

## **SUPPLEMENTARY DATA**

Supplemental information and data regarding the manuscript:

“Prospective study of HPV16 viral load and risk of in situ and invasive squamous cervical cancer” by Karin Sundström, Alexander Ploner, Lisen Arnheim Dahlström, Juni Palmgren, Joakim Dillner, Hans-Olov Adami, Nathalie Ylitalo and Pär Sparén.

This document contains more detailed explanations and motivations for the statistical methodology that was employed in this manuscript (Section 1, Supplemental Figure 1-2 and Supplemental Tables 1-4). It also details results from several sensitivity analyses regarding histological re-classification and period of analysis (Section 2, Supplemental Figure 3-4 and Supplemental Table 5-8).

Table of contents:

1. Statistical Analysis
  - Modeling and adjusting viral load over time
  - Imputation of HPV profiles for the study subjects
  - Classification of HPV16 viral load into tertiles
  - Coefficients of variation for betaglobin and viral load assays
2. Supplementary Results
  - Odds ratios for separated tertiles of HPV16 viral load in CIS and SCC
  - Sensitivity analyses using a) confirmed cases only and b) pre-1995 cases only

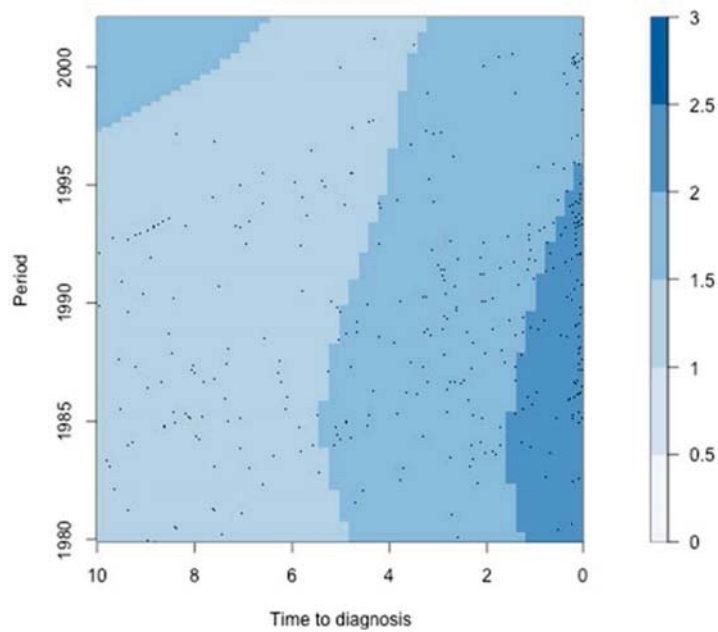
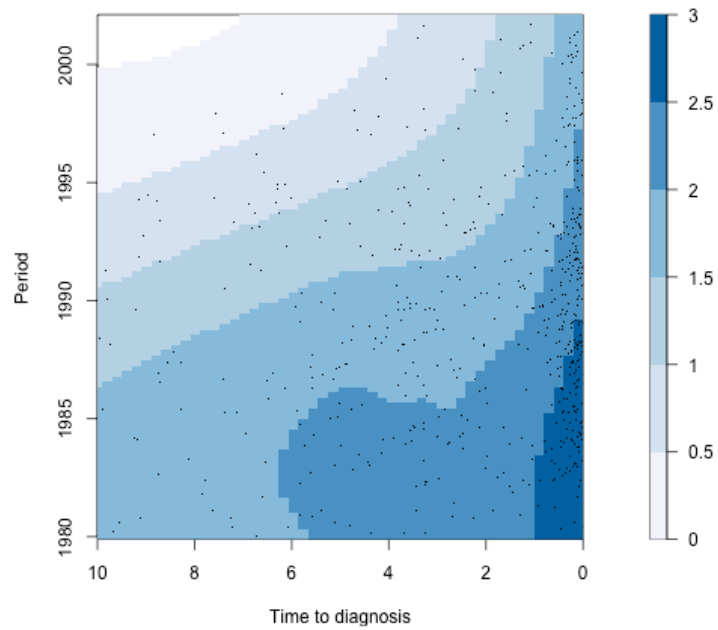
## **1. Statistical Analysis**

### **Modeling and adjusting viral load over time**

During initial data inspection, we found that as expected, viral load levels increased with proximity to diagnosis for both CIS and SCC patients. However, we also found a strong calendar period effect on viral loads that was only apparent in CIS patients, but not in SCC patients (Supp. Fig 1 and 2, next page). Correction for assay drift did not affect these findings.

Supp. Table 1 summarizes the models for both CIS and SCC. Model 1 only contains time to diagnosis, Models 2-4 add each of the other time scales on their own, and Models 5-10 add the different pairwise combinations.

Based on these results, we decided to adjust viral loads in the CIS study for period, and to make no adjustment for SCC. The corresponding models are shown in Supp. Tables 2 and 3.



**Supplemental Figure 1 and 2:** Smoothed viral loads (on log<sub>10</sub> scale) for HPV16 positive women with CIS (Figure 1, top) and SCC (Figure 2, bottom). Vertical axis: date of Pap smear, horizontal axis: time to diagnosis for pap smear; each dot represents one smear. The viral loads indicated by shading are based on a loess smooth of the observed data.

		CIS				SCC		
	Model	Scale	F	p	AIC	F	p	AIC
1	T2D	T2D	14.2	4.3e-11	1775	11.6	6.1e-09	1181
2	T2D+AGE	AGE	2.4	0.049	1774	0.7	0.604	1187
3	T2D+PERIOD	PER	10.0	8.2e-08	1745	1.9	0.092	1182
4	T2D+COHORT	COH	2.2	0.072	1775	1.4	0.210	1185
5	T2D+AGE+PERIOD	AGE	2.7	0.029	1744	0.7	0.576	1186
		PER	9.6	1.6e-07		2.1	0.068	
6	T2D+AGE+COHORT	AGE	2.6	0.036	1764	0.8	0.553	1183
		COH	4.4	0.002		2.5	0.016	
7	T2D+PERIOD+AGE	PER	10.2	5.9e-08	1744	1.9	0.090	1186
		AGE	2.2	0.073		0.9	0.455	
8	T2D+PERIOD+COHORT	PER	10.0	7.9e-08	1747	1.9	0.092	1190
		COH	1.4	0.25		0.7	0.640	
9	T2D+COHORT+AGE	COH	2.3	0.054	1764	1.5	0.184	1183
		AGE	4.7	0.001		2.6	0.037	
10	T2D+COHORT+PERIOD	COH	2.4	0.051	1747	1.4	0.209	1190
		PER	9.0	4.9e-07		1.0	0.417	

**Supp. Table 1:** Effect of different time scales on log10 viral load in HPV16 positive women diagnosed with CIS and SCC. For all models, sequential F-statistics and their respective p-values are reported, together with the AIC for the model.

Variable	Level	Estimate	Std.Error	t-Statistic	p-value
Constant	-	2.98	0.22	13.49	2.20e-36
	0-3m	0.00	-	-	-
T2D	3m-2y	-0.40	0.10	-3.94	9.00e-05
	2-4y	-0.71	0.12	-5.76	1.36e-08
	4-8y	-0.91	0.13	-6.76	3.25e-11
	8-16y	-1.46	0.19	-7.88	1.50e-14
	1975-1980	0.00	-	-	-
Period	1980-1985	-0.12	0.21	-0.61	5.45e-01
	1985-1990	-0.42	0.21	-1.98	4.77e-02
	1990-1995	-0.96	0.22	-4.34	1.70e-05
	1995-2000	-0.79	0.25	-3.19	1.52e-03

**Supp. Table 2:** Parameter estimates with standard errors, t-statistics and p-values for Model 3 (CIS) in Supp. Table 1.

Variable	Level	Estimate	Std.Error	t-Statistic	p-value
Constant	-	2.20	0.12	18.46	6.76e-56
	0-3m	0.00	-	-	-
T2D	3m-2y	-0.23	0.14	-1.60	1.11e-01
	2-4y	-0.72	0.15	-4.74	3.01e-06
	4-8y	-0.96	0.16	-5.84	1.03e-08
	8-16y	-0.94	0.17	-5.43	9.57e-08

**Supp. Table 3:** Parameter estimates with standard errors, t-statistics and p-values for Model 1 (SCC) in Supp. Table 1.

## Imputation of HPV profiles for the study subjects

For each subject and each year, we imputed HPV16 measurements from all observations within a fixed window ( $\pm 5$  years), where observations closer in time received larger weights.

Formally, this entailed that for any individual with observed HPV16 status  $Y_1, \dots, Y_k$  and corresponding viral loads  $L_1, \dots, L_k$  at time points  $t_1, \dots, t_k$ , we estimated the probability of being HPV16 positive at any time point  $t$  as

$$p(t) = \frac{\sum_{d_i \leq w} Y_i / d_i}{\sum_{d_i \leq w} 1 / d_i}$$

where  $d_i = |t - t_i|$  is the absolute distance from the target time point to the observational time points, and  $w$  is the width of the time window. For an empty window, the corresponding probability and measurements were set to missing.

For a single imputation run, we used  $p(t)$  to impute  $Y(t)$  at yearly intervals up to 15 years prior to diagnosis or up to first smear. For an imputed HPV16-negative smear, the corresponding viral load  $L(t)$  was set to zero; for an imputed positive smear, the corresponding viral load was selected randomly from the observed loads  $L_j$  within the current time window, with selection probability proportional to distance in time:

$$p_j(t) \sim 1/d_j$$

with a large finite value substituted for  $d_j=0$ .

**Classification of HPV16 viral load into tertiles**

	<u>Minimum</u>	<u>Threshold</u>	<u>Maximum</u>
<u>CIN3</u>	<u>-1.30</u>	<u>1.95</u>	<u>5.44</u>
<u>SCC</u>	<u>-2.14</u>	<u>2.22</u>	<u>4.82</u>

**Suppl. Table 4: Range and threshold of log<sub>10</sub>(HPV16 viral load) in Table 2 and Figure 3, for CIN3 (threshold at 33% percentile) and SCC (threshold at 67% percentile).**

**Coefficients of variation (CV) for betaglobin and viral load assays**

	<u>Within-test CV</u>	<u>Between-test CV</u>
<u>Betaglobin</u>	<u>1,14%</u>	<u>3,67%</u>
<u>HPV16 viral load</u>	<u>1,39%</u>	<u>4,47%</u>

## 2. Supplementary Results

### Odds ratios/relative risks for individual tertiles of HPV16 viral load in cervical cancer in situ (CIS)

<b>Year before diagnosis</b>	<b>HPV16 negative (reference)</b>	<b>Low VL (OR, 95% CI)</b>	<b>Medium VL (OR, 95% CI)</b>	<b>High VL (OR, 95% CI)</b>
1	1.0	11 (3 - 39)	32 (7 - 134)	40 (7.8 - 202)
2	1.0	10 (3 - 31)	37 (8 - 171)	35 (6.6 - 182)
3	1.0	9 (3 - 23)	17 (6 - 48)	36 (5.9-215)
4	1.0	10 (4 - 29)	20 (6 - 69)	29 (4.8-182)
5	1.0	8 (3 - 20)	18 (6 - 58)	22 (3.8-123)
6	1.0	8 (3 - 17)	12 (4 - 36)	15 (2.4 - 96)
7	1.0	8 (3 - 18)	9 (3 - 27)	13 (1.9 - 90)
8	1.0	11 (4 - 29)	7 (2 - 22)	9 (1.4 - 55)
9	1.0	8 (3 - 22)	6 (2 - 22)	8 (1.1 - 63)
10	1.0	6 (3 - 16)	7 (2 - 24)	6 (0.7 - 48)

**Suppl. Table 5.** Imputed odds ratios for cervical cancer in situ (CIS) according to tertile of HPV16 viral load compared to HPV16-negative women over 10 years before diagnosis.



**Odds ratios/relative risks for individual tertiles of HPV16 viral load  
in invasive squamous cervical cancer (SCC)**

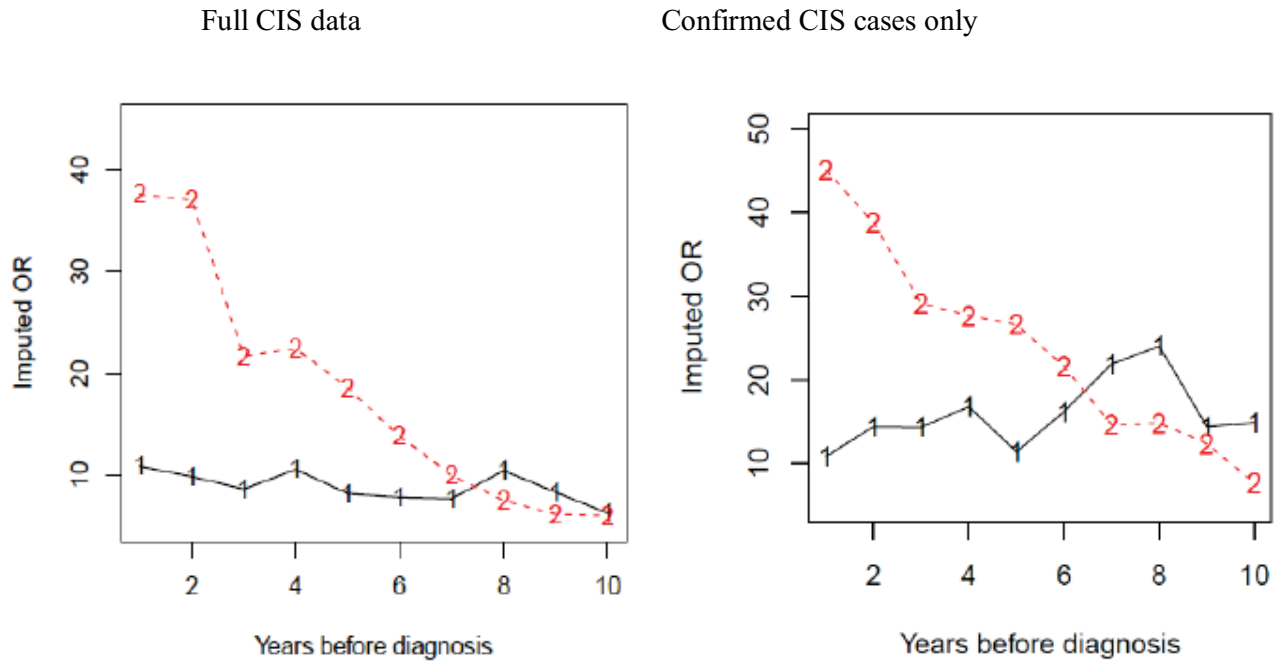
<b>Year before diagnosis</b>	<b>HPV16 negative (reference)</b>	<b>Low VL (OR, 95% CI)</b>	<b>Medium VL (OR, 95% CI)</b>	<b>High VL (OR, 95% CI)</b>
1	1.0	18 (4 - 93)	15 (3 - 69)	63 (8 - 522)
2	1.0	18 (4 - 79)	29 (4 - 211)	58 (7 - 497)
3	1.0	22 (5 - 103)	40 (5 - 324)	51 (6 - 441)
4	1.0	24 (4 - 134)	37 (5 - 287)	32 (4 - 225)
5	1.0	19 (4 - 95)	40 (5 - 304)	24 (4 - 136)
6	1.0	12 (3 - 41)	37 (5 - 272)	21 (3 - 146)
7	1.0	10 (3 - 30)	27 (5 - 159)	24 (3 - 195)
8	1.0	13 (3 - 45)	17 (3 - 88)	11 (2 - 62)
9	1.0	10 (3 - 36)	18 (3 - 105)	7 (2 - 30)
10	1.0	7 (2 - 20)	17 (3 - 95)	12 (2 - 74)

**Suppl. Table 6.** Imputed odds ratios for invasive squamous cell carcinoma (SCC) according to tertile of HPV16 viral load compared to HPV16-negative women over 10 years before diagnosis.

### **Sensitivity analyses**

- A. Full CIS data versus histologically confirmed CIS (n=304 cases)
- B. Full SCC data versus histologically confirmed SCC (n=305 cases)
- C. CIS cases pre-1995 only
- D. SCC cases pre-1995 only

**A. Sensitivity analyses using full CIS data vs confirmed CIS cases only**



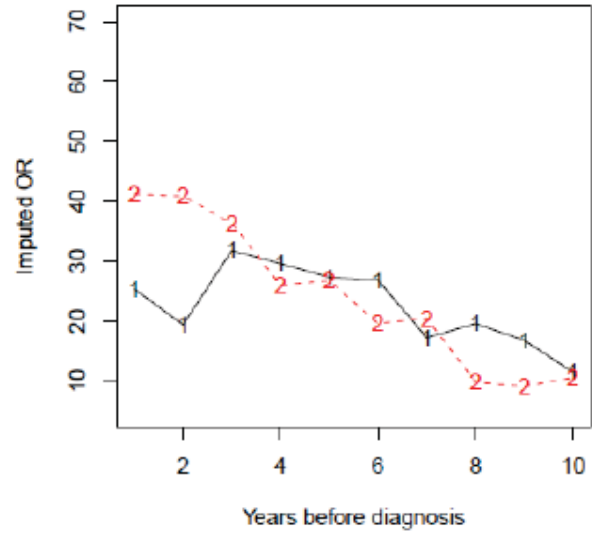
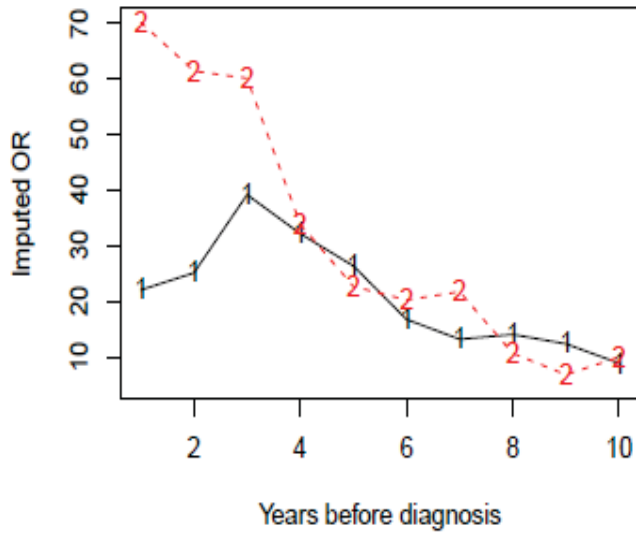
Year	HPV16 negative (reference)	OR for low viral load	OR for medium/high viral load		OR for low viral load (confirmed cases)	OR for medium/high viral load (conf cases)
1	1.0	11 (3 - 36)	38 (11 - 130)		11 (2 - 61)	46 (6 - 323)
2	1.0	10 (3 - 29)	37 (11 - 127)		13 (2 - 74)	41 (6 - 272)
3	1.0	9 (3 - 22)	22 (8 - 59)		14 (3 - 74)	27 (5 - 144)
4	1.0	11 (4 - 29)	22 (8 - 65)		19 (3 - 113)	30 (5 - 188)
5	1.0	8 (3 - 20)	19 (7 - 48)		12 (3 - 50)	24 (5 - 115)
6	1.0	8 (3 - 18)	14 (5 - 37)		16 (4 - 70)	23 (3 - 158)
7	1.0	8 (3 - 17)	10 (4 - 24)		22 (4 - 134)	15 (2 - 89)
8	1.0	10 (4 - 29)	8 (3 - 19)		22 (4 - 135)	14 (2 - 103)
9	1.0	8 (3 - 21)	6 (2 - 17)		15 (3 - 76)	12 (2 - 92)
10	1.0	6 (3 - 16)	6 (2 - 17)		15 (3 - 82)	8 (1 - 48)

**Suppl. Table 7.** Imputed odds ratios for risk of cervical cancer in situ (CIS) according to level of HPV16 viral load (low tertile versus medium/high tertile) compared to HPV16-negative subjects over 10 years before diagnosis.

**B. Sensitivity analysis using full SCC data vs. confirmed SCC cases only**

Full SCC data

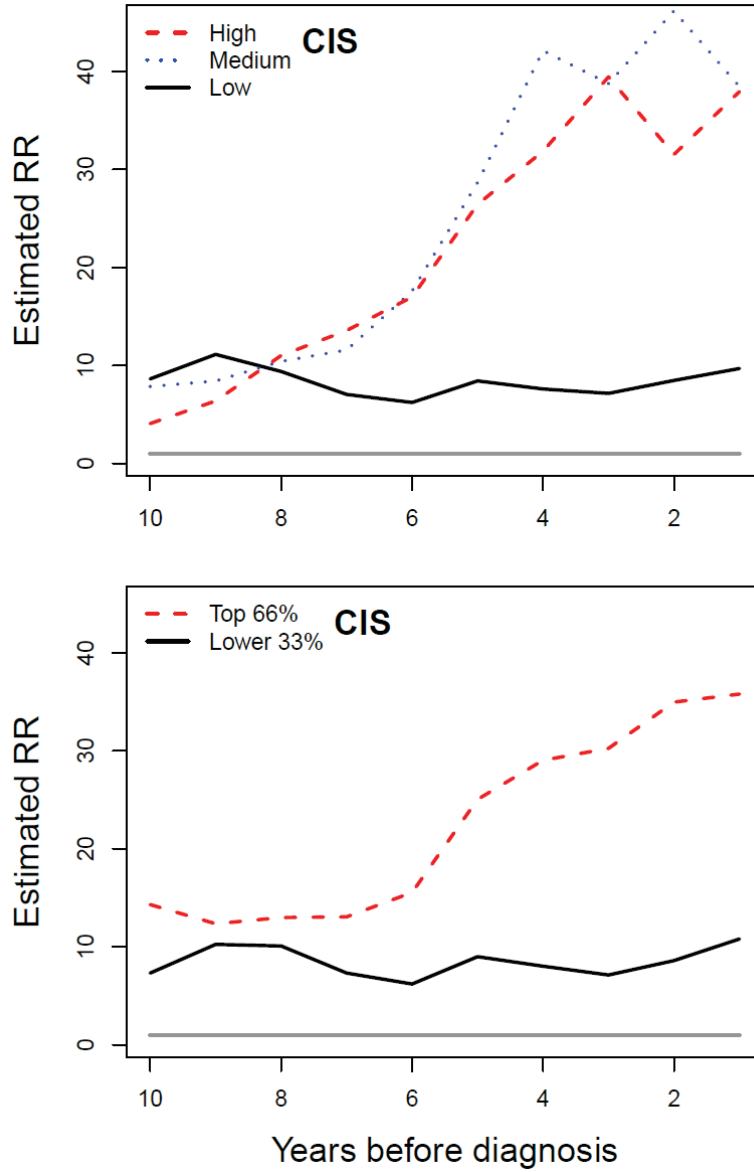
Confirmed SCC cases only



Year	HPV16 negative (reference)	OR for low/medium viral load	OR for high viral load		OR for low/medium viral load (confirmed cases)	OR for high viral load (confirmed cases)
1	1.0	22 (6 - 81)	70 (8 - 628)		25 (5 - 135)	41 (5 - 359)
2	1.0	25 (6 - 104)	61 (7 - 547)		19 (4 - 86)	41 (5 - 348)
3	1.0	39 (8 - 202)	60 (6 - 580)		32 (6 - 180)	36 (4 - 332)
4	1.0	32 (7 - 159)	34 (4 - 271)		30 (5 - 176)	26 (3 - 195)
5	1.0	26 (6 - 112)	23 (4 - 119)		27 (5 - 141)	27 (4 - 197)
6	1.0	17 (6 - 50)	20 (3 - 139)		27 (5 - 152)	20 (2 - 162)
7	1.0	13 (5 - 35)	22 (3 - 154)		17 (5 - 61)	20 (2 - 177)
8	1.0	14 (5 - 38)	11 (2 - 54)		20 (5 - 75)	10 (1 - 66)
9	1.0	12 (5 - 33)	7 (2 - 27)		17 (4 - 63)	9 (1 - 61)
10	1.0	9 (3 - 23)	10 (2 - 57)		11 (3 - 38)	11 (2 - 74)

**Suppl. Table 8.** Imputed odds ratios for risk of invasive squamous cervical cancer (SCC) according to level of HPV16 viral load (low/medium tertiles versus highest tertile) compared to HPV16-negative subjects over 10 years before diagnosis.

C. Suppl. Figure 3. Sensitivity analyses using data pre-1995 only for cancer in situ (CIS)



**D. Suppl. Figure 4. Sensitivity analyses using data pre-1995 only for invasive squamous cell carcinoma (SCC)**

