

RESEARCH

Open Access



# Combining genetic proxies of drug targets and time-to-event analyses from longitudinal observational data to identify target patient populations

Luke Zhang<sup>1</sup>, Prachi Kulkarni<sup>1,2</sup>, Farshad Farshidfar<sup>1</sup>, Whit Tingley<sup>1</sup>, Tim Hoey<sup>1</sup>, Whedy Wang<sup>1</sup>, James R. Priest<sup>1,3</sup> and Sylwia M. Figarska<sup>1\*</sup>

## Abstract

**Background** Human genetics is an important tool for identifying genes as potential drug targets, and the extensive genetic study of cardiovascular disease provides an opportunity to leverage genetics to match specific patient populations to specific drug targets to improve prioritization of patient selection for clinical studies.

**Methods** We selected well described genetic variants in the region of *PCSK9* (rs11591147 and rs562556), *ADRB1* (rs7076938), *ACE* (rs4968782 and rs4363), *GLP1R* (rs10305492) and *ABCC8* (rs757110) for use as proxies for the effects of drugs. Time-to-event analyses were utilized to evaluate their effects on atrial fibrillation (AF) and heart failure (HF) death and/or re-hospitalization using real-world longitudinal dataset. To mitigate the effect of confounding factors for cardiovascular (CV) outcomes, we employed propensity score matching.

**Results** After matching, a genetic proxy for PCSK9 inhibition (rs11591147) improved survival from CV death/heart transplant in individuals following a diagnosis of ischemic heart disease (Hazard Ratio (HR) 0.78,  $P=0.03$ ). A genetic proxy for beta-blockade (rs7076938) improved freedom from rehospitalization or death in individuals with AF (HR 0.92,  $P=0.001$ ), and a genetic proxy of ACE inhibition (rs7076938) improved freedom from rehospitalization for HF or death (HR 0.8,  $P=0.017$ ) and AF (HR 0.85,  $P=0.0014$ ). A protective variant in GLP1R (rs10305492) showed decreased risk of developing HF or CV death after diagnosis of ischemic heart disease (HR=0.82,  $P=0.031$ ) and a protective variant in ABCC8 (rs757110) showed decreased risk of CV mortality since ischemic disease diagnosis (HR=0.88,  $P=0.04$ ) and decreased risk of AF in diabetic patients with ischemic heart disease (HR=0.68,  $P=0.001$ ). Notably, despite smaller cohort sizes after matching, we often observed numerically smaller HRs and reduced  $P$ , indicating more pronounced effects and increased statistical association. However, not all genetic proxies replicated known treatment effects.

**Conclusions** Genetic proxies for well-known drugs corroborate findings from clinical trials in cardiovascular disease. Our results demonstrate a useful analytical approach that leverages genetic evidence from a large cohort with longitudinal outcomes data to effectively select patient populations where specific drug targets may be most effective.

**Keywords** Drug target discovery, Patients selection, in silico trials, Cardiovascular outcomes

\*Correspondence:

Sylwia M. Figarska

sfigarska@tenayathera.com; s.figarska@gmail.com

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Introduction

Drug discovery and clinical development is a high-risk process that may require decades of capital investment prior to clinical use with an estimated median cost of \$1.1 billion for each new drug approved [1]. Cardiovascular diseases with high heterogeneity such as heart failure (HF) frequently require complex and lengthy clinical trials enrolling thousands of patients to demonstrate efficacy [2]. However in some examples, a small effect in a large group may be derived from a large effect in a smaller subgroup of affected individuals [3, 4]. This phenomenon has been observed in a number of cardiovascular therapies, which have shown greater efficacy or clinical utility for specific subgroups of disease following regulatory approval [5, 6]. Recognizing these nuances can guide the development of personalized medicine strategies, facilitating improved treatment effects in more targeted populations. This approach not only significantly reduces development costs but also optimizes therapeutic outcomes, thereby increasing the likelihood of success in drug development.

Regulatory bodies have issued guidance intended to increase the efficiency of drug development and support precision medicine by identifying subgroups of affected individuals where a specific therapeutic approach may be most effective [7]. Oncology has been at forefront of incorporating genetic information into drug discovery by identifying genetic targets which define prognosis in subtypes of large heterogeneous disease entities, and then designing therapies to specifically manipulate these targets [8, 9]. Outside of oncology, human genetics has proven to be a useful tool for selection of drug targets and may also have the potential to enhance the design and conduct of clinical trials in cardiovascular medicine.

Here we describe efforts to develop a novel analytical framework intended to generate a strong pre-clinical hypothesis about the patient population(s) where a therapeutic intervention is most likely to be effective using a real-world clinical dataset and time-to-event analyses. To test and validate our analytical approach for cardiovascular disease, we leverage genetic variants with a known biological effect on gain, loss of function, or protein expression levels, as well as expected associations with cardiovascular phenotypes. With such variants we construct *in silico* trials using real world data and compare the results to the outcomes of well-established randomized clinical trials of approved therapies. We demonstrate the potential utility of our analytical framework using known therapeutic targets across diverse areas of cardiovascular medicine, including lipid biology to atherosclerosis outcomes,

beta-adrenergic signaling to HF and atrial fibrillation (AF) outcomes. Furthermore, we aim to explore effects of antidiabetic medication proxies on cardiovascular outcomes.

## Methods

### Study population, clinical data, and outcomes

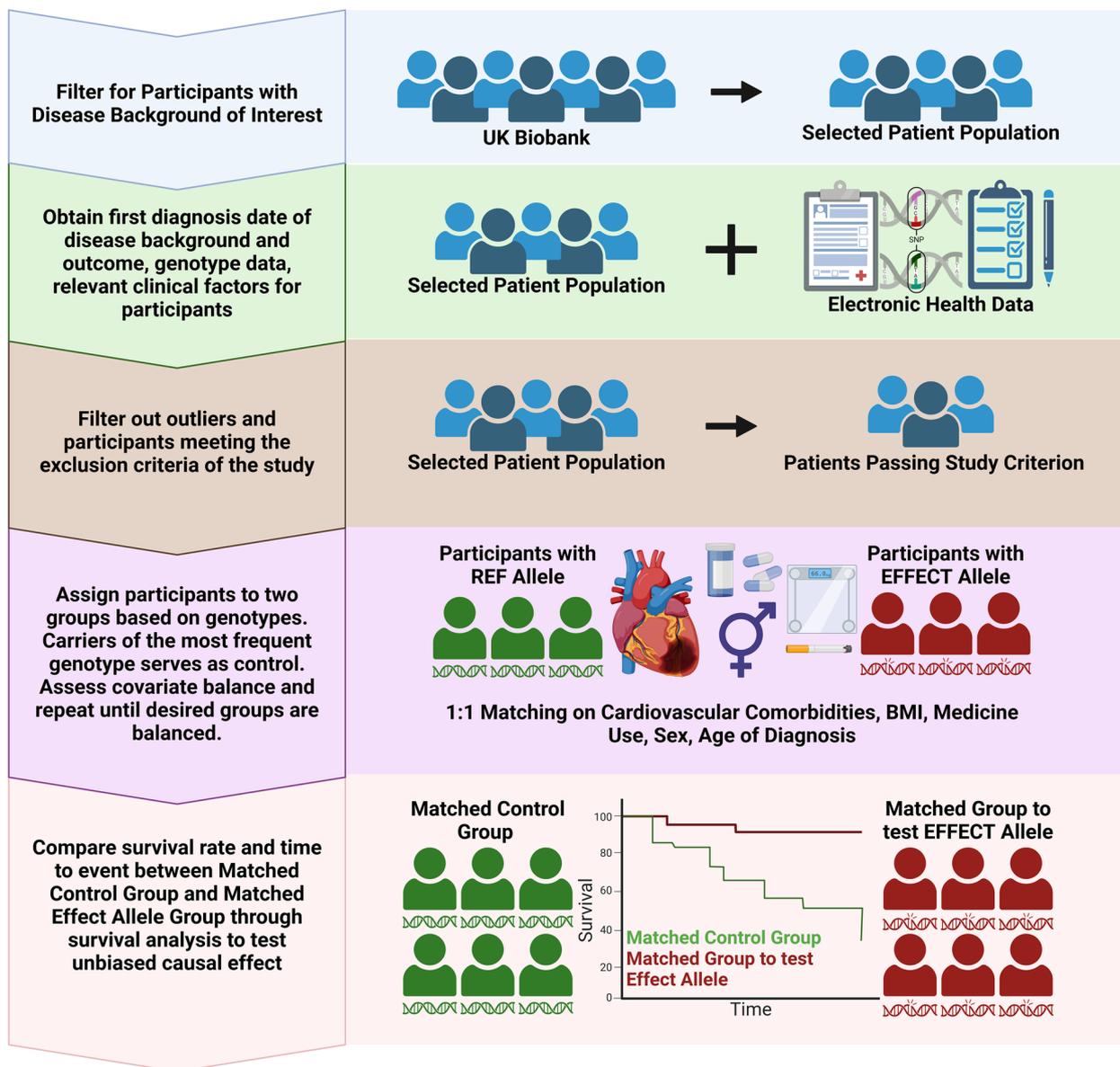
Genetic and clinical data in the UKB cohort were obtained from the UKB (<https://www.ukbiobank.ac.uk>) and is available to researchers through a streamlined application process. The UKB was approved by the North West Multi-Centre Research Ethics Committee and all participants provided written informed consent to participate in the study. We used International Classification of Diseases ICD-9 and ICD-10 codes to define patients' populations, co-morbidities and cardiovascular outcomes. Detailed ICD codes are provided in Supplemental Table 1. We considered two outcomes for this study: a) CV death or heart transplant (CV death/heart transplant), and b) rehospitalization or CV death/heart transplant—composite outcome. Figure 1 presents the analytical framework of participants selection and Fig. 2 summaries all survival analysis performed.

Details on selection of covariates (co-morbidities and medication use) considered for propensity score matching for the specific outcome are described in the supplementary material.

For example, for rs7076938 within the *ADRB1* gene we matched for the following variables: sex, C-Reactive Protein level, smoking status, body mass index (BMI), low-density lipoprotein (LDL) levels, diuretics use, calcium channel blockers use, alpha adrenoceptor blocking drug use, statins use, digoxin use, angiotensin receptor blockers use, anti-coagulant medication use, previous disease diagnosis for aortic valve stenosis, chronic renal failure, chronic obstructive pulmonary disease, HF, type 1 diabetes, type 2 diabetes, ischemic heart disease (IHD), mitral valve disorder, unspecific stroke, and age of AF diagnosis. We excluded individuals using angiotensin-converting enzyme inhibitors or beta-blockers to mitigate confounding effects associated with the excluded medications (Supplemental Fig. 2).

### Statistical approach

All statistical analyses were performed in R (version 4.3.3) using the *MatchIt* for propensity score matching, and *survival* packages for time-to-event analyses. For inclusion, exclusion, and matching of clinical covariates which can vary with time, such as pre-existing conditions, individuals were matched based on covariates at the time of the initial diagnosis. Values for other covariates such as blood pressure, BMI and lipid levels were

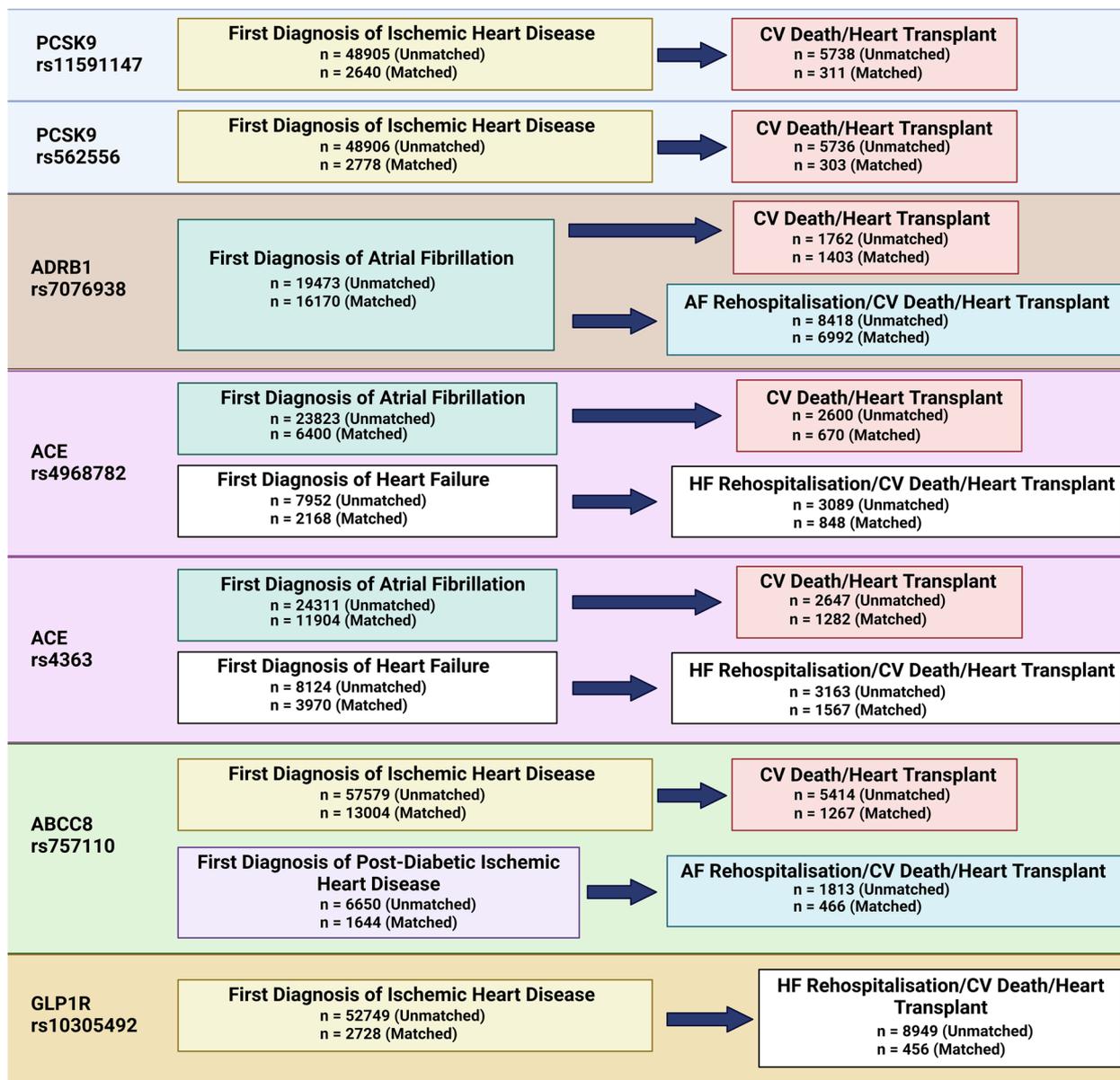


**Fig. 1** Analytical framework. Participants diagnosed with a disease of interest were included in the study. Information on co-morbidities and medication use was obtained and used for propensity score matching procedure. Survival analysis was performed for matched allele carriers

used from the baseline assessment. Medication use was considered from the first available response.

For each genetic variant, individuals with effect allele were defined by variant allele (heterozygous or homozygous) and compared to individuals with the other allele. Covariate balance post-matching was assessed using standardized mean differences in Love plots, generated with the `love.plot()` function in `cobalt` R package, examining the balance measures before and after conditioning. Survival analyses were performed using Cox Proportional Hazards model and visualized with `survplot` as

Kaplan–Meier or Cox Hazards curves as appropriate. For each variant analyzed, we also performed a post-hoc power calculation based upon the available population size, number of events, and follow-up time, using the `SurvSNP` R package (details are provided in the Supplementary Material). The goal was to enhance our understanding of the reliability of our findings and to potentially guide future research endeavors. Additionally, for each single nucleotide polymorphism (SNP) we conducted a phenome-wide association study (PheWAS) using PHENOME Scan ANALYSIS Tool (PHESANT) [10]



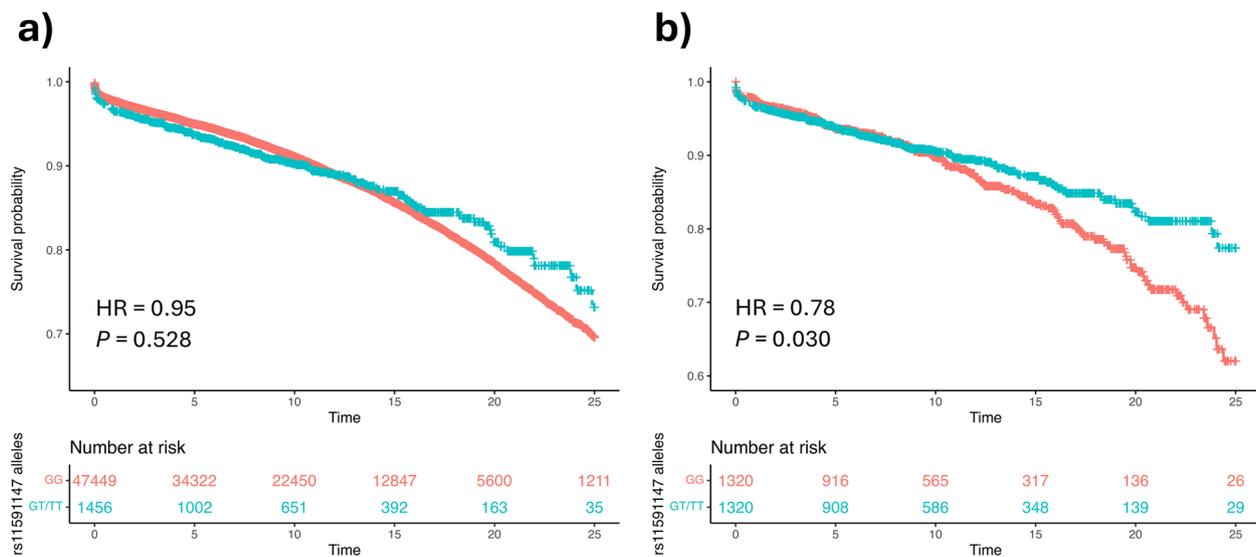
**Fig. 2** Summary of all survival analysis performed for each SNP in the study, indicating disease of interest and considered outcomes

and details of the analysis are provided in the Supplementary Material.

**Genetic variants**

Similar to the approach for instrumental variable selection in Mendelian Randomization, we selected a set of genetic variants with clearly explainable biological impact on a gene which is the target of either an approved or proposed therapeutic modality. Proprotein convertase subtilisin/kexin type 9 (PCSK9) is a well validated genetic target with a clearly defined mechanism of action to reduce circulating cholesterol levels and

multiple successful clinical trials for prevention of atherosclerotic cardiovascular disease (ASCVD). The LoF variant rs11591147 (p.Arg46Leu) represents the effect of PCSK9 inhibition as it disrupts the function of the PCSK9 which results in higher amount of cell surface LDL receptor (LDLR) and internalization of LDL, leading to lower serum LDL levels and subsequently confer protection against ischemic heart disease (IHD) [11–13]. We also tested another PCSK9 missense variant, (rs562556 p.Val474Ile) that is likely to have an impact on PCSK9 activity and previously showed associations with LDL levels and CAD risk [13, 14].



**Fig. 3** Effect of PCSK9 rs11591147 on time to cardiovascular (CV) death/heart transplant from first ischemic heart disease diagnosis (IHD) in (a) unmatched and (b) matched participants. Only individuals of European ancestry were included. Individuals with GT/TT genotypes (proxy for PCSK9 inhibitor treatment) showed better survival for CV death/heart transplant compared to individuals with GG genotype (controls)

*ADRB1* encodes the  $\beta_1$  adrenergic receptor with primary cardiovascular effects in the heart sinus node and juxtaglomerular cells of the kidney to modulate contractility, heart rate, and blood pressure. It is the target of many selective beta blockers used in primary and secondary prevention of ASCVD, HF, hypertension, and AF [15–17], (2024). The variant rs7076938 is a tagging SNP for rs1801253 (Gly389 Arg,  $r^2 = 0.96$  in EUR and  $r^2 = 0.81$  in all populations) that has been shown to alter post-receptor signaling, with the Arg389 receptor (tagged by rs7076938 T) coupling more efficiently with its corresponding G protein (Gs) [18–20] and correlating with higher blood pressure [21, 22]. In our study we tested rs7076938, as it was more significantly associated with SBP in large genome wide association studies (GWAS) [22] and was used as a part of instrument variable in MR analysis representing antihypertensive medication proxy [23].

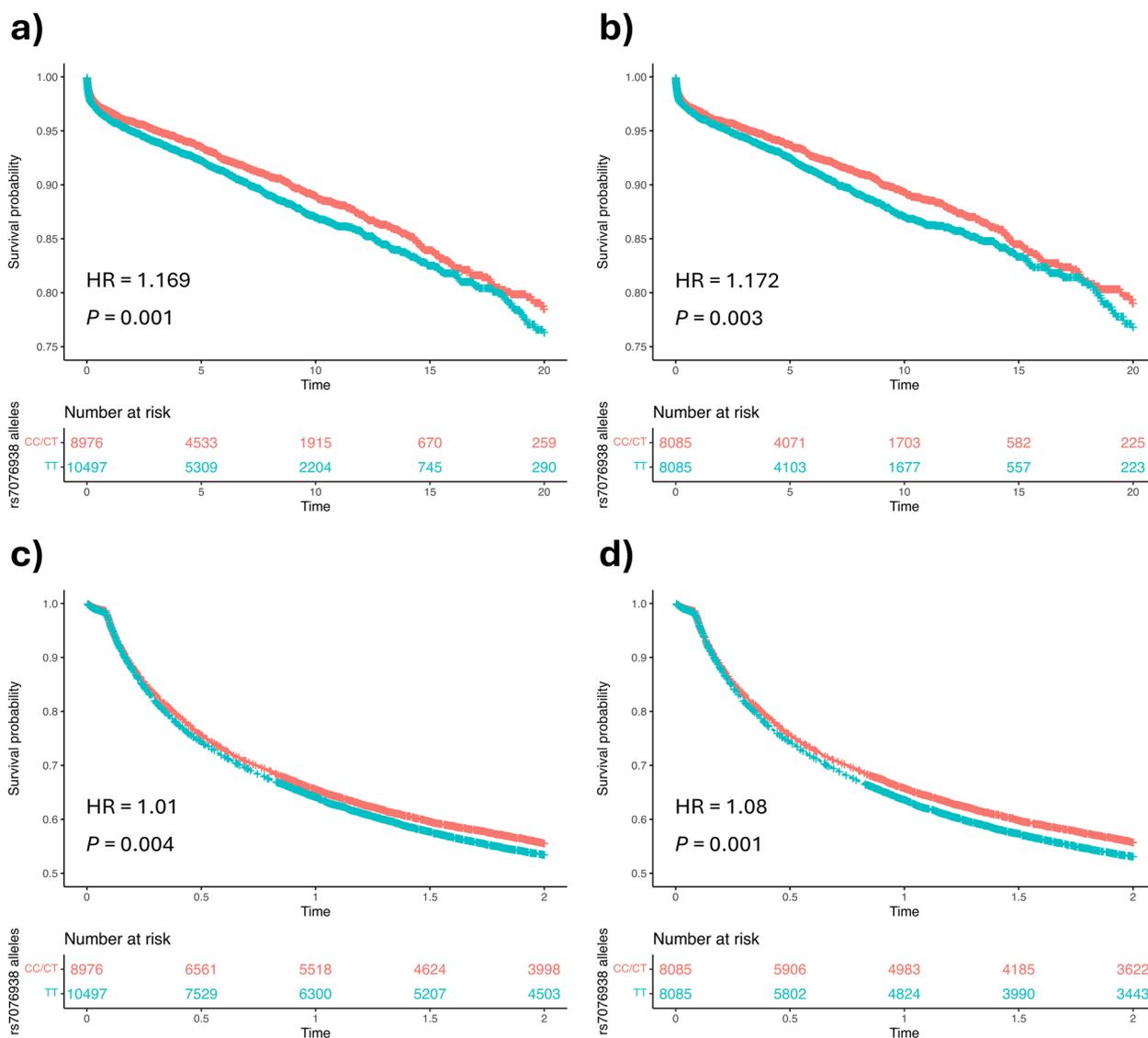
The angiotensin-converting enzyme gene (*ACE*) gene encodes the angiotensin converting enzyme expressed mainly in the lungs which is a key step in the renin/aldosterone-angiotensin system regulating vascular tone and blood pressure. Inhibitors of ACE such as lisinopril are a mainstay of treatment for hypertension, HF, and AF. A genetic variant, rs4968782, located upstream of *ACE* likely affects transcription factor binding, and is a strong eQTL associated with increased *ACE* expression in the lung; therefore mimicking the effect of ACE inhibitors to modulate risk of hypertension. Also, effects of this SNP are stronger than effects of coding variants that were initially linked to circulating ACE levels and

ACE activity [24]. We also tested an intronic variant in *ACE*, rs4363, that previously showed a very strong association [25] with ACE levels reaching significance of  $P = 2 \times 10^{-257}$ . This intronic SNP is highly correlated ( $r^2 = 0.9$  in EUR) with a missense variant rs4343 associated with ACE activity [24].

A common missense variant in the ATP Binding Cassette Subfamily C Member 8 (*ABCC8*) gene encoding a component of the sulfonylurea receptor 1 SUR1 (rs757110, Ala1369Ser or p.A1369S) promotes closure of the ATP-sensitive potassium channel and is associated with increased insulin secretion, thus mimics the effects of sulfonylurea medication [26, 27]. This variant is associated with significantly lower risk of type 2 diabetes and a reduced risk of coronary heart disease [26, 28, 29]. These results suggest that long-term sulfonylurea therapy may reduce risk of future cardiovascular events.

Genetic variant in gene encoding glucagon-like peptide-1 receptor (*GLP1R*) rs10305492 (Ala316 Thr) is associated with lower fasting glucose and reduced risk of type 2 diabetes, similar to the effects of GLP1R agonist therapy. This variant is also associated with lower risk of coronary heart disease [30].

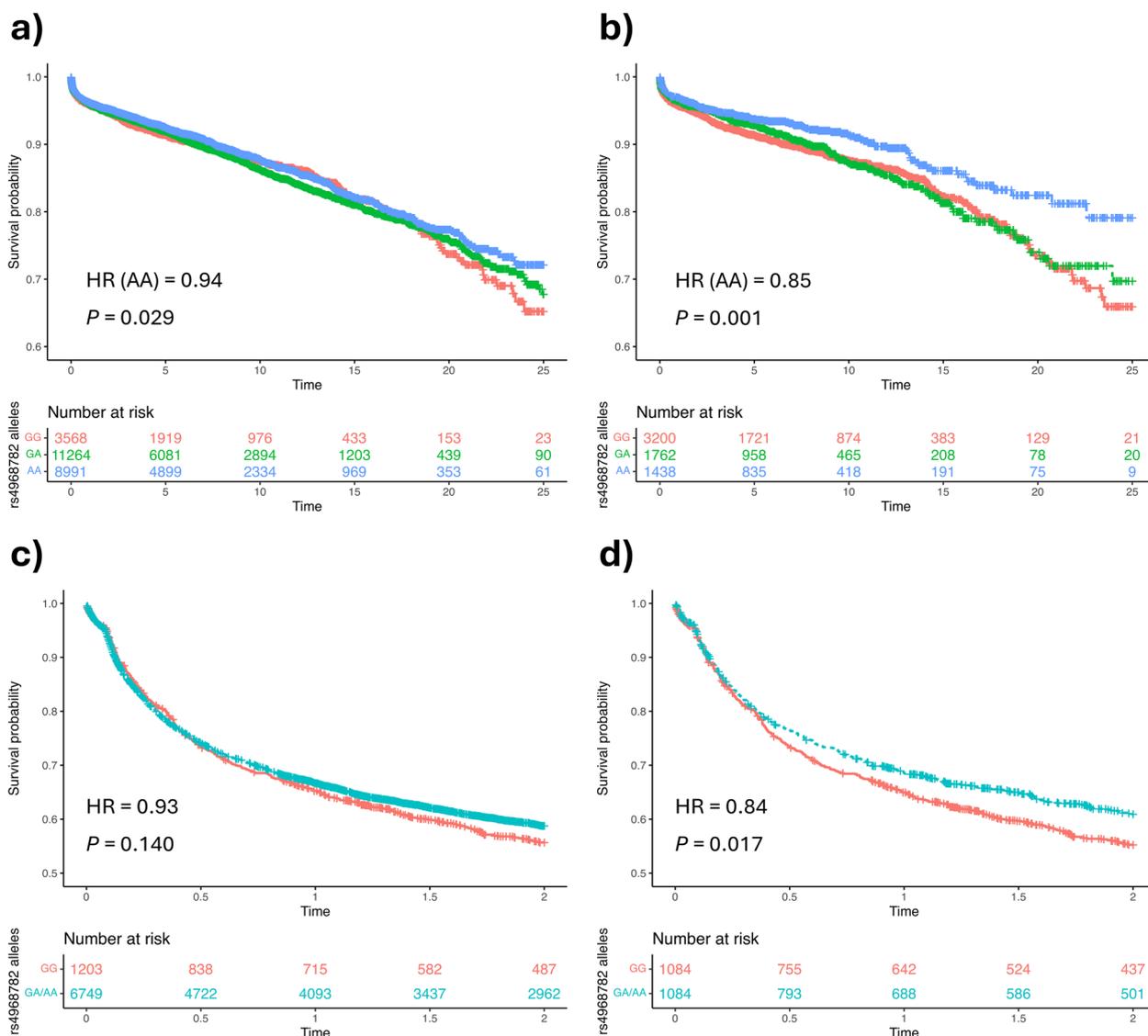
As a first pass diagnostic analysis, we aimed to replicate the previously described case-control associations with the selected variants in the UKB, of which most showed associations with expected phenotypes (Supplemental Table 2). The numbers of allele carriers per variant and event are presented in the figure panels (Figs. 3, 4, 5, 6 and 7) corresponding to each SNP. For patients who did not experience an event, follow-up period was calculated



**Fig. 4** Effect of ADRB1 rs7076938 on time to cardiovascular (CV) death/heart transplant from first atrial fibrillation (AF) diagnosis in (a) unmatched and (b) matched participants, and time to AF rehospitalization or CV death/heart transplant in (c) unmatched and (d) matched participants. Individuals with TT genotype showed increased risk of AF rehospitalization of CV death/heart transplant composite outcome compared to individuals with CC/CT genotypes

from the time of disease diagnosis until the most recent data update from the UKB, version September 2023. Events of non-cardiovascular death were right censored at the time of death. To evaluate the composite outcome of rehospitalization for disease of interest (HF or AF) or CV death/heart transplant, the follow-up period from the time of diagnosis was censored at the number of years where the impact of the SNP was most representative. Hard outcomes typically require longer follow-up to capture enough events, therefore CV death/heart transplant outcome was investigated for a span between 15 and 25

years for different SNPs. Soft outcome such as AF or HF rehospitalization usually require shorter follow-up time as these events occur more frequently, which is why we chose a clinically relevant time of two years for ADRB1 and ACE. ABCC8 and GLP1R, genes affecting glucose metabolism, required a follow-up time of 15 years to capture cardiovascular rehospitalization outcomes. It is a longer time compared to ADRB1 and ACE, probably due to differences in their respective biological pathways affecting outcomes. All the tested variants met the

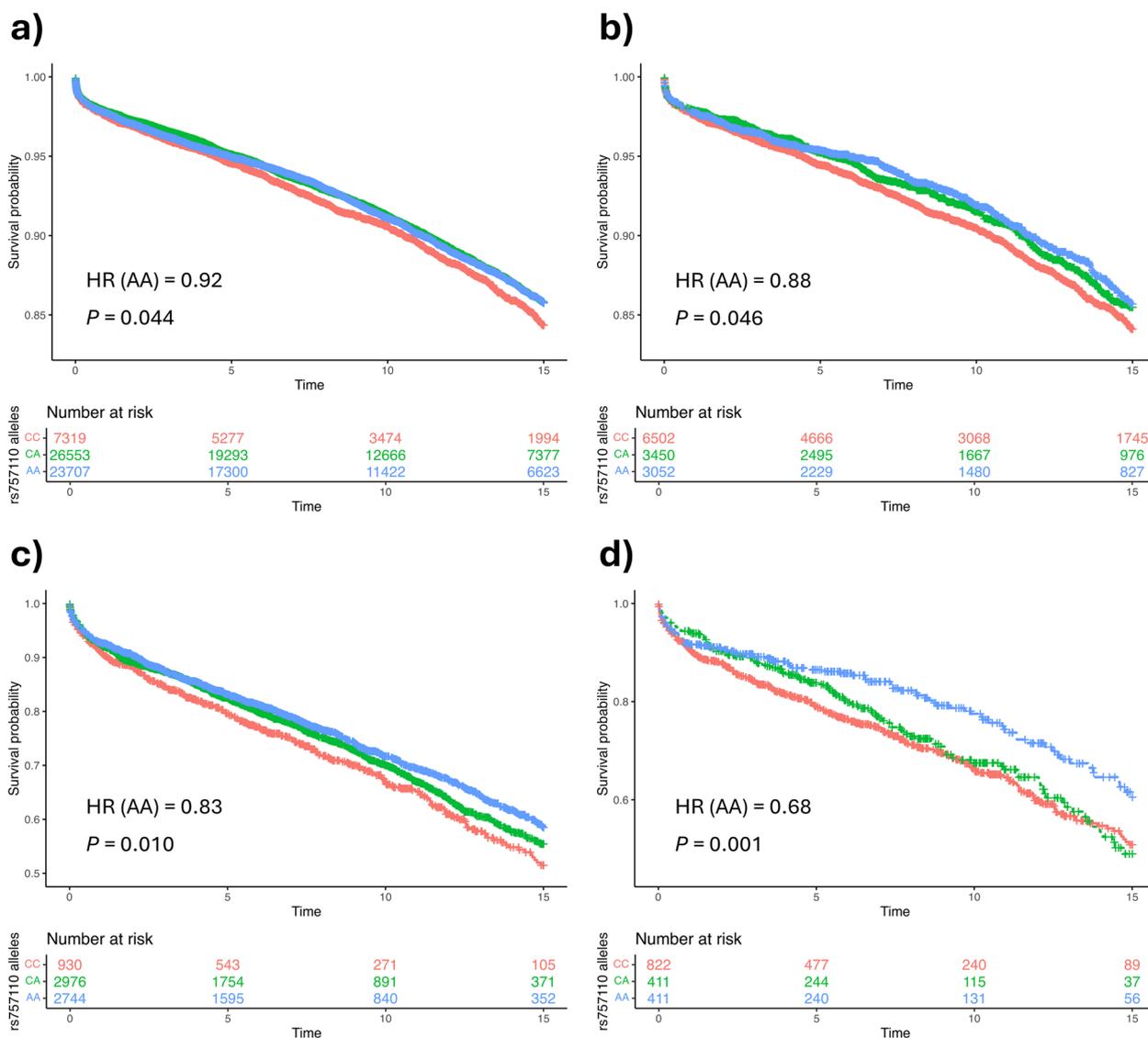


**Fig. 5** Effect of ACE rs4968782 on time to cardiovascular (CV) death/heart transplant from first atrial fibrillation (AF) diagnosis in (a) unmatched and (b) matched individuals of European ancestry, and time to heart failure (HF) rehospitalization or CV death/heart transplant in (c) unmatched and (d) matched participants of any continental ancestry. Carriers of GA or AA genotype showed better survival for rehospitalization and severe cardiovascular outcome compared to individuals with GG genotype

underlying assumptions for proportional hazard risk over time as indicated by log–log plots.

The analytical framework (Fig. 1) consists of selecting the individuals heterozygous or homozygous for the alternative allele for the variant of interest, selection of relevant known clinical risk factors including age at primary diagnosis, clinical comorbidities, and medication use. Cox Proportional Hazard models were performed prior to matching (with adjustments for age at disease diagnosis and sex), and after matching (without adjustments). Given the known associations for each of the

selected variants and the intent of replicating known therapeutic effects, for the purposes of the methods development we defined statistical significance as a nominal *P* less than 0.05 and we did not perform correction for testing of multiple hypotheses. When the number of individuals homozygous for the effect allele was less than 25% of the number of heterozygous individuals, the groups were combined into a single group after matching. For each variant analyzed, we also performed a post-hoc power calculation based upon the available population size, number of events, and follow-up time. A summary



**Fig. 6** Effect of ABCC8 rs757110 on time to cardiovascular (CV) death/heart transplant from first IHD diagnosis in (a) unmatched and (b) matched participants, and on time to composite outcome – AF hospitalization or CV death/heart transplant from first IHD diagnosis in (c) unmatched diabetic patients and (d) matched diabetic participants. Individuals with AA genotype showed decreased risk of AF hospitalization and CV death/heart transplant composite outcome in pre-diabetic IHD patients compared to those with the CC genotype

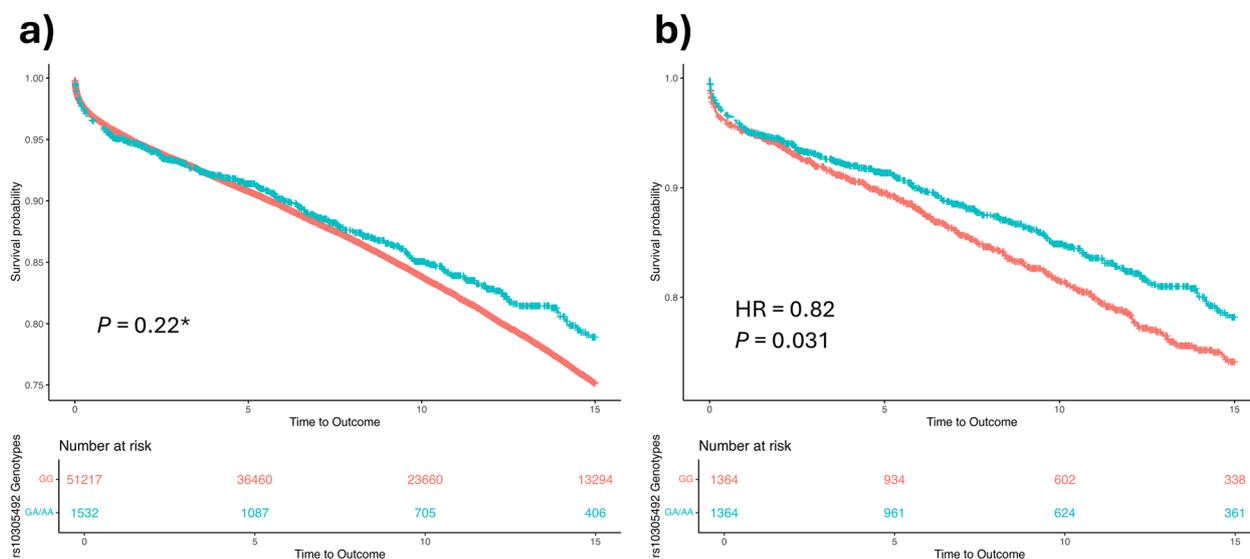
of start and end timepoints for survival analyses conducted for each SNP, as well as the number of patients included in the study, can be seen in Fig. 2. Additionally, power calculations for survival analyses conducted for each SNP can be seen in Supplemental Table 3. Baseline characteristics of individuals included in each analysis for the SNPs are presented in Supplemental Tables 4–14.

**Results**

**PCSK9 rs11591147**

To replicate the successful trials for PCSK9 inhibitors for secondary prevention of ASCVD, we examined

cardiovascular death or heart transplant among individuals after a diagnosis of IHD. The impact of rs11591147 on time to CV death/transplant among genetically determined individuals of European ancestry diagnosed with IHD did not reach statistical significance, exhibiting a hazard ratio (HR) of 0.95 and a P of 0.53. However, after matching on comorbidities existing at the time of diagnosis of ASCVD (Supplemental Fig. 1), medication usage, and cardiovascular comorbidities, carriers of one or more effect alleles (T) demonstrated improved survival compared to individuals with no copies of the effect allele (Fig. 3a). The HR



**Fig. 7** Effect of GLP1R rs10305492 on time to HF hospitalization or cardiovascular (CV) death/heart transplant composite outcome from first ischemic heart disease diagnosis (IHD) in **(a)** unmatched and **(b)** matched participants. Individuals with GA/AA genotypes (proxy for GLP1-R agonist treatment) showed reduced risk of HF hospitalization composite outcome compared to individuals with GG genotype (controls). \*Didn't meet proportional hazard assumption, P calculated with Gehan-Breslow test

markedly decreased to 0.78 with a significant  $P$  of 0.03, indicating a protective effect of PCSK9 rs11591147 against the development of severe cardiovascular outcomes from an IHD diagnosis. The difference between the unmatched and matched HRs underscores the importance of accounting for other causal factors, including some that were implemented as part of a treatment strategy, in determining the true impact of PCSK9 inhibition on cardiovascular outcomes (Fig. 3). Results of covariate balancing for each matching term are presented in Supplemental Fig. 1.

#### PCSK9 rs562556

The impact of rs562556 on time to CV death/transplant from date of first IHD diagnosis among individuals of European ancestry did not reach statistical significance, exhibiting a HR of 1.01 and  $P$  of 0.342. Even after matching on comorbidities (Supplemental Fig. 6) existing at the time of diagnosis of ASCVD, medication usage, and cardiovascular comorbidities, no statistically significant impact was observed by rs562556 (HR = 1.06,  $P = 0.641$ , Supplemental Fig. 5). The lack of statistical significance could also be attributed to relatively low statistical power, as seen in Supplemental Table 3.

#### ADRB1 rs7076938

Analogous to successful trials of beta blockers for treatment of AF, we examined AF progression with the combined outcome of rehospitalization for AF or CV

death/heart transplant [31]. Carriers of TT genotype of rs7076938 had increased risk of CV death/heart transplant since first AF diagnosis (HR = 1.17,  $P = 0.0012$ ) and the effect remained significant in the matched data (HR = 1.17,  $P = 0.0031$ ). Additionally, carriers of the TT genotype showed increased risk of AF rehospitalization or CV death/heart transplant in unmatched (HR = 1.01,  $P = 0.0043$ ) and matched participants (HR = 1.08,  $P = 0.0012$ ). The impact of ADRB1 on cardiovascular outcomes before and after matching can be seen in Fig. 4. Results of covariate balancing can be seen in Supplemental Fig. 2.

#### ACE rs4968782

Carriers of at least one A allele of rs4968782 exhibited decreased risk of CV death/heart transplant from first AF diagnosis in unmatched (HR = 0.94,  $P = 0.029$ ) and matched (HR = 0.85,  $P = 0.0014$ ) participants. When the composite outcome – HF rehospitalization or CV death/heart transplant – was considered, the impact of the A allele was only seen in matched data (HR = 0.84,  $P = 0.017$  vs HR = 0.93,  $P = 0.14$  in unmatched data). It is important to note that the time-to-event analysis of CV death/heart transplant from first AF diagnosis was performed in the individuals of European ancestry as the SNP only showed effects in this subset, while time to composite outcome was performed in the entire population regardless of ancestry. These results underscore the impact of propensity score matching in elucidating the significance of the ACE SNP in cardiovascular outcomes, revealing

previously undetected associations, and enhancing the precision of risk assessments (Fig. 5). Results of covariate balancing are displayed in Supplemental Fig. 3.

#### ACE rs4363

For ACE rs4363, carriers of the A allele did not exhibit decreased risk of CV death/heart transplant from first AF diagnosis (HR = 0.97,  $P = 0.197$  in unmatched, HR = 0.95,  $P = 0.17$  in matched). When the composite outcome – HF rehospitalization or CV death/heart transplant was considered, the impact of the A allele was not seen in either unmatched (HR = 1.01,  $P = 0.89$ ) or matched data (HR = 1.04,  $P = 0.45$ ). It is important to note that the time-to-event analysis of CV death/heart transplant from first AF diagnosis was performed in individuals of European ancestry while time to composite outcome was performed in the entire population, regardless of ancestry, to mirror the analysis of ACE rs4968782. Survival analysis results can be seen in Supplemental Fig. 7, while the results of covariate balancing can be seen in Supplemental Fig. 8. There was limited statistical power for the analysis of ACE rs4363 (Supplemental Table 3).

#### ABCC8 rs757110 Ala1369Ser

In our study carriers of the AA genotype of rs757110 compared to CC genotype had significantly reduced risk of CV outcome after IHD diagnosis (HR of 0.92,  $P = 0.044$ ) in all participants, with more pronounced effect in the matched data (HR of 0.88,  $P = 0.046$ ). Among diabetic individuals we found that the AA genotype was protective against developing AF or CV outcome (HR 0.83,  $P = 0.01$  in unmatched data and HR 0.68,  $P = 0.001$  in the matched data, Fig. 6 and covariate balancing results in Supplemental Fig. 8). Effect of the SNP on ischemic stroke, AF or HF/CV death composite outcome cardiovascular outcomes since type 2 diabetes diagnosis were not statistically significant and are presented in Supplemental Table 15.

#### GLP1R rs10305492 (Ala316 Thr)

We found that carriers of at least one A allele of rs10305492 had significantly lower risk of developing HF or CV composite outcome after IHD diagnosis what was observed after matching procedure (HR 0.82,  $P = 0.031$ , Fig. 7b and Supplemental Fig. 9). This result emphasizes the importance of matching procedure as the protective effect was not captured properly in unmatched data (Fig. 7a). Analysis stratified by diabetic status is presented in Supplemental Table 15 and didn't reach statistical significance due to low statistical power.

PheWAS was conducted for all the SNPs mentioned above, with the aim of finding associations mentioned in previous publications. Analyses were adjusted for age,

sex, and 10 genetic principal components. Both SNPs in PCSK9 showed significant associations with LDL levels ( $\beta = -0.06$ ,  $P < 2e^{-200}$  and  $\beta = 0.013$ ,  $P = 5.6e^{-19}$ , for rs11591147 and rs562556 respectively). Both ADRB1 SNP rs7076938 and ACE rs4968782 were significantly associated with systolic blood pressure  $\beta = 0.012$ ,  $P = 2.17e^{-18}$  and  $\beta = -0.007$ ,  $P = 9.33e^{-07}$  respectively. The ABCC8 SNP rs757110 showed associations with glycated hemoglobin ( $\beta = -0.015$ ,  $P = 2.72e^{-25}$ ) and type 2 diabetes ( $\beta = -0.047$ ,  $P = 1.07e^{-18}$ ), and GLP1R rs10305492 was associated with glucose levels ( $\beta = -0.017$ ,  $P = 2.25e^{-27}$ ). PheWAS results for each individual SNP can be seen in Supplemental Figs. 10–16.

## Discussion

Here we describe a genetic survival analysis which employs a well-characterized genetic instrumental variable as a proxy for potential treatment with a specific drug or target. This approach has the potential to identify treatment effects using real-world outcome data through time-to-event analyses. In examples of targets with approved therapies (PCSK9, ADRB1, ACE, GLP1R and ABCC8), this analytical framework identifies beneficial treatment effects seen in clinical trials. The approach described here is a natural extension of genome-wide association study (GWAS) of clinical outcomes, Mendelian Randomization, and similar time-to-event GWAS [32–34]. Genetic survival analysis offers a critical advantage over standard applications of GWAS or Mendelian Randomization which typically only consider non-genetic covariates such as age and sex. Our proposed approach not only estimates the direction of effect for a specific drug or target in a real-world setting, but it also incorporates clinical covariates, causal factors, and standard of care therapies commonly encountered in clinical trials.

In the life cycle of drug development, the clinical costs of running human trials far exceeds the pre-clinical costs of building a new molecule. Therefore, it is valuable to define as early as possible the disease state and patient population that are likely to benefit most from the new treatment. In cardiovascular medicine this type of 'patient population' information may often be derived from a combination of translational studies in animal models of disease, hypotheses derived from cohort-based clinical research, and exploration of different biomarkers in early phase 1/2 studies. The approach of genetic survival analysis that we present here, may offer the opportunity to explore the potential efficacy of a specific drug target using real-world data and to develop hypotheses about patient selection (co-morbidities, standard treatments), disease status, and biomarkers – in a cost-effective manner that does not require testing in a human

clinical trial. At a high level our findings may add an additional dimension to the utility of genetics in drug development; genetics can be used to identify drug targets but also hold potential to identify the patient populations where those drug targets may have the most benefit.

Cardiovascular diseases such as HF are frequently complex, where a broad diagnosis may arise from a combination of one and often more environmental and genetic causes. Targets identified from genetic data such as GWAS hold potential to guide drug discovery, as these studies have the unique capability for identifying risk loci for a disease without an a priori hypotheses. While GWAS is useful in identifying novel genetic loci that are beyond our current understanding of the disease, cell-based or animal models are unable to fully mimic the human condition, limiting the success rate of translating preclinical findings into clinical practice. Additionally, the genetic factors identified from GWAS for disease susceptibility may not always be the same factors that govern disease progression or long-term outcomes—which are often the endpoints of clinical trials and the most meaningful outcomes to affected individuals and physicians. Therefore, for drug target prioritization, more studies based on clinical samples of affected individuals with real world phenotype data may more accurately reflect the genetic basis of the condition under study. To ensure precise investigation of specific patient populations, our study utilized only Hospital Episode Statistics (HES) data and ICD codes, excluding self-reported cardiovascular complaints without direct medical diagnoses. Further, aligning with best practices in clinical trial enrollment which aims to increase ancestral diversity, our analysis did not limit participants to individuals of European ancestry—except in noted cases—and included all participants, irrespective of ancestry.

Matching is a well-accepted statistical technique to mitigate the influence of measured confounders and constitutes an essential part of clinical trial design often implemented in the form of ‘randomization.’ Matching on the propensity score is often effective at eliminating differences between the allele carrier groups to achieve covariate balance, and consequently often involves discarding units that are not paired with others. In addition to covariate balance, the quality of the match is determined by how many units remain after matching ([35, 36]). In our findings, successful matching often led to a notable reduction in sample size but also a more pronounced effect of the SNP. This suggests that our matching procedure effectively mitigated the influence of confounding factors and enabled us to discern a true effect of the SNP. Put differently, the improvement in power to detect known effects of genetic instruments for PCSK9, ADRB1, ACE, GLP1R and ABCC8 following

matching underscores the methodological robustness of our approach and highlights its compatibility with complexities inherent in real-world longitudinal observational data. It is important to note that matching did not resolve all the complexities of real-world data, as the ACE variant rs4363 did not appear to modify the risk of adverse outcomes in AF or HF, which is not consistent with findings from previous clinical trials [37, 38].

The findings from our approach may also be used to assess potential safety signals and forecast the impact of ongoing trials. Cardiovascular safety has been a concern for antidiabetic treatment for years [39, 40] and therefore in this study we investigated longitudinal effects of genetic proxies or sulfonylureas and GLP1R agonists. Previously the tested genetic variants were associated with decreased risk of IHD in cross sectional analysis [26, 30]. Here we are reporting their protective effect in a longitudinal fashion where we evaluated cardiovascular death risk in individuals previously diagnosed with ischemic heart disease. Results presented for GLP1R precede the ongoing clinical trial SOUL, where CV effects of oral semaglutide in individuals with diabetes and established ASCVD will be assessed [41].

Like other techniques based upon human genetics, the application and interpretability of genetic survival analysis has important limitations. Though we chose genetic instruments largely based on efficacious and widely accepted therapies, our findings are derived from a single large cohort study (UKB) and would benefit greatly from validation with an independent dataset. One important difference between our method and Mendelian Randomization is that instead of a set of SNPs we use only one genetic variant as an instrumental variable. Using a single SNP is a necessary part of the analytical procedure described here in order to assign the simulated treatment groups for subsequent matching and analysis. In theory, multiple SNPs could be used in the form of a polygenic score for a trait as the basis of assigning treatment groups however such an approach is beyond the scope of this manuscript. For simplicity we chose to select coding and non-coding common genetic variants as instrumental variables, which typically have small effect sizes and may be obscured by other clinical factors unless the number of individuals and events are relatively large. Conversely, the power to detect survival benefit may be compromised even with large effect sizes seen with rare variants which may suffer from a small number of individuals and events.

Unlike in a highly controlled clinical trial setting, in a real-world dataset like the UK Biobank, some individuals included in the analysis may not always be receiving what would be considered standard-of-care treatment. To ameliorate important differences in comorbidities and

standard treatments between the groups we relied upon matching. While our findings illustrate that matching on comorbidities and stratification appear to be useful tools for ‘focusing’ the technique to enhance the power and balancing the effects of standard therapies between different SNP treatment groups, matching does not eliminate the possibility of reverse causality or collider bias. Such biases can occur when individuals that carry a particular genetic variant have a measurably different risk of disease susceptibility — often observable in case–control GWAS — which impacts long-term disease outcome [42, 43]. Naturally, the analytical framework may only be applied when a genetic instrumental variable is present in the population of interest. An additional limitation of our current application is its inability to provide a meaningful estimate of treatment effect size for the target, which is among the most critical information derived from clinical trials for regulatory bodies, commercial entities, and payors [44, 45].

A significant strength of the genetic survival analysis work presented here lies in the diversity of genetic variants examined and the broad range of cardiovascular outcomes assessed, including IHD, HF, and AF. This comprehensive approach helps identify specific patient populations that would benefit most from potential treatments. A logical extension of the analytical framework presented here would be to conduct *in silico* trials to select and optimize specific clinical variables and co-morbidities, with the goal of maximizing effect size (and minimizing the cost and duration of a clinical trial) for a new treatment. This approach of stratification might not be limited to clinical variables or co-morbidities but might also logically include other genetic factors such as polygenic scoring or stratification by Mendelian forms of cardiovascular disease.

## Conclusions

The findings presented here suggest that if a genetic proxy for drug efficacy is available, genetic survival analysis could be important in generating meaningful hypotheses for early clinical development of new drugs, as well as for finding patient populations for existing drugs. This set of analyses identifies individuals and diseases that are likely to benefit from treatment of a specific target, and therefore may be a tool to help guide the design of future clinical trials.

### Abbreviations

AF	Atrial Fibrillation
HF	Heart Failure
CV	Cardiovascular
HR	Hazard Ratio
ICD	International Classification of Diseases
IHD	Ischemic Heart Disease
ASCVD	Atherosclerotic Cardiovascular Disease
BMI	Body Mass Index
LDL	Low-Density Lipoprotein
LDLR	Low-Density lipoprotein receptor

SNP	Single Nucleotide Polymorphism
ACE	Angiotensin-Converting Enzyme Gene
PCSK9	Proprotein Convertase Subtilisin/Kexin Type 9
ADRB1	β <sub>1</sub> Adrenergic Receptor
ABCC8	ATP Binding Cassette Subfamily C Member 8
SUR1	Sulfonylurea Receptor 1
GWAS	Genome-Wide Association Study
PheWAS	Phenome-Wide Association Study
HES	Hospital Episode Statistics

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12872-025-04753-1>.

Supplementary Material 1. Supplemental Methods Tables S1-S15. Figures S1-S16.

### Acknowledgements

This research was conducted using the UK Biobank Resource under Application number 84103.

### Clinical trial number

Not applicable.

### Code and examples

<https://github.com/tenayatherapeutics/Genetic-Survival-Analysis-in-UKB/tree/main>.

### Authors' contributions

SMF and JRP designed the study. SMF and LZ performed statistical analysis and prepared figures. SMF, LZ and JRP drafted the paper. All authors reviewed the manuscript.

### Funding

All authors are employees of Tenaya Therapeutics.

### Data availability

Code and Examples: <https://github.com/tenayatherapeutics/Genetic-Survival-Analysis-in-UKB/tree/main>.

## Declarations

### Ethics approval and consent to participate

UK Biobank was approved by the North West Multi-Centre Research Ethics Committee and all participants provided written informed consent to participate.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Tenaya Therapeutics, South San Francisco, CA, USA. <sup>2</sup>University of California San Diego, San Diego, CA, USA. <sup>3</sup>Department of Pediatrics, Stanford University, Stanford, CA, USA.

Received: 13 November 2024 Accepted: 10 April 2025

Published online: 07 May 2025

## References

1. Wouters OJ, McKee M, Luyten J. Errors in Source Data for Study of Drug Development Costs. *JAMA*. 2022;328:1110.
2. Teerlink JR, Diaz R, Felker GM, McMurray JJV, Metra M, Solomon SD, Adams KF, Anand I, Arias-Mendoza A, Biering-Sørensen T, et al. Cardiac

- Myosin Activation with Omecamtiv Mecarbil in Systolic Heart Failure. *N Engl J Med*. 2021;384:105–16.
3. Docherty KF, McMurray JJV, Diaz R, Felker GM, Metra M, Solomon SD, Adams KF, Böhm M, Brinkley DM, Echeverria LE, et al. The Effect of Omecamtiv Mecarbil in Hospitalized Patients as Compared With Outpatients With HF: An Analysis of GALACTIC-HF. *J Card Fail*. 2024;30:26–35.
  4. Felker GM, Solomon SD, Claggett B, Diaz R, McMurray JJV, Metra M, Anand I, Crespo-Leiro MG, Dahlström U, Goncalvesova E, et al. Assessment of Omecamtiv Mecarbil for the Treatment of Patients With Severe Heart Failure: A Post Hoc Analysis of Data From the GALACTIC-HF Randomized Clinical Trial. *JAMA Cardiol*. 2022;7:26–34.
  5. Bernhard B, Heydari B, Abdullah S, Francis SA, Lumish H, Wang W, Jerosch-Herold M, Harris WS, Kwong RY. Effect of six month's treatment with omega-3 acid ethyl esters on long-term outcomes after acute myocardial infarction: The OMEGA-REMODEL randomized clinical trial. *Int J Cardiol*. 2024;399:131698.
  6. Lincroff AM, Brown-Frandsen K, Colhoun HM, Deanfield J, Emerson SS, Esbjerg S, Hardt-Lindberg S, Hovingh GK, Kahn SE, Kushner RF, et al. Semaglutide and Cardiovascular Outcomes in Obesity without Diabetes. *N Engl J Med*. 2023;389:2221–32.
  7. Fda. Enrichment Strategies for Clinical Trials to Support Determination of Effectiveness of Human Drugs and Biological Products Guidance for Industry. 2019.
  8. Druker BJ, Lydon NB. Lessons learned from the development of an abl tyrosine kinase inhibitor for chronic myelogenous leukemia. *J Clin Invest*. 2000;105:3–7.
  9. Ryan CJ, Devakumar LPS, Pettitt SJ, Lord CJ. Complex synthetic lethality in cancer. *Nat Genet*. 2023;55:2039–48.
  10. Millard LAC, Davies NM, Gaunt TR, Smith GD, Tilling K. Software Application Profile: PHESANT: a tool for performing automated phenome scans in UK Biobank. *Int J Epidemiol*. 2018;47:29–35.
  11. Berge KE, Ose L, Leren TP. Missense mutations in the PCSK9 gene are associated with hypocholesterolemia and possibly increased response to statin therapy. *Arterioscler Thromb Vasc Biol*. 2006;26:1094–100.
  12. Cameron J, Holla ØL, Ranheim T, Kulseth MA, Berge KE, Leren TP. Effect of mutations in the PCSK9 gene on the cell surface LDL receptors. *Hum Mol Genet*. 2006;15:1551–8.
  13. Rao AS, Lindholm D, Rivas MA, Knowles JW, Montgomery SB, Ingelsson E. Large-Scale Phenome-Wide Association Study of PCSK9 Variants Demonstrates Protection Against Ischemic Stroke. *Circ Genomic Precis Med*. 2018;11:e002162.
  14. Gai MT, Adi D, Chen XC, Liu F, Xie X, Yang YN, Gao XM, Ma X, Fu ZY, Ma YT, et al. Polymorphisms of rs2483205 and rs562556 in the PCSK9 gene are associated with coronary artery disease and cardiovascular risk factors. *Sci Rep*. 2021;11:11450.
  15. Foody JM, Farrell MH, Krumholz HM. beta-Blocker therapy in heart failure: scientific review. *JAMA*. 287(7):883–9.
  16. Vrablik M, Corsini A, Tümová E. Beta-blockers for Atherosclerosis Prevention: a Missed Opportunity? *Curr Atheroscler Rep*. 2022;24(3):161–9.
  17. ACC/AHA/ACCP/HRS Guideline for the Diagnosis and Management of Atrial Fibrillation: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. 2023;149:E936.
  18. Mason DA, Moore JD, Green SA, Liggett SB. A gain-of-function polymorphism in a G-protein coupling domain of the human beta1-adrenergic receptor. *J Biol Chem*. 1999;274:12670–4.
  19. Sandilands AJ, O'Shaughnessy KM. The functional significance of genetic variation within the beta-adrenoceptor. *Br J Clin Pharmacol*. 2005;60:235–43.
  20. Zhang F, Steinberg SF. S49G and R389G polymorphisms of the beta<sub>1</sub>-adrenergic receptor influence signaling via the cAMP-PKA and ERK pathways. *Physiol Genomics*. 2013;45:1186–92.
  21. Bengtsson K, Melander O, Orho-Melander M, Lindblad U, Ranstam J, Råstam L, Groop L. Polymorphism in the beta(1)-adrenergic receptor gene and hypertension. *Circulation*. 2001;104:187–90.
  22. Surendran P, Feofanova EV, Lahrouchi N, Ntalla I, Karthikeyan S, Cook J, Chen L, Mifsud B, Yao C, Kraja AT, et al. Discovery of rare variants associated with blood pressure regulation through meta-analysis of 1.3 million individuals. *Nat Genet*. 2020;52:1314–32.
  23. Hyman MC, Levin MG, Gill D, Walker VM, Georgakis MK, Davies NM, Marchlinski FE, Damrauer SM. Genetically Predicted Blood Pressure and Risk of Atrial Fibrillation. *Hypertens*. (Dallas, Tex. 1979). 2021;77:376–382.
  24. Chung CM, Wang RY, Chen JW, Fann CSJ, Leu HB, Ho HY, Ting CT, Lin TH, Sheu SH, Tsai WC, et al. A genome-wide association study identifies new loci for ACE activity: potential implications for response to ACE inhibitor. *Pharmacogenomics J*. 2010;10:537–44.
  25. Gudjonsson A, Gudmundsdottir V, Axelsson GT, Gudmundsson EF, Jonsen BG, Launer LJ, Lamb JR, Jennings LL, Aspelund T, Emilsson V, et al. A genome-wide association study of serum proteins reveals shared loci with common diseases. *Nat Commun*. 2022;13:480.
  26. Emdin CA, Klarin D, Natarajan P, Florez JC, Kathiresan S, Khera AV. Genetic Variation at the Sulfonylurea Receptor, Type 2 Diabetes, and Coronary Heart Disease. *Diabetes*. 2017;66:2310–5.
  27. Fatehi M, Raja M, Carter C, Soliman D, Holt A, Light PE. The ATP-sensitive K(+) channel ABCC8 S1369A type 2 diabetes risk variant increases MgATPase activity. *Diabetes*. 2012;61:241–9.
  28. Florez JC, Burt N, De Bakker PIW, Almgren P, Tuomi T, Holmkvist J, Gaudet D, Hudson TJ, Schaffner SF, Daly MJ, et al. Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene region. *Diabetes*. 2004;53:1360–8.
  29. Gloyn AL, Weedon MN, Owen KR, Turner MJ, Knight BA, Hitman G, Walker M, Levy JC, Sampson M, Halford S, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. *Diabetes*. 2003;52:568–72.
  30. Scott RA, Freitag DF, Li L, Chu AY, Surendran P, Young R, Grarup N, Stancáková A, Chen Y, Varga TV, et al. A genomic approach to therapeutic target validation identifies a glucose-lowering GLP1R variant protective for coronary heart disease. *Sci Transl Med*. 2016;8:341ra76.
  31. Steeds RP, Birchall AS, Smith M, Channer KS. An open label, randomised, crossover study comparing sotalol and atenolol in the treatment of symptomatic paroxysmal atrial fibrillation. *Heart*. 1999;82:170–5.
  32. Bi W, Fritsche LG, Mukherjee B, Kim S, Lee S. A Fast and Accurate Method for Genome-Wide Time-to-Event Data Analysis and Its Application to UK Biobank. *Am J Hum Genet*. 2020;107:222–33.
  33. Hukerikar N, Hingorani AD, Asselbergs FW, Finan C, Schmidt AF. Prioritising genetic findings for drug target identification and validation. *Atherosclerosis*. 2024;390.
  34. Pedersen EM, Agerbo E, Plana-Ripoll O, Steinbach J, Krebs MD, Hougaard DM, Werge T, Nordentoft M, Børglum AD, Musliner KL, et al. ADuLT: An efficient and robust time-to-event GWAS. *Nat Commun*. 2023;14:5553.
  35. Fu EL, Groenwold RHH, Zoccali C, Jager KJ, Van Diepen M, and Dekker FW (2019). Merits and caveats of propensity scores to adjust for confounding. *Nephrol Dial Transplant*. 2002;34:1629–35.
  36. Walsh MC, Trentham-Dietz A, Newcomb PA, Gangnon R, Palta M. Use of Propensity Scores to Reduce Case-Control Selection Bias. *Epidemiology*. 2012;23:772–735.
  37. Healey JS, Baranchuk A, Crystal E, Morillo CA, Garfinkle M, Yusuf S, Connolly SJ. Prevention of atrial fibrillation with angiotensin-converting enzyme inhibitors and angiotensin receptor blockers: a meta-analysis. *J Am Coll Cardiol*. 2005;7;45(11):1832–9.
  38. Tai C, Gan T, Zou L, Sun Y, Zhang Y, Chen W, Li J, Zhang J, Xu Y, Lu X, Xu D. Effect of angiotensin-converting enzyme inhibitors and angiotensin II receptor blockers on cardiovascular events in patients with heart failure: a meta-analysis of randomized controlled trials. *BMC Cardiovasc Disord*. 2017;17(1):257.
  39. Davies MJ, Aroda VR, Collins BS, Gabbay RA, Green J, Maruthur NM, Rosas SE, Del Prato S, Mathieu C, Mingrone G, et al. Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*. 2022;45:2753–86.
  40. Low Wang CC, Everett BM, Burman KD, Wilson PWF. Cardiovascular Safety Trials for All New Diabetes Mellitus Drugs: Ten Years of FDA Guidance Requirements to Evaluate Cardiovascular Risk. *Circulation*. 2019;139:1741–3.
  41. McGuire DK, Busui RP, Deanfield J, Inzucchi SE, Mann JFE, Marx N, Mulvagh SL, Poulter N, Engelman MDM, Hovingh GK, et al. Effects of oral semaglutide on cardiovascular outcomes in individuals with type 2 diabetes and established atherosclerotic cardiovascular disease and/or chronic kidney disease: Design and baseline characteristics of SOUL, a randomized trial. *Diabetes Obes Metab*. 2023;25:1932–41.
  42. Karim ME, Petkau J, Gustafson P, Platt RW, Tremlett H, and BeAMS Study Group. Comparison of statistical approaches dealing with time-dependent confounding in drug effectiveness studies. *Stat Methods Med Res*. 2018;27:1709–22.

43. VanderWeele TJ, Tchetgen Tchetgen EJ, Cornelis M, Kraft P. Methodological Challenges in Mendelian Randomization. *Epidemiology*. 2014;25:427–35.
44. Estrada K, Froelich S, Wuster A, Bauer, CR, Sterling T, Clark WT, Ru Y, Trinidad M, Nguyen HP, Luu AR, et al. Identifying therapeutic drug targets using bidirectional effect genes. *Nat Commun*. 2021;12:2224.
45. Sun BB, Kurki MI, Foley CN, Mechakra A, Chen CY, Marshall E, Wilk JB, Ghen CY, Wilk JB, Runz H, et al. Genetic associations of protein-coding variants in human disease. *Nature*. 2022;603:95–102.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.