**Prediction of Length of Stay (LOS) from Patients Data in Sweden**

**Jing Zhang**

12/24/2017

**Abstract**

This project aims to predict the distribution of Length of Stay (LOS) as a function of demographics, disease variables and comorbidities under the consideration of patients left the hospital against medical advice. Also we wanted to fond the proper model to do regression based on the data which contains the data of all patients hospitalized with depression in Sweden from 2008 to 2013.

**Introduction**

Dataset contains the data of all patients hospitalized with depression in Sweden from 2008 to 2013 (Appendix 1). In order to predict the distribution of Length of Stay (LOS) as a function of demographics, disease variables and comorbidities, we need to consider the patients left the hospital against medical advice (LAMA), which is a factor lower the bound of variables. LAMA is also potentially lead to the utilization of resources and the financial cost. We used leaps function to do variables selection and then use linear regression models and survival models (Proportional Hazard & Weibull models) to analyze data. Besides, we analyze the random of LAMA and find that LAMA distributed randomly given most variables but AGE, FEMALE and LOS.

**Results**

**1. Explore the dataset for interesting patterns and relationships among the variables.**

Firstly, we explore dataset by summarizing the Minimum, 1st Quantile, Median, mean, 3rd Quantile, Maximum and sequences of every variable (Appendix2 and Appendix 3). Also we plot histogram to show their distributions (Appendix 4). We found that most of variables distributed binary. Then the co-relationship of variables is represented by coefficients and patterns (Appendix 5), which indicate the correlation between paired variables. We also tried to generate survival curves (Appendix 7). We found that given AGE, NCHRONIC, CHF, ULCER variables, survival curves show obvious differences.

**2. Modeling and regression analysis.**

We first tried linear regression model with normally distribution assumption.

(Appendix 7). After selecting significant variables, the model should be: LOS=155.592895-0.077178\* YEAR +0.078459\*AGE+0.087436\*RACE+0.751830\*FEMALE +0.152238\*NCHRONIC -0.452399\*ALCOHOL+0.309344\*ANEMDEF+1.033791\*CHF-0.184419\* CHRNLUNG-0.615062\* DRUG +0.846711\*HTN-0.241334\*OBESE-1.186226\*LAMA+1.006183\* WGHTLOSS. The most important factors are: YEAR, AGE, RACE, FEMALE, NCHRONIC (Number of chronic), ALCOHOL (Alcohol abuse), ANEMDEF (Deficiency anemias), CHF (Congestive heart failure), DRUG (drug abuse),HTN (Hypertension),LAMA and WGHTLOSS (Weight loss).

Then we used Leads function to do variable selections (Appendix 8). According to RSS plot, we sound that the least variable numbers should be 8. After forward and backward stepwise selection and LASSO model selection, the variables we selected are: AGE, FEMALE, NCHRONIC, ALCOHOL, CHF, DRUG, HTN and LAMA. Age and patients left the hospital against medical advice are most important variables.

From Cox regression, we got the Cox Proportional-Hazards Model (non-parametric model) (Appendix 9) as: LOS=-0.008591\*AGE0.0214957\*RACE0.0381048\*NCHRONIC+ 0.0924597\*ALCOHOL-0.0736864\*ANEMDEF-0.1784027\*CHF+0.3438546\*DRUG-0.220527 \*HTN-0.1251817\*WGHTLOSS, the most significant variables on LOS are: AGE, RACE, NCHRONIC (Number of chronic), ALCOHOL (Alcohol abuse), ANEMDEF (Deficiency anemias), CHF (Congestive heart failure), DRUG (drug abuse), HTN(Hypertension), and WGHTLOSS.

Also we use Weibull regression to get Weibull models (parametric model), the results as Appendix 10 show. We choose AGE, RACE, NCHRONIC (Number of chronic), ALCOHOL (Alcohol abuse), CHF (Congestive heart failure), DRUG (drug abuse), HTN(Hypertension), and WGHTLOSS 8 variables and got the model: -0.01098418\*AGE-0.02370269\*RACE-0.03771904\*NCHRONIC +0.09939747ALCOHOL-0.1891248\*CHF+0.2038557\*DRUG-0.2208518\*HTN-0.1539416\* WGHTLOSS.

After variable selection, we comparing the Proportional Hazard and Weibull models. We found that their coefficients of factors are similar, but in Weibull model, ANEMDEF is not a significant variable. Because standard is needed in the choice of variables when we conduct Weibull model, thus it is not as good as Cox Proportional-Hazards Model. We can say for this case, Cox Proportional-Hazards Model is the best choice.

**3. Checking the random of LAMA variable**

In order to check that those patients who left against medical advice did so at random, we do linear regression for LAMA as a function of other variables (Appendix 11). As the results show, LAMA does not distribute randomly given AGE, FEMALE and LOS. The model we got is: LAMA=-0.0067098\*AGE-0.1440015\*FEMALE-0.0093692\*LOS. We can see that the smaller age attributes to the increased LAMA, and female patients are related to less LAMA than male patients. Also the Longer Length of days are related to the decreased LAMA.

According to the result above, we can control the patients left the hospital against medical advice by control related factors. Also we can focus on these factors to prevent the happening of pre-leavings in hospital. For example, if we want to decreased the LAMA, we can focus on male patients rather than female, we can focus on young patients rather than old ones, and we can pay more attention on new patients rather than patients who already stayed for a long time.

**Appendix**

Appendix 1: variables illustration

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Demographics: | "YEAR" | year | "AGE" | age | "RACE" | race | "FEMALE" | Female |
| Disease variables  and comorbidities: | "NCHRONIC" | Number of chronic | "DIED" | died | "ALCOHOL" | Alcohol abuse | "ANEMDEF" | Deficiency anemias |
| "ARTH" | Rheumatoid arthritis/collagen vascular disease | "BLDLOSS" | Blood loss anemia | "CHF" | Congestive heart failure | "CHRNLUNG" | Chronic pulmonary disease |
| "COAG" | Coagulopathy | "DEPRESS" | Depression | "DRUG" | Drug abuse | "HTN" | Hypertension |
| "LIVER" | Liver disease | "OBESE" | Obesity | "RENLFAIL" | Renal failure | "ULCER" | Peptic ulcer disease excluding bleeding |
|  | "VALVE" | Valvar disease | "WGHTLOSS" | Weight loss |  |  |  |  |
| Others: | "LOS" | Length of stay | "LAMA" | patients left the hospital against medical advice |  |  |  |  |

Appendix 2: Exploration of data set by summarizing

YEAR AGE RACE FEMALE

Min. :2008 Min. :19.00 Min. :1.000 Min. :0.0000

1st Qu.:2011 1st Qu.:32.00 1st Qu.:1.000 1st Qu.:0.0000

Median :2012 Median :45.00 Median :1.000 Median :1.0000

Mean :2012 Mean :45.53 Mean :1.477 Mean :0.5722

3rd Qu.:2013 3rd Qu.:55.00 3rd Qu.:1.000 3rd Qu.:1.0000

Max. :2013 Max. :98.00 Max. :6.000 Max. :1.0000

LOS NCHRONIC DIED

Min. : 0.000 Min. : 1.000 Min. :0.0000000

1st Qu.: 3.000 1st Qu.: 3.000 1st Qu.:0.0000000

Median : 3.000 Median : 4.000 Median :0.0000000

Mean : 4.936 Mean : 4.639 Mean :0.0002897

3rd Qu.: 5.000 3rd Qu.: 6.000 3rd Qu.:0.0000000

Max. :155.000 Max. :19.000 Max. :1.0000000

ALCOHOL ANEMDEF ARTH

Min. :0.0000 Min. :0.00000 Min. :0.00000

1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.00000

Median :0.0000 Median :0.00000 Median :0.00000

Mean :0.1814 Mean :0.08813 Mean :0.01906

3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:0.00000

Max. :1.0000 Max. :1.00000 Max. :1.00000

BLDLOSS CHF CHRNLUNG

Min. :0.000000 Min. :0.00000 Min. :0.0000

1st Qu.:0.000000 1st Qu.:0.00000 1st Qu.:0.0000

Median :0.000000 Median :0.00000 Median :0.0000

Mean :0.008179 Mean :0.03909 Mean :0.1654

3rd Qu.:0.000000 3rd Qu.:0.00000 3rd Qu.:0.0000

Max. :1.000000 Max. :1.00000 Max. :1.0000

COAG DEPRESS DRUG

Min. :0.0000 Min. :0.00000 Min. :0.0000

1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000

Median :0.0000 Median :0.00000 Median :0.0000

Mean :0.0193 Mean :0.03865 Mean :0.2145

3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:0.0000

Max. :1.0000 Max. :1.00000 Max. :1.0000

HTN LIVER OBESE

Min. :0.0000 Min. :0.00000 Min. :0.00000

1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.00000

Median :0.0000 Median :0.00000 Median :0.00000

Mean :0.3761 Mean :0.02414 Mean :0.09885

3rd Qu.:1.0000 3rd Qu.:0.00000 3rd Qu.:0.00000

Max. :1.0000 Max. :1.00000 Max. :1.00000

RENLFAIL LAMA ULCER

Min. :0.00000 Min. :0.000 Min. :0.0000000

1st Qu.:0.00000 1st Qu.:0.000 1st Qu.:0.0000000

Median :0.00000 Median :0.000 Median :0.0000000

Mean :0.04948 Mean :0.329 Mean :0.0002452

3rd Qu.:0.00000 3rd Qu.:1.000 3rd Qu.:0.0000000

Max. :1.00000 Max. :1.000 Max. :1.0000000

VALVE WGHTLOSS

Min. :0.0000 Min. :0.0000

1st Qu.:0.0000 1st Qu.:0.0000

Median :0.0000 Median :0.0000

Mean :0.0177 Mean :0.0271

3rd Qu.:0.0000 3rd Qu.:0.0000

Max. :1.0000 Max. :1.0000

Appendix 3: Frequency of variables

> table(dataFP$RACE) # frequency of race:

1. 2 3 4 5 6

33665 5678 3569 473 264 1219

> table(dataFP$NCHRONIC) # frequency of NCHRONIC:

1 2 3 4 5 6 7 8 9 10 11 12 13

2604 5792 7824 8045 6586 5157 3405 2349 1287 784 493 280 144

14 15 16 17 18 19

72 28 6 6 4 2

> table(dataFP$DIED)# frequency of died:

0 1

44855 13

> table(dataFP$AGE) # frequency of age:

19 20 21 22 23 24 25 26 27 28 29 30 31

837 793 778 851 759 755 716 666 803 699 776 756 1323

32 33 34 35 36 37 38 39 40 41 42 43 44

874 743 782 723 732 896 793 794 880 900 1148 977 949

45 46 47 48 49 50 51 52 53 54 55 56 57

1040 1007 1179 1111 1140 1100 1055 1238 945 1250 893 829 867

58 59 60 61 62 63 64 65 66 67 68 69 70

737 637 581 547 567 416 390 395 391 380 341 328 278

71 72 73 74 75 76 77 78 79 80 81 82 83

235 342 229 226 230 188 207 168 187 137 184 188 147

84 85 86 87 88 89 90 91 92 93 94 95 97

140 136 118 134 71 65 143 16 13 5 8 1 3

98

2

> table(dataFP$LOS) # frequency of LOS

0 1 2 3 4 5 6 7 8 9 10

428 462 6681 21312 1771 3614 3032 355 2009 129 1603

11 12 13 14 15 16 17 18 19 20 21

561 132 594 414 19 325 55 320 105 7 85

22 23 24 25 26 27 28 29 30 31 32

249 8 63 10 39 61 1 110 29 7 25

34 35 36 37 38 40 41 42 43 45 46

38 46 2 15 23 6 3 16 16 5 3

48 50 51 52 53 54 55 56 58 59 60

12 7 2 1 1 12 1 4 2 4 1

61 62 64 65 66 67 69 70 72 84 86

1 2 2 1 1 1 2 3 1 1 2

93 96 99 102 106 109 110 113 114 125 136

1 1 1 1 2 1 1 1 1 1 1

141 150 155

1 2 1

> table(dataFP$ALCOHOL) # frequency of ALCOHOL

0 1

36731 8137

> table(dataFP$ANEMDEF) # frequency of ANEMDEF

0 1

40914 3954

> table(dataFP$ARTH) # frequency of ARTH

0 1

44013 855

> table(dataFP$BLDLOSS) # frequency of BLDLOSS

0 1

44501 367

> table(dataFP$CHF) # frequency of CHF

0 1

43114 1754

> table(dataFP$CHRNLUNG) # frequency of CHRNLUNG

0 1

37445 7423

> table(dataFP$COAG) # frequency of COAG

0 1

44002 866

> table(dataFP$DEPRESS) # frequency of DEPRESS

0 1

43134 1734

> table(dataFP$DRUG) # frequency of DRUG

0 1

35242 9626

> table(dataFP$HTN) # frequency of HTN

0 1

27991 16877

> table(dataFP$LIVER) # frequency of LIVER

0 1

43785 1083

> table(dataFP$OBESE) # frequency of OBESE

0 1

40433 4435

> table(dataFP$RENLFAIL) # frequency of RENLFAIL

0 1

42648 2220

> table(dataFP$LAMA) # frequency of LAMA

0 1

30106 14762

> table(dataFP$ULCER) # frequency of ULCER

0 1

44857 11

> table(dataFP$VALVE) # frequency of VALVE

0 1

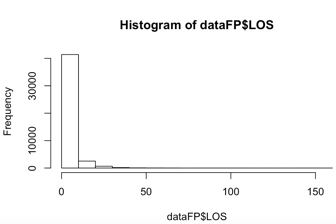
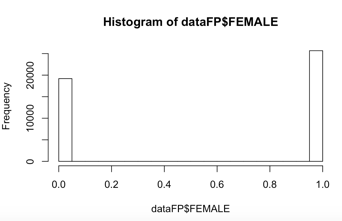
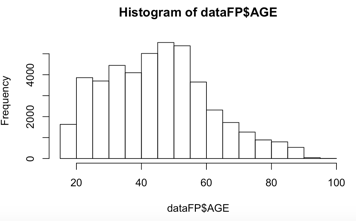
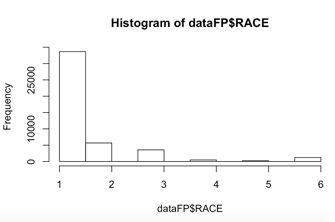
44074 794

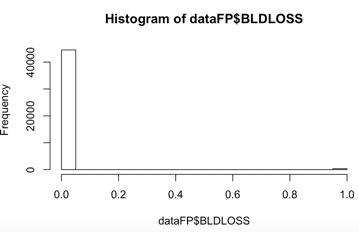
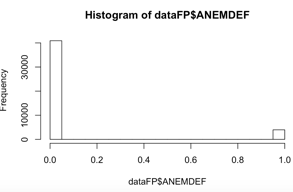
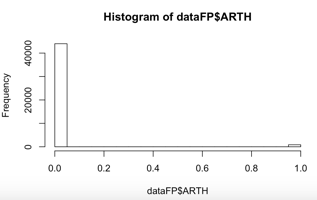
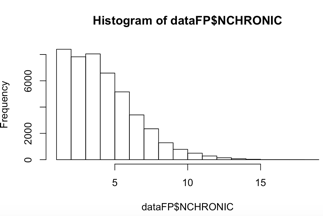
> table(dataFP$WGHTLOSS) # frequency of WGHTLOSS

0 1

43652 1216

Appendix 4: Part of distributions of variables





Appendix 5: Correlationship among variables

YEAR AGE RACE FEMALE

YEAR 1.000000000 -0.010330990 0.0167101170 -0.0116313781

AGE -0.010330990 1.000000000 -0.0739755898 0.0244072803

RACE 0.016710117 -0.073975590 1.0000000000 -0.0062412003

FEMALE -0.011631378 0.024407280 -0.0062412003 1.0000000000

LOS -0.006674896 0.324136367 -0.0101085002 0.0978578749

NCHRONIC 0.071537056 0.367905773 -0.0945432624 -0.0003996073

DIED -0.013770931 0.026797613 -0.0065305810 0.0014870517

ALCOHOL -0.114471613 -0.044524721 -0.0139578152 -0.1385065986

ANEMDEF 0.080481410 0.061134860 0.0031203678 0.0309316562

ARTH 0.010780150 0.031457078 0.0013251647 0.0388116577

BLDLOSS 0.047482559 0.006655961 0.0031042729 -0.0014930801

CHF 0.061515269 0.073557748 -0.0048592130 0.0005621479

CHRNLUNG 0.017410643 0.061484256 -0.0110013288 0.0311325104

COAG 0.058308634 0.019880206 -0.0029184812 -0.0008175159

DEPRESS 0.091318544 -0.006397388 -0.0032733484 0.0032396107

DRUG -0.103253665 -0.164909294 0.0007210443 -0.0965325877

HTN 0.047533798 0.289552506 -0.0119347895 -0.0304517099

LIVER 0.007674187 0.018624057 -0.0022582588 -0.0254348320

OBESE 0.027650516 -0.018925416 -0.0037201952 0.0569682159

RENLFAIL 0.092333892 0.066908209 -0.0025788080 -0.0085603562

LAMA 0.013172932 -0.261610597 0.0169710738 -0.1674285023

ULCER -0.001396358 0.008462504 -0.0016982978 -0.0094774717

VALVE 0.023609329 0.036811541 -0.0042916443 0.0115138265

WGHTLOSS 0.028876719 0.053216779 -0.0033717639 0.0047839483

LOS NCHRONIC DIED ALCOHOL

YEAR -0.006674896 0.0715370557 -0.0137709313 -0.114471613

AGE 0.324136367 0.3679057735 0.0267976131 -0.044524721

RACE -0.010108500 -0.0945432624 -0.0065305810 -0.013957815

FEMALE 0.097857875 -0.0003996073 0.0014870517 -0.138506599

LOS 1.000000000 0.1791447597 0.0188823152 -0.063593577

NCHRONIC 0.179144760 1.0000000000 0.0167156105 0.094632851

DIED 0.018882315 0.0167156105 1.0000000000 -0.004614072

ALCOHOL -0.063593577 0.0946328514 -0.0046140718 1.000000000

ANEMDEF 0.058274982 0.0871613401 0.0039468692 -0.051639525

ARTH 0.014063727 0.0731871043 0.0072054603 -0.023716520

BLDLOSS 0.003597425 0.0076572216 -0.0015460175 -0.032467695

CHF 0.081596896 0.1218546257 0.0168363720 -0.056136078

CHRNLUNG 0.033200434 0.2276897133 0.0029930792 -0.003454631

COAG 0.011295989 0.0370856961 -0.0023883006 0.010068134

DEPRESS 0.004175320 0.0136374936 -0.0034133508 -0.053856882

DRUG -0.110523021 0.0916296916 -0.0025170827 0.256633160

HTN 0.179621128 0.3104337015 0.0084076402 -0.027310239

LIVER 0.009100580 0.0764470705 0.0058551978 0.097837406

OBESE 0.010869541 0.1555749695 -0.0012504800 -0.046185760

RENLFAIL 0.056908512 0.1039070787 0.0081930042 -0.075644655

LAMA -0.192950297 -0.0904042745 -0.0119209898 0.034944353

ULCER 0.007570105 0.0118232159 -0.0002665922 0.003713536

VALVE 0.035068016 0.0658394131 0.0076474946 -0.036411605

WGHTLOSS 0.055045518 0.0425120780 0.0132881031 -0.017998663

ANEMDEF ARTH BLDLOSS CHF

YEAR 0.0804814102 0.010780150 0.047482559 0.0615152695

AGE 0.0611348596 0.031457078 0.006655961 0.0735577475

RACE 0.0031203678 0.001325165 0.003104273 -0.0048592130

FEMALE 0.0309316562 0.038811658 -0.001493080 0.0005621479

LOS 0.0582749823 0.014063727 0.003597425 0.0815968960

NCHRONIC 0.0871613401 0.073187104 0.007657222 0.1218546257

DIED 0.0039468692 0.007205460 -0.001546017 0.0168363720

ALCOHOL -0.0516395247 -0.023716520 -0.032467695 -0.0561360783

ANEMDEF 1.0000000000 0.036028868 0.105322423 0.1259273259

ARTH 0.0360288676 1.000000000 -0.001798123 0.0290812730

BLDLOSS 0.1053224232 -0.001798123 1.000000000 0.0161543986

CHF 0.1259273259 0.029081273 0.016154399 1.0000000000

CHRNLUNG 0.0634446931 0.034021036 -0.007802627 0.1070217719

COAG 0.1089684086 0.014807931 0.021432496 0.0586300329

DEPRESS 0.0637094549 0.027026871 0.011318492 0.0597868144

DRUG -0.0711124764 -0.016453056 -0.032391694 -0.0664759353

HTN 0.1090205565 0.042197020 -0.019434518 0.1422506821

LIVER 0.0284623799 -0.003863613 -0.007833210 0.0094878600

OBESE 0.0303380517 0.009551378 0.008061866 0.0487884907

RENLFAIL 0.2307070746 0.024577106 0.023780440 0.2074446178

LAMA -0.0170519208 -0.013636140 -0.003026378 -0.0218025431

ULCER 0.0001537875 -0.002182598 0.014383370 0.0041866172

VALVE 0.0447412245 0.031985393 0.010332590 0.1656817967

WGHTLOSS 0.1122539249 0.015889899 0.003129538 0.0506091075

CHRNLUNG COAG DEPRESS DRUG

YEAR 0.017410643 0.0583086343 0.0913185436 -0.1032536653

AGE 0.061484256 0.0198802064 -0.0063973875 -0.1649092942

RACE -0.011001329 -0.0029184812 -0.0032733484 0.0007210443

FEMALE 0.031132510 -0.0008175159 0.0032396107 -0.0965325877

LOS 0.033200434 0.0112959889 0.0041753202 -0.1105230212

NCHRONIC 0.227689713 0.0370856961 0.0136374936 0.0916296916

DIED 0.002993079 -0.0023883006 -0.0034133508 -0.0025170827

ALCOHOL -0.003454631 0.0100681340 -0.0538568819 0.2566331604

ANEMDEF 0.063444693 0.1089684086 0.0637094549 -0.0711124764

ARTH 0.034021036 0.0148079312 0.0270268707 -0.0164530558

BLDLOSS -0.007802627 0.0214324957 0.0113184924 -0.0323916937

CHF 0.107021772 0.0586300329 0.0597868144 -0.0664759353

CHRNLUNG 1.000000000 0.0155763734 0.0585414500 0.0136569064

COAG 0.015576373 1.0000000000 0.0281816287 -0.0318828238

DEPRESS 0.058541450 0.0281816287 1.0000000000 -0.0588738738

DRUG 0.013656906 -0.0318828238 -0.0588738738 1.0000000000

HTN 0.117859321 0.0385432427 0.0701193165 -0.1132899560

LIVER 0.028853047 0.1267642039 0.0106573100 0.0402055879

OBESE 0.070796327 -0.0003257815 0.0486613001 -0.0406556213

RENLFAIL 0.059388096 0.0755594319 0.0592922845 -0.0911995555

LAMA -0.013689856 -0.0010075379 0.0006146128 0.0589271630

ULCER 0.012183899 -0.0021968680 0.0042459300 -0.0012482681

VALVE 0.031226668 0.0290890329 0.0213238962 -0.0417334977

WGHTLOSS 0.023575707 0.0633798104 0.0227908033 -0.0320588500

HTN LIVER OBESE RENLFAIL

YEAR 0.0475337982 0.007674187 0.0276505157 0.092333892

AGE 0.2895525063 0.018624057 -0.0189254155 0.066908209

RACE -0.0119347895 -0.002258259 -0.0037201952 -0.002578808

FEMALE -0.0304517099 -0.025434832 0.0569682159 -0.008560356

LOS 0.1796211280 0.009100580 0.0108695413 0.056908512

NCHRONIC 0.3104337015 0.076447070 0.1555749695 0.103907079

DIED 0.0084076402 0.005855198 -0.0012504800 0.008193004

ALCOHOL -0.0273102390 0.097837406 -0.0461857596 -0.075644655

ANEMDEF 0.1090205565 0.028462380 0.0303380517 0.230707075

ARTH 0.0421970199 -0.003863613 0.0095513785 0.024577106

BLDLOSS -0.0194345184 -0.007833210 0.0080618661 0.023780440

CHF 0.1422506821 0.009487860 0.0487884907 0.207444618

CHRNLUNG 0.1178593207 0.028853047 0.0707963269 0.059388096

COAG 0.0385432427 0.126764204 -0.0003257815 0.075559432

DEPRESS 0.0701193165 0.010657310 0.0486613001 0.059292285

DRUG -0.1132899560 0.040205588 -0.0406556213 -0.091199556

HTN 1.0000000000 0.030167459 0.1297675144 0.206629130

LIVER 0.0301674592 1.000000000 0.0135997777 0.019697002

OBESE 0.1297675144 0.013599778 1.0000000000 0.046680685

RENLFAIL 0.2066291303 0.019697002 0.0466806846 1.000000000

LAMA -0.0683196657 -0.003188806 -0.0100379813 -0.015180238

ULCER -0.0004044536 0.016088650 -0.0004164159 0.002991669

VALVE 0.0594417469 0.006426699 0.0036910186 0.062136176

WGHTLOSS 0.0270887066 0.017572267 -0.0194055379 0.058122781

LAMA ULCER VALVE WGHTLOSS

YEAR 0.0131729320 -0.0013963581 0.0236093294 0.028876719

AGE -0.2616105967 0.0084625043 0.0368115409 0.053216779

RACE 0.0169710738 -0.0016982978 -0.0042916443 -0.003371764

FEMALE -0.1674285023 -0.0094774717 0.0115138265 0.004783948

LOS -0.1929502968 0.0075701046 0.0350680164 0.055045518

NCHRONIC -0.0904042745 0.0118232159 0.0658394131 0.042512078

DIED -0.0119209898 -0.0002665922 0.0076474946 0.013288103

ALCOHOL 0.0349443526 0.0037135362 -0.0364116054 -0.017998663

ANEMDEF -0.0170519208 0.0001537875 0.0447412245 0.112253925

ARTH -0.0136361404 -0.0021825982 0.0319853930 0.015889899

BLDLOSS -0.0030263778 0.0143833697 0.0103325901 0.003129538

CHF -0.0218025431 0.0041866172 0.1656817967 0.050609107

CHRNLUNG -0.0136898555 0.0121838994 0.0312266675 0.023575707

COAG -0.0010075379 -0.0021968680 0.0290890329 0.063379810

DEPRESS 0.0006146128 0.0042459300 0.0213238962 0.022790803

DRUG 0.0589271630 -0.0012482681 -0.0417334977 -0.032058850

HTN -0.0683196657 -0.0004044536 0.0594417469 0.027088707

LIVER -0.0031888055 0.0160886502 0.0064266986 0.017572267

OBESE -0.0100379813 -0.0004164159 0.0036910186 -0.019405538

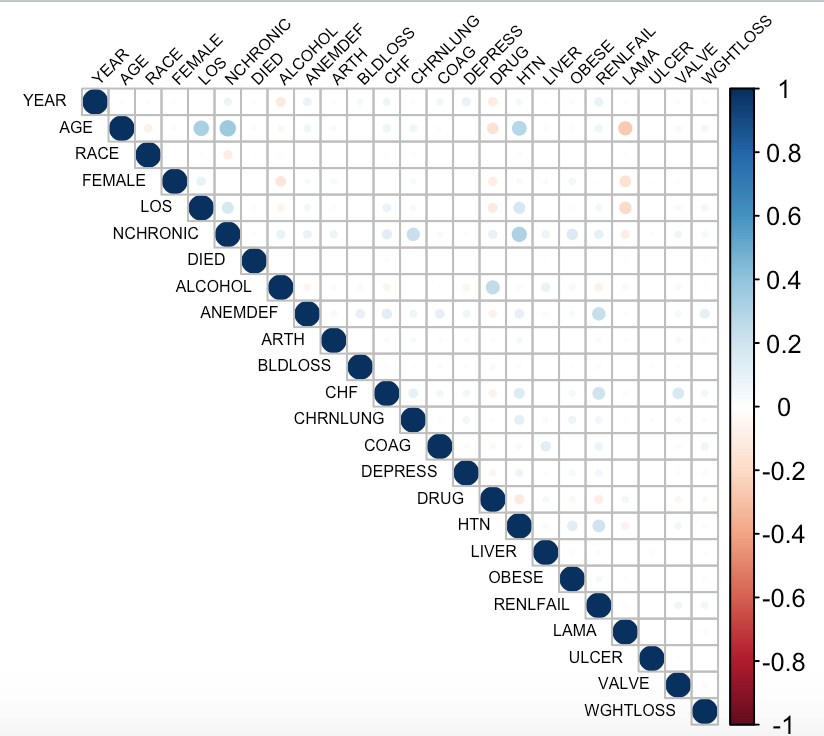
RENLFAIL -0.0151802385 0.0029916691 0.0621361760 0.058122781

LAMA 1.0000000000 -0.0018758167 -0.0008035936 -0.020178709

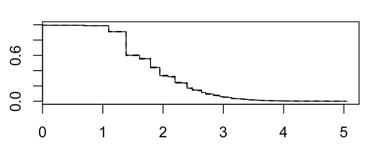
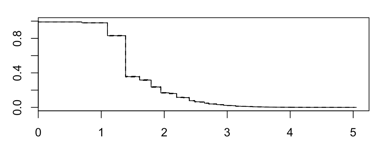
ULCER -0.0018758167 1.0000000000 0.0086956710 0.006153468

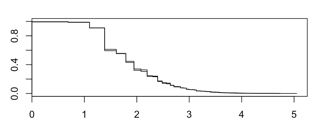
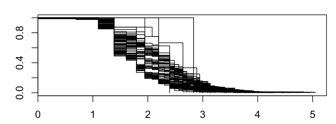
VALVE -0.0008035936 0.0086956710 1.0000000000 0.022362770

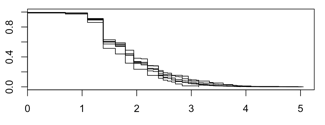
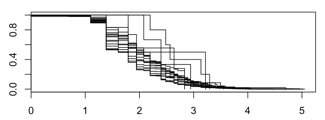
WGHTLOSS -0.0201787093 0.0061534677 0.0223627696 1.000000000

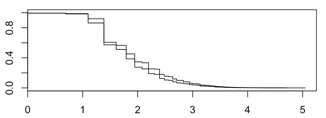
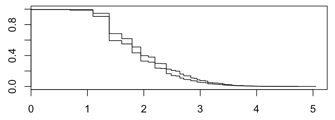


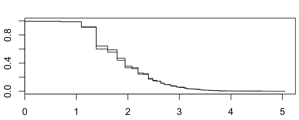
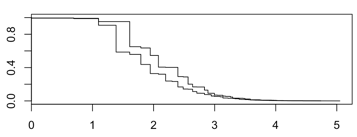
Appendix 6: Kaplan-Meier Survival Curves

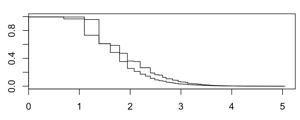
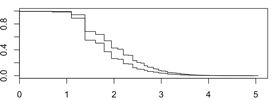
Basic:(LOS~DIED-LAMA)(LOS~DIED)

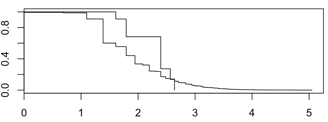
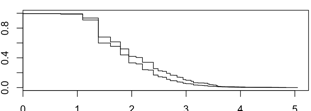
Given Variables(PARTLY):(female)(age)

(race)(nchronic)

(ALCOHOL)(anemdef)

(ARTH)(CHF)

(DRUG)(HTN)

(ULCER)(WGHTLOSS)

Appendix 7: Linear regression results (ANOVA)

Analysis of Variance Table

Response: LOS

Df Sum Sq Mean Sq F value Pr(>F)

YEAR 1 58 58 2.3268 0.127167

AGE 1 137309 137309 5485.2332 < 2.2e-16 \*\*\*

RACE 1 255 255 10.1817 0.001419 \*\*

FEMALE 1 10590 10590 423.0387 < 2.2e-16 \*\*\*

NCHRONIC 1 5864 5864 234.2470 < 2.2e-16 \*\*\*

DIED 1 121 121 4.8210 0.028120 \*

ALCOHOL 1 2824 2824 112.8174 < 2.2e-16 \*\*\*

ANEMDEF 1 1159 1159 46.2871 1.034e-11 \*\*\*

ARTH 1 48 48 1.9179 0.166099

BLDLOSS 1 10 10 0.4108 0.521576

CHF 1 2796 2796 111.6891 < 2.2e-16 \*\*\*

CHRNLUNG 1 133 133 5.3135 0.021165 \*

COAG 1 1 1 0.0408 0.839847

DEPRESS 1 0 0 0.0095 0.922162

DRUG 1 3317 3317 132.4921 < 2.2e-16 \*\*\*

HTN 1 6091 6091 243.3268 < 2.2e-16 \*\*\*

LIVER 1 22 22 0.8703 0.350891

OBESE 1 229 229 9.1570 0.002479 \*\*

RENLFAIL 1 6 6 0.2215 0.637876

LAMA 1 12653 12653 505.4427 < 2.2e-16 \*\*\*

ULCER 1 36 36 1.4329 0.231298

VALVE 1 71 71 2.8280 0.092642 .

WGHTLOSS 1 1173 1173 46.8438 7.786e-12 \*\*\*

Residuals 44844 1122560 25

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> summary(FP.lm)

Call:

lm(formula = log(1 + LOS) ~ ., data = dataFP)

Residuals:

Min 1Q Median 3Q Max

-2.2118 -0.2747 -0.0702 0.1930 3.2710

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 16.4292993 2.8936829 5.678 1.37e-08 \*\*\*

YEAR -0.0076305 0.0014385 -5.304 1.14e-07 \*\*\*

AGE 0.0090285 0.0001588 56.845 < 2e-16 \*\*\*

RACE 0.0072673 0.0021319 3.409 0.000653 \*\*\*

FEMALE 0.0872916 0.0046009 18.973 < 2e-16 \*\*\*

NCHRONIC 0.0208057 0.0010875 19.131 < 2e-16 \*\*\*

DIED 0.2378968 0.1298201 1.833 0.066882 .

ALCOHOL -0.0434921 0.0060755 -7.159 8.27e-13 \*\*\*

ANEMDEF 0.0378481 0.0081930 4.620 3.86e-06 \*\*\*

ARTH -0.0436933 0.0162320 -2.692 0.007109 \*\*

BLDLOSS -0.0025240 0.0247055 -0.102 0.918628

CHF 0.1636273 0.0119708 13.669 < 2e-16 \*\*\*

CHRNLUNG -0.0190591 0.0061510 -3.099 0.001946 \*\*

COAG -0.0166663 0.0163484 -1.019 0.307999

DEPRESS -0.0074684 0.0115992 -0.644 0.519658

DRUG -0.1501076 0.0057962 -25.897 < 2e-16 \*\*\*

HTN 0.1019046 0.0050590 20.143 < 2e-16 \*\*\*

LIVER 0.0235293 0.0146132 1.610 0.107375

OBESE -0.0266072 0.0076060 -3.498 0.000469 \*\*\*

RENLFAIL -0.0011999 0.0108630 -0.110 0.912045

LAMA -0.1622498 0.0049388 -32.852 < 2e-16 \*\*\*

ULCER 0.3569856 0.1410886 2.530 0.011402 \*

VALVE 0.0124705 0.0170332 0.732 0.464096

WGHTLOSS 0.0742028 0.0137409 5.400 6.69e-08 \*\*\*

---

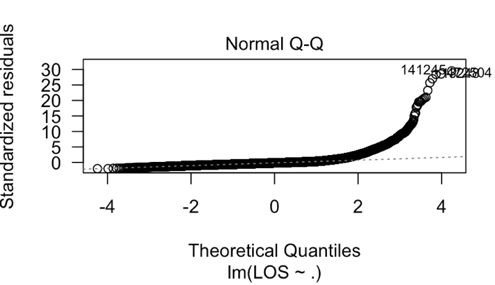
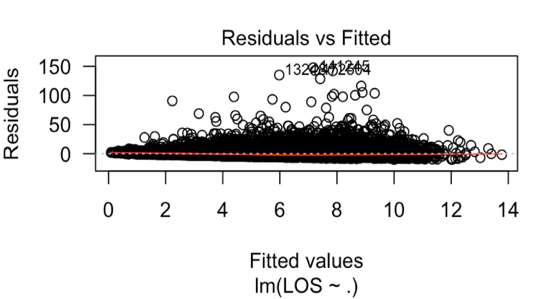
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

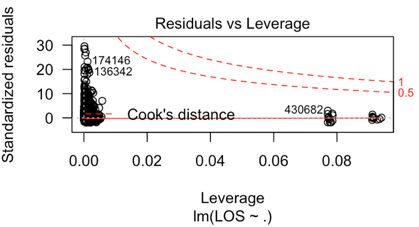
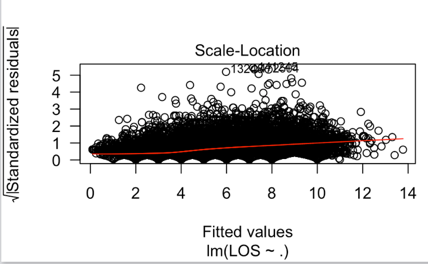
Residual standard error: 0.4676 on 44844 degrees of freedom

Multiple R-squared: 0.2313, Adjusted R-squared: 0.2309

F-statistic: 586.6 on 23 and 44844 DF, p-value: < 2.2e-16

>plot(linearMod2, las = 1)





Appendix 8: Variable selection

YEAR AGE RACE FEMALE NCHRONIC DIED ALCOHOL ANEMDEF ARTH BLDLOSS CHF CHRNLUNG

1 ( 1 ) " " "\*" " " " " " " " " " " " " " " " " " " " "

2 ( 1 ) " " "\*" " " " " " " " " " " " " " " " " " " " "

3 ( 1 ) " " "\*" " " " " " " " " " " " " " " " " " " " "

4 ( 1 ) " " "\*" " " "\*" " " " " " " " " " " " " " " " "

5 ( 1 ) " " "\*" " " "\*" " " " " " " " " " " " " "\*" " "

6 ( 1 ) " " "\*" " " "\*" "\*" " " " " " " " " " " " " " "

7 ( 1 ) " " "\*" " " "\*" "\*" " " " " " " " " " " "\*" " "

8 ( 1 ) " " "\*" " " "\*" "\*" " " " " " " " " " " "\*" " "

COAG DEPRESS DRUG HTN LIVER OBESE RENLFAIL LAMA ULCER VALVE WGHTLOSS

1 ( 1 ) " " " " " " " " " " " " " " " " " " " " " "

2 ( 1 ) " " " " " " " " " " " " " " "\*" " " " " " "

3 ( 1 ) " " " " " " "\*" " " " " " " "\*" " " " " " "

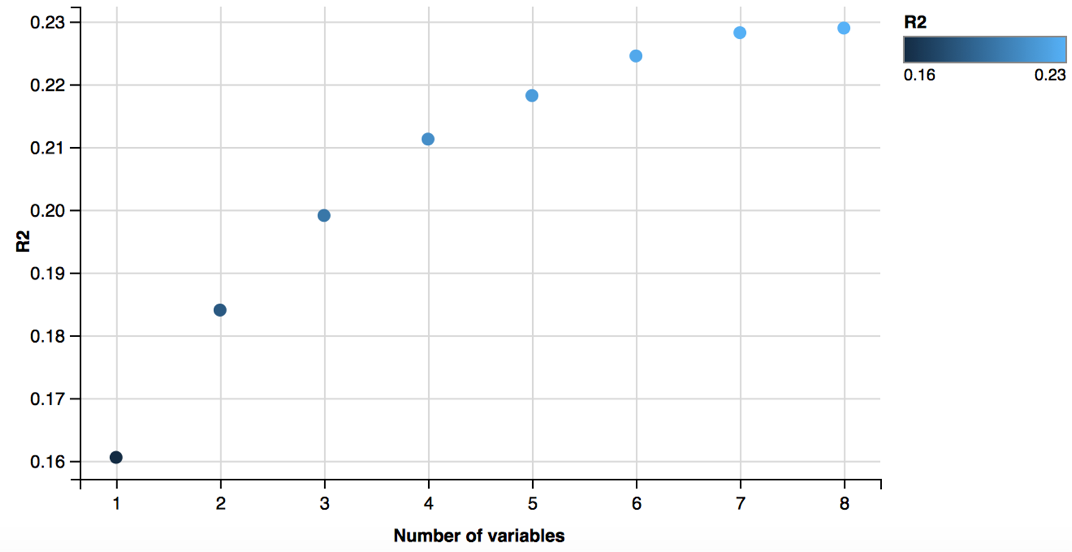
4 ( 1 ) " " " " " " "\*" " " " " " " "\*" " " " " " "

5 ( 1 ) " " " " " " "\*" " " " " " " "\*" " " " " " "

6 ( 1 ) " " " " "\*" "\*" " " " " " " "\*" " " " " " "

7 ( 1 ) " " " " "\*" "\*" " " " " " " "\*" " " " " " "

8 ( 1 ) " " " " "\*" "\*" " " " " " " "\*" " " " " "\*"



> coef(sel.full ,8)

(Intercept) AGE FEMALE NCHRONIC ALCOHOL

1.090889405 0.009179934 0.086604981 0.019123841 -0.039276822

CHF DRUG HTN LAMA

0.166451635 -0.147758214 0.100490867 -0.162660157

> coef(sel.f ,8)

(Intercept) AGE FEMALE NCHRONIC ALCOHOL

1.090889405 0.009179934 0.086604981 0.019123841 -0.039276822

CHF DRUG HTN LAMA

0.166451635 -0.147758214 0.100490867 -0.162660157

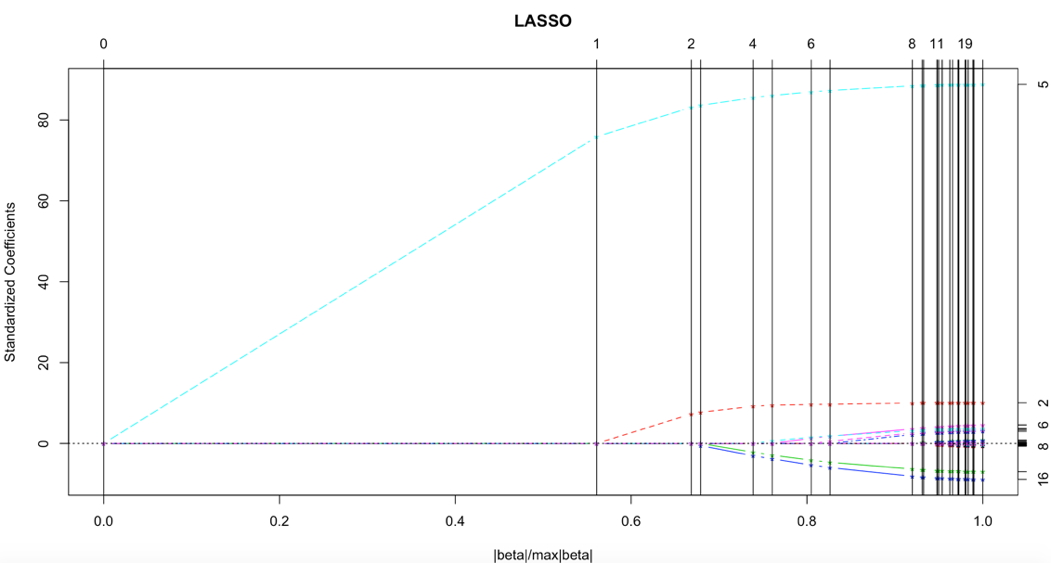
> coef(sel.b ,8)

(Intercept) AGE FEMALE NCHRONIC ALCOHOL

1.090889405 0.009179934 0.086604981 0.019123841 -0.039276822

CHF DRUG HTN LAMA

0.166451635 -0.147758214 0.100490867 -0.162660157



> predict.lars(object=lasso, s=0.85, mode='fraction', type='coefficients')

$s

[1] 0.85

$fraction

[1] 0.85

$mode

[1] "fraction"

$coefficients

YEAR AGE RACE FEMALE LOS

0.000000000 0.002851487 0.000000000 0.005300416 0.076598647

NCHRONIC DIED ALCOHOL ANEMDEF ARTH

0.004383649 0.000000000 0.000000000 0.000000000 0.000000000

BLDLOSS CHF CHRNLUNG COAG DEPRESS

0.000000000 0.025767969 0.000000000 0.000000000 0.000000000

DRUG HTN LIVER OBESE RENLFAIL

-0.074957561 0.021159179 0.000000000 0.000000000 0.000000000

LAMA ULCER VALVE WGHTLOSS

-0.051163346 0.000000000 0.000000000 0.000000000

Appendix 9: Cox Proportional-Hazards Model

Call:

coxph(formula = Surv(log(1 + LOS), 1 - LAMA) ~ ., data = dataFP)

n= 44868, number of events= 30106

coef exp(coef) se(coef) z Pr(>|z|)

YEAR 0.0115032 1.0115696 0.0037145 3.097 0.001956 \*\*

AGE -0.0085919 0.9914449 0.0003997 -21.498 < 2e-16 \*\*\*

RACE -0.0214957 0.9787337 0.0055963 -3.841 0.000123 \*\*\*

FEMALE 0.0284938 1.0289036 0.0122051 2.335 0.019565 \*

NCHRONIC -0.0381048 0.9626121 0.0028000 -13.609 < 2e-16 \*\*\*

DIED -0.1500867 0.8606334 0.2776308 -0.541 0.588785

ALCOHOL 0.0924597 1.0968689 0.0162866 5.677 1.37e-08 \*\*\*

ANEMDEF -0.0736864 0.9289630 0.0209487 -3.517 0.000436 \*\*\*

ARTH 0.0521530 1.0535369 0.0410142 1.272 0.203521

BLDLOSS 0.0057113 1.0057276 0.0636773 0.090 0.928533

CHF -0.1784027 0.8366054 0.0300760 -5.932 3.00e-09 \*\*\*

CHRNLUNG 0.0180407 1.0182044 0.0159483 1.131 0.257973

COAG -0.0141853 0.9859149 0.0425256 -0.334 0.738704

DEPRESS 0.0155131 1.0156340 0.0302871 0.512 0.608510

DRUG 0.3438546 1.4103735 0.0155824 22.067 < 2e-16 \*\*\*

HTN -0.2205272 0.8020958 0.0131303 -16.795 < 2e-16 \*\*\*

LIVER -0.0833083 0.9200674 0.0379539 -2.195 0.028165 \*

OBESE 0.0516911 1.0530504 0.0196776 2.627 0.008617 \*\*

RENLFAIL 0.0203539 1.0205625 0.0276837 0.735 0.462198

ULCER -0.4003371 0.6700941 0.3537693 -1.132 0.257789

VALVE -0.0759137 0.9268962 0.0443139 -1.713 0.086696 .

WGHTLOSS -0.1251817 0.8823365 0.0344959 -3.629 0.000285 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

exp(coef) exp(-coef) lower .95 upper .95

YEAR 1.0116 0.9886 1.0042 1.0190

AGE 0.9914 1.0086 0.9907 0.9922

RACE 0.9787 1.0217 0.9681 0.9895

FEMALE 1.0289 0.9719 1.0046 1.0538

NCHRONIC 0.9626 1.0388 0.9573 0.9679

DIED 0.8606 1.1619 0.4995 1.4830

ALCOHOL 1.0969 0.9117 1.0624 1.1324

ANEMDEF 0.9290 1.0765 0.8916 0.9679

ARTH 1.0535 0.9492 0.9722 1.1417

BLDLOSS 1.0057 0.9943 0.8877 1.1394

CHF 0.8366 1.1953 0.7887 0.8874

CHRNLUNG 1.0182 0.9821 0.9869 1.0505

COAG 0.9859 1.0143 0.9071 1.0716

DEPRESS 1.0156 0.9846 0.9571 1.0777

DRUG 1.4104 0.7090 1.3680 1.4541

HTN 0.8021 1.2467 0.7817 0.8230

LIVER 0.9201 1.0869 0.8541 0.9911

OBESE 1.0531 0.9496 1.0132 1.0945

RENLFAIL 1.0206 0.9799 0.9667 1.0775

ULCER 0.6701 1.4923 0.3350 1.3405

VALVE 0.9269 1.0789 0.8498 1.0110

WGHTLOSS 0.8823 1.1334 0.8247 0.9441

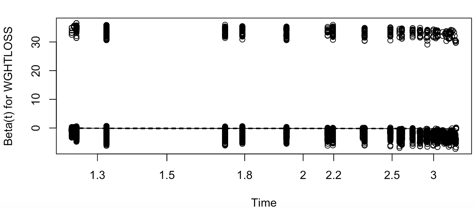
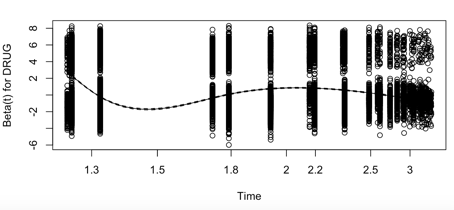
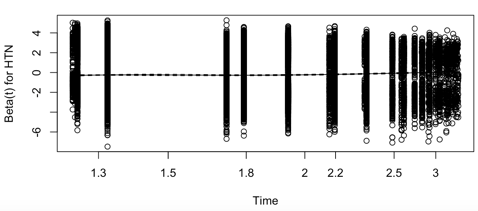
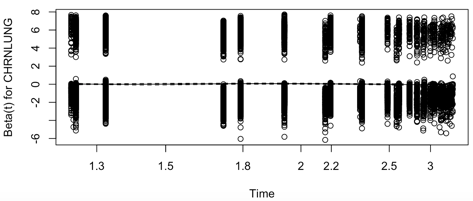
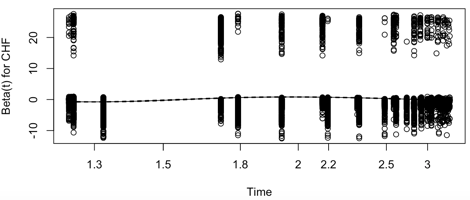
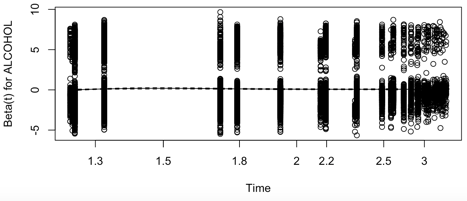
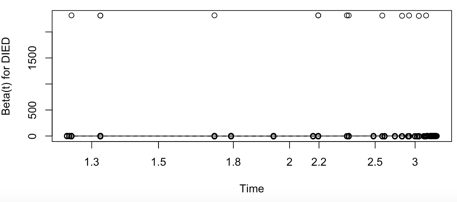
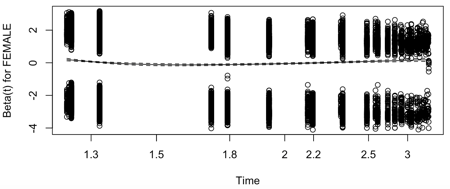
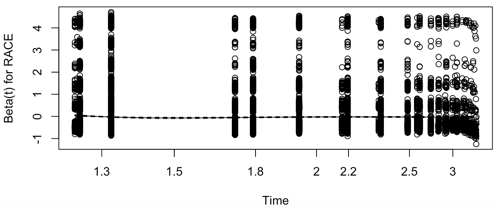
Concordance= 0.619 (se = 0.003 )

Rsquare= 0.06 (max possible= 1 )

Likelihood ratio test= 2764 on 22 df, p=0

Wald test = 2738 on 22 df, p=0

Score (logrank) test = 2770 on 22 df, p=0



Appendix 10: Weibull model

$coef

Estimate SE

lambda 1.649727e-12 1.232576e-11

gamma 3.131094e+00 1.318362e-02

YEAR 1.280414e-02 3.714192e-03

AGE -1.098418e-02 3.987871e-04

RACE -2.370269e-02 5.584291e-03

FEMALE -7.955918e-03 1.221164e-02

NCHRONIC -3.771904e-02 2.795007e-03

DIED -1.563256e-01 2.776345e-01

ALCOHOL 9.939747e-02 1.629199e-02

ANEMDEF -7.352307e-02 2.095640e-02

ARTH 5.547009e-02 4.100648e-02

BLDLOSS -8.230756e-04 6.365445e-02

CHF -1.891248e-01 3.008040e-02

CHRNLUNG 2.606142e-02 1.594526e-02

COAG -4.745270e-03 4.253729e-02

DEPRESS 2.347482e-02 3.028839e-02

DRUG 2.038557e-01 1.558168e-02

HTN -2.208518e-01 1.316477e-02

LIVER -7.332230e-02 3.795225e-02

OBESE 4.888729e-02 1.968660e-02

RENLFAIL 1.639602e-02 2.771417e-02

ULCER 0.000000e+00 0.000000e+00

VALVE -8.213245e-02 4.431074e-02

WGHTLOSS -1.539416e-01 3.447669e-02

$summary

Call:

survreg(formula = Surv(log(2 + LOS), 1 - LAMA) ~ ., data = dataFP,

dist = "weibull")

Value Std. Error z p

(Intercept) 7.24e+00 1.69e+00 4.29 1.8e-05

YEAR -3.33e-03 8.39e-04 -3.98 7.0e-05

AGE 3.42e-03 8.79e-05 38.88 < 2e-16

RACE 7.42e-03 1.26e-03 5.91 3.4e-09

FEMALE 9.93e-03 2.76e-03 3.60 0.00032

NCHRONIC 9.74e-03 6.30e-04 15.46 < 2e-16

DIED 3.50e-02 6.27e-02 0.56 0.57660

ALCOHOL -2.89e-02 3.68e-03 -7.86 3.8e-15

ANEMDEF 1.84e-02 4.73e-03 3.89 0.00010

ARTH -1.24e-02 9.26e-03 -1.33 0.18231

BLDLOSS 3.04e-03 1.44e-02 0.21 0.83237

CHF 4.24e-02 6.80e-03 6.23 4.6e-10

CHRNLUNG -7.94e-03 3.60e-03 -2.20 0.02757

COAG 1.00e-03 9.61e-03 0.10 0.91707

DEPRESS -8.53e-03 6.84e-03 -1.25 0.21237

DRUG -4.37e-02 3.52e-03 -12.40 < 2e-16

HTN 5.84e-02 2.98e-03 19.63 < 2e-16

LIVER 2.33e-02 8.58e-03 2.71 0.00666

OBESE -1.20e-02 4.45e-03 -2.69 0.00708

RENLFAIL -4.66e-03 6.27e-03 -0.74 0.45668

ULCER 0.00e+00 0.00e+00 NA NA

VALVE 2.26e-02 1.00e-02 2.26 0.02412

WGHTLOSS 4.94e-02 7.78e-03 6.35 2.1e-10

Log(scale) -1.49e+00 3.73e-03 -398.63 < 2e-16

Scale= 0.226

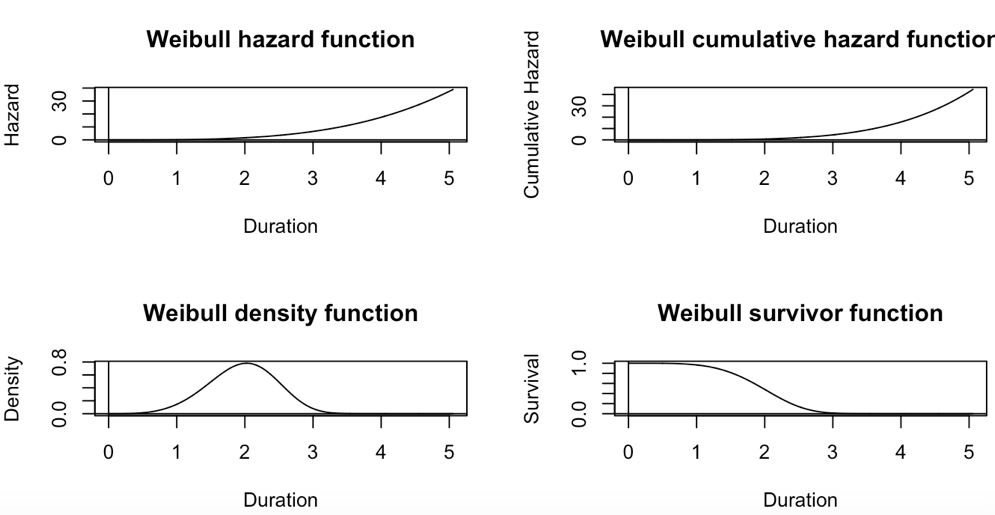
Weibull distribution

Loglik(model)= -27093.4 Loglik(intercept only)= -29401.1

Chisq= 4615.32 on 22 degrees of freedom, p= 0

Number of Newton-Raphson Iterations: 5

n= 44868



Appendix 11: Linear regression of LAMA

Residuals:

Min 1Q Median 3Q Max

-0.6650 -0.3593 -0.2065 0.5180 1.4372

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) -2.9755525 2.7521297 -1.081 0.2796

YEAR 0.0018531 0.0013681 1.354 0.1756

AGE -0.0067098 0.0001513 -44.353 <2e-16 \*\*\*

RACE -0.0005228 0.0020274 -0.258 0.7965

FEMALE -0.1440015 0.0043330 -33.233 <2e-16 \*\*\*

LOS -0.0093692 0.0004173 -22.450 <2e-16 \*\*\*

NCHRONIC 0.0020634 0.0010360 1.992 0.0464 \*

DIED -0.1022116 0.1234399 -0.828 0.4077

ALCOHOL -0.0023514 0.0057798 -0.407 0.6841

ANEMDEF 0.0090023 0.0077911 1.155 0.2479

ARTH -0.0029277 0.0154344 -0.190 0.8496

BLDLOSS -0.0121906 0.0234906 -0.519 0.6038

CHF -0.0042944 0.0113904 -0.377 0.7062

CHRNLUNG 0.0071363 0.0058490 1.220 0.2224

COAG 0.0111999 0.0155446 0.720 0.4712

DEPRESS -0.0047780 0.0110290 -0.433 0.6649

DRUG -0.0049465 0.0055172 -0.897 0.3700

HTN 0.0085567 0.0048231 1.774 0.0761 .

LIVER -0.0095089 0.0138947 -0.684 0.4938

OBESE -0.0129799 0.0072324 -1.795 0.0727 .

RENLFAIL -0.0005474 0.0103289 -0.053 0.9577

ULCER -0.0211431 0.1341527 -0.158 0.8748

VALVE 0.0415165 0.0161949 2.564 0.0104 \*

WGHTLOSS -0.0097453 0.0130720 -0.746 0.4560

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4447 on 44844 degrees of freedom

Multiple R-squared: 0.1049, Adjusted R-squared: 0.1044

F-statistic: 228.4 on 23 and 44844 DF, p-value: < 2.2e-16