MSDS 6372 Project 3

Statistical Prediction of Osteoporosis in Women

Team members:

Albert Alva and Harinder (Harry) Bhasin

Statistical Prediction of Osteoporosis in Women

Introduction

Over 200 million people worldwide suffer from osteoporosis. Approximately 30% of postmenopausal women in the U.S. and Europe have osteoporosis, of which 40% will sustain one or more fragility fractures in their remaining lifetime. With aging of the populations, annual osteoporotic fractures worldwide are estimated to triple by the year 2050. Numerous studies are being done. One research study is done by GLOW (Global Longitudinal Study of Osteoporosis in Women). The goal of this research was to improve understanding of the risk and prevention of osteoporosis-related fractures among female residents of 10 countries who were 55 years of age and older. GLOW enrolled over 60,000 women through over 700 physicians in 10 countries, and conducted annual follow-up for up to 5 years through annual patient questionnaires.

This study will further previous research by analyzing the risk factors collected in the GLOW data with the goal of producing a parsimonious model to predict osteoporosis-related fractures in women 55 years of age and older. SAS 9.4 will be used to describe the data and employ logistic regression methods to confirm the factors that explain risk of future fracture. As a result, we may find that a subset of the risk factors identified in previous studies may adequately predict future fractures.

Descriptive Statistics

Original Dataset Variables

The data included information on 500 women over 55 enrolled in an osteoporosis investigation (table 1).

Table 1

Variable	Name	Value	Description
1	SUB_ID	1 to n	Identification code
2	SITE_ID	1 to 6	Study site
3	PHY_ID	128 unique codes	Physician ID code
4	PRIORFRAC	1=Yes, 0=No	History of prior fracture
5	AGE	Years	Age at enrollment
6	WEIGHT	Kilograms	Weight at enrollment
7	HEIGHT	Centimeters	Height at enrollment
8	BMI	Kg/m2	Body mass index
9	PREMENO	1=Yes, 0=No	Menopause before age 45
10	MOMFRAC	1=Yes, 0=No	Mother had hip fracture
11	ARMASSIST	1=Yes, 0=No	Arms are needed to stand from a chair
12	SMOKE	1=Yes, 0=No	Former or current smoker
13	RATERISK	1=< others, 2=same, 3=> others	Self-reported risk of fracture
14	FRACSCORE	Risk score mentioned below	Fracture risk score
15	FRACTURE	1=Yes, 0=No	Any fracture in first year

The response variable is FRACTURE, the person will have or will not have a fracture in year. The response variable has a binary value of 1=Yes and 0-No.

The Predictor variables include 2 risk factor categories; family / subject health history and current health risks.

Historic Risks: The variables include prior fractures (PRIORFRAC), menopause before age 45(PREMENO) and family history of natural mother suffering fractures (MOMFRAC).

Current Health Risks: The variables are known to be associated with determining current health status. They include age, weight, height, BMI, need of arms to assist standing(ARMASSIST) and smoking status. The RATERISK variable is a self-reported perception of the patient's own risk of fracture; 1= less than peers, 2= equal to peers and 3= greater than peers. Sub_ID, Site_ID

and PHY_ID were not used in the analysis. The FRACSCORE scale was also eliminated from the study due to insufficient information on its validity and reliability.

Correlations

Preliminary analysis showed no strong evidence of lack of independence (table 2). Tests of correlations yielded expected weak correlations among measures within risk categories and weak or no relations in variables between risk categories. All other correlations were generally weak or nonexistent.

Fracture was weakly correlated to prior fractures, age and fracture scores (r= 0.218, 0.208, 0.254, α<.0001). Increases in reported prior fractures and age were associated with increases in fractures.

Table 3

Pearson Correlation Coefficients, N = 500
Prob > |r| under H0: Rho=0

	FRACTURE	PRIORFRAC	AGE	WEIGHT	HEIGHT	BMI	PREMENO	MOMFRAC	ARMASSIST	SMOKE	RATERISK	FRACSCORE	RISKRATE1	RISKRATE2	BMIG1	BMIG2
FRACTURE	1.00000	0.21809 <.0001	0.20765 <.0001	-0.03626 0.4185	-0.13640 0.0022	0.02961 0.5089	0.00876 0.8451	0.10644 0.0173	0.15257 0.0006	-0.03168 0.4797	0.15173 0.0007	0.26448 <.0001	0.01433 0.7492	0.12419 0.0054	0.01240 0.7821	0.02040 0.6491

Measure Creation

Continuous and multilevel variables were re-coded to be represented as dichotomous dummy variables to eliminate measurements of the same trait and to ease the interpretation of the model results by keeping the model additive.

BMI, Height & Weight

Height and weight are used to calculate BMI and it explains over 74% of the variance in weight (r=.860, α<.0001) so they were eliminated from consideration and replaced with BMI.

BMI was further refined to be represented by 2 dummy variables. BMI was first categorized into normal weight (BMI < 25), overweight (25<= BMI < 30) and obese (BMI >= 30). Then, the results were re-coded as shown in table 2.

BMI Group	BMIG1	BMIG2
BMI < 25	0	0
25<= BMI < 30	1	0
BMI >= 30)	0	1

Finally, the categorized BMI variable was re-coded using two dummy variables with normal weight being the reference measure. The three levels of RATERISK were re-coded in the same fashion.

Rate Risk Group	RISKRATE1	RISKRATE2
1	0	0
2	1	0
3	0	1

Initial Full Model

All the variables were included as originally formatted for initial model comparison to the re-coded data model. The modeling used Binary logit with Fisher's scoring optimization technique. These tests for differences in models caused by transforming the variables were evaluated with a null hypothesis of a difference between the models. The ROC comparison curves in figure 1 show very little effect of the data transformation on the model. The area under the curve only decreased by 0.0044 because of the change (ROC area = 0.7204, 0.7161).

The ROC Chi-square Test of Contrasts shows no significant difference between the models with the probability of the Chi-square statistic = .9097(chi-square = 0.1893, df=2). We fail to reject the null hypothesis of contrasts between the models in favor of the alternative of no difference between the models.

Table 4

ROC Contrast Test Results			
Contrast	DF	Chi-Square	Pr > ChiSq
Reference = FULL MODEL	2	0.1893	0.9097

ROC Contrast Estimation and Testing Results by Row						
Contrast	Estimate	Standard Error	95% Wald Confidence Limits		Chi-Square	Pr > ChiSq
Model - FULL MODEL	0.00423	0.00995	-0.0153	0.0237	0.1812	0.6703
Original FULL MODEL - FULL MODEL	0.00434	0.0101	-0.0154	0.0241	0.1862	0.6661

Furthermore, table 3 shows the ROC Contrast Estimation & Test Results by Row. The 95% confidence intervals cross the zero line and the probability of test statistics were 0.6703 and 0.6664 respectively. We cannot be confident the estimate is different from zero. We reject the null hypothesis of contrast by row between the models in favor of the alternative hypotheses of the models not contrasting by row. There is not sufficient evidence to conclude a difference between the models. The additive model will be used for further analysis.

Analysis Full Model

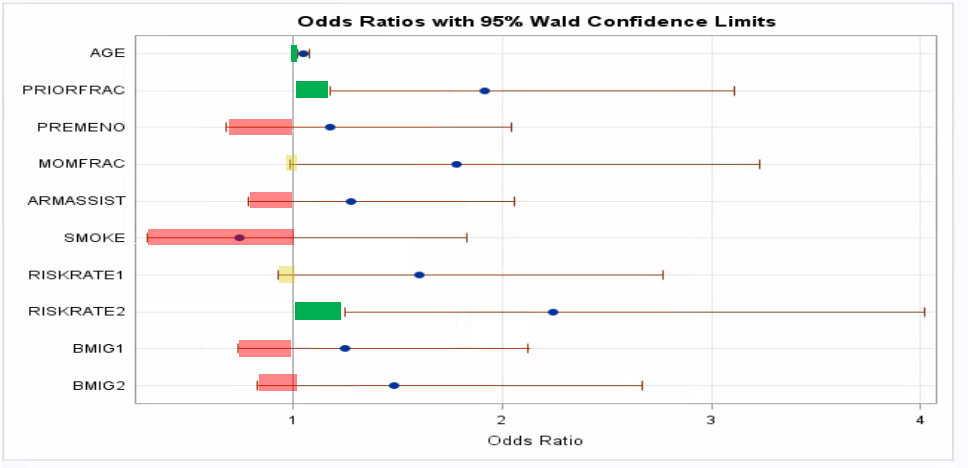
The maximum likelihood estimates(MLE) and odds ratios can be seen on tables 5 & 6 at right. They show the full model consisted of 10 dichotomous variables and 1 continuous variable, age. Not all variables were contributing to the model. The MLE shows the Chi-Square statistic for PREMENO, ARMASSIST, SMOKE, BMIG1, BMIG2 did not reach significance at the α=.05 level at 0.56, 0.33, 0.52, 0.41 and 0.19 respectively(df=1). The parameters with significant contribution were led by AGE with a Chi-square value 4 points higher than the next highest Chi-square, RiskRate2 (11.40, 7.29 df=1) and 8.5 points higher than the lowest Chi-square of contributing parameters, RiskRate1 (2.88).

Tables 5 & 6

Analysis of Maximum Likelihood Estimates					
Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-5.4219	1.0549	26.4190	<.0001
PRIORFRAC	1	0.6492	0.2472	6.8979	0.0086
AGE	1	0.0475	0.0141	11.4026	0.0007
PREMENO	1	0.1639	0.2806	0.3413	0.5591
MOMFRAC	1	0.5770	0.3036	3.6128	0.0573
ARMASSIST	1	0.2416	0.2449	0.9731	0.3239
SMOKE	1	-0.2958	0.4604	0.4129	0.5205
RISKRATE1	1	0.4721	0.2780	2.8838	0.0895
RISKRATE2	1	0.8061	0.2985	7.2935	0.0069
BMIG1	1	0.2236	0.2695	0.6884	0.4067
BMIG2	1	0.3955	0.2995	1.7434	0.1867

The Odds Ratio Estimates and Wald Confidence Intervals graph (figure 2) reinforces the finding of lack of contribution. The measures in red were mentioned above. The metrics in yellow are on the borderline for inclusion since their p values were just above the .05 criteria (0.057, 0.089, df=1). The metrics in green have p values less alpha = .05 thus adding to the model. Using these results as a guide, model selection continued with auto selection methods.

Figure 2



Model Selection

Model selection used forward, backward and stepwise auto selection processes to establish a restricted model. The selection criteria were p< .30 for inclusion and .05 to stay in the model. The parameter contributions are listed in the table below. PRIORFRAC surpassed AGE in Chi-square over the previous analysis and are currently at 23.78 and 25.56 with one degree of freedom. The previously identified non-significant parameters are still above the .05 alpha level.

The summary of the Stepwise Selection below shows the results of each step. The model selected the parameters in ranking order by Chi-square starting with PRIORFRAC and ending with MOMFRAC (23.78 ,3.19). None of the other parameters reached the criteria for entry (P<.3). MOMFRAC(p=.074) was then excluded from the model since it did not meet the criteria to stay (p<.05) . The HL Goodness of Fit test showed the overall model fit since we fail to reject the null hypothesis of good fit (Chi-square12.3357, $\mathrm{DF}= 8$, $\alpha = 0.1368$).

The odds ratio table shows the 4 parameters of the restricted model. In this model, perceiving self as being at higher risk then peers contributed the most followed by prior fractures, RISKRATE1 and AGE (estimates, 2.377, 2.014, 1.731,1.047 (see table 10 for confidence intervals).

Table 7

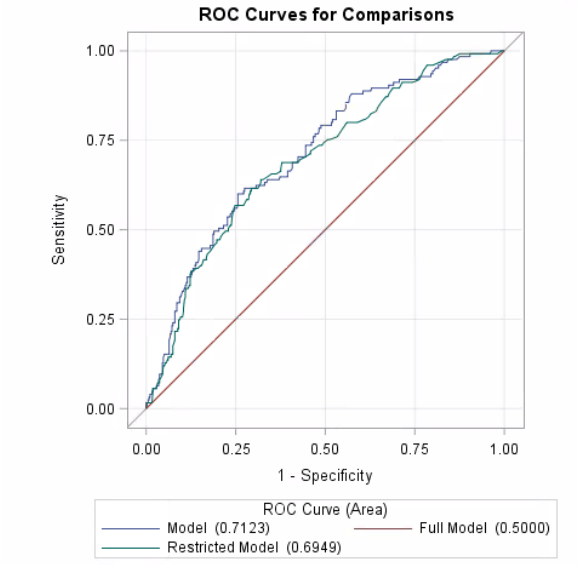
Summary of Stepwise Selection

Step	Effect Entered	Effect Removed	DF	Number In	Score Chi-Square	Wald Chi-Square	Pr > ChiSq
1	PRIORFRAC		1	1	23.7812		<.0001
2	AGE		1	2	11.7015		0.0006
3	RISKRATE2		1	3	5.7049		0.0169
4	RISKRATE1		1	4	4.0258		0.0448
5	MOMFRAC		1	5	3.1932		0.0739
6		MOMFRAC	1	4		3.1546	0.0757

Restricted Model

The restricted model was tested against the full model to ensure it did not contrast from the full model. The ROC curve below shows a loss small loss of area under the curve from 0.7123 to 0.6949 but the tests of contrast and contrast by row had p values sufficient to reject the null hypothesis of contrast present (Chi-square 66.253, df = 2, p<.0001). The statistics for contrast by row were Chi-square 66.2231 and 53.0183 p< .0001.

Figure 3



Tables 8 and 9

ROC Contrast Test Results

Contrast	DF	Chi-Square	Pr > ChiSq
Reference = Full Model	2	66.2533	<.0001

ROC Contrast Estimation and Testing Results by Row

Contrast	Estimate	Standard Error	95% Wald Confidence Limits		Chi-Square	Pr > ChiSq
Model - Full Model	0.2123	0.0261	0.1612	0.2634	66.2231	<.0001
Restricted Model - Full Model	0.1949	0.0268	0.1424	0.2473	53.0183	<.0001

The model

These analyses yielded the following logistic regression model parameters for the prediction of a fracture within one year.

$$In(\mathrm{P(fracture)}/(1-\mathrm{P(fracture)} = \ -4.9905 + .0459 * x_1 + .7002x_2 + \ .5485x_3 + .8657x_4$$

Where x_1 = Age, x_2=Prior Fractures x_3= Self Rating of at same level of risk as peers and x_4= Self Rating of being at higher level of risk than peers.

Conclusion

Osteoporosis will continue to be a significant risk to women of age 55 and above. However, preventive measures and a change in lifestyle choices can be made to reduce the impacts of Osteoporosis. After a certain Age, with prior history, and reducing the risk taking can certainly help women to have minimal risk of having fractures due to Osteoporosis.

Prediction models with the awareness and preventive measures one can reduce the impacts from Osteoporosis.

Appendix

Graphs and Figures

Figure 1.

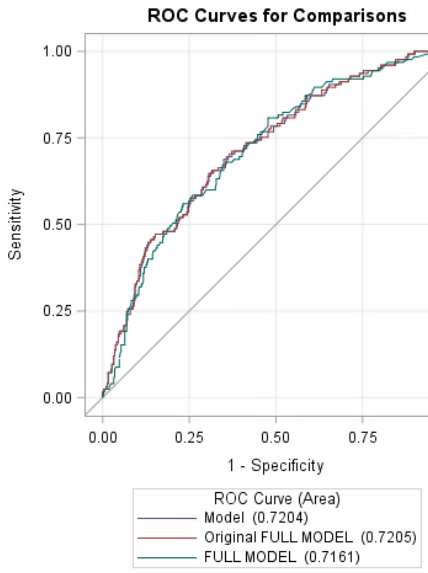


Table 5

Odds Ratio Full Model

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
AGE	1.0000	1.049	1.020	1.078
PRIORFRAC	1.0000	1.914	1.179	3.107
PREMENO	1.0000	1.178	0.680	2.042
MOMFRAC	1.0000	1.781	0.982	3.228
ARMASSIST	1.0000	1.273	0.788	2.058
SMOKE	1.0000	0.744	0.302	1.834
RISKRATE1	1.0000	1.603	0.930	2.765
RISKRATE2	1.0000	2.239	1.247	4.019
BMIG1	1.0000	1.251	0.737	2.121
BMIG2	1.0000	1.485	0.826	2.671

Parameters for entry into stepwise comparisons.

Analysis of Effects Eligible for Entry			
Effect	DF	Score Chi-Square	Pr > ChiSq
AGE	1	21.5600	<.0001
PRIORFRAC	1	23.7812	<.0001
PREMENO	1	0.0384	0.8447
MOMFRAC	1	5.6646	0.0173
ARMASSIST	1	11.6385	0.0006
SMOKE	1	0.5018	0.4787
RISKRATE1	1	0.1027	0.7486
RISKRATE2	1	7.7117	0.0055
BMIG1	1	0.0768	0.7816
BMIG2	1	0.2081	0.6483

Table 10

Model Selection Restricted Model Odds Ratios

Odds Ratio Estimates and Wald Confidence Intervals				
Effect	Unit	Estimate	95% Confidence Limits	
AGE	1.0000	1.047	1.022	1.073
PRIORFRAC	1.0000	2.014	1.256	3.231
RISKRATE1	1.0000	1.731	1.010	2.967
RISKRATE2	1.0000	2.377	1.356	4.165

Correlations

Pearson Correlation Coefficients, N = 500
Prob > |r| under H0: Rho=0

	FRACTURE	PRIORFRAC	AGE	WEIGHT	HEIGHT	BMI	PREMENO	MOMFRAC	ARMASSIST	SMOKE	RATERISK	FRACSCORE	RISKRATE1	RISKRATE2	BMIG1	BMIG2
FRACTURE	1.00000	0.21809 <.0001	0.20765 <.0001	-0.03626 0.4185	-0.13640 0.0022	0.02961 0.5089	0.00876 0.8451	0.10644 0.0173	0.15257 0.0006	-0.03168 0.4797	0.15173 0.0007	0.26448 <.0001	0.01433 0.7492	0.12419 0.0054	0.01240 0.7821	0.02040 0.6491
PRIORFRAC	0.21809 <.0001	1.00000	0.29145 <.0001	-0.02399 0.5925	-0.10220 0.0223	0.03575 0.4251	0.00648 0.8851	0.02219 0.6206	0.19614 <.0001	0.05741 0.2000	0.17484 <.0001	0.48608 <.0001	-0.03690 0.4103	0.17145 0.0001	0.02900 0.5176	0.01742 0.6977
AGE	0.20765 <.0001	0.29145 <.0001	1.00000	-0.27160 <.0001	-0.19265 <.0001	-0.18877 <.0001	-0.15911 0.0004	0.03475 0.4382	0.23832 <.0001	-0.09049 0.0431	-0.04889 0.2752	0.86992 <.0001	0.00897 0.8414	-0.04723 0.2919	0.00270 0.9520	-0.17210 0.0001
WEIGHT	-0.03626 0.4185	-0.02399 0.5925	-0.27160 <.0001	1.00000	0.31597 <.0001	0.84283 <.0001	0.08038 0.0725	-0.06125 0.1715	0.31920 <.0001	0.00291 0.9483	-0.08288 0.0640	-0.16138 0.0003	0.08162 0.0682	-0.11529 0.0099	-0.03894 0.3849	0.78221 <.0001
HEIGHT	-0.13640 0.0022	-0.10220 0.0223	-0.19265 <.0001	0.31597 <.0001	1.00000	0.01857 0.6786	-0.00901 0.8408	0.06963 0.1199	0.07060 0.1148	-0.02437 0.5867	-0.01660 0.7111	-0.16200 0.0003	-0.05781 0.1969	0.01624 0.7171	-0.02765 0.5374	0.03101 0.4890
BMI	0.02961 0.5089	0.03575 0.4251	-0.18877 <.0001	0.84283 <.0001	0.01857 0.6786	1.00000	0.09437 0.0349	-0.07341 0.1011	0.28196 <.0001	0.01637 0.7151	-0.09602 0.0318	-0.09502 0.0336	0.09961 0.0259	-0.13624 0.0023	0.08838 0.0483	0.85887 <.0001
PREMENO	0.00876 0.8451	0.00648 0.8851	-0.15911 0.0004	0.08038 0.0725	-0.00901 0.8408	0.09437 0.0349	1.00000	-0.00917 0.8379	0.07861 0.0791	0.10328 0.0209	0.07592 0.0899	-0.07853 0.0794	0.07237 0.1060	0.02755 0.5387	-0.00919 0.8376	0.09007 0.0441
MOMFRAC	0.10644 0.0173	0.02219 0.6206	0.03475 0.4382	-0.06125 0.1715	0.06963 0.1199	-0.07341 0.1011	-0.00917 0.8379	1.00000	0.00688 0.8781	-0.01282 0.7749	0.12473 0.0052	0.17565 <.0001	0.01009 0.8219	0.10299 0.0213	0.04252 0.3427	-0.08825 0.0486
ARMASSIST	0.15257 0.0006	0.19614 <.0001	0.23832 <.0001	0.31920 <.0001	0.07060 0.1148	0.28196 <.0001	0.07861 0.0791	0.00688 0.8781	1.00000	0.06214 0.1653	0.12270 0.0060	0.57270 <.0001	0.00055 0.9903	0.10629 0.0174	-0.01582 0.7242	0.26311 <.0001
SMOKE	-0.03168 0.4797	0.05741 0.2000	-0.09049 0.0431	0.00291 0.9483	-0.02437 0.5867	0.01637 0.7151	0.10328 0.0209	-0.01282 0.7749	0.06214 0.1653	1.00000	0.00396 0.9296	0.07726 0.0844	-0.01654 0.7121	0.01222 0.7853	-0.03585 0.4238	0.03324 0.4584
RATERISK	0.15173 0.0007	0.17484 <.0001	-0.04889 0.2752	-0.08288 0.0640	-0.01660 0.7111	-0.09602 0.0318	0.07592 0.0899	0.12473 0.0052	0.12270 0.0060	0.00396 0.9296	1.00000	0.08207 0.0667	0.03890 0.3854	0.84797 <.0001	-0.03603 0.4215	-0.06830 0.1272
FRACSCORE	0.26448 <.0001	0.48608 <.0001	0.86992 <.0001	-0.16138 0.0003	-0.16200 0.0003	-0.09502 0.0336	-0.07853 0.0794	0.17565 <.0001	0.57270 <.0001	0.07726 0.0844	0.08207 0.0667	1.00000	-0.00967 0.8292	0.07641 0.0878	-0.03787 0.3981	-0.06646 0.1378
RISKRATE1	0.01433 0.7492	-0.03690 0.4103	0.00897 0.8414	0.08162 0.0682	-0.05781 0.1969	0.09961 0.0259	0.07237 0.1060	0.01009 0.8219	0.00055 0.9903	-0.01654 0.7121	0.03890 0.3854	-0.00967 0.8292	1.00000	-0.49666 <.0001	0.02534 0.5719	0.07705 0.0852
RISKRATE2	0.12419 0.0054	0.17145 0.0001	-0.04723 0.2919	-0.11529 0.0099	0.01624 0.7171	-0.13624 0.0023	0.02755 0.5387	0.10299 0.0213	0.10629 0.0174	0.01222 0.7853	0.84797 <.0001	0.07641 0.0878	-0.49666 <.0001	1.00000	-0.04474 0.3181	-0.10020 0.0251
BMIG1	0.01240 0.7821	0.02900 0.5176	0.00270 0.9520	-0.03894 0.3849	-0.02765 0.5374	0.08838 0.0483	-0.00919 0.8376	0.04252 0.3427	-0.01582 0.7242	-0.03585 0.4238	-0.03603 0.4215	-0.03787 0.3981	0.02534 0.5719	-0.04474 0.3181	1.00000	-0.43429 <.0001
BMIG2	0.02040 0.6491	0.01742 0.6977	-0.17210 0.0001	0.78221 <.0001	0.03101 0.4890	0.85887 <.0001	0.09007 0.0441	-0.08825 0.0486	0.26311 <.0001	0.03324 0.4584	-0.06830 0.1272	-0.06646 0.1378	0.07705 0.0852	-0.10020 0.0251	-0.43429 <.0001	1.00000

Restricted Model

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.9905	0.9027	30.5641	<.0001
AGE	1	0.0459	0.0124	13.6179	0.0002
PRIORFRAC	1	0.7002	0.2412	8.4308	0.0037
RISKRATE1	1	0.5485	0.2750	3.9786	0.0461
RISKRATE2	1	0.8657	0.2862	9.1500	0.0025

```
data glow;
  infile "\\CLIENT\C$\SASDATA\glow500.csv"
delimiter = ","  firstobs = 2;
 input SUB_ID $ SITE_ID $ PHY_ID $ PRIORFRAC AGE WEIGHT
      HEIGHT BMI PREMENO MOMFRAC ARMASSIST SMOKE RATERISK FRACSCORE FRACTURE;
DROP  SUB_ID SITE_ID  PHY_ID;

IF RATERISK = 1 THEN RISKRATE1 = 0;
IF RATERISK = 2 THEN RISKRATE1 = 1;
IF RATERISK = 3 THEN RISKRATE1 = 0;

IF RATERISK = 1 THEN RISKRATE2 = 0;
IF RATERISK = 2 THEN RISKRATE2 = 0;
IF RATERISK = 3 THEN RISKRATE2 = 1;


IF BMI < 25 THEN BMI2 = 0;
IF 25 <= BMI < 30 THEN BMI2=1;
IF BMI >= 30 THEN BMI2= 2;

BMI = BMI2;

if BMI= 0 THEN BMIG1 = 0;
if BMI= 1 THEN BMIG1 = 1;
if BMI= 2 THEN BMIG1 = 0;

if BMI= 0 THEN BMIG2 = 0;
if BMI= 1 THEN BMIG2 = 0;
if BMI= 2 THEN BMIG2 = 1;


if age < 56 then agegroup = 0 ;
if 56 <= age < 61 then agegroup = 1 ;
if 61 <= age < 66 then agegroup = 2;
if 66 <= age < 71 then agegroup = 3;
if 71 <= age < 76 then agegroup = 3 ;
if 76 <= age < 81 then agegroup = 4 ;
if age >= 81 then agegroup = 5;


ident = _n_;
run;
proc print data = glow; run;
proc means data = glow; run;

proc freq;
table PRIORFRAC BMI PREMENO MOMFRAC ARMASSIST SMOKE RATERISK FRACSCORE FRACTURE
proc report data = glow; run;


TITLE 'Correlation';
PROC CORR DATA = GLOW PLOTS() =MATRIX;
VAR
FRACTURE
PRIORFRAC
AGE
WEIGHT
HEIGHT
BMI
PREMENO
MOMFRAC
ARMASSIST
SMOKE
RATERISK
FRACSCORE
RISKRATE1
RISKRATE2
BMIG1
BMIG2
;
RUN;
```

```
/**** comparing original measures to recoded measures / this makes the model additive and not
multiplicative
/*** MODEL COMARISON***/
proc logistic data=GLOW order = data plots=ROC(ID=PROB);
model FRACTURE(EVENT=LAST) =
PRIORFRAC
AGE
WEIGHT
HEIGHT
BMI
PREMENO
MOMFRAC
ARMASSIST
SMOKE
RATERISK
FRACSCORE
RISKRATE1
RISKRATE2
BMIG1
BMIG2
/ RSQUARE;

ROC 'Original FULL MODEL' PRIORFRAC
AGE
WEIGHT
HEIGHT
BMI
PREMENO
MOMFRAC
ARMASSIST
SMOKE
RATERISK
FRACSCORE;

ROC 'FULL MODEL'
PRIORFRAC
PREMENO
MOMFRAC
ARMASSIST
SMOKE
RATERISK
FRACSCORE
RISKRATE1
RISKRATE2
BMIG1
BMIG2;

roccontrast reference('FULL MODEL') / estimate e;
run;

ods graphics on;
title 'Full Model Risk of the Occurrence of Fractures within One Year';
proc logistic data=GLOW order = data plots(only label)=(phat leverage dpc oddsratio);
model FRACTURE(EVENT=LAST) =
PRIORFRAC
AGE
PREMENO
MOMFRAC
ARMASSIST
SMOKE
RISKRATE1
RISKRATE2
BMIG1
BMIG2
/ RISKLIMITS influence LACKFIT;

run;
ods graphics off;
```

```
ods graphics on;
title 'Stepwise Model Selection on Glow Data';
proc logistic data=Glow outest=betas covout plots(only label)=(phat leverage dpc oddsratio);
   model FRACTURE(event='1')=
                        AGE
                        PRIORFRAC
                        PREMENO
                        MOMFRAC
                        ARMASSIST
                        SMOKE
                        RISKRATE1
                        RISKRATE2
                        BMIG1
                        BMIG2

                / selection=stepwise
                  slentry=0.3
                  slstay=0.05
                  details
                  lackfit influence risklimits;
roc "Full Model";
roc 'Restricted Model'
AGE
PRIORFRAC
RISKRATE1
RISKRATE2;

   output out=pred p=phat lower=lcl upper=ucl
          predprob=(individual crossvalidate);
roccontrast reference("Full Model")/ estimate e;

run;
ods graphics off;
```

```
ods graphics on;
title 'Testing Full vs Restricted Model';
proc logistic data=Glow outest=betas covout plots(only label)=(ROC(ID=PROB) oddsratio);
   model FRACTURE(event='1')=
                        AGE
                        PRIORFRAC
                        PREMENO
                        MOMFRAC
                        ARMASSIST
                        SMOKE
                        RISKRATE1
                        RISKRATE2
                        BMIG1
                        BMIG2

                /RSQUARE lackfit influence risklimits;
roc "Full Model";
roc 'Restricted Model'
AGE
PRIORFRAC
RISKRATE1
RISKRATE2;
roccontrast reference("Full Model")/ estimate e;

run;
ods graphics off;

ods graphics on;
title 'Restricted Model';
proc logistic data=Glow outest=betas covout plots(only label)=(ROC(ID=PROB) oddsratio);
```

```
   model FRACTURE(event='1')=
                        AGE
                        PRIORFRAC
                        RISKRATE1
                        RISKRATE2
                    /RSQUARE lackfit influence risklimits;
roc "Full Model";
roc 'priorfrac Model'
AGE
PRIORFRAC;
roccontrast reference("Full Model")/ estimate e;

run;
ods graphics off;
```