PGx_analysis

William Tackett

2024-11-04

Exploratory Data Analysis

```
# read excel file
data <- read_excel("/Users/will/Documents/Documents - Mac (2) 2/UM/Fall24/PGx/Confirmatory PGx Analysis,
data.dict <- read_excel("/Users/will/Documents/Documents - Mac (2) 2/UM/Fall24/PGx/Confirmatory PGx Anal
summary(data)
```

```
##   record_id           Group         Pair_ID           age          ethnicity
## Length:91         Min.   :1.000   Min.   : 1.00   Min.   :30.00   Min.   :0
## Class :character  1st Qu.:2.000   1st Qu.: 7.00   1st Qu.:50.00   1st Qu.:1
## Mode  :character  Median :2.000   Median :11.00   Median :56.50   Median :1
##                   Mean   :1.802   Mean   :10.58   Mean   :57.27   Mean   :1
##                   3rd Qu.:2.000   3rd Qu.:15.00   3rd Qu.:65.75   3rd Qu.:1
##                   Max.   :2.000   Max.   :19.00   Max.   :80.00   Max.   :2
##                                                   NA's   :1
##      race            gender        eligible_gene      dpyd_as
## Min.   :1.000   Min.   :1.000   Min.   :1.000   Min.   :1.000
## 1st Qu.:4.000   1st Qu.:1.000   1st Qu.:1.000   1st Qu.:1.500
## Median :4.000   Median :1.000   Median :1.000   Median :1.500
## Mean   :3.956   Mean   :1.451   Mean   :1.253   Mean   :1.412
## 3rd Qu.:4.000   3rd Qu.:2.000   3rd Qu.:1.500   3rd Qu.:1.500
## Max.   :6.000   Max.   :2.000   Max.   :2.000   Max.   :1.500
##                                                 NA's   :23
##   cancer_type      cancer_stage_and_grade chemo_regimen chemoradiation_yn
## Min.   : 1.000   Min.   :1.000           Min.   :1.0   Min.   :0.0
## 1st Qu.: 1.000   1st Qu.:2.000           1st Qu.:2.0   1st Qu.:0.0
## Median : 3.000   Median :3.000           Median :5.0   Median :0.0
## Mean   : 4.278   Mean   :3.044           Mean   :4.5   Mean   :0.1
## 3rd Qu.: 5.000   3rd Qu.:4.000           3rd Qu.:6.0   3rd Qu.:0.0
## Max.   :14.000   Max.   :4.000           Max.   :7.0   Max.   :1.0
## NA's   :1        NA's   :1               NA's   :1     NA's   :1
## iri_load_dose_1st_cycle iri_load_dose_2nd_cycle iri_total_m2_3rd_cycle
## Min.   : 20.0           Min.   : 20.0           Min.   : 16.0
## 1st Qu.:135.0           1st Qu.:133.8           1st Qu.:132.5
## Median :150.0           Median :135.0           Median :135.0
## Mean   :140.8           Mean   :136.0           Mean   :131.5
## 3rd Qu.:162.5           3rd Qu.:150.0           3rd Qu.:150.0
## Max.   :190.0           Max.   :180.0           Max.   :180.0
## NA's   :68              NA's   :71              NA's   :72
##   cape_bsa_c1     cape_bsa_dose_c2 cape_bsa_dose_c3    fu_bsa_c1
## Min.   : 508.5  Min.   : 508.5   Min.   : 508.5   Min.   : 414.9
## 1st Qu.:1238.5  1st Qu.:1385.0   1st Qu.:1000.6   1st Qu.:1945.0
```

```
## Median :1431.8   Median :1484.4   Median :1431.2   Median :2651.6
## Mean   :1471.0   Mean   :1475.6   Mean   :1340.5   Mean   :2434.7
## 3rd Qu.:1738.2   3rd Qu.:1696.5   3rd Qu.:1572.9   3rd Qu.:2795.5
## Max.   :2276.0   Max.   :2276.0   Max.   :2534.5   Max.   :4431.0
## NA's   :63       NA's   :68       NA's   :75       NA's   :53
## fu_bsa_dose_c2   fu_bsa_dose_c3    TOX_grade3up     treatment_mods
## Min.   : 598.1   Min.   : 598.1   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:1396.8   1st Qu.:1419.7   1st Qu.:0.0000   1st Qu.:0.0000
## Median :2016.2   Median :2087.7   Median :0.0000   Median :1.0000
## Mean   :2094.7   Mean   :1972.5   Mean   :0.4725   Mean   :0.5934
## 3rd Qu.:2404.5   3rd Qu.:2404.2   3rd Qu.:1.0000   3rd Qu.:1.0000
## Max.   :3865.8   Max.   :2999.1   Max.   :1.0000   Max.   :1.0000
## NA's   :74       NA's   :77
## was_treat_mod_due_toxicity    drug_dc          Tx_delay         Dose_change
## Min.   :0.0000                Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:1.0000                1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :1.0000                Median :0.0000   Median :0.0000   Median :0.0000
## Mean   :0.8333                Mean   :0.1149   Mean   :0.2759   Mean   :0.3678
## 3rd Qu.:1.0000                3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000
## Max.   :1.0000                Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
## NA's   :37                    NA's   :4        NA's   :4        NA's   :4
```

```r
head(data)
```

```
## # A tibble: 6 x 28
##   record_id Group Pair_ID   age ethnicity  race gender eligible_gene dpyd_as
##   <chr>     <dbl>   <dbl> <dbl>     <dbl> <dbl>  <dbl>         <dbl>   <dbl>
## 1 1             1       1    44         1     4      1             1       1
## 2 1A            2       1    50         1     4      2             1       1
## 3 2             1       2    52         1     4      1             1       1
## 4 2A            1       2    55         1     4      2             1       1
## 5 2B            1       2    53         1     4      1             1       1
## 6 3             1       3    45         1     4      1             1     1.5
## # i 19 more variables: cancer_type <dbl>, cancer_stage_and_grade <dbl>,
## #   chemo_regimen <dbl>, chemoradiation_yn <dbl>,
## #   iri_load_dose_1st_cycle <dbl>, iri_load_dose_2nd_cycle <dbl>,
## #   iri_total_m2_3rd_cycle <dbl>, cape_bsa_c1 <dbl>, cape_bsa_dose_c2 <dbl>,
## #   cape_bsa_dose_c3 <dbl>, fu_bsa_c1 <dbl>, fu_bsa_dose_c2 <dbl>,
## #   fu_bsa_dose_c3 <dbl>, TOX_grade3up <dbl>, treatment_mods <dbl>,
## #   was_treat_mod_due_toxicity <dbl>, drug_dc <dbl>, Tx_delay <dbl>, ...
```

```r
# recode Group as case_control, 1 --> 0, 2 --> 1
data$case_control <- ifelse(data$Group == 1, 0, 1)
data$Group <- NULL
```

Demographic variables

1.Demographic variables-summary variables for overall cohort and cases and controls, respectively; evaluate
for differences between cases and controls a.Age (Variable name= age) b.Race (Variable name = race)
c.Ethnicity (variable name = ethnicity) d.Eligible gene (variable name = eligible_gene) e.DPYD Activity
score (Variable name = dpyd_as) f.Cancer type (Variable name = cancer_type) g.Cancer stage (Variable
name = cancer_stage_and_grade) h.Chemotherapy regimen (Variable name = chemo_regimen)

```r
# quick summaries
variables <- c("age", "race", "ethnicity", "eligible_gene", "dpyd_as",
               "cancer_type", "cancer_stage_and_grade", "chemo_regimen")
```

```r
for (var in variables) {
  cat(var, ":\n")
  print(summary(data[[var]]))
  cat("\n")
}
```

```
## age :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   30.00   50.00   56.50   57.27   65.75   80.00       1
##
## race :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   4.000   4.000   3.956   4.000   6.000
##
## ethnicity :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0       1       1       1       1       2
##
## eligible_gene :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   1.000   1.000   1.000   1.253   1.500   2.000
##
## dpyd_as :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   1.500   1.500   1.412   1.500   1.500      23
##
## cancer_type :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   1.000   3.000   4.278   5.000  14.000       1
##
## cancer_stage_and_grade :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##   1.000   2.000   3.000   3.044   4.000   4.000       1
##
## chemo_regimen :
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's
##    1.0     2.0     5.0     4.5     6.0     7.0       1
```

```r
# evaluate demographic variables for differences between cases and controls

# test differences for age (continuous)
age_cases <- data %>% filter(case_control == 0) %>% pull(age)
age_controls <- data %>% filter(case_control == 1) %>% pull(age)
t_test_age <- t.test(age_cases, age_controls)
print(t_test_age)
```

```
##
##  Welch Two Sample t-test
##
## data:  age_cases and age_controls
## t = -0.53794, df = 24.606, p-value = 0.5955
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.650198  4.483531
```

```
## sample estimates:
## mean of x mean of y
##  56.00000  57.58333
```

```r
# test differences for race (categorical variable)
race_table <- table(data$race, data$case_control)
chisq_test_race <- chisq.test(race_table)
```

```
## Warning in chisq.test(race_table): Chi-squared approximation may be incorrect
```

```r
print(chisq_test_race)
```

```
##
##  Pearson's Chi-squared test
##
## data:  race_table
## X-squared = 2.1627, df = 3, p-value = 0.5393
```

```r
# test differences for 'ethnicity' (categorical variable)
ethnicity_table <- table(data$ethnicity, data$case_control)
chisq_test_ethnicity <- chisq.test(ethnicity_table)
```

```
## Warning in chisq.test(ethnicity_table): Chi-squared approximation may be
## incorrect
```

```r
print(chisq_test_ethnicity)
```

```
##
##  Pearson's Chi-squared test
##
## data:  ethnicity_table
## X-squared = 0.50423, df = 2, p-value = 0.7772
```

```r
# test differences for 'eligible_gene' (categorical variable)
eligible_gene_table <- table(data$eligible_gene, data$case_control)
mcnemar_test_eligible_gene <- mcnemar.test(table(data$dpyd_as, data$case_control))
print(mcnemar_test_eligible_gene)
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  table(data$dpyd_as, data$case_control)
## McNemar's chi-squared = 0.0625, df = 1, p-value = 0.8026
```

```r
# test differences for 'dpyd_as' (categorical variable)
dpyd_as_table <- table(data$dpyd_as, data$case_control)
mcnemar_test_dpyd_as <- mcnemar.test(table(data$dpyd_as, data$case_control))
print(mcnemar_test_dpyd_as)
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  table(data$dpyd_as, data$case_control)
## McNemar's chi-squared = 0.0625, df = 1, p-value = 0.8026
```

```r
# test differences for 'cancer_type' (categorical variable)
cancer_type_table <- table(data$cancer_type, data$case_control)
chisq_test_cancer_type <- chisq.test(cancer_type_table)
```

```
## Warning in chisq.test(cancer_type_table): Chi-squared approximation may be
## incorrect
```

```r
print(chisq_test_cancer_type)
```

```
##
##  Pearson's Chi-squared test
##
## data:  cancer_type_table
## X-squared = 12.22, df = 9, p-value = 0.2012
```

```r
# test differences for 'cancer_stage_and_grade' (categorical variable)
cancer_stage_table <- table(data$cancer_stage_and_grade, data$case_control)
chisq_test_cancer_stage <- chisq.test(cancer_stage_table)
```

```
## Warning in chisq.test(cancer_stage_table): Chi-squared approximation may be
## incorrect
```

```r
print(chisq_test_cancer_stage)
```

```
##
##  Pearson's Chi-squared test
##
## data:  cancer_stage_table
## X-squared = 1.1104, df = 3, p-value = 0.7746
```

```r
# test differences for 'chemo_regimen' (categorical variable)
chemo_regimen_table <- table(data$chemo_regimen, data$case_control)
chisq_test_chemo_regimen <- chisq.test(chemo_regimen_table)
```

```
## Warning in chisq.test(chemo_regimen_table): Chi-squared approximation may be
## incorrect
```

```r
print(chisq_test_chemo_regimen)
```

```
##
##  Pearson's Chi-squared test
##
## data:  chemo_regimen_table
## X-squared = 2.238, df = 5, p-value = 0.8153
```

2.Dosing information a.Cycle 1 dose for cases and controls, respectively i.Irinotecan (Variable name= iri_load_dose_1st_cycle) ii.Capecitabine (Variable name =cape_bsa_c1) iii.5-Fluorouracil (Variable name=fu_bsa_c1) b.Cycle 2 dose for cases and controls, respectively i.Irinotecan (Variable name=iri_load_dose_2nd_cycle) ii.Capecitabine (Variable name =cape_bsa_dose_c2) iii.5-Fluorouracil (Variable name=fu_bsa_dose_c2) c.Cycle 1 dose for cases and controls, respectively i.Irinotecan (Variable name= iri_total_m2_3rd_cycle) ii.Capecitabine (Variable name =cape_bsa_dose_c3) iii.5-Fluorouracil (Variable name=fu_bsa_dose_c3)

3.Primary endpoint: Incidence of >grade 3 toxicity among cases and controls a.Variable: TOX_grade3up

```r
# calculate proportions of subjects with any grade 3 or higher AE and SAE among the case and control gr
# i.Compare for all
prop.table(table(data$TOX_grade3up, data$case_control), 2)
```

```
##
##              0          1
##   0 0.6111111 0.5068493
##   1 0.3888889 0.4931507
```

```r
mcnemar.test(table(data$TOX_grade3up, data$case_control))
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  table(data$TOX_grade3up, data$case_control)
## McNemar's chi-squared = 19.114, df = 1, p-value = 1.232e-05
```

```r
# calculate proportions of subjects with any grade 3 or higher AE and SAE among the case and control gr
# i.Compare for eligible_gene=1 only
data_elg1 <- data %>% filter(eligible_gene == 1)
prop.table(table(data_elg1$TOX_grade3up, data_elg1$case_control), 2)
```

```
##
##              0         1
##   0 0.6428571 0.5185185
##   1 0.3571429 0.4814815
```

```r
mcnemar.test(table(data_elg1$TOX_grade3up, data_elg1$case_control))
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  table(data_elg1$TOX_grade3up, data_elg1$case_control)
## McNemar's chi-squared = 14.667, df = 1, p-value = 0.0001283
```

```r
# ii. Compare for eligible_gene=2 only
# filter for eligible_gene=2
data_elg2 <- data %>% filter(eligible_gene == 2)
prop.table(table(data_elg2$TOX_grade3up, data_elg2$case_control), 2)
```

```
##
##              0         1
##   0 0.5000000 0.4736842
##   1 0.5000000 0.5263158
```

```r
mcnemar.test(table(data_elg2$TOX_grade3up, data_elg2$case_control))
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  table(data_elg2$TOX_grade3up, data_elg2$case_control)
## McNemar's chi-squared = 3.2727, df = 1, p-value = 0.07044
```

Conditional Logistic Regression

    b. Conditional logistic regression will be used to model the likelihood of any grade 3 or higher AE and SAE as a function of gender and cancer diagnosis

```r
# i. Fit a conditional logistic regression model for all cases and controls
model1 <- clogit(TOX_grade3up ~ gender + cancer_type + strata(age, race, ethnicity, eligible_gene, dpyd
```

```
## Warning in coxexact.fit(X, Y, istrat, offset, init, control, weights = weights,
## : Ran out of iterations and did not converge
```

```r
# post-model
summary(model1)
```

```
## Call:
```

```
## coxph(formula = Surv(rep(1, 91L), TOX_grade3up) ~ gender + cancer_type +
##     strata(age, race, ethnicity, eligible_gene, dpyd_as, chemo_regimen),
##     data = data, method = "exact")
##
##   n= 67, number of events= 30
##    (24 observations deleted due to missingness)
##
##               coef exp(coef)  se(coef)     z Pr(>|z|)
## gender      3.127e+01 3.822e+13 2.640e+04 0.001    0.999
## cancer_type 1.698e+01 2.370e+07 4.260e+03 0.004    0.997
##
##             exp(coef) exp(-coef) lower .95 upper .95
## gender      3.822e+13  2.617e-14         0       Inf
## cancer_type 2.370e+07  4.219e-08         0       Inf
##
## Concordance= 0.8  (se = 0.139 )
## Likelihood ratio test= 3.58  on 2 df,   p=0.2
## Wald test           = 0  on 2 df,   p=1
## Score (logrank) test = 1.72  on 2 df,   p=0.4
```

```r
round(summary(model1)$coef, 4)
```

```
##               coef    exp(coef)  se(coef)      z Pr(>|z|)
## gender      31.2743 3.821618e+13 26395.584 0.0012   0.9991
## cancer_type 16.9810 2.370063e+07  4259.736 0.0040   0.9968
```

```r
# OR and 95% CI
m1.OR.CI <- cbind("OR" = exp(coef(model1)), exp(confint(model1)))
round(m1.OR.CI, 3)
```

```
##                       OR 2.5 % 97.5 %
## gender      3.821618e+13    0    Inf
## cancer_type 2.370063e+07    0    Inf
```

```r
# "we do not have ecog for each patient-had mentioned using cancer stage instead"
model2 <- clogit(TOX_grade3up ~ gender + cancer_stage_and_grade + cancer_type + strata(age, race, ethni
```

```
## Warning in coxexact.fit(X, Y, istrat, offset, init, control, weights = weights,
## : Ran out of iterations and did not converge
```

```r
# post-model
summary(model2)
```

```
## Call:
## coxph(formula = Surv(rep(1, 91L), TOX_grade3up) ~ gender + cancer_stage_and_grade +
##     cancer_type + strata(age, race, ethnicity, eligible_gene,
##     dpyd_as, chemo_regimen), data = data, method = "exact")
##
##   n= 67, number of events= 30
##    (24 observations deleted due to missingness)
##
##                              coef exp(coef)  se(coef)     z Pr(>|z|)
## gender                 9.650e+01 8.112e+41 2.896e+04 0.003    0.997
## cancer_stage_and_grade 1.643e+01 1.362e+07 2.994e+03 0.005    0.996
## cancer_type            3.354e+01 3.683e+14 4.578e+03 0.007    0.994
##
##                         exp(coef) exp(-coef) lower .95 upper .95
```

```
## gender                     8.112e+41  1.233e-42          0      Inf
## cancer_stage_and_grade 1.362e+07  7.342e-08          0      Inf
## cancer_type               3.683e+14  2.715e-15          0      Inf
##
## Concordance= 1  (se = 0 )
## Likelihood ratio test= 6.36  on 3 df,   p=0.1
## Wald test           = 0  on 3 df,   p=1
## Score (logrank) test = 2.42  on 3 df,   p=0.5
```

```r
round(summary(model2)$coef, 4)
```

```
##                              coef     exp(coef)  se(coef)     z Pr(>|z|)
## gender                    96.4994 8.112261e+41 28961.724 0.0033   0.9973
## cancer_stage_and_grade 16.4271 1.362066e+07  2993.516 0.0055   0.9956
## cancer_type               33.5399 3.682964e+14  4578.075 0.0073   0.9942
```

```r
# OR and 95% CI
m2.OR.CI <- cbind("OR" = exp(coef(model2)), exp(confint(model2)))
round(m2.OR.CI, 3)
```

```
##                                   OR 2.5 % 97.5 %
## gender                    8.112261e+41    0    Inf
## cancer_stage_and_grade 1.362066e+07    0    Inf
## cancer_type               3.682964e+14    0    Inf
```

```r
# LRT for two models
logLik1 <- logLik(model1)
logLik2 <- logLik(model2)
LRT_stat <- -2 * (logLik1 - logLik2)
df <- attr(logLik2, "df") - attr(logLik1, "df")
pval <- pchisq(LRT_stat, df = df, lower.tail = FALSE)
# Print results
LRT_results <- list(
  "Likelihood Ratio Test Statistic" = as.numeric(LRT_stat),
  "Degrees of Freedom" = df,
  "p-value" = pval
)
print(LRT_results)
```

```
## $`Likelihood Ratio Test Statistic`
## [1] 2.772589
##
## $`Degrees of Freedom`
## [1] 1
##
## $`p-value`
## 'log Lik.' 0.09589098 (df=2)
```

4. Secondary endpoints Calculate the percentages of dose reductions, delays, or discontinuation among cases and matched controls. McNemar's tests will be used to compare each of these secondary endpoints between cases and controls.

a.Composite endpoint of drug discontinuation, treatment delay, or dosing change secondary to a toxicity event

```r
# i. filter by "was_treat_mod_due_toxicity" and compare "treatment_mods" among all cases and controls
data_treat_mods <- data %>% filter(was_treat_mod_due_toxicity == 1) %>% select(treatment_mods, case_con
```

8

```
prop.table(table(data_treat_mods$treatment_mods, data_treat_mods$case_control), 2)
```

```
##
##     0 1
##   1 1 1
```

```r
# treatment mods = 1 for all in this filtered dataset, so can't perform McNemar's test
data_treat_mods_chi <- chisq.test(table(data_treat_mods$treatment_mods, data_treat_mods$case_control))
print(data_treat_mods_chi)
```

```
##
##   Chi-squared test for given probabilities
##
## data:  table(data_treat_mods$treatment_mods, data_treat_mods$case_control)
## X-squared = 30.422, df = 1, p-value = 3.475e-08
```

```r
# ii. filter by "was_treat_mod_due_toxicity" and compare "treatment_mods" among all cases and controls
data_treat_mods_elig1 <- data %>% filter(was_treat_mod_due_toxicity == 1 & eligible_gene == 1) %>% sele
prop.table(table(data_treat_mods_elig1$treatment_mods, data_treat_mods_elig1$case_control), 2)
```

```
##
##     0 1
##   1 1 1
```

```r
# treatment mods = 1 for all in this filtered dataset, so can't perform McNemar's test
data_treat_mods_elig1_chi <- chisq.test(table(data_treat_mods_elig1$treatment_mods, data_treat_mods_eli
print(data_treat_mods_elig1_chi)
```

```
##
##   Chi-squared test for given probabilities
##
## data:  table(data_treat_mods_elig1$treatment_mods, data_treat_mods_elig1$case_control)
## X-squared = 17.065, df = 1, p-value = 3.613e-05
```

```r
# iii. filter by "was_treat_mod_due_toxicity" and compare "treatment_mods" among all cases and controls
data_treat_mods_elig2 <- data %>% filter(was_treat_mod_due_toxicity == 1 & eligible_gene == 2) %>% sele
```

```
prop.table(table(data_treat_mods_elig2$treatment_mods, data_treat_mods_elig2$case_control), 2)
```

```
##
##      1
##   1 1
```

```r
# data_treat_mods_elig2_chi <- chisq.test(table(data_treat_mods_elig2$treatment_mods, data_treat_mods_e
# print(data_treat_mods_elig2_chi)

print(data_treat_mods_elig2$treatment_mods)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```r
print(data_treat_mods_elig2$case_control)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

  b. Assess each endpoint individually

```r
# i.Filter by "was_treat_mod_due_toxicity", if = 1 then compare "drug_dc" among all cases and controls
data_drug_dc <- data %>% filter(was_treat_mod_due_toxicity == 1) %>% select(drug_dc, case_control)
prop.table(table(data_drug_dc$drug_dc, data_drug_dc$case_control), 2)
```

```
##
##             0          1
##   0 0.50000000 0.90243902
##   1 0.50000000 0.09756098
```

```r
mcnemar.test(table(data_drug_dc$drug_dc, data_drug_dc$case_control))
```

```
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  table(data_drug_dc$drug_dc, data_drug_dc$case_control)
## McNemar's chi-squared = 29.641, df = 1, p-value = 5.199e-08
```

````r
# ii.Filter by "was_treat_mod_due_toxicity", if = 1 then compare "treat_delay" among all cases and cont
data_treat_delay <- data %>% filter(was_treat_mod_due_toxicity == 1) %>% select(Tx_delay, case_control)
prop.table(table(data_treat_delay$Tx_delay, data_treat_delay$case_control), 2)
````

````
##
##               0          1
##   0 1.0000000 0.5121951
##   1 0.0000000 0.4878049
````

````r
mcnemar.test(table(data_treat_delay$Tx_delay, data_treat_delay$case_control))
````

````
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  table(data_treat_delay$Tx_delay, data_treat_delay$case_control)
## McNemar's chi-squared = 19.048, df = 1, p-value = 1.275e-05
````

````r
# iii.Filter by "was_treat_mod_due_toxicity", if = 1 then compare "Dose_change" among all cases and cont
data_dose_change <- data %>% filter(was_treat_mod_due_toxicity == 1) %>% select(Dose_change, case_contro
prop.table(table(data_dose_change$Dose_change, data_dose_change$case_control), 2)
````

````
##
##               0          1
##   0 0.5000000 0.3170732
##   1 0.5000000 0.6829268
````

````r
mcnemar.test(table(data_dose_change$Dose_change, data_dose_change$case_control))
````

````
##
##  McNemar's Chi-squared test with continuity correction
##
## data:  table(data_dose_change$Dose_change, data_dose_change$case_control)
## McNemar's chi-squared = 6.6667, df = 1, p-value = 0.009823
````