Hypothesis test of specific parametric structure in a generalized additive model

Yihe Yang* and Xiaofeng Zhu
Department of Population and Quantitative Health Sciences,
Case Western Reserve University School of Medicine

*Corresponding Author: yxy1234@case.edu

September 10, 2025

Abstract

Before applying flexible nonparametric models such as a generalized additive model (GAM), it is natural to ask whether a simpler parametric form suffices. To address this question, we develop TAPS (Test for Arbitrary Parametric Structure), a framework that integrates estimation and hypothesis testing to evaluate whether a prespecified parametric form adequately captures a covariate effect in a GAM. TAPS accommodates diverse structures, including linearity, piecewise linearity with changepoints, discontinuities with jumps, among many others. It is implemented in the R package `mgcv.taps` built directly on `mgcv`, enabling seamless adoption, broad outcome support, and scalability to biobank-scale data. Using UK Biobank data, we analyze 38 continuous and 8 binary traits to investigate two scientific questions: does the effect of a polygenic risk score (PRS) vary with age beyond a linear interaction, and does retirement at age 65 modify this age-varying effect? We find that age-varying PRS effects are common and often strongly non-linear, and that retirement at 65 significantly modifies these effects for five traits after multiple-testing correction.

Keywords: Generalized additive model; gene–environment interaction; hypothesis test; polygenic risk score; vary-coefficient model;

1 Introduction

Generalized additive model (GAM) (Wood 2017), an extension of the generalized linear model (GLM) (Nelder & Wedderburn 1972), is widely used to model the non-linear effects of covariates on an outcome. Since its first introduction by Hastie & Tibshirani (1990), GAM has undergone substantial methodological and computational development. For instance, the R package `mgcv` exemplifies these advances, which provides a broad class of splines (Wood 2003, 2006) and multiple smoothing parameter selection methods (Wood 2004, 2011), and it has become the de facto standard software for GAM estimation and inference. Beyond methodological innovations, GAMs have also found wide applicability across scientific disciplines. For example, in epidemiology, GAMs have been used to smooth daily incidence data and to reveal trends in childhood type 1 diabetes during the COVID-19 pandemic (McKeigue et al. 2023). In ecology, GAMs have been employed to model the nonlinear effects of environmental factors in chaotic or near-chaotic ecological dynamic systems (Wood 2010). In spatial transcriptomics, GAMs are used to characterize the spatial distribution of cell types across tissues (Shang et al. 2025).

Before applying a nonparametric model such as a GAM, it is natural to ask whether a simpler parametric model can already describe the variation in the data. This is because a parametric model, if correctly specified, is often of greater scientific interest due to its interpretability and simplicity. For example, in genetics, the traditional assumption is that genetic effects of multiple variants across the genome are linear and additive (Hill et al. 2008), but growing evidence of non-linear genetic effects challenges this view (Zuk et al. 2012), sparking debate over whether linear additive models adequately capture the heritability of complex traits. In addition, in gene–environment (G×E) interaction analyses (Zhu et al. 2024, Winkler et al. 2024), it is usually assumed that a simple linear interaction model holds, meaning the genetic effect changes linearly with the environmental variable,

an assumption that can be overly strong and potentially incorrect in practice.

Compared to the wide variety of estimation methods available for GAMs, the tools for hypothesis testing remain relatively limited. Most existing hypothesis test methods focus on assessing whether a smooth function is null (i.e., identically zero) (Wood 2013), rather than whether it conforms to a specific parametric structure of scientific interest. For spline-based methods, hypothesis testing of specific parametric structures is largely restricted to evaluating whether a smooth term follows a polynomial form (Zhang & Lin 2003, Crainiceanu et al. 2005, Nummi et al. 2011). In the context of kernel-based methods, combined estimators have been developed to accommodate arbitrary parametric structures (Olkin & Spiegelman 1987, Fan & Ullah 1999, Mazo & Portier 2021), but extending them to GAMs with multiple smooth terms or non-Gaussian responses remains computationally challenging. To the best of our knowledge, no existing method enables hypothesis testing of arbitrary parametric structures within a GAM, especially in the presence of multiple smooth terms.

We develop the Test for Arbitrary Parametric Structure (TAPS) to assess whether a user-specified parametric structure adequately describes the relationship between a covariate and the response in a GAM. The methodological innovation of TAPS lies in constructing a mixed-effect representation in which the nonparametric component is forced to be orthogonal to any prespecified parametric form, thereby enabling rigorous assessment of whether the parametric structure adequately captures the target function within a GAM. Especially, we develop an R package, `mgcv.taps`, which integrates seamlessly with `mgcv`, one of the most powerful GAM software. This integration enables TAPS to be easily adopted without additional learning cost, to support a wide range of outcome types including survival and ordinal responses (Wood et al. 2016), and to scale efficiently to biobank-scale datasets (Wood et al. 2017). We analyze UK Biobank data (Bycroft et al. 2018) to address two questions: (i) whether the effects of polygenic risk scores (PRSs) on complex traits vary with age beyond

a simple linear interaction, and (ii) whether reaching retirement age at 65 coincides with a structural change in these age-varying effects. Applying TAPS to 38 continuous and 8 binary traits, we found that non-linear age-varying effects of PRSs are widespread, and that five traits exhibit significant trend changes at age 65 after multiple-testing correction.

2 Data and Scientific Questions

2.1 UK Biobank data

The UK Biobank is a large prospective cohort of ∼500,000 individuals aged 40–70 years recruited across the United Kingdom between 2006 and 2010 (Sudlow et al. 2015). Participants completed extensive baseline assessments, including laboratory biomarker assays, standardized physical measurements, and touchscreen questionnaires on lifestyle, diet, and medical history, supplemented by interviews and linkage to electronic health records. Genotypes were phased with SHAPEIT3 and imputed using IMPUTE4 against a combined Haplotype Reference Consortium, UK10K, and 1000 Genomes Phase 3 reference panel, yielding ∼96 million variants (Bycroft et al. 2018). With its scale, harmonized phenotyping, and dense genomic coverage, the UK Biobank is now a cornerstone resource for genetic epidemiology and G×E interaction studies, with ongoing expansions in repeat assessments, multi-omics profiling, and whole-genome sequencing (Abdellaoui et al. 2023).

2.2 Age-varying effect of PRS

The genetic additive model, first formalized by Fisher (1919), is the canonical framework for modeling genotype–phenotype relationships due to its interpretability, parsimony, and practical utility. It can be written as

$$y_i = \sum_{j=1}^{p} g_{ij}\beta_j + \epsilon_i, \tag{1}$$

where y_i is the phenotype of individual i, g_{ij} is the genotype coding for variant j, β_j is its additive effect size, and ϵ_i is a residual error term. Polygenic effects refer to the collection of effect sizes $\beta_1, \ldots, \beta_p$, typically with p large and each β_j small, so that the genetic contribution arises from the aggregation of many weak effects. The linear predictor $\sum_{j=1}^{p} g_{ij}\beta_j$ is then defined as a PRS, representing the linearly cumulative polygenic effects across variants (Zheng et al. 2024).

However, the adequacy of this genetic additive model has been debated. While many studies find that additive effects explain most genetic variance for complex traits, with non-additive effects contributing relatively little at the population level (Hill et al. 2008), other evidence points to systematic departures from additivity, including non-linear genotype–phenotype relationships, dominance, epistasis, and G×E interactions (Zuk et al. 2012). A common specification for G×E interaction is the following linear interaction model:

$$y_i = \gamma x_i + \sum_{j=1}^{p} (\beta_j g_{ij} + \theta_j [g_{ij} \times x_i]) + \epsilon_i, \tag{2}$$

where x_i is an environmental variable, γ is the environmental effect, β_j represents the main effect of variant j, and θ_j its interaction effect with x_i. Because individual variant effects are generally very small, single-variant interaction is very difficult to detect. An alternative approach (Aschard 2016) aggregates multiple variants into a PRS and tests for a universal interaction parameter θ_0 in the model below:

$$y_i = \gamma x_i + \sum_{j=1}^{p} g_{ij}\beta_j + \theta_0 \left[\left(\sum_{j=1}^{p} \beta_j g_{ij} \right) \times x_i \right] + \epsilon_i. \tag{3}$$

Nevertheless, genetic effects have been shown to exhibit context-specific heterogeneity, where their interactions with an individual's environmental or physiological state can depart from linearity (Dahl et al. 2020). For example, Tian et al. (2025) demonstrated that alcohol intake exerts a causal effect on blood pressure only beyond a certain threshold. Hottenga et al. (2005) reported that the genetic effects of certain variants on systolic and diastolic

blood pressure are stratified by age. Motivated by prior evidence of context-stratified genetic effects, our first scientific question is:

> *Among all individuals in the UK Biobank, does the effect of a PRS for a complex trait vary with age, and if so, is this age-varying effect linear or nonlinear?*

Mathematically, we consider the model

$$y_i = \gamma x_i + \sum_{j=1}^{p} g_{ij}\beta_j + \left(\sum_{j=1}^{p} \beta_j g_{ij}\right) \times \theta(x_i) + \epsilon_i, \tag{4}$$

where x_i denotes age and $\theta(\cdot)$ is a smooth function. In the literature, this model is known as the varying coefficient model (Hastie & Tibshirani 1993). Assessing whether $\theta(\cdot)$ follows a simple linear form or a more complex pattern naturally leads to the hypothesis test

$$\mathrm{H}_0 : \theta(x) = \theta_0 x \quad \text{vs.} \quad \mathrm{H}_1 : \theta(x) \text{ is an unrestricted smooth function.} \tag{5}$$

In practice, we will first test whether θ_0 is significantly different from zero and then assess if H_0 is rejected or not, meaning that there is additional variation beyond the linear interaction. As will be introduced later, we develop a new hypothesis testing framework that allows the significance test of θ_0 and the test in (5) to be carried out independently. To our best knowledge, while previous work such as Jiang et al. (2021) has examined age-dependent changes by detecting single variants whose marginal genetic effects vary across age strata, our study takes the first prediction-oriented perspective, asking whether the cumulative genetic effect captured by a PRS departs from a linear interaction with age.

2.3 Retirement-stratified age-varying effect of PRS

A growing number of studies show that retirement affects multiple aspects of late-life health. For example, exploiting the United Kingdom's state pension age as a natural experiment, Furuya & Fletcher (2024) found that eligibility-induced retirement accelerated biological aging by about 0.87 years. In addition, Eibich (2015) reported that retirement improved

self-rated health, reduced stress, and promoted healthier lifestyles in Germany. These studies suggest that retirement is a pivotal life-course event that can alter both biological and functional dimensions of aging.

Following our first one, we secondly investigate the following scientific question:

> *Among individuals near the retirement threshold in the UK Biobank, does the age-varying effect of a PRS exhibit a jump or a change-in-slope at age 65?*

Here, "near the retirement threshold" refers to UK Biobank participants aged $63 < \text{age} < 67$, and "retirement" is defined as an indicator taking the value 1 if age ≥ 65 and 0 otherwise, based on the United Kingdom's state pension age. Mathematically, we consider two commonly-used parametric forms in regression discontinuity design (RDD) and regression kink design (RKD) (Huntington-Klein 2021): a linear jump structure

$$\theta(x) = \alpha_0 + \alpha_1 x + \alpha_J \, \mathbf{1}\{x \geq 65\} + \alpha_K \, (x - 65)_+, \tag{6}$$

and a linear change-of-slope structure

$$\theta(x) = \alpha_0 + \alpha_1 x + \alpha_K \, (x - 65)_+, \tag{7}$$

where α_0, α_1 describe the baseline linear trend, α_J is the jump parameter, and α_K is the change-of-slope parameter; $\mathbf{1}\{x \geq 65\}$ is the indicator function (1 if $x \geq 65$, 0 otherwise), and $(x - 65)_+ = \max(x - 65, 0)$ is the truncated linear function. We test whether $\theta(x)$ can be sufficiently described by the structure (6) or (7) by using the following hypothesis tests:

$$\mathrm{H}_0^{(\text{jump})} : \ \theta(x) \in \mathcal{H}_{\text{jump}} \quad \text{vs.} \quad \mathrm{H}_1 : \theta(x) \text{ is an unrestricted smooth function.} \tag{8}$$

$$\mathrm{H}_0^{(\text{kink})} : \ \theta(x) \in \mathcal{H}_{\text{kink}} \quad \text{vs.} \quad \mathrm{H}_1 : \theta(x) \text{ is an unrestricted smooth function.} \tag{9}$$

where $\mathcal{H}_{\text{jump}}$ and $\mathcal{H}_{\text{kink}}$ are the parametric spaces spanned by the basis functions $(1, x, \mathbf{1}\{x \geq 65\})$ and $(1, x, (x - 65)_+)$, respectively. We first assess whether the structural parameter α_J in (6) or α_K in (7) is statistically significant. If the parameter is significant and the

null hypothesis $\mathrm{H}_0^{(\mathrm{jump})}$ or $\mathrm{H}_0^{(\mathrm{kink})}$ is not rejected, we infer that retirement is associated with a structural change at age 65. If the parameter is significant but $\mathrm{H}_0^{(\mathrm{jump})}$ or $\mathrm{H}_0^{(\mathrm{kink})}$ is rejected, the age-varying effect deviates from the simple jump or kink form, suggesting a more complex non-linear pattern. Conversely, if α_J or α_K is not significant, there is no evidence of a PRS–retirement interaction at age 65, regardless of whether $\mathrm{H}_0^{(\mathrm{jump})}$ or $\mathrm{H}_0^{(\mathrm{kink})}$ is rejected. As will be introduced later, our new hypothesis testing framework ensures that the test for the parametric components is independent of the tests in (8) and (9).

Note that this formulation differs from standard RDD and RKD, which will directly test whether retirement alters a biomarker (e.g., a depression score). Here, our target is the structure of the age-varying PRS effects, and a detected structural change is not interpreted as the causal effect of retirement (Huntington-Klein 2021). We further restrict to individuals near the retirement threshold so that any detected change reflects local differences around the policy cutoff rather than long-term age trends (Huntington-Klein 2021).

3 Method

3.1 Overview of TAPS

To address the two scientific questions in Section 2, we need a framework that (i) flexibly captures non-linear relationships between the PRS and age and (ii) formally tests whether this relationship conforms to prespecified parametric structures, such as linearity or structural breaks at the retirement threshold. The GAM provides a natural basis, enabling smooth, data-driven estimation of age-varying genetic effects while accommodating other covariates parametrically (Wood 2017). Building upon GAM, our proposed TAPS tests whether a smooth effect can be adequately represented by a given parametric structure.

Specifically, the model of a GAM is

$$y_i \sim \mathcal{Y}(\mu_i, \phi_0), \quad g(\mu_i) = \eta_i = \boldsymbol{Z}_i^\top \boldsymbol{\gamma} + f_1(x_{i1}) + \cdots + f_J(x_{iJ}), \tag{10}$$

where y_i is the response belonging to a certain distribution $\mathcal{Y}(\mu_i, \phi_0)$ with mean μ_i and dispersion parameter ϕ_0, $g(\cdot)$ is a known link function, η_i is a linear predictor, $\boldsymbol{Z}_i$ is a vector of strict linear covariates, $\boldsymbol{\gamma}$ is a regression coefficient vector, and $f_j(\cdot)$ is a smooth function of covariate x_{ij}, $j = 1, \ldots, J$. TAPS performs a hypothesis test on whether a target function in (10), such as $f_1(\cdot)$, adheres to a parametric structure. The null and alternative hypotheses that TAPS addresses are

$$\mathrm{H}_0 : f_1 \in \mathcal{H}_A \qquad \text{v.s.} \qquad \mathrm{H}_1 : f_1 \in \mathcal{H} \text{ and } f_1 \notin \mathcal{H}_A, \tag{11}$$

where $\mathcal{H}_A$ is the space of functions with the target parametric structure and $\mathcal{H}$ is a general functional space. In particular, $\mathcal{H}$ should be sufficiently large to accommodate the parametric functions in $\mathcal{H}_A$ (which may even be nonsmooth), but also capable of approximating smooth functions when $\mathcal{H}_A$ does not reflect the true structure.

To achieve this, we propose a new mixed-effects representation of $f_1(x)$:

$$f_1(x_1) = \sum_{s=1}^{q} A_{1s}(x_1)\alpha_{1s} + b_1(x_1), \tag{12}$$

where $\boldsymbol{A}_1(x_1) = (A_{11}(x_1), \ldots, A_{1q}(x_1))^\top$ is a vector of the basis functions of the parametric structure, $\boldsymbol{\alpha}_1 = (\alpha_{11}, \ldots, \alpha_{1q})^\top$ is a fixed vector, and $b_1(x_1)$ is a non-parametric term orthogonal to $\sum_{s=1}^{q} A_{1s}(x_1)\alpha_{1s}$. In this paper, we assume $b_1(x_1)$ to lie in a reproducing kernel Hilbert space (RKHS) and subject to this orthogonality constraint. According to the equivalence between a smooth function in an RKHS and a Gaussian process in L_2 space (Kanagawa et al. 2018), $b_1(x)$ can be regarded as a Gaussian process with a covariance function

$$\mathrm{cov}(b_1(x_{i1}), b_1(x_{j1})) = \phi_1 K_1(x_{i1}, x_{j1}), \tag{13}$$

where $K_1(x_i, x_j)$ is a known bivariate function, and $\phi_1 \geq 0$ is the variance of this Gaussian process. We call the equation (12) a mixed-effect representation of a function where $\sum_{s=1}^{q} A_{1s}(x_1)\alpha_{1s}$ is the fixed effect and $b_1(x_1)$ is the random effect, as it shares the same form as the generalized linear mixed model (GLMM) (Breslow & Clayton 1993). As a result, the hypothesis (11) reduces to

$$\mathrm{H}_0 : \phi_1 = 0 \qquad \text{v.s.} \qquad \mathrm{H}_1 : \phi_1 > 0. \tag{14}$$

The techniques of GLMM, including the Wald test (Wood 2013) and score test (Zhang & Lin 2003), can yield the p-value of (14). Moreover, since the parametric (fixed effect) and nonparametric (random effect) components are orthogonal, we can also directly test the significance of the coefficients in the fixed-effect part, addressing whether the prespecified parametric structure itself is supported by the data.

While the above formulation considers testing for a univariate smooth function $f_1(x)$, the same idea extends naturally to varying-coefficient models:

$$g(\mu_i) = \eta_i = \boldsymbol{Z}_i^\top \boldsymbol{\gamma} + x_{i2} \times \theta_1(x_{i1}) + f_3(x_{i3}) + x_{i5} \times \theta_4(x_{i4}) + \cdots + f_J(x_{iJ}) \tag{15}$$

where the function $\theta_1(x_1)$ in an interaction term is the target of testing. Specifically, TAPS can also evaluate whether there is a complex interaction between two covariates beyond a simple parametric form. Similar to the hypothesis test (11), TAPS can test

$$\mathrm{H}_0 : \theta_1 \in \mathcal{H}_A \qquad \text{v.s.} \qquad \mathrm{H}_1 : \theta_1 \in \mathcal{H} \text{ and } \theta_1 \notin \mathcal{H}_A, \tag{16}$$

where $\mathcal{H}_A$ is a parametric functional space and $\mathcal{H}$ is a general functional space. Furthermore, we consider the same mixed-effects representation of $\theta_1(x_1)$ as (12):

$$\theta_1(x_1) = \sum_{s=1}^{q} A_{1s}(x_1)\alpha_{1s} + b_1(x_1), \tag{17}$$

and we evaluate if $\theta_1(x_1)$ can be sufficiently described by the fixed effect $\sum_{s=1}^{q} A_{1s}(x_1)\alpha_{1s}$ by testing whether the variance of the random effect $b_1(x_1)$ is zero or not. In the context of the

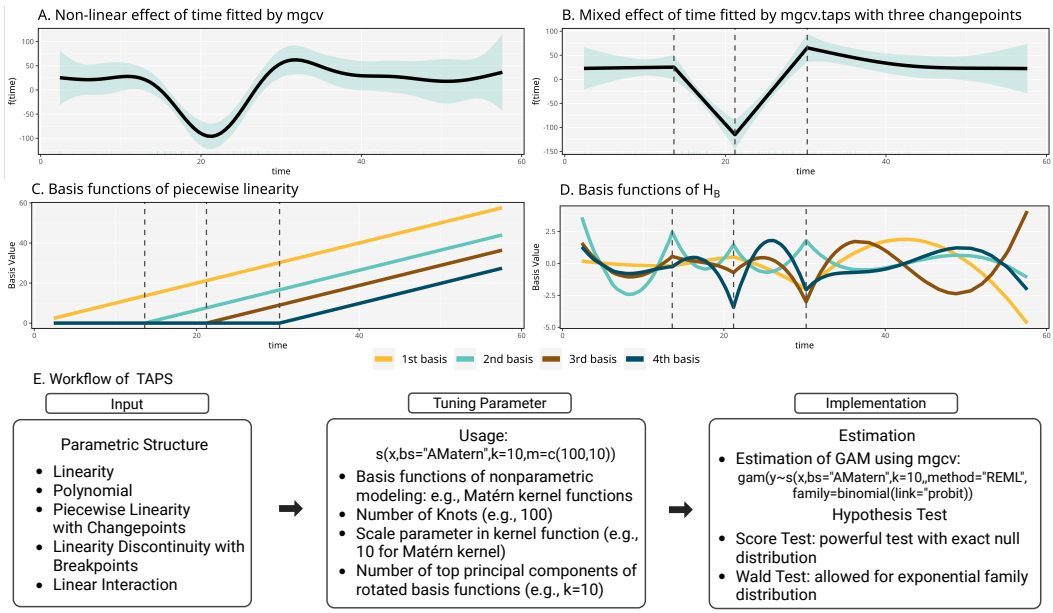

Figure 1: Overview of the TAPS framework. This figure uses the `mcycle` data, which records head acceleration over time in a simulated motorcycle crash experiment, to illustrate the workflow of TAPS. This data is provided by the R package `MASS` and is commonly used for benchmarking smooth curve fitting methods. **A**. The fitted smooth function of time using `mgcv` with default smooth terms. **B**. The fitted smooth function of time using `mgcv.taps` with three user-specified changepoints. **C**. Basis functions of the space of piecewise linearity space $\mathcal{H}_A$. **D**. Basis functions from the kernel-smoothed space $\mathcal{H}_B$, automatically learned via principal component decomposition. **E**. Summary of the TAPS components.

data and scientific questions outlined in Section 2, $x_{i2} = \sum_{j=1}^{p} \beta_j g_{ij}$ is the PRS term, and x_{i1} is age. The null space $\mathcal{H}_A$ takes different forms depending on the question of interest: for the first scientific question, $\mathcal{H}_A$ is the space of linear functions of age, whereas for the second question, $\mathcal{H}_A$ is the parametric space with a jump or changepoint at $x = 65$.

Figure 1A shows a classic example used in literature, the motorcycle acceleration data, which records head acceleration during a simulated crash (Wood et al. 2016). In this example, our goal is to test whether the time trend can be segmented into four phases, with changepoints marked by vertical dashed lines. The null hypothesis assumes that the function lies within a piecewise linear space $\mathcal{H}_A$, whose basis functions are shown in Figure 1C. To test for potential deviations from this structure, TAPS constructed $b_1(x)$ from a space $\mathcal{H}_B = \mathcal{H}/\mathcal{H}_A$, with basis functions adaptively learned from the data, as shown in Figure 1D. Figure 1B

displays the fitted trend under the mixed representation. The first three phases aligned with the piecewise linear form, while the fourth showed a slightly non-linear effect. Figure 1E summarizes the overall TAPS workflow.

3.2 Methodology of TAPS

Methodologically, we address two key challenges in the implementation of TAPS. First, we show how to construct the random effect $b_1(x)$ that is adaptively orthogonal to an arbitrary prespecified $\boldsymbol{A}_1(x)$, guaranteeing an identifiable mixed-effect representation. We have developed the R package `mgcv.taps`, which enables `mgcv` to recognize the new mixed-effect representation (12). Notably, `mgcv` can handle the varying-coefficient model with almost the same usage as the standard GAM. Second, we explore the methodology to conduct the hypothesis test (14) within a GAM model. We design two testing procedures for assessing model terms: a Wald test (Wood 2013) and a score test (Zhang & Lin 2003). The score test is generally more powerful, yet currently limited to outcomes from the exponential family. Whereas the Wald test supports all outcome types implemented in `mgcv`. By offering both options, users can choose the method best suited to their analysis objectives.

The following subsections present the core methodological principles of TAPS, while additional technical details are provided in the **Supplementary Materials**.

3.2.1 Construction of basis functions

The construction of $b_1(x)$ starts from representing it using the RKHS (Wahba 1990, Wainwright 2019), under the alternative hypothesis that $b_1(x)$ is a general smooth function. According to the functional analysis theory, every RKHS $\mathcal{H}_K$ corresponds to a unique semi-positive definite kernel function $K(\cdot,\cdot)$ defined in the Cartesian product space $[0,1]\times[0,1]$.

For any function $b_1 \in \mathcal{H}_K$,

$$\langle b_1, K(\cdot, x)\rangle_{\mathcal{H}_K} = b_1(x), \tag{18}$$

where $\langle \cdot, \cdot \rangle_{\mathcal{H}_K}$ is the inner product in the space $\mathcal{H}_K$. This property ensures that for n observations $\{x_1, \ldots, x_n\} \subset [0, 1]$, there must exist an $(n \times n)$ matrix $\mathbf{K}_1$ with the (i, j)th element being $K(x_{1i}, x_{1j})$ and an $(n \times 1)$ vector $\boldsymbol{\theta}_1 = (\theta_{11}, \ldots, \theta_{1n})^\top$ such that

$$\boldsymbol{b}_1 = \mathbf{K}_1 \boldsymbol{\theta}_1, \tag{19}$$

where $\boldsymbol{b} = (b(x_1), \ldots, b(x_n))^\top$. Note that the domain $[0, 1]$ is adopted purely as a mathematical normalization; in practice, it does not require transforming the covariate values to the $[0, 1]$ range.

We construct the basis functions of $\mathcal{H}_B = \mathcal{H}_K / \mathcal{H}_A$ by finding an $((n + q) \times n)$ orthogonal matrix $\mathbf{Q}$ satisfying the following linear constraint

$$\mathbf{A}_1^\top \mathbf{F}_1 \mathbf{Q}_1 = \mathbf{0}_{q \times n}, \tag{20}$$

where $\mathbf{F}_1 = (\mathbf{A}_1, \mathbf{K}_1)$ is the combination of $\mathbf{A}_1$ and $\mathbf{K}_1$. Here, the orthogonal matrix $\mathbf{Q}_1$ describes a special linear transformation that resorts the columns of $\mathbf{F}_1$, i.e., the basic functions of $\mathcal{H}$, into a new one perpendicular to $\mathbf{A}_1$. As a result, the columns of $\mathbf{F}_1 \mathbf{Q}_1$ can be regarded as the basis functions of $\mathcal{H}_B$ and thus it can represent $\boldsymbol{b}_1 = \mathbf{F}_1 \mathbf{Q}_1 \boldsymbol{\pi}_1$ where $\boldsymbol{\pi}_1$ is an unconstrained vector in $\mathbb{R}^n$. The QR decomposition can yield a candidate of the orthogonal matrix $\mathbf{Q}_1$. Consider the following QR decomposition:

$$\mathbf{F}_1^\top \mathbf{A}_1 = (\mathbf{Q}_{11}, \mathbf{Q}_{12}) \begin{pmatrix} \mathbf{R}_{11} \\ \mathbf{0}_{n \times q} \end{pmatrix}, \tag{21}$$

where $\mathbf{Q}_{11}$ is an $((n + q) \times q)$ orthogonal matrix, $\mathbf{Q}_{12}$ is an $((n + q) \times n)$ orthogonal matrix, and $\mathbf{R}_{11}$ is an $(q \times q)$ upper triangular matrix. Here $\mathbf{Q}_{11}$ and $\mathbf{Q}_{12}$ are the orthogonal basis matrices of $\text{span}(\mathbf{F}_1^\top \mathbf{A}_1)$ and $\overline{\text{span}}(\mathbf{F}_1^\top \mathbf{A}_1)$, respectively, and thus we can choose $\mathbf{Q}_1 = \mathbf{Q}_{12}$.

Mathematically, there is a well-known equivalence between a smooth function in an RKHS induced by a kernel function $K(\cdot,\cdot)$, and a Gaussian process defined on an L_2 space with a covariance function the same as the kernel function $K(\cdot,\cdot)$ (Kanagawa et al. 2018). That is, for a smooth function $b_1(x) \in \mathcal{H}_K$, its n realizations mathematically can be regarded as

$$\boldsymbol{b}_1 \sim \mathcal{N}\left(\mathbf{0},\, \phi_1 \mathbf{K}_1\right). \tag{22}$$

However, $b_1(x)$ is not a free function in $\mathcal{H}_K$: it should be orthogonal to the fixed effect $\mathbf{A}_1^\top \boldsymbol{\alpha}_1$, and hence there is an additional step to construct its prior distribution. Specifically, as $\boldsymbol{b}_1$ falls within $\mathcal{H} = \mathcal{H}_A \cup \mathcal{H}_K$ and $\mathbf{F}_1$ is a basis matrix of $\mathcal{H}$, $\boldsymbol{b}_1$ can be also represented as

$$\boldsymbol{b}_1 = \mathbf{F}_1 \tilde{\boldsymbol{\vartheta}}_1 = \mathbf{A}_1 \tilde{\boldsymbol{\alpha}}_1 + \mathbf{K}_1 \tilde{\boldsymbol{\theta}}_1, \tag{23}$$

where the vector $\tilde{\boldsymbol{\vartheta}}_1 = \left(\tilde{\boldsymbol{\alpha}}_1^\top,\, \tilde{\boldsymbol{\theta}}_1^\top\right)^\top$. The constraint to ensure $\boldsymbol{b}_1$ fully falls within $\mathcal{H}_B$ is

$$\mathbf{A}_1^\top \mathbf{F}_1 \tilde{\boldsymbol{\vartheta}}_1 = \mathbf{0}_{q\times 1}, \tag{24}$$

because $\boldsymbol{b}_1 = \mathbf{F}_1 \tilde{\boldsymbol{\vartheta}}_1$ must be orthogonal to $\mathbf{A}_1 \boldsymbol{\alpha}_1$. According to Wood (2003, Page 99), any $\tilde{\boldsymbol{\vartheta}}_1$ is subject to the constraint (24) can be represented as $\tilde{\boldsymbol{\vartheta}}_1 = \mathbf{Q}_1 \boldsymbol{\pi}_1$, where $\mathbf{Q}_1$ is the basis matrix of $\overline{\text{span}}(\mathbf{F}_1^\top \mathbf{A}_1)$ and $\boldsymbol{\pi}_1$ is a free $(n \times 1)$ vector. As a result, $\boldsymbol{\pi}_1 \sim \mathcal{N}(\mathbf{0}_{n\times 1},\, \phi_1 (\mathbf{Q}_1^\top \tilde{\mathbf{K}}_1 \mathbf{Q}_1)^+)$ where $\tilde{\mathbf{K}}_1 = \begin{pmatrix} \mathbf{0}_{q\times q} & \mathbf{0}_{q\times n} \\ \mathbf{0}_{n\times q} & \mathbf{K}_1 \end{pmatrix}$. Subsequently, since $\boldsymbol{b}_1 = \mathbf{F}_1 \mathbf{Q}_1 \boldsymbol{\pi}_1$, the prior distribution of $\boldsymbol{b}_1$ is

$$\boldsymbol{b}_1 \sim \mathcal{N}\left(\mathbf{0}_{n\times 1},\, \phi_1 \mathbf{F}_1 \mathbf{Q}_1 \left(\mathbf{Q}_1^\top \tilde{\mathbf{K}}_1 \mathbf{Q}_1\right)^+ \mathbf{Q}_1^\top \mathbf{F}_1^\top\right). \tag{25}$$

Thus, testing whether $f_1 = \mathbf{A}_1 \boldsymbol{\alpha}_1$ reduces to testing whether $\phi_1 = 0$.

3.2.2 Dimension reduction

Constructing $\mathbf{B}_1$ requires the QR decomposition of the $((n+q)\times q)$ matrix $\mathbf{F}_1^\top \mathbf{A}_1$, which is costly for large n. To alleviate this, we use two dimension-reduction strategies: (i)

approximate the smooth component with a low-rank RKHS representation (Kammann & Wand 2003), and (ii) apply principal component analysis (PCA) as in Wood (2003). Although $\mathbf{Q}_1^\top \tilde{\mathbf{K}}_1 \mathbf{Q}_1$ in (25) may be negative, the leading eigenvalues retained by PCA are typically far from zero, so negative-definite covariance matrices do not occur in practice.

3.2.3 Score test

We extend the score test proposed by Zhang & Lin (2003) to evaluate whether the smooth deviation term $b_1(x)$ is equal to zero. This test is based on the equivalence between GAM and GLMM. Specifically, the score test is constructed from the following GLMM model:

$$\boldsymbol{Y} = \mathbf{A}\boldsymbol{\alpha} + \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \tag{26}$$

where $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ with $Y_i = \eta_i + (y_i - \mu_i)g'(\mu_i)$, $\mathbf{A}$ is the design matrix for fixed effects $\boldsymbol{\alpha}$, and $\mathbf{B}$ is the design matrix for random effects $\boldsymbol{\beta}$. The noise term follows a Gaussian distribution $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \phi_0 \mathbf{W}^{-1})$, where ϕ_0 is the dispersion parameter, $\mathbf{W} = \mathrm{diag}(W_1, \ldots, W_n)$ and $W_i = 1/(V(\mu_i)g'(\mu_i)^2)$, with $V(\mu)$ denoting the variance function and g the canonical link function. For the GAM (10), $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_J, \mathbf{Z})$, $\mathbf{B} = (\mathbf{F}_1\mathbf{Q}_1, \ldots, \mathbf{F}_J\mathbf{Q}_J)$, $\boldsymbol{\alpha} = (\alpha_1^\top, \ldots, \alpha_J^\top, \boldsymbol{\gamma}^\top)^\top$, and $\boldsymbol{\beta} = (\boldsymbol{\pi}_1^\top, \ldots, \boldsymbol{\pi}_J^\top)^\top$. Since both $\mathbf{B}\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$ are random, it is reasonable to consider them as a new noise $\mathbf{e} = \mathbf{B}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, whose prior distribution is $\mathcal{N}(\mathbf{0}_{n\times 1}, \mathbf{V}(\boldsymbol{\phi}))$ with $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_J)^\top$ and

$$\mathbf{V}(\boldsymbol{\phi}) = \phi_0 \mathbf{W}^{-1} + \sum_{j=1}^{J} \phi_j \mathbf{F}_j \mathbf{Q}_j \mathbf{S}_j \mathbf{Q}_j^\top \mathbf{F}_j^\top, \tag{27}$$

where $\mathbf{S}_j$ is the penalty matrix for the jth smooth term (Wood 2017). For instance, $\mathbf{S}_1 = (\mathbf{Q}_1^\top \tilde{\mathbf{K}}_1 \mathbf{Q}_1)^+$ for the target smooth term f_1.

We first apply the R package `mgcv` to estimate the model parameters. Specifically, the regression coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are estimated by solving:

$$(\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}) = \arg\min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \left\{ \left\| \sqrt{\mathbf{W}}(\boldsymbol{Y} - \mathbf{A}\boldsymbol{\alpha} - \mathbf{B}\boldsymbol{\beta}) \right\|_2^2 + \sum_{j=1}^{J} \frac{1}{\phi_j} \boldsymbol{\beta}^\top \mathbf{F}_j \mathbf{Q}_j \mathbf{S}_j \mathbf{Q}_j^\top \mathbf{F}_j^\top \boldsymbol{\beta} \right\}. \tag{28}$$

Given the current estimate $\hat{\boldsymbol{\alpha}}$, the variance components $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_J)^\top$ are then estimated by minimizing the restricted maximum likelihood (REML) function:

$$\begin{aligned}\mathcal{L}_{\text{REML}}(\boldsymbol{\phi} \mid \hat{\boldsymbol{\alpha}}) = &-\frac{1}{2}(\boldsymbol{Y} - \mathbf{A}\hat{\boldsymbol{\alpha}})^\top \mathbf{V}(\boldsymbol{\phi})^{-1}(\boldsymbol{Y} - \mathbf{A}\hat{\boldsymbol{\alpha}}) \\ &-\frac{1}{2}\log\det(\mathbf{V}(\boldsymbol{\phi})) - \frac{1}{2}\log\det(\mathbf{A}^\top \mathbf{V}(\boldsymbol{\phi})^{-1}\mathbf{A}),\end{aligned} \tag{29}$$

where $\mathbf{V}(\boldsymbol{\phi}) = \phi_0 \mathbf{W}^{-1} + \sum_{j=1}^{J} \phi_j \mathbf{F}_j \mathbf{Q}_j \mathbf{S}_j \mathbf{Q}_j^\top \mathbf{F}_j^\top$. The R package `mgcv` proceeds by iteratively updating $\hat{\boldsymbol{\alpha}}$ (and $\hat{\boldsymbol{\beta}}$) given the current $\hat{\boldsymbol{\phi}}$, and then updating $\hat{\boldsymbol{\phi}}$ based on the updated $\hat{\boldsymbol{\alpha}}$ (and $\hat{\boldsymbol{\beta}}$), until convergence (Wood 2011) .

The score test is based on the score equation $\mathbf{U}(\cdot)$ and Hessian matrix $\mathbf{H}(\cdot,\cdot)$ of the REML function (29) below:

$$\begin{aligned}\mathbf{U}(\phi_j) &= \left.\frac{\partial \mathcal{L}_{\text{REML}}(\boldsymbol{\phi}|\widehat{\boldsymbol{\alpha}})}{\partial \phi_j}\right|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}_0^j} \\ &= -\underbrace{\frac{1}{2}\text{tr}(\mathbf{P}(\boldsymbol{\phi})\mathbf{G}_j)}_{\mathbf{e}(\phi_j)}\Bigg|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}_0^j} + \underbrace{\frac{1}{2}(\boldsymbol{Y} - \mathbf{A}\widehat{\boldsymbol{\alpha}})^\top \mathbf{V}(\boldsymbol{\phi})^{-1}\mathbf{G}_j \mathbf{V}(\boldsymbol{\phi})^{-1}(\boldsymbol{Y} - \mathbf{A}\widehat{\boldsymbol{\alpha}})\Big\}}_{\mathbf{u}(\phi_j)},\end{aligned} \tag{30}$$

$$\mathbf{H}(\phi_j, \phi_k) = \text{E}\left\{\frac{-\partial^2 \mathcal{L}_{\text{REML}}(\boldsymbol{\phi}|\widehat{\boldsymbol{\alpha}})}{\partial \phi_j \partial \phi_k}\right\}\Bigg|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}_0^j} = \frac{1}{2}\text{tr}(\mathbf{P}(\boldsymbol{\phi})\mathbf{G}_j\mathbf{P}(\boldsymbol{\phi})\mathbf{G}_k)\Bigg|_{\boldsymbol{\phi}=\widehat{\boldsymbol{\phi}}_0^j}, \tag{31}$$

where $\widehat{\boldsymbol{\phi}}_0^j = (\hat{\phi}_1, \ldots, \hat{\phi}_{j-1}, 0, \hat{\phi}_{j+1}, \ldots, \hat{\phi}_J)^\top$, $\mathbf{G}_j = \partial \mathbf{V}(\boldsymbol{\phi})/\partial \phi_j = \mathbf{F}_j \mathbf{Q}_j \mathbf{S}_j \mathbf{Q}_j^\top \mathbf{F}_j^\top$, and

$$\mathbf{P}(\boldsymbol{\phi}) = \mathbf{V}(\boldsymbol{\phi})^{-1} - \mathbf{V}(\boldsymbol{\phi})^{-1}\mathbf{A}(\mathbf{A}^\top \mathbf{V}(\boldsymbol{\phi})^{-1}\mathbf{A})^{-1}\mathbf{A}^\top \mathbf{V}(\boldsymbol{\phi})^{-1}. \tag{32}$$

Zhang & Lin (2003) used the moment matching method to show that $\mathbf{u}(\phi_1)$ in (31) asymptotically follows a scaled χ^2 distribution:

$$\mathbf{u}(\phi_1) \sim \kappa \chi^2_\nu, \tag{33}$$

where $\kappa = \mathbf{H}(\phi_1, \phi_1)/(2\mathbf{e}(\phi_1))$ and $\nu = 2\mathbf{e}(\phi_1)^2/\mathbf{H}(\phi_1, \phi_1)$. They thus used $\mathbf{u}(\phi_1)$ as the testing statistics and yielded the p-value by $p(\phi_1) = 1 - F_\chi(\mathbf{u}(\phi_1)/\kappa, \nu)$, where $F_\chi(\cdot, \nu)$ is the cumulative distribution function of the χ^2_ν distribution.

When the sample size n is large, constructing the $(n \times n)$ precision matrix $\mathbf{V}(\boldsymbol{\phi})^{-1}$ and projection matrix $\mathbf{P}(\boldsymbol{\phi})$ is infeasible. We therefore develop an operator-based strategy, defining matrix–vector product operators for $\mathbf{V}(\boldsymbol{\phi})^{-1}$ and $\mathbf{P}(\boldsymbol{\phi})$ to implement residual projection and trace computation efficiently. We further exploit the identity $\mathrm{tr}(\mathbf{A}\mathbf{A}^\top) = \mathrm{tr}(\mathbf{A}^\top\mathbf{A})$ to simplify trace evaluation for the low-rank matrix $\mathbf{G}_1 = \mathbf{F}_1\mathbf{Q}_1\mathbf{S}_1\mathbf{Q}_1^\top\mathbf{F}_1^\top$. With these improvements, the score test runs in about 30 seconds on 10 CPU cores with 100 GB memory for about 400,000 individuals.

3.2.4 Wald test

We adopt the Wald test proposed by Wood (2013) to evaluate whether the smooth deviation term $b_1(x)$ is equal to zero. Specifically, let $\hat{\boldsymbol{\pi}}_1$ denote the estimate of $\boldsymbol{\pi}_1$ under the estimated variance component $\hat{\phi}_1$. The null hypothesis is H_0: $b_1(x) = 0$. The Wald test statistic is given by $\hat{\boldsymbol{b}}_1^\top\mathbf{S}_{b_1}^{-1}\hat{\boldsymbol{b}}_1$, where $\hat{\boldsymbol{b}}_1 = \mathbf{F}_1\mathbf{Q}_1\hat{\boldsymbol{\pi}}_1$ and $\mathbf{S}_{b_1} = \hat{\phi}_1\mathbf{F}_1\mathbf{Q}_1\left(\mathbf{Q}_1^\top\tilde{\mathbf{K}}_1\mathbf{Q}_1\right)^+\mathbf{Q}_1^\top\mathbf{F}_1^\top$. Although this statistic is theoretically assumed to follow a χ^2 distribution, Wood (2013) showed that it is a mixture of χ^2 distributions due to the penalization imposed on $b_1(x)$.

3.2.5 Simulation and other materials

Owing to space constraints, the technical details of TAPS and its evaluation using simulations are presented in the **Supplementary Materials**. Specifically, our simulations investigate the type-I error control of the Wald and score tests when the hypothesized structure is correct, as well as the power of the two methods as the true structure increasingly deviates from the null. In addition, TAPS is a general and powerful framework that extends well beyond genetic epidemiology, making it applicable to a wide range of research domains where GAMs are used. To illustrate its versatility, three additional examples from different fields are provided in the **Supplementary Materials** and at GitHub.

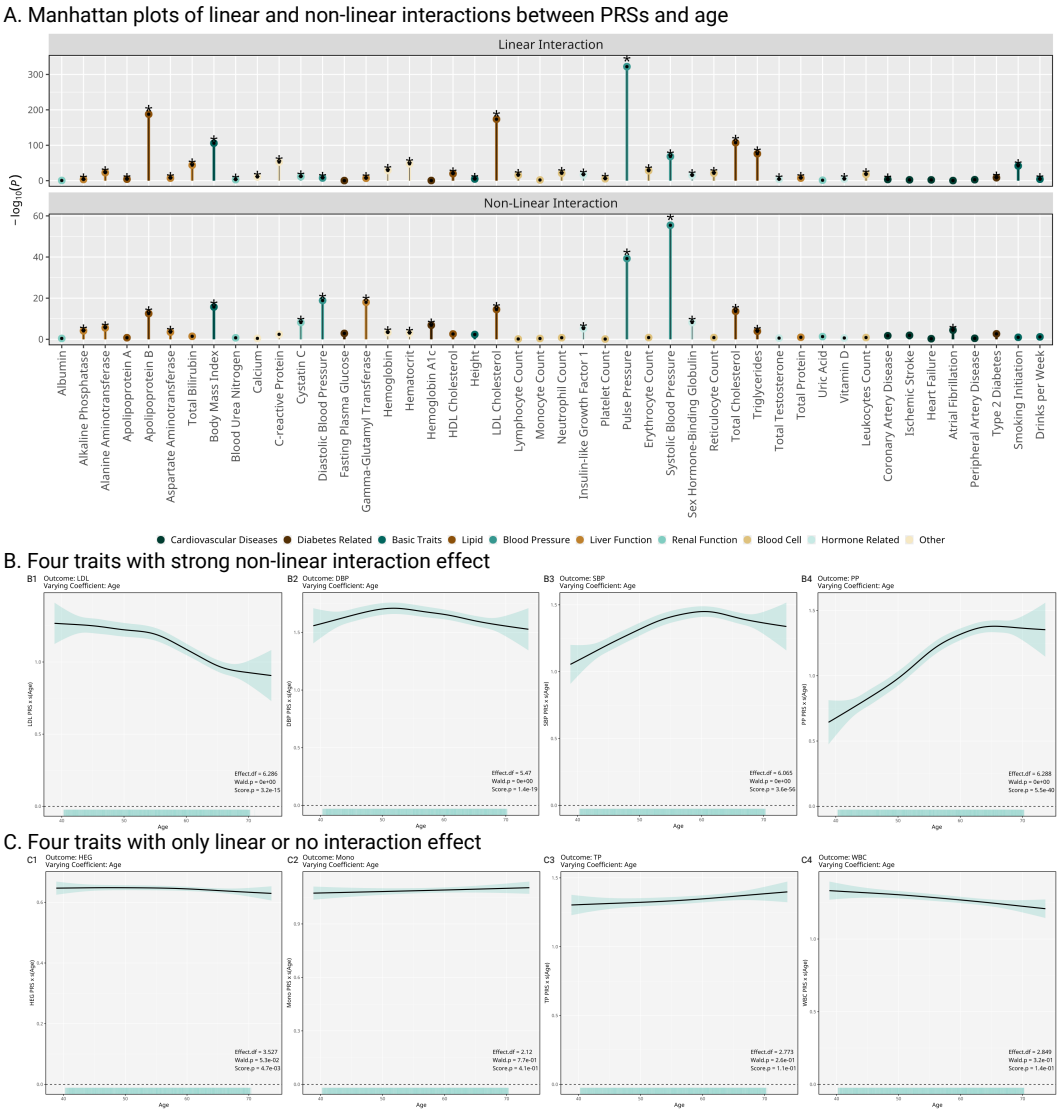

Figure 2: Age-varying coefficient analysis of polygenic risk score across traits. **A**. Manhattan plots of interaction tests evaluating whether polygenic risk score (PRS) effects vary with age. The top panel shows $-\log_{10}(p\text{-values})$ from Wald tests for linear age–PRS interactions, while the bottom panel shows $-\log_{10}(p\text{-values})$ from score tests for non-linear age-varying effects, orthogonal to the linear interaction terms. Asterisk * indicates score test p-value $< 0.05/46$. **B**. Estimated age-varying effect functions for selected traits with strong evidence of nonlinearity. **C**. Estimated age-varying effect functions for traits with linear or no detectable interaction. Each curve represents the fitted age-dependent PRS effect; shaded bands indicate 95% confidence intervals. Each panel reports the effective degrees of freedom (edf), the score test p-value and the Wald test p-value for the variance of random effects.

4 Real Data Analysis

4.1 Linear interaction test of age-varying PRS effect

We applied TAPS to UK Biobank data to investigate whether the age-varying effect of a PRS can be adequately represented by a linear structure, which corresponds to the traditional G×E interaction analysis (Aschard 2016). We analyzed 38 continuous traits and 8 binary traits, including markers of blood pressure, lipids, glucose, liver function, kidney function, type 2 diabetes, and cardiovascular diseases. Our analysis was restricted to participants of genetically inferred European ancestry in the UK Biobank (Privé et al. 2022). The sample sizes ranged from 340,864 (peripheral artery disease) to 411,463 (hypertension), with a median of 392,102. Many traits exhibited skewed distributions and had some outliers; therefore, we applied inverse-rank normalization (IRNT) to these traits and used the corresponding GWAS summary statistics from the Neale Lab (Lab 2018) for each setting. PRSs were estimated using SBayesRC (Zheng et al. 2024) with the default settings and predicted for individuals of European ancestry in the UK Biobank using PLINK (Purcell et al. 2007). In addition, the GAM we worked on is

$$\mathrm{E}(y_i|\mathrm{data}_i) = g^{-1}\left(\mathrm{PRS}_i \times \theta_1(\mathrm{Age}_i) + f_2(\mathrm{Age}_i) + \sum_{j=1}^{10} \mathrm{PC}_{ij}\gamma_j + \mathrm{Sex}\gamma_{11}\right), \tag{34}$$

where y_i is a trait, $g(\cdot)$ is the link function corresponding to the distribution of y_i, PRS_i is its PRS, PC_{ij} is the jth genetic PC, and data_i denotes all the covariates.

As shown in Figure 2A, the top panel presents the Manhattan plot of $-\log_{10}(p)$ values from the t-tests for linear age–PRS interactions, while the bottom panel reports $-\log_{10}(p)$ values from the score tests evaluating non-linear age-varying effects orthogonal to the linear interaction term. Full results are shown in Table S2. Among the 46 traits analyzed, 37 had p-values below the Bonferroni-corrected threshold for the linear interaction term, and 19 had p-values below the same threshold for the non-linear term. All 19 traits with significant

non-linear effects also had significant linear interactions, and there were no traits with non-linear effects significant in the absence of a significant linear interaction. In addition, the majority of traits had significant linear age–PRS interactions, whereas about half of them additionally showed evidence of non-linear interactions. Notably, the most significant score test p-value for non-linear age-varying effects was observed for systolic blood pressure (SBP) (p-value = 3.57E-56), followed by pulse pressure (PP) (p-value = 5.49E-40) and diastolic blood pressure (DBP) (p-value = 1.40E-19).

Figure 2B showed estimated age-dependent PRS effect curves for four representative traits with significant non-linear interaction effects: low-density lipoprotein cholesterol (LDL), DBP, PP, and SBP. A consistent pattern emerged across these examples: the age-varying effect appeared approximately piecewise linear, with inflection points typically occurring between ages 50 and 60. In particular, both DBP and SBP exhibited a rise-and-fall structure: the PRS effect increased during middle age and then declined later in life, with the changepoint for DBP occurring earlier than for PP and SBP. Figure 2C presents traits with either purely linear or no significant age-varying genetic effects. For instance, height (HEG), monocytes (Mono), total proteins (TP), and leukocytes (WBC) showed significant linear interaction effects without evidence of non-linearity. These traits, especially height, served as negative controls, demonstrating that our score test was not overly liberal: it successfully distinguished between linear and non-linear patterns and yielded appropriately conservative results when no age-dependent variation was present.

4.1.1 Follow-up analysis: piecewise linear vs nonlinear PRS effect

As a follow-up to the main analysis, we further examined whether the significant non-linear age-varying effects could be parsimoniously represented by a piecewise linear function with empirically determined changepoints. Specifically, we observed that many age-varying effects of PRSs exhibited an approximately piecewise linear form with one or two changepoints of

Table 1: Test of piecewise linear age-varying effect of a PRS

Trait	Abbreviation	Parametric Part			Nonparametric Part	
		Location	Estimate	p-value	Wald	Score
Systolic Blood Pressure	SBP	Age=58	-0.0288	7.34E-47	0.3962	0.2033
Diastolic Blood Pressure	DBP	Age=52	-0.0222	5.61E-19	0.9693	0.8008
Pules Pressure	PP	Age=62	-0.0353	3.40E-38	0.0785	0.1219

Estimate = fixed-effect estimate of change-of-slope (α_K); p-value = nominal significance of the fixed effect; q-value = FDR-adjusted p-value; Wald and Score = random-effect tests for residual non-linearity after accounting for the fixed effect.

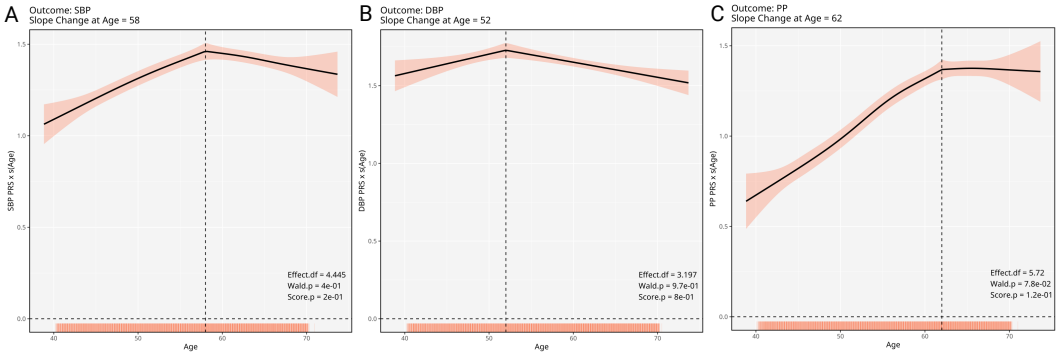

Figure 3: Change-of-slope analyses of age-varying effects of PRSs for three blood pressure traits, with the changepoints of slope determined from empirical inspection. **A**: Systolic blood pressure (SBP), change-of-slope at age 58. **B**: Diastolic blood pressure (DBP), change-of-slope at age 52. **C**: Pulse pressure (PP), change-of-slope at age 62. Each curve shows the fitted age-dependent PRS effect, with shaded bands indicating 95% confidence intervals. Reported within each panel are the effective degrees of freedom (edf), the score test p-value and the Wald test p-value for the variance of random effects.

slope. For example, in Figure 3B1, the age-varying effect of the LDL PRS appears to follow a piecewise linear structure with two changepoints. In contrast, in Figures 3B2–3B4, the three blood pressure traits seem to follow a piecewise linear form with a single changepoint. Motivated by this pattern, we selected SBP, DBP, and PP as examples and tested whether their age-varying effects could be adequately represented by a piecewise linear structure with one changepoint, located at age = 50 for DBP, age = 62 for PP, and age = 58 for SBP.

Table 1 summarizes the structure tests for a single-changepoint piecewise linear specification of the age-varying effects of SBP, DBP, and PP PRSs. For all three traits, the fixed-effect estimates of the change-of-slope (α_K) were negative and highly significant (p-value=

7.34×10^{-47} for SBP, p-value= 5.61×10^{-19} for DBP, and p-value= 3.40×10^{-38} for PP), indicating a consistent attenuation of the genetic effect after the changepoint. In contrast, both the Wald and score tests for the variance of the random effect were non-significant (p-value>0.05) for all traits), suggesting no residual non-linearity after accounting for the fixed change-of-slope. Figure 3 provides a visual illustration of these patterns: the PRS effects increase with age up to the estimated changepoints, age 58 for SBP, age 52 for DBP, and age 62 for PP, and subsequently decline, with no evidence of additional structural complexity beyond the specified piecewise linear structures.

It is worth noting that this follow-up analysis is conceptually similar to the retirement-stratified age-varying PRS effect analysis in Section 2.3. However, here the changepoints of slope were empirically identified from the data rather than fixed at the statutory retirement age, and thus cannot be directly interpreted as retirement-induced structural shifts. Instead, our goal was to assess whether the complex non-linear patterns detected in the main analysis could be reduced to simple piecewise linear forms corresponding to age stratification.

4.2 Structural change tests of retirement-stratified age-varying PRS effect

We applied TAPS to UK Biobank data to statistically assess whether the age-varying effect of a PRS undergoes a structural change at the statutory retirement age of 65 years, considering both a jump and a change-of-slope specification. The analysis focused on participants aged 63–67 from the genetically inferred European ancestry subset of the UK Biobank. We still examined the 38 continuous traits and 8 binary traits. Sample sizes ranged from 52,807 (stroke) to 72,848 (height), with a median of 69,287. Continuous traits were IRNT prior to analysis, and the corresponding GWAS summary statistics from the Neale Lab were used. PRSs were estimated using SBayesRC and predicted for individuals. Moreover, we consider

Table 2: Results of jump and change-of-slope tests of age-varying effects of PRSs at age 65.

Trait	Abbreviation	Parametric Part			Nonparametric Part	
		Test of jump at age = 65				
		Estimate	p-value	q-value	Wald	Score
Apolipoprotein A	APOA	-0.0308	0.0208	0.1910	0.8840	0.4994
Aspartate Aminotransferase	AST	-0.0592	0.0319	0.2637	0.3152	0.1393
Cystatin C	CysC	0.0365	0.0063	**0.0965**	0.9332	0.6665
Diastolic Blood Pressure	DBP	0.0908	0.0146	0.1676	0.9232	0.6316
LDL Cholesterol	LDL	-0.0762	0.0008	**0.0178**	0.7982	0.3057
Total Cholesterol	TC	-0.0628	0.0001	**0.0067**	0.2614	0.1152
		Test of change-of-slope at age = 65				
		Estimate	p-value	q-value	Wald	Score
Albumin	ALB	0.0288	0.0212	0.1952	0.1375	0.0217
Monocyte Count	Mono	-0.0611	0.0069	0.1056	0.9986	0.6349
Total Proteins	TP	-0.1070	0.0012	**0.0560**	0.5793	0.2919
Urate	UA	-0.0543	0.0173	0.1994	0.9976	0.9396
Leukocyte Count	WBC	-0.0794	0.0026	**0.0593**	0.3789	0.2076

Estimate = fixed-effect estimate of jump (α_J) or change-of-slope (α_K); p-value = nominal significance of the fixed effect; q-value = FDR-adjusted p-value; Wald and Score = random-effect tests for residual non-linearity after accounting for the fixed effect. Bold q-values indicate significance at FDR < 0.1.

the same GAM model as (34). Details are provided in **Supplementary Materials**.

Table 2 summarizes the results of the jump and change-of-slope tests for age-varying effects of PRSs at the statutory retirement age of 65. Overall, most of the 46 traits analyzed in the retirement-stratified setting showed no evidence of a structural change in the age-varying effects of PRSs at age = 65. For all traits listed in Table 2, the Wald and score tests of the random effects were non-significant, indicating that once the jump or change-of-slope term was included, no additional non-linear component was detectable. Additionally, no traits were identified in which the fixed-effect parameter was non-significant while the random-effect parameters were significant in the two tests.

In the jump-at-65 analysis, five traits met the nominal threshold of p-value<0.05 for the jump parameter α_J, but only three, cystatin C (CysC; estimate =0.0365, q-value=0.0965), LDL (estimate =-0.0762, q-value=0.0178), and total cholesterol (TC; estimate =-0.0628,

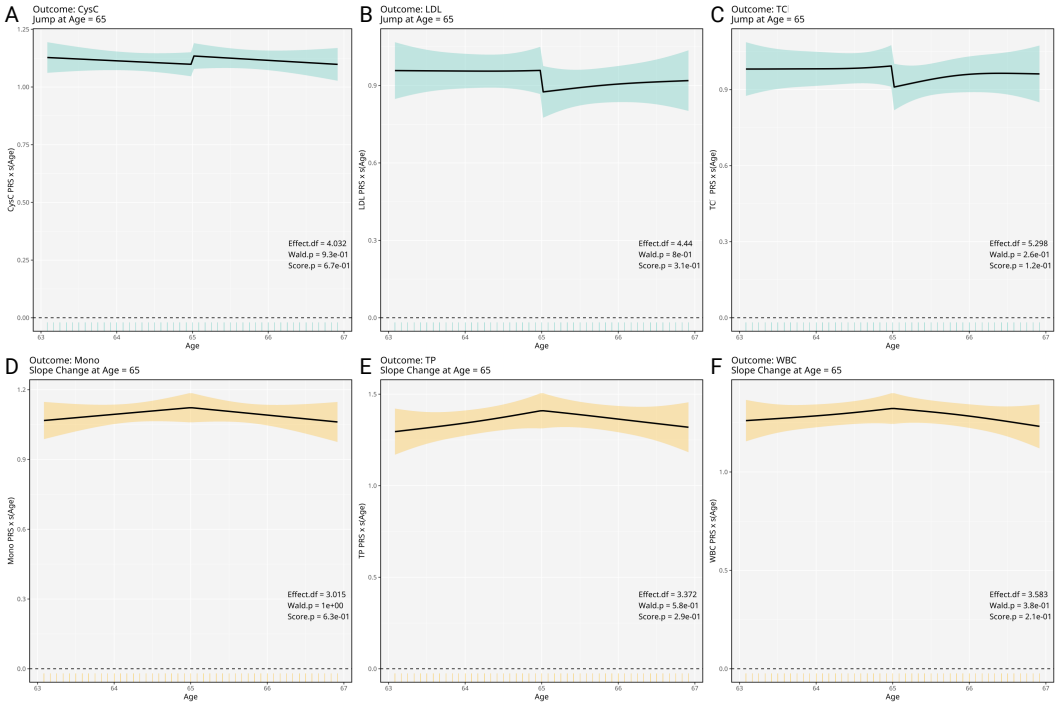

Figure 4: Jump- and change-of-slope analyses of age-varying effects of PRSs at the statutory retirement age of 65 years. **A**–**C**: Representative traits under the jump-at-65 specification. **D**–**F**: Representative traits under the slope-change-at-65 specification. Each curve represents the fitted age-dependent PRS effect; shaded bands indicate 95% confidence intervals. Each panel reports the effective degrees of freedom (edf), the score test p-value and the Wald test p-value for the variance of random effects.

q-value=0.0067), remained significant after controlling the false discovery rate at 10%. From Figure 4A-4C, CysC shows a distinct upward jump at age 65, whereas LDL and TC display clear downward jumps. The remaining traits in this analysis, including apolipoprotein A (APOA), aspartate aminotransferase (AST), and DBP, did not reach the FDR-adjusted significance level.

In the slope-change-at-65 analysis, five traits had change-of-slope parameters α_K with nominal p-value<0.05, but only total proteins (TP; estimate =-0.1070, q-value=0.056) and leukocyte count (WBC; estimate =-0.0794, q-value=0.0593) were significant after FDR adjustment. In Figure 4D–4F, the changes in slope are visually subtle, with most curves exhibiting only slight variations; for instance, TP and WBC show a gradual increase before age 65, followed by a gradual decrease thereafter. Albumin (ALB), monocyte count (Mono),

and urate (UA) were not significant after multiple-testing correction.

5 Discussion

In this paper, we develop TAPS, a computational framework for testing whether a target function in a GAM can be fully described by a prespecified parametric structure. Existing tests are typically restricted to assessing whether a smooth term is zero, rather than whether it follows a scientifically meaningful parametric form. TAPS addresses this gap by formulating the null as membership in a given parametric structure and the alternative as a general smooth function. This formulation directly answers the question: does the function follow the specified structure? Methodologically, TAPS ensures identifiability by decomposing the function into orthogonal parametric (fixed-effect) and nonparametric (random-effect) components, so that the significance of the parametric part can be tested independently of the nonparametric deviation. TAPS provides both Wald and score tests: the Wald test applies to all outcome families supported by `mgcv`, while the score test typically offers greater power. Given its superior Type I error control and higher power, the score test is recommended as the default, with the Wald test as a practical alternative for non-exponential-family outcomes. Finally, we implement TAPS in the R package `mgcv.taps`, tightly integrated with `mgcv`, allowing users to apply it with minimal additional learning, while supporting diverse outcome types and scaling efficiently to biobank-scale data.

In the first analysis, 83% (38/46) exhibited significant age-varying effects of PRSs, indicating that age modifies the genetic architecture for most complex traits analyzed (Sung et al. 2018, Winkler et al. 2024). For half of these 38 traits, the age-varying pattern could not be adequately captured by a simple linear age–PRS interaction, suggesting substantial heterogeneity in how genetic influences evolve across the life course. Similar heterogeneity has been reported for blood pressure, BMI, and metabolic biomarkers, where genetic effects

may accelerate, plateau, or even reverse direction at specific life stages (Hottenga et al. 2005, Guindo-Martínez et al. 2021, Jiang et al. 2021). Several traits were also well approximated by a parsimonious piecewise linear structure, where the linear interaction may hold within certain age ranges but with different slopes, or even opposite signs, between younger and older individuals, as observed for SBP, DBP, and PP. These results suggest that G×E interaction studies could gain power by stratifying analyses by age and other environmental factors (Domingue et al. 2020). Nevertheless, the biological underpinnings of the observed changepoints remain unclear and warrant further investigation.

In the second analysis, we found that retirement at age 65 was associated with structural changes in age-varying PRS effects for only a small subset of traits: LDL, TC, CysC, TP, and WBC. To our knowledge, this is the first evidence that retirement may interact with genetic effects to shape health-related phenotypes. For lipid traits such as LDL and TC, a plausible explanation is that retirement coincides with changes in healthcare coverage or benefit eligibility, potentially altering lipid-lowering treatment patterns (Gallo et al. 2000). For inflammatory and protein markers such as TP and WBC, retirement may reduce chronic psychosocial stress (Eibich 2015), which has been linked to lower systemic inflammation in older adults (Steptoe et al. 2007). These conjectures require validation in targeted trials or longitudinal studies. Compared with prior work (Eibich 2015, Furuya & Fletcher 2024), which examined direct effects of retirement on health indicators, our analysis is novel in testing whether retirement coincides with structural modifications in genetic effects, thereby highlighting a previously unexplored aspect of G×E interaction.

In addition, structural changes were detectable only when focusing on individuals near the retirement threshold (e.g., 63<age<67). In full age-range analyses, these patterns were absent or appeared as simple linear age–PRS interactions. For example, LDL showed no visible jump at age 65 in the broad age model (Figure 2B1), and TP and WBC displayed

only global linear trends in the full sample (Figures 2C2–2C4). This localization is consistent with RDD and RKD theory, which rely on data in a narrow bandwidth around the cutoff (Huntington-Klein 2021). Including observations far from the threshold can dilute local signals and increase bias if the functional form is misspecified. Following RDD/RKD principles, the window must be narrow enough to capture local changes yet large enough to ensure sufficient data. Note that the UK Biobank records participants' month and year of birth along with the date of assessment, allowing age to be derived to the nearest month. Using overly small intervals would therefore yield too few distinct values; we accordingly used $63 < \text{age} < 67$, which balances sample size with local focus around the cutoff.

Our two analyses also highlight important limitations in the current application of PRSs. Specifically, a growing body of literature has shown that the predictive power and risk associations of PRSs differ across age strata (Khan et al. 2022). Our findings suggest that this pattern may arise because the effects of PRSs themselves are continuously age-varying rather than constant over the lifespan. Such patterns are likely driven by the underlying genetic architecture in which individual variant effects are age-dependent (Jiang et al. 2021), and current PRSs therefore capture only age-averaged or absolute accumulated polygenic effects (Chatterjee et al. 2016). We thus speculate that much of the observed age variation in PRS effects may reflect differences in the predictive power of absolute PRSs across age groups, although novel biological mechanisms cannot be excluded. In the future, constructing PRSs based on age-stratified GWAS or explicitly modeling age-varying genetic effects may enable more targeted disease risk prediction across different age groups, thereby enhancing the practical utility of PRSs (Chatterjee et al. 2016).

Several limitations of TAPS warrant further investigation. First, TAPS requires users to specify the parametric structure to be tested, including hyperparameters such as the location of a jump or change-of-slope. We should emphasize that the unity of TAPS is

not designed to discover the best-fitting structure from data, but rather to statistically test whether a given structural form can sufficiently describe the variation of the data, for reasons of interpretability and generalizability. In addition, in some applications like RDD and RKD, as well as the retirement analysis, these locations correspond to cutoffs of treatment assignments and are not subject to tuning (Huntington-Klein 2021). However, we acknowledge the importance of discovering structural forms rather than requiring them to be pre-specified, and future extensions of TAPS could incorporate structural change detection methods (Bai & Perron 1998, Muggeo 2003, Killick et al. 2012, Zhang & Li 2023, Chen et al. 2023).

Acknowledgments

We would like to thank the editors and the anonymous referees for their very careful reviews and constructive suggestions. The first author thanks Dr. Jianxin Pan for multiple insightful discussions. We also thank Dr. Simon Wood for providing the flexible `mgcv` interface, which facilitates user-defined smooth terms and enables seamless methodological extensions. This work was supported by grants HG011052 and HG011052-03S1 (to X.Z.) from the National Human Genome Research Institute (NHGRI), USA.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the author(s) used ChatGPT-4o in order to improve readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Data and Code Availability

All GWAS summary statistics used in this study are publicly available, with their sources listed in Table S1. Given the large number of traits analyzed, we have implemented an interactive Shiny web interface (https://ovlwff-yihe-yang.shinyapps.io/visualization_of_taps/) to visualize all plots. The individual-level data from the UK Biobank used for analyses are available under Application ID: 81097.

The R package `mgcv.taps`, which implements the TAPS method, is available at https://github.com/harryyiheyang/mgcv.taps. All analysis code used in this paper is available at https://github.com/harryyiheyang/mgcv.taps.code.

SUPPLEMENTARY MATERIAL

Supplementary Materials: Supplementary Materials contain the supplementary methodology, simulations, and real data analyses.

Supplementary Tables: Supplementary Tables contain multiple tables recording the results of real data analysis.

References

Abdellaoui, A., Yengo, L., Verweij, K. J. & Visscher, P. M. (2023), '15 years of GWAS discovery: realizing the promise', *Am. J. Hum. Genet.* **110**(2), 179–194.

Aschard, H. (2016), 'A perspective on interaction effects in genetic association studies', *Genet. Epidemiol.* **40**(8), 678–688.

Bai, J. & Perron, P. (1998), 'Estimating and testing linear models with multiple structural changes', *Econometrica* pp. 47–78.

Breslow, N. E. & Clayton, D. G. (1993), 'Approximate inference in generalized linear mixed models', *J. Am. Stat. Assoc.* **88**(421), 9–25.

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K. et al. (2018), 'The uk Biobank resource with deep phenotyping and genomic data', *Nature* **562**(7726), 203–209.

Chatterjee, N., Shi, J. & García-Closas, M. (2016), 'Developing and evaluating polygenic risk prediction models for stratified disease prevention', *Nature Reviews Genetics* **17**(7), 392–406.

Chen, H., Ren, H., Yao, F. & Zou, C. (2023), 'Data-driven selection of the number of change-points via error rate control', *J. Am. Stat. Assoc* **118**(542), 1415–1428.

Crainiceanu, C., Ruppert, D., Claeskens, G. & Wand, M. P. (2005), 'Exact likelihood ratio tests for penalised splines', *Biometrika* **92**(1), 91–103.

Dahl, A., Nguyen, K., Cai, N., Gandal, M. J., Flint, J. & Zaitlen, N. (2020), 'A robust method uncovers significant context-specific heritability in diverse complex traits', *Am. J. Hum. Genet.* **106**(1), 71–91.

Domingue, B. W., Trejo, S., Armstrong-Carter, E. & Tucker-Drob, E. M. (2020), 'Interactions between polygenic scores and environments: Methodological and conceptual challenges', *Sociol. Sci.* **7**, 465–486.

Eibich, P. (2015), 'Understanding the effect of retirement on health: Mechanisms and heterogeneity', *J. Health Econ.* **43**, 1–12.

Fan, Y. & Ullah, A. (1999), 'Asymptotic normality of a combined regression estimator', *J. Multivar. Anal.* **71**(2), 191–240.

Fisher, R. A. (1919), 'XV.—the correlation between relatives on the supposition of Mendelian inheritance.', *Earth Environ. Sci. Trans. R. Soc. Edinb.* **52**(2), 399–433.

Furuya, S. & Fletcher, J. M. (2024), 'Retirement makes you old? Causal effect of retirement on biological age', *Demography* **61**(3), 901–931.

Gallo, W. T., Bradley, E. H., Siegel, M. & Kasl, S. V. (2000), 'Health effects of involuntary job loss among older workers: findings from the health and retirement survey', *J. Gerontol. B. Psychol. Sci. Soc. Sci.* **55**(3), S131–S140.

Guindo-Martínez, M., Amela, R., Bonàs-Guarch, S., Puiggròs, M., Salvoro, C., Miguel-Escalada, I., Carey, C. E., Cole, J. B., Rüeger, S., Atkinson, E. et al. (2021), 'The impact of non-additive genetic associations on age-related complex diseases', *Nat. Commun.* **12**(1), 2436.

Hastie, T. J. & Tibshirani, R. J. (1990), *Generalized additive models*, CRC Press.

Hastie, T. & Tibshirani, R. (1993), 'Varying-coefficient models', *J. R. Stat. Soc. Ser. B Stat. Methodol.* **55**(4), 757–779.

Hill, W. G., Goddard, M. E. & Visscher, P. M. (2008), 'Data and theory point to mainly additive genetic variance for complex traits', *PLoS Genet.* **4**(2), e1000008.

Hottenga, J.-J., Boomsma, D. I., Kupper, N., Posthuma, D., Snieder, H., Willemsen, G. & de Geus, E. J. (2005), 'Heritability and stability of resting blood pressure', *Twin Res. Hum. Genet.* **8**(5), 499–508.

Huntington-Klein, N. (2021), *The effect: An introduction to research design and causality*, Chapman and Hall/CRC.

Jiang, X., Holmes, C. & McVean, G. (2021), 'The impact of age on genetic risk for common diseases', *PLoS Genet.* **17**(8), e1009723.

Kammann, E. & Wand, M. P. (2003), 'Geoadditive models', *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **52**(1), 1–18.

Kanagawa, M., Hennig, P., Sejdinovic, D. & Sriperumbudur, B. K. (2018), 'Gaussian processes and kernel methods: A review on connections and equivalences', *arXiv preprint arXiv:1807.02582* .

Khan, S. S., Page, C., Wojdyla, D. M., Schwartz, Y. Y., Greenland, P. & Pencina, M. J. (2022), 'Predictive utility of a validated polygenic risk score for long-term risk of coronary heart disease in young and middle-aged adults', *Circulation* **146**(8), 587–596.

Killick, R., Fearnhead, P. & Eckley, I. (2012), 'Optimal detection of changepoints with a linear computational cost', *J. Am. Stat. Assoc* **107**(500), 1590–1598.

Lab, N. (2018), 'UK Biobank GWAS round 2 results'. Accessed: 2025-04-06.
URL: *http://www.nealelab.is/uk-biobank/*

Mazo, G. & Portier, F. (2021), 'Parametric versus nonparametric: The fitness coefficient', *Scand. J. Stat.* **48**(4), 1344–1383.

McKeigue, P. M., McGurnaghan, S., Blackbourn, L., Bath, L. E., McAllister, D. A., Caparrotta, T. M. et al. (2023), 'Relation of incident type 1 diabetes to recent covid-19 infection: cohort study using e-health record linkage in scotland', *Diabetes Care* **46**(5), 921–928.

Muggeo, V. M. (2003), 'Estimating regression models with unknown break-points', *Statistics in Medicine* **22**(19), 3055–3071.

Nelder, J. A. & Wedderburn, R. W. (1972), 'Generalized linear models', *J. R. Stat. Soc. A. Stat.* **135**(3), 370–384.

Nummi, T., Pan, J., Siren, T. & Liu, K. (2011), 'Testing for cubic smoothing splines under dependent data', *Biometrics* **67**(3), 871–875.

Olkin, I. & Spiegelman, C. H. (1987), 'A semiparametric approach to density estimation', *J. Am. Stat. Assoc.* **82**(399), 858–865.

Privé, F., Aschard, H., Carmi, S., Folkersen, L., Hoggart, C., O'Reilly, P. F. & Vilhjálmsson, B. J. (2022), 'Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort', *Am. J. Hum. Genet.* **109**(1), 12–23.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., De Bakker, P. I., Daly, M. J. et al. (2007), 'PLINK: a tool set for whole-genome association and population-based linkage analyses', *Am. J. Hum. Genet.* **81**(3), 559–575.

Shang, L., Wu, P. & Zhou, X. (2025), 'Statistical identification of cell type-specific spatially variable genes in spatial transcriptomics', *Nat. Commun.* **16**(1), 1059.

Steptoe, A., Hamer, M. & Chida, Y. (2007), 'The effects of acute psychological stress on circulating inflammatory factors in humans: a review and meta-analysis', *Brain Behav. Immun.* **21**(7), 901–912.

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M. et al. (2015), 'UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age', *PLoS Med.* **12**(3), e1001779.

Sung, Y. J., Winkler, T. W., de Las Fuentes, L., Bentley, A. R., Brown, M. R., Kraja, A. T., Schwander, K., Ntalla, I., Guo, X., Franceschini, N. et al. (2018), 'A large-scale multi-ancestry genome-wide study accounting for smoking behavior identifies multiple significant loci for blood pressure', *Am. J. Hum. Genet.* **102**(3), 375–400.

Tian, H., Patel, A. & Burgess, S. (2025), 'Stratification-based instrumental variable analysis framework for nonlinear effect analysis', *arXiv preprint arXiv:2507.07349* .

Wahba, G. (1990), *Spline Models for Observational Data*, SIAM.

Wainwright, M. J. (2019), *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press.

Winkler, T. W., Wiegrebe, S., Herold, J. M., Stark, K. J., Küchenhoff, H. & Heid, I. M. (2024), 'Genetic-by-age interaction analyses on complex traits in UK Biobank and their potential to identify effects on longitudinal trait change', *Genome Biol.* **25**(1), 300.

Wood, S. (2006), 'Low-rank scale-invariant tensor product smooths for generalized additive mixed models', *Biometrics* **62**(4), 1025–1036.

Wood, S. N. (2003), 'Thin plate regression splines', *J. R. Stat. Soc. Ser. B Stat. Methodol.* **65**(1), 95–114.

Wood, S. N. (2004), 'Stable and efficient multiple smoothing parameter estimation for generalized additive models', *J. Am. Stat. Assoc.* **99**(467), 673–686.

Wood, S. N. (2010), 'Statistical inference for noisy nonlinear ecological dynamic systems', *Nature* **466**(7310), 1102–1104.

Wood, S. N. (2011), 'Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models', *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**(1), 3–36.

Wood, S. N. (2013), 'On p-values for smooth components of an extended generalized additive model', *Biometrika* **100**(1), 221–228.

Wood, S. N. (2017), *Generalized Additive Models: An Introduction with R*, CRC Press.

Wood, S. N., Li, Z., Shaddick, G. & Augustin, N. H. (2017), 'Generalized additive models

for gigadata: modeling the UK black smoke network daily data', *J. Am. Stat. Assoc.* **112**(519), 1199–1210.

Wood, S. N., Pya, N. & Säfken, B. (2016), 'Smoothing parameter and model selection for general smooth models', *J. Am. Stat. Assoc.* **111**(516), 1548–1563.

Zhang, D. & Lin, X. (2003), 'Hypothesis testing in semiparametric additive mixed models', *Biostatistics* **4**(1), 57–74.

Zhang, F. & Li, Q. (2023), 'Segmented correspondence curve regression for quantifying covariate effects on the reproducibility of high-throughput experiments', *Biometrics* **79**(3), 2272–2285.

Zheng, Z., Liu, S., Sidorenko, J., Wang, Y., Lin, T., Yengo, L., Turley, P., Ani, A., Wang, R., Nolte, I. M. et al. (2024), 'Leveraging functional genomic annotations and genome coverage to improve polygenic prediction of complex traits within and between ancestries', *Nat. Genet.* **56**(5), 767–777.

Zhu, X., Yang, Y., Lorincz-Comi, N., Li, G., Bentley, A. R., de Vries, P. S., Brown, M., Morrison, A. C., Rotimi, C. N., Gauderman, W. J. et al. (2024), 'An approach to identify gene-environment interactions and reveal new biological insight in complex traits', *Nat. Commun.* **15**(1), 3385.

Zuk, O., Hechter, E., Sunyaev, S. R. & Lander, E. S. (2012), 'The mystery of missing heritability: Genetic interactions create phantom heritability', *Proc. Natl. Acad. Sci. U.S.A* **109**(4), 1193–1198.