High-risk human papillomavirus status and prognosis in invasive cervical cancer: a nationwide cohort study. Dataset 1

SND-ID: snd1064-1. **Version**: 1.0. **DOI**: https://doi.org/10.5878/rtxk-6790

Citation

Sparén, P. (2019) High-risk human papillomavirus status and prognosis in invasive cervical cancer: a nationwide cohort study. Dataset 1 (Version 1.0) [Data set]. Karolinska Institutet. Available at: https://doi.org/10.5878/rtxk-6790

Creator/Principal investigator(s)

Pär Sparén - Karolinska Institutet, Department of Medical Epidemiology and Biostatistics

Research principal

Karolinska Institutet - Department of Medical Epidemiology and Biostatistics

Description

High-risk human papillomavirus (hrHPV) infection is established as the major cause of invasive cervical cancer (ICC). However, whether hrHPV status in the tumor is associated with subsequent prognosis of ICC is controversial. We aim to evaluate the association between tumor hrHPV status and ICC prognosis using national registers and comprehensive human papillomavirus (HPV) genotyping.

In this nationwide population-based cohort study, we identified all ICC diagnosed in Sweden during the years 2002–2011 (4,254 confirmed cases), requested all archival formalin-fixed paraffin-embedded blocks, and performed HPV genotyping. Twenty out of 25 pathology biobanks agreed to the study, yielding a total of 2,845 confirmed cases with valid HPV results. Cases were prospectively followed up from date of cancer diagnosis to 31 December 2015, migration from Sweden, or death, whichever occurred first. The main exposure was tumor hrHPV status classified as hrHPV-positive and hrHPV-negative. The primary outcome was all-cause mortality by 31 December 2015. Five-year relative survival ratios (RSRs) were calculated, and excess hazard ratios (EHRs) with 95% confidence intervals (CIs) were estimated using Poisson regression, adjusting for education, time since cancer diagnosis, and clinical factors including age at cancer diagnosis and International Federation of Gynecology and Obstetrics (FIGO) stage.

Of the 2,845 included cases, hrHPV was detected in 2,293 (80.6%), and we observed 1,131 (39.8%) deaths during an average of 6.2 years follow-up. The majority of ICC cases were diagnosed at age 30–59 years (57.5%) and classified as stage IB (40.7%). hrHPV positivity was significantly associated with screen-detected tumors, young age, high education level, and early stage at diagnosis ($p < 0.001$). The 5-year RSR compared to the general female population was 0.74 (95% CI 0.72–0.76) for hrHPV-positive cases and 0.54 (95% CI 0.50–0.59) for hrHPV-negative cases, yielding a crude EHR of 0.45 (95% CI 0.38–0.52) and an adjusted EHR of 0.61 (95% CI 0.52–0.71). Risk of all-cause mortality as measured by EHR was consistently and statistically significantly lower for cases with hrHPV-positive tumors for each age group above 29 years and each FIGO stage above IA. The difference in prognosis by hrHPV status was highly robust, regardless of the clinical, histological, and educational characteristics of the cases. The main limitation was that, except for education, we were not able to adjust for lifestyle factors or other unmeasured confounders.

In conclusion, women with hrHPV-positive cervical tumors had a substantially better prognosis than women with hrHPV-negative tumors. hrHPV appears to be a biomarker for better prognosis in cervical cancer independent of age, FIGO stage, and histological type, extending information from already established prognostic factors. The underlying biological mechanisms relating lack of detectable tumor hrHPV to considerably worse prognosis are not known and should be further investigated.

Purpose:

To compile a comprehensive survival and HPV genotyping data and provide a large-scale population-based evaluation of the association between tumor high risk HPV status and prognosis of invasive cervical cancer.

This dataset (ccHPV_RelativeSurvival.dta) comprises 2845 invasive cervical cancer (ICC) cases diagnosed in Sweden during the years 2002-2011, and had valid human papillomavirus (HPV) results assessed from the formalin-fixed, paraffin-embedded (FFPE) blocks.

In order to control the risk of incidental disclosure of personal information, the data available here has been anonymized in the following manner:
• The date of diagnosis has been moved to 2008-07-01 for all subjects.
• Follow-up time has been censored at five years after diagnosis.
• Age at diagnosis and follow-up time after diagnosis have been microaggregated in groups of five subjects (using function microaggregation in R package sdcMicro 2.5.9, available from https://cran.r-project.org/package=sdcMicro)

Analysis of the anonymized data replicates the results presented in main part of the study (Figures 2 & 3, Tables 1-3) with only minor numerical differences, with the following exceptions:
• In Figure 2, relative survival can only be calculated up to five years after diagnosis.
• In Table 1, the number of person years and the mean follow-up time differ considerably due to censoring; the distribution of subjects between age groups varies somewhat due to microaggregation.
• In Figure 3, the excess hazard ratios for age groups 30-44 and 45-59 in Panel A shift noticeably, but without affecting the overall message (comparable reduced risk across all age strata).

The dataset includes 12 variables, eight of which are necessary for the analysis (core variables) and four of which are included for administrative purposes and convenience of coding the analysis (extra variables).
Core variables:
• dx_date: Date of diagnosis
• age: Age (in years) at diagnosis
• x_stage_group: International Federation of Gynecology and Obstetrics (FIGO) stage of tumor, IA; IB; II and III+
• edu_cat: Education (categorical, three levels): 1=low (less than high school); 2=middle (high school); 3=high (university exam and above); 99=missing
• exit_new: End of follow-up (date)
• censor_new: Censoring status: 1=death; 2=censored due to migration, loss of follow-up or end of study
• final_type: Histological type of tumor: SCC=squamous cell carcinoma; AC=adenocarcinoma.
• hr_hpv: High-risk HPV status of tumor (main exposure, binary): 0=hrHPV negative; 1= hrHPV positive

Extra variables:

• entry: Entry date (copy of diagnosis date)
• sex: Gender (all female, for linking to standard population mortality file): 2=female.
• dx_year: Year of diagnosis (for linking to standard population mortality file)

Data contains personal data

No

Language

English

Unit of analysis

Individual

Population

Women diagnosed as cervical cancer during year 2002-2011 in Sweden

Time Method

Longitudinal

Sampling procedure

Other

Biobank is connected to the study

Yes

Variables

11

Number of individuals/objects

2845

Data format / data structure

Numeric

Data collection 1

- Mode of collection: Aggregation
- Time period(s) for data collection: 2002-01-01 – 2015-12-31
- Data collector: Statistics Sweden
- Source of the data: Population group, Biological samples

Geographic spread

Geographic location: Sweden

Responsible department/unit

Department of Medical Epidemiology and Biostatistics

Funding 1

- Funding agency: Swedish Foundation for Strategic Research, SSF
- Funding agency's reference number: KF10-0046; RB 13-0011

Funding 2

- Funding agency: Swedish Research Council
- Funding agency's reference number: 2014-03732; 2017-02346

Funding 3

- Funding agency: The Swedish Cancer Society
- Funding agency's reference number: 110569; 140665

Ethics Review

Stockholm - Ref. 2011/1026-31/4, 2017/1028/32, 02- 189 556, 2012/1028/32, 2011/921-32

Research area

Medical and health sciences (Standard för svensk indelning av forskningsämnen 2011)

Health (CESSDA Topic Classification)

Keywords

Prognosis, Population mortality, Human papillomavirus, Cervical cancer

Publications

Lei J, Ploner A, Lagheden C, Eklund C, Nordqvist Kleppe S, Andrae B, Elfström KM, Dillner J, Sparén P, Sundström K (2018). High-risk human papillomavirus status and prognosis in invasive cervical cancer: A nationwide cohort study. PLoS Med 15(10): e1002666. DOI: 10.1371/journal.pmed.1002666
DOI: https://doi.org/10.1371/j

Accessibility level

Access to data through SND
Access to data is restricted

Use of data

Things to consider when using data shared through SND

Versions

Version 1.0. 2019-10-21

Related research data in SND's catalogue

High-risk human papillomavirus status and prognosis in invasive cervical cancer: a nationwide cohort study. Dataset 2

Download metadata

DataCite

DDI 2.5

DDI 3.3

DCAT-AP-SE 2.0

JSON-LD

PDF

Citation (CSL)

Published: 2019-10-21
Last updated: 2025-01-21