SUPPLEMENTARY MATERIAL

The Burden of Severe Hypercholesterolemia and Familial Hypercholesterolemia in a Population-based Setting in the US

Supplementary Table 1. Exclusion Criteria Based on Secondary Causes of Elevated LDL-C. Subjects Were Excluded if Any of the Following Were Detected Within a 1-year Window Before the Index Date.

Secondary Causes	Definition
Hypothyroidism	TSH >10 mIU/l
Significant Liver Disease	ALP ≥200 IU/l, and total bilirubin >2 mg/dl
Significant Kidney Disease	Creatinine >2.6 mg/dl, eGFR <15 ml/min/BSA, urine protein/creatinine ratio ≥2.95 mg/mg, 24-hour urine protein >3,000 mg/24h, nephrotic syndrome diagnosis codes, or renal failure diagnosis codes
Uncontrolled Diabetes	Hemoglobin A1c >9%
Pregnancy	Pregnancy, abortion, or delivery related diagnosis codes

ALP = alkaline phosphatase; ALT = alanine aminotransferase; AST = aspartate aminotransferase; BSA = body surface area; eGFR = estimated glomerular filtration rate; LDL-C = low-density lipoprotein cholesterol; IU = international unit; TSH = thyroid stimulating hormone.

Supplementary Table 2. Dutch Lipid Clinic Network (DLCN) Scoring System for Familial Hypercholesterolemia*

Criteria and Explanation	Score
Family history	
First-degree relative with known premature coronary and/or vascular disease, or First-degree relative with known LDL-C above the 95th percentile	1
First-degree relative with tendinous xanthomata and/or arcus cornealis, or Children aged less than 18 years with LDL-C above the 95th percentile	2
Clinical history	
Patient with premature CAD†	2
Patient with premature CVD or PAD†	1
Physical examination	
Tendinous xanthomata	6
Arcus cornealis prior to age 45 years	4
LDL-cholesterol levels (mg/dl)	
>325	8
250-325	5
190-249	3
155-189	1
DNA analysis	
Functional mutation in the LDLR, APOB, or PCSK9 gene	8

APOB = apolipoprotein B; CAD = coronary artery disease; CVD = cerebrovascular disease; LDL-C = low-density lipoprotein cholesterol; LDLR = low-density lipoprotein receptor; PAD = peripheral artery disease; PCSK9 = proprotein convertase subtilisin/kexin type 9.

*Diagnosis of familial hypercholesterolemia is based on the total number of points obtained:

- A "**definite**" diagnosis requires >8 points
- A "**probable**" diagnosis requires 6–8 points
- A "**possible**" diagnosis requires 3–5 points

†The term "premature" was originally defined as <55 years in males and <60 years in females. As proposed by some investigators, we used a <65 years as the cutoff in females.

Supplementary Table 3. ICD and CPT Codes Used in Coronary Heart Disease Algorithm

Coronary Heart Disease	ICD-9	ICD-10	CPT
Angina	411.1, 413.9	I20.0, I20.8	
Myocardial infarction	410.00, 410.01, 410.02, 410.10, 410.11, 410.12, 410.20, 410.21, 410.22, 410.30, 410.31, 410.32, 410.40, 410.41, 410.42, 410.50, 410.51, 410.52, 410.60, 410.61, 410.62, 410.70, 410.71, 410.72, 410.80, 410.81, 410.82, 410.90, 410.91, 410.92, 412, 429.71, 429.79	I21.01, I21.02, I21.09, I21.11, I21.19, I21.21, I21.29, I21.3, I21.4, I22.0, I22.1, I22.2, I22.8, I22.9, I23.0, I23.1, I23.2, I23.3, I23.4, I23.5, I23.6, I23.7, I23.8, I24.1, I25.2	
Chronic ischemic heart disease / coronary atherosclerosis	414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.2, 414.3, 414.4, 414.8, 414.9	I25.10, I25.110, I25.111, I25.118, I25.119, I25.5, I25.700, I25.701, I25.708, I25.709, I25.710, I25.711, I25.718, I25.719, I25.720, I25.721, I25.728, I25.729, I25.730, I25.731, I25.738, I25.739, I25.750, I25.751, I25.758, I25.759, I25.760, I25.761, I25.768, I25.769, I25.790, I25.791, I25.798, I25.799, I25.810, I25.811, I25.812, I25.82, I25.83, I25.84, I25.89, I25.9	
Acute ischemic heart disease	411.8, 411.81, 411.89	I24.0, I24.8, I24.9, I25.6	
Percutaneous coronary revascularization	00.66, 36.03, 36.04, 36.06, 36.07, 36.09, V45.82	0270xxx, 0271xxx, 0272xxx, 0273xxx, 02C0xxx, 02C1xxx, 02C2xxx, 02C3xxx, 3E07017, 3E070PZ, 3E07317, 3E073PZ, Z95.5, Z98.61	92920, 92921, 92924, 92925, 92928, 92929, 92933, 92934, 92937, 92938, 92941, 92943, 92944, 92980, 92981, 92982, 92984, 92995, 92996, 92973, 92974
Coronary artery bypass graft surgery	36.10, 36.11, 36.12, 36.13, 36.14, 36.15, 36.16, 36.17, 36.19, 36.2, V45.81	0210xxx, 0211xxx, 0212xxx, 0213xxx, Z95.1	33510, 33511, 33512, 33513, 33514, 33516, 33517, 33518, 33519, 33521, 33522, 33523, 33533, 33534, 33535, 33536

CPT = current procedural terminology; ICD = International Classification of Diseases.

Supplementary Table 4. ICD and CPT Codes Used in Cerebrovascular Disease Algorithm

Cerebrovascular Disease	ICD-9	ICD-10	CPT
Stroke	434.01, 434.91 437.1	I63.30, I63.311, I63.312, I63.313, I63.319, I63.321, I63.322, I63.323, I63.329, I63.331, I63.332, I63.333, I63.339, I63.341, I63.342, I63.343, I63.349, I63.39, I63.50, I63.511, I63.512, I63.513, I63.519, I63.521, I63.522, I63.523, I63.529, I63.531, I63.532, I63.533, I63.539, I63.541, I63.542, I63.543, I63.549, I63.59, I63.8, I63.9, I67.82	
Transient ischemic attack	435.0, 435.1, 435.3, 435.8, 435.9	G45.0, G45.1, G45.2, G45.8, G45.9	
Carotid artery disease	433.01, 433.11, 433.21, 433.31, 433.81, 433.91	I63.00, I63.011, I63.012, I63.013, I63.019, I63.02, I63.031, I63.032, I63.033, I63.039, I63.09, I63.20, I63.211, I63.212, I63.213, I63.219, I63.22, I63.239, I63.29	

CPT = current procedural terminology; ICD = International Classification of Diseases.

Supplementary Table 5. ICD Codes Used in Peripheral Artery Disease Algorithm

Peripheral Artery Disease	ICD-9	ICD-10
Atherosclerotic peripheral artery disease	440.20, 440.21, 440.22, 440.23, 440.24, 440.29	I70.201, I70.202, I70.203, I70.209, I70.211, I70.212, I70.213, I70.219, I70.221, I70.222, I70.223, I70.229, I70.231, I70.232, I70.233, I70.234, I70.235, I70.238, I70.239, I70.241, I70.242, I70.243, I70.244, I70.245, I70.248, I70.249, I70.25, I70.261, I70.262, I70.263, I70.269, I70.291, I70.292, I70.293, I70.299
Exclusion	237.70, 237.71, 237.72, 237.73, 237.79, 443.1, 446.0, 446.4, 446.5, 446.6, 446.7, 710.1, 747.10, 747.11, 747.22, 747.64	I73.1, M30.0, M31.1, M31.30, M31.31, M31.4, M31.5, M31.6, M34.0, M34.1, Q25.1, Q25.21, Q25.29, Q25.3, Q27.32, Q85.00, Q85.01, Q85.02, Q85.03, Q85.09

ICD = International Classification of Diseases.

Supplementary Table 6. Lipid Lowering Medications for SH Cases with Complete Records in the 18-month Periods Before the Index Date, After the Index Date, and Before the Last Follow-up Based on the Indication of Treatment and Age at the Index Date

	Primary Prevention Setting*				Secondary Prevention Setting†			
	None	Non-statin Only	Statin Only	Both	None	Non-statin Only	Statin Only	Both
Cases aged <40 years (n = 717)	n = 690				n = 27			
18-months before the index date	70%	3%	23%	4%	26%	4%	52%	19%
18-months after the index date	45%	3%	42%	10%	19%	0%	56%	26%
18-months before the last follow-up‡	48%	3%	40%	8%	21%	2%	60%	17%
Cases aged 40 - 54 years (n = 2,142)	n = 1,946				n = 196			
18-months before the index date	48%	3%	41%	7%	18%	1%	60%	21%
18-months after the index date	24%	4%	61%	11%	9%	5%	56%	31%
18-months before the last follow-up‡	27%	3%	59%	11%	10%	3%	63%	24%
Cases aged 55 – 69 years (n = 1,797)	n = 1,457				n = 340			
18-months before the index date	36%	3%	51%	10%	9%	0%	64%	27%
18-months after the index date	20%	5%	61%	14%	8%	5%	61%	26%
18-months before the last follow-up‡	21%	4%	58%	16%	12%	3%	63%	22%
Cases aged ≥70 years (n = 708)	n = 400				n = 308			
18-months before the index date	25%	4%	59%	13%	10%	3%	70%	18%
18-months after the index date	20%	4%	61%	15%	12%	6%	63%	19%
18-months before the last follow-up‡	34%	7%	48%	11%	24%	4%	57%	14%

SH = severe hypercholesterolemia.

*Primary prevention is defined as cases with no coronary heart disease, cerebrovascular disease, or peripheral artery disease on or before the index date for the first two time periods, and with no such disease at the follow-up date for the last time period.

†Secondary prevention is defined as cases with coronary heart disease, or cerebrovascular disease, or peripheral artery disease on or before the index date for the first two time periods, and with any such disease at the follow-up date for the last time period.

‡In cases with at least 36 months follow-up.

Supplementary Table 7. Number of Baseline and Incident Event in Cases and Controls with Complete Records

	SH (n = 5,364)	**FH (n = 248)**	**Controls (n = 5,364)**
CHD events	1,163 (21.7%)	189 (76.2%)	926 (17.3%)
Baseline CHD events	714 (13.3%)	146 (58.9%)	542 (10.1%)
New CHD events*	449 (9.7%)	43 (42.2%)	384 (8.0%)
CVD events	589 (11.0%)	49 (19.8%)	475 (8.9%)
Baseline CVD events	200 (3.7%)	14 (5.6%)	185 (3.4%)
New CVD events*	389 (7.5%)	35 (15.0%)	290 (5.6%)
PAD events	176 (3.3%)	25 (10.1%)	168 (3.1%)
Baseline PAD events	71 (1.3%)	9 (3.6%)	73 (1.4%)
New PAD events*	105 (2.0%)	16 (6.7%)	95 (1.8%)
Composite endpoint	1,517 (28.3%)	199 (80.2%)	1,226 (22.9%)
Baseline composite endpoint	871 (16.2%)	149 (60.1%)	686 (12.8%)
New composite endpoint*	646 (14.4%)	50 (50.5%)	540 (11.5%)

CHD = coronary heart disease; CVD = cerebrovascular disease; FH = familial hypercholesterolemia; PAD = peripheral artery disease; SH = severe hypercholesterolemia.

*Excludes people with baseline events.

Supplementary Table 8. Multivariable Adjusted Logistic Regression Model for Baseline Cardiovascular Events in SH Cases with Complete Records in Comparison with Controls

	Coronary Heart Disease		Cerebrovascular Disease		Peripheral Artery Disease		Composite*	
	OR (95% CI)	p Value	OR (95% CI)	p Value	OR (95% CI)	p Value	OR (95% CI)	p Value
Case vs. control	1.47 (1.29 - 1.68)	<0.001	1.10 (0.88 - 1.36)	0.400	0.95 (0.67 - 1.35)	0.785	1.43 (1.26 - 1.62)	<0.001
Age, decade	1.96 (1.85 - 2.07)	<0.001	1.89 (1.73 - 2.07)	<0.001	1.82 (1.58 - 2.10)	<0.001	2.02 (1.92 - 2.13)	<0.001
Sex (male)	2.39 (2.08 - 2.74)	<0.001	1.16 (0.93 - 1.46)	0.184	1.98 (1.39 - 2.83)	<0.001	2.17 (1.91 - 2.47)	<0.001
Race (non-white)	0.90 (0.72 - 1.12)	0.347	0.97 (0.67 - 1.41)	0.879	0.76 (0.40 - 1.43)	0.390	0.88 (0.71 - 1.08)	0.210
Hypertension	2.76 (2.26 - 3.37)	<0.001	2.51 (1.75 - 3.58)	<0.001	2.50 (1.37 - 4.55)	0.003	2.69 (2.25 - 3.21)	<0.001
Diabetes	2.13 (1.86 - 2.45)	<0.001	1.82 (1.45 - 2.28)	<0.001	2.38 (1.65 - 3.42)	<0.001	2.16 (1.90 - 2.46)	<0.001
No tobacco use		<0.001		0.001		<0.001		<0.001
Unknown	0.27 (0.15 - 0.46)		0.07 (0.01 - 0.52)		0.60 (0.08 – 4.56)		0.24 (0.14 - 0.40)	
Tobacco use	1.49 (1.30 - 1.71)		1.31 (1.04 - 1.64)		4.31 (2.63 - 7.07)		1.54 (1.36 - 1.76)	
Normal BMI		0.737		0.172		0.547		0.640
Underweight	1.07 (0.44 - 2.61)		1.57 (0.52 - 4.71)		0.95 (0.12 – 7.55)		1.20 (0.55 - 2.66)	
Unknown	NA		NA		NA		NA	
Overweight	0.90 (0.74 - 1.09)		1.00 (0.75 - 1.33)		0.89 (0.56 - 1.41)		0.88 (0.74 - 1.05)	
Obese	0.99 (0.81 - 1.19)		0.76 (0.56 - 1.03)		0.67 (0.41 - 1.10)		0.90 (0.75 - 1.07)	
Triglyceride ≥150 mg/dl	0.95 (0.82 - 1.09)	0.456	1.08 (0.85 - 1.36)	0.532	0.99 (0.68 - 1.44)	0.960	0.97 (0.85 - 1.11)	0.668
Low HDL-C†	1.40 (1.21 - 1.63)	<0.001	1.21 (0.95 - 1.54)	0.119	1.43 (0.99 - 2.07)	0.060	1.40 (1.22 - 1.61)	<0.001

BMI = body mass index; CI = confidence interval; HDL-C = high-density lipoprotein cholesterol; OR = odds ratio; SH = severe hypercholesterolemia.

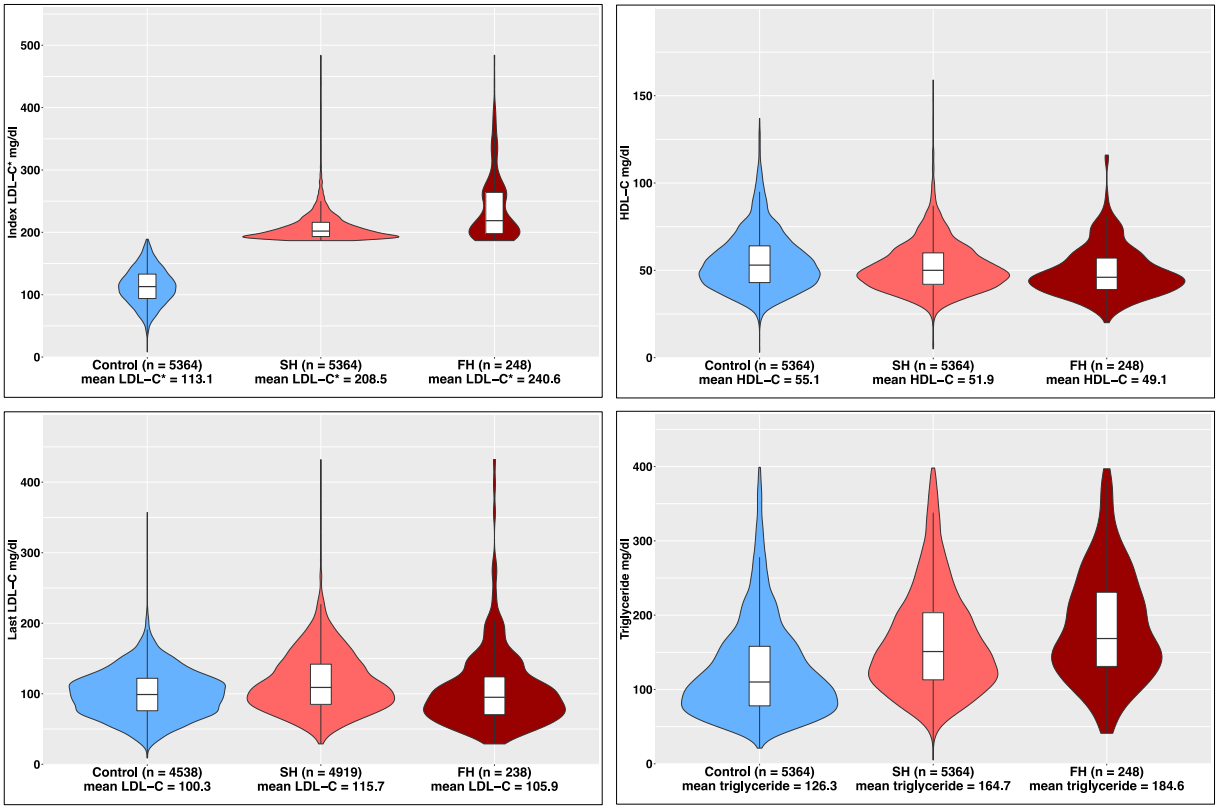
*Defined as the earliest of coronary heart disease, cerebrovascular disease, and peripheral artery disease.
†Low HDL-C was defined as <40 mg/dl for males or <50 mg/dl for females.

Supplementary Table 9. Multivariable Adjusted Models for New Cardiovascular Events in SH Cases with Complete Records Compared with Their Controls.

	Coronary Heart Disease		Cerebrovascular Disease		Peripheral Artery Disease		Composite*	
	HR (95% CI)	p Value	HR (95% CI)	p Value	HR (95% CI)	p Value	HR (95% CI)	p Value
Case vs. control	1.21 (1.05 - 1.39)	0.010	1.30 (1.11 - 1.53)	0.001	1.06 (0.80 - 1.42)	0.678	1.24 (1.10 - 1.39)	<0.001
Age, decade	1.68 (1.58 - 1.78)	<0.001	1.87 (1.75 - 2.00)	<0.001	1.72 (1.52 - 1.94)	<0.001	1.80 (1.71 - 1.90)	<0.001
Sex (male)	1.99 (1.72 - 2.30)	<0.001	1.32 (1.13 - 1.55)	<0.001	1.49 (1.11 - 1.99)	0.008	1.65 (1.46 - 1.86)	<0.001
Race (non-white)	0.66 (0.50 - 0.87)	0.003	1.31 (1.02 - 1.68)	0.032	0.97 (0.58 - 1.60)	0.890	0.93 (0.76 - 1.15)	0.513
Hypertension	1.61 (1.35 - 1.92)	<0.001	1.63 (1.31 - 2.02)	<0.001	2.00 (1.31 - 3.06)	0.001	1.56 (1.35 - 1.80)	<0.001
Diabetes	1.33 (1.14 - 1.56)	<0.001	1.51 (1.28 - 1.78)	<0.001	1.67 (1.23 - 2.26)	<0.001	1.31 (1.14 - 1.49)	<0.001
No tobacco use		<0.001		<0.001		<0.001		<0.001
Unknown	0.82 (0.58 - 1.15)		0.79 (0.52 - 1.20)		0.72 (0.26 - 2.02)		0.81 (0.61 - 1.08)	
Tobacco use	1.32 (1.14 - 1.53)		1.38 (1.17 - 1.63)		2.14 (1.54 - 2.96)		1.41 (1.25 - 1.60)	
Normal BMI		0.286		0.972		0.003		0.577
Underweight	0.79 (0.25 - 2.49)		1.43 (0.53 - 3.87)		4.64 (1.66 - 12.96)		1.30 (0.58 - 2.92)	
Unknown	0.72 (0.10 - 5.20)		NA		NA		0.46 (0.06 - 3.24)	
Overweight	0.91 (0.75 - 1.11)		1.00 (0.81 - 1.24)		0.73 (0.50 - 1.06)		0.91 (0.78 - 1.07)	
Obese	1.08 (0.89 - 1.32)		1.01 (0.81 - 1.26)		0.67 (0.45 - 0.99)		0.98 (0.83 - 1.16)	
Triglyceride ≥150 mg/dl	1.07 (0.92 - 1.25)	0.359	0.97 (0.82 - 1.15)	0.710	1.15 (0.85 - 1.55)	0.382	1.11 (0.98 - 1.26)	0.098
Low HDL-C†	1.16 (0.99 - 1.36)	0.075	1.04 (0.87 - 1.25)	0.655	1.19 (0.86 - 1.64)	0.289	1.12 (0.98 - 1.29)	0.095

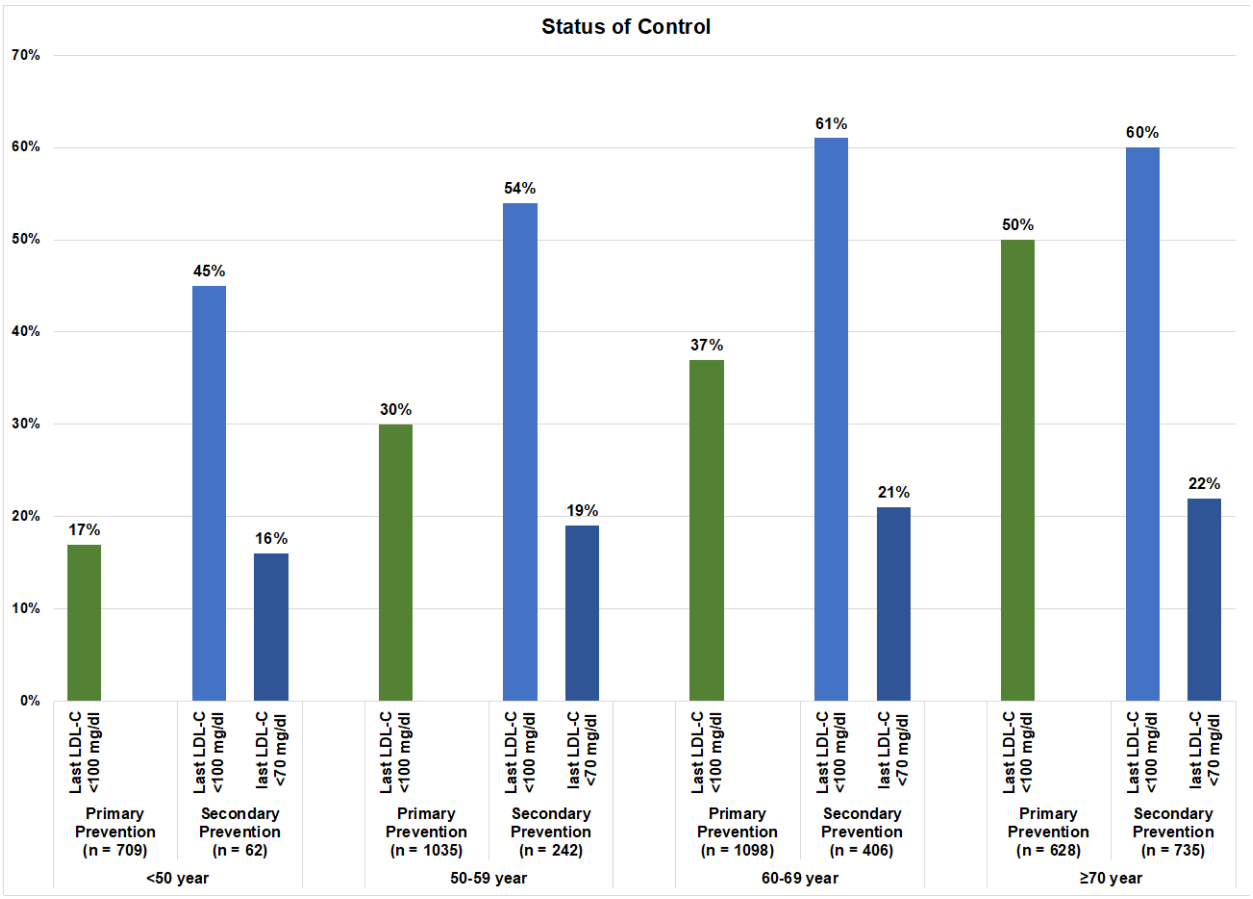
BMI = body mass index; CI = confidence interval; HDL-C = high-density lipoprotein cholesterol; HR = hazard ratio; SH = severe hypercholesterolemia.

*Defined as the earliest of coronary heart disease, cerebrovascular disease, and peripheral artery disease.
†Low HDL-C was defined as <40 mg/dl for males or <50 mg/dl for females.



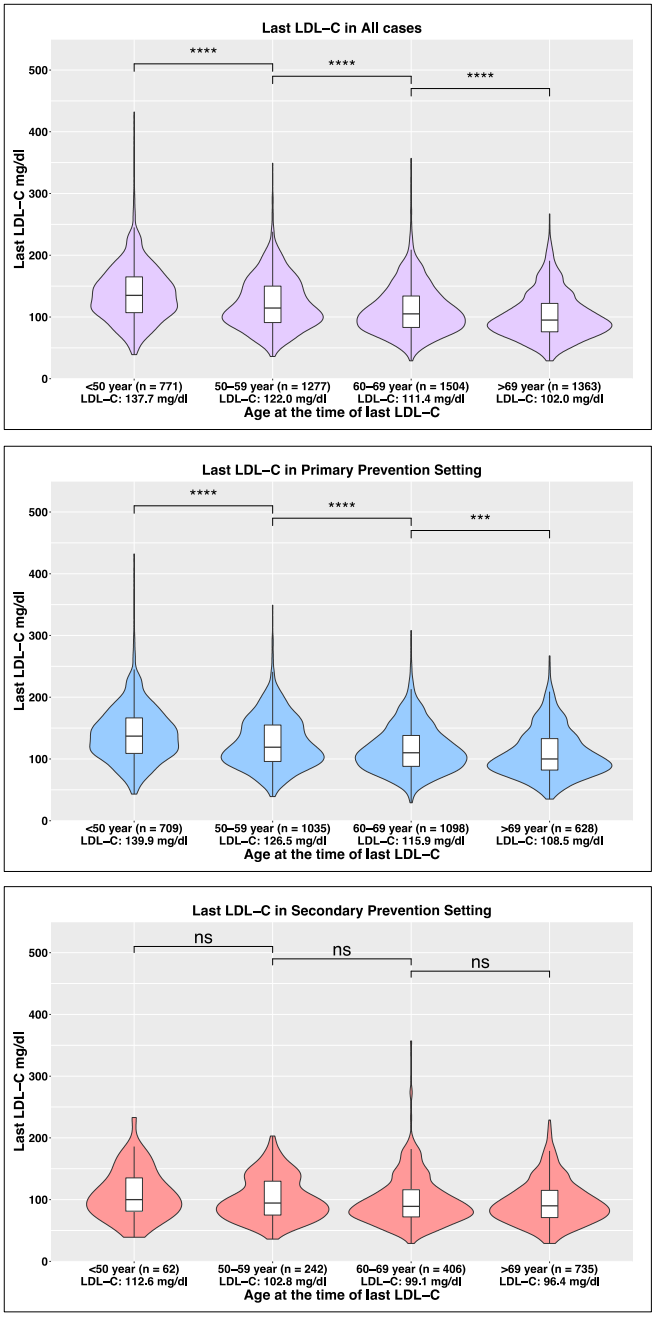
Supplementary Figure 1. Lipid profile in SH, FH, and controls. LDL-C at the index date (top left), HDL-C at the index date (top right), triglyceride at the index date (bottom right), and LDL-C at the last follow-up (bottom left) between controls, SH cases, and FH cases. For those on a statin, the untreated index LDL-C was estimated by multiplying the LDL-C by 1.33.

FH = familial hypercholesterolemia; HDL-C = high-density lipoprotein cholesterol; LDL-C = low-density lipoprotein cholesterol; SH = severe hypercholesterolemia.



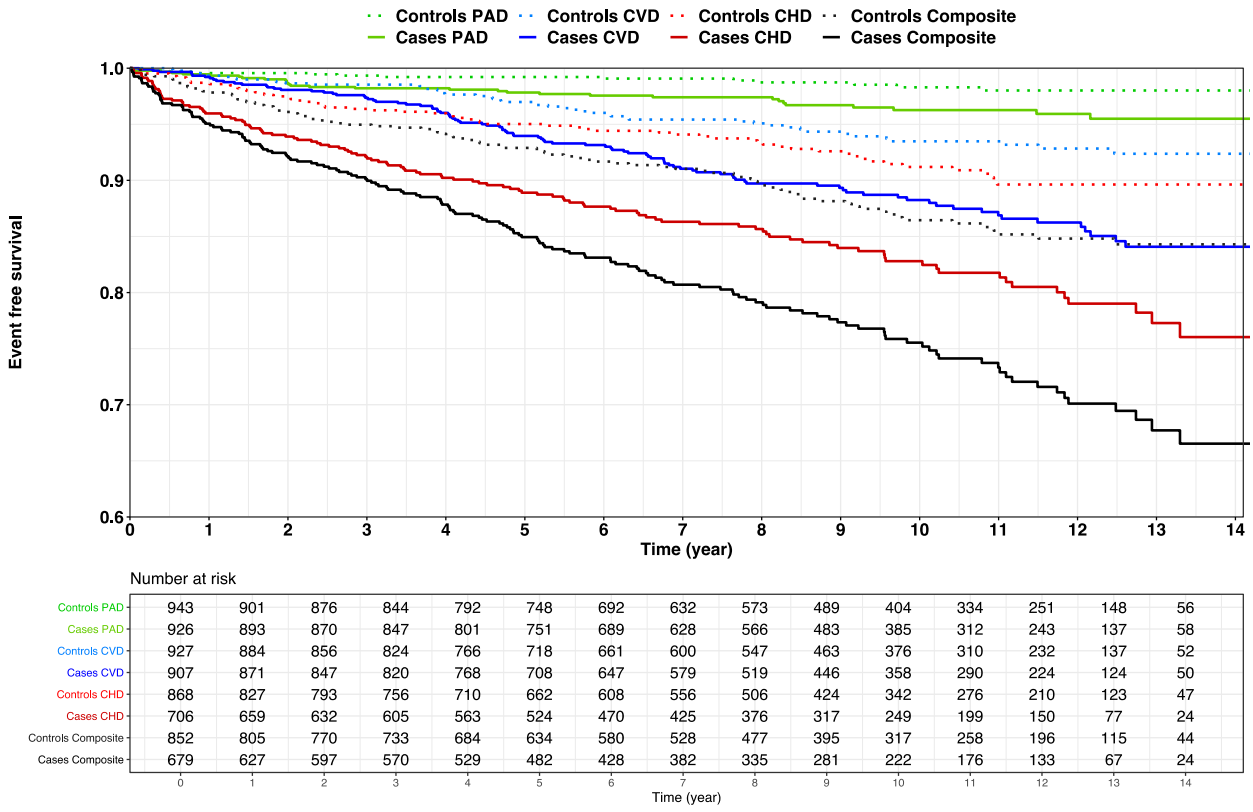
Supplementary Figure 2. Status of LDL-C control in SH cases based on two thresholds in primary vs. secondary prevention and in different age categories. Limited to cases with LDL-C measured at least 6 months after the index date. Primary prevention is defined as cases with no coronary heart disease, cerebrovascular disease, or peripheral artery disease on or before the follow-up date. Secondary prevention is defined as cases with coronary heart disease, or cerebrovascular disease, or peripheral artery disease on or before the follow-up date.

LDL-C = low-density lipoprotein cholesterol.



Supplementary Figure 3. Last measured LDL-C level for SH cases based on the indication of treatment (primary vs. secondary prevention) in different age categories. Limited to cases with complete records and with LDL-C measured at least 6 months after index date. Primary prevention is defined as cases with no coronary heart disease, cerebrovascular disease, or peripheral artery disease on or before the follow-up date. Secondary prevention is defined as cases with coronary heart disease, or cerebrovascular disease, or peripheral artery disease on or before the follow-up date. ***p value <0.001; ****p value <0.0001.

LDL-C = low-density lipoprotein cholesterol; ns = non-significant; SH = severe hypercholesterolemia.



Supplementary Figure 4. Kaplan-Meier plot for survival free of PAD in green (p value: 0.01), CVD in blue (p value: <0.001), CHD in red (p value: <0.001), and composite endpoint in black (p value: <0.001) as well as number at risk of SH cases with a positive family history of premature CHD, CVD, or PAD (American Heart Association definition of FH) and their corresponding controls. The solid line represents the cases, and the dotted line represents the controls.

CHD = coronary heart disease; CVD = cerebrovascular disease; FH = familial hypercholesterolemia; PAD = peripheral artery disease.

Supplementary Methods for estimating Incidence, relative survival, and age recursion model (period) prevalence:

To estimate the incidence of SH and the subset of these who met clinical criteria for FH, incident cases were defined as Olmsted County residents who met the criteria for the first time between 2004 and 2015. Incident cases who were <18 years old, had a missing value in the lipid profile, or did not have a matched control (n=266) were included for estimating incidence and prevalence, but were excluded from all other analyses. Denominators were based on annual REP population counts. Period prevalence was estimated using an age-recursive method, that incorporated estimates of age-specific incidence and relative survival. We first calculated the age and sex-specific incidence rates based on numbers of cases and person years in each "age-sex" bin. Second, the age and sex-specific "relative survival" (the hazard ratio associated with mortality risk when having the condition compared to the mortality risk in the general population), was estimated by using standard life tables for Minnesota white individuals to transform each observed follow-up time into an observed "cumulative hazard" for each incident case, and combining that with their vital status at the end of follow-up, in a negative exponential model with the rate ratio as an exponential scale parameter that could depend on age and sex in a loglinear manner. Finally, the age-sex-specific incidence and age-sex-specific relative hazards were combined using an age-recursive model that calculates age-specific prevalence starting at age zero and increasing a year at a time up to age 89, based on the incidence and differential survival at each successive age. The age-specific prevalence was then used to estimate an overall prevalence adjusted both to the age distribution within the Olmsted population and separately to the US 2010 White population. Standard errors and confidence intervals, both for the age-specific prevalence estimates and for the overall prevalence estimates were obtained using a bootstrap resampling approach, wherein the incident cases were resampled with replacement in a way that allowed the number of cases in each bootstrap sample to have Poisson variation. In effect, this is equivalent to drawing bootstrap samples from the entire set of incident cases and non-incident residents in the community.

Definitions:

λi represent the incidence rate at age i.

πi represent the prevalence at age i.

γi represent the relative mortality at age i (relative to the population)

and θi represent the mortality at age i.

Recursion relationships:

Let φi be the current mortality rate in those who do NOT have the prevalent condition. Then $(1- \pi i) * \varphi i + \pi i * \varphi i * x = \theta i$, where x is the relative hazard of death for the condition relative to those without the condition. Therefore,

$$\Gamma i = \varphi i * x / [(1- \pi i) * \varphi i + \pi i * \varphi i * x]$$

Therefore,

$$\varphi i * x = \gamma i * [(1- \pi i) * \varphi i + \pi i * \varphi i * x]$$

$$x * \varphi i * (1 - \gamma i * \pi i) = \gamma i * (1- \pi i) * \varphi i$$

Thus,

$$x = [\gamma i *(1- \pi i) * \varphi i] / [\varphi i * (1 - \gamma i * \pi i)]$$

Now, for the recursion. We will actually not need "x", the relative hazard for those with and without the condition directly. In words, "The number with the disease at age i+1" = "number with disease at age i" - "number with disease who die" + "number of new cases" The new population size is given by "population at i + 1" = "population at i" - "number of deaths". We arbitrarily allow death to occur first, and then we add incidence into the scheme. So, if we let Ni be the population size at the ith age, we have

$$Ni+1 = Ni * (1 - \theta i)$$

Number alive with disease at beginning of ith epoch: $\pi i * Ni$.

We now allow the force of differential mortality to occur. The number of prevalent cases who remain alive is:

$$(\pi i * Ni) * (1 - \gamma i * \theta i)$$

and the number of prevalent cases who die is:

$$(\pi i * Ni) * \gamma i * \theta i$$

The number of prevalent non-cases who die is the difference between the total number of deaths and this last expression. Thus, this number is:

$$\theta i * Ni - (\pi i * Ni) * \gamma i * \theta i = Ni* \theta i * (1 - \pi i * \gamma i).$$

Note that this implies a constraint that $\gamma i < 1/ \pi i$.

The number of prevalent non-cases who remain alive is therefore:

$$[Ni * (1- \pi i) - Ni* \theta i * (1 - \pi i * \gamma i)].$$

Let us now allow these prevalent non-cases to incur incidence, namely, we will have at the end of the ith epoch:

$$[Ni*(1- \pi i) - Ni* \theta i * (1 - \pi i * \gamma i)] * \lambda i$$

New prevalent cases, and total number of prevalent cases at the end of the epoch will therefore be:

$$(\pi i * Ni) * (1- \gamma i * \theta i) + [Ni*(1- \pi i) - Ni* \theta i * (1- \pi i * \gamma i)] * \lambda i$$

We already know the total number of individuals alive at the end of the epoch, namely $Ni * (1 - \theta i)$. We can now calculate the prevalence at $i + 1$ as:

$$\pi i+1 = [(\pi i * Ni) * (1 - \gamma i * \theta i) + [Ni*(1- \pi i) – Ni * \theta i * (1 - \pi i * \gamma i)] * \lambda i] \div [Ni * (1 - \theta i)]$$

We can eliminate N from this formula and this the recursion formula to get us from prevalence at age i to age i+1.

$$\pi i+1 = [\pi i * (1 - \gamma i * \theta i) + [(1- \pi i) - \theta i * (1 - \pi i * \gamma i)] * \lambda i] \div [(1 - \theta i)]$$

How are the components of the above formula obtained?

λi: From the Poisson incidence model.

θi: We get this from standard life tables.

γi: We get this from a relative survival model, using our incidence cases, their age/sex/calendar year, and their death/follow-up times, and again the Minnesota White survival tables.

Specifically, we calculate $–\log(S0(t))$ for each incidence observation, and each follow-up/death time t. We then consider this transformed time to be proportional to the cumulative hazard with proportionality constant γ. We can then use a censored exponential model.