Reading note: Polygenic prediction of treatment efficacy with causal transfer learning

2025-10-29

Jiacheng Miao, Jin Mu, Xiaoyu Yang, Jason M. Fletcher, Lauren L. Schmitz, Qiongshi Lu. Polygenic prediction of treatment efficacy with causal transfer learning. medRxiv 2025.10.15.25338051; doi: https://doi.org/10.1101/2025.10.15.25338051

Highlights

- Goal: identifying genetic factors that explain heterogeneous treatment effect (HTE).
- Propose a principled statistical framework named **M-Learner** to **identify genetically-driven HTE**.

  My questions:

  1. How RCT data are applied to fine-tune the polygenic efficacy score (PES)? Especially in M-Learner-S?
  2. PES is learned from K pretrained polygenic risk scores (PRSs), where each PRS corresponds to a trait/disease (?). Do we need to choose the traits/diseases such that they are closely related to the response Y_i?
  3. Figure 2 that the correlation between the PRS and true PES is equivalent to the correlation between β_{Gj} and β_{Ij}? Or in other words, why the main term in the GxT model represents PRS?
  4. The proposed method relies on an assumption that the treatment T in the RCT is randomly assigned conditional on $\mathbf{G}_i$ "such that the propensity score of T given G is known and bounded away from 0 and 1". Does it make sense?

Introduction

- Heterogeneous treatment effect (HTE): variation in the effect of a treatment across individuals or subpopulations.

- Detecting HTE remains challenging due to limited sample size in randomized controlled trials (RCTs), often-missing baseline information, and suboptimal statistical methods with limited power.
- Precision medicine. RCTs often have limited sample size and are primarily designed to estimate ATEs rather than HTE. Besides, variables driving the heterogeneity in treatment response are unknown prior to the trial and can be high-dimensional.
- Two primary methods to estimate HTE: subgroup analysis & risk score analysis.
  - Subgroup analysis: stratify samples based on characteristics, followed by statistical comparisons of treatment effects across these groups. This approach is often statistically underpowered due to the limited size of stratified samples.
  - Risk score analysis: combines multiple variables to predict the trial outcome, followed by assessing how the treatment effects change across risk levels defined by this score. Major limitation: risk score developed without using any information on the treatment is not guaranteed to correlate with treatment effectiveness.
- Two key advances to the current practice of HTE inference:
  - Leverages genome-wide genetic variation to predict HTE, "polygenic efficacy".
  - Introduces **M-Learner to quantify genetic influences on HTE**.

Methods

1. Statistical formulation of polygenic efficacy.

Polygenic genotype-by-treatment interaction (GxT) model for individual i

$$Y_i = \sum_{j=1}^{J} G_{ij}\beta_{Gj} + T_i\beta_T + T_i\left(\sum_{j=1}^{J} G_{ij}\beta_{Ij}\right) + \epsilon_i.$$

- Y_i: treatment outcome (e.g. a measure of treatment efficacy or adverse effect);
- G_{ij}: standardized genotype at variant j (mean 0 and variance 1);
- $T_i \in \left\{\sqrt{\frac{1-p}{p}}, \sqrt{\frac{p}{1-p}}\right\}$: standardized treatment indicator with treatment probability p;
- β_{Gj}: additive effect for variant j;
- β_{Ij}: interaction effect for variant j.

Polygenic efficacy score (PES)

- Polygenic efficacy: the aggregated contribution of genome-wide genetic variation to treatment response.
- PES: weighted sum of allele counts with weights given by their GxT effects:

$$\text{PES}_i = \sum_{j=1}^{J} G_{ij}\beta_{Ij}.$$

Connection between conditional average treatment effect (CATE) and PES

Under the polygenic GxT model:

$$\text{CATE}_i = \text{ATE} + C \times \text{PES}_i,$$

- $C := \left(\sqrt{\frac{1-p}{p}} + \sqrt{\frac{p}{1-p}}\right)$: a positive constant related to the probability p of being assigned to treatment group in the trial.
- $\text{ATE} := C\beta_T$: average treatment effect (over the whole population).
- PES explains all the variability in treatment response that is attributable to genetic variation.
- Estimating genetically-driven HTE = estimating the PES.
- Estimating PES in RCT is challenging because of limited sample size and the typically small GxT effect at the SNP level.

Derivation for the connection between PES and CATE

Under the proposed polygenic GxT model, the CATE can be denoted as

$$\begin{aligned}\text{CATE}_i := \text{CATE}(\mathbf{G}_i) &= \mathbb{E}\left[Y_i \mid \mathbf{G}_i, T_i = \sqrt{\frac{1-p}{p}}\right] - \mathbb{E}\left[Y_i \mid \mathbf{G}_i, T_i = -\sqrt{\frac{p}{1-p}}\right] \\ &= \left(\sqrt{\frac{1-p}{p}} + \sqrt{\frac{p}{1-p}}\right)\beta_T + \left(\sqrt{\frac{1-p}{p}} + \sqrt{\frac{p}{1-p}}\right)\sum_{j=1}^{J} G_{ij}\beta_{Ij}.\end{aligned}$$

Under the proposed polygenic GxT model above, the **ATE** in an RCT can be denoted as

$$\text{ATE} = \frac{1}{n}\sum_{i=1}^{n}\text{CATE}(\mathbf{G}_i) = \mathbb{E}\left[Y_i \mid T_i = \sqrt{\frac{1-p}{p}}\right] - \mathbb{E}\left[Y_i \mid T_i = -\sqrt{\frac{p}{1-p}}\right]$$

$$= \left(\sqrt{\frac{1-p}{p}} + \sqrt{\frac{p}{1-p}} \right) \beta_T.$$

The last equation hold because G_{ij} is the standardized genotype at variant j (mean 0, variance 1), that is, $\sum_i G_{ij} = 0$ for all $j = 1, \ldots, J$. Thus, $\text{CATE}_i = \text{ATE} + C \times \text{PES}_i$.

Summary

- **Individual-level causal concept**: polygenic efficacy score (**PES**) is an individualized score that equivalently accounts for CATE, a measure for heterogeneous treatment effect, for individual i.
- **Causal transfer learning from PRS + RCT data to PES**: this score can be estimated from PRS + treatment assignments and outcomes in the RCT using causal machine learning methods.

S1. Supplementary 1. Polygenic risk score (PRS): a statistical genetics concept

Thanks to ChatGPT…

- Most complex traits, such as height, blood pressure, diabetes risk, or schizophrenia, are polygenic, meaning they are **influenced by many genes**, each contributing a **small effect**.
- A **polygenic risk score (PRS)** summarizes these small effects into a single number - representing the **overall genetic liability** of an individual to that **trait**.
- Estimated from genotype-trait association model $Y \sim \xi_1 G_1 + \xi_2 G_2 + \cdots + \xi_J G_J$.

Formula: a weighted sum of risk alleles:

$$\text{PRS}_i = \sum_{j=1}^{J} \xi_j G_{ij}$$

- G_{ij}: of individual i, the genotype at variant j (coded as 0, 1, or 2 for the number of risk alleles).
- β_j: effect size (weight) of SNP j from a GWAS (often a regression coefficient or log-odds ratio)

How it is constructed:

1. Obtain GWAS summary statistics: specifically, each SNP's estimated effect size $\widehat{\beta}_j$ and its p-value.

2. Select SNPs: filtered out redundant SNPs by p-values or adjusting for LD.
3. Compute PRS for each individual: by using their genotypes and the selected weights $\hat{\beta}_j$.
4. Evaluate performance: by correlating the PRS with the observed trait or disease outcome in an independent dataset (out-of-sample evaluation).

Interpretation:

- Higher PRS -> greater genetic predisposition (risk) to the trait/disease.
- Predictive power depends on: heritability of the trait, GWAS sample size, and similarity between training and target population, etc.

Summary

- **PRS ~ genetic characteristics related to a trait**: each PRS can be regarded as a score indicating aggregated genetic contibutions of a certain complex trait/disease.
- PRS is more of a concept with no implication of causality.
- PRS model is pretrained from external GWAS datasets.
- How to choose PRS models? What traits to choose?

2. Overview of M-Learner

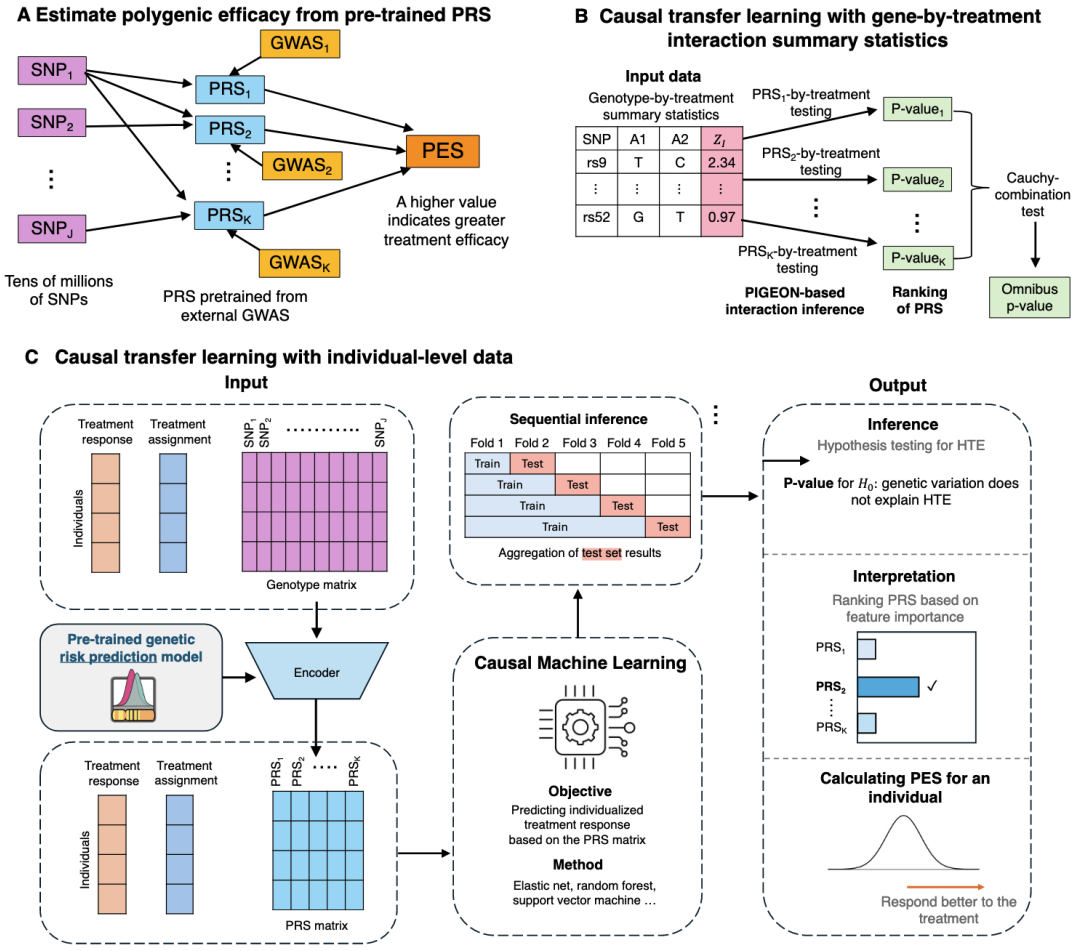

Figure 1: **M-Learner framework for modeling polygenic efficacy and HTE. (A) Genetic representation of polygenic efficacy.** Instead of relying on candidate variants or a single PRS for the trial outcome, M-Learner uses many PRS derived from external GWAS to capture the complex genetic mechanisms underlying HTE. **(B) M-Learner based on summary statistics (M-Learner-S).** Using GxT GWIS summary statistics, M-Learner tests many PRS' interactions with the treatment using PIGEON and integrates them with the Cauchy combination test. The output includes a p-value for testing the HTE and a ranked list of PRS that are informative on the HTE. **(C) M-Learner based on individual-level data (M-Learner-I).** M-Learner applies causal transfer learning to fine-tune pre-trained PRS models for individualized treatment prediction. Sequential inference ensures valid statistical testing. The outputs include a P-value for detecting HTE, a ranked list of PRS contributing to HTE, and individualized predictions of treatment response.

- Two key ideas:
  - M-learner models polygenic efficacy as a function of latent genetic representations

derived from high-dimensional PRS variables (Figure 1A);

- it employs a **causal transfer learning** framework that fine-tunes pretrained PRS models with RCT data to predict treatment response.

Notations:

- $(Y_i(1), Y_i(0))$: potential outcomes under treatment and control for i-th individual.
- $T_i \in \{0, 1\}$: binary treatment assignment.
- $\mathbf{G}_i = (G_{i1}, \ldots, G_{iJ})$: J SNPs.
- $\widehat{\mathbf{PRS}}_i = \left(\widehat{\mathrm{PRS}}_{1i}, \ldots, \widehat{\mathrm{PRS}}_{Ki}\right) \in \mathbb{R}^K$: K pre-trained PRS compressing the information of J SNPs:
  - Each $\widehat{\mathrm{PRS}}_{ki}$ summarizes the polygenic genetic predisposition for a k-th trait/disease. [each refers to the PRS for a trait/disease. multiple traits!]
  - It can be constructed from external GWAS summary statistics.
- **(???!!!)** Assume the treatment is randomly assigned conditional on $\mathbf{G}_i$, such that the propensity score $e(\mathbf{G}_i) = P(T_i = 1 \mid \mathbf{G}_i)$ is known and bounded away from 0 and 1.

M-Learner-I Framework

- Leverages individual-level genotype and treatment data.
- CATE: $\tau(\mathbf{G}_i) = \mathbb{E}[Y_i(1) - Y_i(0) \mid \mathbf{G}_i]$, estimated using causal ML models with a 2-stage procedure:
  1. Fit a predictive model using $\widehat{\mathbf{PRS}}_i$ as input to generate individual-level CATE estimates $\hat{\tau}(\mathbf{G}_i)$.
  2. Evaluate the predictiveness of these estimates using out-of-fold inference; to mitigate overfitting and yields valid out-of-sample estimates of heterogeneity.
- **Input**:
  - Dataset $\{(\mathbf{G}_i, T_i, Y_i)\}_{i=1}^n$;
  - Propensity model $e(\cdot)$ that measures the dependency of genotype $\mathbf{G}$ on treatment assignment T.
  - Number of folds K;
  - A machine learning Learner $\mathcal{L}$.

- **Output:**
  - P-value of testing the null hypothesis H_0: treatment effect is homogeneous across all genetic variation (H_a: heterogeneous treatment effect (HTE) is explained by genetic variation.);
  - The average treatment in each subgroups by predicted PES;
  - PRS feature importance scores.

Step 1: Calculate PRS from SNP data to obtain $\widehat{\mathbf{PRS}}_i \in \mathbb{R}^K$ for each individual.

Step 2: Individual pseudo-outcomes.

Introduce a pesudo-outcome (estimated outcome via inverse probability weighting), an unbiased proxy for the CATE function (?):

$$Y_i^H = \frac{T_i Y_i}{e(\mathbf{G}_i)} - \frac{(1-T_i)Y_i}{1-e(\mathbf{G}_i)}.$$

- It satisfies that $\mathbb{E}\left[Y_i^H \mid \mathbf{G}_i\right] = \tau(\mathbf{G}_i)$ under the consistency assumption.

Step 3: Sequential cross-fitting over a K-fold partition to estimate the individual CATE $\tau(\mathbf{G}_i)$.

Estimate predictive models using sequential cross-fitting (rather than conventional K-fold cross-fitting) "to ensure each prediction is strictly out-of-sample while pooling information across folds with statistical rigor":

1. The data are randomly partitioned into K ordered folds $\{\mathcal{I}_1, \ldots, \mathcal{I}_K\}$;
2. *Train on past folds*: fit the ML model on $\{(\widehat{\mathbf{PRS}}_i, Y_{ji}^H) : j \in \mathcal{I}_1 \cup \cdots \cup \mathcal{I}_{k-1}\}$ to obtain $f_{\hat{\theta}}^{(1:(k-1))}$.
3. *Predict on current fold*: for each $i \in \mathcal{I}_k$, set $\hat{\tau}(\mathbf{G}_i) \leftarrow f_{\hat{\theta}}^{(1:(k-1))}(\mathrm{PRS}_i)$.

Step 4: Assess heterogeneity by the best linear prediction.

To quantify predictiveness and correct for potential bias, perform a **best linear projection (BLP)** of Y_i^H onto $\hat{\tau}(\mathbf{G})$:

$$Y_i^H = \alpha + \beta \hat{\tau}(\mathbf{G}_i) + \epsilon_i,$$

- α: absorb the overall ATE component,
- β: captures how much variation in the pseudo-outcomes is explained by the predicted CATE.
  - $\beta \approx 1$: $\hat{\tau}(\mathbf{G}_i)$ perfectly recovers the true CATE;

- $\beta \approx 0$: $\hat{\tau}(\mathbf{G}_i)$ contains only noise;

Test $H_0 : \beta = 0$, i.e., genetic variation does not explain HTE. Rejection of the null provides statistical evidence for HTE explained by genetic variation.

Step 5: Treatment effect stratification.

To examine how treatment effects differ across genetically defined risk profiles.

- Sort individuals by $\hat{\tau}(\mathbf{G}_i)$ and divide into quantile bins $\{Q_q\}$;
- Compute ATEs in each subgroup:
  - $$\tau_q = \mathbb{E}[Y_i(1) - Y_i(0) | \hat{\tau}(\mathbf{G}_i) \in Q_q].$$
- Increasing (or decreasing) τ_q across quantiles indicates a stronger (or weaker) benefit among individuals with higher predicted responses.

Step 6: Interpretation - PRS importance.

Investigate how each PRS feature individually relates to HTE.

- For each PRS_k, perform single-variable regressions after obtaining $\hat{\tau}(\mathbf{G}_i)$:
  - $\hat{\tau}(\mathbf{G}_i) = \alpha_{0k} + \alpha_k \widehat{\text{PRS}}_{ki} + \eta_{ki}$
  - $\hat{\alpha}_k$: capture marginal association between $\widehat{\text{PRS}}_{ki}$ and the predicted treatment effect $\hat{\tau}(\mathbf{G}_i)$.
  - η_{ki}: random error of individual i corresponde to the k-th trait/disease.
- **PRS importance score**: the absolute t-statistic from the regression
  - $$\text{FeatureImportance}\left(\widehat{\text{PRS}}_{ki}\right) = \frac{|\alpha_k|}{\text{SE}(\alpha_k)}.$$
  - Larger values of $\text{FeatureImportance}\left(\widehat{\text{PRS}}_{ki}\right)$ indicate stronger evidence that $\widehat{\text{PRS}}_{ki}$ explains HTE.

M-Learner-S Framework

- A **PIGEON**-based framework; **PIGEON offers causal inference.**
- Uses GxT summary statistics from **genome-wide interaction studies (GWIS)** in settings with only privacy-preserved summary-level data (Figure 1B).
- First employs a polygenic interaction inference approach to conduct PRS-by-treatment interactions from summary statistics (i.e., interaction effect of PRS and treatment!).
- The resulting interaction p-values can identify PRS that are informative on the HTE.
- Results across multiple PRS are then combined using the Cauchy combination test to yield a global p-value.

Algorithm 2: M-Learner-S Algorithm using GWIS summary statistics as input

---

Input: Z-score from GWIS summary statistics $\texttt{GWIS} = \{Z_{1j} \text{ for } j = 1, \cdots, J\}$, Z-score from $k = 1, \cdots, K$ GWAS summary statistics $\texttt{GWAS}_k = \{Z_{kj} \text{ for } j = 1, \cdots, J\}$. Here, J is the total number of SNPs

Output: P-value for testing the null hypothesis H_0: treatment effect does not differ from genetic variation. A ranked list of PRS (GWAS) that explains the HTE.

Step 1: Apply PIGEON to each GWAS summary staistics $(k = 1, \cdots, K)$ **to compute the PRS×T interaction P-value:**

$$\text{P-value}_k = \texttt{PIGEON}(\texttt{GWIS}, \texttt{GWAS}_k) \text{ for } k = 1, \cdots, K$$

Step 2: Apply Cauchy-combination test to obtain the aggregated p-value
Define the Cauchy combination statistic

$$T_K = \sum_{k=1}^{K} \frac{1}{K} \tan\left[\left(\frac{1}{2} - \text{P-value}_k\right)\pi\right].$$

Under the global null (no HTE), the combined P-value is

$$\text{P-value}_{\text{CCT}} = \frac{1}{2} - \frac{1}{\pi}\arctan\left(T_K\right).$$

Step 3: Ranking the PRS based on the P-value
Let $\text{P-value}_1, \text{P-value}_2, \ldots, \text{P-value}_K$ be the set of PRSxT interaction P-values obtained from Step 1. Define the ordering function q such that

$$\text{P-value}_{q(1)} \leq \text{P-value}_{q(2)} \leq \cdots \leq \text{P-value}_{q(K)}$$

Then the ranked list of PRS corresponds to

$$\text{GWAS}_q(1), \text{GWAS}_q(2), \ldots, \text{GWAS}_{q(K)},$$

where $\text{GWAS}_q(1)$ is the top-ranked GWAS (smallest P-value) and GWAS $q(K)$ is the lowest-ranked.

S2. Supplementary 2

PIGEON (PolygenIc Gene-Environment interactiON)

Miao J, Song G, Wu Y, Hu J, Wu Y, Basu S, Andrews JS, Schaumberg K, Fletcher JM, Schmitz LL, Lu Q. PIGEON: a statistical framework for estimating gene-environment interaction for polygenic traits. Nat Hum Behav. 2025 Aug;9(8):1654-1668. doi: 10.1038/s41562-025-02202-9. Epub 2025 May 23. PMID: 40410536; PMCID: PMC12496094.

- A statistical framework (& software tool) for **quantifying polygenic gene-environment interaction (GxE)** and interaction involving polygenic scores (PGS) more broadly.
- Requires **summary statistics (GWAS and "GWIS")** rather than full individual data.
- Two main objectives:
  1. Estimate the variance component of GxE, i.e., proportion of phenotypic variance explained by gene-environment interaction;
  2. Estimate the "PGS x E" effect, i.e., how an individual's polygenic component interacts with environment or treatment, namely "oracle PGS x E" in the paper.
- Regarding causal inference or treatment effect heterogeneity, PIGEON provides a path to assess heterogeneity in effect by genetic liability, i.e., interaction of polygenic risk and environment/treatment.

Cauchy combination statistics (Cauchy combination test, CCT)

- A p-value combination method to combine multiple correlated p-values into a single global test statistic.
- Suppose we have K hypothesis tests with p-values $p_1, p_2, \cdots, p_K$.
- Test the global null hypothesis H_0 : all p_k correspond to null effects.
- Key idea: transform each p-value using **Cauchy quantile function**, then take a (weighted) average:
  - $$T = \sum_{k=1}^{K} w_k \tan\left[(0.5 - p_k)\pi\right],$$
  - Under the global null H_0, $T \approx \text{Cauchy}(0, 1)$ even if p_k are (mildly) dependent.
  - Thus the combined p-value is $p_{\text{CCT}} = 0.5 - \frac{1}{\pi}\arctan(T)$.
- Why it works (intuition):

  - The Cauchy distribution has very heavy tails, so the combined statistic is dominated by a few small p-values;
  - The Cauchy transformation is robust to correlations because the extreme tail events remain rare even under dependence.

---

3. Scaling laws for HTE estimation with genetic information

- Three strategies for **estimating PES**:
  - Within-sample GWIS approach:
    - Estimate SNP-by-treatment interaction coefficients directly within the RCT and use them as SNP weights for the PES;
    - The estimates are extremely noisy because of limited sample size in the RCT, resulting in lower power.
  - Single-outcome PRS approach:
    - Use a pre-trained PRS for the trial outcome to assess the interaction with the treatment.
    - Its performance depends on the GWAS sample size which determines measurement errors in PRS, and the correlation between this PRS and the true PES, which is equivalent to the correlation between β_{Gj} and β_{Ij}. **(WHY?! Relate to PIGEON?!)**
  - Multi-PRS ensemble approach:
    - Leverage a high-dimensional PRS matrix pretrained from many GWAS, fine-tuned in the RCT sample, to estimate a function that best predicts treatment response.
    - Recommended as it provides improved power to detect genetically-driven HTE.
- Assertion in the paper: Causal transfer learning is able to boosts the power of M-learner when the correlation between the true and estimated PES is at least moderate (i.e. >= 0.5). That is, M-learner achieves high power by aggregating information from a large number of pre-trained PRS even if the true GxT-explained variance is small (approx 3%).

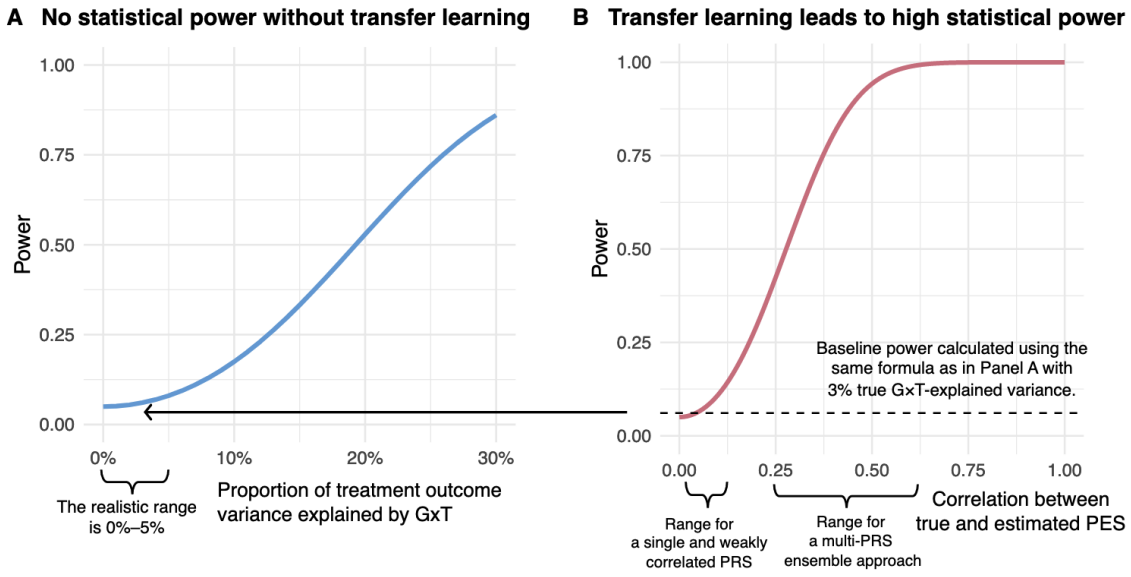

Figure 2: **Power scaling laws for detecting genetically-driven HTE. (A)** For analyses without transfer learning, the graph plots statistical power (y-axis) against the proportion of treatment response variance explained by gene-by-treatment interaction (x-axis). At the realistic range of this value(i.e., 0–5%), statistical power for detecting HTE is negligible. **(B)** With transfer learning, the graph plots power (y-axis) against the correlation between the true and estimated PES (x-axis). When the correlation reaches 0.5, we achieve >90% power. M-Learner achieves high power by aggregating information from a large number of pre-trained PRS.

Derivation for the scaling laws for HTE estimation with genetic variations

- To mathematically show the multi-PRS ensemble strategy yields the most powerful method for the null hypothesis $H_0 : \beta = 0$...

Based on the polygenic GxT model, suppose we have N_{test} testing samples and let S_i be the **estimated PES**:

$$Y_i \sim S_i\alpha_S + T_i\alpha_T + S_iT_i\alpha_I + \delta_i.$$

To evaluate the estimated GxT effect size magnitude and statistical power of hypothesis testing of $H_0 : \alpha_I = 0$, we have (?!)

$$\alpha_I = \frac{\mathrm{Cov}(Y_i, S_iT_i)}{\mathrm{Var}(S_iT_i)} = \frac{\mathrm{Cov}(\mathrm{PES}_i\, T_i, S_iT_i)}{\mathrm{Var}(S_iT_i)} = \ldots$$

- Validate of this equation?
- Skeptical about the derivation...

Standardization, specifically the centering step, of $\mathbf{G}_i$ and T_i reduces the correlation between the main term and the interaction term, but not eliminate it.