Epidemiologic Evidence and Human Papillomavirus Infection as a Necessary Cause of Cervical Cancer

Eduardo L. Franco, Thomas E. Rohan, Luisa L. Villa

As with other malignant neoplasms, epidemiologic and laboratory studies conducted during the past 20 years have shown cervical cancer to be a disease with multifactorial causes and long latency. Unlike most other cancers, however, in which multiple environmental, biologic, and lifestyle determinants contribute independently or jointly to carcinogenesis, cervical cancer has been shown to have a central causal agent, human papillomavirus (HPV) infection (1–3), whose contribution to the risk of the disease is much greater than that of any other recognized determinant (4). On the basis of recent evidence from an international collaborative study (5) of more than 1000 cervical cancer specimens that used a highly sensitive polymerase chain reaction (PCR) protocol, researchers found that the prevalence of HPV DNA in cervical tumors was 93%. This is a higher estimate than had been observed previously in studies that used less meticulous methods for sample collection, preservation, and testing [reviewed in (4)]. Reanalysis of specimens that remained HPV negative revealed that HPV DNA could be detected in other portions of the same specimen or by use of PCR with different primers, thereby raising the prevalence to higher than 95% (5). Similar strategies have also been used to show the presence of HPV DNA in virtually all cases of cervical intraepithelial neoplasia (CIN) (6).

Walboomers and Meije (7) questioned the existence of HPV-negative cervical carcinomas and have argued that the use of the most recent generation of consensus–PCR protocols, which are highly sensitive, allows the identification of a wide spectrum of mucosotropic HPV types, thereby increasing the detection rate to virtually 100%. They proposed (7) that the occurrence of cervical cancer “without involvement of specific HPVs is exceptional or impossible.” Continuous expression of the viral E6 and E7 genes seems to be necessary for cervical carcinogenesis, with additional genetic changes being required to maintain the malignant phenotype (8).

Of the known causes and determinants of cancer, none is considered necessary or sufficient. The suggestion that HPV infection may be the first cause of a human cancer that has been shown to be necessary has obvious implications for primary and secondary prevention of this disease (9). Walboomers and Meijer (7) suggested that the answer to whether or not an HPV-independent causal pathway exists in cervical cancer may be provided by epidemiology.

In this commentary, we analyze the pitfalls of traditional epidemiologic approaches to distinguishing necessary from non-necessary causes of cancer, using as a specific example the role of HPV infection in cervical cancer. We argue that, in traditional epidemiologic designs, misclassification of cumulative exposure to HPV may make it impossible to use the magnitude of the relative risk (RR) estimates for the association between HPV and cervical cancer to differentiate between the necessary- and non-necessary-cause assumptions.

Effect of Misclassification

In theory, it can be expected that, with perfect exposure ascertainment, the RR will tend to infinity (because of a denominator of zero) if cervical cancer cannot arise from routes other than HPV infection. On the other hand, if the assumption that a cause is non-necessary is correct, the cumulative risk of CIN among women not exposed to HPV will not be negligible and, thus, the magnitude of the RR will be lower, all other conditions being held constant. The difference in the magnitude of RRs between these two situations is a key element in judging whether the results from traditional epidemiologic study designs (case–control and cohort studies with a single measurement of exposure) can be used to distinguish a necessary from a non-necessary cause.

Affiliations of authors: E. L. Franco, Departments of Oncology and Epidemiology, McGill University, Montréal, Canada; T. E. Rohan, Department of Population Health Sciences, University of Toronto, and Department of Oncology, McGill University, Montréal; L. L. Villa, Ludwig Institute for Cancer Research, São Paulo, Brazil.

Correspondence to: Eduardo L. Franco, Ph.D., Department of Oncology, McGill University, 546 Pine Ave. West, Montreal, PQ, Canada H2W 1S6 (e-mail: eduardof@oncology.lan.mcgill.ca).

See “Notes” following “References.”

© Oxford University Press
The two RRs (18.9 and 25.6) for the association HPV–CIN with misclassified exposure data in the above cohorts cannot be distinguished on the basis of their confidence intervals (CIs) because, in practice, large cohort studies of the size shown are rarely if ever feasible. A typical approach to measure the strength of the association is to conduct case–control studies nested within these base cohorts. This can be seen by randomly sampling 500 CIN case subjects and 500 control subjects from each of the cohorts, as simulated case–control studies. The resulting odds ratios, computed as estimates of the RRs, and reflect the association between HPV and CIN were calculated for all tables and were used to gauge the impact of misclassification on the ability to distinguish among causality scenarios.

RESULTS

Table 1 shows the two-way frequency tables from four hypothetical cohorts, assuming a 20% prevalence of HPV and a 50% cumulative risk of CIN among HPV-positive women. The first two cohorts were generated under the assumption that women unexposed to HPV infection have a 0.5% cumulative risk of CIN; i.e., HPV was assumed to be a non-necessary cause. These cohorts were obtained first on the basis of perfect HPV classification and then under a 10% bidirectional misclassification (i.e., 10% false positives and 10% false negatives, or sensitivity = specificity = 90%). A second set of cohort tables was produced by assuming that risk of CIN is nonexistent among HPV-negative women—in other words, by assuming that HPV is a necessary cause. True exposure status and misclassified exposure status were then assumed as for the first set of tables. The RR of CIN in the first cohort is 100 (50/0.5 or empirically, as [10000/20000]/[400/80000]) under correct exposure classification. With 10% misclassification, the RR estimate is reduced to 18.9 ([9040/26000]/[1360/74000]). Under the necessary-cause assumption, the nonmisclassified cohort has an RR of infinity (50/0), which is measured as an RR of 25.6 (9000/26000)/[1000/74000]) once 10% misclassification is assumed, with resultant rearrangement of the exposure information.

The two RRs (18.9 and 25.6) for the association HPV–CIN with misclassified exposure data in the above cohorts cannot be distinguished on the basis of their confidence intervals (CIs) because, in practice, large cohort studies of the size shown are rarely if ever feasible. A typical approach to measure the strength of the association is to conduct case–control studies nested within these base cohorts. This can be seen by randomly sampling 500 CIN case subjects and 500 control subjects from each of the cohorts, as simulated case–control studies. The resulting odds ratios, computed as estimates of the RRs, and respective 95% CIs were 28.5 (95% CI = 20.0–40.9) and 38.9 (95% CI = 26.5–57.3) for the non-necessary- and necessary-cause models, respectively. Both estimates are greater than the underlying RRs from the respective base cohorts because of the high frequency of the outcome (CIN) in both hypothetical populations. More important, however, is the fact that, even with such reasonably large case–control samples, each of the CIs included the other model's point estimates, indicating that in practice the two situations could not be distinguished by the traditional epidemiologic approaches. Only much larger case–control studies including 1000 CIN case subjects and 1000 control subjects would generate the level of precision required to differentiate between the RRs for the two causality assumptions in the examples shown in Table 1.

Table 2 portrays the effect of HPV misclassification of different degrees under both causality assumptions for two levels of cumulative risk among the exposed. The prevalence of HPV was assumed to be 20%, as with the previous analysis, and misclassification was based on equal false-negative and false-positive rates. RRs from 24 simulated cohorts are shown, including the four shown in Table 1 for comparison. As expected, RR estimates decrease substantially as a function of misclassification in all four combinations of conditional risk of CIN. With a 50% risk of CIN among the exposed, the magnitude of the RRs under the two causality scenarios becomes virtually indistinguishable at 15% misclassification, although for practical purposes this had already happened with 10% measurement error, as a result of the lack of precision of RR estimates obtained in case–control studies nested within these cohorts. At the lower (5%) level of CIN risk among the exposed, the magnitude of the relations allows the causality assumptions to be differentiated from each other. This is based, however, on an implausibly low cumulative risk of CIN among those exposed to HPV, assuming a moderate RR of 10 for the HPV–CIN relation, with no misclassification.

It is noteworthy that the magnitude of the RR under the necessary-cause assumption is invariant with respect to the risk of CIN among those infected with HPV. This is because there can only be false-negative HPV exposure among CIN case subjects but not false-positive exposure, since by definition HPV infection would have to be present in all cases of CIN under the necessary-cause assumption. This maintains the proportion of exposure constant among case subjects, irrespective of the risk of CIN following exposure.

The latter analyses assume “bidirectional” misclassification with equal rates of false-negative and false-positive exposure ascertainment and exposure prevalence fixed at 20%. A broader
overview of the impact of misclassification on the ability to distinguish between causality assumptions can be seen in Fig. 1, which shows how RRs vary in response to changes in sensitivity and specificity, separately, and for HPV exposure prevalence varying between 2% and 50%. Risk of CIN among the exposed is fixed at 50%, resulting in a baseline RR of 100, if HPV exposure ascertainment is free of error and we assume a non-necessary-cause relation (risk of CIN among the unexposed is 0.5%). At the implausibly low exposure prevalence of 2% (Fig. 1, top), there is less overlapping of the RR curves for the two causality assumptions, with good differentiation between the models. At an HPV prevalence of 10%, there is substantial overlapping of the curves (Fig. 1, middle), with the magnitudes of relations becoming comparable at sensitivity levels of 90% and lower (i.e., false-negative rates of \( \leq 10\% \)). A nearly complete loss of the ability to distinguish between causality assumptions occurs at the higher HPV prevalence of 50% (Fig. 1, bottom). Fig. 1 also shows that the lower the exposure prevalence, the less important the effect of losses in sensitivity in reducing RR estimates. Conversely, specificity takes a more important role at the lower prevalence levels and has an almost negligible effect on the magnitude of the HPV–CIN relation at the relatively high 50% HPV exposure.

**DISCUSSION**

In this series of models that uses both plausible and unrealistic conditions, we have shown that the magnitude of RRs obtained in traditional epidemiologic study designs of HPV and cervical cancer cannot be used to infer whether or not the exposure to HPV infection is a necessary cause of this neoplastic disease. Differentiation between these causal assumptions was possible only when models were based on implausibly low rates of cumulative HPV exposure or implausibly low cumulative

<table>
<thead>
<tr>
<th>HPV as necessary cause?</th>
<th>Cumulative risk, %, of cervical neoplasia</th>
<th>RRs according to degree, %, of misclassification†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Among HPV positive</td>
<td>Among HPV negative</td>
</tr>
<tr>
<td>No</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.5</td>
</tr>
<tr>
<td>Yes</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0</td>
</tr>
</tbody>
</table>

*Exposure to HPV arbitrarily defined as 20%.
†Ratio of risk among HPV-positive subjects to risk among HPV-negative subjects, assuming equal false-negative and false-positive rates in assessing exposure to HPV. No lesion misclassification is assumed.
‡Infinity.

Fig. 1. Relative risk (RR) estimates for the human papillomavirus–cervical intraepithelial neoplasia (HPV–CIN) association as a function of the degree of misclassification in HPV exposure, assuming a 50% cumulative risk of CIN among those exposed to HPV (all curves) and 0.5% or 0% among those unexposed, for non-necessary- and necessary-cause assumptions, respectively. Continuous lines: cohorts based on the assumption that HPV is not a necessary cause of CIN. Broken lines: cohorts based on the necessary-cause assumption for the HPV–CIN association. The three curves in each set represent different specificity levels: 99% (top), 90% (middle), and 80% (bottom). Each graph is based on a different prevalence of HPV exposure: 2% (top), 10% (middle), and 50% (bottom).
risks of CIN among the exposed women (near 5%). The latter value resulted in low RRs for the HPV–CIN association, unlike the level reported in real studies, which typically exceeds 10 (2–4). It should also be mentioned that all of the models were based on perfect outcome classification, a highly untenable assumption in studies of preinvasive cervical lesions. In practice, the effect of mismeasurement of CIN in cohort or case–control studies will be to blur further the distinction between the causal assumptions discussed here.

At relatively moderate levels of misclassification of exposure status, the distinction between necessary and non-necessary causal models is completely lost. The models presented here were mostly based on conditions of relatively high disease incidence to simulate the observed cumulative risk of CIN, a preinvasive cervical lesion that is the outcome of choice in cohort studies of the association between HPV and cervical neoplasia. The same conclusion would have been reached even if we had considered the rarer outcome of invasive cervical cancer. With cervical cancer as the outcome (data not shown), all RRs would have been based on rates with smaller numerators, but the magnitude of the relations would have been the same as that based on CIN, an outcome whose incidence rates are at least 10 times higher than those of cervical cancer in most populations.

Perhaps the most serious problem hampering the validity of much epidemiologic research on risk factors for cancer and other chronic diseases is the effect of measurement error in study variables. In relation to the detection of HPV, the use of PCR techniques and liquid-phase, immunocaptured hybridization have helped to eliminate the incoherence in results caused by the more severe misclassification of HPV status by the first generation of molecular epidemiology studies (10,11). The magnitude of RRs for the association between HPV and cervical cancer increased dramatically after the advent of the latter techniques [reviewed in (4)].

Despite the considerable improvement in laboratory techniques for the detection of HPV DNA, there is one important source of misclassification of HPV exposure status that cannot be readily corrected for by methodologic advances: that caused by the fluctuation in viral load over time. Many cases of HPV infection are transient. Among the first 1200 women who were enrolled in our ongoing cohort study in Brazil and who were tested by use of PCR techniques multiple times for HPV positivity, in only 30% of the participants was the same HPV type detected at enrollment and after 12 months. Conversely, acquisition of HPV infection was documented within 1 year for 20% of the women initially testing negative for HPV (12). Persistent infections were generally, but not always, associated with a high viral load and with oncogenic virus types. They were also more likely to result in high-grade CIN than were transient infections (13). It is clear, therefore, that collection of a single cervical specimen at the time of enrollment in a cohort study or at the time of diagnosis of CIN or of invasive cervical cancer in a case–control study provides little assurance that the laboratory determination of the HPV positivity of that specimen accurately reflects the relevant past exposure to HPV infection that the subject may have had. Infections with low viral load may be labeled erroneously as HPV negative, and a subject with a mildly productive transient infection at the time of testing will be classified as HPV positive in epidemiologic studies based on single-specimen assessment of exposure, regardless of whether the design is cohort or case–control. Such studies will also attribute exposure status to false-positive specimens resulting from contamination. The latter subjects’ unexposed status could be ascertained easily, were exposure to be determined on the basis of cumulative HPV detection in multiple specimens collected over time.

In practice, our inability to understand the possibly necessary causal role of HPV is further aggravated by misclassification of the outcome in cohort studies, which for ethical and practical reasons use preinvasive lesions as end points. Case–control studies of invasive cervical cancer are far less likely to be affected by outcome misclassification, but they are prone to differential exposure misclassification that combines two sources of errors. First, there is a higher cellular yield in a tumor biopsy specimen from a case subject than in exfoliated cells collected with a cotton-tip swab, cytobrush, spatula, or other devices from a control subject’s normal cervical os. This makes HPV exposure assessment among control subjects more likely to result in false negatives than that among case subjects, solely because of cer-
tical sampling bias. Second, the effects of fluctuation in viral load, transience of HPV infection, and other factors inherent to the dynamics of the infection make single testing for the virus less likely to represent past exposure for control subjects than for invasive carcinoma case subjects. The designation ‘‘HPV infection’’ relating to the presence of viral DNA in tumors no longer applies for the latter, since the viral genome is mostly—if not entirely—present in integrated form in cancer cells. To capture the actual exposure experience with the virus that led to cancer would have required sampling the case subject’s cervix at an earlier time when the infection was at a comparable (nonintegrated) state to that of the control subject. The biasing effects of these two errors are in the same positive direction away from the null hypothesis; i.e., they produce RRs that are higher than the one truly underlying the relation between HPV and cervical cancer in the same population.

Fluctuations in viral load and specimen cellularity may also affect the comparison of risk factor profiles between HPV-positive and HPV-negative CIN case subjects by influencing the false-negative rate in specific groups. Burger et al. (14) found that women with HPV-positive CIN had more sexual partners and tended to smoke cigarettes more than HPV-negative patients with CIN. It is conceivable that increased sexual activity with a plurality of partners and higher levels of cigarette smoking both may have facilitated the establishment of more productive lesions, which are less likely to be missed in a single testing opportunity. This argument does not serve to explain differences in clinical behavior between HPV-positive and HPV-negative invasive cervical carcinomas. A few studies, including our own, have found the latter to be associated with poorer survival of patients (15–18). Absence of HPV DNA in these tumors may be associated with low viral load or with loss of viral genomes due to the tumor’s own genetic instability. In any case, the difficulty in identifying the nature of these associations underscores the inability of single-specimen studies to unequivocally show that HPV-negative specimens are not the result of decreased detectability.

Most epidemiologic research on the natural history of HPV infection and cervical cancer has been based on only one measurement of exposure to the virus and its determinants or cofactors and on one measurement of cervical lesion end points. Case–control and cohort investigations have been instrumental in proving that HPV is the primary cause of cervical cancer by use of the approach of determining the baseline status for HPV and other factors and lesion outcomes either simultaneously, retrospectively, or prospectively. Statistical modeling by logistic and proportional hazards regression methods enhances the ability to probe associations in epidemiologic datasets by allowing control of confounding, assessment of interaction among variables, and stratification by design and matching variables and time between onset of exposure and outcome. However, behind the added level of insight that multivariate modeling brings to epidemiologic data analysis, the basic 2 × 2 table correlating HPV exposure and lesions remains the fundamental unit of information used to generate epidemiologic evidence for or against the causality of HPV. Unfortunately, this central 2 × 2 table is usually based on a single-specimen assessment of exposure, which combines the sampling and testing errors typical of one testing opportunity with those resulting from temporal fluctuations in detectability of HPV during the course of infection.

The traditional epidemiologic study designs of single-opportunity assessment of exposure and outcome are not suitable for addressing questions of viral persistence, fluctuation in viral load, regression of cervical lesions, and the dynamics of risk factor changes over time (e.g., acquisition of new sexual partners). To gain an understanding of the role of and mechanism for such dynamic changes in the natural history of the disease, one must conduct studies that collect data repeatedly on risk factors, HPV, and cervical lesions on multiple occasions during follow-up. A longitudinal, repeated-measurement cohort study is required to increase the accuracy and to reduce bias in the assessment of cumulative HPV exposure and outcome history. The longitudinal structure of the resulting datasets can be enormously complex and poses new challenges in data management and analysis (19,20). A number of such investigations have begun in recent years in different populations and include old and new laboratory markers of HPV infection, such as HPV typing (3), serologic response (21), determination of viral load (13,22), and the analysis of molecular variants to better define viral persistence (23,24).

Is there a subset of cervical cancers truly induced by carcinogenic routes other than HPV infection? Do HPV-negative cervical cancers perhaps reflect loss of the HPV genome during disease progression? As discussed in this commentary, unequivocal answers to these questions cannot be obtained by use of traditional epidemiologic study designs based on single assessments of the presence of HPV infection and cervical lesions.

References


NOTES

Partially supported by Public Health Service grant CA70269 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services; and by grant MA13647 from the Medical Research Council of Canada. E. L. Franco is a recipient of a Senior Scholar Award from the “Fonds de la recherche en santé du Québec.”

Manuscript received August 2, 1998; revised December 3, 1998; accepted December 31, 1998.