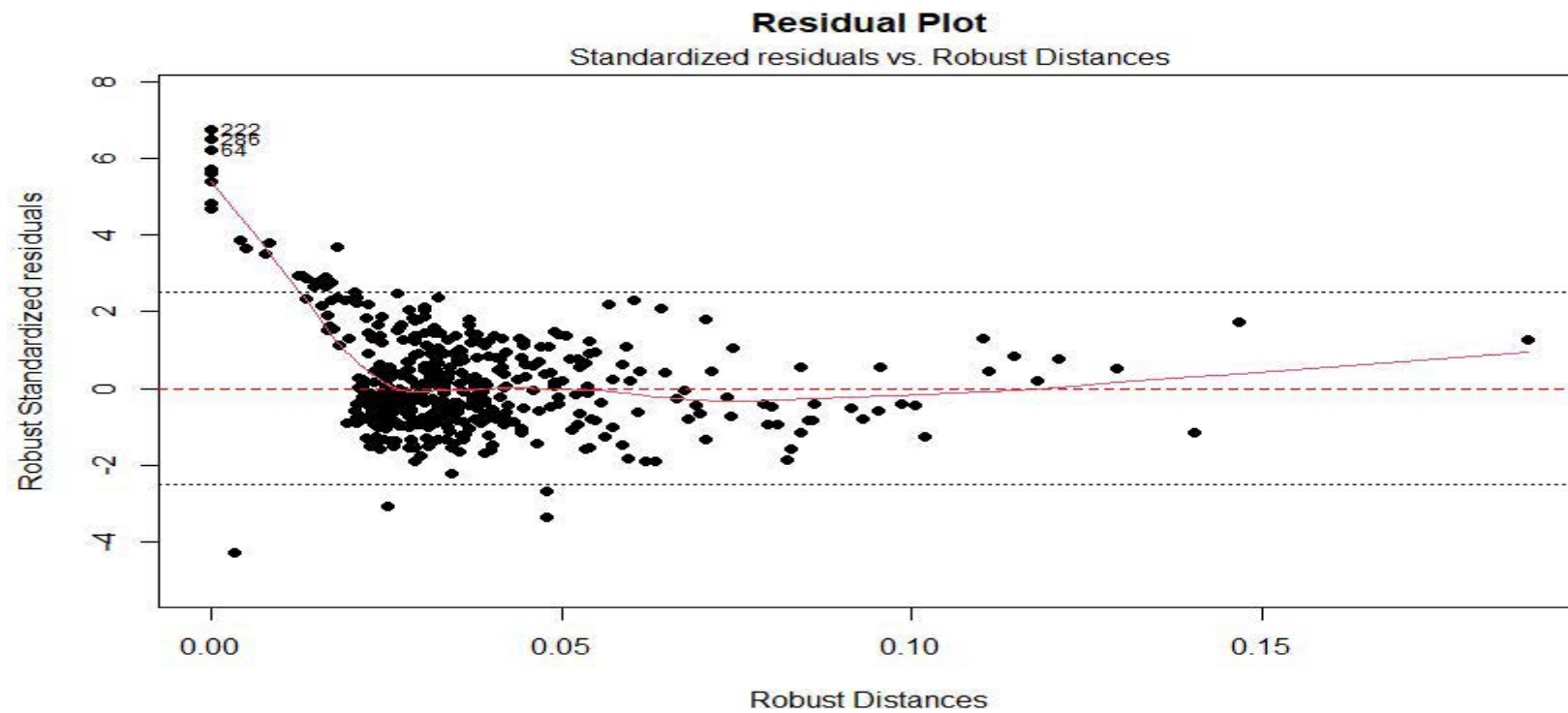


Figure1. model fit results

1. Residuals plot

```
plot(tukey_model11abcde, which = 1, col = "blue", pch = 16, main = "Residual Plot")  
> abline(h = 0, col = "red", lty = 2)
```



2. Residual standard error of the robust model was much lower than the OLS model

```
summary(ols12.15)$sigma
[1] 1881.64
> summary(tukey_model12.15Final)$sigma
[1] 1332.044
```

3. Test of heteroscedasticity –we do not have sufficient evidence to reject the null hypothesis that the variance of the residuals is constant (homoscedastic). The p-value of the BPtest is 0.05501 with 16 degrees of freedom.

```
> > bptest(tukey_model12.15Final)

studentized Breusch-Pagan test

data:  tukey_model12.15Final
BP = 22.001, df = 15, p-value = 0.1078
```

4. Test of multicollinearity

Initial VIF test revealed a moderate level of multicollinearity across occupation groups. Following recoding of variables, the VIF test showed no significant collinearity between independent variables.

```
car::vif(tukey_model11final)

(a) before recode
car::vif(tukey_model11final)
      age_over52      readwritedummy_no      occupdummy_gov
      1.524802      1.782832      4.753868
      occupdummy_private      occupdummy_merch      occupdummy_housewife
      5.034892      1.977299      5.727918
      incomedummy_over3567      nufamdummy_over5      comorbdummy_yes
      1.552517      1.151394      1.336363
      socialsuppdummy_no      lifestyledummy_poor      dietdummy_poor
      1.271712      1.553744      1.498820
      stagedummy_advanced      yearsfudummy_firstyr      sympdummy_symptomatic
      1.366680      1.555568      1.587791

(b) after recode
      age_over52      readwritedummy_no      occup2dum_employed
      1.366715      1.609683      2.490721
```

```

occup2dum_merchant 1.284801
comorbdummy_yes 1.260896
dietdummy_poor 1.542951
sympdummy_symptomatic 1.489381
incomedummy_3567orless 1.547341
socialsuppdummy_no 1.247865
stagedummy_advanced 1.266640
nufamdumy_over5 1.135565
lifestyledummy_poor 1.481054
yearsfdummy_firstyr 1.455039

```

5. The final robust model with model fit results and coefficients

```

tukey_model12.15Final <- lmrob(tc_intd ~ age_over52 + readwritedummy_no + occup2dum_employed + occup2dum_merchant
+ nufamdumy_over5 + comorbdummy_yes + socialsuppdummy_no + lifestyledummy_poor + dietdummy_poor + yearsfdummy_
firstyr + adv_stage + sympdummy_symptomatic + admitdummy_inpt + daysadmission + days.diag_to_tt, data = all_ca_cases_
dataset3)
> summary(tukey_model12.15Final)

Call:
lmrob(formula = tc_intd ~ age_over52 + readwritedummy_no + occup2dum_employed + occup2dum_merchant + nufamdumy_over5 + comorbdummy_yes + socialsuppdummy_no + lifestyledummy_poor + dietdummy_poor + yearsfdummy_firstyr + adv_stage + sympdummy_symptomatic + admitdummy_inpt + daysadmission + days.diag_to_tt, data = all_ca_cases_dataset3, \--> method = "MM")

Residuals:
    Min       1Q   Median       3Q      Max
-5916.8  -908.0  -164.1   1224.5   9026.8

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   -7.0739    222.0823  -0.032  0.974604
age_over52    210.0943    173.9764   1.208  0.227847
readwritedummy_no -222.0755    200.9696  -1.105  0.269753
occup2dum_employed  33.6234    218.5482   0.154  0.877799
occup2dum_merchant  453.9785    354.6270   1.280  0.201163
nufamdumy_over5    492.8157    204.1674   2.414  0.016194 *
comorbdummy_yes    834.9642    167.9186   4.972  9.49e-07 ***
socialsuppdummy_no  627.4826    184.6109   3.399  0.000738 ***
lifestyledummy_poor  224.2978    177.8559   1.261  0.207932
dietdummy_poor    740.2696    185.9441   3.981  8.02e-05 ***
yearsfdummy_firstyr  967.7256    184.4599   5.246  2.42e-07 ***
adv_stage      1324.1189    246.4117   5.374  1.25e-07 ***
sympdummy_symptomatic  39.7220    205.9067   0.193  0.847116

```

```

admitdummy_inpt      860.8666   590.0070   1.459 0.145255
daysadmission       92.0729    55.3280   1.664 0.096797 .
days.diag_to_tt      0.6230     0.9276   0.672 0.502154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Robust residual standard error: 1332
Multiple R-squared:  0.4005,    Adjusted R-squared:  0.3801
Convergence in 28 IRWLS iterations

Robustness weights:
 7 observations c(64,84,222,250,286,310,390)
   are outliers with |weight| = 0 ( < 0.00022);
 29 weights are ~= 1. The remaining 421 ones are summarized as
   Min.   1st Qu.   Median     Mean   3rd Qu.     Max.
0.0004908 0.8541000 0.9413000 0.8751000 0.9818000 0.9989000
Algorithmic parameters:
      tuning.chi          bb      tuning.psi      refine.tol
      1.548e+00          5.000e-01      4.685e+00      1.000e-07
      rel.tol          scale.tol      solve.tol      zero.tol
      1.000e-07          1.000e-10      1.000e-07      1.000e-10
      eps.outlier          eps.x warn.limit.reject warn.limit.meanrw
      2.188e-04          1.386e-09      5.000e-01      5.000e-01
      nResample          max.it      best.r.s      k.fast.s      k.max
           500           50           2           1           200
      maxit.scale      trace.lev          mts      compute.rd fast.s.large.n
           200           0          1000           0           2000
           psi      subsampling          cov
      "bisquare"      "nonsingular"      ".vcov.avar1"
compute.outlier.stats
      "SM"
seed : int(0)

```