### Why the allpbo or alldrg has a small variance?

It dependes on how we do the cross validation, i.e.

- we give them labels as 1 to 10 at the begining of each repetation, or
- within each repetation, we mark the labels 10 times.

For example,

| subj | changescore | treatment assignment | CV group |
|---|---|---|---|
| $i$ | $Y_i$ | $t_i$ | $l_i \in \{1, 2, .., 10\}$ |

If we label all subject as pbo group or drg group, all the change scores $Y_i$s will be included.

For the $j$th repetation, we conduct a 10 fold cross validation and label the subjects in group 1 to 10. We calculate the mean of $Y_i$ whose label is from $k = 1, 2, ..10$ respectively. Then we get the mean values $\bar{Y}_k$.

$$\bar{Y}_k = \frac{1}{n_k} \sum_{l \in k} Y_i$$

where $n_k$ is the number of subjects whose label is $k$. The mean value across the 10 CV is

$$\bar{Y}_j = \frac{1}{10} \sum_{k=1}^{10} \bar{Y}_k = \frac{1}{10} \sum_{k=1}^{10} \frac{1}{n_k} \sum_{l \in k} Y_i \approx \bar{Y}$$

. Therefore, the overall mean value is just the mean value of the dataset's change score. The variance is

$$Var(\bar{Y}_j) = Var(\bar{Y}) = \frac{var(Y)}{n}, \text{ if } n_k\text{s have the same value}$$

We run the whole steps for $N$ times, the mean value of the change score is just

$$\bar{Y}_N = \frac{1}{N} \sum_{j=1}^{N} \bar{Y}_j$$

$$Var(\bar{Y}_N) = \frac{var(Y)}{nN}$$

However, if we resample the labels at each times of the cross validation, the mean value will not change, but the variance can be larger.

```r
set.seed(1)
n = 100
meanY = c()
for(iters in 1:100){
  y = rnorm(n,0,1)
  label = sample(1:10, n, replace = TRUE) # cross validation groups
  meany_cv = c()
  for(i in 1:10){
    meany_cv = c(meany_cv, mean(y[label == i]))
  }
  meanY = c(meanY, mean(meany_cv))
}
mean(meanY)
```

```
## [1] -0.006439615
```

```
sd(meanY)
```

```
## [1] 0.0906425
```

```
set.seed(2)
n = 100
meanY = c()
for(iters in 1:100){
  y = rnorm(n,0,1)
  meany_cv = c()
  for(i in 1:10){
    label = sample(1:10, n, replace = TRUE) # cross validation groups
    meany_cv = c(meany_cv, mean(y[label == i]))
  }
  meanY = c(meanY, mean(meany_cv))
}
mean(meanY)
```

```
## [1] 0.02032141
```

```
sd(meanY)
```

```
## [1] 0.1285985
```