

# UC2 Sentinel project: evaluation of missing pattern of body mass index (BMI) in flu and covid cohorts of Kaiser Permanente Washington EHR

Noorie Hyun

6/19/2024

- Summarize variables with missing values
- visualization of missing patterns of BMI
  - By group (Covid vs. Flu)
  - By the primary outcome of arterial thromboembolism (ATE) event (1: event and 0: no event during 90 days after hospitalization)
- Group 1 diagnosis
  - Absolute Standardized Mean Differences (ASMD) of BMI
    - By the Primary ATE outcome
  - ASMD of BMI2
  - Statistical tests for the null hypothesis of MCAR of BMI
    - Two-Sample Hotelling's T-squared test for missing BMI
    - Two-Sample Hotelling's T-squared test for missing BMI2
  - Little Single global chi-square p-value (Little) cross all missing variables WITH the null hypothesis of MCAR
    - BMI
    - BMI2
- Group 2 diagnosis focusing on MAR
  - BMI
  - BMI2
- Group 3 diagnosis
- Reference

Output was generated with R version 4.4.0 (2024-04-24 ucrt).

This R-markdown result is a part of the Supplementary Material for Pocobelli et al., which is a product of the UC2-sentinel project.

## Summarize variables with missing values

- zip3: zip code
- bmi: median of bmi measured within 90 days prior to hospital admission date
- bmi2: median of bmi measured within 5 years prior to hospital admission date

```
smdi_summarize(ate.ip)
```

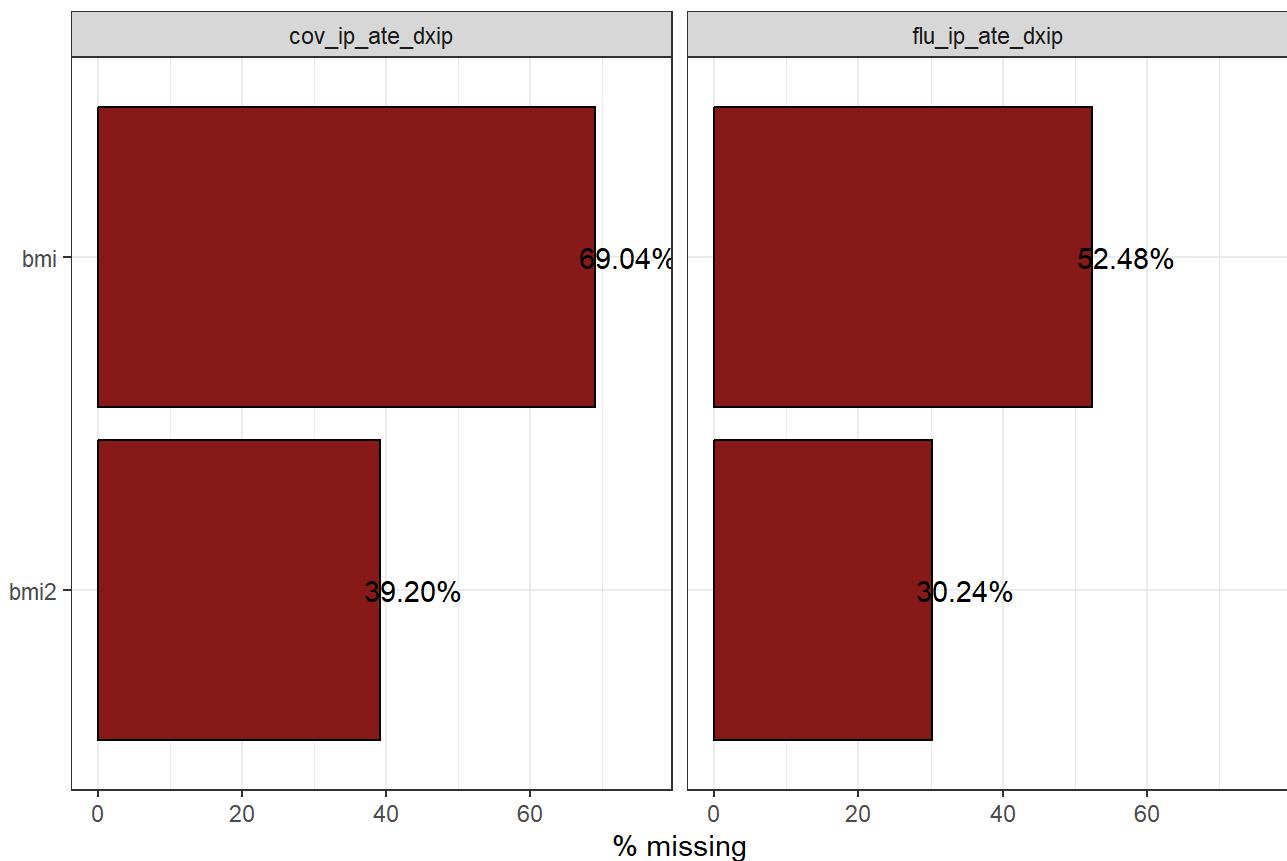
```
## # A tibble: 3 × 4
##   covariate n_miss prop_miss prop_miss_label
##   <chr>      <int>    <dbl> <chr>
## 1 bmi          553      60.6 60.64%
## 2 zip3          326      35.7 35.75%
## 3 bmi2          316      34.6 34.65%
```

## visualization of missing patterns of BMI

By group (Covid vs. Flu)

```
smdi_vis(data=ate.ip, covar=c("bmi", "bmi2"), strata="group")
```

Results stratified by group

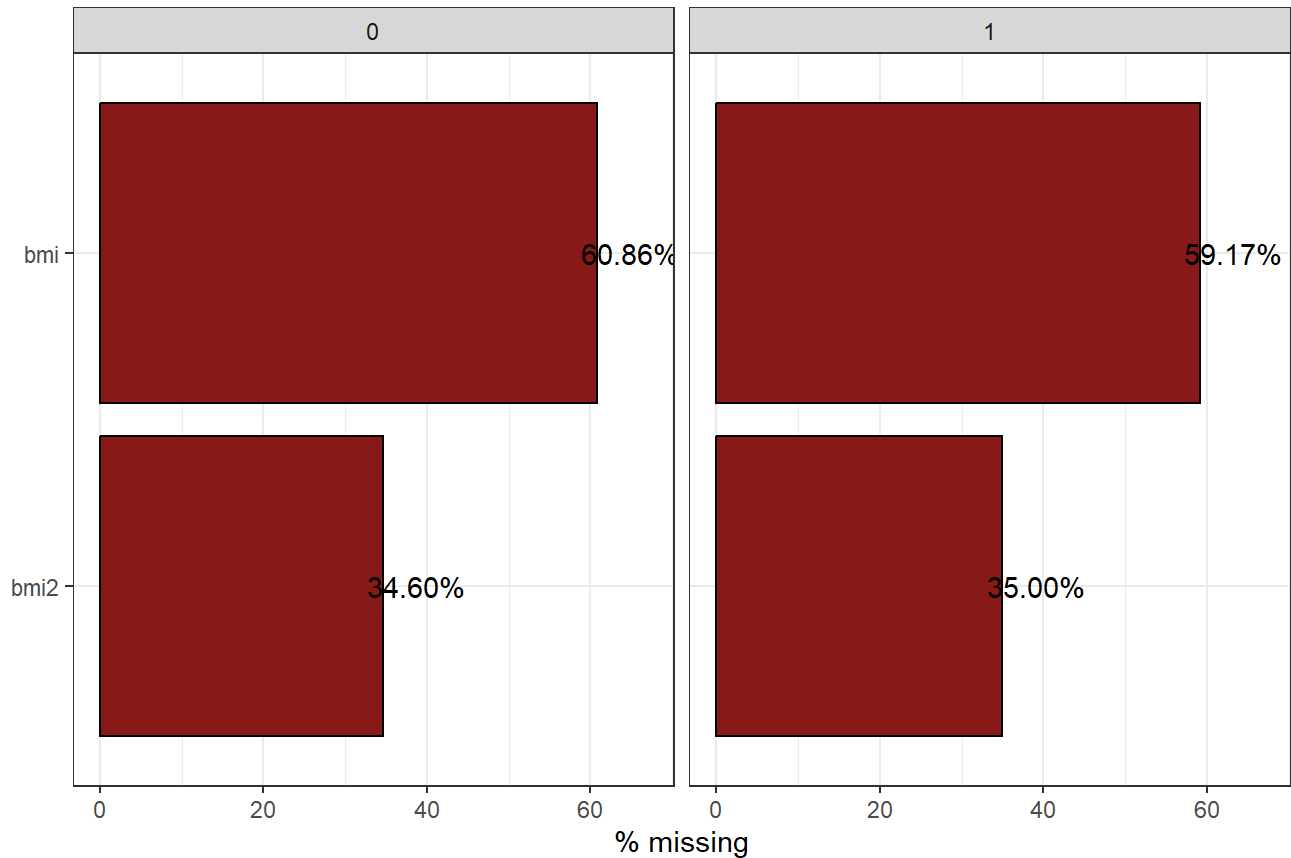


% refer to the number of observations in each stratum of group.

By the primary outcome of arterial thromboembolism (ATE) event (1: event and 0: no event during 90 days after hospitalization)

```
smdi_vis(data=ate.ip, covar=c("bmi", "bmi2"), strata="Event")
```

## Results stratified by Event



% refer to the number of observations in each stratum of Event.

## Group 1 diagnosis

We compute absolute standardized mean differences (ASMDs) of observed patient characteristics. If ASMDs are close to 0 (say,  $<0.1$ ), this implies the missing pattern of BMI does not depend on observed covariates, that is, Missing Completely at Random (MCAR) or Missing Not at Random (MNAR).

## Absolute Standardized Mean Differences (ASMD) of BMI

By the Primary ATE outcome

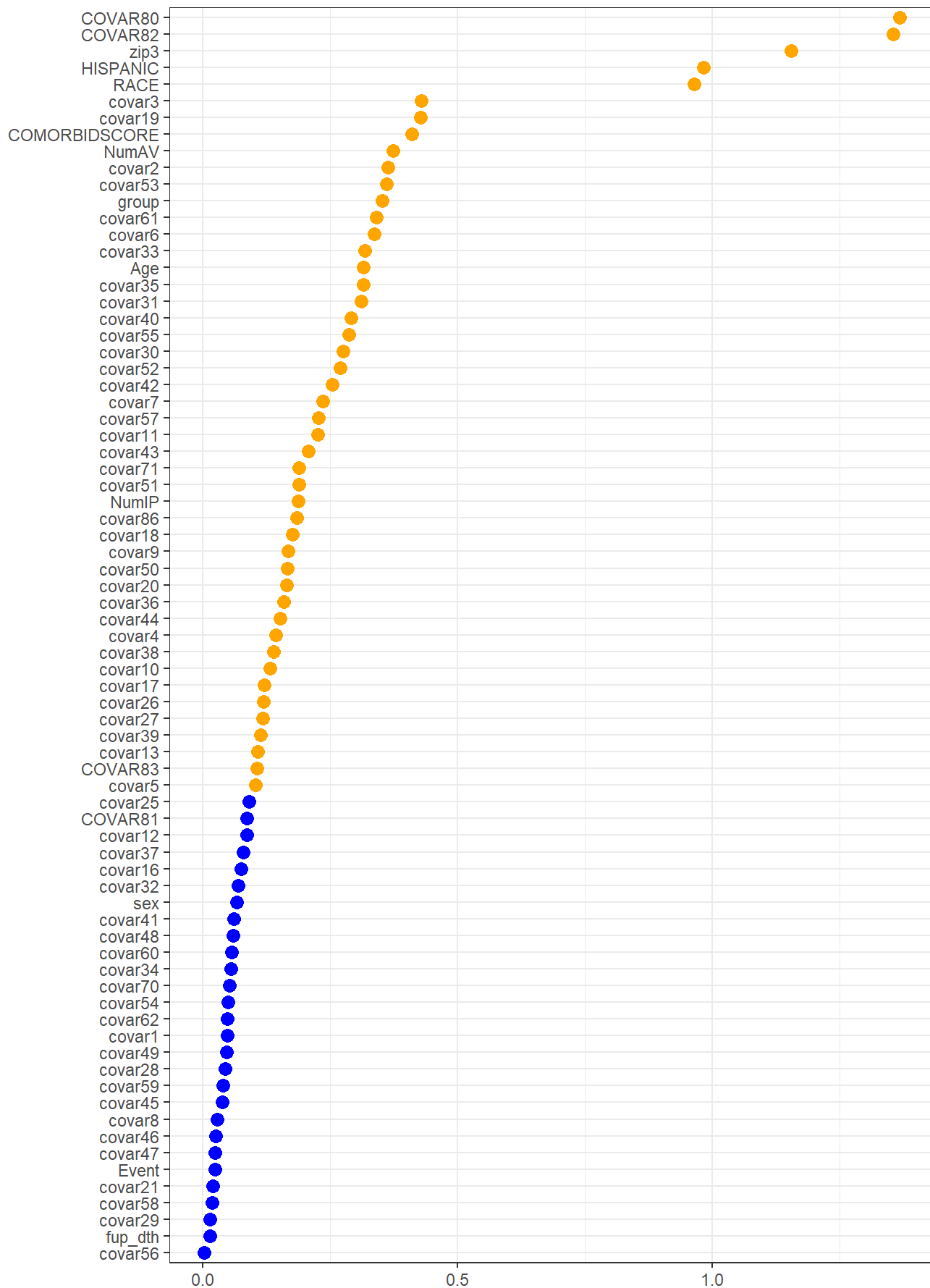
```
bmi2.ind<-which(names(ate.ip)=="bmi2")
out<-smdi_asmd(data=ate.ip[, -bmi2.ind], covar="bmi" , includeNA=FALSE)

out$bmi$asmd_aggregate
```

```
## # A tibble: 1 × 4
##   covariate asmd_median asmd_min asmd_max
##   <chr>      <chr>      <chr>  <chr>
## 1 bmi       0.143      0.002  1.368
```

```
#out$bmi$asmd_table1
out$bmi$asmd_plot
```

## ASMD plot for covariate bmi



Absolute standardized mean difference [ASMD]

ASMD is computed as the asmd between patients with and without observed 'bmi'

## ASMD of BMI2

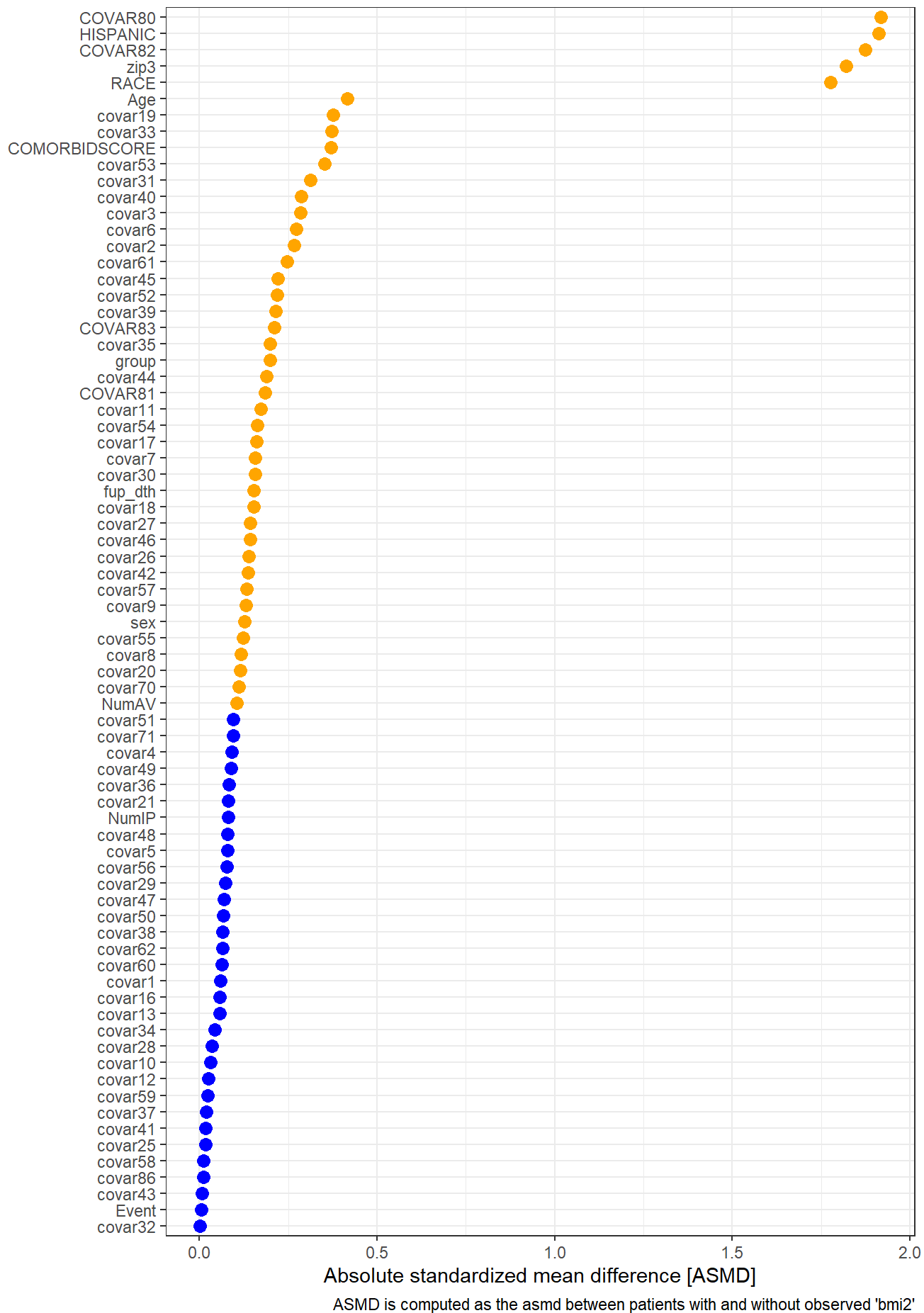
```
bmi.ind<-which(names(ate.ip)=="bmi")
out2<-smdi_asmd(data=ate.ip[,-bmi.ind],covar="bmi2", includeNA=FALSE)

out2$bmi$asmd_aggregate
```

```
## # A tibble: 1 × 4
##   covariate asmd_median asmd_min asmd_max
##   <chr>      <chr>      <chr>   <chr>
## 1 bmi2      0.126      0.001   1.918
```

```
out2$bmi$asmd_plot
```

## ASMD plot for covariate bmi2



## Statistical tests for the null hypothesis of MCAR of BMI

The null hypothesis for Hotelling's and Little's test that the missing pattern of BMI is MCAR.

Two-Sample Hotelling's T-squared test for missing BMI

```
smdi_hotelling(data=ate.ip[, -bmi2.ind], covar="bmi")
```

```
## covariate hotteling_p
## 1      bmi      <.001
```

Two-Sample Hotelling's T-squared test for missing BMI2

```
smdi_hotelling(data=ate.ip[, -bmi.ind], covar="bmi2")
```

```
## covariate hotteling_p
## 1      bmi2      <.001
```

Little Single global chi-square p-value (Little) cross all missing variables WITH the null hypothesis of MCAR

BMI

```
tout<-smdi_little(data=ate.ip[, -bmi2.ind])
tout$p.value
```

```
## [1] 0
```

BMI2

```
tout2<-smdi_little(data=ate.ip[, -bmi.ind])
tout2$p.value
```

```
## [1] 0
```

## Group 2 diagnosis focusing on MAR

This function trains and fits a random forest classification model to predict the missing indicator of BMI given exposure, outcome, and covariates. If the resulting area under the receiver operating characteristic curve (AUC) is meaningfully >0.5, this would give some evidence for MAR against MCAR as being the underlying missingness mechanism.

## BMI

```
output<-smdi_rf(data=ate.ip, covar=c("bmi","bmi2"))  
output$bmi$rf_table
```

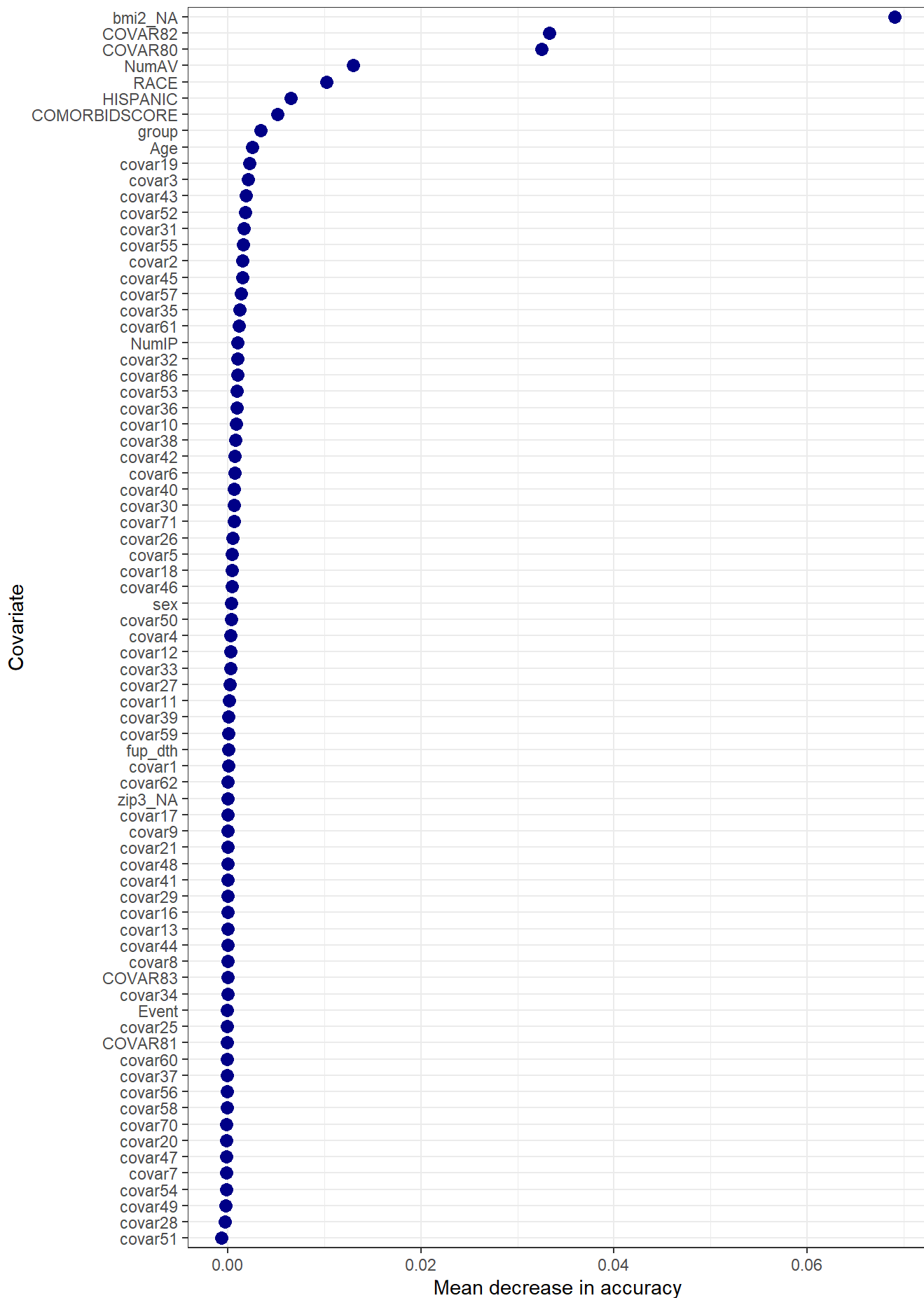
```
## # A tibble: 1 × 2  
##   covariate rf_auc  
##   <chr>      <chr>  
## 1 bmi      0.802
```

### Covariate importance for prediction of BMI missingness

```
output$bmi$rf_plot
```



# Covariate importance for predicting bmi (training set)



## BMI2

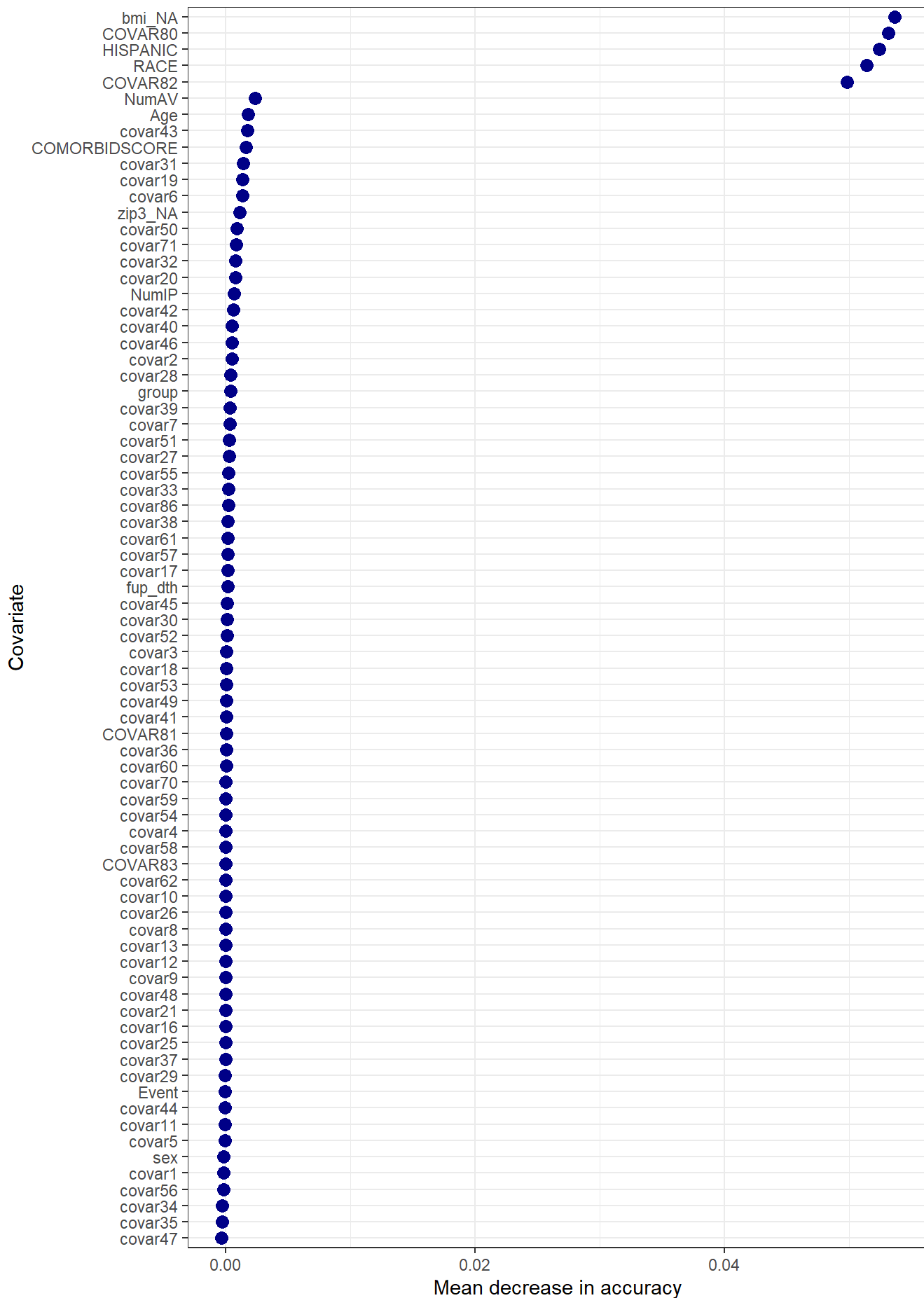
```
output<-smdi_rf(data=ate.ip, covar=c("bmi","bmi2"))  
output$bmi2$rf_table
```

```
## # A tibble: 1 × 2  
##   covariate rf_auc  
##   <chr>      <chr>  
## 1 bmi2      0.885
```

Covariate importance for prediction of BMI2 missingness

```
output$bmi2$rf_plot
```

### Covariate importance for predicting bmi2 (training set)



## Group 3 diagnosis

This function examines the association of the missingness indicator of variable “bmi” and the primary ATE outcome. The function computes both a univariate model and a multi-covariate model adjusted for age, sex, race/ethnicity, dropout status, death status, comorbidity score, number of visits, NUMV. and NUMIP.

As expected, under a MCAR mechanism, there will be no difference in the outcome between patients with and without a value for BMI.

Under MAR, given that missingness can be sufficiently explained by observed covariates, a spurious association in the univariate model will disappear after adjustment.

If the missingness followed any MNAR mechanism, an association was observed regardless of adjustment.

*#Important: don't include variables like ID variables, ZIP codes, dates, etc.*

```
in.covar2<-c("Age","sex","HISPANIC","RACE","group","fup_dth","COMORBIDSCORE", "NumAV", "N
umIP" ,paste0("covar",c(seq(1,13),seq(28,36),seq(40,47),49,51,52,58,70,71)))
```

```
smdi_outcome(data=ate.ip,
              covar=c(in.covar2,"bmi"),
              model=c("logistic"),form_lhs="Event")
```

```
## # A tibble: 1 × 3
##   covariate estimate_univariate estimate_adjusted
##   <chr>      <glue>                <glue>
## 1 bmi      -0.07 (95% CI -0.46, 0.32) -0.10 (95% CI -0.82, 0.60)
```

```
smdi_outcome(data=ate.ip,
              covar=c(in.covar2,"bmi2"),
              model=c("logistic"),form_lhs="Event")
```

```
## # A tibble: 1 × 3
##   covariate estimate_univariate estimate_adjusted
##   <chr>      <glue>                <glue>
## 1 bmi2      0.02 (95% CI -0.39, 0.42) 1.10 (95% CI 0.14, 1.87)
```

## Reference

Janick Weberpals, Sudha R Raman, Pamela A Shaw, Hana Lee, Bradley G Hammill, Sengwee Toh, John G Connolly, Kimberly J Dandreo, Fang Tian, Wei Liu, Jie Li, José J Hernández-Muñoz, Robert J Glynn, Rishi J Desai, smdi: an R package to perform structural missing data investigations on partially observed confounders in real-world evidence studies, JAMIA Open, Volume 7, Issue 1, April 2024, ooae008, <https://doi.org/10.1093/jamiaopen/ooae008> (<https://doi.org/10.1093/jamiaopen/ooae008>)