Factors Associated with Mortality Following Admission to a Hospital for a Gun Shot Wound

Department of Mathematics and Statistics

University of Vermont

Burlington, Vermont, USA

Zihao Huang

Zhuang12@uvm.edu

Abstract:

Each year there are about 30,000 deaths from gunshot wounds. To identify fatal gunshot wounds, in this paper, we focus on the data GSW, that we choose several variables from GSW and simulate the data into logistic regression, then we found the interactions among the variables. Next, we improved the model and then assessed the fit of goodness for this model, finally we display some outliers to discuss.

Keywords: Gunshot wound (GSW), Stata, Logistic Regression

Introduction

“A gunshot wound, also known as (GSW), ballistic trauma, or bullet wound, is a form of physical trauma sustained from the discharge of arms or munitions.” (Wiki)

There 30,000 deaths from gunshot wounds in United States. The accidents are common and recently it becomes a popular topic. To assess the factors of gunshot wounds, we met the trouble first that the gunshot wound data is difficult to obtain. Because since 1994, it is prohibited to collect data of guns and gun control. However, we are able to get the data related to gunshot wounds. Thus, we could obtain the information of patients who shot by guns.

In this paper, we afford to figure out the factors which may be fatal to humans. Firstly, we focus on the data GSW, that we choose several variables from GSW and simulate the data into logistic regression, then we found the interactions among the variables. Next, we improved the model and then assessed the fit of goodness for this model, finally we display some outliers to discuss and we have a conclusion and discussion.

Method

Data

Using the data of GSW, there are approximately 60,000 observations. According to the data, we consider the variables: DIED, AGE, FEMALE, SHOCK, TMPM, AWEEKEND, YEAR, GUN_INT for the study. DIED is binary variable, “0” represents alive after gunshot wound while “1” shows dead. AGE, a continuous variable, is the age of patient while the accident is recorded. FEMALE is a binary variable showing “0” is male and “1” us female. SHOCK is “life-threatening medical condition of low blood perfusion to tissues” (shock, wiki), and “0” represents patient suffered a shock from gunshot, while “1” didn’t. TMPM is a continuous variable to show the severity of gunshot wound. It is a log-likelihood figures with the range -12.77 to 3.54. AWEEKEND is a binary variable, “1” means the gunshot happened at weekend but “0” didn’t. YEAR is a continuous variable recording the year of the accident. In the paper, DIED is used as an outcome.

Statistical Model – logistic regression

“logistic regression, or logit regression” (David A. Freedman, 2009), or logit model is a regression model where the dependent variable (DV) is categorical. “This article covers the case of a binary dependent variable—that is, where it can take only two values, "0" and "1", which represent outcomes such as pass/fail, win/lose, alive/dead or healthy/sick. Cases where the dependent variable has more than two outcome categories may be analyzed in multinomial logistic regression, or, if the multiple categories are ordered, in ordinal logistic regression” (Walker, SH; Duncan, DB, 1967).

Procedure and Result

Firstly, by using code “drop if(variable =.)” and 42,015 observations reserved in the dataset we use univariate model to check each variable we chose. According to the univariate test, we found the variables AGE, FEMALE, TMPM, SHOCK and GSW_INT are significant. We applied these variables into a new model after collapsing YEAR and AWEEKEND (from Table 1.1 to Table 1.3).

For AGE and TMPM are the continuous variables, we would figure out how well they may be applied into this new model. Here we use fractional polynomials for AGE and TMPM. According to the result of Table 1.4, we found m=1 is the best result for AGE, thus, we generate a new variable “AGE1” for AGE.

AGE1=(AGE/10) ^3

For TMPM, they have a range from -12.77 to 3.54, (TMPM) ^3 * ln(TMPM/10) cannot be used. we applied AGE1 into the model.

We explore possible interaction, we add each of the interaction into the model and test the chi-square with df=1 and its p-value listed in Table 1.7. From the result of Table 1.7, we found some interactions are 10 percent significant, then we applied those into the model (Table 1.8). We fit the model by eliminating the non-significant interactions, thus, we got the revised logistic regression model (Table 1.9) and its estimated odds ratio (Table 1.9-1). We check the linearity of continuous variables AGE and TMPM by using Lowess smoother (Table 1.10).

The model is shown below:

$$Y = \beta_0 + \beta_1 \times AGE1 + \beta_2 \times TMPM + \beta_3 \times FEMALE + \beta_4 \times SHOCK + \beta_5 \times GSW\_INT1 + \beta_6 \times GSW\_INT2 + \beta_7 \times GSW\_INT3 + \beta_8 \times GSW\_INT4 + \beta_9 \times AGE1 \times TMPM + \beta_{10} \times TMPM \times SHOCK + \beta_{11} \times TMPM \times GSW\_INT4 + \beta_{12} \times SHOCK \times GSW\_INT2$$

$$\pi = e^{Y} = e^{(\beta0 + \beta1 \times AGE1 + \beta2 \times TMPM + \beta3 \times FEMALE + \beta4 \times SHOCK + \beta5 \times GSW\_INT1 + \beta6 \times GSW\_INT2 + \beta7 \times GSW\_INT3 + \beta8 \times GSW\_INT4 + \beta9 \times AGE1 \times TMPM + \beta10 \times TMPM \times SHOCK + \beta11 \times TMPM \times GSW\_INT4 + \beta12 \times SHOCK \times GSW\_INT2)}$$

Assess the goodness of fit: For table 1.11, the cutpoint is.5 it shows the classification has a 98.55% specificity but it has a 45.02% sensitivity, the model has a good fit and shows that fewer people were predicted to be dead for the gunshot wound. From the table 1.12, this model has a good specificity curve but a weak sensitivity curve. It shows p=.1 may have a good classification. Thus, we tried cutpoint p=.1 and found that with cutpoint=.1 it shows the classification has a 89.26% specificity but it has a 86.09% sensitivity, the model has a good fit but it shows that more people were predicted to be dead for the gunshot wound. In table 1.14, the area under the ROC curve is 0.9422>0.9, we consider the discrimination is outstanding. Hosmer-Lemeshow goodness of fit is conducted and each group contain 1000 observations. P-value=.8706 from the Chi-square with df=998 shows that it has a good fit for this model.

Then we plot four graphs (table 1.17 and table 1.18) to show the relationship among estimated probability, deviation of Distance, deviation of Beta and deviation of Chi-Square. The observations 35397, 2754, 26953, 26858 and10637 (table 1.19) are detected as outlying. We eliminate the outlier one by one: by eliminating the No. 26858 observation, change of coefficient is -.24819028%; by eliminating the No. 2754 observation, change of coefficient is -.27576698%; by eliminating the No. 35397 observation, change of coefficient is .2964495%; by eliminating the No. 26953 observation, change of coefficient is -.15856601% by eliminating the No. 10637 observation, change of coefficient is -.50327473% and table 2 show the model without 5 outliers and its estimated odds ratio.

Conclusion and Limitations

We found the factors related to gunshot wounds are AGE, FEMALE, TMPM, SHOCK and GSW_INT, and model is:

$$Y = \beta_0 + \beta_1\times AGE1 + \beta_2\times TMPM + \beta_3\times FEMALE + \beta_4\times SHOCK + \beta_5\times GSW\_INT1 + \beta_6\times GSW\_INT2 + \beta_7\times GSW\_INT3 + \beta_8\times GSW\_INT4 + \beta_9\times AGE1\times TMPM + \beta_{10}\times TMPM\times SHOCK + \beta_{11}\times TMPM\times GSW\_INT4 + \beta_{12}\times SHOCK\times GSW\_INT2$$

The estimated odds ratio is shown in table 1.16 and the model with reducing 5 observations in table 2. We found the model has good specificity but weak sensitivity in cutpoint .5, but it is great at cut point .1. The model has an excellent discrimination and a good fit when observations are grouped. Then we cancel the outlying observation one by one, and we found it has a good fit for each model. (See table 2) And after canceling 5 outliers it performs better. Finally, the estimated odd ratio tables (Table 2.1) will identify the odds ratio with 4 interactions in this model. And Table 3 will show the 95% confident interval of estimating odd ratios for each of variables and interactions.

However, this paper has several limitations: Firstly, not all the variables in GSW are discussed. For example, GUN_TYP is essential that it may be related to fatal event. Secondly, the model has a poor performance in sensitivity, which means that in general it will predict less people is dead from the gunshot. Thirdly, Examination of outliers are not reliable. The author tried to detect the outlier from high Delta D and Delta Chi-square and Delta Beta, which may be subjective. Finally, it lacks a complete test after eliminating outliers, which may leave for future research.

Reference

1. David Hosmer and Stanley Lemeshow (2016). Applied Logistic Regression (Second Edition)

2. Silverman, Adam (Oct 2005). "Shock: A Common Pathway For Life-Threatening Pediatric Illnesses And Injuries". Pediatric Emergency Medicine Practice. 2 (10).

3. David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 128.

APENDIX

```
. logit died

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood = -11488.397  (backed up)

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(0)        =      -0.00
                                                Prob > chi2       =          .
Log likelihood = -11488.397                     Pseudo R2         =    -0.0000

------------------------------------------------------------------------------
        died |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -2.472212   .0182104  -135.76   0.000    -2.507904    -2.43652
```

Table 1.1: logistic regression of dependent variable

"Logit died" is stored as "A"

```
. logit died age

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood = -11279.513
Iteration 2:   log likelihood = -11259.844
Iteration 3:   log likelihood = -11259.779
Iteration 4:   log likelihood = -11259.779

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(1)        =     457.24
                                                Prob > chi2       =     0.0000
Log likelihood = -11259.779                     Pseudo R2         =     0.0199

------------------------------------------------------------------------------
        died |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0255678   .0011427    22.37   0.000      .023328    .0278075
       _cons |  -3.300689   .0436889   -75.55   0.000    -3.386318   -3.215061
------------------------------------------------------------------------------

. estimates store A1

. lrtest A A1

Likelihood-ratio test                                 LR chi2(1)  =    457.24
(Assumption: A nested in A1)                          Prob > chi2 =    0.0000
```

Table 1.2-1: univariate logistic regression of DIED and AGE

```
. logit died tmpm

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood =  -8667.752
Iteration 2:   log likelihood = -6711.5025
Iteration 3:   log likelihood = -6571.8311
Iteration 4:   log likelihood = -6570.4165
Iteration 5:   log likelihood = -6570.4162

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(1)        =    9835.96
                                                Prob > chi2       =     0.0000
Log likelihood = -6570.4162                     Pseudo R2         =     0.4281
```

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
tmpm	1.136162	.0157704	72.04	0.000	1.105252	1.167071
_cons	.2773033	.0351344	7.89	0.000	.2084411	.3461655

```
. estimates store A2

. lrtest A A2

Likelihood-ratio test                                 LR chi2(1)  =   9835.96
(Assumption: A nested in A2)                          Prob > chi2 =    0.0000
```

Table 1.2-2: univariate logistic regression of DIED and TMPM

```
. logit died shock

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood = -10980.502
Iteration 2:   log likelihood = -10965.574
Iteration 3:   log likelihood = -10965.549
Iteration 4:   log likelihood = -10965.549

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(1)        =    1045.70
                                                Prob > chi2       =     0.0000
Log likelihood = -10965.549                     Pseudo R2         =     0.0455
```

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
shock	1.802039	.0499604	36.07	0.000	1.704119	1.89996
_cons	-2.663872	.0203343	-131.00	0.000	-2.703727	-2.624018

```
. estimates store A3

. lrtest A A3

Likelihood-ratio test                                 LR chi2(1)  =   1045.70
(Assumption: A nested in A3)                          Prob > chi2 =    0.0000
```

Table 1.2-3: univariate logistic regression of DIED and SHOCK

```
. logit died female

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood = -11483.885
Iteration 2:   log likelihood = -11483.867
Iteration 3:   log likelihood = -11483.867

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(1)        =       9.06
                                                Prob > chi2       =     0.0026
Log likelihood = -11483.867                     Pseudo R2         =     0.0004

------------------------------------------------------------------------------
        died |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .1710546   .0558219     3.06   0.002     .0616456    .2804635
       _cons |   -2.49168   .0194269  -128.26   0.000    -2.529756   -2.453604
------------------------------------------------------------------------------

. estimates store A4

. lrtest A A4

Likelihood-ratio test                                 LR chi2(1)  =      9.06
(Assumption: A nested in A4)                          Prob > chi2 =    0.0026
```

Table 1.2-4: univariate logistic regression of DIED and FEMALE

```
. logit died gsw_int1 gsw_int2 gsw_int3 gsw_int4

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood = -10322.575
Iteration 2:   log likelihood = -10240.119
Iteration 3:   log likelihood = -10239.844
Iteration 4:   log likelihood = -10239.844

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(4)        =    2497.11
                                                Prob > chi2       =     0.0000
Log likelihood = -10239.844                     Pseudo R2         =     0.1087

------------------------------------------------------------------------------
        died |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    gsw_int1 |  -1.088773   .0801455   -13.58   0.000    -1.245855   -.9316906
    gsw_int2 |   1.390088   .0712733    19.50   0.000     1.250394    1.529781
    gsw_int3 |  -.8665442   .0676654   -12.81   0.000    -.9991659   -.7339225
    gsw_int4 |  -.6484366   .1643671    -3.95   0.000    -.9705902   -.3262831
       _cons |  -2.036152   .0615932   -33.06   0.000    -2.156873   -1.915432
------------------------------------------------------------------------------

. estimates store A5

. lrtest A A5

Likelihood-ratio test                                 LR chi2(4)  =   2497.11
(Assumption: A nested in A5)                          Prob > chi2 =    0.0000
```

Table 1.2-5: univariate logistic regression of DIED and GSWINT

```
. logit died aweekend

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood = -11485.686
Iteration 2:   log likelihood = -11485.684

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(1)        =       5.42
                                                Prob > chi2       =     0.0199
Log likelihood = -11485.684                     Pseudo R2         =     0.0002
```

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
aweekend	-.0885943	.0381775	-2.32	0.020	-.1634209	-.0137678
_cons	-2.440426	.0225917	-108.02	0.000	-2.484705	-2.396147

```
. estimates store A6

. lrtest A A6

Likelihood-ratio test                                 LR chi2(1)  =      5.42
(Assumption: A nested in A6)                          Prob > chi2 =    0.0199
```

Table 1.2-6: univariate logistic regression of DIED and AWEEKEND

```
. logit died year

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood = -11488.181
Iteration 2:   log likelihood = -11488.181

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(1)        =       0.43
                                                Prob > chi2       =     0.5110
Log likelihood = -11488.181                     Pseudo R2         =     0.0000
```

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
year	-.0041966	.0063842	-0.66	0.511	-.0167094	.0083162
_cons	5.957421	12.82376	0.46	0.642	-19.17668	31.09153

```
. estimates store A7

. lrtest A A7

Likelihood-ratio test                                 LR chi2(1)  =      0.43
(Assumption: A nested in A7)                          Prob > chi2 =    0.5110
```

Table 1.2-7: univariate logistic regression of DIED and YEAR

```
. logit died age female tmpm shock  gsw_int1 gsw_int2 gsw_int3 gsw_int4

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood =  -8897.286
Iteration 2:   log likelihood = -6342.2774
Iteration 3:   log likelihood = -6041.3173
Iteration 4:   log likelihood = -6038.8375
Iteration 5:   log likelihood = -6038.8366
Iteration 6:   log likelihood = -6038.8366

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(8)        =   10899.12
                                                Prob > chi2       =     0.0000
Log likelihood = -6038.8366                     Pseudo R2         =     0.4744
```

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	.0172032	.0016926	10.16	0.000	.0138858	.0205207
female	-.2505775	.0778501	-3.22	0.001	-.4031608	-.0979942
tmpm	1.071244	.0163761	65.42	0.000	1.039147	1.10334
shock	1.059737	.0655743	16.16	0.000	.9312133	1.18826
gsw_int1	-.7404383	.1037683	-7.14	0.000	-.9438203	-.5370562
gsw_int2	.5964517	.0992372	6.01	0.000	.4019503	.790953
gsw_int3	-.8455968	.0889687	-9.50	0.000	-1.019972	-.6712214
gsw_int4	-.8970008	.1990463	-4.51	0.000	-1.287124	-.5068773
_cons	-.0303966	.1007803	-0.30	0.763	-.2279223	.1671291

Table 1.3: Preliminary logistic model

Deviance: **12024.66**. Best powers of **age** among **44** models fit: **0 3**.

Fractional polynomial model comparisons:

age	df	Deviance	Dev. dif.	P (*)	Powers
Not in model	0	12178.971	154.313	0.000	
Linear	1	12077.673	53.015	0.000	1
m = 1	2	12028.961	4.303	0.116	3
m = 2	4	12024.658	—	—	0 3

Table 1.4 Fractional Polynomial for AGE

```
Deviance: 11999.39. Best powers of tmpm among 44 models fit: 3 3.

Fractional polynomial model comparisons:
--------------------------------------------------------------------
tmpm                 df       Deviance    Dev. dif.   P (*)   Powers
--------------------------------------------------------------------
Not in model          0      19426.409    7427.024    0.000
Linear                1      12077.673      78.288    0.000   1
m = 1                 2      12034.221      34.836    0.000   0
m = 2                 4      11999.385         —         —    3 3
```

Table 1.5 Fractional Polynomial for TMPM

```
. logit died tmpm age1 female shock  gsw_int1 gsw_int2 gsw_int3 gsw_int4

Iteration 0:   log likelihood = -11488.397
Iteration 1:   log likelihood = -8884.4004
Iteration 2:   log likelihood = -6325.0023
Iteration 3:   log likelihood = -6016.9451
Iteration 4:   log likelihood = -6014.3044
Iteration 5:   log likelihood = -6014.3032
Iteration 6:   log likelihood = -6014.3032

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(8)        =   10948.19
                                                Prob > chi2       =     0.0000
Log likelihood = -6014.3032                     Pseudo R2         =     0.4765

------------------------------------------------------------------------------
        died |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        tmpm |   1.072762   .0164116    65.37   0.000     1.040596    1.104928
        age1 |   .0030099   .0002391    12.59   0.000     .0025412    .0034786
      female |   -.237344   .0777513    -3.05   0.002    -.3897337   -.0849543
       shock |   1.065509   .0656519    16.23   0.000     .9368336    1.194184
    gsw_int1 |  -.7592492   .1040193    -7.30   0.000    -.9631233    -.555375
    gsw_int2 |   .5750286   .0991012     5.80   0.000     .3807939    .7692633
    gsw_int3 |  -.8419181   .0891342    -9.45   0.000    -1.016618   -.6672182
    gsw_int4 |  -.8655799   .1990511    -4.35   0.000    -1.255713   -.4754469
       _cons |   .3476166   .0887935     3.91   0.000     .1735845    .5216486
------------------------------------------------------------------------------
```

Table 1.6: another preliminary model

INTERACTION	Chi-square (1)	P
age1xtmpm	22.30	<0.0001
age1xfemale	2.38	0.1226
age1xshock	1.95	0.1626
age1xgsw_int1	1.68	0.1945

age1xgsw_int2	0.20	0.6516
age1xgsw_int3	0.03	0.8631
age1xgsw_int4	0.00	0.9470
tmpmxfemale	0.20	0.6557
tmpmxshock	56.05	0.0000
tmpmxgsw_int1	1.87	0.1715
tmpmxgsw_int2	7.87	0.0050
tmpmxgsw_int3	2.30	0.1292
tmpmxgsw_int4	9.93	0.0016
femalexshock	3.01	0.0829
femalexgsw_int1	1.93	0.1645
femalexgsw_int2	6.64	0.0100
femalexgsw_int3	0.56	0.4545
femalexgsw_int4	0.08	0.7791
shockxgsw_int1	0.48	0.4880
shockxgsw_int2	8.86	0.0029
shockxgsw_int3	6.37	0.0116
shockxgsw_int4	0.19	0.6612

Table 1.7: Interactions

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age1	.0015831	.0003873	4.09	0.000	.000824	.0023423
tmpm	1.181331	.0218087	54.17	0.000	1.138587	1.224075
female	-.0464265	.101031	-0.46	0.646	-.2444437	.1515907
shock	.6394478	.0975835	6.55	0.000	.4481876	.8307079
gsw_int1	-.7693518	.1053833	-7.30	0.000	-.9758992	-.5628043
gsw_int2	.5714336	.1261092	4.53	0.000	.3242641	.8186032
gsw_int3	-.8501177	.090339	-9.41	0.000	-1.027179	-.6730564
gsw_int4	-1.468518	.2746755	-5.35	0.000	-2.006872	-.9301642
age1xtmpm	-.0005945	.0001399	-4.25	0.000	-.0008688	-.0003202
tmpmxshock	-.3472609	.0430886	-8.06	0.000	-.4317129	-.2628088
tmpmxgsw_int2	-.0793051	.0408816	-1.94	0.052	-.1594315	.0008213
tmpmxgsw_int4	-.3446125	.1052304	-3.27	0.001	-.5508603	-.1383647
femalexshock	-.2352791	.2000347	-1.18	0.240	-.62734	.1567817
femalexgsw_int2	-.4095404	.1618064	-2.53	0.011	-.7266751	-.0924056
shockxgsw_int2	-.5012552	.1692929	-2.96	0.003	-.8330631	-.1694472
_cons	.5144881	.0939902	5.47	0.000	.3302706	.6987056

Table 1.8: preliminary model with interactions

From table 1.8, we found some interactions are not significant and they made FEMALE not significant, hence, we remove several interactions.

```
Logistic regression                             Number of obs     =     42,015
                                                LR chi2(12)       =   11046.43
                                                Prob > chi2       =     0.0000
Log likelihood = -5965.1823                     Pseudo R2         =     0.4808
```

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age1	.0014505	.0003789	3.83	0.000	.0007079	.0021932
tmpm	1.170753	.0207879	56.32	0.000	1.130009	1.211496
female	-.2265152	.0773656	-2.93	0.003	-.3781491	-.0748814
shock	.6188678	.0952579	6.50	0.000	.4321658	.8055698
gsw_int1	-.7733576	.1047922	-7.38	0.000	-.9787465	-.5679686
gsw_int2	.6434062	.1015107	6.34	0.000	.444449	.8423634
gsw_int3	-.8509295	.0898339	-9.47	0.000	-1.027001	-.6748582
gsw_int4	-1.448569	.2741221	-5.28	0.000	-1.985839	-.9112999
age1xtmpm	-.0006777	.0001323	-5.12	0.000	-.0009371	-.0004183
tmpmxshock	-.3470057	.0430519	-8.06	0.000	-.4313859	-.2626256
tmpmxgsw_int4	-.3274893	.1048273	-3.12	0.002	-.5329472	-.1220315
shockxgsw_int2	-.5471816	.1692899	-3.23	0.001	-.8789837	-.2153795
_cons	.5147329	.0924672	5.57	0.000	.3335006	.6959653

Table 1.9: revised logistic regression model

```
Logistic regression                             Number of obs     =     42,015
                                                LR chi2(12)       =   11046.43
                                                Prob > chi2       =     0.0000
Log likelihood = -5965.1823                     Pseudo R2         =     0.4808
```

died	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age1	1.001452	.0003795	3.83	0.000	1.000708	1.002196
tmpm	3.224419	.067029	56.32	0.000	3.095685	3.358506
female	.7973072	.0616842	-2.93	0.003	.6851284	.9278536
shock	1.856824	.1768772	6.50	0.000	1.54059	2.237971
gsw_int1	.4614611	.0483575	-7.38	0.000	.3757818	.5666754
gsw_int2	1.902952	.1931699	6.34	0.000	1.559631	2.321848
gsw_int3	.4270178	.0383607	-9.47	0.000	.3580793	.5092286
gsw_int4	.2349061	.064393	-5.28	0.000	.1372654	.4020013
age1xtmpm	.9993225	.0001322	-5.12	0.000	.9990634	.9995818
tmpmxshock	.7068013	.0304291	-8.06	0.000	.6496082	.7690298
tmpmxgsw_int4	.720731	.0755523	-3.12	0.002	.5868728	.8851204
shockxgsw_int2	.5785782	.0979474	-3.23	0.001	.4152047	.8062354
_cons	1.673192	.1547153	5.57	0.000	1.395846	2.005644

Table 1.9-1: Estimated Odds Ratio for covariate variables

Now, we check the linearity of continuous variables:

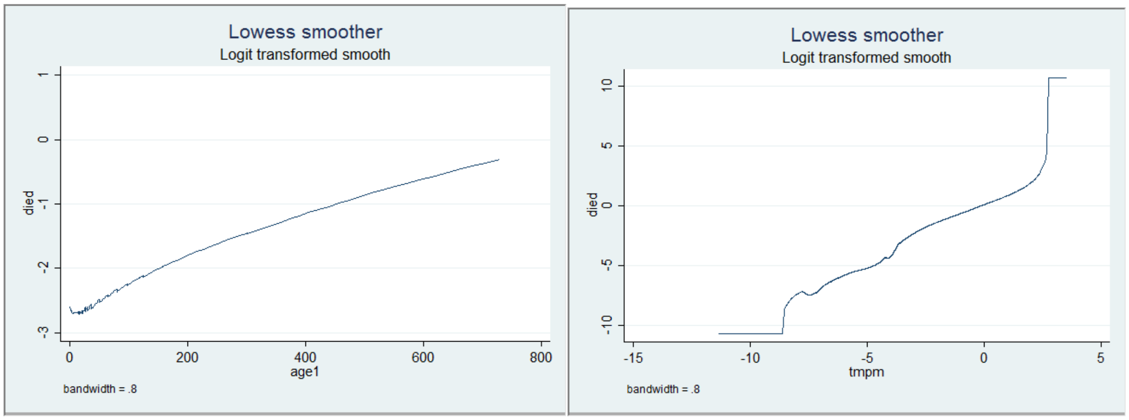

Table 1.10: Lowess smoother of AGE1 and TMPM

```
. estat classification

Logistic model for died

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |      1472           563  |       2035
     -     |      1798         38182  |      39980
-----------+--------------------------+-----------
   Total   |      3270         38745  |      42015

Classified + if predicted Pr(D) >= .5
True D defined as died != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   45.02%
Specificity                     Pr( -|~D)   98.55%
Positive predictive value       Pr( D| +)   72.33%
Negative predictive value       Pr(~D| -)   95.50%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)    1.45%
False - rate for true D         Pr( -| D)   54.98%
False + rate for classified +   Pr(~D| +)   27.67%
False - rate for classified -   Pr( D| -)    4.50%
--------------------------------------------------
Correctly classified                        94.38%
--------------------------------------------------
```

Table 1.11: Classification table with cutpoint=.5

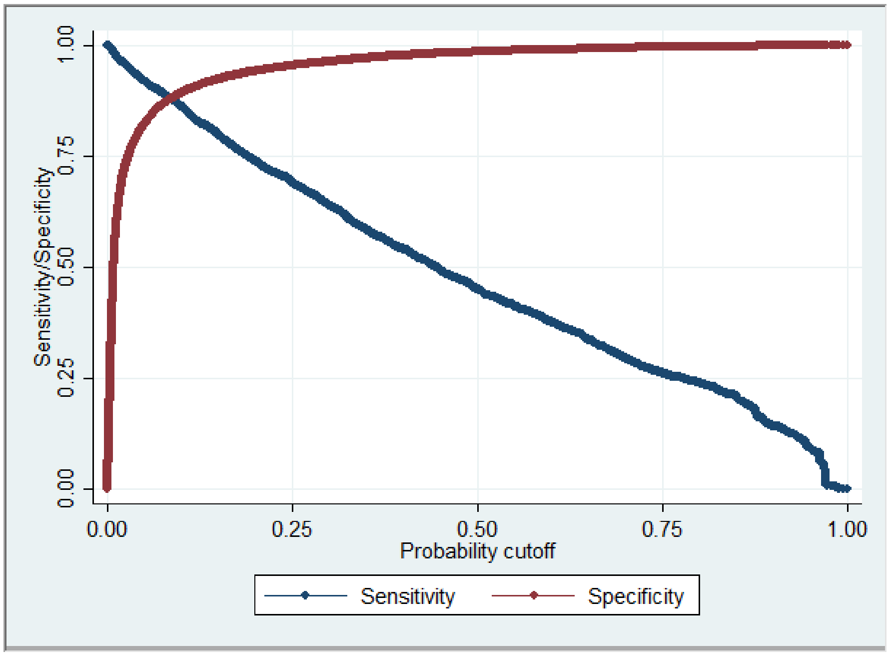

Table 1.12 : plot of ROC curve

```
Logistic model for died

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |      2815          4161  |       6976
     -     |       455         34584  |      35039
-----------+--------------------------+-----------
   Total   |      3270         38745  |      42015

Classified + if predicted Pr(D) >= .1
True D defined as died != 0
--------------------------------------------------
Sensitivity                     Pr( +| D)   86.09%
Specificity                     Pr( -|~D)   89.26%
Positive predictive value       Pr( D| +)   40.35%
Negative predictive value       Pr(~D| -)   98.70%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   10.74%
False - rate for true D         Pr( -| D)   13.91%
False + rate for classified +   Pr(~D| +)   59.65%
False - rate for classified -   Pr( D| -)    1.30%
--------------------------------------------------
Correctly classified                        89.01%
--------------------------------------------------
```

Table 1.13 : Classification table with cutpoint=.1

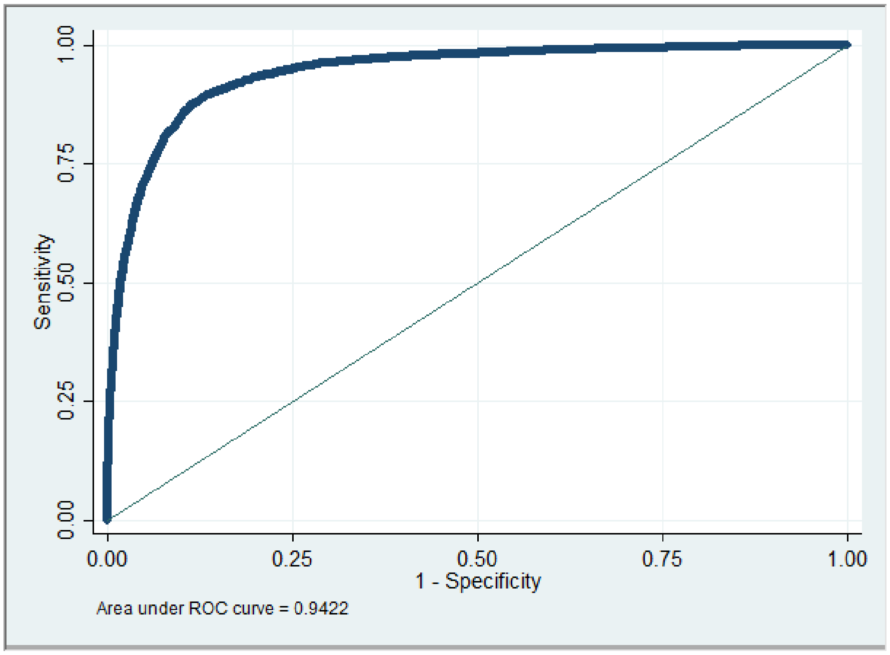

Table 1.14 : Plot of sensitivity versus 1-specificity

number of observations =	42015
number of groups =	1000
Hosmer-Lemeshow chi2(998) =	947.76
Prob > chi2 =	0.8706

Table 1.15: goodness of fit with 1000 groups

Measures of Fit for **logistic** of **died**

Log-Lik Intercept Only:	-11488.397	Log-Lik Full Model:	-5965.182
D(42002):	11930.365	LR(12):	11046.429
		Prob > LR:	0.000
McFadden's R2:	0.481	McFadden's Adj R2:	0.480
Maximum Likelihood R2:	0.231	Cragg & Uhler's R2:	0.549
McKelvey and Zavoina's R2:	0.594	Efron's R2:	0.426
Variance of y*:	8.101	Variance of error:	3.290
Count R2:	0.944	Adj Count R2:	0.278
AIC:	0.285	AIC*n:	11956.365
BIC:	-435213.770	BIC':	-10918.680

```
. logistic died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xt

Logistic regression                             Number of obs     =     42,015
                                                LR chi2(12)       =   11046.43
                                                Prob > chi2       =     0.0000
Log likelihood = -5965.1823                     Pseudo R2         =     0.4808

------------------------------------------------------------------------------
        died | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        age1 |   1.001452   .0003795     3.83   0.000     1.000708    1.002196
        tmpm |   3.224419    .067029    56.32   0.000     3.095685    3.358506
      female |   .7973072   .0616842    -2.93   0.003     .6851284    .9278536
       shock |   1.856824   .1768772     6.50   0.000      1.54059    2.237971
    gsw_int1 |   .4614611   .0483575    -7.38   0.000     .3757818    .5666754
    gsw_int2 |   1.902952   .1931699     6.34   0.000     1.559631    2.321848
    gsw_int3 |   .4270178   .0383607    -9.47   0.000     .3580793    .5092286
    gsw_int4 |   .2349061    .064393    -5.28   0.000     .1372654    .4020013
   age1xtmpm |   .9993225   .0001322    -5.12   0.000     .9990634    .9995818
  tmpmxshock |   .7068013   .0304291    -8.06   0.000     .6496082    .7690298
tmpmxgsw_int4|    .720731   .0755523    -3.12   0.002     .5868728    .8851204
shockxgsw_int2|  .5785782   .0979474    -3.23   0.001     .4152047    .8062354
       _cons |   1.673192   .1547153     5.57   0.000     1.395846    2.005644
------------------------------------------------------------------------------
```

Table 1.16 : estimated odds ratio for each covariate

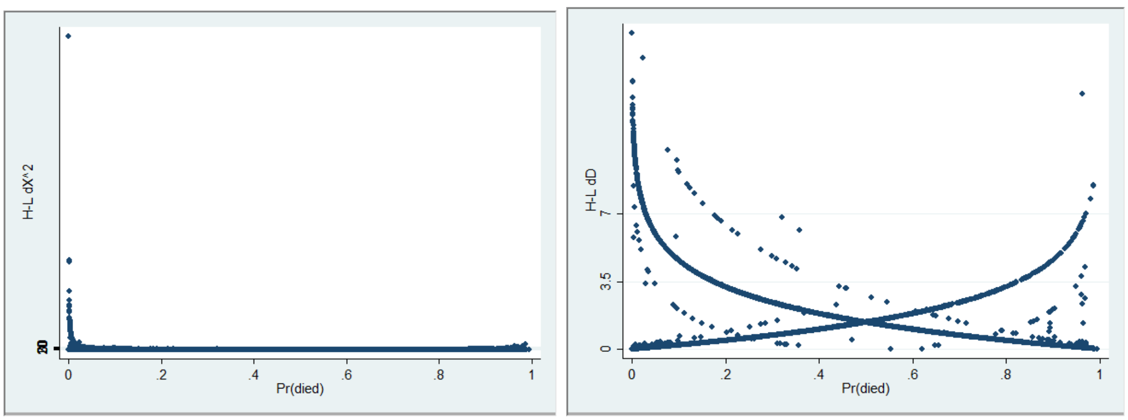

Table 1.17 : plot of Chi-square versus probility of died and plot of Deviance versus probability of died

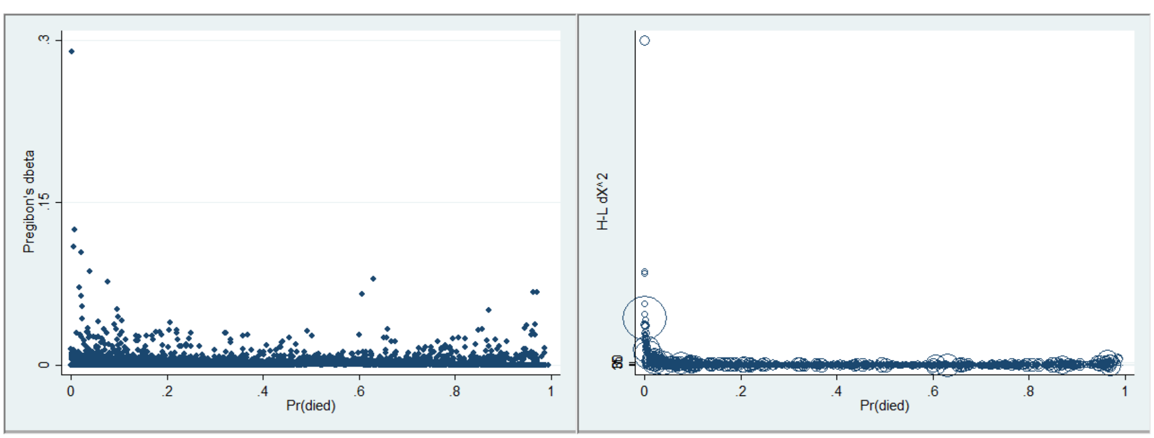

Table 1.18: plot of Distance (delta beta) versus probility of died and plot of Pearson chi-square versus probability of died

We found 5 observations are the outlying n=35397,2754,26953,26858,10637, which have large value of dx (Delta Chi-square) or dd (Delta D) and db (Delta Beta)

41898.

died	age1	tmpm	female	shock	gsw_int1	gsw_int2	gsw_int3	gsw_int4	age1xtmpm	tmpmxs~k
1	195.112	-4.602164	0	0	0	0	0	1	-897.9374	0

tmpmxgs~4	shockx~2	n	p	db	dx	dd	h
-4.602164	0	35397	.0193953	.1040313	50.66276	7.901643	.0020492

41968.

died	age1	tmpm	female	shock	gsw_int1	gsw_int2	gsw_int3	gsw_int4	age1xtmpm	tmpmxs~k
1	4.913	-4.664587	0	0	0	0	0	1	-22.91712	0

tmpmxgs~4	shockx~2	n	p	db	dx	dd	h
-4.664587	0	2754	.0078089	.109228	127.1685	9.713325	.0008582

41971.

died	age1	tmpm	female	shock	gsw_int1	gsw_int2	gsw_int3	gsw_int4	age1xtmpm	tmpmxs~k
1	46.656	-4.929499	0	0	0	0	0	1	-229.9907	0

tmpmxgs~4	shockx~2	n	p	db	dx	dd	h
-4.929499	0	26953	.0076364	.1248648	130.0768	9.759022	.000959

42013.

died	age1	tmpm	female	shock	gsw_int1	gsw_int2	gsw_int3	gsw_int4	age1xtmpm	tmpmxs~k
1	46.656	-6.834318	0	0	0	0	0	1	-318.8619	0

tmpmxgs~4	shockx~2	n	p	db	dx	dd	h
-6.834318	0	26858	.0016371	.2895994	610.1334	12.83577	.0004744

42015.

died	age1	tmpm	female	shock	gsw_int1	gsw_int2	gsw_int3	gsw_int4	age1xtmpm	tmpmxs~k
1	10.648	-6.742759	0	0	1	0	0	0	-71.7969	0

tmpmxgs~4	shockx~2	n	p	db	dx	dd	h
0	0	10637	.0003069	.0149617	3257.229	16.17794	4.59e-06

Table 1.19 : covariate pattern of 5 observations

Goodness-of-fit Test by eliminating No.10637 observation:

```
Logistic model for died, goodness-of-fit test

       number of observations =     42014
 number of covariate patterns =     37341
            Pearson chi2(37328) =     32554.86
                    Prob > chi2 =        1.0000
```

```
      number of observations =      42014
            number of groups =       1000
  Hosmer-Lemeshow chi2(998) =        928.91
                Prob > chi2 =         0.9416
```

Change of AGE1 is -.50327473%

Goodness-of-fit Test by eliminating No.35397 observation:

```
Logistic model for died, goodness-of-fit test

       number of observations =     42014
 number of covariate patterns =     37341
            Pearson chi2(37328) =     35802.11
                    Prob > chi2 =        1.0000
```

```
      number of observations =      42014
            number of groups =       1000
  Hosmer-Lemeshow chi2(998) =       1003.38
                Prob > chi2 =         0.4462
```

Change of AGE1 is .2964495%

Goodness-of-fit Test by eliminating No.2754 observation:

```
Logistic model for died, goodness-of-fit test

       number of observations =     42014
 number of covariate patterns =     37341
             Pearson chi2(37328) =  35752.34
                   Prob > chi2 =      1.0000

       number of observations =     42014
             number of groups =      1000
      Hosmer-Lemeshow chi2(998) =    966.02
                   Prob > chi2 =      0.7607
```

Change of AGE1 is -.27576698%

Goodness-of-fit Test by eliminating No.26953 observation:

```
Logistic model for died, goodness-of-fit test

       number of observations =     42014
 number of covariate patterns =     37341
             Pearson chi2(37328) =  35756.52
                   Prob > chi2 =      1.0000

       number of observations =     42014
             number of groups =      1000
      Hosmer-Lemeshow chi2(998) =    978.98
                   Prob > chi2 =      0.6604
```

Change of AGE1 is -.15856601%

Goodness-of-fit Test by eliminating No.26858 observation:

```
Logistic model for died, goodness-of-fit test

       number of observations =     42014
 number of covariate patterns =     37341
             Pearson chi2(37328) =  35206.73
                   Prob > chi2 =      1.0000

       number of observations =     42014
             number of groups =      1000
      Hosmer-Lemeshow chi2(998) =    994.47
                   Prob > chi2 =      0.5256
```

Change of AGE1 is -.24819028%

```
Logistic regression                          Number of obs     =     42,014
                                             LR chi2(12)       =   11054.45
                                             Prob > chi2       =     0.0000
Log likelihood = -5958.6177                  Pseudo R2         =     0.4812
```

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age1	.0014469	.0003788	3.82	0.000	.0007045	.0021893
tmpm	1.171404	.0208038	56.31	0.000	1.13063	1.212179
female	-.2263135	.0773768	-2.92	0.003	-.3779692	-.0746577
shock	.6160511	.0952319	6.47	0.000	.4293999	.8027022
gsw_int1	-.7735087	.104801	-7.38	0.000	-.9789149	-.5681025
gsw_int2	.6433869	.1015226	6.34	0.000	.4444063	.8423675
gsw_int3	-.850991	.0898411	-9.47	0.000	-1.027076	-.6749056
gsw_int4	-1.373177	.2760003	-4.98	0.000	-1.914127	-.832226
age1xtmpm	-.0006805	.0001323	-5.14	0.000	-.0009399	-.0004211
tmpmxshock	-.3494579	.0430575	-8.12	0.000	-.4338491	-.2650667
tmpmxgsw_int4	-.2722176	.1090255	-2.50	0.013	-.4859036	-.0585315
shockxgsw_int2	-.5480296	.1692124	-3.24	0.001	-.8796798	-.2163794
_cons	.5155281	.0924788	5.57	0.000	.334273	.6967832

Table : logistic regression model eliminating observation No. 26858

```
Logistic regression                               Number of obs     =     42,010
                                                  LR chi2(12)       =   11079.91
                                                  Prob > chi2       =     0.0000
Log likelihood = -5935.6718                       Pseudo R2         =     0.4828
```

died	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age1	.0014375	.0003787	3.80	0.000	.0006952	.0021798
tmpm	1.174337	.020868	56.27	0.000	1.133436	1.215237
female	-.2251146	.0774418	-2.91	0.004	-.3768978	-.0733313
shock	.6108352	.0952181	6.42	0.000	.424211	.7974593
gsw_int1	-.7780113	.1049181	-7.42	0.000	-.9836469	-.5723756
gsw_int2	.6435514	.1016001	6.33	0.000	.4444188	.8426839
gsw_int3	-.8515932	.0899027	-9.47	0.000	-1.027799	-.6753871
gsw_int4	-1.268011	.2824793	-4.49	0.000	-1.821661	-.7143623
age1xtmpm	-.0006872	.0001325	-5.19	0.000	-.0009469	-.0004276
tmpmxshock	-.355557	.0430849	-8.25	0.000	-.4400019	-.2711121
tmpmxgsw_int4	-.1600619	.120078	-1.33	0.183	-.3954105	.0752866
shockxgsw_int2	-.5522522	.1690869	-3.27	0.001	-.8836564	-.2208479
_cons	.5195326	.0925562	5.61	0.000	.3381257	.7009395

died	Odds Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age1	1.001439	.0003793	3.80	0.000	1.000695	1.002182
tmpm	3.235996	.0675287	56.27	0.000	3.106312	3.371093
female	.7984247	.0618315	-2.91	0.004	.6859862	.9292929
shock	1.841969	.1753889	6.42	0.000	1.528384	2.219894
gsw_int1	.4593186	.0481908	-7.42	0.000	.3739449	.5641836
gsw_int2	1.903228	.1933682	6.33	0.000	1.559584	2.322592
gsw_int3	.4267345	.0383646	-9.47	0.000	.3577935	.5089594
gsw_int4	.2813906	.079487	-4.49	0.000	.1617569	.4895042
age1xtmpm	.999313	.0001324	-5.19	0.000	.9990536	.9995725
tmpmxshock	.700783	.0301932	-8.25	0.000	.6440352	.762531
tmpmxgsw_int4	.852091	.1023174	-1.33	0.183	.6734036	1.078193
shockxgsw_int2	.5756519	.0973352	-3.27	0.001	.4132691	.8018386
_cons	1.681242	.1556094	5.61	0.000	1.402317	2.015645

Table 2: logistic regression model eliminating 5 observations

Logistic model for died, goodness-of-fit test

```
       number of observations =     42010
 number of covariate patterns =     37337
             Pearson chi2(37324) =     31827.95
                  Prob > chi2 =        1.0000
```

```
         number of observations =     42010
               number of groups =      1000
     Hosmer-Lemeshow chi2(998) =        902.25
                    Prob > chi2 =         0.9861
```

Age1	TMPM	Logit(p=1)	Odds ratio
x	y	$\beta_0+ \beta_1 x+ \beta_2 y+ \beta_9 xy$	1
x-1	y	$\beta_0+ \beta_1 (x-1)+ \beta_2 y+ \beta_9 (x-1)y$	Exp $(\beta_1+ \beta_9 y)$
x	y-1	$\beta_0+ \beta_1 + \beta_2 (y-1)+ \beta_9 (y-1)x$	Exp $(\beta_2+ \beta_9 x)$
x-1	y-1	$\beta_0+ \beta_1(x-1) + \beta_2 (y-1) + \beta_9 (y-1)(x-1)$	Exp $(\beta_1+ \beta_2+\beta_9 (x+y+1))$

TMPM	SHOCK	Logit(p=1)	Odds ratio
x	1	$\beta_0 + \beta_2$ x+ $\beta_4 + \beta_{10}$ x	1
x-1	1	$\beta_0 + \beta_2$ (x-1)+ $\beta_4 + \beta_{10}$ (x-1)	Exp ($\beta_2 + \beta_{10}$)
x	0	$\beta_0 + \beta_2$ x	Exp ($\beta_4 + \beta_{10}$ x)
x-1	0	$\beta_0 + \beta_2$(x-1)	Exp ($\beta_2 + \beta_4 + \beta_{10}$ x)

TMPM	GSW_INT4	Logit(p=1)	Odds ratio
x	1	$\beta_0 + \beta_2$ x+ $\beta_8 + \beta_{11}$ x	1
x-1	1	$\beta_0 + \beta_2$ (x-1) + $\beta_8 + \beta_{11}$ (x-1)	Exp ($\beta_2 + \beta_4$)
x	0	$\beta_0 + \beta_2$ x	Exp ($\beta_4 + \beta_{11}$ x)
x-1	0	$\beta_0 + \beta_2$(x-1)	Exp ($\beta_2 + \beta_8 + \beta_{11}$ x)

SHOCK	GSW_INT4	Logit(p=1)	Odds ratio
1	1	$\beta_0 + \beta_4 + \beta_8 + \beta_{12}$	1
0	1	$\beta_0 + \beta_8$	Exp ($\beta_4 + \beta_{12}$)
1	0	$\beta_0 + \beta_4$	Exp ($\beta_4 + \beta_8$)
0	0	β_0	Exp ($\beta_4 + \beta_8 + \beta_{12}$)

Table 2.1 Estimated Odds Ratio Tables for interactions

Hence, the Estimated Odds Ratio Tables with confident interval is shown below:

Variable	Odds Ratio	95% Confident Interval
age1	1.001439	1.000695, 1.002182
tmpm	3.235996	3.106312, 3.371093
female	.7984247	.6859862, .9292929
shock	1.841969	1.528384, 2.219894
gsw_int1	.4593186	.3739449, .5641836
gsw_int2	1.903228	1.559584, 2.322592
gsw_int3	.4267345	.3577935, .5089594
gsw_int4	.2813906	.1617569, .4895042

age1	tmpm	Odds Ratio	95% Confident Interval
1	1	3.240651	3.111481 3.375182
1	0	1.001439	1.000695 1.002182
0	1	3.235996	3.106312 3.371093
0	0	1	
tmpm	shock	Odds Ratio	95% Confident Interval
1	1	5.960604	4.979907 7.13443
1	0	3.235996	3.106312 3.371093
0	1	1.841969	1.528384 2.219894
0	0	1	
tmpm	gsw_int4	Odds Ratio	95% Confident Interval
1	1	.9105788	.5243269 1.581368
1	0	3.235996	3.106312 3.371093
0	1	.2813906	.1617569 .4895042
0	0	1	
shock	gsw_int2	Odds Ratio	95% Confident Interval
1	1	3.505687	2.645147 4.646186
1	0	1.841969	1.528384 2.219894
0	1	1.903228	1.559584 2.322592
0	0	1	

Table 3: Estimated Odds Ratio with interactions

APENDIX2:

STATA code:

For univariate test:

Logit died

Estimates store A

(we then add single variable in the model, for example

Logit died age

Estimates store A1

Lrtest A A1)

For revised model:

logit died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xtmpm tmpmxshock tmpmxgsw_int4 shockxgsw_int2

predict p

predict dx, dx2

graph twoway scatter dx p, xlabel(0(.2)1) ylabel(0(10)30)

predict dd, dd

graph twoway scatter dd p, xlabel(0(.2)1) ylabel(0 3.5 7)

predict db, db

graph twoway scatter db p, xlabel(0(.2)1) ylabel(0 .15 .3)

graph twoway scatter dx p [weight=db], xlabel(0(.2)1) ylabel(0 15 30) msymbol(oh)

predict h, h

predict n, n

```
list died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xtmpm tmpmxshock tmpmxgsw_int4 shockxgsw_int2 n p db dx dd h if n==2754 | n==26953 | n==26858 | n==10637 | n==35397

logit died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xtmpm tmpmxshock tmpmxgsw_int4 shockxgsw_int2  if n!= 10637

estat gof

estat gof, group(1000) table

logit died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xtmpm tmpmxshock tmpmxgsw_int4 shockxgsw_int2  if n!= 2754

estat gof

estat gof, group(1000) table

logit died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xtmpm tmpmxshock tmpmxgsw_int4 shockxgsw_int2  if n!= 26953

estat gof

estat gof, group(1000) table

logit died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xtmpm tmpmxshock tmpmxgsw_int4 shockxgsw_int2  if n!= 35397

estat gof

estat gof, group(1000) table

di ((.0014469-.0014505)/.0014505*100)

di ((.0014465-.0014505)/.0014505*100)

logit died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xtmpm tmpmxshock tmpmxgsw_int4 shockxgsw_int2 if n!=26858 & n!=2754 & n!=35397 & n!=26953 &n!=10637

logistic died age1 tmpm female shock gsw_int1 gsw_int2 gsw_int3 gsw_int4 age1xtmpm tmpmxshock tmpmxgsw_int4 shockxgsw_int2 if n!=26858 & n!=2754 & n!=35397 & n!=26953 &n!=10637

estat gof

estat gof, group(1000) table
```