SUPPLEMENTARY MATERIALS

- Autonomic Dysfunction Risk Scores, pages 2-4.
- Supplementary Figure S1 Label, page 5.
- Supplementary Figure S2 Label, page 5.
- Supplementary Figure S1: Diagram of follow-up according to vaccination status.
- Supplementary Figure S2: Distribution of follow-up according to vaccination status.
Autonomic Dysfunction Risk Scores

Increased healthcare utilization may be associated with increased probability of being diagnosed with a syndrome. Using history of hospital contacts, we estimated the risk of being diagnosed with a syndrome for all individuals in the study. This allowed us to evaluate the association between the quadrivalent human papillomavirus vaccine and selected syndromes with autonomic dysfunction in groups of individuals with different baseline probabilities of being diagnosed. We obtained information on all hospital contacts in the five years prior to study entry for all study participants classified according to main chapters of the ICD-10 classification and occurrence (0, 1, 2 or more contacts). We then used logistic regression with lasso regularization to model the probability of being diagnosed with either chronic fatigue syndrome, complex regional pain syndrome or postural orthostatic tachycardia syndrome during the study period given previous hospital contacts. The resulting probabilities (autonomic dysfunction risk scores) were categorized according to deciles in three risk groupings ('low risk': 1st, 2nd, 3rd and 4th deciles, ‘moderate risk’: 5th, 6th and 7th deciles, ‘high risk’: 8th, 9th, and 10th deciles).

When constructing the autonomic dysfunction risk score we considered the entire cohort consisting of 1,375,737 girls and women who were part of the study at some point and had no case diagnosis prior to entering the study. For each person all diagnoses from the 5 years immediately preceding study entry were used to create the risk score. The date used here was the date of admission for non-ambulatory visits. The diagnoses were all encoded using ICD-10 codes.

As there were relatively few cases in the cohort a regression on all ICD-10 codes would likely be infeasible due to overfitting. Instead ICD-10 codes were grouped by which of the 21 ICD-10 chapters they belong to. A 22nd “chapter” was added for special codes used in Denmark that fall outside the 21 chapters. For each individual we would count whether they had 0, 1, or 2+ diagnoses for each of the chapters. All diagnoses except the case diagnoses were included in this. No attempts were made to resolve cases where the same diagnosis might have been entered twice. We then used one hot encoding to transform this to 44 variables that were used in a 10-fold cross-validated logistic lasso regression using the glmnet R package. To find the optimal value of lambda we used the default binomial deviance for the logistic regression. Coefficients for the value of lambda minimizing the binomial deviance were then used to make a predictive model that was applied to the entire cohort.

The predicted values were used to split the dataset into three groups. Most of the 40% with the lowest predicted risk of becoming a case had the exact same predicted risk due to not having any diagnoses in the 5 years preceding study entry. The dataset was then split into a “low” risk group consisting of the 40% with the
lowest predicted risks, a “medium” risk group consisting of the group with predicted risks between the 40th and 70th percentile, and a “high” risk group containing the 30% with the highest predicted risks.

The following table lists the 10 deciles and the interval of estimated risk of becoming a case as determined by the lowest and highest risk among any member of the decile. Since large groups of people with the same diagnostic history appear in the data (in particular a very large group with no diagnoses), we have sorted the dataset according to the disease risk score and the unique individual identifier for tiebreakers when disease risk scores were identical before splitting into deciles.

| Decile | Disease risk range (absolute in %) | Observed cases | Expected cases |
|--------|-----------------------------------|----------------|----------------|
| 1st    | 0.00789-0.0349                    | 46             | 45.99          |
| 2nd    | 0.0349-0.0349                     | 49             | 48.02          |
| 3rd    | 0.0349-0.0349                     | 49             | 48.02          |
| 4th    | 0.0349-0.0367                     | 45             | 48.09          |
| 5th    | 0.0367-0.0425                     | 34             | 56.01          |
| 6th    | 0.0425-0.0515                     | 72             | 65.70          |
| 7th    | 0.0515-0.0612                     | 71             | 74.03          |
| 8th    | 0.0612-0.0915                     | 97             | 99.02          |
| 9th    | 0.0915-0.117                      | 123            | 133.63         |
| 10th   | 0.117-2.13                        | 283            | 251.42         |

The following table presents the adjusted rate ratios (95% CIs) for the association between HPV vaccination and any syndrome in strata of the risk score.

| Autonomic Dysfunction Risk Score | Rate Ratio | Cases | Person-years |
|----------------------------------|------------|-------|--------------|
| 'Low risk', deciles 1-4          | 1.17 (0.81-1.67) | 21    | 192          |
| 'Moderate risk', deciles 5-7     | 0.65 (0.33-1.29)  | 10    | 126          |
| ‘High risk’, deciles 8-10        | 0.91 (0.55-1.50)  | 41    | 327          |
The area under the curve for the model was 0.679 and the receiver operating curve is shown in the Figure.

Figure – Receiver operating characteristic (ROC) curve for the autonomic dysfunction risk score. The predictions were made using the model from the logistic lasso regression. The area under the curve (AUC) is 0.679.
Supplementary Figure S1 Label:

Diagram of follow-up according to vaccination status.

Assignment of follow-up to the referent period and the 365-day risk period in a study participant receiving 3 doses of vaccine with 2 and 4 month intervals between the doses respectively.

Supplementary Figure S2 Label:

Distributions of follow-up according to vaccination status.

Distribution of follow-up according to vaccination status in 869 cases of syndromes with autonomic dysfunction among all Danish born females 10 to 44 years of age during 2007 to 2016.
At-risk status

- Referent period
- 365-day risk period
