THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Narula S, Yusuf S, Chong M, et al. Plasma ACE2 and risk of death or cardiometabolic diseases: a case-cohort analysis. *Lancet* 2020; **396:** 968–76.

Supplementary Appendix

Table of Contents

Supplementary Appendix 1: Supplementary Methods

Selection of Countries

The following information on the PURE study is excerpted from our previously published protocol and subsequent publications which describe the design, sampling, and adjudication in detail.[1] The choice and number of countries selected in PURE reflects a balance between involving a large number of communities in countries at different economic levels, with substantial heterogeneity in social and economic circumstances and policies, and the feasibility of centers to successfully achieve long-term follow-up. Thus, PURE included sites in which investigators are committed to collecting good-quality data for a low-budget study over the planned 10-year follow-up period and did not aim for a strict proportionate sampling of the entire world. The following countries and territories participate in the PURE study: Argentina, Bangladesh, Brazil, Canada, Chile, China, Colombia, Ecuador, India, Iran, Kazakhstan, Kyrgyzstan, Malaysia, Pakistan, Palestine, Peru, Philippines, Poland, Russia, Saudi Arabia, South Africa, Sweden, Tanzania, Turkey, United Arab Emirates, Uruguay, Zimbabwe.

Selection of Communities

Within each country, urban and rural communities were selected based on broad guidelines (see Guidelines for Selection of Countries, Communities, Households, and Individuals Recruited to PURE). A common definition for "community" that is applicable globally is difficult to establish. In PURE, a community was defined as a group of people who have common characteristics and reside in a defined geographic area. A city or large town was not usually considered a single community, rather communities from low-, middle-, and high-income areas were selected from sections of the city and the community area defined according to a geographical measure (e.g., a set of contiguous postal code areas or a group of streets or a village). The primary sampling unit for rural areas in many countries was the village. The reason for inclusion of both urban and rural communities is that for many countries, urban and rural environments exhibit distinct characteristics in social and physical environment, and hence, by sampling both, we ensured considerable variation in societal factors across PURE communities.

The number of communities selected in each country varied, with the aim to recruit communities with substantial heterogeneity in social and economic circumstances balanced against the capacity of local investigators to maintain follow-up. In some countries (e.g., India, China, Canada, and Colombia), communities from several states/provinces were included to capture regional diversity, in policy, socioeconomic status, culture, and physical environment. In other countries (e.g. Iran, Poland, Sweden, and Zimbabwe), fewer communities were selected.

Selections of Households and Individuals

Within each community, sampling was designed to achieve a broadly representative sample of that community of adults aged between 35 and 70 years. The choice of sampling frame within each center was based on both "representativeness" and feasibility of long-term follow-up, following broad study guidelines. Once a community was identified, where possible, common and standardized approaches were applied to the enumeration of households, identification of individuals, recruitment procedures, and data collection.

The method of approaching households differed between regions. For example, in rural areas of India and China, a community announcement was made to the village through contact of a community leader, followed by in-person door-to-door visits of all households. In contrast, in Canada, initial contact was by mail followed by telephone inviting members of the households to a central clinic. Households were eligible if at least 1 member of the household was between the ages of 35 and 70 years and the household members intended to continue living in their current home for a further 4 years.

For each approach, at least 3 attempts at contact were made. All individuals within these households between 35 and 70 years providing written informed consent were enrolled. When an eligible household or eligible individual in a household refused to participate, demographics and self-reported data about CVD risk factors, education, and history of CVD, cancers and deaths in the households within the 2 previous years were recorded.

To ensure standardization and high data quality, we used a comprehensive operations manual, training workshops, DVDs, regular communication with study personnel and standardized report forms. We entered all data in a customized database programmed with range and consistency checks, which was transmitted, electronically to the Population Health Research Institute in Hamilton (Ontario, Canada) where further quality checks were implemented.

Guidelines for Selection of Countries, Communities, Households, and Individuals Recruited to PURE

Countries
1. High-income countries, middle-income countries, and low-income countries, with the bulk of the recruitment from low- and middle-income regions.
2. Committed local investigators with experience in recruiting for population studies.
Communities
1. Select both urban and rural communities. Use the national definition of the country to determine urban and rural communities.
2. Select rural communities that are isolated (distance of >50 km or lack easy access to commuter transportation) from urban centers. However, consider ability to process bloods samples, e.g., villages in rural developing countries should be within 45-min drive of an appropriate facility.
3. Define community to a geographical area, e.g., using postal codes, catchment area of health service/clinics, census tracts, areas bordered by specific streets or natural borders such as a river bank.
4. Consider feasibility for long-term follow-up, e.g., for urban communities, choose sites that have a stable population such as residential colonies related to specific work sites in developing countries. In rural areas, choose villages that have a stable population. Villages at greater distance from urban centers are less susceptible to large migration to urban centers.
5. Enlist a community organization to facilitate contact with the community, eg, in urban areas, large employers (government and private), insurance companies, clubs, religious organizations, clinic or hospital service regions. In rural areas, local authorities such as priests or community elders, hospital or clinic, village leader, or local politician.
Individual
1. Broadly representative sampling of adults 35 to 70 years within each community unit.
2. Consider feasibility for long-term follow-up when formulating community sampling framework, e.g., small percentage random samples of large communities may be more difficult to follow-up because they are dispersed by distance. In rural areas of developing countries that are not connected by telephone, it may be better to sample entire community (i.e., door-to-door systematic sampling).

3. The method of approach of households/individuals may differ between sites. In MIC and HIC, mail, followed up by phone contact may be the practical first means of contact. In LIC, direct household contact through household visits may be the most appropriate means of first contact.
4. Once recruited, all individuals are invited to a study clinic to complete standardized questionnaires and have a standardized set of measurements.

65
66

PURE Data Collection Procedures

Data have been collected at national, community, household, and individual levels with standardized questionnaires (cite). Questions about age, sex, education, smoking status, hypertension, type 2 diabetes, and obesity were identical to those in the INTERHEART and INTERSTROKE studies.[2,3] We obtained blood pressure (BP) measurements in individuals and hypertension was defined as a BP >140/>90 mmHg or in individuals who were already receiving treatment. Fasting glucose was available in most individuals (76%) and type 2 diabetes was defined as those individuals reporting type 2 diabetes or those with a fasting plasma glucose ≥7·0 mmol/L or a HbA1c ≥6·5% or a 2-hour plasma glucose on oral glucose tolerance test ≥11·1 mmol/L. In most of the low-income countries (LIC) and medium-income countries (MIC) there was no central system of death or event registration. Therefore, to arrive at a probable diagnosis or cause of death, we (1) obtained information on prior medical illness and medically certified causes of death, where available, or (2) captured the best available information from reliable sources when medical information was not available. Event documentation was based on information from household interviews and medical records, death certificates and other sources. We also used Verbal Autopsies to ascertain cause of death in addition to medical records, which were reviewed by a health professional. This approach has been used in several studies conducted in LIC and MIC.

To ensure a standard approach and accuracy for the classification of events across all countries and over time, the first 100 CVD events (deaths, myocardial infarction (MI), strokes, heart failure (HF)) for China and India, and 50 cases for other countries were adjudicated locally and by the adjudication chair. If necessary, further training was provided. Every year thereafter, 50 cases for China and India and 25 cases for each of the remaining countries were adjudicated as above.

The standard operating procedure of the PURE study with regard to collection and storage consisted of drawing fasting and non-fasting blood samples from individuals. Samples were subsequently separated into six equal volumes, and frozen immediately at –20°C or –70°C after processing. Samples were shipped in nitrogen vapor tanks from every site to a blood storage site, where they were stored at –160°C in liquid nitrogen (Hamilton). Samples from China, India, Turkey and Malaysia are kept locally because of legislations prohibiting export of biological specimens. Blood samples were previously analyzed for total cholesterol, HDL cholesterol, apoB, apoA1 and glycated hemoglobin.

Selection for the Case-Cohort Analysis

This PURE biomarker analysis is a subsample/sub-study of the original PURE study. Eligibility for the biomarker analysis in PURE needed to maximize the opportunity for assessing novel protein and genetic markers of risk in a statistically appropriate manner, while doing so in a manner that was cost-effective from the perspective of genotyping and multiplex biomarker analysis. As such, the case-cohort design was deemed the most reasonable design. Eligibility at the outset was determined in the following manner:

Inclusion criteria

- Member of a major ethnicity in a residing country (i.e. European Caucasian in Sweden)
- Blood sample available for biomarker and genetic analysis

Exclusion criteria

- Country does not allow export of biological sample (China, India, Turkey, Malaysia) *
- Non-fasting blood only, or missing/inadequate blood sample

* Future targeted validation/replication studies may be possible.

Among the ~55,000 PURE participants with biological specimen stored at our institution locally, "cases" (those not in sub-cohort) were selected if they had at least one major adverse health event , including myocardial infarction (MI), stroke, heart failure (HF), type 2 diabetes (T2D) or death (from all causes, including cardio-vascular diseases). Corresponding members of the 'control' group were selected by frequency-matching according to major country specific ethnicity. This procedure is described as follows:

(1) Incident MI, stroke, HF, T2D, and death were tabulated by major country-specific ethnicity.

(2) The number of controls to be selected in each major country-specific ethnicity is the sum of the adverse health events mentioned above. This approach guaranteed that all cases (by country and ethnicity) had matched controls.

(3) Among the 55,246 PURE participants with biological specimen stored at PHRI, **a random sample of subjects equal to the controls counted in step 2 was selected in each major country-specific ethnicity**. Importantly, *selected controls may have had adverse health events as per case-cohort design*.

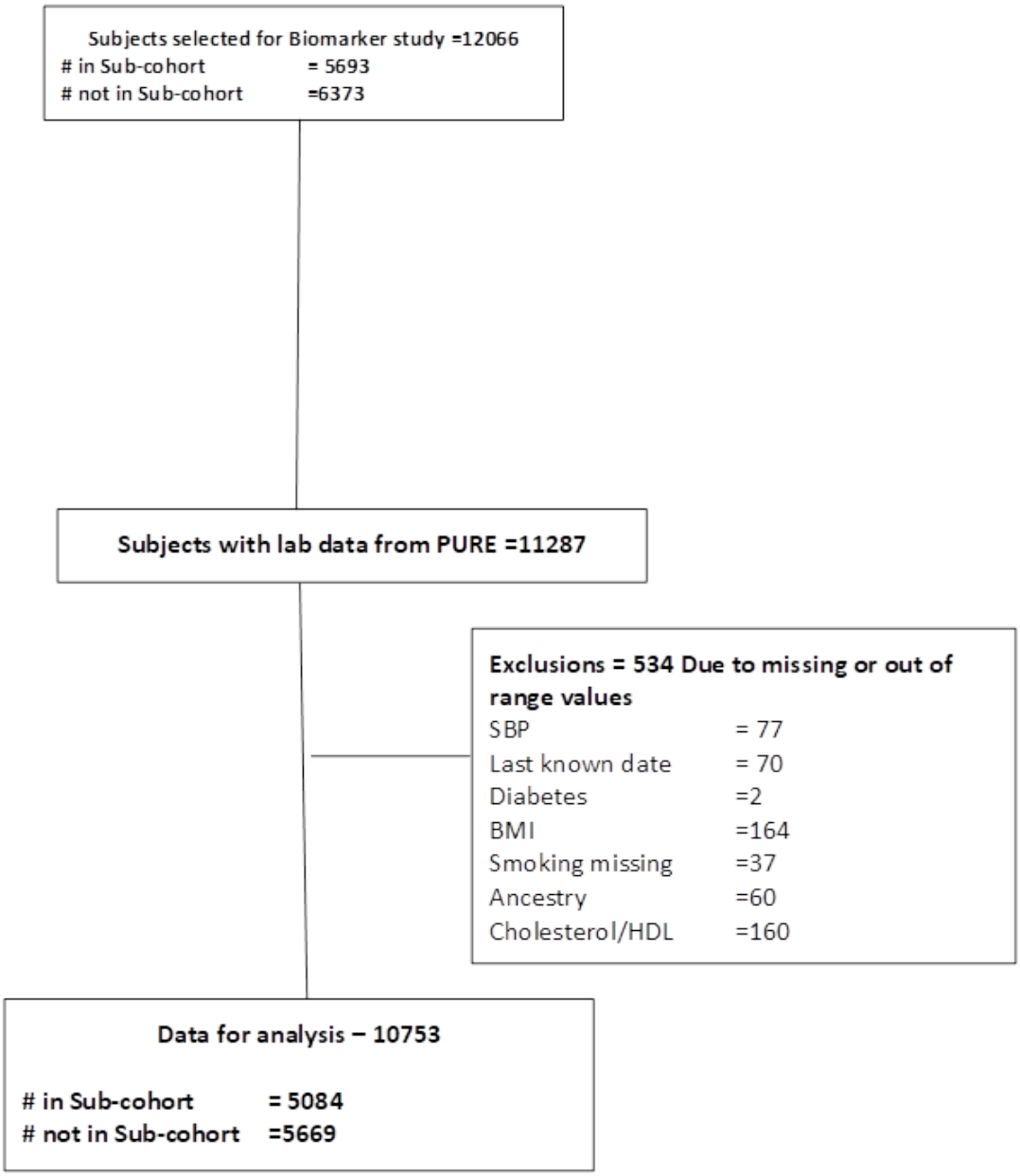
(4) After this initial selection process, 12,066 participants were deemed eligible prior to proteomic quality control measures (described further in further detail below)

(5) A subsequent selection process was undertaken, where individuals whose measurements did not meet quality control standards was undertaken. This yielded 11,287 participants. In other words, after running samples through the proteomics platform 779 participants (12,066 – 11,287), were further deemed ineligible for analysis.

PURE Subjects meeting eligibility criteria = 55,246

Randomly sampled for sub-

participants


Subjects selected for Biomarker study =12066
in Sub-cohort = 5693
not in Sub-cohort =6373
Subjects with lab data from PURE =11287
Exclusions = 534 Due to missing or out of range values
SBP = 77
Last known date = 70
Diabetes =2
BMI =164
Smoking missing =37
Ancestry =60
Cholesterol/HDL =160
Data for analysis – 10753
in Sub-cohort = 5084
not in Sub-cohort =5669

141

142

PURE Biomarker Sub-study Case-Cohort Sampling Depiction:

The above represents a schematization of our case-cohort study.

Incident outcomes collected as part of the PURE Biomarker sub-study include death, myocardial infarction, stroke, diabetes, and heart failure.

*Note: These represent the number of participants available for analysis after accounting for samples that failed genomic and proteomic quality control and/or participants with extreme or missing values in variables of interests

Standardized Event Definitions in PURE

Prospective Follow-up for Cardiovascular Events and Mortality: History of disease was collected at baseline from every participant with standardized questionnaires regarding history of a) hypertension, b) diabetes c) stroke d) angina/myocardial infarction/coronary artery disease e) heart failure f) other heart disease.

Information on specific events (death, myocardial infarction, stroke, heart failure, cancer, hospitalizations, new diabetes, injury, tuberculosis, human immunodeficiency viral infections, malaria, pneumonia, asthma, chronic obstructive pulmonary disease) were obtained from participants or their family members (events were reported by the participants if alive or by a relative if the individual had died). This information was adjudicated centrally in each country by trained physicians using standardized definitions. Because the PURE study involves urban and rural areas from middle- and low-income countries, supporting documents to confirm cause of death and/or event varied in degrees of completion and availability. In most of middle- and low-income countries there was no central system of death or event registration. Therefore, information was obtained about prior medical illness and medically certified cause of death where available, and, second, best available information was captured from reliable sources in those instances where medical information was not available in order to be able to arrive at a probable diagnosis or cause of death. Event documentation was based on information from household interviews and medical records, death certificates and other sources. Verbal autopsies were also used to ascertain cause of death in addition to medical records which were reviewed by a health professional. This approach has been used in several studies conducted in middle- and low-income countries.

To ensure a standard approach and accuracy for classification of events across all countries and over time, the first 100 CVD events (deaths, MI, strokes, heart failure or cancers) for China and India, and 50 cases for other countries were adjudicated both locally and also by the adjudication chair, and if necessary further training was provided. Thereafter, every year, 50 cases for China and India and 25 cases for each of the remaining countries were adjudicated as above.

FATAL EVENTS

Cardiovascular Death – Definitions

01.00 DEATH DUE TO CARDIOVASCULAR EVENTS

01·10 Sudden unexpected Cardiovascular Death (SCVD)

Without evidence of other cause of death, death that occurred suddenly and unexpectedly (examples: witnessed collapse, persons resuscitated from cardiac arrest who later died) or persons seen alive less than 12 hours prior to discovery of death (example persons found dead in his/her bed).

- SCVD is either definite, probable or possible according to the following characteristics:

PURE Adjudication Code	Event Type	Acceptable ICD-10 codes

01·11: Definite	One of the following in persons with: • known cardiovascular disease, or • diabetes with an additional risk factor such as hypertension, smoking, dyslipidemia, micro albuminuria, serum creatinine 50% above upper limit of normal, or • 3 of the above risk factors, or • 2 of the above risk factors in men aged 60 and more and women aged 65 and more	No ICD-10 Code
01·12: Probable	One of the following in persons with: • diabetes, or • 2 of the above risk factors in men aged less than 60 and in women less than 65, or • one of the above risk factor in men aged 60 and more and in women aged 65 and more, or • typical of chest pain or sudden severe dyspnea of less than 20-minute duration preceding the event	
01·13: Possible	In persons without risk factor	
For SCVD, the patient was well or had a stable CVD (example stable angina) when last seen alive. The event of a sudden death occurring during the hospitalization of MI is considered a fatal MI and not sudden death.		

01·30 Fatal Myocardial Infarction (MI)

Symptoms of Myocardial Infarction:

Typical symptoms or suggestive symptoms of MI according to physician are characterized by severe anterior chest pain as tightness, crushing, burning, lasting at least 20 minutes, occurring at rest, or on exertion, that may radiate to the arms or neck or jaw and may be associated with dyspnea, diaphoresis and nausea. However, death associated with nausea and vomiting with or without chest pain not due to another cause may be considered as possible MI if ECG and cardiac markers are not done. These symptoms may have occurred the last month before death.

Fatal myocardial infarction is either definite, probable or possible according to the following characteristics:

PURE Adjudication Code	Event Type	Acceptable ICD-10 codes

01·31: Definite	1. Autopsy demonstrating fresh myocardial infarction and/or recent coronary occlusion, or 2. ECG showing new and definite sign of MI (Minnesota code 1-1-1) or 3. Symptoms typical or atypical or inadequately described but attributed to cardiac origin lasting at least 20 minutes and by troponin or cardiac enzymes (CKMB, CK, SGOT, SLDH) above center laboratory ULN 4. ECG with new ischemic changes (new ST elevation/depression or T wave inversion ≥ 2 mm) and by troponin or cardiac enzymes (CKMB, CK, SGOT, SLDH) above center laboratory ULN	I21- I22
01·32: Probable	1. ECG with sign of probable MI (Minnesota code 1-2-1), or 2. Typical symptoms lasting at least 20 minutes considered of cardiac origin, with only new ST-T changes (new ST elevation/depression or T wave inversion ≥ 1 but < 2mm) without documented increased cardiac markers or enzyme as in PURE definition 1·31 (above), or 3. Increased cardiac enzymes as in PURE definition 1·31 (above) showing a typical pattern of MI as above without symptoms or significant ECG changes	
01·33: Possible	1. ECG with sign of possible MI (Minnesota code 1-3-1) or 2. Typical symptoms or symptoms suggestive of MI according to the physician lasting at least 20 minutes without documented ECG or cardiac marker.	

The Minnesota codes for MI is taken from Rose and Blackburn and published in their book "Evaluation Methods of Cardiovascular Disease WHO 1969".

- Definite MI is Q/R ratio ≥1/3 and Q duration ≥ 0·03 second in one of the following leads: I, II, V2, 3, 4, 5, 6. (code 1-1-1)
- Probable MI is Q/R ratio ≥1/3 and Q duration between 0·02 and 0·03 second in one of the following leads: I, II, V2, 3, 4, 5, 6. (code 1-2-1)
- Possible MI is Q/R ratio between 1/5 and 1/3 and Q duration between 0·02 and 0·03 second in one of the following leads: I, II, V2, 3, 4, 5, 6. (code 1-3-1)

01·40 Fatal Stroke

Fatal stroke is either definite or possible according to the following characteristics:

PURE Adjudication Code	Event Type	Acceptable ICD-10 codes
01·41: Definite	Stroke death is defined as death within 30 days from an acute focal neurological deficit *diagnosed by a physician* and thought to be of vascular origin (without other cause such as brain tumor) with signs and symptoms lasting >= 24 hrs. Stroke death is also considered if death occurred within 24 hrs. of onset of persisting signs and symptoms, or if there is evidence of a recent stroke on autopsy. N.B. • In a subject with a stroke <= 30 days: If death occurred with a pneumonia due to possible aspiration, death will be considered to be due to stroke. • In a subject with a stroke > 30 days: If death occurred with a pneumonia due to possible aspiration, the adjudicator will make a decision according to his/her clinical judgment if death is related to stroke or not. • Subarachnoid hemorrhage death manifested by sudden onset headache with/without focal signs and imaging (CT or MRI) evidence of bleeding primarily in the subarachnoid space is considered a fatal stroke in absence of trauma or brain tumor or malformation • Subdural hematoma death is not considered as a stroke death and may be related to previous trauma or other cause.	I60- I64, I69
01·43: Possible	Death in a participant with a history of sudden onset of focal neurological deficit of one or more limbs, loss of vision or slurred speech lasting about 24 hours.	

01·50 Fatal Congestive Heart Failure

Fatal congestive heart failure is either definite or possible according to the following characteristics:

PURE Adjudication Code	Event Type	Acceptable ICD-10 codes

01·51: Definite	The diagnosis of congestive heart failure may be an autopsy finding in absence of other cause or requires signs (rales, increased jugular venous pressure or ankle edema) or symptoms (nocturnal paroxysmal dyspnea, dyspnea at rest or ankle edema) of congestive heart failure and one or both of the following: • radiological signs of pulmonary congestion, • treatment of heart failure with diuretics *If sudden death occurred in a patient with chronic severe heart failure, it should be adjudicated as fatal congestive heart failure.*	I50
01·52: Probable	Progressive shortness of breath on lying down or at night, improving on sitting up AND any of the following signs or symptoms: swelling of feet, distension of abdomen, progressive cough in a person with known hypertension or a history of previous MI/angina or other heart disease	
01·53: Possible	Progressive shortness of breath on lying down or at night, improving on sitting up AND any of the following signs or symptoms: swelling of feet, distension of abdomen, progressive cough	

01·60 Death Due to Other Cardiovascular Deaths *(other causes [1·10 to 1·50 above] having been excluded)*

PURE Adjudication Code	Event Type	Acceptable ICD-10 codes
01·61	Arterial rupture of aneurysm	I71- I72
01·62	Pulmonary embolism *NOTE: Death associated with pulmonary embolism occurring within 2 weeks after a fracture such as hip, femur should attributed to death due to injury. Refer to Injury, Section 6·0*	I26
01·63	Arrhythmic death (A-V block, sustained ventricular tachycardia in absence of other causes)	I44- I45, I47- I49
01·64	Death after invasive cardiovascular intervention: a perioperative death extending to 30 days after coronary or arterial surgical revascularization and to 7 days after a coronary or arterial percutaneous dilatation (angioplasty) with or without a stent or an invasive diagnostic procedure.	I97

01·65	Congenital heart disease	Q20-Q28
01·66	Heart valve disease (including rheumatic heart disease)	I01, I05- I09, I34- I37
01·67	Endocarditis	I33, I38
01·68	Myocarditis	I40
01·69	Tamponade (pericarditis)	I30, I31, I32
01·70	Other cardiovascular events *(Excluding 1·61 to 1·69 above)* *Valid ICD-10 codes would include the following:* *I11, I12, I13, I23, I24, I25, I27, I28, I42, I51, I52, I65-I68, I73, I74, I96, I98, I99 (Refer to ICD-10 Listing for associated definitions for each code)*	Any valid 'I' (Cardiovascular) ICD-10 code that can be classified as underlying cause of death, not specified above

206

207 NON-FATAL EVENTS

208 Cardiovascular Events – Definitions

209 10·00 NON-FATAL CARDIOVASCULAR EVENTS

210 10·10 Non-Periprocedural Myocardial Infarction (MI)

211 MI is considered either definite, probable or possible according to the following characteristics:

212

PURE Adjudication Code	Event Type	Acceptable ICD-10 codes
10·11: Definite	1. ECG showing new and definite sign of MI (Minnesota code 1-1-1) or 2. Symptoms typical or atypical or inadequately described but attributed to cardiac origin lasting at least 20 minutes and by troponin or cardiac enzymes (CKMB, CK, SGOT, SLDH) above center laboratory ULN	

	3. ECG with new ischemic changes (new ST elevation/depression or T wave inversion ≥ 2 mm) and by troponin or cardiac enzymes (CKMB, CK, SGOT, SLDH) above center laboratory ULN Please note that increased markers may occur in trauma (CK, AST, myoglobin and CK MB to a lesser degree); renal insufficiency, heart failure, pulmonary embolism (troponin), cardioversion (all)	
10·12: Probable	1. ECG with new and probable sign of MI (Minnesota code 1-2-1), or 2. Typical symptoms lasting at least 20 minutes considered of cardiac origin, with only new ST-T changes (new ST elevation/depression or T wave inversion ≥ 1 but < 2mm) without documented increased cardiac markers as in PURE definition 10·11 (above), or 3. Increased cardiac enzymes showing a typical pattern of MI as above without symptoms or significant ECG changes.	I21-I22
10·13: Possible	1. ECG with new and possible sign of MI (Minnesota code 1-3-1), or 2. Typical symptoms lasting 20 minutes and more considered to be of cardiac origin without documented ECG or cardiac marker.	

10·20 Periprocedural Myocardial Infarction

PURE Adjudication Code	Event Type	Acceptable ICD-10 codes
10·21: Definite	1. ECG showing new and definite sign of MI (Minnesota code 1-1-1), or 2. Increased cardiac markers within 48 hours of procedure: • percutaneous coronary intervention: CKMB should be ≥ 5 X ULN or troponin ≥ 5 X above lower level of necrosis OR > 20% increase in cardiac markers if elevated at the beginning of the procedure in a patient with symptoms suggestive of myocardial ischemia • Coronary surgery: Increased cardiac markers CKMB should be ≥ 10X ULN or troponin ≥ 10X above lower limit of necrosis.	I21-I22

The Minnesota codes for MI is taken from Rose and Blackburn and published in their book "Evaluation Methods of Cardiovascular Disease WHO 1969".

- Definite MI is Q/R ratio ≥1/3 and Q duration ≥ 0·03 second in one of the following leads: I, II, V2, 3, 4, 5, 6. (code 1-1-1)
- Probable MI is Q/R ratio ≥1/3 and Q duration between 0·02 and 0·03 second in one of the following leads: I, II, V2, 3, 4, 5, 6. (code 1-2-1)
- Possible MI is Q/R ratio between 1/5 and 1/3 and Q duration between 0·02 and 0·03 second in one of the following leads: I, II, V2, 3, 4, 5, 6. (code 1-3-1)

10·30 Stroke/Transient Ischemic Attack (TIA)

PURE Adjudication Code	Event Type	Acceptable ICD-10 codes
10·31: Definite	Stroke is defined as an acute focal neurological deficit *diagnosed by a physician* and thought to be of vascular origin (without other case such as brain tumor) with signs and symptoms lasting ≥ 24 hrs. N.B. • Subarachnoid hemorrhage manifested by sudden onset headache with/without focal signs and imaging (CT or MRI or lumbar puncture) showing evidence of bleeding primarily in the subarachnoid space is considered a stroke in absence of trauma or brain tumor or malformation • Subdural hematoma is not considered as a stroke and may be related to previous trauma or other cause.	I60-I64, I69
10·33: Possible	Stroke is possible if there is a history of sudden onset of focal neurological deficit of one or more limbs, loss of vision or slurred speech lasting about 24 hours or more	
10·34: TIA	The diagnosis of TIA requires the presence of acute focal neurological deficit thought to be of vascular origin with signs and symptoms lasting less than 24 hours	G45

10·40 Congestive Heart Failure

PURE	Event Type	Acceptable ICD-10 codes

Adjudication Code		
10·41: Definite	The diagnosis of congestive heart failure requires signs (rales, increased jugular venous pressure or ankle edema) or symptoms (nocturnal paroxysmal dyspnea, dyspnea at rest or ankle edema) of congestive heart failure and one or both of the following: • radiological signs of pulmonary congestion, • Treatment of heart failure with diuretics.	I50
10·42: Probable	Progressive shortness of breath on lying down or at night, improving on sitting up AND any of the following signs or symptoms: swelling of feet, distension of abdomen, progressive cough in a person with known hypertension or a history of previous MI/angina or other heart disease	
10·43: Possible	Congestive heart failure is considered possible when there is progressive shortness of breath on lying down or at night, improving on sitting up AND any of the following signs or symptoms: swelling of feet, distension of abdomen, progressive cough	

Additional notes on Statistical Methods

While 12,066 individuals were considered eligible at the outset, following proteomic quality control, an additional 779 samples were considered unsuitable for analysis. As such, this case-cohort study includes 11,287 participants from PURE study. These 779, because they were not suitable for analysis, were deemed ineligible for analysis and as such were believed to have no impact with regards to biasing final estimates. Among these 11,287, a total of 534 subjects were excluded from the study sample (subjects with missing or out of range values in systolic blood pressure, diabetes or smoking status at baseline, BMI, ancestry, last known date and subjects with out of range total cholesterol or HDL). We imputed Total cholesterol, HDL, LDL and triglycerides. The missing data pattern was arbitrary and we used PROC MI to generate 5 imputed data sets with 'FCS' (fully conditional specification) method. We used sex as a classification variable and the data were imputed using age, sex and ancestry. In order to adjust for the effects of antihypertensive medication on systolic blood pressure, we added 15mmHg to the measured systolic blood pressure.[4,5] In order to account for the effects of cholesterol lowering medication on blood lipid concentrations, all individuals on lipid lowering medications were corrected in the following manner: total cholesterol/0·8 and LDL/0·7. HDL and triglycerides were not adjusted.[6]

For the analysis of ACE2 determinants, coefficients of the ordinary least squares regression were pooled using Rubin's rules. The adjusted R^2 for the final model (which included age, sex, smoking, BMI, ancestry, diabetes, blood pressure, and LDL cholesterol) was 0·199. For each cardiovascular outcome and mortality outcome, case-cohort analyses were performed using weighted proportional hazard model and used Self and Prentice weights. PROC MIANLYZE was used to pool the estimates to account for the variability introduced by imputations. The association measure was presented as a hazard ratio per 1 standard deviation (SD) unit increase in the marker, adjusted for the following: age, sex, smoking, BMI, systolic blood pressure, non-HDL cholesterol, and geographic ancestry. Each outcome was also adjusted for diabetes status, however, in the diabetes analysis, individuals with confirmed diabetes status were excluded. We assess the proportionality assumption using Schoenfeld Residual plots over time. Final effect estimates of ACE2 relationship with outcomes are presented with sex, non-HDL cholesterol and BMI as time dependent covariates to the model as these covariates violates the assumption.

Variable Ranking:

Predictors were ranked on the basis of a likelihood ratio chi square statistic. Variable rankings were performed on the first imputed dataset generated. We began by fitting the full model. In this, the full model was fit and compared to the model without each variable. For example, the full model was compared to the same model without non-HDL cholesterol in order to obtain a chi-square statistic. Subsequently, the full model was compared to the model without sex. This was done iteratively for each variable.

Ethical Considerations

All Centers are required to obtain approval from their respective ethics committees (Institutional Review Boards). All subjects' data are confidential and only authorized individuals will have access to study related documents at study Centers. Subjects' identification will be protected for documents (e.g. CRF) transmitted to the Coordinating Office, as well as biomarker and genetic data. Informed consent to obtain the baseline information, to collect and store the genetic and other biological specimens was obtained from all individuals.

ACE2 Assay Validation

Plasma ACE2 levels were measured by proximity extension assay using the Olink Proseek Cardiovascular II96×96 (CVDII) reagent kit (Olink, Uppsala, Sweden), which enable the analysis of 92 CVD-related proteins across 96 individual samples simultaneously. Analytical performance of this panel has been previously validated and can be found elsewhere (https://www.olink.com/products/cvd-ii-panel/). Briefly, coefficient of variation (CV) of the ACE2 assay calculated from linearized NPX values over the limit of detection, is of 8% for the intra-assay (within-run) precision, and of 11% for the inter-assay (between-run) precision. Intra-assay CV <10% and inter-assay CV <15% are considered as optimal (PMID: 12414755). The validated analytical range for ACE2 is from 15·26pg/mL to 62·5ng/mL.

ACE2 Data Quality Controls

Quality controls of biomarker data generated through the CVDII panel, including ACE2, were evaluated in four steps. First, biomarkers were pre-processed using the NPX Manager, a built-in Olink software for quality control. Participant samples were spiked with 4 internal quality controls, including 2 incubation controls, 1 extension control, and 1 amplification/detection control that monitored assay performance. 3 external controls, including an inter-plate control, a negative control, and a pool plasma sample were also included in each plate to monitor for inter- and intra-plate precision. Samples with at least one internal control that deviated more than ±0·3 NPX units from the plate median were flagged and excluded from further analyses. Overall, 649 samples were excluded with a pass quality control rate of 94·8% for this panel. An additional 611 samples (5%) were also identified to have ACE2 levels below the lower limit of quantification (ie. 15·26pg/mL for ACE2). These samples were flagged but retained in following analyses. Second, inter-plate normalization was performed to minimize technical sources of variation between plates. This was achieved by calculating the plate-specific median and overall lot median across all samples. Biomarker levels were subsequently centralized to the overall lot median by taking the difference between the measured biomarker level and the plate-specific median and adding it to the overall median.

Third, the distribution of ACE2 was scrutinized as follows. The shape of the distribution was quantified using skewness and kurtosis metrics. Next, the modality of the distribution was assessed using the Hartigan dip test.[7] Differences in the mean, median, and distribution of levels between sex, ethnic group, and reagent lot were examined. It should be noted that the participant cohort was analyzed using two different reagent lots. As such, quantile normalization was applied to each reagent lot separately and later combined to minimize potential batch effects. Olink provides data on the sensitivity of individual biomarker assays in the CVDII panel to contamination by plasma or serum hemolysate. We defined a biomarker to be hemolysis-sensitive if assay performance was

affected by 5mg/mL or less of haemolysate. ACE2 is not a haemolysis-sensitive biomarker. The mean, median, and distribution of ACE2 was then re-assessed to evaluate quality control performance.

Genetic Analysis

Genotyping on the Thermofisher Axiom Precision Medicine Research Array (r.3) was attempted for 11,683 PURE study participants consenting to research with suitable DNA quantities. 96-plex plates were scanned on the GeneTitan instrument, each including 95 PURE samples and a universal control sample shared across all study plates. Genotype calling was performed according to manufacturer's best practices using a combination of *Axiom Power Tools* and in-house scripts. Genotype calling was performed in three separate batches of approximately equal size (n1=3,778; n2=3,862; n3=3,781) grouping samples by the order in which plates were processed. As per manufacturer recommendations, samples were removed if they had low signal-to-noise contrast (Dish Quality Control < 0·82) or low quality control call rate (QCCR < 0·97).

Genotyping Quality Control

Additional quality control procedures were implemented using PLINK, R, GCTA, and KING softwares with in-house scripts. Sample-level quality control checks included assessments of sample completeness (call rate > 0·95), potential sample mix-ups (discrepancies between reported vs. genetically determined sex and/or ethnicity), genetic duplicates, and sample contamination (excess heterozygosity). Genetically determined ancestry for PURE participants was derived via principal component analysis of directly genotyped common variants alongside a reference set of ancestrally diverse samples (1000Genomes). The top four principal components were plotted to assess clustering of genetic ancestry with self-reported ancestry. Samples exhibiting non-ambiguous discrepancies between genetic and self-reported ancestry were removed. Variant-level quality control checks included assessments of variant completeness (call rate > 0·985), plate and batch effects, non-Mendelian segregation within families (Mendelian errors), Hardy Weinberg Equilibrium deviations (HWE P-value < 1x10-5), and variant frequency (minor allele frequency > 0). After sample and variant quality control procedures, 11,112 samples and 749,783 variants remained. The average genotyping call rate among passing samples was 0·996535.

Imputation Quality Control

For phasing and imputation of genotypes, the 749,783 directly genotyped variants in 11,1112 samples were uploaded to a cloud imputation server hosted by the Sanger Wellcome Trust Institute (https://imputation.sanger.ac.uk/). Genotypes were phased using EAGLE2 and then imputed using PBWT for 95,270,199 variants on the 1000Genomes and UK10K combined reference panel. Quality control of imputed genotypes included removal of variants with: (i) poor imputation quality (INFO<0·30), (ii) violations of Hardy Weinberg equilibrium (P<1x10-5), and (ii) low allele frequency (MAF<0·01). After quality control of imputed genotypes, 2 variants remained. 10,480 (94%) out of the 11,112 samples passing genotyped QC also had passing ACE2 measurements, which were used for all subsequent genetic analyses.

Mendelian Randomization Analyses

Mendelian Randomization analysis is conventionally applied to assess causal effects of circulating biomarkers on disease outcomes with the intent of identifying causal mediators of disease that may represent effective therapeutic targets ("forward Mendelian Randomization"), but extensions such as "reverse Mendelian Randomization" may allow for more specific and sensitive diagnostic biomarkers for disease.[8] We extend this framework to understanding how antecedents of cardiovascular disease may causally affect circulating ACE2 levels. We conducted Mendelian Randomization analyses for the subset of modifiable clinical risk factors associated with

ACE2 levels in the cross-sectional PURE analyses, specifically: body mass index, systolic blood pressure, diabetes status, smoking status.

Genetic Variant Selection

Common genetic variants (minor allele frequency > 0·01) that were strongly associated with clinical risk factors (P<5x10-8) in large European genome-wide association studies were selected as proxies for clinical risk factors. Genetic variants associated with body mass index were derived from Yengo *et al.* (N~700,000)[9]; systolic blood pressure were derived within the UKBiobank study (N~400,000)[10], diabetes status derived from Mahajan *et al.* (N=1,232,901)[11] and smoking initiation status were derived from Liu *et al* (N=1,232,091).[12] Beta coefficients (or log(odds) for discrete risk factors) and standard errors were extracted for each variant.

Harmonization of External Genetic Datasets with PURE

The impact of the pre-selected variants on circulating ACE2 levels was ascertained through a genome-wide association study within a subset of 8699 PURE participants. To mitigate the confounding effect of ethnicity on genetic variation, genetic analyses were restricted to PURE participants of self-reported Latin (n=4058), European (n=3372), or Persian (n=1269) ancestry. African (n=659), South Asian (n=604), East Asian (n=314), and Arab (n=204) participants were excluded from the present analyses because harmonization of PURE with external genetic datasets, which are primarily comprised of European individuals presently, necessitated exclusion of non-European groups. Self-identified Latin and Persian individuals were included in the present genetic analyses because these groups genetically overlap with Europeans. Furthermore, to gauge the transferability of genetically predicted risk factors across these ethnic groups in PURE, we demonstrate similarly strong predictions across European, Latin, and Persian subgroups with a polygenic risk score derived using European weights (Supplementary Table 5).

Pre-selected genetic variants were cross-referenced against the PURE variant set, and matching variants shared between PURE and the external dataset were retained. To account for correlation between genetic variants, variants were first prioritized based on statistical significance and then proximal SNPs correlated with the index variant at (r2 > 0·01 in 1000Genomes Europeans) were iteratively removed to generate an independent set of genetic variants. This process was repeated separately for each clinical risk factor to maximize retention of suitable genetic instruments.

For the remaining independent set of genetic variants present in both external and internal genetic datasets, we conducted genetic association testing within PURE to derive their corresponding effects on plasma ACE2 levels (i.e. outcome betas). The GCTA mixed linear model association with leave-one-chromosome-out method was used assuming an additive model. Covariates included age, sex, recruitment centre, season in which the blood specimen was collected, blood specimen storage time, OLINK processing batch, genotype calling batch, and 10 intra-ethnic genetic principal components. GWAS were conducted separately for European (n=3372), Latin (n=4058), and Persian (n=1269) participants and then combined via inverse-variance-weighted fixed effects meta-analysis with the METAL software.

Mendelian Randomization Framework

Two-sample Mendelian Randomization analyses were conducted wherein the effects of genetic variants on the exposure (i.e. clinical risk factors from external datasets) were obtained from independent datasets from outcome effects (i.e. circulating ACE2 levels within PURE). All MR testing was conducted in R using the TwoSampleMR package (v.0·5·2). Several complementary MR methods were employed including: Inverse Variance Weighted (IVW), Median Weighted, and MR-EGGER regression methods. The IVW method has the greatest statistical power but also makes the most assumptions. Accordingly, MR effect estimates were reported from the IVW model in the

main manuscript only when there were no signs of pleiotropic confounding (horizontal, directional or idiosyncratic). Potential for pleiotropic confounding, whereby an instrument influences the outcome via a mechanism beyond the exposure's mediating effect, was detected using the Cochran's Q and MR-PRESSO tests for global heterogeneity and the EGGER intercept test for directional pleiotropy. The impact of outlying variants (idiosyncratic pleiotropy) on distorting the overall causal effect estimate was evaluated by inspection of leave-one-out-analyses plots and MR-PRESSO. MR-PRESSO testing was run using 1000 simulations; outliers detected by MR-PRESSO were removed and the causal effects were re-estimated without outliers using the aforementioned MR methods. If significant global heterogeneity persisted after outlier removal, the median weighted MR estimate was reported; if directional pleiotropy persisted, the EGGER-regression MR estimate was reported. To facilitate comparison with epidemiological effect estimates, MR effect estimates were scaled to the same units when possible (i.e. body mass index: per 5 kg/m2 increase; systolic blood pressure: per 10 mm Hg increase; diabetes and smoking status: mean difference between cases and controls).

Impact of Common Antihypertensive Drugs on Circulating ACE2

We also applied the same Mendelian Randomization framework to assess the effects of common blood pressure-lowering medications on circulating ACE2 levels. Genetic variants approximating the effects of antihypertensive medication were selected according to previously derived instruments.[13,14] Specifically, genetic variants (i) were located proximally to the genes encoding the therapeutic protein targets and (ii) demonstrated association with circulating protein levels (ACE) and/or blood pressure (ACE/CCB/BB). MR effect estimates were scaled to a 10 mmHg reduction in systolic blood pressure for CCB and BB; for ACE inhibitors, results were expressed per 1 SD decrease in circulating ACE levels.

Sensitivity Analyses

We conducted a series of sensitivity analyses to assess the robustness of MR results. First, to more sensitively identify potential pleiotropic SNPs, we searched phenoscanner for genome-wide significant variants laying outside the intended causal pathway and then repeated MR analyses excluding such variants. Second, we excluded cases with cardiovascular disease or diabetes (N=5234) at baseline and reperformed association testing of genetic instruments on circulating ACE2 levels in this disease-free subcohort (N=3465). Third, we conducted MR analyses using a more stringent pairwise r2 threshold of 0·001 as opposed to 0·01. Fourth, we conducted MR analyses using ethnicity-specific effects on ACE2 levels to evaluate consistency of MR effect estimates across ethnic groups or conversely population-specific effects.

Supplementary Appendix 2: Additional Results -- Supplementary Tables

Supplementary Table 1: Subgroup Analysis to Assess Heterogeneity of Effect of ACE2 on Cardiovascular Outcomes*

Sub group	Number of events in the sub-cohort	Number of events outside the sub-cohort	HR (95% CI)*	P for interaction
	N	N		
Diabetes	.	.		
Yes	65	335	1·10 (0·94 - 1·30)	0·068
No	198	1188	1·25 (1·15 - 1·36)	
Age	.	.		
Age <55 years	88	503	1·13 (1·02 - 1·26)	0·080
Age ≥ 55	175	1020	1·28 (1·16 - 1·41)	
Obese	.	.		
Yes (BMI ≥ 30)	82	475	1·21 (1·07 - 1·37)	0·659
No (BMI< 30)	181	1048	1·24 (1·13 - 1·35)	
Genetic ancestry	.	.		
African	28	57	1·36 (1·08 - 1·72)	0·703
Arab	8	27	0·89 (0·45 - 1·76)	
East Asian	5	63	1·72 (1·17 - 2·53)	
Latin	109	658	1·23 (1·11 - 1·37)	
Persian	27	152	1·18 (0·96 - 1·46)	
South Asian	36	128	1·21 (0·95 - 1·54)	
European	50	438	1·14 (0·97 - 1·34)	
Sex	.	.		
Female	115	649	1·21 (1·09 - 1·34)	0·654
Male	148	874	1·23 (1·11 - 1·37)	

*An exploratory analysis was carried out examining possible heterogenous effect across subgroups as it relates to the cardiovascular outcomes (defined by fatal and non-fatal myocardial infarction, stroke, heart failure, or another fatal cardiovascular disease). In order to account for multiple hypothesis testing, a subgroup required a p-value of 0·05/5 = 0·01 to account for the 5 interaction tests being carried out. Models are adjusted for age, sex, geographic ancestry, smoking, diabetes, body mass index, systolic blood pressure, and non-HDL cholesterol excluding the subgroup variable from the model.

Supplementary Table 2: Comparisons of the associations of ACE 2 and other common continuous risk factors using minimally Adjusted Standardized Coefficients (per SD)

Outcome	HR (95% CI) Death*	HR (95% CI) Cardiovascular Death*	HR (95% CI) Non-CVD death*	HR (95% CI) Myocardial Infarction*	HR (95% CI) Stroke*	HR (95% CI) Heart Failure*	HR (95% CI) Diabetes*
Plasma ACE2	1·41 (1·34 - 1·49)	1·52 (1·38 - 1·68)	1·38 (1·30 - 1·46)	1·37 (1·27 - 1·48)	1·32 (1·21 - 1·45)	1·43 (1·25 - 1·65)	1·68 (1·59 - 1·77)
Systolic Blood Pressure	1·11 (1·06 - 1·17)	1·32 (1·22 - 1·43)	1·03 (0·98 - 1·09)	1·26 (1·18 - 1·35)	1·60 (1·48 - 1·72)	1·48 (1·31 - 1·67)	1·44 (1·36 - 1·51)
non HDL Cholesterol	0·98 (0·93 - 1·03)	1·11 (1·01 - 1·21)	0·93 (0·88 - 0·99)	1·25 (1·17 - 1·35)	1·16 (1·07 - 1·26)	1·03 (0·90 - 1·18)	1·27 (1·21 - 1·34)
BMI	1·02 (0·97 - 1·07)	1·20 (1·09 - 1·32)	0·95 (0·90 - 1·01)	1·30 (1·21 - 1·40)	1·15 (1·05 - 1·25)	1·37 (1·20 - 1·55)	2·08 (1·99 - 2·18)

*Coefficients are hazard ratios per 1 standard deviation increase. Each coefficient presented is minimally adjusted (for age, sex, ancestry). For example, the HR presented for ACE2 is adjusted for age, sex, and ancestry only. The HR presented for blood pressure is adjusted for age, sex, and ancestry only. Continuous risk factors were selected in order to place all variables on a common interpretable scale (per 1 SD) increase. In all mortality related outcomes, ACE2 had the strongest association with the outcome. These results align with the order of our variable ranking plots presented in the main paper.

Supplementary Table 3: Split Sample Estimation of ACE2 Relationship With Cardiovascular Outcomes and Mortality

	Training sample				Validation sample			
Outcome	**Number of subjects in the sub-cohort**	**Number of events in the sub-cohort**	**Number of events outside the sub-cohort**	**HR (95% CI) per 1 SD Fully adjusted**	**Number of subjects in the sub-cohort**	**Number of events in the sub-cohort**	**Number of events outside the sub-cohort**	**HR (95% CI) per 1 SD Fully adjusted**
	N	**N**	**N**		**N**	**N**	**N**	
All cause mortality	2542	136	856	1·39 (1·29 - 1·49)	2542	152	841	1·32 (1·22 - 1·42)
CVD death	2542	41	234	1·37 (1·19 - 1·57)	2542	46	240	1·45 (1·26 - 1·66)
Non CVD death	2542	95	622	1·40 (1·28 - 1·52)	2542	106	601	1·28 (1·17 - 1·40)
MI	2542	73	376	1·26 (1·13 - 1·41)	2542	52	381	1·18 (1·05 - 1·32)
Stroke	2542	39	307	1·16 (1·02 - 1·32)	2542	56	261	1·26 (1·10 - 1·44)
Heart failure	2542	16	115	1·35 (1·11 - 1·65)	2542	22	111	1·17 (0·95 - 1·44)
Diabetes	2301	126	731	1·47 (1·36 - 1·59)	2283	119	739	1·41 (1·30 - 1·53)

We performed a random split estimation of the independent non-overlapping samples. We find that our associations remain consistent and significant across the data splits, with the exception of heart failure whose effect loses statistical significance likely due to the smaller number of heart failure events in the study.

Supplementary Table 4: Impact of Cardiovascular Disease Adjustment on Risk Factor and Demographic Factors on ACE2 Concentration

Variable	Beta in Model Without CVD Adjustment (95% CI)	Beta in Model With Cardiovascular Disease History Adjustment
Sex (Baseline: Female)	0.58 (0.54, 0.61)	0.58 (0.54, 0.61)
Ancestry (Baseline: European)		
South Asian	-0.12 (-0.21, -0.04)	-0.12 (-0.20, -0.04)
Arab	0.06 (-0.06, 0.18)	0.06 (-0.06, 0.18)
Persian	0.12 (0.06, 0.17)	0.11 (0.06, 0.17)
Latin	0.29 (0.25, 0.33)	0.29 (0.25, 0.33)
African	0.46 (0.38, 0.54)	0.46 (0.39, 0.54)
East Asian	0.57 (0.46, 0.68)	0.57 (0.46, 0.68)
BMI, per 5 kg/m^2	0.12 (0.10, 0.14)	0.12 (0.10, 0.14)
Diabetes (Baseline: None)	0.29 (0.23, 0.34)	0.28 (0.23, 0.34)
Age, per decade	0.10 (0.08, 0.12)	0.10 (0.08, 0.12)
SBP, per 10 mmHg	0.039 (0.03, 0.05)	0.038 (0.03, 0.05)
Smoking (Baseline: Ever Smoker)	-0.11 (-0.15, -0.08)	-0.11 (-0.15, -0.07)
LDL-c, per mmol/L	0.027 (0.01, 0.05)	0.026 (0.01, 0.05)

We performed the cross-sectional multivariable regression analysis of ACE2 determinants and found no change in effects after controlling for cardiovascular disease status.

Supplementary Table 5: Impact of Variable Ranking Metric on ACE2 Rank

	Total Mortality	Cardiovascular Death	Non-Cardiovascular Death	Heart Failure	Myocardial Infarction	Stroke
Likelihood Ratio based rank (Chi-sq statistic)	1 (133)	1 (45)	1 (90)	3 (11)	3 (25)	3 (15)
Lin Ying estimate based rank (Chi-sq statistic)	1 (68)	1 (33)	1 (50)	3 (8)	3 (18)	3 (12)

We performed additional analyses for ranking ACE2 on the basis of the absolute magnitude of the Chi-square statistic. Different ranking metrics yielded the same results as it related to the relative rank of ACE2 compared to clinical risk factors.

456
457 **Supplementary Table 6:** Results of Mendelian Randomization Analyses of Clinical Risk Factors vs. circulating
458 ACE2 levels.
459

Trait	MR Method	# Variants	Beta	95% CI	MR P-value	Cochran's Q Heterogeneity P-value	MR-PRESSO Heterogeneity P-value	Egger Intercept P-value
Body Mass Index	**Inverse variance weighted**	**631**	**0·18**	**0·09-0·26**	**3·92E-05**	**0·19**	**0·18**	**0·31**
Body Mass Index	MR Egger	631	0·30	0·26-0·33	0·02	0·19	0·18	0·31
Body Mass Index	Weighted median	631	0·28	0·14-0·43	1·49E-04	NA	0·18	0·31
Systolic Blood Pressure	Inverse variance weighted	395	0·04	-0·01 to 0·09	0·09	0·004	0·002	0·68
Systolic Blood Pressure	MR Egger	395	0·01	-0·13 to 0·15	0·86	0·004	0·002	0·68
Systolic Blood Pressure	**Weighted median**	**395**	**0·04**	**-0·03 to 0·11**	**0·29**	**NA**	**0·002**	**0·68**
LDL Cholesterol*	Inverse variance weighted	93	-0·01	-0·08 to 0·06	0·78	$4·59x10^{-4}$	0·001	0·40
LDL Cholesterol*	MR Egger	93	-0·05	-0·16 to 0·07	0·41	$4·46x10^{-4}$	0·001	0·40
LDL Cholesterol*	**Weighted median**	**93**	**-0·02**	**-0·13 to 0·08**	**0·66**	**NA**	**0·001**	**0·40**

Smoking Initiation	**Inverse variance weighted**	**103**	**0·37**	**-0·06 to 0·80**	**0·09**	**0·20**	**0·21**	**0·38**
Smoking Initiation	MR Egger	103	1·33	-0·87 to 3·54	0·24	0·20	0·21	0·38
Smoking Initiation	Weighted median	103	0·47	-0·13 to 1·07	0·13	NA	0·21	0·38
Type 2 Diabetes**	**Inverse variance weighted**	**247**	**0·29**	**0·05-0·52**	**0·02**	**0·14**	**0·14**	**0·16**
Type 2 Diabetes**	MR Egger	247	-0·05	-0·57 to 0·47	0·85	0·16	0·14	0·16
Type 2 Diabetes**	Weighted median	247	0·12	-0·24 to 0·48	0·51	NA	0·14	0·16

Bolded font indicates the MR method selected for each trait considering the potential presence of heterogeneity or directional pleiotropy.

* In the original MR analysis of LDL cholesterol, MR-PRESSO detected four SNP outliers (rs2642438, rs1169288, rs492602, rs2954029), which led to significant heterogeneity (Cochran Q and MR-PRESSO P-values < 0·001). While attenuated after removal, significant heterogeneity persisted (Cochran Q and MR-PRESSO P-values < 0·001).
**In the original MR analysis of T2D, MR-PRESSO detected two SNP outliers (rs61946386; rs56348580) which led to significant heterogeneity (Cochran Q and MR-PRESSO P-values < 0·001). Information presented in the table reflects MR results after outlier removal.

Supplementary Table 7: Results of Mendelian Randomization Analyses of Antihypertensive Agents vs. circulating ACE2 levels.

Trait	MR Method	# Variants	Beta	95% LCI	95% UCI	MR P-value	Cochran's Q Heterogeneity P-value	MR-PRESSO Heterogeneity P-value	Egger Intercept P-value
ACE inhibitor	**Inverse variance weighted**	**16**	**-0·01**	**-0·04**	**0·02**	**0·56**	**0·69**	**0·76**	**0·78**
ACE inhibitor	MR Egger	16	-0·02	-0·09	0·05	0·61	0·62	0·76	0·78
ACE inhibitor	Weighted median	16	-0·004	-0·05	0·04	0·85	NA	0·76	0·78
Beta Blockers	**Inverse variance weighted**	**6**	**0·31**	**-0·17**	**0·80**	**0·21**	**0·33**	**0·24**	**0·63**
Beta Blockers	MR Egger	6	-0·12	-1·83	1·58	0·89	0·25	0·24	0·63
Beta Blockers	Weighted median	6	0·36	-0·24	0·96	0·24	NA	0·24	0·63
Calcium Channel Blockers	**Inverse variance weighted**	**23**	**0·25**	**-0·04**	**0·54**	**0·09**	**0·13**	**0·13**	**0·14**
Calcium Channel Blockers	MR Egger	23	0·73	0·06	1·40	0·05	0·19	0·13	0·14
Calcium Channel Blockers	Weighted median	23	0·29	-0·12	0·70	0·16	NA	0·13	0·14

Bolded font indicates the MR method selected for each trait considering the potential presence of heterogeneity or directional pleiotropy. Results for the ACE inhibitor analysis are expressed per 1 SD decrease in circulating ACE2 levels. Results are expressed per 10 mmHg decrease in systolic blood pressure except for ACE inhibitor analysis which is expressed per 1 SD decrease in serum ACE concentration.

Supplementary Table 8: Association of European polygenic scores in PURE participants of European (n=3372), Latin (n=4058), and Persian (n=1269) ethnicity

EXPOSURE	# SNPs	TRAIT	SUBGROUP	EFFECT (Beta or Odds Ratio)	95% CI	P-value	% Variance Explained
Systolic Blood Pressure PRS	395	Systolic Blood Pressure* (mmHg)	EUROPEAN	3·25	2·61-3·89	5·49E-23	0·03
			LATIN	4·53	3·77-5·29	3·19E-31	0·03
			PERSIAN	3·27	2·03-4·50	2·76E-07	0·02
Body Mass Index PRS	631	Body Mass Index (kg/m2)	EUROPEAN	1·15	0·97-1·32	6·56E-37	0·05
			LATIN	0·93	0·75-1·11	1·66E-24	0·03
			PERSIAN	1·07	0·78-1·35	7·54E-13	0·05
Smoking Initiation PRS	103	Smoker Status (Ever vs. Never)	EUROPEAN	1·21	1·13-1·30	7·39E-08	0·01**
			LATIN	1·04	0·97-1·11	0·27	0·0005**
			PERSIAN	1·44	1·18-1·76	0·000334	0·03**
Type 2 Diabetes PRS	247	Diabetes Status	EUROPEAN	1·66	1·51-1·83	3·78E-24	0·06**
			LATIN	1·57	1·45-1·70	1·22E-28	0·05**
			PERSIAN	1·48	1·29-1·69	8·89E-09	0·05**

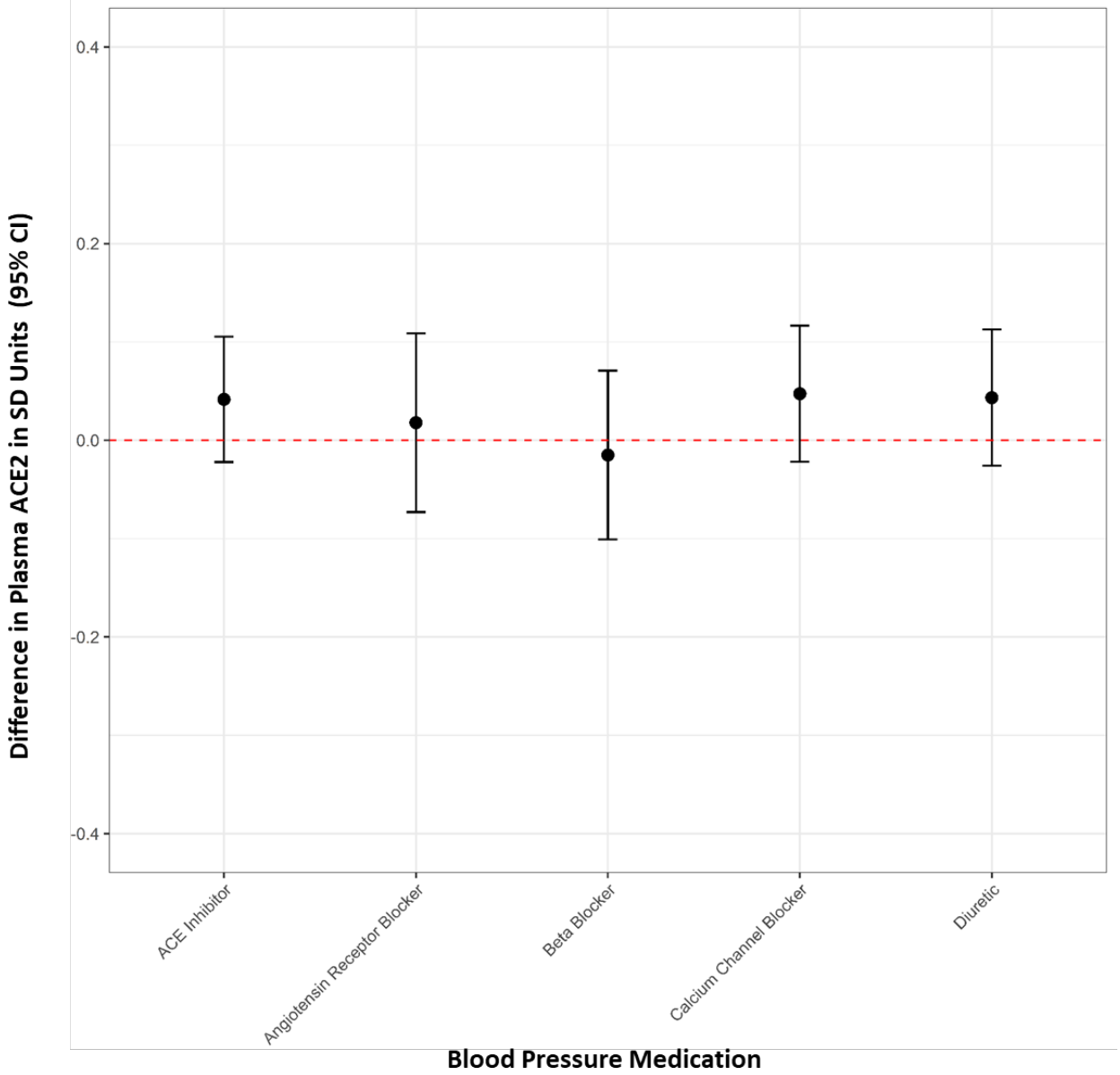
Effects are expressed as change in the trait per 1 standard deviation increase in the exposure PRS.

*Systolic Blood Pressure adjusted for medication use (+15mmHg) as per Evangelou *et al*[5].

** Calculated using Nagelerke's R2

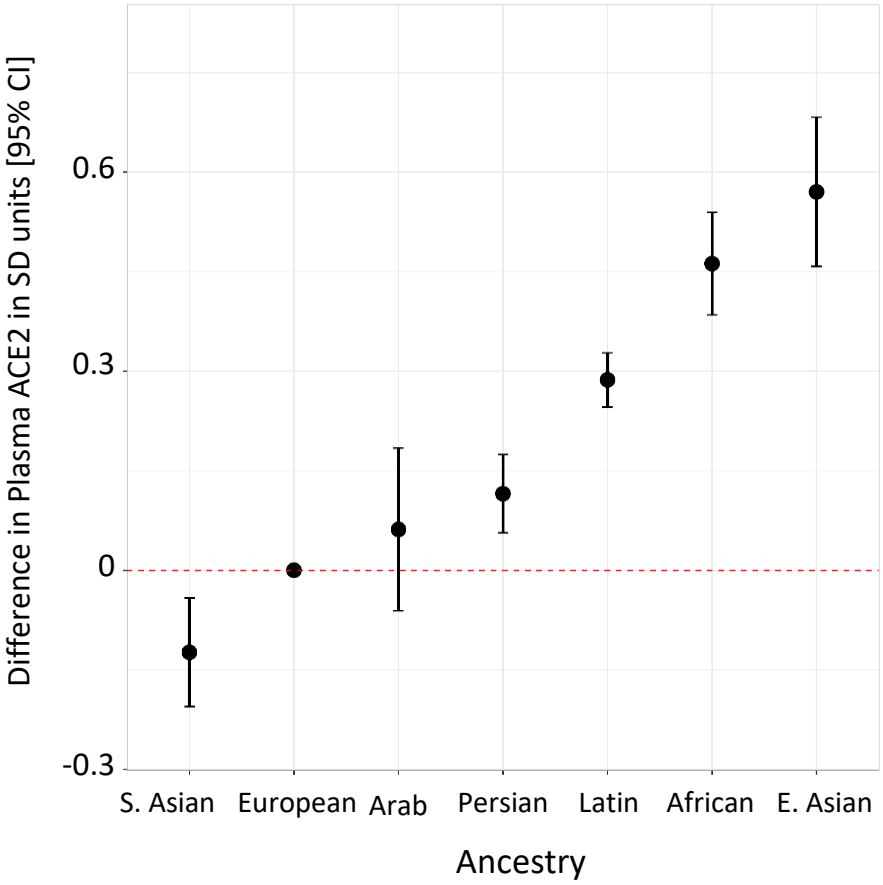
Supplementary Appendix 3: Additional Results – Supplementary Figures

Supplementary Figure 1: Difference in plasma ACE2 by blood pressure medication (Cross-sectional phenotypic analysis)



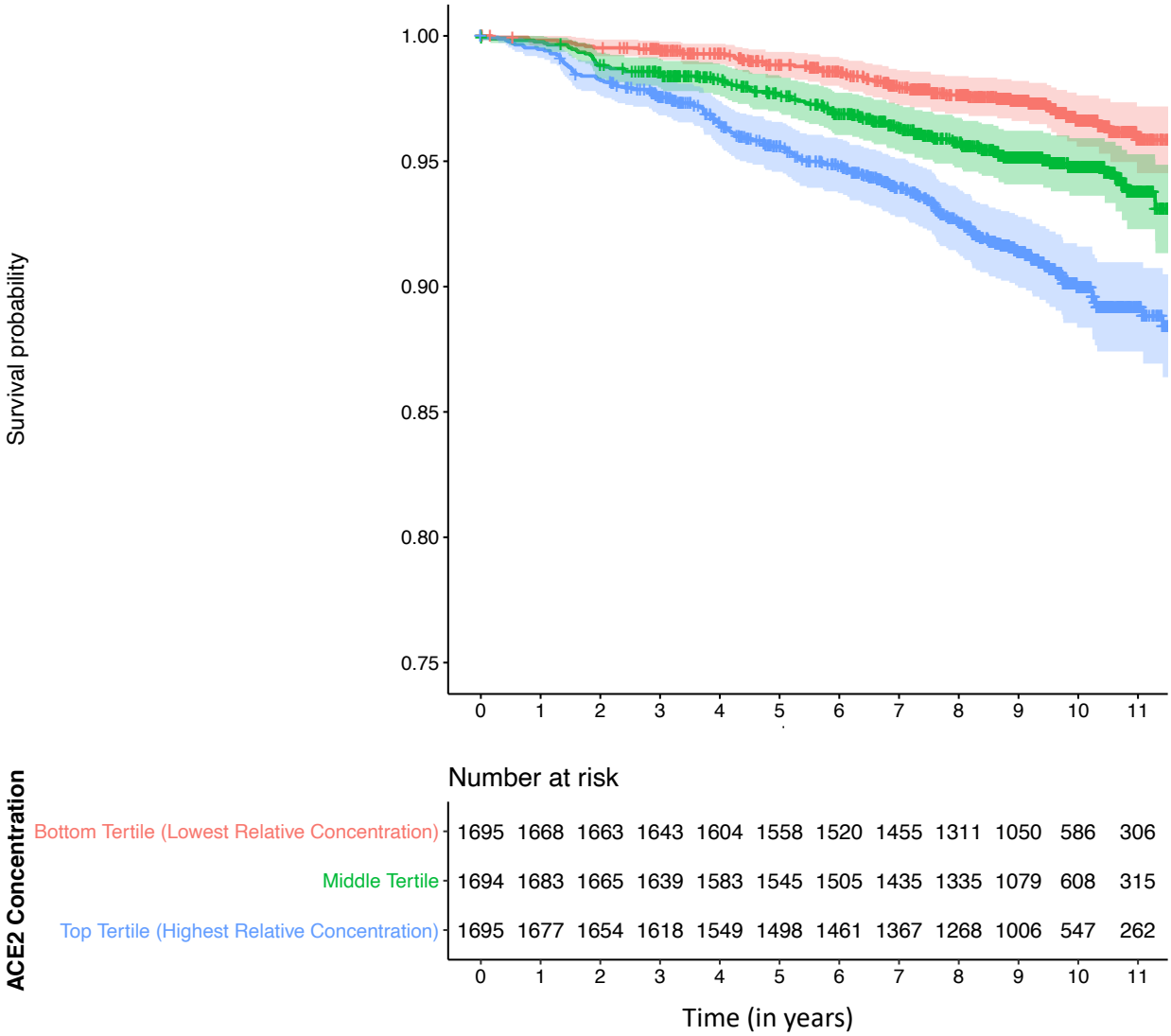
Supplementary Figure 1: The plot presents the adjusted coefficients (mutually adjusted for each other as well as age, sex, blood pressure, diabetes, smoking, diabetes, ancestry, and LDL cholesterol) for blood pressure medication on plasma ACE2 levels.

Supplementary Figure 2: Difference in plasma ACE2 concentration by ancestry



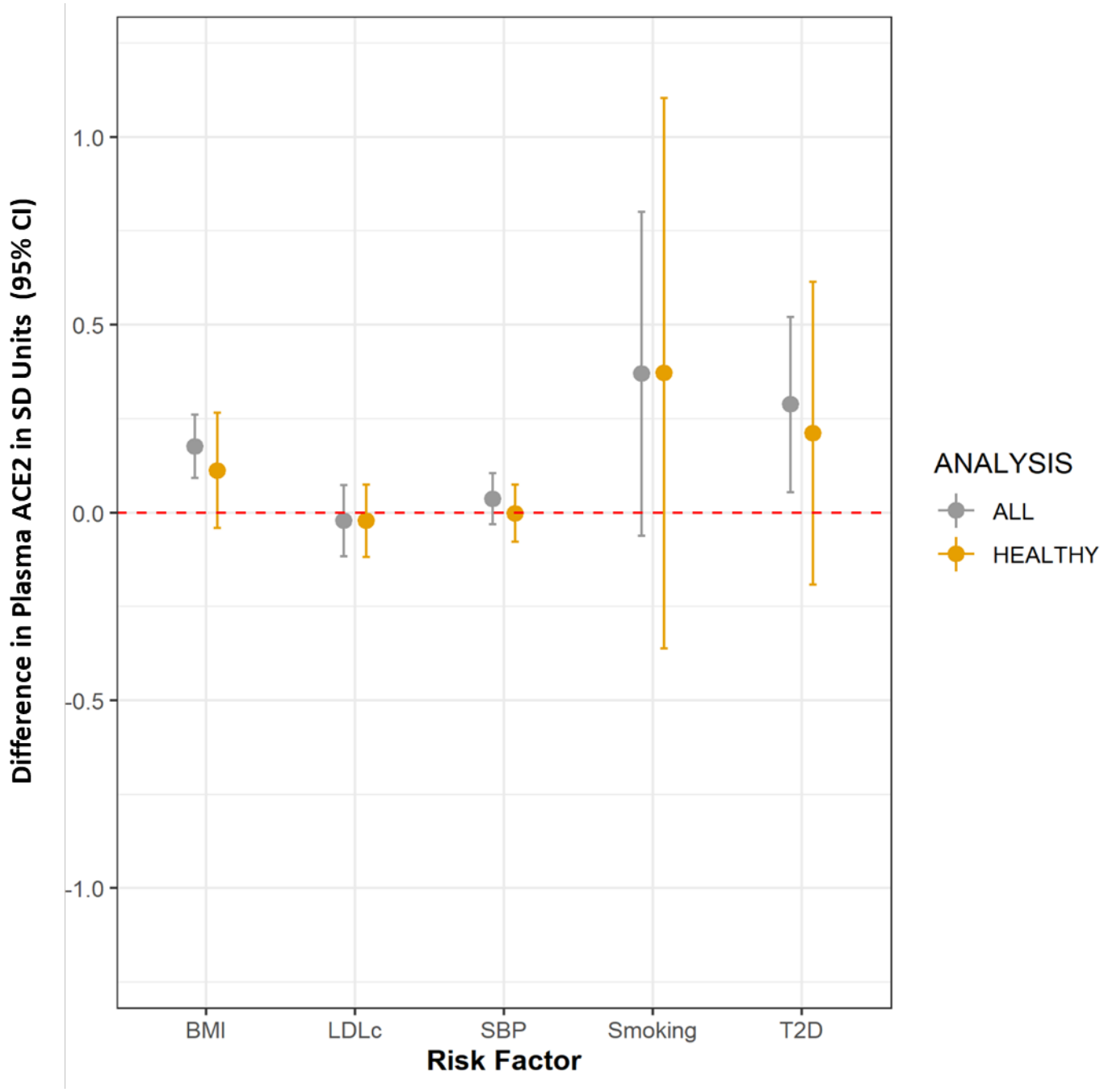
Supplemental Figure 2: Using European levels as the reference group, these adjusted coefficient plots show how each ancestral group varies with respect to ACE2 concentration. Estimates are adjusted for age, sex, blood pressure, LDL-c, BMI, smoking, and diabetes.

Supplementary Figure 3: ACE2 Relationship with Overall Mortality

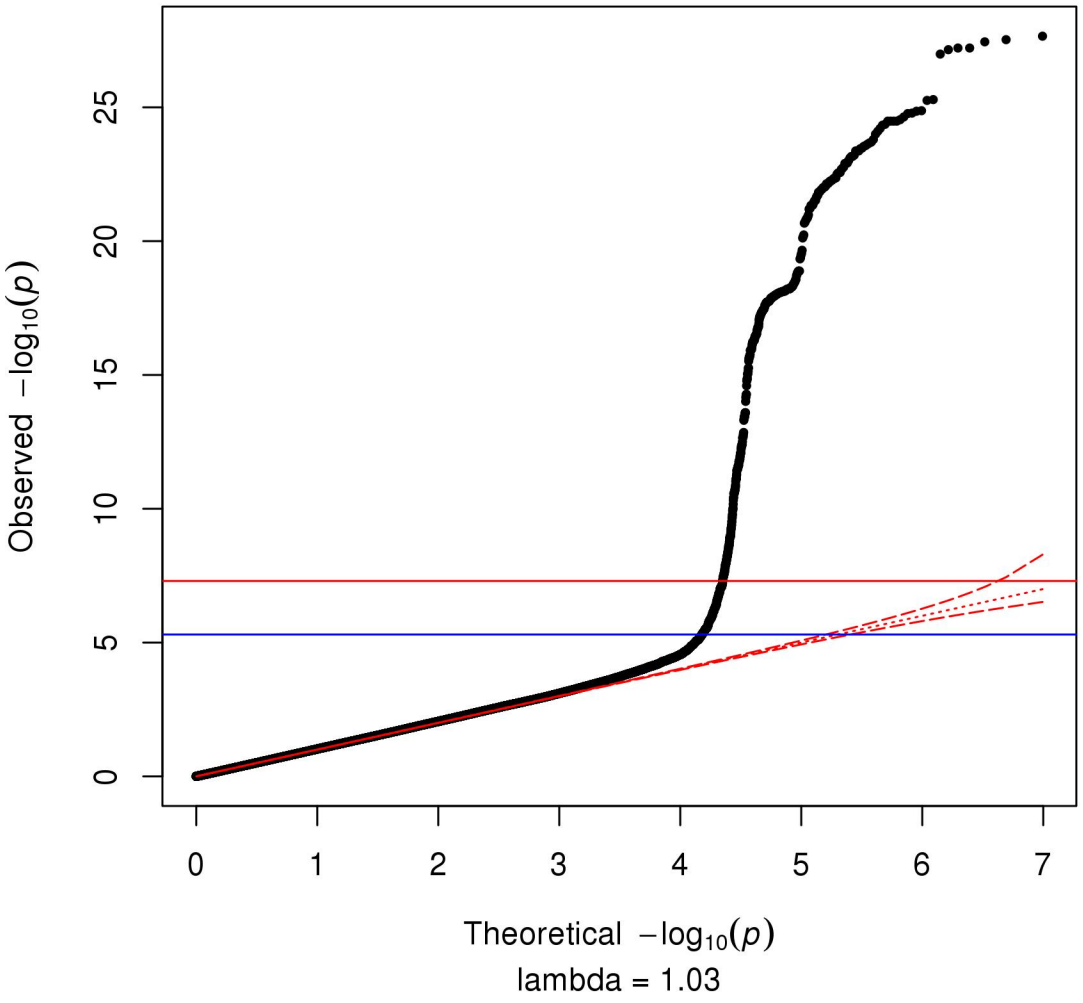


Supplementary Figure 3: Kaplan-Meier of total mortality assessed in the randomly sampled sub-cohort. ACE2 concentration was split in three tertiles. The dashed lines indicate participants who were censored. The red curve corresponds to the lowest plasma ACE2 concentration level (the lowest tertile, n=1695; 50 death events), the green curve corresponds to intermediate levels of ACE2 concentration (the middle tertile, n=1694; 86 death events), and the blue curve corresponds to the highest plasma ACE2 concentration (the highest tertile, n=1695; 152 death events). Individuals with higher concentrations of ACE2 had a higher rate of death relative to those with lower concentrations of ACE2 (log rank p-value<0·0001).

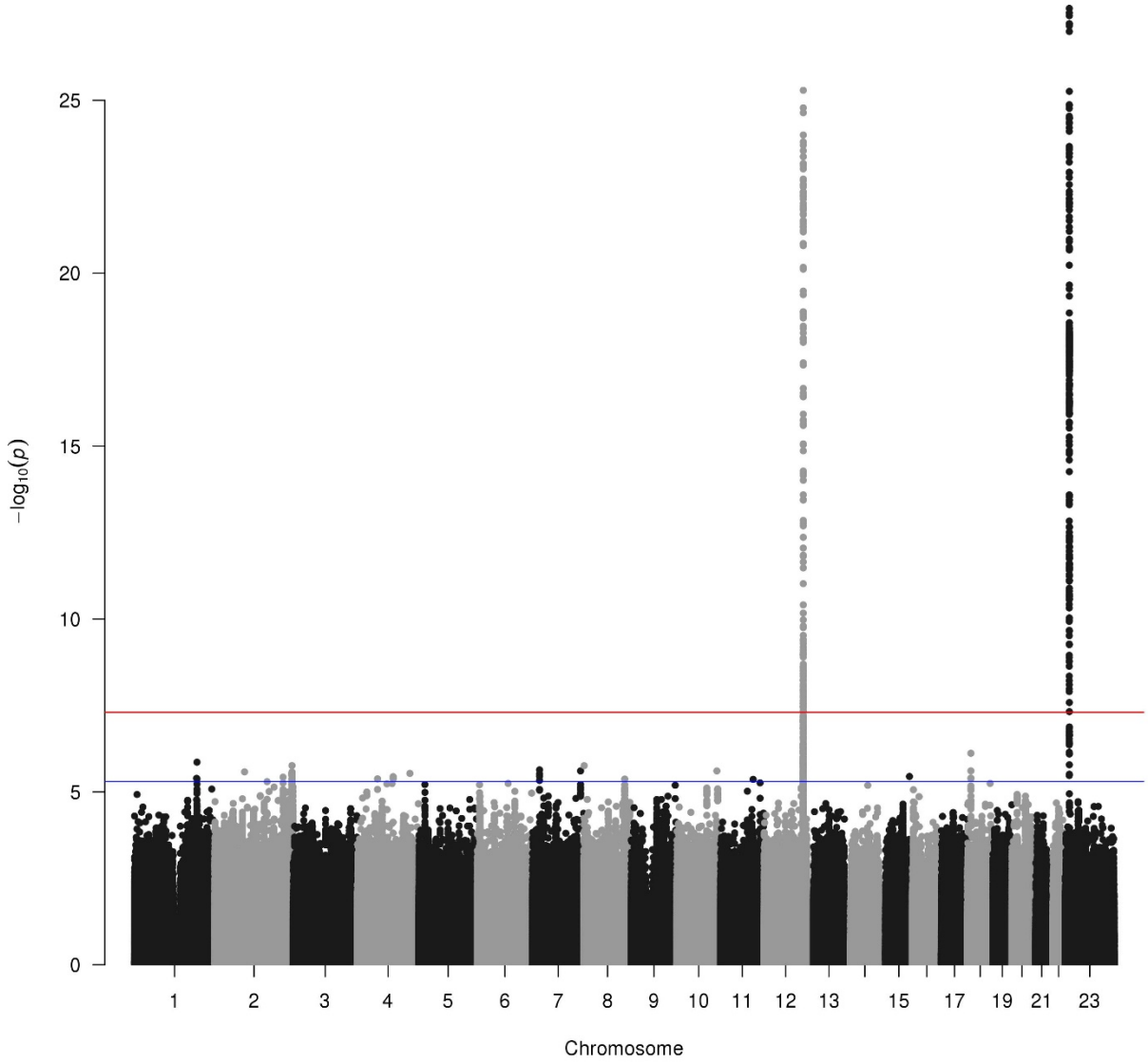
Supplementary Figure 4: Comparison of MR estimates for clinical risk factors vs. ACE2 levels using the full PURE dataset (n=8699) vs. a healthy subset (members of the random subcohort with no history of CVD or diabetes) (n=3465)



Supplementary Figure 5. Quantile-Quantile Plot for the GWAS Meta-analysis of ACE2.

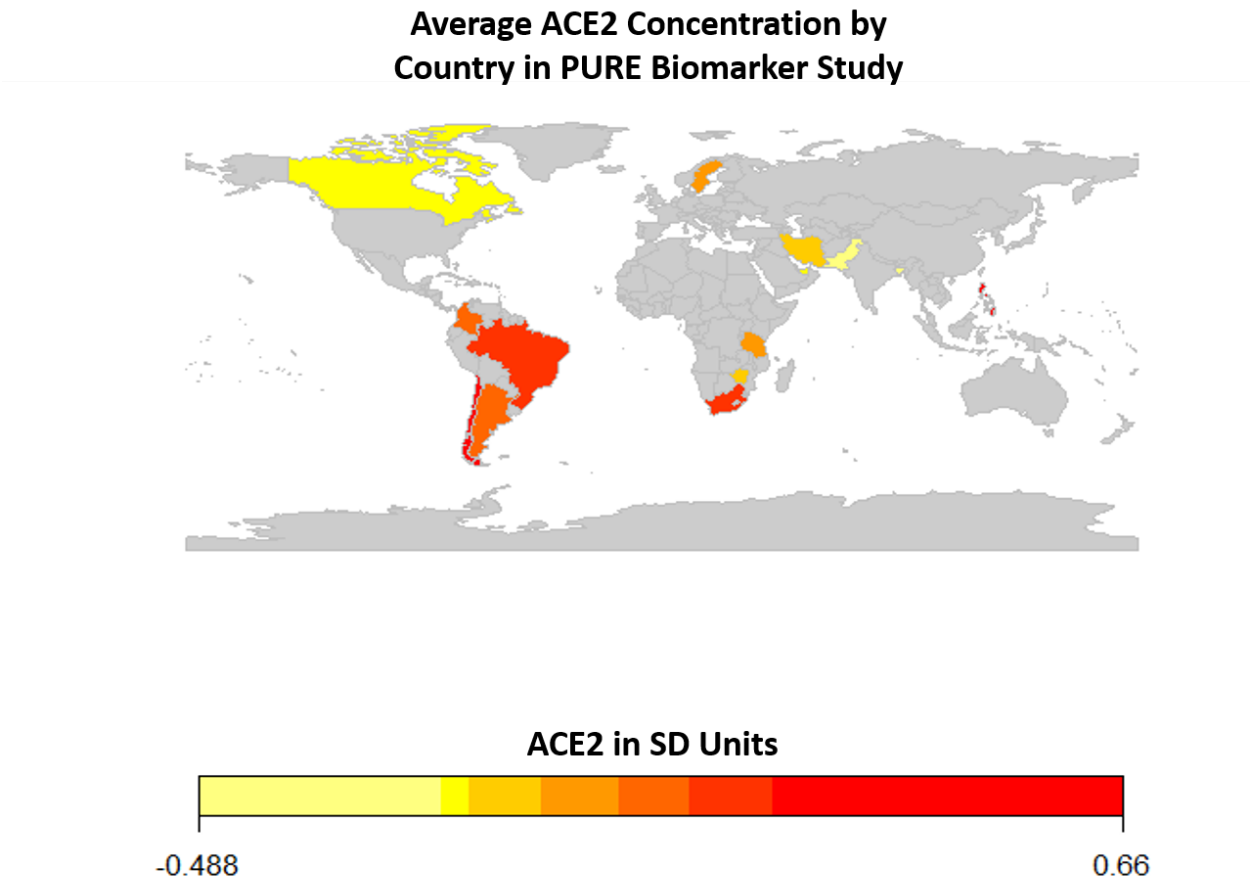


Supplementary Figure 6: Manhattan Plot for the GWAS Meta-analysis of ACE2.

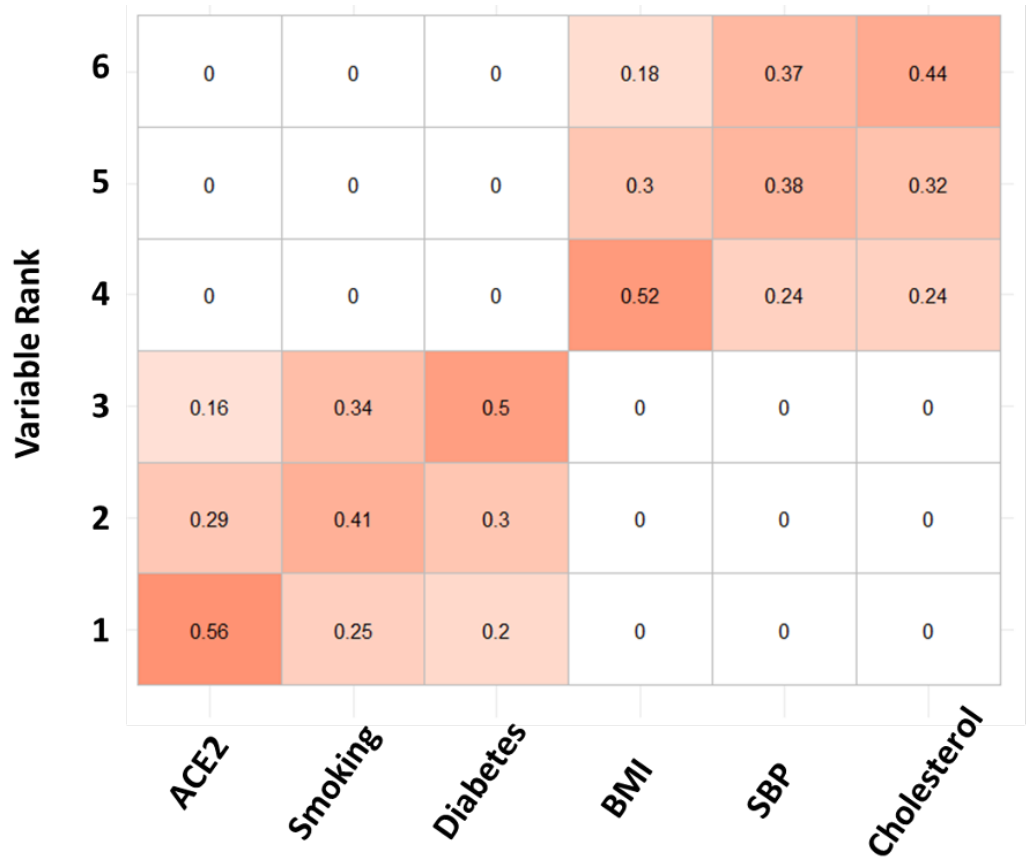


Supplementary Figure 6: Two genome-wide significant ($P<5x10^{-8}$) loci were detected nearby HNF1A on chromosome 12 (rs2464190; $P=5{\cdot}1x10^{-26}$) and ACE2 on chromosome X (rs5936022; $P=2{\cdot}2x10^{-28}$).

Supplementary Figure 7: Map of Countries Included In PURE Biomarker Study and Average ACE2 Levels of Participants in Each Country



Supplementary Figure 8: Bootstrap Validation of Variable Ranks



Supplementary Figure 8: We perform a variable ranking procedure (on the basis of a likelihood ratio chi-square) in 1000 bootstrap generated datasets. The number in each box corresponds to the proportion of times (out of 1000) that particular variable ended up in that rank. As it relates to ACE2, ACE2 was the strongest relative predictor in 555 of the bootstrapped datasets (0·555 was rounded up to 0·56), second strongest in 288, and third strongest in 157 datasets. However, this provides sufficient reason for us to see the importance of those top 3 variables (ACE2, smoking, and diabetes) relative to the bottom 3 (BMI, SBP, and Cholesterol) as none of the bottom 3 variables emerge as a top variable in any of the 1000 datasets. This shows a clear hierarchy of importance as it relates to overall mortality in our group of patients and that ACE2 has a strong relationship with mortality relative to commonly measured risk factors. These ranking distributions were consistent when using the Wald chi square statistics as the basis of ranking (for both analyses using Self-Prentice and Lin Ying derived Wald statistics).

551

552 Works Cited

553 1 Teo K, Chow CK, Vaz M, Rangarajan S, Yusuf S. The Prospective Urban Rural Epidemiology (PURE)
554 study: Examining the impact of societal influences on chronic noncommunicable diseases in low-,
555 middle-, and high-income countries. *American Heart Journal* 2009; **158**: 1-7.e1.

556 2 Yusuf S, Hawken S, Ôunpuu S, *et al.* Effect of potentially modifiable risk factors associated with
557 myocardial infarction in 52 countries (the INTERHEART study): case-control study. *The Lancet* 2004;
558 **364**: 937–52.

559 3 O'Donnell MJ, Xavier D, Liu L, *et al.* Risk factors for ischaemic and intracerebral haemorrhagic stroke
560 in 22 countries (the INTERSTROKE study): a case-control study. *The Lancet* 2010; **376**: 112–23.

561 4 Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of
562 quantitative traits: antihypertensive therapy and systolic blood pressure. *Statistics in Medicine* 2005;
563 **24**: 2911–35.

564 5 the Million Veteran Program, Evangelou E, Warren HR, *et al.* Genetic analysis of over 1 million people
565 identifies 535 new loci associated with blood pressure traits. *Nature Genetics* 2018; **50**: 1412–25.

566 6 Charge Diabetes Working Group, The EPIC-InterAct Consortium, EPIC-CVD Consortium, *et al.*
567 Exome-wide association study of plasma lipids in >300,000 individuals. *Nature Genetics* 2017; **49**:
568 1758–66.

569 7 Hartigan JA, Hartigan PM. The Dip Test of Unimodality. *The Annals of Statistics* 1985; **13**: 70–84.

570 8 Mohammadi-Shemirani P, Sjaarda J, Gerstein HC, *et al.* A Mendelian Randomization-Based Approach
571 to Identify Early and Sensitive Diagnostic Biomarkers of Disease. *Clin Chem* 2019; **65**: 427–36.

572 9 Yengo L, Sidorenko J, Kemper KE, *et al.* Meta-analysis of genome-wide association studies for height
573 and body mass index in ~700000 individuals of European ancestry. *Hum Mol Genet* 2018; **27**: 3641–9.

574 10 Sudlow C, Gallacher J, Allen N, *et al.* UK Biobank: An Open Access Resource for Identifying the
575 Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLoS Med* 2015; **12**.
576 DOI:10.1371/journal.pmed.1001779.

577 11 Mahajan A, Wessel J, Willems SM, *et al.* Refining the accuracy of validated target identification
578 through coding variant fine-mapping in type 2 diabetes. *Nat Genet* 2018; **50**: 559–71.

579 12 Liu M, Jiang Y, Wedow R, *et al.* Association studies of up to 1.2 million individuals yield new insights
580 into the genetic etiology of tobacco and alcohol use. *Nat Genet* 2019; **51**: 237–44.

581 13 Pigeyre M, Sjaarda J, Chong M, *et al.* ACE and Type 2 Diabetes Risk: A Mendelian Randomization
582 Study. *Diabetes Care* 2020; **43**: 835–42.

583 14 Gill D, Georgakis MK, Koskeridis F, *et al.* Use of Genetic Variants Related to Antihypertensive Drugs
584 to Inform on Efficacy and Side Effects. *Circulation* 2019; **140**: 270–9.

585