

# Identifying bias & precision variables

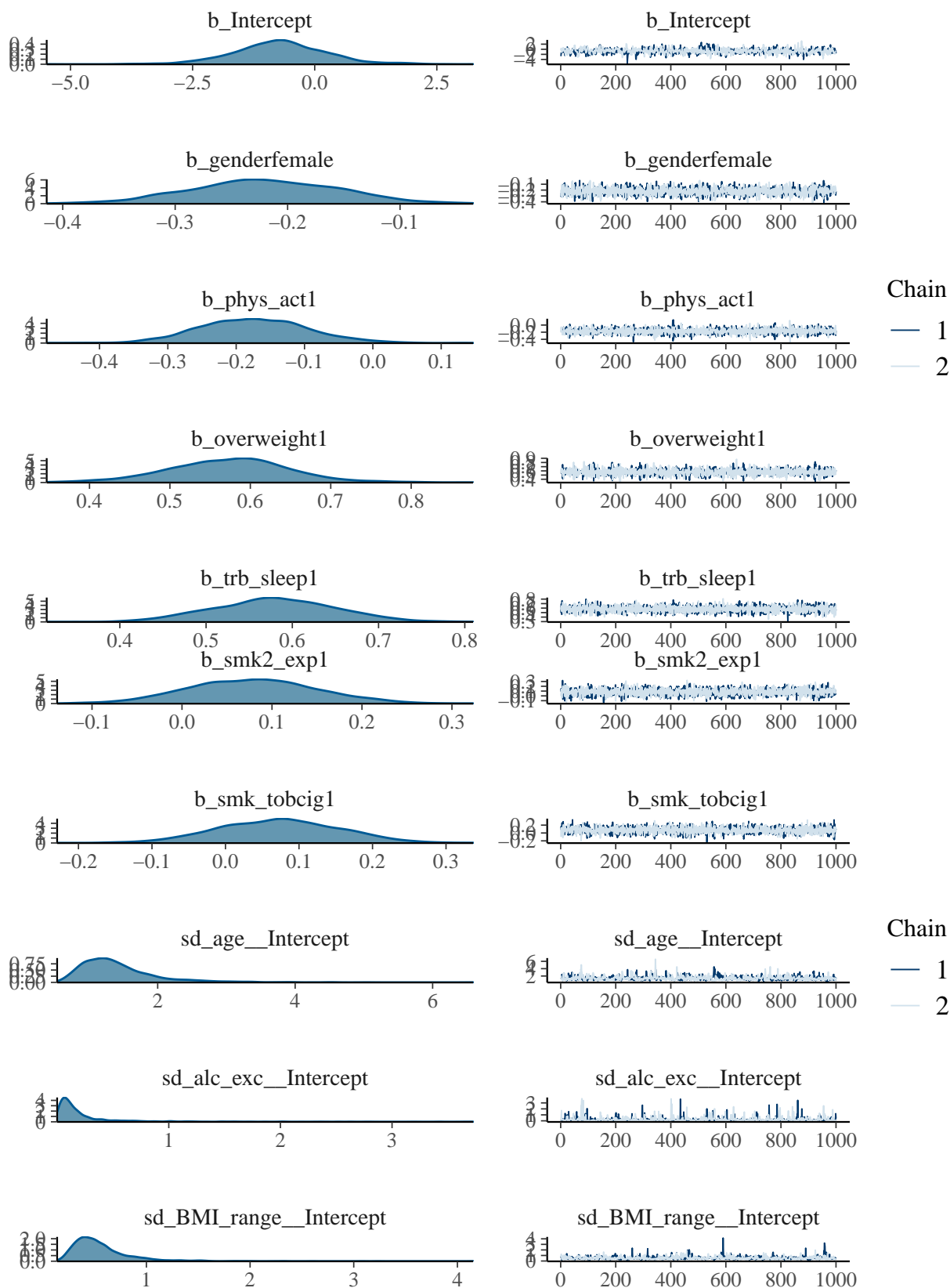
Swen Kuh

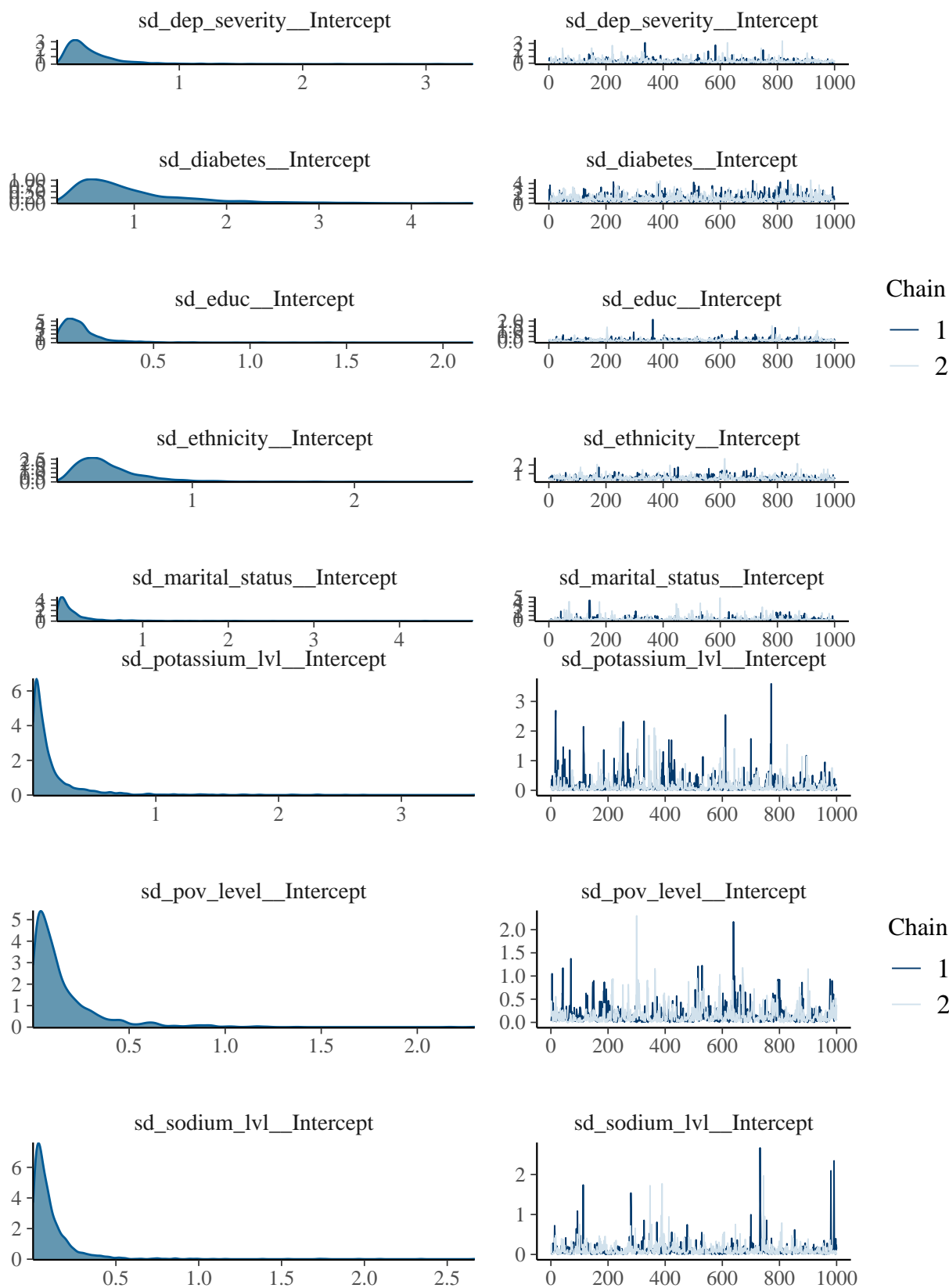
2022-08-29

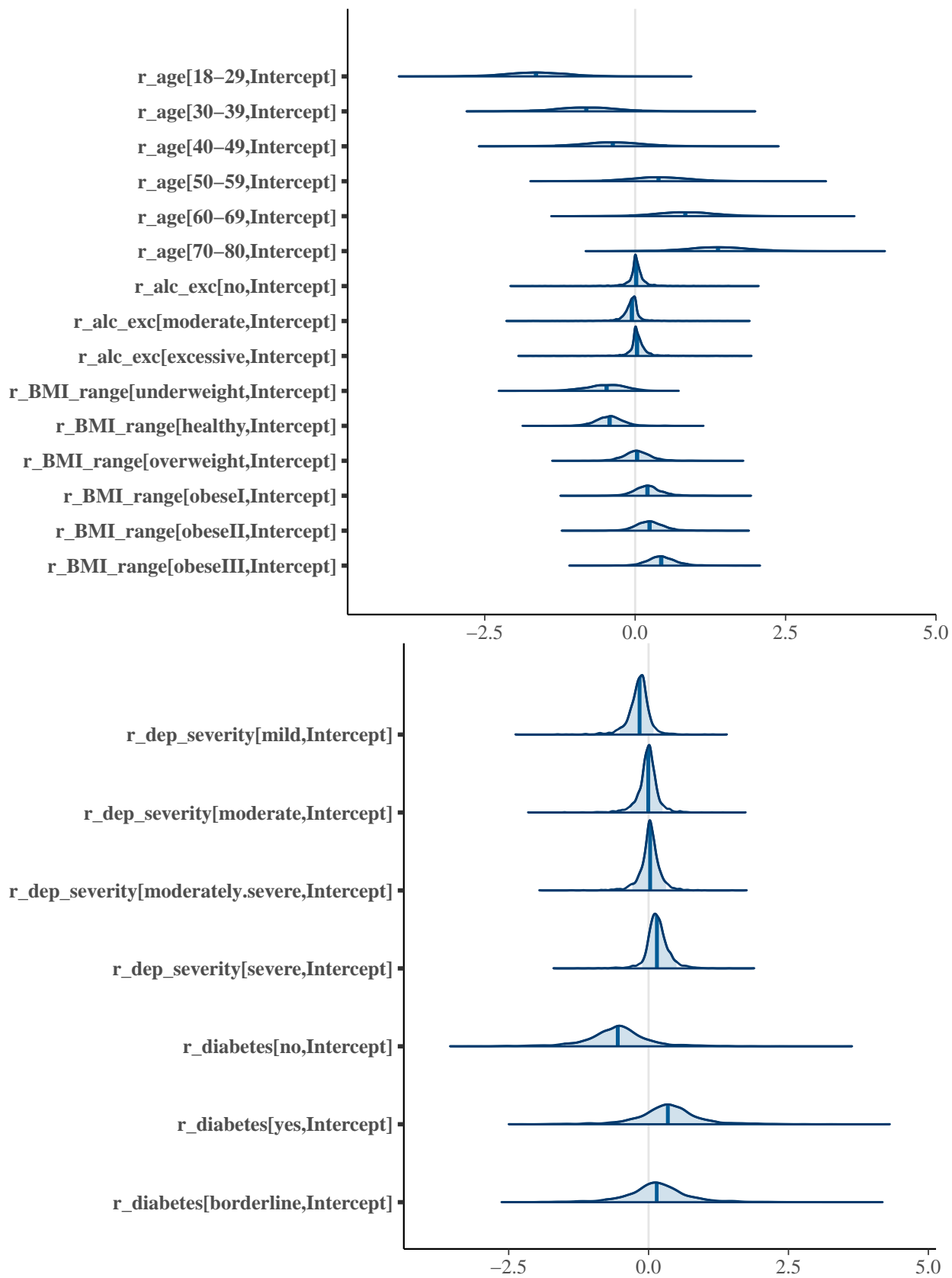
```
# brms ####
nhmodel_outcome_popn = brm(high_bp ~ (1|age) + (1|ethnicity) + gender + (1|educ) +
  (1|marital_status) + phys_act + overweight + (1|alc_exc) + (1|BMI_range) +
  (1|dep_severity) + (1|diabetes) + trb_sleep + (1|pov_level) + smk2_exp +
  smk_tobcig + (1|sodium_lvl) + (1|potassium_lvl),
  data = nhdata,
  seed = 5678,
  chain = 2,
  backend = "cmdstanr", silent = 2,
  cores=2,
  family = bernoulli(link = "logit"),
  control = list(adapt_delta = 0.9),
  file=here('nhanes/data/nhmodel_outcome_popn.rds'))

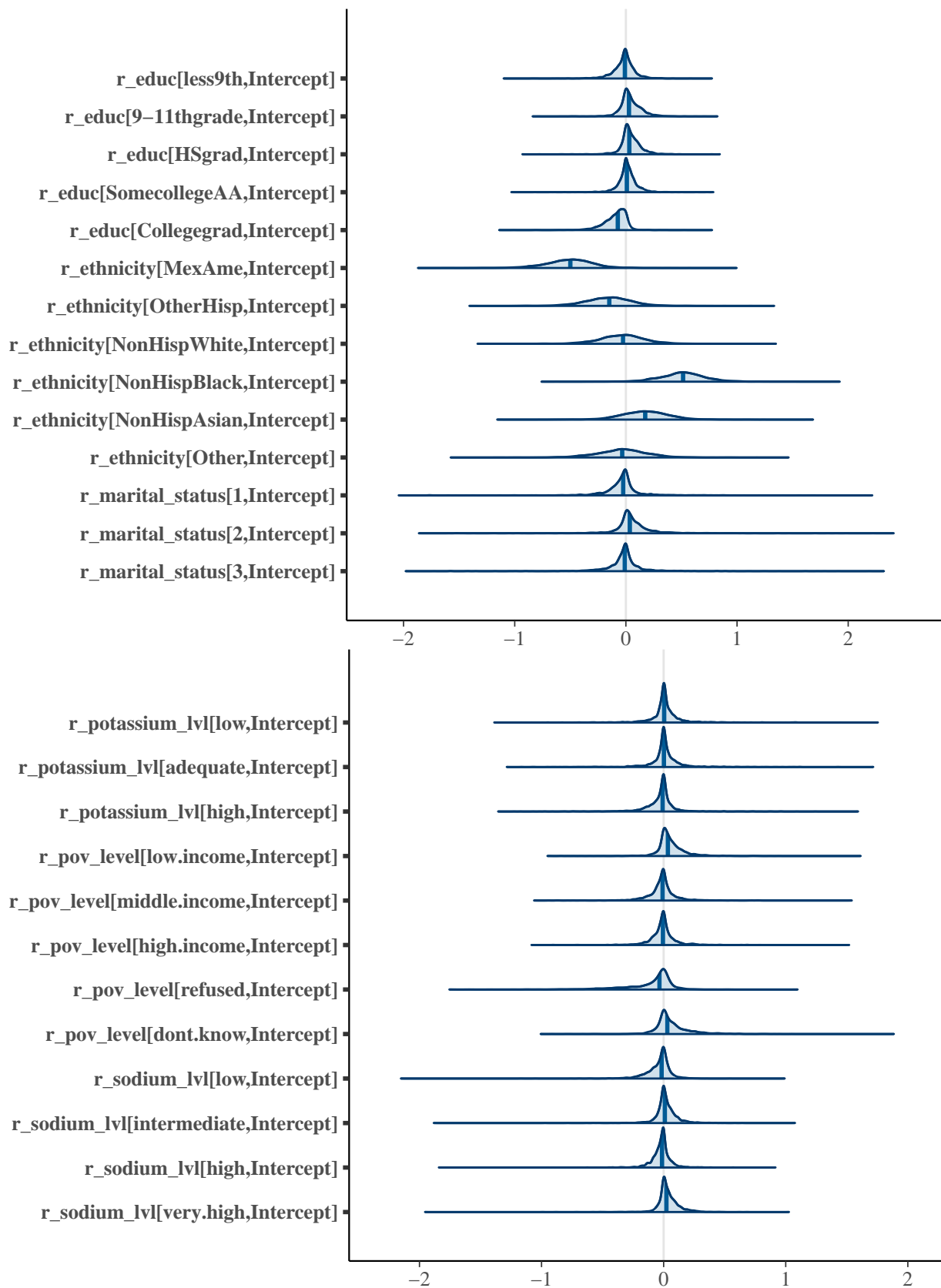
ptm = proc.time()
nhmodel_incl_popn = brm(inclusion ~ (1|age) + (1|ethnicity) + gender + (1|educ) +
  (1|marital_status) + phys_act + overweight + (1|alc_exc) + (1|BMI_range) +
  (1|dep_severity) + (1|diabetes) + trb_sleep + (1|pov_level) + smk2_exp +
  smk_tobcig + (1|sodium_lvl) + (1|potassium_lvl),
  data = nhdata,
  seed = 5678,
  chain = 2,
  backend = "cmdstanr", silent = 2,
  cores=2,
  family = bernoulli(link = "logit"),
  control = list(adapt_delta = 0.9),
  file=here('nhanes/data/nhmodel_incl_popn.rds'))

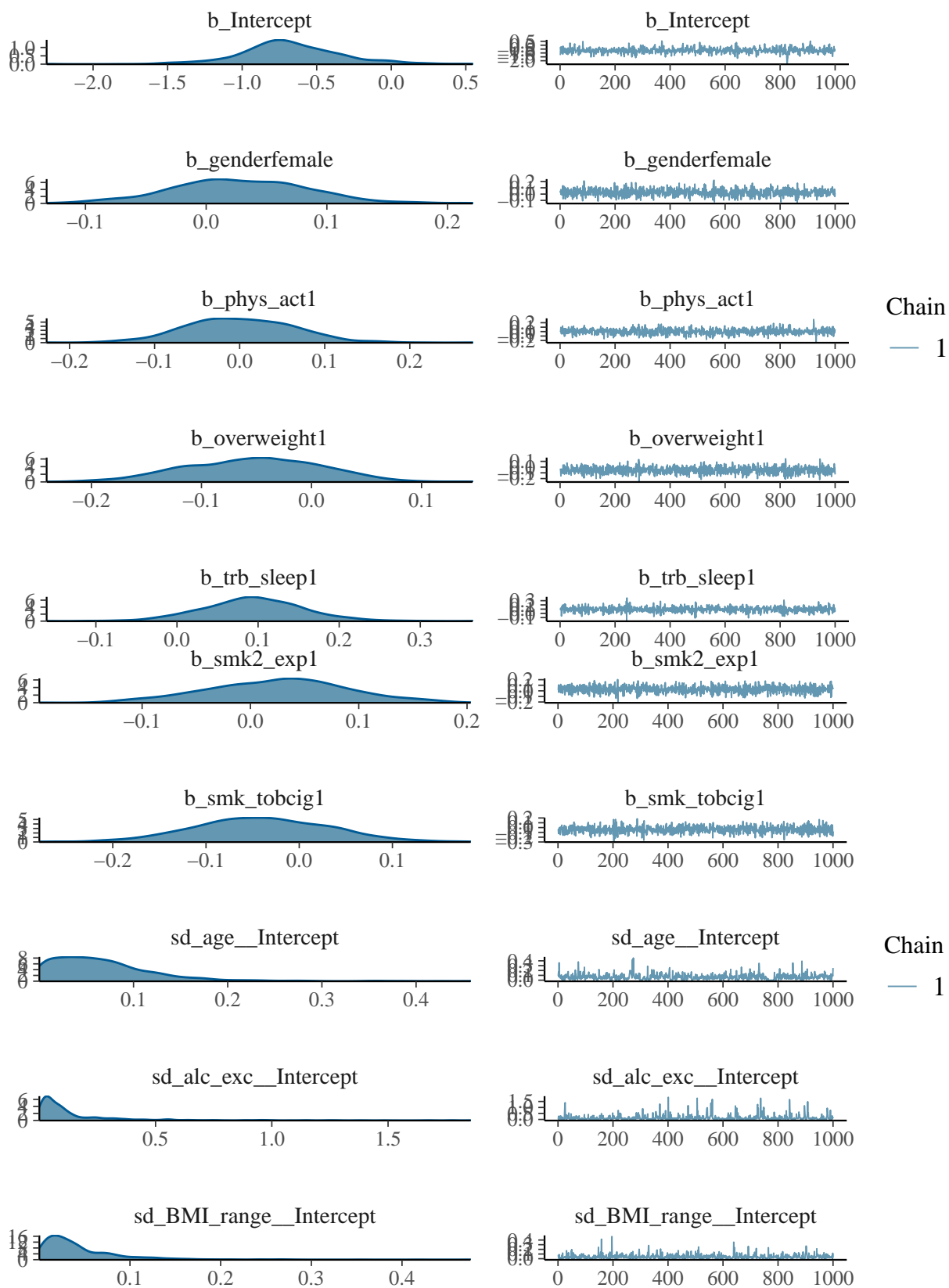
proc.time() - ptm
```

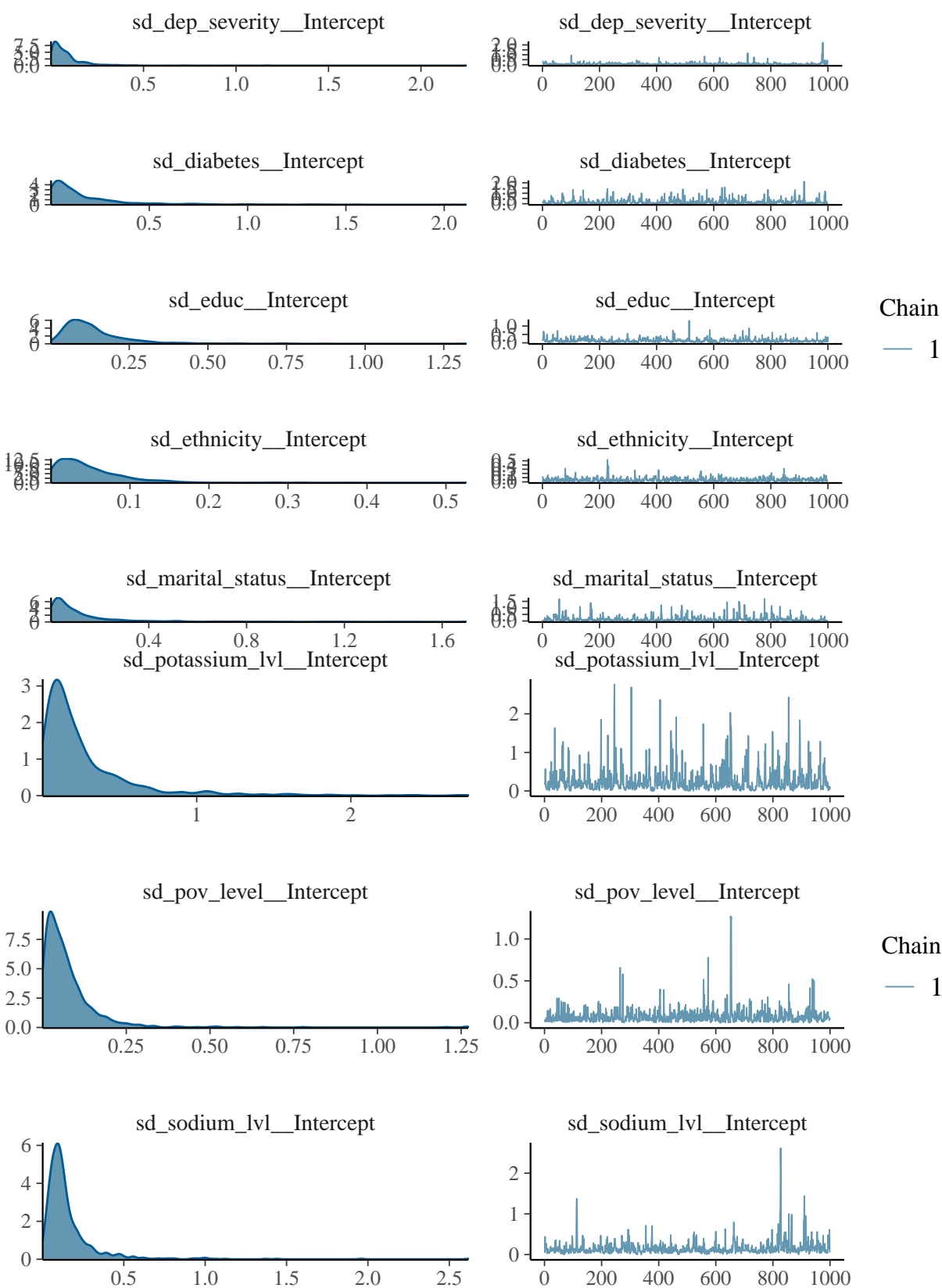


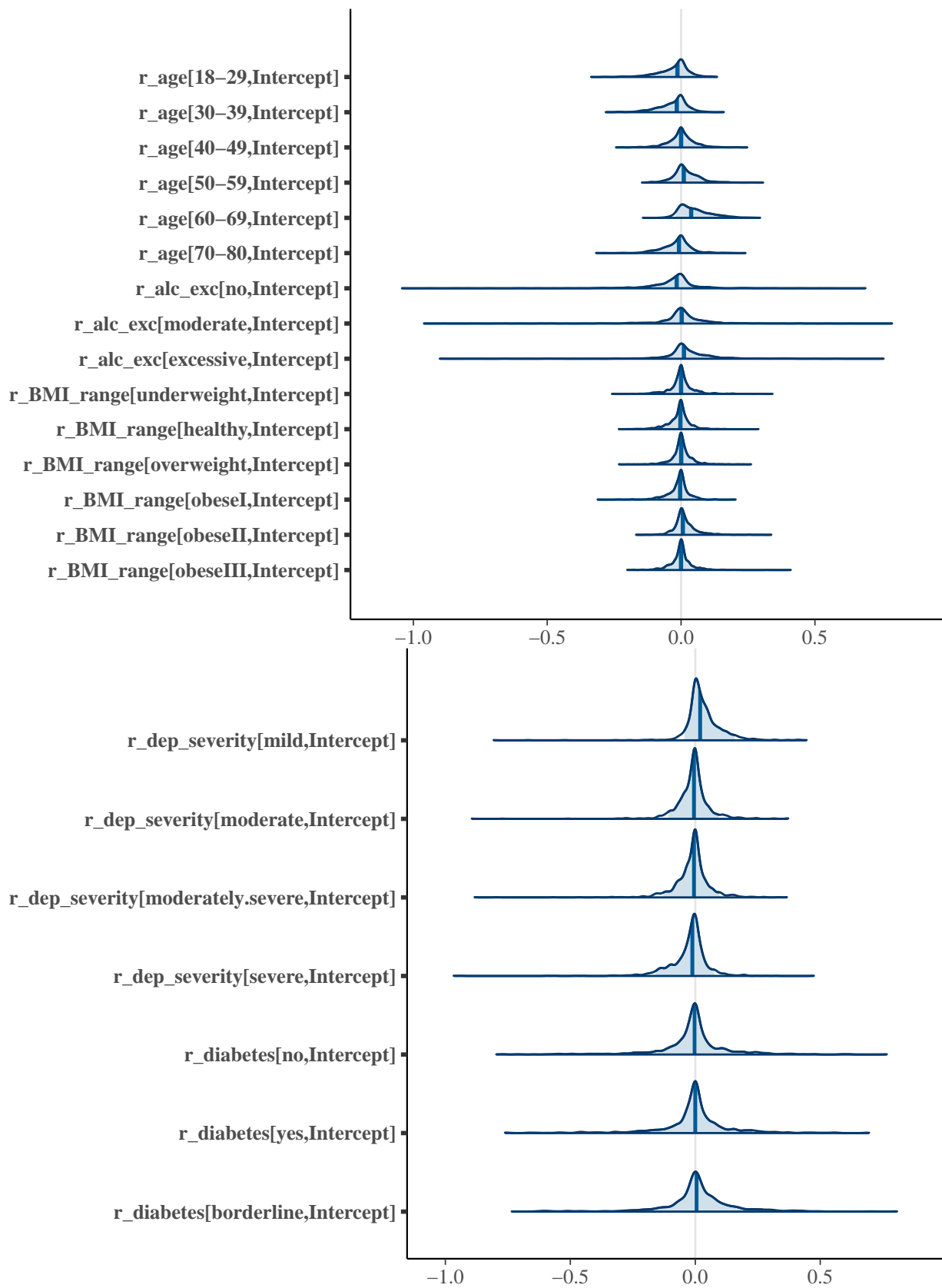




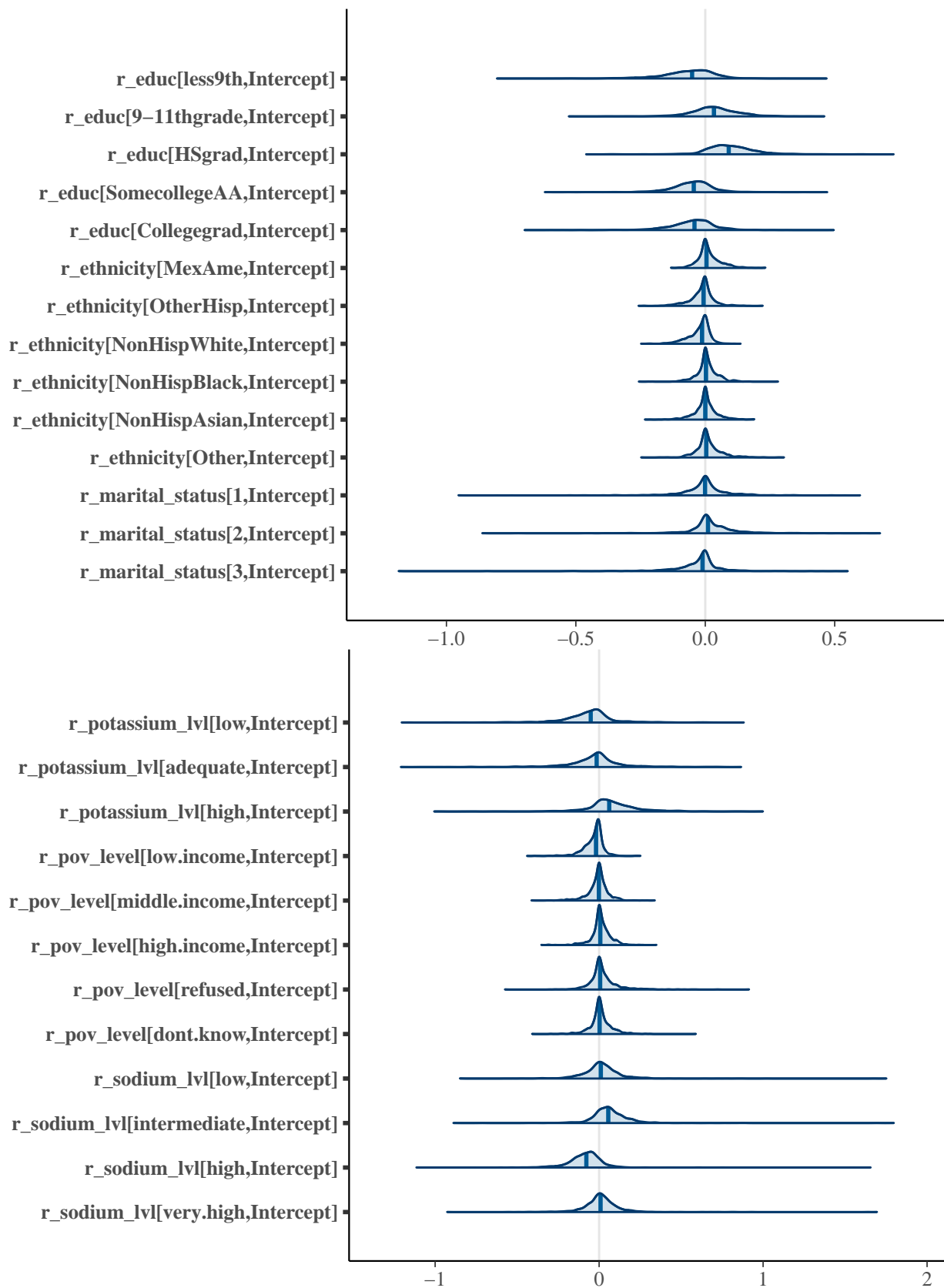












```
## identifying 'significant' variables
# using credible interval and see if it contains 0
cr_outc = nhmodel_outcome_popn %>%
  tidy_draws() %>%
  select(contains(c("b_", "r_"))) %>%
  summarise(across(1:"r_sodium_lvl[very.high,Intercept]", function(x) quantile(x, c(0.1, 0.16, 0.20, 0.80),
    t() %>%
  magrittr::set_colnames(c('q10', 'q16', 'q20', 'q80', 'q84', 'q90')) %>%
  as_tibble(., rownames = "rowname") %>%
  mutate(cri80 = ifelse(q10 < 0 & q90 < 0, 1,
    ifelse(q10 > 0 & q90 > 0, 1, 0)),
    cri68 = ifelse(q16 < 0 & q84 < 0, 1,
    ifelse(q16 > 0 & q84 > 0, 1, 0)),
    cri60 = ifelse(q20 < 0 & q80 < 0, 1,
    ifelse(q20 > 0 & q80 > 0, 1, 0)))

cr_incl = nhmodel_incl_popn %>%
  tidy_draws() %>%
  select(contains(c("b_", "r_"))) %>%
  summarise(across(1:"r_sodium_lvl[very.high,Intercept]", function(x) quantile(x, c(0.1, 0.16, 0.20, 0.80),
    t() %>%
  magrittr::set_colnames(c('q10', 'q16', 'q20', 'q80', 'q84', 'q90')) %>%
  as_tibble(., rownames = "rowname") %>%
  mutate(cri80 = ifelse(q10 < 0 & q90 < 0, 1,
    ifelse(q10 > 0 & q90 > 0, 1, 0)),
    cri68 = ifelse(q16 < 0 & q84 < 0, 1,
    ifelse(q16 > 0 & q84 > 0, 1, 0)),
    cri60 = ifelse(q20 < 0 & q80 < 0, 1,
    ifelse(q20 > 0 & q80 > 0, 1, 0)))

cr_incl %>% filter(cri60 == 1)
```

```
## # A tibble: 5 x 10
##   rowname      q10      q16      q20      q80      q84      q90 cri80 cri68 cri60
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl> <dbl> <dbl> <dbl>
## 1 b_Inter~ -1.05    -0.963    -9.17e-1 -0.430    -0.364    -0.265      1      1      1
## 2 b_trb_s~  0.0148   0.0320   4.45e-2  0.143     0.150     0.170      1      1      1
## 3 r_age[6~ -0.00851 -0.00104  2.42e-4  0.0987   0.112     0.139      0      0      1
## 4 r_educ[~  0.00100  0.0186   2.98e-2  0.160     0.177     0.206      1      1      1
## 5 r_sodiu~ -0.201    -0.165   -1.50e-1 -0.0173  -0.00139  0.0140      0      1      1
```

```
cr_outc %>% filter(cri60 == 1)
```

```
## # A tibble: 22 x 10
##   rowname      q10      q16      q20      q80      q84      q90 cri80 cri68 cri60
##   <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl> <dbl> <dbl> <dbl>
## 1 b_genderf~ -0.305    -0.285    -0.275    -0.163    -0.154    -0.137      1      1      1
## 2 b_phys_ac~ -0.272    -0.253    -0.239    -0.110    -0.0978   -0.0745      1      1      1
## 3 b_overwei~  0.476     0.499     0.511     0.632     0.643     0.665      1      1      1
## 4 b_trb_sle~  0.487     0.508     0.519     0.639     0.650     0.670      1      1      1
## 5 b_smk2_ex~ -0.0113   0.00643   0.0180     0.138     0.149     0.170      0      1      1
## 6 b_smk_tob~ -0.0339  -0.00881  0.00262     0.144     0.157     0.179      0      0      1
## 7 r_age[18~ -2.33     -2.16     -2.08     -1.21     -1.13     -0.983      1      1      1
## 8 r_age[30~ -1.48     -1.31     -1.24     -0.385    -0.304    -0.147      1      1      1
```

```
## 9 r_age[60-~ 0.162 0.334 0.410 1.25 1.34 1.49 1 1 1
## 10 r_age[70-~ 0.696 0.873 0.963 1.80 1.89 2.04 1 1 1
## # ... with 12 more rows

# using "pvalue" approach
(pv0_outc = nhmodel_outcome_popn %>%
  tidy_draws() %>%
  select(contains(c("b_", "r_"))) %>%
  summarise(across(1:"r_sodium_lvl[very.high,Intercept]", function(x)mean(x<0))) %>%
  t() %>%
  magrittr::set_colnames('less0') %>%
  as_tibble(., rownames = "rowname") %>%
  mutate(less0 = as.numeric(less0),
    significance = ifelse(less0 < 0.4 | less0 > 0.6, 1, 0)))

## # A tibble: 55 x 3
##   rowname                less0 significance
##   <chr>                  <dbl>         <dbl>
## 1 b_Intercept           0.786             1
## 2 b_genderfemale        1                 1
## 3 b_phys_act1           0.982             1
## 4 b_overweight1         0                 1
## 5 b_trb_sleep1          0                 1
## 6 b_smk2_exp1           0.138             1
## 7 b_smk_tobcig1         0.195             1
## 8 r_age[18-29,Intercept] 0.996             1
## 9 r_age[30-39,Intercept] 0.934             1
## 10 r_age[40-49,Intercept] 0.774             1
## # ... with 45 more rows

(pv0_incl = nhmodel_incl_popn %>%
  tidy_draws() %>%
  select(contains(c("b_", "r_"))) %>%
  summarise(across(1:"r_sodium_lvl[very.high,Intercept]", function(x)mean(x<0))) %>%
  t() %>%
  magrittr::set_colnames('less0') %>%
  as_tibble(., rownames = "rowname") %>%
  mutate(less0 = as.numeric(less0),
    significance = ifelse(less0 < 0.4 | less0 > 0.6, 1, 0)))

## # A tibble: 55 x 3
##   rowname                less0 significance
##   <chr>                  <dbl>         <dbl>
## 1 b_Intercept           0.965             1
## 2 b_genderfemale        0.31             1
## 3 b_phys_act1           0.507             0
## 4 b_overweight1         0.796             1
## 5 b_trb_sleep1          0.066             1
## 6 b_smk2_exp1           0.345             1
## 7 b_smk_tobcig1         0.716             1
## 8 r_age[18-29,Intercept] 0.657             1
## 9 r_age[30-39,Intercept] 0.701             1
## 10 r_age[40-49,Intercept] 0.49             0
## # ... with 45 more rows
```

```
pv0_outc %>% filter(significance==1)
```

```
## # A tibble: 43 x 3
##   rowname          less0 significance
##   <chr>            <dbl>         <dbl>
## 1 b_Intercept      0.786           1
## 2 b_genderfemale    1             1
## 3 b_phys_act1       0.982           1
## 4 b_overweight1     0             1
## 5 b_trb_sleep1      0             1
## 6 b_smk2_exp1       0.138           1
## 7 b_smk_tobcig1     0.195           1
## 8 r_age[18-29,Intercept] 0.996           1
## 9 r_age[30-39,Intercept] 0.934           1
## 10 r_age[40-49,Intercept] 0.774           1
## # ... with 33 more rows
```

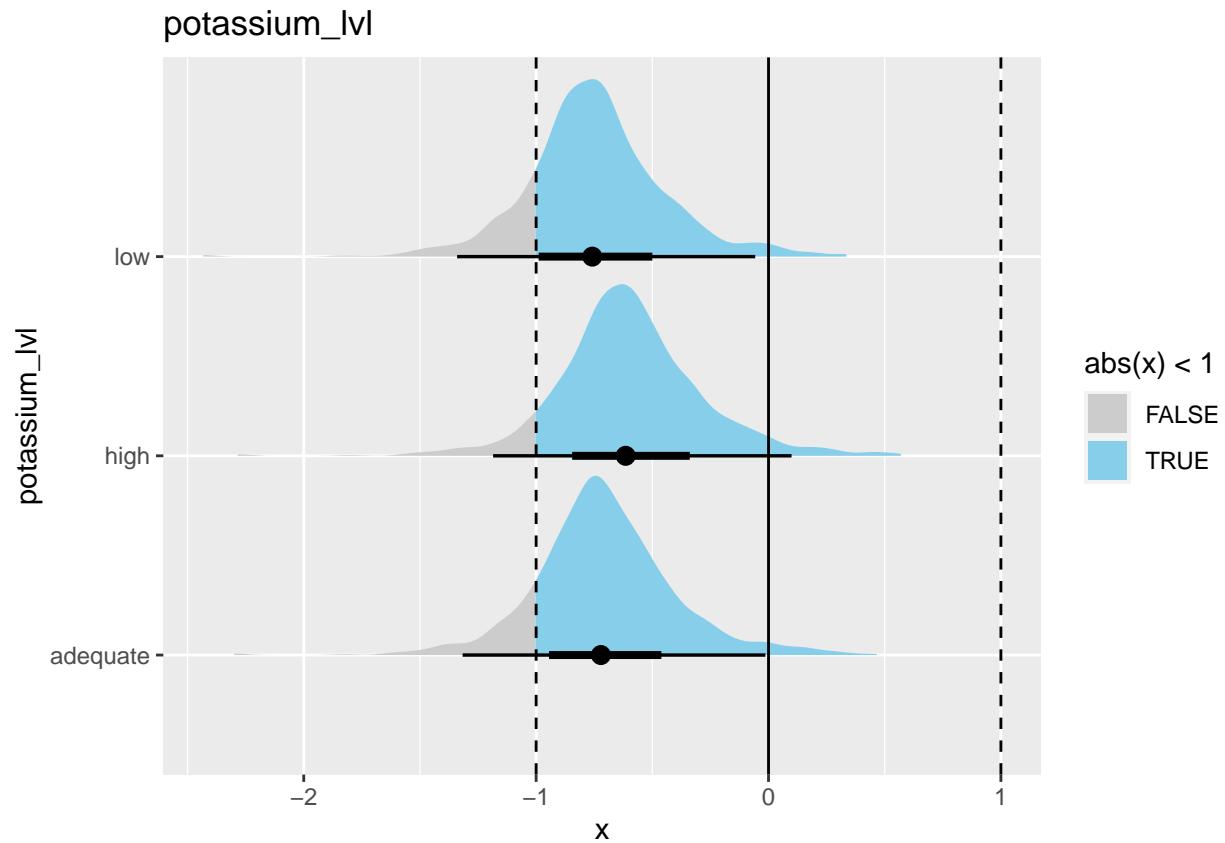
```
pv0_incl %>% filter(significance==1)
```

```
## # A tibble: 30 x 3
##   rowname          less0 significance
##   <chr>            <dbl>         <dbl>
## 1 b_Intercept      0.965           1
## 2 b_genderfemale    0.31            1
## 3 b_overweight1     0.796           1
## 4 b_trb_sleep1      0.066           1
## 5 b_smk2_exp1       0.345           1
## 6 b_smk_tobcig1     0.716           1
## 7 r_age[18-29,Intercept] 0.657           1
## 8 r_age[30-39,Intercept] 0.701           1
## 9 r_age[50-59,Intercept] 0.364           1
## 10 r_age[60-69,Intercept] 0.192           1
## # ... with 20 more rows
```

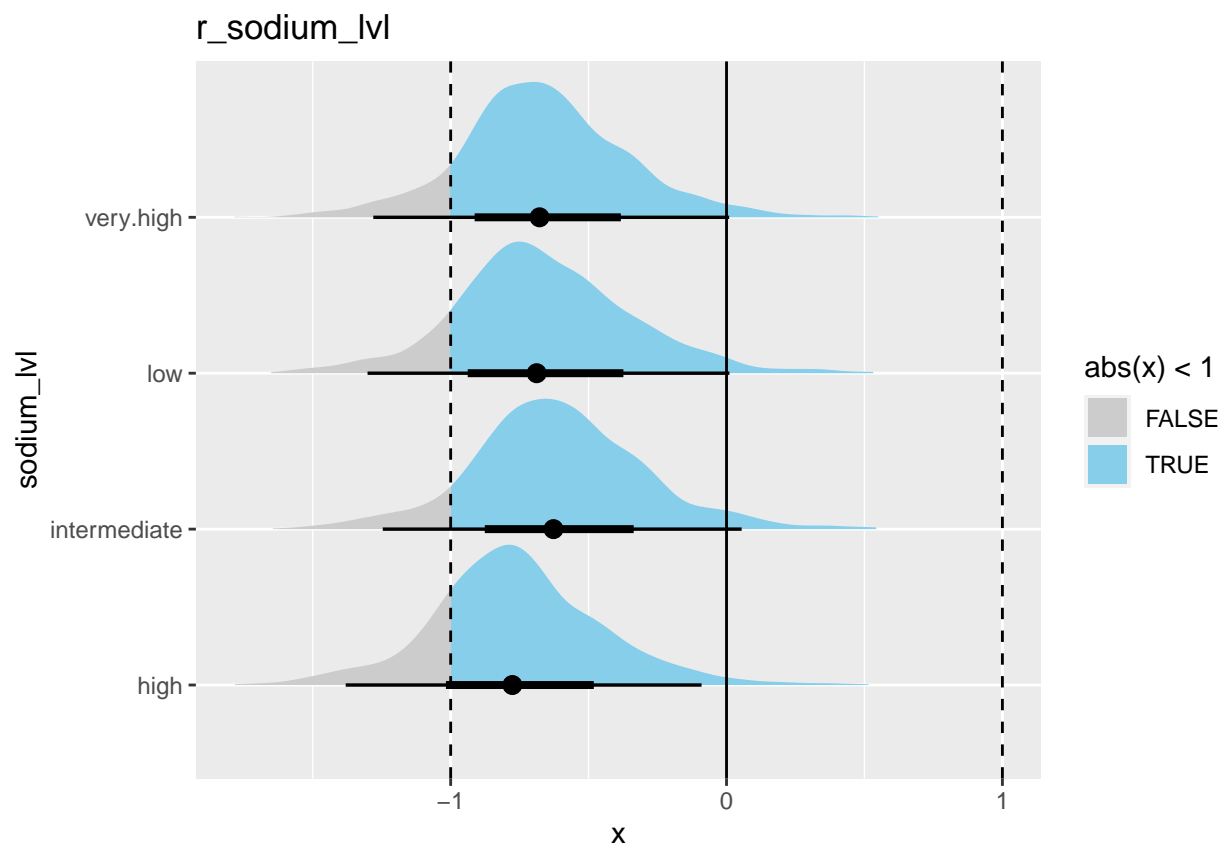
```
# different way of plotting ####
```

```
# potassium_level
```

```
nhmodel_incl_popn %>%
  spread_rvars(b_Intercept, r_potassium_lvl[potassium_lvl,]) %>%
  mutate(condition_mean = b_Intercept + r_potassium_lvl) %>%
  ggplot(aes(y = potassium_lvl, dist = condition_mean, fill = stat(abs(x) < 1))) +
  stat_dist_halfeye() +
  geom_vline(xintercept = c(-1, 1), linetype = "dashed") +
  geom_vline(xintercept = 0) +
  scale_fill_manual(values = c("gray80", "skyblue")) +
  ggtitle('potassium_lvl')
```



```
nhmodel_incl_popn %>%
  spread_rvars(b_Intercept, r_sodium_lvl[sodium_lvl,]) %>%
  mutate(condition_mean = b_Intercept + r_sodium_lvl) %>%
  ggplot(aes(y = sodium_lvl, dist = condition_mean, fill = stat(abs(x) < 1))) +
  stat_dist_halfeye() +
  geom_vline(xintercept = c(-1, 1), linetype = "dashed") +
  geom_vline(xintercept = 0) +
  scale_fill_manual(values = c("gray80", "skyblue")) +
  ggtitle('r_sodium_lvl')
```



```
nhmodel_outcome_popn %>%
  spread_rvars(b_Intercept, r_age[age,]) %>%
  mutate(condition_mean = b_Intercept + r_age) %>%
  ggplot(aes(y = age, dist = condition_mean, fill = stat(abs(x) < 1))) +
  stat_dist_halfeye() +
  geom_vline(xintercept = c(-1, 1), linetype = "dashed") +
  scale_fill_manual(values = c("gray80", "skyblue")) +
  ggtitle('age')
```

