Appendix 4: Calculation of Power for permutation tests [posted as supplied by author]

To estimate the power of the permutation test for a specified relative risk restricted to a specified window, we first estimated the empirical distribution of the test statistic (the minimum p-value among all the standard tests over the set of subsets) under the null hypothesis from the set of permutations. We estimated the miscarriage rates in the placebo arm consistent with the observed total number of miscarriages and the relative risk and set of pregnancies at increased risk specified under the alternative hypothesis; for the intervention arm, we took the product of the miscarriage rate in the placebo arm and the relative risk as the rate in the intervention arm in the set of pregnancies at increased risk, and the rate in placebo arm as the rate in for the other pregnancies in the intervention arm. We simulated miscarriage in the pregnancies using the probability of miscarriage described above.

For each simulation, we calculated the minimum P-value among all the tests in each data set simulated. The power for a 1-sided α test is the percentage of simulations with minimum P-value below the quantile of the null distribution. Powers of the permutation test are simulated for relative risks from 1.0 to 4.0 with the increasing step of 0.02 (RR = 1.00, 1.02, 1.04, …, 3.96, 3.98, 4.00). For any give power *P*, the minimal relative risk which can be detected using the permutation test is defined by the smallest *RR* with power ≥ *P.*

The power for the standard test is defined by the probability Pr($Z > z_{1-\alpha}$ / relative risk in a given window > 1), where Z is the Wald-type test statistic defined in Appendix 2.1.2 and $z_{1-\alpha}$ is the 1- α quantile of the standard normal distribution.

Power comparisons for permutation tests of miscarriage effect. Supplemental Tables 1a-1e show the benefit of the permutation test. For example, when the excess risk is restricted to pregnancies conceived between 0 and 29 days after vaccination, the standard test of difference in proportions can detect a lower relative risk (2.40) in the window with power of 0.95 than can the permutation test (2.88) with power of 90% (Supplemental Table 1c). But if the subsets of pregnancies conceived in the [0,14], [0,44], [0,59], or [0,89] intervals of time between vaccination and conception were assumed to be the windows of elevated risk in the HPV arm, the corresponding minimum detectable relative risks are higher (3.08, 2.64, 2.78, and 2.98) when the risk is elevated only in the [0,29] window. The minimum detectable risks become far higher as the overlap of pregnancies in the true window and in the presumed window decreases. Thus, the permutation test provides assurance that power will be maintained when the precise window of excess risk is unknown at the cost of reduced power when the precise window is correctly specified. Minimum detectable risk for power of 0.99, 0.95, 0.9, 0.8 and 0.5 are shown in Supplemental Tables 1a to 1e.