Infant Infection Timing Models

```
library(ggplot2)
library(devtools)
library(gridExtra)
library(ggpubr)
library(data.table)
library(Metrics)
library(tidyverse)
library(GGally)

source("Infant_functions.R")
source("Infant_predict.R")
source("Infant_plotting.R")

knitr::opts_chunk$set(fig.width=16, fig.height=8)
```

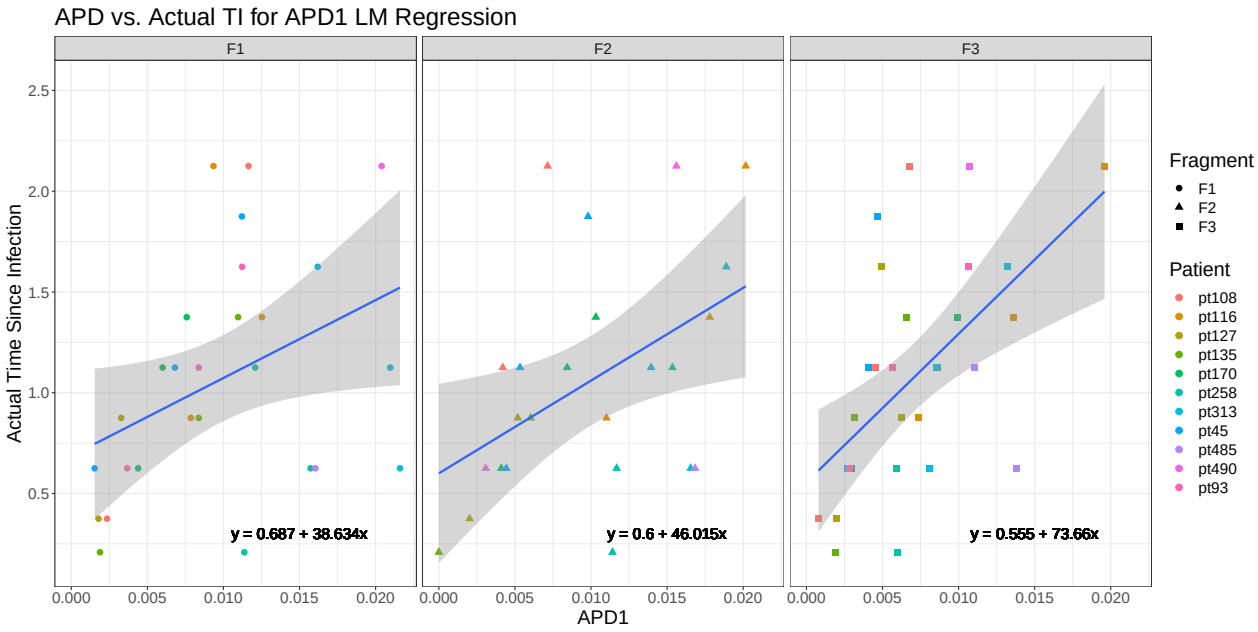
```
# Set working directory
setwd("/Users/magdalenarussell/Documents/Matsen_group/infection-timing/Infant_Model")

# Load existing data
training_data = read.csv("../_ignore/AllRunsAvg.csv")
testing_data = read.csv("../_ignore/AllRunsAvg.csv")
```

Note: for all of the following analyses, I have assigned the actual time since infection for all patients to be the midpoint of the third trimester (i.e. we are assigning birth to be time zero, so the actual infection time here is -0.125 years.)

The following plots show the models created for the data using an APD cutoff of 0.01. First you will see the actual time since infection versus APD regression (using the indicated new modelling strategy). You will see the model statistics below these plots (i.e. model equation coefficients (also show on the plots in the form of an equation) and p values for each coefficient). Second you will see the relationship between the estimated time since infection and the actual time since infection. The perfect correlation (show as the blue spotted line) depicts the y = x line where actual time since infection = estimated time since infection. For these plots, I am showing a LM regression through the points with a confidence interval to show the trend of these estimated times. However, these estimated times were calculated using the indicated new modelling strategy (not always LM regression). Lastly, you will see, again, a plot comparing the estimated time since infection and the actual time since infection. Here, the plots show the trends per individual.

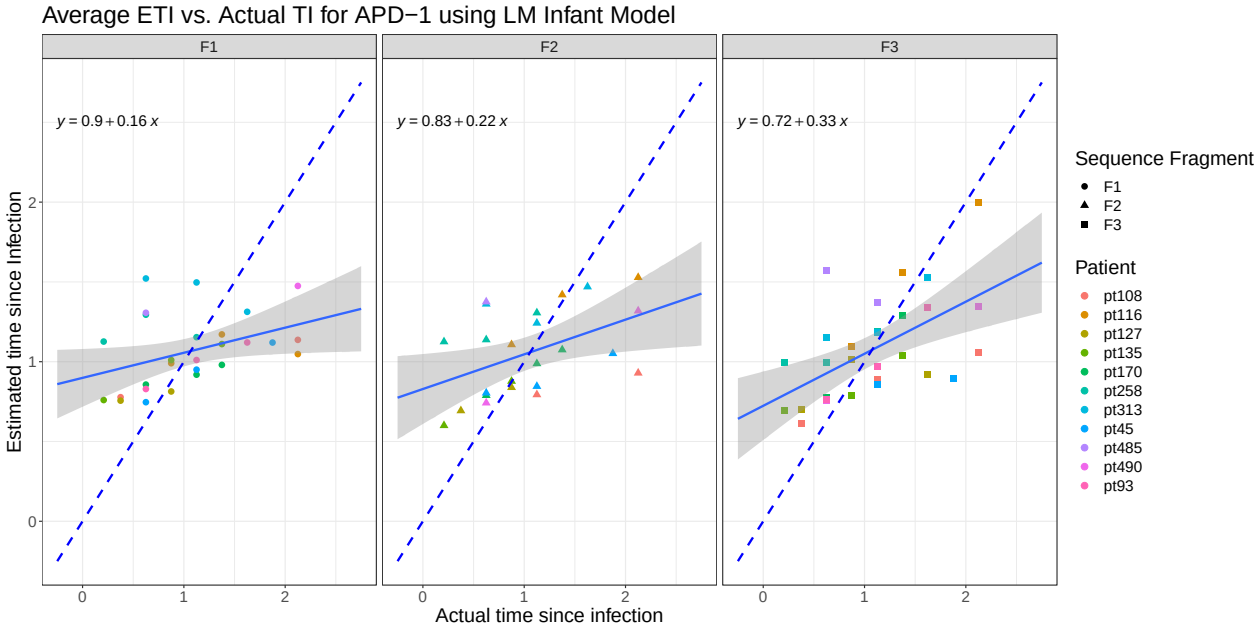
```
for (ty in c("LM", "LM_origin", "LAD", "LAD_origin", "LM_GEE")){
    together = compile_data(train_data = training_data, new_data = testing_data, apd = 1, type = ty)
    plot_APD_TI_regression(together, apd = 1, type = ty)
    show_model_statistics(apd = 1, type= ty)
    plot_ETI_TI_regression(together, apd = 1, type = ty)
}
```

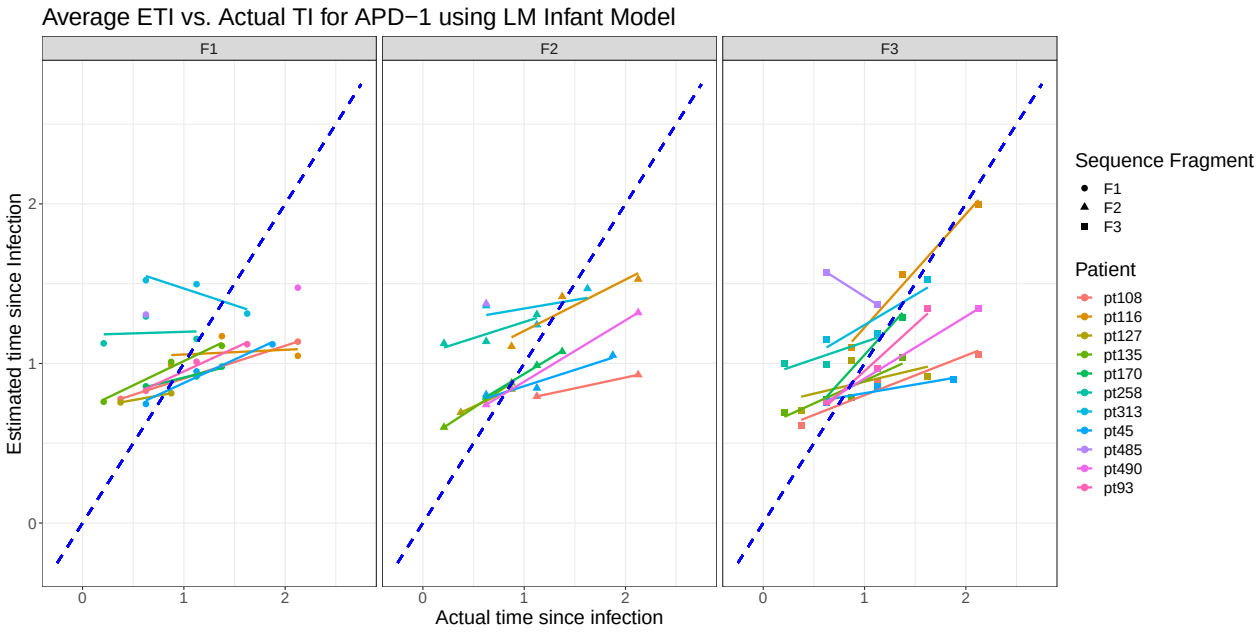
```
## Saving 16 x 8 in image
```



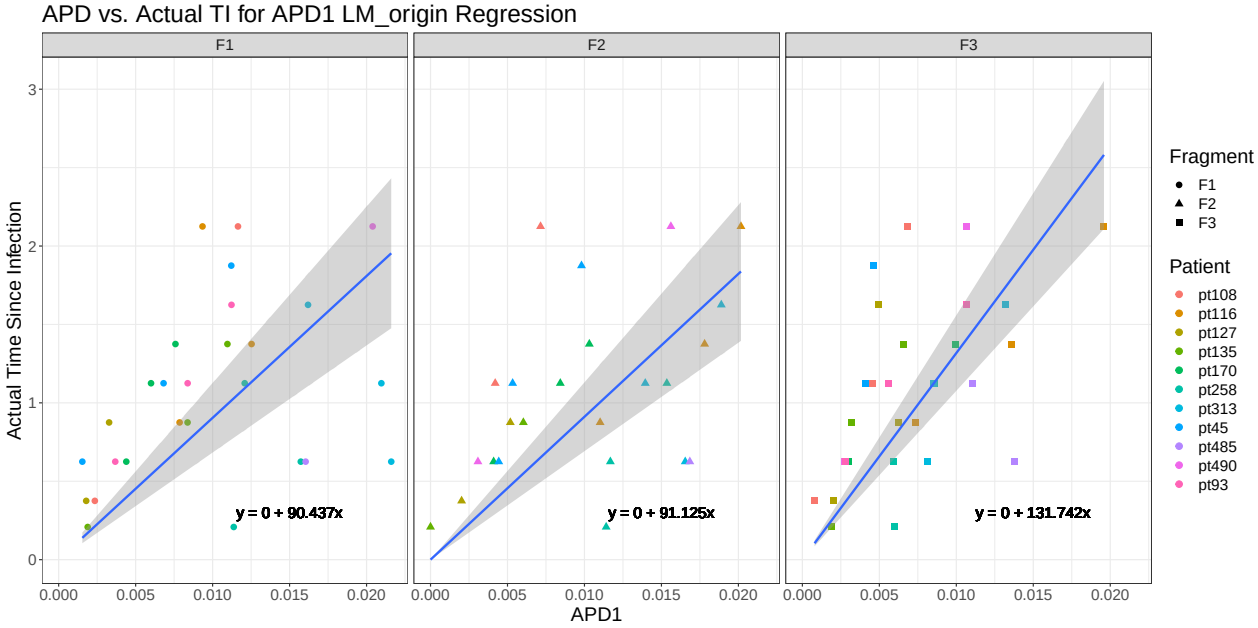

```
## [1] "LM Model for APD1 and Fragment F1 Summary Statistics:"
##                Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)  0.6868725   0.204124 3.364976 0.002473165
## avg_apd1    38.6337771  17.886224 2.159974 0.040561325
## [1] "LM Model for APD1 and Fragment F2 Summary Statistics:"
##                Estimate Std. Error  t value   Pr(>|t|)
## (Intercept)  0.6001113  0.2144139 2.798844 0.01046298
## avg_apd1    46.0145096 18.6146676 2.471949 0.02165105
## [1] "LM Model for APD1 and Fragment F3 Summary Statistics:"
##                Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)  0.5554609  0.1612772 3.444138 0.0017649364
## avg_apd1    73.6599533 19.6573417 3.747198 0.0007911118
```

```
## Saving 16 x 8 in image
```



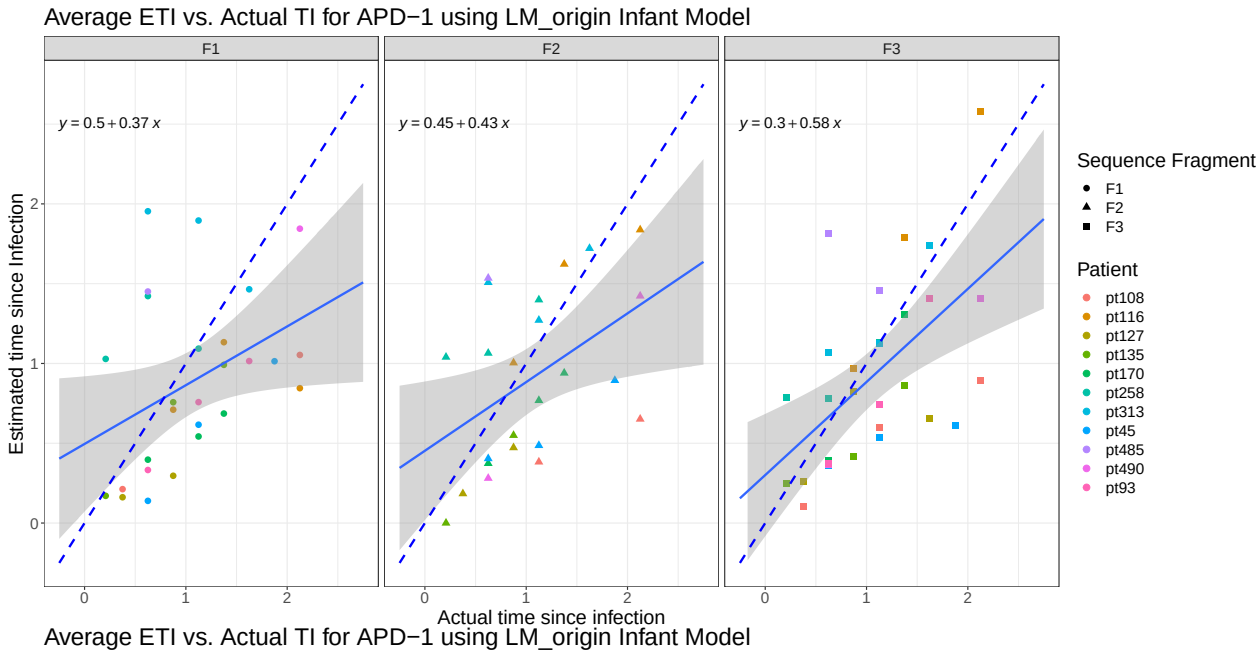


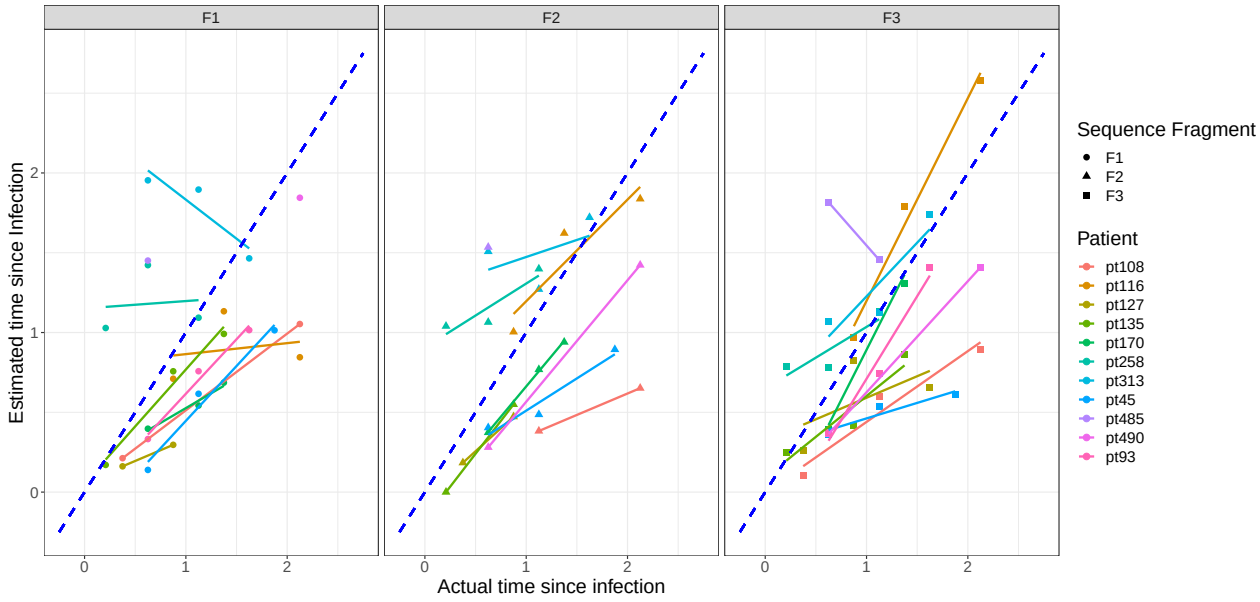

```
## Saving 16 x 8 in image
```



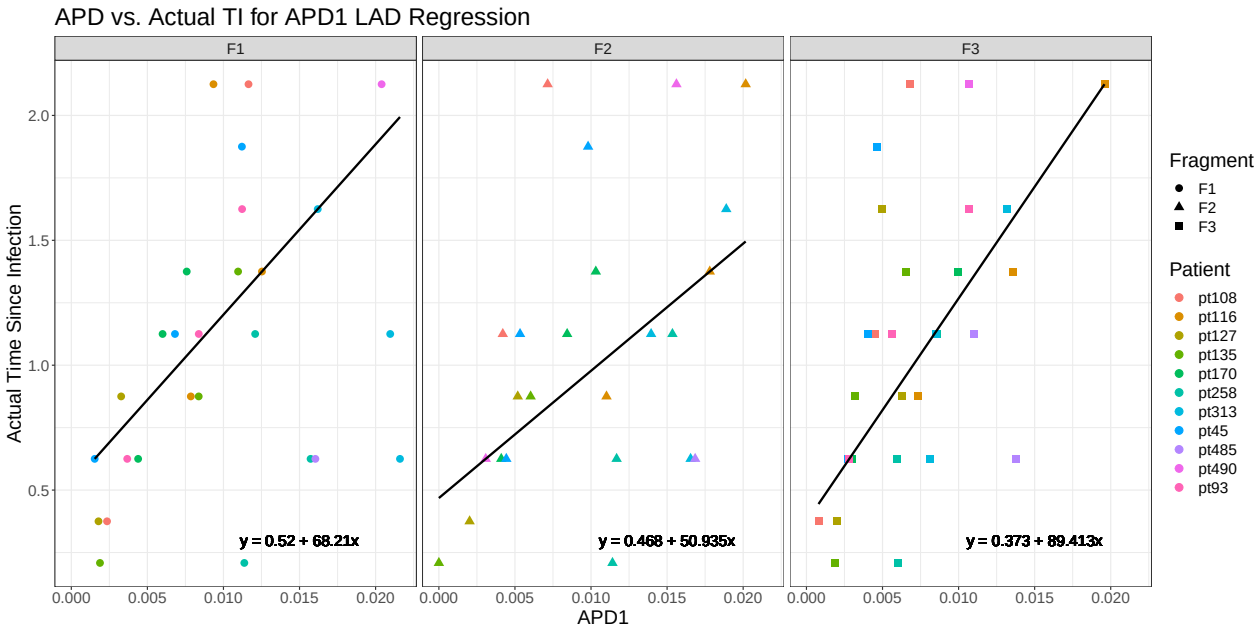

```
## [1] "LM_origin Model for APD1 and Fragment F1 Summary Statistics:"
##          Estimate Std. Error  t value     Pr(>|t|)
## avg_apd1 90.43658   10.76303 8.402518 6.960345e-09
## [1] "LM_origin Model for APD1 and Fragment F2 Summary Statistics:"
##          Estimate Std. Error  t value     Pr(>|t|)
## avg_apd1 91.12475   10.60674 8.591209 1.235489e-08
## [1] "LM_origin Model for APD1 and Fragment F3 Summary Statistics:"
##          Estimate Std. Error  t value     Pr(>|t|)
## avg_apd1 131.7423   11.78763 11.17632 3.227911e-12
```

```
## Saving 16 x 8 in image
```



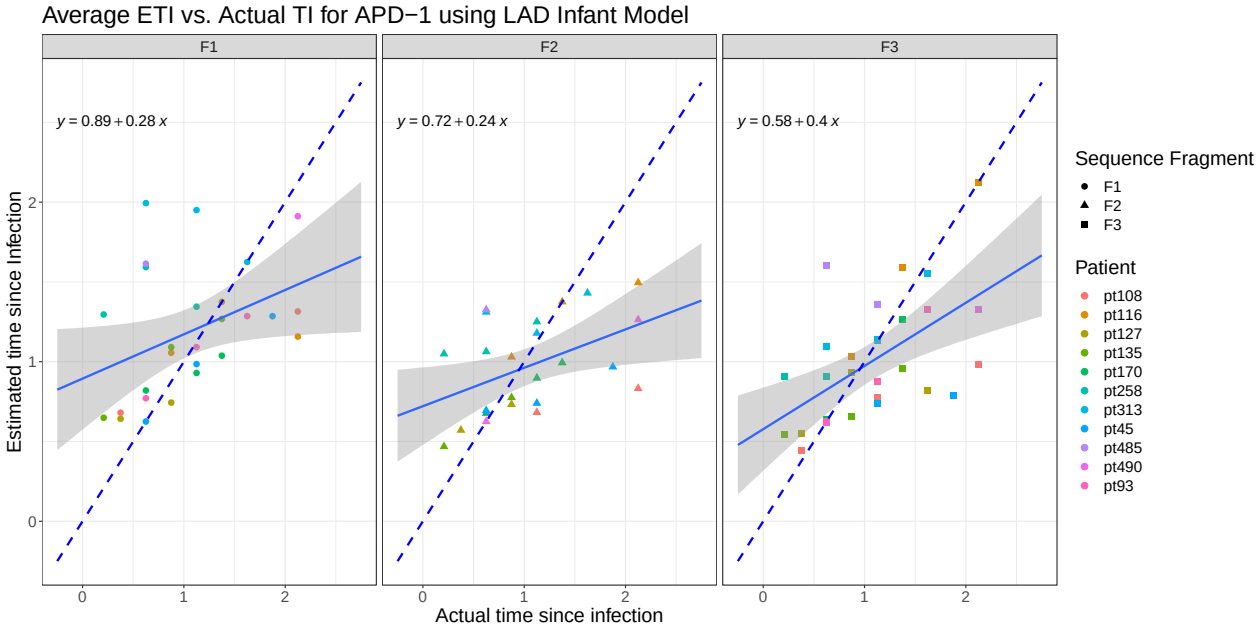


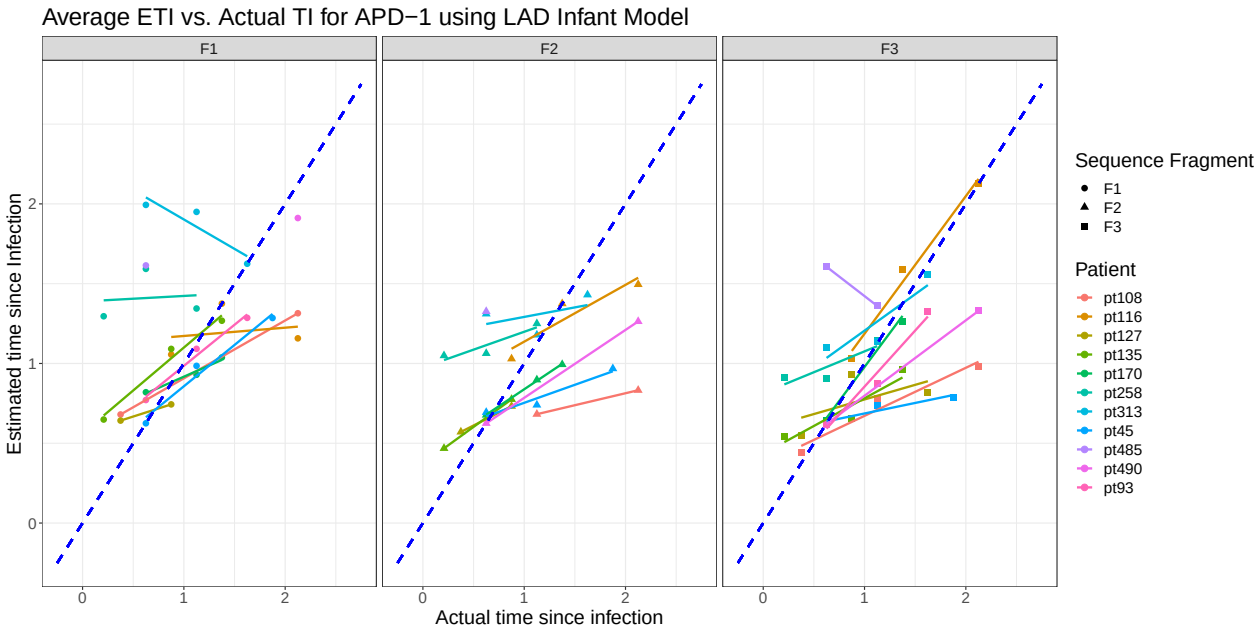

```
## Saving 16 x 8 in image
```



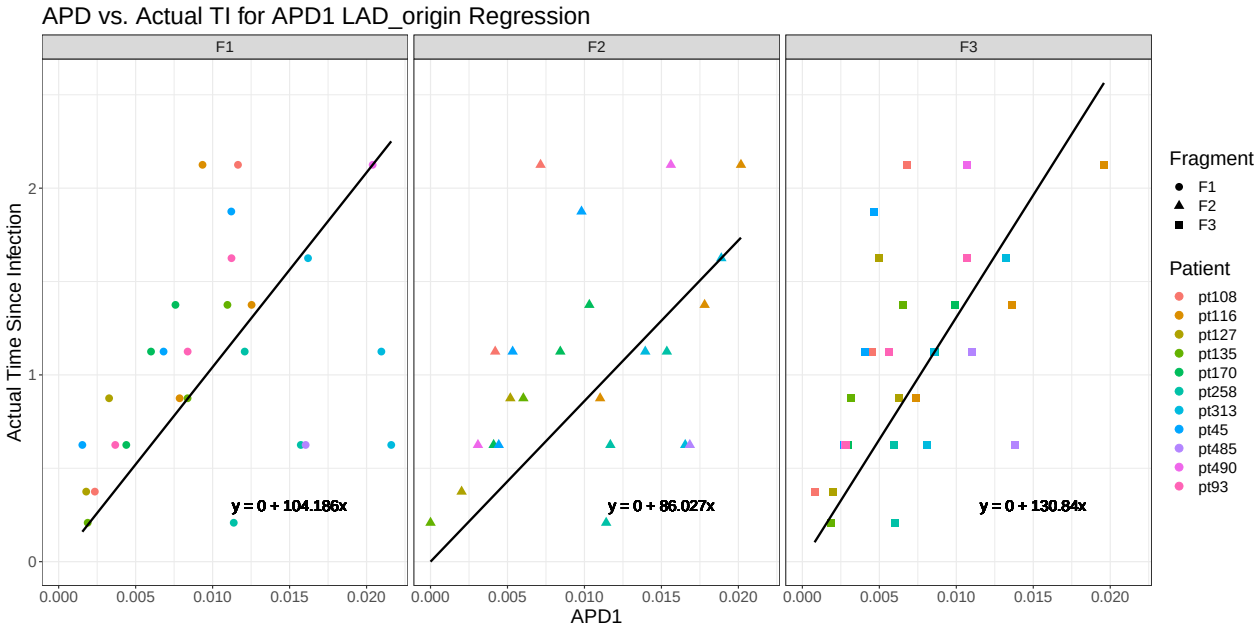

```
## [1] "LAD Model for APD1 and Fragment F1 Summary Statistics:"
##                Estimate  Std.Error  Z value    p-value
## (Intercept)  0.5201617  0.3100915 1.677446 0.09345531
## avg_apd1    68.2096252 27.1715490 2.510333 0.01206175
## [1] "LAD Model for APD1 and Fragment F2 Summary Statistics:"
##                Estimate  Std.Error  Z value    p-value
## (Intercept)  0.4681131  0.3108028 1.506142 0.13203072
## avg_apd1    50.9345741 26.9828060 1.887668 0.05907056
## [1] "LAD Model for APD1 and Fragment F3 Summary Statistics:"
##                Estimate  Std.Error  Z value      p-value
## (Intercept)  0.3732041  0.2183118 1.709500 0.0873582913
## avg_apd1    89.4130783 26.6090256 3.360254 0.0007787091
```

```
## Saving 16 x 8 in image
```



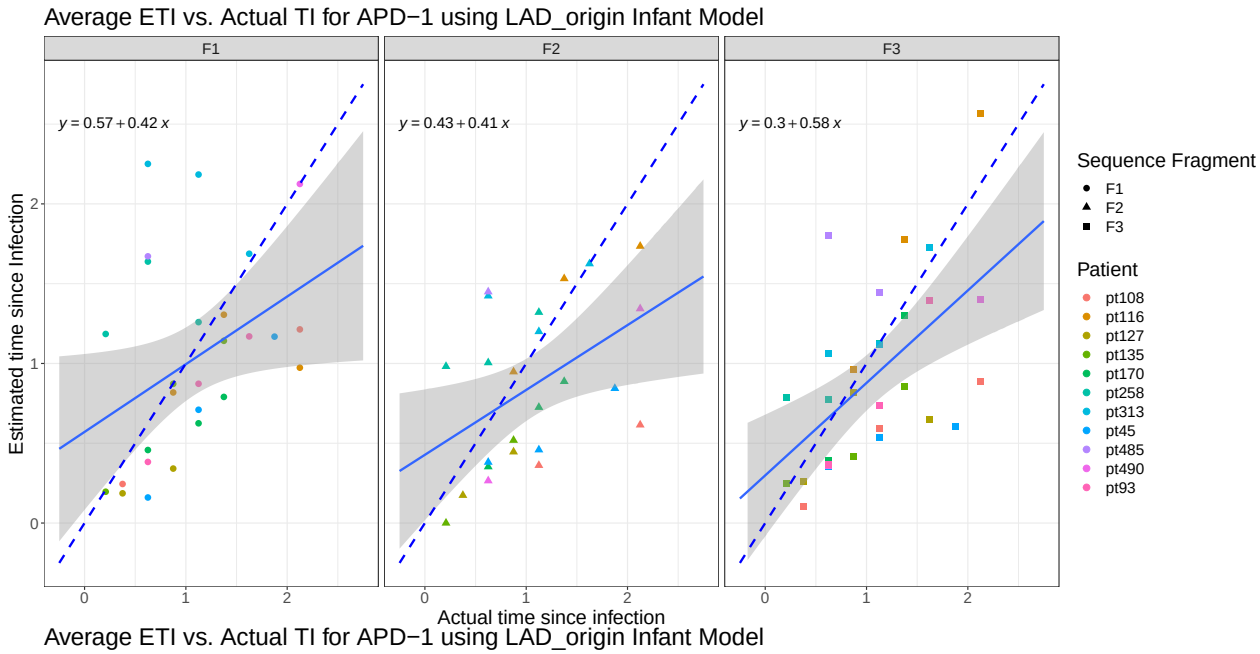


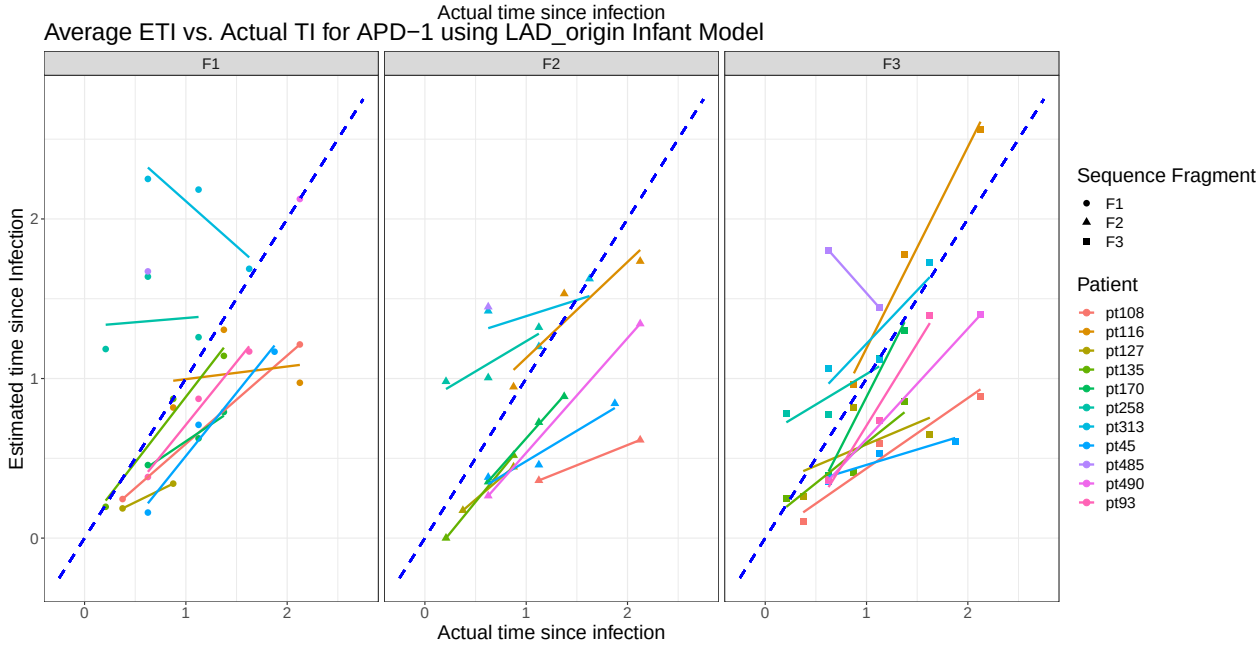

```
## Saving 16 x 8 in image
```



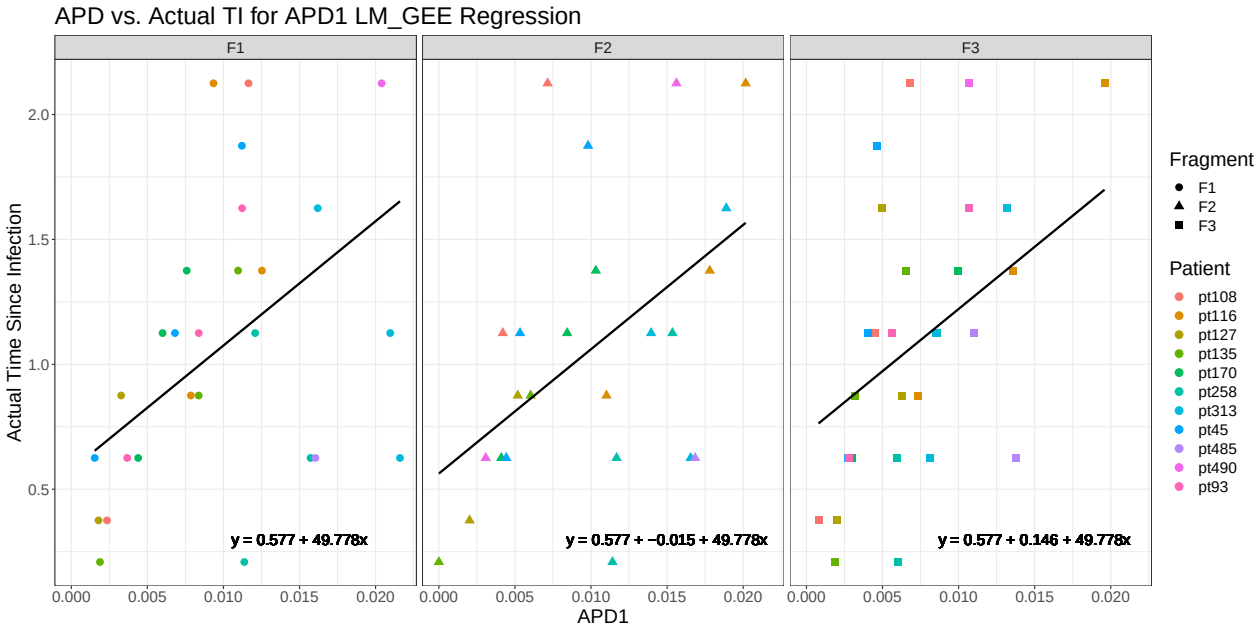

```
## [1] "LAD_origin Model for APD1 and Fragment F1 Summary Statistics:"
##          Estimate Std.Error  Z value    p-value
## avg_apd1 104.1856  16.23346 6.417952 1.3812e-10
## [1] "LAD_origin Model for APD1 and Fragment F2 Summary Statistics:"
##          Estimate Std.Error  Z value      p-value
## avg_apd1 86.02727  16.80154 5.120203 3.052072e-07
## [1] "LAD_origin Model for APD1 and Fragment F3 Summary Statistics:"
##          Estimate Std.Error  Z value      p-value
## avg_apd1 130.8403  17.24915 7.585319 3.316704e-14
```

```
## Saving 16 x 8 in image
```



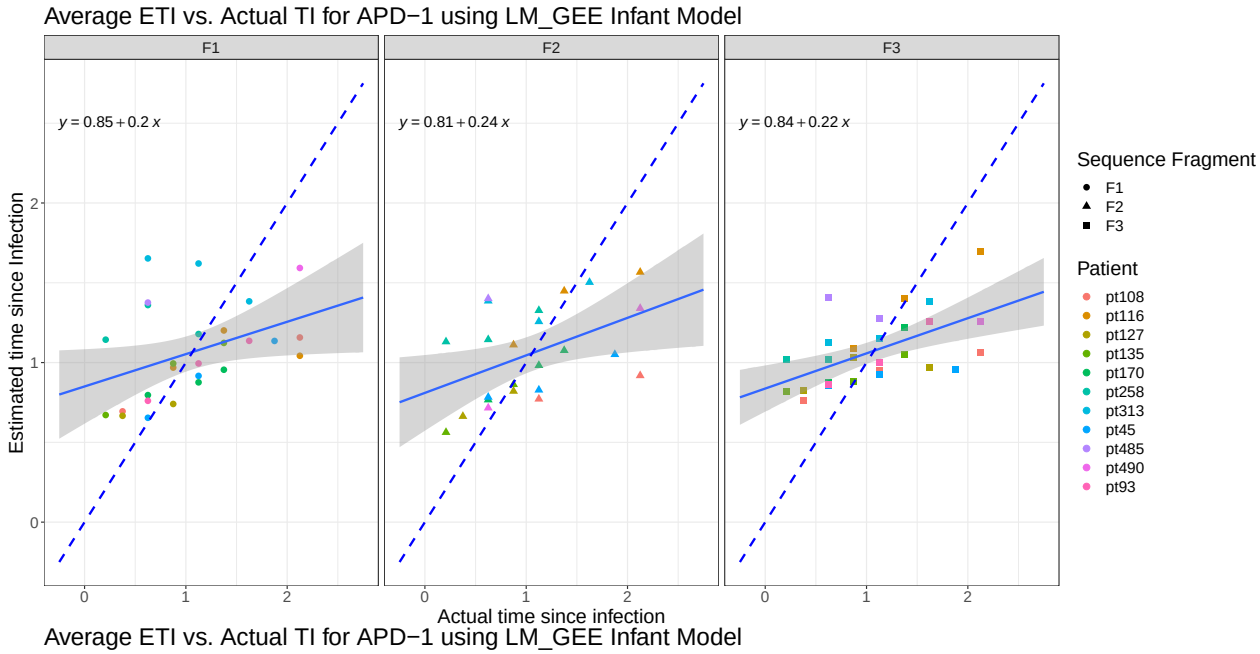


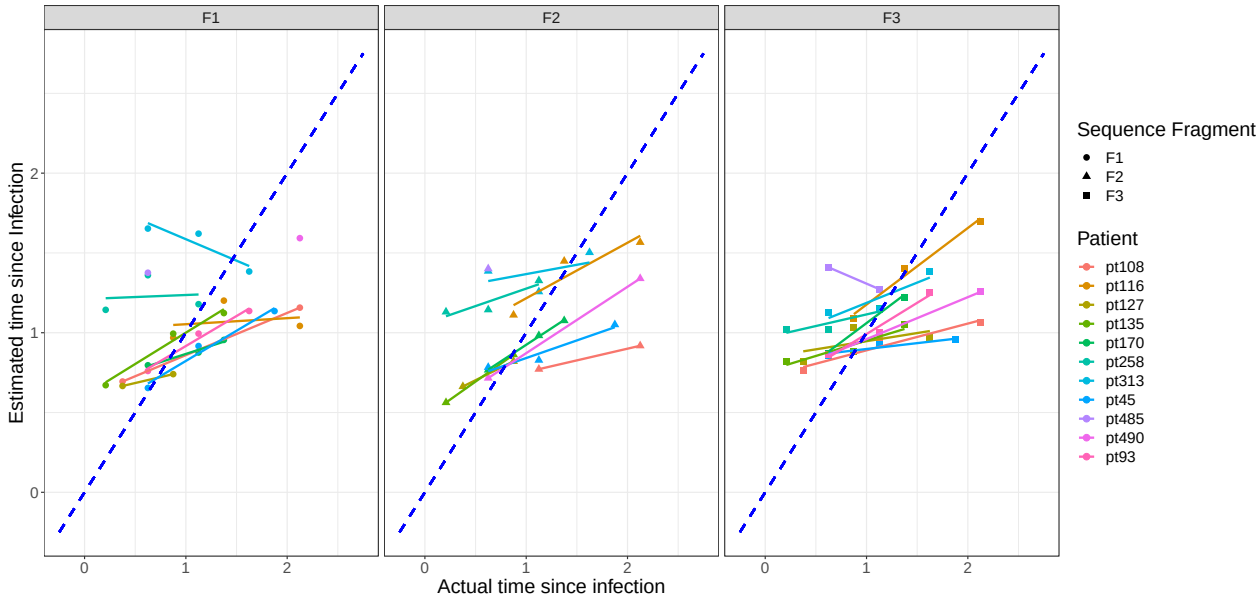

```
## Saving 16 x 8 in image
```




```
## [1] "LM_GEE Model for APD1 Summary Statistics:"
##                         Estimate Naive S.E.    Naive z Robust S.E.
## (Intercept)          0.57741090  0.1429107  4.0403616  0.11171322
## avg_apd1            49.77758358 10.6470437  4.6752493 15.16451484
## factor(fragment)F2 -0.01482986  0.1419875 -0.1044448  0.05995904
## factor(fragment)F3  0.14614874  0.1364798  1.0708449  0.05297275
##                      Robust z
## (Intercept)         5.1686893
## avg_apd1            3.2825042
## factor(fragment)F2 -0.2473333
## factor(fragment)F3  2.7589416
##        (Intercept)           avg_apd1 factor(fragment)F2
##       2.357415e-07       1.028895e-03       8.046503e-01
## factor(fragment)F3
##       5.798890e-03
```

```
## Saving 16 x 8 in image
```

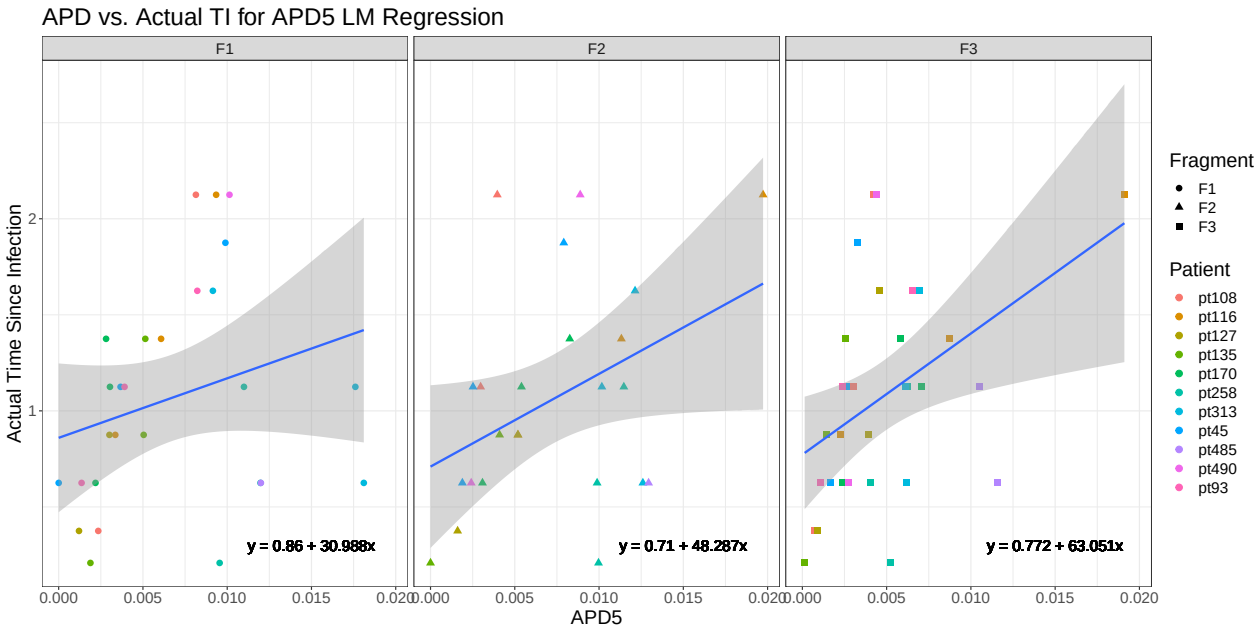




Here you will see the same plotting scheme except, now, the following plots show the models created for the data using an APD cutoff of 0.05.

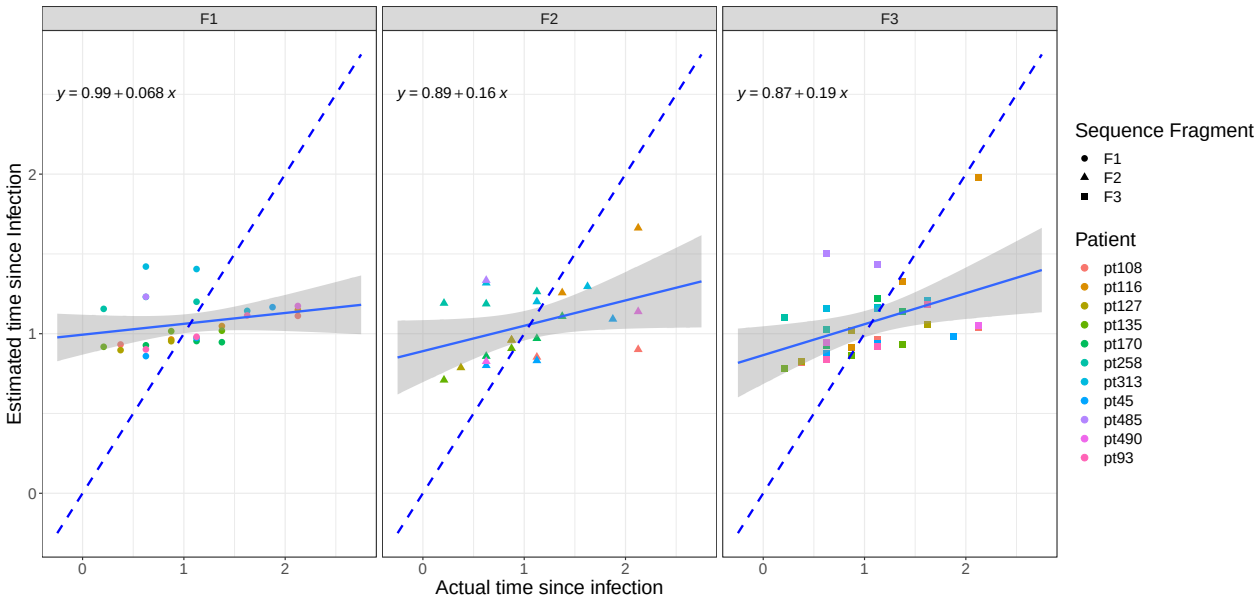
```
for (ty in c("LM", "LM_origin", "LAD", "LAD_origin", "LM_GEE")){
    together = compile_data(train_data = training_data, new_data = testing_data, apd = 5, type = ty)
    plot_APD_TI_regression(together, apd = 5, type = ty)
    show_model_statistics(apd = 5, type= ty)
    plot_ETI_TI_regression(together, apd = 5, type = ty)
}
```

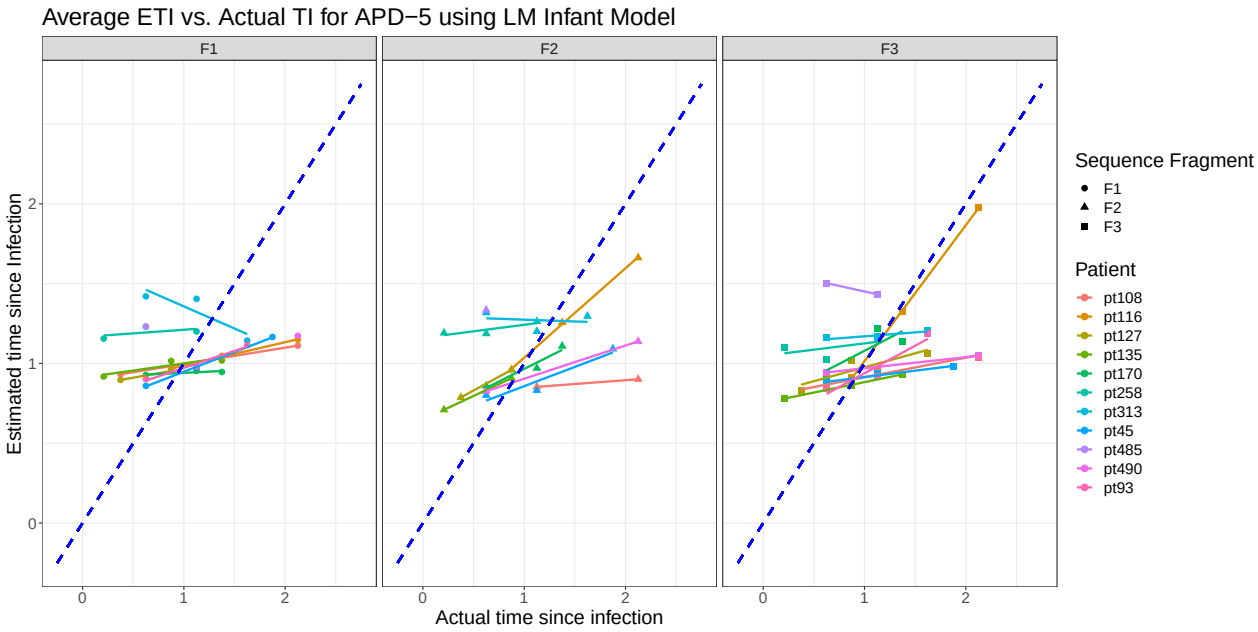
```
## Saving 16 x 8 in image
```



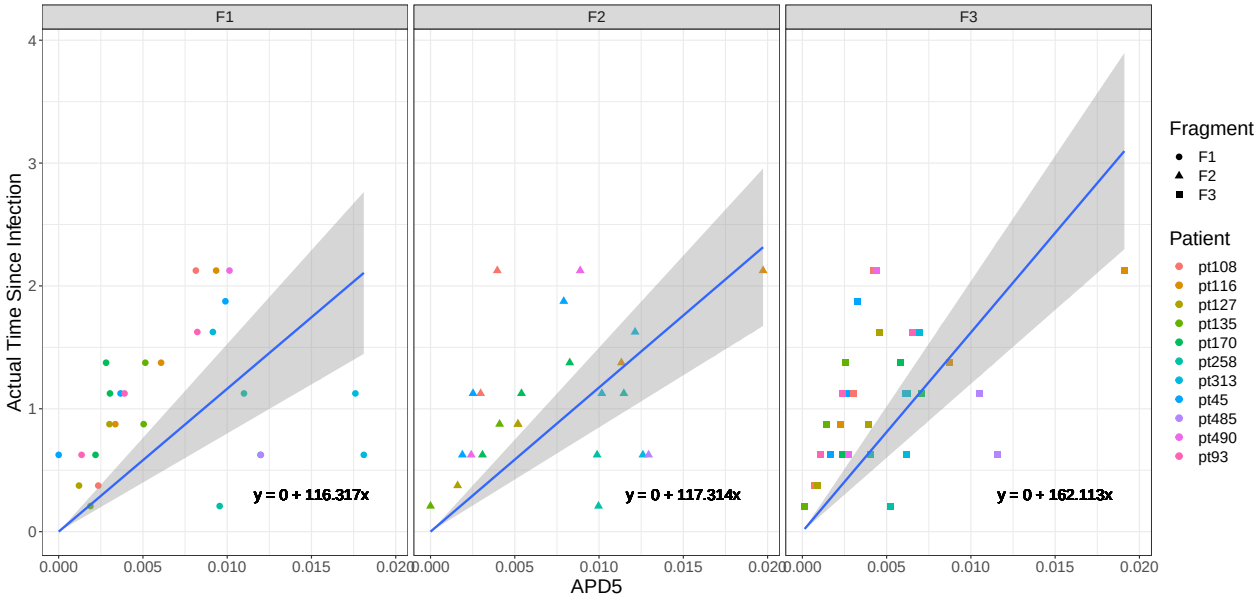

```
## [1] "LM Model for APD5 and Fragment F1 Summary Statistics:"
##                Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)  0.8595803  0.1880881 4.570094 0.0001135604
## avg_apd5    30.9880608 22.9409226 1.350777 0.1888674317
## [1] "LM Model for APD5 and Fragment F2 Summary Statistics:"
##                Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)  0.7097526  0.2042171 3.475481 0.002146657
## avg_apd5    48.2872289 23.6799716 2.039159 0.053625692
## [1] "LM Model for APD5 and Fragment F3 Summary Statistics:"
##                Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)  0.7723701  0.1457767 5.298309 1.107161e-05
## avg_apd5    63.0511604 23.8721359 2.641203 1.316952e-02
```

```
## Saving 16 x 8 in image
```



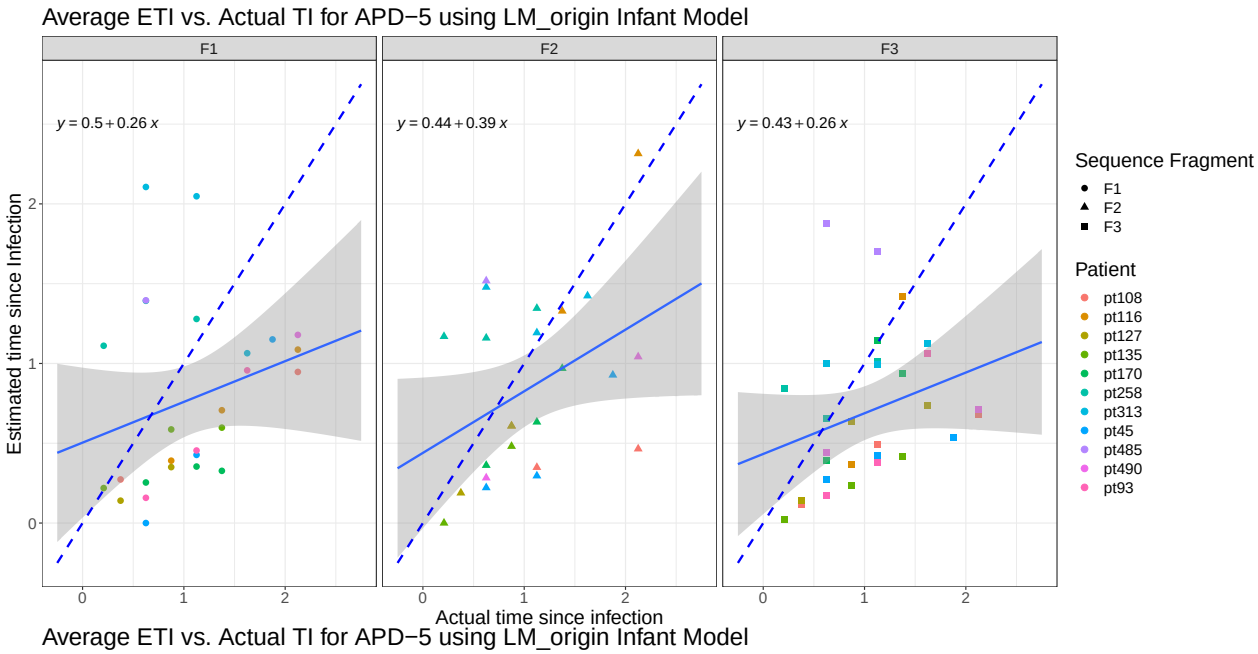


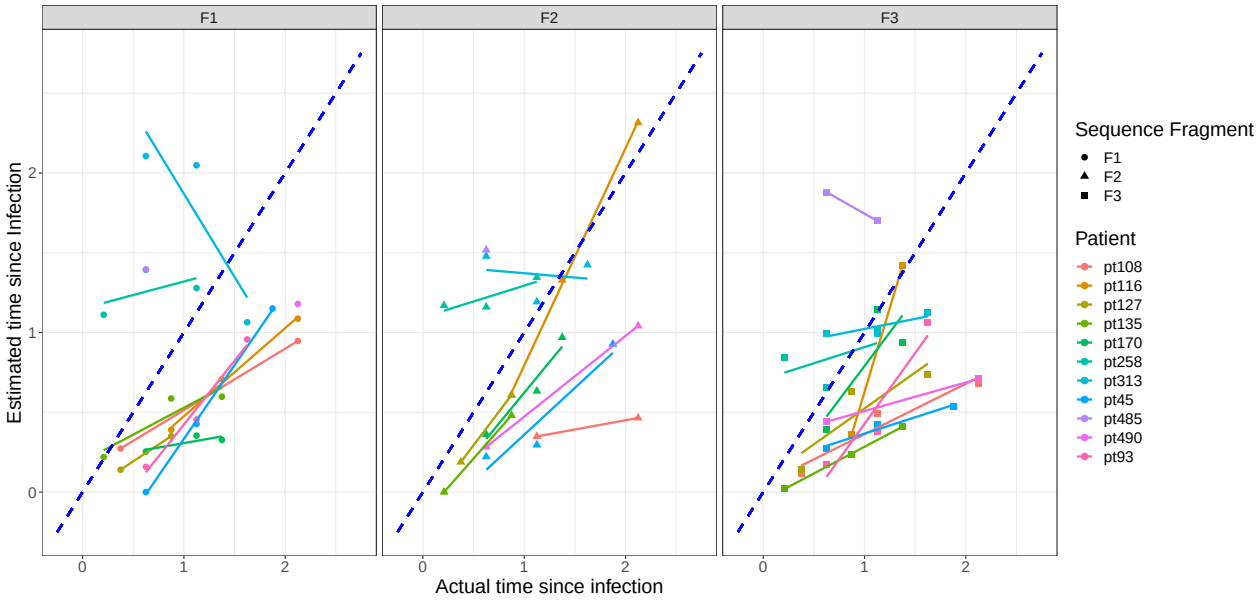

```
## Saving 16 x 8 in image
```



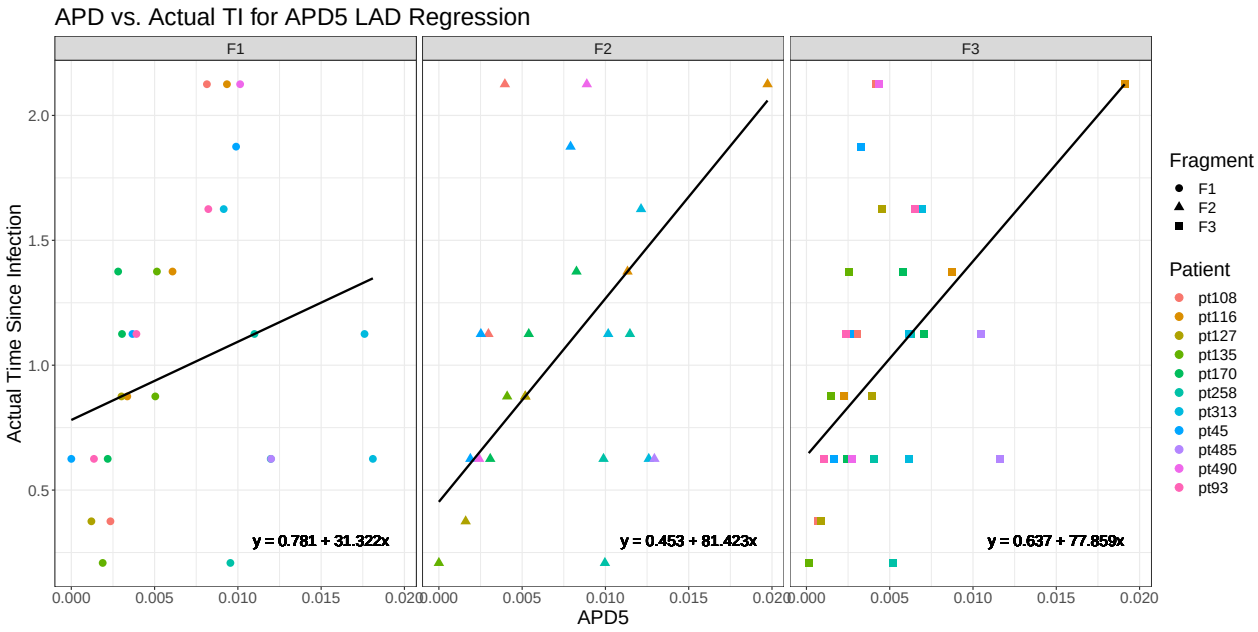

```
## [1] "LM_origin Model for APD5 and Fragment F1 Summary Statistics:"
##          Estimate Std. Error  t value     Pr(>|t|)
## avg_apd5 116.3167   17.70788 6.568642 5.772489e-07
## [1] "LM_origin Model for APD5 and Fragment F2 Summary Statistics:"
##          Estimate Std. Error  t value     Pr(>|t|)
## avg_apd5 117.3145   15.69606 7.474136 1.349512e-07
## [1] "LM_origin Model for APD5 and Fragment F3 Summary Statistics:"
##          Estimate Std. Error  t value    Pr(>|t|)
## avg_apd5 162.1126   20.47231 7.918627 7.73886e-09
```

```
## Saving 16 x 8 in image
```



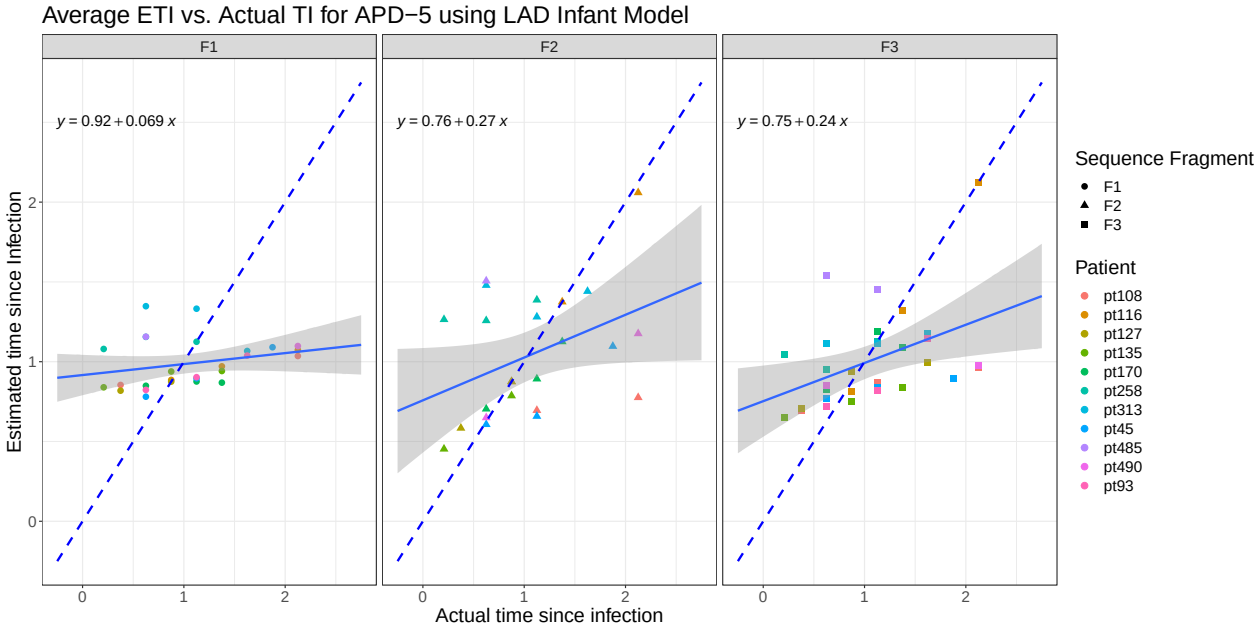


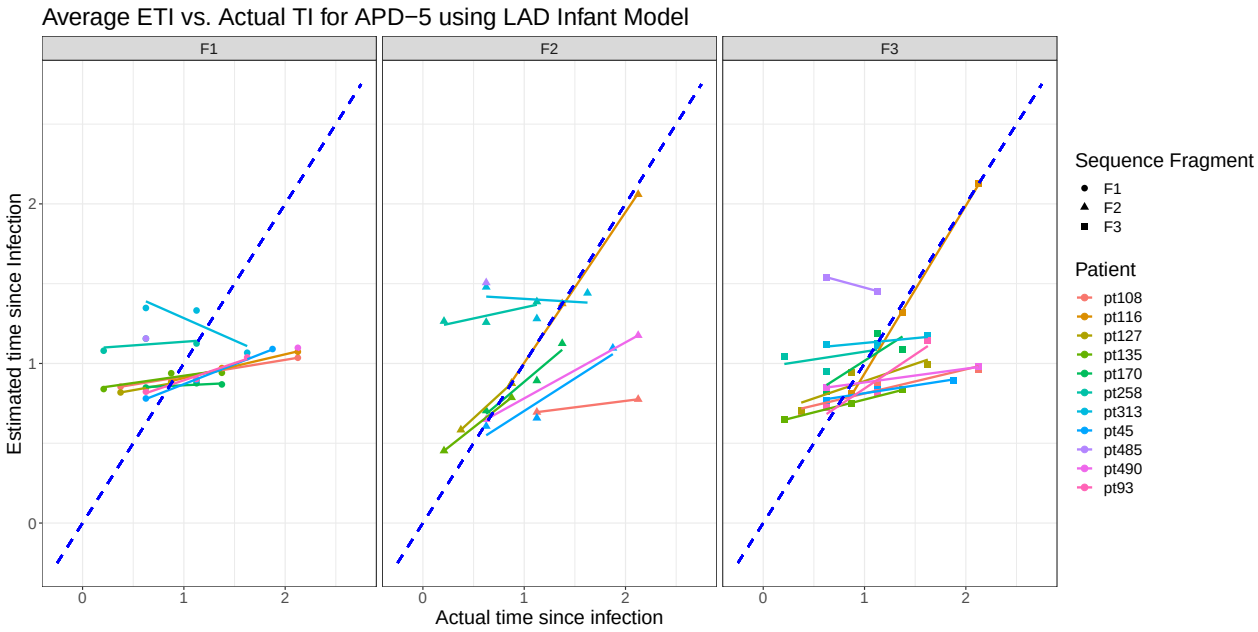

```
## Saving 16 x 8 in image
```



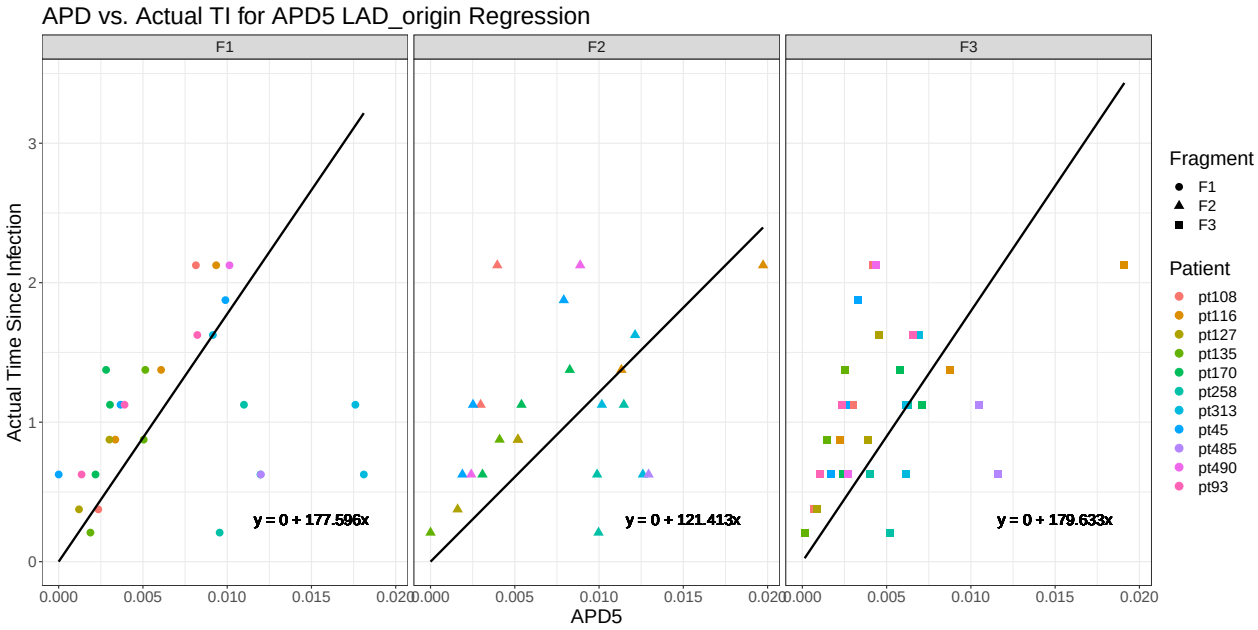

```
## [1] "LAD Model for APD5 and Fragment F1 Summary Statistics:"
##                Estimate  Std.Error   Z value     p-value
## (Intercept)  0.7806487  0.2995866 2.6057532 0.009167249
## avg_apd5    31.3216953 36.5402881 0.8571825 0.391344035
## [1] "LAD Model for APD5 and Fragment F2 Summary Statistics:"
##                Estimate  Std.Error  Z value    p-value
## (Intercept)  0.4528843  0.2877176 1.574058 0.11547401
## avg_apd5    81.4233551 33.3622682 2.440582 0.01466361
## [1] "LAD Model for APD5 and Fragment F3 Summary Statistics:"
##                Estimate  Std.Error  Z value     p-value
## (Intercept)  0.6373845  0.2151962 2.961876 0.003057712
## avg_apd5    77.8592682 35.2401532 2.209391 0.027147475
```

```
## Saving 16 x 8 in image
```



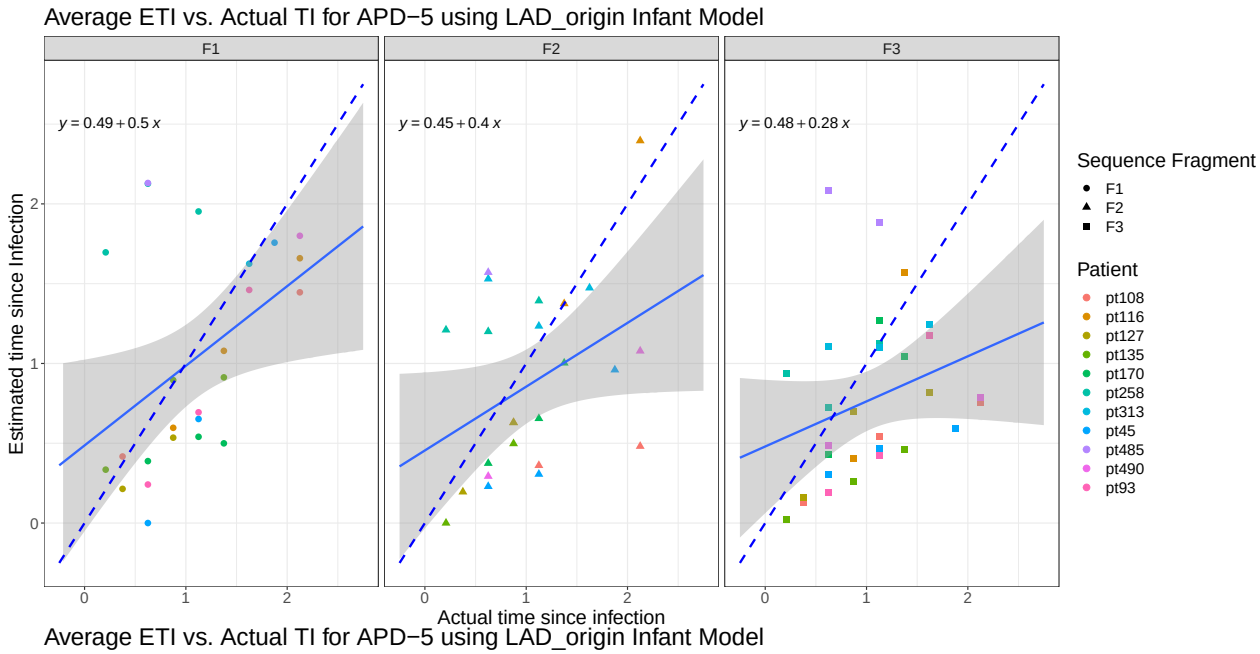


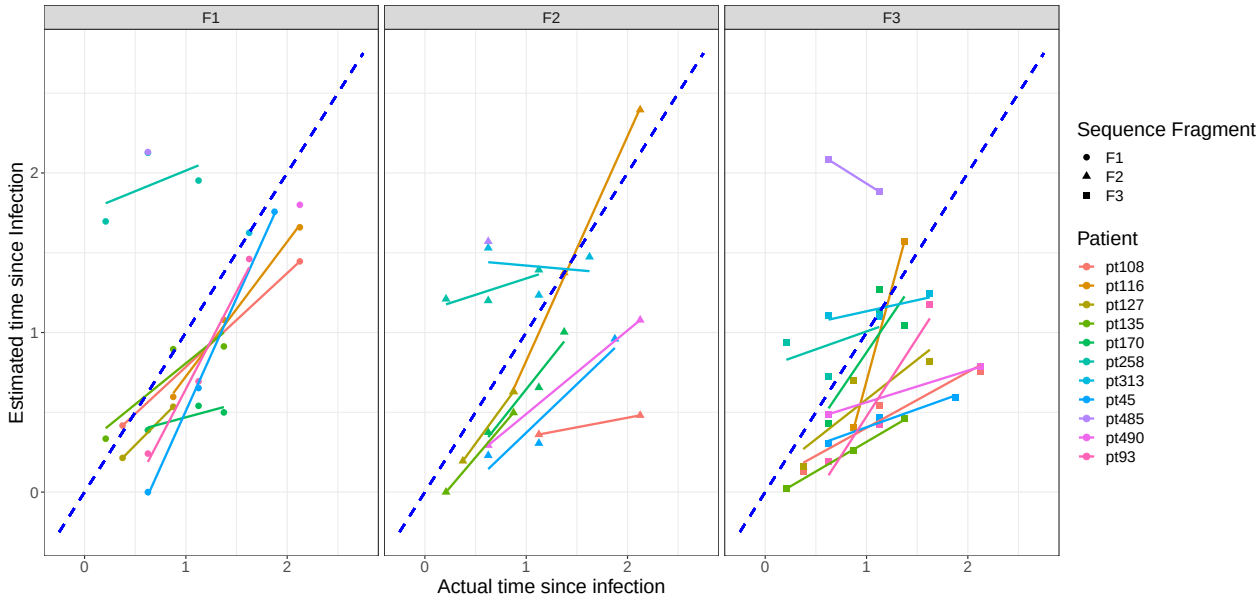

```
## Saving 16 x 8 in image
```



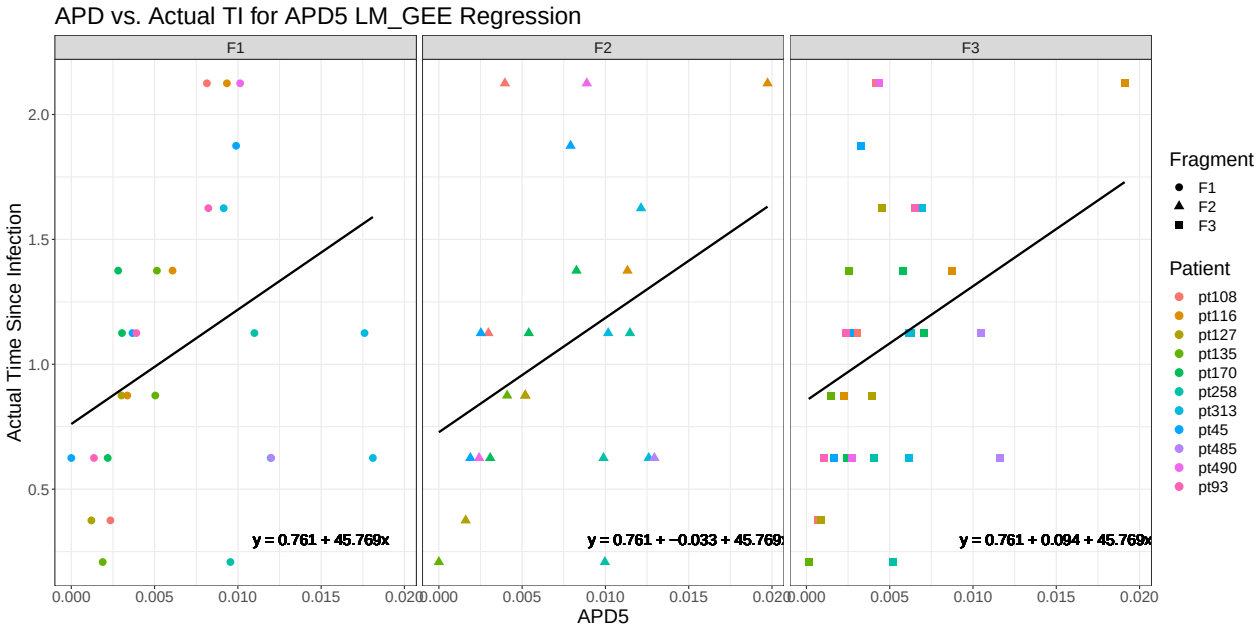

```
## [1] "LAD_origin Model for APD5 and Fragment F1 Summary Statistics:"
##          Estimate Std.Error  Z value      p-value
## avg_apd5 177.5964   29.5629 6.007409 1.885118e-09
## [1] "LAD_origin Model for APD5 and Fragment F2 Summary Statistics:"
##          Estimate Std.Error  Z value      p-value
## avg_apd5 121.4133  24.63826 4.927837 8.314501e-07
## [1] "LAD_origin Model for APD5 and Fragment F3 Summary Statistics:"
##          Estimate Std.Error  Z value      p-value
## avg_apd5 179.6327  32.27193 5.566221 2.603229e-08
```

```
## Saving 16 x 8 in image
```



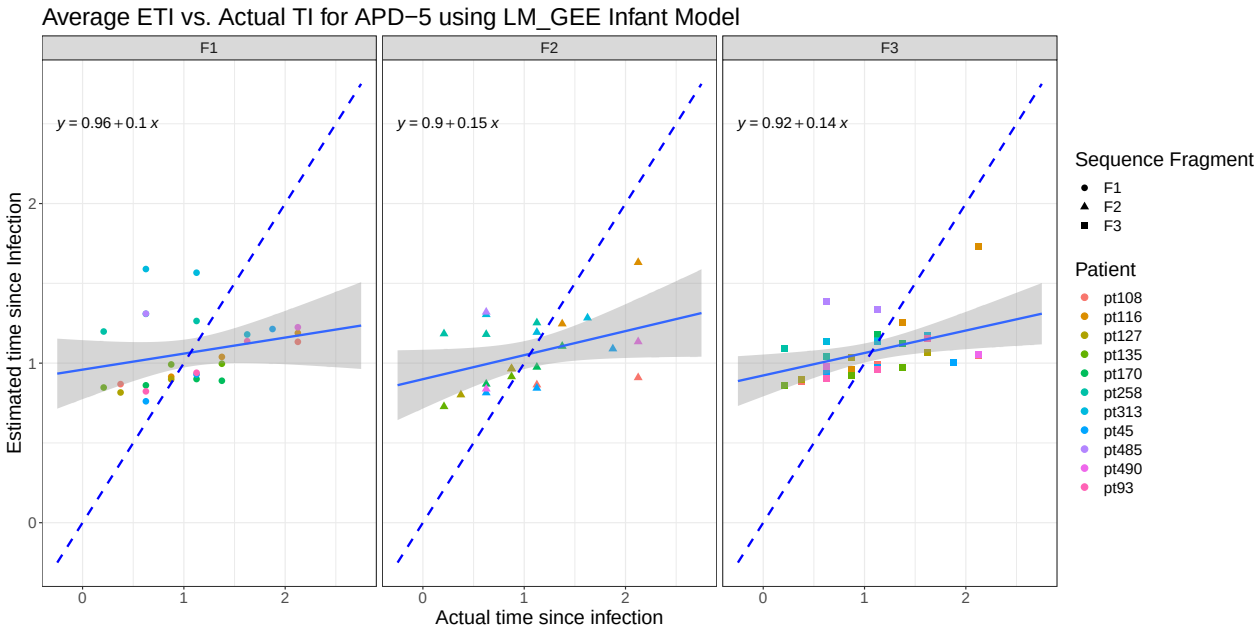


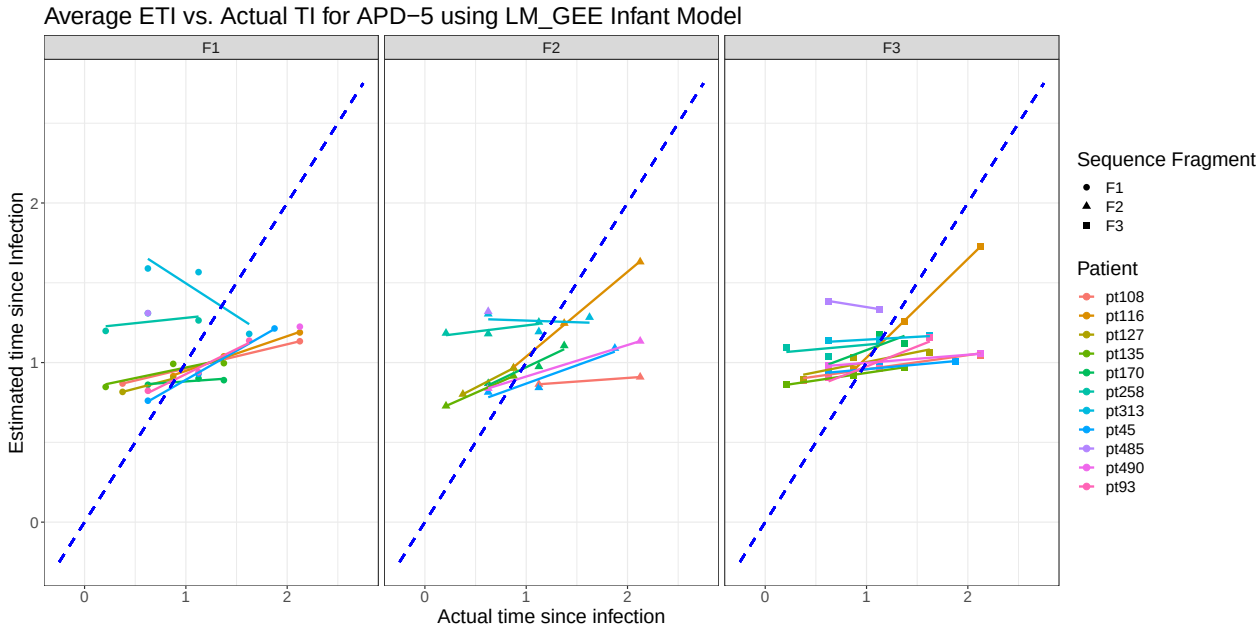

```
## Saving 16 x 8 in image
```




```
## [1] "LM_GEE Model for APD5 Summary Statistics:"
##                          Estimate Naive S.E.    Naive z Robust S.E.  Robust z
## (Intercept)           0.76095229  0.1362235  5.5860586  0.09444327  8.057242
## avg_apd5             45.76861490 13.3981106  3.4160500 20.27227328  2.257695
## factor(fragment)F2   -0.03298185  0.1500156 -0.2198561  0.06374178 -0.517429
## factor(fragment)F3    0.09407500  0.1428497  0.6585594  0.04668233  2.015216
##         (Intercept)           avg_apd5 factor(fragment)F2
##        8.881784e-16       2.396467e-02       6.048567e-01
## factor(fragment)F3
##        4.388199e-02
```

```
## Saving 16 x 8 in image
```

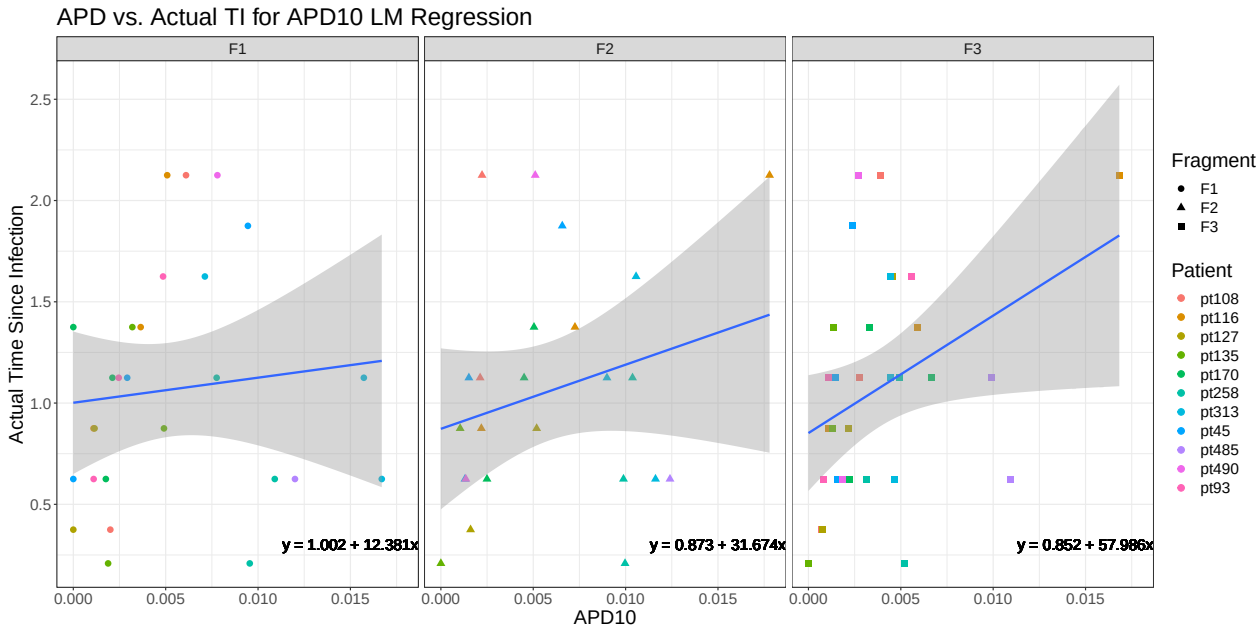




Here, again, you will see the same plotting scheme except, now, the following plots show the models created for the data using an APD cutoff of 0.1.

```
for (ty in c("LM", "LM_origin", "LAD", "LAD_origin", "LM_GEE")){
    together = compile_data(train_data = training_data, new_data = testing_data, apd = 10, type = ty)
    plot_APD_TI_regression(together, apd = 10, type = ty)
    show_model_statistics(apd = 10, type= ty)
    plot_ETI_TI_regression(together, apd = 10, type = ty)
}
```

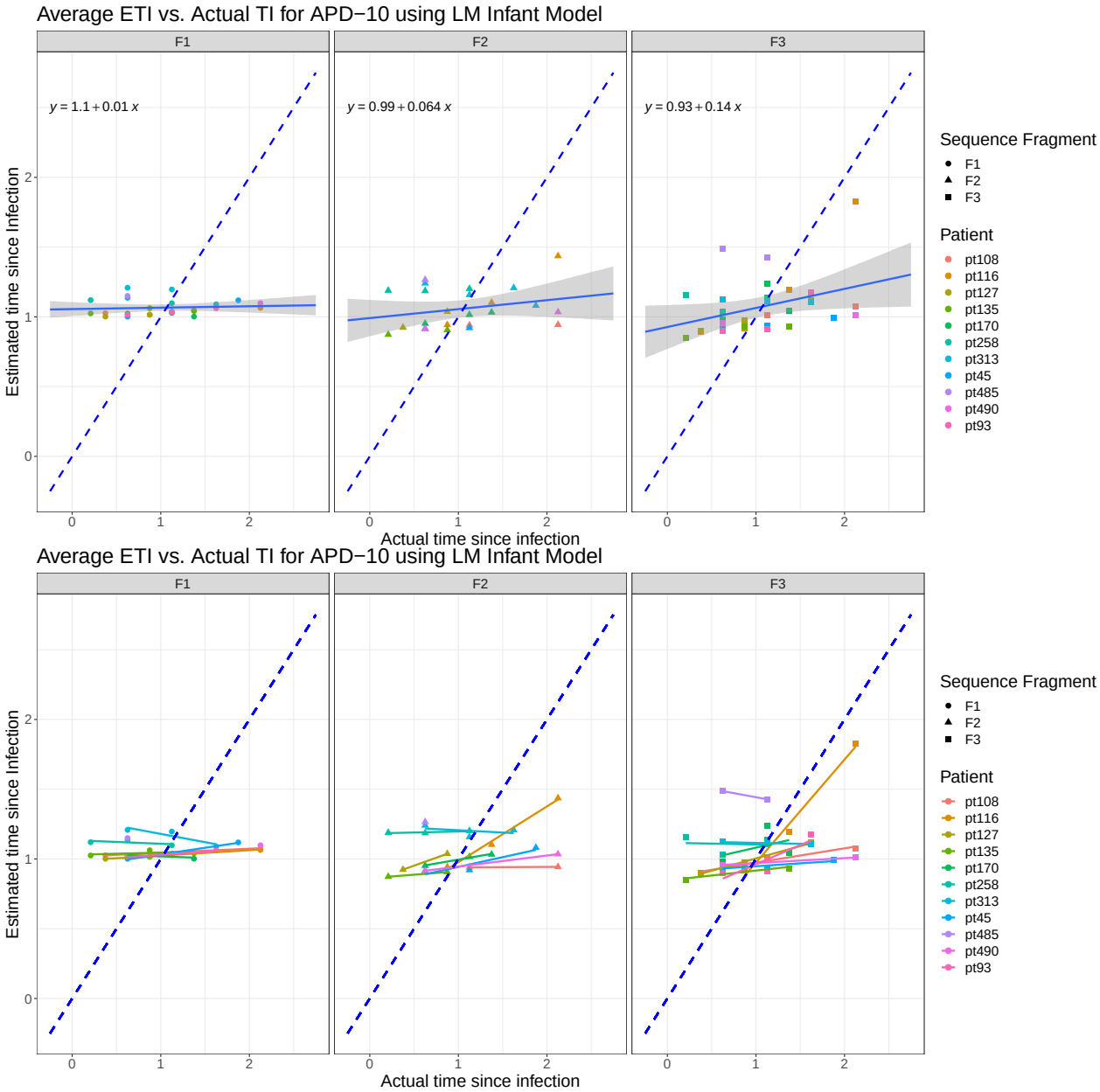
```
## Saving 16 x 8 in image
```



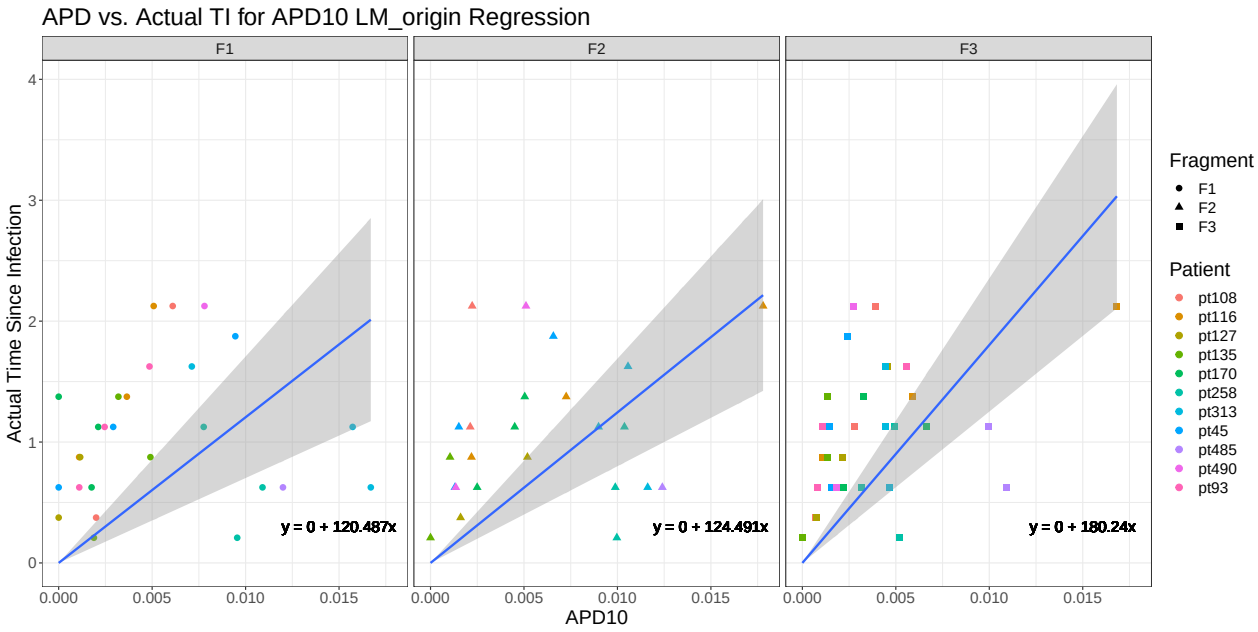

```
## [1] "LM Model for APD10 and Fragment F1 Summary Statistics:"
##                Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  1.001571  0.1707357 5.8662067 4.045402e-06
```

```
## avg_apd10   12.380843 24.5210732 0.5049063 6.180467e-01
## [1] "LM Model for APD10 and Fragment F2 Summary Statistics:"
##               Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)  0.8727834   0.191634 4.554429 0.0001557378
## avg_apd10   31.6735168  25.771600 1.229009 0.2320527190
## [1] "LM Model for APD10 and Fragment F3 Summary Statistics:"
##               Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)  0.8519093  0.1397386 6.096450 1.222884e-06
## avg_apd10   57.9861322 27.0534250 2.143393 4.060003e-02
```

```
## Saving 16 x 8 in image
```

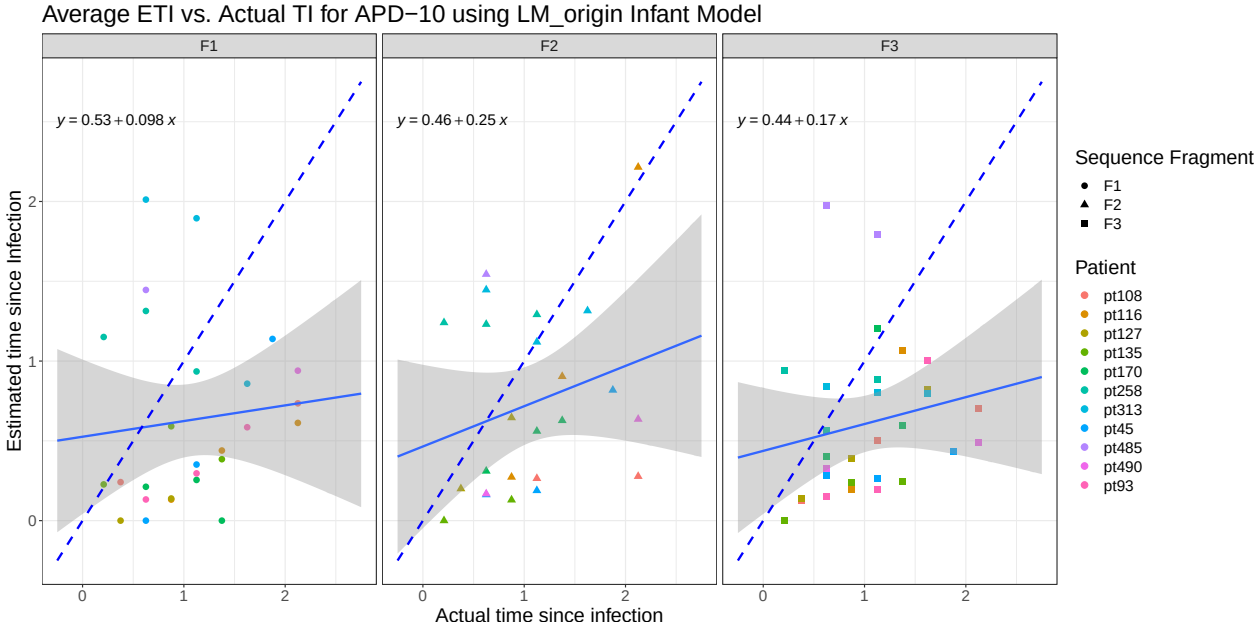


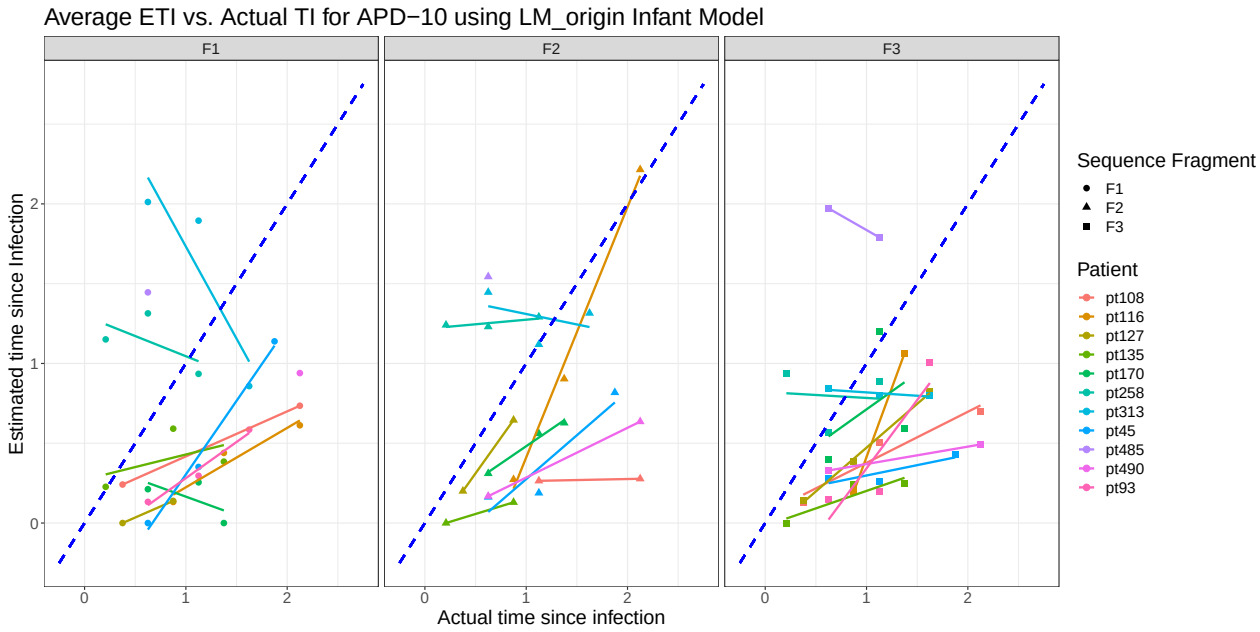

```
## Saving 16 x 8 in image
```



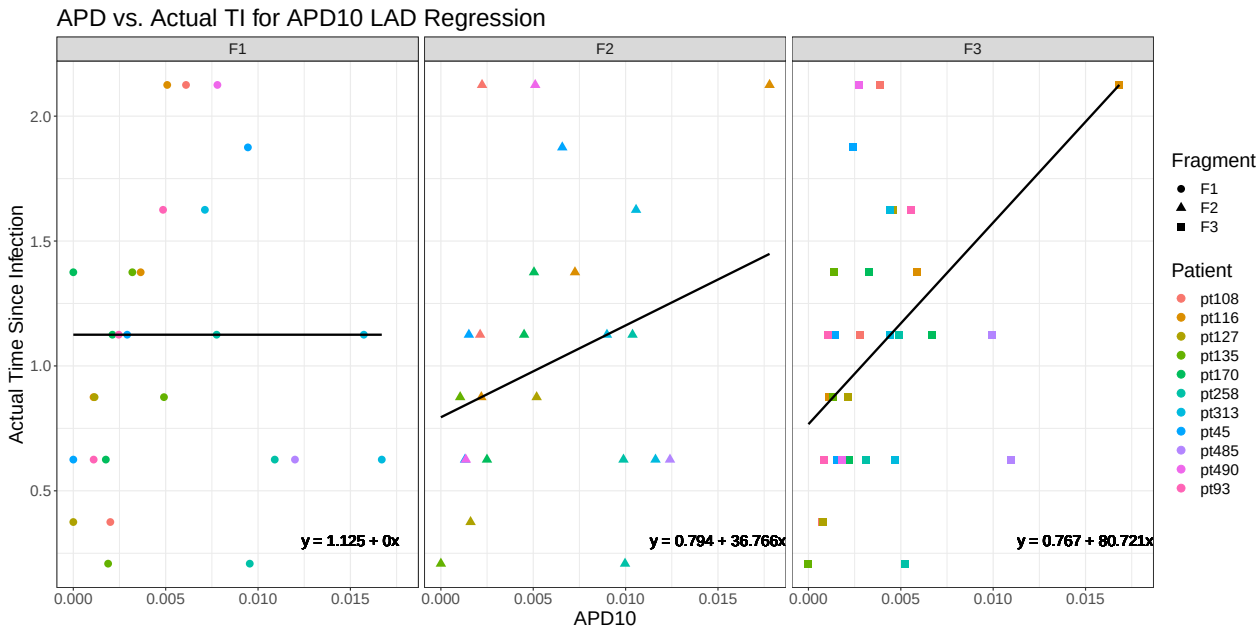

```
## [1] "LM_origin Model for APD10 and Fragment F1 Summary Statistics:"
##           Estimate Std. Error  t value     Pr(>|t|)
## avg_apd10 120.4874   24.45261 4.927385 4.070286e-05
## [1] "LM_origin Model for APD10 and Fragment F2 Summary Statistics:"
##           Estimate Std. Error  t value     Pr(>|t|)
## avg_apd10 124.4912   21.50463 5.789045 6.741141e-06
## [1] "LM_origin Model for APD10 and Fragment F3 Summary Statistics:"
##           Estimate Std. Error  t value     Pr(>|t|)
## avg_apd10 180.2404    26.9682 6.683442 2.100821e-07
```

```
## Saving 16 x 8 in image
```



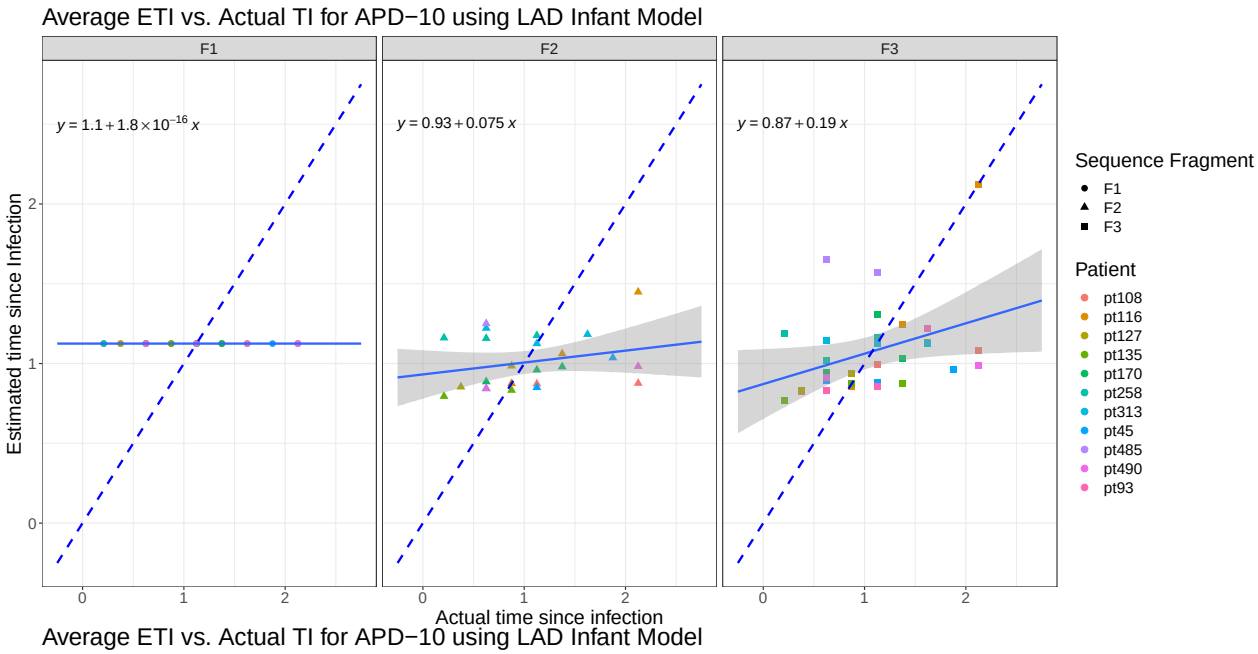


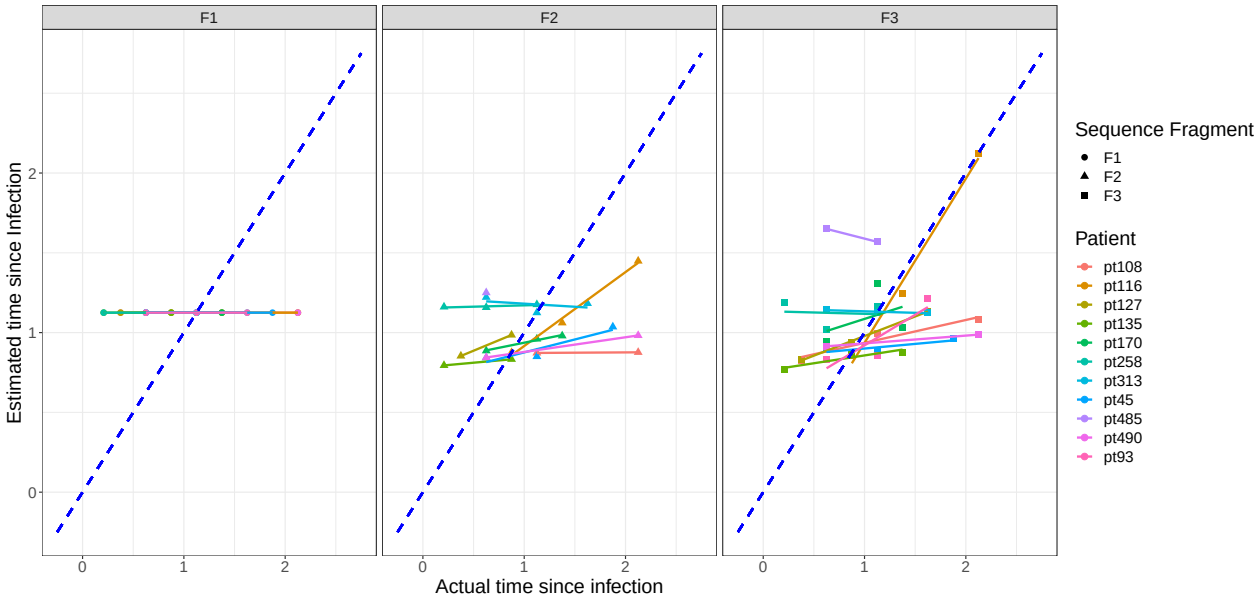

```
## Saving 16 x 8 in image
```



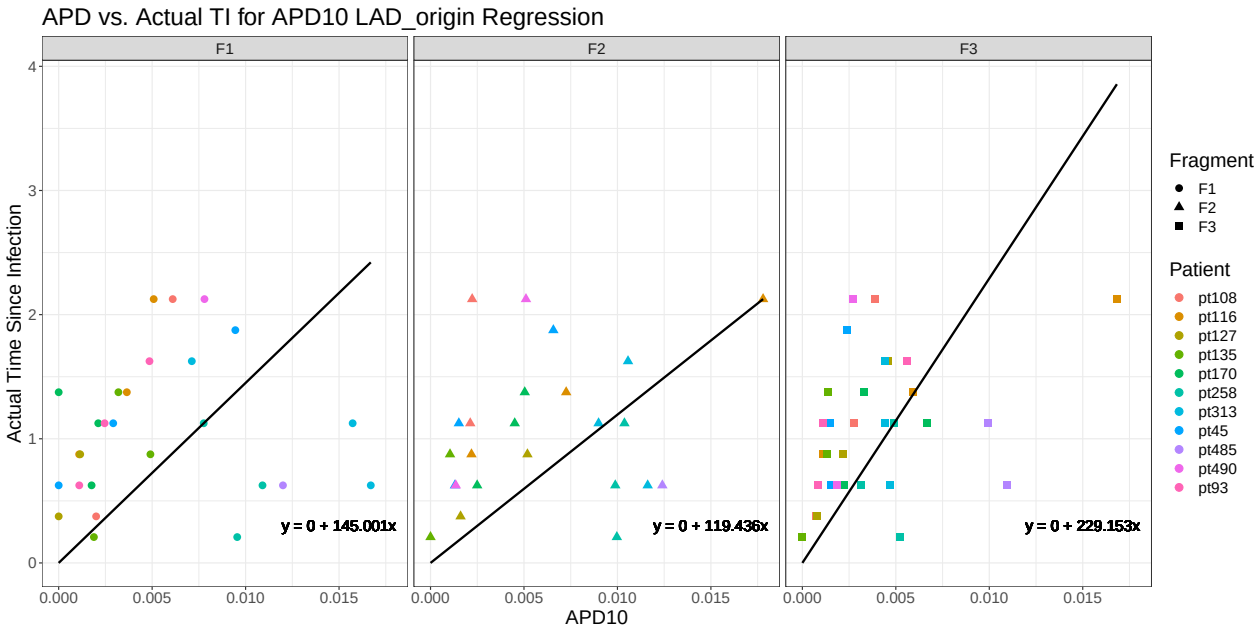

```
## [1] "LAD Model for APD10 and Fragment F1 Summary Statistics:"
##             Estimate  Std.Error Z value    p-value
## (Intercept)    1.125  0.2719231  4.1372 3.5157e-05
## avg_apd10      0.000 39.0536182  0.0000 1.0000e+00
## [1] "LAD Model for APD10 and Fragment F2 Summary Statistics:"
##              Estimate  Std.Error   Z value     p-value
## (Intercept)  0.794464  0.2896571 2.7427742 0.006092256
## avg_apd10   36.765869 38.9540844 0.9438258 0.345258683
## [1] "LAD Model for APD10 and Fragment F3 Summary Statistics:"
##              Estimate  Std.Error  Z value      p-value
## (Intercept)  0.766699  0.2127834 3.603190 0.0003143358
## avg_apd10   80.721062 41.1949189 1.959491 0.0500553444
```

```
## Saving 16 x 8 in image
```



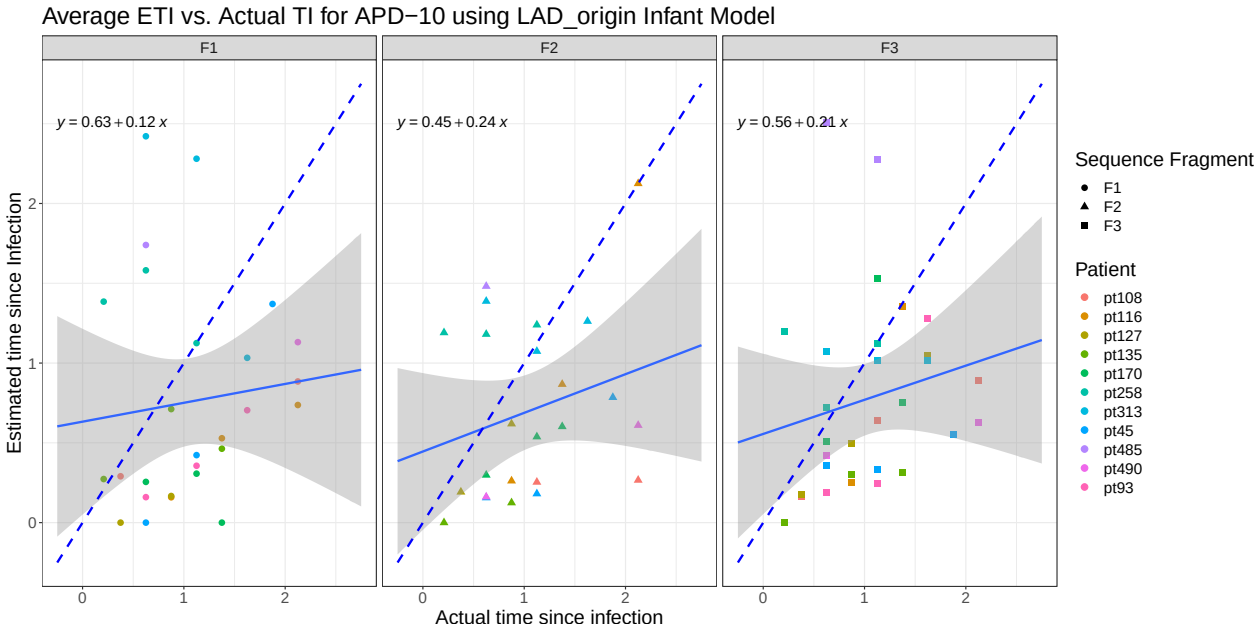


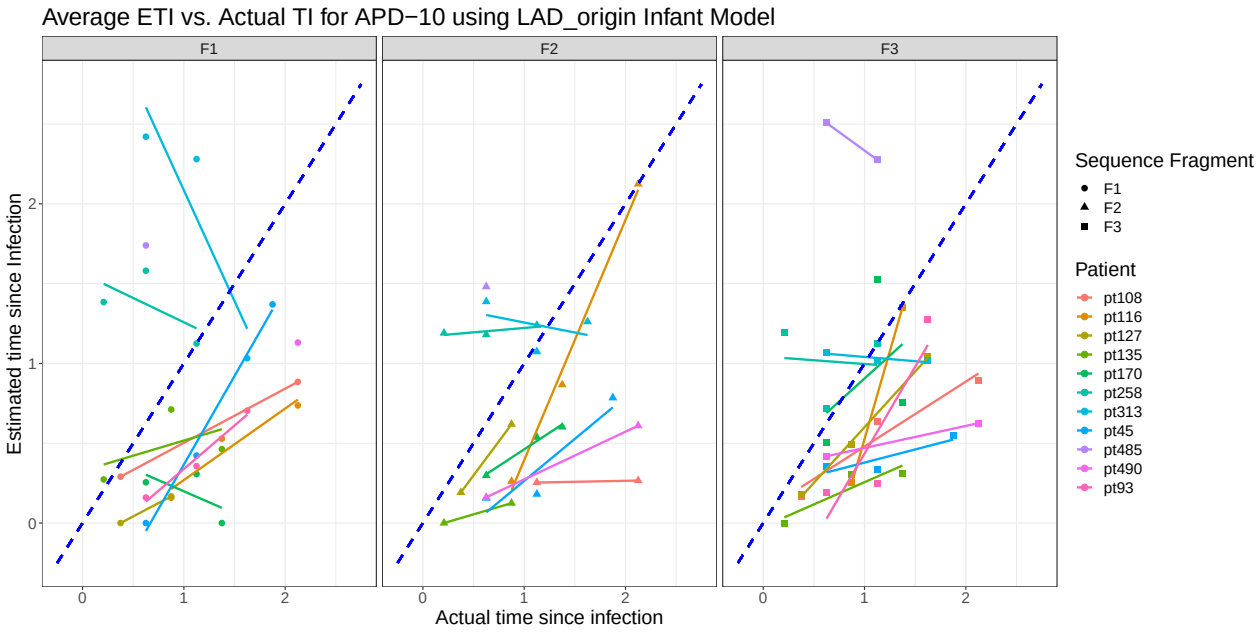

```
## Saving 16 x 8 in image
```



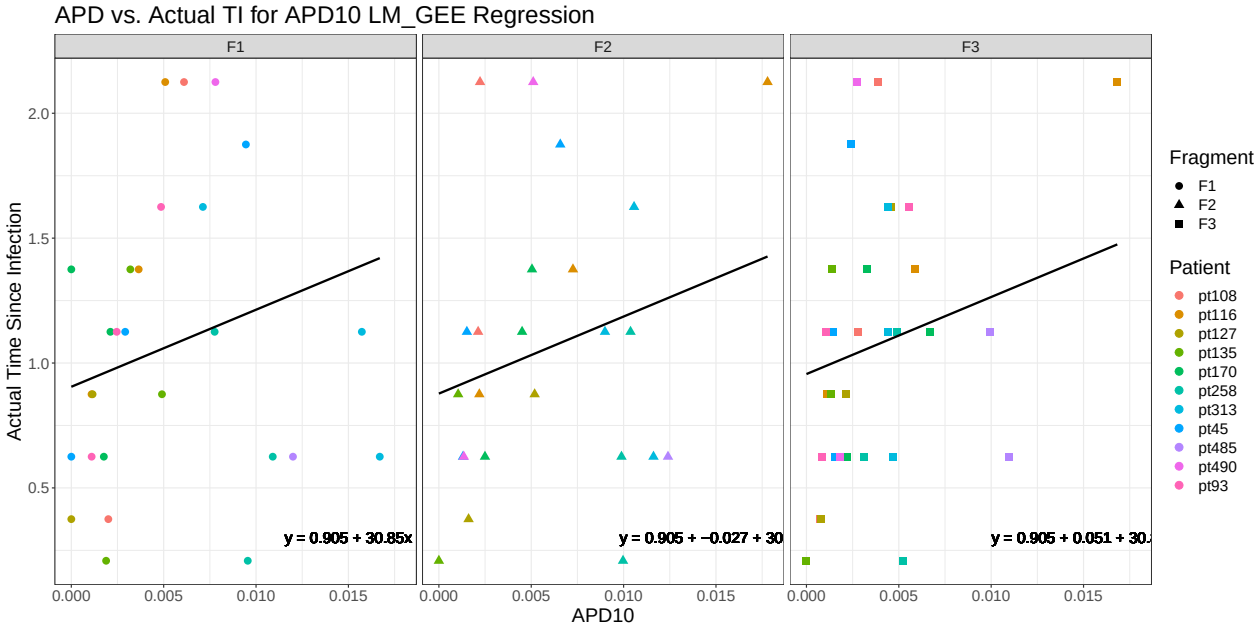

```
## [1] "LAD_origin Model for APD10 and Fragment F1 Summary Statistics:"
##           Estimate Std.Error  Z value      p-value
## avg_apd10 145.0006  42.65267 3.399566 0.0006749289
## [1] "LAD_origin Model for APD10 and Fragment F2 Summary Statistics:"
##           Estimate Std.Error Z value      p-value
## avg_apd10 119.4363  34.54921 3.45699 0.0005462442
## [1] "LAD_origin Model for APD10 and Fragment F3 Summary Statistics:"
##           Estimate Std.Error  Z value    p-value
## avg_apd10 229.1527  43.69364 5.244531 1.5668e-07
```

```
## Saving 16 x 8 in image
```






```
## Saving 16 x 8 in image
```




```
## [1] "LM_GEE Model for APD10 Summary Statistics:"
##                          Estimate Naive S.E.    Naive z Robust S.E.
## (Intercept)           0.90492138  0.1318987  6.8607305  0.08967801
## avg_apd10            30.85048485 14.6829337  2.1011118 22.73931647
## factor(fragment)F2 -0.02729840  0.1565716 -0.1743509  0.05492514
## factor(fragment)F3  0.05088403  0.1480864  0.3436104  0.03893931
##                       Robust z
## (Intercept)         10.0907840
## avg_apd10            1.3567024
## factor(fragment)F2 -0.4970111
## factor(fragment)F3  1.3067523
##        (Intercept)          avg_apd10 factor(fragment)F2
```

```
##          0.0000000          0.1748758          0.6191812
## factor(fragment)F3
##          0.1912968
```

```
## Saving 16 x 8 in image
```