Methods

Scenario 1

When D_2 is the time to event, it includes two variables (T, δ), where T is the time to event and δ is the indicator of non-censoring. We generate T from the Cox proportional hazards regression model (3), where $\lambda(T|G, X)$ is the hazard function with risk factors of G and X and $2T$ is an arbitrary baseline hazard function. We generate δ from the uniform distribution $U(0, \tau)$, where τ is chosen to yield the fixed proportion c_2 of non-censored subsequent events.

Scenario 2

We simulate Z with a normal distribution of mean ηG and variance 1. We set η to 0.354 and 0.514 to represent that 5% and 10% variance of Z is explained by G.

Calculation of the true marginal association

The true marginal association (i.e., HR or OR) of G on D_2 in scenario 1 can be calculated by the following counterfactual method. First, we generate a sample of the general population with values on G, X, and D_1; the sample size is large, here 200,000,000 (so that the "new dataset" defined below will be sufficiently large for estimating the true marginal association with ignorable variation). We then obtain the subset with $D_1 = 1$ when the target population is population 2, or the subset with $D_1 = 1$ and $S = 1$ when the target population is population 3. In either case, we denote the subset by $\mathcal{A}$. Let N be the sample size of $\mathcal{A}$. Second, we create a new dataset that consists of three sets of samples, each of size N. We generate the first set of samples by taking all subjects in $\mathcal{A}$, replacing their values of G by 0, and generating D_2 based on the original values of X and the new values of G. We generate the second and third sets of samples similarly except that we set all G's to be 1 and 2, respectively. Finally, we calculate the association estimate using the new values of G and D_2 in the new dataset. This way, we obtain

1

the true association estimate in the target population that is characterized by the observed distribution of X in the same population. The marginal effect of G on D_2 or Z on D_2 in scenario 2 is estimated in a similar manner as described above, except that new values of Z are generated after setting G to values of 0, 1, and 2.

Evaluation metrics

The percentage bias is $(\bar{\theta} - \hat{\theta}_{true})/\hat{\theta}_{true} \times 100$, where $\hat{\theta}_{true}$ is the true marginal association estimated by the counterfactual method and $\bar{\theta}$ is the average of the naïve estimates $\hat{\theta}_i$ across B replicates, i.e., $\bar{\theta} = \sum_{i=1}^{B} \hat{\theta}_i / B$. The coverage of the 95% confidence interval (CI) is the proportion of times the 95% CI $\hat{\theta}_i \pm Z_{0.975} SE(\hat{\theta}_i)$ includes $\hat{\theta}_{true}$, where $SE(\hat{\theta}_i)$ is the estimated standard error for $\hat{\theta}_i$ within each replicate and $Z_{0.975}$ is the 0.975 quantile of the standard normal distribution. We set the number of replicates B to 5000 so that the estimated coverage of the 95% CI should fall between 0.944 and 0.956 with 95% probability.

Table S1. Results of the estimated hazard ratio (HR) for a genetic variant that associates with risk of first event, survival, and risk of a subsequent CHD event (scenario 1)

Conditional HR		Marginal HR														
		Population 2							Population 3							
X on D_2	G on D_2	True		Naive estimator					Bias	True		Naive estimator					Bias
$\exp\beta_X$	$\exp\beta_G$	HR	SE	HR	SE	SEE	COV	POW	%	HR	SE	HR	SE	SEE	COV	POW	%
3	1	1	0.005	1	0.044	0.044	0.95	0.049	0.12	1	0.006	1	0.044	0.043	0.95	0.049	0.36
	1.3	1.27	0.006	1.27	0.052	0.051	0.951	1	-0.11	1.27	0.007	1.23	0.05	0.05	0.883	0.999	-2.99
	1.6	1.54	0.008	1.53	0.061	0.061	0.945	1	-0.24	1.56	0.009	1.45	0.058	0.058	0.529	1	-7.25
	2	1.91	0.01	1.88	0.078	0.077	0.931	1	-1.5	1.93	0.012	1.72	0.07	0.07	0.195	1	-10.82
	3	2.79	0.017	2.71	0.127	0.128	0.914	1	-2.61	2.81	0.019	2.37	0.103	0.105	0.034	1	-15.52
5	1	1	0.005	1	0.044	0.043	0.948	0.053	-0.24	1	0.005	1	0.043	0.043	0.951	0.047	-0.42
	1.3	1.25	0.006	1.23	0.049	0.05	0.925	0.999	-2.08	1.26	0.007	1.18	0.049	0.048	0.605	0.978	-6.64
	1.6	1.49	0.008	1.44	0.058	0.057	0.857	1	-3.37	1.51	0.009	1.34	0.054	0.054	0.155	1	-11.2
	2	1.8	0.009	1.71	0.068	0.069	0.759	1	-4.95	1.84	0.011	1.54	0.062	0.062	0.012	1	-15.83
	3	2.56	0.015	2.32	0.105	0.105	0.432	1	-9.14	2.62	0.018	1.99	0.085	0.085	0	1	-23.98
10	1	0.99	0.005	1	0.043	0.043	0.946	0.055	0.61	0.99	0.005	1	0.043	0.043	0.947	0.055	0.67
	1.3	1.23	0.006	1.15	0.046	0.046	0.659	0.938	-6.09	1.25	0.007	1.12	0.046	0.046	0.25	0.79	-10.21
	1.6	1.46	0.007	1.29	0.051	0.05	0.116	1	-11.62	1.49	0.008	1.23	0.049	0.049	0.002	0.999	-17.78
	2	1.72	0.009	1.45	0.056	0.056	0.006	1	-15.95	1.79	0.011	1.35	0.053	0.053	0	1	-24.71
	3	2.33	0.012	1.78	0.071	0.071	0	1	-23.79	2.51	0.016	1.6	0.064	0.065	0	1	-35.99

The naive estimator was obtained from Cox regression. The true value was obtained from the counterfactual method. SE is the empirical standard error. SEE is the mean of standard error estimates. COV is the coverage of the 95% CI. POW is the power when $\exp\beta_G \neq 1$ and the type 1 error when $\exp\beta_G = 1$. Bias (%) is the difference between the estimated marginal HR and the true marginal HR, divided by the true marginal HR.

Table S2. Results of the estimated hazard ratio (HR) for a genetic variant that encodes a biomarker that associates with risk of first event, survival, and risk of a subsequent CHD event (scenario 2)

Conditional OR			Marginal OR														
			Population 2								Population 3						
G on Z	X on D_2	Z on D_2	True		Naïve estimator					Bias	True		Naïve estimator					Bias
η	$\exp\beta_X$	$\exp\beta_Z$	OR	SE	OR	SE	SEE	COV	POW	%	OR	SE	OR	SE	SEE	COV	POW	%
0.354	3	1	1	0.005	1	0.044	0.043	0.948	0.051	-0.36	1	0.006	1	0.043	0.044	0.95	0.05	-0.35
		1.3	1.08	0.005	1.09	0.045	0.046	0.953	0.525	0.64	1.09	0.006	1.07	0.046	0.046	0.923	0.393	-1.95
		1.6	1.16	0.006	1.16	0.048	0.048	0.952	0.949	0.05	1.18	0.007	1.13	0.048	0.047	0.852	0.849	-3.71
		2	1.25	0.006	1.24	0.051	0.05	0.947	0.999	-0.87	1.25	0.007	1.19	0.049	0.049	0.802	0.991	-4.58
		3	1.39	0.007	1.36	0.054	0.054	0.928	1	-1.7	1.4	0.008	1.28	0.052	0.052	0.454	1	-8.07
	5	1	1.01	0.005	1	0.043	0.043	0.947	0.049	-0.75	0.99	0.005	1	0.043	0.043	0.948	0.05	0.64
		1.3	1.08	0.005	1.07	0.045	0.045	0.947	0.412	-0.65	1.08	0.006	1.06	0.045	0.045	0.931	0.287	-1.93
		1.6	1.15	0.006	1.13	0.046	0.047	0.936	0.861	-1.67	1.16	0.006	1.11	0.046	0.046	0.823	0.688	-4.19
		2	1.23	0.006	1.2	0.048	0.049	0.885	0.992	-2.97	1.23	0.007	1.16	0.047	0.047	0.695	0.938	-5.88
		3	1.36	0.007	1.29	0.052	0.051	0.726	1	-5.24	1.35	0.008	1.23	0.048	0.049	0.322	0.999	-9.26
	10	1	1	0.005	1	0.043	0.043	0.947	0.054	-0.07	1	0.005	1	0.043	0.043	0.953	0.047	0.06
		1.3	1.07	0.005	1.05	0.044	0.044	0.917	0.223	-2.17	1.07	0.006	1.04	0.044	0.044	0.885	0.156	-3.16
		1.6	1.14	0.006	1.09	0.045	0.045	0.79	0.572	-4.51	1.15	0.006	1.07	0.044	0.044	0.651	0.401	-6.27
		2	1.21	0.006	1.13	0.045	0.045	0.651	0.865	-6.16	1.22	0.007	1.1	0.045	0.045	0.322	0.693	-9.42
		3	1.33	0.007	1.19	0.043	0.043	0.138	0.998	-10.36	1.33	0.007	1.16	0.047	0.046	0.057	0.95	-13.08
0.514	3	1	1	0.005	1	0.044	0.043	0.948	0.051	-0.36	1	0.006	1	0.043	0.044	0.95	0.05	-0.35
		1.3	1.14	0.006	1.13	0.047	0.047	0.951	0.832	-0.48	1.13	0.006	1.11	0.046	0.047	0.932	0.704	-1.84
		1.6	1.25	0.006	1.24	0.051	0.05	0.941	1	-1.01	1.25	0.007	1.2	0.049	0.049	0.849	0.995	-3.74
		2	1.38	0.007	1.36	0.054	0.055	0.946	1	-1.08	1.36	0.007	1.29	0.052	0.052	0.762	1	-5.02
		3	1.61	0.008	1.56	0.062	0.062	0.888	1	-2.8	1.62	0.009	1.44	0.056	0.057	0.133	1	-11.52
	5	1	1.01	0.005	1	0.043	0.043	0.947	0.049	-0.75	0.99	0.005	1	0.043	0.043	0.948	0.05	0.64
		1.3	1.12	0.005	1.11	0.046	0.046	0.944	0.704	-1.1	1.12	0.006	1.09	0.046	0.046	0.896	0.523	-2.76
		1.6	1.23	0.006	1.2	0.049	0.049	0.921	0.994	-2	1.23	0.007	1.16	0.049	0.048	0.703	0.944	-5.77
		2	1.34	0.007	1.3	0.052	0.052	0.847	1	-3.61	1.36	0.008	1.23	0.049	0.05	0.323	0.999	-9.26
		3	1.57	0.008	1.45	0.057	0.057	0.472	1	-7.66	1.58	0.009	1.35	0.053	0.054	0.016	1	-14.86
	10	1	1	0.005	1	0.043	0.043	0.947	0.054	-0.07	1	0.005	1	0.043	0.043	0.953	0.047	0.06
		1.3	1.11	0.005	1.07	0.045	0.044	0.858	0.422	-3.46	1.12	0.006	1.06	0.045	0.044	0.768	0.304	-4.94
		1.6	1.21	0.006	1.13	0.045	0.046	0.623	0.882	-6.4	1.22	0.007	1.11	0.045	0.045	0.362	0.697	-8.98
		2	1.31	0.006	1.19	0.047	0.047	0.365	0.995	-8.62	1.32	0.007	1.16	0.046	0.046	0.077	0.95	-12.5
		3	1.51	0.007	1.29	0.043	0.044	0.003	1	-14.82	1.53	0.009	1.23	0.048	0.048	0	1	-19.6

The naïve estimator was obtained from Cox regression. The true value was obtained from the counterfactual method. SE is the empirical standard error. SEE is the mean of standard error estimates. COV is the coverage of the 95% CI. POW is the power when $\exp\beta_Z \neq 1$ and the type 1 error when $\exp\beta_Z = 1$. Bias (%) is the difference between the estimated marginal HR and the true marginal HR, divided by the true marginal HR. η = 0.354 and 0.514 represent that 5% and 10% variance of Z is explained by G, respectively.

Table S3. Results of the estimated hazard ratio (HR) for a non-genetic biomarker that associates with risk of first event, survival, and risk of a subsequent CHD event (scenario 2)

	Conditional OR				Marginal OR													
					Population 2						Population 3							
G on Z	X on D_2	Z on D_2	True		Naive estimator					Bias	True		Naive estimator					Bias
η	$\exp\beta_X$	$\exp\beta_Z$	OR	SE	OR	SE	SEE	COV	POW	%	OR	SE	OR	SE	SEE	COV	POW	%
0.354	3	1	1	0.004	1	0.028	0.027	0.942	0.053	0.34	1	0.004	1	0.028	0.027	0.943	0.058	0.18
		1.3	1.26	0.005	1.27	0.034	0.035	0.951	1	0.41	1.28	0.005	1.23	0.034	0.034	0.646	1	-4.25
		1.6	1.54	0.006	1.53	0.042	0.042	0.941	1	-0.53	1.55	0.007	1.45	0.04	0.04	0.301	1	-6.74
		2	1.89	0.007	1.86	0.051	0.052	0.913	1	-1.64	1.88	0.008	1.71	0.048	0.049	0.082	1	-9.07
		3	2.67	0.011	2.56	0.073	0.074	0.723	1	-3.87	2.7	0.012	2.3	0.067	0.069	0	1	-14.68
	5	1	1	0.004	1	0.027	0.027	0.95	0.051	-0.25	1	0.004	1	0.027	0.027	0.951	0.047	0.22
		1.3	1.25	0.005	1.23	0.034	0.034	0.918	1	-1.48	1.26	0.005	1.18	0.033	0.032	0.289	1	-6.75
		1.6	1.5	0.006	1.43	0.04	0.039	0.655	1	-4.16	1.52	0.006	1.34	0.037	0.037	0.007	1	-11.67
		2	1.77	0.007	1.68	0.046	0.046	0.494	1	-5.3	1.82	0.008	1.53	0.042	0.043	0	1	-15.9
		3	2.46	0.01	2.18	0.06	0.062	0.007	1	-11.65	2.52	0.011	1.93	0.055	0.056	0	1	-23.24
	10	1	1.01	0.004	1	0.027	0.027	0.945	0.051	-0.74	1	0.004	1	0.027	0.027	0.946	0.051	-0.18
		1.3	1.23	0.005	1.15	0.031	0.031	0.282	0.999	-6.53	1.25	0.005	1.12	0.03	0.03	0.022	0.987	-10.12
		1.6	1.45	0.006	1.28	0.034	0.035	0.004	1	-11.44	1.48	0.006	1.23	0.033	0.033	0	1	-17.22
		2	1.7	0.006	1.43	0.038	0.039	0	1	-15.77	1.75	0.007	1.34	0.037	0.037	0	1	-23.59
		3	2.28	0.009	1.71	0.043	0.044	0	1	-24.99	2.34	0.01	1.57	0.042	0.044	0	1	-32.96
0.514	3	1	1	0.004	1	0.027	0.027	0.944	0.053	0.28	1	0.004	1	0.027	0.027	0.942	0.058	0.13
		1.3	1.27	0.005	1.27	0.034	0.034	0.951	1	-0.34	1.27	0.005	1.23	0.033	0.033	0.761	1	-3.24
		1.6	1.55	0.006	1.53	0.041	0.041	0.933	1	-1.03	1.56	0.006	1.45	0.039	0.039	0.224	1	-7.17
		2	1.88	0.007	1.85	0.05	0.05	0.911	1	-1.48	1.9	0.008	1.71	0.046	0.047	0.037	1	-9.76
		3	2.66	0.01	2.55	0.071	0.071	0.683	1	-4.04	2.7	0.012	2.29	0.066	0.067	0	1	-14.93
	5	1	1	0.004	1	0.026	0.027	0.951	0.047	-0.29	1	0.004	1	0.027	0.027	0.95	0.048	0.26
		1.3	1.25	0.004	1.23	0.032	0.033	0.903	1	-1.77	1.26	0.005	1.18	0.031	0.032	0.34	1	-6.2
		1.6	1.48	0.005	1.43	0.037	0.038	0.757	1	-3.34	1.52	0.006	1.34	0.036	0.036	0.003	1	-11.66
		2	1.78	0.007	1.68	0.045	0.045	0.392	1	-5.86	1.82	0.008	1.53	0.041	0.042	0	1	-15.69
		3	2.44	0.009	2.17	0.057	0.059	0.007	1	-11.41	2.49	0.011	1.92	0.053	0.055	0	1	-22.73
	10	1	1.01	0.004	1	0.026	0.026	0.943	0.051	-0.69	1	0.004	1	0.026	0.026	0.947	0.052	-0.16
		1.3	1.23	0.004	1.15	0.031	0.03	0.28	1	-6.45	1.24	0.005	1.12	0.029	0.03	0.023	0.993	-9.9
		1.6	1.45	0.005	1.28	0.033	0.033	0.003	1	-11.49	1.47	0.006	1.22	0.032	0.032	0	1	-16.78
		2	1.7	0.006	1.43	0.037	0.037	0	1	-16.16	1.74	0.007	1.34	0.035	0.036	0	1	-23.04
		3	2.25	0.008	1.69	0.039	0.04	0	1	-24.7	2.35	0.01	1.57	0.042	0.043	0	1	-33.41

The naïve estimator was obtained from Cox regression. The true value was obtained from the counterfactual method. SE is the empirical standard error. SEE is the mean of standard error estimates. COV is the coverage of the 95% CI. POW is the power when $\exp\beta_Z \neq 1$ and the type 1 error when $\exp\beta_Z = 1$. Bias (%) is the difference between the estimated marginal HR and the true marginal HR, divided by the true marginal HR. $\eta = 0.354$ and 0.514 represent that 5% and 10% variance of Z is explained by G, respectively.

Table S4. Results of the estimated odds ratio (OR) for a genetic variant that associates with risk of first event, survival, and risk of a subsequent CHD event (scenario 1)

Conditional OR		Marginal OR																	
		Population 1				Population 2								Population 3					
X on D_2	G on D_2	Naive estimator		True		Naive estimator					Bias	True		Naive estimator					Bias
$\exp\beta_X$	$\exp\beta_G$	OR	POW	OR	SE	OR	SE	SEE	COV	POW	%	OR	SE	OR	SE	SEE	COV	POW	%
3	1	1	0.053	1	0.003	1	0.045	0.045	0.948	0.05	-0.57	1	0.003	1	0.045	0.045	0.949	0.051	-0.26
	1.3	1.3	1	1.25	0.003	1.24	0.052	0.052	0.952	0.999	-0.22	1.26	0.004	1.21	0.052	0.051	0.869	0.993	-3.43
	1.6	1.6	1	1.48	0.004	1.47	0.061	0.061	0.949	1	-0.29	1.5	0.005	1.41	0.058	0.059	0.673	1	-6.14
	2	1.99	1	1.77	0.004	1.76	0.074	0.073	0.945	1	-0.73	1.8	0.005	1.65	0.068	0.069	0.44	1	-8.34
	3	2.98	1	2.45	0.006	2.41	0.108	0.109	0.94	1	-1.54	2.5	0.007	2.2	0.096	0.097	0.175	1	-12.04
5	1	1	0.047	1	0.002	1	0.045	0.045	0.946	0.053	-0.08	1	0.003	1	0.045	0.045	0.951	0.049	0.17
	1.3	1.29	1	1.2	0.003	1.19	0.05	0.05	0.95	0.982	-1.03	1.21	0.003	1.16	0.049	0.049	0.825	0.928	-4.29
	1.6	1.58	1	1.38	0.003	1.36	0.056	0.057	0.941	1	-1.42	1.41	0.004	1.3	0.054	0.054	0.534	1	-7.47
	2	1.97	1	1.62	0.004	1.58	0.065	0.066	0.921	1	-2.06	1.65	0.004	1.48	0.062	0.061	0.228	1	-10.7
	3	2.92	1	2.17	0.005	2.09	0.091	0.09	0.855	1	-3.61	2.23	0.006	1.88	0.08	0.08	0.025	1	-15.62
10	1	1	0.055	1	0.003	1	0.045	0.045	0.945	0.055	-0.19	1	0.003	1	0.045	0.045	0.949	0.049	-0.39
	1.3	1.27	1	1.18	0.003	1.15	0.05	0.049	0.91	0.893	-2.4	1.18	0.003	1.12	0.048	0.048	0.783	0.767	-4.94
	1.6	1.54	1	1.35	0.004	1.29	0.055	0.054	0.804	1	-4.56	1.36	0.004	1.24	0.053	0.052	0.424	0.998	-8.79
	2	1.89	1	1.55	0.005	1.46	0.06	0.061	0.693	1	-5.87	1.56	0.005	1.38	0.057	0.057	0.142	1	-11.88
	3	2.73	1	2.06	0.007	1.86	0.078	0.078	0.328	1	-9.59	2.08	0.007	1.7	0.071	0.071	0.002	1	-18.43

The naïve estimator was obtained from logistic regression. The true value was obtained from the counterfactual method. SE is the empirical standard error. SEE is the mean of standard error estimates. COV is the coverage of the 95% CI. POW is the power when $\exp\beta_G \neq 1$ and the type 1 error when $\exp\beta_G = 1$. Bias (%) is the difference between the estimated marginal OR and the true marginal OR, divided by the true marginal OR.

Table S5. Results of the estimated odds ratio (OR) for a genetic variant that encodes a biomarker that associates with risk of first event, survival, and risk of a subsequent CHD event (scenario 2)

			Conditional OR		Marginal OR															
			Population 1		Population 2								Population 3							
G on Z	X on D_2	Z on D_2	Naïve estimator		True		Naïve estimator					Bias	True		Naïve estimator					Bias
η	$\exp\beta_X$	$\exp\beta_Z$	OR	POW	OR	SE	OR	SE	SEE	COV	POW	%	OR	SE	OR	SE	SEE	COV	POW	%
0.354	3	1	1	0.051	1	0.003	1	0.044	0.045	0.948	0.052	-0.16	1	0.003	1	0.045	0.045	0.952	0.049	-0.24
		1.3	1.1	0.552	1.08	0.003	1.08	0.047	0.047	0.953	0.412	-0.09	1.09	0.003	1.07	0.046	0.047	0.938	0.338	-1.6
		1.6	1.18	0.963	1.15	0.003	1.14	0.049	0.049	0.947	0.866	-0.68	1.15	0.004	1.12	0.049	0.048	0.921	0.768	-2.08
		2	1.27	1	1.21	0.003	1.2	0.05	0.051	0.948	0.993	-0.89	1.22	0.004	1.18	0.05	0.05	0.843	0.966	-3.98
		3	1.47	1	1.32	0.003	1.3	0.054	0.054	0.932	1	-1.73	1.33	0.004	1.25	0.052	0.053	0.68	1	-6.08
	5	1	1	0.049	1	0.002	1	0.044	0.045	0.954	0.047	-0.19	1	0.003	1	0.044	0.045	0.952	0.048	-0.18
		1.3	1.1	0.545	1.07	0.002	1.06	0.046	0.047	0.956	0.284	-0.58	1.07	0.003	1.05	0.045	0.046	0.932	0.215	-1.96
		1.6	1.18	0.958	1.12	0.002	1.11	0.048	0.048	0.947	0.694	-0.56	1.13	0.003	1.09	0.047	0.047	0.899	0.553	-2.92
		2	1.27	1	1.18	0.003	1.17	0.05	0.05	0.943	0.948	-1.19	1.19	0.003	1.14	0.049	0.049	0.833	0.848	-4.23
		3	1.45	1	1.28	0.003	1.26	0.052	0.053	0.936	1	-1.6	1.29	0.003	1.21	0.051	0.051	0.702	0.995	-5.92
	10	1	1	0.051	1	0.003	1	0.044	0.045	0.95	0.049	-0.34	1	0.003	1	0.044	0.045	0.956	0.042	0.17
		1.3	1.09	0.487	1.06	0.003	1.05	0.045	0.046	0.948	0.154	-0.84	1.06	0.003	1.04	0.045	0.046	0.934	0.154	-1.94
		1.6	1.16	0.938	1.11	0.003	1.09	0.047	0.047	0.937	0.508	-1.84	1.11	0.003	1.08	0.046	0.047	0.89	0.385	-3.04
		2	1.25	0.999	1.16	0.003	1.14	0.048	0.049	0.922	0.842	-2.27	1.16	0.003	1.11	0.048	0.048	0.828	0.7	-4.31
		3	1.42	1	1.26	0.003	1.22	0.052	0.052	0.896	0.997	-2.92	1.26	0.004	1.18	0.051	0.05	0.71	0.976	-5.78
0.514	3	1	1	0.051	1	0.003	1	0.044	0.045	0.948	0.052	-0.16	1	0.003	1	0.045	0.045	0.952	0.049	-0.24
		1.3	1.14	0.86	1.12	0.003	1.12	0.048	0.048	0.954	0.723	-0.28	1.13	0.004	1.1	0.047	0.048	0.925	0.616	-2.13
		1.6	1.27	1	1.22	0.003	1.21	0.052	0.051	0.944	0.995	-0.53	1.23	0.004	1.18	0.051	0.05	0.864	0.972	-3.5
		2	1.42	1	1.32	0.003	1.31	0.055	0.055	0.939	1	-1.13	1.33	0.004	1.26	0.053	0.053	0.766	1	-5.04
		3	1.74	1	1.5	0.004	1.46	0.06	0.061	0.905	1	-2.5	1.51	0.004	1.38	0.057	0.058	0.448	1	-8.37
	5	1	1	0.049	1	0.002	1	0.044	0.045	0.954	0.047	-0.19	1	0.003	1	0.044	0.045	0.952	0.048	-0.18
		1.3	1.14	0.842	1.1	0.002	1.09	0.048	0.047	0.949	0.518	-0.66	1.1	0.003	1.08	0.047	0.047	0.92	0.402	-2.34
		1.6	1.27	1	1.18	0.003	1.17	0.05	0.05	0.944	0.952	-1.2	1.19	0.003	1.14	0.049	0.049	0.828	0.852	-4.12
		2	1.42	1	1.27	0.003	1.25	0.053	0.053	0.935	1	-1.59	1.28	0.003	1.21	0.052	0.051	0.703	0.993	-5.83
		3	1.72	1	1.42	0.003	1.4	0.058	0.058	0.923	1	-1.96	1.44	0.004	1.32	0.054	0.055	0.501	1	-7.85
	10	1	1	0.051	1	0.003	1	0.044	0.045	0.95	0.049	-0.34	1	0.003	1	0.044	0.045	0.956	0.042	0.17
		1.3	1.13	0.807	1.09	0.003	1.07	0.047	0.047	0.94	0.369	-1.46	1.09	0.003	1.06	0.047	0.046	0.911	0.267	-2.58
		1.6	1.25	1	1.16	0.003	1.14	0.049	0.049	0.914	0.84	-2.42	1.16	0.003	1.11	0.048	0.048	0.815	0.684	-4.52
		2	1.38	1	1.24	0.003	1.2	0.051	0.051	0.876	0.994	-3.34	1.25	0.004	1.17	0.05	0.05	0.669	0.957	-6.32
		3	1.65	1	1.39	0.004	1.34	0.055	0.056	0.839	1	-3.96	1.41	0.004	1.28	0.054	0.054	0.394	1	-9.01

The naïve estimator was obtained from logistic regression. The true value was obtained from the counterfactual method. SE is the empirical standard error. SEE is the mean of standard error estimates. COV is the coverage of the 95% CI. POW is the power when $\exp\beta_Z \neq 1$ and the type 1 error when $\exp\beta_Z = 1$. Bias (%) is the difference between the estimated marginal OR and the true marginal OR, divided by the true marginal OR. η = 0.354 and 0.514 represent that 5% and 10% variance of Z is explained by G, respectively.

Table S6. Results of the estimated odds ratio (OR) for a non-genetic biomarker that associates with risk of first event, survival, and risk of a subsequent CHD event (scenario 2)

Conditional OR					Marginal OR															
			Population 1		Population 2								Population 3							
			Naïve estimator		True		Naïve estimator					Bias	True		Naïve estimator					Bias
G on Z	X on D2	Z on D2	OR	POW	OR	SE	OR	SE	SEE	COV	POW	%	OR	SE	OR	SE	SEE	COV	POW	%
η	$\exp\beta_X$	$\exp\beta_Z$																		
0.354	3	1	1	0.055	1	0.002	1	0.029	0.028	0.945	0.056	0.19	1	0.003	1	0.028	0.028	0.951	0.05	-0.09
		1.3	1.3	1	1.24	0.003	1.24	0.035	0.035	0.952	1	0.1	1.26	0.003	1.21	0.034	0.035	0.784	1	-3.32
		1.6	1.6	1	1.49	0.003	1.47	0.042	0.042	0.932	1	-0.95	1.5	0.004	1.41	0.04	0.041	0.439	1	-5.93
		2	1.99	1	1.78	0.004	1.75	0.051	0.052	0.913	1	-1.72	1.82	0.005	1.65	0.049	0.05	0.097	1	-9.36
		3	2.97	1	2.49	0.006	2.41	0.077	0.078	0.83	1	-3.18	2.55	0.007	2.2	0.07	0.072	0.006	1	-13.8
	5	1	1	0.043	1	0.002	1	0.028	0.028	0.956	0.045	0.18	1	0.002	1	0.028	0.028	0.951	0.048	0.06
		1.3	1.29	1	1.2	0.002	1.19	0.033	0.034	0.943	1	-0.84	1.21	0.002	1.16	0.033	0.033	0.662	0.999	-4.32
		1.6	1.58	1	1.38	0.002	1.36	0.04	0.039	0.916	1	-1.45	1.42	0.003	1.3	0.038	0.038	0.153	1	-8.19
		2	1.97	1	1.62	0.003	1.58	0.047	0.047	0.871	1	-2.35	1.67	0.004	1.48	0.043	0.044	0.014	1	-11.41
		3	2.89	1	2.18	0.005	2.1	0.066	0.066	0.77	1	-3.75	2.24	0.005	1.89	0.059	0.059	0	1	-15.56
	10	1	1	0.054	1	0.002	1	0.029	0.028	0.948	0.052	0.03	1	0.002	1	0.029	0.028	0.947	0.052	-0.24
		1.3	1.27	1	1.18	0.002	1.15	0.033	0.033	0.867	0.998	-2.23	1.18	0.003	1.12	0.033	0.032	0.58	0.985	-4.84
		1.6	1.54	1	1.34	0.003	1.28	0.037	0.037	0.69	1	-4.09	1.36	0.003	1.24	0.035	0.035	0.105	1	-8.86
		2	1.88	1	1.54	0.003	1.45	0.042	0.042	0.482	1	-5.6	1.56	0.004	1.37	0.039	0.04	0.006	1	-11.97
		3	2.69	1	1.99	0.005	1.86	0.056	0.056	0.368	1	-6.72	2.03	0.005	1.7	0.051	0.052	0	1	-16.41
0.514	3	1	1	0.056	1	0.002	1	0.028	0.028	0.944	0.056	0.15	1	0.003	1	0.027	0.028	0.951	0.05	-0.09
		1.3	1.3	1	1.24	0.003	1.25	0.034	0.034	0.951	1	0.08	1.26	0.003	1.21	0.033	0.034	0.761	1	-3.47
		1.6	1.6	1	1.48	0.003	1.47	0.041	0.041	0.943	1	-0.75	1.5	0.004	1.41	0.04	0.04	0.377	1	-6.16
		2	1.99	1	1.78	0.004	1.75	0.05	0.051	0.903	1	-1.79	1.81	0.004	1.65	0.047	0.048	0.11	1	-8.89
		3	2.96	1	2.48	0.006	2.41	0.076	0.076	0.852	1	-2.8	2.52	0.007	2.2	0.07	0.071	0.011	1	-12.85
	5	1	1	0.045	1	0.002	1	0.027	0.028	0.954	0.045	0.14	1	0.002	1	0.027	0.028	0.95	0.049	0.03
		1.3	1.29	1	1.2	0.002	1.19	0.033	0.033	0.939	1	-0.87	1.21	0.002	1.16	0.032	0.032	0.635	1	-4.39
		1.6	1.58	1	1.38	0.002	1.36	0.038	0.038	0.916	1	-1.47	1.41	0.003	1.3	0.036	0.036	0.171	1	-7.82
		2	1.96	1	1.63	0.003	1.58	0.045	0.045	0.84	1	-2.7	1.66	0.003	1.48	0.042	0.042	0.019	1	-11.19
		3	2.89	1	2.19	0.004	2.11	0.064	0.064	0.762	1	-3.66	2.24	0.005	1.9	0.058	0.058	0	1	-15.35
	10	1	1	0.053	1	0.002	1	0.028	0.028	0.949	0.051	0	1	0.002	1	0.028	0.028	0.949	0.051	-0.2
		1.3	1.27	1	1.18	0.002	1.15	0.032	0.032	0.864	0.999	-2.29	1.18	0.003	1.12	0.031	0.031	0.565	0.988	-4.82
		1.6	1.54	1	1.34	0.003	1.28	0.036	0.036	0.715	1	-3.79	1.35	0.003	1.24	0.034	0.034	0.102	1	-8.51
		2	1.88	1	1.54	0.003	1.46	0.041	0.041	0.473	1	-5.54	1.56	0.004	1.38	0.039	0.039	0.006	1	-11.93
		3	2.69	1	2.01	0.005	1.87	0.056	0.055	0.334	1	-6.81	2.04	0.005	1.71	0.051	0.051	0	1	-15.93

The naïve estimator was obtained from logistic regression. The true value was obtained from the counterfactual method. SE is the empirical standard error. SEE is the mean of standard error estimates. COV is the coverage of the 95% CI. POW is the power when $\exp\beta_Z \neq 1$ and the type 1 error when $\exp\beta_Z = 1$. Bias (%) is the difference between the estimated marginal OR and the true marginal OR, divided by the true marginal OR. η = 0.354 and 0.514 represent that 5% and 10% variance of Z is explained by G, respectively.

Table S7. Results of the estimated odds ratio (OR) for a genetic variant (scenario 1) in case-control studies

Conditional OR		Marginal OR													
		Population 1		Population 2						Population 3					
X on D_2	G on D_2	Naive estimator		True		Naive estimator			Bias	True		Naive estimator			Bias
$\exp \beta_X$	$\exp \beta_G$	OR	POW	OR	SE	OR	COV	POW	%	OR	SE	OR	COV	POW	%
3	1	1	0.05	1	0.003	1	0.954	0.043	-0.37	1	0.003	1	0.955	0.046	-0.16
	1.3	1.3	1	1.25	0.003	1.24	0.954	1	-0.3	1.26	0.004	1.21	0.517	1	-3.51
	1.6	1.6	1	1.48	0.004	1.47	0.951	1	-0.29	1.5	0.005	1.41	0.073	1	-6.07
	2	1.99	1	1.77	0.004	1.76	0.931	1	-0.62	1.8	0.005	1.65	0.006	1	-8.12
	3	2.98	1	2.45	0.006	2.42	0.913	1	-1.17	2.5	0.007	2.22	0	1	-11.48
5	1	1	0.052	1	0.002	1	0.952	0.047	-0.08	1	0.003	1	0.949	0.048	0.27
	1.3	1.29	1	1.2	0.003	1.19	0.924	1	-0.87	1.21	0.003	1.16	0.372	1	-4.21
	1.6	1.58	1	1.38	0.003	1.36	0.872	1	-1.42	1.41	0.004	1.3	0.013	1	-7.47
	2	1.97	1	1.62	0.004	1.58	0.798	1	-2.06	1.65	0.004	1.48	0	1	-10.58
	3	2.93	1	2.17	0.005	2.09	0.52	1	-3.56	2.23	0.006	1.88	0	1	-15.44
10	1	1	0.057	1	0.003	1	0.953	0.046	-0.29	1	0.003	1	0.951	0.044	-0.39
	1.3	1.27	1	1.18	0.003	1.15	0.739	1	-2.4	1.18	0.003	1.12	0.248	1	-4.86
	1.6	1.54	1	1.35	0.004	1.29	0.301	1	-4.49	1.36	0.004	1.24	0.001	1	-8.71
	2	1.89	1	1.55	0.005	1.46	0.088	1	-5.93	1.56	0.005	1.38	0	1	-11.95
	3	2.75	1	2.06	0.007	1.86	0	1	-9.64	2.08	0.007	1.7	0	1	-18.29

The case-control design with 12500 cases and 12500 controls was adopted for populations 1, 2, and 3. The naïve estimator was obtained from logistic regression. The true value was obtained from the counterfactual method. SE is the empirical standard error. SEE is the mean of standard error estimates. COV is the coverage of the 95% CI. POW is the power when $\exp \beta_G \neq 1$ and the type 1 error when $\exp \beta_G = 1$. Bias (%) is the difference between the estimated marginal OR and the true marginal OR, divided by the true marginal OR.

Figure S1. Illustration of different effect sizes

(a)   **(b)**
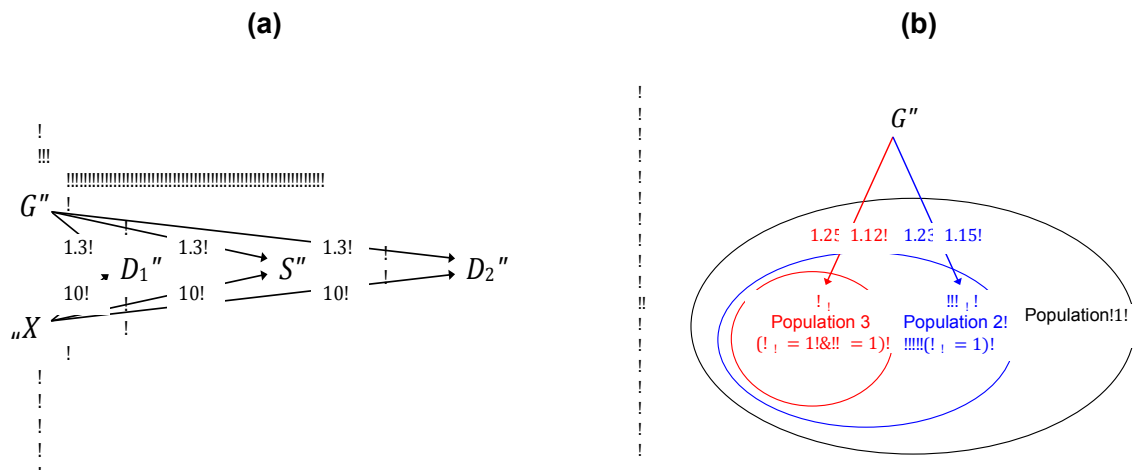
(a) The outcomes D_1, S, and D_2 are generated by the logistic regression models (1) and (2) and the Cox model (3), respectively, with a conditional effect of 1.3 for the SNP of interest (G) and a conditional effect of 10 for all remaining factors (X) in all three models.

(b) Conditioning on having the first CHD event ($D_1 = 1$, population 2), the true marginal hazard ratio (HR) is estimated to be 1.23 by the counterfactual method and the naïve marginal HR estimate is 1.15 by the standard Cox model, resulting in a bias of 6.09% towards the null; conditioning on both having and surviving the first CHD event ($D_1 = 1$ and $S = 1$, population 3), the true and naïve marginal HR estimates are 1.25 and 1.12, respectively, resulting in a bias of 10.21% towards the null.

Figure S2. Bias as the overall disease rate in the general population (c_1), rate of non-censored subsequent events (c_2), death rate (c_S), and SNP minor allele frequency (π) vary
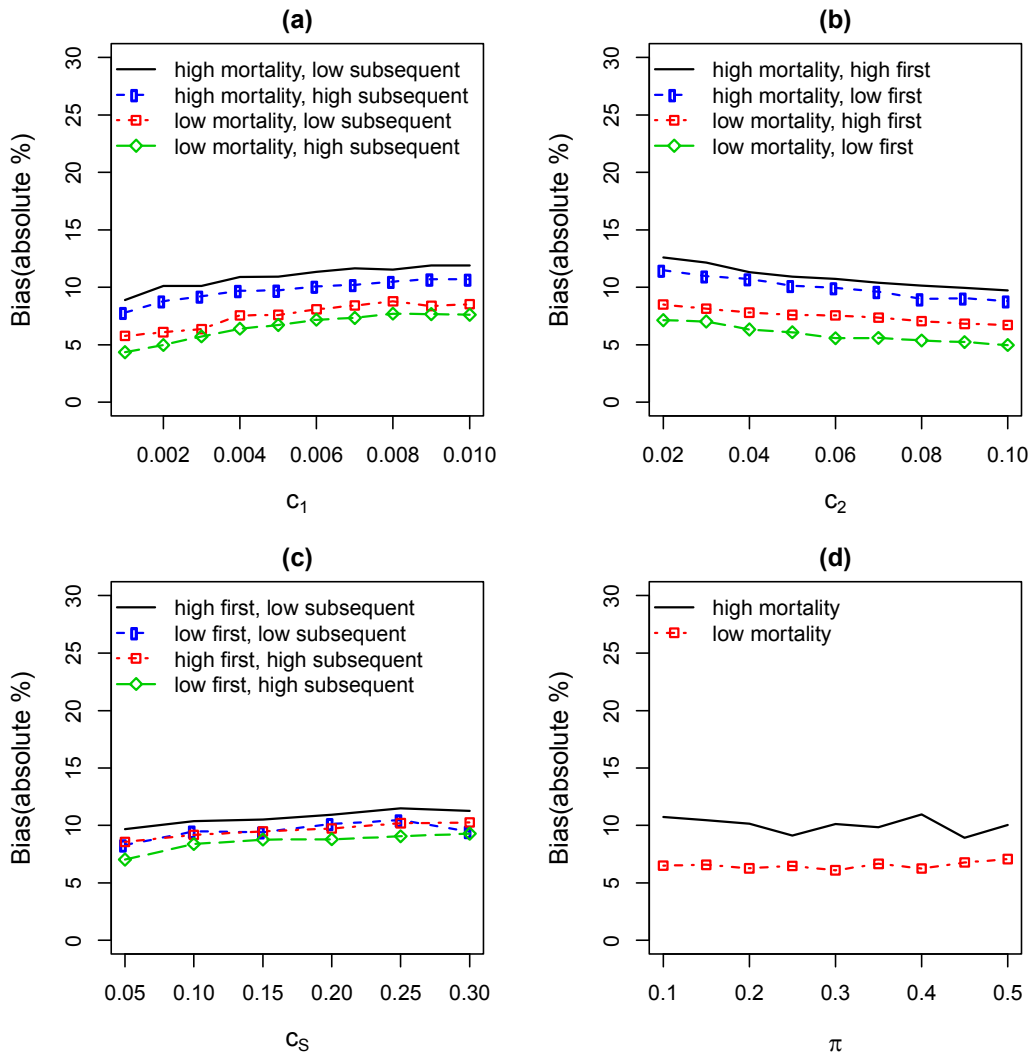
All results were obtained under scenario 1 (where the genetic variant associates with risk of first event, survival, and risk of subsequent event) when D_2 is time to event and the hazard ratios for G and X are 1.3 and 10, respectively. "high mortality": $c_S = 0.2$. "low mortality": $c_S = 0.0$. "high first": $c_1 = 0.005$. "low first": $c_1 = 0.002$. "high subsequent": $c_2 = 0.1$. "low subsequent": $c_2 = 0.05$. For (a)-(c), $\pi = 0.3$. For (d), $c_1 = 0.002$ and $c_2 = 0.05$.

Figure S3. Results of the estimated odds ratio (OR) for a genetic variant (scenario 1) in cohort and case-control studies
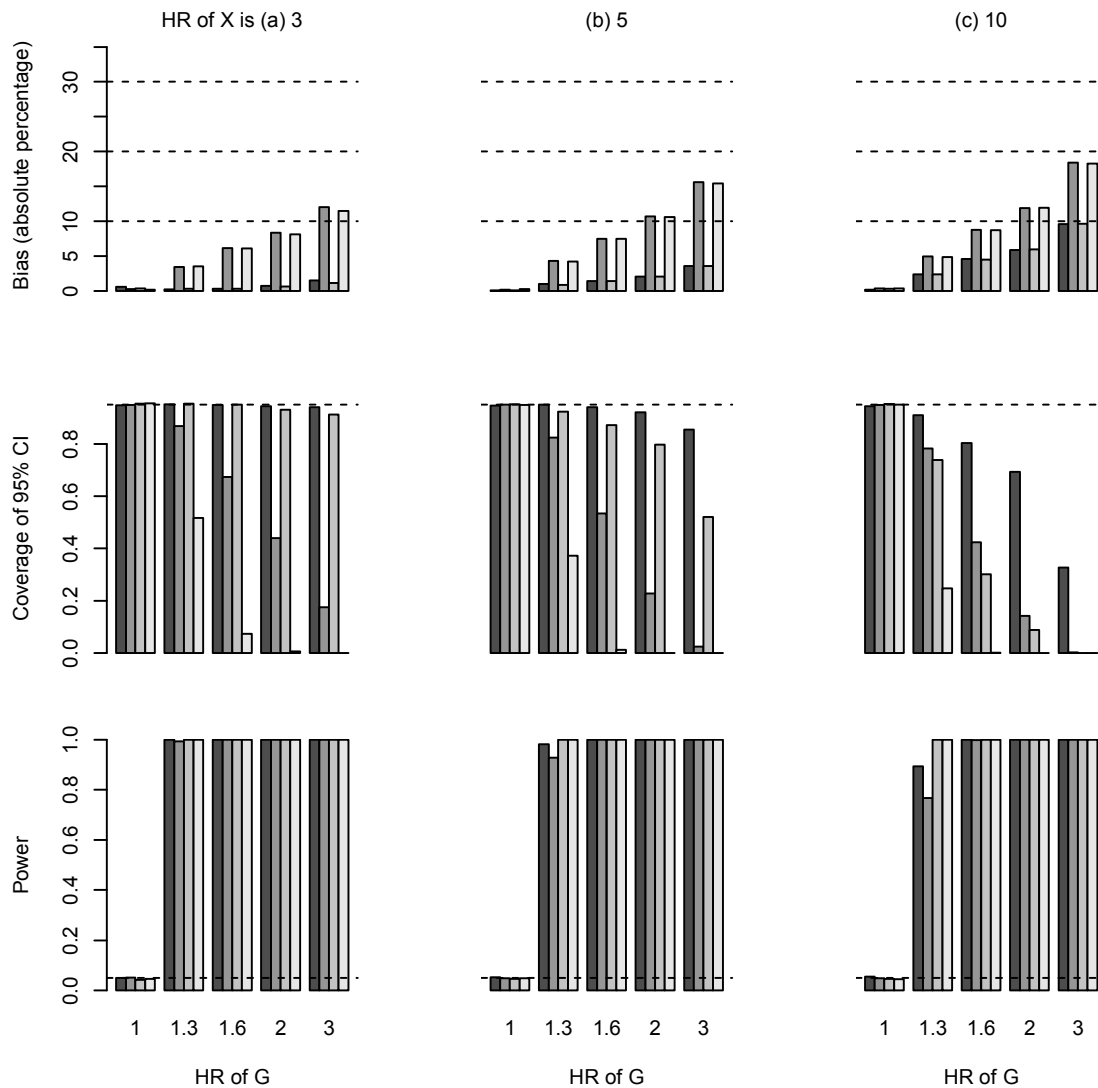
The sample size is 25000 for both study designs. The four bars (from left to right) at each HR of G pertain to population 2 (selection of subjects with fatal or non-fatal first events) in cohort studies, population 3 (selection of subjects with non-fatal first events) in cohort studies, population 2 in case-control studies, and population 3 in case-control studies. The dashed line in the middle panel indicates the expected coverage of 0.95. The dashed line in the lower panel indicates the nominal significance level of 0.05. Power under the HR of 1 for G means type 1 error.

12

Appendix

GENIUS-CHD Consortium full co-author list with affiliations to be supplied