        2 |      166     6.18         54     9.26          69.17
        3 |      712    26.53        243    41.68          55.71
        4 |      104     3.87         46     7.89          61.90
----------+-----------------------------------------------------
    Total |     2684   100.00        745   127.79          69.64
                               (n = 583)
```

```
--------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       28    51.85         20    55.56          90.32
        2 |        6    11.11          5    13.89          60.00
        3 |       15    27.78         12    33.33          78.95
        4 |        5     9.26          3     8.33          62.50
----------+-----------------------------------------------------
    Total |       54   100.00         40   111.11          81.03
                                   (n = 36)

. by failure:xttab loc2 if year==2

--------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     1512    62.51        327    68.99          79.04
        2 |      122     5.04         37     7.81          46.21
        3 |      614    25.38        203    42.83          50.33
        4 |      171     7.07         46     9.70          60.64
----------+-----------------------------------------------------
    Total |     2419   100.00        613   129.32          66.17
                                  (n = 474)

--------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        5    21.74          3    25.00          55.56
        2 |        8    34.78          5    41.67         100.00
        3 |        9    39.13          5    41.67          69.23
        4 |        1     4.35          1     8.33         100.00
----------+-----------------------------------------------------
    Total |       23   100.00         14   116.67          79.49
                                   (n = 12)
```

```
. by failure:xttab loc2 if year==3

-------------------------------------------------------------------------------
-> failure = 0

                  Overall              Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      764    63.25        208    72.73          81.62
        2 |       73     6.04         21     7.34          50.00
        3 |      273    22.60        104    36.36          53.22
        4 |       98     8.11         30    10.49          64.90
----------+-----------------------------------------------------
    Total |     1208   100.00        363   126.92          70.27
                               (n = 286)

-------------------------------------------------------------------------------
-> failure = 1

                  Overall              Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        6    60.00          6    66.67         100.00
        2 |        1    10.00          1    11.11         100.00
        4 |        3    30.00          2    22.22         100.00
----------+-----------------------------------------------------
    Total |       10   100.00          9   100.00         100.00
                                 (n = 9)

. by failure:xttab loc2 if year==4

-------------------------------------------------------------------------------
-> failure = 0

                  Overall              Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      578    75.06        154    78.97          85.88
        2 |       26     3.38         11     5.64          76.47
        3 |      147    19.09         60    30.77          52.31
        4 |       19     2.47         11     5.64          67.86
----------+-----------------------------------------------------
    Total |      770   100.00        236   121.03          76.07
                               (n = 195)

-------------------------------------------------------------------------------
-> failure = 1

                  Overall              Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1    25.00          1    33.33         100.00
        2 |        1    25.00          1    33.33          50.00
        4 |        2    50.00          2    66.67          66.67
----------+-----------------------------------------------------
    Total |        4   100.00          4   133.33          70.83
```

```
                              (n = 3)

. by failure:xttab loc2 if year==5

----------------------------------------------------------------------------
-> failure = 0

                Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      353     77.92         89     83.18          84.86
        2 |       12      2.65          5      4.67         100.00
        3 |       79     17.44         35     32.71          47.02
        4 |        9      1.99          4      3.74         100.00
----------+-----------------------------------------------------
    Total |      453    100.00        133    124.30          75.92
                            (n = 107)

----------------------------------------------------------------------------
--
-> failure = 1

                Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        3     75.00          2    100.00          75.00
        3 |        1     25.00          1     50.00          33.33
----------+-----------------------------------------------------
    Total |        4    100.00          3    150.00          61.11
                              (n = 2)

. by failure:xttab loc2 if year==6

----------------------------------------------------------------------------
-> failure = 0

                Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      182     79.82         50     87.72          85.45
        3 |       42     18.42         20     35.09          51.85
        4 |        4      1.75          1      1.75         100.00
----------+-----------------------------------------------------
    Total |      228    100.00         71    124.56          76.19
                             (n = 57)
```

```
-------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        2    66.67          1    50.00         100.00
        4 |        1    33.33          1    50.00         100.00
----------+-----------------------------------------------------
    Total |        3   100.00          2   100.00         100.00
                                 (n = 2)

. by failure:xttab loc2 if year==7

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       90    82.57         21    87.50          86.54
        3 |       19    17.43          9    37.50          52.78
----------+-----------------------------------------------------
    Total |      109   100.00         30   125.00          76.41
                                (n = 24)

-------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1   100.00          1   100.00         100.00
----------+-----------------------------------------------------
    Total |        1   100.00          1   100.00         100.00
                                 (n = 1)

. by failure:xttab loc2 if year==8

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        9    75.00          4    80.00          90.00
        3 |        3    25.00          2    40.00          60.00
----------+-----------------------------------------------------
    Total |       12   100.00          6   120.00          80.00
                                 (n = 5)
```

```
---------------------------------------------------------------------------
-> failure = 1

                 Overall             Between            Within
     loc2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        2     50.00          1    100.00          50.00
        3 |        2     50.00          1    100.00          50.00
----------+-----------------------------------------------------
    Total |        4    100.00          2    200.00          50.00
                                  (n = 1)

. by failure:xttab type2 if year==1

---------------------------------------------------------------------------
-> failure = 0

                 Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     2428     90.46        533     91.42          99.51
        2 |      256      9.54         53      9.09          93.77
----------+-----------------------------------------------------
    Total |     2684    100.00        586    100.51          98.99
                                (n = 583)

---------------------------------------------------------------------------
-> failure = 1

                 Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       50     92.59         34     94.44         100.00
        2 |        4      7.41          2      5.56         100.00
----------+-----------------------------------------------------
    Total |       54    100.00         36    100.00         100.00
                                 (n = 36)

. by failure:xttab type2 if year==2

---------------------------------------------------------------------------
-> failure = 0

                 Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     2320     95.91        445     93.88          99.66
        2 |       99      4.09         31      6.54          90.00
----------+-----------------------------------------------------
    Total |     2419    100.00        476    100.42          99.03
                                (n = 474)
```

```
-----------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       23    100.00        12    100.00         100.00
----------+-----------------------------------------------------
    Total |       23    100.00        12    100.00         100.00
                                (n = 12)

. by failure:xttab type2 if year==3

-----------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     1139     94.29       268     93.71          99.13
        2 |       69      5.71        22      7.69          83.13
----------+-----------------------------------------------------
    Total |     1208    100.00       290    101.40          97.92
                               (n = 286)

-----------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       10    100.00         9    100.00         100.00
----------+-----------------------------------------------------
    Total |       10    100.00         9    100.00         100.00
                                 (n = 9)

. by failure:xttab type2 if year==4

-----------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      715     92.86       181     92.82         100.00
        2 |       55      7.14        14      7.18         100.00
----------+-----------------------------------------------------
    Total |      770    100.00       195    100.00         100.00
                               (n = 195)
```

```
-----------------------------------------------------------------------------
-> failure = 1

                 Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        4   100.00          3   100.00         100.00
----------+-----------------------------------------------------
    Total |        4   100.00          3   100.00         100.00
                                   (n = 3)

. by failure:xttab type2 if year==5

-----------------------------------------------------------------------------
-> failure = 0

                 Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      408    90.07         94    87.85         100.00
        2 |       45     9.93         13    12.15         100.00
----------+-----------------------------------------------------
    Total |      453   100.00        107   100.00         100.00
                                 (n = 107)

-----------------------------------------------------------------------------
-> failure = 1

                 Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        4   100.00          2   100.00         100.00
----------+-----------------------------------------------------
    Total |        4   100.00          2   100.00         100.00
                                   (n = 2)

. by failure:xttab type2 if year==6

-----------------------------------------------------------------------------
-> failure = 0

                 Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      200    87.72         48    84.21         100.00
        2 |       28    12.28          9    15.79         100.00
----------+-----------------------------------------------------
    Total |      228   100.00         57   100.00         100.00
                                  (n = 57)
```

```
-------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1     33.33         1     50.00         100.00
        2 |        2     66.67         1     50.00         100.00
----------+-----------------------------------------------------
    Total |        3    100.00         2    100.00         100.00
                                 (n = 2)

. by failure:xttab type2 if year==7

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       87     79.82        19     79.17         100.00
        2 |       22     20.18         5     20.83         100.00
----------+-----------------------------------------------------
    Total |      109    100.00        24    100.00         100.00
                                (n = 24)

-------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1    100.00         1    100.00         100.00
----------+-----------------------------------------------------
    Total |        1    100.00         1    100.00         100.00
                                 (n = 1)

. by failure:xttab type2 if year==8

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        8     66.67         3     60.00         100.00
        2 |        4     33.33         2     40.00         100.00
----------+-----------------------------------------------------
    Total |       12    100.00         5    100.00         100.00
                                 (n = 5)
```

```
-----------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    type2 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        2 |        4   100.00          1   100.00         100.00
----------+-----------------------------------------------------
    Total |        4   100.00          1   100.00         100.00
                                   (n = 1)

. by failure:xttab race3 if year==1

-----------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     2217    86.74        476    86.55         100.00
        2 |      339    13.26         74    13.45         100.00
----------+-----------------------------------------------------
    Total |     2556   100.00        550   100.00         100.00
                                 (n = 550)

-----------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       42    77.78         28    77.78         100.00
        2 |       12    22.22          8    22.22         100.00
----------+-----------------------------------------------------
    Total |       54   100.00         36   100.00         100.00
                                  (n = 36)

. by failure:xttab race3 if year==2

-----------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |     2061    88.53        393    86.75         100.00
        2 |      267    11.47         60    13.25         100.00
----------+-----------------------------------------------------
    Total |     2328   100.00        453   100.00         100.00
                                 (n = 453)
```

```
-------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       15    65.22          9    75.00         100.00
        2 |        8    34.78          3    25.00         100.00
----------+-----------------------------------------------------
    Total |       23   100.00         12   100.00         100.00
                               (n = 12)

. by failure:xttab race3 if year==3

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      988    87.82        237    89.10         100.00
        2 |      137    12.18         29    10.90         100.00
----------+-----------------------------------------------------
    Total |     1125   100.00        266   100.00         100.00
                              (n = 266)

-------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        8    88.89          7    87.50         100.00
        2 |        1    11.11          1    12.50         100.00
----------+-----------------------------------------------------
    Total |        9   100.00          8   100.00         100.00
                                (n = 8)

. by failure:xttab race3 if year==4

-------------------------------------------------------------------------------
-> failure = 0

                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      657    89.02        163    88.59         100.00
        2 |       81    10.98         21    11.41         100.00
----------+-----------------------------------------------------
    Total |      738   100.00        184   100.00         100.00
                              (n = 184)
```

```
------------------------------------------------------------------------------
-> failure = 1

                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        4   100.00          3   100.00         100.00
----------+-----------------------------------------------------
    Total |        4   100.00          3   100.00         100.00
                                 (n = 3)

. by failure:xttab race3 if year==5

------------------------------------------------------------------------------
-> failure = 0
                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      394    88.94         91    88.35         100.00
        2 |       49    11.06         12    11.65         100.00
----------+-----------------------------------------------------
    Total |      443   100.00        103   100.00         100.00
                               (n = 103)

------------------------------------------------------------------------------
-> failure = 1
                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        4   100.00          2   100.00         100.00
----------+-----------------------------------------------------
    Total |        4   100.00          2   100.00         100.00
                                 (n = 2)

. by failure:xttab race3 if year==6

------------------------------------------------------------------------------
-> failure = 0
                  Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |      192    84.96         48    85.71         100.00
        2 |       34    15.04          8    14.29         100.00
----------+-----------------------------------------------------
    Total |      226   100.00         56   100.00         100.00
                                (n = 56)
```

```
--------------------------------------------------------------------------------
-> failure = 1
                Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        3    100.00         2    100.00         100.00
----------+-----------------------------------------------------
    Total |        3    100.00         2    100.00         100.00
                                   (n = 2)

. by failure:xttab race3 if year==7

--------------------------------------------------------------------------------
-> failure = 0

                Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |       70     67.96        16     72.73         100.00
        2 |       33     32.04         6     27.27         100.00
----------+-----------------------------------------------------
    Total |      103    100.00        22    100.00         100.00
                                  (n = 22)

--------------------------------------------------------------------------------
-> failure = 1

                Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        1    100.00         1    100.00         100.00
----------+-----------------------------------------------------
    Total |        1    100.00         1    100.00         100.00
                                   (n = 1)

. by failure:xttab race3 if year==8

--------------------------------------------------------------------------------
-> failure = 0

                Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        9     75.00         4     80.00         100.00
        2 |        3     25.00         1     20.00         100.00
----------+-----------------------------------------------------
    Total |       12    100.00         5    100.00         100.00
                                   (n = 5)
```

```
-------------------------------------------------------------------------------
-> failure = 1

                    Overall             Between            Within
    race3 |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
        1 |        4    100.00          1    100.00         100.00
----------+-----------------------------------------------------
    Total |        4    100.00          1    100.00         100.00
                                 (n = 1)
```

BIBLIOGRAPHY

Albrektsson, T., Zarb, G., Worthington, P., Eriksson, A.R.: The long-term efficacy of currently used implants: a review and proposed criteria of success. *Int J Oral Maxillofac Implants* (1986) 1:11-25.

Becktor JP, Eckert SE, Isaksson S, Keller EE.: The influence of mandibular dentition on implant failures I bone-grafted edentulous maxillae. Int J Oral Maxillofac Implants (2002) Jan-Feb;17(1):69-77.

Buser D, Merieske-Stern R, Bernard JP, Behneke A, Behneke N, Hirt HP, et. al.: Long-term evaluation of non-submerged ITI implants. Part 1: 8-year life table analysis of a prospective multi-center study with 2359 implants. Clin Oral Implants Res (1997) 8:161-172.

Brocard, D., Barthet, P., Baysse, E., Duffort, J.F., Eller, P., Justumus, P., et. al.: A multicenter report on 1,022 consecutively placed ITI implants; a 7-year longitudinal study. *Int J Oral Maxillofac Implants* (2000) 15:691-700.

Carlin, J.B., et. al,: Tutorial in Biostatistics-Analysis of Binary Outcomes in Longitudinal Studies Using Weighted Estimating Equations and Discrete-Time Survival Methods: Prevalence and Incidence of Smoking in an Adolescent Cohort. Statistics In Medicine (1999) 1:2655-2679.

Chuang, S.K., Tian, L., Wei, L.J., and Dobson, T.B.: Kaplan-Meier Analysis of Dental Implant Survival: A Strategy for Estimating Survival with Clustered Observations. J Dent Res (2001) 80(11):2016-2020,

Eckert SE, Wollan PC: Retrospective review of 1170 endosseous implants placed in partially edentulous jaws. J Prosthet Dent (1998) 79:415-421.

Encyclopedia of Biostatistics: Logistic Regression (p2316-2327).

Efron B.: The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association (1977) 72: 557-565.

Goto M, Jin-Nouchi S, Ihara K, Katsuki T: Longitudinal follow-up of osseointegrated implants in patients with resected jaws. Int J Oral Maxillofac Implants (2002) Mar-Apr;17(2):225-30.

Haas, R., Mensdorff-Pouilly N., Mailath, G., Watzek, G.: Survival of 1,920 IMZ implants followed for up to 100 months. *Int J Oral Maxillofac Implants* (1996) 11:581-588.

Haas R, Polak C, Furhauser R, Mailath-Pokorny G, Dortbudak O, Watzek G: A long-term follow-up of 76 Branemark single-tooth implants. Clin Oral Implants Res 2002 Feb;13(1):38 43.

Herrmann I, Lekholm U, Holm S, Karlsson S.: Impact of implant interdependency when evaluating success rates: a statistical analysis of multicenter results. Int J Prosthodont (1999) Mar-Apr;12(2):160-6.

Hising, P., Bolin, A., Branting, C.: Reconstruction of severely resorbed alveolar ridge crests with dental implants using a bovine bone mineral for augmentation. *Int J Oral Maxillofac Implants* (2001) 16:90-97.

Hosmer, D. and Lemeshow, S.: *Applied Logistic Regression* (1989), Wiley, New York.

Huber, P.J. "The behavior of maximum likelihood estimation under non-standard conditions", in Fifth Berkeley Symposium in Mathematical Statistics and Probability, University of California Press, Berkeley, CA. (1967), pp-221-233.

Ivanoff, C.J., Grondahl, K., Sennerby, L.,Lekholm, U.: Influence of variations in implant diameters: a 3- to 5-year retrospective clinical report. Int J Oral Maxillofac Implants (1999) 11:291-298.

Jemt T, Chai J, Harnett J, Heath MR, Hutton JE, Johns RB, et. al.: A 5-year prospective multicenter follow-up report on overdentures supported by osseointegrated implants. Int J Oral Maxillofac Implants (1996) 11:291-298.

Kalbfleisch, J.D., and Prentice, R.L., the *Statistical Analysis of Failure Time Data* (1980), New York:Wiley.

Lambert PM, Morris HF, Ochi S.: The influence of smoking on 3-year clinical success of osseointegrated dental implants. Ann Periodontol (2000) Dec;5(1):79-89.

Lazzara, R., Siddiqui, A.A., Binon, P., Feldman, S.A., Weiner, R., Phillips, R., et. al.: Retrospective multicenter analysis of 3i endosseous implants placed over a five-year period. *Clin Oral Implants Res* (1996) 7:73-83.

Lee, E.W., Wei, L.J., and Amato, D.A.: "Cox-Type Regression Analysis for Large Numbers of Small Groups of Correlated Failure Time Observations" in *Survival Analysis:State of the Art*, eds. (1992) J.P. Klein and P.K. Goel, Dordrecht:Kluwer Acedemic, pp. 237-247

Lekholm, U., Gnu, J., Henry, Higuchi K., Linden, U., Bergstom,C., et. al.: Survival of the Branemark implant in partially edentulous jaws: a 10-year prospective multicenter study. *Int J Oral Maxillofac Implants* (1999)14:639-645.

Manz M.C.: Factors associated with radiographic vertical bone loss around implants placed in a clinical study.:Ann Periodontal (2000) Dec;5(1):137-51

Mau, J.: On statistics of success and loss for implants. *Int Dent J* (1993) 43:254-261.

McCullough, P. and Nelder, J.A.: *Generalized Linear Models.* (1983) Chapman & Hall, London.

Morris HF, Ochi S.: Influence of two different approaches to reporting implant survival outcomes for five different prosthodontic applications. Ann Periodontol (2000) Dec;5(1):90 100.

Morris HF, Ochi S.: Influence of research center on overall survival outcomes at each phase of treatment. Ann Periodontol (2000) Dec;5(1):129-36

Morris HF, Ochi S, Wrinkler S.: Implant survival in patients with type 2 diabetes: placement to 36 months. Ann Periodontol (2000) Dec;5(1):157-65.

Ochi S.: The Dental Implant Clinical Research Group study: study design and statistical methods utilized. Ann Periodontol (2000) Dec;5(1):12-4.

Orenstein IH, Petrazzuolo B, Morris HF, Ochi S.: Variables affecting survival of single-tooth hydroxyapatite-coated implants in anterior maxillae at 3 years. Ann Periodontol (2000) Dec;5(1):68-78.

Orenstein IH, Tarnow DP, Morris HF, Ochi S: Three-year post-placement survival of implants mobile at placement. Ann Periodontol (2000) Dec;5(1):32-41.

Peto, R. and Peto, J. Asymptotically efficient rank invariant procedures: *Journal of the Royal Statistical Society, A*, (1972) 135, 185-207.

Rosenquist, B., Grenthe, B.: Immediate placement of implants into extraction sockets: implant survival. *Int J Oral Maxillofac Implants* (1996) 11:205-209.

Spiekerman, C.F. and Lin, D.Y.: Marginal Models for Multivariate Failure Time Data. *Journal of the American Statistical Association*, September (1998), Vol. 93, No. 443, Theory and Methods

Spray JR, Black CG, Morris HF, Ochi S.: The influence of bone thickness on facial marginal bone response:stage 1 placement through stage 2 uncovering. Ann Periodontol (2000) Dec;5(1):119-28

Survival Analysis and Epidemiological Tables (STATA Manual release 8-Page 218, 2004 Revised Edition

Tong DC, Rioux K, Drangsholt M, Beirne OR.: A review of survival rates for implants placed in grafted maxillary sinuses using meta-analysis. Int J Oral Maxillofacial Implants (1998) Mar Apr,13(2):175-82.

Wei, L.J., Lin, D.Y., and Weisfeld, L.: "Regression Analysis of Multivariate Failure Time Data by Modelling Marginal Distributions." *Journal of the American Statistical Association*. (1989) 84, 1065-1073

Wheeler, S.L.: Eight-year clinical retrospective study of titanium plasma-sprayed and hydroxyapatite-coated cylinder implants. *Int J Oral Maxillofac Implants* (1996) 11:340-350.

White, H.: "Maximum likelihood estimation of misspecified models". Econometrica, (1980) 50, 1-25.

Widmark, G., Andersson, B., Carlsson, G.E., Lindvall, A.M., Ivanoff, C.J.: Rehabilitation of patients with severely resorbed maxillae by means of implants with or without bone grafts: a 3- to 5-year follow-up clinical report. *Int J Oral Maxillofac Implants* (2001) 16:73-79.

Winkler S, Morris HF, Ochi S.: Implant survival to 36 months as related to length and diameter. Ann Periodontol (2000) Dec;5(1):22-31.

Wu, M. and Ware, J.H.: "On the use of repeated measurements in regression analysis with dichotomous responses" *Biometrics* (1979) 35:513-521.