A4Q3

Jiaqi Bi

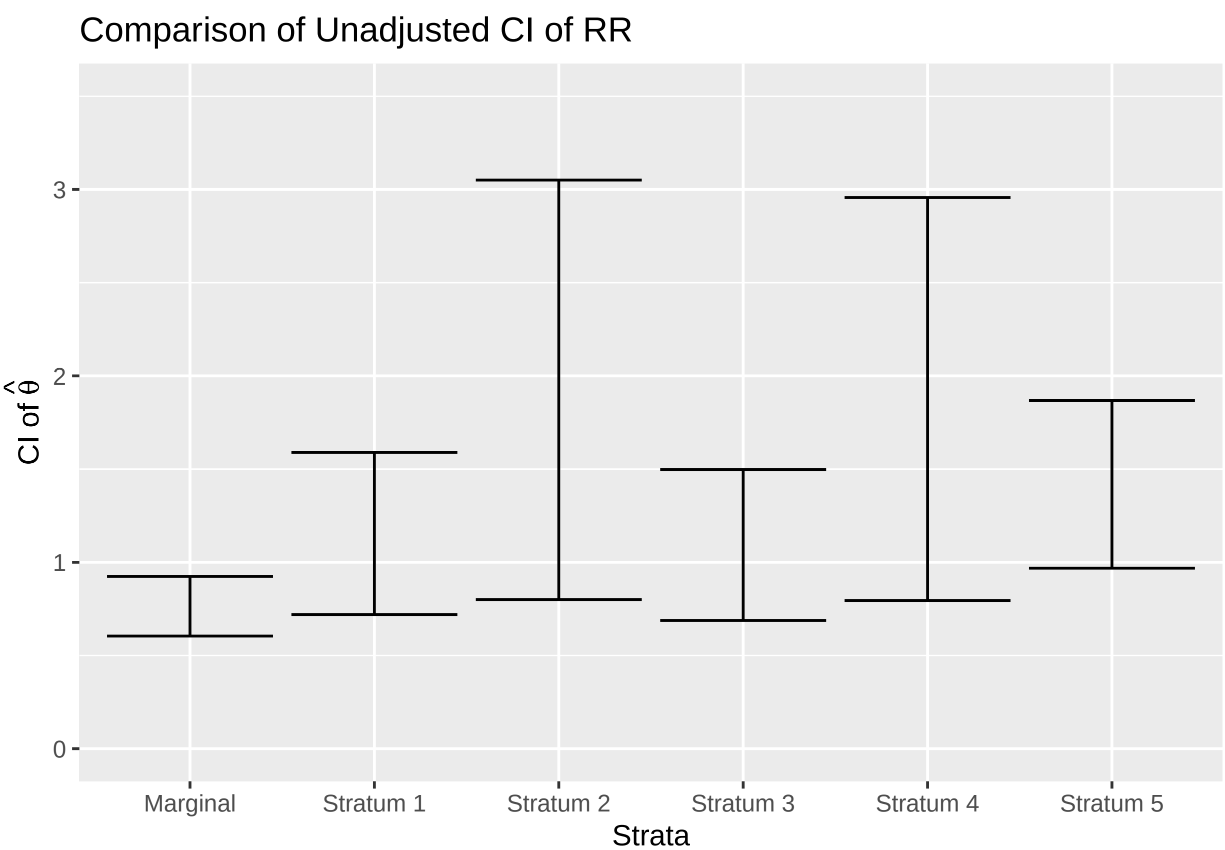
Stratified Unadjusted and Marginal

```
library(dplyr)
library(ggplot2)
library(tidyverse)
library(latex2exp)
un_est <- function(a, b, c, d) {

  n1= a + c
  n2 = b + d
  m1 = a + b
  m2 = c + d

  p1 = a/(a+c)
  p2 = b/(b+d)
  RD = p1 - p2
  RR = p1/p2
  OR = (p1/(1-p1))/(p2/(1-p2))

  log.RR = log(RR)
  log.OR = log(OR)

  var.RD = (p1 * (1 - p1))/n1 + (p2 * (1-p2))/n1
  var.logRR = 1/a - 1/n1 + 1/b + 1/n2
  var.logOR = 1/a + 1/b + 1/c + 1/d

  CI.RD = c(RD - 1.96 * sqrt(var.RD), RD + 1.96 * sqrt(var.RD))
  CI.logRR = c(log.RR - 1.96 * sqrt(var.logRR),
               log.RR + 1.96 * sqrt(var.logRR))
  CI.logOR = c(log.OR - 1.96 * sqrt(var.logOR),
               log.OR + 1.96 * sqrt(var.logOR))
  CI.RR = exp(CI.logRR)
  CI.OR = exp(CI.logOR)
```

```
  return(list(
    RD=RD,
    RR=RR,
    OR=OR,
    log.RR=log.RR,
    log.OR=log.OR,
    var.RD=var.RD,
    var.logRR=var.logRR,
    var.logOR=var.logOR,
    CI.RD=CI.RD,
    CI.RR=CI.RR,
    CI.OR=CI.OR
  ))
}

## Stratum 1
res.str1 <- un_est(18, 162, 25, 252)
## Stratum 2
res.str2 <- un_est(12, 26, 123, 431)
## Stratum 3
res.str3 <- un_est(27, 121, 104, 475)
## Stratum 4
res.str4 <- un_est(7, 21, 3, 25)
## Stratum 5
res.str5 <- un_est(14, 353, 7, 359)
## Marginal
a <- c(18, 12, 27, 7, 14)
b <- c(162, 26, 121, 21, 353)
c <- c(25, 123, 104, 3, 7)
d <- c(252, 431, 475, 25, 359)
a <- sum(a)
b <- sum(b)
c <- sum(c)
d <- sum(d)
res.mar <- un_est(a, b, c, d)

df <- tibble(res.str1$CI.RR, res.str2$CI.RR, res.str3$CI.RR, res.str4$CI.RR, res.str5$CI
df <- t(df)
df <- as.data.frame(df)
colnames(df) <- c("lower", "upper")
rownames(df) <- c("Stratum 1", "Stratum 2", "Stratum 3", "Stratum 4", "Stratum 5", "Marg
df <- rownames_to_column(df)
df <- rename(df, c("Strata" = "rowname"))
```

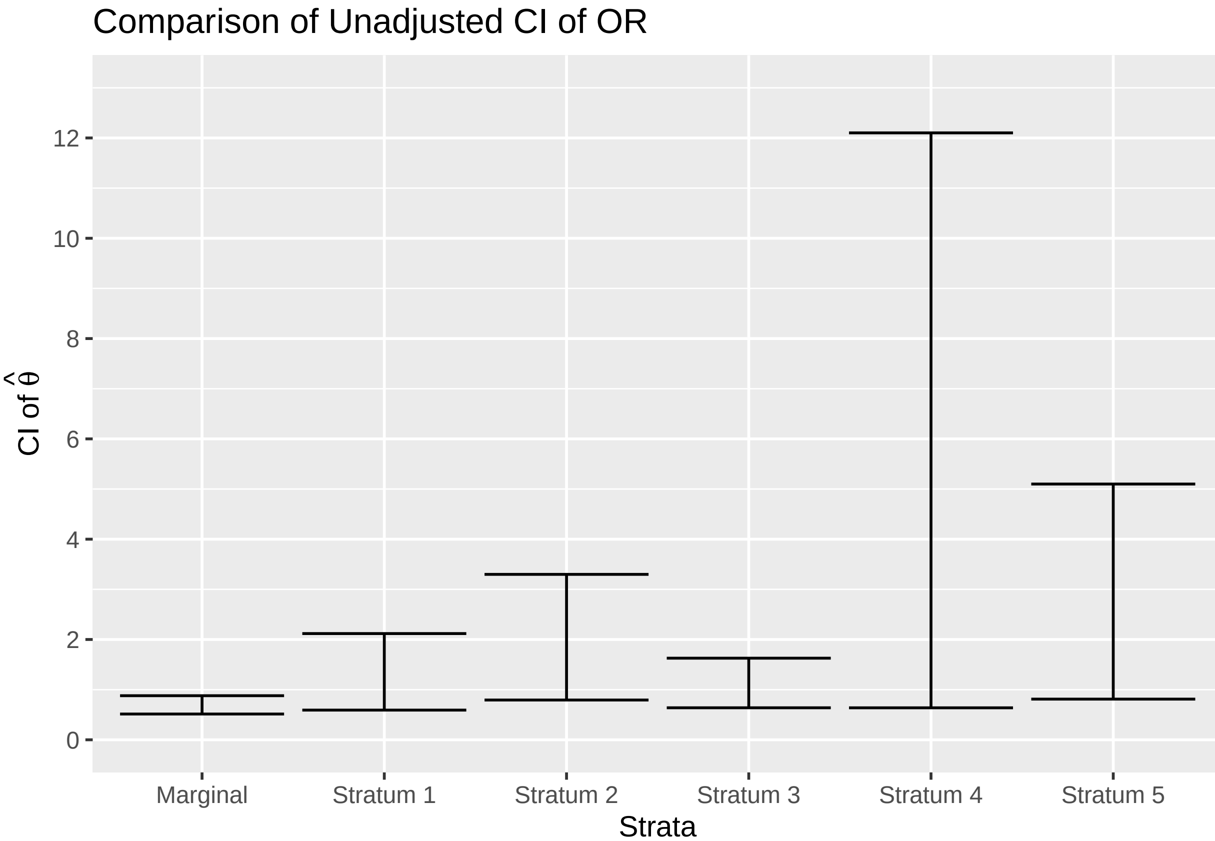
```
df.OR <- tibble(res.str1$CI.OR, res.str2$CI.OR, res.str3$CI.OR, res.str4$CI.OR, res.str5
df.OR <- t(df.OR)
df.OR <- as.data.frame(df.OR)
colnames(df.OR) <- c("lower", "upper")
rownames(df.OR) <- c("Stratum 1", "Stratum 2", "Stratum 3", "Stratum 4", "Stratum 5", "M
df.OR <- rownames_to_column(df.OR)
df.OR <- rename(df.OR, c("Strata" = "rowname"))
```

```
## Plot RR CI unadjusted
ggplot(df, aes(x=Strata)) +
  geom_errorbar(aes(ymin = lower, ymax = upper)) +
  scale_y_continuous(limits = c(0, 3.5)) +
  xlab("Strata") +
  ylab(TeX("CI of $\\hat{\\theta}$")) +
  ggtitle("Comparison of Unadjusted CI of RR")
```




```
## Plot OR CI Unadjusted
ggplot(df.OR, aes(x=Strata)) +
  geom_errorbar(aes(ymin = lower, ymax = upper)) +
  scale_y_continuous(limits = c(0, 13), breaks = seq(0,13,2)) +
  xlab("Strata") +
  ylab(TeX("CI of $\\hat{\\theta}$")) +
```

```
  ggtitle("Comparison of Unadjusted CI of OR")
```



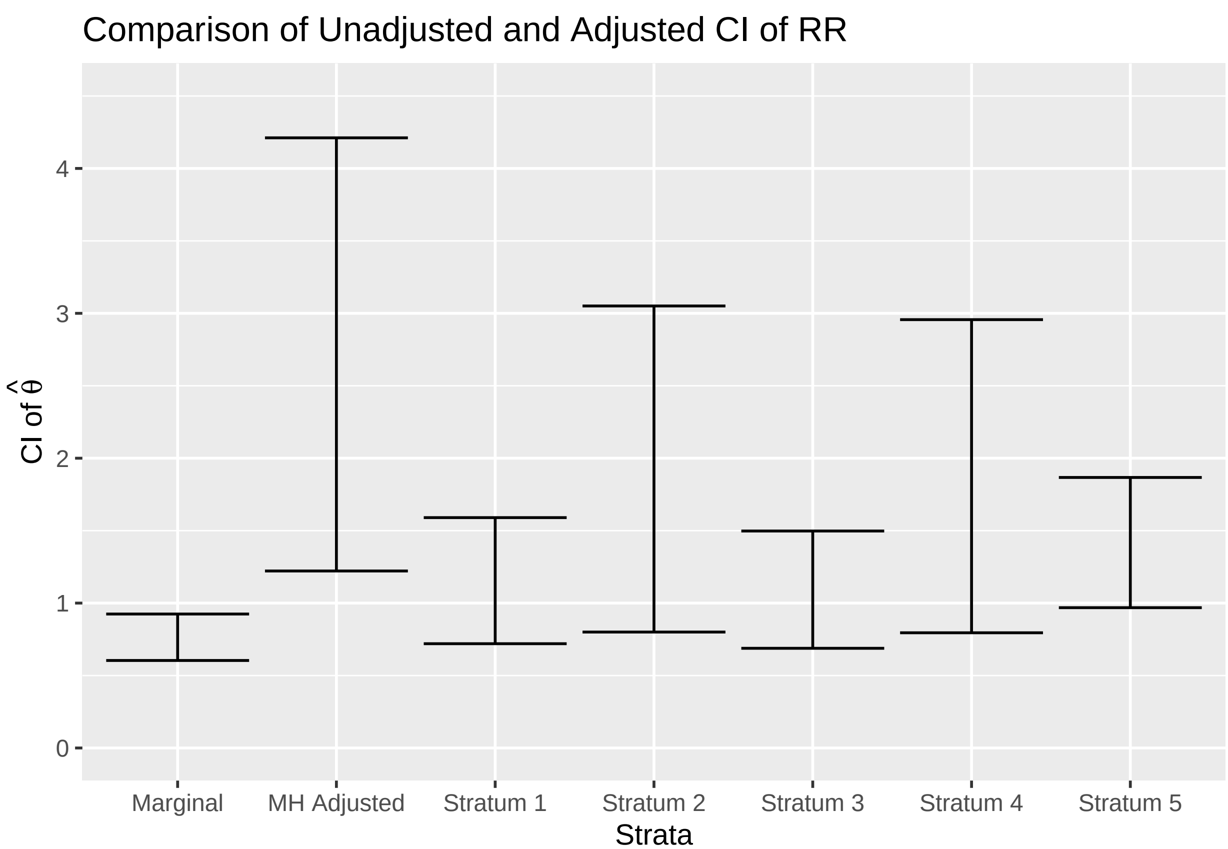
MH Stratified-Adjusted

```
voltage <- array(
  c(18, 162, 25, 252,
    12, 26, 123, 431,
    27, 121, 104, 475,
    7, 21, 3, 25,
    14, 353, 7, 359),
  dim = c(2, 2, 5),
  dimnames = list(response = c("Case", "Control"),
                  treatment = c("<100m", ">100m"),
                  strata = c("Study 1", "Study 2", "Study 3", "Study 4", "Study 5"))
)

ak <- voltage[1,1,]
bk <- voltage[1,2,]
ck <- voltage[2,1,]
dk <- voltage[2,2,]
strata.spe.data <- data.frame(ak, bk, ck, dk)
```

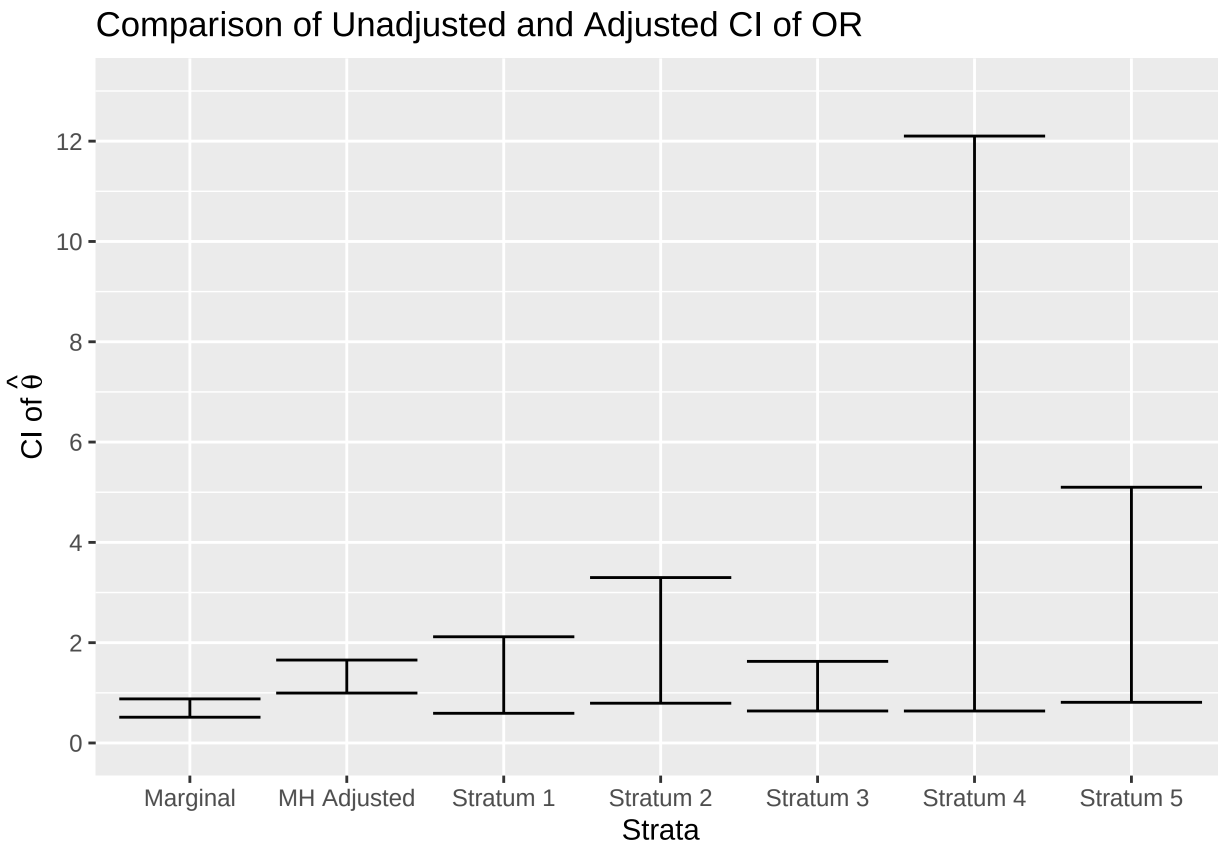
```r
MH.adj_est <- function(ak, bk, ck, dk) {

  ## Pre-check on input validation
  suppressWarnings(
   if(is.vector(any(ak, bk, ck, dk)) == FALSE) {
    stop("Input invalid")
  }
  )

  ## Sample size calculation
  Nk = ak + bk + ck + dk
  n1k = ak + ck
  n2k = bk + dk
  m1k = ak + bk
  m2k = ck + dk

  ## Calculate MH OR
  nume.OR <- sum((ak * dk)/Nk)
  deno.OR <- sum((bk * ck)/Nk)
  OR.mh <- nume.OR/deno.OR

  ## Calculate MH RR
  nume.RR <- sum((ak * n2k)/Nk)
  deno.RR <- sum((ak * n1k)/Nk)
  RR.mh <- nume.RR/deno.RR

  ## Stratified-adjusted MH test and CI
  #### Under Null test-based variance, inherently incorrect
  chi.square.MH <- ((sum(ak-n1k * (m1k/Nk)))^2)/(sum((m1k * m2k * n1k * n2k)/(Nk^2 * (Nk
  #### 5 Sums
  S1 <- sum((ak * dk)/Nk); S2 <- sum((bk * ck)/Nk); S3 <- sum(((ak + dk)*ak*dk)/Nk^2);
  S4 <- sum(((bk + ck)*bk*ck)/Nk^2); S5 <- sum(((ak+dk)*bk*ck+(bk+ck)*ak*dk)/Nk^2)
  #### RR CI
  log.RR = log(RR.mh)
  var.logRR.mh = (log.RR^2)/(chi.square.MH^2)
  CI.logRR.MH = c(log.RR - 1.96 * sqrt(var.logRR.mh),
                  log.RR + 1.96 * sqrt(var.logRR.mh))
  CI.RR = exp(CI.logRR.MH)
  #### OR CI
  log.OR = log(OR.mh)
  var.logOR.mh = S3/(2^S1^2) + S5/(2*S1*S2) + S4/(2*S2^2)
  CI.logOR.mh = c(log.OR - 1.96 * sqrt(var.logOR.mh),
                  log.OR + 1.96 * sqrt(var.logOR.mh))
  CI.OR = exp(CI.logOR.mh)
```

```r
  ## Output
  return(list(
    OR.mh = OR.mh,
    RR.mh = RR.mh,
    var.RR = var.logRR.mh,
    var.OR = var.logOR.mh,
    CI.RR = CI.RR,
    CI.OR = CI.OR
  ))
}
MH.adj_est(ak, bk, ck, dk)

adj.df.RR <- MH.adj_est(ak, bk, ck, dk)$CI.RR
adj.df.OR <- MH.adj_est(ak, bk, ck, dk)$CI.OR
```

```r
## Combined graph of Unadjusted & adjusted
df[nrow(df) + 1,] <- c("MH Adjusted", adj.df.RR)
df.OR[nrow(df.OR) + 1,] <- c("MH Adjusted", adj.df.OR)
df$lower <- as.numeric(df$lower)
df$upper <- as.numeric(df$upper)
df.OR$lower <- as.numeric(df.OR$lower)
df.OR$upper <- as.numeric(df.OR$upper)

## Plot RR CI unadjusted vs. adjusted
ggplot(df, aes(x=Strata)) +
  geom_errorbar(aes(ymin = lower, ymax = upper)) +
  scale_y_continuous(limits = c(0, 4.5), breaks = seq(0,6,1)) +
  xlab("Strata") +
  ylab(TeX("CI of $\\hat{\\theta}$")) +
  ggtitle("Comparison of Unadjusted and Adjusted CI of RR")
```




```r
## Plot OR CI Unadjusted vs. adjusted
ggplot(df.OR, aes(x=Strata)) +
  geom_errorbar(aes(ymin = lower, ymax = upper)) +
  scale_y_continuous(limits = c(0, 13), breaks = seq(0, 13, 2)) +
  xlab("Strata") +
  ylab(TeX("CI of $\\hat{\\theta}$")) +
  ggtitle("Comparison of Unadjusted and Adjusted CI of OR")
```



Conclusion

Based on those graphs that compare different stratum and unadjusted vs. adjusted CIs generated through R code, we can conclude that marginal unadjusted analysis return a smaller estimate and narrower CI. MH adjusted analysis gives higher range of RR than unadjusted analysis, it also returns higher estimate on the measure of both RR and OR. It may be caused by the imbalanecs in the number of observing objects from each stratum, that some of strata have more participants than others.

Table 1: Marginal and Stratified Unadjusted Measure

	Stratum					
Measure	1	2	3	4	5	Marginal
RD	0.027	0.032	0.003	0.243	0.171	-0.078
$V(RD)$	0.011	0.001	0.002	0.046	0.022	0.001
95% CI for RD	(-0.18,0.23)	(-0.03,0.09)	(-0.09,0.10)	(-0.18,0.66)	(-0.12,0.46)	(-0.14,0.01)
RR	1.070	1.562	1.015	1.533	1.345	0.747
$\log(RR)$	0.067	0.446	0.015	0.427	0.296	-0.291
$V(\log(RR))$	0.041	0.117	0.039	0.112	0.028	0.012
95% CI for RR	(0.72,1.59)	(0.80,3.05)	(0.69,1.50)	(0.80,2.96)	(0.97,1.87)	(0.60,0.92)
OR	1.120	1.617	1.019	2.778	2.034	0.672
$\log(OR)$	0.113	0.481	0.019	1.022	0.710	-0.397
$V(\log(OR))$	0.106	0.132	0.057	0.564	0.220	0.019
95% CI for OR	(0.59,2.12)	(0.79, 3.30)	(0.64,1.63)	(0.64,12.10)	(0.81, 5.10)	(0.51, 0.88)

Table 2: MH Adjusted Measure

Measure	MH Estimate	95% CI
OR	1.283	(1.06,1.55)
$V(OR_{MH})$	0.017	
RR	2.268	(1.22,4.21)
$V(RR_{MH})$	0.100	