SUPPLEMENTAL METHODS

In order to calculate variance estimates for different combinations of risk factors simply by summing their variance values, we have to be able to assume independence of the risk estimates for the variables in question. To test for independence, we assessed whether or not the off-diagonal elements of the covariance matrix of estimates in the DOVE, HOPE, NCOCS, and USC sites substantially affected the total variance for each combination of risk variables. We ran a logistic regression modeling OCP use, tubal ligation, endometriosis, family history of ovarian cancer, parity, and risk score quintile against ovarian cancer status by site for each of the four sites. For each site, we estimated the covariance matrix for the parameter estimates.

For each site, we generated every unique combination of the genetic risk score and environmental variables, and calculated the variance for each combination using the previously calculated covariance matrices. We then calculated the variance using only the diagonal of the covariance matrices (*i.e.*, only variance with no covariance), thus assuming risk estimates for the variables are independent of each other. We compared the two sets of variance estimates, in order to assess the validity of the assumption of independence. The two sets of variance values were very similar for all four sites.

An additional complication in calculating the variance genetic score arose because only the individual SNP allele risk estimates and variances had been published and not the behavior of the genetic score.

For each combination of SNP alleles within the DOVE, HOPE, NCOCS, and USC sites we calculated a genetic risk score by multiplying the published odds ratio for each allele into a combined risk variable (assuming that the effect of each SNP is independent), and then categorizing this variable into quintiles based on the frequency of the underlying combination among controls. In order to obtain the odds ratio for each of these quintiles, we averaged the combined genetic risk odds ratios for all the combinations within the given quintile, weighted by their frequency within the quintile. We then scaled the quintile odds ratios to be in reference to the lowest risk quintile by dividing the odds ratios by the lowest quintile odds ratio.

Obtaining variance estimates for the quintile odds ratios using only published data was not as straightforward, because which quintile a given combination is placed in is not independent of the published odds ratios. We ran a simulation where normally distributed random values were added onto the log odds ratios for each SNP, and the quintile groupings for each combination of SNP alleles were recalculated using these new odds ratios. Then we calculated the quintile variance estimates by multiplying the squared average values for each SNP in the quintile by the variance values for each SNP, and then summing those values together. We ran this simulation 1000 times to assess the stability of the variance values.

The variance values were quite stable, and thus we were able to make the assumption of independence between the quintile categories and the variance of the SNP log odds ratios for the purpose of calculating the variance of the quintile estimates. We therefore calculated the quintile variance estimates for each quintile by multiplying average SNP value within the quintile and SNP variance for each SNP, and then summed these values.