Survival Analysis of Mortality of Adjuvant Chemotherapy for Colon Cancer

Yijia Jiang, Ziyan Xu

2022-12-04

1. Data Tidy

Recall that there are two records for each patient indicated by the event type (etype) variable, where etype == 1 refers to the event of a recurrence and etype == 2 indicates death. In order to answer our first research question, which is to study the time until death, we must create a marginal model by subsetting the colon data to only include the event of mortality.

```
# import the dataset from survival package
colon <- survival::colon

# tidy mortality dataset
colon.death <- colon %>% filter(etype == '2') %>%
               dplyr::select(-c(id,study,etype, node4)) %>%
               drop_na() %>%
               mutate(rx = as.factor(rx))
```

2. Kaplan-Meier Survival Estimate

```
death.fit <- survfit(Surv(time,status) ~ rx, data = colon.death)
options(max.print = 10000)
print(summary(death.fit))
```

```
## Call: survfit(formula = Surv(time, status) ~ rx, data = colon.death)
##
##                 rx=Obs
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##   113    305       1    0.997 0.00327        0.990        1.000
##   125    304       1    0.993 0.00462        0.984        1.000
##   145    303       1    0.990 0.00565        0.979        1.000
##   164    302       1    0.987 0.00651        0.974        1.000
##   166    301       1    0.984 0.00727        0.969        0.998
##   187    300       1    0.980 0.00795        0.965        0.996
##   208    299       1    0.977 0.00857        0.960        0.994
##   215    298       1    0.974 0.00915        0.956        0.992
```

```
##   218   297       1    0.970 0.00969        0.952        0.990
##   238   296       1    0.967 0.01020        0.947        0.987
##   241   295       1    0.964 0.01068        0.943        0.985
##   242   294       1    0.961 0.01113        0.939        0.983
##   253   293       1    0.957 0.01157        0.935        0.980
##   259   292       2    0.951 0.01238        0.927        0.975
##   264   290       1    0.948 0.01277        0.923        0.973
##   275   289       1    0.944 0.01314        0.919        0.970
##   289   288       1    0.941 0.01349        0.915        0.968
##   311   287       1    0.938 0.01384        0.911        0.965
##   313   286       1    0.934 0.01417        0.907        0.963
##   322   285       1    0.931 0.01450        0.903        0.960
##   331   284       1    0.928 0.01481        0.899        0.957
##   365   283       1    0.925 0.01512        0.895        0.955
##   372   282       1    0.921 0.01542        0.892        0.952
##   381   281       1    0.918 0.01571        0.888        0.949
##   384   280       2    0.911 0.01627        0.880        0.944
##   390   278       1    0.908 0.01653        0.876        0.941
##   409   277       1    0.905 0.01680        0.873        0.938
##   411   276       1    0.902 0.01705        0.869        0.936
##   413   275       2    0.895 0.01755        0.861        0.930
##   417   273       1    0.892 0.01779        0.858        0.927
##   421   272       1    0.889 0.01802        0.854        0.925
##   433   271       1    0.885 0.01825        0.850        0.922
##   437   270       1    0.882 0.01847        0.846        0.919
##   438   269       1    0.879 0.01869        0.843        0.916
##   459   267       1    0.875 0.01891        0.839        0.913
##   462   266       1    0.872 0.01913        0.835        0.910
##   464   265       1    0.869 0.01933        0.832        0.908
##   465   264       2    0.862 0.01974        0.824        0.902
##   474   262       1    0.859 0.01994        0.821        0.899
##   485   261       1    0.856 0.02013        0.817        0.896
##   499   260       1    0.852 0.02032        0.813        0.893
##   506   259       1    0.849 0.02051        0.810        0.890
##   510   258       1    0.846 0.02069        0.806        0.887
##   537   257       1    0.842 0.02087        0.803        0.884
##   563   256       2    0.836 0.02122        0.795        0.879
##   570   254       1    0.833 0.02139        0.792        0.876
##   576   253       1    0.829 0.02156        0.788        0.873
##   587   252       1    0.826 0.02172        0.785        0.870
##   591   251       1    0.823 0.02188        0.781        0.867
##   594   250       1    0.819 0.02204        0.777        0.864
##   595   249       1    0.816 0.02220        0.774        0.861
##   599   248       1    0.813 0.02235        0.770        0.858
##   612   247       1    0.810 0.02250        0.767        0.855
##   622   246       1    0.806 0.02265        0.763        0.852
##   659   245       1    0.803 0.02279        0.760        0.849
##   663   244       1    0.800 0.02294        0.756        0.846
##   665   243       1    0.796 0.02308        0.752        0.843
##   670   242       1    0.793 0.02321        0.749        0.840
##   673   241       1    0.790 0.02335        0.745        0.837
##   685   240       1    0.787 0.02348        0.742        0.834
##   687   239       1    0.783 0.02361        0.738        0.831
##   692   238       1    0.780 0.02374        0.735        0.828
```

```
##    709    237        1    0.777 0.02387        0.731         0.825
##    716    236        1    0.773 0.02400        0.728         0.822
##    717    235        1    0.770 0.02412        0.724         0.819
##    718    234        1    0.767 0.02424        0.721         0.816
##    721    233        1    0.764 0.02436        0.717         0.813
##    743    232        1    0.760 0.02447        0.714         0.810
##    753    231        1    0.757 0.02459        0.710         0.807
##    758    230        1    0.754 0.02470        0.707         0.804
##    760    229        1    0.750 0.02481        0.703         0.801
##    761    228        1    0.747 0.02492        0.700         0.798
##    770    227        1    0.744 0.02503        0.696         0.794
##    774    226        1    0.740 0.02513        0.693         0.791
##    775    225        1    0.737 0.02523        0.689         0.788
##    832    224        1    0.734 0.02533        0.686         0.785
##    833    223        1    0.731 0.02543        0.682         0.782
##    840    222        1    0.727 0.02553        0.679         0.779
##    845    221        1    0.724 0.02563        0.675         0.776
##    854    220        1    0.721 0.02572        0.672         0.773
##    863    219        1    0.717 0.02581        0.669         0.770
##    874    218        1    0.714 0.02590        0.665         0.767
##    883    217        1    0.711 0.02599        0.662         0.764
##    887    216        1    0.708 0.02608        0.658         0.761
##    901    215        1    0.704 0.02616        0.655         0.757
##    924    214        1    0.701 0.02625        0.651         0.754
##    928    213        1    0.698 0.02633        0.648         0.751
##    929    212        1    0.694 0.02641        0.645         0.748
##    936    211        1    0.691 0.02649        0.641         0.745
##    949    210        1    0.688 0.02657        0.638         0.742
##    957    209        1    0.685 0.02664        0.634         0.739
##    961    208        1    0.681 0.02672        0.631         0.736
##    963    207        1    0.678 0.02679        0.627         0.733
##    966    206        1    0.675 0.02686        0.624         0.729
##    976    205        1    0.671 0.02693        0.621         0.726
##   1021    204        1    0.668 0.02700        0.617         0.723
##   1031    203        1    0.665 0.02707        0.614         0.720
##   1048    202        1    0.661 0.02713        0.610         0.717
##   1070    201        1    0.658 0.02720        0.607         0.714
##   1079    200        1    0.655 0.02726        0.604         0.711
##   1083    199        1    0.652 0.02732        0.600         0.707
##   1101    198        1    0.648 0.02738        0.597         0.704
##   1133    197        1    0.645 0.02744        0.593         0.701
##   1134    196        1    0.642 0.02749        0.590         0.698
##   1136    195        1    0.638 0.02755        0.587         0.695
##   1139    194        1    0.635 0.02760        0.583         0.692
##   1159    193        1    0.632 0.02766        0.580         0.688
##   1166    192        1    0.629 0.02771        0.577         0.685
##   1178    191        1    0.625 0.02776        0.573         0.682
##   1195    190        1    0.622 0.02780        0.570         0.679
##   1198    189        1    0.619 0.02785        0.566         0.676
##   1209    188        1    0.615 0.02790        0.563         0.673
##   1216    187        1    0.612 0.02794        0.560         0.669
##   1230    186        1    0.609 0.02798        0.556         0.666
##   1237    185        1    0.606 0.02803        0.553         0.663
##   1246    184        1    0.602 0.02807        0.550         0.660
```

```
##  1262    183       1    0.599 0.02811        0.546        0.657
##  1272    182       1    0.596 0.02814        0.543        0.653
##  1290    181       1    0.592 0.02818        0.540        0.650
##  1295    180       1    0.589 0.02821        0.536        0.647
##  1304    179       1    0.586 0.02825        0.533        0.644
##  1313    178       1    0.583 0.02828        0.530        0.641
##  1314    177       1    0.579 0.02831        0.526        0.637
##  1327    176       1    0.576 0.02834        0.523        0.634
##  1363    175       1    0.573 0.02837        0.520        0.631
##  1375    174       1    0.569 0.02840        0.516        0.628
##  1434    173       1    0.566 0.02842        0.513        0.625
##  1437    172       1    0.563 0.02845        0.510        0.621
##  1447    171       1    0.559 0.02847        0.506        0.618
##  1482    170       1    0.556 0.02849        0.503        0.615
##  1530    169       1    0.553 0.02851        0.500        0.612
##  1548    168       1    0.550 0.02853        0.496        0.608
##  1656    167       1    0.546 0.02855        0.493        0.605
##  1679    166       1    0.543 0.02857        0.490        0.602
##  1692    165       1    0.540 0.02858        0.487        0.599
##  1723    164       1    0.536 0.02860        0.483        0.596
##  1745    163       1    0.533 0.02861        0.480        0.592
##  1772    162       1    0.530 0.02862        0.477        0.589
##  1788    161       1    0.527 0.02863        0.473        0.586
##  1790    160       1    0.523 0.02864        0.470        0.583
##  1818    157       1    0.520 0.02866        0.467        0.579
##  1875    150       1    0.516 0.02867        0.463        0.576
##  1896    148       1    0.513 0.02869        0.460        0.572
##  1907    145       1    0.509 0.02871        0.456        0.569
##  1915    144       1    0.506 0.02873        0.453        0.565
##  1950    141       1    0.502 0.02875        0.449        0.562
##  2077    131       1    0.498 0.02878        0.445        0.558
##  2083    130       1    0.495 0.02882        0.441        0.554
##  2085    129       1    0.491 0.02885        0.437        0.551
##  2133    117       1    0.487 0.02890        0.433        0.547
##  2171    105       1    0.482 0.02900        0.428        0.542
##  2213     92       1    0.477 0.02915        0.423        0.537
##  2257     81       1    0.471 0.02938        0.417        0.532
##  2284     76       1    0.465 0.02964        0.410        0.527
##  2287     75       1    0.458 0.02988        0.403        0.521
##  2351     66       1    0.452 0.03023        0.396        0.515
##  2527     47       1    0.442 0.03107        0.385        0.507
##  2552     42       1    0.431 0.03207        0.373        0.499
##  2789     16       1    0.404 0.03981        0.333        0.490
## 
##                 rx=Lev 
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    24    294       1    0.997 0.00340        0.990        1.000
##    56    293       1    0.993 0.00479        0.984        1.000
##    93    292       1    0.990 0.00586        0.978        1.000
##   122    291       1    0.986 0.00676        0.973        1.000
##   129    290       1    0.983 0.00754        0.968        0.998
##   133    289       1    0.980 0.00825        0.964        0.996
##   150    288       1    0.976 0.00889        0.959        0.994
##   165    287       1    0.973 0.00949        0.954        0.992
```

```
##   171    286       2    0.966 0.01057        0.945        0.987
##   191    284       1    0.963 0.01107        0.941        0.985
##   206    283       1    0.959 0.01154        0.937        0.982
##   219    282       2    0.952 0.01242        0.928        0.977
##   222    280       1    0.949 0.01283        0.924        0.974
##   226    279       1    0.946 0.01323        0.920        0.972
##   232    278       1    0.942 0.01361        0.916        0.969
##   257    277       1    0.939 0.01398        0.912        0.967
##   283    276       1    0.935 0.01434        0.908        0.964
##   314    275       2    0.929 0.01502        0.900        0.958
##   316    273       1    0.925 0.01535        0.896        0.956
##   323    272       1    0.922 0.01566        0.892        0.953
##   342    271       1    0.918 0.01597        0.888        0.950
##   343    270       1    0.915 0.01627        0.884        0.947
##   349    269       1    0.912 0.01656        0.880        0.945
##   355    268       1    0.908 0.01684        0.876        0.942
##   356    267       1    0.905 0.01712        0.872        0.939
##   362    266       1    0.901 0.01739        0.868        0.936
##   366    265       1    0.898 0.01765        0.864        0.933
##   376    264       1    0.895 0.01791        0.860        0.930
##   382    263       1    0.891 0.01816        0.856        0.927
##   402    262       1    0.888 0.01841        0.852        0.925
##   406    261       1    0.884 0.01865        0.849        0.922
##   420    260       1    0.881 0.01889        0.845        0.919
##   422    259       1    0.878 0.01912        0.841        0.916
##   430    258       1    0.874 0.01934        0.837        0.913
##   438    257       1    0.871 0.01957        0.833        0.910
##   439    256       1    0.867 0.01978        0.829        0.907
##   443    255       1    0.864 0.02000        0.826        0.904
##   472    254       1    0.861 0.02020        0.822        0.901
##   475    253       1    0.857 0.02041        0.818        0.898
##   486    252       1    0.854 0.02061        0.814        0.895
##   499    251       1    0.850 0.02081        0.811        0.892
##   512    250       1    0.847 0.02100        0.807        0.889
##   522    249       1    0.844 0.02119        0.803        0.886
##   546    248       1    0.840 0.02137        0.799        0.883
##   553    247       1    0.837 0.02156        0.796        0.880
##   559    246       1    0.833 0.02174        0.792        0.877
##   573    245       1    0.830 0.02191        0.788        0.874
##   580    244       1    0.827 0.02208        0.784        0.871
##   582    243       1    0.823 0.02225        0.781        0.868
##   589    242       1    0.820 0.02242        0.777        0.865
##   602    241       1    0.816 0.02258        0.773        0.862
##   608    240       1    0.813 0.02274        0.770        0.859
##   628    239       1    0.810 0.02290        0.766        0.856
##   629    238       1    0.806 0.02306        0.762        0.853
##   642    237       1    0.803 0.02321        0.758        0.850
##   643    236       1    0.799 0.02336        0.755        0.846
##   647    235       1    0.796 0.02351        0.751        0.843
##   664    234       1    0.793 0.02365        0.747        0.840
##   669    233       1    0.789 0.02379        0.744        0.837
##   675    232       1    0.786 0.02393        0.740        0.834
##   678    231       1    0.782 0.02407        0.737        0.831
##   684    230       1    0.779 0.02420        0.733        0.828
```

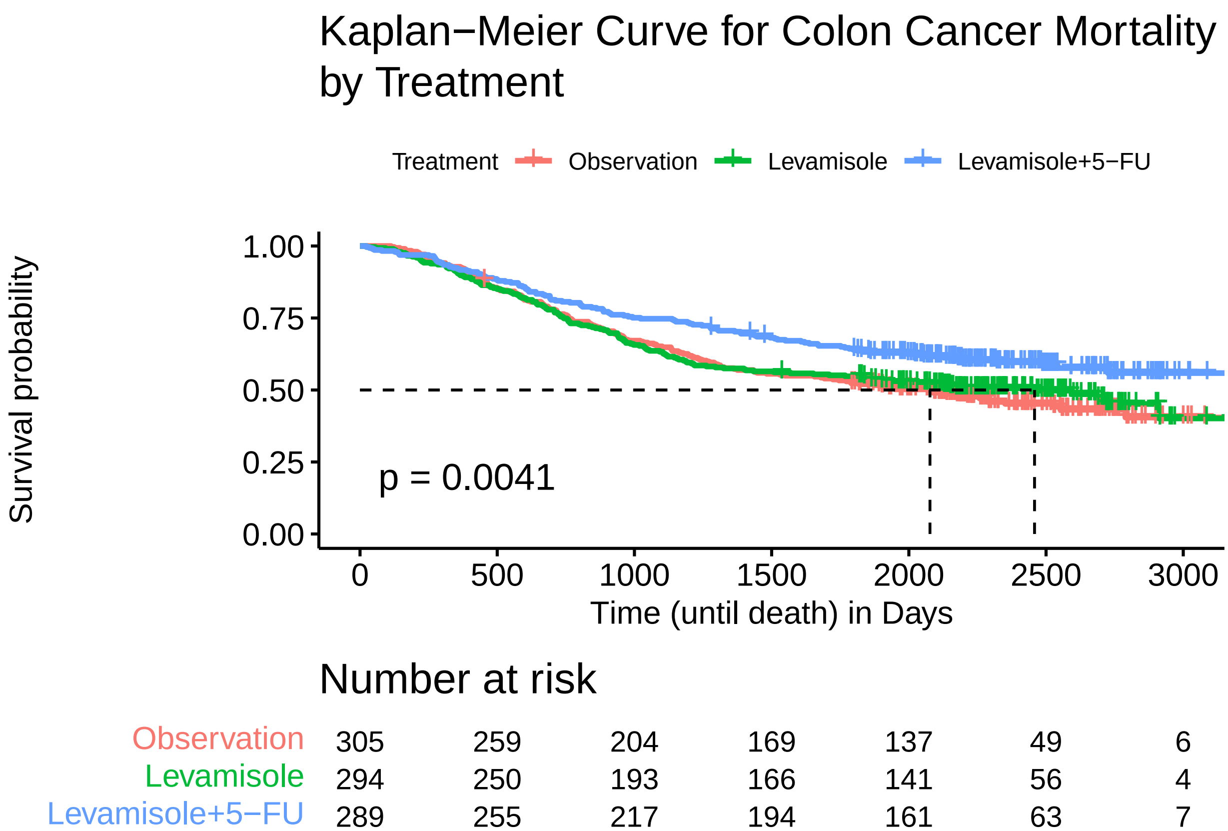
```
##   706    229         1    0.776 0.02433         0.729         0.825
##   708    228         1    0.772 0.02446         0.726         0.822
##   709    227         1    0.769 0.02459         0.722         0.818
##   720    226         1    0.765 0.02472         0.718         0.815
##   723    225         1    0.762 0.02484         0.715         0.812
##   729    224         1    0.759 0.02496         0.711         0.809
##   730    223         1    0.755 0.02508         0.708         0.806
##   739    222         1    0.752 0.02520         0.704         0.803
##   743    221         1    0.748 0.02531         0.700         0.800
##   755    220         1    0.745 0.02542         0.697         0.796
##   759    219         2    0.738 0.02564         0.690         0.790
##   764    217         1    0.735 0.02575         0.686         0.787
##   766    216         1    0.731 0.02585         0.682         0.784
##   797    215         1    0.728 0.02596         0.679         0.781
##   806    214         1    0.724 0.02606         0.675         0.777
##   833    213         1    0.721 0.02615         0.672         0.774
##   846    212         1    0.718 0.02625         0.668         0.771
##   858    211         1    0.714 0.02635         0.664         0.768
##   875    210         1    0.711 0.02644         0.661         0.765
##   890    209         1    0.707 0.02653         0.657         0.761
##   902    208         1    0.704 0.02662         0.654         0.758
##   905    207         1    0.701 0.02671         0.650         0.755
##   909    206         1    0.697 0.02679         0.647         0.752
##   938    205         1    0.694 0.02688         0.643         0.749
##   939    204         1    0.690 0.02696         0.640         0.745
##   940    203         1    0.687 0.02704         0.636         0.742
##   942    202         1    0.684 0.02712         0.633         0.739
##   944    201         1    0.680 0.02720         0.629         0.736
##   952    200         1    0.677 0.02728         0.625         0.732
##   961    199         2    0.670 0.02742         0.618         0.726
##   968    197         1    0.667 0.02749         0.615         0.723
##   969    196         1    0.663 0.02756         0.611         0.720
##   986    195         1    0.660 0.02763         0.608         0.716
##   997    194         1    0.656 0.02770         0.604         0.713
##  1018    193         1    0.653 0.02776         0.601         0.710
##  1034    192         1    0.650 0.02782         0.597         0.707
##  1037    191         1    0.646 0.02789         0.594         0.703
##  1041    190         1    0.643 0.02795         0.590         0.700
##  1046    189         1    0.639 0.02800         0.587         0.697
##  1055    188         1    0.636 0.02806         0.583         0.693
##  1092    187         1    0.633 0.02812         0.580         0.690
##  1103    186         1    0.629 0.02817         0.576         0.687
##  1105    185         1    0.626 0.02822         0.573         0.684
##  1112    184         1    0.622 0.02827         0.569         0.680
##  1117    183         1    0.619 0.02832         0.566         0.677
##  1122    182         1    0.616 0.02837         0.562         0.674
##  1145    181         1    0.612 0.02842         0.559         0.671
##  1154    180         1    0.609 0.02846         0.556         0.667
##  1161    179         1    0.605 0.02850         0.552         0.664
##  1178    178         1    0.602 0.02855         0.549         0.661
##  1186    177         1    0.599 0.02859         0.545         0.657
##  1191    176         1    0.595 0.02863         0.542         0.654
##  1207    175         1    0.592 0.02866         0.538         0.651
##  1215    174         1    0.588 0.02870         0.535         0.647
```

```
##  1219    173       1    0.585 0.02874        0.531        0.644
##  1262    172       1    0.582 0.02877        0.528        0.641
##  1295    171       1    0.578 0.02880        0.524        0.638
##  1325    170       1    0.575 0.02883        0.521        0.634
##  1399    169       1    0.571 0.02886        0.518        0.631
##  1405    168       1    0.568 0.02889        0.514        0.628
##  1434    167       1    0.565 0.02892        0.511        0.624
##  1509    166       1    0.561 0.02894        0.507        0.621
##  1568    164       1    0.558 0.02897        0.504        0.618
##  1652    163       1    0.554 0.02899        0.500        0.614
##  1709    162       1    0.551 0.02901        0.497        0.611
##  1768    161       1    0.548 0.02903        0.493        0.608
##  1829    158       1    0.544 0.02906        0.490        0.604
##  1839    156       1    0.541 0.02908        0.486        0.601
##  1850    155       1    0.537 0.02910        0.483        0.597
##  1851    154       1    0.534 0.02912        0.479        0.594
##  1885    151       1    0.530 0.02914        0.476        0.590
##  1932    148       1    0.526 0.02916        0.472        0.587
##  2023    140       1    0.523 0.02919        0.469        0.583
##  2079    134       1    0.519 0.02924        0.465        0.579
##  2128    127       1    0.515 0.02929        0.460        0.575
##  2152    118       1    0.510 0.02936        0.456        0.571
##  2171    114       1    0.506 0.02945        0.451        0.567
##  2458     63       1    0.498 0.03005        0.442        0.560
##  2593     40       1    0.485 0.03178        0.427        0.552
##  2683     32       1    0.470 0.03421        0.408        0.542
##  2718     25       1    0.451 0.03766        0.383        0.532
##  2910      9       1    0.401 0.05794        0.302        0.533
## 
##                 rx=Lev+5FU 
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##    23    289       1    0.997 0.00345        0.990        1.000
##    34    288       1    0.993 0.00488        0.984        1.000
##    45    287       1    0.990 0.00596        0.978        1.000
##    52    286       1    0.986 0.00687        0.973        1.000
##    79    285       1    0.983 0.00767        0.968        0.998
##   127    284       1    0.979 0.00839        0.963        0.996
##   138    283       1    0.976 0.00904        0.958        0.994
##   141    282       1    0.972 0.00965        0.954        0.991
##   144    281       1    0.969 0.01022        0.949        0.989
##   251    280       1    0.965 0.01075        0.945        0.987
##   269    279       1    0.962 0.01126        0.940        0.984
##   271    278       1    0.958 0.01174        0.936        0.982
##   274    277       1    0.955 0.01219        0.931        0.979
##   276    276       1    0.952 0.01263        0.927        0.977
##   279    275       1    0.948 0.01305        0.923        0.974
##   283    274       1    0.945 0.01345        0.919        0.971
##   293    273       1    0.941 0.01384        0.914        0.969
##   302    272       1    0.938 0.01422        0.910        0.966
##   304    271       1    0.934 0.01458        0.906        0.963
##   324    270       1    0.931 0.01493        0.902        0.961
##   326    269       1    0.927 0.01527        0.898        0.958
##   340    268       1    0.924 0.01560        0.894        0.955
##   355    267       1    0.920 0.01592        0.890        0.952
```

```
##   363    266       1    0.917 0.01623        0.886         0.949
##   389    265       1    0.913 0.01654        0.882         0.946
##   400    264       1    0.910 0.01683        0.878         0.944
##   428    263       1    0.907 0.01712        0.874         0.941
##   430    262       1    0.903 0.01740        0.870         0.938
##   441    261       1    0.900 0.01767        0.866         0.935
##   448    260       1    0.896 0.01794        0.862         0.932
##   454    259       1    0.893 0.01820        0.858         0.929
##   460    258       1    0.889 0.01846        0.854         0.926
##   484    257       1    0.886 0.01871        0.850         0.923
##   498    256       1    0.882 0.01895        0.846         0.920
##   503    255       1    0.879 0.01919        0.842         0.917
##   529    254       1    0.875 0.01943        0.838         0.914
##   550    253       1    0.872 0.01965        0.834         0.911
##   576    252       1    0.869 0.01988        0.830         0.908
##   578    251       1    0.865 0.02010        0.827         0.905
##   580    250       1    0.862 0.02031        0.823         0.902
##   592    249       1    0.858 0.02052        0.819         0.899
##   601    248       1    0.855 0.02073        0.815         0.896
##   603    247       1    0.851 0.02093        0.811         0.893
##   609    246       1    0.848 0.02113        0.807         0.890
##   614    245       1    0.844 0.02133        0.804         0.887
##   616    244       1    0.841 0.02152        0.800         0.884
##   641    243       1    0.837 0.02171        0.796         0.881
##   642    242       1    0.834 0.02189        0.792         0.878
##   666    241       1    0.830 0.02207        0.788         0.875
##   674    240       1    0.827 0.02225        0.785         0.872
##   692    239       2    0.820 0.02260        0.777         0.866
##   693    237       1    0.817 0.02276        0.773         0.862
##   696    236       1    0.813 0.02293        0.769         0.859
##   712    235       1    0.810 0.02309        0.766         0.856
##   736    234       1    0.806 0.02325        0.762         0.853
##   765    233       1    0.803 0.02341        0.758         0.850
##   802    232       2    0.796 0.02371        0.751         0.844
##   806    230       1    0.792 0.02386        0.747         0.841
##   811    229       1    0.789 0.02400        0.743         0.837
##   844    228       1    0.785 0.02415        0.740         0.834
##   862    227       1    0.782 0.02429        0.736         0.831
##   884    226       1    0.779 0.02442        0.732         0.828
##   887    225       2    0.772 0.02469        0.725         0.822
##   905    223       1    0.768 0.02482        0.721         0.818
##   911    222       1    0.765 0.02495        0.717         0.815
##   916    221       1    0.761 0.02508        0.714         0.812
##   961    220       1    0.758 0.02520        0.710         0.809
##   977    219       1    0.754 0.02532        0.706         0.806
##   993    218       1    0.751 0.02544        0.703         0.802
##  1022    217       1    0.747 0.02556        0.699         0.799
##  1138    216       1    0.744 0.02567        0.695         0.796
##  1145    215       1    0.740 0.02579        0.692         0.793
##  1151    214       1    0.737 0.02590        0.688         0.790
##  1193    213       1    0.734 0.02601        0.684         0.786
##  1201    212       1    0.730 0.02611        0.681         0.783
##  1212    211       1    0.727 0.02622        0.677         0.780
##  1246    210       1    0.723 0.02632        0.673         0.777
```

```
## 1276    209       2    0.716 0.02652        0.666        0.770
## 1279    207       1    0.713 0.02661        0.663        0.767
## 1302    205       1    0.709 0.02671        0.659        0.764
## 1306    204       1    0.706 0.02681        0.655        0.760
## 1365    203       1    0.702 0.02690        0.652        0.757
## 1387    202       1    0.699 0.02699        0.648        0.754
## 1388    201       1    0.695 0.02708        0.644        0.751
## 1424    199       1    0.692 0.02717        0.641        0.747
## 1439    198       1    0.688 0.02725        0.637        0.744
## 1446    197       1    0.685 0.02734        0.633        0.741
## 1495    195       1    0.681 0.02742        0.630        0.737
## 1511    194       1    0.678 0.02750        0.626        0.734
## 1521    193       1    0.674 0.02759        0.622        0.731
## 1550    192       1    0.671 0.02766        0.619        0.727
## 1607    191       1    0.667 0.02774        0.615        0.724
## 1620    190       1    0.664 0.02782        0.612        0.721
## 1637    189       1    0.660 0.02789        0.608        0.717
## 1668    188       1    0.657 0.02796        0.604        0.714
## 1671    187       1    0.653 0.02803        0.601        0.711
## 1752    186       1    0.650 0.02810        0.597        0.707
## 1767    185       1    0.646 0.02817        0.593        0.704
## 1783    184       1    0.643 0.02823        0.590        0.701
## 1798    183       1    0.639 0.02830        0.586        0.697
## 1812    181       1    0.636 0.02836        0.583        0.694
## 1831    176       1    0.632 0.02843        0.579        0.690
## 1856    174       1    0.628 0.02850        0.575        0.687
## 1995    163       1    0.625 0.02858        0.571        0.683
## 2021    158       1    0.621 0.02867        0.567        0.680
## 2052    152       1    0.617 0.02877        0.563        0.676
## 2127    139       1    0.612 0.02891        0.558        0.672
## 2174    129       1    0.607 0.02907        0.553        0.667
## 2197    120       1    0.602 0.02926        0.548        0.663
## 2318     97       1    0.596 0.02961        0.541        0.657
## 2482     70       1    0.588 0.03039        0.531        0.650
## 2542     51       1    0.576 0.03190        0.517        0.642
## 2725     33       1    0.559 0.03539        0.493        0.633
```

```
ggsurvplot(death.fit, conf.int = F, break.time.by = 500, pval = TRUE,
           font.x.size = 12, font.y.size = 12, font.legend.size = 9, surv.median.line = "hv",
           legend.title = "Treatment", legend.labs = c("Observation", "Levamisole","Levamisole+5-FU"),
           title = "Kaplan-Meier Curve for Colon Cancer Mortality \nby Treatment",
           xlab = "Time (until death) in Days",
           risk.table = T, risk.table.height = 0.25, risk.table.fontsize = 4,
           tables.theme = theme_cleantable())
```



From the plot above, there is some indication that patients who received the adjuvant treatment with levamisole plus fluorouracil (Lev + 5FU) have a higher survival probability than patients with no further treatment and patients who received the treatment with levamisole alone. The median survival time for observation group and levamisole group are approximately 2100 days and 2200 days, respectively. However, until the end of the trial, the survival probability of Levamisole + 5-FU treatment group is greater than 50% as we fail to observe the curve crossing the 50% line.

3. Log-Rank Test

Noticing the difference of survival probability between the three treatment groups, we conduct a Log-rank hypothesis test to test the null hypothesis of no difference among the three treatments in the mortality model.

```
survdiff(Surv(time, status) ~ rx, data = colon.death)
```

```
## Call:
## survdiff(formula = Surv(time, status) ~ rx, data = colon.death)
##
##                N Observed Expected (O-E)^2/E (O-E)^2/V
## rx=Obs       305      164      143     3.170      4.75
## rx=Lev       294      149      138     0.868      1.28
## rx=Lev+5FU 289      117      149     6.957     10.67
##
##  Chisq= 11  on 2 degrees of freedom, p= 0.004
```

From this log-rank test, we get a p-value that is closed to 0.004, which is significant at a 0.05 level. We want to conclude that there is a significant difference among the three treatments in the mortality model.

4. Cox PH Model

4.1 Model Selection

We now use automatic stepwise selection with Akaike information criterion (AIC) to determine the covariates that best represent an appropriate cox proportional hazards model for the event of death.

```r
# fit all the variables in the model
d.model.full <- coxph(Surv(time, status) ~ ., data = colon.death)

# stepwise selection with AIC criterion
d.model.aic <- step(d.model.full, direction = "both", k = 2)
```

```
## Start:  AIC=5440.31
## Surv(time, status) ~ rx + sex + age + obstruct + perfor + adhere +
##     nodes + differ + extent + surg
##
##            Df    AIC
## - perfor    1 5438.3
## - sex       1 5438.5
## - adhere    1 5440.0
## - differ    1 5440.2
## <none>        5440.3
## - age       1 5440.6
## - obstruct  1 5442.1
## - surg      1 5443.6
## - rx        2 5446.4
## - extent    1 5455.3
## - nodes     1 5502.9
##
## Step:  AIC=5438.33
## Surv(time, status) ~ rx + sex + age + obstruct + adhere + nodes +
##     differ + extent + surg
##
##            Df    AIC
## - sex       1 5436.5
## - adhere    1 5438.0
## - differ    1 5438.2
## <none>        5438.3
## - age       1 5438.6
## - obstruct  1 5440.1
## + perfor    1 5440.3
## - surg      1 5441.6
## - rx        2 5444.4
## - extent    1 5453.3
## - nodes     1 5501.0
##
```

```
## Step:  AIC=5436.47
## Surv(time, status) ~ rx + age + obstruct + adhere + nodes + differ +
##     extent + surg
##
##            Df    AIC
## - adhere    1 5436.2
## - differ    1 5436.3
## <none>        5436.5
## - age       1 5436.8
## - obstruct  1 5438.3
## + sex       1 5438.3
## + perfor    1 5438.5
## - surg      1 5439.7
## - rx        2 5442.4
## - extent    1 5451.5
## - nodes     1 5499.1
##
## Step:  AIC=5436.18
## Surv(time, status) ~ rx + age + obstruct + nodes + differ + extent +
##     surg
##
##            Df    AIC
## <none>        5436.2
## + adhere    1 5436.5
## - differ    1 5436.5
## - age       1 5436.9
## + sex       1 5438.0
## - obstruct  1 5438.1
## + perfor    1 5438.2
## - surg      1 5439.4
## - rx        2 5442.3
## - extent    1 5452.6
## - nodes     1 5498.3
```

The resulting model with the lowest AIC is: Surv(time, status) ~ obstruct + differ + extent + surg + nodes + age + rx

Next, we used the Analysis of Deviance procedure to get the proper Likelihood Ratio Test to confirm if each of the covariates selected by the stepwise selection method is significant to include in the Cox Proportional Model.

```
anova(d.model.aic)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(time, status)
## Terms added sequentially (first to last)
##
##           loglik   Chisq Df Pr(>|Chi|)
## NULL     -2767.9
## rx       -2762.2 11.4100  2   0.003329 **
## age      -2761.9  0.6730  1   0.411993
## obstruct -2759.8  4.1666  1   0.041229 *
## nodes    -2723.9 71.8138  1  < 2.2e-16 ***
## differ   -2721.9  3.9678  1   0.046378 *
```

```
## extent   -2712.7 18.3458  1  1.842e-05 ***
## surg     -2710.1  5.2615  1   0.021802 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We can see that p-values for covariates “obstruct”, “differ”, “extent”, “surg”, “nodes” and “rx” are much smaller than 0.05, except for “age”, indicating that they have a significant effect on time until death. Therefore, we will include these 6 covariates in our Cox PH model.

Therefore, we obtain the resulting model, which is Surv(time, status) ~ obstruct + differ + extent + surg + nodes + rx.

4.2 Model Diagnostics

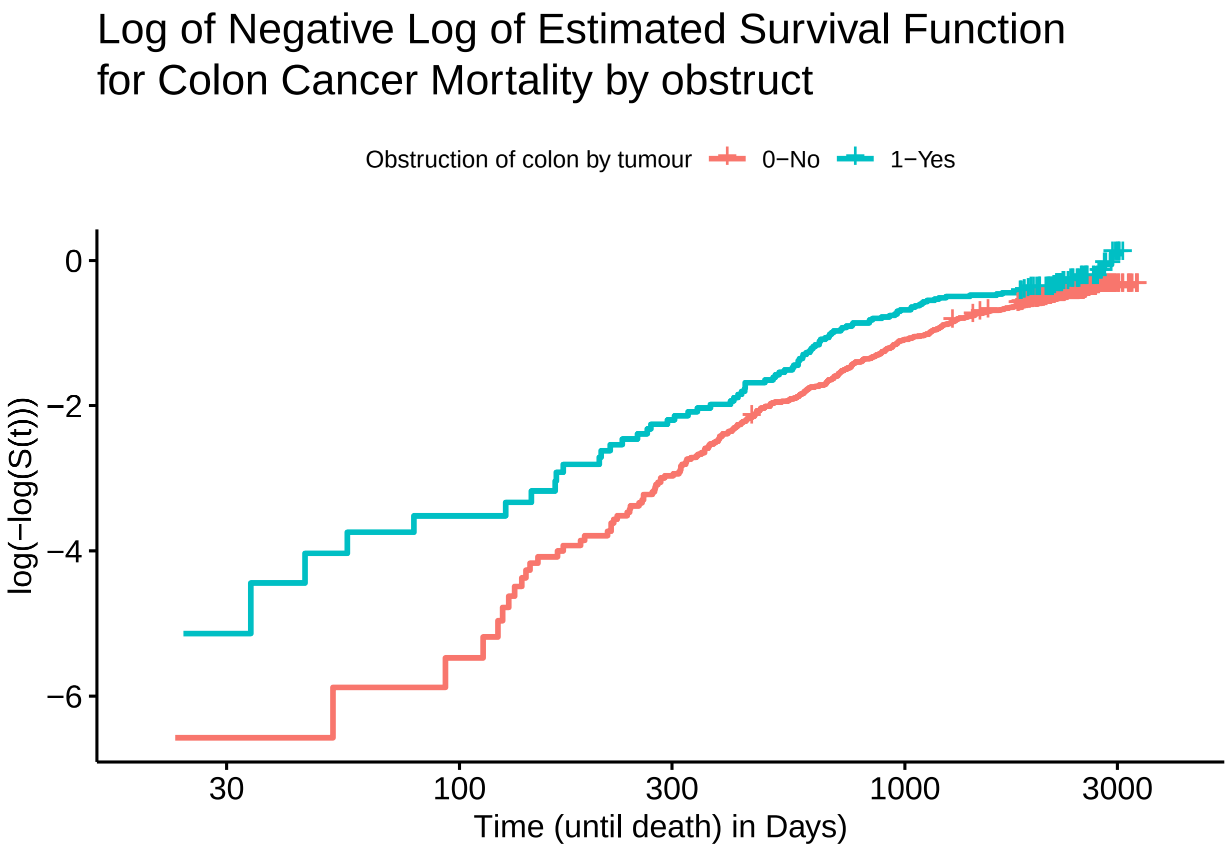
4.2.1 Check proportionality of hazard ratios

Log of Negative Log of Estimated Survival Function

To check the proportional hazards assumption for this model, we make diagnostic plots using log of negative log of estimated survival function. First comes to covariate obstruct.

```
d.obstruct.fit <- survfit(Surv(time, status) ~ obstruct, data = colon.death)
cloglog_obstruct <- ggsurvplot(d.obstruct.fit, conf.int = F, font.x.size = 12, font.y.size = 12, font.le
                               fun = "cloglog",
                               xlim = c(20, 4000),
                               xlab = "Time (until death) in Days)",
                               title = "Log of Negative Log of Estimated Survival Function \nfor Colon C
                               legend.title = "Obstruction of colon by tumour",
                               legend.labs = c("0-No", "1-Yes"))

cloglog_obstruct
```

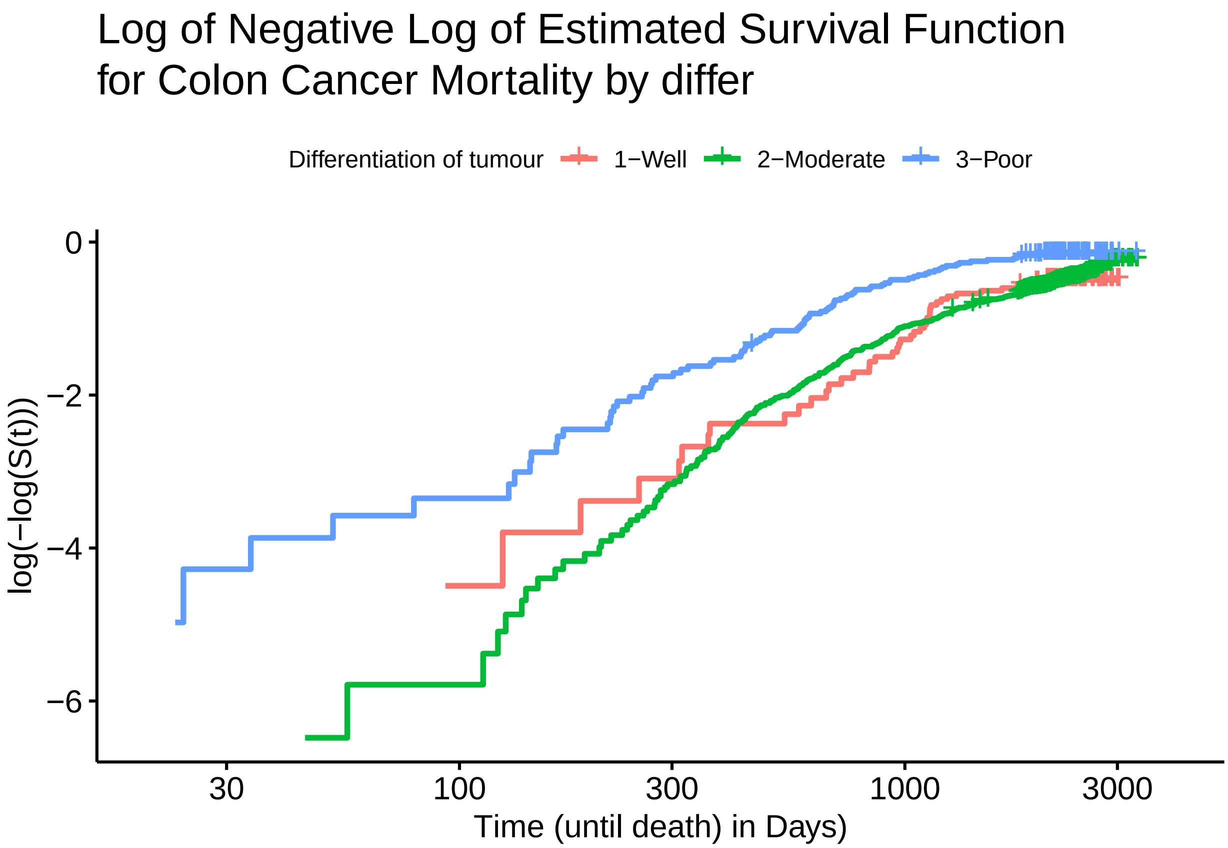


According to the plot, the distance between two obstruct curves begin to narrow after 2000 days which causes some concern that the assumption might be violated. However, it is also reasonable to assume that the curves are wider apparent earlier in the study since there are less occurrences of death before 2000 days. Hence we believe it is best to ignore the noisiness of the plot since the curves are roughly parallel after 2000 days. Thus, the cox proportional hazards assumption is valid to use for "obstruct".

We continue to plot the C-log-log plot for the covariate differ.

```
d.differ.fit <- survfit(Surv(time, status) ~ differ, data = colon.death)
cloglog_differ <- ggsurvplot(d.differ.fit, conf.int = F, font.x.size = 12, font.y.size = 12, font.legen
                            fun = "cloglog",
                            xlim = c(20, 4000),
                            xlab = "Time (until death) in Days)",
                            title = "Log of Negative Log of Estimated Survival Function \nfor Colon Ca
                            legend.title = "Differentiation of tumour",
                            legend.labs = c("1-Well", "2-Moderate","3-Poor"))

cloglog_differ
```

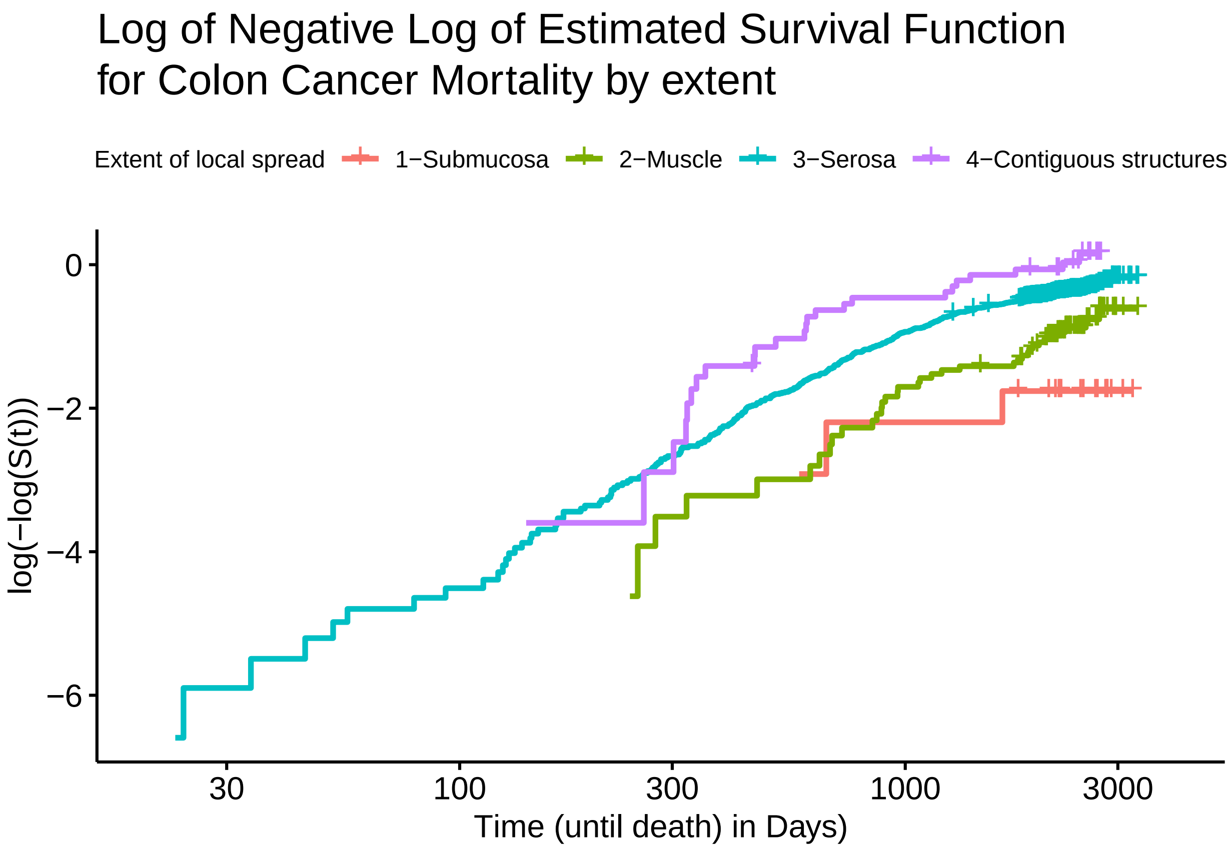


According to the plot, the curves in the C-log-log plot are crossing over after 300 days. The assumption of proportional hazard ratio among the three differ groups is violated as we fail to see three parallel lines against log time.

We continue to plot the C-log-log plot for the covariate extent.

```
d.extent.fit <- survfit(Surv(time, status) ~ extent, data = colon.death)
cloglog_extent <- ggsurvplot(d.extent.fit, conf.int = F, font.x.size = 12, font.y.size = 12, font.legend
                             fun = "cloglog",
                             xlim = c(20, 4000),
                             xlab = "Time (until death) in Days)",
                             title = "Log of Negative Log of Estimated Survival Function \nfor Colon Ca
                             legend.title = "Extent of local spread",
                             legend.labs = c("1-Submucosa", "2-Muscle", "3-Serosa", "4-Contiguous struc

cloglog_extent
```

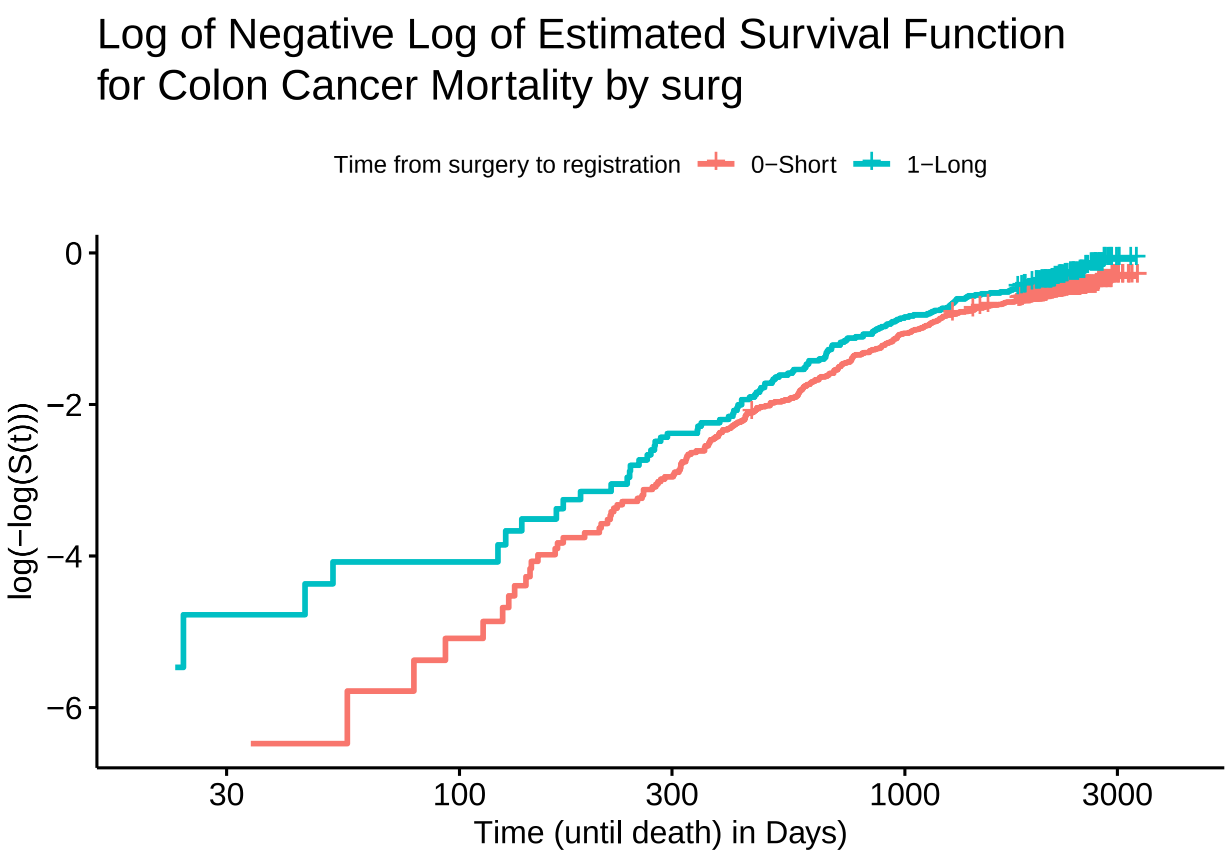


According to the plot, the curves in the C-log-log plot are crossing over after 100 days. The assumption of proportional hazard ratio of extent is violated as we fail to see four parallel lines against log time.

We continue to plot the C-log-log plot for the covariate surg.

```
d.surg.fit <- survfit(Surv(time, status) ~ surg, data = colon.death)
cloglog_surg <- ggsurvplot(d.surg.fit, conf.int = F, font.x.size = 12, font.y.size = 12, font.legend.si
                           fun = "cloglog",
                           xlim = c(20, 4000),
                           xlab = "Time (until death) in Days)",
                           title = "Log of Negative Log of Estimated Survival Function \nfor Colon Canc
                           legend.title = "Time from surgery to registration",
                           legend.labs = c("0-Short", "1-Long"))

cloglog_surg
```

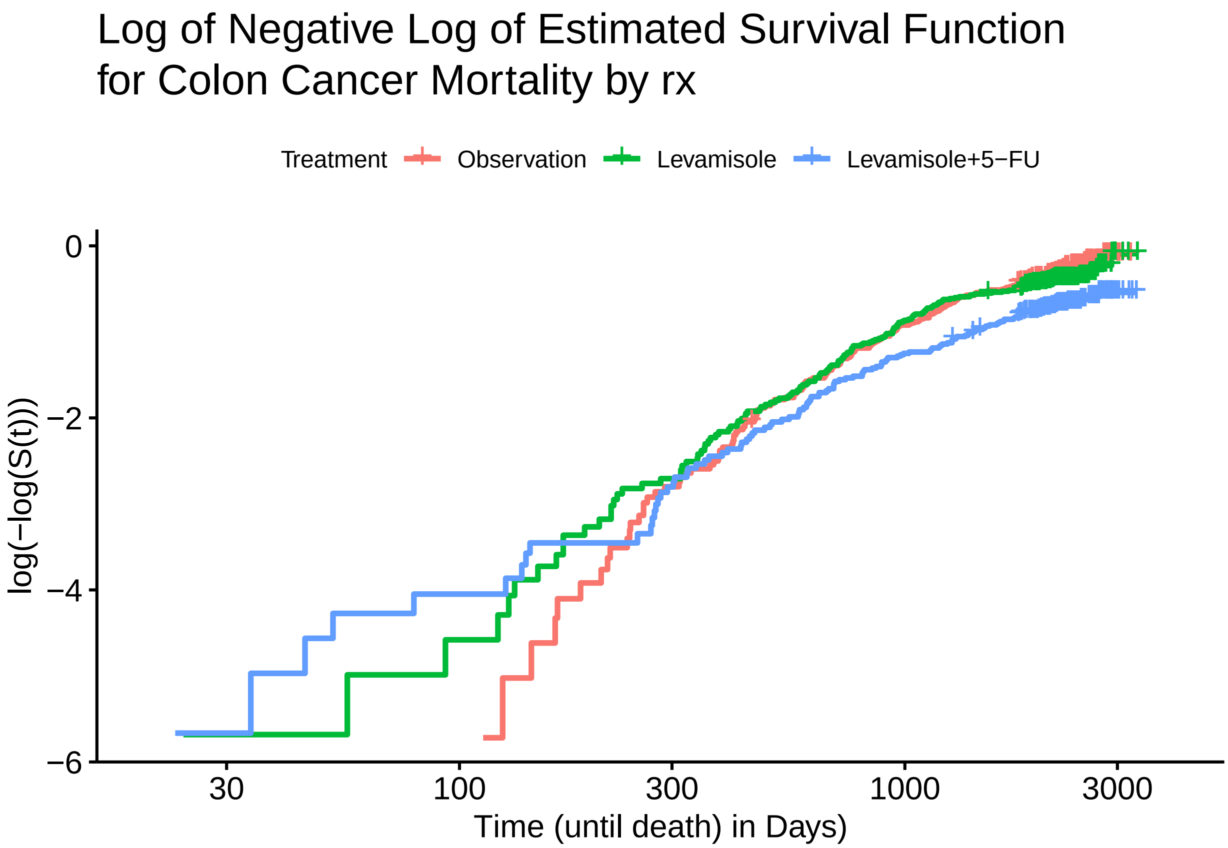


According to the plot, the distance between two surg curves begin to narrow after 100 days and becomes parallel to each other. We can assume that the cox proportional hazards assumption is valid to use for `surg`.

We continue to plot the C-log-log plot for the covariate rx.

```
d.surg.fit <- survfit(Surv(time, status) ~ rx, data = colon.death)
cloglog_rx <- ggsurvplot(d.surg.fit, conf.int = F, font.x.size = 12, font.y.size = 12, font.legend.size
                         fun = "cloglog",
                         xlim = c(20, 4000),
                         xlab = "Time (until death) in Days)",
                         title = "Log of Negative Log of Estimated Survival Function \nfor Colon Cancer
                         legend.title = "Treatment",
                         legend.labs = c("Observation", "Levamisole","Levamisole+5-FU"))

cloglog_rx
```



According to the plot, the distance between three treatment curves begin to narrow after 2000 days which causes some concern that the assumption might be violated. However, it is also reasonable to assume that the curves are wider apparent earlier in the study since there are less occurrences of death before 2000 days. Hence we believe it is best to ignore the noisiness of the plot since the curves are roughly parallel after 2000 days. Thus, the cox proportional hazards assumption is valid to use for this covariate.

```
# combine all the cloglog plots and save them to a pdf file
splots <- list()
splots[[1]] <- cloglog_obstruct
splots[[2]] <- cloglog_differ
splots[[3]] <- cloglog_extent
splots[[4]] <- cloglog_surg
splots[[5]] <- cloglog_rx
cloglog_plot = arrange_ggsurvplots(splots, print = FALSE, ncol = 2, nrow = 3)

ggsave(cloglog_plot,file = "./plot/d.C-log-log-plots.pdf",width = 12,height = 15)
ggsave(cloglog_plot,file = "./plot/d.C-log-log-plots.png",width = 12,height = 15)
```

Schoenfeld residuals

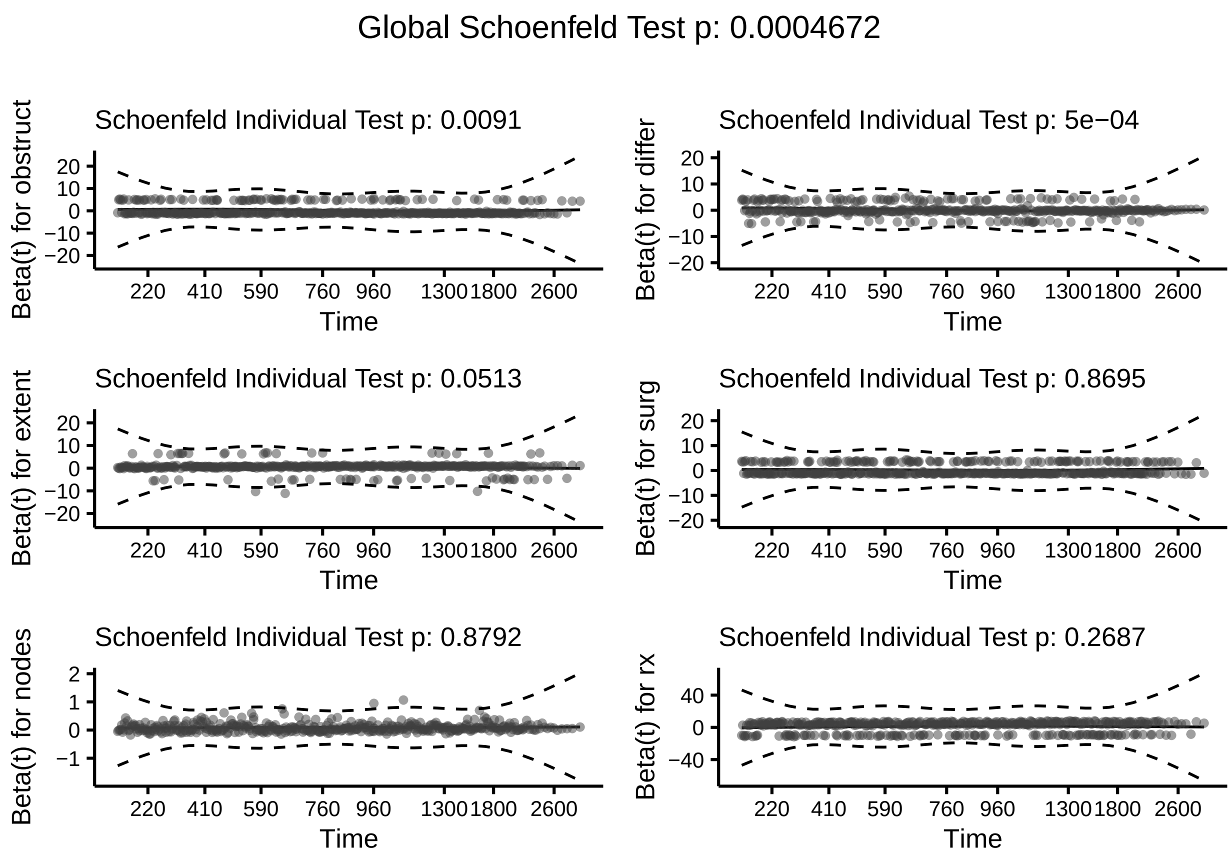
```
czph <- cox.zph(coxph(Surv(time, status) ~ obstruct + differ + extent + surg + nodes + rx, data = colon
czph
```

```
##             chisq df       p
## obstruct   6.7958  1 0.00914
```

```
## differ    12.2795  1 0.00046
## extent     3.7997  1 0.05126
## surg       0.0270  1 0.86951
## nodes      0.0231  1 0.87925
## rx         2.6285  2 0.26868
## GLOBAL    26.1823  7 0.00047
```

```
schoenfeld_plot <- ggcoxzph(czph, font.main = 10, font.x = 10, font.y = 10, font.tickslab = 8,
                            point.alpha = 0.5, point.col = "grey25")

schoenfeld_plot
```

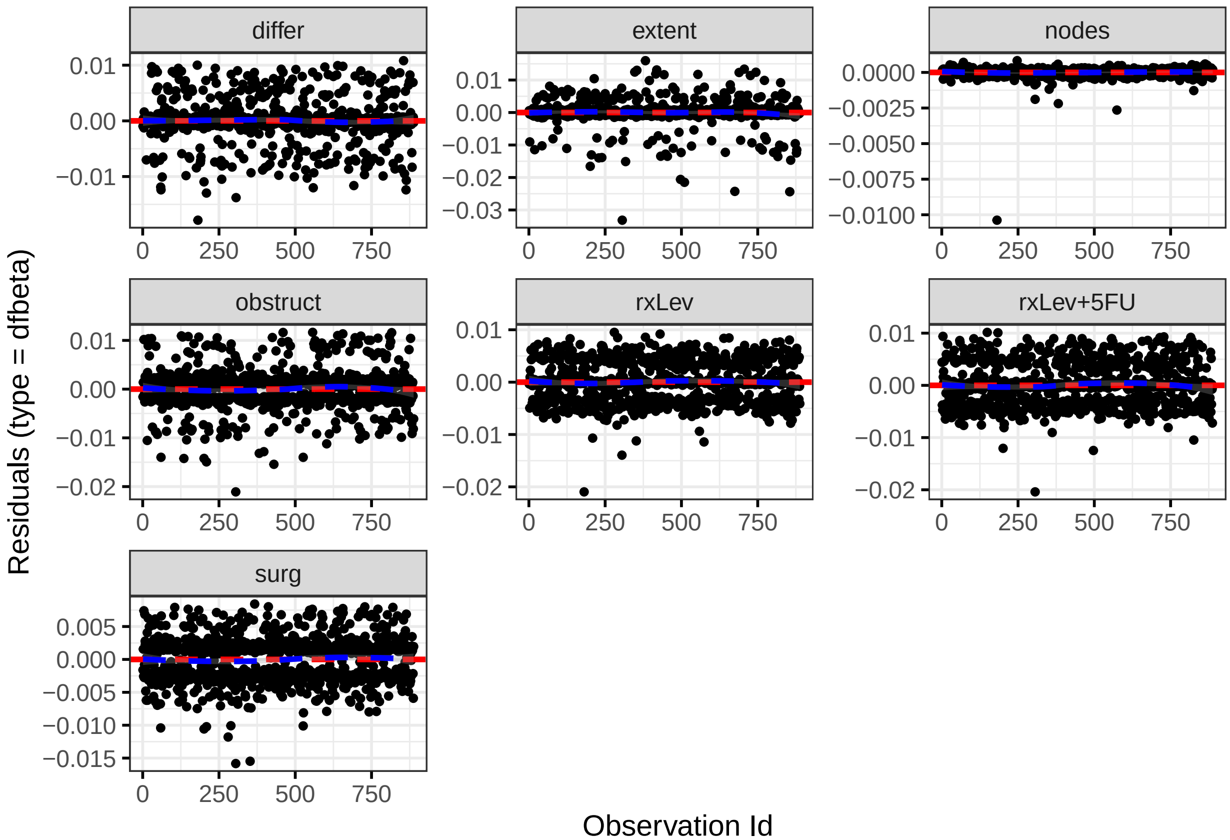



```
ggsave("./plot/d.schoenfeld residual_plots.pdf", arrangeGrob(grobs = schoenfeld_plot))
ggsave("./plot/d.schoenfeld residual_plots.png", arrangeGrob(grobs = schoenfeld_plot))
```

From the output above, the tests for covariates "obstruct", "differ" are statistically significant, and the global test is also statistically significant. Therefore, we can assume the violation of proportional hazards on these covariates, which requires corrections of non-proportional hazard ratio.

4.2.2 Test influential observations

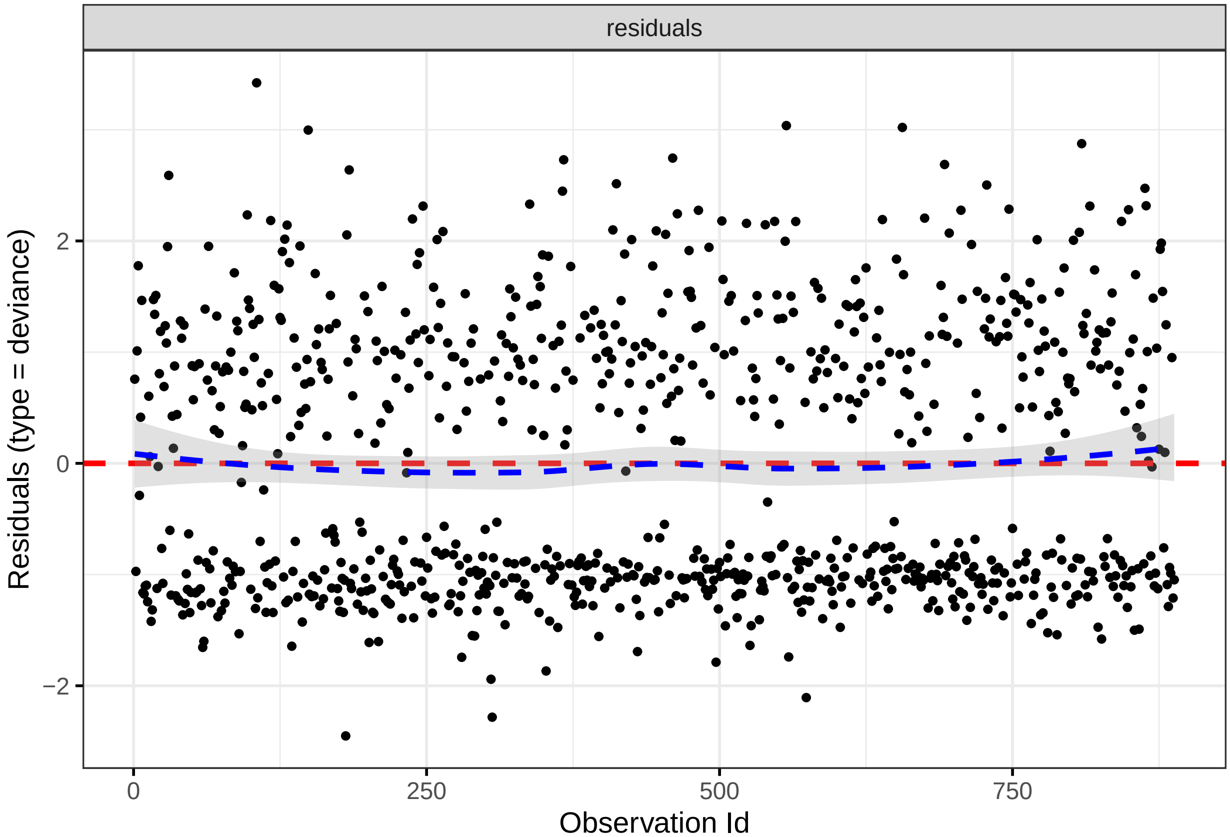
```
# check outliers in term of dfbeta
outlier_dfbeta <- ggcoxdiagnostics(coxph(Surv(time, status) ~ obstruct + differ + extent + surg + nodes
                                  type = "dfbeta", linear.predictions = FALSE, ggtheme = theme_bw())
outlier_dfbeta
```




```
ggsave(outlier_dfbeta, file = "./plot/d.outlier_dfbeta.pdf")
ggsave(outlier_dfbeta, file = "./plot/d.outlier_dfbeta.png")

outlier_deviance <- ggcoxdiagnostics(coxph(Surv(time, status) ~ obstruct + differ + extent + surg + node
                                    type = "deviance", linear.predictions = FALSE, ggtheme = theme_bw())

outlier_deviance
```



It's also possible to check outliers by visualizing the influence of each point, in terms of DFBETA - the impact on the coefficient of covariates in the model were that specific point to be removed from the data set. In general, we can identify the observations with positive and negative DFBETAs. Comparing the magnitudes of the largest dfbeta values to the regression coefficients suggests that none of the observations is terribly influential individually, even though some of the dfbeta values for extent are large compared with the others.

4.3 Corrections for violation of the PH Assumption

```
d.model.inter = coxph(Surv(time, status) ~ rx + obstruct + differ + extent + surg + nodes + tt(obstruct)

summary(d.model.inter)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ rx + obstruct + differ +
##     extent + surg + nodes + tt(obstruct) + tt(differ), data = colon.death,
##     tt = function(x, t, ...) x * log(t))
##
##   n= 888, number of events= 430
##
##                    coef exp(coef)  se(coef)      z Pr(>|z|)
## rxLev         -0.067966  0.934292  0.114157 -0.595 0.551597
## rxLev+5FU     -0.356181  0.700346  0.121620 -2.929 0.003405 **
## obstruct       3.005889 20.204167  0.913252  3.291 0.000997 ***
## differ         3.324329 27.780355  0.794159  4.186 2.84e-05 ***
```

```
## extent         0.500140  1.648951  0.116581  4.290 1.79e-05 ***
## surg           0.249440  1.283307  0.106085  2.351 0.018707 *
## nodes          0.086542  1.090397  0.009386  9.221  < 2e-16 ***
## tt(obstruct) -0.428125  0.651730  0.140517 -3.047 0.002313 **
## tt(differ)   -0.485106  0.615632  0.120666 -4.020 5.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## rxLev           0.9343    1.07033    0.7470    1.1686
## rxLev+5FU       0.7003    1.42787    0.5518    0.8889
## obstruct       20.2042    0.04949    3.3735  121.0047
## differ         27.7804    0.03600    5.8580  131.7429
## extent          1.6490    0.60645    1.3121    2.0722
## surg            1.2833    0.77924    1.0424    1.5799
## nodes           1.0904    0.91710    1.0705    1.1106
## tt(obstruct)    0.6517    1.53438    0.4948    0.8584
## tt(differ)      0.6156    1.62435    0.4860    0.7799
##
## Concordance= 0.671  (se = 0.013 )
## Likelihood ratio test= 137.1  on 9 df,   p=<2e-16
## Wald test            = 160.4  on 9 df,   p=<2e-16
## Score (logrank) test = 166.1  on 9 df,   p=<2e-16
```

```
anova(d.model.inter)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(time, status)
## Terms added sequentially (first to last)
##
##               loglik      Chisq Df Pr(>|Chi|)
## NULL         -2767.9
## rx           -4789.5 -4043.2284  2  1.0000000
## obstruct     -4787.4     4.1985  1  0.0404592 *
## differ       -4781.2    12.3563  1  0.0004395 ***
## extent       -4767.0    28.4124  1  9.804e-08 ***
## surg         -4765.2     3.7231  1  0.0536662 .
## nodes        -4716.8    96.7371  1  < 2.2e-16 ***
## tt(obstruct) -4632.3   169.0432  1  < 2.2e-16 ***
## tt(differ)   -2699.4  3865.8481  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We observe significance from Schoenfeld residuals, so we modify the model by adding the interaction of covariates with function of time.

4.4 Final model

With the inclusion of treatment, our final model is given by: Surv(time, status) ~ rx + obstruct + differ + extent + surg + nodes + tt(obstruct) + tt(differ) , data = colon.death, tt = function(x,t, . . .) x*log(t)).

```r
d.model.final = coxph(Surv(time, status) ~ rx + obstruct + differ + extent + surg + nodes + tt(obstruct)

summary(d.model.final)
```

```
## Call:
## coxph(formula = Surv(time, status) ~ rx + obstruct + differ +
##     extent + surg + nodes + tt(obstruct) + tt(differ), data = colon.death,
##     tt = function(x, t, ...) x * log(t))
##
##   n= 888, number of events= 430
##
##                   coef exp(coef)  se(coef)      z Pr(>|z|)
## rxLev        -0.067966  0.934292  0.114157 -0.595 0.551597
## rxLev+5FU    -0.356181  0.700346  0.121620 -2.929 0.003405 **
## obstruct      3.005889 20.204167  0.913252  3.291 0.000997 ***
## differ        3.324329 27.780355  0.794159  4.186 2.84e-05 ***
## extent        0.500140  1.648951  0.116581  4.290 1.79e-05 ***
## surg          0.249440  1.283307  0.106085  2.351 0.018707 *
## nodes         0.086542  1.090397  0.009386  9.221  < 2e-16 ***
## tt(obstruct) -0.428125  0.651730  0.140517 -3.047 0.002313 **
## tt(differ)   -0.485106  0.615632  0.120666 -4.020 5.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##              exp(coef) exp(-coef) lower .95 upper .95
## rxLev           0.9343    1.07033    0.7470    1.1686
## rxLev+5FU       0.7003    1.42787    0.5518    0.8889
## obstruct       20.2042    0.04949    3.3735  121.0047
## differ         27.7804    0.03600    5.8580  131.7429
## extent          1.6490    0.60645    1.3121    2.0722
## surg            1.2833    0.77924    1.0424    1.5799
## nodes           1.0904    0.91710    1.0705    1.1106
## tt(obstruct)    0.6517    1.53438    0.4948    0.8584
## tt(differ)      0.6156    1.62435    0.4860    0.7799
##
## Concordance= 0.671  (se = 0.013 )
## Likelihood ratio test= 137.1  on 9 df,   p=<2e-16
## Wald test            = 160.4  on 9 df,   p=<2e-16
## Score (logrank) test = 166.1  on 9 df,   p=<2e-16
```

5. Discussion

Our final model is given by: Surv(time, status) ~ rx + obstruct + differ + extent + surg + nodes + tt(obstruct) + tt(differ).

We can now observe whether treatments helps improve the survival rate in colon cancer patients from negative coefficients of different treatment groups.

Our result indicates that the hazard ratio for the group treated with rxLev is 0.934, corresponding to a 6.6% decrease in risk of death compared to the observation group. Since the p-value for the coefficient is 0.552

which is larger than 0.05, there is not significant evidence to indicate a survival difference between the group treated with rxLev and the observation group. Additionally, we are 95% confident that the true risk of death is between [0.747, 1.169]. Therefore, we conclude that the treatment rxLev is not effective on improving the survival probability in colon cancer patients.

On the other hand, our result indicates that the hazard ratio for the group treated with rxLev+5FU is 0.700, corresponding to a 30% decrease in risk of death compared to the observation group. Since the p-value for the coefficient is 0.003 which is samller than 0.05, there is significant evidence to indicate a survival difference between the group treated with rxLev+5FU and the observation group. Additionally, we are 95% confident that the true risk of death is between [0.552, 0.889]. Therefore, we conclude that the treatment rxLev+5FU is effective on improving the survival probability in colon cancer patients.

Therefore, we conclude that the treatment Levamisole+5-FU is effective as an adjuvant Chemotherapy on improving the survival rate in colon cancer patients.