# Sample Reuse

Jianhui Gao

2023-06-27

## Recovery Rate

Table 1: **Number of genome-wide significant SNPs recovered by standard and SynSurr GWAS across increasing ablation of the target phenotype.** The oracle method establishes the number of genome-wide significant (GWS) variants ($p < 5 \times 10^{-8}$) that would be identified in the absence of missingness. Missingness was introduced by ablating 25%, 50%, 75%, and 90% of the target phenotypes. Standard and SynSurr GWAS were performed on each of the ablated data sets.

| Missing Rate (%) | Height | | | | FEV1 | | | |
|---|---|---|---|---|---|---|---|---|
| | $n_{obs}$ | Oracle | Standard | SynSurr | $n_{obs}$ | Oracle | Standard | SynSurr |
| 25 | 262,105 | 7,177 | 4,896(68.22%) | 5,169(72.02%) | 231,388 | 974 | 546(56.06%) | 587(60.26%) |
| 50 | 174,737 | 7,177 | 2,742(38.21%) | 3,319(46.24%) | 154,259 | 974 | 278(28.54%) | 305(31.31%) |
| 75 | 87,368 | 7,177 | 834(11.62%) | 1,325(18.46%) | 77,129 | 974 | 60(6.16%) | 75(7.70%) |
| 90 | 34,947 | 7,177 | 192(2.68%) | 320(4.46%) | 30,852 | 974 | 0(0%) | 2(0.21%) |

## False Discovery Rate

Table 2: **False positive discoveries of the standard GWAS and the SynSurr GWAS across various missing rates.** The false positive discovery was rate defined as the number of SNPs that are significant in the standard or SynSurr GWAS, but not in the oracle GWAS.
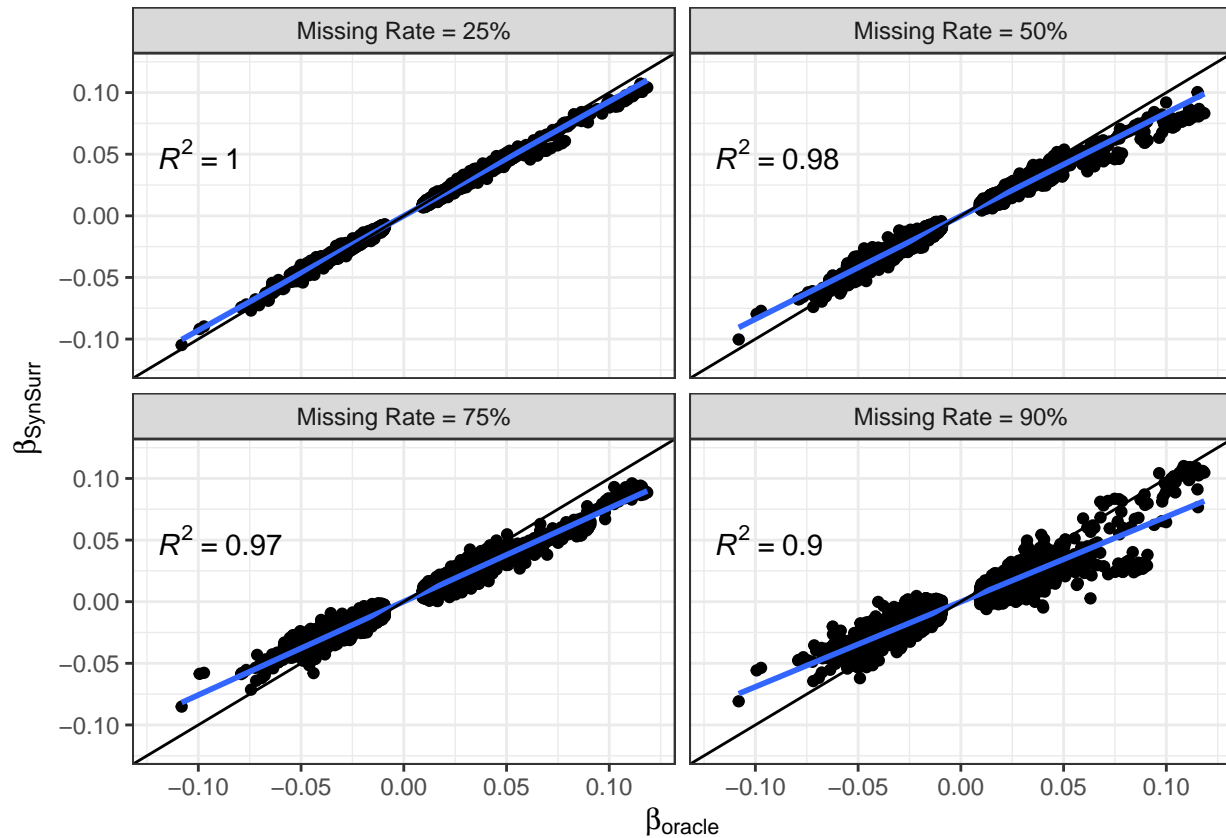
| Missing Rate(%) | Height | | FEV1 | |
|---|---|---|---|---|
| | Standard | SynSurr | Standard | SynSurr |
| 25 | 75(0.017%) | 82(0.018%) | 19(0.004%) | 40(0.009%) |
| 50 | 18(0.004%) | 27(0.006%) | 6(0.001%) | 7(0.001%) |
| 75 | 1(0%) | 4(0%) | 0(0%) | 2(0%) |
| 90 | 0(0%) | 1(0%) | 0(0%) | 0(0%) |

## Effect Size Estimate

### Height

```
## 
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
## 
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

## Warning: The dot-dot notation (`..rr.label..`) was deprecated in ggplot2 3.4.0.
## i Please use `after_stat(rr.label)` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

## `geom_smooth()` using formula = 'y ~ x'
```



# FEV1

```
## `geom_smooth()` using formula = 'y ~ x'
```

Missing Rate = 25%

Missing Rate = 50%

$R^2 = 0.99$

$R^2 = 0.98$

Missing Rate = 75%

Missing Rate = 90%

$R^2 = 0.95$

$R^2 = 0.89$

$\beta_{\text{SynSurr}}$

$\beta_{\text{oracle}}$

0.10

0.05

0.00

−0.05

−0.10

−0.10   −0.05   0.00   0.05   0.10