Test Reliability Is Critically Important to Molecular Epidemiology: An Example from Studies of Human Papillomavirus Infection and Cervical Neoplasia

Mark H. Schiffman and Arthur Schatzkin

Abstract

To demonstrate that it is critically important to achieve excellent test reliability before conducting full-scale molecular epidemiological studies, data were compared from two consecutive case-control studies of human papillomavirus (HPV) infection and cervical intraepithelial neoplasia. The major methodological difference between the two studies was the much greater reliability of the HPV test used in the second study. Although the first study used an assay considered state-of-the-art at that time, mediocre test reliability led to (a) a weakened association between HPV and risk of cervical intraepithelial neoplasia, (b) a weakened association between known risk factors for cervical intraepithelial neoplasia and HPV prevalence, (c) failure to demonstrate that HPV infection explains the known risk factors for cervical intraepithelial neoplasia, and (d) a marked reduction in the estimated proportion of cervical intraepithelial neoplasia attributable to HPV infection. With an improved assay, the second study strongly supported the idea that HPV infection is an intermediate end point explaining the known epidemiology of cervical intraepithelial neoplasia. Based on this experience and supportive theoretical considerations, we recommend that researchers optimize the reliability of innovative assays before application to full-scale molecular epidemiological projects.

Goal of the Article

In collaborative “molecular epidemiology” studies, the importance of measurement error is often undervalued by both laboratory scientists and epidemiologists. To illustrate this point, a comparison is presented of data from two consecutive case-control studies of HPV infection and cervical intraepithelial neoplasia. The comparison demonstrates how critical measurement error can be in molecular epidemiology, even when other epidemiological concerns have been addressed.

This article is intended for epidemiologists and laboratory scientists who are considering a collaborative molecular epidemiological study, incorporating newly developed, state-of-the-art laboratory assays into full-scale population studies. These collaborative studies are increasingly frequent in cancer research. Epidemiologists are eager to translate the powerful advances in understanding the molecular pathogenesis of cancer into assays that can be used in epidemiology and screening studies. Many molecular biologists wish to promote the clinical application of their advances.

Successful collaborative studies between molecular biologists and epidemiologists require a careful, joint consideration of methodological priorities, to avoid crushing the project with often competing demands deriving from the two disciplines. For example, should the study rely on more laborious, optimal testing techniques or can more rapid “shortcuts” be used? Should the project staff collect optimal biospecimens (e.g., biopsies) or those types of specimens more acceptable to the patient and clinician? Is it important to complete a large study to achieve good statistical “power” or is it wiser to assure quicker project completion and lower cost?

In our experience, the two most important epidemiological concerns in molecular epidemiological studies are assay reliability and the relatively unbiased selection of controls. This article will demonstrate the critical importance of assay reliability, defined as the ability of the assay to generate consistent, comparable results when applied to many clinical specimens tested over the course of an epidemiological project, which can take months or even years. The supportive examples will be drawn from two investigations of HPV and cervical intraepithelial neoplasia. The underlying supportive theory of the discussion has been discussed in several previous publications (1–3).

HPV Infection and Cervical Neoplasia

Some background details concerning this area of cancer research are necessary to make sense of the examples. Epidemiological studies have consistently observed associations of sexual factors with risk of cervical cancer and its preinvasive precursor lesion, cervical intraepithelial neoplasia (4, 5). The major sexual risk factors have been shown to be lifetime number of sexual partners and age at first intercourse, with lifetime number of sexual partners being the most important single factor (5). These epidemiological observations have motivated the search for a venerally transmitted causative agent. HPV infection was first suggested to be that central etiological agent by zur Hausen et al. (6), who used DNA hybridization methods to detect HPV types 16 and 18 in a small group of cervical cancer specimens. DNA hybridization methods remain the primary means of testing for HPV infection. More than 70 types of HPV have been defined, of which about 20 are found in cervical specimens. It is now generally accepted that most cervical cancer specimens contain DNA of HPV types 16, 18, or a few other types (7). Cervical intraepithelial neoplasia specimens generally contain either the cancer-associated types or other types (e.g., 6, 11, 42) not found in cancers.

The early case series of HPV infection and cervical cancer were small, had informally chosen control specimens, and did not attempt to assess confounding by other covariates (8). However, these early studies used relatively accurate DNA hybridization analyses of tumor biopsies and derived the correct answer that HPV infection is strongly associated with risk of cervical cancer (later extended to intraepithelial neoplasia as well).

When epidemiologists sought to confirm the results of the early case series in rigorously controlled population studies, it was not possible to rely on HPV testing of biopsies, because of the ethical restriction on biopsying nondiseased women. There followed a several-year period when HPV measurement technology was first adapted to permit the more rapid testing of scant, noninvasively obtained cervical specimens collected at the time of routine gynecological examination by scrape or lavage. Many different combinations of specimen collection and HPV-testing methods were used (9). Interlaboratory comparisons of these varying strategies were generally discouraging, indicating poor assay reliability (10). We conducted our first case-control study of HPV infection and cervical intraepithelial neoplasia during this developmental period, using HPV-testing methods thought to be state-of-the-art at the time but soon proving to be inaccurate (11).

Improved HPV-testing methods were developed and validated after the conclusion of the first project (12, 13). In a second case-control study of HPV infection and cervical intraepithelial neoplasia (14), with a study design virtually identical with the first project but with an improved HPV-testing method, we redressed the same analytic questions that had motivated the original effort (15). We reasoned...
that, if HPV infection is the key venereal agent that explains the association of sexual behavior with risk of cervical intraepithelial neoplasia, the following should be true: (a) HPV infection should be strongly associated with risk of cervical intraepithelial neoplasia; (b) the risk of HPV infection, in turn, should be associated with the same measurements of sexual behavior known to influence the risk of intraepithelial neoplasia; (c) the often-observed associations of the sexual variables with risk of cervical intraepithelial neoplasia should be explained by HPV infection, i.e., statistical adjustment for HPV infection should eliminate the association of intraepithelial neoplasia risk with sexual behavior; and (d) the attributable proportion of cervical intraepithelial neoplasia related to HPV infection should be high, if infection truly is a key intermediate end point on the causal pathway to cervical intraepithelial neoplasia (1). As shown below, the second case-control study confirmed all of these hypotheses (14).

Based on a growing body of epidemiological and laboratory evidence, HPV is now widely accepted to be a central etiological factor for cervical neoplasia (16).

The point of this article is to show how moderate measurement error during the first study severely limited our ability to study the epidemiology of HPV and cervical intraepithelial neoplasia for each of the four statistical points listed above. The article will show that the epidemiological methods of the two studies were comparable, discuss the repeatability of the HPV measurements used in each study, and contrast the results obtained in the two projects. Finally, a few implications of the comparison will be discussed.

Materials and Methods

We conducted the first case-control study of HPV infection and cervical intraepithelial neoplasia in 1986–1987 at three Washington, DC, area hospitals. The design has been discussed in detail elsewhere (17). In brief, the case-control study was nested within a large cross-sectional screening of 2820 women receiving routine cervical cytological (Pap smear) screening in 13 different obstetrics and gynecology clinics. We successfully recruited 85% of eligible women. The Pap smear diagnosis was used to classify subjects as controls (normal or benign reactive changes, n = 2517) or cases (276 subjects with low-grade intraepithelial neoplasia and 27 with high-grade intraepithelial neoplasia were combined for this presentation). To measure HPV infection, a 3-ml cervicovaginal lavage was tested by Southern blot DNA hybridization techniques. Because of the expense of the assay, we tested a sample of the subjects, including 269 (89%) of the cases, and controls matched 2:1 to the cases on age group (± 5 years), race, clinic, and appointment date. We subsequently excluded from the analysis all controls with a medical history of cervical intraepithelial neoplasia or cancer, leaving 400 controls in the analytic data set.

Recruitment for the second case-control study was conducted in 1989–1990 at seven Kaiser-Permanente obstetrics and gynecology clinics in Portland, OR (14). About 22,000 women receiving routine Pap smear screening were recruited, and participation rates among eligible women approached 95%. Again, cervical cytological diagnoses were used to define subjects as cases of cervical intraepithelial neoplasia or cancer. For the second case-control study, we tested all of the high-grade cases of intraepithelial neoplasia (n = 50), and we selected randomly 450 cases of low-grade intraepithelial neoplasia (72% of the total) to complete a 500–woman case group. For controls, we selected randomly a 500–woman (3%) sample of the 17,824 women in the screening with normal cervical cytological diagnoses and no known medical history of cervical intraepithelial neoplasia or cancer.

The same consultant cytopathologist collaborated on both case-control studies; thus, the case definitions were comparable in the two studies. However, the HPV-testing methods were much different. In the second study, we used an improved 10-ml cervicovaginal lavage to collect more uniformly adequate DNA specimens. The specimens were tested by an L1 consensus primer polymerase chain reaction technique, developed by Manos et al. (12). This polymerase chain reaction method amplified DNA from most of the 70 known (and some still unidentified) HPV types, permitting sensitive detection of HPV in minute quantities of clinical specimen. In advance of the study, the polymerase chain reaction method compared favorably in an interlaboratory experiment (13) to an optimal Southern blot DNA hybridization (not the flawed Southern blot technique used in the first study).

In summary, the major methodological difference between the two projects involved the exposure measurement (HPV test). Other details were quite similar, with two possibly noteworthy exceptions. First, case-control matching was used in the first study alone. However, statistical control for the same factors (age, ethnicity, and clinic) in the second project did not change the conclusions, suggesting that this methodological difference is not likely to be crucial to the point being made here. Second, the population in the first study was younger and poorer than in the Portland project, leading to a higher true HPV prevalence among controls (18, 19) and reduced crude relative risks associating sexual behaviors with risks of HPV infection and cervical intraepithelial neoplasia. This second, more important caveat is discussed further below.

Results

Repeatability of HPV Tests in the Two Projects. Quality control repeat specimens from the first case-control study demonstrated mediocre repeatability of the HPV-testing procedure (Table 1). Concurrently, we observed poor interlaboratory agreement in a comparison of four laboratories (including our study collaborator) performing the same type of HPV test on identical aliquots of specimen DNA (10). There was interlaboratory disagreement regarding the presence of HPV infection in 14 (35%) of the 40 test specimens. Additional disagreements as to HPV type were substantial. We could not directly assess the accuracy of the DNA hybridization method we were using, because there was no reference standard for HPV DNA detection. We inferred, from the poor intra- and interlaboratory repeatability of the method, that the test was producing substantially misclassified and suspect data.

The intralaboratory repeatability of the polymerase chain reaction method used in the second case-control experiment was found to be much better, although it must be noted that masked repeat specimens in Portland were tested in the same batches, while in the DC project, repeat specimens were run in separate batches. This difference biases the repeatability data to apparently better repeatability in Portland. With this caveat, the differences are striking (Table 1). In addition, in the preliminary methodological pretest comparing the method to optimal Southern blot hybridization (the most trusted test for HPV typing), all specimens classified as HPV positive by Southern blot hybridization were also HPV positive by polymerase chain reaction, and agreement regarding HPV type was nearly complete (13). Based on these data, we concluded that the data generated in the second case-control study were likely to be much less misclassified than in the previous study. No direct comparison of the testing methods from the two studies has been performed.

Results from the Two HPV Studies. The comparisons of the two studies' results are shown in Tables 2–4. In Table 2, the odds ratios are shown from each of the two studies for the associations of HPV...
This reexamination of the DC population has demonstrated a sample of the DC control population, using the newer HPV method study, in accordance with three other large data sets published at Table 2, were also dramatic. In the first study, for example, the odds low socioeconomic status, oral contraceptive use, and young age. several additional univariate risk factors for HPV infection, including reaction test method (21). In the second study, moreover, we observed results of another recent analysis using the same polymerase chain expected strong association of sexual behavior and HPV infection was intraepithelial neoplasia.

approximately the same time that also suffered from testing error (3). No other risk factors for HPV infection were found. In contrast, the HPV infection and lifetime number of sexual partners, the main sexual risk factor for cervical intraepithelial neoplasia. Only a weak associ invasive neoplasia.

Table 2 Odds ratios associating HPV infection with risk of cervical intraepithelial neoplasia in two case-control studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Controls</th>
<th>Cases</th>
<th>Odds ratio (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>%</td>
<td>No.</td>
<td>%</td>
</tr>
<tr>
<td>Case</td>
<td>1986-1987</td>
<td>HPV negative</td>
<td>279</td>
</tr>
<tr>
<td>HPV positive</td>
<td>74</td>
<td>18.5</td>
<td>125</td>
</tr>
<tr>
<td>Inadequate sample</td>
<td>47</td>
<td>11.8</td>
<td>17</td>
</tr>
<tr>
<td>Total</td>
<td>400</td>
<td></td>
<td>269</td>
</tr>
<tr>
<td>Case</td>
<td>1989-1990</td>
<td>HPV negative</td>
<td>375</td>
</tr>
<tr>
<td>HPV positive</td>
<td>80</td>
<td>16.0</td>
<td>381</td>
</tr>
<tr>
<td>Inadequate sample</td>
<td>45</td>
<td>9.0</td>
<td>30</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td></td>
<td>500</td>
</tr>
</tbody>
</table>

* Most inadequate samples resulted from problems in collection, not testing.

Table 3 Odds ratios associating HPV infection and lifetime number of sexual partners in the control groups of two studies of cervical intraepithelial neoplasia

<table>
<thead>
<tr>
<th>Lifetime no. of sexual partners</th>
<th>No. of subjects</th>
<th>% HPV positive</th>
<th>Odds ratio (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>1986-1987</td>
<td>1</td>
<td>64</td>
</tr>
<tr>
<td>2</td>
<td>53</td>
<td>20.8</td>
<td>2.1 (0.8-6.0)</td>
</tr>
<tr>
<td>3-4</td>
<td>81</td>
<td>25.9</td>
<td>2.8 (1.1-7.2)</td>
</tr>
<tr>
<td>5-9</td>
<td>71</td>
<td>25.4</td>
<td>2.8 (1.1-7.2)</td>
</tr>
<tr>
<td>10+</td>
<td>75</td>
<td>18.7</td>
<td>1.9 (0.7-5.0)</td>
</tr>
<tr>
<td>Study 2</td>
<td>1989-1990</td>
<td>1</td>
<td>107</td>
</tr>
<tr>
<td>2</td>
<td>55</td>
<td>16.4</td>
<td>6.8 (1.8-26.2)</td>
</tr>
<tr>
<td>3-5</td>
<td>109</td>
<td>22.0</td>
<td>9.8 (2.8-33.6)</td>
</tr>
<tr>
<td>6-9</td>
<td>66</td>
<td>28.8</td>
<td>14.0 (4.0-49.7)</td>
</tr>
<tr>
<td>10+</td>
<td>68</td>
<td>25.0</td>
<td>11.6 (3.2-41.2)</td>
</tr>
</tbody>
</table>

The attributable proportions of cervical intraepithelial neoplasia explainable by HPV infection were 36.2% for the first study and 77.1% for the second. The attributable proportions were calculated according to the formula:

\[
\text{Attributable proportion} = \frac{\% \text{ HPV positivity among cases} \times 1 - \frac{1}{\text{Relative risk}}}{\text{Relative risk}}
\]

The conclusions drawn from the two studies, based on these statistics, would clearly be different. In the first study HPV infection would appear to be a risk factor of some importance, but not the central etiological agent. Based on the second study, a central, causal etiological role appears more likely. Of note, the prevalence of HPV infection in cases and the attributable risk increased even further, to near 90%, when misclassification of the disease end point was reduced by expert pathology panel reviewers unaware of the HPV infection status (14).

Discussion

First, to mention the limitations of the comparison, the data would be even more convincing if we had been able to retest the exact specimens from the first study using the newer HPV test. Some of the differences in results could be explained by differences in the two study populations. Specifically, the Portland population is at much lower risk than the DC population for cervical neoplasia, and the

Table 4 Effect of adjustment for HPV infection on the association between lifetime number of sexual partners and risk of cervical intraepithelial neoplasia

<table>
<thead>
<tr>
<th>Lifetime no. of sexual partners</th>
<th>Cases</th>
<th>Controls</th>
<th>Crude odds ratio (95% confidence interval)</th>
<th>Adjusted odds ratio* (95% confidence interval)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
<td>1986-1987</td>
<td>1</td>
<td>25</td>
<td>69</td>
</tr>
<tr>
<td>2</td>
<td>47</td>
<td>61</td>
<td>2.1 (1.2-3.9)</td>
<td>2.2 (1.2-4.0)</td>
</tr>
<tr>
<td>3-4</td>
<td>69</td>
<td>90</td>
<td>2.1 (1.2-3.7)</td>
<td>2.0 (1.1-3.6)</td>
</tr>
<tr>
<td>5-9</td>
<td>71</td>
<td>79</td>
<td>2.5 (1.4-4.3)</td>
<td>2.4 (1.3-4.3)</td>
</tr>
<tr>
<td>10+</td>
<td>48</td>
<td>89</td>
<td>1.5 (0.8-2.7)</td>
<td>1.5 (0.8-2.8)</td>
</tr>
<tr>
<td>Study 2</td>
<td>1989-1990</td>
<td>1</td>
<td>40</td>
<td>113</td>
</tr>
<tr>
<td>2</td>
<td>34</td>
<td>58</td>
<td>1.7 (0.9-2.9)</td>
<td>1.0 (0.5-1.9)</td>
</tr>
<tr>
<td>3-5</td>
<td>127</td>
<td>116</td>
<td>3.1 (2.0-4.8)</td>
<td>1.1 (0.6-1.9)</td>
</tr>
<tr>
<td>6-9</td>
<td>116</td>
<td>70</td>
<td>4.7 (2.9-7.5)</td>
<td>1.5 (0.9-2.7)</td>
</tr>
<tr>
<td>10+</td>
<td>116</td>
<td>74</td>
<td>4.4 (2.8-7.0)</td>
<td>1.6 (0.9-2.8)</td>
</tr>
</tbody>
</table>

* Adjusted for HPV DNA detection.

Excludes subjects with missing questionnaire or HPV data.
Portland women report fewer average lifetime numbers of sexual partners. This difference in populations is not negligible but is unlikely to explain the major divergence of the two studies. The recent testing of specimens from previously untested controls in the Washington, DC, population, using the new HPV method, demonstrated epidemiological associations previously missed there. Moreover, a recent study confirmed the central role of HPV infection in explaining sexual risk factors for cervical neoplasia, in both a low-risk (like Portland) and a high-risk (like the Washington, DC) population (22).

Although consecutive studies can never be compared with absolute certainty of comparability, the data strongly indicate the importance of test reliability in molecular epidemiological studies. This point has been proven more broadly by the entire field of HPV epidemiology, which has made progress in parallel with advances in HPV-testing methods (20). Thus, it is worthwhile to discuss a few of the possible implications of the comparison presented here.

First, the comparison demonstrated that even moderate amounts of misclassification of dichotomous variables, a category that includes many biomarkers and screening tests, can dramatically affect epidemiological associations (3, 23). What true relative risks underlie the odds ratios of 1.5 and 2.0 that we epidemiologists routinely find using other assays (or even questionnaires) that may prove as inaccurate as the HPV test used in the first study?

Epidemiologists might do well to focus on measurement error as a primary issue when they critique ground-breaking clinical studies that use state-of-the-art measurement techniques. We epidemiologists tend to criticize early clinical studies by listing a set of predictable common concerns, such as subtler biases of design, statistical power calculations, and proper analytic methods, are more minor issues in molecular epidemiology, which focuses typically on detecting new associations with high relative risks.

**References**

Test Reliability Is Critically Important to Molecular Epidemiology: An Example from Studies of Human Papillomavirus Infection and Cervical Neoplasia

Mark H. Schiffman and Arthur Schatzkin


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/54/7_Supplement/1944s

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/54/7_Supplement/1944s.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.