[CANCER RESEARCH 60, 6027–6032, November 1, 2000]

A Prospective Study Showing Long-Term Infection with Human Papillomavirus 16 before the Development of Cervical Carcinoma *in Situ*[1]

Nathalie Ylitalo,[2] Agnetha Josefsson, Mads Melbye, Per Sörensen, Morten Frisch, Per Kragh Andersen, Pär Sparén, Margit Gustafsson, Patrik Magnusson, Jan Pontén,[3] Ulf Gyllensten,[4] and Hans-Olov Adami[4]

Department of Medical Epidemiology, Karolinska Institutet, 171 77 Stockholm, Sweden [N. Y., P. Sp., H-O. A.]; Department of Epidemiology Research, Statens Serum Institut, 2300 Copenhagen, Denmark [M. M., P. Sö., M. F., P. K. A.]; Department of Genetics and Pathology, Uppsala University, 751 23 Uppsala, Sweden [A. J., M. G., P. M., J. P., U. G.]; Department of Epidemiology, Harvard University, Boston, Massachusetts 02115 [H-O. A.]

ABSTRACT

Human papillomavirus 16 (HPV16) is a predominant cause of cervical neoplasia. However, no population-based study with long-term follow-up has clarified the temporal relationship between HPV16 infection and occurrence of carcinoma *in situ*, or the importance of recurrent or persistent infection. This nested case-control study was carried out in a population-based cohort of women participating in cytological screening whose initial smear, taken in 1969–1995, was normal. During up to 26 years of follow-up, carcinoma *in situ* was diagnosed in 484 eligible women. Archival smears from these women were compared with smears from 619 individually matched controls. After DNA extraction, a highly sensitive PCR system was used to detect HPV16. Among case women, the prevalence of HPV16 positivity was 56% at the time of diagnosis. The relative risk of cervical carcinoma *in situ* increased from 3.6 (95% confidence interval, 1.2–11.0) 13 years before diagnosis to 11.1 (95% confidence interval, 5.5–22.2) 1 year before diagnosis. Having a positive smear at entry to the cohort increased risk >5-fold, whereas having persistent infection with HPV in two subsequent smears increased risk 30-fold. We estimated that among HPV16-positive women, the median incubation period from infection to carcinoma *in situ* was 7–12 years. We conclude that evidence of persistent and/or recurrent infection is associated with a drastically higher risk of cervical carcinoma *in situ* than occasional infection with HPV16.

INTRODUCTION

During the last 20 years, strong experimental and epidemiological evidence has linked infection with certain types of HPV[5] to the development of cervical cancer (1), the third most common cancer among women worldwide (2). This association, established consistently all around the world, has particularly been shown for HPV16, which accounts for at least 50% of all cervical HPV infections (3).

Although viral persistence seems to be crucial for the development of cervical neoplasia (4–7), the mechanisms for viral persistence remain largely unknown (8, 9). Because most earlier epidemiological studies had a case-control design (1, 9), they determined HPV status only at the time of diagnosis of the cases, whereas HPV persistence can be studied only in prospective studies. Few such prospective studies have been performed, and they have generally been small, have had short follow-up, have been carried out among highly selected groups of women, and have studied dysplasia rather than invasive carcinoma of the cervix or its immediate precursor, carcinoma *in situ* (1, 4–6, 10, 11). Although results from these studies have been confirmatory, questions remain about persistence, recurrence, and the incubation period between infection and neoplastic development.

In the present study, we took advantage of the favorable conditions for follow-up studies in Sweden to examine the temporal association between HPV16 infection and cervical carcinoma *in situ*. This was done by analyzing archival smears, taken repeatedly up to 26 years before diagnosis of cervical carcinoma *in situ* from cases and matched controls, nested in a population-based cohort of women participating in cytological screening in Sweden.

SUBJECTS AND METHODS

Subjects. A cytological screening program covering all women 30–49 years of age began in Uppsala County in Sweden in 1967 (12). A total of 732,287 smears from 146,889 women were stored at the University Hospital, and all information was computerized from 1969 through 1995. Using this cytology register, we defined a cohort comprising all women with at least one registered smear, provided that they were (*a*) registered with a first normal smear (PAP = 1); (*b*) born in Sweden; (*c*) <50 years of age at entry into the cohort; and (*d*) eligible for an interview at study start (1 January, 1996; Ref. 13). The time of the first registered smear defined the entry into the cohort. A total of 105,760 women fulfilled the inclusion criteria for the cohort.

All incident cases ($n = 504$) of squamous cell cervical carcinoma *in situ* in the cohort were identified through computerized linkage between the study cohort and the virtually 100% complete National Cancer Registry from 1969 through 1995 (14). Notification to the registry is mandatory both for *in situ* and invasive cervical cancer (15). For each case, five controls, individually matched by date of first registered smear (±90 days) and by year of birth, were randomly selected from the study cohort. We did not match on number of smears because this variable is likely associated with sexual practices (*e.g.*, the need for contraceptives, or treatment for sexually transmitted diseases) and hence with the probability of HPV infection.

Eligible controls had no history of *in situ* or invasive cervical carcinoma or any known hysterectomy before the date of diagnosis for their corresponding matched case; information about hysterectomy was obtained during telephone interviews with ~85% of the participants (13). Among the controls, eight (2%) had undergone hysterectomy before the diagnosis of carcinoma *in situ* of their corresponding case and were subsequently excluded together with their matched case. Because we did not match on number of smears, some first controls, randomly selected from the set of five, had only one smear taken during the study period. In such instances, we included a second control, randomly chosen from the remaining matched controls originally selected, to increase statistical power.

Cytological and Histological Review. The first smears for all eligible cases and controls were reviewed by a skilled cytotechnician, blinded to the case-control status. Cases regarded as not having a normal first smear (PAP = 1) were excluded. Controls without a normal first smear were replaced by another randomly selected control.

To confirm the diagnosis of squamous cell carcinoma *in situ*, the histological specimens from all 504 incident cases in the cohort were reviewed, except for 54 where the specimens could not be found. After review, we excluded five cases (three with invasive adenocarcinoma and two with microinvasive squamous cell carcinoma) and their matched controls. Each case and her matched control(s) constituted a risk set. After cytological and histological review, the study group consisted of 495 risk sets, including 495 cases, 495 first controls, and 154 second controls.

Received 4/12/00; accepted 9/1/00.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported by a grant from the NIH, Bethesda, MD (NR 1 RO1 CA61197-01A3), by grants from the Swedish Cancer Society, and by the Danish National Science Foundation.

[2] To whom requests for reprints should be addressed, at Department of Medical Epidemiology, Karolinska Institutet, Box 281, S-171 77 Stockholm, Sweden. Phone: 46 8 728 6140; Fax: 46 8 314957; E-mail: Nathalie.Ylitalo@mep.ki.se.

[3] Deceased.

[4] These two investigators have contributed equally to this work.

[5] The abbreviation used is: HPV, human papillomavirus.

HPV Analyses. All smears from eligible cases and controls during the period from entry into the cohort until the date of the case's diagnosis of cervical carcinoma *in situ* were analyzed for HPV. To determine HPV status more accurately for each control at the time of carcinoma *in situ* diagnosis in their matched case, we also, when available, included the controls' first smear taken after the carcinoma *in situ* diagnosis of the cases. Before analyses, all smears were sorted according to risk set and coded. Thus, those who performed the analyses were blinded to the case-control status. All smears in a risk set were analyzed at the same time.

After DNA extraction by described methods (16, 17), we analyzed the smears for HPV with a quantitative PCR system, based on real-time detection of accumulated fluorescence (TaqMan), using the 5′-exonuclease assay with a nonextendible hybridization probe with fluorescent dye linked to both the 5′ and 3′ ends (18, 19). When the probe is intact, the emission from the 5′ dye is quenched by the 3′ dye. During the extension phase of the PCR, the Taq polymerase cleaves the hybridization probe, producing a fluorescent signal proportional to the amount of PCR product generated. The sensitivity of the PCR system for analyses of archival smears was largely dependent on the DNA quality and the potential presence of inhibitors during the PCR reaction. The addition of BSA to the PCR reaction removed the inhibition caused by the Pap stain, thus yielding a sensitivity comparable to high molecular weight DNA. The characteristics of this PCR-based assay for HPV typing have been described in detail elsewhere (17). To prevent contamination during DNA extraction and PCR analysis, we used dedicated rooms for these procedures. The PCR was prepared in a room with one area for clean reagents (without DNA) and another for adding DNA template to the PCR reagents. PCR amplifications were performed in a second room. Our single-tube nested system did not require opening the tubes between the PCR reactions, further minimizing the risk of cross-contamination.

All smears were analyzed for the presence of HPV16 by amplification of a 180-bp fragment of the E1 open-reading frame in the presence of an HPV16-specific fluorescent hybridization probe (17, 20). In addition, the amplification of a human gene fragment (294 bp) from the β-actin locus indicated whether HPV16-negative samples were falsely negative because of an insufficient amount of DNA for amplification. Hence, smears testing negative for both β-actin and HPV16 were excluded (325 smears for cases, 358 smears for controls), but smears negative for β-actin and positive for HPV16 were not (137 smears for cases, 28 smears for controls). Consequently, 11 risk sets were lost because either the case or both controls had no usable smears. For the remaining statistical analyses, we had 484 risk sets, including 484 cases, 484 first controls, and 135 second controls.

Statistical Analyses. The association between HPV16 and time before diagnosis was considered separately among cases and controls. For each case, we used the longitudinal HPV data to generate response variables indicating HPV16 status in different time intervals before diagnosis: 1 if one or more HPV16-positive smears; 0 if only HPV16-negative smears; and missing if no smears in a certain time interval. The probability of being HPV16-positive was then estimated in a model with HPV16 status in the different time intervals as dependent variables. Because of the possible correlation between responses from the same individual, the probability was estimated using generalized estimating equations (GEE method) with the autoregressive (AR-1) correlation structure in a logistic regression model. Tests for linearity of the effect of time were performed using the fact that estimates of time effects are asymptotically normally distributed (21). Among controls, we considered the relationship between prevalence of HPV16 and age at the time of smear. In each age group ($<$20, 20–24, 25–29, 30–34, 35–39, $\geq$40 years), we generated response variables indicating HPV16 status, using the same definition as described for cases. The probability of being HPV16-positive for the controls was then estimated in a model with HPV16 status in the different age groups as dependent variables, again using the GEE method in a logistic regression model.

Odds ratios with 95% confidence intervals served as measures of relative risk. Using conditional logistic regression (22), we analyzed the risk association between HPV16 positivity in the first smear and cervical carcinoma *in situ*, stratified by age at time of first smear ($<$25, 25–29, $\geq$30 years). In addition, among women having two or more smears, we estimated the relationship between risk of cervical carcinoma *in situ* and HPV16 status in the last two smears taken before diagnosis, again using conditional logistic regression.

To relate the risk association between HPV16 positivity at different years before diagnosis and cervical carcinoma *in situ*, we used the following multiple imputation method: For each woman, the probability of being HPV16-positive at different years t before diagnosis was estimated based on the average of the four (or fewer) closest smears in a selected neighborhood around the year t considered. If t_1–t_k denote time points before diagnosis and HPV_1–HPV_k denote the corresponding HPV16 status, and $HPV_i = 1$ if the smear is positive and $HPV_i = 0$ if negative, then estimates of the probability were:

$$P(t) = \frac{\sum_i \frac{HPV_i}{d_i}}{\sum_i \frac{1}{d_i}}$$

where the sum is taken over the four (or fewer) closest smears in the chosen neighborhood ($\pm$ 4 years) around the year t and $d_i = |t_i - t|$ is the absolute distance. In neighborhoods with no smears, the probability was defined as missing. For each case and control, estimation was performed separately for values of t varying from 1 to 15 years before diagnosis or until time of first smear. In Fig. 1 we give an example of how the probability $P(t)$ was estimated. For each year, the estimated probability $P(t)$ of being HPV16-positive was used to generate a single binary exposure variable, $Z(t)$, which took the value 1 with probability $P(t)$ and 0 otherwise (missing if missing probability). Using conditional logistic regression model with the covariate $Z(t)$, we analyzed the risk association between cervical carcinoma *in situ* and HPV16 status separately for each year before diagnosis. The simulation procedure was performed 100 times, and the estimated risk association between HPV16 positivity at different years before diagnosis and carcinoma *in situ* was the average of the separate estimates. The variance of the estimated risk was calculated as the average of within-imputation variance plus the between-imputation variance (23). Confidence intervals were then assigned as usual. To evaluate the robustness of the results, different neighborhoods were applied ($\pm$ 3 year, no restriction). This did not affect the results.

To estimate the incubation period from HPV16 infection to carcinoma *in situ*, we calculated the nonparametric Kaplan-Meier estimator among the group of cases with an HPV16-positive final smear during the last year before diagnosis. In the analysis of the probability "survival" function, the time scale used was backwards from diagnosis to estimated conversion from HPV16-negative to -positive. The time of conversion was estimated in two different ways: The first estimation was based on the midpoint between the last (*i.e.*, most recent) registered HPV16-negative smear and the closest subsequent HPV16-positive smear. Cases with no HPV16-negative smears were censored at the time of their first smear. The second estimation was based on the midpoint between the first (*i.e.*, the earliest) registered HPV16-negative smear and the closest subsequent HPV16-positive smear. Here, cases were censored at the time of the first smear if that smear was HPV16-positive. The Kaplan-Meier estimates were calculated for both estimated times of conversion because these were expected to give lower and upper limits of the true probability. In these analyses, we stratified for age at diagnosis, grouped as ($<$35, $\geq$35 years).

We used SAS procedures PHREG (24) for conditional logistic

Fig. 1. Example showing how to estimate probability of being HPV-positive using different time windows ($\pm$ 4 years). HPV status (0 or 1) at different years before diagnosis is indicated. The probability of being HPV-positive 10 years ($t = 10$) before diagnosis is calculated using the expression described in "Subjects and Methods." Using a neighborhood of $\pm$ 4 years: $P(10) = (0/1 + 1/3 + 0/2 + 0/4)/(1/1 + 1/3 + 1/2 + 1/4) = (1/3)/(25/12) = 4/25$.

regression analyses, GENMOD (25) for GEE analyses, and LIFETEST (26) for Kaplan-Meier estimates.

RESULTS

Characteristics of Participants. The characteristics of the study population and the distribution of smears before diagnosis are summarized in Table 1. The cases and their first controls had a median age at diagnosis of 35 years (range, 20–70 years) and were followed for a median time of 8 years (range, <1 to 25 years). The second controls had a median age of 34 years and were followed for a median time of 5 years. The total number of smears analyzed for HPV16 was 4034: 2228 for the cases (median, 4; range, 1–25); 1479 for the first controls (median, 2; range, 1–15); and 327 for the second controls (median, 2; range, 1–7). As seen in Table 1, the majority of smears in surplus among cases were taken as confirmatory smears during the last year before the diagnosis of carcinoma *in situ* was established. In the other time intervals, the distribution of smears among cases and first controls were more similar. Therefore, in all subsequent analyses of the risk association between HPV16 and cervical carcinoma *in situ*, except for the analyses of HPV16 status in the first smear, we disregarded smears taken during the last year before diagnosis. The total number of HPV16-positive smears was 922 for the cases, 120 for the first controls, and 20 for the second controls (Table 1).

HPV16 Prevalence. Among controls, the constant background age-specific prevalence was estimated as 0.10 (0.09–0.12). Among cases, a steadily increasing probability of being HPV16-positive was observed as the time of diagnosis approached (Fig. 2). Approximately 16–18 years before diagnosis of carcinoma *in situ*, the probability of HPV16 positivity among cases did not differ from the estimated background prevalence of 0.10 in the control group, whereas immediately before diagnosis, the probability had increased to 0.56. We accepted the test for linearity of the time effect (on logit-scale), and the test for trend was highly significant ($P < 0.0001$).

Table 1 *Characteristics of the 1103 participants in a nested case-control study of cervical carcinoma in situ in Sweden 1969–1995*

	Cases (n = 484)	First controls (n = 484)	Second controls (n) = 135)
Age at diagnosis (yr), median (range)	35 (21–70)	35 (20–70)	34 (20–59)
Age at entry (yr), median (range)	26 (15–49)	26 (15–49)	27 (17–49)
Time in the study, (yr) median (range)	8 (<1–25)	8 (<1–25)	5 (<1–23)
Smears			
Total number	2228	1479	327
Median (range)	4 (1–25)	2 (1–15)	2 (1–7)
75% percentile	6	4	3
Distribution of smears (yr before diagnosis)			
<1	773	313	112
1–5	590	433	109
6–10	496	423	70
11–15	247	194	20
>15	122	116	16
No. participants with smears (yr before diagnosis)			
<1	384	251	89
1–5	321	286	81
6–10	299	286	81
11–15	147	137	16
>15	81	80	10
No. participants with HPV16-positive smears (no. smears)			
0	178	391	115
1	86	75	20
2	54	14	0
≥3	166	4	0

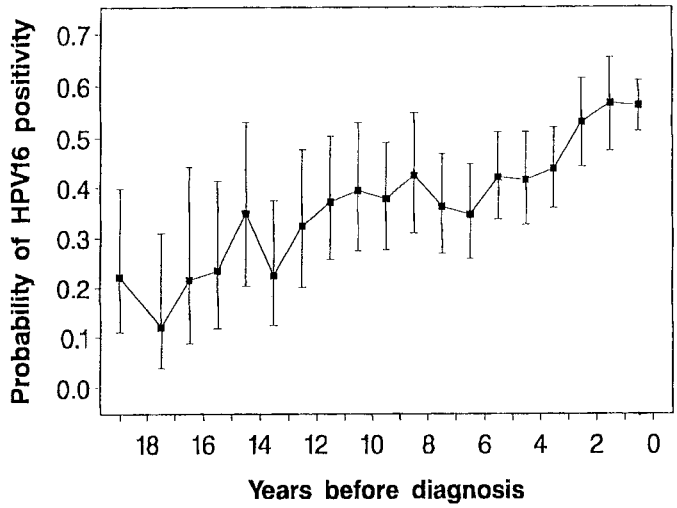

Fig. 2. Estimated probability of HPV16 positivity among cases at each year before diagnosis. *Bars*, confidence interval at each year.

Table 2 *Odds ratios and 95% confidence intervals of cervical carcinoma in situ in relation to HPV16 status*

	No. of cases/ controls	OR (95% CI)[a,b]
HPV 16 status in the first smear		
Age <25 years		
Neg	144/221	1 (Ref.)
Pos	53/15	5.4 (2.6–11.1)
Age 25–29 years		
Neg	99/168	1 (Ref.)
Pos	47/10	7.4 (3.1–17.9)
Age ≥30 years		
Neg	110/192	1 (Ref.)
Pos	30/11	5.3 (2.2–13.0)
Total	483/617[c]	$P = 0.434$[d]
HPV 16 status in the last two smears before diagnosis[e]		
Neg-Neg	108/235	1 (Ref.)
Pos-Neg	26/11	4.9 (2.1–11.3)
Neg-Pos	54/14	9.7 (4.3–21.8)
Pos-Pos	75/4	31.2 (10.6–91.8)

[a] Smears taken the last year before diagnosis have been excluded.
[b] OR, odds ratio; CI, confidence interval; Neg, negative; Pos, positive; Ref, within reference values.
[c] One risk set (including one case and two controls) was not included because of missing information on HPV status.
[d] Test for interaction by age.
[e] Only analyzed for women with 2 or more smears.

Among the 422 interviewed cases (~90%), neither smoking status (ever *versus* never or current *versus* ex- or nonsmokers) nor oral contraceptive use (never *versus* ever or current *versus* ex- or nonusers) before diagnosis affected the probability of being HPV16-positive at different time intervals (data not shown).

Risk Associations. We examined the association between HPV16 status in the first smear (taken at the same time point before diagnosis for cases and their matched controls) and the risk for cervical carcinoma *in situ*, stratifying by age at time of first smear. Regardless of age, having an HPV16-positive first smear was related to a more than 5-fold increased risk for a subsequent diagnosis of cervical carcinoma *in situ* (Table 2).

To further clarify the importance of persistence and transience we examined the presence of HPV16 in the two most recent smears before diagnosis, disregarding smears taken during the last year before diagnosis. Women with a positive first smear and a negative second smear had a 5-fold increased risk compared with women with only HPV16-negative smears (Table 2). An almost 10-fold increased risk was observed among women who initially had a negative smear and

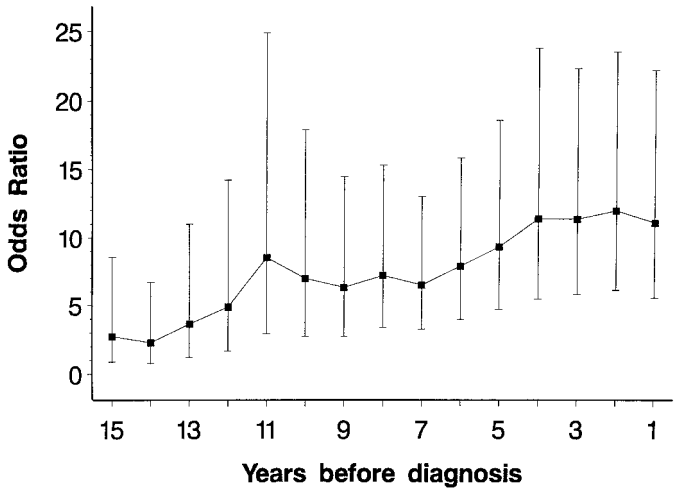

Fig. 3. Odds ratio for cervical carcinoma *in situ* in relation to HPV16 status estimated for each year before diagnosis, using a multiple imputation method. The curve is generated using smears in the neighborhood of ± 4 years. *Bars*, confidence interval at each year.

a subsequent positive smear, compared with HPV16-negative women. Infection with HPV16 in both smears was associated with a substantially increased risk of 31.2 (95% confidence interval, 10.6–91.8).

Subsequently, we calculated risk estimates for each year before diagnosis, using the multiple imputation method. Regardless of whether the analyses were performed without restrictions or with restrictions in certain time windows, a fairly consistent trend was observed. The HPV16-positive individuals were at increased risk up to 13 years before diagnosis (Fig. 3). Adjustment for potential confounding by smoking, oral contraceptive use, and number of sexual partners among the interviewed subjects did not affect risk estimates materially (data not shown).

Estimated Incubation Period. The time from HPV16 infection to diagnosis was estimated among cases who were HPV16-positive at diagnosis. In Fig. 4, the time since first HPV16-positive smear is indicated among cases, stratified according to age at diagnosis (<35, ≥35 years). Among women ≥35 years at diagnosis, ~50% had become HPV16-positive between 7 and 12 years before diagnosis. The corresponding estimated incubation period for women diagnosed before age 35 was 5–8 years.

DISCUSSION

Ideally, the natural history of cervical carcinoma *in situ* would be studied in a prospective study following many thousands of healthy women, initially without cytological changes, with frequently repeated measurements until their eventual diagnosis of carcinoma *in situ*. Apart from the ethical dilemma with this approach, such a study would probably have to continue for approximately two decades to capture the entire biological spectrum from normal cytology to carcinoma *in situ*; costs would be prohibitive and logistic problems substantial.

To overcome these difficulties, we used archival smears, taken repeatedly up to 26 years before diagnosis of carcinoma *in situ*, from case women and matched control women nested in a population-based cohort of women attending cytological screening in Sweden. Because of the nested design, our study preserves the validity of the underlying cohort study. Furthermore, HPV analyses were performed with a highly sensitive PCR-based system, by personnel blinded for case-control status. Since contamination is unlikely because of stringent procedures for DNA extraction and HPV analyses and because the HPV detection system has a high sensitivity and specificity, we chose to include β-actin-negative smears that were HPV16-positive in the analyses. Indeed, our risk estimates did not change when all β-actin-negative smears were excluded from the analyses (data not shown). Furthermore, we had detailed covariate information on a large proportion of the women (13), which enabled us to adjust at each year before diagnosis for the potential confounding influence of smoking, oral contraceptive use, and number of sexual partners. Potential limitations include missing information regarding hysterectomy for approximately one-third of the controls. However, we have no reason to believe that these women would differ with respect to prevalence of hysterectomy from the interviewed women, among whom only 2% (8 of 422) had undergone a hysterectomy. This implies that approximately five included, not interviewed controls may received hysterectomies. It is unlikely that inclusion of these controls has affected the risk estimates.

The cases had more registered smears than the controls. Reassuringly, the difference in number of smears was largely explained by the large number of confirmatory smears among the cases taken during the last year before diagnosis. By excluding all smears registered during the last year from the analyses of the risk associations, we reduced potential bias attributable to higher probability for HPV detection among cases. Furthermore, we used a multiple imputation method to determine HPV16 status at a given time point, thus taking into consideration and controlling for the fact that cases and their matched controls had smears taken at different time points before diagnosis. Consequently, we were able to maintain a matched approach in our risk estimations, preserving the power obtained by using all risk sets.

Our finding of HPV16 DNA in 56% of carcinomas *in situ* at the time of diagnosis is in line with previous studies both on *in situ* and invasive cervical carcinomas (3). In recent years, a few prospective studies with short follow-up have indicated that persistent, rather than transient, infection with oncogenic HPV types is associated with increased risk for cervical carcinoma *in situ* (4–6). In our population-based study where information on HPV status had been collected up to 26 years before diagnosis, we found strong evidence that persistent and/or recurrent infection with HPV16 is associated with substantially higher risk than occasional infection. Thus, having an HPV16-positive first smear, on average 8 years before diagnosis, increased risk more

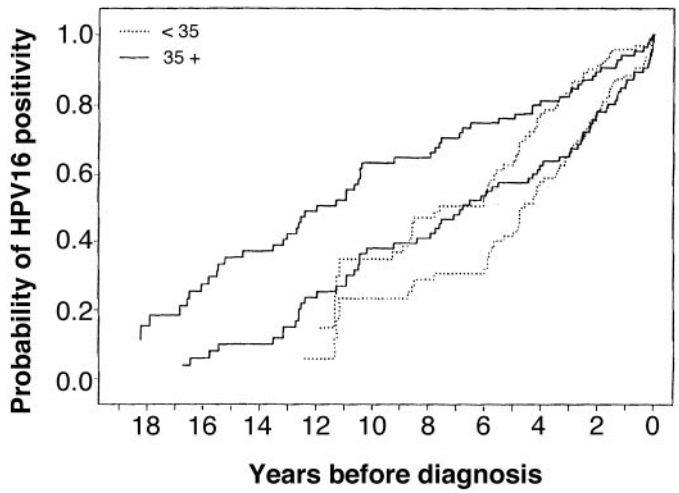

Fig. 4. Kaplan-Meier estimate giving the probability of HPV16 positivity among cases at different years before diagnosis. This analysis includes only cases who were HPV16-positive at time of diagnosis. Upper and lower limits for the time interval at each probability value are indicated for women ≥35 years at diagnosis (*solid line*) and <35 years at diagnosis (*dotted line*).

than 5-fold, whereas having an HPV16-positive last smear increased risk almost 10-fold and being HPV16-positive in the last two smears increased risk 30-fold. In line with these findings, persistent and/or recurrent infection was a rare observation among our control women who did not develop carcinoma *in situ*. In a recent case-control study by Wallin *et al.* (27), HPV DNA status was established in PAP smears before diagnosis of invasive cervical carcinoma. Similar to our results, a positive HPV status in both the first and last samples was associated with a highly increased risk. Moreover, Wallin *et al.* (27) detected the same HPV type in the first smear and the biopsy specimens in all case women with positive HPV status at entry, further supporting the hypothesis that viral persistence is crucial for cervical carcinogenesis.

Factors associated with the establishment of persistent infection may explain why HPV infection is such a common sexually transmitted disease among young women (28) and why so few of these women ever develop carcinoma *in situ* and ultimately invasive cervical cancer (29). However, the determinants of persistence are still largely unknown. The typically successful control of HPV infection appears to depend on the interaction between the virus, the infected host cells, and the host immune system (30, 31). Probably the most convincing evidence that the immune system plays a central role comes from studies of women infected with HIV and those undergoing renal transplantation. With increasing immunosuppression, these women develop high prevalences of oncogenic HPV infections and subsequently carcinoma *in situ* (1, 32). Logically, behavioral cofactors for HPV persistence should exist as well, but there presently are few data to support this assumption. In particular, exogenous factors such as cigarette smoking and oral contraceptive use have been mentioned, but not conclusively settled, as cofactors influencing HPV persistence (4–6). In our study, risk estimates were not affected by adjustment at each year before diagnosis for smoking behavior and oral contraceptive use.

It has long been suspected that the time from HPV infection to carcinoma *in situ* may be several years, but to date no study has ever been able to more precisely estimate the length of this period. In our study, HPV16-positive women were already at increased risk of being diagnosed with cervical carcinoma *in situ* 13–15 years before they were actually diagnosed. However, because HPV16 had only been measured at a limited number of time points before diagnosis, we cannot exclude the possibility that some women seemingly having persistent infection may have lost the virus and subsequently become reinfected by the same viral type during the interval between smears. Thus, based on our results, we cannot conclude whether the higher frequency of HPV16 infection seen among cases so many years before diagnosis is the result of persistent infection or reinfection with HPV16. This problem of interpretation can be reduced, but not eliminated, through more frequent HPV measurements in a prospective study. However, only by studying the microheterogeneity of HPV16 detected in the smears can we be sure that the same HPV16 subtype has persisted over a long period. On the other hand, in favor of persistence are our findings that adjustment for sexual activity at each time point before diagnosis of carcinoma *in situ* had no effect on either the prevalence of HPV among cases or the risk association between cases and controls.

We acknowledge that our attempt to estimate an incubation period for case women who were HPV16-positive at diagnosis can be criticized. This period may be underestimated because HPV16 infection is likely to be present some time before detection, but it could also be overestimated because carcinoma *in situ*, an asymptomatic condition, may exist some time before detection. The net result of these two competing mechanisms is presently unknown. Nevertheless, we feel confident in our incubation estimates because of the close concordance between the values obtained by two extreme definitions for time of initial HPV infection. Younger case women are bound to have had, on average, a faster progression to carcinoma *in situ* than older women simply because of shorter interval from start of exposure. In contrast, many women 35 years or older are likely to have passed their sexually most active years with different partners and thus the period at highest risk of exposure to HPV16. In this latter group of women, the average incubation period from HPV16 infection to carcinoma *in situ* was estimated to be between 7 and 12 years. However, for 20% of the women, the incubation period may have been 15 years or longer.

In conclusion, our data imply that infection with HPV16 is associated with a highly increased risk of cervical carcinoma *in situ*, detectable up to 13 years before diagnosis. The risk increased dramatically among women with persistent or recurrent HPV16 infection in two sequential smears. We estimate that the average incubation period from initial HPV16 infection to diagnosis of carcinoma *in situ* is between 7 and 12 years but that it may be up to two decades for some women.

ACKNOWLEDGMENTS

We acknowledge valuable comments on the design, analysis, and interpretation of this study by Drs. Tom Rohan and Matthew Zack.

REFERENCES

1. IARC. Human papillomaviruses. Monographs on the Evaluation of Carcinogenic Risks to Humans. IARC Sci. Publ., *64:* 35–260, 1995.
2. Parkin, D. M., Pisani, P., and Ferlay, J. Estimates of the worldwide incidence of 25 major cancers in 1990. Int. J. Cancer, *80:* 827–841, 1999.
3. Bosch, F. X., Manos, M. M., Munoz, N., Sherman, M., Jansen, A. M., Peto, J., Schiffman, M. H., Moreno, V., Kurman, R., Shah, K. V., and the International Biological Study on Cervical Cancer Study Group. Prevalence of human papillomavirus in cervical cancer: a worldwide perspective. J. Natl. Cancer Inst., *87:* 796–802, 1995.
4. Hildesheim, A., Schiffman, M. H., Gravitt, P. E., Glass, A. G., Greer, C. E., Zhang, T., Scott, D. R., Rush, B. B., Lawler, P., Sherman, M. E., Kurman, R. J., and Manos, M. M. Persistence of type-specific human papillomavirus infection among cytologically normal women. J. Infect. Dis., *169:* 235–240, 1994.
5. Ho, G. Y. F., Burk, R. D., Klein, S., Kadish, A. S., Chang, C. J., Palan, P., Basu, J., Tachezy, R., Lewis, R., and Romney, S. Persistent genital human papillomavirus infection as a risk factor for persistent cervical dysplasia. J. Natl. Cancer Inst., *87:* 1365–1371, 1995.
6. Remmink, A. J., Walboomers, J. M. M., Helmerhorst, T. J. M., Voorhorst, F. J., Rozendaal, L., Risse, E. K. J., Meijer, C. J. L. M., and Kenemans, P. The presence of persistent high-risk HPV genotypes in dysplastic cervical lesions is associated with progressive disease; natural history up to 36 months. Int. J. Cancer, *61:* 306–311, 1995.
7. Liaw, K-L., Glass, A. G., Manos, M. M., Greer, C. E., Scott, D. R., Sherman, M., Burk, R. D., Kurman, R. J., Wacholder, S., Rush, B. B., Cadell, D. M., Lawler, P., Tabor, D., and Schiffman, M. Detection of human papillomavirus DNA in cytologically normal women and subsequent cervical squamous intraepithelial lesions. J. Natl. Cancer Inst., *91:* 954–960, 1999.
8. Schneider, A., and Koutsky, L. A. Natural history and epidemiological features of genital HPV infection. *In:* N. Munoz and F. X. Bosch, (eds.), The Epidemiology of Human Papillomavirus and Cervical Cancer, pp. 25–52, IARC. Scientific Publ. No. 119. Lyon, France: IARC, 1992.
9. Munoz, N., and Bosch, F. X. Current views on the epidemiology of HPV and cervical cancer. *In:* C. Lacey (ed.), Papillomavirus Reviews: Current Research on Papillomaviruses, pp. 227–237. Leeds, UK: FRCP, Leeds University Press, 1996.
10. Munoz, N., Bosch, X., and Kaldor, J. M. Does human papillomavirus cause cervical cancer? The state of the epidemiological evidence. Br. J. Cancer, *57:* 1–5, 1988.
11. Koutsky, L. A., Holmes, K. K., Critchlow, C. W., Stevens, C. E., Paavonen, J., Beckmann, A. M., DeRouen, T. A., Galloway, D. A., Vernon, D., and Kiviat, N. B. A cohort study of the risk of cervical intraepithelial neoplasia grade 2 or 3 in relation to papillomavirus infection. N. Engl. J. Med., *327:* 1272–1278, 1992.
12. Gustafsson, L., Sparén, P., Gustafsson, M., Wilander, E., Bergström, R., and Adami, H-O. Efficiency of organised and opportunistic cytological screening for cancer in situ of the cervix. Br. J. Cancer, *72:* 498–505, 1995.
13. Ylitalo, N., Sörensen, P., Josefsson, A., Frisch, M., Sparén, P., Pontén, J., Gyllensten, U., Melbye, M., and Adami, H-O. Smoking and oral contraceptives as risk factors for cervical carcinoma in situ. Int. J. Cancer, *81:* 357–365, 1999.
14. Socialstyrelsen. Epidemiologiskt centrum. Cancer incidence in Sweden, The Swedish Cancer Registry. 1995.
15. National Board of Health and Welfare. Socialstyrelsens cirkulär angående anmälan till cancerregistret. Stockholm: National Board of Health and Welfare, 1968.
16. Chua, K. L., and Hjerpe, A. Polymerase chain reaction analysis of human papillomavirus in archival cytologic smears. Anal. Quant. Cytol. Histol., *17:* 221–229, 1995.

17. Josefsson, A., Livak, K., and Gyllensten, U. Detection and quantitation of human papillomavirus by using the fluorescent 5′ exonuclease assay. J. Clin. Microbiol., *37:* 490–496, 1999.
18. Livak, K. J., Marmaro, J., and Todd, J. A. Towards fully automated genomewide polymorphism screening. Nat. Genet., *9:* 341–342, 1995.
19. Holland, P. M., Abrahamson, R. D., Watson, R., and Gelfand, D. H. Detection of specific polymerase chain reaction product by utilizing the 5′-3′ exonuclease activity of *Thermus aquaticus* DNA polymerase. Proc. Natl. Acad. Sci. USA, *88:* 7276–7280, 1991.
20. Ylitalo, N., Bergström, T., and Gyllensten, U. Detection of genital human papillomavirus by single-tube nested PCR and type-specific oligonucleotide hybridization. J. Clin. Microbiol., *33:* 1822–1828, 1995.
21. Diggle, P. J., Liang, K-Y., and Zeger, S. L. Analysis of longitudinal data. Oxford: Oxford University Press, 1994.
22. Breslow, N. E., and Day, N. E. Statistical Methods in Cancer Research. 1. The Analysis of Case-Control Studies. IARC Scientific Publ. No. 32. Lyon, France: IARC, 1980.
23. Rubin, D. The Encyclopedia of Biostatistics, P. Armitage and T. Colton (eds.), Vol. 4, pp. 2772–2779. New York: John Wiley and Sons, 1998.
24. SAS Institute. The PHREG procedure. *In:* SAS/STAT Software: Changes and Enhancements through Release 6.12, pp. 807–884. Cary, NC: SAS Institute, 1996.
25. SAS Institute. The GENMOD procedure. *In:* SAS/STAT Software: Changes and Enhancements through Release 6.12, pp. 21–42. Cary, NC: SAS Institute, 1996.
26. SAS Institute. The LIFETEST procedure. *In:* SAS/STAT Users Guide, Version 6, 4th Ed., Vol. 2, pp. 1027–1069. Cary, NC: SAS Institute, 1989.
27. Wallin, K-L., Wiklund, F., Ångström, T., Bergman, F., Stendahl, U., Wadell, G., Hallmans, G., and Dillner, J. Type-specific persistence of human papillomavirus DNA before the development of invasive cervical cancer. N. Engl. J. Med., *341:* 1633–1638, 1999.
28. Bauer, H. M., Ting, Y., Greer, C. E., Chambers, J. C., Tashiro, C. J., Chimera, J., Reingold, A., and Manos, M. M. Genital human papillomavirus infection in female university students as determined by a PCR-based method. JAMA, *265:* 472–477, 1991.
29. Ho, G. Y. F., Bierman, R., Beardsley, L., Chang, C. L., and Burk, R. D. Natural history of cervicovaginal papillomavirus infection in young women. N. Engl. J. Med., *338:* 423–428, 1998.
30. Schiffman, M. H. New epidemiology of human papillomavirus infection and cervical neoplasia. J. Natl. Cancer Inst., *87:* 1345–1347, 1995.
31. Tsukui, T., Hildesheim, A., Schiffman, M. H., Lucci, J., III, Contois, D., Lawler, P., Rush, B. B., Lorincz, A. T., Corrigan, A., Burk, R. D., Qu, W., Marshall, M. A., Mann, D., Carrington, M., Clerici, M., Shearer, G. M., Carbone, D. P., Scott, D. R., Houghten, R. A., and Berzofsky, J. A. Interleukin 2 production *in vitro* by peripheral lymphocytes in response to human papillomavirus-derived peptides: correlation with cervical pathology. Cancer Res., *56:* 3967–3974, 1996.
32. Melbye, M., Smith, E., Wohlfahrt, J., Osterlind, A., Orholm, M., Bergmann, O. J., Mathiesen, L., Darragh, T. M., and Palefsky, J. M. Anal and cervical abnormality in women—prediction by human papillomavirus tests. Int. J. Cancer, *68:* 559–565, 1996.

A Prospective Study Showing Long-Term Infection with Human Papillomavirus 16 before the Development of Cervical Carcinoma *in Situ*

Nathalie Ylitalo, Agnetha Josefsson, Mads Melbye, et al.

Cancer Res 2000;60:6027-6032.