Robust Tests for Additive Gene-Environment Interaction in Case-Control Studies Using Gene-Environment Independence

Gang Liu, Seunggeun Lee, Alice W. Lee, Anna H. Wu, Elisa V. Bandera, Allan Jensen, Mary Anne Rossing, Kirsten B. Moysich, Jenny Chang-Claude, Jennifer Doherty, Aleksandra Gentry-Maharaj, Lambertus Kiemeney, Simon A. Gayther, Francesmary Modugno, Leon Massuger, Ellen L. Goode, Brooke Fridley, Kathryn L. Terry, Daniel W. Cramer, Susan J. Ramus, Hoda Anton-Culver, Argyrios Ziogas, Jonathan P. Tyrer, Joellen M. Schildkraut, Susanne K. Kjaer, Penelope M. Webb, Roberta B. Ness, Usha Menon, Andrew Berchuck, Paul D. Pharoah, Harvey Risch, Celeste Leigh Pearce and Bhramar Mukherjee

Correspondence to Dr. Bhramar Mukherjee, Department of Biostatistics and Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109 (email: bhramar@umich.edu)

Author affiliations: Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA (Gang Liu, Seunggeun Lee, and Bhramar Mukherjee); Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles, California, USA (Alice W. Lee, Anna H. Wu, Malcolm C. Pike, Celeste Leigh Pearce); Department of Cancer Epidemiology, Division of Population Sciences, Moffitt Cancer Center, Tampa, Florida, USA (Catherine M. Phelan); Cancer Prevention and Control, Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey, USA (Elisa V. Bandera); Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark (Allan Jensen, Susanne K. Kjaer); Program in

Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA (Mary Rossing); Department of Epidemiology, University of Washington, Seattle, Washington, USA (Mary Rossing); Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York, USA (Kirsten B. Moysich); Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany (Jennifer Chang-Claude); University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany (Jennifer Chang-Claude); Department of Epidemiology, The Geisel School of Medicine at Dartmouth, Hanover, NH, USA (Jennifer Doherty); Women's Cancer, Institute for Women's Health, University College London, London, United Kingdom (Aleksandra Gentry-Maharaj); Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands (Lambertus Kiemeney); Department of Obstetrics, Gynecology, and Reproductive Sciences, Division of Gynecologic Oncology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA (Francesmary Modugno); Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania, USA (Francesmary Modugno); Womens Cancer Research Program, Magee-Womens Research Institute and University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, USA (Francesmary Modugno); Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands (Leon Massuger); Department of Health Sciences Research, Division of Epidemiology, Mayo Clinic, Rochester, Minnesota, USA (Ellen L. Goode); Kansas IDeA Network of Biomedical Research Excellence Bioinformatics Core, University of Kansas Cancer Center, Kansas City, Kansas, USA (Brooke Fridley); Obstetrics and Gynecology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts,

USA (Kathryn L. Terry, Daniel W. Cramer); Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA (Kathryn L. Terry, Daniel W. Cramer); Genetic Epidemiology Research Institute, UCI Center for Cancer Genetics Research and Prevention, School of Medicine, Department of Epidemiology, University of California Irvine, Irvine, California, USA (Hoda Anton-Culver, Argyrios Ziogas); Department of Public Health and Primary Care, Center for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK (Jonathan Tyrer, Paul D. Pharoah); Department of Public Health Sciences, The University of Virginia, Charlottesville, Virginia, USA (Joellen M. Schildkraut); Department of Gynecology, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark (Susanne K. Kjaer); Queensland Institute of Medical Research, Brisbane, Australia (Penelope M. Webb); University of Texas School of Public Health, Houston, Texas, USA (Roberta B. Ness); Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA (Malcolm C. Pike); Department of Women's Cancer, EGA Institute for Women's Health, University College London, London, United Kingdom (Usha Menon); Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina, USA (Andrew Berchuck); Department of Oncology, Center for Cancer Genetic Epidemiology, University of Cambridge, University of Cambridge, Cambridge, United Kingdom (Paul D. Pharoah); Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA (Harvey Risch); and Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan, USA (Celeste Leigh Pearce).

This work was supported by: the US National Cancer Institute (R01 CA076016); the COGS project is funded through a European Commission's Seventh Framework Programme grant (agreement number 223175 HEALTH F2 2009-223175); the Genetic Associations and Mechanisms in Oncology (GAME-ON): a NCI Cancer Post-GWAS Initiative (U19-CA148112); the Ovarian Cancer Association Consortium is supported by a grant from the Ovarian Cancer Research Fund thanks to donations by the family and friends of Kathryn Sladek Smith (PPD/RPCI.07).

It was also supported by the National Institutes of Health (P30 CA14089, R01 CA61132, P01 CA17054, N01 PC67010, R03 CA113148, N01 CN025403, and R03 CA115195 [USC], K07 CA095666, R01 CA83918, K22 CA138563, and P30 CA072720 [NJO], R01 CA122443, P30 CA15083, and P50 CA136393R01 [MAY], R01 CA112523 and R01 CA87538 [DOV], R01 CA058860 [UCI]; R01 CA063678, R01 CA074850, and R01 CA080742 [CON], R01 CA76016 [NCO], R01 CA54419 and P50 CA105009 [NEC], R01 CA61107 [MAL], and R01 CA095023, R01 CA126841, M01 RR000056, P50 CA159981, and K07 CA80668 [HOP]); California Cancer Research Program (0001389V20170 and 2110200 [USC]); German Federal Ministry of Education and Research of Germany, Programme of Clinical Biomedical Research (01GB9401 [GER]); German Cancer Research Centre (GER); Danish Cancer Society (94 222 52 [MAL]); Mermaid I [MAL]; Eve Appeal/Oak Foundation (UKO); Cancer Institute of New Jersey (NJO); the National Institute for Health Research University College London Hospitals Biomedical Research Centre (UKO); US Army Medical Research and Materiel Command (W81XWH-10-1-02802 [NEC], DAMD17-02-1-0669 [HOP], DAMD17-02-1-0666 [NCO], and DAMD17-01-1-0729 [AUS]); Roswell Park Alliance Foundation (HOP); Cancer Councils of New South Wales, Victoria, Queensland,

South Australia and Tasmania (Multi-State Application Numbers 191, 211 and 182 [AUS]); Cancer Foundation of Western Australia (AUS); National Health and Medical Research Council of Australia (199600 and 400281 [AUS]); Mayo Foundation (MAY); Minnesota Ovarian Cancer Alliance (MAY); Fred C. and Katherine B. Andersen Foundation (MAY); Radboud University Medical Centre (NTH); Lon V Smith Foundation (LVS-39420 [UCI]); National Institute of Environmental Health Sciences (T32 ES013678 for A.W.L.); National Health and Medical Research Council of Australia (for G.C.T. and P.W.).

Research reported in this publication was partially supported by NCI award number P30 CA008748 (PI: Thompson) to Memorial Sloan Kettering Cancer Center. It was also supported by the National Cancer Institute of the National Institutes of Health (P30 CA046592). Lastly, this work was also supported by the National Science Foundation (NSF DMS 1406712) and the National Institutes of Health (NIH ES 20811). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: None declared.

Running head: Adaptive Test for Additive Interaction

Abbreviations:

CML: constrained maximum likelihood;

EB: empirical Bayes;

GRS: genetic risk score;

GWAS: genome-wide association studies;

LRT: likelihood ratio test;

MLE: maximum likelihood estimates;

OCP: oral contraceptive pill;

RERI: relative excess risk due to interaction;

SNP: single nucleotide polymorphism;

UML: unconstrained maximum likelihood;

WGRS: weighted genetic risk score.

ABSTRACT

There have been recent proposals advocating the use of additive gene-environment interaction instead of the widely used multiplicative scale, as a more relevant public health measure. Using gene-environment independence enhances the power for testing multiplicative interaction in case-control studies. However, under departure from this assumption, substantial bias in the estimates and inflated Type I error in the corresponding tests can occur. This paper extends the empirical Bayes (EB) approach previously developed for multiplicative interaction that trades off between bias and efficiency in a data-adaptive way, to the additive scale. An EB estimator of *Relative Excess Risk due to Interaction* is derived and the corresponding Wald test is proposed with general regression setting under a retrospective likelihood framework. We study the impact of gene-environment association on the resultant test with case-control data. Our simulation studies suggest that the EB approach uses the gene-environment independence assumption in a data-adaptive way and provides power gain compared to the standard logistic regression analysis and better control of Type I error when compared to the analysis assuming gene-environment independence. We illustrate the methods with data from the Ovarian Cancer Association Consortium.

Keywords: Bias-Variance Trade-off**,** Effect Modification, Empirical Bayes, Genetic Risk Score, Relative Excess Risk, Shrinkage.

Word Count: Abstract: 186 words, Body: 4460 words

INTRODUCTION

There has been an increasing interest in searching for gene by environment interaction (G x E) in the post genome-wide association studies (GWAS) era with limited success (1-5). A number of methods have been proposed for efficient search of G x E effects that use the gene-environment independence assumption (2, 6-10). Almost all of these studies have focused on testing/estimation of multiplicative interaction, perhaps due to the fact that standard logistic regression is the most commonly used tool for analyzing case-control data (11-13). However, it has been suggested in the literature that additive interaction is a more relevant public health measure (3, 14, 15). If the environmental exposure, say, E, can potentially be modified via an intervention, the additive gene x environment interaction measure can quantify the differences in the number of cases prevented if the intervention was offered in a prioritized way, across strata defined by genetic risk. This characterization helps with policy questions when limited access to an intervention are available. Moreover, the additive measure of interaction corresponds more closely to the notion of mechanistic or causal measures of interaction (16, 17). Although not commonly recognized, it is possible to *test* for additive interaction in a logistic regression model using case-control data. While a direct estimate of additive interaction on a risk difference scale cannot be obtained from case-control data, an alternative parameter, the *relative excess risk due to interaction* (RERI), can be represented in terms of relative risks. Assuming that the disease is rare, relative risks can be approximated by corresponding odds ratios and thus RERI can be viewed as a function of both main effects and multiplicative interaction parameters in a logistic regression model. Standard Delta theorem can be applied to provide asymptotic variance and subsequently a Wald test for the null hypotheses RERI=0 can be

conducted (18-20). The fact that RERI=0 if and only if the additive null holds provides us a way to test for interaction on the additive scale by testing H_0: RERI=0. More recently, Han et.al (21) developed a likelihood ratio test (LRT) for H_0: RERI=0, applying the retrospective likelihood framework proposed by Chatterjee and Carroll (22) that permits the incorporation of the G-E independence assumption, and leads to a more powerful test than the previously proposed Wald test, in modest sample sizes, for both the unconstrained and constrained ML method. However, it is not clear how to extend the LRT in an EB-type adaptive framework and thus we proceeded with combining estimates of RERI instead of deriving a combination LRT.

In this paper, we first consider the binary G, E scenario to illustrate our method for testing additive interaction in case-control studies. We provide closed form expressions of the maximum likelihood estimates (MLE) and Wald test of the RERI parameter without (unconstrained MLE) and with assuming gene-environment independence (constrained MLE). We then extend the empirical Bayes-type shrinkage approach for multiplicative G x E interaction proposed by Mukherjee et.al (6) to estimate RERI and test for additive interaction. An adaptively weighted estimator of RERI that combines the constrained and unconstrained estimators is proposed to trade-off between bias and efficiency. Finally, we extend the method to handle a completely general regression setting using the retrospective profile likelihood based framework in (22). We conduct a simulation study to compare the performance of various tests and illustrate our method by applying it to study the interaction between oral contraceptive pill (OCP) use and previously identified genetic factors in a large consortium of case-control studies of ovarian cancer.

METHODS

We first consider a simple setup of an unmatched case-control study with a dichotomous genetic factor G and a dichotomous environmental exposure E. Let E=1 (E=0) denote an exposed (unexposed) individual and G=1 (G=0) denote whether an individual is a carrier (non-carrier) of the susceptible genetic marker. Let D denote the disease status, where D=1 (D=0) stands for an affected (unaffected) individual. Let N_0 and N_1 be the number of selected controls and cases, respectively. The data can be represented in the form of a 2×4 table as displayed in Web Appendix 1.

Let $\boldsymbol{r_0} = (r_{01}, r_{02}, r_{03}, r_{04})$ and $\boldsymbol{r_1} = (r_{11}, r_{12}, r_{13}, r_{14})$ denote the vector of observed cell frequencies in the controls and the cases, respectively. Let $r_G = r_{03} + r_{04}$ denote the frequency of *G*=1 and $r_E = r_{02} + r_{04}$ denote the frequency of E=1 among controls. Let $\boldsymbol{p_0} = (p_{01}, p_{02}, p_{03}, p_{04})$ and $\boldsymbol{p_1} = (p_{11}, p_{12}, p_{13}, p_{14})$ denote the true population parameters of the cell probabilities corresponding to a particular G-E configuration in the underlying control and case populations respectively. Let $p_G = p_{03} + p_{04}$ denote the marginal prevalence of G=1 among controls and $p_E = p_{02} + p_{04}$ denote the marginal prevalence of E=1 among controls. The observed vectors of the cell counts can be viewed as random draws from two independent multinomial distributions in controls and cases respectively, namely, $\boldsymbol{r_0}$~Multinomial $(N_0, \boldsymbol{p_0})$ and $\boldsymbol{r_1}$~Multinomial$(N_1, \boldsymbol{p_1})$.

Let us introduce the following notation for the key parameters of interest. Let $OR_E = \frac{P(D=1|E=1,G=0)}{P(D=0|E=1,G=0)} \Big/ \frac{P(D=1|E=0,G=0)}{P(D=0|E=0,G=0)} = p_{01}p_{12}/p_{02}p_{11}$ denote the odds ratio associated with E for non-susceptible individuals (G=0), $OR_G = \frac{P(D=1|G=1,E=0)}{P(D=0|G=1,E=0)} \Big/ \frac{P(D=1|G=0,E=0)}{P(D=0|G=0,E=0)} = p_{01}p_{13}/p_{03}p_{11}$ denote the odds ratio associated with G for unexposed individuals (E=0) and $OR_{GE} =$

$\frac{P(D=1|E=1,G=1)}{P(D=0|E=1,G=1)} / \frac{P(D=1|E=0,G=0)}{P(D=0|E=0,G=0)} = p_{01}p_{14}/p_{04}p_{11}$ denote the joint odds ratio associated with the sub-group G=1 and E=1 compared to the reference group of G=0 and E=0. The multiplicative interaction parameter ψ is defined as:

$$\psi = \frac{OR_{GE}}{OR_G OR_E} = \frac{p_{02}p_{03}p_{11}p_{14}}{p_{01}p_{04}p_{12}p_{13}} = \frac{\frac{p_{11}p_{14}}{p_{12}p_{13}}}{\exp(\theta_{GE})}, \text{where } \theta_{GE} = \log\frac{p_{01}p_{04}}{p_{02}p_{03}}.$$

The parameter θ_{GE} represents the log odds ratio between G and E among the controls, characterizing the gene-environment association. In the additive scale, the measure of interaction is defined as:

$$\begin{aligned} \mathrm{p_{additive}} = {} & [\mathrm{P(D=1|E=1,G=1) - P(D=1|E=0,G=0)}] \\ & - [\mathrm{P(D=1|E=1,G=0) - P(D=1|E=0,G=0)}] \\ & - [\mathrm{P(D=1|E=0,G=1) - P(D=1|E=0,G=0)}] \\ = {} & \mathrm{P(D=1|E=1,G=1) - P(D=1|E=1,G=0) - P(D=1|E=0,G=1) +} \\ & \mathrm{P(D=1|E=0,G=0)} \quad (1) \end{aligned}$$

Dividing (1) throughout by $\mathrm{P(D=1|E=0,G=0)}$ we obtain a new measure relative excess risk due to interaction (RERI)

$$RERI_{RR} = RR_{GE} - RR_G - RR_E + 1. \qquad (2)$$

When the disease is rare, OR approximates RR. Hence, we have

$$RERI_{OR} \approx OR_{GE} - OR_G - OR_E + 1. \qquad (3)$$

Note that by (1) and (3), testing $\mathrm{H_0: p_{additive} = 0}$ is equivalent to testing $\mathrm{H_0}: RERI_{RR} = 0$, which is typically translated into $\mathrm{H_0}: RERI_{OR} = 0$ in a case-control study as described in VanderWeele (23). After defining the above relevant parameters of interest, we use the definition of RERI in

equation (3) in terms of ORs to proceed with inference under case-control sampling assuming the disease is rare for all configurations of G and E.

Unconstrained maximum likelihood estimation

The unconstrained maximum-likelihood (UML) estimate for all OR parameters mentioned above are obtained by simply substituting p_{dj} with its MLE, $\hat{p}_{dj} = r_{dj}/N_d$, implying,

$$\hat{\psi}_{uml} = \frac{\widehat{OR}_{GE}}{\widehat{OR}_G \widehat{OR}_E} = \frac{r_{02}r_{03}r_{11}r_{14}}{r_{01}r_{04}r_{12}r_{13}}, \qquad \hat{\sigma}^2_{uml} = Var(\log(\hat{\psi}_{uml})) = \sum_{d=0}^{1}\sum_{j=1}^{4}\frac{1}{r_{dj}}$$

The G-E association log odds ratio in controls can also be estimated as $\hat{\theta}_{GE} = \log\frac{r_{01}r_{04}}{r_{02}r_{03}}$.

The UML estimate of RERI can be easily obtained by plugging the corresponding estimated ORs in an unconstrained model into equation (3) and by the invariance property of MLE, serves as a consistent and asymptotically unbiased estimate of RERI regardless of the gene-environment independence assumption.

$$\widehat{RERI}_{uml} = \frac{r_{01}r_{14}}{r_{11}r_{04}} - \frac{r_{01}r_{13}}{r_{11}r_{03}} - \frac{r_{01}r_{12}}{r_{11}r_{02}} + 1 \quad (4)$$

Note that $\boldsymbol{r_0}$ and $\boldsymbol{r_1}$ are realizations from two independent multinomial distributions, and we can employ Delta method (Web Appendix 2) to obtain the asymptotic variance of $\widehat{RERI}_{\text{uml}}$, which is the same as noted in (17-19). The Wald test for interaction is based on the standardized Z statistic $Z_{uml} = \widehat{RERI}_{uml}/\sqrt{\widehat{Var}\left(\widehat{RERI}_{uml}\right)}$ which follows a N (0,1) distribution under the null RERI=0.

Constrained maximum likelihood estimation

Under G-E independence among controls, i.e. $\theta_{GE} = 0$ and rare disease assumptions, Zhang et.al (24) proposed the constrained MLEs (CML) for $\boldsymbol{p_0}$ and $\boldsymbol{p_1}$ as follows: $\hat{p}_{01} = \frac{(r_{01}+r_{03})(r_{01}+r_{02})}{N_0^2}$,

$\hat{p}_{02} = \frac{(r_{01}+r_{02})(r_{02}+r_{04})}{N_0^2}, \hat{p}_{03} = \frac{(r_{01}+r_{03})(r_{03}+r_{04})}{N_0^2}, \hat{p}_{04} = \frac{(r_{02}+r_{04})(r_{03}+r_{04})}{N_0^2}$ and $\hat{p}_{1j} = \frac{r_{1j}}{N_1}, j = 1,2,3,4.$

We obtain the corresponding OR estimates by substituting p_{dj} with its constrained MLE under G-E independence, $\widehat{OR}_E = \frac{r_{12}(r_{01}+r_{03})}{r_{11}(r_{02}+r_{04})}$, $\widehat{OR}_G = \frac{r_{13}(r_{01}+r_{02})}{r_{11}(r_{03}+r_{04})}$, $\widehat{OR}_{GE} = \frac{r_{14}(r_{01}+r_{02})(r_{01}+r_{03})}{r_{11}(r_{02}+r_{04})(r_{03}+r_{04})}$ and $\hat{\psi}_{\text{cml}} = \frac{r_{11}r_{14}}{r_{12}r_{13}}$, $\hat{\sigma}^2_{cml} = Var(\log(\hat{\psi}_{cml})) = \sum_{j=1}^{4} \frac{1}{r_{1j}}$. Note that the estimated multiplicative interaction parameter $\hat{\psi}$ is a function of only $\boldsymbol{r_1}$, and is identical to the case-only estimator.

The CML estimate of RERI can be computed by plugging the estimated ORs under the constraint into equation (3). Formally, the CML estimator for RERI is given by

$$\widehat{RERI}_{cml} = \frac{(r_{01}+r_{03})(r_{01}+r_{02})r_{14}}{(r_{02}+r_{04})(r_{03}+r_{04})r_{11}} - \frac{(r_{01}+r_{02})r_{13}}{(r_{03}+r_{04})r_{11}} - \frac{(r_{01}+r_{03})r_{12}}{(r_{02}+r_{04})r_{11}} + 1. \quad (5)$$

Under G-E independence assumption among controls, the CML estimator is consistent and asymptotically unbiased for the true RERI parameter. It is more precise than the UML estimator of RERI in equation (4) based on our simulations. The asymptotic variance of the CML estimator can also be approximated by Delta method, which is shown in Web Appendix 3. The Wald test for RERI in a constrained model again uses the standardized Z statistic $Z_{cml} = \widehat{RERI}_{cml} / \sqrt{\widehat{Var}\left(\widehat{RERI}_{cml}\right)}$, and the power of the test is slightly lower than LRT for additive interaction in (21) as will be illustrated through our simulations. Under violation of gene-environment independence assumption, $\theta_{GE} \neq 0$, the CML estimate is asymptotically biased for the true RERI parameter and the tests are invalid.

Empirical Bayes estimation

Mukherjee et.al (6) proposed an empirical Bayes (EB) estimator of the multiplicative interaction which shrinks the UML and CML estimators in a data-adaptive way. It relaxes G-E independence assumption and makes a trade-off between bias and efficiency. Formally, the EB estimator of multiplicative interaction is given by

$$\log(\hat{\psi}_{EB}) = \frac{\hat{\sigma}^2_{uml}}{\hat{\theta}^2_{GE}+\hat{\sigma}^2_{uml}}\log(\hat{\psi}_{cml}) + \frac{\hat{\theta}^2_{GE}}{\hat{\theta}^2_{GE}+\hat{\sigma}^2_{uml}}\log(\hat{\psi}_{uml}) \,, \quad (6)$$

where $\hat{\psi}_{cml} = \frac{r_{11}r_{14}}{r_{12}r_{13}}$, $\hat{\psi}_{uml} = \frac{r_{02}r_{03}r_{11}r_{14}}{r_{01}r_{04}r_{12}r_{13}}$, $\hat{\sigma}^2_{uml} = \sum_{d=0}^{1}\sum_{j=1}^{4}\frac{1}{r_{dj}}$ and $\hat{\theta}_{GE} = \log\frac{r_{01}r_{04}}{r_{02}r_{03}}$.

We employ the same idea of adaptive weighting and propose the EB estimator for RERI as,

$$\widehat{RERI}_{EB} = \frac{(\widehat{RERI}_{uml}-\widehat{RERI}_{cml})^2}{\widehat{Var}(\widehat{RERI}_{uml})+(\widehat{RERI}_{uml}-\widehat{RERI}_{cml})^2}\widehat{RERI}_{uml} + \frac{\widehat{Var}(\widehat{RERI}_{uml})}{\widehat{Var}(\widehat{RERI}_{uml})+(\widehat{RERI}_{uml}-\widehat{RERI}_{cml})^2}\widehat{RERI}_{cml}$$

$$= \widehat{RERI}_{uml} + K(\widehat{RERI}_{cml} - \widehat{RERI}_{uml}) \quad (7),$$

where $K = V(V + \hat{\kappa}\hat{\kappa}^T)^{-1}$ is a shrinkage factor of the same form as defined in Chen et.al (25) with $\hat{\kappa} = \widehat{RERI}_{uml} - \widehat{RERI}_{cml}$ and $V = \widehat{Var}\left(\widehat{RERI}_{uml}\right)$. To explain the intuitive rationale behind the estimator, observe that as $\widehat{\theta_{GE}} \to 0$, i.e. as the data provide the evidence in favor of G-E independence, $\widehat{RERI}_{uml} - \widehat{RERI}_{cml} \to 0$, the estimator puts more weight on CML estimator to gain more efficiency, and as $\widehat{\theta_{GE}} \to \infty$. i.e. as the G-E dependence becomes stronger in control population, $\widehat{RERI}_{uml} - \widehat{RERI}_{cml}$ becomes larger, then the EB estimator puts more weight on UML estimator to reduce bias. In large samples, the EB estimator converges to the UML estimate and thus is asymptotically unbiased for the true RERI parameter (6). The asymptotic variance of $\widehat{RERI}_{EB}$ is derived by Delta method (See Web Appendix 4), assuming $\widehat{Var}\left(\widehat{RERI}_{uml}\right)$ as a constant relative to the order of magnitude of the point estimates (6). We use Wald test for the EB estimator based on the standardized Z statistic $Z_{EB} = \widehat{RERI}_{EB}/\sqrt{\widehat{Var}\left(\widehat{RERI}_{EB}\right)}$.

Remark 1. We also considered two other forms of adaptive weights. One is to modify the shrinkage factor *K* in (7) and let $\widehat{k^*} = \hat{\theta}_{GE}$ instead of $\widehat{RERI}_{uml} - \widehat{RERI}_{cml}$, namely, $\widehat{RERI}_{EB1} = \widehat{RERI}_{uml} + K^*(\widehat{RERI}_{cml} - \widehat{RERI}_{uml})$, where $K^* = V\left(V + \widehat{\kappa^*}\widehat{\kappa^*}^T\right)^{-1}$. The other is to plug in the EB estimates, $\widehat{OR}_{EB}$, obtained from using the retrospective likelihood framework in (6) as implemented in R package CGEN (6, 22, 25) directly into equation (3), namely, $\widehat{RERI}_{EB2} = \widehat{OR}_{GE} - \widehat{OR}_G - \widehat{OR}_E + 1$, where all estimated ORs are EB estimates proposed under the multiplicative model. The EB estimator we proposed in equation (7) demonstrates superior performance among the three choices, based on our simulation study.

Remark 2: As shown in Chen et.al (25), the asymptotic theory for CML and consequently EB is non-regular under the independence assumption. The Delta method does not technically apply for estimating the asymptotic variance. Theoretically, the test statistic also fails to be asymptotically normal under G-E independence (25, 26). However, in practice, the estimated variance derived by the Delta Method approximates the empirical variance very well as noted in the simulation studies (see Web Appendix 5, Web Tables 1-2 and Web Figures 1-2). Under G-E dependence, EB estimate converges in large sample to UML estimate and thus to the true RERI parameter and standard likelihood asymptotics holds (6).

Profile likelihood framework for general regression setting

Consider the retrospective likelihood considered in Chatterjee and Carroll (22), Mukherjee et.al (6) and as implemented in the R package CGEN:

$$P(G, E, \boldsymbol{Z}|D) = \frac{P(D = 1|G, E, \boldsymbol{Z})P(G|E, \boldsymbol{Z})P(E, \boldsymbol{Z})}{\sum_{G,E,\boldsymbol{Z}} P(D = 1|G, E, \boldsymbol{Z})P(G|E, \boldsymbol{Z})P(E, \boldsymbol{Z})} \qquad (8)$$

The three ingredients of the above retrospective likelihood are:

(a) The logistic regression disease risk model of interest with multiplicative GEI parameter:

$\text{logit P}(D = 1|G, E, \boldsymbol{Z}) = \beta_0 + \beta_G G + \beta_E E + \beta_{GE} G \times E + \boldsymbol{\beta_Z^T Z}$, where $\boldsymbol{Z}$ denotes other covariates.

(b) $\text{logit P}(G|E, \boldsymbol{Z}) = \theta_0 + \theta_{GE} E + \boldsymbol{\theta_{GZ}^T Z}$. While this is the gene model used for UML, allowing G-E dependence, in the CML method, $\text{P}(G|E, \boldsymbol{Z})$ reduces to $\text{P}(G|\boldsymbol{Z})$ under the assumption of G-E independence conditional on $\boldsymbol{Z}$, implying $\theta_{GE} \equiv 0$.

(c) The distribution $\text{P}(E, \boldsymbol{Z})$ is allowed to be completely non-parametric. We then maximize the retrospective likelihood using existing routines in CGEN to obtain $\widehat{\beta}_{uml}$ and $\widehat{\beta}_{cml}$, the vector of all the parameter estimates of the disease risk model in (a), namely, $(\beta_0, \beta_G, \beta_E, \beta_{GE}, \boldsymbol{\beta_Z})$.

When it comes to defining RERI with a general *G* and *E* variable adjusting for covariates **Z**, particularly with case-control data, as described in VanderWeele (23), let us denote by $RERI_{OR}(E_0, E_1, G_0, G_1)$ the relative excess risk due to interaction by replacing risk ratios with corresponding odds ratios in the RERI expression in (3) as typically done in a case-control study. With general continuous and ordinal exposures one has to consider the magnitude of change in exposure for which one is examining the interaction. Let us consider the situation when environmental risk factor changes from E_0 to E_1 and genetic risk factor changes from G_0 to G_1 but other covariates $\boldsymbol{z}$ are held constant. Formally, it is defined as

$$
\begin{aligned}
&RERI_{OR}(E_0, E_1, G_0, G_1) \\
&= OR(G_1, E_1) - OR(G_1, E_0) - OR(G_0, E_1) + 1 \\
&= \exp\{\beta_G(G_1 - G_0) + \beta_E(E_1 - E_0) + \beta_{GE}(G_1 \times E_1 - G_0 \times E_0)\} \\
&\qquad - \exp\{\beta_E(E_1 - E_0) + \beta_{GE} G_0 \times (E_1 - E_0)\} \\
&\qquad - \exp\{\beta_G(G_1 - G_0) + \beta_{GE}(G_1 - G_0) \times E_0\} + 1
\end{aligned}
$$

$= f(\beta_G, \beta_E, \beta_{GE}) \approx RERI(E_0, E_1, G_0, G_1)$ (9)

This last approximation of risk ratios by odds ratios holds when the outcome is rare in each stratum defined by the two exposures or when controls are selected from the entire population, not just the non-cases (27). More generally, if G and E are both categorical factors with I and J levels with coefficients corresponding to different levels of each factor, then $\beta_G, \beta_E, \beta_{GE}$ in equation (9) become (I-1), (J-1) and (I-1)(J-1) dimensional vectors instead of scalars. Note that $\widehat{RERI}_{uml} = f(\widehat{\boldsymbol{\beta}}_{uml})$ and $\widehat{RERI}_{cml} = f(\widehat{\boldsymbol{\beta}}_{cml})$, can be viewed as function of UML and CML estimates of relative risk parameters, where *f* is the function in equation (9). The variance of $\widehat{RERI}_{uml}$ and $\widehat{RERI}_{cml}$ can be calculated by Delta method. The EB estimator of RERI is same as in equation (7) and its estimated variance is calculated by Delta method using the joint distribution of $(\widehat{\boldsymbol{\beta}}_{\boldsymbol{uml}}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{cml}})$ as proposed by Mukherjee et.al (6) (Web Appendix 6). The Wald tests for the three estimators are all based on the standardized Z statistic. We have provided general codes to test for RERI at (28).

Example: Analysis of G x E interactions in case-control studies of ovarian cancer

Epithelial ovarian cancer is one of the most common malignancies of the female reproductive tract. Approximately 14,240 women died from ovarian cancer in 2016 in the United States, causing more deaths than any other cancer of the female reproductive system. There are several well-established non-genetic risk factors for ovarian cancer (29-35), and recent genome-wide association studies have identified and replicated 18 variants that influence disease risk (36). To this end, the Ovarian Cancer Association Consortium (OCAC) has undertaken an effort to study interactions focusing on the 18 confirmed single nucleotide polymorphisms (SNPs) and seven

well-established risk factors: race, history of endometriosis, first degree family history of ovarian cancer, oral contraceptive pill (OCP) use, parity, tubal ligation, and age. In our illustrative analysis, we focus on OCP x SNP interaction and use genetic data from 15 OCAC studies that also have data on epidemiologic risk factors.

Each SNP is coded as the number of risk alleles a subject carried and all subsequent analysis assumed this additive genetic susceptibility model. Published ORs of the 18 confirmed loci in Web Table 3 are from analyses presented in Collaborative Oncological Gene-Environment Study (37-44). As a parsimonious and succinct way of summarizing the effects of genetic variants across all loci for each subject, we construct a "genetic risk score" (GRS) variable as the sum of the risk allele counts across all loci and a “weighted genetic risk score” (WGRS) as the weighted sum, where the weight for each individual SNP is determined by the published log OR in large meta-analysis. Polygenic risk scores have been used for risk stratification in multiple G x E papers recently (3,45). Analysis of marginal effect for GRS and WGRS is shown in Web Table 4. Each environmental factor is coded as a categorical variable as described in Web Table 5. The merged G × E dataset has a sample size of 11,661 subjects with European ancestry, with 4,135 cases and 7,526 controls from 13 study sites (Web Table 6).

To illustrate our inference for interactions between OCP use (1 =ever and 0 =never) and genetic risk factors we consider both single SNP x OCP and (W)GRS x OCP interaction. For single SNP analysis, we consider the top two hits in the 18 confirmed loci, i.e. rs62274042 (SNP1) and rs10962691 (SNP2) as reported in Web Table 3. We used additive coding for our SNP x OCP analysis. For GRS and WGRS, we use the quartiles in controls to define a categorical variable with

four categories. The analysis model adjusts for study site and all other environmental risk factors except race.

Simulation design

In our simulation study, we first investigate the Type I error, standard power at level α and power at empirical α (empirical Type I error is used to report power in situations where Type I error is not maintained) of Wald tests for $\widehat{RERI}_{uml}$, $\widehat{RERI}_{cml}$ and $\widehat{RERI}_{EB}$ under various alternative values of RERI across a spectrum of scenarios, varying the strength of G-E association, main effects of G and E, minor allele frequency of G, prevalence of exposure E, test size and sample sizes. We compare the power of Wald test for $\widehat{RERI}_{cml}$ with the previously proposed LRT for additive interaction under G-E independence (21). We also explore estimation properties like the absolute relative bias and MSE of the three estimators as well as those of two alternative proposals, $\widehat{RERI}_{EB1}$ and $\widehat{RERI}_{EB2}$. Note that both RERI and multiplicative interaction parameters are obtained from the underlying true logistic regression model

$$\text{logit P}(\text{D} = 1|G, E) = \beta_0 + \beta_E E + \beta_G G + \beta_{GE} GE,$$

where RERI=$\exp(\beta_G + \beta_E + \beta_{GE}) - \exp(\beta_G) - \exp(\beta_E) + 1$, and $\psi = \exp(\beta_{GE})$, so that the value of RERI is well-defined given ψ and vice versa, once the main effect parameters $OR_G = \exp(\beta_G)$ and $OR_E = \exp(\beta_E)$ are specified.

We set prevalence of G and E in controls, $p_G = (0.1, 0.2, 0.3)$ and $p_E = (0.3, 0.4, 0.5)$; the main effects $OR_G = (1.1, 1.2, 1.3)$; $OR_E = (1.3, 1.5, 1.7)$; sample size $N_0 = N_1 = (4000, 20000)$; size of test $\alpha = (0.05, 5\times10^{-6})$; the strength of G-E association, $\exp(\theta_{GE})$, change from 0.8 to 1.2 at a grid of 0.1 and RERI change from 0 to 1.5 with a grid of 0.1. The number of simulated

datasets is 1000 when $\alpha = 0.05$ and is 10^6 when $\alpha = 5\times10^{-6}$. The population parameters of cell probability $\boldsymbol{p_0}$ and $\boldsymbol{p_1}$ are defined by solving the equations in Web Appendix 7 (9, 46):

We generate data independently from the two multinomial distributions corresponding to the case and control populations, according to the above probabilities with number of cases and control as $N_0,\ N_1$, respectively. We also considered another simulation setting to mimic a large-scale genomewide search of interactions where we use random distribution for the parameters corresponding to the set of null markers. We first compute the UML, CML and EB estimators using equations (4), (5), and (7) and then compare their Type I error, power, power at empirical α , absolute relative bias and MSE. Type I error over 1000 replications. Power are estimated by the proportion of null hypothesis $\mathrm{H_0}$: $RERI = 0$ rejected at the given level of significance α, i.e. the proportion of times $|Z|>Z_{1-\alpha/2}$, where Z is Wald test statistic. Power at empirical α is a modified power which utilizes an empirical P value threshold as the rejection rule to control the Type I error around the given significance level when the Type I error at the desired nominal level is not maintained. The absolute relative bias is calculated by averaging $\left|\widehat{RERI} - RERI\right|/RERI$ and MSE is calculated by averaging $\left(\widehat{RERI} - RERI\right)^2$.

RESULTS

Ovarian cancer data example

The distributions of GRS and WGRS in cases and controls are displayed in Web Figure 3. Relative to the control distributions, the upper tails of the case distributions are shifted slightly rightward. We calculate UML, CML and EB estimators of interactions in both multiplicative and additive scale. The estimates, corresponding CIs and P-values of Wald test are shown in Table 1. In SNP1×OCP analysis, the strength of G-E association is modest: $\exp(\theta_{\mathrm{GE}})$ =1.07 (95% CI

[0.94,1.21]), EB estimate of RERI is -0.16 with 95% CI [-0.50,0.18], where the weight on $\widehat{RERI}_{uml}$ is 43%. In SNP2×OCP analysis, the G-E association seems weaker with $\exp(\theta_{\mathrm{GE}})$=0.96 (95% CI [0.83,1.11]). EB estimate of RERI is 0.04 with 95% CI [-0.11,0.18], with its weight on $\widehat{RERI}_{uml}$ decreasing to 11%. The confidence intervals corresponding to $\widehat{RERI}_{EB}$ are narrower compared to the corresponding intervals for $\widehat{RERI}_{uml}$. The point estimate $\widehat{RERI}_{EB}$ lies between $\widehat{RERI}_{uml}$ and $\widehat{RERI}_{cml}$, reflecting the combined efficiency-robustness feature of the EB estimator. In WGRS×OCP analysis we report interactions associated with a change of OCP from 0 to 1 (ever users to never users) and WGRS from the lowest to the highest quartile (as defined through distribution of WGRS in controls) the multiplicative measure of interaction $\hat{\psi}_{EB}$ is not significant at α=0.05 but $\widehat{RERI}_{EB}$ departs from 0 significantly with EB estimate of RERI -0.52(95% CI [-0.91, -0.13]) and has a very small P-value, 0.009.

To visually present the results, we fit a standard logistic regression model including the main effects of OCP use and quartiles of WGRS as a categorical factor, and an interaction term for WGRS×OCP adjusting for study sites and other risk factors. Figure 1 shows the odds ratio of OCP and corresponding CI stratified by WGRS. The odds ratio of OCP is 0.61 (0.50,0.74) in the lowest WGRS quartile and 0.51 (0.43,0.60) in the highest quartile. The overlapping CIs indicate a non-significant multiplicative interaction. Additionally, if we assume that approximately 1.3 percent of women will be diagnosed with ovarian cancer at some point during their lifetime (47) and 70% women will use OCP at some point in their life in this population (estimated from the OCAC data), we present the estimated lifetime risk of ovarian cancer and corresponding 95% CI within each WGRS stratum in Figure 2, for OCP users and non-users. Estimates of lifetime absolute risk for OCP users is 0.75% (0.57%, 0.98%) and 1.23% (1.00%, 1.51%) for OCP non-users in the lowest

WGRS stratum with a difference of 0.48% (0.02%, 0.94%) and the corresponding numbers were 1.40% (1.08%, 1.81%) and 2.72% (2.05%, 3.60%) with a difference of 1.32% (0.24%, 2.52%) for subjects in the highest WGRS stratum, showing why the test for RERI is significant.

Results from the Simulation Study

Type I error. Web Table 7 presents Type I errors for different tests of RERI. One can observe that UML maintains nominal level α across different choices of θ_{GE}. An inflated Type I error associated with CML is observed when G-E independence assumption is violated. EB test is valid when $\exp(\theta_{\mathrm{GE}})$=1 and has a modest inflation on Type I error when G is associated with E. The maximal observed Type I error of EB at α=0.05 is 0.099 when sample size is 40,000, test size is 0.05 and $\exp(\theta_{\mathrm{GE}})$=1.1. Web Figure 4 presents how Type I error varies with $\exp(\theta_{\mathrm{GE}})$ for the three estimators. The Type I error of CML is very sensitive to the G-E association but the performance of EB is relatively robust with marked reduction in Type I error compared to CML. The findings remain similar for different choices of p_G, p_E, OR_G and OR_E (Web Tables 8-9).

Results from additional simulation mimicking a Genomewide Association Study: To justify the use of EB estimator in genomewide assessment of G-E interaction, we conduct another simulation study similar to that in Reference (8), which generates 2000 cases and controls with 1 causal marker together with M-1 null markers where M is 10,000. G-E independence parameter θ_{GE} in controls have a random mixture distribution with point mass around independence and $\mathrm{p_{ind}}$ is the proportion of null loci that follow G-E independence. The detailed simulation setting is presented in Web Appendix 8. The expected nominal level for both familywise error rate and expected number of false positives is 0.05 when G-E independence holds. However, if there is G-E dependence for a proportion of markers, Bonferroni correction cannot guarantee the nominal

level for EB and CML. As shown in Table 2, when 99% of the markers are independent, EB maintains familywise Type I error rate of 0.06 and expected number of false positives of 0.06. The performance of CML is significantly worse with familywise error rate of 99% and expected number of false positives 3.76.

Power. Figure 3 shows the power curves of Wald test for three estimators with $\mathrm{H}_0: RERI = 0$ under different strengths of G-E association (Web Tables 10-15). It is hard to compare the estimated powers directly from the figure as the inflated Type I error of CML and EB leads to the misleading high power values. Hence, we assess the power at empirical α for CML and EB, which controls the corresponding Type I error at 0.05. UML is the most efficient when $\exp(\theta_{\mathrm{GE}})$=0.8, CML is the most efficient when $\exp(\theta_{\mathrm{GE}})$=1 and 1.2, and EB power always lies in between. For a sample numerical comparison, let us compare the powers of the three approaches at RERI=0.5 to represent one typical scenario. When $\exp(\theta_{\mathrm{GE}})$=0.8, the empirical power of EB (0.275) is 41% lower than UML (0.672), meanwhile CML has nearly 0 power. When $\exp(\theta_{\mathrm{GE}})$=1, the empirical power of EB (0.870) is 25% higher than UML (0.693) but 10% lower than CML (0.970). When $\exp(\theta_{\mathrm{GE}})$=1.2, the empirical power of EB (0.718) is slightly higher than UML (0.714) but 28% lower than CML (0.993). We then compare the power of Wald test for $\widehat{RERI}_{cml}$ with LRT for additive interaction shown in Web Figure 5. The power of LRT is uniformly slightly higher than the Wald test with true value of RERI varying from 0 to 0.5 with a grid of 0.1. Absolute relative bias and MSE results are relegated to Web Appendix 9, Web Tables 16-19, Web Figure 6.

DISCUSSION

In this paper, we extend the EB estimator of gene-environment interaction proposed earlier on the multiplicative scale to additive scale in case-control studies. The EB estimator exploits G-E independence assumption to perform a trade-off between bias and efficiency. The simulation study showed that the test based on the EB estimator can provide a good control of Type I error and it is always intermediate between UML and CML with respect to power, relative bias and mean squared error. In the ovarian cancer data example, we conducted a (W)GRS×OCP analysis to illustrate the application of the proposed method. We found a significant additive (W)GRS×OCP interaction but insignificant multiplicative interaction at α=0.05.

As an inherent limitation of case-control studies, only the relative risk can be estimated, e.g. RERI, instead of the underlying direct measure, e.g. $\mathrm{p}_{\mathrm{additive}}$ in equation (1), because p_{11} can only be estimated from cohort data. However, general population incidence data from cohort studies can be combined with case-control risk-factor models to estimate absolute risks in population-based case-control studies (48), as we carried out in Figure 2. If the rare disease assumption for each configuration of G and E does not hold, approximating RR by OR in case-control studies will not be accurate and thus the proposed estimate of RERI may depart from the truth. By using the retrospective maximum likelihood estimates, using prior guesses for disease prevalence and adaptive combinations like EB procedure we can make our inference less biased under violation of the rare disease and gene-environment independence assumptions.

There is increasingly more interest in inference for additive interaction using case-control data. Tchetgen –Tchetgen et.al (49) described a general approach to test for G x E additive interaction exploiting G-E independence which is robust to possible misspecification of main effects in the

outcome regression. Han et.al (50) proposed a score test for UML and CML estimators of genetic associations under the additive null. In the future, it is of analytical interest to establish an EB version of adaptive score test and adaptive LRT as most of the recent work has been in terms of combining estimators but not tests.

ACKNOWLEDGEMENTS

Author affiliations: Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA (Gang Liu, Seunggeun Lee, and Bhramar Mukherjee); Department of Preventive Medicine, University of Southern California Keck School of Medicine, Los Angeles, California, USA (Alice W. Lee, Anna H. Wu, Malcolm C. Pike, Celeste Leigh Pearce); Department of Cancer Epidemiology, Division of Population Sciences, Moffitt Cancer Center, Tampa, Florida, USA (Catherine M. Phelan); Cancer Prevention and Control, Rutgers Cancer Institute of New Jersey, New Brunswick, New Jersey, USA (Elisa V. Bandera); Department of Virus, Lifestyle and Genes, Danish Cancer Society Research Center, Copenhagen, Denmark (Allan Jensen, Susanne K. Kjaer); Program in Epidemiology, Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA (Mary Rossing); Department of Epidemiology, University of Washington, Seattle, Washington, USA (Mary Rossing); Department of Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York, USA (Kirsten B. Moysich); Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany (Jennifer Chang-Claude); University Cancer Center Hamburg (UCCH), University Medical Center Hamburg-Eppendorf, Hamburg, Germany (Jennifer Chang-Claude); Department of Epidemiology, The Geisel School of Medicine at Dartmouth, Hanover, New Hampshire, USA (Jennifer Doherty);

Women's Cancer, Institute for Women's Health, University College London, London, United Kingdom (Aleksandra Gentry-Maharaj); Radboud University Medical Center, Radboud Institute for Health Sciences, Nijmegen, The Netherlands (Lambertus Kiemeney); Department of Obstetrics, Gynecology, and Reproductive Sciences, Division of Gynecologic Oncology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA (Francesmary Modugno); Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania, USA (Francesmary Modugno); Womens Cancer Research Program, Magee-Womens Research Institute and University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania, USA (Francesmary Modugno); Radboud University Medical Center, Radboud Institute for Molecular Life Sciences, Nijmegen, The Netherlands (Leon Massuger); Department of Health Sciences Research, Division of Epidemiology, Mayo Clinic, Rochester, Minnesota, USA (Ellen L. Goode); Kansas IDeA Network of Biomedical Research Excellence Bioinformatics Core, University of Kansas Cancer Center, Kansas City, Kansas, USA (Brooke Fridley); Obstetrics and Gynecology Center, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA (Kathryn L. Terry, Daniel W. Cramer); Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA (Kathryn L. Terry, Daniel W. Cramer); Genetic Epidemiology Research Institute, UCI Center for Cancer Genetics Research and Prevention, School of Medicine, Department of Epidemiology, University of California Irvine, Irvine, California, USA (Hoda Anton-Culver, Argyrios Ziogas); Department of Public Health and Primary Care, Center for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Cambridge, UK (Jonathan Tyrer, Paul D. Pharoah); Department of Public Health Sciences, The University of Virginia, Charlottesville, Virginia, USA (Joellen M. Schildkraut); Department of Gynecology,

Rigshospitalet, University of Copenhagen, Copenhagen, Denmark (Susanne K. Kjaer); Queensland Institute of Medical Research, Brisbane, Australia (Penelope M. Webb); University of Texas School of Public Health, Houston, Texas, USA (Roberta B. Ness); Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA (Malcolm C. Pike); Department of Women's Cancer, EGA Institute for Women's Health, University College London, London, United Kingdom (Usha Menon); Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, North Carolina, USA (Andrew Berchuck); Department of Oncology, Center for Cancer Genetic Epidemiology, University of Cambridge, University of Cambridge, Cambridge, United Kingdom (Paul D. Pharoah); Department of Chronic Disease Epidemiology, Yale School of Public Health, New Haven, Connecticut, USA (Harvey Risch); and Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan, USA (Celeste Leigh Pearce).

This work was supported by: the US National Cancer Institute (R01 CA076016); the COGS project is funded through a European Commission's Seventh Framework Programme grant (agreement number 223175 HEALTH F2 2009-223175); the Genetic Associations and Mechanisms in Oncology (GAME-ON): a NCI Cancer Post-GWAS Initiative (U19-CA148112); the Ovarian Cancer Association Consortium is supported by a grant from the Ovarian Cancer Research Fund thanks to donations by the family and friends of Kathryn Sladek Smith (PPD/RPCI.07).

It was also supported by the National Institutes of Health (P30 CA14089, R01 CA61132, P01 CA17054, N01 PC67010, R03 CA113148, N01 CN025403, and R03 CA115195 [USC], K07 CA095666, R01 CA83918, K22 CA138563, and P30 CA072720 [NJO], R01 CA122443, P30

CA15083, and P50 CA136393R01 [MAY], R01 CA112523 and R01 CA87538 [DOV], R01 CA058860 [UCI]; R01 CA063678, R01 CA074850, and R01 CA080742 [CON], R01 CA76016 [NCO], R01 CA54419 and P50 CA105009 [NEC], R01 CA61107 [MAL], and R01 CA095023, R01 CA126841, M01 RR000056, P50 CA159981, and K07 CA80668 [HOP]); California Cancer Research Program (0001389V20170 and 2110200 [USC]); German Federal Ministry of Education and Research of Germany, Programme of Clinical Biomedical Research (01GB9401 [GER]); German Cancer Research Centre (GER); Danish Cancer Society (94 222 52 [MAL]); Mermaid I [MAL]; Eve Appeal/Oak Foundation (UKO); Cancer Institute of New Jersey (NJO); the National Institute for Health Research University College London Hospitals Biomedical Research Centre (UKO); US Army Medical Research and Materiel Command (W81XWH-10-1-02802 [NEC], DAMD17-02-1-0669 [HOP], DAMD17-02-1-0666 [NCO], and DAMD17-01-1-0729 [AUS]); Roswell Park Alliance Foundation (HOP); Cancer Councils of New South Wales, Victoria, Queensland, South Australia and Tasmania (Multi-State Application Numbers 191, 211 and 182 [AUS]); Cancer Foundation of Western Australia (AUS); National Health and Medical Research Council of Australia (199600 and 400281 [AUS]); Mayo Foundation (MAY); Minnesota Ovarian Cancer Alliance (MAY); Fred C. and Katherine B. Andersen Foundation (MAY); Radboud University Medical Centre (NTH); Lon V Smith Foundation (LVS-39420 [UCI]); National Institute of Environmental Health Sciences (T32 ES013678 for A.W.L.); National Health and Medical Research Council of Australia (for G.C.T. and P.W.).

Research reported in this publication was partially supported by NCI award number P30 CA008748 (PI: Thompson) to Memorial Sloan Kettering Cancer Center. It was also supported by the National Cancer Institute of the National Institutes of Health (P30 CA046592). Lastly, this

work was also supported by the National Science Foundation (NSF DMS 1406712) and the National Institutes of Health (NIH ES 20811). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

We thank all the individuals who took part in this study and all the researchers, clinicians and technical and administrative staff who have made possible the many studies contributing to this work. In particular, we thank: Dr. D. Bowtell, Dr. A. DeFazio, Dr. D. Gertig, Dr. A. Green, Dr. P. Parsons, Dr. N. Hayward, and Dr. D. Whiteman (AUS); the staff of the genotyping unit, Dr. S LaBoissiere and F Robidoux (Genome Quebec); Dr. U. Eilber (GER); Dr. I. Jacobs, Dr. M. Widschwendter, Dr. E. Wozniak, N. Balogun, Dr. A. Ryan and J. Ford (UKO).

Conflict of interest: None declared.

REFERENCES

1. Hutter C, Chang-Claude J, Slattery M, et al. Characterization of Gene-Environment Interactions for Colorectal Cancer Susceptibility Loci. *Cancer Research*. 2012; 72(8): 2036-2044.
2. Hsu L, Jiao S, Dai J, et al. Powerful Cocktail Methods for Detecting Genome-Wide Gene-Environment Interaction. *Genetic Epidemiology*. 2012; 36(3): 183-194.
3. Garcia-Closas M, Rothman N, Figueroa J, et al. Common Genetic Polymorphisms Modify the Effect of Smoking on Absolute Risk of Bladder Cancer. *Cancer Research*. 2013; 73(7): 2211-2220.
4. Figueiredo J, Hsu L, Hutter C, et al. Genome-Wide Diet-Gene Interaction Analyses for Risk of Colorectal Cancer. *PLoS Genetics*. 2014; 10(4): p.e1004228.
5. Lewinger J, Morrison J, Thomas D, et al. Efficient Two-Step Testing of Gene-Gene Interactions in Genome-Wide Association Studies. *Genetic Epidemiology*. 2013; 37(5): 440-451.
6. Mukherjee B and Chatterjee N. Exploiting Gene-Environment Independence for Analysis of Case-Control Studies: An Empirical Bayes-type Shrinkage Estimator to Trade-Off between Bias and Efficiency. *Biometrics*. 2008; 64(3): 685-694.
7. Murcray C, Lewinger, J and Gauderman W. Gene-Environment Interaction in Genome-Wide Association Studies. *American Journal of Epidemiology*. 2008; 169(2): 219-226.
8. Mukherjee B, Ahn J, Chatterjee N, et al. Testing Gene-Environment Interaction in Large-Scale Case-Control Association Studies: Possible Choices and Comparisons. *American Journal of Epidemiology,* 2012; 175(3): 177-190.

9. Boonstra P, Mukherjee B, Chatterjee N, et al. Tests for Gene-Environment Interactions and Joint Effects with Exposure Misclassification. *American Journal of Epidemiology*. 2016; 183(3): 237-247.
10. Thomas D. Methods for Investigating Gene-Environment Interactions in Candidate Pathway and Genome-Wide Association Studies. *Annual Review of Public Health*. 2010; 31(1): 21-36.
11. Prentice R and Pyke R. Logistic Disease Incidence Models and Case-Control Studies. *Biometrika*.1979; 66(3): 403.
12. Piegorsch W, Weinberg C and Taylor J. Non-hierarchical logistic models and case-only designs for assessing susceptibility in population-based case-control studies. *Statist. Med.*1994; 13(2): 153-162.
13. Umbach D and Weinberg C. Designing and analyzing case-control studies to exploit independence of genotype and exposure. *Statist. Med.* 1997; 16(15): 1731-1743.
14. Du M, Zhang X, Hoffmeister M, et al. No Evidence of Gene-Calcium Interactions from Genome-Wide Analysis of Colorectal Cancer Risk. *Cancer Epidemiology Biomarkers & Prevention*. 2014; 23(12): 2971-2976.
15. Joshi A, Lindstrom S, Husing A, et al. Additive Interactions Between Susceptibility Single-Nucleotide Polymorphisms Identified in Genome-Wide Association Studies and Breast Cancer Risk Factors in the Breast and Prostate Cancer Cohort Consortium. *American Journal of Epidemiology*. 2014; 180(10): 1018-1027.
16. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Philadelphia: Wolters Kluwer Health; 2015: 71-87.

17. VanderWeele T. A Word and That to Which it Once Referred. *Epidemiology*. 2011; 22(4): 612-613.
18. Hosmer D and Lemeshow S. Confidence Interval Estimation of Interaction. *Epidemiology*. 1992; 3(5): 452-456.
19. Zou G. On the Estimation of Additive Interaction by Use of the Four-by-two Table and Beyond. *American Journal of Epidemiology*. 2008; 168(2): 212-224.
20. VanderWeele T. Sample Size and Power Calculations for Additive Interactions. *Epidemiologic Methods*. 2012; 1(1): 159-188.
21. Han S, Rosenberg P, Garcia-Closas M, et al. Likelihood Ratio Test for Detecting Gene (G)-Environment (E) Interactions Under an Additive Risk Model Exploiting G-E Independence for Case-Control Data. *American Journal of Epidemiology*. 2012; 176(11): 1060-1067.
22. Chatterjee N and Carroll R. Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika*. 2005; 92(2): 399-418.
23. VanderWeele TJ. Explanation in causal inference: methods for mediation and interaction. New York, NY: Oxford University Press; 2015: 255-260.
24. Zhang L, Mukherjee B, Ghosh M, et al. Accounting for error due to misclassification of exposures in case–control studies of gene–environment interaction. *Statist. Med*. 2008; 27(15): 2756-2783.
25. Chen YH, Chatterjee N and Carroll R. Shrinkage estimators for robust and efficient inference in haplotype-based case-control studies. *Journal of the American Statistical Association.* 2009; 104: 220-233.

26. Leeb H and Pötscher BM. Sparse estimators and the oracle property, or the return of Hodges' estimator. *Journal of Econometrics.* 2008;142(1):201–211.
27. Knol M, Vandenbroucke J, Scott P, et al. What Do Case-Control Studies Estimate? Survey of Methods and Assumptions in Published Case-Control Research. *American Journal of Epidemiology*. 2008; 168(9): 1073-1081.
28. Github website. https://github.com/GreysonL/RERI/releases. Updated January 12, 2017. Accessed May 12, 2017.
29. Beral V, Doll R, Hermon C, et al. Ovarian cancer and oral contraceptives: collaborative reanalysis of data from 45 epidemiological studies including 23,257 women with ovarian cancer and 87,303 controls. *Lancet*. 2008; 371(9609):303–314.
30. Pike MC, Pearce CL, Peters R, et al. Hormonal factors and the risk of invasive ovarian cancer: a population-based case-control study. *Fertil Steril*. 2004; 82(1):186–195.
31. Whiteman DC, Murphy MF, Cook LS, et al. Multiple births and risk of epithelial ovarian cancer. *J Natl Cancer Inst*. 2000; 92(14):1172–1177.
32. Tung KH, Goodman MT, Wu AH, et al. Reproductive factors and epithelial ovarian cancer risk by histologic type: a multiethnic case-control study. *American Journal of Epidemiology*. 2003; 158(7):629–638.
33. Cibula D, Widschwendter M, Majek O, et al. Tubal ligation and the risk of ovarian cancer: review and meta-analysis. *Hum Reprod Update*. 2011; 17(1):55–67.
34. Pearce CL, Templeman C, Rossing MA, et al. Association between endometriosis and risk of histological subtypes of ovarian cancer: a pooled analysis of case-control studies. *Lancet Oncol*. 2012; 13(4): 385-394.

35. Auranen A, Pukkala E, Makinen J, et al. Cancer incidence in the first- degree relatives of ovarian cancer patients. *Br J Cancer*. 1996; 74(2):280–284.
36. Pearce CL, Rossing M, Lee, et al. Combined and Interactive Effects of Environmental and GWAS-Identified Risk Factors in Ovarian Cancer. *Cancer Epidemiology Biomarkers & Prevention.* 2013; 22(5): 880-890.
37. Kuchenbaecker, et al. Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat Genet*. 2015; 47(2): 164-171.
38. Bolton KL, et al. Common variants at 19p13 are associated with susceptibility to ovarian cancer. *Nat Genet*. 2010; 42:880–884.
39. Goode EL, et al. A genome-wide association study identifies susceptibility loci for ovarian cancer at 2q31 and 8q24. *Nat Genet*. 2010; 42:874–879.
40. Song H, et al. A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet*. 2009; 41:996–1000.
41. Pharoah PD, et al. GWAS meta-analysis and replication identifies three new susceptibility loci for ovarian cancer. *Nat Genet.* 2013; 45:362–370. 370e361–370e362.
42. Permuth-Wey J, et al. Identification and molecular characterization of a new ovarian cancer susceptibility locus at 17q21.31. *Nat Commun*. 2013; 4:1627.
43. Bojesen SE, et al. Multiple independent variants at the TERT locus are associated with telomere length and risks of breast and ovarian cancer. *Nat Genet*. 2013; 45:371–384. 384e371–384e372.
44. Couch FJ, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet*. 2013; 9:e1003212.

45. Li S, Zhao J, Loos RJ, et al. Cumulative effects and predictive value of common obesity-susceptibility variants identified by genome-wide association studies. *Am J Clin Nutr*. 2010; 91(1): 184–190.

46. Mukherjee B, Ahn J, Chatterjee N, et al. Tests for gene-environment interaction from case-control data: a novel study of Type I error, power and designs. *Genetic Epidemiology.* 2008; 32(7): 615-626.

47. National Cancer Institute, Surveillance, Epidemiology, and End Results Program, SEER Stat Fact Sheets—Ovarian Cancer. http://seer.cancer.gov/statfacts/html/ovary.html. Published April 14, 2017. Accessed May 12, 2017.

48. Risch HA, Yu H, Lu L, et al. Detectable symptomatology preceding the diagnosis of pancreatic cancer and absolute risk of pancreatic cancer diagnosis. *American Journal of Epidemiology*. 2015; 182(1): 26-34.

49. Tchetgen E, Sofer T and Wong B. A General Approach to Detect Gene (G)-environment (E) Additive Interaction Leveraging G-E Independence in Case-control Studies. [online] Collection of Biostatistics Research Archive. Available at: http://biostats.bepress.com/harvardbiostat/paper177/ [Accessed 30 Jun. 2014].

50. Han S, Rosenberg P, Chatterjee N, et al. An exposure-weighted score test for genetic associations integrating environmental risk factors. *Biometrics*. 2015; 71(3): 596-605.

Table 1. Estimates and 95% confidence interval corresponding to SNP/GRS x Oral Contraceptive Pill Use Interaction under Both Multiplicative and Additive Scale with accompanying P-values from Wald Tests

Interaction	**Multiplicative (ψ)**			**Additive (RERI)**		
SNP1[a]×OCP[b]	Estimate[c]	95% CI	P-value	Estimate[d]	95% CI	P-value
UML	0.94	0.73, 1.22	0.645	-0.25	-0.60, 0.10	0.162
CML	1.06	0.88, 1.28	0.548	-0.09	-0.33, 0.14	0.432
EB	1.00	0.78, 1.29	0.970	-0.16	-0.50, 0.18	0.348
SNP2[a]×OCP						
UML	0.93	0.82, 1.05	0.255	0.08	-0.18, 0.34	0.552
CML	0.94	0.85, 1.04	0.224	0.03	-0.18, 0.25	0.757
EB	0.94	0.85, 1.04	0.222	0.04	-0.11, 0.18	0.598
GRS[d]×OCP						
UML	0.82	0.65, 1.02	0.073	-0.64	-1.01, -0.27	0.001
CML	0.92	0.77, 1.08	0.305	-0.43	-0.68, -0.18	0.001
EB	0.86	0.69, 1.07	0.197	-0.54	-0.93, -0.16	0.005
WGRS[d]×OCP						
UML	0.90	0.76, 1.06	0.212	-0.61	-0.99, -0.23	0.002
CML	0.95	0.83, 1.08	0.417	-0.40	-0.67, -0.14	0.003
EB	0.93	0.81, 1.08	0.366	-0.52	-0.91, -0.13	0.009

Abbreviations: CML, constrained maximum-likelihood; EB, empirical Bayes; GRS, genetic risk score; RERI, relative excess risk due to interaction; UML, unconstrained maximum-likelihood; WGRS, weighted genetic risk score.

[a] SNP1 denotes rs62274042 and SNP2 denotes rs10962691. Marginal disease odds ratios corresponding to these SNPs are 1.45 (1.37, 1.54) and 1.25 (1.20, 1.30) respectively.

[b] OCP=1 if the individual ever used OCP and OCP=0 if never.

[c] The analysis is based on subjects with European ancestry, using data on 4,135 cases and 7,526 controls from 13 study sites from the Ovarian Cancer Association Consortium. The model adjusts for history of endometriosis, first degree family history of ovarian cancer, parity, tubal ligation, age and study site.

[d] (W)GRS is a categorical variable defined by quartiles of WGRS in controls, e.g. (W)GRS=3 if it is above the 75th percentile in controls and (W)GRS=0 if it is below the 25th percentile in controls. The minimal, 25th, 50th, 75th percentiles and the maximum are 3, 11, 12, 14 ,22 for GRS and 0.32, 1.33, 1.53, 1.75 and 2.86 for WGRS. In this table, we only present the coefficient of the interaction term corresponding to a change of OCP from 0 to 1 and of WGRS from 0 to 3.

Table 2. Empirical Familywise Type I Error Rate at 5% overall level of significance, and Expected Number of False Positives corresponding to UML, CML and EB Wald Tests

	Proportion of markers satisfying gene-environment independence (p_{ind}) [a]					
	0.95	0.99	0.995	0.9975	0.9995	1.00
Empirical Familywise Type I error [b]						
UML	0.084	0.072	0.062	0.071	0.041	0.058
CML	1.000	0.994	0.966	0.745	0.874	0.064
EB	0.138	0.056	0.045	0.038	0.042	0.035
Expected number of false positives [c]						
UML	0.085	0.073	0.062	0.071	0.042	0.059
CML	23.451	3.761	2.814	1.050	0.937	0.067
EB	0.150	0.060	0.045	0.039	0.044	0.035

Abbreviations: CML, constrained maximum-likelihood; EB, empirical Bayes; RERI, relative excess risk due to interaction; UML, unconstrained maximum-likelihood.

[a] The population-level G-E association structure among null loci is assumed to be of the form of a mixture distribution reflecting that a large fraction, i.e., $\mathrm{p_{ind}}$, of the SNPs, indeed, are independent of E in the population, whereas the remaining $(1 - p_{ind})$ of SNPs show some departures from the independence assumption following a N (0, sd=log(1.5)/2) distribution.

[b] The Wald test is for RERI=0 under a large-scale genomewide G x E scan simulation scenario with 10000 markers and 2000 cases and controls. Empirical familywise type I error is estimated as the empirical proportion of data sets declaring at least 1 null marker to be significant using level of significance α/10000. This estimates the probability of at least one false positive under the global null.

[c] Expected number of false positives is estimated as the average number of falsely rejected null hypotheses, averaged over 1000 data sets.

Figure 1. Odds ratio of oral contraceptive pill and corresponding 95% CI within each quartile of the weighted genetic risk score. The odds ratios are estimated from a standard logistic regression adjusting for history of endometriosis, first degree family history of ovarian cancer, parity, tubal ligation, age and study site.

Figure 2. Predicted probability of ovarian cancer and corresponding 95% CI within each quartile of the weighted genetic risk score comparing oral contraceptive pill users and non-users. The relative risk parameters are obtained from a standard logistic regression model adjusting for history of endometriosis, first degree family history of ovarian cancer, parity, tubal ligation, age and study site. We assume that approximately 1.3 percent of women will be diagnosed with ovarian cancer at some point during their lifetime and 70% women will use oral contraceptive pill at some point in their life. The predicted probabilities are estimated by fixing other covariates at their most frequent value.

Figure 3. Power curves of unconstrained maximum-likelihood (UML), constrained maximum-likelihood (CML) and empirical Bayes (EB) Wald test for relative excess risk due to interaction (RERI) under different strength of G-E association: data are generated on 4000 cases and 4000 controls with fixed parameters $p_G = 0.2$, $p_E = 0.3$, $OR_G = 1.2$, $OR_E = 1.5$. RERI changes from 0 to 1.5 with a grid level of 0.1, corresponding multiplicative interaction changes from 0.94 to 1.78. The top panels (A, B, C) correspond to the raw power, whereas the bottom panels (D, E, F) correspond to the power at empirical α. The left, center, and right panels correspond to different values of the G-E association odds ratio, i.e. exp(θ_{GE})=0.8, 1.0, 1.2.

Web Appendix 1. Data for an Unmatched Case-control Study with a Dichotomous Genetic Factor and a Dichotomous Environmental Exposure

	G=0		*G*=1		
	E=0	*E*=1	*E*=0	*E*=1	
D=0	r_{01}	r_{02}	r_{03}	r_{04}	N_0
D=1	r_{11}	r_{12}	r_{13}	r_{14}	N_1

Recall that $\boldsymbol{r_0} \sim$ Multinomial $(N_0, \boldsymbol{p_0})$ and $\boldsymbol{r_1} \sim$ Multinomial $(N_1, \boldsymbol{p_1})$, hence the estimate of variance matrix of $\boldsymbol{r_0}$ is

$$\hat{V}_0 = \begin{bmatrix} r_{01}(N_0 - r_{01}) & -r_{01}r_{02} & -r_{01}r_{03} & -r_{01}r_{04} \\ -r_{01}r_{02} & r_{02}(N_0 - r_{02}) & -r_{02}r_{03} & -r_{02}r_{04} \\ -r_{01}r_{03} & -r_{02}r_{03} & r_{03}(N_0 - r_{03}) & -r_{03}r_{04} \\ -r_{01}r_{04} & -r_{02}r_{04} & -r_{03}r_{04} & r_{04}(N_0 - r_{04}) \end{bmatrix} / N_0$$

Similarly, the estimate of variance matrix of $\boldsymbol{r_1}$ is

$$\hat{V}_1 = \begin{bmatrix} r_{11}(N_1 - r_{11}) & -r_{11}r_{12} & -r_{11}r_{13} & -r_{11}r_{14} \\ -r_{11}r_{12} & r_{12}(N_1 - r_{12}) & -r_{12}r_{13} & -r_{12}r_{14} \\ -r_{11}r_{13} & -r_{12}r_{13} & r_{13}(N_1 - r_{13}) & -r_{13}r_{14} \\ -r_{11}r_{14} & -r_{12}r_{14} & -r_{13}r_{14} & r_{14}(N_1 - r_{14}) \end{bmatrix} / N_1$$

And $\boldsymbol{r_0}$ and $\boldsymbol{r_1}$ are independent.

Web Appendix 2. Variance of the UML estimator

$$\widehat{\mathrm{RERI}}_{\mathrm{uml}} = \frac{r_{01}r_{14}}{r_{11}r_{04}} - \frac{r_{01}r_{13}}{r_{11}r_{03}} - \frac{r_{01}r_{12}}{r_{11}r_{02}} + 1 = f_{UML}(\boldsymbol{r_0}, \boldsymbol{r_1})$$

$$\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{uml}}) = \left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_0}}\right)^T \hat{V}_0 \left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_0}}\right) + \left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_1}}\right)^T \hat{V}_1 \left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_1}}\right)$$

Note that,

$$\frac{\partial f_{UML}}{\partial \boldsymbol{r_0}} = \frac{r_{01}}{r_{11}}\left[\frac{A}{r_{01}}, \frac{r_{12}}{r_{02}^2}, \frac{r_{13}}{r_{03}^2}, -\frac{r_{14}}{r_{04}^2}\right]^T \text{ and } \frac{\partial f_{UML}}{\partial \boldsymbol{r_1}} = -\frac{r_{01}}{r_{11}}\left[\frac{A}{r_{11}}, \frac{1}{r_{02}}, \frac{1}{r_{03}}, -\frac{1}{r_{04}}\right]^T,$$

$$\text{where } A = \frac{r_{14}}{r_{04}} - \frac{r_{13}}{r_{03}} - \frac{r_{12}}{r_{02}}.$$

Using the fact that $r_{01}+r_{02}+r_{03}+r_{04}=N_0$ and $r_{11}+r_{12}+r_{13}+r_{14}=N_1$, we can simplify the variance expressions further and obtain

$$\left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_0}}\right)^T \hat{V}_0 \left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_0}}\right) = \left(\frac{r_{01}}{r_{11}}\right)^2 \left(\frac{A^2}{r_{01}}+\frac{r_{12}^2}{r_{02}^3}+\frac{r_{13}^2}{r_{03}^3}+\frac{r_{14}^2}{r_{04}^3}\right)$$

$$\left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_1}}\right)^T \hat{V}_1 \left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_1}}\right) = \left(\frac{r_{01}}{r_{11}}\right)^2 \left(\frac{A^2}{r_{11}}+\frac{r_{12}}{r_{02}^2}+\frac{r_{13}}{r_{03}^2}+\frac{r_{14}}{r_{04}^2}\right).$$

Hence,

$$\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{uml}}) = \left(\frac{r_{01}}{r_{11}}\right)^2 \left[\left(\frac{1}{r_{01}}+\frac{1}{r_{11}}\right)A^2+\frac{r_{12}^2}{r_{02}^3}+\frac{r_{13}^2}{r_{03}^3}+\frac{r_{14}^2}{r_{04}^3}+\frac{r_{12}}{r_{02}^2}+\frac{r_{13}}{r_{03}^2}+\frac{r_{14}}{r_{04}^2}\right]$$

Web Appendix 3. Variance of the CML estimator

$$\widehat{\mathrm{RERI}}_{\mathrm{cml}} = \frac{(r_{01}+r_{03})(r_{01}+r_{02})r_{14}}{(r_{02}+r_{04})(r_{03}+r_{04})r_{11}} - \frac{(r_{01}+r_{02})r_{13}}{(r_{03}+r_{04})r_{11}} - \frac{(r_{01}+r_{03})r_{12}}{(r_{02}+r_{04})r_{11}} + 1$$

Let $r_G = r_{03}+r_{04}, r_E = r_{02}+r_{04}$, then one can rewrite $\widehat{\mathrm{RERI}}_{\mathrm{cml}}$ as,

$$\widehat{\mathrm{RERI}}_{\mathrm{cml}} = \frac{r_{14}(N_0-r_E)(N_0-r_G)}{r_{11}r_E r_G} - \frac{r_{13}(N_0-r_G)}{r_{11}r_G} - \frac{r_{12}(N_0-r_E)}{r_{11}r_E} + 1 = f_{CML}(\boldsymbol{r_0^*}, \boldsymbol{r_1})$$

$where\ \boldsymbol{r_0^*} = (r_G, r_E)^T$

r_G, r_E follow a binomial distribution $Bin(N_0, p_{03}+p_{04})$ and $Bin(N_0, p_{02}+p_{04})$. So $\widehat{Var}(r_G) = r_G(N_0-r_G)/N_0, \widehat{Var}(r_E) = r_E(N_0-r_E)/N_0$ $and\ \widehat{Cov}(r_G, r_E) = \widehat{Cov}(r_{03}+r_{04}, r_{02}+r_{04}) = \widehat{Cov}(r_{02}, r_{03}) + \widehat{Cov}(r_{03}, r_{04}) + \widehat{Cov}(r_{02}, r_{04}) + \widehat{Var}(r_{04}) = (r_{01}r_{04}-r_{02}r_{03})/N_0 = B/N_0$.

$$\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{cml}}) = \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_0^*}}\right)^T \hat{V}_0^* \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_0^*}}\right) + \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_1}}\right)^T \hat{V}_1 \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_1}}\right),$$

$$where\ \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_0^*}}\right)^T \hat{V}_0^* \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_0^*}}\right) = \frac{N_0}{r_{11}^2}[2BCD + r_G(N_0-r_G)C^2 + r_E(N_0-r_E)D^2]$$

$$C = r_{13} - r_{14}\frac{N_0-r_E}{r_E}, D = r_{12} - r_{14}\frac{N_0-r_G}{r_G}\ and\ \frac{\partial f_{CML}}{\partial \boldsymbol{r_1}} = [\frac{F}{r_{11}}, -\frac{N_0-r_E}{r_{11}r_E}, -\frac{N_0-r_G}{r_{11}r_G},$$

$$\frac{(N_0 - r_E)(N_0 - r_G)}{r_{11} r_E r_G}]^T, F = r_{12}\frac{N_0 - r_E}{r_E} + r_{13}\frac{N_0 - r_G}{r_G} - r_{14}\frac{(N_0 - r_G)(N_0 - r_E)}{r_G r_E}, \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_1}}\right)^T \hat{V}_1$$

$$\left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_1}}\right) = \frac{1}{r_{11}^2}\left[\frac{F^2}{r_{11}} + r_{12}\left(\frac{N_0 - r_E}{r_E}\right)^2 + r_{13}\left(\frac{N_0 - r_G}{r_G}\right)^2 + r_{14}\left(\frac{(N_0 - r_G)(N_0 - r_E)}{r_G r_E}\right)^2\right].$$

Hence,

$$\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{cml}}) = \frac{N_0}{r_{11}^2}[2BCD + r_G(N_0 - r_G)C^2 + r_E(N_0 - r_E)D^2] + \frac{1}{r_{11}^2}\left[\frac{F^2}{r_{11}} + r_{12}\left(\frac{N_0 - r_E}{r_E}\right)^2\right.$$

$$\left. + r_{13}\left(\frac{N_0 - r_G}{r_G}\right)^2 + r_{14}\left(\frac{(N_0 - r_G)(N_0 - r_E)}{r_G r_E}\right)^2\right]$$

Web Appendix 4. Variance of the EB estimator

$$\widehat{RERI}_{EB} = \frac{\left(\widehat{RERI}_{uml} - \widehat{RERI}_{cml}\right)^2}{\widehat{Var}_{uml} + \left(\widehat{RERI}_{uml} - \widehat{RERI}_{cml}\right)^2}\widehat{RERI}_{uml} + \frac{\widehat{Var}_{uml}}{\widehat{Var}_{uml} + \left(\widehat{RERI}_{uml} - \widehat{RERI}_{cml}\right)^2}\widehat{RERI}_{cml}$$

Here we consider $\widehat{Var}_{uml}$ as a constant, then $\widehat{RERI}_{EB}$ is a function of $\widehat{RERI}_{uml}$ and $\widehat{RERI}_{cml}$.

Again, we apply delta method to derive the corresponding variance.

$$\widehat{Var}\left(\widehat{\mathrm{RERI}}_{\mathrm{EB}}\right) = Var(f(\widehat{\mathrm{RERI}}_{\mathrm{uml}}, \widehat{\mathrm{RERI}}_{\mathrm{cml}}))$$

$$= \begin{bmatrix}\frac{\partial \widehat{\mathrm{RERI}}_{\mathrm{EB}}}{\partial \widehat{\mathrm{RERI}}_{\mathrm{uml}}} & \frac{\partial \widehat{\mathrm{RERI}}_{\mathrm{EB}}}{\partial \widehat{\mathrm{RERI}}_{\mathrm{cml}}}\end{bmatrix}\begin{bmatrix}\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{uml}}) & \widehat{\mathrm{Cov}}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}}, \widehat{\mathrm{RERI}}_{\mathrm{cml}}\right) \\ \widehat{\mathrm{Cov}}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}}, \widehat{\mathrm{RERI}}_{\mathrm{cml}}\right) & \widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{cml}})\end{bmatrix}\begin{bmatrix}\frac{\partial \widehat{\mathrm{RERI}}_{\mathrm{EB}}}{\partial \widehat{\mathrm{RERI}}_{\mathrm{uml}}} \\ \frac{\partial \widehat{\mathrm{RERI}}_{\mathrm{EB}}}{\partial \widehat{\mathrm{RERI}}_{\mathrm{cml}}}\end{bmatrix}$$

$$= a^2\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{uml}}) + b^2\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{cml}}) + 2ab\widehat{\mathrm{Cov}}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}}, \widehat{\mathrm{RERI}}_{\mathrm{cml}}\right)$$

$$where\ a = \frac{\partial \widehat{\mathrm{RERI}}_{\mathrm{EB}}}{\partial \widehat{\mathrm{RERI}}_{\mathrm{uml}}} = \frac{3\widehat{\mathrm{RERI}}_{\mathrm{uml}}^2 - 4\widehat{\mathrm{RERI}}_{\mathrm{uml}}\widehat{\mathrm{RERI}}_{\mathrm{cml}} + \widehat{\mathrm{RERI}}_{\mathrm{cml}}^2}{\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{uml}}) + \left(\widehat{\mathrm{RERI}}_{\mathrm{uml}} - \widehat{\mathrm{RERI}}_{\mathrm{uml}}\right)^2} -$$

$$\frac{2\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{uml}})\widehat{\mathrm{RERI}}_{\mathrm{cml}}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}} - \widehat{\mathrm{RERI}}_{\mathrm{cml}}\right) + 2\widehat{\mathrm{RERI}}_{\mathrm{uml}}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}} - \widehat{\mathrm{RERI}}_{\mathrm{cml}}\right)^3}{\left(\widehat{Var}(\widehat{\mathrm{RERI}}_{\mathrm{uml}}) + \left(\widehat{\mathrm{RERI}}_{\mathrm{uml}} - \widehat{\mathrm{RERI}}_{\mathrm{uml}}\right)^2\right)^2}$$

$$b = \frac{\partial \widehat{\mathrm{RERI}}_{\mathrm{EB}}}{\partial \widehat{\mathrm{RERI}}_{\mathrm{cml}}} = \frac{\widehat{Var}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}}\right)^2 - \widehat{Var}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}}\right)\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}} - \widehat{\mathrm{RERI}}_{\mathrm{uml}}\right)^2}{\left(\widehat{Var}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}}\right) + \left(\widehat{\mathrm{RERI}}_{\mathrm{uml}} - \widehat{\mathrm{RERI}}_{\mathrm{uml}}\right)^2\right)^2}$$

And by using similar argument,

$$\widehat{\mathrm{Cov}}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}}, \widehat{\mathrm{RERI}}_{\mathrm{cml}}\right) = \left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_0}}\right)^T \hat{V}_0 \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_0}}\right) + \left(\frac{\partial f_{UML}}{\partial \boldsymbol{r_1}}\right)^T \hat{V}_1 \left(\frac{\partial f_{CML}}{\partial \boldsymbol{r_1}}\right)$$

$\frac{\partial f_{UML}}{\partial \boldsymbol{r_0}}, \frac{\partial f_{UML}}{\partial \boldsymbol{r_1}}$ and $\frac{\partial f_{CML}}{\partial \boldsymbol{r_1}}$ are already calculated in previous derivation, substituting

$$\frac{\partial f_{CML}}{\partial \boldsymbol{r_0}} = \frac{[H, I, J, K]^T}{r_{11}}, where\ H = \frac{r_{14}(2N_0 - r_G - r_E)}{r_G r_E} - \frac{r_{13}}{r_G} - \frac{r_{12}}{r_E}, I = r_{12}\frac{N_0 - r_E}{r_E^2} - \frac{r_{13}}{r_G} +$$

$$r_{14}\frac{(N_0 - r_E)(r_{04} - r_{01})}{r_E^2 r_G}, J = r_{13}\frac{N_0 - r_G}{r_G^2} - \frac{r_{12}}{r_E} + r_{14}\frac{(N_0 - r_G)(r_{04} - r_{01})}{r_G^2 r_E} \text{ and}$$

$$K = r_{12}\frac{N_0 - r_E}{r_E^2} + r_{13}\frac{N_0 - r_G}{r_G^2} - r_{14}\frac{(N_0 - r_G)(N_0 - r_E)}{r_G^2 r_E} - r_{14}\frac{(N_0 - r_G)(N_0 - r_E)}{r_E^2 r_G}$$

Finally, we obtain the covariance

$$\widehat{\mathrm{Cov}}\left(\widehat{\mathrm{RERI}}_{\mathrm{uml}}, \widehat{\mathrm{RERI}}_{\mathrm{cml}}\right) = \frac{\mathrm{r}_{01}}{r_{11}^2}[AH + \frac{r_{12}}{r_{02}}I + \frac{r_{13}}{r_{03}}J - \frac{r_{14}}{r_{04}}K - \frac{AF}{r_{11}} + \frac{r_{12}}{r_{02}}\frac{N_0 - r_E}{r_E} +$$

$\frac{r_{13}}{r_{03}}\frac{N_0 - r_G}{r_G} + \frac{r_{14}}{r_{04}}\frac{(N_0 - r_G)(N_0 - r_E)}{r_G r_E}]$

Web Appendix 5. Estimated and Empirical Variance of Three Estimators

Although, theoretically Delta method doesn't apply under G-E independence, practically, the estimated variance from delta method approximate the simulation results very well. Hence we consider both CML and EB estimators are ready to use. We illustrate this point by a similar simulation to the power analysis one in the paper. We set p(G)=0.2, p(E)=0.3, OR(G)=1.2, OR(E)=1.5, number of cases and controls are both 2000. We change the true values of RERI from

-0.3 to 0.3 and exp(θ_{GE}) from 0.9 to 1.1 with a grid of 0.05. Then we calculate the empirical variance of estimated RERI and the average of estimated variance by the proposed method across 5,000 datasets. The Web Table 1 shows the two variances of three estimators. The empirical variance of EB deviates a little from the variance obtained by formula because we consider $\widehat{Var}\left(\widehat{RERI}_{uml}\right)$ as a constant in the derivation.

Given the fact that the estimated variance is close to the empirical variance, we then numerically evaluated the distribution of the test statistics corresponding to CML and EB to a reference normal distribution under large sample size. In this situation, we set the numbers of cases and controls to be 5000, and additionally fix RERI=0 and $\exp(\theta_{GE}) = \exp\left(\frac{1}{\sqrt{N_0+N_1}}\right) = \exp(0.01)$, which is the contiguous departure from the independence assumption that you have indicated in your comment. We evaluate the empirical distribution of CML and EB test statistics over 5,000 simulation replications (See Web Figure 1). Shapiro-Wilk tests do not reject the null hypothesis that the distributions of CML and EB are normal with p=0.607 and p=0.307, respectively. The QQ plots of the empirical quantile and theoretical standard normal quantile also indicate strong agreement (See Web Figure 2). Additionally, we calculate the empirical MSE under different θ_{GE} of order $N^{-\frac{1}{k}}$ with k=2,3,4,5 using the same setting (See Web Table 2).

Web Table 1. Comparison of Proposed Variance Formula and Empirical Variance

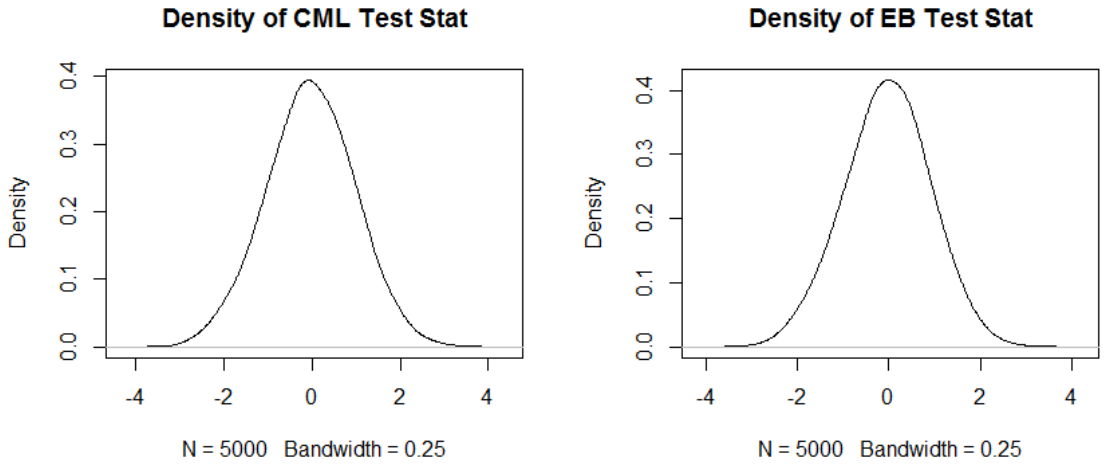
RERI	exp(θ_{GE})	UML1*	UML2*	CML1*	CML2*	EB1*	EB2*
-0.3	0.90	0.0555	0.0550	0.0249	0.0238	0.0445	0.0416
	0.95	0.0546	0.0542	0.0248	0.0240	0.0398	0.0361
	1.00	0.0537	0.0541	0.0248	0.0244	0.0377	0.0348
	1.05	0.0533	0.0538	0.0249	0.0251	0.0383	0.0360
	1.10	0.0526	0.0544	0.0250	0.0254	0.0404	0.0393
0.3	0.90	0.0798	0.0813	0.0316	0.0321	0.0639	0.0603
	0.95	0.0780	0.0774	0.0323	0.0320	0.0561	0.0518
	1.00	0.0762	0.0733	0.0329	0.0324	0.0524	0.0480
	1.05	0.0749	0.0764	0.0336	0.0342	0.0528	0.0499
	1.10	0.0739	0.0759	0.0345	0.0352	0.0563	0.0545

*1 denotes the average of 5000 estimated variances by proposed formula and 2 denotes the empirical variance of 5000 point estimates.

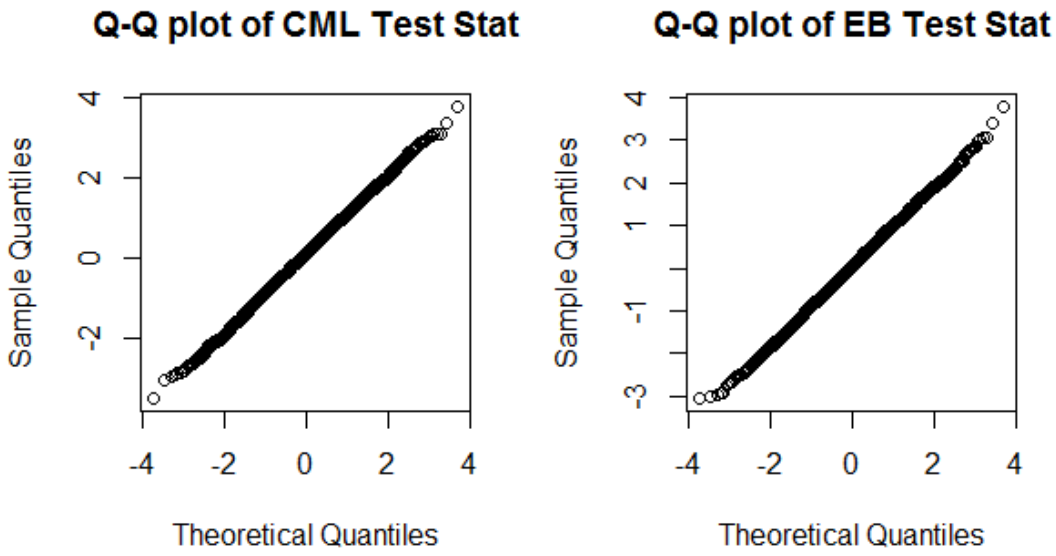
Web Table 2. MSE of UML, CML and EB estimators over 5000 replications with different choices of θ_{GE}

$\theta_{GE} \propto N^{-\frac{1}{k}}$	UML	CML	EB
k=2	0.0255	0.0113	0.0159
k=3	0.0245	0.0162	0.0176
k=4	0.0244	0.0348	0.0236
k=5	0.0237	0.0694	0.0280

Web Figure 1: Density of CML and EB Test Statistics



Web Figure 2: QQ Plots of CML and EB Test Statistics



Web Appendix 6. UML, CML and EB Estimators in General Cases

$$
\begin{aligned}
&RERI(E_0, E_1, G_0, G_1) \\
&\approx \exp\{\beta_G(G_1 - G_0) + \beta_E(E_1 - E_0) + \beta_{GE}(G_1{\times}E_1 - G_0{\times}E_0)\} \\
&\qquad - \exp\{\beta_E(E_1 - E_0) + \beta_{GE}G_0{\times}(E_1 - E_0)\} \\
&\qquad - \exp\{\beta_G(G_1 - G_0) + \beta_{GE}(G_1 - G_0){\times}E_0\} + 1 \\
&= f(\beta_G, \beta_E, \beta_{GE})
\end{aligned}
$$

So $RERI(E_0, E_1, G_0, G_1)$ is a function of $\beta_G, \beta_E, \beta_{GE}$ given E_0, E_1, G_0, G_1. The derivatives of RERI with respect to $\beta_G, \beta_E, \beta_{GE}$ are

$$
\begin{aligned}
f'_G = \frac{\partial RERI}{\partial \beta_G} &= [\exp\{\beta_G(G_1 - G_0) + \beta_E(E_1 - E_0) + \beta_{GE}(G_1{\times}E_1 - G_0{\times}E_0)\} \\
&\quad - \exp\{\beta_G(G_1 - G_0) + \beta_{GE}(G_1 - G_0){\times}E_0\}](G_1 - G_0) \\
f'_E = \frac{\partial RERI}{\partial \beta_E} &= [\exp\{\beta_G(G_1 - G_0) + \beta_E(E_1 - E_0) + \beta_{GE}(G_1{\times}E_1 - G_0{\times}E_0)\} \\
&\quad - \exp\{\beta_E(E_1 - E_0) + \beta_{GE}G_0{\times}(E_1 - E_0)\}](E_1 - E_0) \\
f'_{GE} = \frac{\partial RERI}{\partial \beta_{GE}} &= \exp\{\beta_G(G_1 - G_0) + \beta_E(E_1 - E_0) + \beta_{GE}(G_1{\times}E_1 - G_0{\times}E_0)\}(G_1{\times}E_1 - G_0{\times}E_0) \\
&\quad - \exp\{\beta_E(E_1 - E_0) + \beta_{GE}G_0{\times}(E_1 - E_0)\}G_0{\times}(E_1 - E_0) \\
&\quad - \exp\{\beta_G(G_1 - G_0) + \beta_{GE}(G_1 - G_0){\times}E_0\}(G_1 - G_0){\times}E_0
\end{aligned}
$$

More generally, if G or E is a factor with different coefficients for different levels, then RERI has the form below:

$$
\begin{aligned}
&RERI(E_0, E_1, G_0, G_1) \\
&= \frac{\exp(\beta_{G_1} + \beta_{E_1} + \beta_{G_1E_1}) - \exp(\beta_{G_1} + \beta_{E_0} + \beta_{G_1E_0}) - \exp(\beta_{G_0} + \beta_{E_1} + \beta_{G_0E_1}) + \exp(\beta_{G_0} + \beta_{E_0} + \beta_{G_0E_0})}{\exp(\beta_{G_0} + \beta_{E_0} + \beta_{G_0E_0})} \\
&= \exp\{(\beta_{G_1} - \beta_{G_0}) + (\beta_{E_1} - \beta_{E_0}) + (\beta_{G_1E_1} - \beta_{G_0E_0})\} - \exp\{(\beta_{G_1} - \beta_{G_0}) + (\beta_{G_1E_0} - \beta_{G_0E_0})\} \\
&\qquad + \exp\{(\beta_{E_1} - \beta_{E_0}) + (\beta_{G_0E_1} - \beta_{G_0E_0})\} + 1 \\
&= f(\beta_G, \beta_E, \beta_{GE})
\end{aligned}
$$

If any of βs with subscript 0 corresponds to the reference group in the model, then we set it to zero. The derivatives of RERI in this situation is similar to those derived above but with β, G, E, GE as vectors instead of scalars. For example, if G is coded as 0-1-2 with 0 as reference group, then $(\beta_{G_2} - \beta_{G_1})$ can be expressed as $\beta_G(G_2 - G_1) = (\beta_{G_1}, \beta_{G_2})[(0,1) - (1,0)]^T$ and $(\beta_{G_2} - \beta_{G_0})$ can be expressed as $(\beta_{G_1}, \beta_{G_2})[(0,1) - (0,0)]^T$.

According to Mukherjee and Chatterjee (6), the joint distribution of the MLE $(\widehat{\boldsymbol{\beta}}_{\boldsymbol{uml}}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{cml}})$ is asymptotically multivariate normal. The variance-covariance matrix can be estimated from "CGEN" package and it has the form below.

$$
Var\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{uml}}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{cml}}\right) = \begin{bmatrix} \boldsymbol{V_1} & \boldsymbol{\Sigma} \\ \boldsymbol{\Sigma} & \boldsymbol{V_2} \end{bmatrix}
$$

where $\widehat{\boldsymbol{\beta}}_{\boldsymbol{uml}}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{cml}}$ are both k × 1 vectors and $\boldsymbol{V_1}, \boldsymbol{V_2}, \boldsymbol{\Sigma}$ are k × k matrices, where k is the number of coefficients of $(\beta_G, \beta_E, \beta_{GE})$ in the logistic regression disease risk model. By applying Delta method, we obtain

(a) The variance of UML estimator of RERI is $B_1^T V_1 B_1$, where B_1 is the derivative vector of $RERI_{uml}$.
(b) The variance of CML estimator of RERI is $B_0^T V_0 B_0$, where B_0 is the derivative vector of $RERI_{cml}$.
(c) The EB estimator of RERI is

$$\frac{(\widehat{RERI}_{uml} - \widehat{RERI}_{cml})^2}{(\widehat{RERI}_{uml} - \widehat{RERI}_{cml})^2 + B_1^T V_1 B_1} \widehat{RERI}_{uml} + \frac{B_1^T V_1 B_1}{(\widehat{RERI}_{uml} - \widehat{RERI}_{cml})^2 + B_1^T V_1 B_1} \widehat{RERI}_{cml}$$

Note that $\partial\widehat{RERI}_{EB}$ is a function of $\widehat{RERI}_{uml}$ and $\widehat{RERI}_{cml}$. Let R_1 denote $\widehat{RERI}_{uml}$ and R_0 denote $\widehat{RERI}_{cml}$.

$$\frac{\partial\widehat{RERI}_{EB}}{\partial R_1} = \frac{3R_1^2 - 4R_1R_0 + R_0^2}{B_1^T V_1 B_1 + (R_1 - R_0)^2} - 2\frac{B_1^T V_1 B_1 R_0(R_1 - R_0) + R_1(R_1 - R_0)^3}{[B_1^T V_1 B_1 + (R_1 - R_0)^2]^2},$$

$$\frac{\partial\widehat{RERI}_{EB}}{\partial R_0} = \frac{(B_1^T V_1 B_1)^2 - B_1^T V_1 B_1 (R_1 - R_0)^2}{[B_1^T V_1 B_1 + (R_1 - R_0)^2]^2}$$

$$\mathrm{Cov}\left(\widehat{RERI}_{uml}, \widehat{RERI}_{cml}\right) = Cov\left(f\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{uml}}\right), f\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{cml}}\right)\right)$$

$$= \left[\frac{\partial f\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{uml}}\right)}{\partial \widehat{\boldsymbol{\beta}}_{\boldsymbol{uml}}}\right]^T Cov\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{uml}}, \widehat{\boldsymbol{\beta}}_{\boldsymbol{cml}}\right) \frac{\partial f\left(\widehat{\boldsymbol{\beta}}_{\boldsymbol{cml}}\right)}{\partial \widehat{\boldsymbol{\beta}}_{\boldsymbol{cml}}} = B_1^T \, \Sigma B_0$$

Hence, the variance of $\widehat{RERI}_{EB}$ is $A^T \begin{bmatrix} B_1^T V_1 B_1 & B_1^T \, \Sigma B_0 \\ B_1^T \, \Sigma B_0 & B_0^T V_0 B_0 \end{bmatrix} A$, where $A^T = [\frac{\partial\widehat{RERI}_{EB}}{\partial R_1}, \frac{\partial\widehat{RERI}_{EB}}{\partial R_0}]$.The Wald test is still based on standardized Z statistic.

Web Table 3. Odds Ratios for Marginal Associations of 18 Susceptibility Variants with Ovarian Cancer Risk from Previous Studies

SNP	Chromosome	Position	Reference Allele(s)	Tested Allele	Tested Allele Frequency	Published OR[a] 95% CI
rs58722170	1p34.3	38096421	G	C	0.22	1.07 (1.03, 1.11)
rs10069690	5p15.33	1279790	C	T	0.26	1.09 (1.05, 1.13)
chr10:21878831:D	10p12.31	21878831	allele1	T	0.30	1.09 (1.06, 1.13)
rs17329882	4q26	119949960	A	C	0.24	1.09 (1.06, 1.13)
rs1879586	17q21.31	43567337	C	G	0.17	1.13 (1.09, 1.18)
rs56318008	1p36	22470407	C	T	0.15	1.11 (1.06, 1.15)
rs4808075	19p13.11	17390291	T	C	0.30	1.12 (1.09, 1.16)
chr9:136138765:D	9q34.2	136138765	allele2	G	0.15	1.15 (1.10, 1.20)
rs7207826	17q21.32	46500673	T	C	0.27	1.13 (1.09,1.16)
rs76837345	8q21.13	82668818	A	G	0.07	1.20 (1.13, 1.27)
rs62274042	3q25.31	156435952	G	A	0.05	1.45 (1.37,1.54)
rs635634	9q34.2	136155000	C	T	0.20	1.12 (1.08, 1.16)
rs3744763	17q12	36090885	A	G	0.59	1.06 (1.03, 1.09)
chr17:29181220:I	17q11.2	29181220	AT	A	0.72	1.11 (1.07, 1.15)
rs6755777	2q31.1	177043326	G	T	0.32	1.13 (1.09,1.16)
rs1400482	8q24.21	129541931	A	G	0.87	1.18 (1.13,1.23)
rs116133110	6q22.1	28480635	C	T	0.69	1.07 (1.03, 1.10)
rs10962691 (rs117224476)[b]	9q22.2	16915105	G	C	0.79	1.25 (1.20, 1.30)

Abbreviations: allele1, TCCCTTC; allele2, GCGCCCACCACTA.

[a] ORs are obtained from analysis in Collaborative Oncological Gene-Environment Study (COGS) (26-33).

[b] rs117224476 was the strongest hit in previous studies, however, it was not genotyped in current study. rs10962691 is the closest proxy of this SNP.

Web Table 4. Odds Ratios for Marginal Associations of Categorical GRS and WGRS in Subjects with European Ancestry with 4,135 Cases and 7,526 Controls from 13 Sites

Risk Score Categories	OR	95% CI	P value
GRS [1][a]	Ref		
GRS [2]	1.14[b]	(1.01,1.29)	3.56×10^{-2}
GRS [3]	1.32	(1.19,1.46)	8.48×10^{-8}
GRS [4]	1.90	(1.71,2.10)	2.95×10^{-33}
WGRS [1][a]	Ref		
WGRS [2]	1.20[c]	(1.07,1.36)	2.13×10^{-3}
WGRS [3]	1.38	(1.23,1.55)	4.76×10^{-8}
WGRS [4]	1.97	(1.76,2.20)	9.04×10^{-33}

Abbreviations: GRS, genetic risk score; WGRS, weighted genetic risk score; Ref, reference group.
[a] GRS [k] and WGRS [k] denote the k^{th} category of the corresponding risk score in controls. The 0%,25%, 50%, 75% and 100% percentiles are 3, 11, 12, 14, 22 for GRS; 0.32, 1.33, 1.53, 1.75, 2.86 for WGRS.
[b] ORs are estimated from logistic regression models including main effects for GRS/WGRS, adjusted for study sites and 26 principal components.

Web Table 5: Odds Ratios for Marginal Associations of Seven Environmental Exposures with Ovarian Cancer Risk with 7,783 Cases and 13,750 Controls from 15 Sites

Exposure	Control	Case	Unadjusted OR[a]	Adjusted OR[b]	95 % CI		P value
OC never	4227	3197	Ref	Ref			
OC ever	9523	4586	0.64	0.60	0.56	0.64	5.02×10^{-51}
Fullbirths>=1	11623	5783	Ref	Ref			
Fullbirths=0	2127	2000	1.89	1.73	1.60	1.87	2.10×10^{-45}
Age							
(69,74)	1283	707	Ref	Ref			
(50,54)	1631	1020	1.13	1.39	1.22	1.59	8.76×10^{-7}
(54,59)	2157	1330	1.12	1.42	1.26	1.61	2.26×10^{-8}
(59,64)	2038	1207	1.07	1.34	1.18	1.52	4.01×10^{-6}
(64,69)	1669	950	1.03	1.16	1.02	1.32	2.08×10^{-2}
<=50	4335	2187	0.92	1.06	0.94	1.19	3.61×10^{-1}
>74	637	382	1.09	1.06	0.90	1.26	4.75×10^{-1}
Race							
Black	366	196	Ref	Ref			
HW	353	260	1.38	1.36	1.05	1.74	1.74×10^{-2}
NHW	12440	6876	1.03	1.06	0.88	1.28	5.55×10^{-1}
Asian	314	287	1.71	1.41	1.09	1.82	8.81×10^{-3}
other	277	164	1.11	1.04	0.78	1.37	8.07×10^{-1}
tublig=Yes	3078	1300	Ref	Ref			
tublig=No	10672	6483	1.44	1.39	1.29	1.51	1.62×10^{-16}
endom=No	12915	6966	Ref	Ref			
endom=Yes	835	817	1.81	1.64	1.48	1.83	6.30×10^{-20}
famhist=No	13355	7318	Ref	Ref			
famhist=Yes	395	465	2.15	2.19	1.90	2.52	6.76×10^{-27}

Abbreviations: endom, history of endometriosis; famhist, family history; Fullbirths, number of full-term births; HW, Hispanic White; NHW, Non-Hispanic White; tublig, tubal ligation; Ref, reference group.

[a] ORs are obtained from univariate logistic regression models, as well as contingency tables.

[b] ORs are estimated from logistic regression models including main effects for environmental exposures, and adjusted for study sites.

Web Table 6: Description of Study Sites Included in the Analysis

Study Name	Study Abbreviation	Country	Number of Cases (E)	Number of Controls (E)	Number of Cases (G×E)	Number of Controls (G×E)
Australian Ovarian Cancer Study	AUS	Australia	1037	1383	791	1078
Connecticut Ovary Study	CON	USA	357	551	0	0
Diseases of the Ovary and their Evaluation	DOV	USA	884	1674	700	1326
Germany Ovarian Cancer Study	GER	Germany	175	526	137	374
Hawaii Ovarian Cancer Study	HAW	USA	277	407	39	72
Hormones and Ovarian Cancer Prediction	HOP	USA	465	1457	328	957
Danish Malignant Ovarian Tumor Study	MAL	Denmark	432	1465	236	602
Mayo Clinic Ovarian Cancer Case Control Study	MAY	USA	201	0	177	0
North Carolina Ovarian Cancer Study	NCO	USA	689	1003	468	682
New England-based Case-Control Study of Ovarian Cancer	NEC	USA	1187	1763	335	557
New Jersey Ovarian Cancer Study	NJO	USA	206	441	0	0
Nijmegen Polygene Study & Nijmegen Biomedical Study	NTH	Netherlands	240	475	231	461
UC Irvine Ovarian Cancer Study	UCI	USA	280	263	104	120
UK Ovarian Cancer Population Study	UKO	UK	351	731	288	630
Los Angeles County Case-Control Studies of Ovarian Cancer	USC	USA	1002	1611	301	667

Web Appendix 7. Equations with respect to Cell Probability p_0 and $p_1$in Simulation

$$\exp(\theta_{GE}) = \frac{p_{01}(p_{01} + p_G + p_E - 1)}{(1 - p_G - p_{01})(1 - p_E - p_{01})}$$

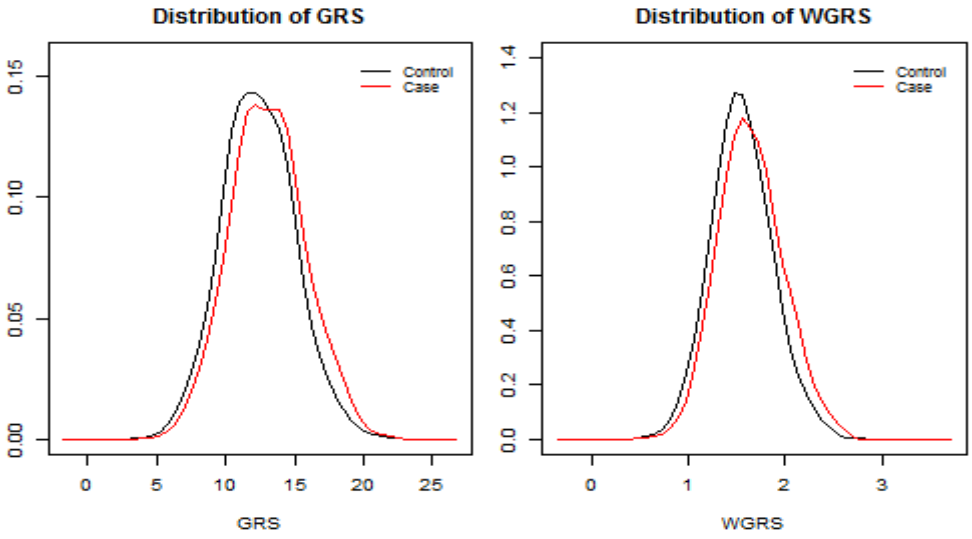
$p_{02} = 1 - p_G - p_{01},\ p_{03} = 1 - p_E - p_{01},\ p_{04} = 1 - p_{01} - p_{02} - p_{03}$

Let $p = p_{01} + p_{02}OR_E + p_{03}OR_G + p_{04}OR_{GE}$,

$p_{11} = p_{01}/p\ ,\ p_{12} = p_{02}OR_E/p\ ,\ p_{13} = p_{03}OR_G/p\ ,\ p_{14} = p_{04}OR_{GE}/p\ ,$

where $OR_{GE} = RERI + OR_G + OR_E - 1$ according to equation (3) given RERI, OR_G and OR_E.

Web Figure 3: Distributions of GRS and WGRS in Cases and Controls in Subjects with European Ancestry

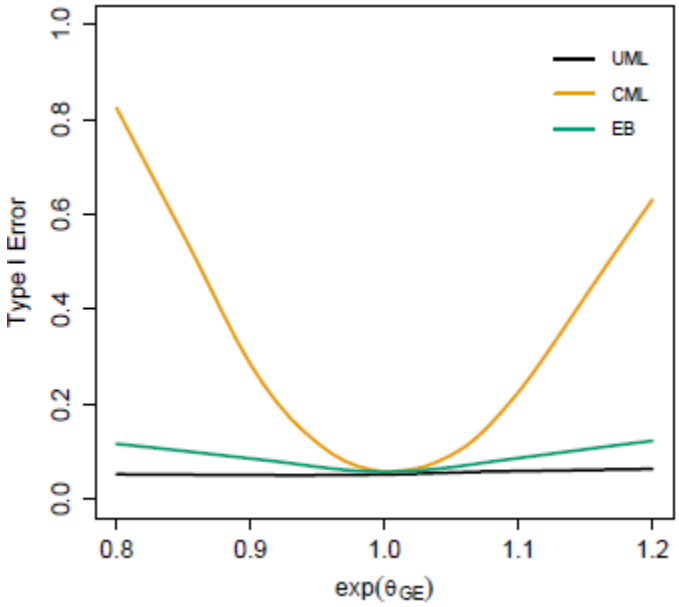


Web Table 7. Type I Error for Different Estimators of RERI Under Four Sample Size and Test Size Settings When $\mathrm{p_G} = 0.2, \mathrm{p_E} = 0.3, \mathrm{OR_G} = 1.2, \mathrm{OR_E} = 1.5$

Setting	$\exp(\theta_{GE})$	0.8	0.9	1	1.1	1.2
N=4000[a], α=0.05	UML	0.049	0.053	0.048	0.047	0.059
	CML	0.822	0.27	0.053	0.219	0.645
	EB	0.085	0.092	0.049	0.078	0.098
N=20000, α=0.05	UML	0.056	0.057	0.05	0.051	0.053
	CML	1	0.854	0.051	0.768	1
	EB	0.061	0.086	0.031	0.099	0.075
N=4000, α =5×10^{-6}	UML	8.00×10^{-6}	7.00×10^{-6}	4.00×10^{-6}	4.00×10^{-6}	3.00×10^{-6}
	CML	4.66×10^{-2}	7.83×10^{-4}	3.00×10^{-6}	1.11×10^{-4}	5.71×10^{-3}
	EB	9.52×10^{-4}	1.42×10^{-4}	1.00×10^{-6}	2.10×10^{-5}	2.36×10^{-4}
N=20000, α =5×10^{-6}	UML	4.00×10^{-6}	7.00×10^{-6}	2.00×10^{-6}	3.00×10^{-6}	5.00×10^{-6}
	CML	9.66×10^{-1}	6.67×10^{-2}	7.00×10^{-6}	2.72×10^{-2}	7.31×10^{-1}
	EB	1.30×10^{-5}	8.68×10^{-4}	4.00×10^{-6}	6.13×10^{-4}	2.80×10^{-5}

[a] N denotes the sample size of cases and controls, $N_0 = N_1 = N$.

Web Figure 4. Type I error for Wald tests of RERI under different θ_{GE} when $p_G = 0.3, p_E = 0.3, OR_G = 1.2, OR_E = 1.5, N = 4000,\ \alpha = 0.05$.



Web Table 8: Type I error for UML, CML and EB estimator of RERI
Fix: $p_G = 0.2, p_E = 0.3, \mathrm{N}_0 = N_1 = 4000, \alpha = 0.05$
Vary: $OR_E = (1.3, 1.5, 1.7)$, $OR_G = (1.1, 1.2, 1.3)$, $\exp(\theta_{GE}) = (0.8, 1, 1.2)$

$\exp(\theta_{GE})$	OR_E	OR_G	UML	CML	EB
	1.3	1.1	0.056	0.766	0.102
	1.3	1.2	0.044	0.792	0.090
	1.3	1.3	0.056	0.822	0.084
	1.5	1.1	0.056	0.790	0.104
0.8	1.5	1.2	0.058	0.822	0.098
	1.5	1.3	0.062	0.832	0.088
	1.7	1.1	0.064	0.804	0.110
	1.7	1.2	0.064	0.814	0.112
	1.7	1.3	0.060	0.834	0.098
	1.3	1.1	0.076	0.040	0.022
	1.3	1.2	0.056	0.050	0.048
	1.3	1.3	0.044	0.060	0.046
	1.5	1.1	0.052	0.048	0.048
1	1.5	1.2	0.054	0.054	0.046
	1.5	1.3	0.058	0.046	0.038
	1.7	1.1	0.058	0.044	0.038
	1.7	1.2	0.062	0.046	0.040
	1.7	1.3	0.054	0.044	0.038
	1.3	1.1	0.056	0.624	0.102
	1.3	1.2	0.064	0.634	0.112
	1.3	1.3	0.052	0.656	0.104
	1.5	1.1	0.046	0.632	0.092
1.2	1.5	1.2	0.066	0.624	0.100
	1.5	1.3	0.058	0.642	0.118
	1.7	1.1	0.066	0.632	0.088
	1.7	1.2	0.070	0.638	0.114
	1.7	1.3	0.064	0.642	0.096

Web Table 9: Type I error for UML, CML and EB estimator of RERI
Fix: $OR_G = 1.2,\ OR_E = 1.5, \mathrm{N}_0 = N_1 = 4000, \alpha = 0.05$
Vary: $p_G = (0.1, 0.2, 0.3), p_E = (0.3, 0.4, 0.5),\ \exp(\theta_{GE}) = (0.8, 1, 1.2)$

$\exp(\theta_{GE})$	p_G	p_E	UML	CML	EB
	0.1	0.3	0.048	0.612	0.114
	0.2	0.3	0.058	0.822	0.098
	0.3	0.3	0.058	0.890	0.100
	0.1	0.4	0.046	0.644	0.092
0.8	0.2	0.4	0.054	0.820	0.076
	0.3	0.4	0.058	0.900	0.080
	0.1	0.5	0.042	0.624	0.092
	0.2	0.5	0.074	0.828	0.090
	0.3	0.5	0.072	0.896	0.078
	0.1	0.3	0.042	0.064	0.044
	0.2	0.3	0.054	0.054	0.046
	0.3	0.3	0.042	0.064	0.046
	0.1	0.4	0.064	0.050	0.038
1	0.2	0.4	0.060	0.082	0.064
	0.3	0.4	0.062	0.072	0.044
	0.1	0.5	0.044	0.058	0.044
	0.2	0.5	0.054	0.052	0.044
	0.3	0.5	0.076	0.038	0.040
	0.1	0.3	0.060	0.372	0.090
	0.2	0.3	0.066	0.624	0.100
	0.3	0.3	0.064	0.730	0.116
	0.1	0.4	0.038	0.428	0.082
1.2	0.2	0.4	0.068	0.658	0.082
	0.3	0.4	0.054	0.766	0.094
	0.1	0.5	0.058	0.408	0.106
	0.2	0.5	0.056	0.632	0.094
	0.3	0.5	0.050	0.760	0.084

Web Appendix 8. Type I Error of EB Test in Large Genomewide Assessment Studies

To justify the EB estimator in genomewide assessment of GxE, we conduct another simulation study similar to that in References (8), which generates 1 significant marker together with M-1 null markers and the G-E odds ratios in control are different between causal makers and null markers. The detailed simulation setting is as below.

we consider the situation with 2,000 cases and 2,000 controls with the number of markers M =10,000. The prevalence of binary E is 0.5 throughout. All main effect parameters are assumed to be unity, namely, $OR_G = OR_E = 1$ across all scenarios, which is another big difference from the simulation in paper. We assume a situation with only 1 causal locus having true interaction with E and others null with no interaction effect. At the causal locus, the prevalence of G is set at 0.35. G-E odds ratio among controls for the causal locus is set at 3 values, namely, exp(θ_{GE})=1.0, 0.9, 1.1, corresponding to independence, negative, and positive dependence. The interaction parameter at the causal locus RERI=exp(β_{GE})-1 is varied from 0.1 to 1. Among the M – 1 null loci, without any interaction effects with E, the prevalence of G is assumed to be Uniform (0.2, 0.5). The population-level G-E association structure among null loci is assumed to be of the form of a mixture distribution reflecting that a large fraction, e.g., $\mathrm{p_{ind}}$, of the SNPs, indeed, is independent of E in the population, whereas the remaining SNPs show some departures from the independence assumption. We generated the log odds ratio of the G-E association in controls corresponding to null loci as $\theta_{\mathrm{G^0E}} \sim \mathrm{p_{ind}}\delta_0 + (1-\mathrm{p_{ind}})N(0, sd = \log(1.5)/2)$. Here, δ_0 is a point mass at 0 reflecting G-E independence. The standard deviation (sd) parameter of the normal distribution part of the mixture is chosen such that, of the θ_{GE} values that depart from independence, 95% fall within $\pm\log(1.5)$. We vary the simulation parameter $\mathrm{p_{ind}}$ from 0.95 to

1.0 to create G and E dependence among more null markers. The family-wise type I error rate is estimated as the empirical proportion of data sets declaring at least 1 null marker to be significant using level of significance α/M. The expected number of false positives is estimated as the average number of falsely rejected null hypotheses, averaged over 1000 data sets. We can see the type I error of EB test still perform well in genomewide studies.

Web Table 10: Power for UML, CML and EB estimator of RERI with $\mathbf{H_0}$: $RERI = 0$
Fix: $p_G = 0.2, p_E = 0.3, OR_G = 1.2, OR_E = 1.5, \mathrm{N}_0 = N_1 = 4000, \alpha = 0.05$
Vary: $RERI = (0,1, \text{step} = 0.1)$, $\exp(\theta_{GE}) = (0.8, 1, 1.2)$

$\exp(\theta_{GE})$	RERI	UML	CML	CML*	EB	EB*
	0	0.049	0.822	0.056	0.085	0.047
	0.1	0.069	0.534	0.008	0.045	0.023
	0.2	0.147	0.253	0.002	0.060	0.030
	0.3	0.317	0.084	0.000	0.155	0.067
	0.4	0.507	0.038	0.000	0.273	0.145
0.8	0.5	0.672	0.133	0.000	0.442	0.275
	0.6	0.820	0.336	0.000	0.637	0.439
	0.7	0.917	0.594	0.005	0.798	0.617
	0.8	0.957	0.800	0.033	0.903	0.786
	0.9	0.989	0.932	0.116	0.968	0.907
	1	0.997	0.980	0.278	0.992	0.967
	0	0.039	0.053	0.053	0.030	0.030
	0.1	0.069	0.135	0.135	0.091	0.091
	0.2	0.175	0.350	0.350	0.258	0.258
	0.3	0.320	0.659	0.659	0.495	0.495
	0.4	0.509	0.858	0.858	0.712	0.712
1	0.5	0.693	0.970	0.970	0.870	0.870
	0.6	0.854	0.997	0.997	0.941	0.941
	0.7	0.916	0.999	0.999	0.974	0.974
	0.8	0.978	1.000	1.000	0.994	0.994
	0.9	0.995	1.000	1.000	0.997	0.997
	1	0.997	1.000	1.000	0.999	0.999
	0	0.061	0.645	0.042	0.098	0.049
	0.1	0.083	0.877	0.187	0.177	0.105
	0.2	0.161	0.981	0.482	0.317	0.207
	0.3	0.351	0.999	0.779	0.486	0.356
	0.4	0.540	1.000	0.944	0.665	0.522
1.2	0.5	0.714	1.000	0.993	0.802	0.718
	0.6	0.858	1.000	0.998	0.920	0.862
	0.7	0.943	1.000	1.000	0.962	0.944
	0.8	0.973	1.000	1.000	0.990	0.973
	0.9	0.995	1.000	1.000	0.999	0.995
	1	0.997	1.000	1.000	0.999	0.998

* denotes the power at empirical α
Empirical α for CML are 10^{-5} and 10^{-4} under $\exp(\theta_{GE})$=0.8 and 1.2.
Empirical α for EB is 0.02 under $\exp(\theta_{GE})$=0.8 and 1.2.

Web Table 11: Power for UML, CML and EB estimator of RERI with $\mathbf{H_0}: \psi = 0$
Fix: $p_G = 0.2, p_E = 0.3, OR_G = 1.2, OR_E = 1.5, \mathrm{N}_0 = N_1 = 4000, \alpha = 0.05$
Vary: $RERI = (0,1, \text{step} = 0.1),\ \exp(\theta_{GE}) = (0.8, 1, 1.2)$

$\exp(\theta_{GE})$	RERI	UML	CML	CML*	EB	EB*
	0	0.071	0.941	0.172	0.161	0.121
	0.1	0.038	0.818	0.053	0.075	0.053
	0.2	0.061	0.597	0.008	0.065	0.049
	0.3	0.149	0.317	0.002	0.073	0.051
	0.4	0.262	0.160	0.001	0.129	0.089
0.8	0.5	0.411	0.060	0.000	0.245	0.183
	0.6	0.574	0.062	0.000	0.367	0.291
	0.7	0.703	0.127	0.000	0.488	0.407
	0.8	0.818	0.283	0.002	0.654	0.550
	0.9	0.902	0.498	0.009	0.773	0.707
	1	0.944	0.702	0.027	0.879	0.821
	0	0.074	0.118	0.118	0.084	0.084
	0.1	0.051	0.049	0.049	0.035	0.035
	0.2	0.084	0.117	0.117	0.084	0.084
	0.3	0.128	0.273	0.273	0.188	0.188
	0.4	0.251	0.534	0.534	0.397	0.397
1	0.5	0.404	0.755	0.755	0.590	0.590
	0.6	0.574	0.911	0.911	0.768	0.768
	0.7	0.703	0.970	0.970	0.857	0.857
	0.8	0.828	0.993	0.993	0.932	0.932
	0.9	0.910	1.000	1.000	0.960	0.960
	1	0.951	1.000	1.000	0.981	0.981
	0	0.087	0.366	0.011	0.068	0.038
	0.1	0.060	0.661	0.055	0.100	0.054
	0.2	0.074	0.878	0.169	0.159	0.100
	0.3	0.154	0.971	0.395	0.282	0.192
	0.4	0.278	0.996	0.681	0.404	0.297
1.2	0.5	0.435	1.000	0.856	0.557	0.436
	0.6	0.567	1.000	0.957	0.669	0.570
	0.7	0.717	1.000	0.991	0.809	0.726
	0.8	0.825	1.000	0.999	0.891	0.828
	0.9	0.917	1.000	1.000	0.946	0.916
	1	0.948	1.000	1.000	0.967	0.947

* denotes the power at empirical α
Empirical α for CML are 10^{-5} and 5×10^{-5} under $\exp(\theta_{GE})$=0.8 and 1.2.
Empirical α for EB are 0.03 and 0.02 under $\exp(\theta_{GE})$=0.8 and 1.2.

Web Table 12: Power of Wald test and LRT for $RERI_{cml}$
Fix: $p_G = 0.2,\ p_E = 0.3, OR_G = 1.2, OR_E = 1.5, \mathrm{N}_0 = N_1 = 4000, \alpha = 0.05$
Vary: $RERI = (0,1)$ with step 0.1, $\exp(\theta_{GE}) = (0.8, 1, 1.2)$

$\exp(\theta_{GE})$	RERI	Wald	LRT
0.8	0	0.822	0.813
	0.1	0.534	0.520
	0.2	0.253	0.244
	0.3	0.084	0.078
	0.4	0.038	0.043
	0.5	0.133	0.140
	0.6	0.336	0.353
	0.7	0.594	0.616
	0.8	0.800	0.813
	0.9	0.932	0.934
	1	0.980	0.983
1	0	0.053	0.055
	0.1	0.135	0.145
	0.2	0.350	0.371
	0.3	0.659	0.676
	0.4	0.858	0.867
	0.5	0.970	0.974
	0.6	0.997	0.999
	0.7	0.999	0.999
	0.8	1.000	1.000
	0.9	1.000	1.000
	1	1.000	1.000
1.2	0	0.645	0.663
	0.1	0.877	0.886
	0.2	0.981	0.982
	0.3	0.999	0.999
	0.4	1.000	1.000
	0.5	1.000	1.000
	0.6	1.000	1.000
	0.7	1.000	1.000
	0.8	1.000	1.000
	0.9	1.000	1.000
	1	1.000	1.000

Web Table 13: Power for UML, CML and EB estimator of RERI with $\mathbf{H_0}$: $RERI = 0$
Fix: $\mathrm{RERI} = 0.5, OR_G = 1.2, OR_E = 1.5, \mathrm{N}_0 = N_1 = 4000, \alpha = 0.05$
Vary: $p_G = (0.1, 0.2, 0.3)$, $p_E = (0.3, 0.4, 0.5)$, $\exp(\theta_{GE}) = (0.8, 1, 1.2)$

$\exp(\theta_{GE})$	p_E	p_G	UML	CML	EB
	0.3	0.1	0.350	0.072	0.147
	0.3	0.2	0.672	0.133	0.442
	0.3	0.3	0.811	0.177	0.628
	0.4	0.1	0.459	0.088	0.243
0.8	0.4	0.2	0.776	0.182	0.571
	0.4	0.3	0.881	0.254	0.761
	0.5	0.1	0.537	0.139	0.320
	0.5	0.2	0.821	0.242	0.666
	0.5	0.3	0.898	0.283	0.792
	0.3	0.1	0.416	0.837	0.638
	0.3	0.2	0.693	0.970	0.870
	0.3	0.3	0.824	0.993	0.925
	0.4	0.1	0.532	0.877	0.713
1	0.4	0.2	0.788	0.986	0.915
	0.4	0.3	0.903	0.995	0.954
	0.5	0.1	0.563	0.892	0.761
	0.5	0.2	0.811	0.988	0.915
	0.5	0.3	0.911	0.998	0.971
	0.3	0.1	0.444	0.998	0.620
	0.3	0.2	0.714	1.000	0.802
	0.3	0.3	0.857	1.000	0.902
	0.4	0.1	0.550	1.000	0.693
1.2	0.4	0.2	0.786	1.000	0.875
	0.4	0.3	0.904	1.000	0.934
	0.5	0.1	0.583	1.000	0.714
	0.5	0.2	0.821	1.000	0.876
	0.5	0.3	0.909	1.000	0.936

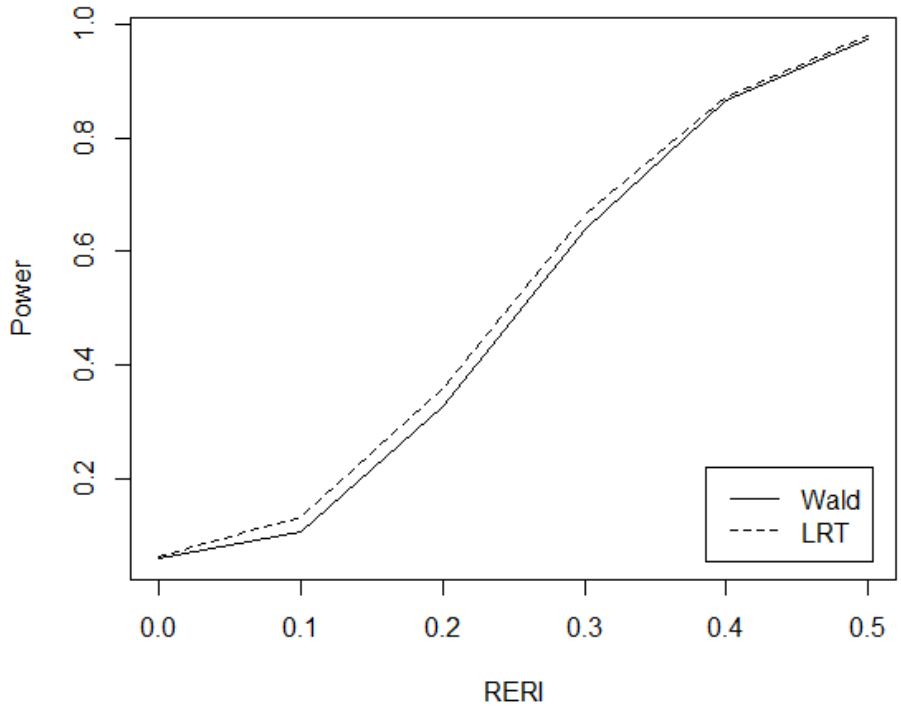
Web Table 14: Power for UML, CML and EB estimator of RERI with $\mathbf{H_0}$: $RERI = 0$
Fix: $\mathrm{RERI} = 0.5, p_G = 0.2,\ p_E = 0.3, \mathrm{N}_0 = N_1 = 4000, \alpha = 0.05$
Vary: $OR_E = (1.3, 1.5, 1.7)$, $OR_G = (1.1, 1.2, 1.3)$, $\exp(\theta_{GE}) = (0.8, 1, 1.2)$

$\exp(\theta_{GE})$	OR_E	OR_G	UML	CML	EB
0.8	1.3	1.1	0.803	0.294	0.597
	1.3	1.2	0.740	0.193	0.529
	1.3	1.3	0.707	0.172	0.483
	1.5	1.1	0.725	0.171	0.505
	1.5	1.2	0.672	0.133	0.442
	1.5	1.3	0.636	0.100	0.403
	1.7	1.1	0.657	0.107	0.397
	1.7	1.2	0.598	0.080	0.356
	1.7	1.3	0.577	0.064	0.344
1	1.3	1.1	0.796	0.993	0.931
	1.3	1.2	0.761	0.988	0.903
	1.3	1.3	0.748	0.979	0.886
	1.5	1.1	0.749	0.984	0.893
	1.5	1.2	0.693	0.970	0.870
	1.5	1.3	0.669	0.961	0.848
	1.7	1.1	0.667	0.958	0.842
	1.7	1.2	0.634	0.946	0.844
	1.7	1.3	0.597	0.938	0.806
1.2	1.3	1.1	0.842	1.000	0.901
	1.3	1.2	0.795	1.000	0.870
	1.3	1.3	0.788	1.000	0.856
	1.5	1.1	0.761	1.000	0.833
	1.5	1.2	0.714	1.000	0.802
	1.5	1.3	0.693	1.000	0.791
	1.7	1.1	0.712	1.000	0.808
	1.7	1.2	0.683	1.000	0.784
	1.7	1.3	0.639	1.000	0.745

Web Table 15: Power for UML, CML and EB estimator of RERI with $\mathbf{H_0}$: $RERI = 0$
Fix: $\mathrm{RERI} = 0.5, OR_G = 1.2, OR_E = 1.5,\ p_G = 0.2,\ p_E = 0.3$
Vary: $\mathrm{N}_0 = N_1 = (400, 4000),\ \alpha = (0.005, 0.05),\ \exp(\theta_{GE}) = (0.8, 0.9, 1, 1.1, 1.2)$

$\mathrm{N}_0 = N_1, \alpha$	$\exp(\theta_{GE})$	UML	CML	EB
	0.8	0.672	0.133	0.442
4000	0.9	0.693	0.697	0.676
0.05	1	0.693	0.970	0.870
	1.1	0.721	1.000	0.846
	1.2	0.714	1.000	0.802
	0.8	0.300	0.023	0.123
4000	0.9	0.321	0.328	0.182
0.005	1	0.346	0.838	0.600
	1.1	0.368	0.995	0.617
	1.2	0.386	1.000	0.512
	0.8	0.030	0.041	0.012
400	0.9	0.032	0.090	0.043
0.05	1	0.044	0.167	0.100
	1.1	0.049	0.287	0.150
	1.2	0.051	0.427	0.213
	0.8	0.000	0.000	0.000
400	0.9	0.000	0.005	0.002
0.005	1	0.000	0.018	0.004
	1.1	0.000	0.028	0.012
	1.2	0.000	0.070	0.020

Web Figure 5. Comparison of Power between LRT and Wald Test for CML estimator when $p_G = 0.3, p_E = 0.3, OR_G = 1.2, OR_E = 1.5, N = 4000, \alpha = 0.05$, exp($\theta_{GE}$)=1 and RERI changes from 0 to 0.5 with a grid level of 0.1.



Web Appendix 9. Discussions on absolute relative bias and MSE

In this section, we first illustrate that $\widehat{RERI}_{EB}$ is preferred to $\widehat{RERI}_{EB1}$ and $\widehat{RERI}_{EB2}$ as described in Remark 1, in terms of MSE, then we compare the absolute relative bias (ARB) and MSE of $\widehat{RERI}_{uml}$, $\widehat{RERI}_{cml}$ and $\widehat{RERI}_{EB}$ (Web Tables 16-19). Compared to $\widehat{RERI}_{EB}$, $\widehat{RERI}_{EB1}$ puts more weight on the CML estimator and it has a larger MSE in 26 of 48 parameter settings. In contrast, $\widehat{RERI}_{EB2}$ puts more weight on UML estimator and it has a larger MSE in 32 of 48 parameter settings. Thus, we choose $\widehat{RERI}_{EB}$ as the proposed empirical Bayes estimator. Next, we compare the ARB of $\widehat{RERI}_{uml}$, $\widehat{RERI}_{cml}$ and $\widehat{RERI}_{EB}$ under different strengths of G-E association. Without loss of generality, let us see the ARB at RERI=0.5 for example. UML is always an unbiased estimator, and the ARB of EB (0.157) is 20% less than that of CML (0.773) when $\exp(\theta_{\mathrm{GE}})$=0.8. Finally, we compare the MSE of the three estimators at RERI=0.5. When $\exp(\theta_{\mathrm{GE}})$=0.8, the MSE of EB (0.103) is 56% of CML (0.182) but 12% higher than UML (0.092). When $\exp(\theta_{\mathrm{GE}})$=1, the MSE of EB (0.054) is 65% of UML (0.083) but 50% higher than CML (0.036). In summary, EB is always intermediate between UML and CML, and it leans towards either UML or CML which has a smaller ARB and MSE depending on $\exp(\theta_{\mathrm{GE}})$ (Web Figure 6).

Web Table 16: Absolute relative bias, variance and MSE of $\widehat{RERI}_{EB}$, $\widehat{RERI}_{EB1}$ and $\widehat{RERI}_{EB2}$
Fix: $OR_G = 1.2,\ OR_E = 1.5,\ p_G = 0.2,\ p_E = 0.3, \mathrm{N}_0 = N_1 = 4000$
Vary: $\exp(\theta_{GE}) = (0.8, 1, 1.2)$, $RERI = (0, 1.5)$ with a grid level of 0.1

$\exp(\theta_{GE})$	RERI	Absolute Relative Bias			Variance			MSE		
		EB	EB1	EB2	EB	EB1	EB2	EB	EB1	EB1
0.8	0	Inf	Inf	Inf	0.037	0.032	0.040	0.078	0.079	0.083
	0.1	0.671	1.244	0.500	0.040	0.035	0.044	0.085	0.085	0.090
	0.2	0.347	0.668	0.243	0.043	0.036	0.047	0.090	0.090	0.096
	0.3	0.225	0.462	0.143	0.043	0.036	0.048	0.091	0.091	0.098
	0.4	0.181	0.378	0.110	0.045	0.037	0.051	0.096	0.096	0.104
	0.5	0.154	0.325	0.089	0.051	0.041	0.058	0.108	0.108	0.118
	0.6	0.140	0.295	0.079	0.053	0.042	0.061	0.113	0.114	0.123
	0.7	0.119	0.263	0.061	0.055	0.042	0.063	0.117	0.117	0.128
	0.8	0.098	0.236	0.042	0.058	0.043	0.067	0.123	0.122	0.135
	0.9	0.098	0.229	0.044	0.064	0.047	0.074	0.136	0.135	0.150
	1	0.084	0.211	0.031	0.068	0.049	0.079	0.143	0.142	0.158
	1.1	0.082	0.205	0.031	0.068	0.048	0.079	0.143	0.146	0.159
	1.2	0.076	0.195	0.025	0.075	0.051	0.087	0.157	0.158	0.175
	1.3	0.074	0.192	0.024	0.076	0.050	0.090	0.162	0.162	0.182
	1.4	0.071	0.186	0.021	0.082	0.054	0.097	0.174	0.175	0.195
	1.5	0.067	0.180	0.017	0.082	0.052	0.097	0.173	0.178	0.194
1	0	Inf	Inf	Inf	0.023	0.019	0.024	0.045	0.038	0.047
	0.1	0.026	0.026	0.014	0.023	0.019	0.024	0.046	0.038	0.048
	0.2	0.001	0.002	0.006	0.023	0.019	0.025	0.047	0.039	0.050
	0.3	0.004	0.004	0.001	0.025	0.020	0.027	0.051	0.041	0.054
	0.4	0.000	0.001	0.004	0.028	0.023	0.030	0.056	0.045	0.061
	0.5	0.007	0.007	0.004	0.031	0.025	0.034	0.062	0.050	0.069
	0.6	0.001	0.000	0.003	0.032	0.026	0.035	0.064	0.051	0.070
	0.7	0.000	0.001	0.002	0.033	0.026	0.036	0.065	0.051	0.073
	0.8	0.005	0.005	0.002	0.036	0.028	0.040	0.072	0.055	0.081
	0.9	0.004	0.004	0.002	0.035	0.027	0.039	0.069	0.055	0.078
	1	0.001	0.001	0.001	0.036	0.028	0.041	0.072	0.056	0.082
	1.1	0.000	0.000	0.002	0.042	0.033	0.048	0.084	0.065	0.096
	1.2	0.002	0.002	0.004	0.040	0.031	0.046	0.080	0.063	0.093
	1.3	0.002	0.002	0.003	0.041	0.032	0.047	0.082	0.064	0.095
	1.4	0.005	0.005	0.007	0.047	0.036	0.055	0.094	0.072	0.110
	1.5	0.005	0.004	0.006	0.051	0.039	0.059	0.101	0.077	0.118
	0	Inf	Inf	Inf	0.034	0.031	0.033	0.073	0.076	0.071
	0.1	0.736	1.234	0.617	0.035	0.032	0.035	0.075	0.080	0.073
	0.2	0.380	0.655	0.307	0.039	0.035	0.039	0.084	0.088	0.082
	0.3	0.264	0.466	0.206	0.039	0.036	0.040	0.085	0.091	0.083
	0.4	0.206	0.372	0.155	0.044	0.039	0.044	0.094	0.100	0.093

	0.5	0.168	0.312	0.121	0.047	0.041	0.048	0.100	0.107	0.099
1.2	0.6	0.139	0.268	0.095	0.050	0.044	0.051	0.106	0.113	0.106
	0.7	0.128	0.247	0.085	0.053	0.046	0.055	0.113	0.122	0.113
	0.8	0.124	0.237	0.082	0.046	0.042	0.047	0.102	0.121	0.099
	0.9	0.106	0.213	0.065	0.053	0.048	0.055	0.115	0.133	0.112
	1	0.106	0.209	0.065	0.055	0.050	0.058	0.122	0.143	0.120
	1.1	0.098	0.197	0.058	0.057	0.051	0.059	0.125	0.150	0.122
	1.2	0.088	0.184	0.049	0.060	0.055	0.063	0.132	0.158	0.129
	1.3	0.083	0.177	0.044	0.065	0.059	0.067	0.140	0.170	0.137
	1.4	0.080	0.171	0.041	0.066	0.060	0.069	0.145	0.178	0.142
	1.5	0.080	0.170	0.042	0.077	0.068	0.082	0.168	0.201	0.167

Web Table 17: Absolute relative bias, variance and MSE of $\widehat{RERI}_{uml}$, $\widehat{RERI}_{cml}$ and $\widehat{RERI}_{EB}$
Fix: $OR_G = 1.2,\ OR_E = 1.5,\ p_G = 0.2,\ p_E = 0.3, \mathrm{N}_0 = N_1 = 4000$
Vary: $\exp(\theta_{GE}) = (0.8, 0.9, 1, 1.1, 1.2)$, $RERI = (0,1.5)$ with a grid level of 0.1

$\exp(\theta_{GE})$	RERI	Absolute Relative Bias			Variance			MSE		
		UML	CML	EB	UML	CML	EB	UML	CML	EB
0.8	0	Inf	Inf	Inf	0.033	0.013	0.033	0.067	0.135	0.075
	0.1	0.082	3.410	0.680	0.034	0.014	0.034	0.070	0.144	0.079
	0.2	0.030	1.770	0.362	0.037	0.014	0.037	0.075	0.153	0.084
	0.3	0.034	1.206	0.236	0.039	0.014	0.039	0.080	0.160	0.089
	0.4	0.012	0.943	0.196	0.039	0.015	0.039	0.083	0.173	0.094
	0.5	0.015	0.773	0.157	0.046	0.016	0.046	0.092	0.182	0.103
	0.6	0.011	0.664	0.136	0.046	0.017	0.046	0.095	0.192	0.107
	0.7	0.011	0.585	0.120	0.053	0.018	0.053	0.105	0.203	0.117
	0.8	0.010	0.526	0.107	0.055	0.018	0.055	0.110	0.214	0.123
	0.9	0.009	0.480	0.098	0.056	0.020	0.056	0.114	0.225	0.127
	1	0.012	0.439	0.087	0.060	0.022	0.060	0.121	0.235	0.135
	1.1	0.010	0.410	0.083	0.056	0.021	0.056	0.121	0.246	0.135
	1.2	0.008	0.387	0.080	0.068	0.022	0.067	0.136	0.260	0.151
	1.3	0.006	0.367	0.077	0.072	0.026	0.072	0.144	0.277	0.161
	1.4	0.008	0.347	0.071	0.074	0.025	0.074	0.150	0.285	0.167
	1.5	0.011	0.328	0.065	0.081	0.027	0.080	0.161	0.295	0.177
0.9	0	Inf	Inf	Inf	0.032	0.013	0.025	0.065	0.051	0.057
	0.1	0.025	1.669	0.649	0.034	0.014	0.026	0.069	0.056	0.061
	0.2	0.009	0.841	0.314	0.034	0.014	0.027	0.071	0.057	0.063
	0.3	0.001	0.584	0.222	0.037	0.015	0.028	0.075	0.061	0.067
	0.4	0.001	0.450	0.170	0.039	0.016	0.030	0.080	0.064	0.070
	0.5	0.003	0.370	0.138	0.042	0.016	0.032	0.085	0.068	0.075
	0.6	0.006	0.314	0.116	0.045	0.018	0.034	0.091	0.072	0.080
	0.7	0.006	0.275	0.101	0.048	0.019	0.036	0.096	0.075	0.084
	0.8	0.002	0.251	0.094	0.048	0.020	0.037	0.100	0.080	0.088
	0.9	0.002	0.229	0.086	0.051	0.021	0.039	0.106	0.084	0.092
	1	0.005	0.208	0.076	0.056	0.023	0.043	0.114	0.089	0.099
	1.1	0.002	0.196	0.074	0.057	0.025	0.044	0.118	0.095	0.104
	1.2	0.002	0.185	0.070	0.061	0.024	0.046	0.125	0.098	0.110
	1.3	0.003	0.174	0.065	0.063	0.026	0.048	0.131	0.103	0.115
	1.4	0.005	0.163	0.060	0.070	0.028	0.054	0.141	0.107	0.123
	1.5	0.004	0.157	0.058	0.076	0.030	0.057	0.150	0.114	0.130
1	0	Inf	Inf	Inf	0.030	0.014	0.019	0.061	0.028	0.040
	0.1	0.020	0.018	0.029	0.033	0.015	0.021	0.066	0.029	0.044
	0.2	0.007	0.008	0.013	0.036	0.015	0.023	0.071	0.031	0.047
	0.3	0.006	0.006	0.010	0.036	0.016	0.023	0.073	0.032	0.048
	0.4	0.011	0.011	0.014	0.040	0.018	0.026	0.079	0.035	0.053
	0.5	0.002	0.003	0.005	0.041	0.018	0.026	0.083	0.036	0.054

1	0.6	0.002	0.003	0.005	0.043	0.019	0.028	0.087	0.038	0.058
	0.7	0.002	0.003	0.005	0.049	0.022	0.031	0.095	0.042	0.063
	0.8	0.001	0.000	0.001	0.048	0.022	0.031	0.098	0.043	0.064
	0.9	0.005	0.004	0.002	0.053	0.023	0.033	0.105	0.046	0.069
	1	0.001	0.001	0.001	0.055	0.027	0.036	0.110	0.051	0.074
	1.1	0.001	0.003	0.004	0.060	0.026	0.039	0.118	0.052	0.079
	1.2	0.001	0.003	0.004	0.063	0.027	0.040	0.124	0.054	0.082
	1.3	0.010	0.009	0.007	0.064	0.030	0.041	0.129	0.059	0.086
	1.4	0.005	0.003	0.002	0.070	0.030	0.044	0.138	0.061	0.091
	1.5	0.005	0.004	0.003	0.069	0.032	0.045	0.141	0.064	0.093
	0	Inf	Inf	Inf	0.032	0.014	0.026	0.062	0.050	0.054
	0.1	0.010	1.476	0.557	0.034	0.016	0.028	0.066	0.053	0.058
	0.2	0.016	0.751	0.275	0.033	0.016	0.027	0.067	0.054	0.058
	0.3	0.014	0.541	0.213	0.035	0.017	0.028	0.071	0.061	0.062
	0.4	0.001	0.407	0.154	0.036	0.018	0.029	0.074	0.062	0.064
	0.5	0.005	0.340	0.131	0.039	0.018	0.031	0.079	0.067	0.069
1.1	0.6	0.001	0.285	0.107	0.043	0.021	0.034	0.085	0.070	0.073
	0.7	0.005	0.257	0.099	0.044	0.021	0.035	0.089	0.075	0.077
	0.8	0.000	0.227	0.086	0.046	0.023	0.038	0.094	0.079	0.082
	0.9	0.003	0.209	0.081	0.048	0.024	0.039	0.099	0.084	0.086
	1	0.005	0.195	0.076	0.054	0.026	0.044	0.107	0.090	0.094
	1.1	0.005	0.183	0.072	0.055	0.028	0.045	0.111	0.096	0.097
	1.2	0.004	0.170	0.067	0.056	0.028	0.045	0.115	0.099	0.100
	1.3	0.009	0.167	0.069	0.061	0.031	0.050	0.123	0.109	0.109
	1.4	0.010	0.159	0.066	0.063	0.034	0.052	0.128	0.117	0.114
	1.5	0.008	0.150	0.062	0.069	0.035	0.057	0.138	0.121	0.122
	0	Inf	Inf	Inf	0.031	0.015	0.031	0.061	0.108	0.068
	0.1	0.021	2.850	0.688	0.033	0.016	0.033	0.064	0.113	0.071
	0.2	0.005	1.478	0.359	0.033	0.016	0.033	0.066	0.120	0.073
	0.3	0.000	1.019	0.250	0.035	0.018	0.035	0.070	0.129	0.078
	0.4	0.004	0.783	0.188	0.039	0.019	0.039	0.076	0.136	0.084
	0.5	0.002	0.645	0.156	0.042	0.020	0.042	0.081	0.145	0.090
1.2	0.6	0.003	0.552	0.132	0.043	0.023	0.045	0.084	0.154	0.094
	0.7	0.003	0.491	0.122	0.047	0.025	0.047	0.090	0.166	0.101
	0.8	0.004	0.443	0.111	0.049	0.026	0.049	0.095	0.176	0.106
	0.9	0.005	0.405	0.102	0.049	0.026	0.049	0.098	0.185	0.109
	1	0.001	0.371	0.092	0.051	0.027	0.051	0.102	0.193	0.114
	1.1	0.006	0.351	0.090	0.056	0.030	0.056	0.110	0.209	0.124
	1.2	0.009	0.332	0.088	0.056	0.031	0.056	0.112	0.222	0.128
	1.3	0.009	0.315	0.084	0.062	0.035	0.062	0.122	0.236	0.137
	1.4	0.006	0.297	0.078	0.065	0.037	0.065	0.128	0.245	0.144
	1.5	0.010	0.287	0.078	0.066	0.038	0.066	0.132	0.261	0.151

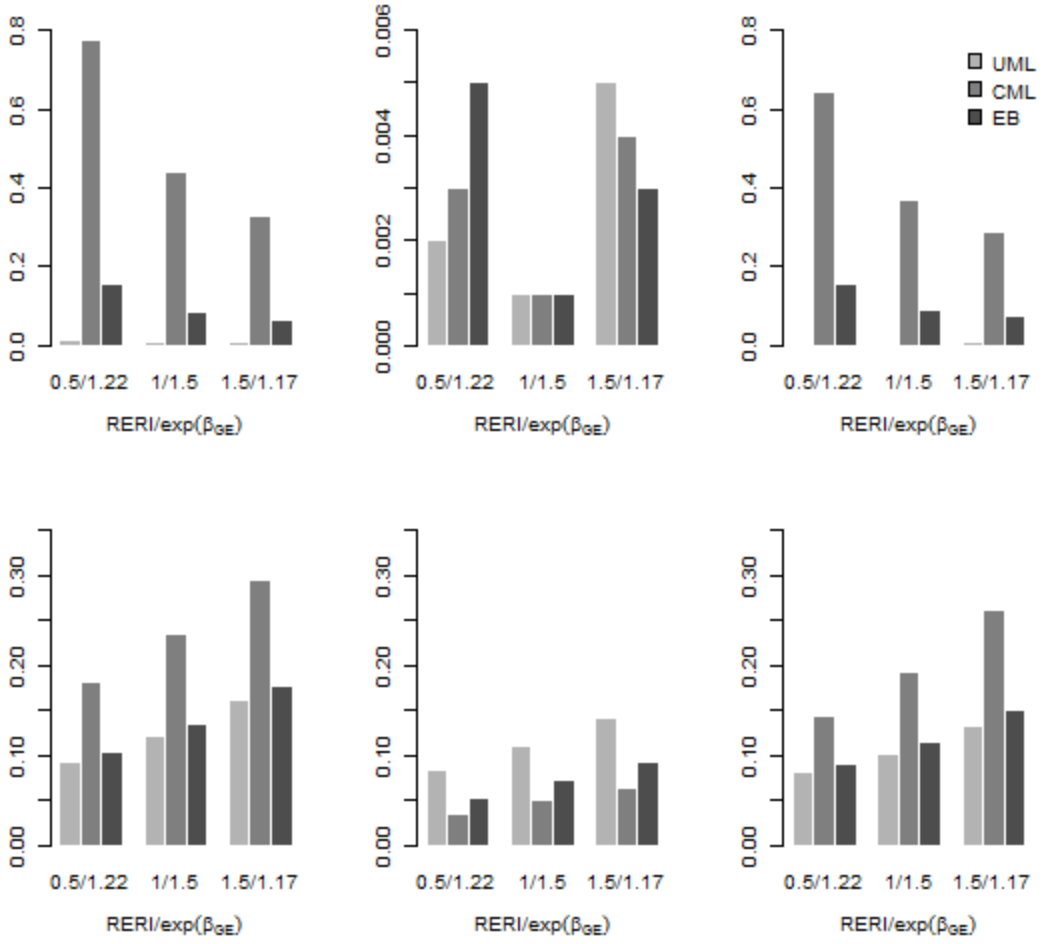
Web Table 18: Absolute relative bias, variance and MSE of $\widehat{RERI}_{uml}$, $\widehat{RERI}_{cml}$ and $\widehat{RERI}_{EB}$
Fix: $RERI = 0.5,\ p_G = 0.2,\ p_E = 0.3, \mathrm{N}_0 = N_1 = 4000$
Vary: $OR_E = (1.3, 1.5, 1.7), OR_G = (1.1, 1.2, 1.3),\ \exp(\theta_{GE}) = (0.9, 1, 1.1)$

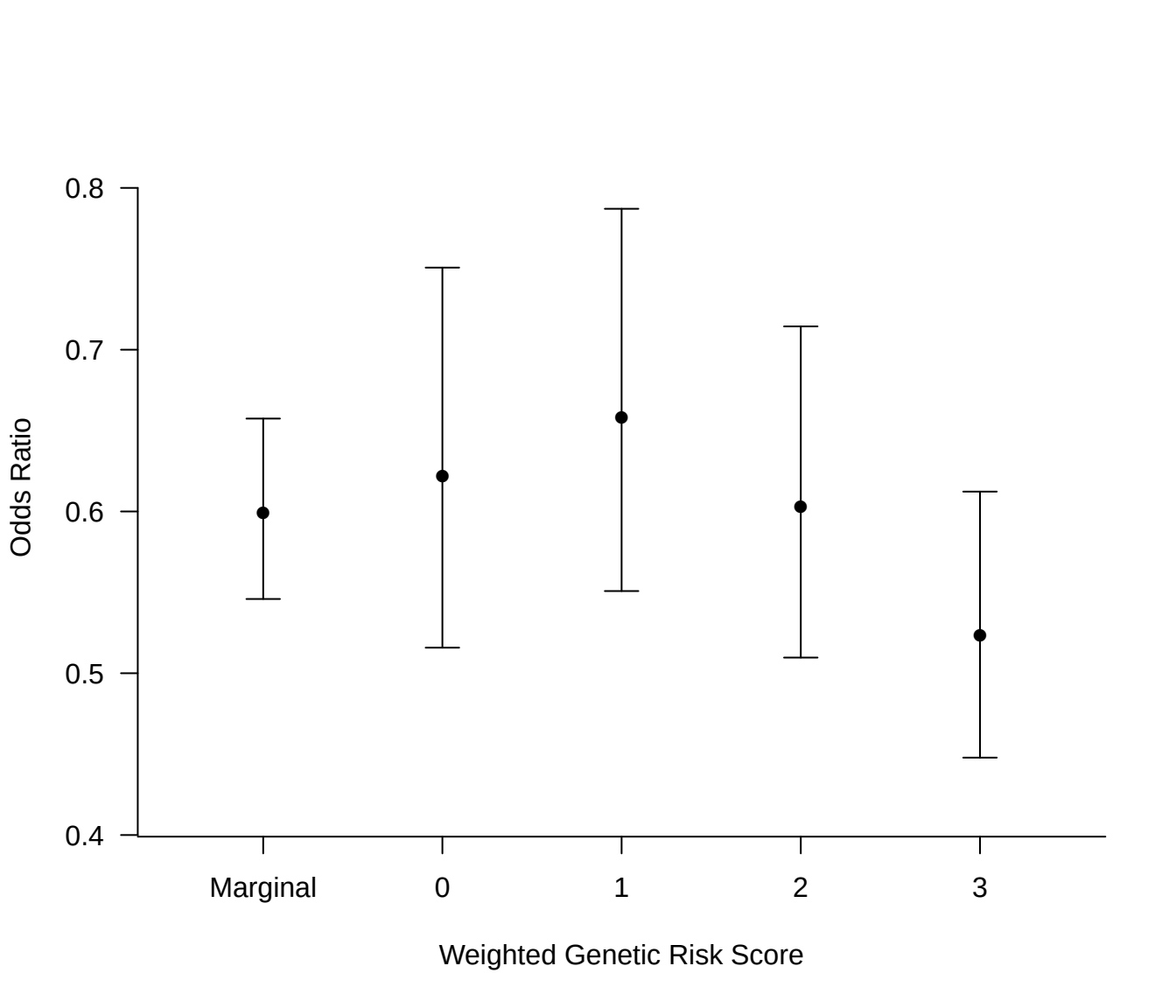
$\exp(\theta_{GE})$	OR_E	OR_G	ARB			Variance			MSE		
			UML	CML	EB	UML	CML	EB	UML	CML	EB
0.9	1.3	1.1	0.004	0.330	0.129	0.032	0.013	0.024	0.065	0.054	0.058
	1.3	1.2	0.001	0.344	0.131	0.035	0.015	0.026	0.071	0.059	0.063
	1.3	1.3	0.003	0.362	0.138	0.040	0.015	0.031	0.080	0.064	0.071
	1.5	1.1	0.005	0.351	0.131	0.039	0.015	0.030	0.079	0.062	0.070
	1.5	1.2	0.003	0.370	0.138	0.042	0.016	0.032	0.085	0.068	0.075
	1.5	1.3	0.008	0.381	0.138	0.048	0.017	0.037	0.095	0.072	0.083
	1.7	1.1	0.007	0.378	0.140	0.045	0.018	0.035	0.092	0.072	0.081
	1.7	1.2	0.002	0.399	0.150	0.049	0.020	0.037	0.099	0.080	0.087
	1.7	1.3	0.000	0.419	0.158	0.052	0.021	0.040	0.107	0.086	0.094
1	1.3	1.1	0.005	0.005	0.007	0.032	0.015	0.020	0.065	0.029	0.043
	1.3	1.2	0.004	0.005	0.007	0.037	0.015	0.023	0.073	0.031	0.047
	1.3	1.3	0.003	0.003	0.006	0.037	0.017	0.023	0.075	0.034	0.050
	1.5	1.1	0.004	0.003	0.000	0.038	0.017	0.024	0.077	0.034	0.051
	1.5	1.2	0.002	0.003	0.005	0.041	0.018	0.026	0.083	0.036	0.054
	1.5	1.3	0.001	0.001	0.004	0.044	0.021	0.028	0.090	0.040	0.059
	1.7	1.1	0.005	0.005	0.002	0.044	0.020	0.027	0.089	0.040	0.058
	1.7	1.2	0.002	0.002	0.002	0.046	0.021	0.029	0.094	0.042	0.062
	1.7	1.3	0.000	0.000	0.003	0.052	0.023	0.033	0.104	0.045	0.069
1.1	1.3	1.1	0.009	0.302	0.120	0.031	0.015	0.025	0.062	0.053	0.055
	1.3	1.2	0.009	0.317	0.126	0.033	0.016	0.027	0.068	0.058	0.059
	1.3	1.3	0.009	0.333	0.130	0.037	0.018	0.030	0.074	0.064	0.065
	1.5	1.1	0.004	0.323	0.125	0.036	0.018	0.029	0.074	0.062	0.064
	1.5	1.2	0.005	0.340	0.131	0.039	0.018	0.031	0.079	0.067	0.069
	1.5	1.3	0.008	0.358	0.139	0.043	0.021	0.034	0.087	0.073	0.076
	1.7	1.1	0.003	0.349	0.134	0.043	0.020	0.034	0.086	0.072	0.075
	1.7	1.2	0.010	0.372	0.146	0.045	0.022	0.036	0.092	0.079	0.081
	1.7	1.3	0.007	0.383	0.147	0.051	0.023	0.040	0.101	0.084	0.088

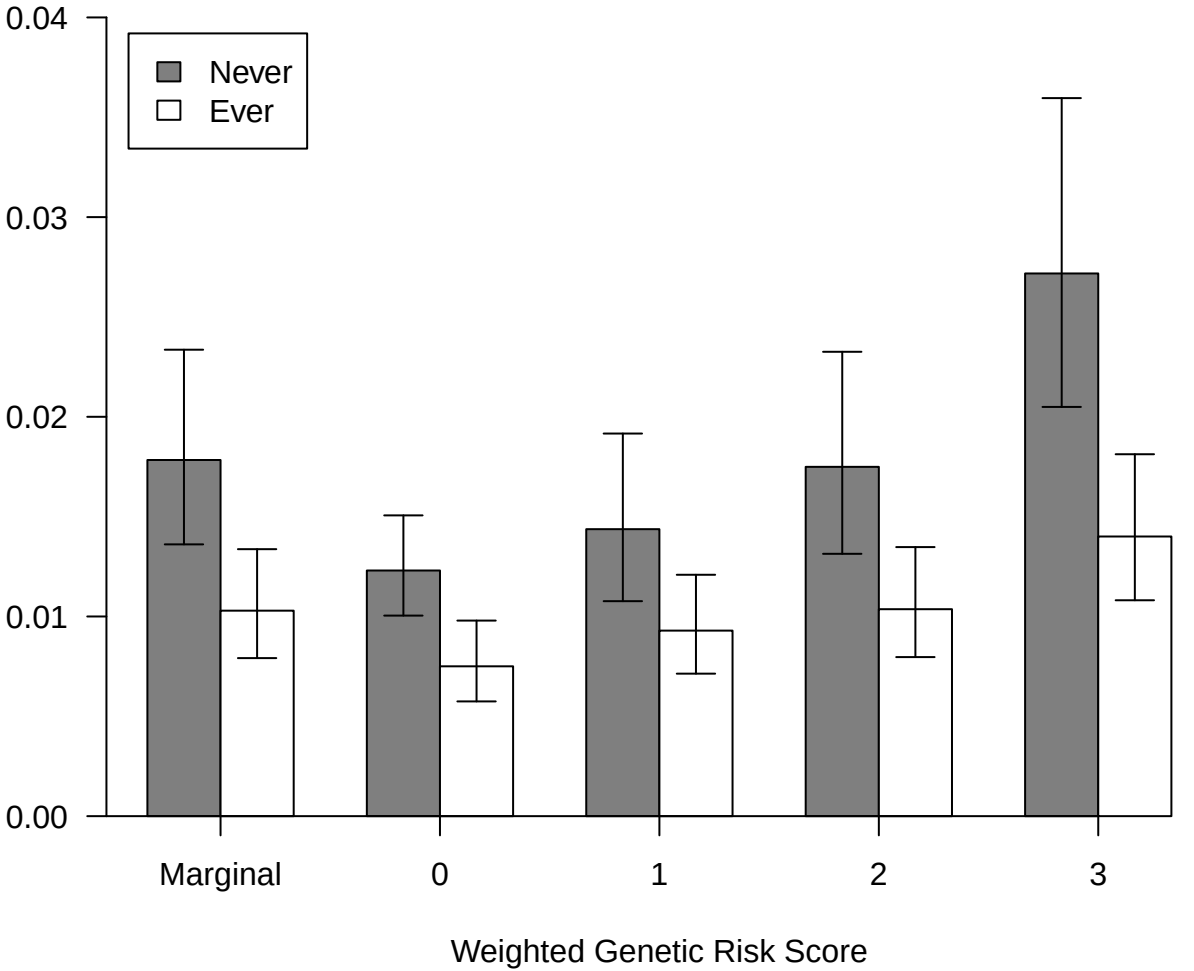
Web Table 19: Absolute relative bias, variance and MSE of $\widehat{RERI}_{uml}$, $\widehat{RERI}_{cml}$ and $\widehat{RERI}_{EB}$
Fix: $RERI = 0.5, OR_E = 1.5, OR_G = 1.2, \mathrm{N}_0 = N_1 = 4000$
Vary: $p_G = (0.1, 0.2, 0.3),\ p_E = (0.3, 0.4, 0.5)\,,\ \exp(\theta_{GE}) = (0.9, 1, 1.1)$

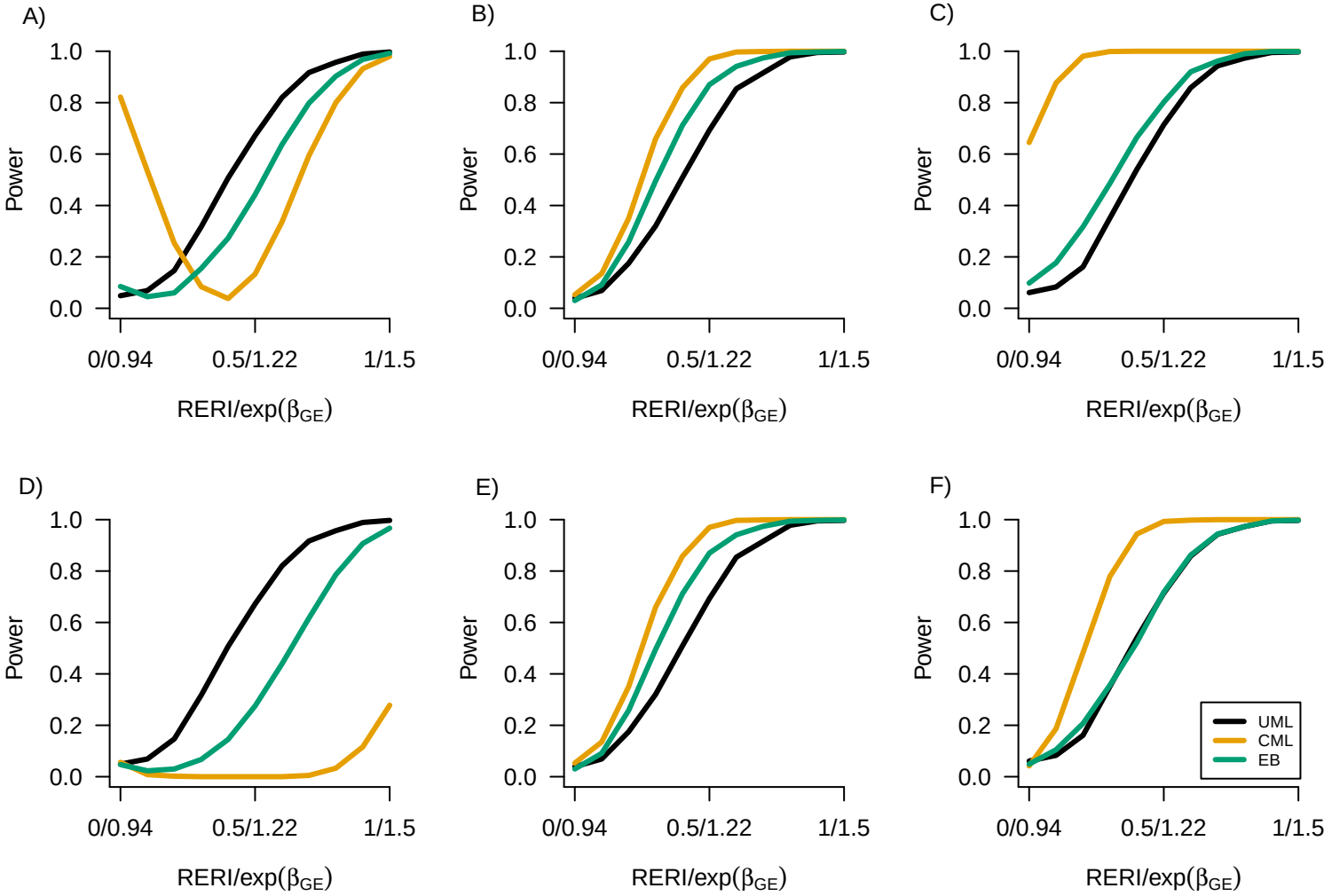
$\exp(\theta_{GE})$	p_E	p_G	ARB			Variance			MSE		
			UML	CML	EB	UML	CML	EB	UML	CML	EB
0.9	0.3	0.1	0.000	0.381	0.163	0.083	0.029	0.056	0.166	0.095	0.130
	0.3	0.2	0.003	0.370	0.138	0.042	0.016	0.032	0.085	0.068	0.075
	0.3	0.3	0.008	0.357	0.133	0.030	0.013	0.025	0.061	0.058	0.057
	0.4	0.1	0.018	0.363	0.135	0.068	0.025	0.049	0.135	0.084	0.109
	0.4	0.2	0.018	0.334	0.113	0.038	0.016	0.031	0.073	0.059	0.066
	0.4	0.3	0.020	0.324	0.097	0.027	0.012	0.023	0.053	0.050	0.049
	0.5	0.1	0.013	0.333	0.130	0.059	0.027	0.043	0.118	0.080	0.096
	0.5	0.2	0.003	0.333	0.121	0.032	0.014	0.025	0.063	0.057	0.057
	0.5	0.3	0.011	0.315	0.102	0.023	0.012	0.020	0.045	0.048	0.043
1	0.3	0.1	0.017	0.010	0.008	0.082	0.031	0.049	0.161	0.064	0.104
	0.3	0.2	0.002	0.003	0.005	0.041	0.018	0.026	0.083	0.036	0.054
	0.3	0.3	0.008	0.002	0.009	0.033	0.014	0.021	0.063	0.027	0.042
	0.4	0.1	0.019	0.000	0.007	0.065	0.026	0.040	0.130	0.054	0.085
	0.4	0.2	0.013	0.011	0.007	0.036	0.016	0.022	0.070	0.032	0.046
	0.4	0.3	0.006	0.006	0.002	0.025	0.012	0.016	0.049	0.024	0.034
	0.5	0.1	0.006	0.001	0.005	0.057	0.027	0.037	0.115	0.054	0.078
	0.5	0.2	0.009	0.004	0.005	0.033	0.015	0.021	0.063	0.030	0.043
	0.5	0.3	0.007	0.002	0.006	0.023	0.011	0.015	0.045	0.023	0.031
1.1	0.3	0.1	0.002	0.356	0.136	0.074	0.035	0.056	0.150	0.101	0.118
	0.3	0.2	0.005	0.340	0.131	0.039	0.018	0.031	0.079	0.067	0.069
	0.3	0.3	0.004	0.320	0.118	0.032	0.014	0.026	0.061	0.054	0.055
	0.4	0.1	0.007	0.332	0.136	0.057	0.028	0.043	0.119	0.086	0.096
	0.4	0.2	0.016	0.326	0.134	0.034	0.018	0.028	0.067	0.061	0.060
	0.4	0.3	0.014	0.311	0.121	0.026	0.013	0.022	0.050	0.049	0.047
	0.5	0.1	0.023	0.322	0.146	0.056	0.029	0.042	0.113	0.083	0.093
	0.5	0.2	0.014	0.290	0.101	0.031	0.015	0.025	0.062	0.052	0.054
	0.5	0.3	0.009	0.287	0.099	0.023	0.012	0.020	0.045	0.044	0.042

Web Figure 6. Absolute relative bias and MSE for UML, CML, EB estimator of RERI under nine settings: data generated on 4000 cases and 4000 controls with 1000 replications under RERI=0.5, 1, 1.5, corresponding multiplicative interaction 1.22, 1.5, 1.78 and fixed parameters $p_G = 0.2$, $p_E = 0.3$, $OR_G = 1.2$, $OR_E = 1.5$. The top panel corresponds to the absolute relative bias, whereas the bottom panel corresponds to the MSE. The left, center, and right panels correspond to different values of the G-E odds ratio, i.e. exp(θ_{GE})=0.8, 1, 1.2.




0.8
0.7
0.6
0.5
0.4
Odds Ratio
Marginal
0
1
2
3
Weighted Genetic Risk Score


0.04
0.03
0.02
0.01
0.00
Never
Ever
Marginal
0
1
2
3
Weighted Genetic Risk Score


A)
B)
C)
D)
E)
F)
Power
1.0
0.8
0.6
0.4
0.2
0.0
0/0.94
0.5/1.22
1/1.5
RERI/exp(β_GE)
UML
CML
EB