Appendix 3: Permutation Methods [posted as supplied by author]

Permutation test to address lack of knowledge of time of conception during which vaccination might confer risk. Any extra risk of miscarriage conferred by the vaccine may be limited to subsets of pregnancies defined by time between vaccination and conception. The originally planned analyses of pregnancy outcomes considered all pregnancies, regardless of time since vaccination, or were restricted to pregnancies conceived within a few weeks of vaccination. Pre-specifying the wrong subset of concern can reduce power in two ways, however. When the subset at putative risk includes pregnancies outside the period of excess risk, pregnancies at no increased risk dampen the effect of pregnancies at excess risk, thereby reducing the apparent effect size and power. When the putative subset does not include all pregnancies at excess risk, power is lower, even though the estimate of effect remains unbiased when the strength of the effect is the same for all pregnancies in the subset. Expert consultants identified at the request of the safety monitoring committees (AW and GM) were unable to pre-specify the set of pregnancies, defined by time between vaccination and conception, during which an effect of vaccination on miscarriage risk would be most likely to occur. Therefore, our primary analysis uses a permutation test[1 2] that allowed a range of subsets of pregnancies at elevated risk due to vaccination against HPV.

The null hypothesis of the permutation test is that the vaccine confers no additional risk of miscarriage for any set of pregnancies, regardless of time between conception and vaccination. The alternative hypothesis is that the vaccine confers additional risk of miscarriage for at least one subset of

pregnancies conceived during a specified time interval after vaccination. The permutation test ensures that the claimed or nominal chance of rejection of the null hypothesis when the null hypothesis holds is accurate. The permutation test has reasonable power loss compared with a standard test where the subset of pregnancies at increased risk is correctly specified, and still maintains power, even when the assumed subset of pregnancies at increased risk is far from correct. Because the tests for overlapping subsets can have strong positive correlations, power loss from specifying a broad range of possible subsets is less than in the familiar family-wise error rate ("multiple comparisons") setting where the tests are usually approximately independent.

The test statistic for the permutation test is the largest Z-statistic among tests of equality of proportion in subsets of pregnancies defined by pre-specified time between vaccination and estimated conception date. The P-value from a permutation test is the quantile (percentile) of the observed test statistic in the distribution of the test statistic under the null hypothesis. Here, a set of random permutations generate the empirical null distribution from test statistics calculated from a set of virtual studies, each with assignment of study units to one of two virtual treatment arms.

In each permutation, each study unit is assigned a random assignment to one arm so that the number of units in each arm equals the number of units assigned to that arm in the study. The test statistic is calculated with the groups defined by the labels; each group consists of a mixture of births to

women who received Cervarix or Hepatitis A vaccine. We used an algorithm to calculate the permutation P-value:

1. We created subsets of pregnancies defined by number of days from estimated conception. We considered subsets of pregnancies conceived in windows defined by time between vaccination and conception with a 2 days resolution (-60, -58, -56, ... , -2, 0, 2, 4, ..., 56, 58, 60] from 60 days before to 60 days after estimated conception, and 7 days resolution more than 60 days away from estimated date of conception (..., -81, -74, -67; 67, 74, 81, ...). Only subsets with at least 10 pregnancies in each arm were considered.

2. The permutation test was based on K=10,000 permutation selections. For each permutation, we

1.1. Randomly assign N_1 pregnancies to group X and the remaining $N_2 = N - N_1$ to group Y, where and N_2 are the numbers of pregnancies in the two arms in the analysis.

1.2. Calculate the one-sided P-values for the test of association between study arm and risk of miscarriage in the pregnancies in group X with the pregnancies of group Y with estimated date of conception within each window using one of two test-statistics:

1.2.1. A robust Wald-type comparison of two proportions from

McCullagh and Nelder.[3]

$$z = \frac{\log(\hat{\Theta}R)}{\text{s.e.}[\log(\hat{\Theta}R)]}, \text{ where log is natural log,}$$

$$\hat{\Theta}R = \frac{\left(a+\frac{1}{2}\right)\left(d+\frac{1}{2}\right)}{\left(b+\frac{1}{2}\right)\left(c+\frac{1}{2}\right)} \text{ and}$$

$$\text{s.e.}[\log(\hat{\Theta}R)] = \left[\frac{1}{\left(a+\frac{1}{2}\right)} + \frac{1}{\left(b+\frac{1}{2}\right)} + \frac{1}{\left(c+\frac{1}{2}\right)} + \frac{1}{\left(d+\frac{1}{2}\right)}\right]^{\frac{1}{2}}$$

where a, b, c and d are the numbers in the two-by-two table of study arm by pregnancy outcome (miscarriage vs. no miscarriage).

or

1.2.2. A conditional chi-squared score test based on counts with expected value conditional on total count. The chi-square based on counts has form $\chi^2 = \sum_{i=0,1} \frac{(O_i - E_i)^2}{E_i}$, where O_i is the number of events in group i and $E_i = \frac{D_i}{D_0 + D_1}(O_0 + O_1)$, is the expected proportion of the total events in group i under the null hypothesis, which is the product of person time in group i and the overall rate combining the two arms. The contribution to D_i from a woman randomized to arm i can is the total number of days during which the woman is at risk and in the stratum.

1.3. The minimum P-value over all the windows is the k^{th} observation in the null permutation distribution

Calculate the permutation P-value as 1/K times the sum of 1 plus the number of observations in the null permutation distribution (from step 1) that are below the observed test statistic of the permutation test. For example, if the minimum P-value observed among the correlated tests is less than the lowest of the K permutation test statistics, the permutation P-value is 1/K; if the minimum P-value lies between the lowest and next to lowest of the K permutation test statistics, the permutation P-value is 2/K.

1 Chatterjee N, Kalaylioglu Z, Moslehi R, Peters U, Wacholder S. Powerful multilocus tests of genetic association in the presence of gene-gene and gene-environment interactions. *Am J Hum Genet* 2006;79:1002-16.

2 Chapman J, Clayton D. Detecting association using epistatic information. *Genet Epidemiol* 2007;31:894-909.

3 McCullough P, Nelder JA. Generalized linear models. Chapman and Hall, 1989.