5-2018

Tibial Fracture Nonunions Following Intramedullary Nailing: An Application of Logistic Regression

Katherine Wu
Lehigh University

ECKARDT SENIOR THESIS

TIBIAL FRACTURE NONUNIONS FOLLOWING INTRAMEDULLARY NAILING:

AN APPLICATION OF LOGISTIC REGRESSION

Katie Wu

ABSTRACT

This paper utilizes the principles of logistic regression analysis to study patient factors that are strongly linked to the outcome of intramedullary nail surgery. This surgery is performed on fractured tibial bones to bring the bone back to union, or a state of connection again. The alternative result of the surgery is nonunion, or a state in which the bone does not heal properly. This paper draws from a recent publication in the Journal of Orthopedic Trauma, "Tibial fracture nonunion and time to healing following reamed intramedullary nailing" (Dailey, Wu, Wu, McQueen, & Court-Brown, 2018).

INTRODUCTION

This paper utilizes the principles of logistic regression analysis to determine and analyze patient factors that are most influential in determining the outcome of reamed nailing in tibial injuries. From a statistical standpoint, the processes of model building, model selection, and residual analysis in logistic regression are explored. These techniques are implemented on 1,006 patient records from 1985 to 2007 collected from the State Hospital in Carstairs, Scotland (a Scottish Level 1 trauma centre). A total of 1,590 adult tibial fractures were obtained, but after specifying inclusion criteria to ensure that each observation had enough information, 1,006 cases remained. The primary outcome studied in this analysis is the result of the intramedullary nail procedure on a tibial fracture. This is a binary result: union (coded 0, the bone properly heals) or non-union (coded 1, the bone does not properly heal).

The next section of the paper will describe the basics of a tibial fracture to create a foundation of basic domain knowledge prior to implementing statistical procedures. The following section will describe the fundamentals of logistic regression. Then, a step-by-step process will be outlined to demonstrate the process of creating, choosing, and analyzing a logistic regression model to represent the tibial data. A discussion of the results and significance of the findings concludes the paper.

Figure 1: Open and closed fractures

TIBIAL INJURIES

A tibial fracture is synonymous to a shinbone fracture and occurs between the ankle and knee. Because this bone is so long, a major force is required to break this bone. Figure 1 illustrates what a fracture of this bone looks like. One can either suffer a closed fracture (shown as the "stable" fracture) or an open fracture. An open fracture penetrates the skin and typically creates much more damage to muscles, ligaments and tendons around the fracture site (OrthoInfo).

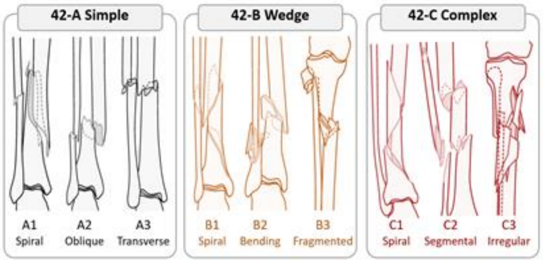


Figure 2: AO Classification scheme

Aside from the classification of open or closed, there are various other medical cataloguing schemes that are followed. There is the Müller AO Classification of long bone fractures, and for the open fractures, there is the Gustilo Classification of open trauma wounds. The AO Classification scheme is illustrated in Figure 2. Both of these schemes classify the injuries by a letter type and a number type.

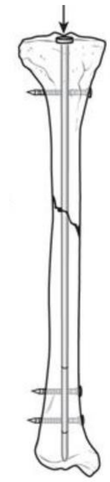


Figure 3: Intramedullary nailing

On a more general level, injuries can be given an Injury Severity Score (ISS), an anatomical scoring system that attempts to numerically identify the level of severity for an injury. This variable takes values from 0 to 75, with 75 indicating a non-survivable injury.

Some patients with a tibial injury must have a fasciotomy, a procedure that involves an incision of the fascia and is often implemented when an injury causes some sort of loss of circulation. All of the patients that will be studied in this analysis have undergone the intramedullary nailing procedure in attempts to heal the bone. This surgical procedure, shown in Figure 3, inserts a medal rod from the knee to the marrow canal of the bone to keep the fracture in place (OrthInfo). Because this procedure is not ideal of children, the patients included in this analysis are all over the age of eighteen.

LOGISTIC REGRESSION

Logistic regression allows the relationship between explanatory variables and a discrete, often binary, response variable to be modeled effectively. The goal of utilizing this form of regression is to create the most parsimonious and interpretable model to define the interaction between the covariates and the outcome. This type of regression differs from typical linear regression for one main reason: the outcome variable must remain within the range [0,1] in logistic regression, while the outcome variable for linear regression has the range (-∞, ∞). Because of this transformation, the assumption of normally distributed error terms from linear regression is no longer held. A specific link function must be chosen (logit, probit, complementary log-log) depending on the structure of the error terms. In this case, we choose to follow the most widely used logit transformation corresponding to errors following the standard logistic distribution. Considering a univariate case, if we utilize

$$\pi(x) = \frac{e^{\beta_0+\beta_1 x}}{1+\ e^{\beta_0+\beta_1 x}}$$

to represent *E(Y|x)*, the expected value of the outcome given the value of the independent variable, we see that the logit transformation can be defined as

$$g(x) = ln\left[\frac{\pi(x)}{1-\pi(x)}\right].$$

Setting this equal to the regular linear combination of the intercept term and the covariate, we obtain:

$$ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 +\ \beta_1 x.$$

We now see that the logit is linear and can range from (-∞, ∞). The basic model that will be utilized in this analysis is simply a multivariate extension of this model, allowing for *p* different independent variables, as shown in the equation below.

$$ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \beta_0 + \beta_1 x + \beta_2 x + \beta_3 x + \cdots + \beta_p x$$

APPLICATION TO THE DATA

This analysis utilizes the 1,006 filtered patient records from 1985 – 2007 from the State Hospital in Carstairs, Scotland. Using Statistical Analysis Software (SAS), a logistic regression model is created, fit, and analyzed to determine influential factors in determining the union/nonunion outcome of intramedullary nailing of a tibial fracture.

MODEL BUILDING & SELECTION

Univariable Analysis

Prior to fitting any model, we first must examine the underlying behavior of the individual independent covariates. Initially, we start with the following variables:

- **CARSTAIRS** (identification)
- **AGE** (recorded age)
- **SEX** (binary: male or female)
- **AOLETTER** (AO letter category)
- **AOCODE1** (AO # category)
- **GUSTILONUM** (Gustilo # category)
- **GUSTILLET** (Gustilo letter category)
- **IMPLANT** (binary: yes/no)
- **FASCIOTOMY** (binary: yes/no)
- **VALID** (binary: yes/no)
- **ID** (identification)
- **INJAGE** (continuous calculated age)
- **CLOSED** (binary: yes/no)
- **SMOKES** (binary: yes/no)
- **INJCODE** (categorical injury code)
- **NONUNION** (binary: yes/no)
- **ISSTOTAL** (continuous ISS score)
- **IMN** (binary: yes/no)

After initial univariable analysis, we notice that CARSTAIRS and ID are a nonrelevant values specific to identifying the patient record at the hospital, IMPLANT and IMN are unnecessary since every patient has an implant (IMPLANT) and has undergone the intramedullary nail surgery (IMN), VALID is unimportant to the regression as every single patient record has been marked as a valid case after filtering, and AGE is a potentially incorrect recording, so the calculated INJAGE (determined from each patient's date of birth and date of record) should be used instead. We continue to study the remaining variables for significance.

Of the remaining variables, only INJAGE and ISSTOTAL are continuous variables. The remaining variables are categorical or binary and will be treated as such.

For the continuous variables, univariable analysis consists of studying the distribution of the variable and running a univariable logistic regression on NONUNION. When this is done, the likelihood ratio chi-square test statistics and Wald test statistics for testing the global null $\beta=0$ are examined for significance. For example, when the ISSTOTAL variables is run as the sole continuous covariate in determining the NONUNION outcome, we obtain the following equation and statistics:

$$ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = -2.5428 + 0.0651x$$

Likelihood Ratio Chi-Square: 8.0406 (p-value: 0.0046)

Wald Statistic: 9.1670 (p-value: 0.0025)

Both of these tests study whether or not this model is significantly better at predicting the NONUNION outcome than a model where the β values are equal to zero (an intercept only model). We see that these statistics follow a chi-square distribution with degrees of freedom that create significant p-values at $\alpha = 0.05$. Thus, we conclude that this covariate is univariably significant.

For the categorical variables, univariable analysis consists of studying the contingency table of outcomes for each category and studying the likelihood ratio chi-square test statistic (as before with the continuous variables). For example, GUSTILLET is a categorical variable determining which letter category of Gustilo classification the open fraction belongs to. If we look at the frequencies of unions/nonunions within each category, we see the following table:

Table of GUSTILLET by NONUNION

Frequency Percent Row Pct Col Pct	NONUNION		
GUSTILLET(GUSTILLET)	0	1	Total
0	33 23.57 76.74 40.24	10 7.14 23.26 17.24	43 30.71
1	33 23.57 75.00 40.24	11 7.86 25.00 18.97	44 31.43
2	15 10.71 28.85 18.29	37 26.43 71.15 63.79	52 37.14
3	1 0.71 100.00 1.22	0 0.00 0.00 0.00	1 0.71
Total	82 58.57	58 41.43	140 100.00

Frequency Missing = 866

We can see that in the GUSTILLET category 3, there are 0 cases of a nonunion outcome and only 1 case of a union outcome. Due to the very little information in this category, quasi-complete separation would occur as the SAS program attempts to iteratively converge the Newton-Raphson algorithm to estimate the regression coefficients. Because of this, we must collapse the 3 category with the 2 category. Once this is done, there is now more sufficient information within this variable. Additionally, the likelihood ratio chi-square value is 31.3391 with a p-value of <0.0001. Therefore, we conclude significance at a level of $\alpha = 0.05$.

For additional details regarding the univariable process, see the included appendix file.

Model Building

We now look at the various methods we can take to execute the model building step. Specifically, we will examine 2 methods: stepwise selection and iteratively building the model with domain knowledge.

Stepwise selection is a process which adds the variable with the most significant likelihood ratio test when compared with the null model. It then adds the next most significant variable to the model, and then checks that none of the variables in the model have become insignificant. If they have, they are removed, and additional variables are entered based on significance. This stepwise procedure is a combination of forward and backward selection. The threshold requirements for the significance of the test statistics are different for entry and staying. Because we want to ensure that our model has as many variables as necessary for a meaningful result, very liberal constraints were created with an enter requirement of only 0.40 and a stay requirement of 0.45. However, when these are enforced, we see the results in the table below.

Summary of Stepwise Selection								
	Effect							
Step	Entered	Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq	Variable Label
1	GUSTILLET		2	1	20.9912		<.0001	GUSTILLET
2	INJCODE		7	2	16.6792		0.0196	
3	AOCODE1		2	3	4.9406		0.0846	AOCODE1
4		INJCODE	7	2		5.9471	0.5459	
5	INJCODE		7	3	13.6160		0.0584	
6		INJCODE	7	2		5.9471	0.5459	

This procedure ignores the clinically important variables and creates a somewhat non-interpretable and non-implementable model. The only variables left in the model at the end are GUSTILLET and AOCODE1.

The other method utilized in this analysis is based on building a model with domain knowledge and the univariable study. Based on the results from the univariable analysis, all variables with test statistics significant at a level of α = 0.25 are added to the initial model. Additionally, variables that are known to be clinically significant according to domain experts are added. Initially, this includes:

- SEX*
- AOLETTER
- AOCODE1
- GUSTILONUM
- GUSTILLET
- INJAGE*
- CLOSED SMOKES*
- INJCODE
- ISSTOTAL

*Included initially due to potential clinical significance even though they did not match the 0.25 significance level

The model is fit with all of these variables. Their individual significance in terms of contribution to the overall model is examined through their Wald statistics, a statistic that follows the chi-square distribution. Those variables that do not carry significance in the overall model are removed. Likelihood ratio tests are conducted between the model without the variables and the model with variables to justify removal. This process of removing variables and checking significances is also complicated by checking for interaction effects. A particular covariate may be non-significant on its own, but when combined with another, becomes a very important predictor.

Thus, as variables are added and removed, interaction effects must also be tested and kept in mind.

Additionally, continuous variables must be examined to see if a power transformation on the variable could perhaps create a better fit. In this case, INJAGE is examined individually by creating a univariate logistic regression of INJAGE on NONUNION and another one of INJAGE^2 on NONUNION. The results indicate that INJAGE^2 creates a more significant model than INJAGE. Further, due to domain knowledge, it was thought that perhaps splitting INJAGE into different categories would create meaningful and interpretable results. A model was created in which INJAGE is divided into 3 groups: (1) 18 – 30, (2) 20 – 60, (3) 60+. This new categorical variable is represented as AG.

Model Selection

Once these procedures are completed, the following competing models can be compared.

Model Covariates	Pearson Chi-Sq	Deviance	Likelihood Ratio, B=0	Wald Statistic	G (to full)	DF
All univariably significant covariates	99.9644 (0.2922)	98.5648 (0.3268)	56.2279 (<0.0001)	27.2999 (0.0979)	39.364 (*5.991)	19
AGE^2, AOCODE1, GUSTILLET, CLOSED	82.7751 (0.5481)	93.0508 (0.2579)	31.1939 (<0.0001)	21.7773 (<0.0001)	36.622 (*24.996)	6
AOCODE1, GUSTILLET	4.5865 (0.3324)	4.8395 (0.3042)	30.8334 (<0.0001)	21.5756 (0.0002)	36.983 (*27.587)	4
AGE^2, AOLETTER, CLOSED	***1032.9588 (0.0239)***	***553.6228 (1.0000)***	***139.4414 (<0.0001)***	***118.6115 (<0.0001)***	***499.967 (*27.587)***	***4***
AG, AOLETTER, CLOSED	18.4873 (0.1017)	20.5393 (0.0575)	144.1576 (<0.0001)	121.3017 (<0.0001)	495.251 (*26.296)	5

For details on these conclusions, see the appendix file

The process of choosing an optimal model involves some subjectivity. The model with all univariably significant covariates is not optimal due to concerns of unnecessary variables and overfitting. The model including AGE^2, AOCODE1, GUSTILLET, CLOSED is not optimal because the GUSTILLET variable only applies to open fractures. As a result, the number of cases that this model works on is significantly reduced and less useful. The same argument can be applied as a reasoning to why the AOCODE1, GUSTILLET model is not chosen. Lastly, the AG, AOLETTER, CLOSED model is not chosen because it is more mathematically robust to treat a continuous variable as such rather than splitting it into buckets and losing a portion of additional information. Thus, we choose the AGE^2, AOLETTER, CLOSED model.

ASSESSING THE FIT OF THE MODEL

Goodness-of-Fit Statistics

We will examine 3 different goodness-of-fit statistics: Deviance, Pearson Chi-Square, and Hosmer-Lemeshow. The deviance statistic is based upon individual deviance residuals:

$$d(y_j, \pi_j) = -\sqrt{2m_j|\ln(1-\hat{\pi}_j)|}$$

Here, y_j is the response of covariate pattern *j*. A covariate pattern is a unique combination of covariates. There are *J* total covariate patterns and m_j total cases in covariate pattern *j*. The deviance statistic is the sum of the squares of these residuals.

$$D = \sum_{j=1}^{J} d(y_{j,} \hat{\pi}_j)^2$$

Similarly, the Pearson chi-square statistic is based on individual Pearson residuals:

$$r(y_j, \hat{\pi}_j) = \frac{(y_j - m_j\hat{\pi}_j)}{\sqrt{m_j\hat{\pi}_j(1 - \hat{\pi}_j)}}$$

And the Pearson chi-square statistic is just the squared sum of these residuals:

$$X^2 = \sum_{j=1}^{J} r(y_{j,} \hat{\pi}_j)^2$$

Both the deviance and Pearson chi-square statistic follow the chi-square distribution with degrees of freedom equal to *(J – (p+1))*, the number of total covariate patterns minus the number of covariates plus one.

The Hosmer-Lemeshow statistic breaks the instances down into groups, often times deciles. The number of groups is called *g*. From there, this goodness-of-fit statistic is calculated and approximately follows a chi-square distribution with degrees of freedom equal to *(g-2)*, the number of total groups minus two.

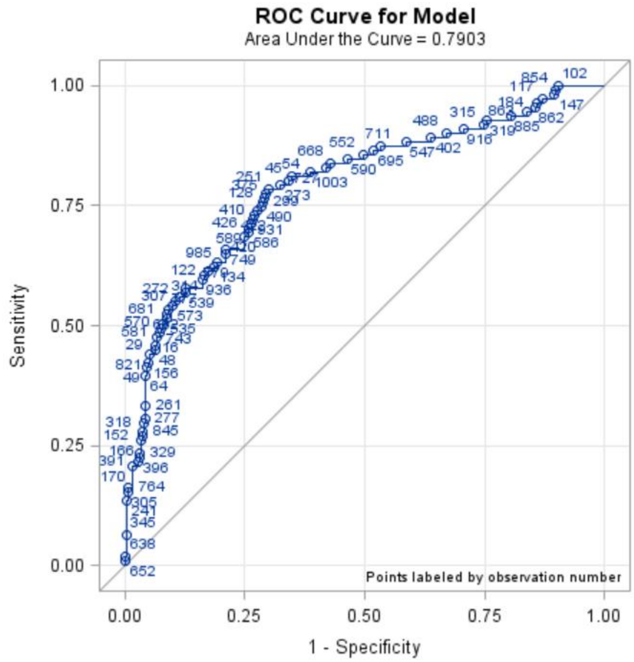
The results from utilizing this model create a deviance statistic of 553.6228 (p-value of 1.0000), a Pearson chi-square statistic of 1032.9588 (p-value of 0.0239), and a Hosmer-Lemeshow statistic of 6.0057 (p-value of 0.6466). Both the deviance and Hosmer-Lemeshow statistics indicate that this model fits well. Again, this model selection stage is relatively subjective, and we conclude that these are adequate to determine this model is a reasonable fit.

Testing the Global Null

Additionally, the statistics in this model that test the null hypothesis that β = 0 are both highly significant. These statistics are not checking the individual covariates (as they were before), but they are checking the overall model. The likelihood ratio test statistic is 139.4414 with a p-value of <0.0001. The Wald statistic is 118.6115 with a p-value of <0.0001. Therefore, we can conclude that both these tests show significance of this model outperforming an intercept-only model.

Receiver Operating Curve

To see how this model can classify the data, we can examine the receiver operating curve below.



The y-axis models the true positive rate, the ratio of correctly classified positives to total positives. The x-axis models the false positive rate, the ratio of incorrectly classified positives to total positives. Utilizing different cut-points in predicted probabilities for determining whether the outcome is union or nonunion, each point represents a different threshold. The 45 degree line represents the classification of a random classifier. Clearly, the model outperforms this random classifier. The area under the curve (AUC) is equal of 0.7903. According to traditional academic ranking, an AUC of 0.7 – 0.8 is considered “fair”, while an AUC from 0.8 – 0.9 is considered “good.” Our model is bordering the threshold between “fair” and “good,” which is acceptable, especially considering the messy reality of data.

Influential Points

We now must examine the outlying data points in this analysis and determine if any of them are influential enough to alter the fit of the model. To do this, we will look at 3 main statistics:

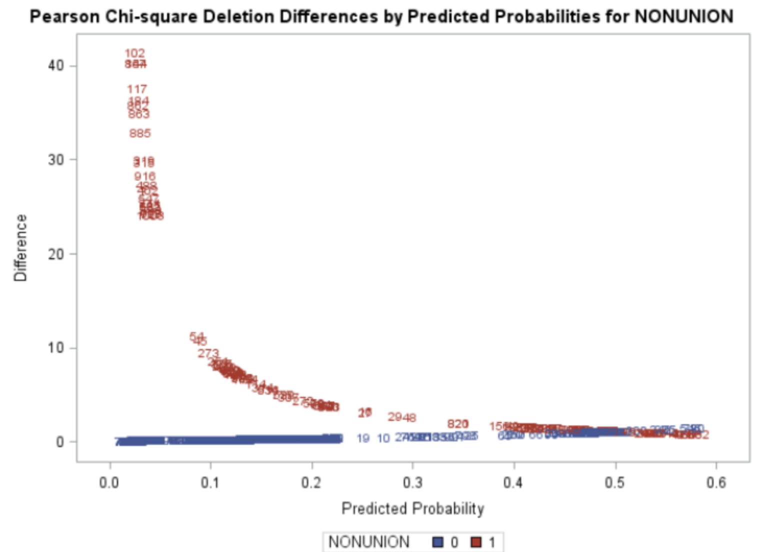
- Pearson chi-square deletion differences
- Deviance deletion differences
- Leverage

Pearson chi-square deletion differences show the change in the value of the Pearson chi-square goodness-of-fit statistic due to the deletion of observations with specific covariate patterns.

Deviance deletion differences similarly show the differences in the value of the deviance goodness-of-fit statistic due to the deletion of observations with specific covariate patterns.

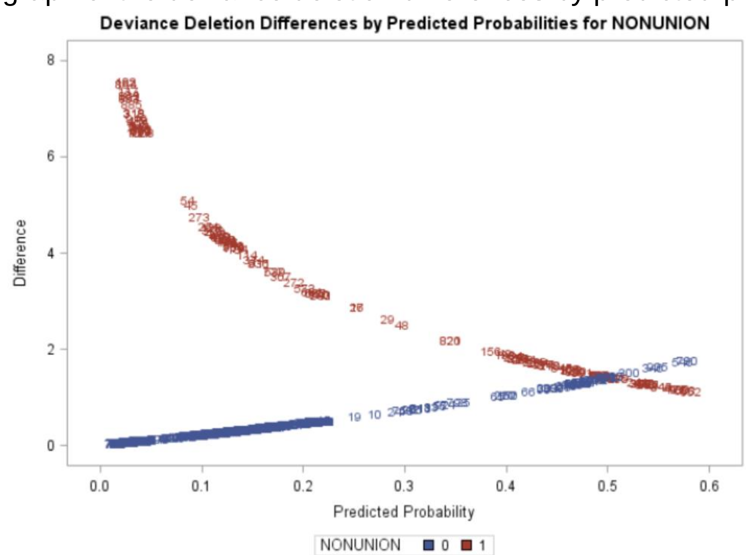
In linear regression, the leverage values are the diagonal terms in the hat matrix that represent the differences in distance from the predicted values and the mean. In the case of logistic regression, the terminology of distance does not hold in the same way, but the points still indicate predicted values further from the actual values.

In all three cases, visual analyses of these graphical representations help to identify potentially influential points. Below, we see the graph of the Pearson chi-square deletion differences by predicted probabilities.



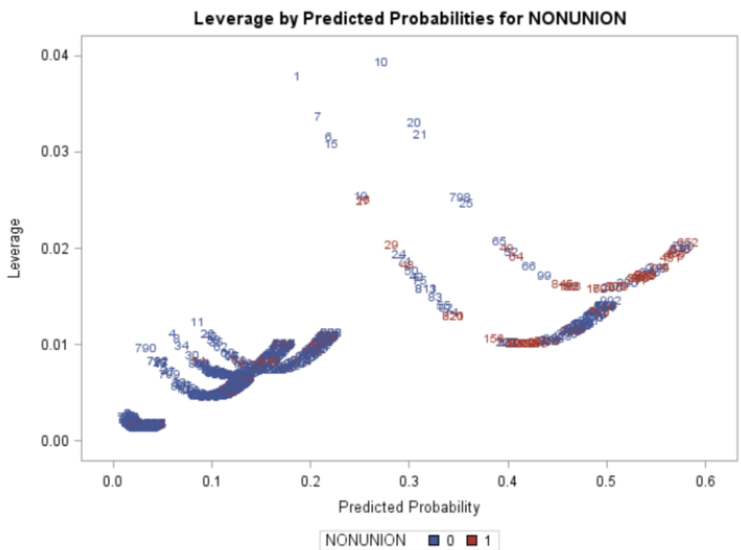
To be safe, anything with a difference greater than 15 is considered an outlier. Those observation numbers of these points are recorded.

Below, we see the graph of the deviance deletion differences by predicted probabilities.



Here, values above 5 are considered to be outliers. The observation numbers of these points are recorded.

Below, we see the graph of the leverage by predicted probabilities.



Values above 0.025 are considered to be outliers here, and their observation numbers are recorded.

To see if these points significantly affected the fit of the model, the logistic regression model with the same covariates is fit again, this time to the data excluding the outlying observations. We see that the data fits extremely well:

Deviance: 374.4317 (p-value: 1.0000)
Pearson chi-square: 498.5497 (p-value: 1.0000)
Hosmer-Lemeshow: 13.0527 (p-value: 0.1100)
Likelihood ratio: 213.7112 (p-value: <0.0001)
Wald: 135.1123 (p-value: <0.0001)

Additionally, we see that the estimates for the coefficients remain very similar in both models. The signs of the coefficients also do not change, which indicates that these outliers did not influence the final model decision. The model with the original data is on the left, and the model without outliers is on the right.

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.1940	0.1744	46.8635	<.0001
AGE2		1	-0.00019	0.000076	6.5997	0.0102
AOLETTER	1	1	-0.9612	0.1572	37.3702	<.0001
AOLETTER	2	1	0.3233	0.1552	4.3394	0.0372
CLOSED	0	1	0.9505	0.1150	68.2953	<.0001

Analysis of Maximum Likelihood Estimates						
Parameter		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept		1	-1.5598	0.2191	50.6875	<.0001
AGE2		1	-0.00020	0.000101	3.8275	0.0504
CLOSED	0	1	1.2976	0.1411	84.5152	<.0001
AOLETTER	1	1	-1.6698	0.2254	54.8859	<.0001
AOLETTER	2	1	0.6273	0.1831	11.7340	0.0006

Therefore, we conclude that since the model still fits extremely well, these outliers were not influential enough to sway the determination of the optimal model. We conclude that this model including AGE^2, AOCODE1, and CLOSED is the optimal model in determining the outcome of union or nonunion in tibial fractures.

CONCLUSION

We conclude that age, the AO number classification, and whether the fracture is open or closed are all the most important factors that a patient can have in order to predict the end result of the intramedullary nailing surgery. This parallels nicely with the clinical assumptions that the nature of the fracture (open or closed) and age should play a large factor in the ultimate outcome of the surgery.

Logically, this also makes sense. It is much more likely that a younger person will be able to fully recover from this injury and less likely that an elderly person will be able to recover from such a traumatic injury. It is also much more likely that a fracture that does not penetrate the skin and involve extra complications will heal much faster than one that does. Additionally, we see that the numerical categories in the Müller AO Classification scheme that indicate the severity of the injury play an important role. This is also very reasonable as a more severe injury is much more likely to result in nonunion.

This model can potentially be used to assess a patient's probability of nonunion if given information regarding the patient's age, injury severity classification (AO number), and injury type (open/closed). Not only does this final model have the statistical and mathematical support behind it, but it also has clinical, domain-level evidence to be a rational model.

REFERNCES

Dailey, H. L., Wu, K. A., Wu, P., Mcqueen, M. M., & Court-Brown, C. M. (2018). Tibial fracture nonunion and time to healing following reamed intramedullary nailing. Journal of Orthopedic Trauma, 1. doi:10.1097/bot.0000000000001173

Hosmer, D. W., & Lemeshow, S. (2000). Applied logistic regression. Oxford: Wiley-Blackwell.

OrthoInfo. (n.d.). Tibia (Shinbone) Shaft Fractures. Retrieved May 1, 2018, from https://orthoinfo.aaos.org/en/diseases--conditions/tibia-shinbone-shaft-fractures