Clustering of Multiple Human Papillomavirus Infections in Women From a Population-Based Study in Guanacaste, Costa Rica

Salvatore Vaccarella,¹ Silvia Franceschi,¹ Rolando Herrero,² Mark Schiffman,³ Ana Cecilia Rodriguez,² Allan Hildesheim,³ Robert D. Burk,⁴ and Martyn Plummer¹

¹International Agency for Research on Cancer, Lyon, France; ²Proyecto Epidemiológico Guanacaste, Fundación Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud, San José, Costa Rica; ³Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland; and ⁴Albert Einstein College of Medicine, New York, New York

Objective. To evaluate clustering patterns of prevalent infection with multiple human papillomavirus (HPV) types in 8365 nonhysterectomized women from the Guanacaste Study of HPV Natural History.

Methods. HPV testing was performed on cervical cells by MY09/M11 L1 degenerate consensus primer polymerase chain reaction method, with dot-blot hybridization for genotyping. Logistic regression was used to model type-specific HPV positivity, adjusted for age, lifetime number of sexual partners, and specific HPV type prevalence. Woman-level random effects were added to represent unobservable risk factors common to all HPV types.

Results. The observed-to-expected ratio for infections with 2 types was 1.16 (95% credible interval: 1.11–1.21), and for ≥ 3 types was 1.04 (95% credible interval: .96–1.13). The tendency of HPV types to cluster increased significantly with the genetic similarity of L1 regions. *P* value $< .01$ was observed for 2 HPV pairs: HPV-62 and -81 were found together more, while HPV-51 and -71 were found together less often than expected.

Conclusions. We found a small degree of aggregation between any HPV types and lack of clustering between specific carcinogenic types. Our data indirectly provide reassurance on lack of misclassification for the large majority of HPV types in multiple infections detected by the MY09/11 method and genotyped using dot-blot hybridization.

Concurrent infections with more than one genital human papillomavirus (HPV) type are relatively frequent in the general female population [1–7]. On account of the sexually transmitted nature of HPV, women infected with one HPV type are also often infected with additional types [2–9]. However, it is not clear yet whether particular combinations of HPV types have a tendency to cluster together above or below what would be expected by sexual transmission. This is relevant for the evaluation of the effects of HPV prophylactic vaccines. The removal of certain HPV types through vaccination

could, in theory, indirectly increase or decrease the prevalence of other untargeted types.

This issue is complicated by the fact that several HPV testing methods exist, each displaying a different sensitivity to detect specific HPV types and multiple HPV infections [10–12]. Polymerase chain reaction (PCR) systems that use a single primer sequence might be inefficient in amplifying multiple types. Diagnostic artifacts could affect the detection of HPV types in multiple infections. A previous analysis of the International Agency for Research on Cancer HPV Prevalence Surveys (IHPS) using GP5+/6+ PCR assays showed that genotyping with an enzyme immunoassay (EIA), though not with reverse line blot (RLB) procedure, led to an overestimation of certain combinations of HPV types [1]. We therefore evaluated the pattern of prevalent infection with multiple HPV types in 8365 nonhysterectomized women at enrollment into the Guanacaste Study of HPV Natural History (GSHNH), a population-based cohort study carried out in Costa Rica.

Received 29 December 2010; accepted 22 March 2011.

Potential conflicts of interest: none reported.

Correspondence: Salvatore Vaccarella, PhD, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon cedex 08, France (vaccarella@iarc.fr).

The Journal of Infectious Diseases 2011;204:385–90

© The Author 2011. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

0022-1899 (print)/1537-6613 (online)/2011/2043-0011\$14.00

DOI: 10.1093/infdis/jir286

METHODS

Study Population

A population-based cohort, named Proyecto Epidemiológico Guanacaste, was assembled between 1993 and 1994 in Guanacaste, Costa Rica, to study the natural history of HPV and cervical neoplasia. Details of this study have been previously reported [13, 14]. Briefly, from a random sample of census tracts of this mainly rural population (240 000 inhabitants), 11 742 potentially eligible subjects aged 18 years and older were identified, 10 738 women were eligible and invited to participate, and 10 049 women (94% of eligible women) agreed to visit one of the study clinics. Women who agreed to participate signed informed consent forms approved by an Ad Hoc Institutional Review Board of Costa Rica and the US National Cancer Institute. Interviews included questions on demographic factors, medical history, and behaviors (sexual, reproductive, and smoking), addressing possible risk factors for cervical cancer. From the present analyses, we excluded 665 women who had undergone total hysterectomies before enrollment, leaving 9384 women. The baseline data were used in the present report.

Gynecological Examination and Specimen Collection

At enrollment, women underwent a pelvic examination by a small team of experienced nurses. Exfoliated cervical cells for conventional and liquid-based cytology (ThinPrep, Hologic Corporation [formerly Cytec]) were collected using a Cervex broom device (Rovers). A second specimen of exfoliated cells was also taken for HPV testing and stored in specimen transport medium.

HPV DNA Detection Techniques

HPV testing was performed by a MY09/M11 L1 degenerate consensus primer PCR (MY09/M11 PCR) method [10] with AmpliTaq Gold polymerase [15]. This system includes additional primers for HPV51 (HMB01) and 2 primers for HPV35 (5'-CGGCCCAACGGAACTGATC, 5'-GCACAAGGCCATA-ATAATGG). After amplification, PCR products were analyzed by gel electrophoresis, transferred to nylon filters, and then hybridized overnight by use of radiolabeled generic probes for HPV (types 11, 16, 18, 51, 73, and 81 combined). Dot-blot hybridization was used for HPV genotyping; probes were specific for the following 45 genital HPV types [10, 15–16]: 2, 6, 11, 13, 16, 18, 26, 31–35, 39–40, 42–45, 51–59, 61–62, 64, 66–74, 81–85, and 89. Hybridized probes were detected by use of a streptavidin–horseradish peroxidase conjugate (Amersham) and a commercial chemiluminescence detection kit (Amersham). A specimen was classified as HPV positive but was uncharacterized if it tested positive for HPV DNA by the radiolabeled generic probe mix but was not positive for HPV DNA by the type-specific probe. Three experienced investigators interpreted each dot-blot result, and discrepancies were resolved by consensus.

Sequence Alignment and Phylogenetic Analysis

Prototype sequences for specific HPV types were obtained from the National Center for Biotechnology Information/GenBank database. Subsequently, the DNA sequence data of L1 genes were aligned with ClustalW2 using default options [17, 18]. Similarity between HPV types was measured by the percent identity scores in the best sequence alignment. Use of the L1 open reading frame has been shown to be robust for classification of closely related HPV types and species [19].

Statistical Analyses

The analysis of clustering of HPV types was restricted to HPV types with a prevalence of at least .5% in the whole study population. Twenty-four types were thus considered, including 11 carcinogenic types (HPV16, -18, -31, -33, -35, -39, -45, -51, -52, -56, and -58) and 13 other types (HPV6, -53, -54, -61, -62, -66, -70, -71, -73, -81, -83, -84, and -85) [20, 21].

The statistical methodology to assess clustering of multiple HPV infections is similar to that of a previous analysis of the IARC HPV Prevalence Surveys [1]. Briefly, multivariate logistic regression was used to model type-specific HPV positivity. Models were controlled for age (<25, 25–34, 35–44, 45–54, ≥55 years), lifetime number of sexual partners, and specific HPV type prevalence, as indicated. Since the data have a hierarchical structure, with HPV infections nested within women, multilevel models were used, with woman-level random effects. In this context, a random effect is an unobserved quantity that varies across units in a population. Random effects allow women with the same observable risk factors to have different levels of risk of HPV positivity. Woman-level random effects represent all sources of residual variations in risk of HPV infection other than those already represented by covariates. Models were fitted using a Bayesian approach, using Markov Chain Monte Carlo simulation. Estimates are reported as posterior means and 95% credible intervals (CIs), unless otherwise noted. Discrepancies between the data and the model were measured using observed-to-expected (O/E) ratios for the counts of multiple infections. Goodness of fit of the model was assessed by posterior predictive 2-sided *P* values [22], with the number of pairwise joint HPV infections as a test quantity.

The 2-way analysis of associations among 24 HPV types implies the generation of 276 *P* values. In order to minimize errors due to multiple comparisons and for consistency with the previous analysis of the IHPS, the threshold for significance of *P* values to test was set to .01.

RESULTS

Out of 9384 women who came to visit one of the clinics, 8424 women underwent a pelvic examination and had valid HPV results and information on lifetime number of sexual partners. Forty-nine women with a histologically confirmed diagnosis of

Figure 1. Type-specific human papillomavirus (HPV) prevalence in 8365 women, Guanacaste Study of HPV Natural History.

cervical intraepithelial neoplasia grade 3 (CIN3) and 10 women with invasive cervical cancer were excluded, thus leaving 8365 women for this analysis. The mean age of the study participants was 40.4 years. Overall, HPV prevalence was 23.4%, and multiple infections were found in 35.6% of HPV-positive women.

Figure 1 shows the proportion of individual HPV types found in single and multiple infections. Among carcinogenic types, HPV16 was most often (50.0%) found alone, whereas HPV35 and -45 were detected in combination with other types in $\geq 70\%$ of infected women.

Table 1 shows the observed and expected numbers of women with single and multiple HPV infections (1 type, 2 types, and ≥ 3 types) under 3 different statistical models of increasing

complexity. The basic model included only age and specific HPV type prevalence as covariates. The O/E ratio for infection with 2 HPV types was 1.20 (95% CI, 1.13–1.28) and for infection with ≥ 3 HPV types was 4.52 (95% CI, 4.04–5.04). The adjusted model included additionally a woman's lifetime number of sexual partners (1, ≥ 2) as a covariate. With the adjusted model, the O/E ratio was 1.17 (95% CI, 1.09–1.24) for 2 HPV types and 3.61 (95% CI, 3.20–4.10) for ≥ 3 HPV types. The full model included random intercepts for individual women, representing unobserved host or environmental risk factors for HPV infection that are common to all types. With the full model, the O/E ratio for infections with 2 types was 1.16 (95% CI, 1.11–1.21) and for infections with ≥ 3 types was 1.04 (95% CI, .96–1.13).

Table 1. Observed-to-Expected Ratio of Multiple Infections With the 24 Most Common Human Papillomavirus (HPV) Types, According to Various Models, Guanacaste Study of HP Natural History.

Number of HPV Types	O	Basic Model		Adjusted Model		Full Model	
		E ^a	O/E Ratio (95% CI) ^a	E ^b	O/E Ratio (95% CI) ^b	E ^c	O/E Ratio (95% CI) ^c
0	6496	6003.6	1.08 (1.07–1.09)	6046.3	1.07 (1.06–1.09)	6459.8	1.01 (1.00–1.01)
1	1258	1976.2	0.64 (.62–.65)	1910.7	0.66 (.64–.68)	1356.1	0.93 (.90–.96)
2	408	340.1	1.20 (1.13–1.28)	351.5	1.17 (1.09–1.24)	352.8	1.16 (1.11–1.21)
3–7	203	45.1	4.52 (4.04–5.04)	56.5	3.61 (3.20–4.10)	196.3	1.04 (.96–1.13)

NOTE. O, observed; E, expected; CI, credible interval.

^a Controlling for age and study area.

^b As ^a plus lifetime number of sexual partners.

^c As ^b plus individual random effects.

Figure 2. *P* values for joint human papillomavirus (HPV) infections, as estimated by the Full Model.

Although the difference between observed and expected counts was reduced substantially after the addition of individual random effects, particularly for infections with ≥ 3 HPV types, a small excess of infections with 2 types was still observed.

Figure 2 shows *P* values for tests of the hypothesis of no association between HPV types, under the full model. A *P* value $< .01$ was observed for 2 pairs of HPV types: HPV62 and -81 were found together significantly more often than expected (O/E ratio, 2.81; 99% confidence interval based on the Poisson distribution, 1.39–5.02), while HPV51 and -71 were found together significantly less often than expected (O/E ratio, .07; 99% confidence interval based on the Poisson distribution, .00–.52).

Table 2 displays the results on the tendency of closely related HPV types to cluster together. Among the HPV pairs, 222 showed a percent identity of the DNA sequences in the L1 region between 59% and 69%, 49 pairs had a percent identity score between 70% and 79%, and 5 pairs had a percent identity score of 80% or above. The O/E ratio was .91 (99% CI, .80–1.05) for pairs with percent identity $< 70\%$, 1.07 (99% CI, .93–1.24) for

pairs with percent identity between 70% and 79%, and 1.22 (99% CI, 1.01–1.49) for pairs with percent identity $\geq 80\%$. The formal test for linear trend of O/E ratios was significant ($P = .0089$). The analyses were stratified by age groups (< 30 , 30–49, and ≥ 50 years), producing substantially similar results.

DISCUSSION

Overall, in the present analysis of the Guanacaste Study of HPV Natural History, which used the MY09/11 PCR method for HPV testing, infections with multiple HPV types occurred more often than would be expected by chance. However, the excess of multiple HPV infections was small, after controlling for sources of common correlation between HPV types. For the large majority of HPV types, including those targeted by the current prophylactic HPV vaccines, there was no evidence of a tendency to cluster positively or negatively in multiple infections. However, the tendency to cluster increased weakly but significantly with the genetic similarity of the L1 region. One pair (HPV62

Table 2. Observed-to-Expected Ratio of Coinfections With Human Papillomavirus (HPV) Pairs, Aggregated According to Percent Identity of DNA of L1 Gene, Guanacaste Study of HPV Natural History.

% Identity	Number of pairs	MY09/11		
		Number of Coinfections With HPV Pairs		Observed/expected ratio (99% credible interval) ^a
		Observed	Expected ^a	
59–69 ^b	222	1106	1212.2	0.91 (.80–1.05)
70–79	49	270	251.9	1.07 (.93–1.24)
$\geq 80^c$	5	26	21.3	1.22 (1.01–1.49)
Linear trend test				$P = .0089$

NOTE. ^a Asestimated by the Full Model.

^b Including HPV pair 51/71.

^c Including HPV pair 62/81.

with -81) was found together significantly more, and one pair (HPV51 with -71) was found together significantly less often than expected ($P < .01$).

A lack of evidence that particular HPV types are more or less likely to be detected together was also reported for a large sample of young women from Guanacaste, Costa Rica [7]. That study used SPF10/LiPA for HPV testing, a method that does not detect some HPV types that are relatively common, such as HPV71 and -61. General clustering of HPV types has often been observed previously [1–7, 9]. These studies have mainly used GP5+/6+ or MY09/11 PCR methods for HPV testing. The tendency to cluster has generally been explained by the fact that all HPV types are sexually transmitted and, therefore, are associated with the same risk factors. Indeed, multiple HPV infections are usually found to be more common in women with a higher number of lifetime sexual partners [23], and this was also observed in a previous analysis of the present study population [13]. Adjustment for sexual behavior is challenging, as much information is not available or derives from poorly measured questionnaire variables. A feature of our analysis was the inclusion into the logistic regression model of random effects at the woman level that allowed us to explicitly account for the common correlation between HPV infections [1, 8, 24–25] due to the unobservable risk factors shared by all HPV types. Of note, woman-level random effects do not distinguish between unobserved risk factors due to sexual behavior or partner behavior and other individual risk factors, such as immunological susceptibility. Therefore, although the common correlation between HPV types is expected to be largely driven by sexual behaviors, the cause of this variation remains unknown.

It is possible that diagnostic artifacts, as opposed to other explanations including biological interaction between HPV types, could be the reason for the small observed excess of multiple HPV infections that could not be attributed to the common correlation between HPV infections. In the IHPS study [1], HPV positivity was assessed by GP5+/6+ primer-mediated PCR, and genotyping was done using EIA or RLB. GP5+/6+ primers are known to be less efficient in the amplification of multiple types (and, therefore, less sensitive for detection of multiple HPV infections) and of certain specific types, ie, HPV53 and -61, compared with MY09/11 primer PCR [1]. In that study, a significant excess of 4 HPV pairs (HPV18/45, -31/35, -33/35, and -33/58) was explained by a certain degree of cross-hybridization between genetically related types that occurred only when the EIA, but not the RLB method, was used as genotyping procedure [1]. Genetically related HPV types have similar probe regions, which might favor cross-hybridization. No clear evidence of cross-hybridization was found in the present study from Guanacaste. However, we observed a small degree of aggregation between genetically related HPV types. The only 2 significant associations observed in the present study, however, involved relatively common HPV types and went in

opposite directions: HPV62 and -81, types that are closely related, were found together more often than expected, while HPV51 and -71, types that are not closely related, were found together less often than expected.

In conclusion, understanding whether certain HPV types tend to cluster is of importance to evaluate cross-protection in current HPV vaccines and natural history studies based on the detection of the virus. The present analysis of the GSHNH study that used MY09/11 for HPV testing found a small degree of aggregation between any HPV types and a lack of clustering between specific carcinogenic types. Our present data indirectly provide reassurance on lack of misclassification for the large majority of HPV types in multiple infections detected by the MY09/11 method and genotyped using dot-blot hybridization.

Funding

The Guanacaste Project was supported by the National Cancer Institute, National Institutes of Health, Department of Health and Human Services contracts (NO1-CP-21081, NO1-CP-33061, NO1-CP-40542, and NO1-CP-506535). A. H. and M. S. were supported by the Intramural Research Program of the National Cancer Institute. R. D. B. was supported by the National Cancer Institute (grant CA78527) and used the facilities available through the Einstein Cancer Research Center and the Center for AIDS Research.

References

1. Vaccarella S, Franceschi S, Snijders PJ, Herrero R, Meijer CJ, Plummer M. Concurrent infection with multiple human papillomavirus types: pooled analysis of the IARC HPV Prevalence Surveys. *Cancer Epidemiol Biomarkers Prev* **2010**; 19:503–10.
2. Thomas KK, Hughes JP, Kuypers JM, et al. Concurrent and sequential acquisition of different genital human papillomavirus types. *J Infect Dis* **2000**; 182:1097–102.
3. Liaw KL, Hildesheim A, Burk RD, et al. A prospective study of human papillomavirus (HPV) type 16 DNA detection by polymerase chain reaction and its association with acquisition and persistence of other HPV types. *J Infect Dis* **2001**; 183:8–15.
4. Rousseau MC, Pereira JS, Prado JC, Villa LL, Rohan TE, Franco EL. Cervical coinfection with human papillomavirus (HPV) types as a predictor of acquisition and persistence of HPV infection. *J Infect Dis* **2001**; 184:1508–17.
5. Chaturvedi AK, Myers L, Hammons AF, et al. Prevalence and clustering patterns of human papillomavirus genotypes in multiple infections. *Cancer Epidemiol Biomarkers Prev* **2005**; 14:2439–45.
6. Mendez F, Muñoz N, Posso H, et al. Cervical coinfection with human papillomavirus (HPV) types and possible implications for the prevention of cervical cancer by HPV vaccines. *J Infect Dis* **2005**; 192: 1158–65.
7. Chaturvedi A, Katki H, Hildesheim A, et al. Human papillomavirus infection with multiple types: pattern of coinfection and risk of cervical disease. *J Infect Dis* **2011**; 203:910–20.
8. Plummer M, Schiffman M, Castle PE, Maucourt-Boulch D, Wheeler CM. A 2-year prospective study of human papillomavirus persistence among women with a cytological diagnosis of atypical squamous cells of undetermined significance or low-grade squamous intraepithelial lesion. *J Infect Dis* **2007**; 195:1582–9.
9. Franco EL, Villa LL, Sobrinho JP, et al. Epidemiology of acquisition and clearance of cervical human papillomavirus infection in women

- from a high-risk area for cervical cancer. *J Infect Dis* **1999**; 180:1415–23.
10. Qu W, Jiang G, Cruz Y, et al. PCR detection of human papillomavirus: comparison between MY09/MY11 and GP5+/GP6+ primer systems. *J Clin Microbiol* **1997**; 35:1304–10.
 11. van Doorn LJ, Quint W, Kleter B, et al. Genotyping of human papillomavirus in liquid cytology cervical specimens by the PGMV line blot assay and the SPF(10) line probe assay. *J Clin Microbiol* **2002**; 40:979–83.
 12. Klug SJ, Molijn A, Schopp B, et al. Comparison of the performance of different HPV genotyping methods for detecting genital HPV types. *J Med Virol* **2008**; 80:1264–74.
 13. Herrero R, Schiffman MH, Bratti C, et al. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste project. *Rev Panam Salud Publica* **1997**; 1:362–75.
 14. Bratti MC, Rodriguez AC, Schiffman M, et al. Description of a seven-year prospective study of human papillomavirus infection and cervical neoplasia among 10000 women in Guanacaste, Costa Rica. *Rev Panam Salud Publica* **2004**; 15:75–89.
 15. Castle PE, Schiffman M, Gravitt PE, et al. Comparisons of HPV DNA detection by MY09/11 PCR methods. *J Med Virol* **2002**; 68:417–23.
 16. Safaeian M, Quint K, Schiffman M, et al. *Chlamydia trachomatis* and risk of prevalent and incident cervical premalignancy in a population-based cohort. *J Natl Cancer Inst* **2010**; 102:1794–804.
 17. Larkin MA, Blackshields G, Brown NP, et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**; 23:2947–8.
 18. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **1994**; 22:4673–80.
 19. Bernard HU, Burk RD, Chen Z, van Doorslaer K, Hausen H, de Villiers EM. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. *Virology* **2010**; 401:70–9.
 20. Bouvard V, Baan R, Straif K, et al. A review of human carcinogens—part B: biological agents. *Lancet Oncol* **2009**; 10:321–2.
 21. Schiffman M, Clifford G, Buonaguro FM. Classification of weakly carcinogenic human papillomavirus types: addressing the limits of epidemiology at the borderline. *Infect Agent Cancer* **2009**; 4:8.
 22. Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analyses, 2nd ed. Boca Raton, FL: Chapman & Hall/CRC Press, **2004**.
 23. Vaccarella S, Franceschi S, Herrero R, et al. Sexual behavior, condom use and HPV: pooled analysis of the International Agency for Research on Cancer HPV Prevalence Surveys. *Cancer Epidemiol Biomarkers Prev* **2006**; 15:326–33.
 24. Xue X, Gange SJ, Zhong Y, et al. Marginal and mixed-effects models in the analysis of human papillomavirus natural history data. *Cancer Epidemiol Biomarkers Prev* **2010**; 19:159–69.
 25. Plummer M, Vaccarella S, Franceschi S. Multiple human papillomavirus infections: the exception or the rule? *J Infect Dis* **2011**; 203:891–3.