Web Material

Web Appendix 1. Details in data cleaning of the CPRD and the JUPITER trial

(1) The CPRD:

Step 1. Data cleaning and defining the variables.

Given possible data entry error and some extreme laboratory values and measurements, we applied the trimming rules (i.e., excluding the data if the values are outside of limits) listed in Table A to clean data. We then defined the variables of interests within 2 years prior to the index date. If multiple entries were available, we selected the data closest to the index date.

Special consideration was made for hsCRP. Infection can sharply increase the hsCRP values; however, elevated hsCRP induced by infection is not equivalent to those related to clinically relevant inflammation in cardiovascular diseases which is the main value of interest for this study. To define an hsCRP value for the CPRD patients, we only selected data from the hsCRP tests with values of 20 mg/L or less within 2 years prior to the index date. We further excluded the CPRD patients with recent infection, which was defined as having an hsCRP test result of >20 mg/L within 90 days prior to the index date.

Table A. Trimming rules for CPRD laboratory values and measurements

Test or Measurement	Lower limit	Upper limit
BMI (Kg/m^2)	12.00	60.00
Systolic blood pressure (mmHg)	60.00	280.00
Diastolic blood pressure (mmHg)	40.00	140.00
hsCRP (mg/L)	0.01	1000.00
Fasting plasma glucose (mmol/L)	1.00	50.00
HDL-C (mg/dL)	10.00	200.00
LDL-C (mg/dL)	1.00	600.00
Total cholesterol (mg/dL)	60.00	800.00
Triglycerides (mg/dL)	15.00	10,000.00
Creatinine (mg/dL)	0.15	20.00

Step 2. Handling the extreme values of the multiply imputed data

After multiply imputing missing data for the CPRD population, some patients possibly have the imputed values outside of the trimming ranges. To deal with this, we truncated the values using the limits of the trimming rules except for hsCRP (Table A). For example, we limited the maximum value of BMI to 60. As an example, for a given patient with imputed BMI of 65, he/she would have BMI of 60 for future analysis. For hsCRP, we truncated values at 0.01 for lower limit and at 20 for upper limit.

(2) The JUPITER trial:

After examining JUPITER data, we found some unlikely test results, likely due to data entry errors, which conflict with trial inclusion criteria. Thus, we also applied limits, listed in Table B, to re-code the JUPITER trial data. Values outside of the limits were re-coded as missing.

In JUPITER, the baseline hsCRP value was defined as the average of two hsCRP test values before randomization which could result in a baseline hsCRP lower than 2 mg/L. In our study, we took different approaches. For the patients who had the mean values of two hsCRP tests between 2 and 20 mg/L, we used the same value as the baseline hsCRP. For other patients, we chose whichever test was between 2 and 20 mg/L. For those with both two hsCRP test values of <2 or >20 mg/L, we coded the baseline hsCRP as missing.

As a result, there were 1,128 (6.3%) patients in the JUPITER trial with at least one variable coded as missing. In the unweighted analysis, all JUPITER patients were included (N=17,802). However, in the weighted analysis, only JUPITER patients without missing data were included (N=16,674).

Table B. Limits for JUPITER laboratory values and measurements

Test or Measurement	Lower limit	Upper limit
BMI (Kg/m^2)	12.00	60.00
Systolic blood pressure (mmHg)	n/a	190.00
Diastolic blood pressure (mmHg)	n/a	100.00
Fasting plasma glucose (mmol/L)	n/a	7.00
LDL-C (mg/dL)	n/a	130.00
Triglycerides (mg/dL)	n/a	500.00

Web Appendix 2. Secondary Analysis using Weighting by Inverse Probability

We had used “a simplified version” of weighting by inverse probability of missingness to address selection due to missing data in the target population. In order to overcome the non-monotone missing pattern in our study, we created one binary variable (Yes/No) to indicate whether a patient had missing data for ≥1 variable of interest in all CPRD patients. A weighting method for non-monotone missing data would model the arbitrary missing-data patterns. This is why we called our weighting method a simplified version. After creating one overall missing indicator, we used a logistic model to estimate the probability of being missing (PM) based on measured variables. Patients who met exclusion criteria were excluded after estimating PM. We assigned missingness weight as 1/(1-PM) for patients with complete data and 0 for patients with missing data. After weighting by missingness, the patients with complete data were comparable to the entire eligible CPRD population with respect to measured variables such as age and sex.

We then stacked the complete case cohort and the JUPITER trial data, and assigned missingness weight as 1 for the JUPITER participants for the modeling purpose. We ran another weighted logistic regression model based on missingness weight and all effect modifiers of interests to obtain the selection probability of being in RCT (PS). Selection weight was calculated as ((1- PS)/PS) for the JUPITER participants, and selection weight was truncated at 20. Within the reweighted JUPITER trial, we then used Cox proportional hazard models to estimate hazard ratios and their 95% confidence intervals (CIs) accounting for the weighting (robust variance estimator) for the comparison of major cardiovascular events between those randomized to rosuvastatin versus placebo. We calculated the weighted risk difference using a non-parametric method accounting for non-CVD competing causes of death, and obtained standard deviation from 200 bootstrap datasets to calculate 95% CIs.

After accounting for selection into the trial, the baseline characteristics of the JUPITER participants became comparable to the entire eligible CPRD population (Table C below). Within the reweighted JUPITER trial, we estimated the hazard ratio and risk difference of major cardiovascular events between those randomized to rosuvastatin versus placebo. After standardizing to the CPRD population, the rosuvastatin benefits in JUPITER based on weighting by inverse probability of missingness were more similar to the results after standardizing to the complete case cohort than to the results based on multiple imputation (Table D and Figure A below). Fully exploring the differences of using multiple imputation versus weighting by inverse probability of missingness with respect to generalizability would require a more sophisticated implementation of the weighting method but we think this is outside the scope of our paper.

Table C. Baseline Characteristics among the Complete Case Cohort weighted by inverse missingness probability and the JUPITER trial, before and after weighting by selection into trial[a, b]

Characteristics	Complete Case Cohort Weighted by Inverse Missingness Probability		Weighted the JUPITER Trial		Absolute Standardized Difference[c]
	n	(%)	n	(%)	
No.	2,190	(100.0)	15,131	(100.0)	--
Male	1,359	(62.0)	9,584	(63.3)	0.0269
Current Smoker	364	(16.6)	2,512	(16.6)	0.0007
Anti-hypertensive drugs	1,069	(48.8)	7,189	(47.5)	0.0262
Aspirin	239	(10.9)	1,585	(10.5)	0.0140
Moderate CKD[d]	495	(22.6)	3,464	(22.9)	0.0066
	Median	(IQR)	Median	(IQR)	
Age, years	64	(58-74)	64	(57-74)	0.0329
BMI, kg/m^2	27.9	(24.6-31.5)	27.7	(24.9-31.3)	0.0126
hs-CRP, mg/L	5.0	(3.0-7.4)	4.9	(3.1-7.4)	0.0097
DBP, mmHg	80	(74-86)	80	(74-87)	0.0440
SBP, mmHg	138	(128-146)	138	(128-147)	0.0111
HDL-C, mg/dL	50	(43-62)	51	(42-62)	0.0032
LDL-C, mg/dL	112	(99-123)	112	(100-121)	0.0116
Triglycerides, mg/dL	126	(104-168)	124	(107-162)	0.0007
Total Cholesterol, mg/dL	190	(178-205)	191	(177-205)	0.0107
Glucose, mmol/L	5.1	(4.8-5.5)	5.2	(4.8-5.6)	0.0236

Table D. Hazard ratio (HR) and risk difference (RD) of major cardiovascular events in JUPITER after standardizing the complete case cohort which was weighted by inverse missingness probability. Selection weight was truncated at 20.

	Point Estimate	95% CI
HR	0.67	0.43, 1.06
1-yr RD	-0.41	-1.16, 0.35
2-yr RD	-0.49	-1.89, 0.90
3-yr RD	-2.61	-5.14, -0.08
4-yr RD	-4.86	-8.20, -1.52

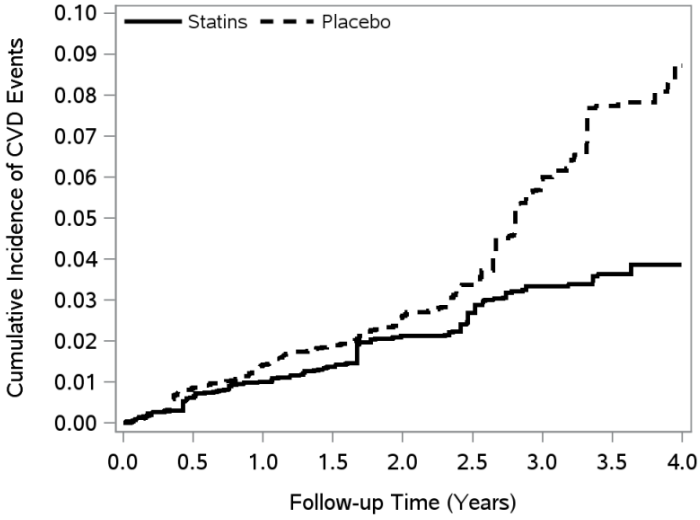


Figure A. Cumulative incidences of the primary endpoint (major cardiovascular events) by treatment in the JUPITER trial after standardizing to all eligible CPRD population using weighting to account for missingness.

Web Table 1. Baseline Characteristics among the Imputed Cohort and the JUPITER trial among patients aged younger than 80 years, before and after weighting (no weight truncation)[a]

Characteristics	Imputed Cohort[b]		JUPITER Trial							
			Original Trial		Weighted Trial		Weighted Rosuvastatin Group		Weighted Placebo Group	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
No.	325,522	(100.0)	16,964	(100.0)	15,785	(100.0)	7,949	(100.0)	7,836	(100.0)
Male	224,005	(68.8)	10,604	(62.5)	10,816	(68.5)	5,395	(67.9)	5,421	(69.2)
Current Smoker	72,708	(22.3)	2,776	(16.4)	3,639	(23.1)	1,824	(22.9)	1,815	(23.2)
Anti-hypertensive drugs	110,737	(34.0)	8,298	(48.9)	5,555	(35.2)	2,772	(34.9)	2,783	(35.5)
Aspirin	20,854	(6.4)	3,076	(18.1)	1,008	(6.4)	512	(6.4)	496	(6.3)
Moderate CKD[c]	77,157	(23.7)	2,920	(17.2)	3,603	(22.8)	1,895	(23.8)	1,708	(21.8)
	Median	(IQR)	Median	(IQR)	Median	(IQR)	Median	(IQR)	Median	(IQR)
Age, years	62	(55-69)	65	(60-70)	62	(56-69)	63	(56-69)	62	(55-68)
BMI[d]	27	(23.9-30.4)	28.4	(25.4-32.0)	27	(24.1-30.3)	27.1	(24.1-30.4)	27	(24.1-30.2)
hs-CRP, mg/L	4.7	(3.1-7.5)	4.1	(2.8-6.5)	4.7	(3.0-7.7)	4.7	(3.1-7.7)	4.6	(3.0-7.6)
DBP, mmHg	80	(74-86)	80	(75-87)	80	(74-87)	80	(74-88)	80	(75-87)
SBP, mmHg	137	(126-146)	134	(124-145)	137	(126-148)	138	(126-148)	136	(126-147)
HDL-C, mg/dL	55	(44-67)	49	(40-59)	53	(43-68)	54	(43-68)	53	(43-67)
LDL-C, mg/dL	109	(94-120)	108	(94-119)	109	(93-121)	109	(94-121)	109	(93-120)
Triglycerides, mg/dL	116	(85-163)	119	(86-170)	115	(82-164)	114	(81-162)	116	(83-166)
Total Cholesterol, mg/dL	189	(172-206)	185	(169-200)	190	(172-207)	190	(172-207)	190	(172-206)
Glucose, mmol/L	5.2	(4.6-5.8)	5.2	(4.9-5.7)	5.2	(4.8-5.6)	5.2	(4.8-5.6)	5.2	(4.8-5.6)

Abbreviation: BMI, body mass index; CKD, chronic kidney disease; CPRD, clinical practice research datalink; hs-CRP, high-sensitivity C reactive protein; DBP, diastolic blood pressure; SBP, systolic blood pressure; IQR, interquartile range; JUPITER, justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

[a] The logistic regression model used to predict the probability of being selected in the imputed cohort included the following variables: age, age^2, age^3, age^4, male, male×age, male×age^2, smoker, smoker×age, smoker×age^2, smoker×age^3, CKD, CKD×male, CKD×age, CKD×age^2, anti-hypertensive drugs, anti-hypertensive drugs×age, anti-hypertensive drugs×age^2, anti-hypertensive drugs×CKD, aspirin, age^{x}aspirin, age^2×aspirin, age^3×aspirin, aspirin×anti-hypertensive drugs, aspirin×CKD, aspirin×smoker, BMI, BMI×male, BMI×age, hsCRP, hsCRP^2, hsCRP×CKD, DBP, DBP×CKD, SBP, SBP^2, SBP^3, SBP×CKD, SBP×age, SBP×anti-hypertensive drugs, SBP^2×anti-hypertensive drugs, SBP^3×anti-hypertensive drugs, HDL-C, HDL-C×CKD, HDL-C×aspirin, HDL-C×male, HDL-C×DBP, HDL-C×SBP, LDL-C, LDL-C^2, LDL-C^3, LDL-C^4, LDL-C×smoker, glucose, glucose×age, glucose×age^2, glucose×age^3, glucose×anti-hypertensive drugs, glucose×BMI, glucose×aspirin, glucose×LDL-C, glucose×DBP, log(triglycerides), log(triglycerides)^2, log(triglycerides)^3, log(triglycerides)×smoker, total cholesterol, and (total cholesterol)^2.

[b] We multiply imputed 20 datasets. We used one randomly selected imputed dataset to select variables included in the selection model and to examine balance in baseline characteristics between the target population and JUPITER. Thus, we presented the results from this randomly selected imputed dataset in the table.

[c] Moderate CKD was defined as the estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m^2. eGFR was calculated using serum creatinine values and the Modification of Diet in Renal Disease Study (MDRD) equation. Patients on dialysis or with eGFR <15 mL/min/1.73 m^2 were excluded from the study population.

[d] Weighted (kg)/height (m^2)

Web Table 2. Baseline Characteristics among the Complete Case Cohort and the JUPITER trial among patients aged younger than 80 years, before and after weighting (no weight truncation)[a]

Characteristics	Imputed Cohort		JUPITER Trial							
			Original Trial		Weighted Trial		Weighted Rosuvastatin Group		Weighted Placebo Group	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
No.	2,437	(100.0)	16,964	(100.0)	15,625	(100.0)	7,769	(100.0)	7,856	(100.0)
Male	1,454	(59.7)	10,604	(62.5)	9,390	(60.1)	4,597	(59.2)	4,793	(61.0)
Current Smoker	430	(17.6)	2,776	(16.4)	2,740	(17.5)	1,423	(18.3)	1,317	(16.8)
Anti-hypertensive drugs	1,313	(53.9)	8,298	(48.9)	8,482	(54.3)	4,164	(53.6)	4,317	(55.0)
Aspirin	246	(10.1)	3,076	(18.1)	1,609	(10.3)	805	(10.4)	804	(10.2)
Moderate CKD[b]	441	(18.1)	2,920	(17.2)	2,836	(18.2)	1,424	(18.3)	1,412	(18.0)
	Median	(IQR)	Median	(IQR)	Median	(IQR)	Median	(IQR)	Median	(IQR)
Age, years	62	(57-69)	65	(60-70)	62	(57-68)	62	(57-68)	62	(56-68)
BMI[c]	29	(25.3-32.7)	28	(25.4-32.0)	28.7	(25.6-32.4)	28.6	(25.5-32.5)	28.7	(25.7-32.4)
hs-CRP, mg/L	5	(3.0-7.0)	4	(2.8-6.5)	4.9	(3.0-7.2)	4.9	(3.1-7.3)	4.8	(2.9-7.1)
DBP, mmHg	80	(74-86)	80	(75-87)	80	(74-86)	80	(74-86)	80	(74-86)
SBP, mmHg	136	(125-144)	134	(124-145)	135	(126-145)	135	(125-145)	135	(126-144)
HDL-C, mg/dL	50	(43-62)	49	(40-59)	51	(42-62)	51	(42-62)	50	(41-62)
LDL-C, mg/dL	112	(101-120)	108	(94-119)	112	(99-121)	112	(99-122)	112	(99-121)
Triglycerides, mg/dL	128	(105-175)	119	(86-170)	125	(106-169)	125	(106-167)	125	(106-170)
Total Cholesterol, mg/dL	193	(178-205)	185	(169-200)	192	(177-206)	193	(178-206)	192	(177-206)
Glucose, mmol/L	5	(4.8-5.5)	5	(4.9-5.7)	5.2	(4.8-5.6)	5.2	(4.8-5.6)	5.2	(4.8-5.6)

Abbreviation: BMI, body mass index; CKD, chronic kidney disease; CPRD, clinical practice research datalink; hs-CRP, high-sensitivity C reactive protein; DBP, diastolic blood pressure; SBP, systolic blood pressure; IQR, interquartile range; JUPITER, justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

[a] The logistic regression model used to predict the probability of being selected in the imputed cohort included the following variables: age, age^2, male, male×age, male×age^2, smoker, smoker×age, smoker×male, CKD, CKD×male, CKD×age, CKD×age^2, anti-hypertensive drugs, aspirin, aspirin×anti-hypertensive drugs, BMI, BMI^2, hsCRP, hsCRP^2, hsCRP^3, hsCRP^4, hsCRP^5, hsCRP^6, hsCRP^7, DBP, SBP, SBP^2, SBP^3, SBP×anti-hypertensive drugs, HDL-C, HDL-C×smoker, LDL-C, LDL-C×male, glucose, log(triglycerides), log(triglycerides)^2, log(triglycerides)^3, log(triglycerides)^4, and total cholesterol.

[b] Moderate CKD was defined as the estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m^2. eGFR was calculated using serum creatinine values and the Modification of Diet in Renal Disease Study (MDRD) equation. Patients on dialysis or with eGFR <15 mL/min/1.73 m^2 were excluded from the study population.

[c] Weighted (kg)/height (m^2)

Web Table 3. Frequency of age group in the JUPITER trial and the imputed cohort

Age, years	CPRD		JUPITER	
	n	%	n	%
50-54	68,161	18.8	874	5.0
55-59	42,539	11.8	2,781	15.8
60-64	76,961	21.3	3,797	21.5
65-69	50,470	14.0	4,634	26.3
70-74	37,768	10.4	3,037	17.2
75-79	31,236	8.6	1,720	9.8
80+	54,699	15.1	806	4.6

Web Table 4. Baseline Characteristics among the Imputed Cohort and the JUPITER trial, before and after weighting (no weight truncation)[a]

Characteristics	Imputed Cohort[b]		JUPITER Trial							
			Original Trial		Weighted Trial		Weighted Rosuvastatin Group		Weighted Placebo Group	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
No.	383,418	(100.0)	17,802	(100.0)	16,960	(100.0)	8,258	(100.0)	8,702	(100.0)
Male	245,691	(64.1)	11,001	(61.8)	10,797	(63.7)	5,244	(63.5)	5,553	(63.8)
Current Smoker	77,909	(20.3)	2,821	(15.8)	3,363	(19.8)	1,640	(19.9)	1,723	(19.8)
Anti-hypertensive drugs	148,576	(38.8)	8,846	(49.7)	6,755	(39.8)	3,324	(40.3)	3,431	(39.4)
Aspirin	35,871	(9.4)	3,313	(18.6)	1,590	(9.4)	733	(8.9)	856	(9.8)
Moderate CKD[c]	110,490	(28.8)	3,257	(18.3)	4,824	(28.4)	2,360	(28.6)	2,464	(28.3)
	Median	(IQR)	Median	(IQR)	Median	(IQR)	Median	(IQR)	Median	(IQR)
Age, years	64	(57-74)	66	(60-71)	64	(57-74)	65	(57-74)	64	(57-75)
BMI, kg/m^2	26.6	(23.5-30.1)	28.3	(25.3-32.0)	26.6	(23.5-29.7)	26.6	(23.6-29.9)	26.5	(23.5-29.5)
hs-CRP, mg/L	4.8	(3.1-7.6)	4.1	(2.8-6.5)	4.7	(3.1-7.7)	4.8	(3.1-7.6)	4.7	(3.0-7.7)
DBP, mmHg	80	(74-86)	80	(75-87)	80	(73-86)	80	(74-86)	80	(72-86)
SBP, mmHg	138	(126-147)	134	(124-145)	138	(128-148)	139	(128-150)	136	(128-147)
HDL-C, mg/dL	56	(45-68)	49	(40-60)	57	(45-69)	57	(45-69)	56	(45-69)
LDL-C, mg/dL	109	(93-120)	108	(94-119)	109	(92-120)	108	(93-120)	109	(91-119)
Triglycerides, mg/dL	115	(83-159)	118	(85-169)	116	(82-162)	115	(81-160)	116	(83-164)
Total Cholesterol, mg/dL	189	(172-206)	185	(169-200)	190	(172-206)	191	(172-207)	189	(172-206)
Glucose, mmol/L	5.2	(4.6-5.8)	5.2	(4.8-5.7)	5.2	(4.8-5.6)	5.1	(4.8-5.6)	5.2	(4.8-5.6)

Abbreviation: BMI, body mass index; CKD, chronic kidney disease; CPRD, clinical practice research datalink; hs-CRP, high-sensitivity C reactive protein; DBP, diastolic blood pressure; SBP, systolic blood pressure; IQR, interquartile range; JUPITER, justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

[a] The logistic regression model used to predict the probability of being selected in the imputed cohort included the following variables: age, age^2, age^3, age^4, age^5, male, male×age, male×age^2, male×age^3, male×age^4, smoker, smoker×age, smoker×male, CKD, CKD×male, CKD×age, CKD×age^2, CKD×age^3, anti-hypertensive drugs, anti-hypertensive drugs×age, anti-hypertensive drugs×age^2, anti-hypertensive drugs×male, anti-hypertensive drugs×CKD, aspirin, age×aspirin, age^2×aspirin, aspirin×anti-hypertensive drugs, BMI, BMI^2, BMI×CKD, BMI×smoker, BMI×male, BMI^2×male, hsCRP, hsCRP^2, hsCRP^3, hsCRP^4, hsCRP×smoker, hsCRP×male, hsCRP^2×male, hsCRP^3×male, DBP, DBP×SBP, SBP, SBP^2, SBP^3, SBP×CKD, SBP^2×CKD, SBP^3×CKD, SBP×anti-hypertensive drugs, SBP^2×anti-hypertensive drugs, SBP^3×anti-hypertensive drugs, HDL-C, HDL-C^2, HDL-C^3, HDL-C^4, HDL-C^5,HDL-C×CKD, HDL-C×smoker, HDL-C×male, LDL-C, LDL-C^2, LDL-C^3, LDL-C^4, glucose, glucose×anti-hypertensive drugs, glucose×BMI, glucose×age, glucose×age^2, glucose×LDL-C, log(triglycerides), log(triglycerides)^2, log(triglycerides)^3, log(triglycerides)×HDL-C, and total cholesterol.

[b] We multiply imputed 20 datasets. We used one randomly selected imputed dataset to select variables included in the selection model and to examine balance in baseline characteristics between the target population and JUPITER. Thus, we presented the results from this randomly selected imputed dataset in the table.

[c] Moderate CKD was defined as the estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 ^{2}. eGFR was calculated using serum creatinine values and the Modification of Diet in Renal Disease Study (MDRD) equation. Patients on dialysis or with eGFR <15 mL/min/1.73 m^2 were excluded from the study population.

Web Table 5. Baseline Characteristics among the Imputed Cohort and the JUPITER trial, before and after weighting (weight truncation at 20) [a]

Characteristics	Imputed Cohort[b]		JUPITER Trial				Absolute Standardized Difference[c]
			Original Trial		Weighted Trial		
	n	(%)	n	(%)	n	(%)	
No.	383,418	(100.0)	17,802	(100.0)	16,165	(100.0)	
Male	245,691	(64.1)	11,001	(61.8)	10,236	(63.3)	0.0158
Current Smoker	77,909	(20.3)	2,821	(15.8)	3,316	(20.5)	0.0048
Anti-hypertensive drugs	148,576	(38.8)	8,846	(49.7)	6,451	(39.9)	0.0236
Aspirin	35,871	(9.4)	3,313	(18.6)	1,540	(9.5)	0.0059
Moderate CKD[d]	110,490	(28.8)	3,257	(18.3)	4,256	(26.3)	0.0557
	Median	(IQR)	Median	(IQR)	Median	(IQR)	
Age, years	64	(57-74)	66	(60-71)	65	(57-74)	0.0189
BMI, kg/m^2	26.6	(23.5-30.1)	28.3	(25.3-32.0)	26.5	(23-29.8)	0.0105
hs-CRP, mg/L	4.8	(3.1-7.6)	4.1	(2.8-6.5)	4.8	(3.1-7.7)	0.0001
DBP, mmHg	80	(74-86)	80	(75-87)	80	(73-86)	0.0108
SBP, mmHg	138	(126-147)	134	(124-145)	138	(127-150)	0.0385
HDL-C, mg/dL	56	(45-68)	49	(40-60)	57	(45-68)	0.0146
LDL-C, mg/dL	109	(93-120)	108	(94-119)	109	(93-120)	0.0232
Triglycerides, mg/dL	115	(83-159)	118	(85-169)	113	(81-161)	0.0081
Total Cholesterol, mg/dL	189	(172-206)	185	(169-200)	190	(172-206)	0.0006
Glucose, mmol/L	5.2	(4.6-5.8)	5.2	(4.8-5.7)	5.2	(4.8-5.6)	0.0380

Abbreviation: BMI, body mass index; CKD, chronic kidney disease; CPRD, clinical practice research datalink; hs-CRP, high-sensitivity C reactive protein; DBP, diastolic blood pressure; SBP, systolic blood pressure; IQR, interquartile range; JUPITER, justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

[a] The logistic regression model used to predict the probability of being selected in the imputed cohort included the following variables: age, age^2, age^3, age^4, age^5, male, male×age, male×age^2, male×age^3, male×age^4, smoker, smoker×age, smoker×male, CKD, CKD×male, CKD×age, CKD×age^2, CKD×age^3, anti-hypertensive drugs, anti-hypertensive drugs×age, anti-hypertensive drugs×age^2, anti-hypertensive drugs×male, anti-hypertensive drugs×CKD, aspirin, age×aspirin, age^2×aspirin, aspirin×anti-hypertensive drugs, BMI, BMI^2, BMI×CKD, BMI×smoker, BMI×male, BMI^2×male, hsCRP, hsCRP^2, hsCRP^3, hsCRP^4, hsCRP×smoker, hsCRP×male, hsCRP^2×male, hsCRP^3×male, DBP, DBP×SBP, SBP, SBP^2, SBP^3, SBP×CKD, SBP^2×CKD, SBP^3×CKD, SBP×anti-hypertensive drugs, SBP^2×anti-hypertensive drugs, SBP^3×anti-hypertensive drugs, HDL-C, HDL-C^2, HDL-C^3, HDL-C^4, HDL-C^5,HDL-C×CKD, HDL-C×smoker, HDL-C×male, LDL-C, LDL-C^2, LDL-C^3, LDL-C^4, glucose, glucose×anti-hypertensive drugs, glucose×BMI, glucose×age, glucose×age^2, glucose×LDL-C, log(triglycerides), log(triglycerides)^2, log(triglycerides)^3, log(triglycerides)×HDL-C, and total cholesterol.

b We multiply imputed 20 datasets. We used one randomly selected imputed dataset to select variables included in the selection model and to examine balance in baseline characteristics between the target population and JUPITER. Thus, we presented the results from this randomly selected imputed dataset in the table.

[c] Absolute standardized difference was calculated between the weighted JUPITER trial and the target population using SAS Macro stddiff% (http://support.sas.com/resources/papers/proceedings12/335-2012.pdf).

[d] Moderate CKD was defined as the estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m^2. eGFR was calculated using serum creatinine values and the Modification of Diet in Renal Disease Study (MDRD) equation. Patients on dialysis or with eGFR <15 mL/min/1.73 m^2 were excluded from the study population.

Web Table 6. Baseline Characteristics among the Complete Case Cohort and the JUPITER trial, before and after weighting (no weight truncation)[a]

Characteristics	Complete Case Cohort		JUPITER Trial					
			Original Trial		Weighted Trial		Weighted Rosuvastatin Group		Weighted Placebo Group	
	n	(%)	n	(%)	n	(%)	n	(%)	n	(%)
No.	2,677	(100.0)	17,802	(100.0)	16,416	(100.0)	8,101	(100.0)	8,315	(100.0)
Male	1,549	(57.9)	11,001	(61.8)	9,560	(58.2)	4,666	(57.6)	4,895	(58.9)
Current Smoker	453	(16.9)	2,821	(15.8)	2,789	(17.0)	1,421	(17.5)	1,368	(16.5)
Anti-hypertensive drugs	1,515	(56.6)	8,846	(49.7)	9,254	(56.4)	4,570	(56.4)	4,684	(56.3)
Aspirin	322	(12.0)	3,313	(18.6)	1,991	(12.1)	977	(12.1)	1,014	(12.2)
Moderate CKD[b]	567	(21.2)	3,257	(18.3)	3,536	(21.5)	1,724	(21.3)	1,812	(21.8)
	Median	(IQR)	Median	(IQR)	Median	(IQR)	Median	(IQR)	Median	(IQR)
Age, years	63	(58-71)	66	(60-71)	63	(58-71)	64	(58-70)	63	(58-71)
BMI, kg/m^2	28.4	(24.9-32.3)	28.3	(25.3-32.0)	28.3	(25.2-32.0)	28.2	(25.2-32.0)	28.3	(25.2-32.0)
hs-CRP, mg/L	5	(3.0-7.0)	4.1	(2.8-6.5)	4.9	(3.0-7.3)	4.9	(3.0-7.3)	4.9	(3.0-7.3)
DBP, mmHg	80	(73-85)	80	(75-87)	80	(72-86)	80	(73-86)	80	(72-85)
SBP, mmHg	136	(125-144)	134	(124-145)	135	(126-145)	135	(126-146)	134	(125-144)
HDL-C, mg/dL	51	(43-62)	49	(40-60)	51	(42-63)	51	(42-63)	51	(42-62)
LDL-C, mg/dL	112	(101-120)	108	(94-119)	112	(99-121)	112	(99-122)	112	(98-121)
Triglycerides, mg/dL	126	(104-168)	118	(85-169)	125	(105-166)	125	(106-165)	124	(105-168)
Total Cholesterol, mg/dL	191	(178-205)	185	(169-200)	192	(177-206)	193	(178-206)	191	(176-205)
Glucose, mmol/L	5.1	(4.8-5.5)	5.2	(4.8-5.7)	5.2	(4.8-5.6)	5.2	(4.8-5.6)	5.2	(4.8-5.6)

Abbreviation: BMI, body mass index; CKD, chronic kidney disease; CPRD, clinical practice research datalink; hs-CRP, high-sensitivity C reactive protein; DBP, diastolic blood pressure; SBP, systolic blood pressure; IQR, interquartile range; JUPITER, justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

[a] The logistic regression model used to predict the probability of being selected in the imputed cohort included the following variables: age, age^2, male, male×age, male×age^2, smoker, smoker×age, smoker×male, CKD, CKD×male, CKD×age, CKD×age^2, anti-hypertensive drugs, aspirin, aspirin×anti-hypertensive drugs, BMI, BMI^2, hsCRP, hsCRP^2, hsCRP^3, hsCRP^4, hsCRP^5, hsCRP^6, hsCRP^7, DBP, SBP, SBP^2, SBP^3, SBP×anti-hypertensive drugs, HDL-C, HDL-C×smoker, LDL-C, LDL-C×male, glucose, log(triglycerides), log(triglycerides)^2, log(triglycerides)^3, log(triglycerides)^4, and total cholesterol.

[b] Moderate CKD was defined as the estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m^2. eGFR was calculated using serum creatinine values and the Modification of Diet in Renal Disease Study (MDRD) equation. Patients on dialysis or with eGFR <15 mL/min/1.73 m^2 were excluded from the study population.

Web Table 7. Baseline Characteristics among the Complete Case Cohort and the JUPITER trial, before and after weighting (weight truncation at 20) [a]

Characteristics	Complete Case Cohort		JUPITER Trial				Absolute Standardized Difference[b]
			Original Trial		Weighted Trial		
	n	(%)	n	(%)	n	(%)	
No.	2,677	(100.0)	17,802	(100.0)	16,235	(100.0)	
Male	1,549	(57.9)	11,001	(61.8)	9,465	(58.3)	-0.0088
Current Smoker	453	(16.9)	2,821	(15.8)	2,747	(16.9)	0.0000
Anti-hypertensive drugs	1,515	(56.6)	8,846	(49.7)	9,208	(56.7)	-0.0025
Aspirin	322	(12.0)	3,313	(18.6)	1,978	(12.2)	-0.0048
Moderate CKD[c]	567	(21.2)	3,257	(18.3)	3,418	(21.1)	0.0031
	Median	(IQR)	Median	(IQR)	Median	(IQR)	
Age, years	63	(58-71)	66	(60-71)	63	(58-70)	0.0100
BMI, kg/m^2	28.4	(24.9-32.3)	28.3	(25.3-32.0)	28.3	(25.2-32.0)	-0.0108
hs-CRP, mg/L	5.0	(3.0-7.0)	4.1	(2.8-6.5)	4.9	(3.0-7.3)	-0.0348
DBP, mmHg	80	(73-85)	80	(75-87)	80	(73-86)	-0.0240
SBP, mmHg	136	(125-144)	134	(124-145)	135	(126-145)	-0.0031
HDL-C, mg/dL	51	(43-62)	49	(40-60)	51	(42-62)	0.0248
LDL-C, mg/dL	112	(101-120)	108	(94-119)	112	(99-121)	0.0014
Triglycerides, mg/dL	126	(104-168)	118	(85-169)	125	(105-166)	-0.0105
Total Cholesterol, mg/dL	191	(178 -205)	185	(169-200)	192	(177-206)	-0.0217
Glucose, mmol/L	5.1	(4.8-5.5)	5.2	(4.8-5.7)	5.2	(4.8-5.6)	-0.0377

Abbreviation: BMI, body mass index; CKD, chronic kidney disease; CPRD, clinical practice research datalink; hs-CRP, high-sensitivity C reactive protein; DBP, diastolic blood pressure; SBP, systolic blood pressure; IQR, interquartile range; JUPITER, justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin; LDL-C, low-density lipoprotein cholesterol; HDL-C, high-density lipoprotein cholesterol.

[a] The logistic regression model used to predict the probability of being selected in the imputed cohort included the following variables: age, age^2, male, male×age, male×age^2, smoker, smoker×age, smoker×male, CKD, CKD×male, CKD×age, CKD×age^2, anti-hypertensive drugs, aspirin, aspirin×anti-hypertensive drugs, BMI, BMI^2, hsCRP, hsCRP^2, hsCRP^3, hsCRP^4, hsCRP^5, hsCRP^6, hsCRP^7, DBP, SBP, SBP^2, SBP^3, SBP×anti-hypertensive drugs, HDL-C, HDL-C×smoker, LDL-C, LDL-C×male, glucose, log(triglycerides), log(triglycerides)^2, log(triglycerides)^3, log(triglycerides)^4, and total cholesterol.

[b] Absolute standardized difference was calculated between the weighted JUPITER trial and the target population using SAS Macro stddiff% (http://support.sas.com/resources/papers/proceedings12/335-2012.pdf).

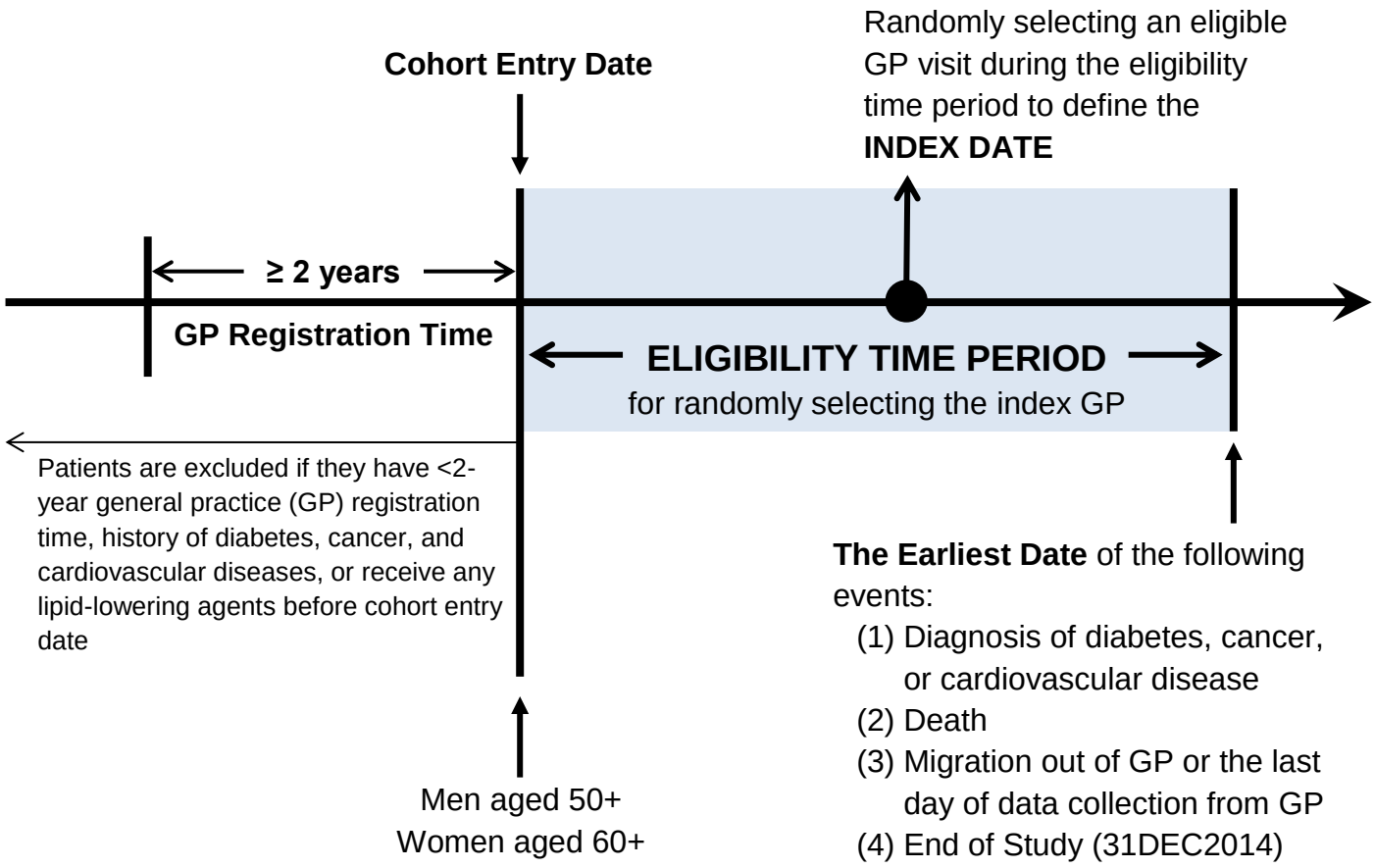
[c] Moderate CKD was defined as the estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m^2. eGFR was calculated using serum creatinine values and the Modification of Diet in Renal Disease Study (MDRD) equation. Patients on dialysis or with eGFR <15 mL/min/1.73 m^2 were excluded from the study population.

Web Table 8. Treatment Effect (Hazard Ratio (HR), 95% Confidence Interval (CI)) of Rosuvastatin in the JUPITER trial after standardized to the imputed cohort with and without weight truncated at 20, across 20 imputed datasets

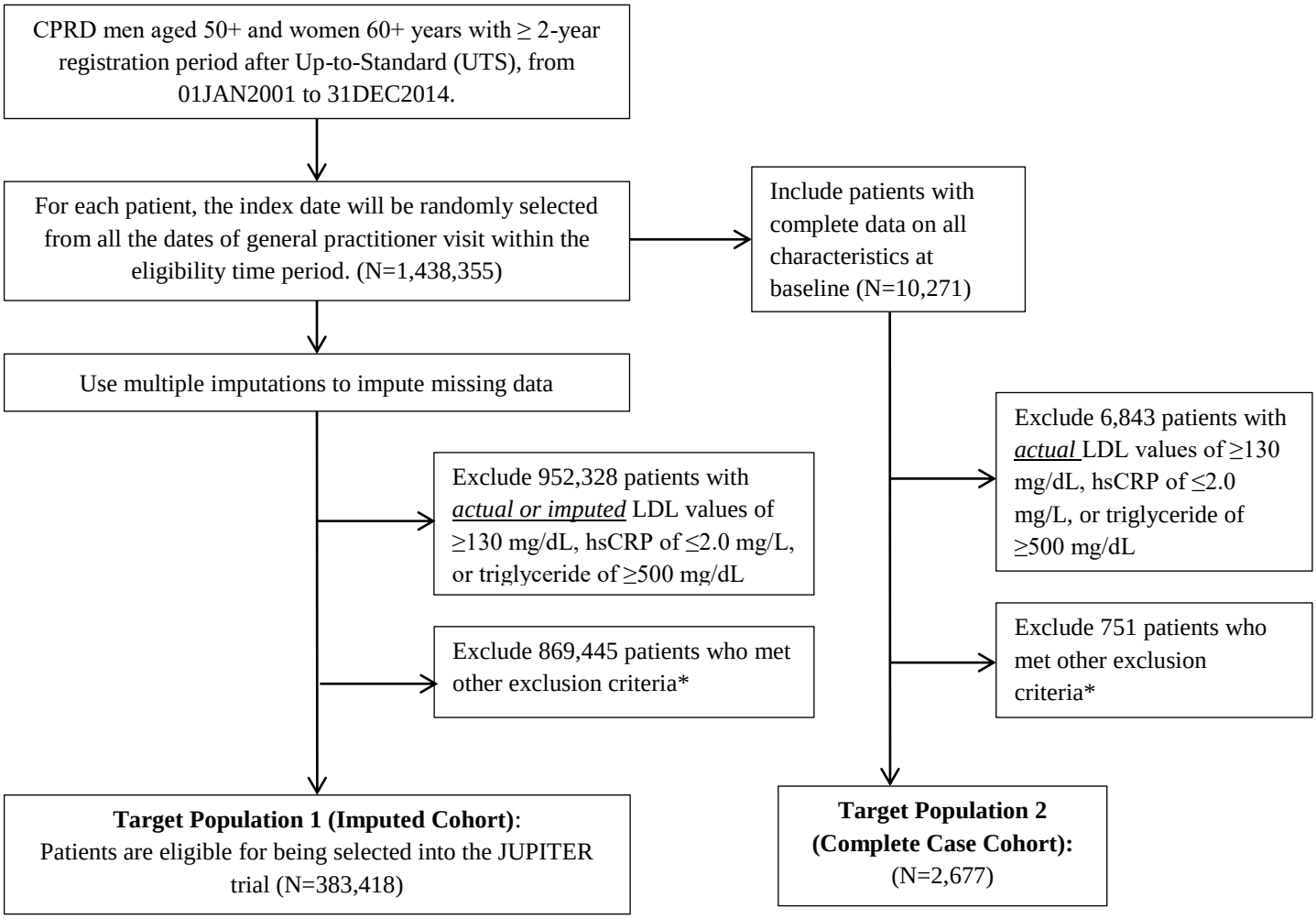
Imputed Dataset	Untruncated		Weight truncated at 20	
	HR	(95% CI)	HR	(95% CI)
1	0.75	(0.45, 1.25)	0.71	(0.45, 1.11)
2	0.74	(0.44, 1.23)	0.70	(0.45, 1.10)
3	0.74	(0.45, 1.23)	0.71	(0.45, 1.11)
4	0.76	(0.45, 1.29)	0.71	(0.45, 1.11)
5	0.74	(0.44, 1.24)	0.71	(0.45, 1.11)
6	0.74	(0.44, 1.24)	0.71	(0.45, 1.11)
7	0.74	(0.45, 1.23)	0.70	(0.45, 1.10)
8	0.74	(0.45, 1.23)	0.70	(0.45, 1.10)
9	0.74	(0.45, 1.24)	0.70	(0.45, 1.10)
10	0.75	(0.44, 1.26)	0.70	(0.45, 1.10)
11	0.75	(0.45, 1.26)	0.71	(0.45, 1.11)
12	0.74	(0.44, 1.25)	0.71	(0.45, 1.11)
13	0.75	(0.45, 1.27)	0.71	(0.45, 1.11)
14	0.75	(0.45, 1.26)	0.71	(0.45, 1.11)
15	0.75	(0.45, 1.25)	0.70	(0.45, 1.10)
16	0.75	(0.45, 1.25)	0.71	(0.45, 1.11)
17	0.75	(0.45, 1.25)	0.71	(0.45, 1.11)
18	0.73	(0.44, 1.22)	0.70	(0.45, 1.10)
19	0.75	(0.45, 1.25)	0.70	(0.45, 1.10)
20	0.73	(0.44, 1.21)	0.70	(0.45, 1.10)

Web Table 9. Treatment Effect (Hazard Ratio (HR), 95% Confidence Interval (CI)) of Rosuvastatin in the JUPITER trial after standardized to the imputed cohort among patients aged younger than 80 years, with and without weight truncated at 20, across 20 imputed datasets

Imputed Dataset	Untruncated		Weight truncated at 20	
	HR	(95% CI)	HR	(95% CI)
1	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
2	0.55	(0.39, 0.77)	0.55	(0.39, 0.77)
3	0.56	(0.40, 0.78)	0.55	(0.40, 0.77)
4	0.55	(0.39, 0.77)	0.55	(0.39, 0.76)
5	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
6	0.56	(0.40, 0.78)	0.55	(0.40, 0.77)
7	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
8	0.55	(0.39, 0.77)	0.55	(0.39, 0.76)
9	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
10	0.55	(0.40, 0.78)	0.55	(0.39, 0.77)
11	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
12	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
13	0.55	(0.40, 0.78)	0.55	(0.39, 0.77)
14	0.56	(0.40, 0.78)	0.55	(0.40, 0.77)
15	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
16	0.55	(0.40, 0.77)	0.55	(0.40, 0.77)
17	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
18	0.55	(0.40, 0.77)	0.55	(0.39, 0.77)
19	0.55	(0.39, 0.77)	0.55	(0.39, 0.76)
20	0.56	(0.40, 0.78)	0.55	(0.40, 0.77)



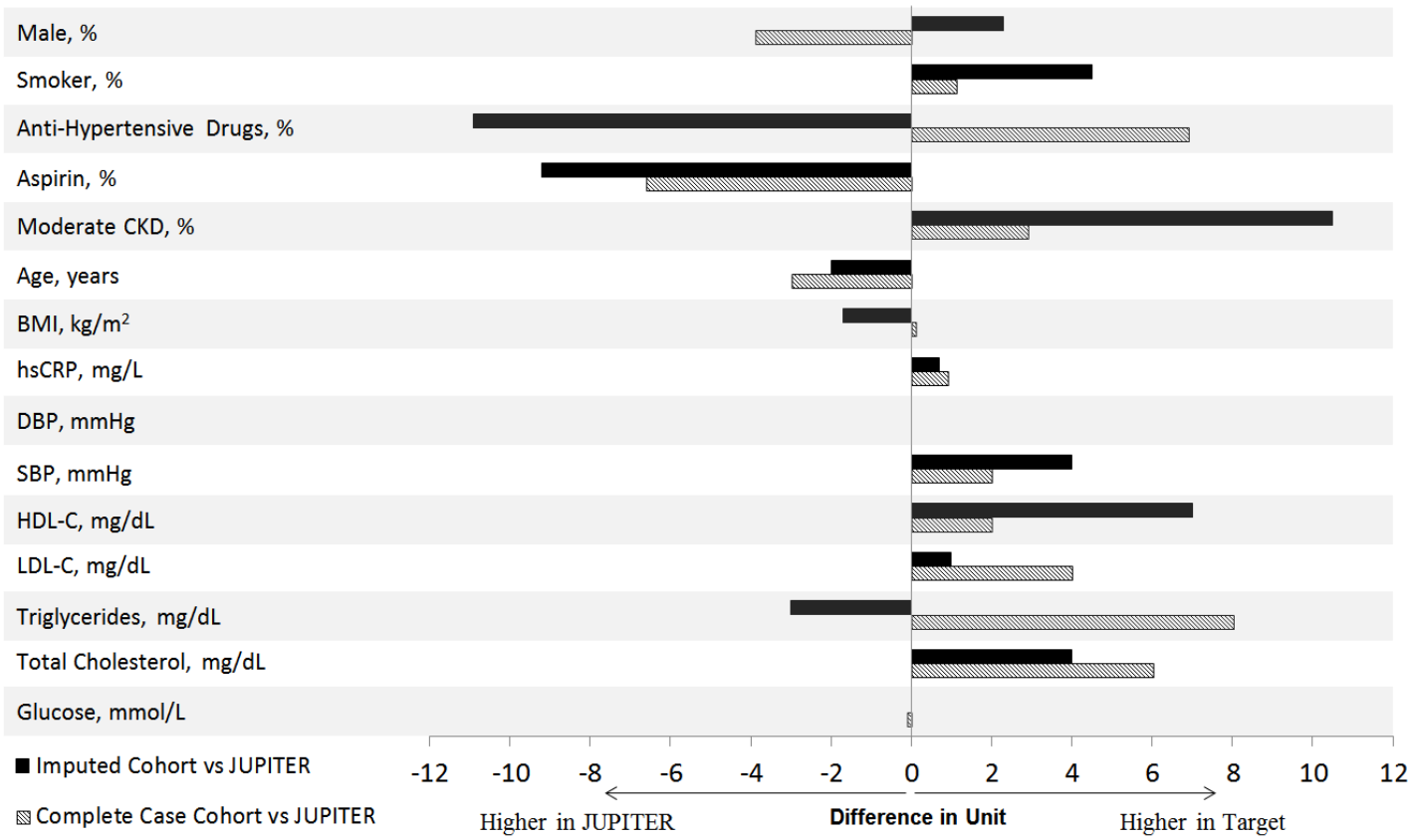
Web Figure 1. Illustration of concept of selecting the target population from the CPRD



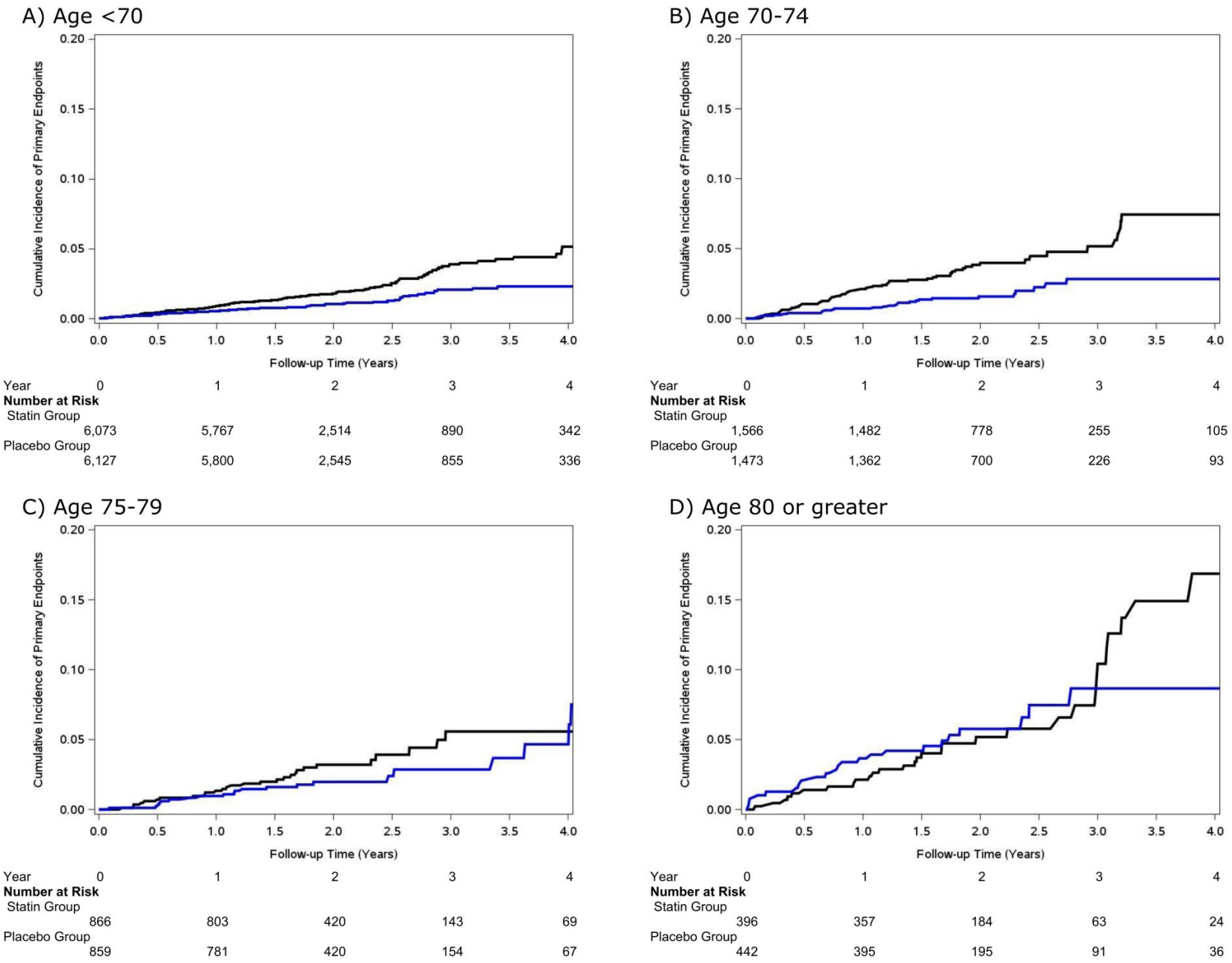
*Other exclusion criteria included:

(1) Females using hormone replacement therapy
(2) Diagnosis of drug or alcohol abuse
(3) Use of immunosuppressant agents or glucocorticoids
(4) Diagnosis of Inflammatory conditions such as severe arthritis, lupus, or inflammatory bowel disease
(5) Diagnosis of severe mental conditions such as schizophrenia
(6) DBP >100mmHg or SBP >190mmHg
(7) Diagnosis of liver diseases
(8) Alanine aminotransferase level that was more than twice the upper limit of the normal range
(9) Uncontrolled hypothyroidism
(10) A creatine kinase level that was more than three times the upper limit of the normal range
(11) CRP >20 mg/L within 90 days prior to the index date, indicating recent severe bacterial infection
(12) Serum Creatinine >2 mg/dL, diagnosis of ESRD, or on dialysis

Web Figure 2. The flowchart of the target population selection



Web Figure 3. Comparison of Baseline Characteristics in Two Target Population and the JUPITER trial. Difference was calculated as in percentage for dichotomous variables and in median for continuous variables between a target population and the JUPITER trial.



Web Figure 4. Plots of cumulative incidence for primary endpoints in the JUPITER trial, stratified by treatment (blue line for rosuvastatin and black line for placebo) and age groups