The known unknowns of HPV natural history

Patti E. Gravitt1,2

1Perdana University Graduate School of Medicine, Serdang, Malaysia. 2Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA.

The discovery that certain high-risk strains of human papillomavirus (HR-HPV) cause nearly 100% of invasive cervical cancer has spurred a revolution in cervical cancer prevention by promoting the development of viral vaccines. Although the efficacy of these vaccines has already been demonstrated, a complete understanding of viral latency and natural immunity is lacking, and solving these mysteries could help guide policies of cervical cancer screening and vaccine use. Here, we examine the epidemiological and biological understanding of the natural history of HPV infection, with an eye toward using these studies to guide the implementation of cervical cancer prevention strategies.

Introduction

The discovery that certain high-risk strains of human papillomavirus (HR-HPV) cause nearly 100% of invasive cervical cancer (1) has spurred a revolution in cervical cancer prevention. Randomized controlled trials (RCTs) have evaluated the efficacy of two prophylactic vaccines: Gardasil (also known as HPV4 or quadrivalent vaccine) targets two of the most carcinogenic HPV genotypes (HPV16 and HPV18), as well as two types responsible for more than 90% of anogenital warts (HPV6 and HPV11), and Cervarix (also known as HPV2 or bivalent vaccine) targets HPV16 and HPV18 alone. The results of the vaccine trials have been summarized in detail elsewhere (2–10), demonstrating nearly complete protection against anogenital warts (HPV6 and HPV11), and Cervarix (also known as HPV4 or quadrivalent vaccine) targets HPV16 and HPV18 alone. Despite the depth of our understanding of HPV and CIN natural history, two key parameters remain elusive: (a) do serum antibodies detected after natural infection confer protection against reinfection with the same HPV genotype? and (b) does a DNA-negative state represent complete virologic clearance or immunologic control of infection below detection limits of current HPV assays (i.e., viral latency)? Because of the uncertainty around these parameters, decision models often involve sensitivity analyses in which assumptions about natural immunity and, to a lesser extent, viral latency are varied, and most have found that the prediction from the models is highly sensitive to these assumptions (34–36).

Methodologic limitations continue to stymie our ability to provide unequivocal epidemiologic evidence to more precisely inform the decision models. A central limitation is the literal interpretation of HPV infection status based solely on DNA detection in exfoliated cervical cell samples. Typically, HPV DNA–positive women are considered “HPV infected” and HPV DNA–negative women are considered “HPV uninfected.” Thus, in natural history studies with repeated HPV DNA measures over time (usually at 4-, 6-, or 12-month intervals), women can be observed to have multiple transitions based on cervical DNA detection even in the absence of disease progression (37, 38). These transition states and the potential underlying complexity of the directly observed prospective DNA data are represented in Figure 2. For illustrative purposes, HPV DNA in this example represents HPV16 type–specific infection. A woman in state a is HPV uninfected and is sexually inexperienced. The transition from state a to state b (HPV16 DNA positive) is determined by the woman’s risk of sexual HPV exposure. Estimates from studies of young women initiating first sexual activity suggest that the initial a→b transition is common, with cumulative transition probability greater than 30% after 24 months of first sexual exposure (39). Natural history studies also suggest that 90% of women will transition from state b to state c (HPV16 DNA negative) within 2 years of initial HPV DNA detection (23, 40). These transitions are classically interpreted as incident and cleared HPV infection, respectively. Increasingly, natural history data demonstrate that women transition from state c back to type–specific DNA positivity (state d) (37, 41, 42). The greatest uncertainty surrounds the interpretation of this c→d transition. Typically, a woman in state c is considered to have cleared the previously detected HPV infection and is usually presumed to be uninfected in this state. However, due to the uncertainty surrounding our understanding of natural immunity against reinfection and HPV latency, state c can, in fact, reasonably represent at least three independent states: c1 represents a woman with prior infection who has completely cleared HPV and is now HPV unin-
review series

**Figure 1**
Natural history model of HPV and cervical cancer. HPV is acquired via sexual intercourse ("incidence"), but the majority of HPV is "cleared" within 2 years in most women. Approximately 60% of women with HPV DNA detected will develop serum antibodies against HPV (HPV seropositive), and if cellular samples are collected during peak viral production, mild cytologic abnormalities may be detected on Pap smears. A minority of HPV infections persist, and individuals with persistent high-risk HPV are at a substantial risk of developing cervical precancer, or CIN3. The CIN3 lesions are the targets of screening, because more than one-third of these will progress to invasive cervical cancer within 10–20 years. The dashed lines reflect the uncertainty in the natural history of HPV. Namely, it is unclear whether anti-HPV antibody developed following natural HPV infection protects against reinfection, and whether loss of HPV detection reflects virologic clearance or establishment of viral latency.

**Figure 2**
HPV DNA transition states in natural history studies. State a: sexually inexperienced, HPV uninfected; state b: HPV infected, HPV DNA positive; state c: HPV DNA negative; state c1: HPV uninfected, immune; state c2: HPV uninfected, susceptible; state c3: HPV infected, HPV DNA negative, immune; state d: new or recurrent HPV DNA positive. See text for further explanation.
is reasonable to hypothesize that the relative proportion of \( c_1 \) to \( c_3 \) decreases with increasing age, due to cumulative exposure of prior HPV infection concurrent with decreasing sexual risk behavior. The following sections will reexamine the epidemiologic and biologic evidence for protective antibody-mediated immunity resulting from natural infection and HPV latency, with careful consideration of the underlying \( c \)-state variability described above.

**Development of protective immunity following natural HPV infection**

Studies that have attempted to address whether serum antibodies developed following natural HPV infection show protection against reinfection have shown inconsistent results. Protection from reinfection among seropositive individuals has been detected in three studies of college-aged women (43–45) and in one of three population-based studies with average age greater than 30 years (22, 38, 46). Differences in the serological assay between studies may explain some of this inconsistency (22, 47). However, a recent analysis of HPV incidence among baseline seropositive versus seronegative women in the placebo arm of the trial of the quadrivalent vaccine Gardasil (Merck) in mid-adult women (Protocol 019) showed direct evidence of protection in younger, but not older, women (ref. 48 and Figure 3). Women aged 26–34 years showed a lower rate of new type-specific detection among the seropositive compared with seronegative women (1.0 versus 5.7 per 100 person-years, respectively), suggesting substantial protection against reinfection as was seen in the observational studies of college-aged women discussed above. However, the rate of new type-specific DNA detection among the older (aged 35–45) seropositive women was slightly higher than that among seronegative women of the same age (2.8 vs. 2.1 per 100 person-years, respectively), consistent with the overall lack of association between baseline serostatus and subsequent HPV DNA detection in the older cohorts. Do we conclude from these data that serum-detected antibodies in older women fail to protect against reinfection? Decision models have often considered waning HPV immunity with aging (49, 50); however, an alternative explanation for these data comes from a more careful consideration of the influence of the \( c_1 \rightarrow c_3 \) transition of undetectable to detectable DNA in a latently HPV-infected woman.

**HPV latency versus acquisition**

New sexual exposures and latent reactivation are often invoked as explanation for the second peak HPV prevalence observed in older women, particularly in Latin America (51) (Figure 4). Some epidemiologic studies report a strong association between new sexual partners and HPV incidence and conclude that sexual risk behavior drives most HPV infection, even in older women (26, 38).

For example, Trotter et al. (38) reported an increased risk of new HPV DNA detection among women reporting a new sexual partner both in the total cohort and in analyses restricted to women over age 40 years with more than two lifetime sexual partners. They reasoned that since they observed an increased relative risk of new HPV DNA detection in older women with recent new sex partners, latent reactivation was unlikely, because new sex partners should not be a risk factor for reactivation (32). Care must be taken, however, to avoid overinterpretation of a relative measure of association. A strong risk factor (such as a new sexual partner) will demonstrate an increased relative risk even in situations where that factor does not explain a majority of outcomes (e.g., there is a high risk of breast cancer among BRCA1 carriers, but these genetic mutations explain only a small proportion of all breast cancer incidence; ref. 52). In another large cohort study, population-attributable risk estimates based on exposure prevalence suggested that new sexual partners explained only 21% of HPV incidence (53). Another 21% was explained by increasing lifetime number of sexual partners (i.e., women most at risk for reactivation), and 12% was explained by decreased in vitro immune response to HPV virus-like particles or mitogen. These data support the view that new sexual exposures carry a high risk of HPV infection even in older women, but the proportion of infections that stem from sexual acquisition versus reactivation of prior infection are difficult to estimate.

The results from Merck’s mid-adult women study discussed above also support the notion that newly detected HPV infection is more likely attributable to reactivation in older women compared to younger women (48). In this study, the lower HPV incidence observed in older versus younger seronegative women suggests

**Figure 3**

Incidence of 4 HPV genotypes per 100 person-years among women in the placebo arm of the Gardasil vaccine trial in mid-adult women. All women in the analysis were DNA negative for the relevant vaccine HPV type at baseline. Blue, rate of new DNA detection among women HPV seronegative for the same HPV type; red, rate of new DNA detection among women HPV seropositive for the same HPV type. Adapted with permission from *Sexually Transmitted Diseases* (48).

**Figure 4**

Age-specific HPV prevalence in women with normal cytology from five world regions. Reprinted with permission from *Vaccine* (23).
an average decline in new sexual exposures with increasing age, as has been reported previously (54). It seems unlikely that the relatively higher HPV incidence in older seropositive versus younger seropositive women reflects a uniquely high sexual risk among the older seropositives. However, if there were an increased risk of reactivation of HPV in older women, only the women with previous HPV6, -11, -16, or -18 infections would be at risk for this event. If reactivation explained a higher proportion of incident events in the older women, it would be disproportionately manifest in the seropositive group.

HPV in immune suppressed populations

Some of the most compelling data supporting HPV latency have come from clinical observations and natural history studies of severely immune-suppressed women. Two large prospective cohort studies of HIV-infected and HIV-uninfected women, the Women’s Interagency HIV Study (WIHS) and the HIV Epidemiology Research Study (HERS), provided indirect evidence that a proportion of the increase in detectable HPV DNA at genital sites in HIV-infected women results from latent reactivation (42, 55). Strickler et al. (55) compared the rate of new HPV DNA detection in women who reported 18 months of sexual abstinence with the rate of new HPV DNA detection in sexually active women. New HPV DNA was detected in both HIV-uninfected and HIV-infected women who were sexually inactive, and risk of new HPV DNA detection increased in the HIV-infected women with increasing immune suppression to a high of 22% in the most severely immune suppressed (CD4+ T cell count <200/μl). Importantly, a non-negligible 5% of the HIV-uninfected and sexually inactive women had new cervical HPV DNA detection. A similar analysis comparing new HPV DNA detection in sexually active and inactive women was conducted in the HERS study (42). Here, recurrence and reactivation rates for HPV16 were 2.7 of 100 person-years and 3.3 of 100 person-years, respectively, and were nearly three times higher among HIV-infected compared with HIV-uninfected individuals (30). While residual confounding by sexual behavior may have been minimized in these studies, it cannot be completely ruled out. However, the strong association between HPV risk and increasing immune suppression supports a more direct biological effect of HIV infection on HPV natural history. In addition, we and others have shown that the increase in HPV prevalence and new DNA detection in HIV-infected women occurs early in HIV infection (56, 57). For example, in our study in Zimbabwe, a significant 5-fold increased risk of new HPV DNA multiple type detection was observed within 90 days of HIV acquisition (57). This observation is supported by increasing evidence that acute HIV infection causes massive memory T cell depletion in mucosal tissues (58–60). These data suggest not only a key role for immune memory in the control of HPV infections below limits of detection, but also suggest that the increased prevalence of HPV in the HIV-infected woman is not fully explained by shared sexual risks, since the effect was not observed in the six months preceding HIV acquisition in the seroconverting women (57).

The clinical observations of increased HPV-associated lesions following organ transplantation further support a role of immunologic control of lifetime persistent HPV infections. The increased risk of HPV-associated cancers in transplant populations is similar to that in HIV-infected populations (61). In addition, the appearance of fulminate warts in immune-suppressed individuals after a
long period of “dormancy” suggests that virus can remain in the body subclinically for long periods of time, presumably controlled by a competent memory immune response (62). Because transplant populations are not considered at high risk of sexually transmitted infection, these observations would certainly support an important role for immunologic competence in controlling HPV infection.

**Molecular evidence for a latent state of papillomavirus infection**

Despite the evidence described above, human studies are unable to directly demonstrate establishment of latency and induction of reactivation from the latent state. However, animal models of papillomavirus infection have provided sufficient experimental evidence for papillomavirus latency (63). A prototypical example is cottontail rabbit papillomavirus (CRPV), which infects the hairy skin epithelium of its host. Experimental inoculation of the skin of the rabbit with CRPV will result in papilloma formation, followed by spontaneous regression within 8–10 weeks (64). Persistence of viral DNA is common after papilloma resolution, but at a dramatically lower copy number (1 viral gene copy per 40–1,000 cells after regression versus 75 viral gene copies per cell in growing papillomavirus) (64). Furthermore, viral latency can be induced in the cottontail rabbit using low-dose CRPV inoculation. This latent infection is characterized by low-copy CRPV DNA detected in biopsies from clinically and histologically normal tissues with no papilloma formation (65). However, these latent infections were virologically competent; 45.8% of latent sites formed papillomas after mild skin irritation, and 10.5% of untreated sites spontaneously developed papillomas. These experimental data are analogous to the detection of HPV DNA in normal laryngeal epithelium in patients with recurrent respiratory papillomatosis (RRP) who are in remission (66, 67). These patients experience recurrent papilloma formation in the respiratory tract, presumably from an inability to control HPV6 or HPV11 infection in these tissues. Latent virus in the CRPV model and in RRP patients has also been shown to have a different gene expression profile (60, 68). For example, Zhang et al. (60) used the CRPV model and found that latent CRPV expressed E1 mRNA, but not the HPV early genes E6 and E7. These transcripts, which can act as oncogenes, were induced with UV light activation; 40% of latent sites had detectable E6/E7 mRNA, and 27% formed papillomas. Recent studies in the cottontail oral papillomavirus (COPV) model provide similar evidence for latency, though early viral gene transcription in the latent COPV sites included E6/E7 mRNA, while late viral protein expression was not detectable (63). The same study used laser capture microdissection to localize the site of latency to the epidermal stem cell. Taken together, results from the CRPV and COPV experimental models (along with analogous observations in RRP patients) show that (a) papillomavirus DNA can persist in the host at the site of infection at low copies after lesion resolution; (b) CRPV mRNA in latent infections is characterized by E1, but not E6/E7, expression; (c) latent CRPV infections can form papillomas following induction by UV light or mechanical irritation; (d) a small fraction of papillomas will spontaneously arise from latent sites without direct induction; and (e) the site of latency may be the epidermal stem cell.

Based on the CRPV studies, a model for cervical papillomavirus latency is proposed in Figure 5. HPV infects basal epithelial cells at sites of microtrauma, likely a normal consequence of sexual intercourse. Infected basal cells induced to differentiate to fill the wound will result in active papillomavirus infection, as has been previously described. A few infected basal stem cells will retain HPV episomes, but do not differentiate, and these infected cells are unlikely to be sampled using standard exfoliative techniques employed in most epidemiologic studies, which sample only the surface epithelium. Thus, HPV in a basal stem cell may remain undetectable until triggered to differentiate by undetermined stimuli such as wound repair and hormonal regulation (63, 69). With high cellular turnover in the cervical epithelium, one may expect relatively constant detection of HPV in this model. However, it is likely that a reactivated infection subsequent to latently infected stem cell differentiation is immediately recognized and brought under control by a functional memory immune response resulting in very short duration of active infection, as is the case with herpesvirus infection (70). This is consistent with a general paradigm shift away from the belief that recruitment of memory T cells from the peripheral blood or lymph node is required following acute challenge, toward a recognition that a very large number of effector memory T cells reside in peripheral tissues (71). Evidence from the CRPV model supports this notion. CD8+ T cells isolated from the epithelial site of latent CRPV infection have been shown to expand in culture following IL-2 induction, suggesting that memory T cells are resident in the epithelium, where they can quickly respond to periodic reactivation and reduce the duration of productive infection (64). This model would be consistent with the increase in new HPV DNA detection immediately following memory T cell depletion in acute HIV and in sexually abstinent women with chronic HIV infection (55, 57). The relatively lower rates of recurrent HPV DNA detection in healthy populations suggest that immunologic control is the rule, but that background reactivation is common and likely associated with short duration of detectable viral DNA, as has been shown to be the case for HSV-2 recurrence (72). This is supported by observations from the placebo arm of the Gardasil trial, which followed young women aged 18–26 years, where the average rate of recurrent type-specific DNA detection following two consecutive negative DNA swabs was 8% over 36 months (37).

**Implications of the proposed model of HPV latency**

More data will be required to fully evaluate the validity of this proposed model of HPV latency. Study designs with frequent cervical or cervicovaginal sampling in various populations (e.g., HIV-infected, pre/post-transplantation, perimenopausal women) will provide critical insight into the dynamics of HPV DNA detection that occur between the sampling intervals that drive our natural history models. For example, a study that has looked at HPV DNA detection in adolescent women with frequent sampling intervals (weekly vaginal self-sample) clearly demonstrated short-term variability in type-specific HPV detection (41).

Interpretation of determinants of HPV DNA detection in these and more traditional study designs must continue to acknowledge the underlying uncertainty in the c state following loss of HPV DNA detection (formerly referred to as clearance). When evaluating a possible c−→c+d transition, statistical methods must accommodate the fact that newly detected DNA resulting from reactivation is conditioned on prior infection. In addition, novel biomarkers, such as cervical cytokine profiling and full HPV genome sequencing, could be used to assist in epidemiologic causal inference in the absence of direct biomarkers of new versus recurrent HPV DNA.

Why should this uncertainty matter to the clinician faced with decisions about HPV vaccination and HPV-based cervical cancer screening in their patients? The benefit of vaccination in a sexually active woman is uncertain (20), and on a population level, HPV
vaccination with increasing time since sexual debut (correlated with aging) rapidly loses cost-effectiveness (73). This loss of cost-effectiveness is driven by a combination of reduced frequency of new sexual exposures with age and the lack of efficacy of vaccine in women already infected with vaccine-associated HPV types (54, 74, 75). The argument that on an individual level nearly all women will benefit from HPV vaccination, since less than 1% of women have evidence of prior infection to all four vaccine types, is overly simplistic (76). The actual benefit to an individual (defined as reduced risk of cervical cancer) will be a function of the combined probability of prior infection, as well as the risk of new acquisition, risk of persistence/disease progression, risk of missed lesions in screening, and risk of progression to cervical cancer. This conditional probability will approach zero for most women. Indeed, the efficacy reported for prevention of any grade of CIN in women without regard to prior infection in the Gardasil trial in mid-adult women was a nonsignificant 5.5% (95% CI: 19 to 25) (74, 77). With such minimal benefit, harm to vaccinating a woman with a current vaccine-associated HPV infection must be kept at practically nil.

What is the risk of CIN following a reactivated HPV infection compared with a new acquisition? Because we are unable to differentiate the e states, the answer is unclear. Since older women with newly detected DNA have risks for CIN2/3 progression similar to those of younger women (78), it is unlikely that reactivated infections pose a unique risk for disease. The more important clinical management issue related to the uncertainty regarding new acquisition and reactivation is the psychosocial impact. One need only a cursory internet search on “my HPV test came back positive” to understand the anxiety associated with HPV testing in women who consider themselves to be at low risk of sexually transmitted infections (e.g., women in long-term, monogamous relationships).

A message that repeated infections with types a woman harbored in her past “are largely explained by the acquisition of new sexual partners” (38) may inappropriately implicate sexual infidelity in the relationship. A stronger consensus in the HPV clinical and epidemiologic community about HPV latency is needed to ensure that women impacted by HPV-based cervical cancer prevention efforts and their primary care providers have accurate information to guide their decision making and counseling efforts. In summary, while we are privileged to have two wildly successful interventions to prevent invasive cervical cancer, implementation of these interventions will require intensive education for both the public and health care providers. When uncertainties about the natural history of HPV infection cloud the development of educational messages and policy necessary to ensure broad dissemination of these interventions, it is inappropriate to view the need for resolution of the remaining uncertainties as superfluous.

Acknowledgments
The author receives funding from the National Cancer Institute (R01 CA123467 and R21 CA156537). I thank Michelle Silver for her invaluable assistance in the preparation of the manuscript. I also thank Morgan Marks and Alison Liu for their helpful comments during the preparation of this manuscript.

Address correspondence to: Patti E. Gravitt, Perdana University Graduate School of Medicine, Block D1, MAEPS Building, MARDI Complex, Jalan MAEPS Perdana, 43400 Serdang, Selangor Darul, Ehsan, Malaysia. Phone: 6.03.8941.8646; Fax: 6.03.8941.7706; E-mail: pgravitt@perdanauniversity.edu.my.


http://www.ici.org


