analy_theta

Haodong Li

2022-11-28

1. Experiment and question

The LEADER trial is a multi-centre randomised trial to determine Liraglutide (GLP-1) effects on cardiovascular events. So the original treatment variable is Liraglutide/Placebo and outcome consists of several cardiovascular events. It is known that GLP-1 users are also often on statins and other cardiovascular drugs. In this project, we want to explore the heterogeneity of liraglutide treatment effect among concomitant medication (especially statin) users, with diabetes progression as primary outcome, cardiovascular events as second outcome.

2. Overview of data

```
outcome = 'diab'; t = 24
df <- get_data(outcome, t)

nodes <- list(W = setdiff(names(df), c("Y", "A")),
              A = 'A',
              Y = 'Y')

p_before = ncol(df)-2
df <- process_missing(df, nodes)$data
print(paste0("data dimension (before): n=", nrow(df), ", p=", p_before))
```

```
## [1] "data dimension (before): n=4169, p=51"
```

```
print(paste0("data dimension (after): n=", nrow(df), ", p=", ncol(df)-2))
```

```
## [1] "data dimension (after): n=4169, p=61"
```

```
# Table of Baseline Characteristics
df_w_summary <- df_w %>%
  filter(INSNVFL == FALSE) %>%
  select(-c("USUBJID", "INSNVFL")) %>%
  mutate(A = ifelse(A == 1, "Liraglutide", "Placebo"))

df_w_summary <- labelled::remove_labels(df_w_summary)
tbl <- table1(~. | A, data = df_w_summary, caption = "Baseline characteristics in LEADER")
# kable(as.data.frame(tbl), longtable=TRUE, booktabs=TRUE)
```

```
tbl %>% t1kable(longtable=TRUE) %>%
  kable_styling(., font_size = 9, latex_options = "hold_position", position = "center")
```

Table 1: Baseline characteristics in LEADER

	Liraglutide	Placebo	Overall
	(N=2038)	(N=2131)	(N=4169)
AGE			
Mean (SD)	64.5 (7.22)	64.5 (7.13)	64.5 (7.17)
Median [Min, Max]	64.0 [50.0, 87.0]	64.0 [50.0, 88.0]	64.0 [50.0, 88.0]
Missing	1 (0.0%)	0 (0%)	1 (0.0%)
SEX			
F	764 (37.5%)	829 (38.9%)	1593 (38.2%)
M	1274 (62.5%)	1302 (61.1%)	2576 (61.8%)
RACE			
ASIAN	199 (9.8%)	210 (9.9%)	409 (9.8%)
BLACK	185 (9.1%)	223 (10.5%)	408 (9.8%)
OTHER	89 (4.4%)	76 (3.6%)	165 (4.0%)
WHITE	1565 (76.8%)	1622 (76.1%)	3187 (76.4%)
COUNTRY			
Africa	64 (3.1%)	86 (4.0%)	150 (3.6%)
America	956 (46.9%)	1011 (47.4%)	1967 (47.2%)
Asia	337 (16.5%)	330 (15.5%)	667 (16.0%)
Europe	615 (30.2%)	644 (30.2%)	1259 (30.2%)
Pacific	66 (3.2%)	60 (2.8%)	126 (3.0%)
SMOKER			
CURRENT SMOKER	236 (11.6%)	246 (11.5%)	482 (11.6%)
NEVER SMOKED	853 (41.9%)	886 (41.6%)	1739 (41.7%)
PREVIOUS SMOKER	949 (46.6%)	999 (46.9%)	1948 (46.7%)
NYHACLAS			
NYHA CLAS NA	1672 (82.0%)	1760 (82.6%)	3432 (82.3%)
NYHA CLASS I	75 (3.7%)	78 (3.7%)	153 (3.7%)
NYHA CLASS II	244 (12.0%)	244 (11.5%)	488 (11.7%)
NYHA CLASS III	47 (2.3%)	49 (2.3%)	96 (2.3%)
DIABDUR			
Mean (SD)	15.5 (7.91)	15.3 (8.01)	15.4 (7.96)
Median [Min, Max]	14.2 [0.100, 54.9]	14.3 [0.100, 61.0]	14.3 [0.100, 61.0]
Missing	4 (0.2%)	3 (0.1%)	7 (0.2%)
ANTDBFL			
Insulin-OAD	361 (17.7%)	377 (17.7%)	738 (17.7%)
Insulin+OAD(s)	1677 (82.3%)	1754 (82.3%)	3431 (82.3%)
AHYPERFL			
Yes	192 (9.4%)	189 (8.9%)	381 (9.1%)
No	1846 (90.6%)	1942 (91.1%)	3788 (90.9%)
INCPASSN			
Mean (SD)	5.00 (0.0585)	4.99 (0.0717)	5.00 (0.0656)
Median [Min, Max]	5.00 [4.00, 5.00]	5.00 [4.00, 5.00]	5.00 [4.00, 5.00]
BMIBL			
Mean (SD)	32.8 (6.18)	33.0 (6.47)	32.9 (6.33)
Median [Min, Max]	32.0 [19.7, 66.5]	32.0 [18.0, 81.0]	32.0 [18.0, 81.0]
Missing	3 (0.1%)	2 (0.1%)	5 (0.1%)
PULSEBL			
Mean (SD)	72.8 (11.3)	72.8 (11.6)	72.8 (11.4)
Median [Min, Max]	72.0 [37.0, 172]	72.0 [35.0, 121]	72.0 [35.0, 172]
SYSBPBL			

Mean (SD)	136 (18.7)	136 (17.9)	136 (18.3)
Median [Min, Max]	135 [74.0, 222]	135 [72.5, 227]	135 [72.5, 227]
DIABPBL			
Mean (SD)	76.1 (10.4)	76.0 (10.3)	76.1 (10.3)
Median [Min, Max]	76.5 [43.5, 128]	76.5 [38.5, 123]	76.5 [38.5, 128]
HBA1CBL			
Mean (SD)	8.93 (1.61)	8.84 (1.52)	8.88 (1.56)
Median [Min, Max]	8.50 [5.10, 18.5]	8.50 [5.80, 16.7]	8.50 [5.10, 18.5]
HDL1BL			
Mean (SD)	1.17 (0.322)	1.16 (0.314)	1.17 (0.318)
Median [Min, Max]	1.13 [0.270, 3.21]	1.11 [0.0800, 3.32]	1.11 [0.0800, 3.32]
Missing	40 (2.0%)	33 (1.5%)	73 (1.8%)
LDL1BL			
Mean (SD)	2.29 (0.915)	2.30 (0.935)	2.30 (0.925)
Median [Min, Max]	2.12 [0.360, 8.23]	2.14 [0.200, 7.93]	2.12 [0.200, 8.23]
Missing	40 (2.0%)	33 (1.5%)	73 (1.8%)
CHOL1BL			
Mean (SD)	4.35 (1.11)	4.35 (1.15)	4.35 (1.13)
Median [Min, Max]	4.17 [1.86, 9.35]	4.17 [1.81, 12.6]	4.17 [1.81, 12.6]
Missing	40 (2.0%)	33 (1.5%)	73 (1.8%)
RC			
Mean (SD)	0.883 (0.507)	0.882 (0.565)	0.883 (0.538)
Median [Min, Max]	0.750 [-0.910, 4.87]	0.750 [0.150, 11.0]	0.750 [-0.910, 11.0]
Missing	40 (2.0%)	33 (1.5%)	73 (1.8%)
RCoverHDL			
Mean (SD)	0.864 (0.717)	0.868 (0.800)	0.866 (0.760)
Median [Min, Max]	0.669 [-0.820, 10.0]	0.663 [0.0753, 14.7]	0.667 [-0.820, 14.7]
Missing	40 (2.0%)	33 (1.5%)	73 (1.8%)
TRIG1BL			
Mean (SD)	2.03 (1.36)	2.03 (1.65)	2.03 (1.52)
Median [Min, Max]	1.66 [0.440, 13.3]	1.65 [0.360, 36.1]	1.65 [0.360, 36.1]
Missing	40 (2.0%)	33 (1.5%)	73 (1.8%)
CREATBL			
Mean (SD)	92.0 (43.7)	91.1 (43.3)	91.5 (43.5)
Median [Min, Max]	81.5 [33.0, 609]	81.0 [20.0, 726]	81.0 [20.0, 726]
EGFMDRBC			
Mean (SD)	77.0 (27.8)	77.7 (28.4)	77.4 (28.1)
Median [Min, Max]	76.2 [7.30, 200]	77.0 [6.80, 419]	76.5 [6.80, 419]
RETINSEV			
no retinopathy	32 (1.6%)	41 (1.9%)	73 (1.8%)
non-proliferative	435 (21.3%)	397 (18.6%)	832 (20.0%)
proliferative	154 (7.6%)	162 (7.6%)	316 (7.6%)
retinsev NA	1417 (69.5%)	1531 (71.8%)	2948 (70.7%)
GERDBLFL			
Yes	943 (46.3%)	910 (42.7%)	1853 (44.4%)
No	1095 (53.7%)	1221 (57.3%)	2316 (55.6%)
PPIFL			
Yes	475 (23.3%)	495 (23.2%)	970 (23.3%)
No	1563 (76.7%)	1636 (76.8%)	3199 (76.7%)
H2BLFL			
Yes	83 (4.1%)	75 (3.5%)	158 (3.8%)
No	1955 (95.9%)	2056 (96.5%)	4011 (96.2%)
MIFL			
Yes	652 (32.0%)	658 (30.9%)	1310 (31.4%)
No	1386 (68.0%)	1473 (69.1%)	2859 (68.6%)

STROKEFL			
Yes	345 (16.9%)	370 (17.4%)	715 (17.2%)
No	1693 (83.1%)	1761 (82.6%)	3454 (82.8%)
REVASFL			
Yes	822 (40.3%)	900 (42.2%)	1722 (41.3%)
No	1216 (59.7%)	1231 (57.8%)	2447 (58.7%)
STENFL			
Yes	539 (26.4%)	546 (25.6%)	1085 (26.0%)
No	1499 (73.6%)	1585 (74.4%)	3084 (74.0%)
CHDFL			
Yes	182 (8.9%)	192 (9.0%)	374 (9.0%)
No	1856 (91.1%)	1939 (91.0%)	3795 (91.0%)
IHDFL			
Yes	522 (25.6%)	550 (25.8%)	1072 (25.7%)
No	1516 (74.4%)	1581 (74.2%)	3097 (74.3%)
CHFFL			
Yes	291 (14.3%)	293 (13.7%)	584 (14.0%)
No	1747 (85.7%)	1838 (86.3%)	3585 (86.0%)
KIDFL			
Yes	627 (30.8%)	623 (29.2%)	1250 (30.0%)
No	1411 (69.2%)	1508 (70.8%)	2919 (70.0%)
MICFL			
Yes	563 (27.6%)	619 (29.0%)	1182 (28.4%)
No	1475 (72.4%)	1512 (71.0%)	2987 (71.6%)
HYPFL			
Yes	488 (23.9%)	515 (24.2%)	1003 (24.1%)
No	1550 (76.1%)	1616 (75.8%)	3166 (75.9%)
LVSDFL			
Yes	472 (23.2%)	495 (23.2%)	967 (23.2%)
No	1566 (76.8%)	1636 (76.8%)	3202 (76.8%)
PADFL			
Yes	274 (13.4%)	300 (14.1%)	574 (13.8%)
No	1764 (86.6%)	1831 (85.9%)	3595 (86.2%)
CVRISK			
High	1717 (84.2%)	1762 (82.7%)	3479 (83.4%)
Medium	321 (15.8%)	369 (17.3%)	690 (16.6%)
HBA1CGRN			
Mean (SD)	1.56 (0.497)	1.55 (0.498)	1.55 (0.497)
Median [Min, Max]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]
DDURGRN			
Mean (SD)	1.68 (0.467)	1.66 (0.472)	1.67 (0.470)
Median [Min, Max]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]	2.00 [1.00, 2.00]
Missing	4 (0.2%)	3 (0.1%)	7 (0.2%)
statin_use			
Yes	1558 (76.4%)	1591 (74.7%)	3149 (75.5%)
No	480 (23.6%)	540 (25.3%)	1020 (24.5%)
antihypertensives			
Yes	1693 (83.1%)	1751 (82.2%)	3444 (82.6%)
No	345 (16.9%)	380 (17.8%)	725 (17.4%)
betab			
Yes	1186 (58.2%)	1219 (57.2%)	2405 (57.7%)
No	852 (41.8%)	912 (42.8%)	1764 (42.3%)
minera			
Yes	135 (6.6%)	141 (6.6%)	276 (6.6%)

No	1903 (93.4%)	1990 (93.4%)	3893 (93.4%)
adp			
Yes	1511 (74.1%)	1567 (73.5%)	3078 (73.8%)
No	527 (25.9%)	564 (26.5%)	1091 (26.2%)
vkantag			
Yes	135 (6.6%)	137 (6.4%)	272 (6.5%)
No	1903 (93.4%)	1994 (93.6%)	3897 (93.5%)
caantag			
Yes	713 (35.0%)	708 (33.2%)	1421 (34.1%)
No	1325 (65.0%)	1423 (66.8%)	2748 (65.9%)
thiazide			
Yes	393 (19.3%)	380 (17.8%)	773 (18.5%)
No	1645 (80.7%)	1751 (82.2%)	3396 (81.5%)
loopdiur			
Yes	433 (21.2%)	486 (22.8%)	919 (22.0%)
No	1605 (78.8%)	1645 (77.2%)	3250 (78.0%)

3. CATE estimation

```
# generate risk table
# risk and cate est
# temp = cbind(df, "mua_hat" = df_fit$mua_hat, "tau" = df_fit$tau) %>%
#   group_by(A, statin_use) %>%
#   summarise(EY = mean(mua_hat), "cate_d" = mean(tau))
#
# temp$statin_use <- as.logical(temp$statin_use)
# cate_t1 = temp$EY[which(temp$statin_use & temp$A == 1)] - temp$EY[which(temp$statin_use & temp$A == 0
# cate_t0 = temp$EY[which(!temp$statin_use & temp$A == 1)] - temp$EY[which(!temp$statin_use & temp$A ==
# temp$cate_t = c(cate_t0,cate_t1,cate_t0,cate_t1)
# temp$cate_d = rep(c(mean(temp$cate_d[c(1,3)]), mean(temp$cate_d[c(2,4)])), 2)
# #empirical risk
# tmp_Y = df %>%
#   group_by(A, statin_use) %>%
#   mutate(statin_use  = as.logical(statin_use)) %>%
#   summarise(count_Y = sum(Y),
#             group_N = n(),
#             risk_emp = sum(Y)/n())
# temp <- left_join(temp, tmp_Y, by = c("A", "statin_use"))
cm_names <- c("statin_use", "antihypertensives", "betab", "minera", "adp",
              "vkantag", "caantag", "thiazide", "loopdiur")

get_df_risk <- function(df, df_fit, cm_names){
  df_risk = data.frame()
  for (cm in cm_names){
    temp = cbind(df, "mua_hat" = df_fit$mua_hat, "tau" = df_fit$tau) %>%
    group_by(across(all_of(c("A", cm)))) %>%
    summarise(EY = mean(mua_hat), "cate_d" = mean(tau))

    cate_t1 = temp$EY[which(temp[,cm] == "TRUE" & temp$A == 1)] -
      temp$EY[which(temp[,cm] == "TRUE" & temp$A == 0)]
    cate_t0 = temp$EY[which(temp[,cm] == "FALSE" & temp$A == 1)] -
      temp$EY[which(temp[,cm] == "FALSE" & temp$A == 0)]
    temp$cate_t = c(cate_t0,cate_t1,cate_t0,cate_t1)
```

```
    temp$cate_d = rep(c(mean(temp$cate_d[c(1,3)]), mean(temp$cate_d[c(2,4)])), 2)
    #empirical risk
    tmp_Y = df %>%
      group_by(across(all_of(c("A", cm)))) %>%
      summarise(count_Y = sum(Y),
                group_N = n(),
                risk_emp = sum(Y)/n())
    temp <- left_join(temp, tmp_Y, by = c("A", cm))

    tmp <- temp %>% rename("risk_n" = "EY") %>%
      select(all_of(c("A", cm, "count_Y", "group_N",
                      "risk_emp","risk_n", "cate_t", "cate_d")))

    tmp <- tmp %>% mutate(across(all_of(c(cm)), ~paste0(str_sub(., 1,1),"_", str_sub(cm, 1,6))))
    colnames(tmp) <- c("A", "subgroup", "count_Y", "group_N",
                       "risk_emp","risk_n", "cate_t", "cate_d")
    df_risk <- rbind(df_risk, tmp)
  }
  return(df_risk)
}

tmp <- get_df_risk(df, df_fit, cm_names)
```

```
tmp[, -c(7,8)] %>%
  mutate(across(where(is.numeric), ~round(., 4))) %>%
  kable("latex", booktabs = T, linesep = c("", "", "", "\\addlinespace"),
        caption = "Subgroup risk estimates in LEADER", align=rep('c', 6)) %>%
  # collapse_rows(columns = 1, latex_hline = "major", valign = "middle") %>%
  column_spec(1:6, width = "0.6in") %>%
  kable_styling(font_size = 9)
```

Table 2: Subgroup risk estimates in LEADER

A	subgroup	count_Y	group_N	risk_emp	risk_n
0	F_statin	111	540	0.2056	0.1983
0	T_statin	300	1591	0.1886	0.1811
1	F_statin	82	480	0.1708	0.1591
1	T_statin	189	1558	0.1213	0.1408
0	F_antihy	73	380	0.1921	0.1912
0	T_antihy	338	1751	0.1930	0.1842
1	F_antihy	44	345	0.1275	0.1556
1	T_antihy	227	1693	0.1341	0.1430
0	F_betab	184	912	0.2018	0.1879
0	T_betab	227	1219	0.1862	0.1836
1	F_betab	113	852	0.1326	0.1497
1	T_betab	158	1186	0.1332	0.1419
0	F_minera	386	1990	0.1940	0.1854
0	T_minera	25	141	0.1773	0.1854
1	F_minera	256	1903	0.1345	0.1458
1	T_minera	15	135	0.1111	0.1362
0	F_adp	105	564	0.1862	0.1878
0	T_adp	306	1567	0.1953	0.1846
1	F_adp	70	527	0.1328	0.1494
1	T_adp	201	1511	0.1330	0.1437
0	F_vkanta	378	1994	0.1896	0.1856
0	T_vkanta	33	137	0.2409	0.1831
1	F_vkanta	254	1903	0.1335	0.1461
1	T_vkanta	17	135	0.1259	0.1322
0	F_caanta	279	1423	0.1961	0.1873
0	T_caanta	132	708	0.1864	0.1817
1	F_caanta	180	1325	0.1358	0.1455
1	T_caanta	91	713	0.1276	0.1445
0	F_thiazi	333	1751	0.1902	0.1842
0	T_thiazi	78	380	0.2053	0.1913
1	F_thiazi	208	1645	0.1264	0.1446
1	T_thiazi	63	393	0.1603	0.1475
0	F_loopdi	328	1645	0.1994	0.1864
0	T_loopdi	83	486	0.1708	0.1823
1	F_loopdi	214	1605	0.1333	0.1471
1	T_loopdi	57	433	0.1316	0.1378

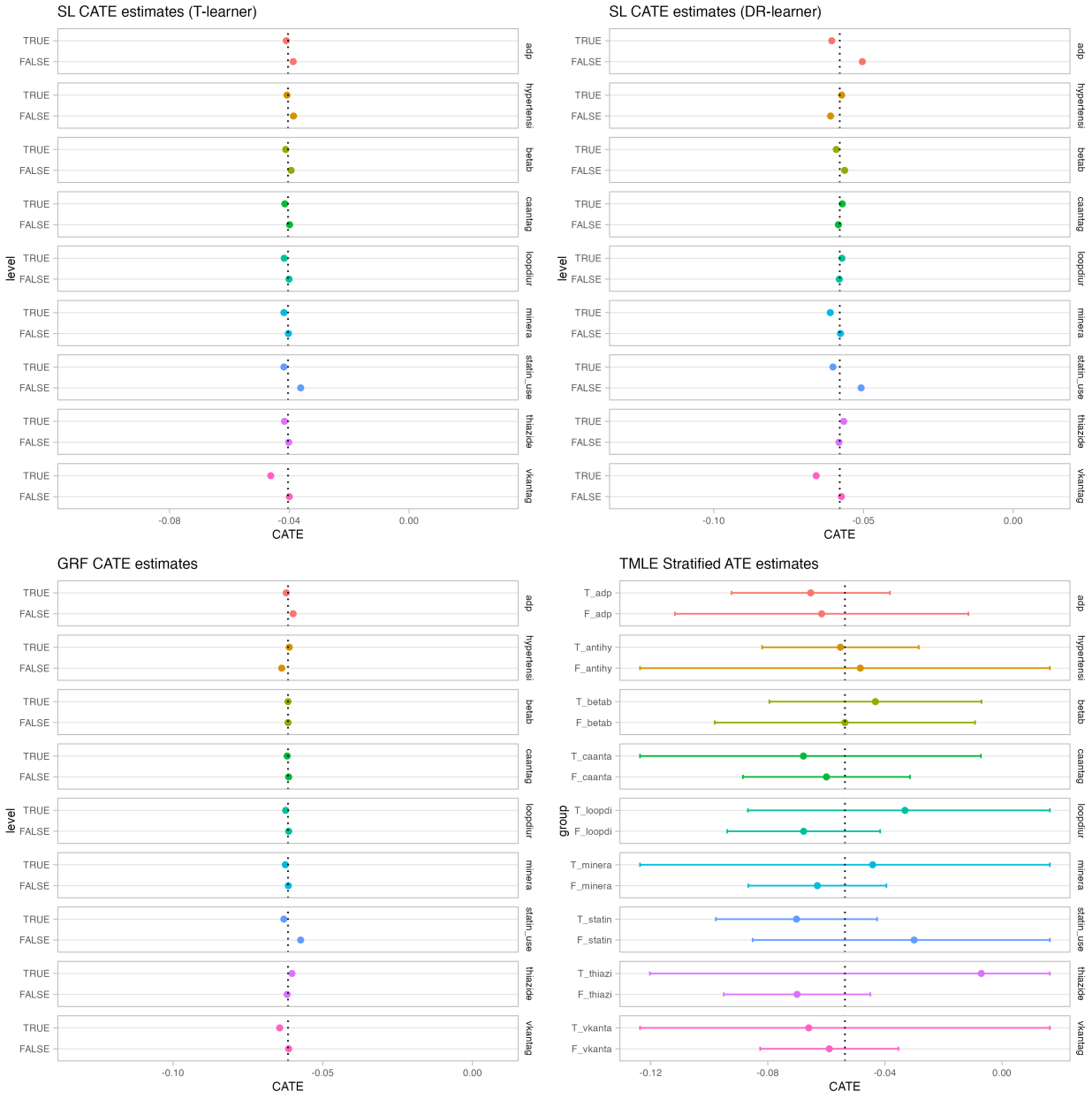

Figure 1: CATE estiamtes

4. VTE estimation

```
# VTE
# df_fit$tau <- df_fit$mu1_hat - df_fit$mu0_hat
aipw_vte <- VTE(df_fit, method="AIPW")
tmle_vte <- VTE(df_fit, method="TMLE")
```

```
aipw_vte
```

```
##
## Call:
## VTE(data = df_fit, method = "AIPW")
```

```
##
## Estimator: AIPW
## Estimate Values:
##            Estimate   Std.Error     CI_L        CI_U Wald.value  Wald.pval
## ATE     -0.06141248  0.01087298 -0.08272 -0.04010144    31.9019 1.6216e-08 ***
## rootVTE  0.20028062  0.01158241  0.17758  0.22298215   299.0060 < 2.22e-16 ***
## VTE      0.04011233  0.00231973  0.03557  0.04465900   299.0060 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
tmle_vte
```

```
##
## Call:
## VTE(data = df_fit, method = "TMLE")
##
## Estimator: TMLE
## Estimate Values:
##            Estimate   Std.Error     CI_L        CI_U Wald.value  Wald.pval
## ATE     -0.06021730  0.01083373 -0.08145 -0.03898319    30.8949 2.7239e-08 ***
## rootVTE  0.12140251  0.02729289  0.06791  0.17489657    19.7859 8.6618e-06 ***
## VTE      0.01473857  0.00331343  0.00824  0.02123288    19.7859 8.6618e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5. VIM estimation

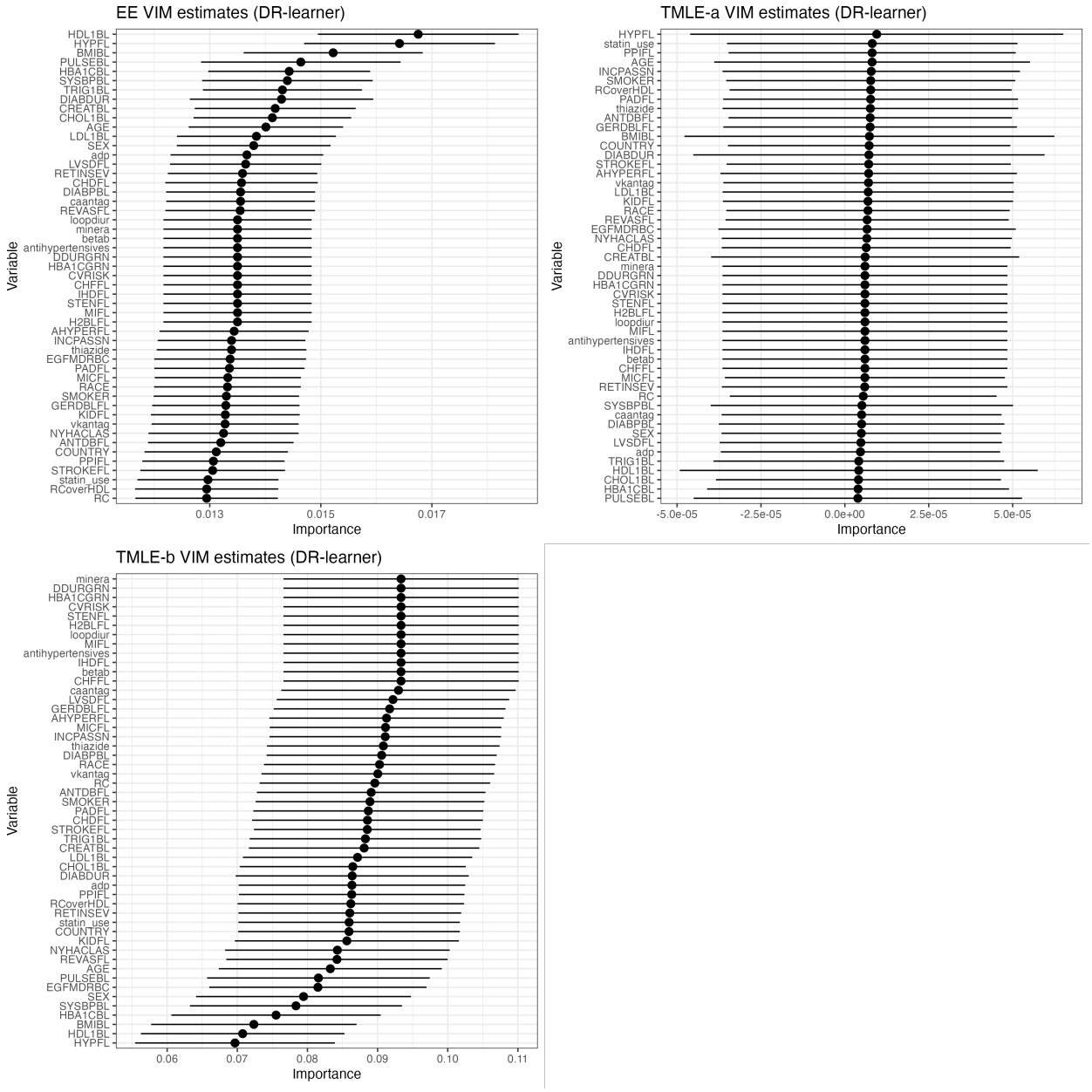

Figure 2: VIM estiamtes