Quantifying Uncertainty in Polygenic Risk Scores Using Conformalized Quantile Regression

Chen Wang[1,2], Fan Wang[1], Malgorzata Bogdan[3], Marco Masala[4], Edoardo Fiorillo[4], Marcella Devoto[4], Francesco Cucca[4,5], Dan Belsky[6], Iuliana Ionita-Laza[1,7,#]

[1]Department of Biostatistics, Columbia University, New York, USA
[2]Division of Nephrology, Department of Medicine, Columbia University, New York, USA
[3]Department of Mathematics, University of Wrocław, Poland
[4]Institute for Genetic and Biomedical Research, National Research Council, Italy
[5]Department of Biomedical Sciences, University of Sassari, Sassari, Italy
[6]Department of Epidemiology and Butler Columbia Aging Center, Columbia University, New York, USA
[7]Department of Statistics, Lund University, Lund, Sweden
[#]ii2135@columbia.edu

Summary

Polygenic risk scores (PRS) are widely used in post-GWAS analyses to predict complex traits across humans, animals, and plants. While significant progress has been made in developing new PRS methods, much less attention has been given to quantifying the uncertainty associated with these predictions. In this work, we explore a method for uncertainty quantification based on quantile regression. When paired with conformal prediction, this approach enables the construction of prediction intervals with guaranteed coverage, offering lower and upper bounds within which the phenotype is likely to fall with high probability. We apply this framework to data from the UK Biobank and the ProgeNIA/SardiNIA studies, showing that the resulting prediction intervals: (1) exhibit reduced uncertainty compared to standard methods, (2) maintain valid coverage under minimal model assumptions, and (3) provide more realistic estimates by allowing for asymmetry and individual-specific interval lengths. Overall, we present a novel framework for uncertainty quantification in PRS analyses and highlight the importance of incorporating uncertainty into predictive modeling.

Introduction

Polygenic risk score (PRS) prediction is an important post-GWAS analysis whose goal is to predict an individual's genetic risk for a particular trait or disease[1]. Unlike traditional risk factors, PRS can identify individuals at high (or low) risk from birth and therefore such individuals can, in principle, be prioritized (or not) for early lifestyle interventions. Despite tremendous work in this area, PRS are not yet ready to be deployed in the clinic[2]. One major issue is the limited prediction ability of such scores, which may be improved by incorporating additional information from biomarkers and risk factors for the disease of interest[3]. Another major limitation is that prediction uncertainty is not usually accounted for which affects the performance of these scores in clinical risk stratification[4]. Uncertainty in PRS prediction is inherent due to uncertainty in SNP effect size

estimates, possibly inaccurate modeling assumptions (e.g. linear additive model), unmodelled factors such as environment and gene-by-environment interactions, population structure, and chance. Therefore, the point estimate normally reported is insufficient and does not convey the underlying variability. Reporting and accounting for uncertainty is essential as that can affect medical decision making, while providing more transparency about the limitations of PRS for prediction purposes.

We investigate here the problem of uncertainty quantification for PRS from a novel perspective based on quantile regression that is a leading methodology for constructing prediction intervals and provides important insights beyond what can be learned from linear regression models[5–7]. In the usual PRS prediction task the aim is to predict the mean of a phenotype (Y) given the genotype profile of an individual (X) assuming a linear model. The resulting PRS is a linear combination of individual genotypes weighted by estimated genetic effects from a linear regression (LR) model in GWAS. This is essentially a point prediction, meant to reflect the typical or average phenotypic outcome given a particular genotype profile, i.e. $E(Y|X)$. A natural alternative to linear regression for quantitative traits is based on quantile regression (QR) that extends mean-based LR to the analysis of conditional quantiles of phenotype Y given genotypes X[7], $Q_Y(\tau|X)$ for any quantile level $\tau \in (0,1)$; when $\tau = 0.5$, $Q_Y(\tau|X)$ corresponds to the conditional median of Y given genotypes X. QR consists of fitting a series of linear regression models for any quantile level $\tau \in (0,1)$, providing estimates of quantile-specific regression coefficients. QR is an overall well-powered and flexible approach that allows detection of genetic variants with both additive and more complex non-additive effects[7,8]. While variants with additive effects are well detected by LR, those with more complex effects tend to show heterogeneity across quantile levels and are better detected by QR techniques[7]. Importantly, QR allows us to move away from point predictions to prediction intervals that naturally account for the uncertainty in the point estimation (**Fig. 1(a)**). We further combine the QR methodology with conformal prediction[9], a machine learning technique to build prediction intervals that helps improve coverage of the regular QR prediction intervals. Unlike existing methods for uncertainty quantification which focus on uncertainty around the predicted phenotypic mean[4], the QR-based methods account for variability across the entire phenotype distribution. We refer to this new framework as QR-PRS.

We illustrate QR-PRS in applications to 64 quantitative traits with a detailed analysis for C-reactive protein, using the UK Biobank (UKBB) data[10]. We validate the results in applications to an independent study, ProgeNIA/SardiNIA, a longitudinal study of a cohort of Sardinian subjects. The resulting prediction intervals, while revealing high variability around the PRS estimates in general, have appealing features when compared to LR-based intervals, including lower magnitude of uncertainty in risk stratification, asymmetry and varying lengths across individuals. Beyond the current implementation and application, the proposed framework can serve as a foundational framework for future development of more complex models for uncertainty quantification.

Methods

QR-PRS construction and prediction intervals

QR-PRS is constructed using the pruning + thresholding (P+T) method[11,12]. In what follows we assume we have separate training/validation/test datasets, with specific details on these datasets being provided during specific applications.

1) *Pruning*: Pruning is performed to reduce correlation among variants based on linkage disequilibrium (LD) estimated from the training data with a threshold of $r^2 < 0.1$.
2) *Thresholding*: A significance threshold is applied to filter out variants with potentially no effect on the trait. A simple threshold of 0.05 is applied to the quantile-specific QR p-values[7] (see below on how to obtain these p-values using tools from QR). Note that this thresholding step can select different sets of variants for QR-PRS at different quantile levels, as some variants with heterogeneous effects will only be significant at certain quantile levels.
3) *QR-PRS computation*: QR-PRS for each individual in the test dataset is calculated as the sum of the weighted effect alleles across M SNPs selected by pruning and thresholding above. The quantile-specific effect sizes estimated from the training data are used as weights of the selected variants, reflecting their contribution to the trait. For individual i and M SNPs with indices $(1), \dots, (M)$ selected at quantile level τ we compute $\text{QR-PRS}_\tau(i) = \sum_{j=1}^{M} X_{i(j)} \hat{\beta}_j(\tau)$. Note that we want to ensure quantile non-crossing, i.e. that the predicted quantiles $\text{QR-PRS}_\tau(i)$ at quantile levels $\tau < \tau'$ are monotonic. We use here a common post-hoc adjustment based on sorting the predicted quantiles at a fine set of quantiles that ensures monotonicity[13,14].

To estimate quantile-specific coefficients $\beta(\tau)$ used in steps 2 and 3 above, we fit linear QR models on the training data. Specifically, for the jth variant and specific quantile level $\tau \in (0,1)$ we fit the following conditional QR model:

$$Q_Y(\tau \mid X_j, C) = X_j \beta_j(\tau) + C\alpha(\tau),$$

where C corresponds to covariates we want to adjust for; $\beta_j(\tau)$ and $\alpha(\tau)$ are quantile-specific coefficients and can be estimated by minimizing the pinball loss function[15]. This optimization problem can be solved efficiently using existing tools. For testing, we employ a commonly used hypothesis testing tool in QR, the rank score test[16,17] to test the null hypothesis $H_0: \beta_j(\tau) = 0$ and obtain a p-value for each SNP at any given quantile level τ.

Plug-in prediction intervals. By fitting these models, we will be able to predict low and high phenotypic quantiles and build prediction intervals for each individual i of the form $\left(\text{QR-PRS}_{\alpha_{lo}}(i), \text{QR-PRS}_{\alpha_{hi}}(i)\right)$. We refer to them as plug-in prediction intervals in contrast to conformal prediction intervals we discuss next.

Conformal prediction intervals. One potential issue is that plug-in QR prediction intervals are not guaranteed correct coverage due to various factors such as model misspecification, imprecision in estimation of conditional quantiles, and unmodelled important factors such as environment and gene-by-environment interactions. Conformal prediction is a machine learning technique to build prediction intervals that may account for unmodelled randomness in the phenotype[9]. Combining conformal prediction with QR is a promising approach as it can lead to more realistic

prediction intervals that are adaptive to heteroskedasticity in the data, while providing finite sample coverage guarantees[5]. We therefore build conformal versions of the prediction intervals above and refer to them as $(\text{CQR-PRS}_{\alpha_{lo}}(i), \text{CQR-PRS}_{\alpha_{hi}}(i))$.

Specifically, we fit QR models on the training data for some upper and lower quantile values α_{lo} and α_{hi}. We then compute conformity scores on a separate calibration set $E_i = max\left(\text{QR-PRS}_{\alpha_{lo}}(i) - Y_i, Y_i - \text{QR-PRS}_{\alpha_{hi}}(i)\right)$ for each individual i in the calibration set. Given a new individual $n+1$ we construct the conformalized prediction interval for Y_{n+1} as $\left(\text{QR-PRS}_{\alpha_{lo}}(n+1) - Q_{1-\alpha}(E), \text{QR-PRS}_{\alpha_{hi}}(n+1) + Q_{1-\alpha}(E)\right)$, where $1-\alpha$ is the desired coverage level and $Q_{1-\alpha}(E)$ is the $(1-\alpha)th$ empirical quantile of conformity scores E. This construction leads to prediction intervals with marginal coverage guarantee[9]. We refer to these prediction intervals as conformal prediction intervals (CQR).

Computation of $\boldsymbol{P(Y > t|X)}$

Given the ability to estimate conditional quantiles $Q_Y(\tau|X)$ at different quantile levels τ as described above, we can then compute $P(Y > t|X) = 1 - F(t|X)$ for any threshold of interest t, where $F(Y|X)$ is the cumulative distribution function (CDF). Specifically, the conditional quantile function is related to the CDF by the following formula: $Q_Y(\tau|X) = \inf\{y: F(y|X) \geq \tau\}$, or equivalently $F(t|X) = Q_Y^{-1}(t|X)$. Therefore, given $Q_Y(\tau|X)$ for a grid of quantile levels $\tau \in (0,1)$, we can use a smooth interpolation method such as b-splines[18] to approximate the CDF.

LR-PRS and credible interval construction

Uncertainty quantification for PRS based on LR (LR-PRS) is computed following the approach described in Ding et al.[4] based on LDpred2[19], a Bayesian method to estimate the posterior effect size distribution of each SNP by accounting for LD. Specifically, we first ran LDpred2 per autosomal chromosome with a grid of different values for several parameters, including heritability (0.7, 1.0, and 1.4 times the heritability estimated with LD score regression), polygenicity (evenly spaced on a logarithmic scale between 10^{-4} and 1), and sparsity (1 or 0 indicating whether or not a sparse model is applied) using a training dataset, as in Ding et al.[4]. The model (parameter set) with the highest R^2 between the phenotype and the genetic prediction in a separate validation dataset was selected as the best PRS model to estimate the LR-PRS. For the optimal parameter setting, we ran LDpred2 a second time on the training dataset, with the Markov Chain Monte Carlo (MCMC) sampler with 100 burn-in iterations to produce B=500 posterior samples of effect sizes. Then, for each individual in the test dataset we computed a PRS estimate from each posterior sample to approximate the posterior distribution of PRS for that individual and empirically computed a ρ-level credible interval.

Simulation settings

We evaluate coverage for the QR prediction intervals (both plug-in and conformalized) using simulations under two settings as described next.

Homogenous settings. We utilized unimputed genotype data from the UKBB white British samples. For quality control, we excluded individuals with more than 10% missing genotypes, included only SNPs with a minor allele frequency (MAF) greater than or equal to 0.05 and

genotyping rate of at least 90%, and filtered out SNPs with a Hardy-Weinberg equilibrium p-value less than 10^{-4}. We selected 300,000 independent individuals, and after quality control 348,911 SNPs remained. Note that the evaluation of coverage is done on a separate test dataset with 26,377 independent white British individuals.

Given the genotype matrix $\mathbf{X}$, heritability h_g^2, and proportion of causal variants p_{causal}, standardized effects and phenotypes are generated as follows

$$\beta_j \sim \begin{cases} N\left(0, \dfrac{h_g^2}{M p_{\text{causal}}}\right) & \text{with probability } p_{\text{causal}} \\ 0 & \text{with probability } 1 - p_{\text{causal}} \end{cases}.$$

$$(y_1, \dots, y_N)^\top \sim N\left(\mathbf{X}\boldsymbol{\beta}, \left(1 - h_g^2\right)\mathbf{I}_N\right).$$

We simulate data under a range of parameters: $h_g^2 \in \{0.05, 0.1, 0.25, 0.5, 0.8\}$ and $p_{\text{causal}} \in \{0.001, 0.01, 0.1, 1\}$, for a total of 20 simulation settings.

Heterogeneous settings. As above we generate standardized effect sizes and phenotypes as follows:

$$\beta_j \sim \begin{cases} N\left(0, \frac{h_g^2}{M p_{\text{causal}}}\right) & \text{with probability } p_{\text{causal}} \\ 0 & \text{with probability } 1 - p_{\text{causal}} \end{cases}.$$

For $m = m_1 + m_2$ causal variants, we assume that m_1 variants have homogenous effects and m_2 variants have heterogeneous effects, i.e. local effects. For a variant $l = (m_1 + 1) \dots m$ with heterogeneous effects, $b_l(\tau)$ is a local effect defined as

$$b_l(\tau) = \begin{cases} \beta_l \times \tau & \text{if } \tau > 0.7 \\ 0 & \text{otherwise} \end{cases}.$$

Hence $b_l(\tau)$ is non-zero only for quantile level $\tau > 0.7$. We then generate the phenotype Y as follows:

$$Q_Y(\tau \mid X) = \sum_{j=1}^{m_1} X_j \beta_j + \sum_{l=1}^{m_2} X_{m_1+l} b_{m_1+l}(\tau).$$

To generate Y, we use the inverse transform sampling. We simulate β using a range of parameters: $h_g^2 \in \{0.05, 0.1, 0.25, 0.5, 0.8\}$ and $p_{\text{causal}} \in \{0.001, 0.01, 0.1, 1\}$. We vary m_1 and m_2 by considering several proportions of causal variants with heterogeneous effects among all causal variants ({0.01, 0.1, 0.5}).

Results

Simulations

We evaluate the empirical coverage of the constructed QR-PRS prediction intervals, both plug-in $\left(\text{QR-PRS}_\alpha(i), \text{QR-PRS}_{1\text{-}\alpha}(i)\right)$ and conformal $\left(\text{CQR-PRS}_\alpha(i), \text{CQR-PRS}_{1\text{-}\alpha}(i)\right)$, for different values of $\alpha \in \{0.025, 0.1, 0.2, 0.3, 0.4\}$ under the two simulation scenarios described above for

individuals in the independent test data (with 26,377 white British individuals). Specifically, we compute the proportion of individuals whose QR-PRS/CQR-PRS prediction intervals overlap their true phenotype value. The plug-in QR-PRS prediction intervals tend to be miscalibrated (**Fig. S1-S2**), while the conformalized versions (CQR-PRS) are well calibrated across the different settings (**Fig. S3-S4**).

Applications to C-reactive protein in UKBB

We first show results on applications to C-reactive protein (CRP). The training dataset for PRS derivation is the white British cohort (n=325,306 unrelated individuals and ~8.6 million SNPs) in UKBB where the original LR and QR GWAS summary statistics are derived. The test dataset includes the European samples in UKBB that are not used in the training set (n=39,643) for which the PRS performance will be assessed.

We perform QR at different quantile levels $\tau \in (0.025, 0.975)$, adjusting for several covariates including age, sex, batch, and 10 principal components (PC) of genetic variation (**Fig. S5).** The estimated quantile-specific coefficients are used to compute quantile specific PRS (QR-PRS_τ) for every individual in the test data.

In addition to the typical LR-PRS distribution, we now obtain quantile-specific QR-PRS distributions across a grid of quantile levels $\tau \in (0.025,0.975)$ (**Fig. S6**; see **Fig. S11** for CRP in ProgeNIA/SardiNIA). $\text{QR-PRS}_{0.5}$ (corresponding to median regression) is highly correlated with the mean-based LR-PRS (**Fig. S7**; see **Fig. S12** for CRP in ProgeNIA/SardiNIA) as expected. CRP has similar positive but weak correlation levels with LR-PRS and QR-PRS for the central quantiles $\tau \in (0.4,0.6)$. While QR-PRS corresponding to the central quantiles are highly correlated among themselves (and with LR-PRS), correlations become weaker with QR-PRS at more extreme quantiles (e.g. $\tau = 0.025$ or $\tau = 0.975$).

We construct QR prediction intervals $\left(\text{QR-PRS}_\alpha(i), \text{QR-PRS}_{1\text{-}\alpha}(i)\right)$ (and conformalized versions) and use them to perform more detailed risk stratification than possible based on just the single value PRS. For example, if we consider a high-risk threshold t (e.g. the 90th percentile of the $\text{QR-PRS}_{0.5}$ distribution), we can classify individuals into more refined risk categories, as proposed in Ding et al.[4].

In particular, individuals with $\text{QR-PRS}_{0.5}$ above the high-risk threshold t can be further divided into

(1) certain-above t: the interval $\left(\text{QR-PRS}_\alpha(i), \text{QR-PRS}_{1\text{-}\alpha}(i)\right)$ is above t.
(2) uncertain-above t: the interval $\left(\text{QR-PRS}_\alpha(i), \text{QR-PRS}_{1\text{-}\alpha}(i)\right)$ contains t.

Similarly, individuals with $\text{QR-PRS}_{0.5}$ below the high-risk threshold t can be divided into:

(1) uncertain-below t: the interval $\left(\text{QR-PRS}_\alpha(i), \text{QR-PRS}_{1\text{-}\alpha}(i)\right)$ contains t.
(2) certain-below t: the interval $\left(\text{QR-PRS}_\alpha(i), \text{QR-PRS}_{1\text{-}\alpha}(i)\right)$ is below t.

We take $\alpha = 0.025$ for illustration in the rest of the analyses. Examples of individuals in each of the four categories are shown in **Fig. 1(b)**. For each individual, we also compute tail probabilities $p = P(Y > t|X)$ for a given threshold t (90th percentile of $\text{QR-PRS}_{0.5}$). For the four individuals in **Fig. 1(b)**, these probabilities correspond well to our intuition, i.e. certain-above t: $p = 1$; uncertain-above t: $p = 0.61$; uncertain-below t: $p = 0.47$; and certain-below t: $p = 0$.

It is also important to note that, unlike conventional models, QR-PRS prediction intervals are not constrained to be symmetric and are allowed to vary in length from individual to individual (**Fig. 1(c)**), which is a key difference with the Ding et al.[4] credible intervals. In particular, individuals in the certain-above/certain-below t risk categories have asymmetric and narrower prediction intervals relative to individuals in the uncertain-above/uncertain-below t risk groups (**Fig. S8**). Furthermore, individuals with similar $\text{QR-PRS}_{0.5}$ values can have different prediction intervals which is also in contrast with the method in Ding et al.[4]. Such cases can arise if, for example, some of these individuals carry specific variants with effects only at upper or lower quantiles of the phenotype distribution (and negligible effects on the mean phenotype). Such heterogeneous effects across quantile levels may reflect underlying gene-by-environment or context-specific effects[7]. Therefore, the inclusion of such variants in the QR-PRS construction allows us to implicitly capture some unmodelled interactions or context-dependent associations. This is an important point because, while gene-by-environment interactions are assumed to play a crucial role in determining biomarkers and quantitative traits[20], they are difficult to model explicitly given the large number of possible interacting factors and the fact that only few such factors can be assessed in any given study.

Similar to Ding et al.[4] we can calculate what proportion of individuals in a "high-risk" category based on point prediction $\text{QR-PRS}_{0.5}$ being greater than a threshold t are classified as belonging to certain-above t vs. uncertain-above t groups. And, similarly, for individuals in a "low-risk" category based on $\text{QR-PRS}_{0.5}$, we can calculate what proportion of them are in certain-below t vs. uncertain-below t groups. Results will depend on whether we use the plug-in vs. conformal prediction intervals.

Plug-in QR-PRS prediction intervals: Of those individuals with "high" $\text{QR-PRS}_{0.5}$ a small percentage are in the certain-above t group; i.e. for those at or above the 90th percentile of the $\text{QR-PRS}_{0.5}$ distribution, only 2.8% fall into the certain-above t group; for those at or above the 80th percentile of the $\text{QR-PRS}_{0.5}$ distribution, only 5% fall into the certain-above t group (**Table 1**). Roughly 30% (40%) of those classified as certain-above *t* had actual CRP values above the 90th (80th) percentile of the UKBB distribution. Although this proportion may seem low, unmodelled environmental factors such as diet, physical activity, body weight and lifestyle have a substantial contribution to CRP.
It is also interesting to identify individuals at the other extreme, i.e. those individuals with "low" $\text{QR-PRS}_{0.5}$. Among those at or below the 90th (80th) percentile of the $\text{QR-PRS}_{0.5}$ distribution, 41% (26.4%) are in the certain-below t category with 93.4% (86.9%) of them with actual CRP values below the 90th (80th) percentile of the UKBB distribution.

Furthermore, the categorization of an individual in one of the four risk groups agrees well with the computed probabilities $P(Y > t|X)$ (**Table 1**).

Conformal QR-PRS prediction intervals: For the conformalized versions, the intervals become wider as expected and therefore fewer individuals can be classified in the certain-above and certain-below t groups. Specifically, of those individuals with "high" $\text{QR-PRS}_{0.5}$, i.e. at or above the 90th (80th) percentile of the $\text{QR-PRS}_{0.5}$ distribution, only 1.5% (2.5%) are in the certain-above t group (**Table 1**). Roughly 27% (38%) of those classified as certain-above *t* had actual CRP values above the 90th (80th) percentile of the UKBB distribution. Among those at or below the 90th (80th) percentile of the $\text{QR-PRS}_{0.5}$ distribution, 30.8% (18%) are in the certain-below t category with 93.7% (87.6%) of them with actual CRP values below the 90th (80th) percentile of the UKBB distribution.

Comparison with risk stratification based on LR-PRS credible intervals: The proportions in certain-above and certain-below groups for QR-PRS/CQR-PRS are meaningfully larger than those based on the LR-PRS 95% credible intervals[4] being above or below the 90th (80th) percentile of the LR-PRS distribution. Specifically, 0.0% (0.1%) of the individuals with "high" LR-PRS are in the certain-above t group (**Table 1**), concordant with the results in Ding et al.[4]. For certain-below, only 7% (3%) are in the certain-below t category for LR-PRS (consistent with the results in Ding et al.[4]) with 96.7% (94.4%) of them with actual CRP values below the 90th (80th) percentile of the UKBB distribution.

Tail probabilities $P(Y > t|X)$ have different distributions for individuals in different risk categories (**Fig. 2(a)**). The correlation between CRP and $P(Y > t|X)$ is positive (**Fig. 2(b)**) but not particularly strong, likely due to environmental factors playing an important role for CRP. On the other hand, the correlation between $\text{QR-PRS}_{0.5}$ and $P(Y > t|X)$ is stronger, suggesting that $\text{QR-PRS}_{0.5}$ provides useful information about $P(Y > t|X)$ (**Fig. 2(c)**). Additionally, the probabilities $P(Y > t|X)$ from QR-PRS and LR-PRS are highly correlated (**Fig. 2(d)**).

Risk stratification based on prediction intervals as above is not the same as that based on the point predictions $\text{QR-PRS}_{0.5}$. Although individuals in the certain-above t category have significantly higher $\text{QR-PRS}_{0.5}$ and actual CRP values relative to individuals in the uncertain-above t category, there are individuals in the uncertain-above t group with larger $\text{QR-PRS}_{0.5}$ values than individuals in the certain-above t group (**Fig. S9**).

Applications to 64 traits

In addition to CRP, we applied QR-PRS to 63 other quantitative traits with a range of heritability values (**Table S1**). We show that across these diverse traits the conformal prediction intervals have valid (marginal) coverage, while the plug-in prediction intervals from QR and the LR credible intervals can suffer from undercoverage (**Fig. 3**).

Although, as expected, we observe a pattern of high uncertainty in risk stratification across all traits, both QR-PRS and CQR-PRS identify a significantly greater number of individuals in the

certain-above *t* and certain-below *t* categories compared to LR-PRS (**Fig. 4(a,c), Table S2**). Notably, the proportion of individuals in the certain-above and certain-below groups is positively and significantly correlated with the trait's heritability (**Fig. S10**).

Validations using the ProgeNIA/SardiNIA project data

We utilize a second, independent dataset from the ProgeNIA/SardiNIA project[21,22], a longitudinal study of a cohort of Sardinian subjects (n=6,159) for validation. Of the 64 quantitative traits we analyzed in the UKBB, 28 are available to us from the ProgeNIA/SardiNIA project. The results confirm the high uncertainty in risk stratification, and validate the finding that inference based on prediction intervals derived in the QR framework results in more individuals being assigned to the certain-above *t* and certain-below *t* risk categories relative to the counterpart LR framework (**Fig. 4(b,d), Table S3**).

Discussion

There is great interest in the human, plant and animal genetics communities to build trait prediction models based on genome-wide genotype data. For human genetics, the emphasis has been on employing PRS for risk stratification in clinical settings. However, this has proven to be challenging due to complexities underlying risk to disease including the important role of environmental effects and their complex interactions with genetic factors. Furthermore, although scientists are well aware of the inherent imprecision and uncertainty around PRS estimates, this uncertainty is rarely reported or accounted for in analyses. However, taking uncertainty into account can lead to better risk stratification for individuals in high/low risk groups[4], while also providing more accurate information on the current limitations of PRS prediction.

We discuss here the need to assess uncertainty around the point PRS estimate and describe a natural way to quantify uncertainty based on building valid prediction intervals by combining QR and conformal prediction. The resulting intervals have correct marginal coverage under minimal assumptions, and they are more realistic relative to more conventional methods to assess uncertainty, i.e. they can be asymmetric and have varying length across individuals. In contrast, credible intervals as described in Ding et al. assess uncertainty in the estimated conditional mean $E(Y|X)$, and they do not capture the full variability in the phenotype Y around $E(Y|X)$. Furthermore, these intervals may be miscalibrated in practice due to violation of various model assumptions, such as genetic architecture and LD patterns.

Although QR-PRS is by design applicable to quantitative traits, it serves as a framework to illustrate how uncertainty quantification can be achieved by constructing more realistic prediction intervals via quantile regression. Moreover, disease biomarkers and intermediate cellular phenotypes[23] are important in clinical medicine and pharmacogenomics. Similarly, for plants and animals constructing PRS for quantitative traits can be even more useful than for human studies for purposes such as selective breeding and conservation efforts[24–26]. Beyond quantitative traits, a future direction of interest will be to aggregate multiple QR-PRS at different quantile levels and for different disease-related endophenotypes or biomarkers to derive meta

PRS scores for disease traits of interest with improved accuracy[27], similar to ideas previously implemented in the context of LR[3,28]. Preliminary analyses suggest that including PRS for biomarkers may improve the accuracy over a disease-only PRS, in some cases substantially so (**Fig. S13**, **S14**).

There are a few limitations of the proposed QR-PRS score and these will be the subject of future work. One inherent limitation is that QR requires individual level data to train the quantile regression models, and thus cannot leverage existing summary statistics. However, given the large biobanks available across the world with individual level data and diverse biomarkers, including molecular, clinical and imaging biomarkers, and the increasing trend to collect quantitative traits across human, plant and animal studies, this is not a major concern. A more important limitation is that our current PRS construction is based on a simple pruning + thresholding method and has not been particularly optimized. A range of more sophisticated techniques including ridge regression and Bayesian regression methods[11,12] could be incorporated in the quantile framework but that requires non-trivial effort in the context of quantile regression, especially from a computational perspective, and we leave that for future work. Finally, while conformal prediction guarantees marginal coverage (when evaluated across all individuals), conditional coverage is not guaranteed. In particular, coverage within subgroups of individuals, such as those with high trait values, may not be guaranteed. There is ongoing active work around constructing conditionally valid prediction intervals that can be useful in this setting[29].

Acknowledgements. This research has been conducted using the UK Biobank Resource under Application Number 41849. We thank Atlas Khan for help with the UKBB data processing.

Data Availability. The individual-level genotype and phenotype data are available to approved researchers through the UK Biobank web portal at https://www.ukbiobank.ac.uk/. The LR and QR GWAS summary statistics for the biomarkers reported here are publicly available at https://doi.org/10.5281/zenodo.11095249.

Code Availability. Scripts for QR-PRS and CQR-PRS analyses are available at https://github.com/Iuliana-Ionita-Laza/QRPRS. The LR-PRS and PRS uncertainty estimation described in Ding et al. was performed using LDpred2 from https://github.com/privefl/bigsnpr and the code from https://github.com/bogdanlab/prs-uncertainty. Scripts for QR GWAS analysis are available at https://github.com/Iuliana-Ionita-Laza/QRGWAS. The quantile regression was performed using the R package quantreg (https://cran.r-project.org/package=quantreg) version 5.94. The rank score test was performed using the R package QRank (https://cran.r-project.org/package=QRank) version 1.0. Analyses with these R packages were performed in R release 4.2.2. The linear regression analysis and P+T PRS were performed using the PLINK 2.0 (https://www.cog-genomics.org/plink/2.0/).

References

1. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
2. Lewis, C. M. & Vassos, E. Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* **12**, 44 (2020).
3. Sinnott-Armstrong, N. *et al.* Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* **53**, 185–194 (2021).
4. Ding, Y. *et al.* Large uncertainty in individual polygenic risk score estimation impacts PRS-based risk stratification. *Nat. Genet.* **54**, 30–39 (2022).
5. Romano, Y., Patterson, E. & Candes, E. Conformalized Quantile Regression. in *Advances in Neural Information Processing Systems* vol. 32 (Curran Associates, Inc., 2019).
6. Pouplin, T., Jeffares, A., Seedat, N. & Van Der Schaar, M. Relaxed quantile regression: prediction intervals for asymmetric noise. in *Proceedings of the 41st International Conference on Machine Learning* vol. 235 40951–40981 (JMLR.org, Vienna, Austria, 2024).
7. Wang, C. *et al.* Genome-wide discovery for biomarkers using quantile regression at biobank scale. *Nat. Commun.* **15**, 6460 (2024).
8. Snaebjarnarson, A. S. *et al.* Complex effects of sequence variants on lipid levels and coronary artery disease. *Cell* **186**, 4085-4099.e15 (2023).
9. Angelopoulos, A. N. & Bates, S. A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification. Preprint at https://doi.org/10.48550/arXiv.2107.07511 (2022).
10. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).

11. Purcell, S. M. *et al.* Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature* **460**, 748–752 (2009).

12. Dudbridge, F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLOS Genet.* **9**, e1003348 (2013).

13. Chernozhukov, V., Fernández-Val, I. & Galichon, A. Quantile and Probability Curves Without Crossing. *Econometrica* **78**, 1093–1125 (2010).

14. Chernozhukov, V., Fernández-Val, I. & Galichon, A. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96**, 559–575 (2009).

15. Koenker, R. & Bassett, G. Regression Quantiles. *Econometrica* **46**, 33–50 (1978).

16. Koenker, R. *Quantile Regression*. vol. 38 (Cambridge university press, 2005).

17. Song, X. *et al.* QRank: a novel quantile regression tool for eQTL discovery. *Bioinformatics* **33**, 2123–2130 (2017).

18. Boor, C. D. *A Practical Guide to Splines*. (Springer New York, 1978).

19. Privé, F., Arbel, J. & Vilhjálmsson, B. J. LDpred2: better, faster, stronger. *Bioinformatics* **36**, 5424–5431 (2021).

20. Mackay, T. F. C. & Anholt, R. R. H. Pleiotropy, epistasis and the genetic architecture of quantitative traits. *Nat. Rev. Genet.* **25**, 639–657 (2024).

21. Pilia, G. *et al.* Heritability of cardiovascular and personality traits in 6,148 Sardinians. *PLoS Genet.* **2**, e132 (2006).

22. Sidore, C. *et al.* Genome sequencing elucidates Sardinian genetic architecture and augments association analyses for lipid and blood inflammatory markers. *Nat. Genet.* **47**, 1272–1281 (2015).

23. Homilius, M. *et al.* Perturbational phenotyping of human blood cells reveals genetically determined latent traits associated with subsets of common diseases. *Nat. Genet.* **56**, 37–50 (2024).

24. Wray, N. R., Kemper, K. E., Hayes, B. J., Goddard, M. E. & Visscher, P. M. Complex Trait Prediction from Genome Data: Contrasting EBV in Livestock to PRS in Humans. *Genetics* **211**, 1131–1141 (2019).

25. Fuller, Z. L. *et al.* Population genetics of the coral Acropora millepora: Toward genomic prediction of bleaching. *Science* **369**, eaba4674 (2020).

26. Wang, X. *et al.* Integrative genomics reveals the polygenic basis of seedlessness in grapevine. *Curr. Biol.* **34**, 3763-3777.e5 (2024).

27. Fakoor, R., Kim, T., Mueller, J., Smola, A. J. & Tibshirani, R. J. Flexible Model Aggregation for Quantile Regression. *J. Mach. Learn. Res.* **24**, 1–45 (2023).

28. Abraham, G. *et al.* Genomic risk score offers predictive performance comparable to clinical risk factors for ischaemic stroke. *Nat. Commun.* **10**, 5819 (2019).

29. Chernozhukov, V., Wüthrich, K. & Zhu, Y. Distributional conformal prediction. *Proc. Natl. Acad. Sci.* **118**, e2107794118 (2021).

30. Denny, J. C. *et al.* Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat. Biotechnol.* **31**, 1102–1111 (2013).

31. Wu, P. *et al.* Mapping ICD-10 and ICD-10-CM Codes to Phecodes: Workflow Development and Initial Evaluation. *JMIR Med. Inform.* **7**, e14325 (2019).

32. Zhou, W. *et al.* Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat. Genet.* **50**, 1335–1341 (2018).

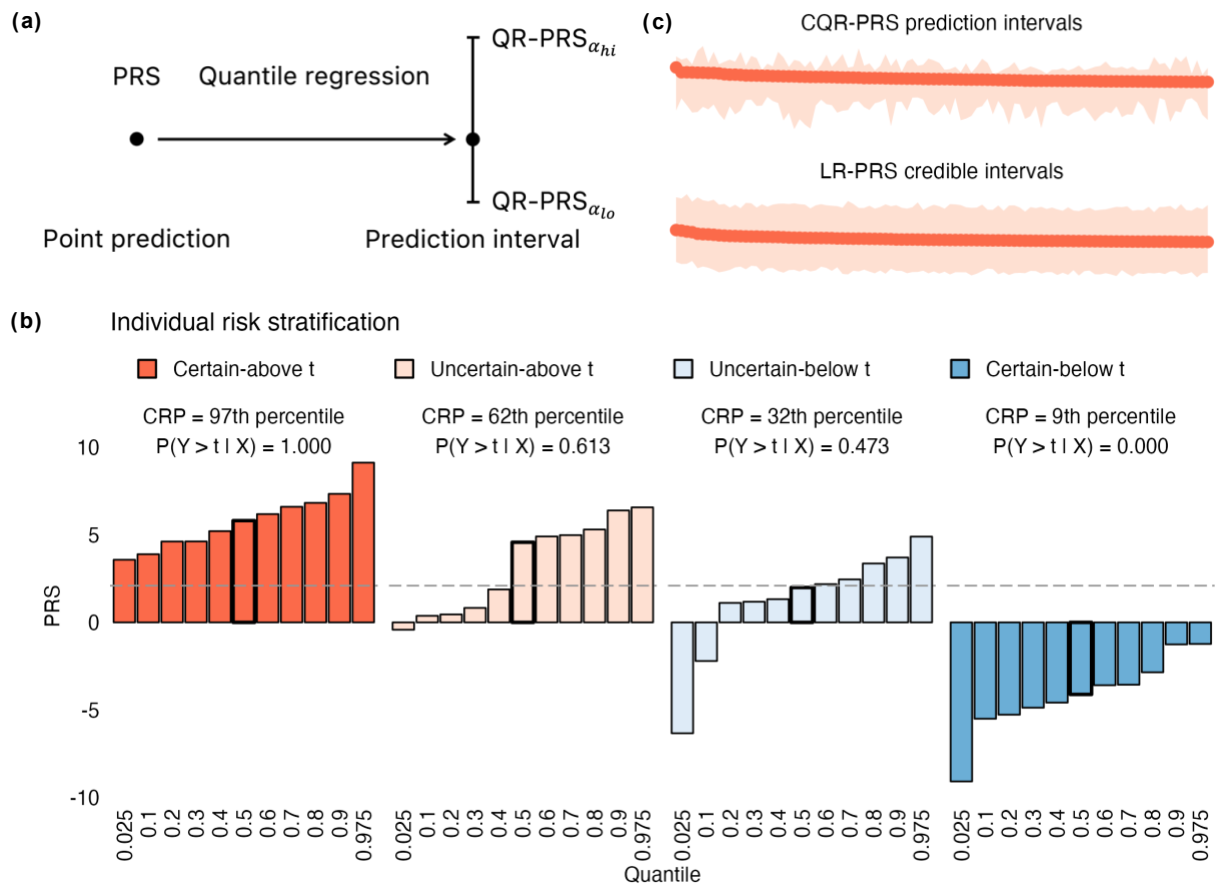

Figure 1: **From point prediction to interval prediction for PRS.** (a) Quantile regression naturally moves the emphasis from point prediction in conventional PRS to prediction interval that gives predictions of lower and upper quantiles of the conditional phenotype distribution. (b) Four individuals, each in a different risk category based on their predicted C-reactive protein quantiles across a grid of quantile levels $\tau \in (0.025, 0.975)$ (each bar corresponds to a predicted quantile). The threshold, $t = 90^{\text{th}}$ percentile of $\text{QR-PRS}_{0.5}$ (top 10%), is shown as a gray horizontal line. For every individual we also report their actual C-reactive protein (CRP) percentile and $P(Y > t|X)$. (c) Prediction intervals ($\text{CQR-PRS}_{0.025}, \text{CQR-PRS}_{0.975}$) and 95% credible intervals are shown for top 100 individuals in uncertain-above categories for CQR-PRS and LR-PRS.

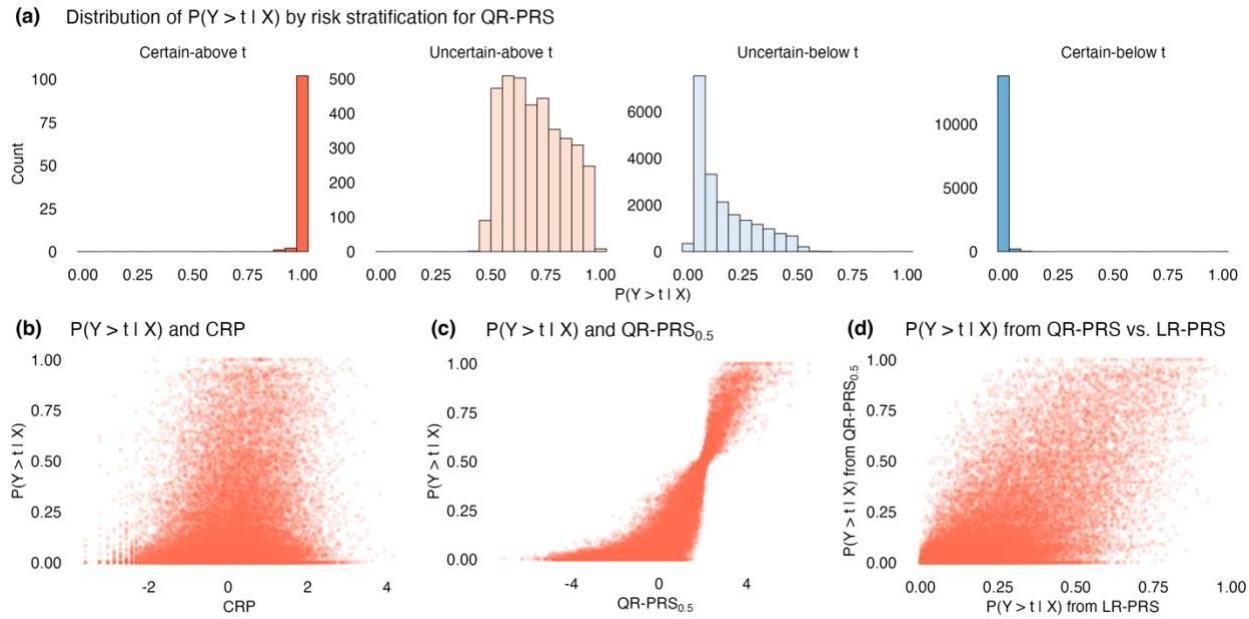

Figure 2: Distribution of $P(Y > t|X)$ for individuals in different risk categories and correlation to CRP and $\mathrm{QR\text{-}PRS}_{0.5}$. Panel (a) corresponds to QR-PRS prediction intervals when using 10% as threshold for "high" $\mathrm{QR\text{-}PRS}_{0.5}$. Panels (b)-(d) show the relationship between $P(Y > t|X)$, CRP, and $\mathrm{QR\text{-}PRS}_{0.5}$: (b) $P(Y > t|X)$ and actual CRP, $r = 0.23$; (c) $P(Y > t|X)$ and $\mathrm{QR\text{-}PRS}_{0.5}$, $r = 0.79$; (d) $P(Y > t|X)$ from QR-PRS vs. LR-PRS, $r = 0.67$.

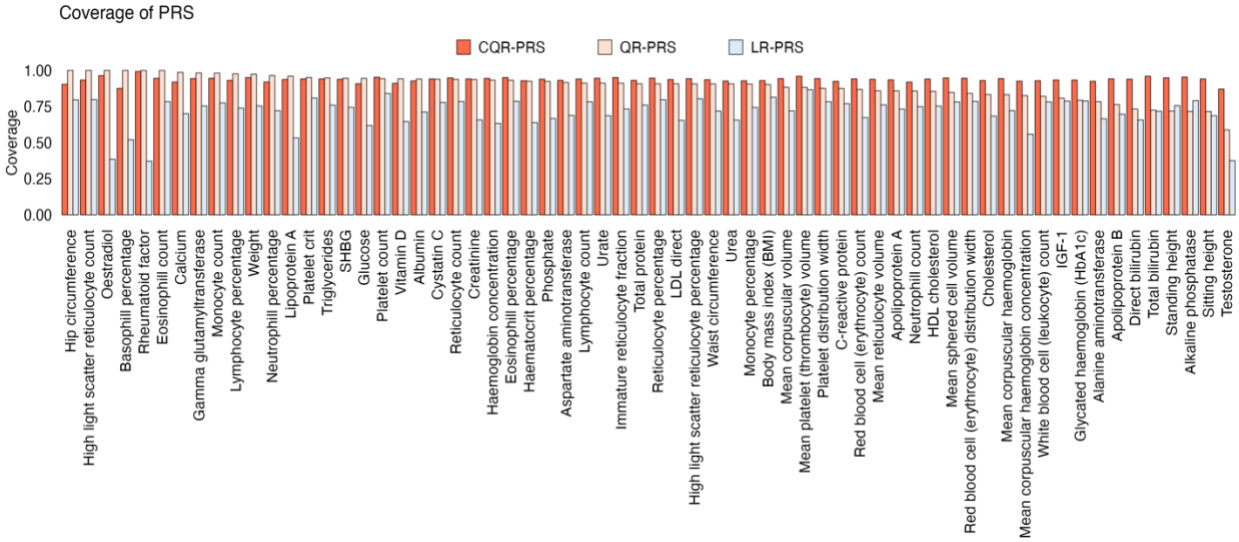

Figure 3: Coverage for prediction/credible intervals from QR-PRS, CQR-PRS, LR-PRS across 64 traits in UKBB. Coverage is computed as the proportion of individuals whose prediction intervals (QR-PRS, CQR-PRS) and 95% credible intervals (LR-PRS) overlap their true phenotype value.

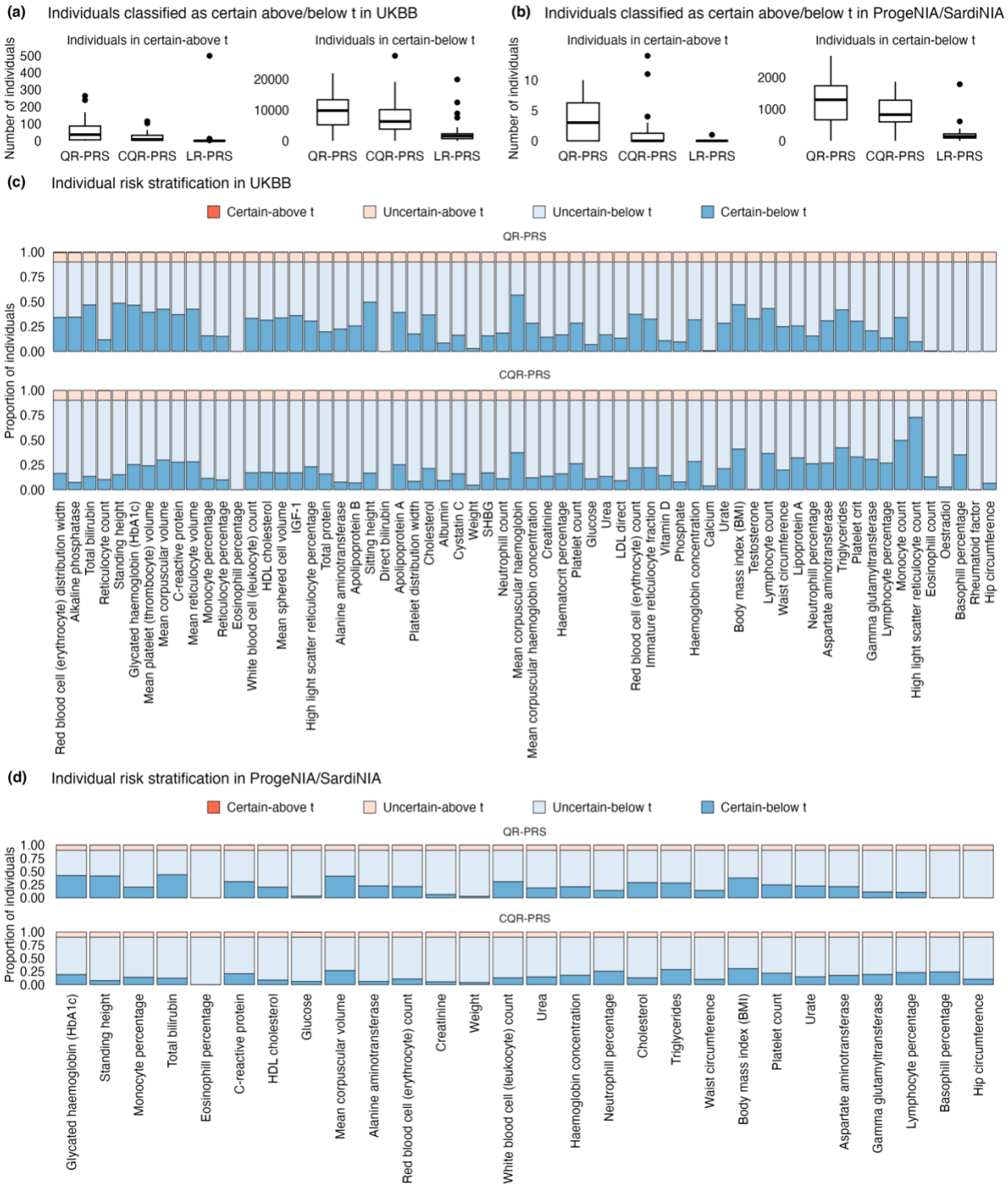

Figure 4: Risk stratification based on LR-PRS and QR-PRS across 64 traits. (a) Number of individuals classified as certain-above/-below t by QR-PRS, CQR-PRS and LR-PRS in UKBB. (b) Number of individuals classified as certain-above/-below t by QR-PRS, CQR-PRS and LR-PRS in ProgeNIA/SardiNIA. (c) Risk stratification based on QR-PRS/CQR-PRS across 64 traits in UKBB. (d) Risk stratification based on QR-PRS/CQR-PRS across 28 traits in ProgeNIA/SardiNIA. The threshold t corresponds to the 90th $\text{QR-PRS}_{0.5}$ percentile.

Threshold	Stratification	QR-PRS				CQR-PRS				LR-PRS			
		N	Fraction	% CRP above/ below t	P(Y>t|X)	N	Fraction	% CRP above/ below t	P(Y>t|X)	N	Fraction	% CRP above/ below t	P(Y>t|X)
Top 10%	Certain-above t	105	2.8%	28.6%	99.4%	55	1.5%	27.3%	99.6%	1	0.0%	0.0%	98.0%
Top 10%	Uncertain-above t	3683	97.2%	19.4%	70.4%	3733	98.5%	19.6%	68.0%	3787	100.0%	21.1%	62.2%
Top 10%	Uncertain-below t	20075	58.9%	89.5%	16.1%	23583	69.2%	89.9%	17.4%	31717	93.0%	90.8%	20.6%
Top 10%	Certain-below t	14011	41.1%	93.4%	0.4%	10503	30.8%	93.7%	0.5%	2369	7.0%	96.7%	1.4%
Top 20%	Certain-above t	378	5.0%	41.3%	99.3%	188	2.5%	38.3%	99.4%	4	0.1%	50.0%	98.4%
Top 20%	Uncertain-above t	7197	95.0%	30.3%	72.6%	7387	97.5%	30.7%	70.3%	7571	99.9%	33.3%	64.4%
Top 20%	Uncertain-below t	22289	73.6%	81.3%	18.7%	24854	82.0%	81.7%	20.4%	29400	97.0%	83.0%	24.5%
Top 20%	Certain-below t	8010	26.4%	86.9%	0.5%	5445	18.0%	87.6%	0.5%	899	3.0%	94.4%	1.5%

Table 1: **Risk stratification (QR-PRS, CQR-PRS, and LR-PRS) for C-reactive protein (CRP) in UKBB.** Among individuals with $\mathrm{QR\text{-}PRS}_{0.5}$/LR-PRS in top 10% and top 20%, we report the proportion of individuals "certain-above *t*" and "uncertain-above *t*". Similarly, for individuals with $\mathrm{QR\text{-}PRS}_{0.5}$/LR-PRS below top 10% and top 20%, we report the proportion of individuals "certain-below *t*" and "uncertain-below *t*". For QR-PRS/CQR-PRS, these definitions are based on the plug-in/conformal prediction intervals $\left((\mathrm{C})\mathrm{QR\text{-}PRS}_{0.025}(i), (\mathrm{C})\mathrm{QR\text{-}PRS}_{0.975}(i)\right)$. For LR-PRS, the definitions are based on the 95% credible intervals for LR-PRS. For each risk category we also report the proportion of individuals with actual CRP values above or below the top 10% and top 20%, and the mean value of $P(Y > t|X)$ for individuals in the category.

Supplemental Material

Combining multiple biomarker PRS for disease risk prediction

Although QR-PRS is designed for quantitative traits such as biomarkers, these scores have clinical implications and can, in principle, help build better prediction methods for clinical diseases of interest. Specifically, we can build a composite or meta PRS for disease traits by integrating quantile-specific QR-PRS for multiple biomarkers. For a given quantile level τ, we can construct a composite predictor for a target disease as follows:

$$\text{meta-PRS}_{\tau} = \beta_1 PRS_{1,\tau} + \beta_2 PRS_{2,\tau} + \cdots + \beta_M PRS_{M,\tau} + \beta_D PRS_D,$$

where M is the number of biomarkers to be combined and D is the target disease for which a PRS already exists. The β coefficients can be estimated by regularized regression techniques such as least absolute shrinkage and selection operator (LASSO) or ridge regression. The aggregation can be more general and include QR-PRS at different quantile levels.

Combining multiple QR-PRS for different biomarkers can improve disease risk prediction for disease clinical outcomes

Although QR-PRS is by definition applicable to quantitative traits only, we can combine multiple PRS scores for biomarkers and disease into an overall meta-PRS for disease risk prediction. We constructed a meta-PRS by combining the QR-PRS of 64 biomarkers with the disease-specific PRS for ten disease phenotypes studied in Sinnott-Armstrong et al.[3] (**Table S4**) using a penalized regression model. We used phecode to define case-control status for the ten disease phenotypes[30–32]. A ridge regression model was trained on unrelated white British samples from UKBB using the glmnet package in R. To reduce overfitting, the regularization parameter of ridge regression was optimized through five-fold cross-validation (CV) on the training set. The final model was subsequently applied to a separate dataset of unrelated European ancestry samples (non-British, n = 39,838) from UKBB to evaluate the predictive performance, the area under the receiver operating characteristic curve (AUROC), of the meta-PRS (**Fig. S13**). Overall, the meta-PRS leads to improved AUROC values vs. disease PRS, especially for several diseases including CKD (0.563 vs. 0.471) and Gout (0.678 vs. 0.635). The risk effects of meta-PRS scores were measured using odds ratio (OR) and 95% confidence interval (CI) among all individuals and among the top decile of individuals at high risk predicted by different PRS models (**Fig. S14**), confirming the superior performance of the meta-PRS over disease PRS alone.

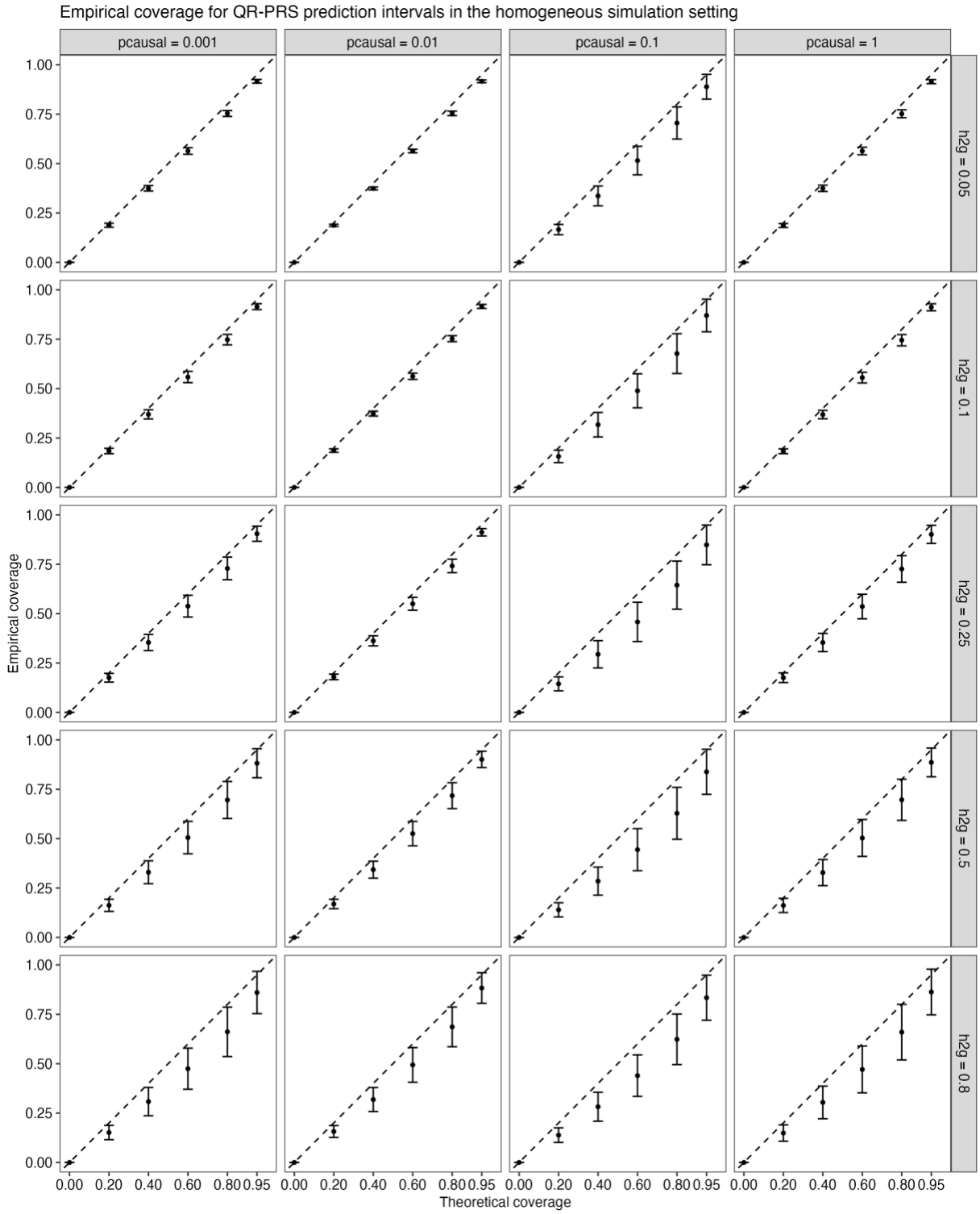

Figure S1: Empirical coverage vs. theoretical coverage for QR-PRS prediction intervals in the homogeneous simulation setting, with varying heritability and polygenicity parameters.

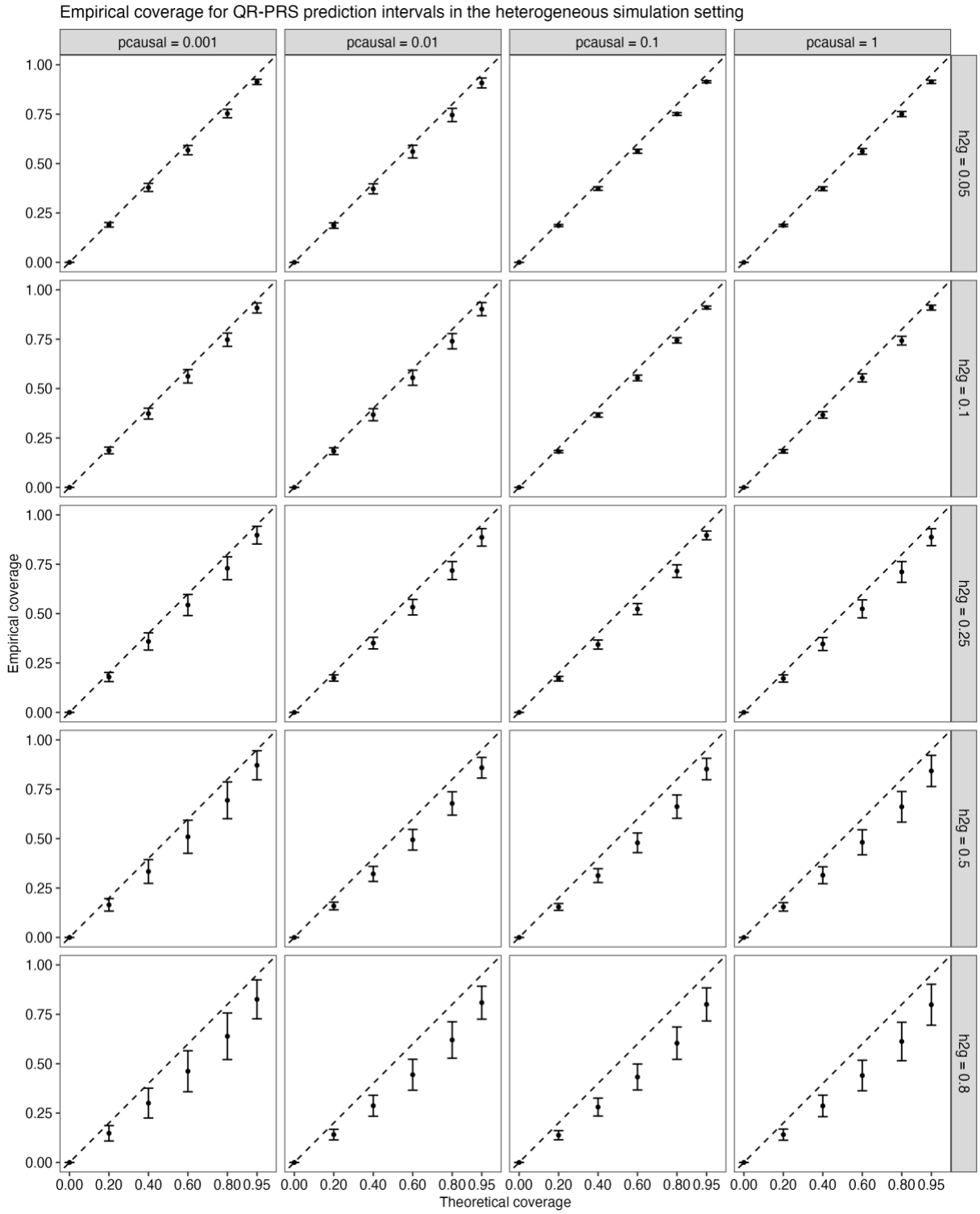

Figure S2: Empirical coverage vs. theoretical coverage for QR-PRS prediction intervals in the heterogeneous simulation setting ($m_1 : m_2 = 1 : 1$), with varying heritability and polygenicity parameters.

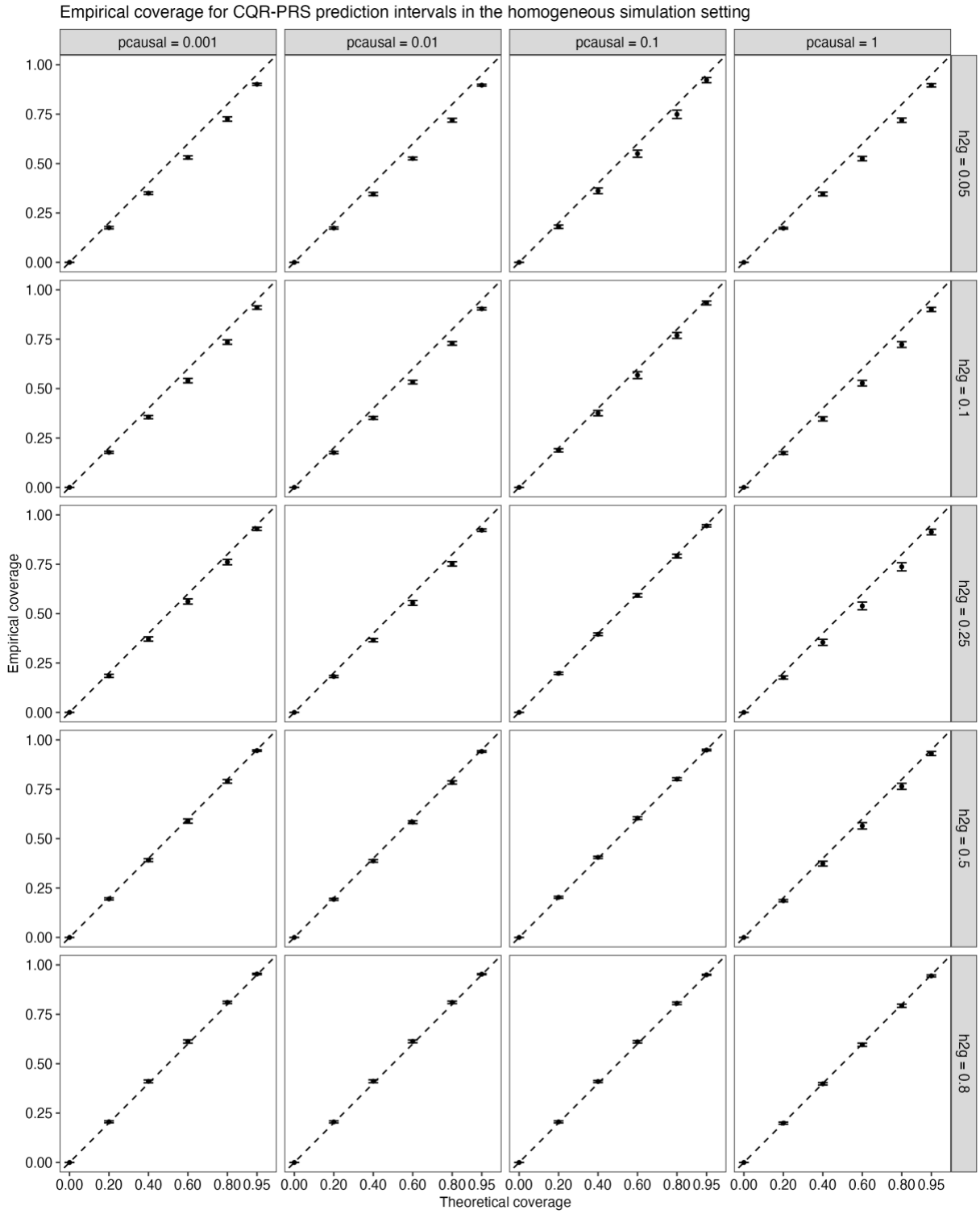

Figure S3: Empirical coverage vs. theoretical coverage for CQR-PRS prediction intervals in the homogeneous simulation setting, with varying heritability and polygenicity parameters.

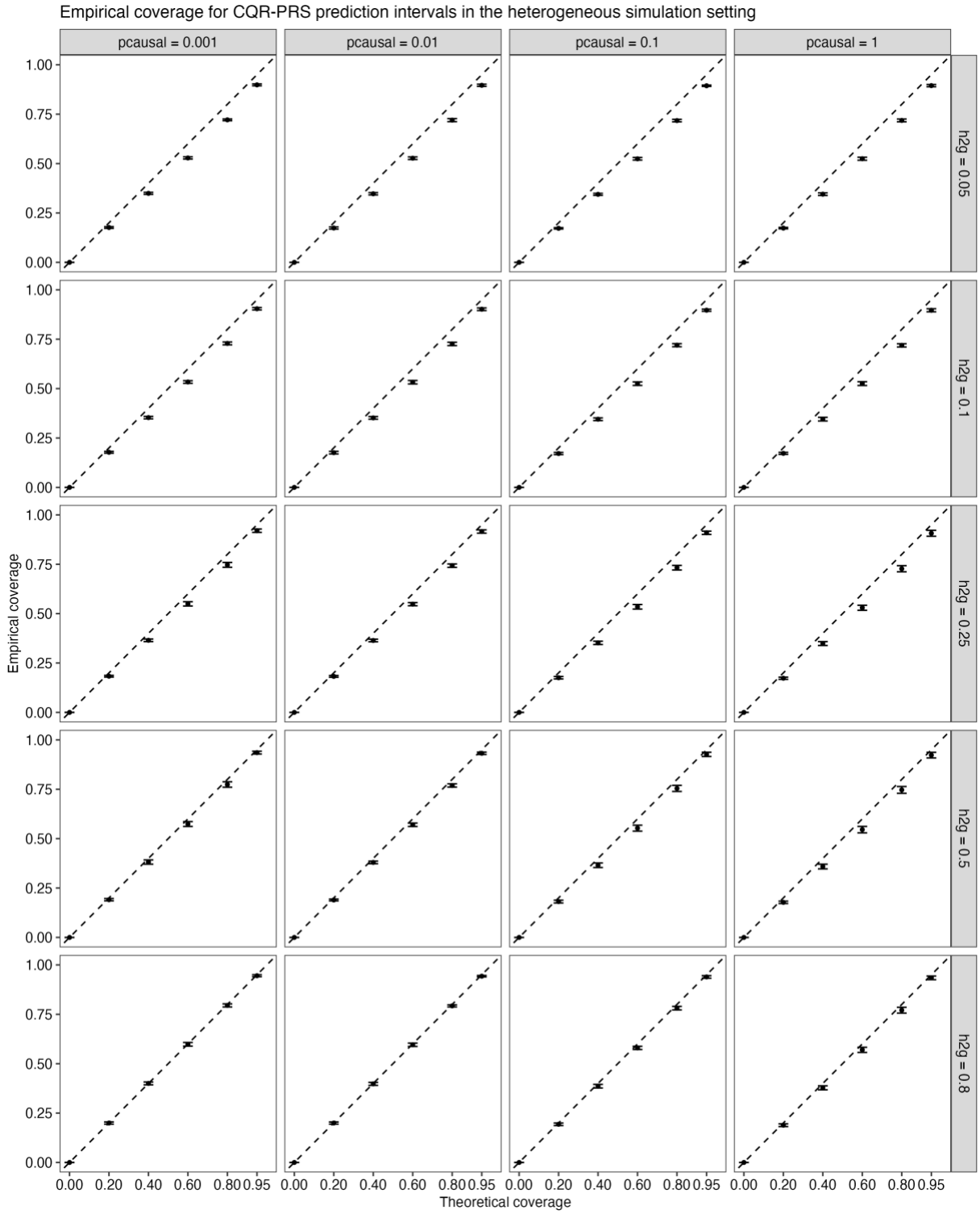

Figure S4: Empirical coverage vs. theoretical coverage for CQR-PRS prediction intervals in the heterogeneous simulation setting ($m_1 : m_2 = 1 : 1$), with varying heritability and polygenicity parameters.

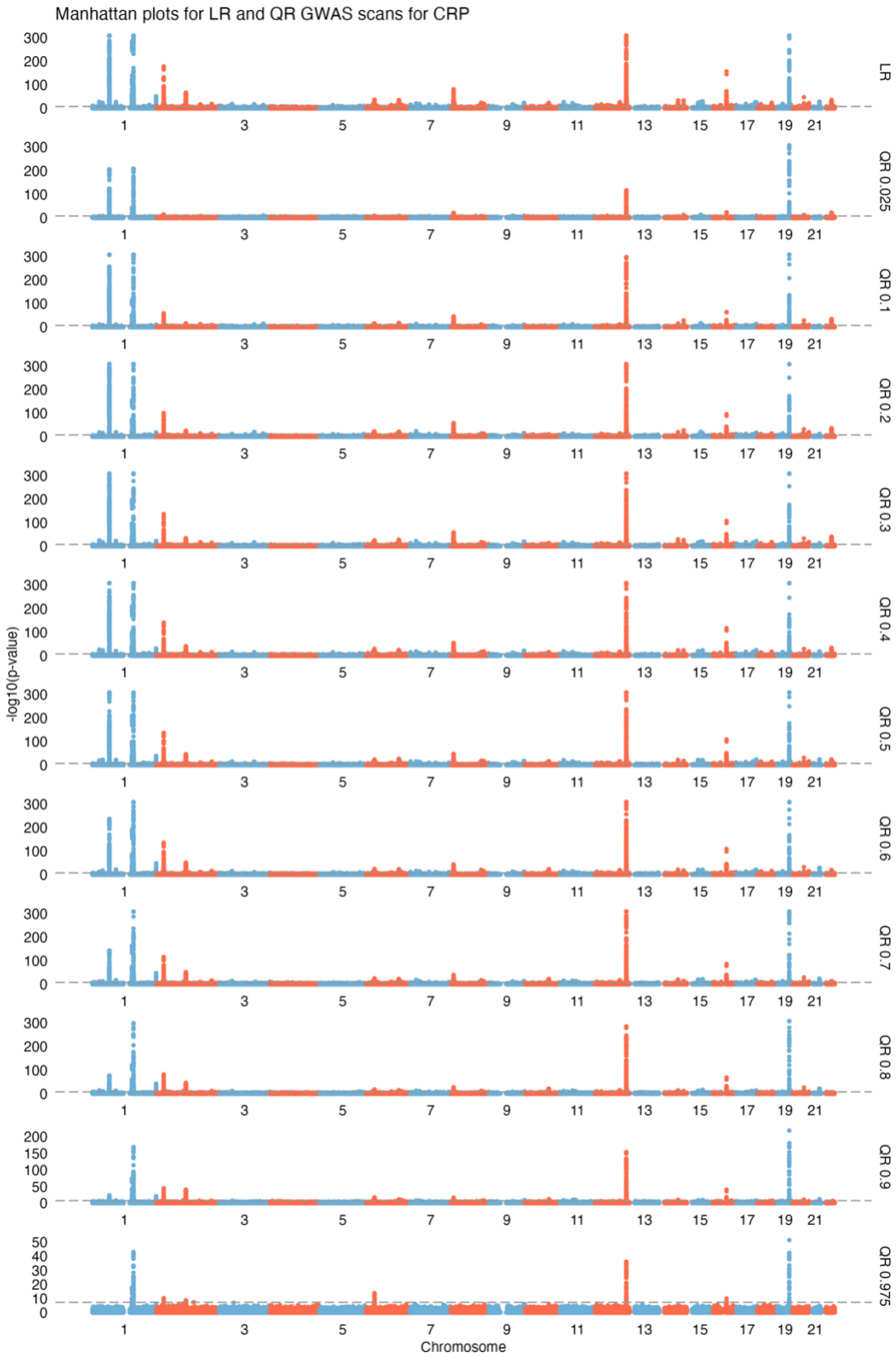

Figure S5: **Manhattan plots for LR and QR GWAS (CRP) in UKBB.** Quantile regression is run across a grid of quantile levels $\tau \in (0.025, 0.975)$.

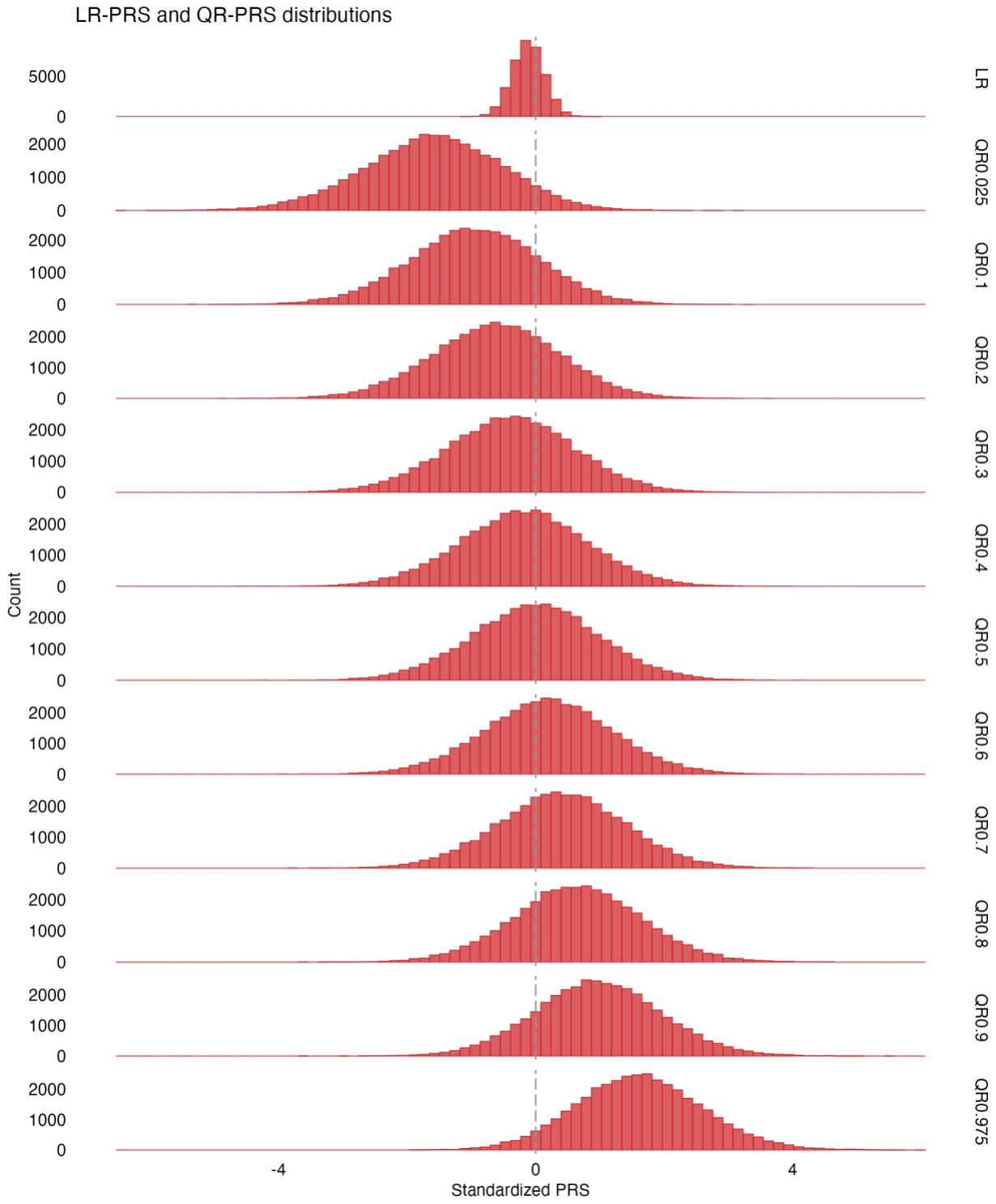

Figure S6: **LR-PRS and QR-PRS distribution (CRP) in UKBB**. PRS distribution for the mean-based LR-PRS, and individual quantile QR-PRS across a grid of quantile levels $\tau \in (0.025, 0.975)$. The different QR-PRS are standardized with respect to the $\text{QR-PRS}_{0.5}$ distribution, by subtracting the mean of $\text{QR-PRS}_{0.5}$ and dividing by the standard deviation.

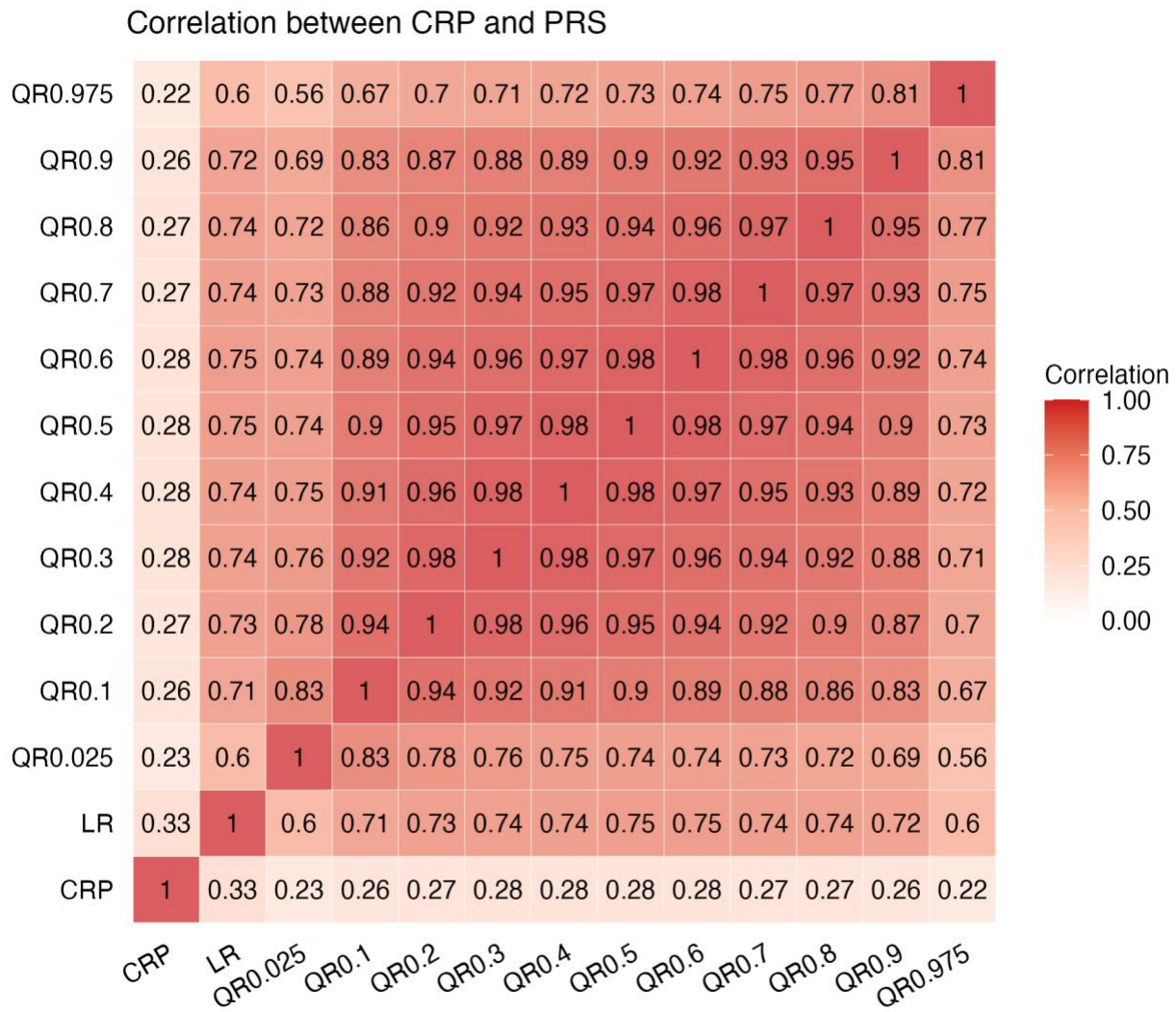

Figure S7: **Pearson correlations among actual CRP values, LR-PRS, and QR-PRS across a grid of quantile levels $\tau \in (0.025, 0.975)$ in UKBB.**

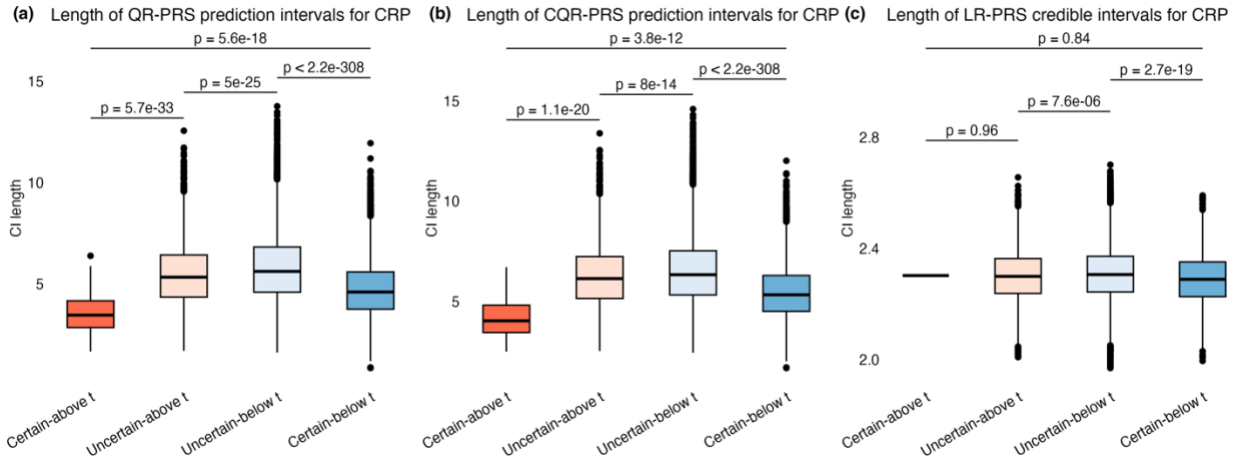

Figure S8: QR-PRS/LR-PRS interval lengths in different risk categories. (a-b) QR/CQR prediction intervals are significantly smaller for certain-above/certain-below *t* categories relative to uncertain-above/uncertain-below *t.* (c) Distribution of lengths of LR credible intervals for different risk categories. Note that the scales of the QR-PRS and LR-PRS scores are different since LR-PRS is a shrinkage-based method unlike QR-PRS.

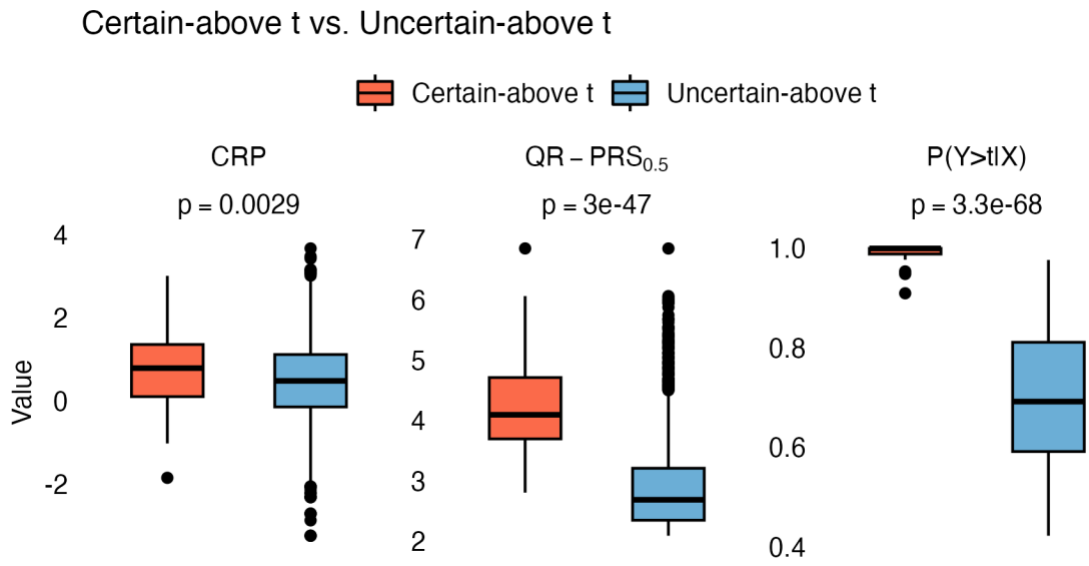

Figure S9: Certain-above *t* vs. uncertain-above *t* individuals (CRP) in UKBB. Distributions of actual CRP values, $\text{QR-PRS}_{0.5}$ values, and $P(Y > t|X)$ where $t = 90^{\text{th}}$ percentile of $\text{QR-PRS}_{0.5}$ in "certain-above *t*" vs. "uncertain-above *t*" individuals.

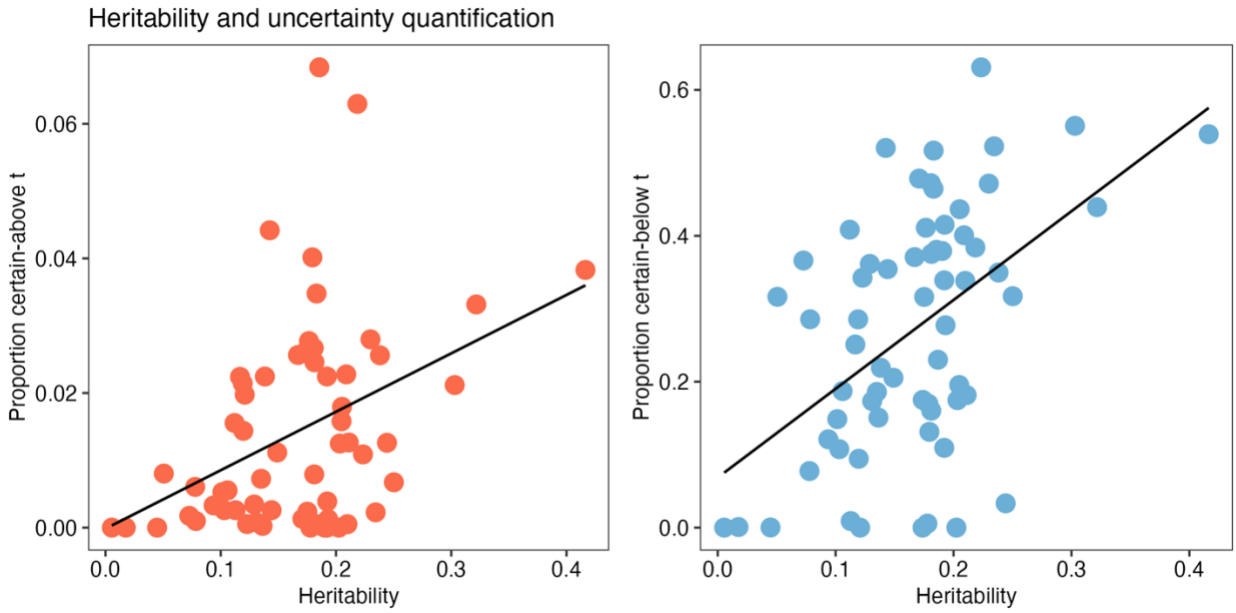

Figure S10: Relationship between heritability for a trait and uncertainty quantification in UKBB. Left: Proportion of individuals in certain-above t based on QR-PRS vs. SNP heritability as estimated by LDSC ($r = 0.39, p = 1.4E - 03$). Right: Proportion of individuals in certain-below t based on QR-PRS vs. SNP heritability as estimated by LDSC ($r = 0.50, p = 3.0E - 05$).

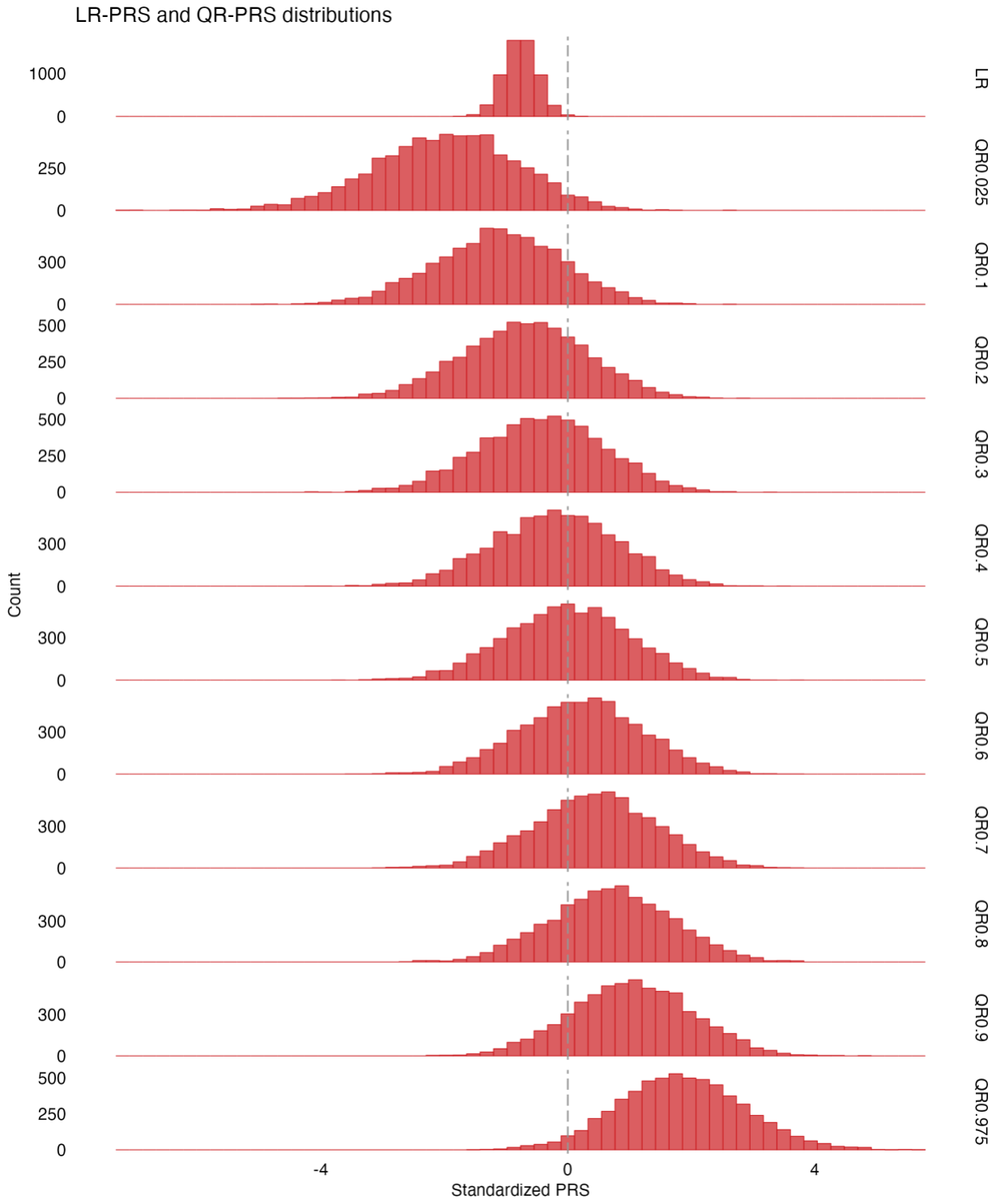

Figure S11: **LR-PRS and QR-PRS distribution (CRP) in ProgeNIA/SardiNIA**. PRS distribution for the mean-based LR-PRS, and individual quantile QR-PRS across a grid of quantile levels $\tau \in (0.025, 0.975)$. The different QR-PRS are standardized with respect to the $\text{QR-PRS}_{0.5}$ distribution, by subtracting the mean of $\text{QR-PRS}_{0.5}$ and dividing by the standard deviation.

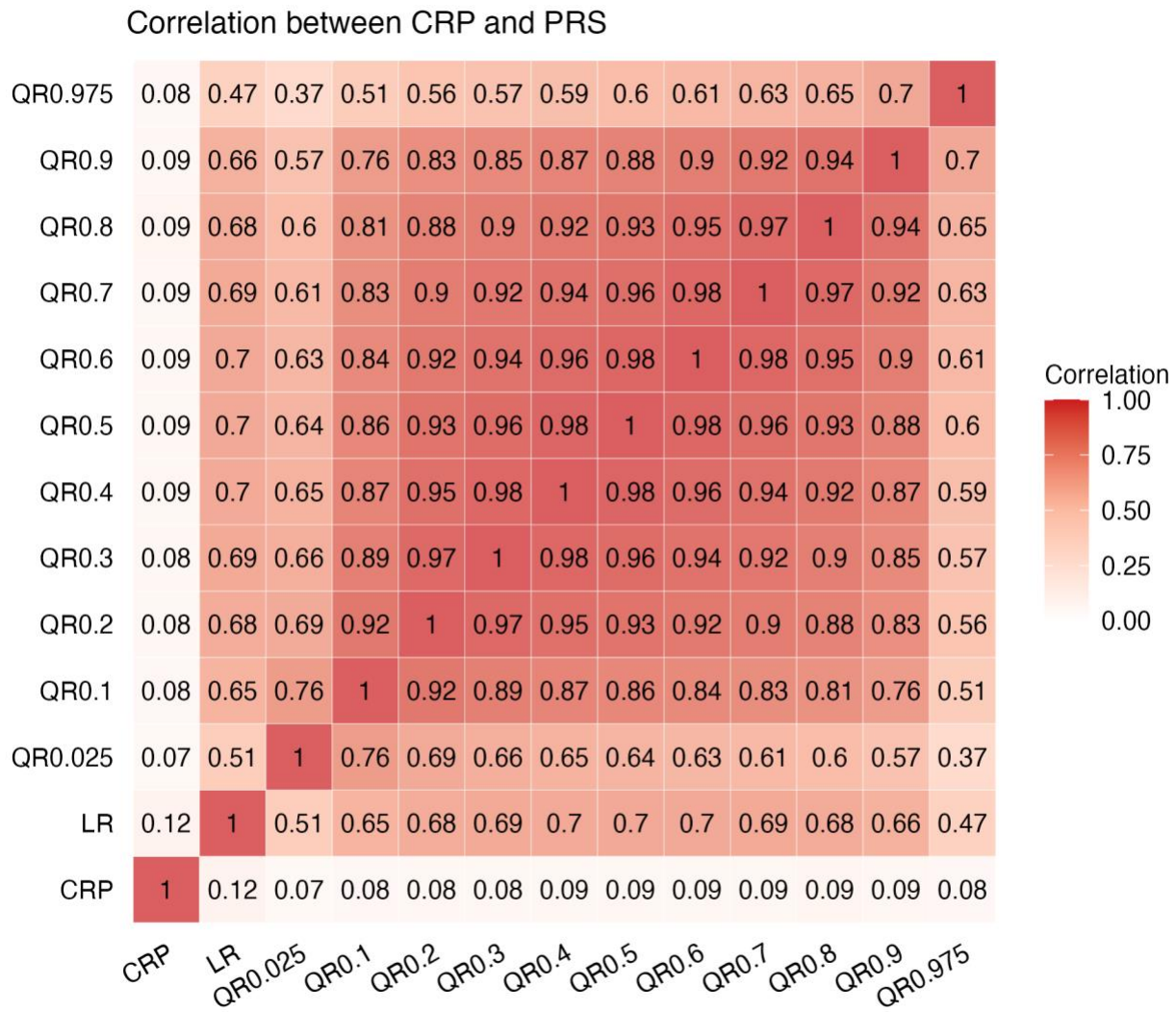

Figure S12: **Pearson correlations among actual CRP values, LR-PRS, and QR-PRS across a grid of quantile levels $\tau \in (0.025, 0.975)$ in ProgeNIA/SardiNIA.**

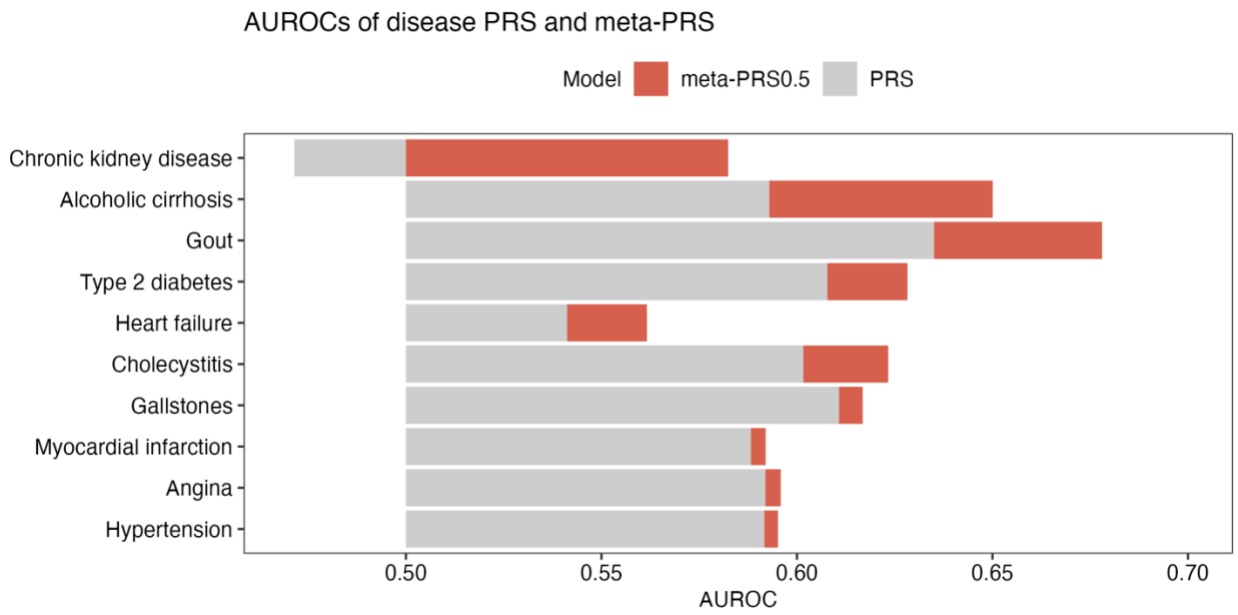

Figure S13: AUROC of meta-PRS for disease prediction in UKBB. AUROC values for ten diseases for disease PRS and meta-$\mathrm{PRS}_{0.5}$ are shown.

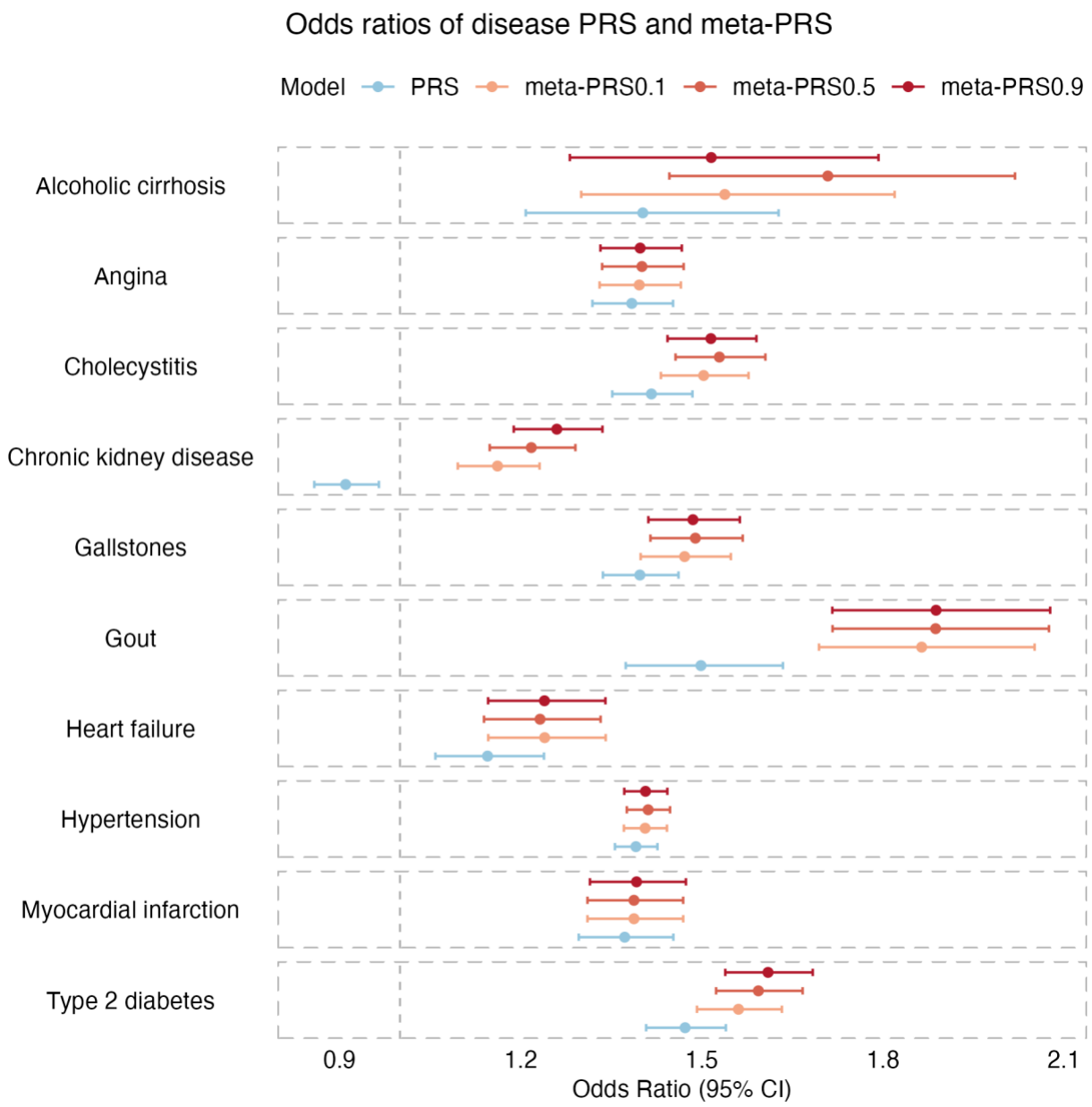

Figure S14: Odds ratio (OR) and 95% confidence interval (CI) for disease prediction in UKBB based on disease PRS and meta-PRS at different quantile levels.

UKB field	Trait	Abbreviation	N training	N test	Heritability
20015	Sitting height	SH	325379	39657	0.303
21001	Body mass index (BMI)	BMI	325306	39643	0.234
21002	Weight	W	325419	39662	0.244
30000	White blood cell (leukocyte) count	WBC	316708	38593	0.167
30010	Red blood cell (erythrocyte) count	RBC	316712	38594	0.192
30020	Haemoglobin concentration	HGB	316711	38594	0.144
30030	Haematocrit percentage	HCT	316712	38594	0.135
30040	Mean corpuscular volume	MCV	316710	38594	0.230
30050	Mean corpuscular haemoglobin	MCH	316709	38594	0.223
30060	Mean corpuscular haemoglobin concentration	MCHC	316706	38594	0.051
30070	Red blood cell (erythrocyte) distribution width	RDW	316710	38594	0.186
30080	Platelet count	PLT	316711	38594	0.250
30090	Platelet crit	PCT	316708	38594	0.210
30100	Mean platelet (thrombocyte) volume	MPV	316707	38594	0.322
30110	Platelet distribution width	PDW	316707	38594	0.205
30120	Lymphocyte count	LY#	316158	38518	0.171
30130	Monocyte count	MO#	316158	38518	0.190
30140	Neutrophill count	NE#	316158	38518	0.149
30150	Eosinophill count	EO#	316158	38518	0.178
30180	Lymphocyte percentage	LY%	316162	38519	0.136
30190	Monocyte percentage	MO%	316162	38519	0.174
30200	Neutrophill percentage	NE%	316162	38519	0.131
30210	Eosinophill percentage	EO%	316162	38519	0.174
30220	Basophill percentage	BA%	316162	38519	0.045
30240	Reticulocyte percentage	RET%	311586	37846	0.179
30250	Reticulocyte count	RET#	311587	37846	0.180
30260	Mean reticulocyte volume	MRV	311587	37845	0.181
30270	Mean sphered cell volume	MSCV	311587	37846	0.181
30280	Immature reticulocyte fraction	IRF	311587	37845	0.129
30290	High light scatter reticulocyte percentage	HLR%	311587	37846	0.192
30300	High light scatter reticulocyte count	HLR#	311587	37846	0.192
30600	Albumin	ALB	284977	34773	0.120
30610	Alkaline phosphatase	ALP	311170	37959	0.219
30620	Alanine aminotransferase	ALT	311036	37936	0.117
30630	Apolipoprotein A	APOA1	283281	34542	0.205
30640	Apolipoprotein B	APOB	309650	37764	0.119
30650	Aspartate aminotransferase	AST	310021	37817	0.123
30660	Direct bilirubin	DBIL	264854	32397	0.121
30670	Urea	UREA	310962	37930	0.106
30680	Calcium	CALD	284875	34768	0.113
30690	Cholesterol	CHOL	311166	37961	0.112
30700	Creatinine	CREA	311004	37938	0.181
30710	C-reactive protein	CRP	310483	37874	0.177
30720	Cystatin C	CYSC	311134	37959	0.211
30730	Gamma glutamyltransferase	GGT	311007	37935	0.187
30740	Glucose	GLU	284675	34709	0.078
30750	Glycated haemoglobin (HbA1c)	HbA1c	311027	37949	0.183
30760	HDL cholesterol	HDL	284853	34753	0.238
30770	IGF-1	IGF-1	309486	37767	0.209
30780	LDL direct	LDL	310574	37876	0.101
30790	Lipoprotein A	LP-a	247437	30660	0.078
30800	Oestradiol	OEST	48805	6997	0.018
30810	Phosphate	PHOS	284431	34689	0.103
30820	Rheumatoid factor	RF	27778	3214	0.006
30830	SHBG	SHBG	282241	34403	0.203
30840	Total bilirubin	TBIL	309868	37799	0.143
30850	Testosterone	TEST	282219	34267	0.073
30860	Total protein	TP	284668	34741	0.138
30870	Triglycerides	TG	310908	37921	0.183
30880	Urate	UA	310756	37921	0.175
30890	Vitamin D	VITD	297470	36374	0.094
48	Waist circumference	WC	325825	39714	0.193
49	Hip circumference	HC	325788	39710	0.202
50	Standing height	H	325667	39699	0.416

Table S1: **Sample sizes of LR-PRS/QR-PRS training and test datasets and SNP heritability for each of 64 traits in UK Biobank.** SNP heritability for each trait was estimated using LDSC.

UKB field	Abbr.	Certain-above t									Certain-below t								
		QR-PRS			CQR-PRS			LR-PRS			QR-PRS			CQR-PRS			LR-PRS		
		N	Fraction	% Trait above t	N	Fraction	% Trait above t	N	Fraction	% Trait above t	N	Fraction	% Trait below t	N	Fraction	% Trait below t	N	Fraction	% Trait below t
30070	RDW	264	6.8%	25.4%	41	1.1%	31.7%	0	0.0%	0.0%	13226	38.1%	92.8%	6374	18.4%	93.5%	1682	4.8%	95.8%
30610	ALP	239	6.3%	33.5%	11	0.3%	63.6%	1	0.0%	100.0%	13119	38.4%	94.3%	2820	8.3%	96.6%	2897	8.5%	97.2%
30840	TBIL	167	4.4%	48.5%	2	0.1%	50.0%	9	0.2%	88.9%	17706	52.0%	96.9%	5093	15.0%	98.7%	12451	36.6%	98.6%
30250	RET#	152	4.0%	32.9%	115	3.0%	32.2%	0	0.0%	0.0%	4471	13.1%	94.6%	3909	11.5%	94.5%	1496	4.4%	97.1%
50	H	152	3.8%	35.5%	8	0.2%	25.0%	13	0.3%	38.5%	19268	53.9%	96.3%	6018	16.8%	98.5%	7499	21.0%	98.4%
30750	HbA1c	132	3.5%	26.5%	13	0.3%	15.4%	0	0.0%	0.0%	17652	51.7%	93.7%	9584	28.1%	95.0%	1561	4.6%	96.7%
30100	MPV	128	3.3%	43.8%	28	0.7%	57.1%	10	0.3%	80.0%	15258	43.9%	96.3%	9341	26.9%	97.2%	4222	12.2%	98.2%
30040	MCV	108	2.8%	32.4%	38	1.0%	23.7%	4	0.1%	50.0%	16379	47.2%	94.6%	11582	33.3%	95.5%	3022	8.7%	98.0%
30710	CRP	105	2.8%	28.6%	55	1.5%	27.3%	1	0.0%	0.0%	14011	41.1%	93.4%	10503	30.8%	93.7%	2369	7.0%	96.7%
30260	MRV	101	2.7%	32.7%	32	0.8%	34.4%	0	0.0%	0.0%	16082	47.2%	94.1%	10670	31.3%	94.9%	1453	4.3%	98.3%
30190	MO%	101	2.6%	32.7%	38	1.0%	31.6%	0	0.0%	0.0%	6070	17.5%	95.6%	4418	12.7%	95.7%	1230	3.5%	98.0%
30240	RET%	99	2.6%	29.3%	37	1.0%	35.1%	1	0.0%	0.0%	5785	17.0%	94.5%	3739	11.0%	94.7%	1335	3.9%	97.2%
30210	EO%	100	2.6%	32.0%	58	1.5%	34.5%	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	1201	3.5%	97.3%
30000	WBC	99	2.6%	31.3%	17	0.4%	41.2%	1	0.0%	0.0%	12880	37.1%	93.6%	6621	19.1%	94.6%	1566	4.5%	97.1%
30760	HDL	89	2.6%	29.2%	17	0.5%	41.2%	0	0.0%	0.0%	10936	35.0%	94.9%	6093	19.5%	95.7%	2489	8.0%	98.2%
30270	MSCV	93	2.5%	35.5%	7	0.2%	28.6%	0	0.0%	0.0%	12781	37.5%	94.9%	6410	18.8%	95.8%	1540	4.5%	97.9%
30770	IGF-1	86	2.3%	38.4%	10	0.3%	50.0%	1	0.0%	0.0%	13621	40.1%	94.8%	6456	19.0%	96.2%	1813	5.3%	98.3%
30290	HLR%	85	2.2%	37.6%	45	1.2%	51.1%	0	0.0%	0.0%	11547	33.9%	94.2%	8763	25.7%	94.8%	1642	4.8%	98.1%
30860	TP	78	2.2%	32.1%	55	1.6%	32.7%	0	0.0%	0.0%	6850	21.9%	94.1%	5506	17.6%	94.6%	604	1.9%	97.7%
30620	ALT	85	2.2%	24.7%	6	0.2%	16.7%	0	0.0%	0.0%	8569	25.1%	93.0%	2917	8.5%	94.2%	313	0.9%	97.4%
30640	APOB	81	2.1%	33.3%	2	0.1%	0.0%	0	0.0%	0.0%	9703	28.5%	95.1%	2633	7.7%	96.9%	3045	9.0%	98.2%
20015	SH	84	2.1%	28.6%	0	0.0%	0.0%	5	0.1%	20.0%	19658	55.1%	94.7%	6637	18.6%	96.7%	4428	12.4%	97.5%
30660	DBIL	64	2.0%	53.1%	1	0.0%	0.0%	3	0.1%	66.7%	0	0.0%	0.0%	0	0.0%	0.0%	8871	30.4%	97.4%
30630	APOA1	62	1.8%	33.9%	13	0.4%	30.8%	2	0.1%	0.0%	13569	43.6%	94.2%	8695	28.0%	94.8%	1716	5.5%	97.6%
30110	PDW	61	1.6%	36.1%	12	0.3%	25.0%	1	0.0%	100.0%	6791	19.6%	95.7%	3257	9.4%	96.7%	2369	6.8%	98.1%
30690	CHOL	59	1.6%	23.7%	9	0.2%	44.4%	0	0.0%	0.0%	13955	40.8%	94.4%	8128	23.8%	95.5%	1218	3.6%	98.0%
30600	ALB	50	1.4%	30.0%	49	1.4%	32.7%	0	0.0%	0.0%	2954	9.4%	94.1%	3217	10.3%	94.2%	400	1.3%	96.8%
30720	CYSC	48	1.3%	27.1%	46	1.2%	26.1%	1	0.0%	0.0%	6206	18.2%	94.1%	6078	17.8%	94.1%	2256	6.6%	97.5%
21002	W	50	1.3%	42.0%	103	2.6%	36.9%	2	0.1%	100.0%	1188	3.3%	96.0%	1826	5.1%	95.5%	2053	5.8%	97.8%
30830	SHBG	43	1.2%	32.6%	49	1.4%	32.7%	1	0.0%	100.0%	5410	17.5%	94.5%	5837	18.9%	94.6%	2050	6.6%	97.0%
30140	NE#	43	1.1%	34.9%	8	0.2%	37.5%	0	0.0%	0.0%	7122	20.5%	93.7%	4275	12.3%	94.4%	945	2.7%	96.3%
30050	MCH	42	1.1%	21.4%	5	0.1%	20.0%	1	0.0%	100.0%	21916	63.1%	94.3%	14363	41.4%	95.4%	3297	9.5%	97.8%
30060	MCHC	31	0.8%	16.1%	3	0.1%	0.0%	0	0.0%	0.0%	10990	31.6%	92.5%	4673	13.5%	93.4%	78	0.2%	97.4%
30700	CREA	30	0.8%	23.3%	28	0.7%	25.0%	0	0.0%	0.0%	5474	16.0%	93.1%	5171	15.1%	93.2%	1120	3.3%	95.8%
30030	HCT	28	0.7%	21.4%	24	0.6%	25.0%	0	0.0%	0.0%	6462	18.6%	93.8%	6192	17.8%	93.8%	921	2.7%	97.0%
30080	PLT	26	0.7%	38.5%	18	0.5%	44.4%	2	0.1%	0.0%	11019	31.7%	96.1%	10124	29.1%	96.3%	2732	7.9%	98.3%
30740	GLU	21	0.6%	14.3%	64	1.8%	14.1%	0	0.0%	0.0%	2422	7.8%	93.4%	3828	12.3%	93.2%	55	0.2%	94.5%
30670	UREA	21	0.6%	33.3%	11	0.3%	36.4%	0	0.0%	0.0%	6388	18.7%	92.8%	5091	14.9%	93.2%	239	0.7%	96.2%
30780	LDL	20	0.5%	20.0%	5	0.1%	40.0%	0	0.0%	0.0%	5070	14.9%	95.2%	3441	10.1%	95.9%	1953	5.7%	98.1%
30010	RBC	15	0.4%	33.3%	5	0.1%	20.0%	0	0.0%	0.0%	14423	41.5%	93.9%	8507	24.5%	95.3%	1902	5.5%	98.2%
30280	IRF	13	0.3%	46.2%	1	0.0%	100.0%	0	0.0%	0.0%	12313	36.2%	94.6%	8437	24.8%	95.2%	650	1.9%	97.4%
30890	VITD	12	0.3%	25.0%	34	0.9%	26.5%	0	0.0%	0.0%	3963	12.1%	95.0%	5215	15.9%	94.7%	356	1.1%	97.5%
30810	PHOS	9	0.3%	44.4%	5	0.1%	40.0%	0	0.0%	0.0%	3359	10.8%	93.7%	2747	8.8%	93.6%	290	0.9%	97.9%
30020	HGB	10	0.3%	10.0%	5	0.1%	0.0%	0	0.0%	0.0%	12309	35.4%	93.9%	10996	31.7%	94.2%	1083	3.1%	97.1%
30680	CALD	9	0.3%	22.2%	50	1.4%	28.0%	1	0.0%	100.0%	281	0.9%	93.6%	1305	4.2%	93.0%	299	1.0%	95.0%
30880	UA	9	0.2%	22.2%	5	0.1%	0.0%	0	0.0%	0.0%	10794	31.6%	93.6%	8037	23.5%	94.2%	2415	7.1%	97.6%
21001	BMI	9	0.2%	44.4%	5	0.1%	20.0%	0	0.0%	0.0%	18648	52.3%	94.3%	16216	45.5%	94.6%	2089	5.9%	98.0%
30850	TEST	6	0.2%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	11292	36.6%	92.0%	98	0.3%	98.0%	41	0.1%	95.1%
30120	LY#	5	0.1%	60.0%	3	0.1%	100.0%	1	0.0%	0.0%	16588	47.9%	93.9%	14054	40.5%	94.2%	1560	4.5%	97.1%
48	WC	5	0.1%	40.0%	2	0.1%	50.0%	0	0.0%	0.0%	9916	27.7%	94.0%	7905	22.1%	94.4%	1032	2.9%	97.5%
30790	LP-a	3	0.1%	0.0%	10	0.3%	20.0%	499	16.3%	50.5%	7882	28.6%	95.5%	9847	35.7%	95.4%	19891	72.1%	96.7%
30200	NE%	3	0.1%	33.3%	24	0.6%	25.0%	1	0.0%	0.0%	6011	17.3%	92.9%	10090	29.1%	92.7%	562	1.6%	95.7%
30650	AST	2	0.1%	50.0%	2	0.1%	50.0%	0	0.0%	0.0%	11659	34.3%	93.3%	10140	29.8%	93.6%	556	1.6%	95.9%
30870	TG	2	0.1%	50.0%	3	0.1%	66.7%	2	0.1%	50.0%	15856	46.5%	94.1%	16017	46.9%	94.1%	1401	4.1%	98.0%
30090	PCT	2	0.1%	50.0%	3	0.1%	33.3%	1	0.0%	0.0%	11748	33.8%	95.1%	12774	36.8%	94.9%	2407	6.9%	97.9%
30730	GGT	1	0.0%	0.0%	1	0.0%	0.0%	0	0.0%	0.0%	7857	23.0%	94.8%	11623	34.0%	94.5%	2157	6.3%	97.6%
30180	LY%	1	0.0%	0.0%	4	0.1%	25.0%	0	0.0%	0.0%	5227	15.1%	95.3%	10374	29.9%	94.5%	728	2.1%	97.3%
30130	MO#	0	0.0%	0.0%	0	0.0%	0.0%	1	0.0%	100.0%	13146	37.9%	96.6%	19138	55.2%	95.8%	1562	4.5%	99.0%
30300	HLR#	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	3726	10.9%	97.2%	27557	80.9%	93.5%	1655	4.9%	98.5%
30150	EO#	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	191	0.6%	98.4%	5000	14.4%	94.6%	1315	3.8%	97.6%
30800	OEST	0	0.0%	0.0%	2	0.3%	0.0%	0	0.0%	0.0%	5	0.1%	100.0%	195	3.1%	92.3%	0	0.0%	0.0%
30220	BA%	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	7	0.0%	100.0%	13557	39.1%	92.9%	12	0.0%	100.0%
30820	RF	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	1	0.0%	100.0%	0	0.0%	0.0%
49	HC	0	0.0%	0.0%	2	0.1%	50.0%	0	0.0%	0.0%	0	0.0%	0.0%	2614	7.3%	96.6%	1142	3.2%	97.9%

Table S2: Number and proportion of certain-above/-below *t* individuals based on QR-PRS, CQR-PRS and LR-PRS across 64 traits in UKBB. The threshold *t* corresponds to the 90th $\text{QR-PRS}_{0.5}$ or LR-PRS percentile, respectively. N is the number of individuals in the certain-above or -below group. Fraction is the proportion of certain-above/-below *t* among the individuals above/below *t*. %Trait above/below *t* is the proportion of individuals with the actual trait value above/below top 10% of the corresponding trait distribution in the certain-above/-below *t* category.

UKB field	Abbr.	Certain-above t									Certain-below t								
		QR-PRS			CQR-PRS			LR-PRS			QR-PRS			CQR-PRS			LR-PRS		
		N	Fraction	% Trait above t	N	Fraction	% Trait above t	N	Fraction	% Trait above t	N	Fraction	% Trait below t	N	Fraction	% Trait below t	N	Fraction	% Trait below t
30750	HbA1c	6	1.7%	16.7%	0	0.0%	0.0%	0	0.0%	0.0%	1476	46.8%	93.9%	670	21.2%	93.9%	98	3.1%	96.9%
50	H	10	1.6%	20.0%	0	0.0%	0.0%	1	0.2%	0.0%	2546	45.9%	95.7%	469	8.5%	98.1%	614	11.1%	98.2%
30190	MO%	9	1.5%	33.3%	0	0.0%	0.0%	0	0.0%	0.0%	1238	22.6%	93.9%	850	15.5%	94.4%	138	2.5%	96.4%
30840	TBIL	8	1.3%	62.5%	0	0.0%	0.0%	0	0.0%	0.0%	2686	48.5%	95.2%	752	13.6%	96.8%	1788	32.3%	96.6%
30210	EO%	7	1.2%	42.9%	3	0.5%	33.3%	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	139	2.6%	95.0%
30710	CRP	7	1.1%	14.3%	4	0.7%	0.0%	0	0.0%	0.0%	1863	34.0%	91.5%	1266	23.1%	92.4%	220	4.0%	96.4%
30760	HDL	7	1.1%	28.6%	1	0.2%	0.0%	0	0.0%	0.0%	1248	22.5%	93.8%	535	9.7%	95.7%	312	5.6%	98.4%
30740	GLU	7	1.1%	28.6%	14	2.3%	14.3%	0	0.0%	0.0%	194	3.5%	93.3%	375	6.8%	92.8%	4	0.1%	100.0%
30040	MCV	6	1.0%	16.7%	0	0.0%	0.0%	0	0.0%	0.0%	2525	45.7%	93.7%	1643	29.7%	94.4%	359	6.5%	96.4%
30620	ALT	6	1.0%	16.7%	1	0.2%	0.0%	0	0.0%	0.0%	1376	24.8%	92.4%	382	6.9%	92.4%	26	0.5%	88.5%
30010	RBC	5	0.8%	20.0%	0	0.0%	0.0%	0	0.0%	0.0%	1318	23.9%	93.4%	660	12.0%	94.2%	196	3.5%	96.9%
30700	CREA	5	0.8%	20.0%	4	0.6%	25.0%	0	0.0%	0.0%	385	7.0%	90.9%	347	6.3%	90.5%	137	2.5%	97.1%
21002	W	4	0.6%	25.0%	11	1.8%	27.3%	0	0.0%	0.0%	169	3.1%	95.3%	248	4.5%	94.8%	215	3.9%	97.2%
30000	WBC	3	0.5%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	1876	33.9%	93.9%	805	14.6%	94.9%	174	3.1%	96.0%
30670	UREA	3	0.5%	0.0%	2	0.3%	0.0%	0	0.0%	0.0%	1152	20.8%	92.6%	898	16.2%	92.5%	18	0.3%	94.4%
30020	HGB	2	0.3%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	1291	23.4%	95.4%	1099	19.9%	95.4%	110	2.0%	95.5%
30200	NE%	1	0.2%	100.0%	3	0.5%	100.0%	0	0.0%	0.0%	856	15.6%	93.2%	1555	28.3%	92.5%	64	1.2%	96.9%
30690	CHOL	1	0.2%	0.0%	1	0.2%	0.0%	0	0.0%	0.0%	1772	32.0%	93.9%	797	14.4%	95.6%	94	1.7%	95.7%
30870	TG	1	0.2%	0.0%	1	0.2%	0.0%	0	0.0%	0.0%	1728	31.2%	95.0%	1771	32.0%	95.3%	145	2.6%	97.2%
48	WC	1	0.2%	100.0%	0	0.0%	0.0%	0	0.0%	0.0%	865	15.6%	93.8%	620	11.2%	93.7%	121	2.2%	99.2%
21001	BMI	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	2308	41.7%	94.5%	1867	33.7%	94.9%	176	3.2%	98.3%
30080	PLT	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	1505	27.2%	96.5%	1342	24.3%	96.7%	388	7.0%	97.4%
30880	UA	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	1374	24.8%	93.0%	922	16.6%	93.0%	235	4.2%	94.0%
30650	AST	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	1305	23.6%	91.3%	1084	19.6%	91.3%	42	0.8%	95.2%
30730	GGT	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	678	12.2%	92.0%	1200	21.7%	92.6%	222	4.0%	97.7%
30180	LY%	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	624	11.4%	95.4%	1407	25.6%	93.8%	72	1.3%	98.6%
30220	BA%	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	1444	26.9%	95.5%	0	0.0%	0.0%
49	HC	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	0	0.0%	0.0%	640	11.6%	96.1%	131	2.4%	100.0%

Table S3: Number and proportion of certain-above/-below *t* individuals based on QR-PRS, CQR-PRS and LR-PRS across 28 traits in ProgeNIA/SardiNIA. The threshold *t* corresponds to the 90th $\text{QR-PRS}_{0.5}$ or LR-PRS percentile, respectively. N is the number of individuals in the certain-above or -below group. Fraction is the proportion of certain-above/-below *t* among the individuals above/below *t*. %Trait above/below *t* is the proportion of individuals with the actual trait value above/below top 10% of the corresponding trait distribution in the certain-above/-below *t* category.

Phenotype	Abbreviation	Phecode	Training set			Test set		
			N	N cases	N controls	N	N cases	N controls
Alcoholic cirrhosis	AC	317.11	305491	801	304690	37048	134	36914
Angina	Angina	411.3	313922	16429	297493	38385	1744	36641
Chronic kidney disease	CKD	585	322691	10541	312150	39391	1161	38230
Cholecystitis	Cholecystitis	574	324908	14646	310262	39667	1545	38122
Gallstones	Gallstones	574.1	323540	13278	310262	39493	1371	38122
Gout	Gout	274.1	326031	3482	322549	39792	415	39377
Heart failure	HF	428	325613	5704	319909	39749	630	39119
Hypertension	HTN	401.1	325643	72483	253160	39764	7842	31922
Myocardial infarction	MI	411.2	308697	11204	297493	37857	1216	36641
Type 2 diabetes	T2D	250.2	325178	17445	307733	39710	2026	37684

Table S4: Disease phenotypes and sample sizes in the training and test datasets for the meta-PRS model.