Practice of Epidemiology

Generalizing Randomized Clinical Trial Results: Implementation and Challenges Related to Missing Data in the Target Population

Jin-Liern Hong, Michele Jonsson Funk, Robert LoCasale, Sara E. Dempster, Stephen R. Cole, Michael Webster-Clark, Jessie K. Edwards, and Til Stürmer*

* Correspondence to Dr. Til Stürmer, Department of Epidemiology, UNC Gillings School of Global Public Health, University of North Carolina at Chapel Hill, 135 Dauer Drive, 2101 McGavran-Greenberg Hall, CB #7435, Chapel Hill, NC 27599-7435 (e-mail: til.sturmer@post.harvard.edu).

Statins are indicated in patients with elevated levels of high-sensitivity C-reactive protein and normal low-density lipoprotein cholesterol based on results of the multicountry trial, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) (2003–2008), but the benefit in real-world populations remains unknown. We sought to generalize JUPITER results to trial-eligible population using data from the UK Clinical Practice Research Datalink (CPRD), 2001–2014. We multiply imputed missing baseline characteristics for the CPRD population and selected the trial-eligible participants as the target population based on observed and imputed values. Trial participants were weighted to be representative of the CPRD population ($n = 383,418$) based on individual predicted probability of selection into the trial. Trial participants were also standardized to the CPRD population without missing values ($n = 2,677$). In JUPITER, rosuvastatin reduced cardiovascular risk with a 3-year risk difference of −2.0% (95% confidence interval (CI): −2.9, −1.1). The rosuvastatin effect was muted in the first 2 years but remained strong at 3 years after standardizing to the imputed CPRD population (3-year risk difference = −2.7%; 95% CI: −5.8, 0.4) and the CPRD population without missing data (3-year risk difference = −1.7%; 95% CI: −3.5, 0.1). The study serves as an illustration of possible approaches to understanding generalizability of trials using real-world databases given limitations due to missing data on inclusion/exclusion criteria.

cardiovascular diseases; generalizability; JUPITER trial; randomized clinical trial; statins

Abbreviations: BMI, body mass index; CI, confidence interval; CPRD, Clinical Practice Research Datalink; CVD, cardiovascular disease; HDL-C, high-density lipoprotein cholesterol; hs-CRP, high-sensitivity C-reactive protein; JUPITER, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin; LDL-C, low-density lipoprotein cholesterol; RCT, randomized clinical trial.

Statins (3-hydroxy-3-methylglutaryl coenzyme A reductase inhibitors) are widely prescribed for the primary prevention of cardiovascular disease (CVD) (1). Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) showed that rosuvastatin effectively reduced the risk of CVD in individuals not at high risk for CVD because of elevated low-density lipoprotein cholesterol (LDL-C) but with elevated levels of high-sensitivity C-reactive protein (hs-CRP) (2). As a well-designed and properly conducted randomized clinical trial (RCT), JUPITER has strong internal validity. However, the external validity of the JUPITER trial has been called into question because the subjects enrolled in the JUPITER trial may not represent the real-world population of interest (3). The heterogeneous effect of rosuvastatin in the JUPITER trial (4–6) and the general selection of trial participants would need to be taken into account when generalizing the expected benefit of rosuvastatin to a representative population of real-world patients who meet JUPITER inclusion criteria.

To generalize RCT results to real-world populations, we need to standardize the distribution of factors affecting treatment efficacy in the RCT population to that observed in the real-world population. With multiple effect modifiers, standardization quickly becomes intractable. Cole and Stuart (7) therefore proposed a weighting approach based on predicting the

individual probability of being in the trial versus the target population as a function of multiple factors and to reweight the RCT participants to be representative of the target population. This idea is a generalization of the same concept as propensity score weighting to control for confounding (8, 9).

While compelling in theory, in practice, data on RCT inclusion/exclusion criteria and effect modifiers are likely to be missing in a portion of the target population, because measures for systemic inflammation (such as hs-CRP) and for chronic kidney disease are not routinely recorded in existing real-world data. Physicians are more likely to order these tests for patients at higher risks for CVD. This leads to a selection of the study population representing the target population. Ignoring individuals in the target population with missing values on effect modifiers could thus result in the wrong target population.

The objective of this study was to estimate the effect of rosuvastatin initiation on the risk of CVD events in the UK population who would have been eligible for the JUPITER trial, as represented in the Clinical Practice Research Datalink (CPRD) population (10), by generalizing the results of the JUPITER trial to the target population and addressing both selection into the trial and issues related to the identification of a study population that represents the target population.

METHODS

Study population

The JUPITER trial (ClinicalTrials.gov identifier: NCT00239681) was a randomized, double-blind, placebo-controlled trial of rosuvastatin (20 mg daily) for the primary prevention of CVD events, enrolling 17,802 patients with subclinical systemic inflammation but without hypercholesterolemia from 26 countries between 2003 and 2008 (2). Briefly, eligible patients were men aged ≥ 50 years or women aged ≥ 60 years who had elevated hs-CRP levels but did not have hyperlipidemia (i.e., hs-CRP ≥ 2.0 mg/L, LDL-C <130 mg/dL, and triglycerides <500 mg/dL). The exclusion criteria included a history of CVD, diabetes, and cancer. The primary endpoint was the occurrence of a first major CVD event, defined as nonfatal myocardial infarction, nonfatal stroke, hospitalization for unstable angina, arterial revascularization procedure, or confirmed death from cardiovascular causes.

The target population that we chose in this study consisted of patients in the CPRD who would have been eligible to participate in the JUPITER trial. The concept of the algorithm used to identify the eligible patients is described as follows and in Web Figure 1 (available at https://academic.oup.com/aje). The flow chart of study cohort selection is included in Web Figure 2.

We selected all men aged ≥ 50 years and women aged ≥ 60 years with ≥ 2 years of registration with a general practitioner after the practice came "Up-to-Standard" (i.e., the practice data met the CPRD's quality criteria), during the study period from January 1, 2001, through December 31, 2014. The date of meeting both age and registration-time requirements was defined as cohort entry date. Patients were required to have no history of CVD, diabetes, or cancer and to have received no prescriptions of any lipid-lowering drugs before or on the cohort entry date. CVD and cancer events were defined as 1 diagnosis Read code,

and diabetes was defined by a diagnosis Read code, receipt of any antihyperglycemic drug, or abnormal glucose or hemoglobin A1C test results (fasting plasma glucose ≥ 7.0 mmol/L or hemoglobin A1C of $\geq 6.5\%$). The eligibility time period was defined from the cohort entry date to the earliest date of the following events: diagnosis of CVD, diabetes, or cancer (except for basal-cell or squamous-cell carcinoma of the skin); initiation of any lipid-lowering agents; death; the last day of data collection from the general practice; migration out of the general practice; or the end of study (December 31, 2014). For each patient, we randomly selected an index date from all the dates of general practitioner visits within the eligibility time period. General practitioner visit was defined as having a medical record with consultation types of Clinic or Surgery Consultation visit from the Clinical Table in the CPRD.

Covariates and missing data

In the CPRD, the baseline characteristics and laboratory measurements were assessed for each patient, using data within 24 months prior to the index date, including demographic information, body mass index (BMI), tobacco smoking status, laboratory tests, medical history, medication use, and indicators of health-care use. To account for possible data-entry errors in the CPRD and JUPITER, we applied trimming rules for all laboratory tests and measurement results, and we made further considerations for hs-CRP values given that underlying hs-CRP values could be masked by infection, both of which we outline in detail in Web Appendix 1.

We imputed missing baseline characteristics and laboratory measurements among the eligible CPRD patients, using multiple imputation (11–15). Multiple imputation relies on the assumption that data are missing at random. This means, for example, that missingness of hs-CRP may depend on observed patient characteristics but does not depend on the (unknown) actual value of the hs-CRP or other unobserved data within strata of measured predictors. We used a multivariate normal imputation model fit by Markov Chain Monte Carlo simulation to impute all missing data in 20 imputed data sets. Each imputed data set was then analyzed separately, and the results from the 20 analyses were combined using Rubin's formula to give overall estimates while accounting for the variability across the imputed data sets (14).

After imputing missing baseline characteristics for all CPRD patients, we selected the patients who satisfied the trial criteria based on actual or imputed values (e.g., LDL-C value of <130 mg/dL, hs-CRP value of ≥ 2.0 mg/L, and triglyceride value of <500 mg/dL) in each imputed data set. These cohorts were called the imputed cohorts, the main study populations to represent the target. To assess the potential selection of the target population due to missing values, we examined another study population consisting of patients with complete data only. From all eligible CPRD patients, we selected only patients who had complete data on BMI, smoking status, LDL-C, high-density lipoprotein cholesterol (HDL-C), triglycerides, total cholesterol, systolic and diastolic blood pressure, hs-CRP, serum creatinine, and glucose at the index date and then included those who met the inclusion criteria of the JUPITER trial based on actual laboratory measurements. Following the usual terminology in

Table 1.   Baseline Characteristics Before and After Multiply Imputing Missing Values Among Clinical Practice Research Datalink Participants (United Kingdom, 2001–2014) Before Applying the Inclusion/Exclusion Criteria for "Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin" (Multiple Countries, 2003–2008)

| Variable[a] | Original CPRD Population (n = 1,438,355) | | | After Multiple Imputation[b,c] | |
| | Distribution | | Missing | Distribution | |
	%	Median (IQR)	%	%	Median (IQR)
Age, years		63 (57–72)	0		63 (57–72)
Male sex	57.6		0	57.6	
BMI[d]		26.8 (23.9–30.3)	57.6		26.7 (23.6–30.1)
Current smoker	21.0		35.2	18.0	
hs-CRP, mg/L		4.0 (2.0–7.0)	88.2		3.8 (2.0–6.8)
DBP, mm Hg		80 (74–87)	23.3		80 (74–87)
SBP, mm Hg		139 (128–149)	23.3		138 (127–149)
HDL-C, mg/dL		54 (46–66)	65.1		56 (46–68)
LDL-C, mg/dL		135 (112–159)	70.0		135 (112–159)
Triglycerides, mg/dL		124 (91–169)	68.2		121 (89–168)
Total cholesterol, mg/dL		218 (193–244)	57.8		219 (193–245)
Glucose, mmol/L		5.1 (4.8–5.6)	87.1		5.3 (4.6–6.0)
Serum Creatinine, mg/dL		1.0 (0.8–1.1)	45.6		1.0 (0.8–1.1)

Abbreviations: BMI, body mass index; CPRD, Clinical Practice Research Datalink; DBP, diastolic blood pressure; HDL-C, high-density lipoprotein cholesterol; hs-CRP, high-sensitivity C reactive protein; IQR, interquartile range; JUPITER, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure.

[a] The variables chosen for multiple imputation and for the model of selection into the trial were based on their high relevance to cardiovascular diseases. Several laboratory tests were used in the exclusion criteria of JUPITER but were not selected here because of large missingness in the CPRD and not being strong predictors for cardiovascular disease.

[b] The distributions of the variables were very similar across the 20 imputed data sets. Thus, we presented the results from a randomly selected imputed data set in the table.

[c] The imputation model included the following variables: age, male sex, BMI, weight within 2 years prior to the index date, previous weight recorded before 2 years before the index date, tobacco smoking status within 2 years prior to the index date, prior smoking status recorded before 2 years before the index date, diagnosis related to smoking, smoking cessation prescriptions, diagnosis of alcohol abuse, DBP, SBP, LDL-C, HDL-C, log transformation of triglycerides, total cholesterol, log transformation of hs-CRP, glucose, serum creatinine, use of any nonsteroidal antiinflammatory drugs, use of angiotensin-converting-enzyme inhibitor, use of angiotensin II receptor blockers, use of beta blockers, use of loop or nonloop diuretics, use of calcium channel blockers, diagnosis of renal disease, calendar year, and number of general practitioner encounters. In addition, occurrence of major cardiovascular disease and rosuvastatin initiation were prospectively assessed within 2 years after the index date and were included in the multiple imputation model to improve the performance of imputation.

[d] Weight (kg)/height (m)^2.

missing data settings, we refer to this selected study population as the complete case cohort.

Statistical analysis

To generalize the JUPITER trial results to the target population, we estimated the probability of selection into JUPITER by combining JUPITER and the target population into a single data set. Given the large size of the JUPITER trial and the target population, we estimated the probability of selection into JUPITER using a logistic regression model that included all factors potentially affecting selection and higher-order product terms for the joint distribution (7). The higher-order and interaction terms included in the logistic model were selected to achieve marginal balance of covariates between the trial and target populations. Next, we used the standardized mortality or morbidity ratio method to assign selection weights to each JUPITER trial participant depending on their individual characteristics. These selection weights were calculated as the odds of 1 minus the predicted probability of selection into the trial and scaled by multiplying with the (marginal) odds of being in the trial (9). Within the reweighted JUPITER trial, we then used Cox proportional hazard models (with robust variance estimator) to estimate hazard ratios and their 95% confidence intervals for the comparison of major cardiovascular events between those randomized to rosuvastatin versus placebo. We calculated the weighted risk difference using a nonparametric method accounting for non-CVD competing causes of death, and we obtained standard deviations from 200 bootstrap data sets to calculate 95% confidence intervals (16). To

Table 2. Baseline Characteristics Before and After Selection Weighting[a,b] Among the Imputed Cohort From the Clinical Practice Research Datalink (United Kingdom, 2001–2014) and the Participants from "Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin" (Multiple Countries, 2003–2008)

Characteristic	Imputed Cohort[c] (n = 383,418)		JUPITER Trial							
			Original Trial (n = 17,802)		Weighted Trial[d] (n = 16,165)		Weighted Rosuvastatin Group (n = 8,015)		Weighted Placebo Group (n = 8,150)	
	%	Median (IQR)	%	Median (IQR)	%	Median (IQR)	%	Median (IQR)	%	Median (IQR)
Age, years		64 (57–74)		66 (60–71)		65 (57–74)		65 (57–74)		64 (57–74)
Male sex	64.1		61.8		63.3		63.2		63.4	
BMI[e]		26.6 (23.5–30.1)		28.3 (25.3–32.0)		26.5 (23.6–29.8)		26.6 (23.6–30.0)		26.5 (23.6–29.7)
Current smoker	20.3		15.8		20.5		20.3		20.7	
Antihypertensive drugs	38.8		49.7		39.9		39.9		39.9	
Aspirin	9.4		18.6		9.5		9.0		10.1	
Moderate CKD[f]	28.8		18.3		26.3		26.9		25.7	
hs-CRP, mg/L		4.8 (3.1–7.6)		4.1 (2.8–6.5)		4.8 (3.1–7.7)		4.8 (3.1–7.6)		4.8 (3.1–7.7)
DBP, mm Hg		80 (74–86)		80 (75–87)		80 (73–86)		80 (74–87)		80 (73–86)
SBP, mm Hg		138 (126–147)		134 (124–145)		138 (127–150)		139 (128–150)		136 (126–148)
HDL-C, mg/dL		56 (45–68)		49 (40–60)		57 (45–68)		57 (45–69)		56 (45–68)
LDL-C, mg/dL		109 (93–120)		108 (94–119)		109 (93–120)		108 (93–120)		109 (92–120)
Triglycerides, mg/dL		115 (83–159)		118 (85–169)		113 (81–161)		114 (80–158)		113 (82–163)
Total cholesterol, mg/dL		189 (172–206)		185 (169–200)		190 (172–206)		191 (173–206)		190 (172–205)
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)

Abbreviations: BMI, body mass index; CKD, chronic kidney disease; CPRD, Clinical Practice Research Datalink; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HDL-C, high-density lipoprotein cholesterol; hs-CRP, high-sensitivity C reactive protein; IQR, interquartile range; JUPITER, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure.

[a] Weights were truncated at 20. We chose our truncation weight of 20 based on comparing the distribution of covariates between the trial-eligible CPRD population and the weighted trial population as well as the variance of the treatment effect estimate across various levels of truncation. The results without weight truncation are presented in Web Table 4.

[b] 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, antihypertensive drugs, antihypertensive drugs × age, antihypertensive drugs × age^2, antihypertensive drugs × male, antihypertensive drugs × CKD, aspirin, age × aspirin, age^2 × aspirin, aspirin × antihypertensive 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 × antihypertensive drugs, SBP^2 × antihypertensive drugs, SBP^3 × antihypertensive 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 × antihypertensive 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.

[c] We multiply imputed 20 data sets. We used one randomly selected imputed data set 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 data set in the table.

[d] Absolute standardized differences in baseline characteristics between the weighted JUPITER trial and the target population are presented in Web Table 5 and were less than 5%. We calculated absolute standardized differences using SAS (SAS Institute, Inc.) Macro stddiff% (29).

[e] Weight (kg)/height $(m)^2$.

[f] Moderate CKD was defined as the eGFR <60 mL/minute/1.73 m^2. eGFR was calculated using serum creatinine values and the Modification of Diet in Renal Disease Study equation. Patients on dialysis or with eGFR <15 mL/minute/1.73 m^2 were excluded from the study population.

Table 3. Baseline Characteristics Before and After Selection Weighting[a,b] Among the Complete Case Cohort From the Clinical Practice Research Datalink (United Kingdom, 2001–2014) and the Participants From "Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin" (Multiple Countries, 2003–2008)

Characteristic	Complete Case Cohort (n = 2,677)		JUPITER Trial							
			Original Trial (n = 17,802)		Weighted Trial[c] (n = 16,235)		Weighted Rosuvastatin Group (n = 8,069)		Weighted Placebo Group (n = 8,166)	
	%	Median (IQR)	%	Median (IQR)	%	Median (IQR)	%	Median (IQR)	%	Median (IQR)
Age, years		63 (58–71)		66 (60–71)		63 (58–70)		64 (58–70)		63 (58–70)
Male sex	57.9		61.8		58.3		57.6		59.0	
BMI[d]		28.4 (24.9–32.3)		28.3 (25.3–32.0)		28.3 (25.2–32.0)		28.3 (25.2–32.0)		28.4 (25.2–32.0)
Current smoker	16.9		15.8		16.9		17.6		16.3	
Antihypertensive drugs	56.6		49.7		56.7		56.3		57.1	
Aspirin	12.0		18.6		12.2		12.0		12.4	
Moderate CKD[e]	21.2		18.3		21.1		21.2		20.9	
hs-CRP, mg/L		5.0 (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, mm Hg		80 (73–85)		80 (75–87)		80 (73–86)		80 (73–86)		80 (73–86)
SBP, mm Hg		136 (125–144)		134 (124–145)		135 (126–145)		135 (126–146)		134 (125–145)
HDL-C, mg/dL		51 (43–62)		49 (40–60)		51 (42–62)		51 (42–63)		50 (42–62)
LDL-C, mg/dL		112 (101–120)		108 (94–119)		112 (99–121)		112 (99–122)		112 (99–121)
Triglycerides, mg/dL		126 (104–168)		118 (85–169)		125 (105–166)		125 (106–165)		125 (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)

Abbreviations: BMI, body mass index; CKD, chronic kidney disease; CPRD, Clinical Practice Research Datalink; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; HDL-C, high-density lipoprotein cholesterol; hs-CRP, high-sensitivity C reactive protein; IQR, interquartile range; JUPITER, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin; LDL-C, low-density lipoprotein cholesterol; SBP, systolic blood pressure.
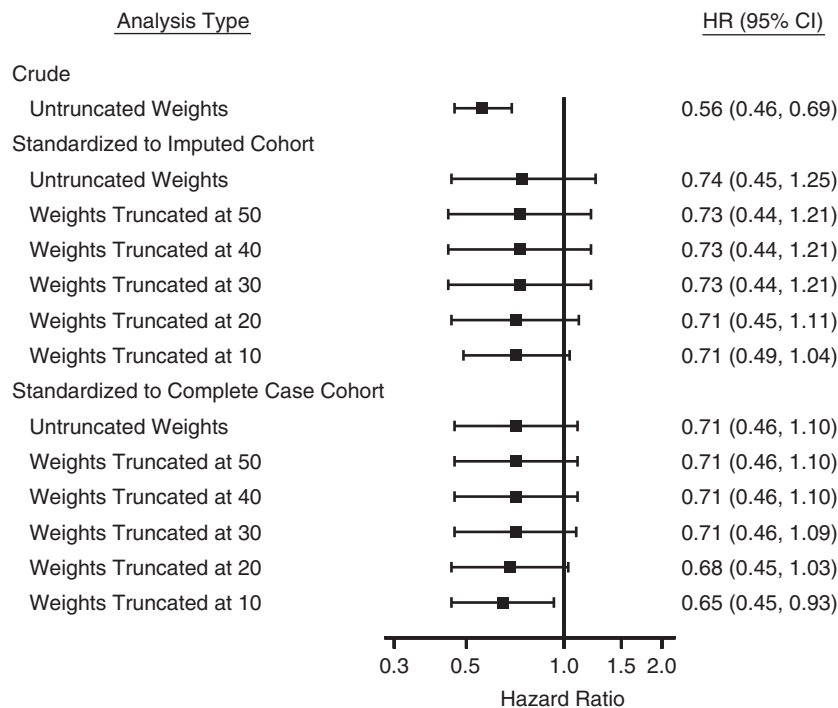
[a] Weights were truncated at 20. We chose our truncation weight of 20 based on comparing the distribution of covariates between the trial-eligible CPRD population and the weighted trial population as well as the variance of the treatment effect estimate across various levels of truncation. The results without weight truncation are presented in Web Table 6.

[b] The logistic regression model used to predict the probability of being selected in the complete case 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, antihypertensive drugs, aspirin, aspirin × antihypertensive drugs, BMI, BMI^2, hsCRP, $hsCRP^2$, $hsCRP^3$, $hsCRP^4$, $hsCRP^5$, $hsCRP^6$, $hsCRP^7$, DBP, SBP, SBP^2, SBP^3, SBP × antihypertensive 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.

[c] Absolute standardized differences in baseline characteristics between the weighted JUPITER trial and the target population are presented in Web Table 7 and were less than 5%. We calculated absolute standardized differences using SAS (SAS Institute, Inc.) Macro stddiff% (29).

[d] Weight (kg)/height $(m)^2$.

[e] Moderate CKD was defined as the eGFR <60 mL/minute/1.73 m^2. eGFR was calculated using serum creatinine values and the Modification of Diet in Renal Disease Study equation. Patients on dialysis or with eGFR <15 mL/minute/1.73 m^2 were excluded from the study population.

Analysis Type	HR (95% CI)
Crude	
Untruncated Weights	0.56 (0.46, 0.69)
Standardized to Imputed Cohort	
Untruncated Weights	0.74 (0.45, 1.25)
Weights Truncated at 50	0.73 (0.44, 1.21)
Weights Truncated at 40	0.73 (0.44, 1.21)
Weights Truncated at 30	0.73 (0.44, 1.21)
Weights Truncated at 20	0.71 (0.45, 1.11)
Weights Truncated at 10	0.71 (0.49, 1.04)
Standardized to Complete Case Cohort	
Untruncated Weights	0.71 (0.46, 1.10)
Weights Truncated at 50	0.71 (0.46, 1.10)
Weights Truncated at 40	0.71 (0.46, 1.10)
Weights Truncated at 30	0.71 (0.46, 1.09)
Weights Truncated at 20	0.68 (0.45, 1.03)
Weights Truncated at 10	0.65 (0.45, 0.93)

Hazard Ratio: 0.3 0.5 1.0 1.5 2.0

Figure 1. Treatment effect (hazard ratio (HR) and 95% confidence interval (CI)) of rosuvastatin in the multiple-country trial, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) (2003–2008), before and after standardizing to the imputed cohort and the complete case cohort with selection weight truncated at 20. For the results from JUPITER standardized to the imputed cohort, the hazard ratios were first estimated in JUPITER standardized to each imputed data set (Web Table 8) and then combined into overall estimates to account for variability.

reduce the influence of large weights and to increase the precision of the estimate, we implemented weight truncation of predicted probabilities (17). Because only 806 of JUPITER participants were aged 80 years or older, we also reestimated the generalized treatment effect among patients less than 80 years of age.

All statistical analyses were performed with SAS, version 9.3 (SAS Institute, Inc., Cary, North Carolina). This study was approved by the Institutional Review Board at the University of North Carolina at Chapel Hill and by the Independent Scientific Advisory Committee for Medicines and Healthcare products Regulatory Agency database research in the United Kingdom.

RESULTS

In the CPRD, we identified a total of 1,438,355 men aged ≥50 years and women aged ≥60 years who had no medical history of diabetes, cancer, or CVD at cohort entry. As expected, many patients had missing data for BMI, smoking status, and relevant laboratory tests (Table 1). Blood pressure was the best recorded variable, with 23% missing, followed by smoking status, with 35% missing. Fifty-eight percent of patients had missing BMI and 58%–70% had missing cholesterol data. Values for hs-CRP were missing in 88%. After imputing missing data, the proportion of individuals identified as current smokers decreased from 21% to 18%, but the distribution of all

other variables was comparable with data before imputation. Among these approximately 1.4 million CPRD patients, only 10,271 (0.7%) patients had complete data on BMI, smoking status, and all relevant laboratory tests.

After imputing missing data, a total of 383,418 patients (27% of the initial cohort) would have been eligible for the JUPITER trial. Table 2 presents the baseline characteristics of the imputed cohort and the JUPITER trial participants. Compared with the imputed cohort, the JUPITER trial participants tended to have higher BMI but lower hs-CRP and HDL-C levels, and they were more likely to receive aspirin and antihypertensive drugs, to be female, and to be free of moderate chronic kidney disease. As expected, the baseline characteristics of the weighted trial participants were very similar to that of the imputed cohort. Also expected, weighting did not affect randomization, as demonstrated by similar covariate distributions in the weighted rosuvastatin and placebo arms.

Among the 10,271 CPRD patients with complete data on all relevant variables, 2,677 (26%) patients would have been eligible to participate in the trial. Compared with the imputed cohort, individuals in the complete case cohort were less likely to be male, to be current smokers, or to have moderate chronic kidney disease. The patients in the complete case cohort were also more likely to receive antihypertensive drugs or aspirin and to have higher BMI and worse lipid levels. Web Figure 3 presents visualized comparison for baseline characteristics among the JUPITER trial and two target populations. Table 3 compares the baseline

characteristics between the complete case cohort and the JUPITER trial participants. The JUPITER trial participants were more likely to be older and male, to have lower hs-CRP and lipid levels, and to receive aspirin, but were less likely to receive antihypertensive drugs.

The hazard ratio for the effect of rosuvastatin initiation on primary CVD prevention was 0.56 (95% confidence interval (CI): 0.46, 0.69) based on the intention-to-treat analysis in JUPITER (2), and it was attenuated to 0.74 (95% CI: 0.45, 1.25) after weighting to the imputed cohort. We estimated a hazard ratio of 0.71 (95% CI: 0.45, 1.11) after truncating the selection weights at 20 (Figure 1). After weighting the trial participants to the complete case cohort, the hazard ratio of the rosuvastatin effect was 0.71 (95% CI: 0.46, 1.10), which was similar to the estimate obtained by weighting to the imputed cohort, despite the differences in characteristics between the two cohorts. The hazard ratio remained 0.68 (95% CI: 0.45, 1.03) with weight truncated at 20. In a test for interaction between treatment and follow-up time, there was no clear violation of the proportional hazards assumption in the JUPITER trial and after standardizing to the complete case cohort. However, the proportional hazards assumption was violated in the JUPITER trial after standardizing to the imputed cohort due to a few patients in the placebo group who had large weights (>10) and developed CVD at approximately 3 years. We still report the hazard ratio but also report cumulative incidences, which might be more valid in this setting.

Figure 2 and Table 4 show the plots and estimates of crude and weighted cumulative incidences of CVD events, respectively. Rosuvastatin was associated with an absolute reduction in the cumulative incidence of CVD events of 3.3 percentage points (95% CI: 2.1, 4.5) at 4 years (Figure 2A). After weighting to the imputed cohort, the benefits of rosuvastatin did not emerge until the end of 2 years of treatment (Figure 2B), but with 5.8 percentage points, the risk reduction was larger at 4 years (95% CI: 1.0, 10.6). The trending of cumulative incidence was similar between the results of JUPITER (Figure 2A) and the results after standardizing to the complete case cohort (Figure 2C). The absolute reduction in the cumulative incidence of CVD events associated with rosuvastatin became overt at the end of the first year from randomization and was 3.7 percentage points (95% CI: 1.2, 6.2) at 4 years.

When generalizing trial results to patients aged younger than 80 years in the imputed cohort (baseline characteristics are shown in Web Tables 1 and 2), the estimate of the rosuvastatin effects was stronger and more precise (hazard ratio = 0.55; 95% CI: 0.40, 0.77, without applying weight truncation, Figure 3). The benefits of rosuvastatin became overt after the first year from randomization (Figure 4) and the risk reduction was 2.7 percentage points (95% CI: 1.8, 3.6) at 4 years (Table 5). Stronger benefits of rosuvastatin were also observed when generalizing to those aged <80 years in the complete case cohort (hazard ratio = 0.61; 95% CI: 0.42, 0.88, without applying weight truncation). Similar trending of cumulative incidence were observed with 2.9 percentage points (95% CI: 0.6, 5.3) at 4 years.

DISCUSSION

In this study, we showed that generalizing the JUPITER trial results to the UK target population who would have been
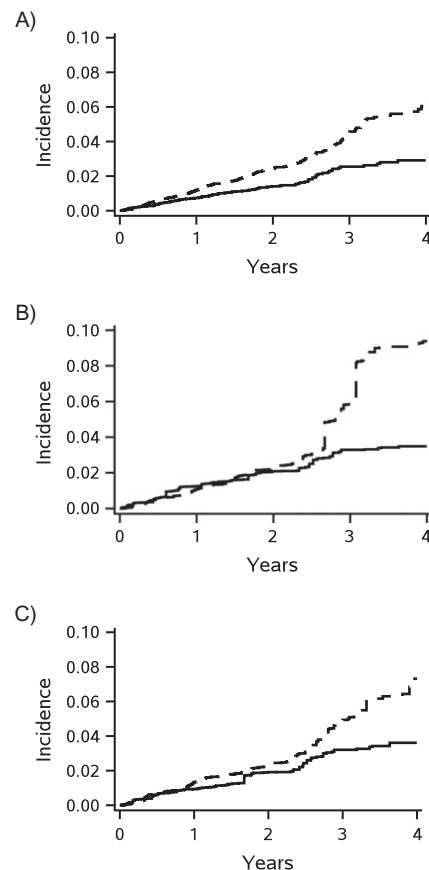
Figure 2. Cumulative incidences of the primary endpoint (major cardiovascular events) according to treatment in the multiple-country trial, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) (2003–2008), before (A) and after standardizing to the imputed cohort (B) and the complete case cohort (C), with selection weight truncated at 20 and accounting for competing risk of any death. The cumulative incidence curves were plotted based on 20 imputed data sets (solid line for rosuvastatin group and dashed line for placebo group).

eligible to be enrolled in the trial was possible and led to a finding of slightly attenuated but still beneficial effects of rosuvastatin on primary prevention of CVD in patients without hyperlipidemia. Reweighting the trial population allowed us to estimate a more realistic treatment effect in the target population while still benefitting from the advantages of randomization.

Our generalized estimate is still based on the trial population, including any effect of the trial on adherence and persistence—what is often referred to as efficacy rather than effectiveness. The only difference between the original trial result and our generalized trial result is that we made the trial population similar to the target population by upweighting individuals less likely to be enrolled. We did so by implementing weighting methods that are widely used to achieve marginal balance of characteristics across treatment cohorts for confounding control (i.e., propensity scores) (18, 19). These methods are infrequently used, however, in other settings, including for generalizing trial results.

While relying on simple principles and statistical techniques, the implementation of our method was hampered by the fact that
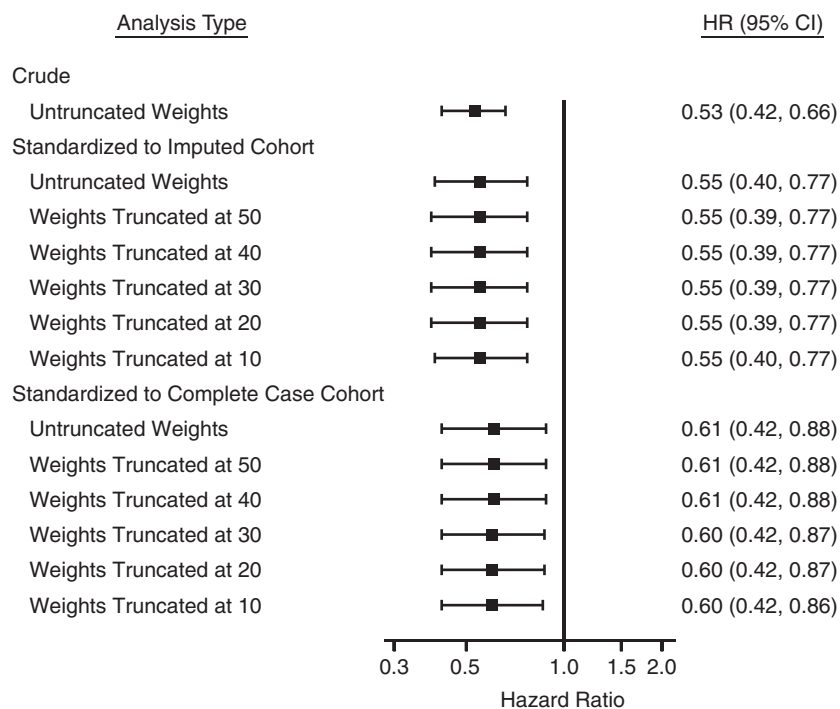
Table 4. Risk Difference of Major Cardiovascular Events for Rosuvastatin in "Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin" (Multiple Countries, 2003–2008), Before and After Standardizing to the Imputed Cohort and the Complete Case Cohort From the Clinical Practice Research Datalink (United Kingdom, 2001–2014), with Selection Weight Truncated at 20

Year From Randomization	JUPITER		JUPITER, Standardized to the Imputed Cohort		JUPITER, Standardized to the Complete Case Cohort	
	Risk Difference, %	95% CI	Risk Difference, %	95% CI	Risk Difference, %	95% CI
1	−0.5	−0.7, −0.2	0.2	−0.5, 0.9	−0.4	−1.0, 0.2
2	−1.0	−1.5, −0.6	−0.3	−1.3, 0.8	−0.4	−1.6, 0.7
3	−2.0	−2.9, −1.1	−2.7	−5.8, 0.4	−1.7	−3.5, 0.1
4	−3.3	−4.5, −2.1	−5.8	−10.6, −1.0	−3.7	−6.2, −1.2

Abbreviation: CI, confidence interval; JUPITER, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin.

key inclusion criteria were available only in a selected subgroup of the target population. We addressed this selection problem by imputing missing values. We found that only a very small proportion of the CPRD patients had complete data on all relevant clinical variables, and that the imputed cohort and complete case cohort were quite different. The results indicated that selection of study populations without accounting for missingness on effect modifiers would fail to represent the target population from the real world.

Approaches dealing with missing data on confounders in nonexperimental studies have been extensively investigated with respect to bias ([20], [21]). Common methods such as complete case analysis and use of missing indicators have been shown to yield biased estimates in realistic settings. In contrast, multiple-imputation methods lead to unbiased estimates under a more plausible assumption, the missing-at-random assumption. While confounding control is not exactly the same as generalizability, the methods to deal with missing values are likely

Analysis Type		HR (95% CI)
Crude		
Untruncated Weights		0.53 (0.42, 0.66)
Standardized to Imputed Cohort		
Untruncated Weights		0.55 (0.40, 0.77)
Weights Truncated at 50		0.55 (0.39, 0.77)
Weights Truncated at 40		0.55 (0.39, 0.77)
Weights Truncated at 30		0.55 (0.39, 0.77)
Weights Truncated at 20		0.55 (0.39, 0.77)
Weights Truncated at 10		0.55 (0.40, 0.77)
Standardized to Complete Case Cohort		
Untruncated Weights		0.61 (0.42, 0.88)
Weights Truncated at 50		0.61 (0.42, 0.88)
Weights Truncated at 40		0.61 (0.42, 0.88)
Weights Truncated at 30		0.60 (0.42, 0.87)
Weights Truncated at 20		0.60 (0.42, 0.87)
Weights Truncated at 10		0.60 (0.42, 0.86)

Hazard Ratio

Figure 3. Treatment effect (hazard ratio (HR) and 95% confidence interval (CI)) of rosuvastatin in the multiple-country trial, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) (2003–2008), before and after standardizing to the complete case cohort and complete case cohort among patients aged <80 years without selection weight truncation. For the results from JUPITER standardized to the imputed cohort, the hazard ratios were first estimated in JUPITER standardized to each imputed data set (Web Table 9) and then combined into overall estimates to account for variability.
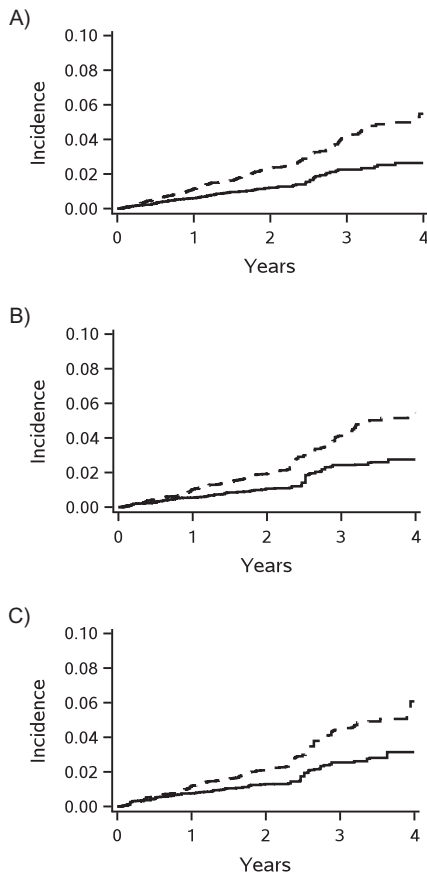
Figure 4. Cumulative incidences of the primary endpoint (major cardiovascular events) according to treatment in the multiple-country trial, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin (JUPITER) (2003–2008), before (A) and after standardizing to the imputed cohort (B) and the complete case cohort (C) among patients aged <80 years without selection weight truncation and with accounting for competing risk of any death. The cumulative incidence curves were plotted based on 20 imputed data sets (solid line for rosuvastatin group and dashed line for placebo group).

to share many characteristics across these settings. In our study, the missing-at-random assumption is likely violated to some extent for many factors, but it seemed plausible for routine laboratory tests and a measure of subclinical inflammation (i.e., hsCRP) in our setting. Previous studies of the UK primary care databases have suggested that BMI and blood pressure measurements were likely missing at random, and that prevalence of current smoking was similar to that observed in the national survey (22–24). In addition, our study examined the older population who were free of major diseases (i.e., diabetes, CVD, and cancer). Among these relatively healthy older patients, laboratory tests may be more likely to be part of their routine examinations rather than due to anticipated abnormalities. Patients who were older or visited general practice more regularly may be more likely to receive routine examinations. The prediction of missing data based on measured patient characteristics varied widely, and more work is needed to identify situations where multiple imputation will not work. Multiple imputation has been shown to be useful, however, even in settings with modest prediction (25).

Weighing by inverse probability of missingness is an alternative approach to dealing with missing data; however, standard applications require a monotone missing pattern when there are multiple missing variables (26, 27), which was not the case in our data. We therefore implemented a simplified version of weighting as a secondary analysis. To bypass the assumption of monotone missingness, we created an overall missing indicator and reweighted the complete case population to the entire eligible population (Web Appendix 2). The results were similar to the results after standardizing to the complete case cohort.

The effect of rosuvastatin on primary prevention of CVD became attenuated after standardizing the trial results to the imputed or complete case cohort. This shift toward the null can be explained in part by more subjects aged ≥80 years and higher values of hs-CRP and HDL-C in the target population. Greater benefits of rosuvastatin on the relative-risk scale were observed in the JUPITER trial for patients with HDL-C ≤50 mg/dL than for those with HDL-C >50 mg/dL (hazard ratio = 0.50 vs. 0.73) and those with lower hs-CRP values at baseline (≤3.5 mg/L) than for those with moderate (3.5–5 mg/L) or higher (>5 mg/L)

Table 5. Risk Difference of Major Cardiovascular Events for Rosuvastatin Among Patients Aged Less Than 80 Years in "Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin" (Multiple Countries, 2003–2008), Before and After Standardizing to the Imputed Cohort and the Complete Case Cohort From the Clinical Practice Research Datalink (United Kingdom, 2001–2014), without Selection Weight Truncation

Year From Randomization	JUPITER		JUPITER, Standardized to the Imputed Cohort		JUPITER, Standardized to the Complete Case Cohort	
	Risk Difference, %	95% CI	Risk Difference, %	95% CI	Risk Difference, %	95% CI
1	−0.5	−0.8, −0.3	−0.5	−0.8, −0.2	−0.4	−1.0, 0.1
2	−1.1	−1.5, −0.6	−0.9	−1.3, −0.5	−0.8	−1.6, −0.1
3	−2.0	−2.8, −1.1	−1.7	−2.8, −0.6	−1.9	−3.5, −0.3
4	−2.8	−4.1, −1.6	−2.7	−3.6, −1.8	−2.9	−5.3, −0.6

Abbreviation: CI, confidence interval; JUPITER, Justification for the Use of Statins in Prevention: an Intervention Trial Evaluating Rosuvastatin.

levels (hazard ratio = 0.38 vs. 0.80 vs. 0.65), although these differences were not statistically significant. However, only 806 JUPITER participants were aged 80 years or older (Web Table 3), leading to an imprecise estimate of the treatment effect in this subgroup and limiting our ability to generalize the trial results to patients in that age group in the target population. Thus, we also generalized the trial results to those aged <80 years in the target population and found greater benefits of rosuvastatin in those relatively younger population.

We also observed a rosuvastatin effect on primary CVD prevention on the risk-difference scale after generalizing to the imputed cohort and complete case cohort (Figure 2 and Table 4). The JUPITER trial results and the results reweighted to the complete case cohort showed the effect of rosuvastatin on CVD prevention following the first year from randomization. In contrast, the risk reduction associated with rosuvastatin in the JUPITER trial generalized to the imputed cohort was not observed until the end of the second year from randomization and abruptly increased after 3 years. This is because the large increases in the cumulative incidence after 2.5 years of follow-up were observed only in the placebo group, resulting from a few patients who were over the age of 80 years with some less-prevalent characteristics (e.g., nonuse of aspirin and any antihypertensive drugs or higher HDL-C levels), and thus had large weights and also had incident cardiovascular disease. This pattern—that the rosuvastatin benefits on CVD were strong only after prolonged treatment—was also seen among patients aged ≥80 years in the original (i.e., unweighted) JUPITER trial but not for patients aged <70, 70–74, and 75–79 years (Web Figure 4). In addition, when generalizing the trial results to the patients aged <80 years in the imputed cohort, the rosuvastatin effects also emerged at 1 year.

Our study has a number of limitations. We need to consider the validity of the assumption that we captured all characteristics related to selection into the trial and to treatment effect heterogeneity. This assumption is analogous to the assumption of no unmeasured confounding when estimating treatment effects in a nonexperimental study (7). The CPRD includes a variety of data, allowing us to identify potential variables as comprehensively as possible; however, owing to the inherent limitations of the CPRD, race/ethnicity is not available. The JUPITER trial enrolled patients from various racial/ethnic groups and from 26 countries. It is possible that both factors would affect the probability of selection into the trial. Although race/ethnicity has been shown to affect LDL-C responses to statin treatment (28), the subgroup analysis of the JUPITER trial showed no difference of racial/ethnic groups and regions on the treatment effect of rosuvastatin. Thus, the assumption of "no unmeasured effect modifiers" might still be plausible in this setting. We also assumed that we correctly specified the logistic models used to predict the probability of selection, although the true models are unknown. The weighted results could be sensitive to the model specification. In this study, we tested the selection models by examining marginal balance of covariates. In the future, more flexible methods, such as a generalized additive model or machine learning, may help relax this assumption of model specification. In addition, our generalized results were limited within the CPRD population to patients who would have been eligible for JUPITER. The JUPITER results cannot be generalized to subjects who are not represented in JUPITER due

to the positivity assumption. Although we selected the target population strictly following the inclusion/exclusion criteria of JUPITER, there is a possibility of some violation of the positivity assumption. Despite the lack of an upper age limit in JUPITER, the oldest trial participant was 97 years of age, making it impossible to generalize trial results to patients older than that. Last, our estimate of the rosuvastatin effect on CVD after generalizing to the imputed cohort was very imprecise due to large weights in some patients. We applied weight truncation to increase the precision of the treatment effect estimate, but it made the weighted trial population slightly less representative of the target population. Despite these limitations, this generalization method based on weighting can still be an effective solution to translate the results of an RCT to real-world population.

In summary, after addressing selection into the JUPITER trial, the relative treatment effect of rosuvastatin on CVD events generalized to the UK target population eligible to enroll the JUPITER trial was slightly less pronounced but still present, especially with prolonged duration of treatment. Our study found that only a small fraction of the eligible CPRD patients had complete data on all effect modifiers and provides evidence for the need to impute missing data when generalizing trial results to target populations.

REFERENCES

1. Stone NJ, Robinson JG, Lichtenstein AH, et al. 2013 ACC/AHA guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular risk in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014; 63(25 Pt B):2889–2934.
2. Ridker PM, Danielson E, Fonseca FA, et al. Rosuvastatin to prevent vascular events in men and women with elevated C-reactive protein. *N Engl J Med*. 2008;359(21):2195–2207.
3. Kappagoda CT, Amsterdam EA. Another look at the results of the JUPITER trial. *Am J Cardiol*. 2009;104(11):1603–1605.
4. Glynn RJ, Koenig W, Nordestgaard BG, et al. Rosuvastatin for primary prevention in older persons with elevated C-reactive protein and low to average low-density lipoprotein cholesterol levels: exploratory analysis of a randomized trial. *Ann Intern Med*. 2010;152(8):488–496.
5. Ridker PM, MacFadyen J, Cressman M, et al. Efficacy of rosuvastatin among men and women with moderate chronic kidney disease and elevated high-sensitivity C-reactive protein: a secondary analysis from the JUPITER (Justification for the Use of Statins in Prevention-an Intervention Trial Evaluating Rosuvastatin) trial. *J Am Coll Cardiol*. 2010;55(12):1266–1273.
6. Ridker PM, MacFadyen JG, Fonseca FA, et al. Number needed to treat with rosuvastatin to prevent first cardiovascular events and death among men and women with low low-density lipoprotein cholesterol and elevated high-sensitivity C-reactive protein: justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin (JUPITER). *Circ Cardiovasc Qual Outcomes*. 2009;2(6):616–623.
7. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172(1):107–115.
8. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70(1):41–55.
9. Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology*. 2003;14(6):680–686.
10. The Clinical Practice Research Datalink (CPRD). https://www.cprd.com/home/. Accessed November 15, 2016.
11. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. *Stat Med*. 1991; 10(4):585–598.
12. Schemper M, Smith TL. Efficient evaluation of treatment effects in the presence of missing covariate values. *Stat Med*. 1990;9(7):777–784.
13. Schemper M, Heinze G. Probability imputation revisited for prognostic factor studies. *Stat Med*. 1997;16(1–3):73–80.
14. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York, NY: John Wiley & Sons; 2004.
15. Little RJ. Regression with missing X's: a review. *J Am Stat Assoc*. 1992;87(420):1227–1237.
16. Cole SR, Lau B, Eron JJ, et al. Estimation of the standardized risk difference and ratio in a competing risks framework: application to injection drug use and progression to AIDS after initiation of antiretroviral therapy. *Am J Epidemiol*. 2015;181(4):238–245.
17. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol*. 2008;168(6):656–664.
18. Stürmer T, Joshi M, Glynn RJ, et al. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol*. 2006;59(5):437–447.
19. Stürmer T, Rothman KJ, Avorn J, et al. Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution–a simulation study. *Am J Epidemiol*. 2010; 172(7):843–854.
20. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol*. 1991;134(8):895–907.
21. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol*. 1995;142(12):1255–1264.
22. Bhaskaran K, Forbes HJ, Douglas I, et al. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open*. 2013;3(9): e003389.
23. Booth HP, Prevost AT, Gulliford MC. Validity of smoking prevalence estimates from primary care electronic health records compared with national population survey data for England, 2007 to 2011. *Pharmacoepidemiol Drug Saf*. 2013; 22(12):1357–1361.
24. Marston L, Carpenter JR, Walters KR, et al. Issues in multiple imputation of missing data for large general practice clinical databases. *Pharmacoepidemiol Drug Saf*. 2010;19(6): 618–626.
25. Rubin DB. Multiple imputation after 18+ years. *J Am Stat Assoc*. 1996;91(434):473–489.
26. Tsiatis AA. *Semiparametric Theory and Missing Data*. New York, NY: Springer; 2006.
27. Horton NJ, Kleinman KP. Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat*. 2007;61(1): 79–90.
28. Yood MU, McCarthy BD, Kempf J, et al. Racial differences in reaching target low-density lipoprotein goal among individuals treated with prescription statin therapy. *Am Heart J*. 2006; 152(4):777–784.
29. Yang D, Dalton JE. A Unified Approach to Measuring the Effect Size Between Two Groups Using SAS. http://support.sas.com/resources/papers/proceedings12/335-2012.pdf. Accessed November 15, 2016.