

MULTI-TRAIT LIABILITY MODEL FOR GENE DISCOVERY AND PREDICTION IN LARGE BIOBANKS



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



WHY STUDY THE GENETICS OF COMPLEX TRAITS?

Why do some people develop a disorder and others do not?

Part of the answer lies in biology and measuring that biological component



WHY STUDY THE GENETICS OF COMPLEX TRAITS?

Why do some people develop a disorder and others do not?

Part of the answer lies in biology and measuring that biological component

Two approaches can be taken:

- Understanding the biological bases of a disorder (gene discovery)
- Identifying individuals with an elevated risk of disorder (genetic prediction)



WHY STUDY THE GENETICS OF COMPLEX TRAITS?

Why do some people develop a disorder and others do not?

Part of the answer lies in biology and measuring that biological component

Two approaches can be taken:

- Understanding the biological bases of a disorder (gene discovery)
- Identifying individuals with an elevated risk of disorder (genetic prediction)

Statistical challenge of extracting weak signals from very high-dimensional data

This talk is focused around a method that utilise both family history information and correlated disorders



FROM GENOME TO DATA

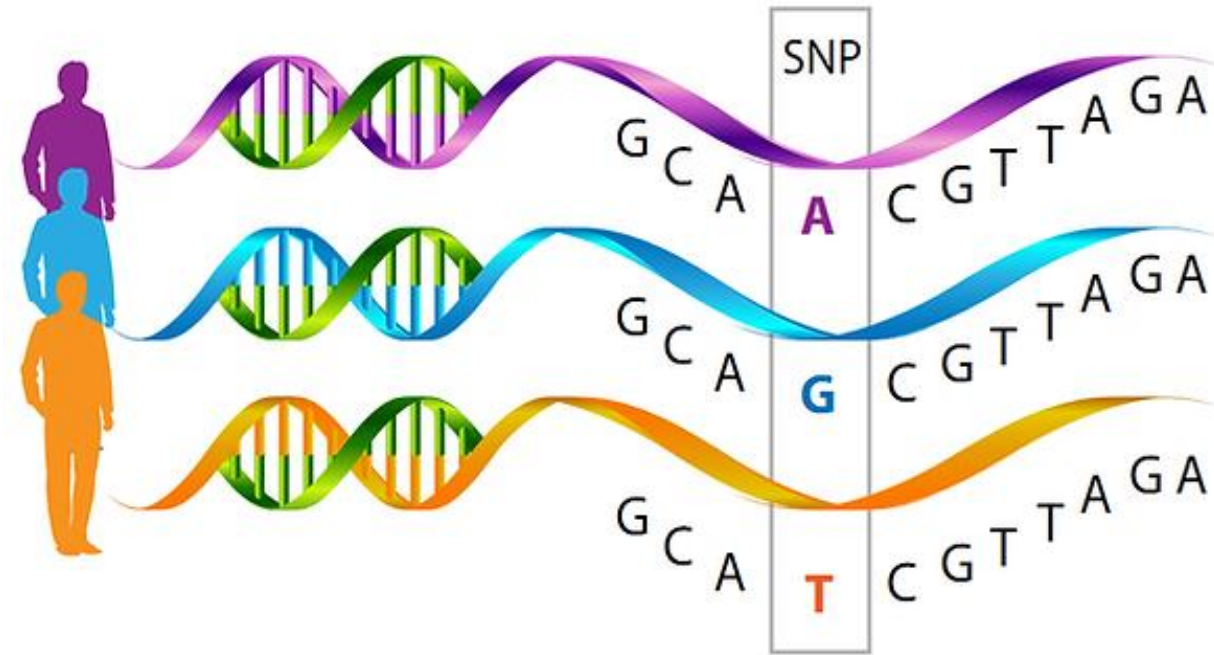
The human genome: ~3 billion base pairs

Most positions are identical across individuals

Single Nucleotide Polymorphism (SNP):

- positions where individuals differ

A lot of analysis in human genetics is based on SNPs



<https://www.nutrigeneticsspecialists.com/single-post/2017/03/27/what-is-a-snp>

FROM GENOME TO DATA

A genome is coded as a vector of SNPs

Each position is coded as 0, 1, or 2

- SNPs follow $\text{binom}(2, p)$
- p : minor allele frequency

Genotype
Matrix

$$\begin{matrix} & \overbrace{\hspace{10em}}^M \\ N \left\{ \begin{bmatrix} 1 & \dots & 2 \\ 2 & & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 0 \\ 0 & & 1 \end{bmatrix} \right. \end{matrix}$$



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



FROM GENOME TO DATA

A genome is coded as a vector of SNPs

Each position is coded as 0, 1, or 2

- SNPs follow $\text{binom}(2, p)$
- p : minor allele frequency

Very high dimensional data:

Typical genotype matrix dimensions

- N is the 100.000s
- M is in the millions

Joint modelling is not feasible

Genotype
Matrix

$$\begin{matrix} & \overbrace{\hspace{10em}}^M \\ N \left\{ \begin{bmatrix} 1 & \dots & 2 \\ 2 & & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 0 \\ 0 & & 1 \end{bmatrix} \right. \end{matrix}$$



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



LIABILITY THRESHOLD MODEL (LTM)

Liability threshold model elegantly links continuous variation with binary outcomes.

Every individual has a (hidden) liability,
 $\ell = \ell_g + \ell_e$, with $\ell_g \sim N(0, h^2)$ independent of $\ell_e \sim N(0, 1 - h^2)$, h^2 is the heritability:

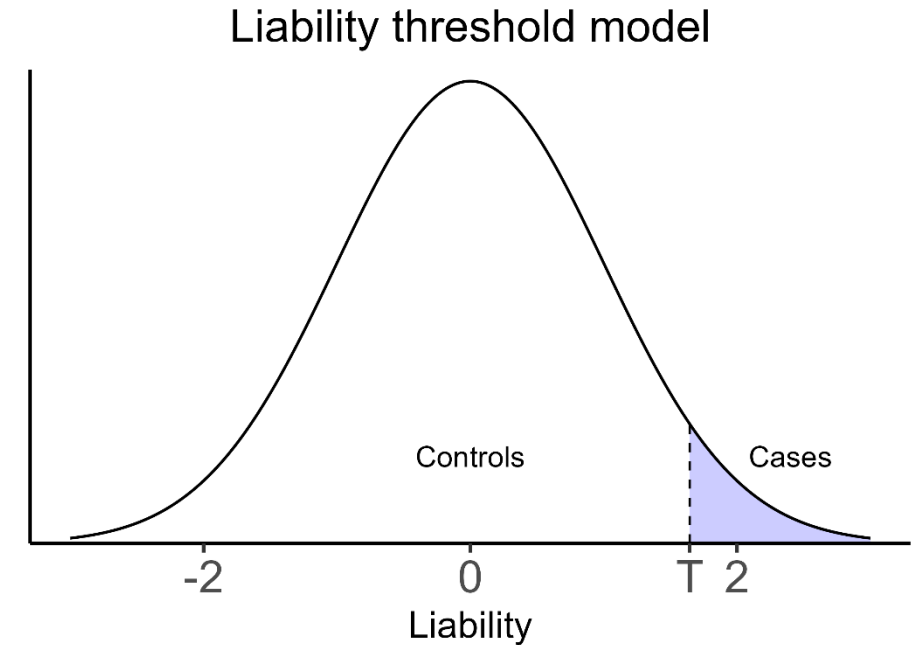
$$\ell \sim N(0,1)$$

Case-Control status:

$$y = \begin{cases} 1, & T < \ell \\ 0, & \text{otherwise} \end{cases}$$

Where T is given by $P(\ell > T) = k$

k is the life-time prevalence



GENOME-WIDE ASSOCIATION STUDY

For N individuals, M SNPs, and k covariates - Linear regression GWAS – a SNP-by-SNP scan:

$$y = \underbrace{\beta G_j}_{SNP} + \underbrace{X\gamma}_{covariates} + \underbrace{\varepsilon}_{noise}, \quad \varepsilon \sim N(0, \sigma^2 I_N)$$

Test: $H_0: \beta = 0$ against $H_A: \beta \neq 0$ for each SNP



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



GENOME-WIDE ASSOCIATION STUDY

For N individuals, M SNPs, and k covariates - Linear regression GWAS – a SNP-by-SNP scan:

$$y = \underbrace{\beta G_j}_{\text{SNP}} + \underbrace{X\gamma}_{\text{covariates}} + \underbrace{\varepsilon}_{\text{noise}}, \quad \varepsilon \sim N(0, \sigma^2 I_N)$$

Test: $H_0: \beta = 0$ against $H_A: \beta \neq 0$ for each SNP

Summary statistics are effect size estimate, test statistic, and p-value given by:

$$\hat{\beta} = \frac{G_j^T y}{G_j^T G_j}, \quad \text{var}(\hat{\beta}) \approx \frac{\sigma^2}{2np(1-p)}, \quad \text{where } T = \frac{\hat{\beta}}{SE(\hat{\beta})} \sim t_{N-k-1} \quad \text{and } p = 2(1 - \Phi(|T|))$$



GENOME-WIDE ASSOCIATION STUDY

For N individuals, M SNPs, and k covariates - Linear regression GWAS – a SNP-by-SNP scan:

$$y = \underbrace{\beta G_j}_{\text{SNP}} + \underbrace{X\gamma}_{\text{covariates}} + \underbrace{\varepsilon}_{\text{noise}}, \quad \varepsilon \sim N(0, \sigma^2 I_N)$$

Test: $H_0: \beta = 0$ against $H_A: \beta \neq 0$ for each SNP

Summary statistics are effect size estimate, test statistic, and p-value given by:

$$\hat{\beta} = \frac{G_j^T y}{G_j^T G_j}, \quad \text{var}(\hat{\beta}) \approx \frac{\sigma^2}{2np(1-p)}, \quad \text{where } T = \frac{\hat{\beta}}{SE(\hat{\beta})} \sim t_{N-k-1} \quad \text{and } p = 2(1 - \Phi(|T|))$$

Multiple testing correction with Bonferroni correction, assuming 1 million independent SNPs

Genome-wide significant threshold: $p < 5 \times 10^{-8}$



GENOME-WIDE ASSOCIATION STUDY

For N individuals, M SNPs, and k covariates - Linear regression GWAS – a SNP-by-SNP scan:

$$y = \underbrace{\beta G_j}_{\text{SNP}} + \underbrace{X\gamma}_{\text{covariates}} + \underbrace{\varepsilon}_{\text{noise}}, \quad \varepsilon \sim N(0, \sigma^2 I_N)$$

Test: $H_0: \beta = 0$ against $H_A: \beta \neq 0$ for each SNP

Summary statistics are effect size estimate, test statistic, and p-value given by:

$$\hat{\beta} = \frac{G_j^T y}{G_j^T G_j}, \quad \text{var}(\hat{\beta}) \approx \frac{\sigma^2}{2np(1-p)}, \quad \text{where } T = \frac{\hat{\beta}}{SE(\hat{\beta})} \sim t_{N-k-1} \quad \text{and } p = 2(1 - \Phi(|T|))$$

Multiple testing correction with Bonferroni correction, assuming 1 million independent SNPs

Genome-wide significant threshold: $p < 5 \times 10^{-8}$

Power is maximised when y is the genetic liability



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



SINGLE-TRAIT LTFGRS

Liability threshold-family genetic risk score

(LT-FGRS) estimates genetic liability and

extends the LTM.



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



SINGLE-TRAIT LTFGRS

Liability threshold-family genetic risk score

(LT-FGRS) estimates genetic liability and extends the LTM.

Family history can be accounted for by:

$$(\ell_g, \ell, \ell_{p_1}, \ell_{p_2})^T \sim N(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} h^2 & h^2 & 0.5h^2 & 0.5h^2 \\ h^2 & 1 & 0.5h^2 & 0.5h^2 \\ 0.5h^2 & 0.5h^2 & 1 & 0 \\ 0.5h^2 & 0.5h^2 & 0 & 1 \end{pmatrix}$$

Covariance structure depend on heritability, h^2 , and expected genetic overlap



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



SINGLE-TRAIT LTFGRS

Liability threshold-family genetic risk score

(LT-FGRS) estimates genetic liability and extends the LTM.

Family history can be accounted for by:

$$(\ell_g, \ell, \ell_{p_1}, \ell_{p_2})^T \sim N(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} h^2 & h^2 & 0.5h^2 & 0.5h^2 \\ h^2 & 1 & 0.5h^2 & 0.5h^2 \\ 0.5h^2 & 0.5h^2 & 1 & 0 \\ 0.5h^2 & 0.5h^2 & 0 & 1 \end{pmatrix}$$

Covariance structure depend on heritability, h^2 , and expected genetic overlap

We desire to estimate:

$$\widehat{\ell}_g = E[\ell_g | Y], \quad \text{with } Y = (y, y_{p_1}, y_{p_2})^T$$

Conditioning on Y leads to restrictions on the liabilities and a high dimensional truncated normal



SINGLE-TRAIT LTFGRS

Liability threshold-family genetic risk score

(LT-FGRS) estimates genetic liability and extends the LTM.

Family history can be accounted for by:

$$(\ell_g, \ell, \ell_{p_1}, \ell_{p_2})^T \sim N(0, \Sigma)$$

$$\Sigma = \begin{pmatrix} h^2 & h^2 & 0.5h^2 & 0.5h^2 \\ h^2 & 1 & 0.5h^2 & 0.5h^2 \\ 0.5h^2 & 0.5h^2 & 1 & 0 \\ 0.5h^2 & 0.5h^2 & 0 & 1 \end{pmatrix}$$

Covariance structure depend on heritability, h^2 , and expected genetic overlap

We desire to estimate:

$$\widehat{\ell}_g = E[\ell_g | Y], \quad \text{with } Y = (y, y_{p_1}, y_{p_2})^T$$

Conditioning on Y leads to restrictions on the liabilities and a high dimensional truncated normal

We estimate $\widehat{\ell}_g$ with either

- i) Gibbs Sampler (sampling)
- ii) Pearson-Aitken selection formula
(iterative analytical)



MULTI-TRAIT LTFGRS

Consider two disorders, t_1 and t_2 with

$$\ell_{t_1} = (\ell_g, \ell, \ell_{p_1}, \ell_{p_2})^T \sim N(0, \Sigma^{(t_1)})$$

And ℓ_{t_2} the same, but for t_2 .



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



MULTI-TRAIT LTFGRS

Consider two disorders, t_1 and t_2 with

$$\ell_{t_1} = (\ell_g, \ell, \ell_{p_1}, \ell_{p_2})^T \sim N(0, \Sigma^{(t_1)})$$

And ℓ_{t_2} the same, but for t_2 .

Then define

$$l = (\ell_{t_1}, \ell_{t_2})^T \sim N(0, \tilde{\Sigma})$$

With

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma^{(t_1)} & \Sigma^{(t_1 t_2)} \\ \Sigma^{(t_1 t_2)} & \Sigma^{(t_2)} \end{pmatrix}$$

Where $\Sigma^{(t_i)}$ is given as before



MULTI-TRAIT LTFGRS

Consider two disorders, t_1 and t_2 with

$$\ell_{t_1} = (\ell_g, \ell, \ell_{p_1}, \ell_{p_2})^T \sim N(0, \Sigma^{(t_1)})$$

And ℓ_{t_2} the same, but for t_2 .

Then define

$$l = (\ell_{t_1}, \ell_{t_2})^T \sim N(0, \tilde{\Sigma})$$

With

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma^{(t_1)} & \Sigma^{(t_1 t_2)} \\ \Sigma^{(t_1 t_2)} & \Sigma^{(t_2)} \end{pmatrix}$$

Where $\Sigma^{(t_i)}$ is given as before

We define entry-wise:

$$(\Sigma^{(t_1 t_2)})_{ij} = K_{ij} \rho_{t_1 t_2} \sqrt{h_{t_1}^2 h_{t_2}^2}$$

with genetic correlation $\rho_{t_1 t_2}$ and expected genetic overlap K_{ij}



MULTI-TRAIT LTFGRS

Consider two disorders, t_1 and t_2 with

$$\ell_{t_1} = (\ell_g, \ell, \ell_{p_1}, \ell_{p_2})^T \sim N(0, \Sigma^{(t_1)})$$

And ℓ_{t_2} the same, but for t_2 .

Then define

$$l = (\ell_{t_1}, \ell_{t_2})^T \sim N(0, \tilde{\Sigma})$$

With

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma^{(t_1)} & \Sigma^{(t_1 t_2)} \\ \Sigma^{(t_1 t_2)} & \Sigma^{(t_2)} \end{pmatrix}$$

Where $\Sigma^{(t_i)}$ is given as before

We define entry-wise:

$$(\Sigma^{(t_1 t_2)})_{ij} = K_{ij} \rho_{t_1 t_2} \sqrt{h_{t_1}^2 h_{t_2}^2}$$

with genetic correlation $\rho_{t_1 t_2}$ and expected genetic overlap K_{ij}

$$K_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0.5 & \text{if } i, j \text{ 1}^{\text{st}} \text{ degree relatives} \\ 0.25 & \text{if } i, j \text{ 2}^{\text{nd}} \text{ degree relatives} \\ 0.125 & \text{if } i, j \text{ 3}^{\text{rd}} \text{ degree relatives} \\ 0 & \text{Otherwise} \end{cases}$$

MULTI-TRAIT LTFGRS

We are not restricted to two disorders:

$$l = (\ell_{t_1}, \ell_{t_2}, \dots, \ell_{t_{\tilde{N}}})^T \sim N(0, \tilde{\Sigma})$$

With

$$\tilde{\Sigma} = \begin{pmatrix} \Sigma^{(t_1)} & \dots & \Sigma^{(t_1 t_{\tilde{N}})} \\ \vdots & \ddots & \vdots \\ \Sigma^{(t_{\tilde{N}} t_1)} & \dots & \Sigma^{(t_{\tilde{N}})} \end{pmatrix}$$

Where $\Sigma^{(t_i)}$ is given as before

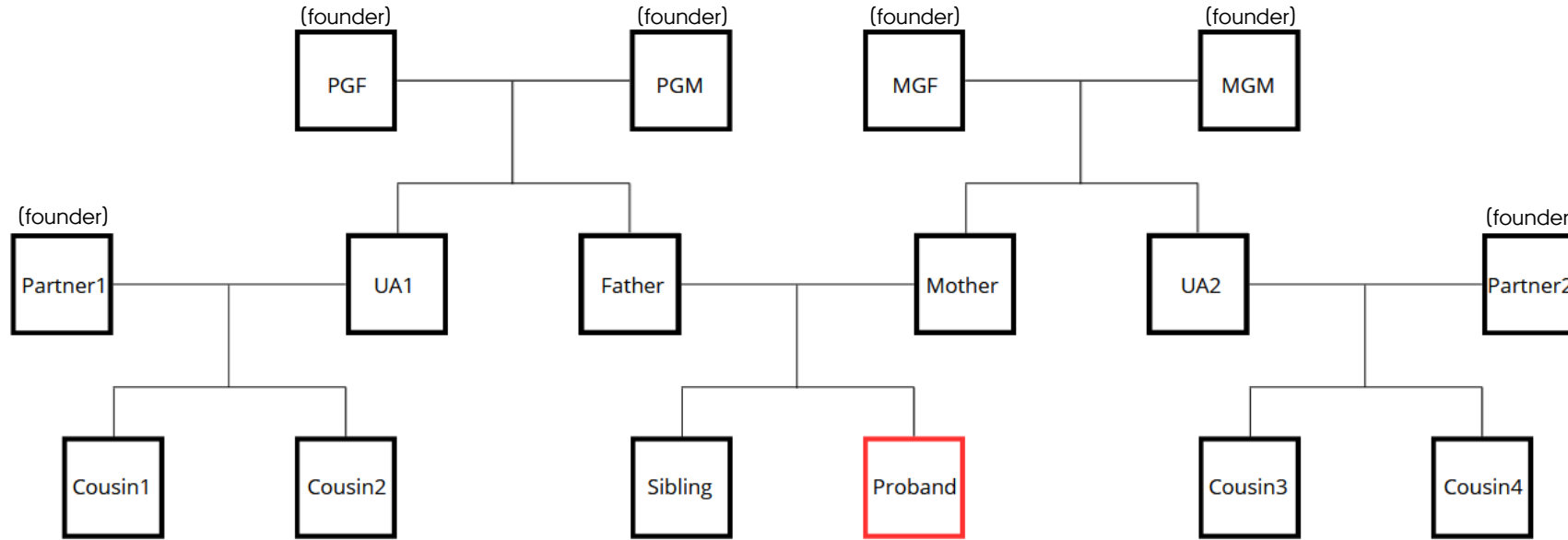
We define entry-wise:

$$(\Sigma^{(t_n t_m)})_{ij} = K_{ij} \rho_{t_n t_m} \sqrt{h_{t_n}^2 h_{t_m}^2}$$

with genetic correlation $\rho_{t_n t_m}$ and expected genetic overlap

$$K_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0.5 & \text{if } i, j \text{ 1}^{\text{st}} \text{ degree relatives} \\ 0.25 & \text{if } i, j \text{ 2}^{\text{nd}} \text{ degree relatives} \\ 0.125 & \text{if } i, j \text{ 3}^{\text{rd}} \text{ degree relatives} \\ 0 & \text{Otherwise} \end{cases}$$

SIMULATION SETUP - GENOTYPES



We set:

$$N = 1,000,000$$

$$M = 10,000$$

$$MAF_j \sim Unif(0, 1/2)$$

For founders are drawn from:

$$G_j \sim bin(2, MAF_j) \text{ for } j = 1, \dots, M$$

Transmission from parent to offspring:

$$\begin{array}{lcl} 0 & \rightarrow & 0 \\ 1 & \rightarrow & bern(0.5) \\ 2 & \rightarrow & 1 \end{array}$$



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



SIMULATION SETUP - PHENOTYPES

We simulate **3 phenotypes**

Heritability: $h^2 = 0.5$

genetic correlation: $\rho_G = 0.25$

Effect sizes: $C = 1,000$ draws from
 $\beta = (\beta_1, \beta_2, \beta_3)^T \sim N(0, \Sigma)$

$$\Sigma = \begin{pmatrix} h^2 & -\rho_G & \rho_G \\ -\rho_G & h^2 & \rho_G \\ \rho_G & \rho_G & h^2 \end{pmatrix}$$

Overlap:

Traits share 50% or 100% of the same locations

Genetic liability:

$$\ell_g = \frac{\beta_k^T G_i}{\sqrt{C}} \sim N(0, h^2)$$

G_i is (scaled) SNP vector of individual i



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



SIMULATION SETUP - PHENOTYPES

We simulate **3 phenotypes**

Heritability: $h^2 = 0.5$

genetic correlation: $\rho_G = 0.25$

Effect sizes: $C = 1,000$ draws from
 $\beta = (\beta_1, \beta_2, \beta_3)^T \sim N(0, \Sigma)$

$$\Sigma = \begin{pmatrix} h^2 & -\rho_G & \rho_G \\ -\rho_G & h^2 & \rho_G \\ \rho_G & \rho_G & h^2 \end{pmatrix}$$

Overlap:

Traits share 50% or 100% of the same locations

Genetic liability:

$$\ell_g = \frac{\beta_k^T G_i}{\sqrt{C}} \sim N(0, h^2)$$

G_i is (scaled) SNP vector of individual i

The full liability is then

$$\ell = \ell_g + \ell_e$$

With

$$\ell_e \sim N(0, 1 - h^2)$$

Case-control status assigned with

$$Y = \begin{cases} 1, & \ell > T \\ 0, & \ell \leq T \end{cases}$$

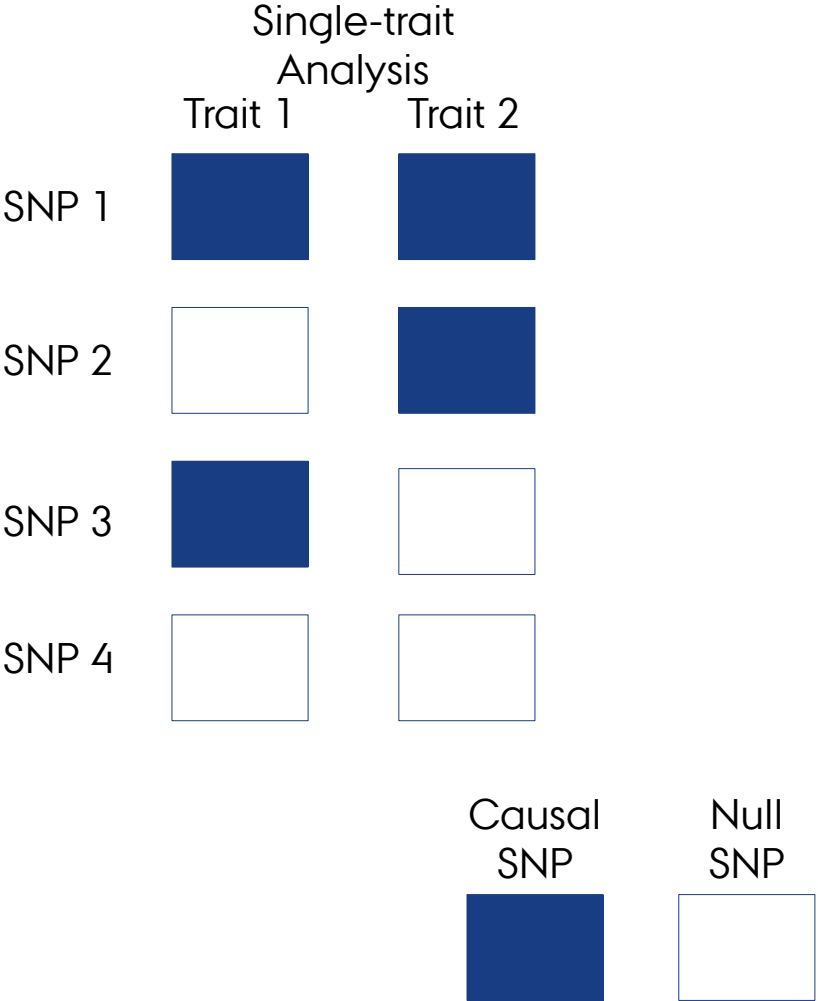
Where T satisfy

$$P(\ell > T) = K$$

Life-time prevalence of the trait ($K = 0.05$)



SHARED AND TRAIT-SPECIFIC EFFECTS



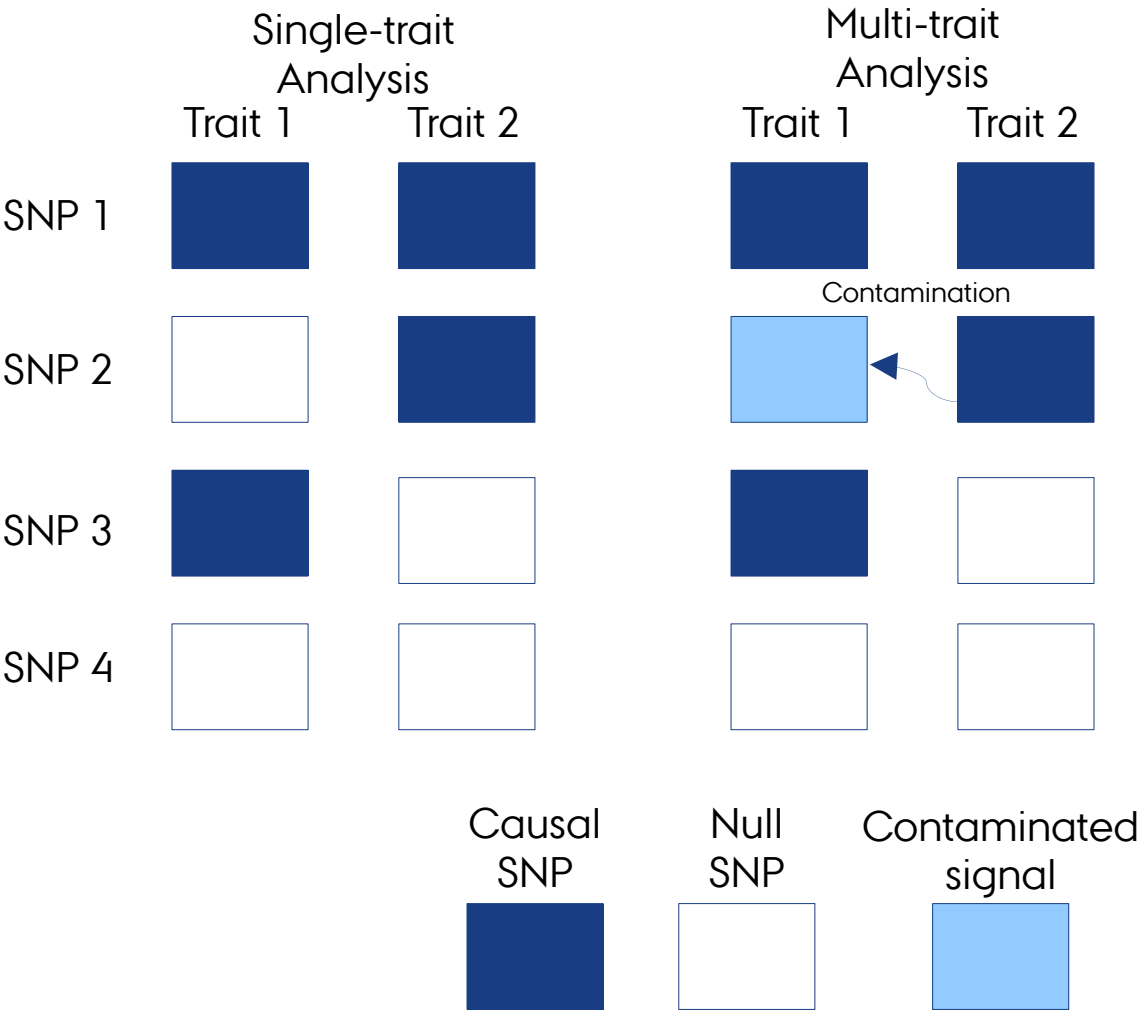
SHARED AND TRAIT-SPECIFIC EFFECTS

Others have shown information leakage exists when conditioning on other traits

Does partial overlap lead to "signal contamination" in GWAS?

Is specificity to the primary trait reduced in prediction?

"Signal Contamination"

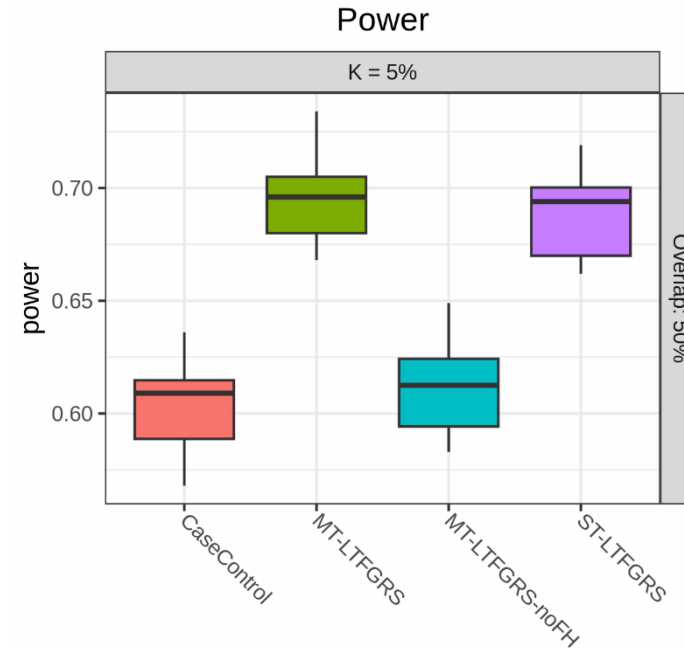


SIMULATION RESULTS: GWAS

Power: $\frac{1}{C} \sum_{i=1}^M 1_{p_i < 5 \times 10^{-8}}$

Including FH is an improvement over Case-Control status

MT offers modest improvement



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



SIMULATION RESULTS: GWAS

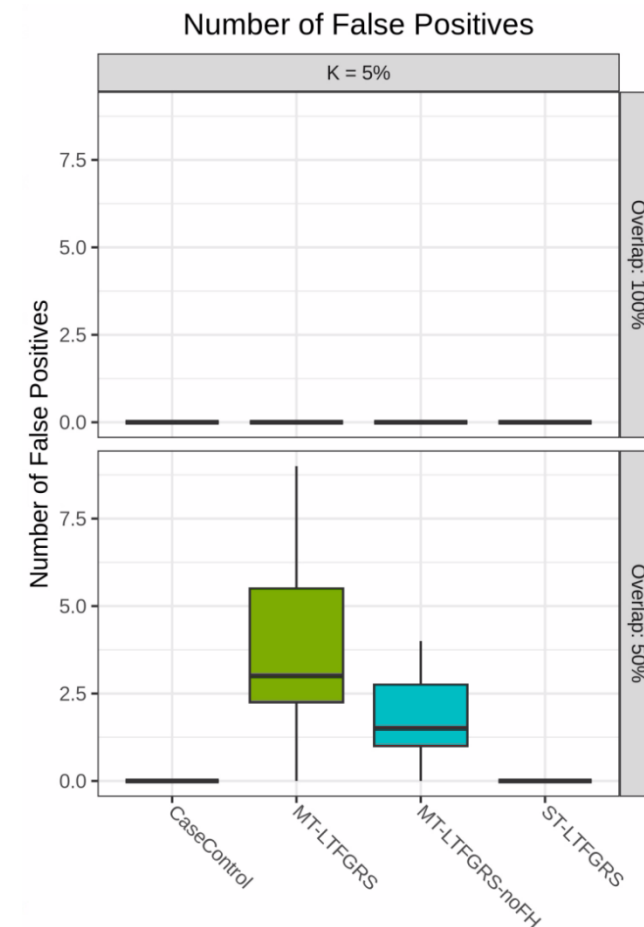
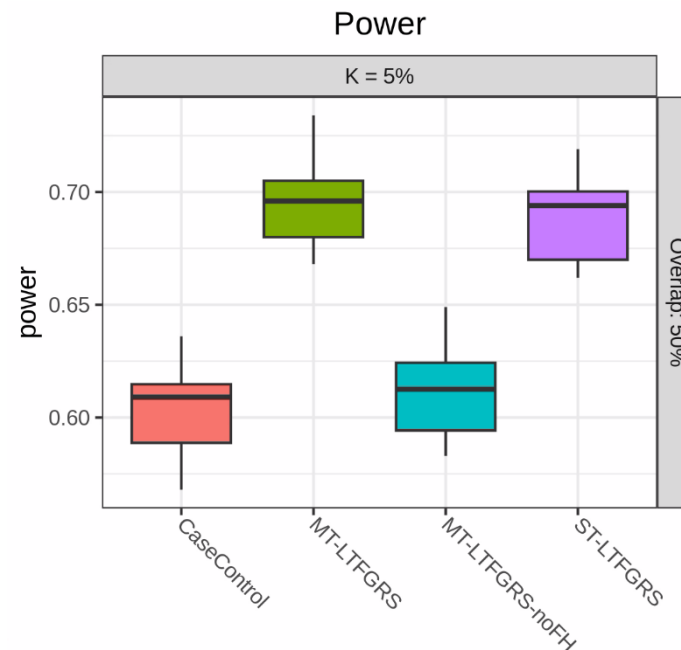
Power: $\frac{1}{C} \sum_{i=1}^M 1_{p_i < 5 \times 10^{-8}}$

Including FH is an improvement over Case-Control status

MT offers modest improvement

False Positives:

Multi-trait phenotypes show signs of ‘Signal Contamination’, even at modest power gains

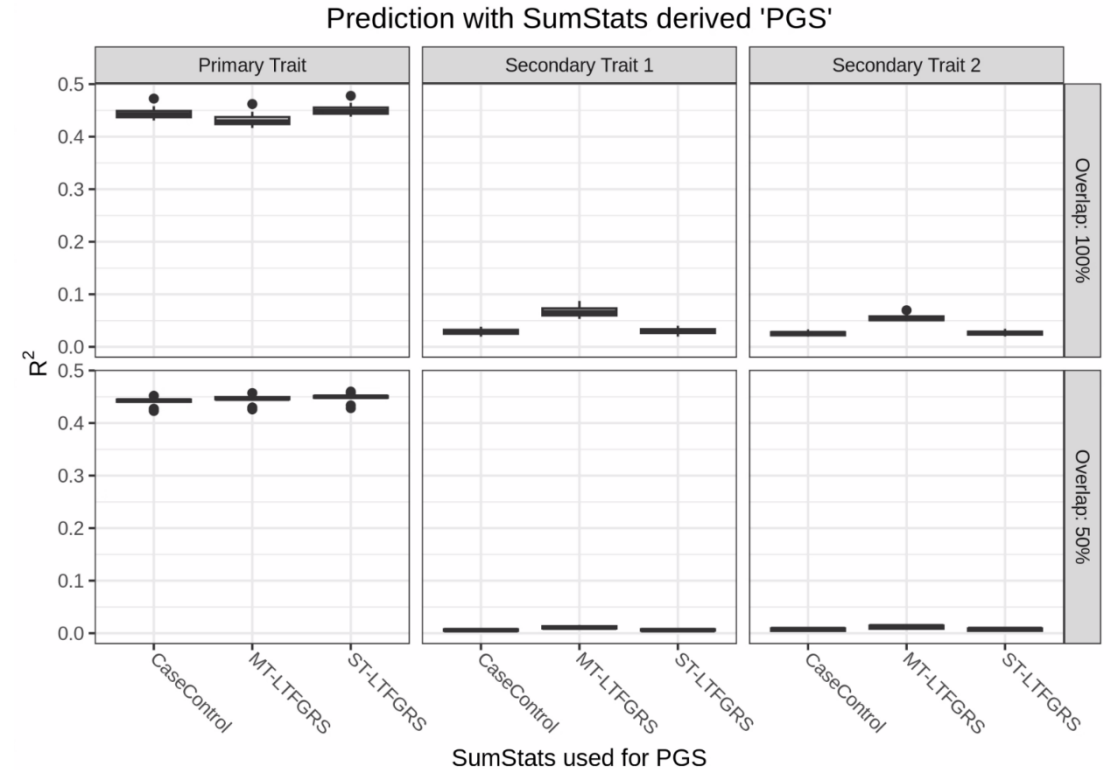


SIMULATION RESULTS: PREDICTION

Simple PGS's from summary statistics:

$$PGS = \tilde{\beta}^T \tilde{G}_i$$

Where $\tilde{\beta}$ are SNPs with $p < 5 \times 10^{-8}$



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



SIMULATION RESULTS: PREDICTION

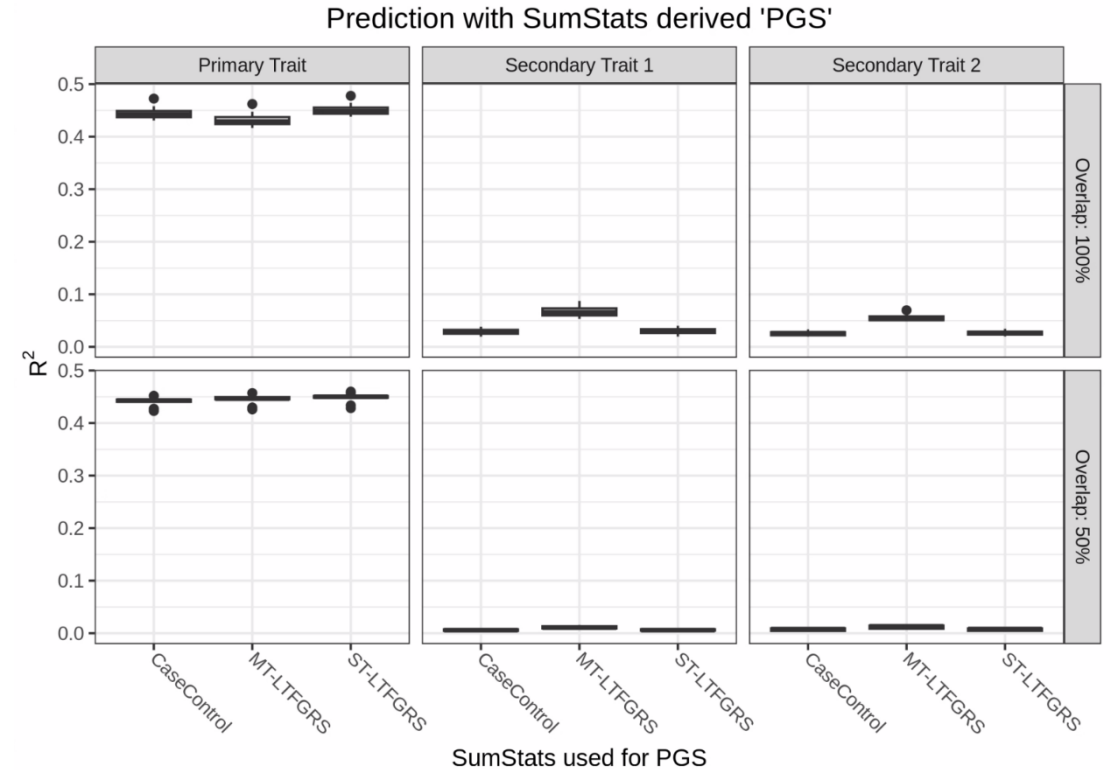
Simple PGS's from summary statistics:

$$PGS = \tilde{\beta}^T \tilde{G}_i$$

Where $\tilde{\beta}$ are SNPs with $p < 5 \times 10^{-8}$

Primary: R^2 values close to theoretical max (h^2)

Secondary: R^2 is elevated for MT-LTFGRS



SIMULATION RESULTS: PREDICTION

Simple PGS's from summary statistics:

$$PGS = \tilde{\beta}^T \tilde{G}_i$$

Where $\tilde{\beta}$ are SNPs with $p < 5 \times 10^{-8}$

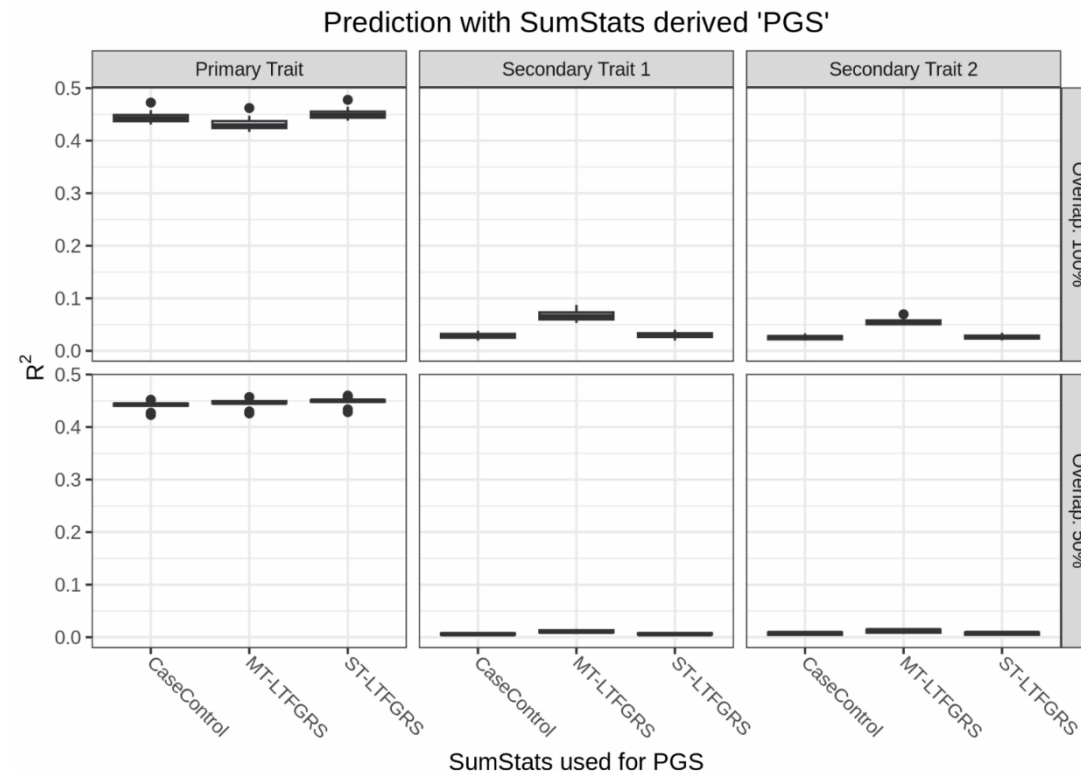
Primary: R^2 values close to theoretical max (h^2)

Secondary: R^2 is elevated for MT-LTFGRS

Simulation conclusions – GWAS:

power may be improved

signal contamination is real



SIMULATION RESULTS: PREDICTION

Simple PGS's from summary statistics:

$$PGS = \tilde{\beta}^T \tilde{G}_i$$

Where $\tilde{\beta}$ are SNPs with $p < 5 \times 10^{-8}$

Primary: R^2 values close to theoretical max (h^2)

Secondary: R^2 is elevated for MT-LTFGRS

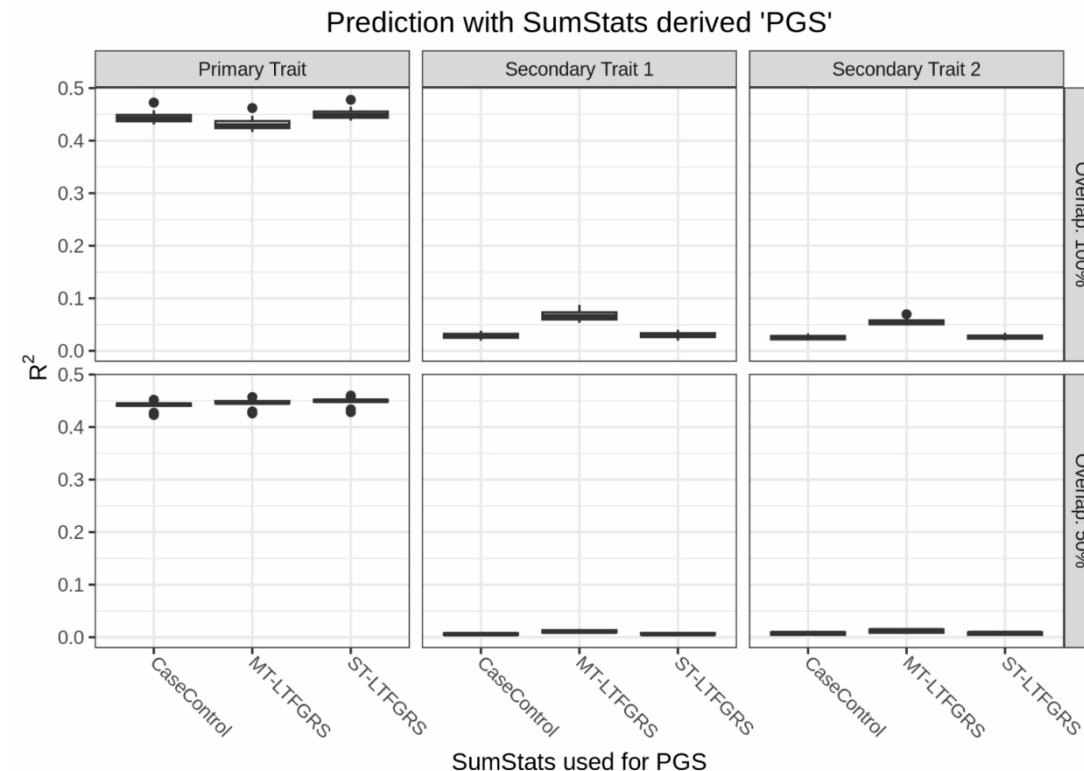
Simulation conclusions – GWAS:

power may be improved

signal contamination is real

Simulation conclusions – Prediction:

MT sumstats may influence downstream analysis in opaque ways



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy

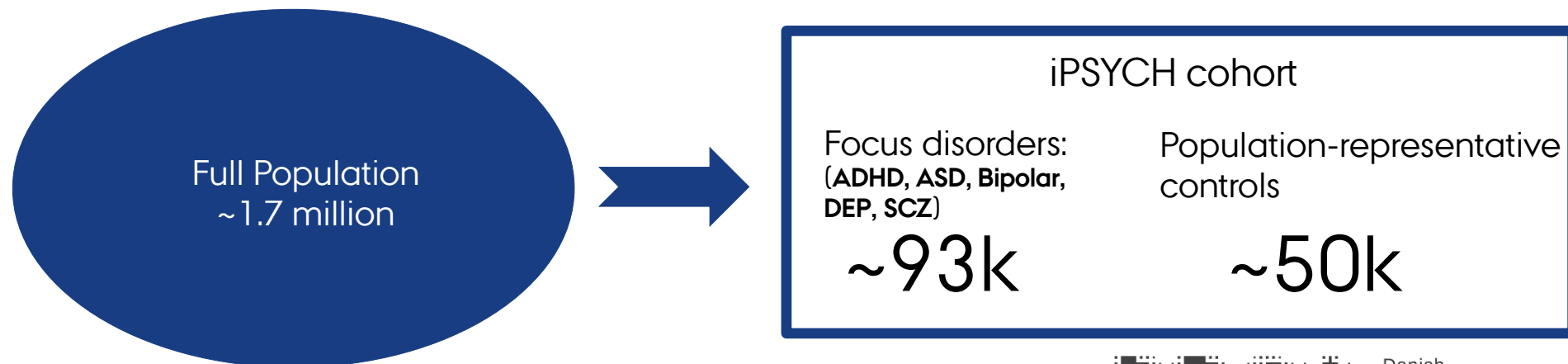


IPSYCH OVERVIEW

- Singletons (no twins) born in Denmark between May 1st, 1981 and December 31st, 2008
- the mother is known
- the child is alive and living in Denmark by their first birthday.

~ 1.7 million individuals satisfy this condition.

iPSYCH sample ALL cases with a target disorder (~93k), and ~50k controls



ADHD GWAS IN IPSYCH

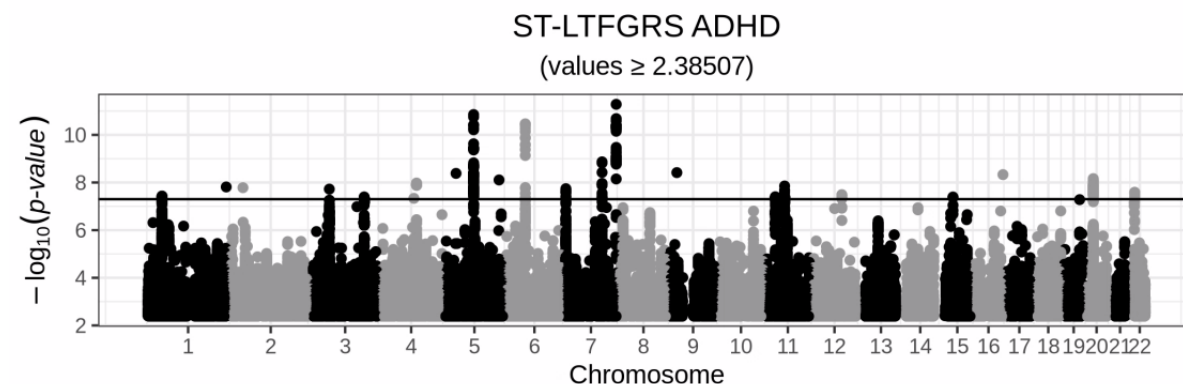
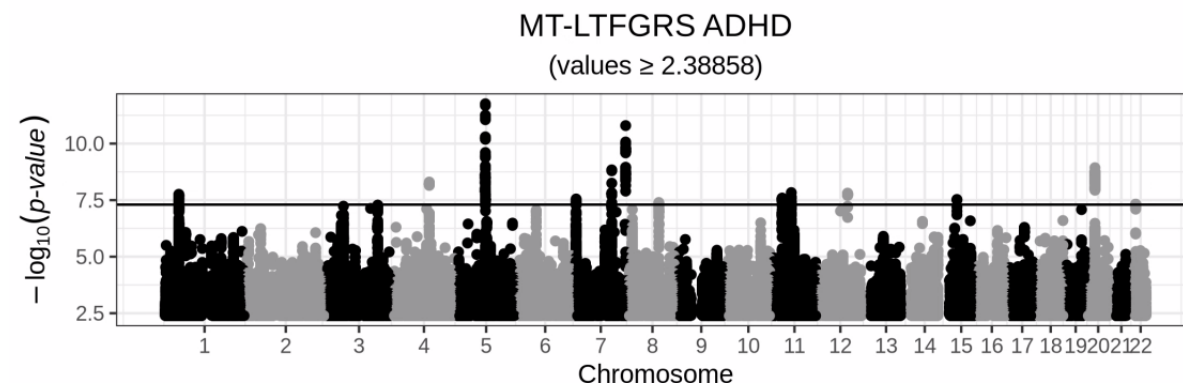
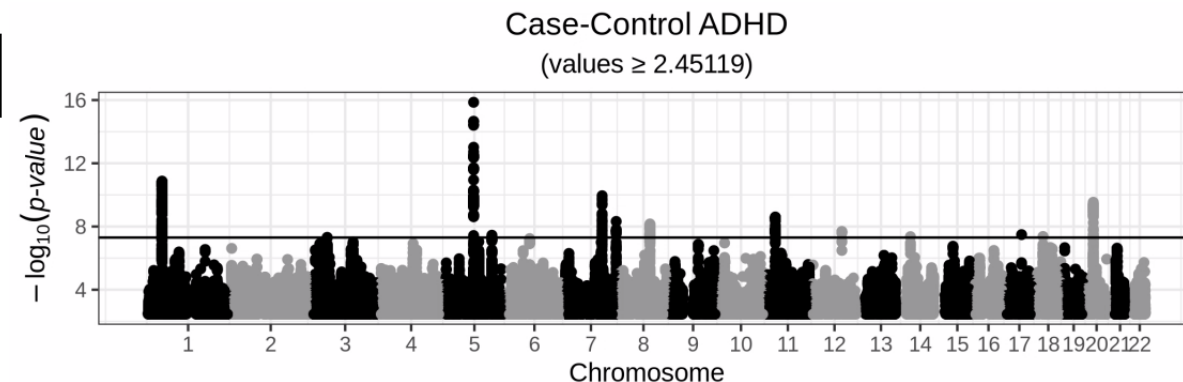
MT-LTFGRS uses information from Autism, Bipolar, Depression, and Schizophrenia too

We perform linear regression GWAS on

- Unrelated individuals
- European genetic ancestry
- Outcomes: MT-LTFGRS, Case-Control, ST-LTFGRS

There is something extra, but neither ST-LTFGRS or MT-LTFGRS offer large improvements in power

This is a 'first-pass' analysis with simple models.



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy

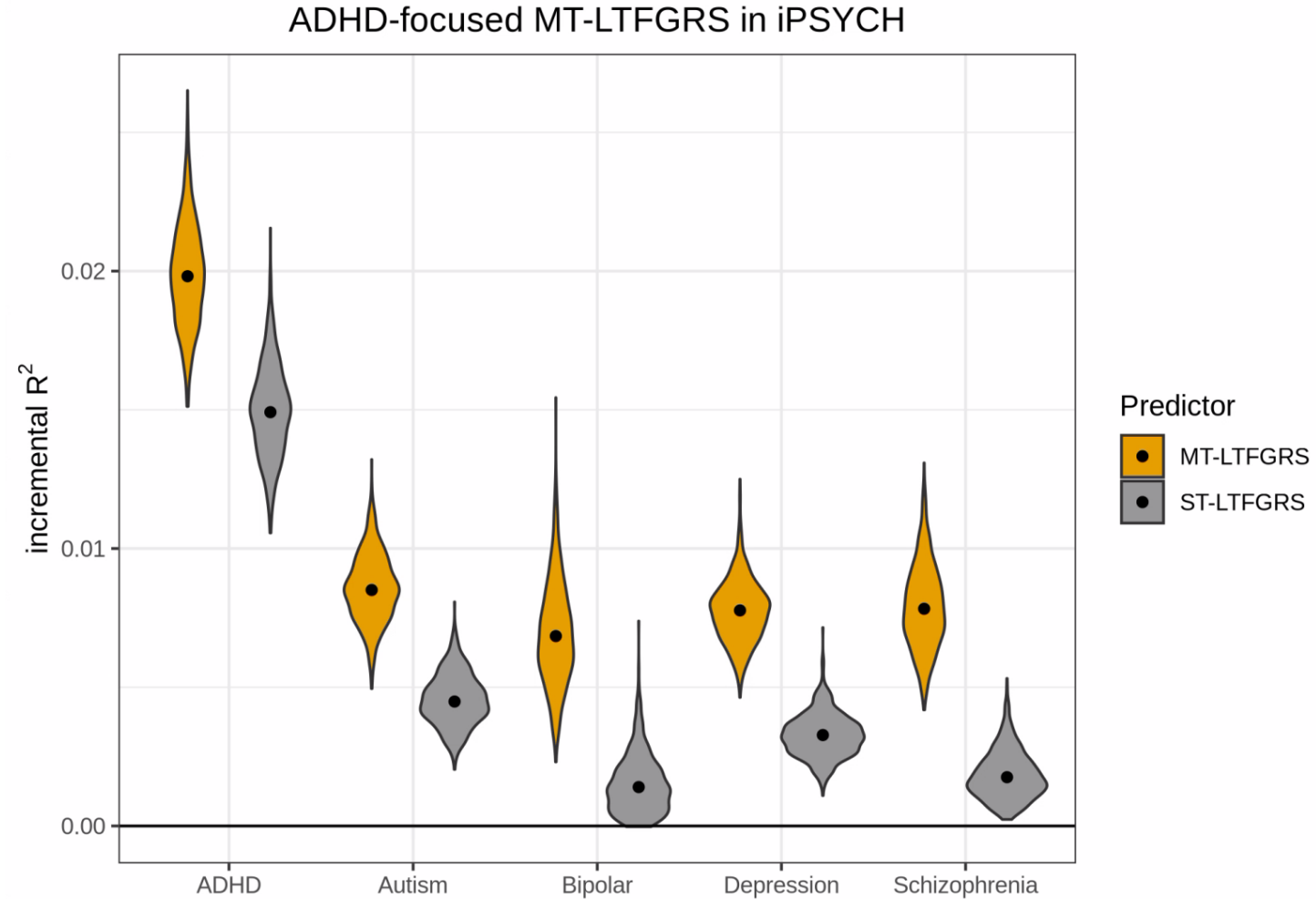


ADHD PREDICTION WITH MT-LTFGRS

proband information is excluded

For all iPSYCH outcomes:

Target outcome $\sim$ base covariates + ADHD-LTFGRS



ADHD PREDICTION WITH MT-LTFGRS

proband information is excluded

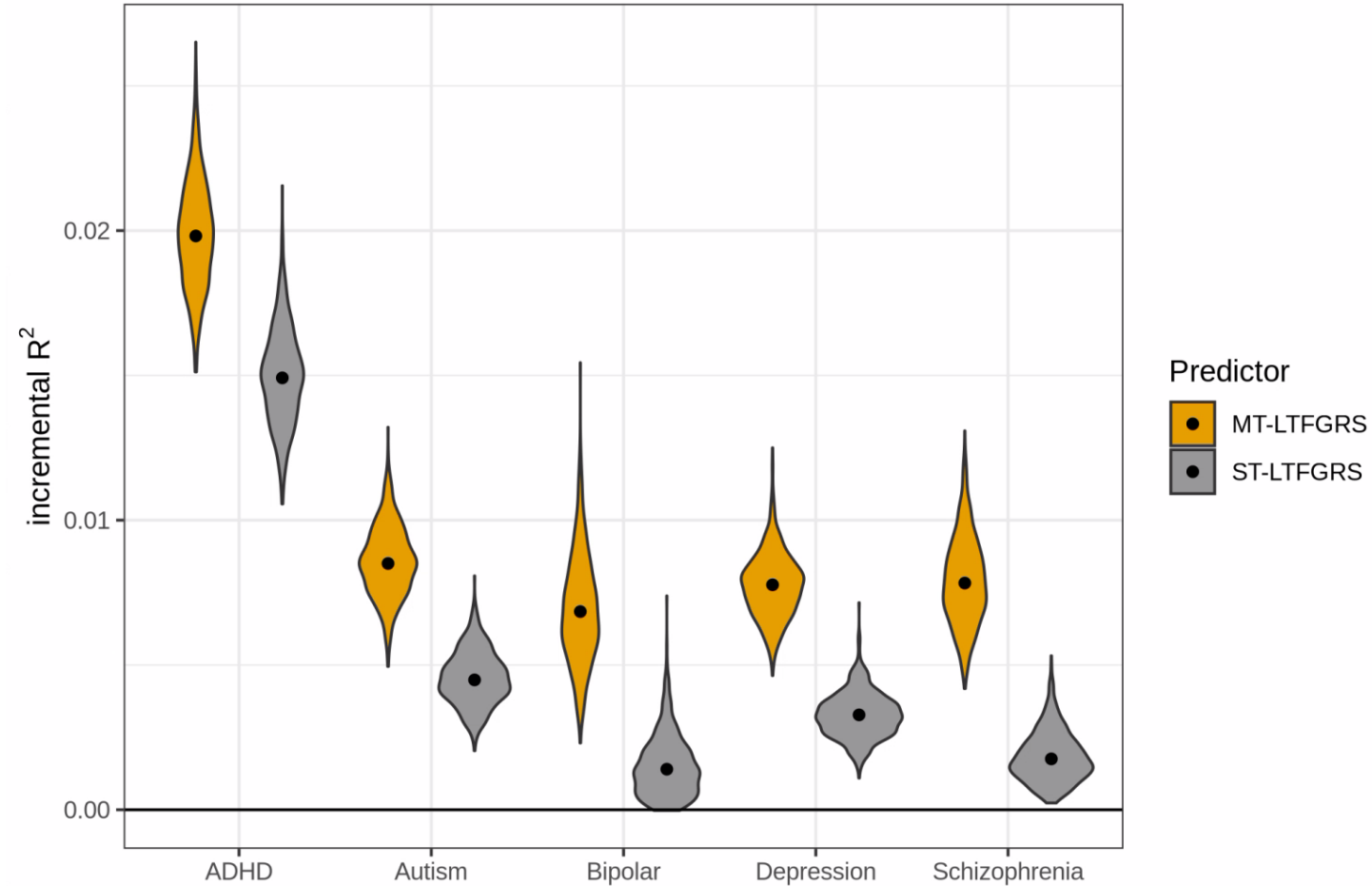
For all iPSYCH outcomes:

Target outcome $\sim$ base covariates + ADHD-LTFGRS

R^2 measures how much variance can be explained on the liability scale

Incremental R^2 measure improvement above base covariates.

ADHD-focused MT-LTFGRS in iPSYCH



CONCLUSION

Developed a Multi-trait extension of the single-trait model

Showed that

- i) power in GWAS can be increased,
- ii) prediction is improved
- iii) some potentially issues exist

Ascertainment is currently not considered in simulations, but are present in real data



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

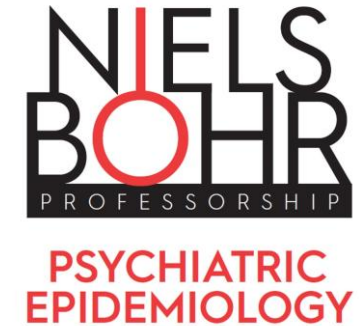
EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy



ACKNOWLEDGEMENTS



Aarhus University

- Bjarni Vilhjálmsson
- Florian Privé
- Esben Agerbo
- Jette Steinbach
- Oleguer Plana-Ripoll
- Clara Albiñana
- Zhihong Zhu
- Ole Hansen
- Anders Jespersen
- Menno Witteween

Aarhus University

- John McGrath
- Preben B Mortensen
- Liselotte Petersen
- Zeynep Yilmaz
- Carsten B Pedersen
- Doug Speed
- Anders Børglum
- Jakob Grove
- Katherine Musliner
- Asger Hobolth

The Klarman Family Foundation Anorexia Nervosa Genetics Initiative

University of Queensland

- Naomi Wray

UCLA

- Noah Zaitlen

IPB Roskilde

- Andrew Schork
- Morten Dybdahl Krebs
- Thomas Werge

FinnGen

- Mark Daly
- Andrea Ganna

github.com/EmilMiP/LTFGRS



NATIONAL CENTRE FOR REGISTER-BASED RESEARCH

DEPARTMENT OF PUBLIC HEALTH

DSTS
5. MAY 2026

EMIL MICHAEL PEDERSEN
POSTDOC



Danish
Data Science
Academy

