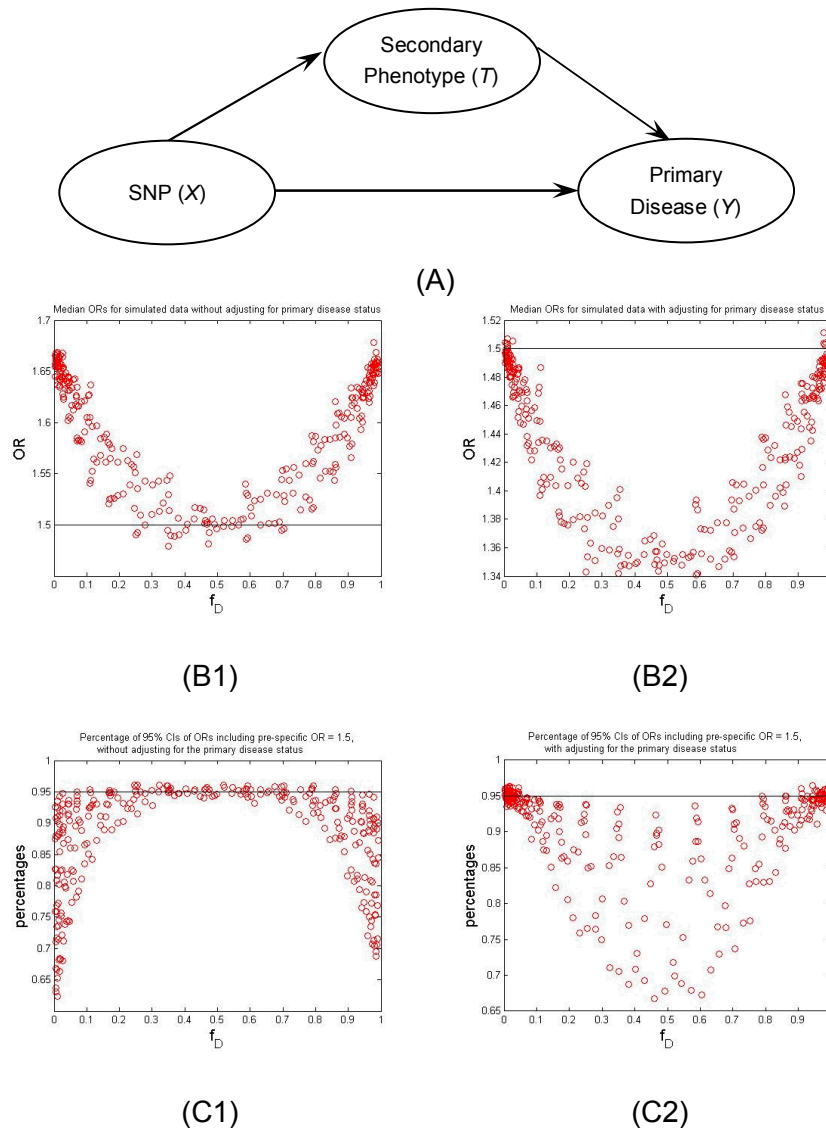


**SUPPLEMENTARY MATERIALS**

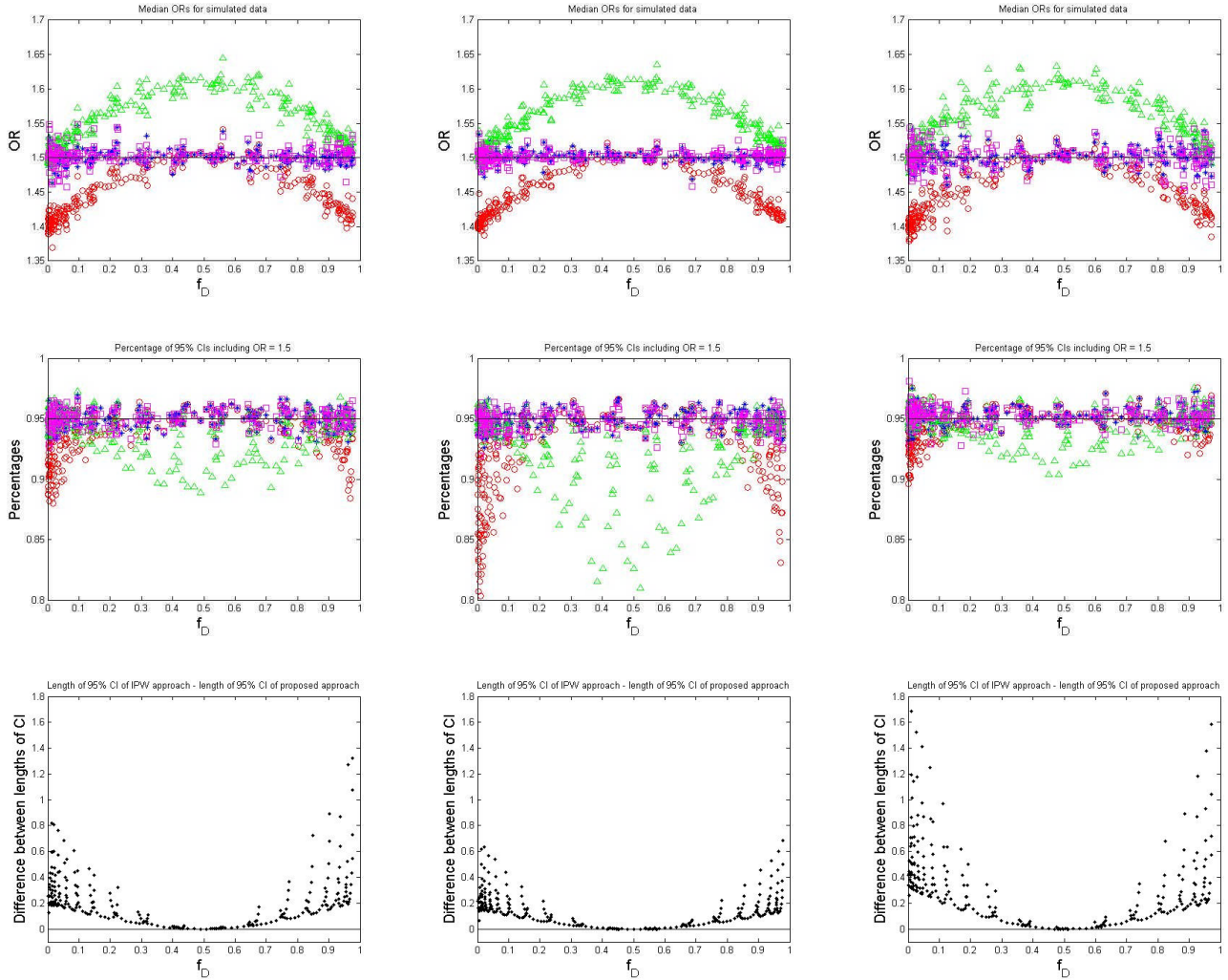
**Figure S1. Demonstration of bias using simulation studies. The results were based on 1,000 replicates, each with 1,000 cases and 1,000 controls. The true OR of the SNP associated with secondary phenotype was 1.5.**



SNP = single nucleotide polymorphisms  
 CI = confidence interval

(A) Network structure corresponding to the associations among the SNP, the primary disease, and the secondary phenotype. (B1) Median ORs obtained by using logistic regression without adjusting for the primary disease status. The straight line gives the pre-specified OR = 1.5. (B2) Median ORs obtained by using logistic regression with adjusting for the primary disease status. The straight line gives the pre-specified OR = 1.5. (C1) Percentage of 95% CIs including OR = 1.5 based on 1,000 replicates. The OR for each replicate was obtained without adjusting for primary disease status. (C2) Percentage of 95% CIs including OR = 1.5. The OR for each replicate was obtained with adjusting for primary disease status.

**Figure S2. Simulation results for three genetic models with protective secondary phenotype, based on 1,000 replicates, each with 1,000 cases and 1,000 controls. The true OR of the SNP associated with secondary phenotype was 1.5.**



(A) Dominant

(B) Additive

(C) Recessive

CI = confidence interval

Symbol “○” represents the results obtained by using logistic regression without adjusting for the primary disease status. Symbol “△” represents the results obtained by using logistic regression with adjusting for the primary disease status. Symbol “□” represents the results obtained by using the extended IPW regression approach. Symbol “\*” represents the results obtained by using the bias correction approach proposed in this article.