LD Score Regression Practical

International Statistical Genetics Workshop

IBG, University of Colorado Boulder

Tuesday, 5th March 2024

MadhurBain Singh

singhm18@vcu.edu

Virginia Institute for Psychiatric and Behavior Genetics

Virginia Commonwealth University, Richmond, VA

Thanks to Benjamin M. Neale.

Copy the scripts

```
mkdir  ~/ldsc
cp  /home/madhur/2024/Day-2/LDSC_Tutorial/*  ~/ldsc/.
```

Qualtrics link

This link is also on top of the R script: *ldsc_practical_ms_bn.R*

https://qimr.az1.qualtrics.com/jfe/form/SV_5hjQbGJvygTkAZM

LDSC Practical Outline

1. Continuous phenotype – BMI (body mass index)
2. Introduction to/Review of the Liability Threshold Model.
3. Bonus 1: Binary phenotype – MDD (major depressive disorder)
4. Bonus 2: Examine the impact of population case prevalence on the liability-scale heritability of a binary (case vs. control) phenotype.

Phenotype 1 – Body Mass Index (BMI)

Take-home points

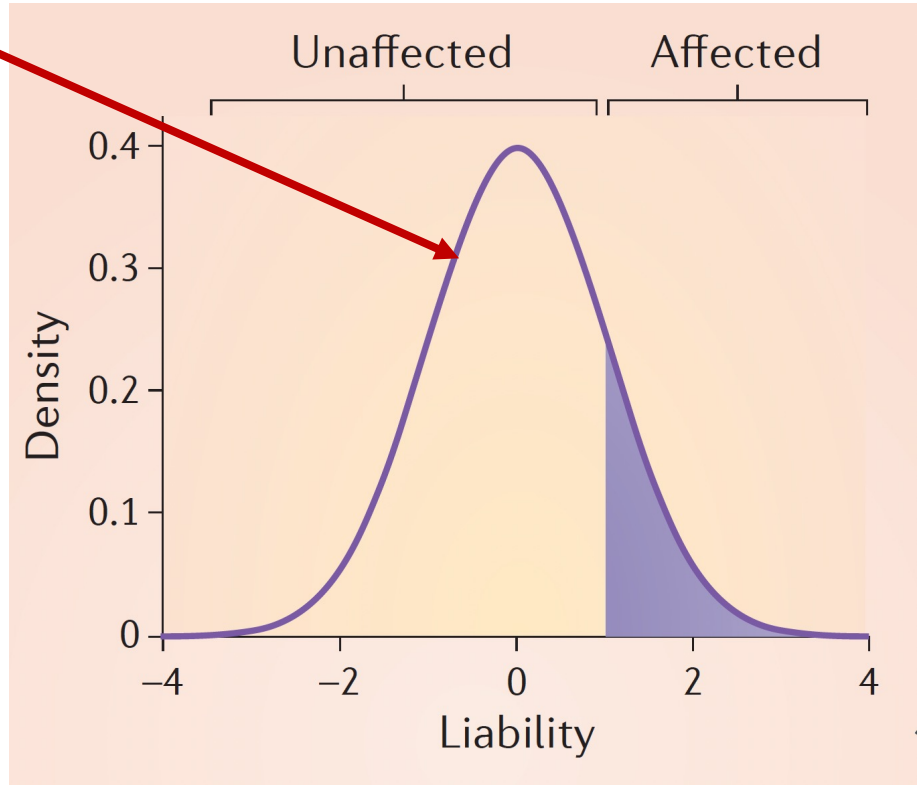
- LDSC allows us to estimate SNP heritability using only the GWAS sum stats.
- Importantly, LDSC helps us differentiate true genetic signal (polygenicity) from confounding (population stratification) in GWAS.

Important Point to Note

- LDSC uses the provided sum stats to estimate the heritability/variance per SNP. The program then uses the estimate of the per-SNP heritability to compute the total variance across all SNPs in the genome [i.e., the SNP heritability].
- **If the sum stats used to run LDSC analyses are NOT a random subset of all SNPs across the genome, the estimated per-SNP variance and, thus, the SNP-based heritability will be biased.**

Liability Threshold Model

- Individuals have a **latent continuous liability** underlying manifest complex* binary traits.
- The liability is assumed to have **normal distribution** in the population, with a mean of 0 and a variance of 1.

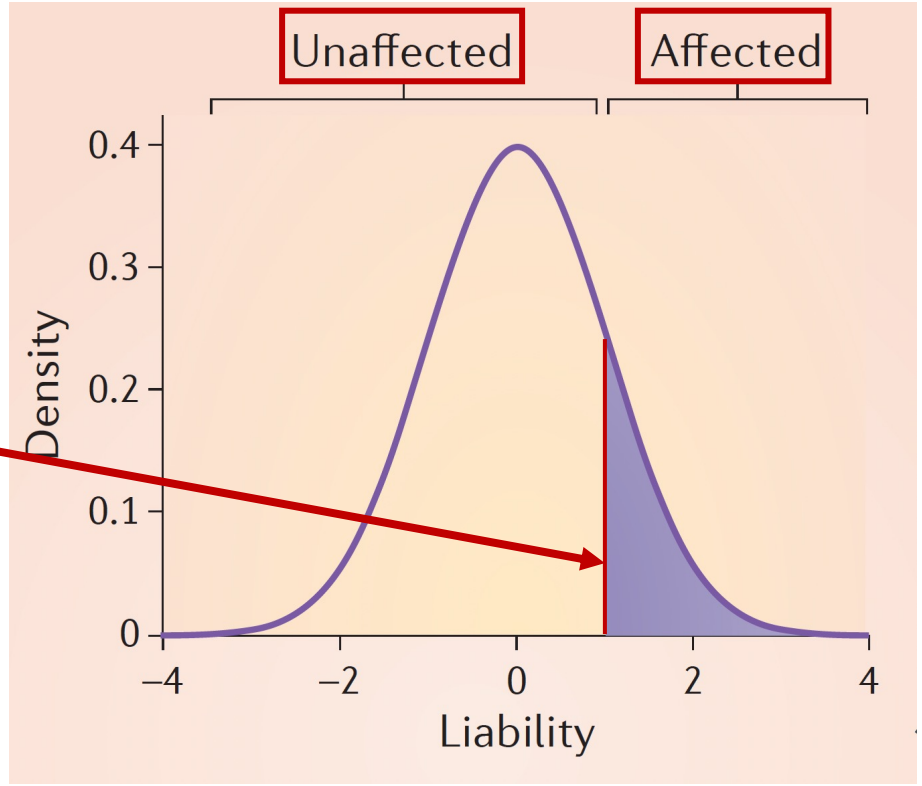


*multifactorial phenotypes with many genetic and environmental risk factors

Witte, J., Visscher, P. & Wray, N. *Nat Rev Genet* **15**, 765–776 (2014).
https://doi.org/10.1038/nrg3786

Liability Threshold Model

- Individuals have a **latent continuous liability** underlying manifest complex binary traits.
- The liability is assumed to have **normal distribution** in the population, with a mean of 0 and a variance of 1.
- There exists a certain **threshold, *t***, on the liability scale
  - All individuals above this threshold exhibit the trait. [“affected/cases”]
  - All individuals below this threshold *do not* exhibit the trait. [“unaffected/controls”]



Witte, J., Visscher, P. & Wray, N. *Nat Rev Genet* **15**, 765–776 (2014).
https://doi.org/10.1038/nrg3786

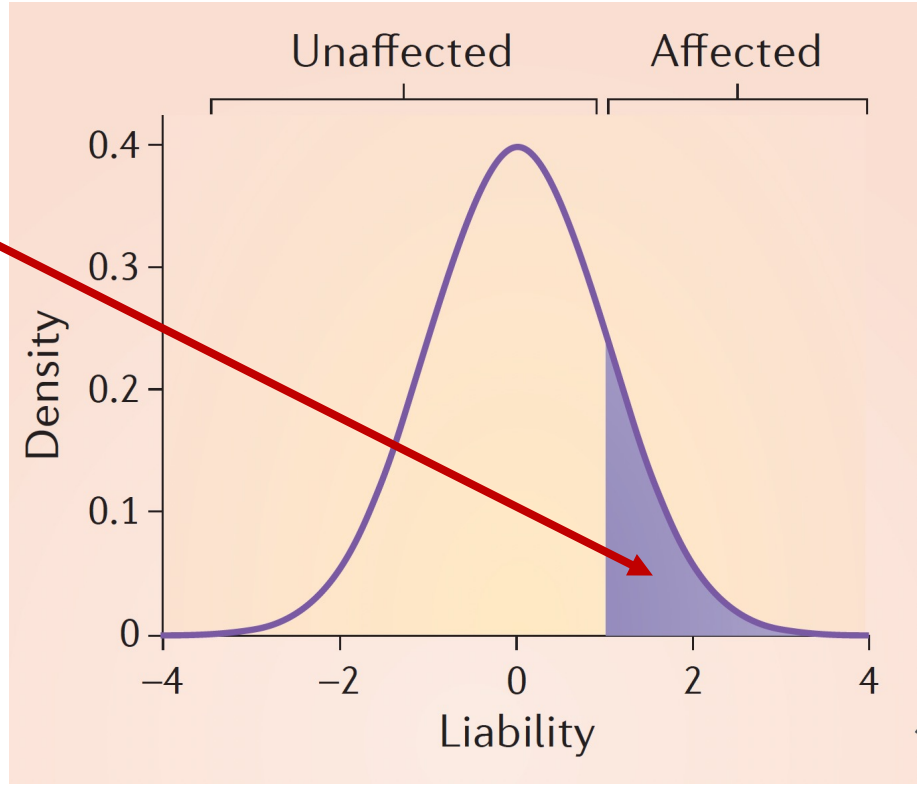
Liability Threshold Model

- **<u>Population prevalence (K)</u>** = Area under the distribution curve to the right of the threshold, t.

- If we know the population prevalence, we may estimate t.
  - By using the probit function.

$$probit(p) = \Phi^{-1}(p)$$

NB: probit is the inverse of the cumulative distribution function of the standard normal distribution (ϕ).



Witte, J., Visscher, P. & Wray, N. *Nat Rev Genet* **15**, 765–776 (2014).
https://doi.org/10.1038/nrg3786

From population prevalence of a binary disease status
To threshold on the latent continuous liability

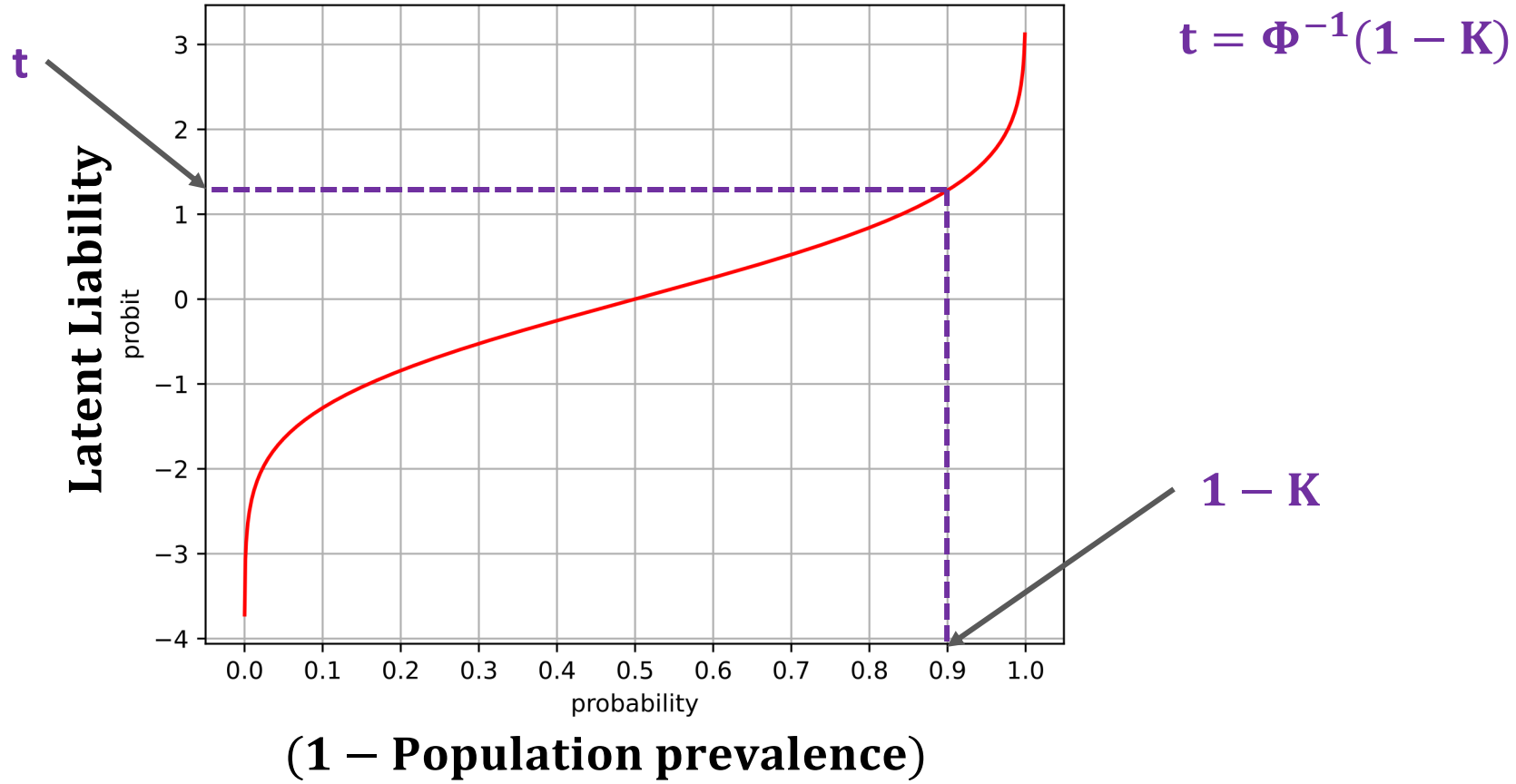


Image credit: https://en.wikipedia.org/wiki/Probit#/media/File:Probit_plot.svg

From genetic risk burden to disease risk on the observed scale

Multiplicative Scale

At a given locus with risk variant *B*, the absolute risk of disease (*D*) increases by a **multiplicative factor (relative risk, RR)** in comparison to the risk among non-carriers (k_{bb})

Genotype ***bb*** → $\Pr(D \mid bb) = k_{bb}$

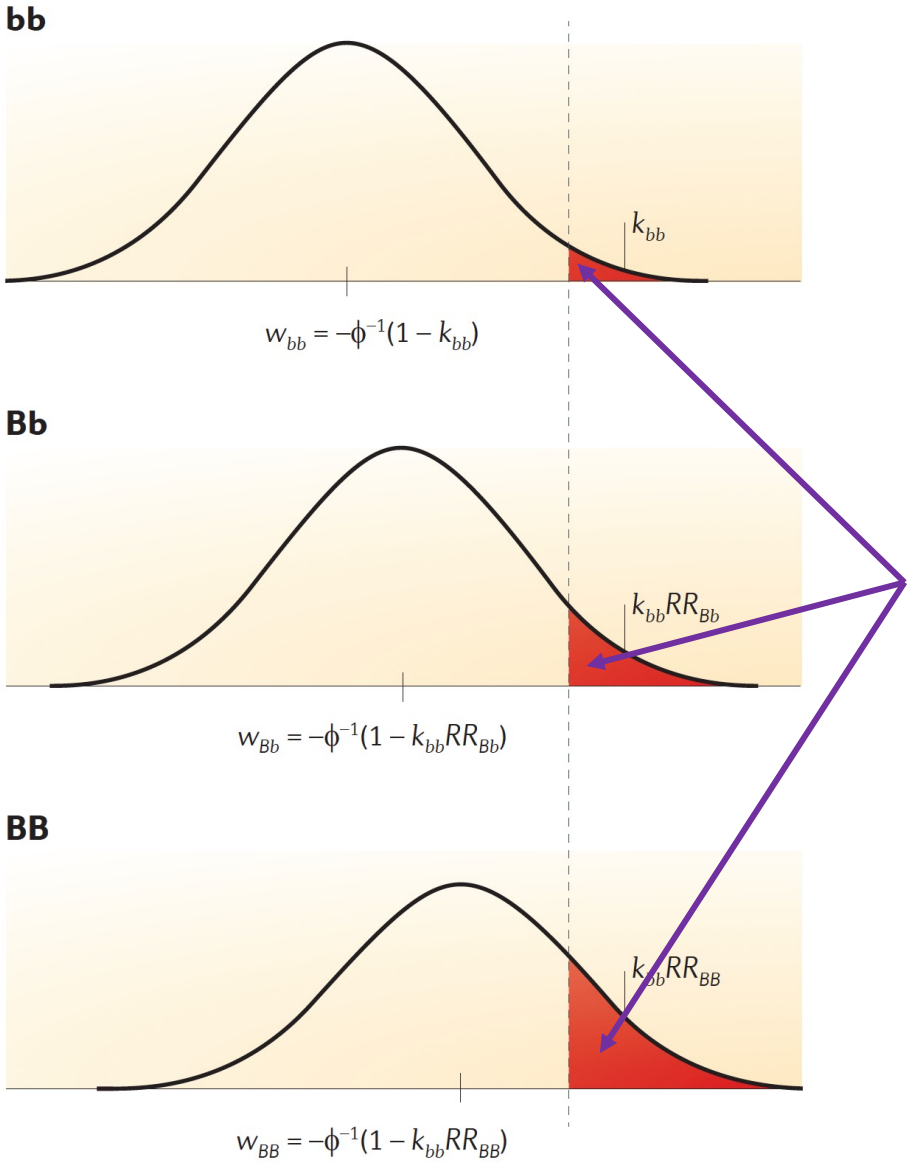
Genotype ***Bb*** → $\Pr(D \mid Bb) = k_{bb} * RR_{Bb}$

Genotype ***BB*** → $\Pr(D \mid BB) = k_{bb} * RR_{BB}$

- Non-linear relationship

Witte, J., Visscher, P. & Wray, N. *Nat Rev Genet* **15**, 765–776 (2014).
https://doi.org/10.1038/nrg3786

Observed scale (visualized under the liability threshold model)



Observed Scale

- **Multiplicative**
- **Non-Linear**

The observed disease risk (prevalence among individuals with that genotype) increases multiplicatively.

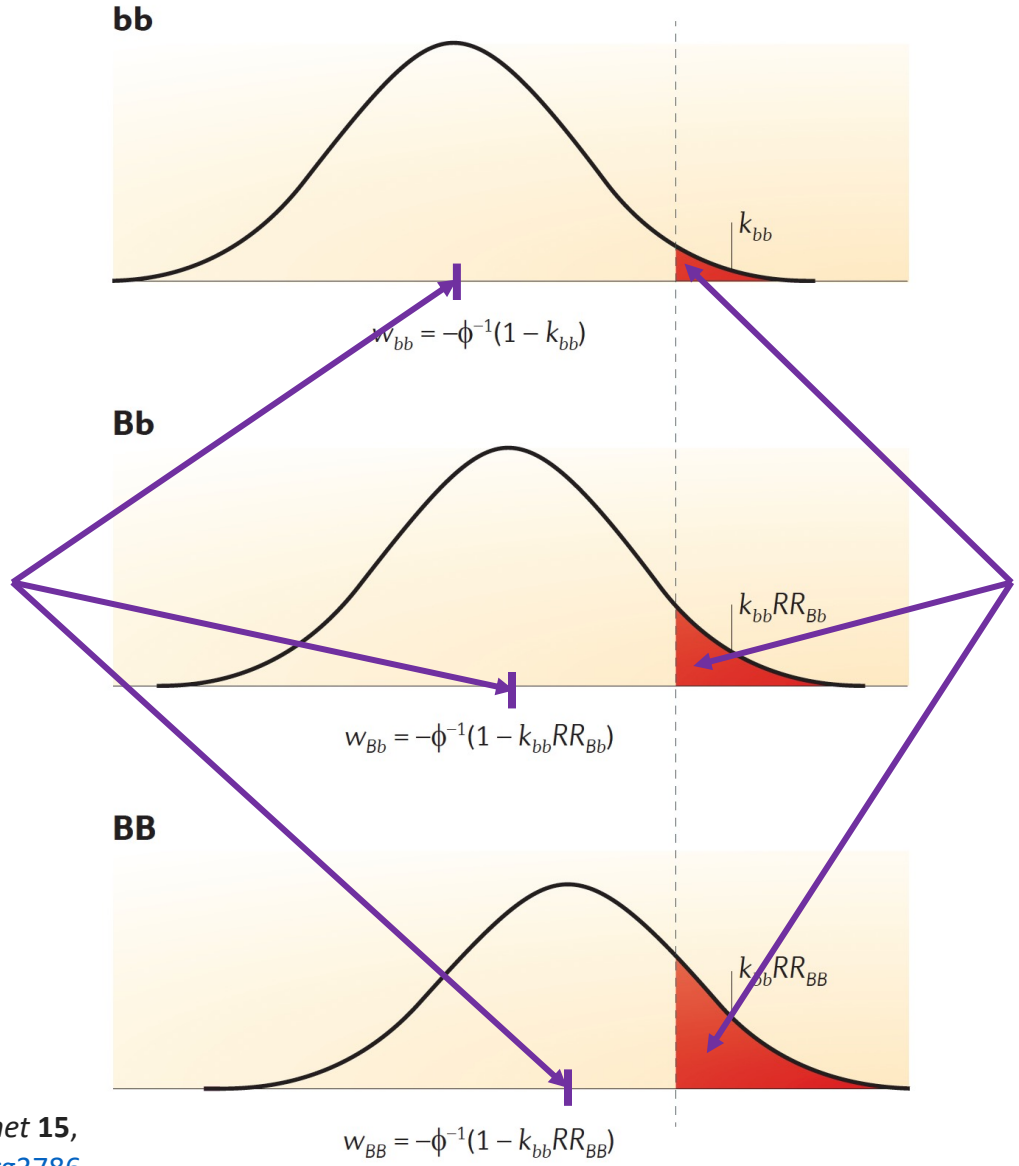
Witte, J., Visscher, P. & Wray, N. *Nat Rev Genet* **15**, 765–776 (2014). https://doi.org/10.1038/nrg3786

Observed scale to Liability scale

Liability Scale

- **Additive**
- **Linear**

The mean latent liability (among individuals with that genotype) increases additively.



Observed Scale

- **Multiplicative**
- **Non-Linear**

The observed disease risk (prevalence among individuals with that genotype) *still* increases multiplicatively.

Witte, J., Visscher, P. & Wray, N. *Nat Rev Genet* **15**, 765–776 (2014). https://doi.org/10.1038/nrg3786

From observed-scale heritability to liability-scale heritability

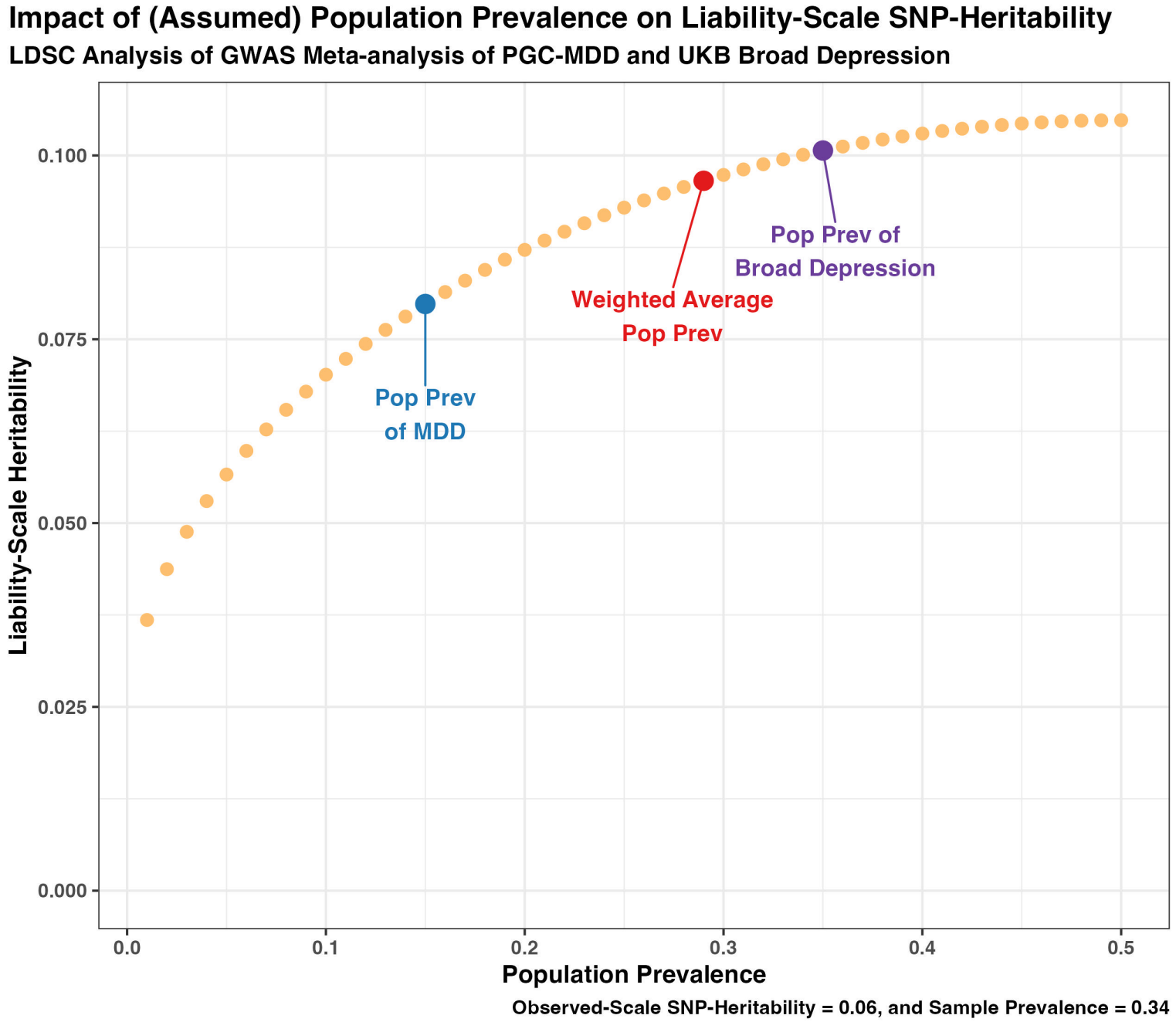
$$h_l^2 = \hat{h}_o^2 \frac{K\,(1-K)}{z^2} \frac{K(1-K)}{P(1-P)}$$

We need to specify two parameters:

- **Sample prevalence (*P*)**
- **Population prevalence (*K*)**
  - Note that *z* is computed from *K*, so need not be specified.

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). *Am J Hum Genet*, *88*(3), 294–305. https://doi.org/10.1016/j.ajhg.2011.02.002

Bonus Phenotype 2 – Major Depressive Disorder (MDD)
Time 15 minutes


Impact of (Assumed) Population Prevalence on Liability-Scale SNP-Heritability
LDSC Analysis of GWAS Meta-analysis of PGC-MDD and UKB Broad Depression
Liability-Scale Heritability
0.100
0.075
0.050
0.025
0.000
0.0
0.1
0.2
0.3
0.4
0.5
Population Prevalence
Pop Prev of MDD
Weighted Average Pop Prev
Pop Prev of Broad Depression
Observed-Scale SNP-Heritability = 0.06, and Sample Prevalence = 0.34

References

Witte, J. S., Visscher, P. M., & Wray, N. R. (2014). The contribution of genetic variants to disease depends on the ruler. *Nature Reviews Genetics*, *15*(11), 765-776. https://doi.org/10.1038/nrg3786

Lee, Sang H., Wray, Naomi R., Goddard, Michael E., & Visscher, Peter M. (2011). Estimating Missing Heritability for Disease from Genome-wide Association Studies. *The American Journal of Human Genetics*, *88*(3), 294-305. https://doi.org/10.1016/j.ajhg.2011.02.002