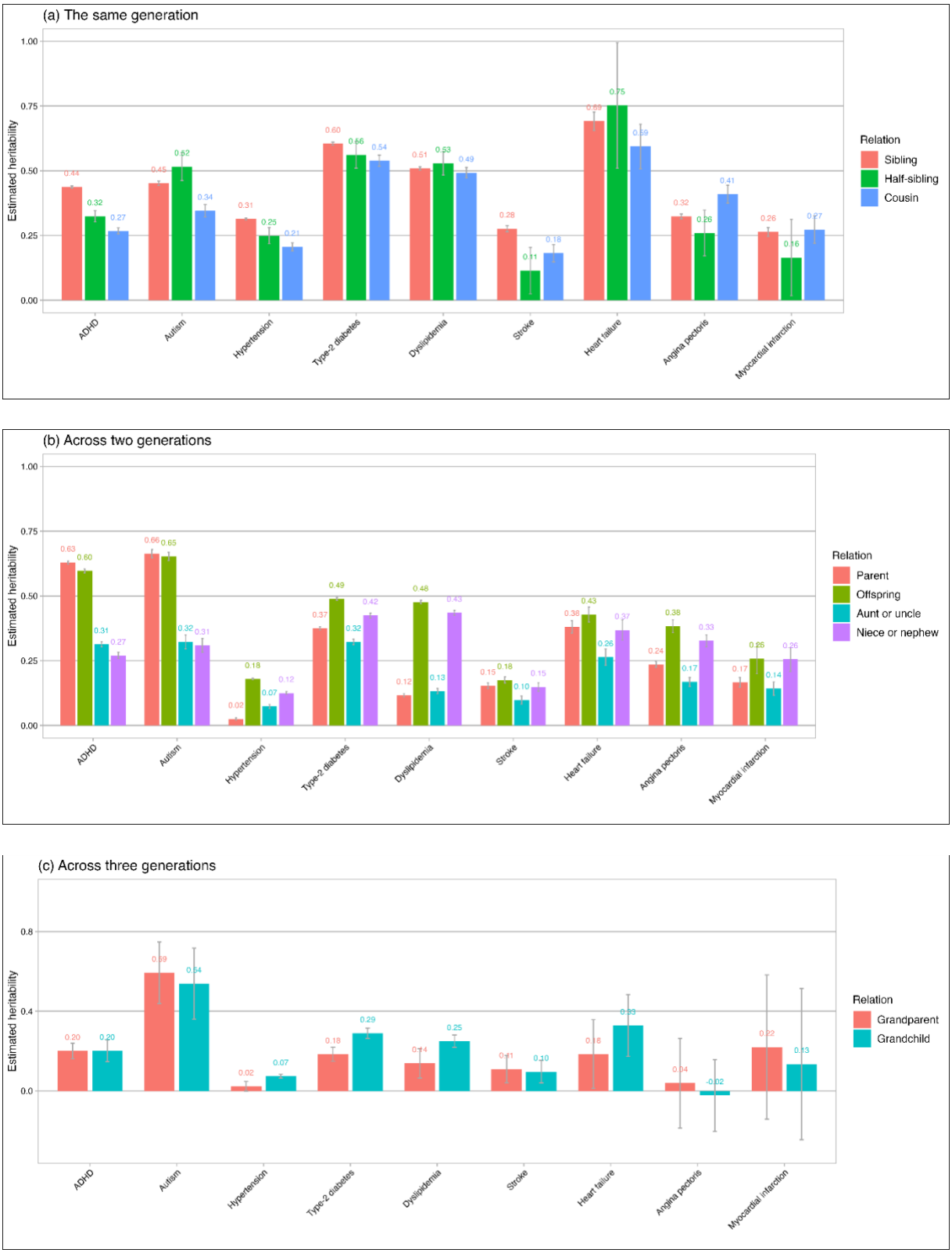
Supplementary eFigures

Supplementary methods

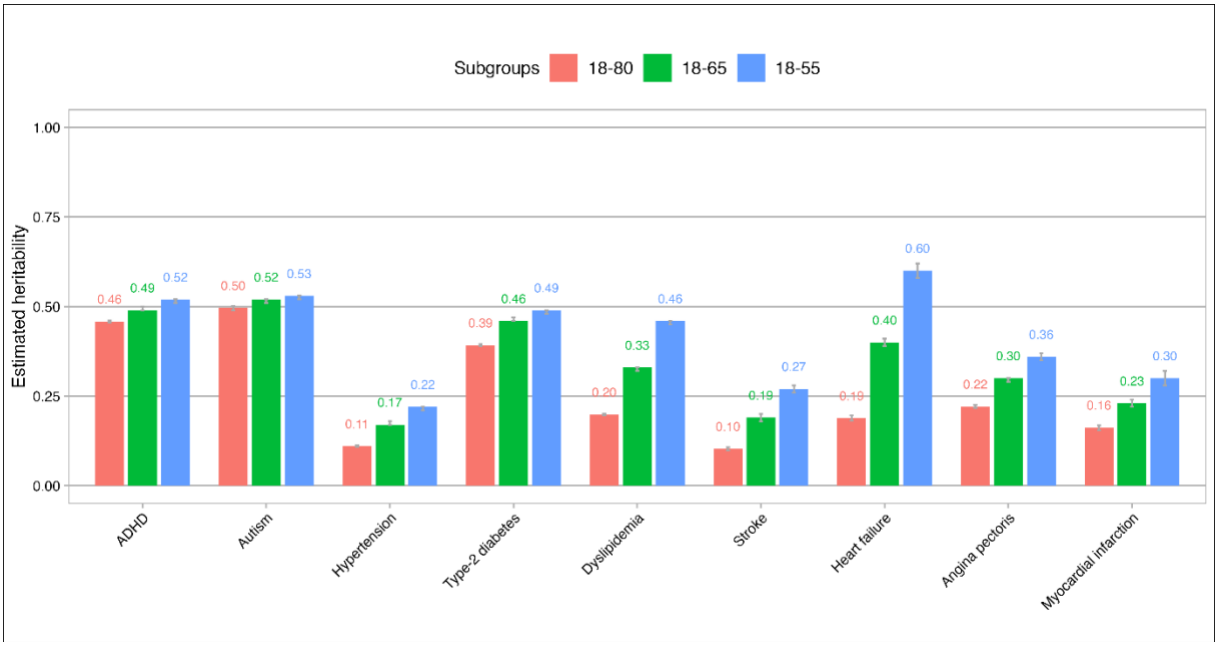
Supplementary eFigure 1. Heritabilities with 95% confidence intervals of all conditions estimated from relatives in different generations



Note:

ADHD, attention-deficit/hyperactivity disorder

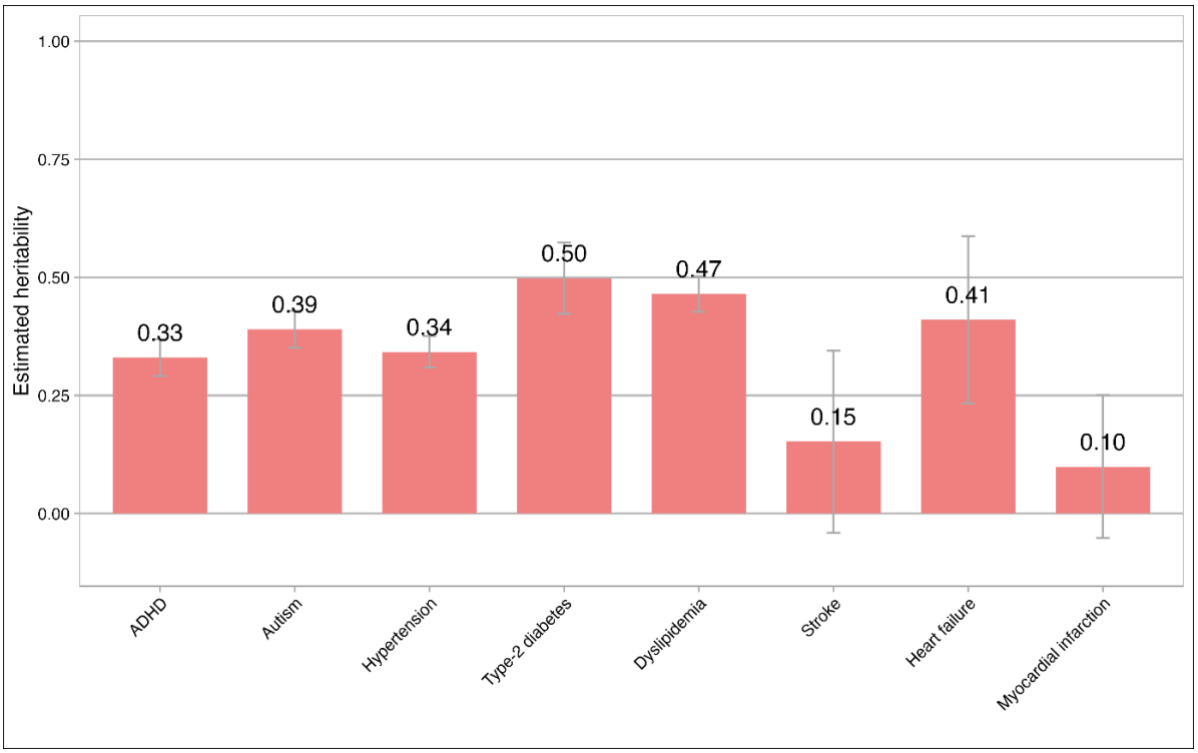
Supplementary eFigure 2. Estimated heritabilities with 95% confidence intervals of ADHD, autism, and cardiometabolic conditions in different age subgroups



Note:

ADHD, attention-deficit/hyperactivity disorder

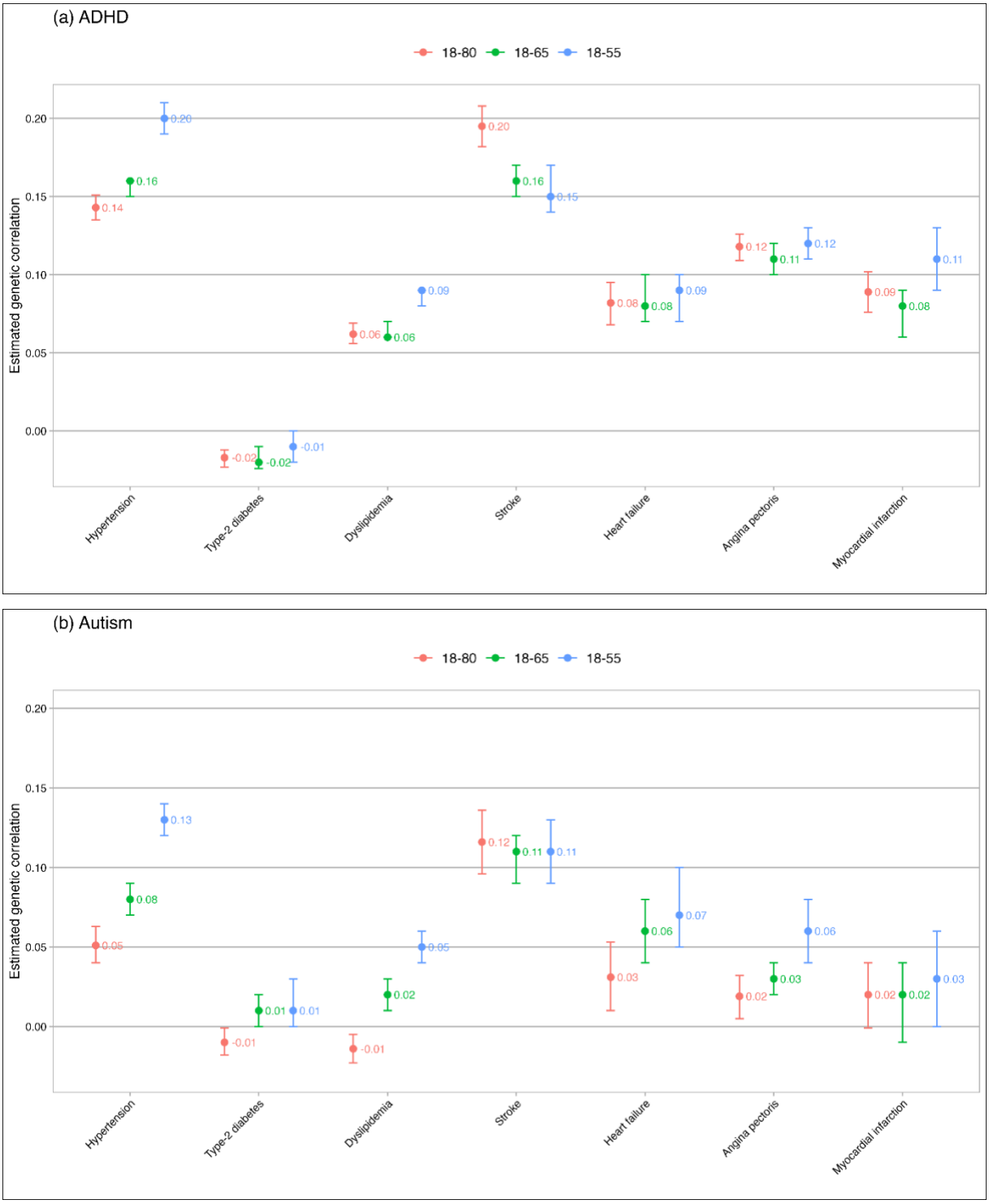
Supplementary eFigure 3. Estimated heritabilities with 95% confidence intervals of ADHD, autism, and cardiometabolic conditions from a Dutch cohort Lifelines (for validation)



Note: ADHD, attention-deficit/hyperactivity disorder

The results for ADHD and autism were derived from Vos et al.[1] using questionnaire-based assessments of symptoms.

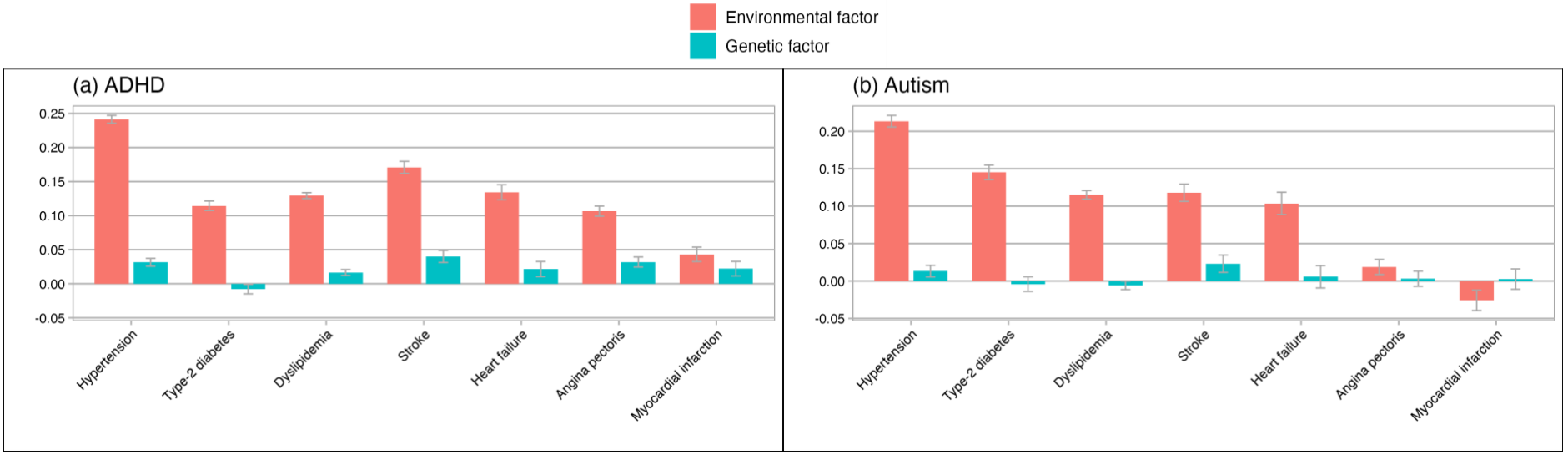
Supplementary eFigure 4. Estimated genetic correlations with 95% confidence intervals between ADHD, autism and cardiometabolic conditions in different age subgroups



Note:

ADHD, attention-deficit/hyperactivity disorder

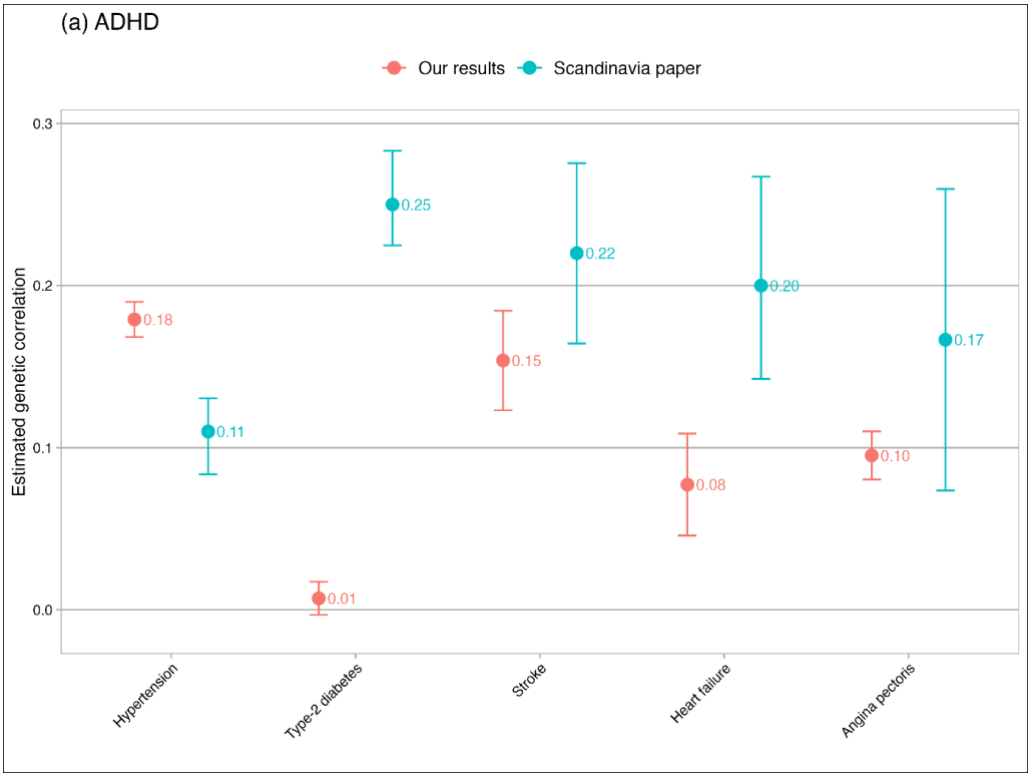
Supplementary eFigure 5. Genetic and environmental contributions to the phenotypic covariance with 95% confidence intervals between ADHD, autism and cardiometabolic conditions

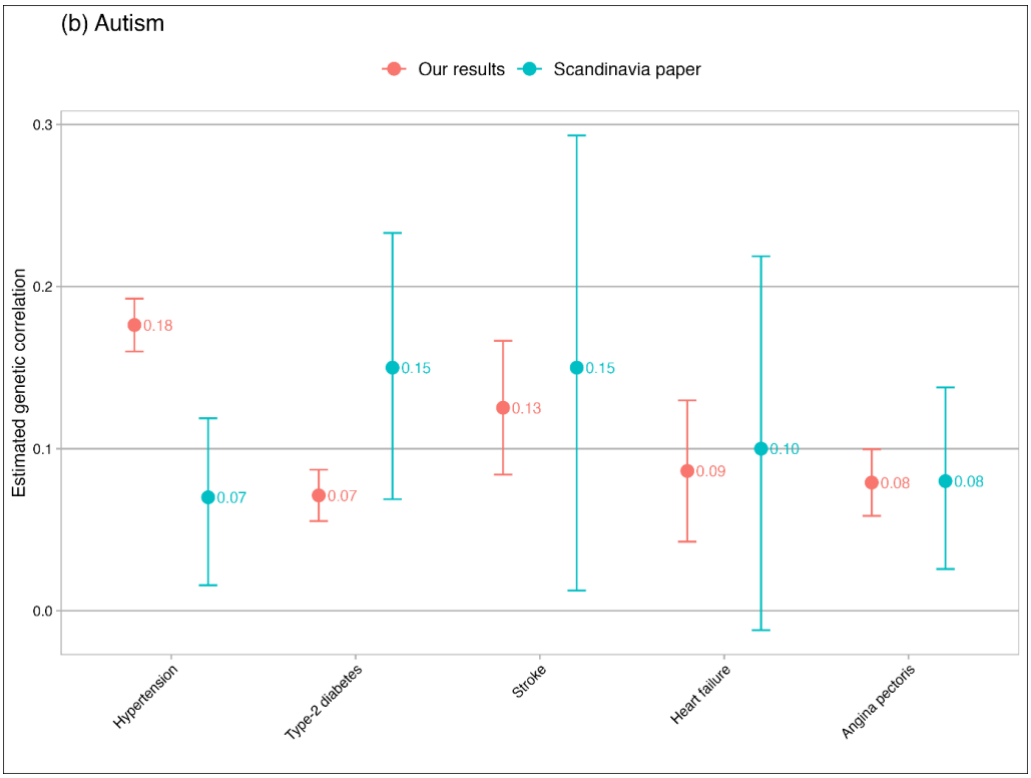


Note:

ADHD, attention-deficit/hyperactivity disorder.

Supplementary eFigure 6. Genetic correlations with 95% confidence intervals compared to a previous study among Scandinavians[17]

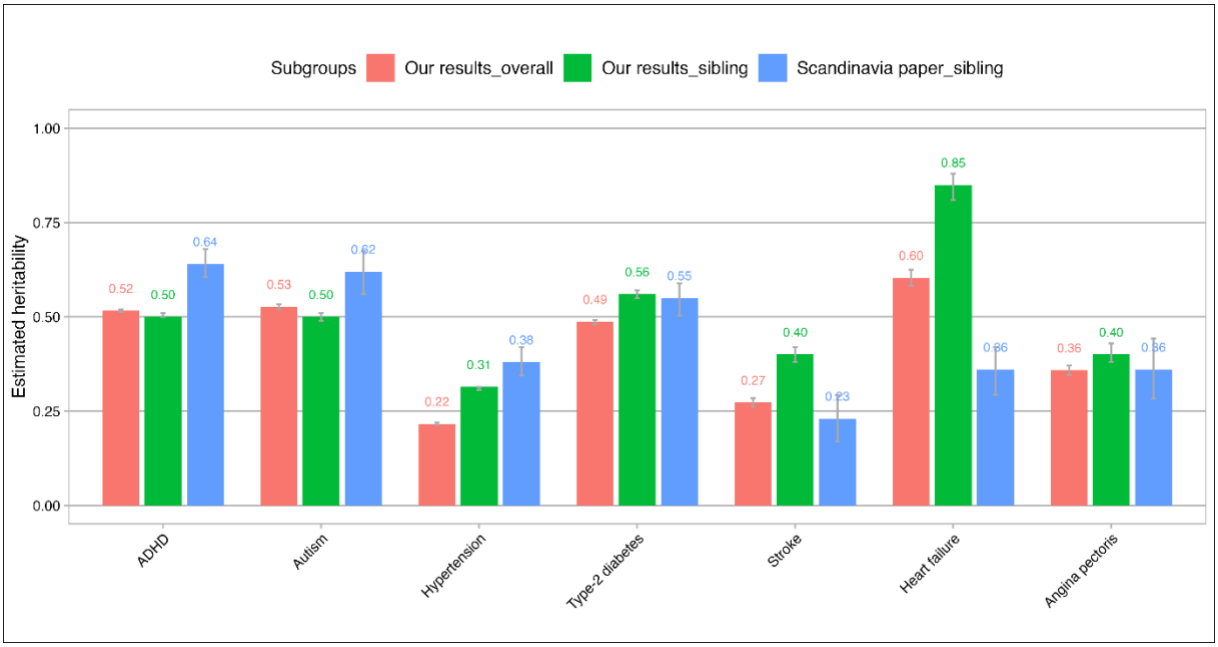




Note:

We used genetic correlations estimated among parent-offsprings in a subgroup aged 18-55 years (birth years 1965-2002), to compare with those reported in the previous Scandinavian study [17], which estimated the genetic correlations using parent-offspring pairs born between 1981 and 2005.

Supplementary eFigure 7. Heritabilities with 95% confidence intervals compared to a previous family study among Scandinavians[17]



Note:

We used heritability estimates in a subgroup aged 18-55 years (birth years 1965-2002), to compare with those reported in the previous Scandinavian study [17], which estimated the heritability using sibling pairs born between 1965 and 2016.

Supplementary methods

1. Additional data source (for validation): Lifelines

Lifelines is a multi-disciplinary prospective population-based cohort study examining in a unique three-generation design the health and health-related behaviours of 167,729 persons living in the North of the Netherlands.[2,3] It employs a broad range of investigative procedures in assessing the biomedical, socio-demographic, behavioural, physical and psychological factors which contribute to the health and disease of the general population, with a special focus on multi-morbidity and complex genetics. The participants were recruited between 2006 and 2013 at baseline and followed up from 2014 on-ward. The Lifelines protocol was approved by the University Medical Centre Groningen (UMCG) Medical ethical committee under number 2007/152 and all participants signed an informed consent form.[4]

1.1 Study population

We validated the results of heritabilities of cardiometabolic conditions using the baseline data of Lifelines. All adult participants with relevant data were included. This yielded a total sample of 152,009 adult individuals between 18 and 80 years of age with a mean age of 44.6 years (SD: 12.9) and 41.5% were male. More than half (56.7%) had at least one first-degree relative (1DR) in the dataset, while 13.7% had at least one second-degree relative (2DR) in the dataset. Prevalences of cardiometabolic conditions are shown in **eTable 1** and ranged from 0.5% for heart failure to 32.6% for dyslipidemia.

1.2 Ascertainment of cardiometabolic conditions

In Lifelines, we identified six cardiometabolic conditions, including hypertension, type-2 diabetes, dyslipidemia, stroke, heart failure, and myocardial infarction.

Hypertension

Blood pressure was measured using an automatic sphygmomanometer. We defined hypertension based on any of the following criteria:

1) systolic blood pressure ≥ 140mmHg;

2) diastolic blood pressure ≥ 90 mmHg;

3) use of blood pressure lowering drugs based on Anatomical Therapeutic Chemical (ATC) codes (C02, C03, C07, C08, C09 and G04CA03).

Type-2 diabetes

We defined type-2 diabetes based on any of the following criteria:

1) use of glucose-lowering drugs based on ATC codes (A10);

2) fasting glucose > 7 mmol/l or HbA1c > 6.5%;

3) self-reported type-2 diabetes validated by self-reported use of glucose-lowering drugs.

To avoid misclassifying other types of diabetes, participants meeting the above definition of type-2 diabetes but also meeting any of the following criteria were coded as missing:

1) participants under 30 years of age who did not self-report diabetes or had missing self-reported diabetes data;

2) participants who reported type-1 or other type of diabetes;

3) participants with diabetes onset before the age of 30 years or currently under 30 years old who treated their diabetes only with insulin (based on ATC codes or self-report);

4) pregnant participants.

Dyslipidemia

We defined dyslipidemia based on any of the following criteria:[5-8]

1) total cholesterol ≥ 6.5 mmol/L;

2) low-density lipoprotein cholesterol > 4.91 mmol/L;

3) high-density lipoprotein cholesterol < 1.03 mmol/L (men) or < 1.29 mmol/L (women);

4) triglycerides ≥ 2.3 mmol/L;

5) use of lipid modifying drugs based on ATC code C10.

Stroke

We defined stroke as self-reported stroke.

Heart failure

We defined heart failure as self-reported heart failure validated by either supporting drug use based on ATC codes (C09, C03DA), or self-reported therapy for heart failure (medication, pacemaker or heart transplant).

Myocardial infarction

We defined myocardial infarction as self-reported myocardial infarction with drug use (platelet aggregation inhibitors/antithrombotic drugs) or electrocardiography signs corresponding to myocardial infarction.[9]

2. Statistical Analysis

All statistical analyses were conducted for main results using CBS data. The analyses for familial aggregation and heritability of cardiometabolic conditions were conducted in Lifelines for validation.

2.1 Familial (co-)aggregation

To examine the familial (co-)aggregation, recurrence risk ratio (λ_R) was estimated. λ_R is the ratio of the prevalence of the condition in relatives of affected participants (e.g., siblings) compared to the prevalence in the general population,[10] with λ_R > 1 indicating positive familial (co-)aggregation (e.g., elevated risk in those with positive family history).

$$\lambda_R = \frac{K_r}{K} \quad \text{(A1)}$$

Where K is the prevalence of the condition in the whole study population, and K_r is the prevalence in individuals with affected relatives.

We employed plug-in methods to estimate the marginal K and K_r. First, we fitted logistic regression models regressing the condition outcome on the exposure (e.g., whether a sibling was affected by the condition). Models were adjusted for sex, number of relatives, and age (modeled as age and age^2 to account for non-linear effects) of both the individual and the sibling. The interaction between exposure and age (age and age^2) and the interaction between exposure and sex were included to account for effect modifiers.

$$\log\left(\frac{\Pr(C=1)}{1-\Pr(C=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 + \beta_9 X_9 + \beta_{10} X_{10} \quad \text{(A2)}$$

Where C = condition status of the individual (1 for cases, 0 for controls); X_1 = condition status of the relative (1 if the relative is affected, 0 if not); X_2 = sex of the individual; X_3 = number of the type of relatives; X_4 = age of the individual; X_5 = age^2 of the individual; X_6 = age of the relative (if there is more than one relative, e.g., siblings, the mean age was used); X_7 = age^2 of the relative; X_8 = condition status of the relative × age of the individual; X_9 = condition status of the relative × age^2 of the individual; X_{10}= condition status of the relative × sex of the individual.

Second, we estimated the average prevalence of the condition among the total population by plugging-in the fitted logistic regression to the observed total population; and the marginal prevalence of the condition among individuals with an affected sibling by plugging-in the fitted logistic regression to the observed total population with the exposure set to 1 for all observations.[11,12] To account for familial clustering and the resulting within-family correlations, 95% confidence intervals (CIs) were estimated for each prevalence using a robust sandwich estimator.[13] Standard errors (SE) of λ_R were estimated using the delta method.[14] These analyses were conducted using the R package "marginaleffects".[15,16]

2.2 Heritability (h^2) and genetic correlation (r_g)

We used equation B1 to estimate $h^2_{relative}$ and equation B2 to calculate the approximate SE in situations where individuals and their relatives have the same condition, and individuals have one affected relative.[17] These methods are generalized for scenarios where individuals have both parents affected by the same condition, as shown in B3 for h^2 and B4 for SE.

$$h^2 = \frac{T - T_R \times \sqrt{1-(1-\frac{T}{i})\times(T^2-T_R^2)}}{a_R \times (i+(i-T)\times T_R^2)} \quad \text{(B1)}$$

$$SE(h^2) \approx \frac{1}{a_R} \times \sqrt{[\frac{s.e(K)^2}{y^2} \times \left(\frac{1}{i} + a_R \times h^2 \times (i-T)\right)^2 + \frac{s.e(K_R)^2}{i^2 \times y_R^2}]} \quad \text{(B2)}$$

$$h^2 = \frac{2T - \sqrt{2}T_R \times \sqrt{2-(1-\frac{T}{i})\times(T^2-T_R^2)}}{2i+(i-T)\times T_R^2} \quad \text{(B3)}$$

$$SE(h^2) \approx \sqrt{[\frac{s.e(K)^2}{y^2} \times \left(\frac{1}{i} + h^2 \times (i-T)\right)^2 + \frac{s.e(K_R)^2}{i^2 \times y_R^2}]} \quad \text{(B4)}$$

Where T is the liability threshold of the condition in the general population, T_R is the liability threshold of the condition based on affected family members, i is the mean liability of the condition in the population calculated as $i = \frac{y}{K}$; K is the lifetime probability of the condition in the population and y is the height of the normal curve at threshold T. Under the liability threshold model, those with phenotypic liability, $Z \sim N(0,1)$, greater than the threshold T have the condition such that distribution $p(Z > T) = K$. a_R is the average coefficient of additive

genetic relationship between relatives (0.5 for first-, 0.25 for second-, and 0.125 for third-degree relatives), K_R is the lifetime probability of the condition in individuals with affected family members.

We used equation C1 to estimate shared heritability $r_{cf}h_ch_{f,relative}$ and equation C2 to estimate the approximate SE for situations where individuals have one condition (c), their relatives have the other condition (f), and individuals have one affected relative.

Equations C3 and C4 are for situations in which individuals have both parents affected by the condition (f).

$$r_{cf}h_ch_f = \frac{T_c - T_{cf}\sqrt{1-(1-\frac{T_f}{i_f})(T_c^2-T_{cf}^2)}}{a_R(i_f+(i_f-T_f)T_{cf}^2)} \quad \text{(C1)}$$

$$SE(r_{cf}h_ch_f) \approx \frac{1}{a_R}\sqrt{[\frac{s.e_{K_f}{}^2}{y_f{}^2}\left(\frac{1}{i_f}+a_Rr_{cf}h_ch_f(i_f-T_f)\right)^2+\frac{1}{i_f^2}(\frac{s.e_{K_c}{}^2}{y_c{}^2}+\frac{s.e_{K_{cf}}{}^2}{y_{cf}{}^2})]} \quad \text{(C2)}$$

$$r_{cf}h_ch_f = \frac{2T_c-\sqrt{2}T_{cf}\sqrt{2-(1-\frac{T_f}{i_f})(T_c^2-T_{cf}^2)}}{2i_f+(i_f-T_f)T_{cf}^2} \quad \text{(C3)}$$

$$SE(r_{cf}h_ch_f) \approx \sqrt{[\frac{s.e_{K_f}{}^2}{y_f{}^2}\left(\frac{1}{i_f}+r_{cf}h_ch_f(i_f-T_f)\right)^2+\frac{1}{i_f^2}(\frac{s.e_{K_c}{}^2}{y_c{}^2}+\frac{s.e_{K_{cf}}{}^2}{y_{cf}{}^2})]} \quad \text{(C4)}$$

We used equations D1 and D2 to estimate genetic correlation r_g and SE for each type of relative.

$$r_g = \frac{r_{cf}h_ch_f}{h_ch_f} \quad \text{(D1)}$$

$$SE(r_g) \approx r_g\sqrt{\frac{s.e(r_{cf}h_ch_f)^2}{r_{cf}h_ch_f{}^2}+\frac{(h_ch_f\sqrt{\frac{0.25\times(h_c^2h_f^2\sqrt{\frac{s.e(h_c^2)^2}{h_c^2{}^2}+\frac{s.e(h_f^2)^2}{h_f^2{}^2}})^2}{(h_c^2h_f^2)^2}})^2}{h_c^2h_f^2}} \quad \text{(D2)}$$

2.3 Contributions of genetic and environmental factors

We used equations E1 and E2 to estimate phenotypic (tetrachoric) correlations r_p , and equations E2, E3, and E4 to calculate the genetic factor G and environmental factor E for each type of relative.

$$r_p = \frac{OR^{\frac{3}{4}}-1}{OR^{\frac{3}{4}}+1} \quad \text{(E1)}^{18}$$

Where OR was estimated from logistic regressions.

$$\log\left(\frac{\Pr(C=1)}{1-\Pr(C=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad \text{(E2)}$$

Where C = condition status of the individual (e.g., 1 for having hypertension, 0 for not having hypertension); X_1 = the other condition status of the individual (e.g., 1 for having autism spectrum disorder, 0 for not having autism spectrum disorder); X_2 = sex of the individual; X_3 = age of the individual; X_4 = age^2 of the individual.

$$E = cov(E_c, E_f) = cov(P_c, P_f) - G \quad \text{(E3)}$$

$$cov(P_c, P_f) = r_p \times \sqrt{Var(P_c) \times Var(P_f)} \quad \text{(E4)}$$

$$G = cov(G_c, G_f) = r_g \times \sqrt{Var(G_c) \times Var(G_f)}$$

$$Var(G) = h^2 \times Var(P) \quad \text{(E5)}$$

In liability threshold model, $Var(P_c) = Var(P_f) = 1$. r_g is genetic correlation between one condition (c) and the other condition (f) calculated from equation D1, h^2 is the heritability calculated from equation B1.

References

1. Vos M, Wang R, Rommelse NN, Snieder H, Larsson H, Hartman CA. Familial co-aggregation and shared familiality among neurodevelopmental problems and with aggressive behavior, depression, anxiety, and substance use. *Psychological Medicine*. 2024:1-13.
2. Sijtsma A, Rienks J, van der Harst P, Navis G, Rosmalen JGM, Dotinga A. Cohort Profile Update: Lifelines, a three-generation cohort study and biobank. *International Journal of Epidemiology*. Dec 13 2021;51(5):E295-E302. doi:10.1093/ije/dyab257
3. Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *International Journal of Epidemiology*. Aug 2015;44(4):1172-1180. doi:10.1093/ije/dyu229
4. Stolk RP, Rosmalen JG, Postma DS, et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *Eur J Epidemiol*. 2008;23(1):67-74. doi:10.1007/s10654-007-9204-4
5. Mach F, Baigent C, Catapano AL, et al. 2019 ESC/EAS Guidelines for the management of dyslipidaemias: lipid modification to reduce cardiovascular risk. *Eur Heart J*. Jan 1 2020;41(1):111-188. doi:10.1093/eurheartj/ehz455
6. Triatin RD, Chen Z, Ani A, et al. Familial co-aggregation and shared genetics of cardiometabolic disorders and traits: data from the multi-generational Lifelines Cohort Study. *Cardiovascular Diabetology*. 2023;22(1):282.
7. Clinic; C. Hypercholesterolemia. 2025. https://my.clevelandclinic.org/health/diseases/23921-hypercholesterolemia
8. Bilen O, Virani SS, Nambi V. Hyperlipidemia. 2018;
9. van der Ende MY, Hartman MHT, Schurer RAJ, et al. Prevalence of electrocardiographic unrecognized myocardial infarction and its association with mortality. *Int J Cardiol*. Sep 15 2017;243:34-39. doi:10.1016/j.ijcard.2017.05.063
10. Risch N. Linkage strategies for genetically complex traits. I. Multilocus models. *American journal of human genetics*. 1990;46(2):222.
11. Muller CJ, MacLehose RF. Estimating predicted probabilities from logistic regression: different methods correspond to different target populations. *Int J Epidemiol*. Jun 2014;43(3):962-70. doi:10.1093/ije/dyu029
12. Dickerman BA, Hernan MA. Counterfactual prediction is not only for causal inference. *Eur J Epidemiol*. Jul 2020;35(7):615-617. doi:10.1007/s10654-020-00659-8
13. Zeileis A, Köll S, Graham N. Various Versatile Variances: An Object-Oriented Implementation of Clustered Covariances in R. *Journal of Statistical Software*. Oct 2020;95(1):1-36. doi:10.18637/jss.v095.i01
14. Oehlert GW. A note on the delta method. *The American Statistician*. 1992;46(1):27-29.
15. Arel-Bundock V GN, Heiss A (Forthcoming). How to Interpret Statistical Models Using marginaleffects in R and Python. *Journal of Statistical Software*.
16. Vincent Arel-Bundock NG, Andrew Heiss. Marginal Effects Zoo. https://marginaleffects.com/CITATION.html
17. Wray NR, Gottesman, II. Using summary data from the danish national registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. *Front Genet*. 2012;3:118. doi:10.3389/fgene.2012.00118
18. Digby PGN. Approximating the Tetrachoric Correlation-Coefficient. *Biometrics*. 1983;39(3):753-757. doi:Doi 10.2307/2531104