



## Genetics-driven risk predictions leveraging the Mendelian randomization framework

Daniel Sens, Liubov Shilova, Ludwig Gräf, et al.

*Genome Res.* 2024 34: 1276-1285 originally published online September 27, 2024

Access the most recent version at doi:[10.1101/gr.279252.124](https://doi.org/10.1101/gr.279252.124)

---

**References** This article cites 56 articles, 4 of which can be accessed free at:  
<http://genome.cshlp.org/content/34/9/1276.full.html#ref-list-1>

**Open Access** Freely available online through the *Genome Research* Open Access option.

**Creative Commons License** This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

**Email Alerting Service** Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

---

To subscribe to *Genome Research* go to:  
<https://genome.cshlp.org/subscriptions>

---

© 2024 Sens et al.; Published by Cold Spring Harbor Laboratory Press

## Method

# Genetics-driven risk predictions leveraging the Mendelian randomization framework

Daniel Sens,<sup>1,2</sup> Liubov Shilova,<sup>1,2,3</sup> Ludwig Gräf,<sup>1,4</sup> Maria Grebenschchikova,<sup>1,5</sup> Bjoern M. Eskofier,<sup>1,3</sup> and Francesco Paolo Casale<sup>1,2,4</sup>

<sup>1</sup>Institute of AI for Health, Helmholtz Zentrum München—German Research Center for Environmental Health, 85764 Neuherberg, Germany; <sup>2</sup>Helmholtz Pioneer Campus, Helmholtz Zentrum München—German Research Center for Environmental Health, 85764 Neuherberg, Germany; <sup>3</sup>Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054 Erlangen, Germany; <sup>4</sup>School of Computation, Information and Technology, Technical University of Munich, 85748 Garching, Germany; <sup>5</sup>School of Management, Technical University of Munich, 80333 Munich, Germany

Accurate predictive models of future disease onset are crucial for effective preventive healthcare, yet longitudinal data sets linking early risk factors to subsequent health outcomes are limited. To overcome this challenge, we introduce a novel framework, Predictive Risk modeling using Mendelian Randomization (PRiMeR), which utilizes genetic effects as supervisory signals to learn disease risk predictors without relying on longitudinal data. To do so, PRiMeR leverages risk factors and genetic data from a healthy cohort, along with results from genome-wide association studies of diseases of interest. After training, the learned predictor can be used to assess risk for new patients solely based on risk factors. We validate PRiMeR through comprehensive simulations and in future type 2 diabetes predictions in UK Biobank participants without diabetes, using follow-up onset labels for validation. Moreover, we apply PRiMeR to predict future Alzheimer's disease onset from brain imaging biomarkers and future Parkinson's disease onset from accelerometer-derived traits. Overall, with PRiMeR we offer a new perspective in predictive modeling, showing it is possible to learn risk predictors leveraging genetics rather than longitudinal data.

[Supplemental material is available for this article.]

Large biobanks such as UK Biobank (UKB) (Sudlow et al. 2015), the German National Cohort (Wichmann et al. 2016), and others (Leitsalu et al. 2015; Abul-Husn et al. 2019), have unlocked access to extensive health metrics and risk factors in healthy individuals, enabling disease risk predictions for prevention. Yet, limited follow-up data can hinder risk predictive modeling (Pingault et al. 2018), particularly for less prevalent diseases.

Mendelian randomization (MR) is pivotal for identifying causal links between risk factors and health outcomes, utilizing genetic data across different cohorts (Smith and Ebrahim 2003; Sanderson et al. 2022). For instance, MR has elucidated the causal impact of risk factors such as cholesterol levels on cardiovascular disease, invalidating the protective role of high-density lipoprotein (HDL) cholesterol (Voight et al. 2012) and confirming the adverse effects of low-density lipoprotein (LDL) (Holmes et al. 2015). As interest grows in using MR for preventive healthcare (Glass et al. 2013; Chioloro 2018; Dixon et al. 2020; Yuan and Larsson 2020; Xu et al. 2022), we explore its potential for disease risk predictions as an alternative to longitudinal studies.

We present Predictive Risk modeling using Mendelian Randomization (PRiMeR), a novel framework for learning disease risk predictors through nonlinear functions of multiple risk factors leveraging genetic effects. To achieve this, PRiMeR utilizes risk factors and genetic data from a healthy cohort (Fig. 1A), and results from genome-wide association studies (GWAS) of diseases of interest (Fig. 1B). During training, PRiMeR fine-tunes the risk predictor

function to ensure that the genetic effects on both the predictor and the disease outcome are aligned across selected genetic variants (Fig. 1C), upholding MR's foundational principles. Once trained, disease risk in new patients can be assessed solely using risk factors (Fig. 1D). It is important to note that although PRiMeR employs the MR framework for disease risk predictions, it does not constitute a test for causality.

We validate PRiMeR's risk predictions through extensive simulations and in a type 2 diabetes (T2D) prediction task, leveraging follow-up labels for validation. Finally, we apply PRiMeR to identify a brain imaging predictor of Alzheimer's disease (AD) risk and an accelerometer-based predictor of Parkinson's disease (PD) risk.

## Results

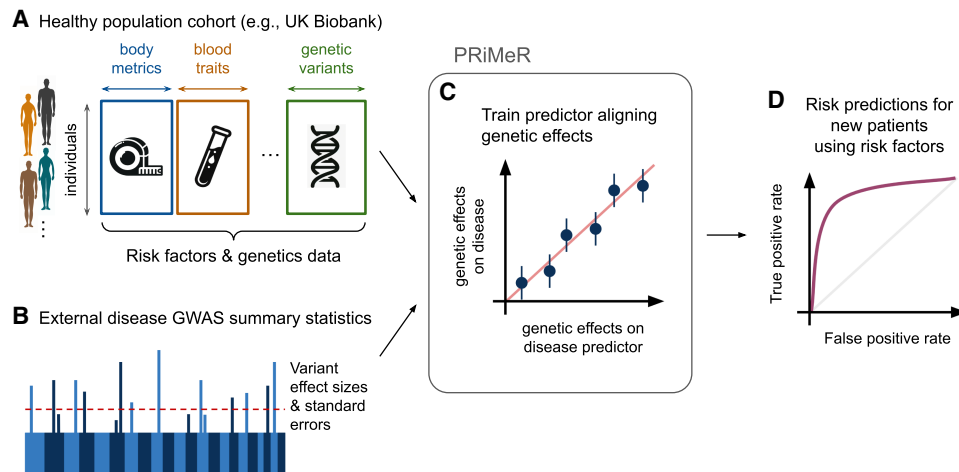
### Predictive risk modeling using Mendelian randomization

Traditionally, two-sample MR utilizes GWAS summary statistics of a risk factor (exposure, e.g., LDL cholesterol) and a disease (outcome, e.g., cardiovascular disease) from different cohorts to assess the directional effect of the exposure on the outcome (Fig. 2A). MR operates under the premise that, given essential assumptions, if an exposure causally influences an outcome, then the effects of exposure-associated variants on the outcome should be directly proportional to their effects on the exposure, with the slope of this proportionality quantifying the directional effect (Sanderson et al. 2022). Technically, for  $S$  independent exposure-associated variants, the directional effect  $\hat{\alpha}$  is estimated through inverse

**Corresponding author:** [francescopaolo.casale@helmholtz-munich.de](mailto:francescopaolo.casale@helmholtz-munich.de)

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.279252.124>. Freely available online through the *Genome Research* Open Access option.

© 2024 Sens et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.



**Figure 1.** Overview of the PRiMeR framework for disease risk prediction. (A) PRiMeR utilizes matched health metrics and genetic data from a cohort of healthy individuals. (B) It integrates these with disease-specific GWAS summary statistics from an external cohort. (C) The framework trains risk predictors to align genetic effects with those observed in disease outcomes, maintaining adherence to two-sample MR principles. (D) Posttraining, the model's accuracy in predicting disease risk is evaluated, for example, through the receiver operating characteristic curve against actual follow-up disease onset data.

variance weighting (IVW) regression (Burgess et al. 2013), where the genetic effects on the outcome ( $\beta_o \in R^S$ ) are regressed on the genetic effects on the exposure ( $\beta_e \in R^K$ ), while accounting for their standard errors ( $s_o \in R^S$ ) (Fig. 2B). A critical assumption of MR is the absence of horizontal pleiotropy; that is, exposure-associated genetic variants must affect the outcome solely via the exposure, without affecting alternative pathways (Verbanck et al. 2018).

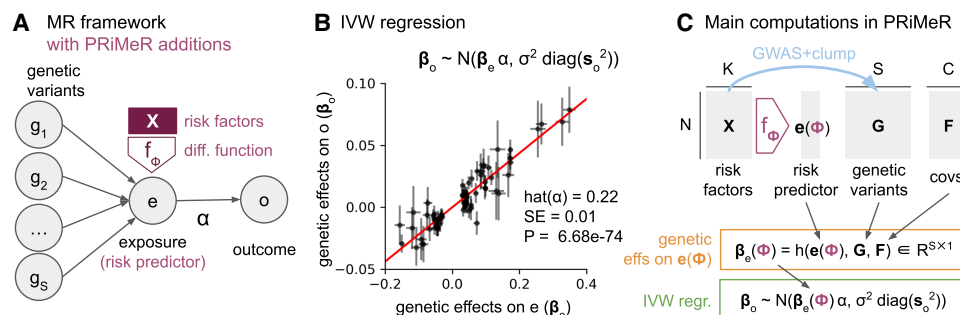
In this work, we investigate the use of the two-sample MR framework to learn disease risk predictors, enabling predictive modeling when longitudinal data are missing or scarce. To clarify how this could be feasible, we provide an illustrative example: Suppose directional effects  $\hat{\alpha}_1, \dots, \hat{\alpha}_K$  from  $K$  candidate risk factors  $x_1, \dots, x_K$  on a disease outcome are determined through two-sample MR. These effects can be aggregated to construct a linear risk predictor  $f(x) = \sum_{k \in C} \hat{\alpha}_k x_k$ , where  $C$  represents the set of significant directional effects.

With PRiMeR, we extend this concept to learn nonlinear risk predictors combining multiple risk factors, leveraging individual-level data from a genetic cohort of healthy individuals and dis-

ease-specific GWAS summary statistics. To do so, we introduce a differentiable function  $f$  parametrized by  $\phi$  aggregating multiple risk factors into a single risk predictor (Fig. 2A). The predictor is then fine-tuned to optimize the IVW regression. Briefly, for  $N$  individuals,  $K$  risk factors  $X \in R^{N \times K}$ ,  $C$  covariates  $F \in R^{N \times C}$ , and  $S$  independent genetic variants  $G \in R^{N \times S}$  associated with at least one of the  $K$  risk factors, the IVW regression loss can be computed as follows (Fig. 2C):

1. Compute aggregate risk predictor  $e(\phi) \in R^{N \times 1}$  from  $X$  using  $f_\phi$ .
2. Compute genetic effects  $\beta_e(\phi) \in R^{S \times 1}$  on the aggregate risk predictor as the marginal regression weights of each variant  $G_{:,s}$  on  $e(\phi)$  accounting for covariates  $F$ . This step mirrors the risk factor GWAS step in standard MR.
3. Compute IVW regression loss based on risk predictor genetic effects  $\beta_e(\phi)$ , and disease outcome statistics  $\beta_o$  and  $s_o$ ; that is,  $\mathcal{L}_{IVW}(\phi, \alpha, \sigma^2) = -\log \mathcal{N}(\beta_o | \beta_e(\phi)\alpha, \sigma^2 \text{diag}(s_o^2))$ .

As all these steps are differentiable,  $\phi$  can be learned through gradient-based optimization of the IVW loss (Methods). To select



**Figure 2.** Mathematical and computational details of the PRiMeR framework. (A) Diagram illustrating the core MR assumptions, where genetic variants ( $g_1, \dots, g_s$ ) influence exposure ( $e$ ), which in turn affects outcome ( $o$ ) with directional effect  $\alpha$ . Additions unique to PRiMeR are highlighted in purple: a risk predictor is computed as a differentiable function  $f_\phi$  (parametrized by  $\phi$ ) of risk factors  $X$ . (B) Illustration of IVW regression, where genetic variant effects on outcome ( $\beta_o$ ) are regressed on the aggregate risk predictor ( $\beta_e$ ), accounting for their standard errors ( $s_o$ ). (C) Main computations in PRiMeR, including computation of the risk predictor  $e(\phi)$ , the estimation of genetic effects on the risk predictor  $\beta_e(\phi)$ , and the computation of the IVW regression loss. The function  $h(e(\phi), G, F)$  returns marginal regression weights of each variant  $G_{:,s}$  on  $e(\phi)$  accounting for covariates  $F$ . As all these steps are differentiable,  $f_\phi$  can be learned through gradient-based optimization of the IVW regression loss.

independent risk factor-associated variants for our analyses, we performed univariate GWAS analyses for each risk factor followed by a multivariate clumping procedure (Methods). We considered the following nonlinear function of risk factors:

$$f_{\phi}(x) = \sum_{k=1}^K a_k g(b_k x_k + c_k),$$

with parameters  $\phi = \{a_1, \dots, a_K, b_1, \dots, b_K, c_1, \dots, c_K\}$ , where  $g$  is a nonlinear increasing warping function, a choice that enables modeling potential nonlinearities while being simple and clinically plausible. Such a shape function is commonly used in risk prediction as it captures the scenario where contributions from single risk factors remain minimal until a critical threshold and then escalate (Wainberg et al. 2019; Liang et al. 2020; Zhao et al. 2023).

### Validation of PRiMeR using simulated data

We evaluated the proposed PRiMeR framework through a series of simulations derived from UKB, encompassing 309,846 unrelated European individuals. We focused on 26 blood traits observed as potential risk factors in healthy individuals, simulating scenarios where subsets of these traits affect future health outcomes. Importantly, the aggregate risk was simulated as a linear combination of contributions from these factors, each transformed by a nonlinear increasing warping function to represent contributions that activate beyond specific thresholds (Methods). Our simulation framework allowed us to examine the efficacy of PRiMeR under various conditions, including the presence of horizontal pleiotropy and varying degrees of risk factor influence on outcome variance.

We compared PRiMeR with its linear variant (PRiMeR-LIN) and linear risk predictors based on two-sample univariate and multivariable Mendelian randomization (UVMR-based and MVMR-based, respectively; Methods). Beyond MR-derived models, we included a longitudinal reference model (LRM) trained directly on individual-level follow-up labels as a performance benchmark (Methods). Our evaluation maintained a strict two-sample framework, preventing any overlap between the cohorts used for determining genetic effects on risk factors and outcomes. We measured the accuracy of the risk predictions for all methods using Spearman's correlation coefficient, comparing estimated risk scores versus simulated ones in a held-out validation set. To ensure the calibration of our evaluation procedure, we verified models' performance was equivalent to random chance in control simulations without a directional effect (Supplemental Fig. A1).

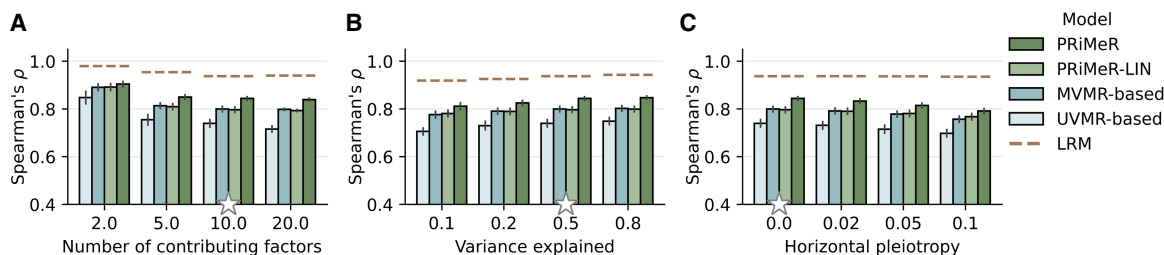
The findings from our simulations underscore the robustness and versatility of PRiMeR across a wide range of scenarios. Specifically, PRiMeR's performance in estimating risk remained stable when increasing the number of causal risk factors (Fig. 3A). This superior performance persisted across different values of the variance explained by the risk factors (Fig. 3B) and when simulating horizontal pleiotropy (Fig. 3C; Methods). We also assessed the robustness of PRiMeR across differently transformed contributions of single risk factors (Supplemental Fig. A2; Methods) and varying numbers of observed genetic variants strongly associated with the risk factors (Supplemental Fig. A3; Methods).

Finally, we note that although the LRM offers the highest predictive accuracy, PRiMeR's performance can be competitive if follow-up data are sparse (Supplemental Fig. A4). Collectively, these results highlight PRiMeR's potential as a powerful tool for predictive modeling when follow-up labels are scarce.

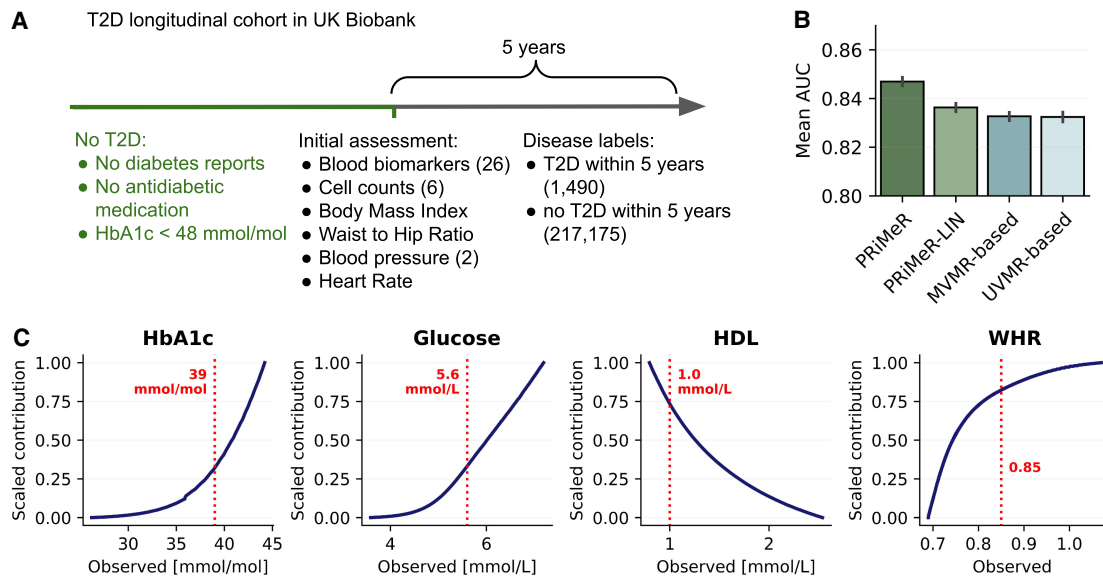
### Validation of PRiMeR in predicting 5-year type 2 diabetes risk

Next, we evaluated the prediction accuracy of PRiMeR in a real-world setting, considering a T2D data set derived from the UKB data set. Specifically, we aimed to predict 5-year T2D risk by leveraging risk factors and genetic data from 218,665 UKB individuals with no reported T2D at the initial assessment (Methods), and external GWAS summary statistics for T2D (Mahajan et al. 2018). As input risk factors, we used 37 traits previously linked to diabetes risk (Edlitz and Segal 2022), including metabolic, anthropometric, and cardiovascular metrics (Fig. 4A). We used 6077 independent genetic variants associated with at least one of the risk traits at the genome-wide significance level ( $P < 5 \times 10^{-8}$ ; Methods).

PRiMeR outperformed baseline MR methods, achieving an average AUC of 0.847 ( $\pm 0.002$ ) against 0.836 ( $\pm 0.002$ ) obtained using the MVMR-based predictor ( $P < 10^{-4}$ ) (Fig. 4B; Supplemental Fig. A5). Additionally, we evaluated MR-based model predictions against a polygenic risk score (PRS) model that relies exclusively on genetic data for predictions (Thompson et al. 2024; Methods), unlike MR-based models which use risk factors for prediction. Notably, the PRS model markedly underperformed compared to MR-based models, recording an AUC of 0.647 ( $\pm 0.002$ ) (Supplemental Fig. A5). Although a supervised reference model expectedly yielded the best performance when trained on full individual-level follow-up data, PRiMeR demonstrated competitiveness in scenarios with low numbers of follow-up labels (Supplemental Fig. A4).



**Figure 3.** Assessment of disease risk prediction accuracy using simulated data. Comparison of model accuracy in recovering the simulated aggregate risk factor measured by Spearman's correlation coefficient. Compared are PRiMeR, its linear variant (PRiMeR-LIN), a predictor based on multivariable MR (MVMR-based), a predictor based on univariate MR (UVMR-based), and the supervised model accessing individual-level follow-up labels (LRM; Methods), varying the number of contributing risk factors (A), the fraction of outcome variance explained by the risk factors (B), and the fraction of outcome variance explained by horizontal pleiotropy (C). Stars denote standard values held constant while other parameters were varied. Error bars indicate standard errors across 10 replicate experiments.



**Figure 4.** Validation of PRiMeR in predicting 5-year T2D risk. (A) Schematic representation of UKB T2D cohort, showing the inclusion criteria, the 37 risk factors included in our analysis, and the definition of the 5-year T2D onset labels. (B) Comparison of the mean area under the receiver operating characteristic curve (AUC) scores for 5-year T2D onset labels obtained using PRiMeR, its linear variant (PRiMeR-LIN), a predictor based on multivariable MR (MVMR-based) and univariate MR (UVMR-based). Error bars denote standard errors across 50 random train/test splits (Methods). (C) Scaled contributions to risk learned by PRiMeR as function of observed values for glycated hemoglobin (HbA1c), glucose, HDL, and waist-to-hip ratio (WHR). Risk reference thresholds are annotated in red.

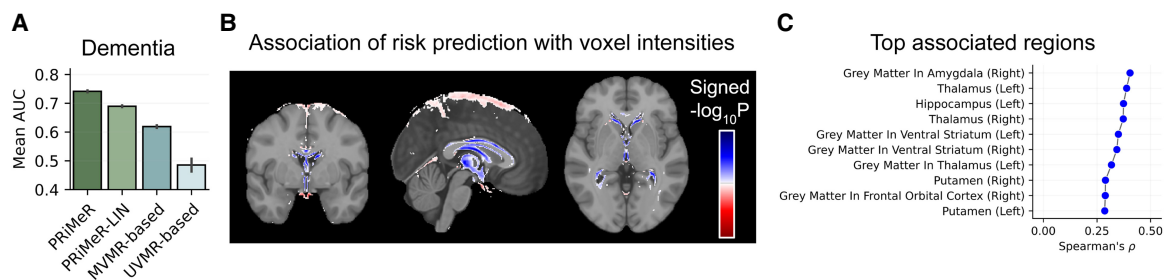
The risk predictor derived from PRiMeR robustly aligns with established clinical knowledge, underscored by its correlation with individual factors (Supplemental Fig. A6; Edlitz and Segal 2022). Notably, the nonlinear relationships identified by PRiMeR align with clinical expectations; for example, the risk contributions from glycated hemoglobin and glucose show significant increases nearing clinical risk thresholds (Fig. 4C). Overall, these results showcase the accuracy and clinical plausibility of risk predictors learned through PRiMeR in a real data setting.

#### Application of PRiMeR to predict 5-year Alzheimer's disease risk from brain imaging biomarkers

We applied PRiMeR to identify imaging biomarkers predictive of 5-year AD risk focusing on 31,552 unrelated European individuals in the UKB with brain imaging data. As imaging risk factors, we

considered 70 subcortical and gray matter volume traits from T1 MRI having at least five independent genome-wide significant signals, for a total of 353 independent genetic variants associated with at least one of these traits. As external AD GWAS results, we used AD GWAS summary statistics from Wightman et al. (2021).

All multivariable MR models exceeded the performance expected by chance (Fig. 5A), with PRiMeR achieving markedly higher accuracy compared to linear counterparts (PRiMeR AUC at  $0.741 \pm 0.003$  vs. PRiMeR-LIN at  $0.690 \pm 0.003$  vs. MVMR-based at  $0.629 \pm 0.004$ ) (Fig. 5B). A thorough analysis of the key imaging features pinpointed by PRiMeR for AD predictions underscored their correlation with reductions in gray matter and subcortical volume across various regions (Supplemental Fig. A7), particularly in the midbrain (Fig. 5B,C), consistent with known AD pathology (Knopman et al. 2021). Overall, these results underscore PRiMeR's effectiveness in utilizing genetic data for accurate risk prediction in the context of diseases with lower prevalence, such as AD.



**Figure 5.** Application of PRiMeR to predict 5-year AD risk. (A) Comparative performance of PRiMeR against baseline MR models using average AUC for 5-year AD predictions using follow-up labels. (B) Heatmap of the signed  $-\log_{10} P$ -value of association between voxel intensities and the AD risk predictor scores, overlaid on the MNI152 template (Miller et al. 2016; Alfaro-Almagro et al. 2018; <https://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152Nlin6>). Areas where increased risk predictor scores correlate with significant increased (decreased) voxel intensities are highlighted in red (blue) (Bonferroni-adjusted  $P < 0.05$ ). (C) Spearman's correlation coefficients between the AD risk predictor and individual MRI traits in the validation set. Results for the top 10 associated regions are displayed, with associations for all analyzed regions available in Supplemental Figure A7.

## Application of PRiMeR to predict 5-year Parkinson's disease risk from accelerometer features

We applied PRiMeR to learn the risk of 5-year PD, focusing on 69,670 unrelated European individuals in the UKB with accelerometer data. As risk factors, we considered 38 accelerometer-derived biomarkers having at least one independent genome-wide significant signal (Methods), and we considered external PD GWAS summary statistics from the FinnGen cohort (Kurki et al. 2023).

PRiMeR and PRiMeR-LIN achieved the highest accuracy (AUC of  $0.787 \pm 0.003$  and  $0.784 \pm 0.003$ , respectively). In contrast, the MVMR-based and UVMR-based predictors showed significantly lower accuracies with AUCs of  $0.521 \pm 0.006$  and  $0.519 \pm 0.006$ , respectively (Supplemental Fig. A8). When compared to a PRS model, PRiMeR demonstrated superior predictive performance (AUC of  $0.787 \pm 0.003$  vs.  $0.624 \pm 0.001$ ) (Supplemental Fig. A8). The analysis of key accelerometer features revealed their strong association with sleep duration and physical activity levels (Supplemental Fig. A9), confirming recent findings (Schalkamp et al. 2023).

## Discussion

In this study, we demonstrate that the two-sample MR framework can be extended to enable disease risk predictions using genetic information, without relying on longitudinal data. We introduce PRiMeR, a method for genetics-based risk predictions that leverages results from disease GWAS as supervisory signals for training risk predictors. The introduced approach is especially valuable given that genetic biobanks boast extensive health metrics but can lack longitudinal disease onset data for specific diseases, especially for those with lower incidence rates.

We validated PRiMeR through simulations and applications to predict diabetes from cardiovascular health indicators, AD from brain imaging biomarkers, and PD from accelerometer data. In simulations, PRiMeR outperformed baseline models and demonstrated robustness to horizontal pleiotropy, where genetic variants influence outcomes through alternative pathways, creating outliers in the regression of genetic effects central to two-sample MR. In real data applications, PRiMeR effectively recapitulated several established risk factors. In the T2D application, higher risk correlated with higher body mass index and waist-to-hip ratio (Belkina and Denis 2010; Burhans et al. 2018), lower levels of sex hormone-binding globulin (SHBG) (Huang et al. 2023), and an imbalance in cholesterol levels (Lewis and Steiner 1996; Sparks et al. 2012; Vergès 2015)—that is, higher LDL and lower HDL (Supplemental Fig. A6). In the brain imaging application, lower volumes in the amygdala, thalamus, and hippocampus correlated with higher AD risk (Supplemental Fig. A7), aligning with known disease pathogenesis (Vereecken et al. 1994; de Jong et al. 2008; Poulin et al. 2011). Finally, lower sleep duration and reduced physical activity were linked to increased PD risk (Supplemental Fig. A9), confirming known early stage symptoms (Xu et al. 2010; Lysen et al. 2019; Schalkamp et al. 2023).

Despite its advantages, PRiMeR is not without limitations. PRiMeR requires genetic cohorts for risk factors and outcomes to be sampled from the same population, and failure to meet these criteria may lead to challenges due to variations in linkage disequilibrium patterns (The International HapMap Consortium et al. 2007), necessitating the integration of robust instrument selection

strategies, such as variant fine-mapping (Cai et al. 2023). Although the simple nonlinearity implemented in PRiMeR enables interpretability, it may fail to capture more complex relationships between risk factors. Although extensions to more complex parametric forms or flexible neural network functions would mitigate this, managing overfitting will represent a key challenge. Future work to address this may involve extending PRiMeR's Bayesian framework by integrating recent developments in deep probabilistic models (Kingma and Welling 2014; Nikolentzos et al. 2023). Furthermore, incorporating explicit mechanisms to counter weak instrument bias (Wang and Kang 2022) and horizontal pleiotropy (Sanderson et al. 2022) are critical areas for further development.

Finally, although our focus has been on predicting disease risk without reliance on longitudinal data, utilizing a causal inference framework for disease prediction may provide a viable method to mitigate confounding in longitudinal data sets (Pingault et al. 2018), potentially enhancing the generalization of risk predictors across different data sets. The potential of PRiMeR to facilitate this exploration opens exciting avenues for future research, particularly as more cohorts with deeper phenotype data become available.

As we look to the future, we identify three key areas where PRiMeR can make a significant impact. It offers promising solutions for diseases with low prevalence and well-developed GWAS, such as Alzheimer's, Parkinson's, amyotrophic lateral sclerosis, and bipolar disorder. These applications are critical where traditional longitudinal analyses are often limited by small sample sizes, particularly for specialized biomarker modalities—for example, only 19 individuals developed AD within 5 years in the T1 brain MRI cohort in our analysis. Additionally, PRiMeR has potential applications in underdiagnosed diseases such as attention deficit hyperactivity disorder, depression, and fatty liver disease, all of which have robust GWAS but lack reliable diagnostic labels for longitudinal analysis. Finally, we are poised to extend PRiMeR's application to molecular genetic data sets, such as bulk and single-cell expression quantitative trait loci data sets (Lonsdale et al. 2013; van der Wijst et al. 2020; Yazar et al. 2022), where longitudinal information is typically not available.

## Methods

### Predictive risk modeling utilizing Mendelian randomization

#### *Two-sample Mendelian randomization and inverse variance weighting*

Two-sample MR leverages summary statistics of GWAS of a risk factor (exposure) and a health outcome to infer the causal effect of the exposure on the outcome. Assuming  $S$  independent genetic variants associated with the analyzed exposure, this can be estimated through the IVW regression (Burgess et al. 2013):

$$\beta_o \sim \mathcal{N}(\beta_e \alpha, \sigma^2 \text{diag}(s_o^2)), \quad (1)$$

where  $\mathcal{N}$  denotes the multivariate normal distribution,  $\beta_o \in R^S$  denotes the variant effects on the outcome,  $s_o \in R^S$  is the standard errors,  $\beta_e \in R^S$  is the variant effects on the exposure,  $\alpha$  is the regression slope, and  $\sigma^2$  is the variance of the regression error. Within this framework, the causal effect of the exposure on the outcome is the maximum likelihood estimator of the regression slope  $\alpha$ , that

$$\text{is, } \hat{\alpha} = \frac{\sum_{i=1}^S w_i \beta_{e,i} \beta_{o,i}}{\sum_{i=1}^S w_i (\beta_{e,i})^2} \text{ with standard error } \text{SE}(\hat{\alpha}) = \frac{1}{\sqrt{\sum_{i=1}^S w_i (\beta_{e,i})^2}},$$

where  $w_i = \frac{1}{(s_{o,i})^2}$ . MR is an instrumental variable analysis method (Angrist et al. 1996), using the genetic variants associated with the exposure as instruments. As such, it relies on key assumptions (Sanderson et al. 2022): (1) The chosen instruments are robustly associated with the exposure; (2) the instruments are independent of any confounders that may influence both the exposure and the outcome; and (3) the instruments influence the outcome only through the exposure, that is, no horizontal pleiotropy. Moreover, for the causal effect estimate to be valid, the exposure and outcome statistics need to be estimated on independent cohorts sampled from the same population (Zhao et al. 2019).

### Predictive risk modeling utilizing Mendelian randomization

In classical two-sample MR,  $\beta_e$  is retrieved from the GWAS of a single risk factor. However, given access to individual-level data and multiple risk factors, it is feasible to define an aggregate risk factor as a function of these factors and compute  $\beta_e$  by regressing genetic instruments against this aggregate risk factor. Importantly, if the function  $f_\phi$  parametrized by  $\phi$  is differentiable, the corresponding genetic effects on the aggregate risk factor  $\beta_e(\phi)$  are also differentiable, enabling the learning of an aggregate risk function  $f_\phi$  by directly optimizing the IVW regression loss through gradient descent. For  $N$  individuals,  $K$  risk factors  $X \in R^{N \times K}$ ,  $C$  covariates  $F \in R^{N \times C}$ , and  $S$  independent genetic variants  $G \in R^{N \times S}$ , each associated with at least one of the  $K$  risk factors, the IVW regression loss can be computed as

$$\mathcal{L}_{\text{IVW}}(\phi, \alpha, \sigma^2) = -\log \mathcal{N}(\beta_o \mid \beta_e(\phi), \alpha, \sigma^2 \text{diag}(s_o)),$$

where the genetic effects of the aggregated risk factor  $\beta_e(\phi)$  are computed as the marginal regression weights of each variant  $G_{\cdot,1}, \dots, G_{\cdot,S}$  on  $f_\phi(X)$  accounting for covariates  $F$  (Supplemental Information). As  $\mathcal{L}_{\text{IVW}}$  is fully differentiable in  $\phi, \alpha, \sigma^2$ , the predictor function  $f_\phi$  is end-to-end trainable. Regarding the analytical form of  $f$ , we opted for a linear combination of nonlinear increasing warping functions of single risk factors:

$$f_\phi(X) = \sum_{k=1}^K a_k \text{ELU}(b_k X_k + c_k),$$

with parameters  $\phi = \{a_1, \dots, a_K, b_1, \dots, b_K, c_1, \dots, c_K\}$  and where  $\text{ELU}(\cdot)$  is the exponential linear unit function (Clevert et al. 2015). This formulation assumes contributions from single risk factors remain minimal until a critical threshold is reached, after which they escalate (Wainberg et al. 2019; Liang et al. 2020; Zhao et al. 2023). Note that PRiMeR reduces to multivariable MR when selecting a linear function for  $f_\phi$  (Supplemental Information; Burgess and Thompson 2015; Sanderson et al. 2020), underscoring the robust foundation and adaptability of our approach. An overview of related methods to PRiMeR is detailed in Supplemental Information.

### Bayesian model and optimization

To enhance PRiMeR's robustness in scenarios with a limited number of genetic variants robustly associated with the analyzed risk factors, we implemented a Bayesian inference approach. This involved introducing priors over the parameters  $\phi$  and optimizing the log marginal likelihood of the IVW model. For parameters

where analytical integration was infeasible, mean-field variational inference was utilized to derive the evidence lower bound (ELBO). Optimization of the ELBO was achieved through gradient descent using the Adam optimizer, incorporating the reparametrization trick to enable backpropagation through the expectation term of the ELBO. This approach aligns with standard practices in variational inference methods that leverage gradient descent (Kingma and Welling 2014; Ranganath et al. 2014; Engelmann et al. 2024). The learning rate for all experiments was fixed at 0.01, and we consistently applied gradient clipping with a norm bound of one while training for 1000 epochs. Risk predictions were obtained as the mean of the variational posterior of the model. Comprehensive details on our Bayesian model and variational inference procedure can be found in Supplemental Information. Finally, we note that prior to all experiments, risk factors were normalized using a rank-inverse Gaussian transformation, a widely used phenotype transformation for GWAS analyses (McCaw et al. 2020). Our PRiMeR framework was implemented in PyTorch (Paszke et al. 2019).

### Selection of genetic variants

To identify genetic variants associated with risk factors, we first conducted a univariate GWAS for each risk factor followed by a multivariate clumping procedure. GWAS analysis utilized linear regression via GCTA (fastGWA-Ir functionality) (Jiang et al. 2019), adjusting for sex, age, UKB array type, and the top 20 genetic principal components. Adjusting for the top 20 genetic principal components is a standard practice to correct for population structure in genetic analyses of unrelated Europeans (Price et al. 2006). After GWAS, we applied clumping on the minimum  $P$ -value statistics across all traits using PLINK (Purcell et al. 2007), with parameters fixed to a  $P$ -value threshold of  $5 \times 10^{-8}$ , an  $r^2$  linkage disequilibrium cutoff of 0.05, and a clumping window of 5000 kb, following Zhu et al. (2018). This procedure ensured that selected variants are approximately independent and associated with at least one of the risk factors at genome-wide significant level ( $P < 5 \times 10^{-8}$ ).

### Comparison models

In our study, we assess the performance of PRiMeR in comparison to predictive models based on univariate Mendelian randomization (UVMR-based) and multivariate Mendelian randomization (MVMR-based). Both UVMR-based and MVMR-based prediction methods apply a linear risk prediction function  $f(X) = \sum_{k=1}^K a_k X_{\cdot,k}$ ,

where  $X$  represents risk factors and  $a_k$  represents the estimated causal effects. UVMR-based determines  $a_k$  from MR-estimated causal effects  $\hat{\alpha}_k$  for each risk factor  $X_k$ , requiring at least five genome-wide significant genetic instruments, assigning  $a_k=0$  for factors with a lower number of instruments or nonsignificant causal effects (Bonferroni-corrected  $P < 0.05$ ). Conversely, MVMR-based determines  $a_k$  jointly across all risk factors, through a multivariate regression of the risk factor effect size matrix  $B_e \in R^{S \times K}$  on the outcome effect sizes  $\beta_o$  (Burgess and Thompson 2015; Sanderson 2021). Additionally, we contrast PRiMeR's performance with its linear counterpart, PRiMeR-LIN, adopting a linear disease risk prediction function  $f$ . As performance benchmarks, we also included supervised models trained directly on individual-level follow-up labels. We primarily compared PRiMeR against a LRM using the same risk prediction function as PRiMeR. Additionally, we conducted comparisons with ElasticNet, RandomForest, and XGBoost models. Hyperparameters for all models with access to individual-level follow-up labels were optimized using an inner fivefold cross-validation procedure (Supplemental Information). In simulation

scenarios, these models aimed to minimize the mean squared error, whereas in the T2D study, the objective was to minimize the binary cross-entropy loss.

## Simulations

### Data set generation

In our simulation study, we used 26 blood traits from 309,865 unrelated Europeans from the UKB data set, as potential risk factors. The data set is available at <https://biobank.ndph.ox.ac.uk/ukb/> after a registration and approval process. We crafted scenarios where a subset of these traits exerted a causal influence on the health outcome, with individual risk contributions following a nonlinear increasing warping function—initially remaining negligible until surpassing a certain threshold, beyond which they increased linearly. The health outcome was generated as the sum of a linear combination of these nonlinearly transformed risk factors, a horizontal pleiotropy effect, and Gaussian noise. The horizontal pleiotropy effect was simulated as a direct genetic contribution from a subset of the variants associated with the blood traits. We systematically varied key parameters, such as the number of causal risk factors, and the proportions of the outcome variance explained by the risk factors and the horizontal pleiotropy effect, respectively. Additionally, we created a scenario in which we controlled the sharpness of the risk function, and whether it saturates after surpassing a certain threshold, instead of growing linearly. We jointly validated different values of sharpness and saturation, ranging from very flat risks to very abrupt risk increases, and from fully unbounded risks to those asymptotically approaching an upper bound. Finally, to assess the robustness of our model under conditions of limited genetic data, we explored scenarios with fewer genetic variants by randomly subsampling from the full set of variants associated with blood traits. For each simulation parameter configuration, we considered 10 repeat experiments, utilizing distinct random seeds. Detailed descriptions of our simulation approach can be found in [Supplemental Information](#).

### Evaluation framework

Adhering to a two-sample framework, the data were divided evenly into risk factor and outcome cohorts, with this consistent split maintained throughout all simulated scenarios. Through the genetic variant selection procedure detailed above, 2904 independent genetic variants were identified in the risk factor cohort. Using the outcome cohort, the effects of these genetic variants on the outcome were estimated—this step substitutes the real data analysis process of obtaining genetic effects on the outcome from external GWAS results. We trained PRiMeR using 80% of the risk factor cohort and evaluated the risk prediction accuracy on the remaining 20%. Prediction accuracy was assessed by calculating Spearman's correlation coefficient between the predicted and simulated risk values. Within this evaluation framework, we compare PRiMeR against PRiMeR-LIN, as well as UVMR-based and MVMR-based predictors. We also compared PRiMeR with models trained on individual-level follow-up data, including the LRM, ElasticNet, RandomForest, and XGBoost models ([Supplemental Fig. A4](#)). For these models, we considered inner fivefold cross-validation for hyperparameter selection (grid of explored values in [Supplemental Information](#)). Standard errors for all metrics were calculated from the results of 10 repeat experiments. To ensure the calibration of the evaluation procedure, Spearman's correlation coefficients of all MR-based models were verified to be compatible with zero in simulations without causal links ([Supplemental Fig. A1](#)).

## Diabetes risk predictions

### Cohort definition

We utilized PRiMeR to predict 5-year T2D risk from 37 established risk factors (Edlitz and Segal 2022). For the outcome genetic effects and standard errors, we considered the external T2D GWAS summary statistics from Mahajan et al. (2018), which excluded the UKB cohort and can be obtained from <http://www.type2diabetesgenetics.org/>. For the risk factor cohort, we considered 218,665 unrelated Europeans from UKB who did not have diabetes at the time of assessment. After matching variants across the two data sets and excluding palindromic variants, the genetic variant selection procedure described above identified 6077 independent genetic variants associated with at least one of the 37 traits. More info on the longitudinal cohort definition can be found in [Supplemental Information](#).

### Evaluation framework

We compared PRiMeR with MR-based models (PRiMeR-LIN, UVMR-based, and MVMR-based) and models with direct access to individual-level follow-up data (LRM, ElasticNet, RandomForest, and XGBoost) ([Supplemental Information](#)). Additionally, we included a PRS predictor, computed externally by Thompson et al. (2024) and made available in UKB through field 26285, for comparison. To assess T2D risk prediction accuracy, we employed the AUC, using actual 5-year T2D risk as labels (derived from fields 41280 and 41270 using ICD10 code E11). To ensure robust significance testing and estimate standard errors, we conducted 50 repeat experiments with random 80%/20% splits for training and testing. Standard errors for all metrics were computed across these experiments, along with *t*-tests to assess performance improvements.

### Interpretation of the learned T2D biomarker

To assess the risk predictor learned by PRiMeR in the T2D experiments, we employed analysis. Firstly, univariate associations were quantified using Spearman's correlation coefficients between the risk predictor and each input risk factor in out-of-sample individuals ([Supplemental Fig. A6](#)). Secondly, to evaluate the model's ability to capture nonlinear relationships among the selected risk factors, we visualized the learned contributions (normalized between 0 and 1) against observed values of these factors ([Fig. 4C](#)). We observed that all reported values are highly consistent across all 50 repeat experiments.

## Imaging biomarkers of dementia

### Cohort definition

We employed two-sample MR methods to estimate the 5-year AD risk based on brain T1 MRI features. For the risk factor cohort, we selected 31,552 unrelated Europeans from UKB with T1 brain imaging data. Out of the 153 brain volume features that are available in UKB, we included 70 brain T1 MRI traits with at least five associated variants  $P < 5 \times 10^{-8}$  as risk factors in all MR models, for which we identified a total of 385 genetic variants. More info on the risk factor cohort can be found in [Supplemental Information](#). For outcome genetic effects, we utilized external GWAS summary stats for AD in unrelated Europeans from Wightman et al. (2021), which we downloaded from [https://ctg.cncr.nl/software/summary\\_statistics](https://ctg.cncr.nl/software/summary_statistics).

## Evaluation

We compared PRiMeR with PRiMeR-LIN, as well as UVMR-based and MVMR-based predictors. We assessed AD risk prediction accuracy by AUC using actual 5-year AD risk as labels (derived from field 131036). Across 31,552 unrelated Europeans from UKB with T1 brain imaging data, only 19 developed AD within 5 years. All models were trained on 80% of the healthy risk factor cohort, while the remaining 20% along with 19 individuals with reported AD were used as a test set. To robustly test for significance and estimate standard errors, we conducted 50 repeat experiments, each employing different random 80%/20% splits.

## Interpretation of the learned AD biomarker

For the interpretation of the learned risk predictor, we conducted a voxel-based association analysis using T1-weighted MRI scans that were registered to the MNI152 template (Grabner et al. 2006; Miller et al. 2016; Alfaro-Almagro et al. 2018; <https://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152Nlin6>). Each voxel's intensity was regressed against the out-of-sample individual's risk predictor scores, adjusting for sex, age, UKB array type, and the top 20 genetic principal components. This linear association test yielded  $P$ -values for the contribution of the risk predictor to each voxel, which were transformed into a heatmap overlay on the MNI152 template using signed  $P$ -values ( $-\log_{10}(P) \cdot \text{sign}(\beta)$ ). Blue regions on the heatmap indicate areas of volume decrease associated with increased risk predictor values, highlighting potential areas that correlate with higher AD risk (Fig. 4B). Furthermore, we quantified Spearman's correlation coefficient between the risk predictor and each input imaging trait within the held-out validation set (Fig. 4C; Supplemental Fig. A7). Overall, both analyses displayed remarkable consistency across all experimental repeats, underscoring the reliability of our findings.

## Accelerometer-based biomarkers for Parkinson's disease risk prediction

### Cohort definition

We used PRiMeR to predict 5-year PD risk from accelerometer-derived features from Schalkamp et al. (2023). For outcome genetic effects and standard errors, we used external PD GWAS summary statistics from the FinnGen cohort, which is available at [https://www.finnngen.fi/en/access\\_results](https://www.finnngen.fi/en/access_results) (Freeze 11; G6\_PARKINSON). Our risk factor cohort included 69,670 from UKB without PD at the time of accelerometer data collection (field 90003). After matching variants across data sets and removing palindromic variants, we identified 45 independent variants associated ( $P < 5 \times 10^{-8}$ ) with at least one of the 38 accelerometer traits, which were used as input features for our predictors. Due to the lower number of instruments, we repeated the experiment with a relaxed significance threshold for the inclusion of variants in genetics-based predictive modeling ( $P < 10^{-6}$ ), which yielded 185 variants and confirmed the robustness of our results across both thresholds (Supplemental Fig. A10). The Supplemental Information provides more details on the longitudinal cohort definition and the exact names of the considered features.

### Evaluation framework

We compared the performance of PRiMeR against PRiMeR-LIN, as well as UVMR-based and MVMR-based predictors. Additionally, we included a PRS predictor, computed externally by Thompson et al. (2024) and made available in UKB through field 26260, for comparison. We assessed PD risk predictions, using the area under the receiver operating characteristic curve (AUC ROC) as our pri-

mary metric, based on actual 5-year PD risk labels (field 131022). Our cohort comprised 69,670 unrelated Europeans from the UKB with accelerometer data, of whom 128 developed PD within 5 years. We trained all models on 80% of healthy participants from the risk factor cohort and tested on the remaining 20%, along with the 128 PD cases. To ensure the robustness of our results, we show the standard error across 50 random 80%/20% splits.

### Assessment of risk contribution

We computed Spearman's correlation coefficients between the predicted risk and the input accelerometer features within the test set (Supplemental Fig. A9). The values remained highly consistent across all 50 repeat experiments, highlighting the robustness of the results.

### Use of artificial intelligence

In the preparation of this manuscript, we utilized the large language model GPT-4 (<https://chat.openai.com/>) for editing assistance, including language polishing and clarification of text. Although this tool assisted in refining the manuscript's language, it was not used to generate contributions to the original research, data analysis, or interpretation of results. All final content decisions and responsibilities rest with the authors.

### Software availability

An open-source software implementation of PRiMeR and all baseline methods are available at GitHub (<https://github.com/AIH-SGML/PRiMeR>), Zenodo (<https://doi.org/10.5281/zenodo.13632773>), and as Supplemental Code.

## Competing interest statement

The authors declare no competing interests.

## Acknowledgments

We thank Julien Gagneur for his feedback on the manuscript. This research has been conducted using the UK Biobank Resource (Application number 87065). F.P.C. and D.S. were funded by the Free State of Bavaria's Hightech Agenda through the Institute of AI for Health (AIH). L.S. and B.M.E. acknowledge the support of the Friedrich-Alexander-Universität Erlangen-Nürnberg under the joint research school Munich School for Data Science (MUDS).

**Author contributions:** F.P.C. conceived the study and supervised the work. D.S., L.G., and F.P.C. implemented the methods. D.S., L.S., L.G., M.G., and F.P.C. analyzed the data. M.G. contributed to the preprocessing of the accelerometry data and initial prediction analyses. B.M.E. contributed critical insights to the design and interpretation of the accelerometer biomarker analysis. D.S., L.S., L.G., and F.P.C. wrote the initial draft, with all authors contributing to subsequent revisions and refinements of the manuscript.

## References

- Abul-Husn NS, Soper ER, Odgis JA, Cullina S, Bobo D, Moscati A, Rodriguez JE, CBIPM Genomics Team, Regeneron Genetics Center, Loos RJF, et al. 2019. Exome sequencing reveals a high prevalence of BRCA1 and BRCA2 founder variants in a diverse population-based biobank. *Genome Med* **12**: 2. doi:10.1186/s13073-019-0691-1
- Alfaro-Almagro F, Jenkinson M, Bangerter NK, Andersson JLR, Griffanti L, Douaud G, Sotiropoulos SN, Jbabdi S, Hernandez-Fernandez M, Vallee E, et al. 2018. Image processing and Quality Control for the first

- 10,000 brain imaging datasets from UK Biobank. *Neuroimage* **166**: 400–424. doi:10.1016/j.neuroimage.2017.10.034
- Angrist JD, Imbens GW, Rubin DB. 1996. Identification of causal effects using instrumental variables. *J Am Stat Assoc* **91**: 444–455. doi:10.1080/01621459.1996.10476902
- Belkina AC, Denis GV. 2010. Obesity genes and insulin resistance. *Curr Opin Endocrinol Diabetes Obes* **17**: 472–477. doi:10.1097/MED.0b013e328333c5c48
- Burgess S, Thompson SG. 2015. Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. *Am J Epidemiol* **181**: 251–260. doi:10.1093/aje/kwu283
- Burgess S, Butterworth A, Thompson SG. 2013. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol* **37**: 658–665. doi:10.1002/gepi.21758
- Burhans MS, Hagman DK, Kuzma JN, Schmidt KA, Kratz M. 2018. Contribution of adipose tissue inflammation to the development of type 2 diabetes mellitus. *Compr Physiol* **9**: 1–58. doi:10.1002/cphy.c170040
- Cai M, Wang Z, Xiao J, Hu X, Chen G, Yang C. 2023. XMAP: cross-population fine-mapping by leveraging genetic diversity and accounting for confounding bias. *Nat Commun* **14**: 6870. doi:10.1038/s41467-023-42614-7
- Chiolero A. 2018. Why causality, and not prediction, should guide obesity prevention policy. *Lancet Public Health* **3**: e461–e462. doi:10.1016/S2468-2667(18)30158-0
- Clevert D-A, Unterthiner T, Hochreiter S. 2016. Fast and accurate deep network learning by exponential linear units (ELUs). In *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, 2016, Conference Track Proceedings* (ed. Bengio Y, LeCun Y). <http://arxiv.org/abs/1511.07289>.
- de Jong LW, van der Hiele K, Veer IM, Houwing JJ, Westendorp RGJ, Bollen ELEM, de Bruin PW, Middelkoop HAM, van Buchem MA, van der Grond J. 2008. Strongly reduced volumes of putamen and thalamus in Alzheimer's disease: an MRI study. *Brain* **131**: 3277–3285. doi:10.1093/brain/awn278
- Dixon P, Hollingworth W, Harrison S, Davies NM, Davey Smith G. 2020. Mendelian randomization analysis of the causal effect of adiposity on hospital costs. *J Health Econ* **70**: 102300. doi:10.1016/j.jhealeco.2020.102300
- Edlitz Y, Segal E. 2022. Prediction of type 2 diabetes mellitus onset using logistic regression-based scorecards. *eLife* **11**: e71862. doi:10.7554/eLife.71862
- Engelmann JP, Palma A, Tomczak JM, Theis FJ, Casale FP. 2024. Mixed models with multiple instance learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 3664–3672. PMLR.
- Glass TA, Goodman SN, Hernán MA, Samet JM. 2013. Causal inference in public health. *Annu Rev Public Health* **34**: 61–75. doi:10.1146/annurev-publhealth-031811-124606
- Grabner G, Janke AL, Budge MM, Smith D, Pruessner J, Collins DL. 2006. Symmetric atlas and model based segmentation: an application to the hippocampus in older adults. *Med Image Comput Comput Assist Interv* **9**: 58–66. doi:10.1007/11866763\_8
- Holmes MV, Asselbergs FW, Palmer TM, Drenos F, Lanktree MB, Nelson CP, Dale CE, Padmanabhan S, Finan C, Swerdlow DI, et al. 2015. Mendelian randomization of blood lipids for coronary heart disease. *Eur Heart J* **36**: 539–550. doi:10.1093/eurheartj/ehs571
- Huang R, Wang Y, Yan R, Ding B, Ma J. 2023. Sex hormone binding globulin is an independent predictor for insulin resistance in male patients with newly diagnosed type 2 diabetes mellitus. *Diabetes Ther* **14**: 1627–1637. doi:10.1007/s13300-023-01445-x
- The International HapMap Consortium, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, Belmont JW, Boudreau A, Hardenbol P, et al. 2007. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**: 851–861. doi:10.1038/nature06258
- Jiang L, Zheng Z, Qi T, Kemper KE, Wray NR, Visscher PM, Yang J. 2019. A resource-efficient tool for mixed model association analysis of large-scale data. *Nat Genet* **51**: 1749–1755. doi:10.1038/s41588-019-0530-8
- Kingma DP, Welling M. 2014. Auto-encoding variational Bayes. In *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, 2014, Conference Track Proceedings* (ed. Bengio Y, LeCun Y). <http://arxiv.org/abs/1312.6114>
- Knopman DS, Amieva H, Petersen RC, Chételat G, Holtzman DM, Hyman BT, Nixon RA, Jones DT. 2021. Alzheimer disease. *Nat Rev Dis Primers* **7**: 33. doi:10.1038/s41572-021-00269-y
- Kurki MI, Karjalainen J, Palta P, Sipilä TP, Kristiansson K, Donner KM, Reeve MP, Laivuori H, Aavikko M, Kaunisto MA, et al. 2023. Finngen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**: 508–518. doi:10.1038/s41586-022-05473-8
- Leitsalu L, Haller T, Esko T, Tammesoo M-L, Alavere H, Snieder H, Perola M, Ng PC, Mägi R, Milani L, et al. 2015. Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *Int J Epidemiol* **44**: 1137–1147. doi:10.1093/ije/dyt268
- Lewis GF, Steiner G. 1996. Acute effects of insulin in the control of VLDL production in humans. Implications for the insulin-resistant state. *Diabetes Care* **19**: 390–393. doi:10.2337/diacare.19.4.390
- Liang F, Liu F, Huang K, Yang X, Li J, Xiao Q, Chen J, Liu X, Cao J, Shen C, et al. 2020. Long-term exposure to fine particulate matter and cardiovascular disease in China. *J Am Coll Cardiol* **75**: 707–717. doi:10.1016/j.jacc.2019.12.031
- Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, Hasz R, Walters G, Garcia F, Young N, et al. 2013. The genotype-tissue expression (GTEx) project. *Nat Genet* **45**: 580–585. doi:10.1038/ng.2653
- Lysen TS, Darweesh SKL, Ikram MK, Luik AI, Ikram MA. 2019. Sleep and risk of parkinsonism and Parkinson's disease: a population-based study. *Brain* **142**: 2013–2022. doi:10.1093/brain/awz113
- Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, Payne AJ, Steinthorsdottir V, Scott RA, Garup N, et al. 2018. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* **50**: 1505–1513. doi:10.1038/s41588-018-0241-6
- McCaw ZR, Lane JM, Saxena R, Redline S, Lin X. 2020. Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. *Biometrics* **76**: 1262–1272. doi:10.1111/biom.13214
- Miller KL, Alfaro-Almagro F, Bangerter NK, Thomas DL, Yacoub E, Xu J, Bartsch AJ, Jbabdi S, Sotiropoulos SN, Andersson JLR, et al. 2016. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci* **19**: 1523–1536. doi:10.1038/nn.4393
- Nikolentzos G, Vazirgiannis M, Xypolopoulos C, Lingman M, Brandt EG. 2023. Synthetic electronic health records generated with variational graph autoencoders. *NPJ Digit Med* **6**: 83. doi:10.1038/s41746-023-00822-x
- Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimeshain N, Antiga L, et al. 2019. PyTorch: an imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems* (ed. Wallach H, et al.), Vol. 32. Curran Associates, Inc. [https://proceedings.neurips.cc/paper\\_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf).
- Pingault J-B, O'Reilly PF, Schoeler T, Ploubidis GB, Rijdsdijk F, Dudbridge F. 2018. Using genetic data to strengthen causal inference in observational research. *Nat Rev Genet* **19**: 566–580. doi:10.1038/s41576-018-0020-3
- Poulin SP, Dautoff R, Morris JC, Barrett LF, Dickerson BC, Alzheimer's Disease Neuroimaging Initiative. 2011. Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiatry Res* **194**: 7–13. doi:10.1016/j.pscychres.2011.06.014
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. 2006. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**: 904–909. doi:10.1038/ng1847
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559–575. doi:10.1086/519795
- Ranganath R, Gerrish S, Blei D. 2014. Black box variational inference. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics* (ed. Kaski S, Corander J), Vol. 33 of Proceedings of Machine Learning Research, pp. 814–822. PMLR, Reykjavik, Iceland.
- Sanderson E. 2021. Multivariable Mendelian randomization and mediation. *Cold Spring Harb Perspect Med* **11**: a038984. doi:10.1101/cshperspect.a038984
- Sanderson E, Smith GD, Windmeijer F, Bowden J. 2020. Corrigendum to: an examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol* **49**: 1057. doi:10.1093/ije/dyaa101
- Sanderson E, Glymour MM, Holmes MV, Kang H, Morrison J, Munafò MR, Palmer T, Schouling CM, Wallace C, Zhao Q, et al. 2022. Mendelian randomization. *Nat Rev Methods Primers* **2**: 6. doi:10.1038/s43586-021-00092-5
- Schalkamp A-K, Peall KJ, Harrison NA, Sandor C. 2023. Wearable movement-tracking data identify Parkinson's disease years before clinical diagnosis. *Nat Med* **29**: 2048–2056. doi:10.1038/s41591-023-02440-2
- Smith GD, Ebrahim S. 2003. "Mendelian randomization": can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol* **32**: 1–22. doi:10.1093/ije/dyg070
- Sparks JD, Sparks CE, Adeli K. 2012. Selective hepatic insulin resistance, VLDL overproduction, and hypertriglyceridemia. *Arterioscler Thromb Vasc Biol* **32**: 2104–2112. doi:10.1161/ATVBAHA.111.241463
- Sudlog 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**: e1001779. doi:10.1371/journal.pmed.1001779
- Thompson DJ, Wells D, Selzam S, Peneva I, Moore R, Sharp K, Tarran WA, Beard EJ, Riveros-Mckay F, Giner-Delgado C, et al. 2024. A systematic evaluation of the performance and properties of the UK Biobank Polygenic Risk Score (PRS) Release. *PLoS One* **19**: e0307270. doi:10.1371/journal.pone.0307270
- van der Wijst M, de Vries DH, Groot HE, Trynka G, Hon CC, Bonder MJ, Stegle O, Nawijn MC, Idaghmour Y, van der Harst P, et al. 2020. The single-cell eQTLGen consortium. *eLife* **9**: e52155. doi:10.7554/eLife.52155
- Verbanck M, Chen C-Y, Neale B, Do R. 2018. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet* **50**: 693–698. doi:10.1038/s41588-018-0099-7
- Vereecken TH, Vogels OJ, Nieuwenhuys R. 1994. Neuron loss and shrinkage in the amygdala in Alzheimer's disease. *Neurobiol Aging* **15**: 45–54. doi:10.1016/0197-4580(94)90143-0
- Vergès B. 2015. Pathophysiology of diabetic dyslipidemia: where are we? *Diabetologia* **58**: 886–899. doi:10.1007/s00125-015-3525-8
- Voight BF, Peloso GM, Orho-Melander M, Frikke-Schmidt R, Barbalic M, Jensen MK, Hindy G, Hólm H, Ding EL, Johnson T, et al. 2012. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* **380**: 572–580. doi:10.1016/S0140-6736(12)60312-2
- Wainberg M, Mahajan A, Kundaje A, McCarthy MI, Ingelsson E, Sinnott-Armstrong N, Rivas MA. 2019. Homogeneity in the association of body mass index with type 2 diabetes across the UK Biobank: a Mendelian randomization study. *PLoS Med* **16**: e1002982. doi:10.1371/journal.pmed.1002982
- Wang S, Kang H. 2022. Weak-instrument robust tests in two-sample summary-data Mendelian randomization. *Biometrics* **78**: 1699–1713. doi:10.1111/biom.13524
- Wichmann H-E, Hörlein A, Ahrens W, Nauck M. 2016. The biobank of the German National Cohort as a resource for epidemiologic research. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **59**: 351–360. doi:10.1007/s00103-015-2305-4
- Wightman DP, Jansen IE, Savage JE, Shadrin AA, Bahrami S, Holland D, Rongve A, Borte S, Winsvold BS, Drange OK, et al. 2021. A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* **53**: 1276–1282. doi:10.1038/s41588-021-00921-z
- Xu Q, Park Y, Huang X, Hollenbeck A, Blair A, Schatzkin A, Chen H. 2010. Physical activities and future risk of Parkinson disease. *Neurology* **75**: 341–348. doi:10.1212/WNL.0b013e3181ea1597
- Xu Y, Wang C, Li Z, Cai Y, Young O, Lyu A, Zhang L. 2022. A machine learning model for disease risk prediction by integrating genetic and non-genetic factors. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 868–871. IEEE, Las Vegas, NV.
- Yazar S, Alquicira-Hernandez J, Wing K, Senabouth A, Gordon MG, Andersen S, Lu Q, Rowson A, Taylor TRP, Clarke L, et al. 2022. Single-cell eQTL mapping identifies cell type-specific genetic control of autoimmune disease. *Science* **376**: eabf3041. doi:10.1126/science.abf3041
- Yuan S, Larsson SC. 2020. An atlas on risk factors for type 2 diabetes: a wide-angled Mendelian randomisation study. *Diabetologia* **63**: 2359–2371. doi:10.1007/s00125-020-05253-x
- Zhao Q, Wang J, Spiller W, Bowden J, Small DS. 2019. Two-sample instrumental variable analyses using heterogeneous samples. *Stat Sci* **34**: 317–333. doi:10.1214/18-STS692
- Zhao J, Stockwell T, Naimi T, Churchill S, Clay J, Sherk A. 2023. Association between daily alcohol intake and risk of all-cause mortality: a systematic review and meta-analyses. *JAMA Netw Open* **6**: e236185. doi:10.1001/jamanetworkopen.2023.6185
- Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, Robinson MR, McGrath JJ, Visscher PM, Wray NR, et al. 2018. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun* **9**: 224. doi:10.1038/s41467-017-02317-2

Received March 4, 2024; accepted in revised form September 3, 2024.