

# Chapter 18: Statistical Issues in the Design and Analysis of Studies of Human Papillomavirus and Cervical Neoplasia

Sholom Wacholder

---

**Appropriately sophisticated statistical approaches are crucial for addressing the increasingly complex set of critical questions that follow from the recognition that human papillomavirus (HPV) is a necessary causal factor for cervical cancer. Cervical cancer researchers have defined the major stages of cervical carcinogenesis, with HPV infection as the necessary cause. Focus of etiologic studies is shifting from establishing causality to determining risk factors for HPV persistence and neoplastic progression using serially collected biomarkers. Prevention-oriented epidemiology and trials of new screening strategies and vaccines will rely on surrogate endpoints because we cannot let women develop cancer when it can be prevented. Future epidemiologic and prevention studies of HPV infection and cervical carcinogenesis will exploit subtle pathologic distinctions and will employ improved measurements of complex molecular biologic phenomena. The anticipated statistical issues are highlighted in this discussion. [J Natl Cancer Inst Monogr 2003; 31:125–30]**

---

The simple contingency table in which a disease (yes/no) is related to an exposure (present/absent) is the fundamental data representation in epidemiology. The advancing knowledge of cervical carcinogenesis demonstrated in the other chapters in this monograph shows, however, that human papillomavirus (HPV) infection and cervical neoplasia are not simple dichotomies and the precise classification boundaries are debatable; that each has errors that may depend on the other; that each can be transient; and that surrogate outcomes, not invasive cervical cancer, must be used as the endpoint in prospective studies. These difficulties permeate this chapter, where some of the most important statistical issues facing new studies of HPV and cervical neoplasia are noted, with an expanded discussion of a few. Issues are classified according to the same themes represented in the other chapters: etiology and prevention.

## ETIOLOGY

### Defining Exposure to Oncogenic HPV

There are approximately 15 oncogenic types of HPV that are established as separate causal exposures (1); most of the rest of the more than 40 types of HPV that infect the cervix can be ignored in the study of cervical cancer. Within the HPV type, moreover, variants with relatively minor polymorphic variation may also affect risks of neoplasia (2). HPV infections of type 16, the predominant type, are common enough to be studied separately; however, each of the remaining oncogenic types is much less common. We do not yet know the optimal strategy for deciding whether even to include a given type as oncogenic, especially if it is rare and its effect on risk is much less than that of many of the ones already labeled as oncogenic. Nor do we

know whether and how to consider each or subsets of each type or variant separately.

### Multiple Types of Infection

Approximately 20–30% of infected women have multiple types detectable (3). The presence of any pair in a woman is positively correlated because of a common, sexual route of transmission. Although understanding the joint effects of infection with multiple types will be critical to natural history studies and to planning of trials of vaccines that do not target all types, no prospective studies to date are large enough to fully analyze possible effect modification by types.

### Transience and Recurrence of Infection

Even if there were only one HPV type, it would still be difficult to define incidence and prevalence of HPV infections, which are typically transient and can recur quickly. The reliable reference standard of molecular testing for prevalent HPV-DNA in expert laboratories unavoidably classifies many women as falsely positive or negative, compared with the ideal parameters of infection, such as persistent infection or any exposure during her lifetime, that we often wish to estimate. One problematic example of false positives is a consequence of sensitive and specific molecular techniques that detect HPV-DNA at a level that is too low to have clear clinical significance and is only weakly associated with prevalent or incident disease (4). Classifying lifetime exposure using available HPV-DNA or serologic measurements is certain to generate false negatives (*see* chapter 12, Iftner and Villa).

### Non-Molecular Classification

The clinical and microscopic signs of HPV infection are misclassified more often than the molecular tests. HPV infection is often not apparent upon visual examination of the cervix, including colposcopy or even upon cytologic sampling (Pap tests) (5).

### Impact of Misclassification

The profound impact of misclassification on studies of HPV infection is well documented (6,7), with statistical lessons for other measurements (e.g., nutrition) that cannot be as easily improved. It would be a useful exercise to document more fully

---

*Correspondence to:* Sholom Wacholder, Ph.D., Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, EPS 8046, Bethesda, MD 20892-7244 (e-mail: wacholder@nih.gov).

*See "Notes" following "References."*

*Journal of the National Cancer Institute Monographs*, No. 31, © Oxford University Press 2003; all rights reserved.

the impact of varying degrees of misclassification in an area where the underlying association is so profound.

### *Newer Assays*

As an added layer of complexity in measuring exposure to HPV, researchers are incorporating novel, complex assays such as viral expression micro-arrays, and topographic mapping of specific lesions by position on the cervix. This work and similar efforts in other areas are in the field, without solutions to important statistical issues in hand.

### **Defining Disease Endpoints**

Clinically, cervical cancer is known to arise from identifiable precursors (cervical intraepithelial neoplasia or CIN). Precursors are broadly divided into low-grade (CIN1, the histopathologic evidence of HPV infection) and high-grade (CIN3, including carcinoma *in situ*). In some countries, including the United States, CIN2 is defined as an intermediate diagnosis that is treated for safety as if it were CIN3, even though CIN2 describes a heterogeneous category of lesions that are more likely than CIN3 lesions to regress.

### *Transitions*

Determination of the factors associated with transitions between the stages of cervical carcinogenesis can enhance understanding of its natural history; the distinctions between adjacent stages are unavoidably much more prone to misclassification than the distinction of cancer from normal controls. HPV also causes many equivocal changes that can be confused with benign inflammatory and reactive processes, and these equivocal changes are recognized more commonly than clear-cut CIN on Pap smears. Accordingly, misclassification of cytology and histopathology is a big problem (8,9). For example, in studying risk factors for progression of HPV infection to CIN2 and CIN3, identification and exclusion of CIN2 is sometimes not feasible; consequently, CIN2 is often included in the case group, even though some cases represent only acute HPV infection and not precancer.

### *Differential Misclassification*

The definitions and classifications of HPV exposure and cervical disease are so closely related that the probability of misclassification of one depends on whether the other is present. New tools for addressing differential misclassification of disease by exposure and of exposure by disease are needed, particularly when studying certain viral risk factors like the effect of viral load, itself linked to both ease of molecular detection and presence of diagnosable disease (8–10).

### *Terminology*

Finally, in comparing studies from around the world it is important to realize that diagnostic terms differ by geographic region. The same names are applied to different entities (e.g., CIN1 is more stringently and less sensitively diagnosed in Scandinavia than in the United States) and different terms to the same entities (e.g., the names for minor cytologic changes are not strictly comparable or even directly translatable between the United States and many other countries) (11). A calibration study where a team of experts apply a generally accepted defi-

nition to stratified, randomly selected slides from different studies can mitigate this potential problem for pooled analyses.

### **HPV as the Necessary Cause of CIN and Cancer**

#### *Cofactors*

Virtually every case of cervical cancer and CIN3 contains oncogenic HPV. Given the enormous relative risk in the years after even a single sensitive and specific HPV measurement, the etiologic fraction of cervical cancer and CIN3 HPV approaches 100%. Therefore, there are no other “independent” risk factors, only cofactors that modify risk among HPV-infected women. Because the HPV-negative stratum contains almost no true cases, the usual approaches to confounding and effect modification, including Mantel–Haenszel, stratification and standard modeling, do not apply.

#### *Timing*

The assessment of the effects of cofactors is most informative when careful attention is paid to the timing of effect and relation to the primary role of HPV. Is a putative risk factor for neoplasia instead only associated with acquiring HPV infection, perhaps by sexual behavior or by reducing immune response? Does presence of the putative risk factor increase the likelihood of persistence or progression after infection occurs? For example, HPV infection is the causal intermediate for the best known epidemiologic risk factor for cervical cancer, “lifetime number of sexual partners,” which is not an independent risk factor (12). However, other factors such as smoking, multiparity, and long-term oral contraceptive use are evidently both related (perhaps not causally) to the chance of becoming infected and cofactors for progression (13–15). Less firmly established cofactors, such as other sexually transmitted infections, chronic inflammation, or dietary ones, are plausibly related to HPV infection. Clarifying the possible dual role of these agents in acquiring infection and in subsequent cervical carcinogenesis will require carefully designed natural history studies and analyses informed by biologic reasoning.

#### *Misclassification by Cofactors*

Because the association between HPV infection and risk of cancer is so strong, it is critical to evaluate the possibility of misclassification of HPV infection or cytologically defined disease by other variables. For example, smoking might be an indicator of risk of a subsequent oncogenic HPV infection that leads to cancer in a cohort of initially HPV infected women, even though the infection detected at baseline itself never progressed. Moreover, some determinants of progression, such as oral contraceptive use, may influence the appearance of lesions and perhaps increase cytologic sensitivity (15).

#### *Summary*

Overall, statistical modeling in etiologic studies of HPV and multi-stage cervical carcinogenesis is so complex that rote approaches must be abandoned.

### **Descriptive Data**

SEER<sup>1</sup> incidence, American Cancer Society region-specific or race-specific mortality data, and international comparisons can be useful in showing where aggressive intervention in in-

creasing rate of effective Pap screening is likely to be worthwhile. One must be careful, however, not to try to infer too much about the etiology of cervical cancer from ecologic studies because incidence and mortality are highly dependent on screening and censoring treatment of precursor lesions. Differing patterns of age-specific type-specific HPV prevalence in population-based screening studies are of great interest but cannot be understood without consideration of female and male sexual patterns, effects of aging on immunity, and the possibility of birth-cohort effects. Although we do not have enough events yet to perform age-period-cohort analyses in our large cohort studies, we expect that within the next several years such analyses will be helpful in suggesting possible sources of regional differences in age-specific rates.

### Case–Control Studies

The choice of cases in case–control studies of HPV and cervical carcinogenesis is not conceptually difficult. The key consideration is that case definition be appropriate to the question being asked. For example, to determine risk factors for HPV infection, incident cases of infection among women recruited initially as virgins would be an optimal choice because the rate of infection is high immediately after the beginning of sexual activity. To study risk factors for progression, women with newly diagnosed CIN3 or cancer might be the best choice; it may not be essential in these studies to test the cases for HPV, given that the rare test negatives are more likely to be false negatives than true negatives.

The proper choice of controls and analytic strategy, however, are conceptually difficult problems, especially when the goal is to determine risk factors for progression (*see* chapter 2, Schiffman and Krüger-Kjaer). A crucial question is whether the oncogenic-HPV negative controls can contribute to analyses of progression, or whether all analyses should be restricted to cases and HPV-positive controls. Using only controls that are oncogenic HPV positive at the time of selection is itself problematic because we cannot determine how long they have been infected. Although HPV infections are most prevalent among young women beginning sexual activity and tend to disappear within 2 years, most women diagnosed with CIN3 or cancer are approximately 30 years of age or in their 40s, respectively, and have probably been infected for many years. Thus, in a case–control study without previously collected biospecimens, it is impossible to determine the controls' HPV status at the time the putative cofactor would affect risk. Serology assays to define HPV exposure may be helpful, but they are currently not sensitive for any lifetime infection, and neither time at acquisition nor duration of infection can be made precise (*see* chapter 5, Wang and Hildesheim). Effect estimates will be distorted by association of cofactors with acquisition or persistence of infection, which seems likely for variables like using oral contraceptives or smoking.

#### *Consequences of One-Time HPV Exposure Ascertainment in Controls*

The key point is that the effects on risk of time since infection and duration of infection are not estimable from a standard case–control design with one measurement of HPV. For example, suppose we want to calculate an odds ratio that retrospectively approximates a prospective relative risk from a theoretical cohort of women newly infected with HPV, i.e., to estimate the

relative risk faced by a woman newly diagnosed with an oncogenic type of infection in a clinic. An age-matched case–control design will yield an odds ratio that overestimates the risk from being HPV infected on one occasion in the past because many controls called *unexposed* will have already cleared a transient infection by the time of testing.

### Cohort Designs

#### *Serial Measurements*

Because single time-point measurements of HPV exposure cannot be used to estimate the prospective relative risk of cervical cancer associated with HPV infection, there is a need for longitudinal studies that include repeated or serial measurements of HPV and disease outcomes, in order to capture changes in exposure and disease progression/regression. Generalized estimating equations, multistage, and Markov models have been used to analyze such longitudinal studies (16). Incorporating measures of disease and exposure with correlated errors when measured contemporaneously and with correlated errors between pairs of measurements from different collection times will be a major area of future applied research.

#### *Rapid Changes in Infection and Disease Status*

A challenging question will be how to overcome the limitations imposed by dynamic, subtle transitions that are too rapid to be observed even with dense repeated measurements. HPV infection and CIN1 may come and go, or an intermediate stage might both arise and regress in the interval between consecutive observations (17,18). At the least, missing values widen the intervals of observation, and can present difficulties of missing data even in the simpler problem of estimating effects of screening, as when participants miss a scheduled Pap smear appointment (18).

#### *Censoring*

Clinicians treat CIN2 and some cases of persistent CIN1, leading to censoring of downstream events; the censoring can be related to covariates under study (e.g., oral contraceptive use). Some of these problems can be modeled, but reliance on assumptions of independent censoring or data missing at random rather than related to exposure or probability of transitions can lead to conclusions that are not robust. Recognition of when violations of assumptions are or are not tolerable for the interpretation of results will require understanding of statistical techniques, study setting, and natural history of infection.

### PREVENTION: SCREENING

#### Measures of Agreement

There are now several competing options for cervical cancer screening, based on a single technique or a combination of two (19). A fundamental issue that has renewed importance in the face of multiple screening options is reproducibility of the screening result; variability arises not only from differences in classification of the same biospecimen or image by different analytic methods or human readers, but also from variation in the entire collection process. Similarly, it has proven difficult to improve on the mediocre reproducibility of conventional methods such as cytology and colposcopy (20). Molecular assays

based on DNA are more reproducible, but have other deficiencies, particularly lower specificity.

### Transportability of Study Findings

The effectiveness of screening in a community will relate not only to its performance under ideal conditions but even more so to performance in the field (21). Choosing the best specific screening protocol, whether based on pathology, presence of virus, colposcopy or some combination for a given setting, will depend on rates of both false positive and false negative misclassification of a single screen; correlation between errors for repeated interpretations of the same screening modalities and among the combinations, if used; diagnostic criteria for CIN2 and CIN3; correlation between misclassification of screen and diagnostic endpoint; and regional differences in frequency of cofactors.

The incidence and prevalence of disease fundamentally affect positive and negative predictive value (PPV and NPV), and therefore the performance of a regimen in a different setting, even if everything else is similar. Moreover, in comparing screening regimens tested in different settings or in determining the optimal one to apply somewhere else, one must consider the generalizability of sensitivity and specificity estimates from a study in a controlled setting to a clinic in a very different socioeconomic milieu in another population at another time. The generalizability of each of these parameters requires comparable definitions and cutpoints of screening variables and of disease in the field. For these reasons alone, it is naïve to assume that sensitivity and specificity from one study can automatically be reproduced elsewhere. Further, the mix of subgroups of disease is likely to be different from setting to setting; for example, the distribution of infections that progress at different rates can affect overall performance of a regimen.

Further, one modality may outperform another in one or more studies due to the specific implementations of the modalities, rather than as a reflection of intrinsic differences between the modalities. For example, differences in screening frequency and timing, target ages, cut points for referral and post-referral strategies all affect performance. Moreover, the optimal strategy screening strategy will vary by setting, depending on the medical care system and availability of financial resources. Optimizing and evaluating and monitoring a screening program for a setting is a complex problem that requires interactions among pathologists, clinicians and statisticians in ways that have not occurred to date for any prevention program.

### Choice of Surrogate Endpoint

Invasive cervical cancer is often not the best and sometimes not even an available disease endpoint in prevention studies. In case-control studies, the presence of invasion could affect the results of molecular assays, for example, differentiation markers. In prospective designs, including randomized controlled trials, we cannot follow women until they develop cancer when an effective screen is available.

The usefulness of surrogate endpoints is evident (22), but the appropriate surrogate endpoint is unclear. CIN3 (which includes carcinoma *in situ*) seems to be a valid surrogate endpoint for invasive cancer. CIN3 and cancer share the same risk factors (23), CIN3 tends to persist, and, most importantly, screening programs show that a decrease in CIN3 leads to a corresponding decrease in invasive cancer. In some regions, however, women

are treated at the lower threshold of CIN2, which includes the worst-appearing acute lesions as well as incipient CIN3 in an unpredictable blend whose exact composition depends on local diagnostic stringency. For U.S. prevention studies, CIN2 is a useful disease endpoint, because it represents the “real-life” treatment threshold. However, CIN1 is an unreliable diagnosis and a much weaker surrogate endpoint (9,24). For example, a reduction of a fraction of CIN1 or even CIN2 might not lead to an equivalent reduction in cancer because an intervention (e.g., therapeutic vaccination) might reduce low grade CIN by simply speeding the clearance of the many CIN1 already destined to regress and, therefore, would not affect the small fraction destined to progress to cancer.

Recently, with the recognition that persistent oncogenic HPV infection is the necessary cause of CIN3 and cancer, there has been interest in adopting the virologic endpoint of HPV persistence as a surrogate for cancer (25). A difficult statistical issue is defining persistence (*see* chapter 2, Schiffman and Krüger-Kjaer). HPV persistence must be defined at least on a type-specific level. Perhaps a variant-specific definition would be even better at assuring that the original infection was not cleared and a new one acquired; however, a type- or variant-specific definition of persistence would require research assays rather than the kit combining all oncogenic types into a single positive/negative result typically used in clinics. Another issue is fixing the time for persistence as defined by repeated DNA detection. The shape of the risk curve for progression as a function of duration of persistence is just emerging from early data. Most infections clear within 2 years (26); although many clinicians and patients are likely to want to be more aggressive and use a shorter follow-up prior to excisional treatment, a period longer than 2 years would be ideal for study clarity. Statisticians helping to establish rigorous evidence-based screening guidelines will likely face a narrowing window of opportunity, as increasingly aggressive clinical practices lead to still earlier censoring.

### Screening: Verification Bias and Alloyed Gold Standards

Many studies of new technologies in primary screening do not look for disease via colposcopy and multiple biopsies except in those with a positive result by one or more tests because the risks and inconvenience of disease testing are not compensated by the yield of disease, except perhaps in underserved populations (27). Because it is not clear how many of the screen-negative women truly are diseased, sensitivity, specificity, and NPV cannot be estimated.

There are several approaches to address this problem. Simply performing colposcopy on a strictly random sample of screened women without clinical indication minimizes the number of women who are harmed by unnecessary procedures. For the results to be valid, colposcopy of the women whose screen was positive and whose screen was negative must be equivalent and ideally would be blinded, and the sampling fraction must be considered in estimating sensitivity, specificity, and PPV and their precision. Multiple screening tests applied concomitantly will send women who are negative for each test but positive for the other to colposcopy; however, women who screen positive for one test are more likely to have a positive colposcopy than women screened negative for both, leading to an underestimate of sensitivity. On the other hand, omitting women who were not colposcoped from sensitivity calculations (28) or considering these women to be disease free, despite the lack of colposcopic

verification, gives optimistic estimates of sensitivity unless the specificity of the screening strategy which deems a woman positive if any of the individual tests is positive is 100%. Indirect estimation methods based on latent class models or other statistical techniques allow computation of diagnostic parameters when more than one test is applied and disease confirmation is restricted to positive cases by either test (29,30), but usually require an untenable assumption that the errors of the tests not be correlated.

### Screening: Lead Time Bias

Demonstration of increased sensitivity improvement is not sufficient to show that a screening tool is improved. A better assay of a precursor lesion may be a poorer screen if the extra lesions found are less likely to progress to cancer. For example, HPV-DNA testing that is more sensitive than the conventional Pap smear might detect tiny CIN3 lesions that are likely to regress, not invade. This is still only a theoretical concern, however; there was no evidence for such regression of small CIN3 lesions in a 24-month trial recently reported (31).

Similarly, increased sensitivity from adjunctive testing using HPV is not necessarily advantageous. A nominal increase in the baseline sensitivity of cytology can occur by chance when using adjunctive testing with HPV or other methods, even if the new test were totally random with respect to the disease being evaluated; methods for addressing this are described in chapter 13 (Franco and Ferenczy) (32). Even a real gain in sensitivity from the adjunct screening tests can have an unacceptable cost of specificity.

Inspection of the ROC curve (33) comparing individual and combined regimens (including “both positive” or “either positive” test strategies) is useful for assessing the trade-offs from different ways of using tests. Youden’s test (34) gives equal weight to sensitivity and specificity, thereby putting more value on the typically rarer false negatives than the more common false positives. Youden’s test and measures that use likelihood ratios, like those in Macaskill et al. (35), may be misleading to those who do not realize that conclusions might be altered by different weightings of loss from false positive and false negative errors.

### Cooperation with Health Decision Analysts

Decision making requires natural history data in order to model and optimize prevention strategies (see chapter 15, Goldie). Epidemiologists and statisticians need to recognize the data and analyses required by decision analysis. Moreover, theory-based decision analysis needs to be judged against empirical data. The need for interaction is evident.

### CONCLUSION

The potential translational impact of the breakthroughs in HPV epidemiology over the past few years is unmatched in other areas of cancer epidemiology. The very factors that led to these advances generate novel methodological and statistical questions and provide the opportunity to apply theoretical developments to an important practical problem. The existence of a single family of viruses responsible for virtually all tumors in the cervix and the fact that the viruses clear and do not lead to cancer in most women make prospective, serial biospecimen collection imperative. Identification of cofactors that explain why infections clear rapidly, persist, or progress requires careful control selection in case-control studies. We need to develop methods

for identifying cofactors, for describing natural history, and for suggesting prevention regimens that capitalize on the fact that several distinct stages in the carcinogenesis pathway from normal to cancer can be identified clinically in the cervix, uniquely among common tumors. Moreover, the great benefit that the true endpoints, invasive cancer and death, can be prevented in nearly all women leads to dependent censoring and forces us to consider the implications of using intermediate endpoints. Misclassification problems are particularly acute because the exposure, HPV, is virtually synonymous with early stage disease and because errors in cytology and in detecting oncogenic virus are correlated with one another. Although cytologic screening for prevention is a proven success, there is still a need for development of new methods for the important clinical epidemiology research on assessing optimal screening strategies that involve viral testing. Remarkably, an area where associations are so strong that patterns can emerge from complexity and a public health benefit is so enormous provides one of best interdisciplinary windows into carcinogenesis and cancer prevention.

### REFERENCES

- (1) Munoz N, Bosch FX, de Sanjose S, Herrero R, Castellsague X, Shah KV, et al. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003;348:518–27.
- (2) Xi LF, Koutsky LA, Galloway DA, Kuypers J, Hughes JP, Wheeler CM, et al. Genomic variation of human papillomavirus type 16 and risk for high grade cervical intraepithelial neoplasia. *J Natl Cancer Inst* 1997;89:796–802.
- (3) Herrero R, Hildesheim A, Bratti C, Sherman ME, Hutchinson M, Morales J, et al. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. *J Natl Cancer Inst* 2000;92:464–74.
- (4) Lorincz AT, Castle PE, Sherman ME, Scott DR, Glass AG, Wacholder S, et al. Viral load of human papillomavirus and risk of CIN3 or cervical cancer. *Lancet* 2002;360:228–9.
- (5) Schiffman MH. Recent progress in defining the epidemiology of human papillomavirus infection and cervical neoplasia [Commentary]. *J Natl Cancer Inst* 1992;84:394–8.
- (6) Franco EL. The sexually transmitted disease model for cervical cancer: incoherent epidemiologic findings and the role of misclassification of human papillomavirus infection. *Epidemiology* 1991;2:98–106.
- (7) Schiffman MH, Schatzkin A. Test reliability is critically important to molecular epidemiology: an example from studies of human papillomavirus infection and cervical neoplasia. *Cancer Res* 1994;54:1944s–7s.
- (8) Sherman ME, Schiffman MH, Lorincz AT, Manos MM, Scott DR, Kurman RJ, et al. Towards objective quality assurance in cervical cytopathology: correlation of cytopathologic diagnoses with detection of high-risk human papillomavirus types. *Am J Clin Pathol* 1994;102:182–7.
- (9) Stoler MH, Schiffman M for the ALTS Group. Inter-observer reproducibility of cervical cytologic and histologic diagnoses: realistic estimates from the ASCUS-LSIL Triage Study (ALTS). *JAMA* 2001;285:1500–5.
- (10) Sherman ME, Cox JT, Schiffman M and the ALTS Study Group. Effects of age and human papilloma viral load on colposcopy triage: data from the randomized Atypical Squamous Cells of Undetermined Significance/Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS). *J Natl Cancer Inst* 2002;94:102–7.
- (11) Scott DR, Hagmar B, Maddox P, Hjerpe A, Dillner J, Cuzick J, et al. Use of human papillomavirus DNA testing to compare equivocal cervical cytologic interpretations in the United States, Scandinavia, and the United Kingdom. *Cancer* 2002;96:14–20.
- (12) Schiffman MH, Bauer HM, Hoover RN, Glass AG, Cadell DM, Rush BB, et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. *J Natl Cancer Inst* 1993;85:958–64.
- (13) Munoz N, Franceschi S, Bosetti C, Moreno V, Herrero R, Smith JS, et al. Role of parity and human papillomavirus in cervical cancer: the IARC multicentric case-control study. *Lancet* 2002;359:1093–101.

- (14) Moreno V, Bosch FX, Munoz N, Meijer CJ, Shah KV, Walboomers JM, et al. Effect of oral contraceptives on risk of cervical cancer in women with human papillomavirus infection: the IARC multicentric case-control study. *Lancet* 2002;359:1085-92.
- (15) Castle PE, Wacholder S, Lorincz AT, Scott DR, Sherman ME, Glass AG, et al. A prospective study of high-grade cervical neoplasia risk among human papillomavirus infected women. *J Natl Cancer Inst* 2002;94:1406-14.
- (16) Commenges D. Inference for multi-state models from interval-censored data. *Stat Methods Med Res* 2002;11:167-82.
- (17) Moscicki AB, Hills N, Shiboski S, Powell K, Jay N, Hanson E, et al. Risks for incident human papillomavirus infection and low-grade squamous intraepithelial lesion development in young females. *JAMA* 2001;285:2995-3002.
- (18) Castle PE, Wacholder S, Sherman ME, Lorincz AT, Scott DR, Glass AG, et al. Absolute risk of a subsequent abnormal Pap among oncogenic human papillomavirus DNA-positive, cytologically normal women. *Cancer* 2002;95:2145-51.
- (19) Sherman ME, Lorincz AT, Scott DR, Glass AG, Mielzynska I, et al. Baseline cytopathology, human papillomavirus testing, and risk for cervical neoplasia: a 10-year cohort analysis. *J Natl Cancer* 2003;95:46-52.
- (20) Smith A, Sherman ME, Scott DR, Tabbarra SO, Dworkin L, Olson J, et al. Review of the Bethesda System Atlas does not improve reproducibility or accuracy in the classification of atypical squamous cells of undetermined significance. *Cancer* 2000;90:201-6.
- (21) Bosch FX, Munoz N. The viral etiology of cervical cancer. *Virus Res* 2002;89:183-90.
- (22) Schatzkin A, Freedman L, Schiffman MH, Dawsey S. Validation of intermediate endpoints in cancer research. *J Natl Cancer Inst* 1990;82:1746-52.
- (23) Shah KV, Kessis TD, Shah F, Gupta JW, Shibata D, Jones RW. Human papillomavirus investigation of patients with cervical intraepithelial neoplasia 3, some of whom progressed to invasive cancer. *Int J Gynecol Pathol* 1996;15:127-30.
- (24) Cox JT, Schiffman M, Solomon DS for the ALTS Group. Prospective follow-up suggests similar risk of subsequent high-grade CIN among women with CIN 1 or negative colposcopically-directed biopsy. *Am J Obstet Gynecol*. 2003. In press.
- (25) Nobbenhuis MA, Walboomers JM, Helmerhorst TJ, Rozendaal L, Remmink AJ, et al. Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study. *Lancet* 1999;354:20-5.
- (26) Richardson H, Kelsall G, Tellier P, Voyer H, Abramamowicz M, Ferenczy A, et al. The natural history of type-specific human papillomavirus infections in female university students. 2003. In press.
- (27) Belinson J, Qiao YL, Pretorius R, Zhang WH, Elson P, Li L, et al. Shanxi Province Cervical Cancer Screening Study: a cross-sectional comparative trial of multiple techniques to detect cervical neoplasia. *Gynecol Oncol* 2001;83:439-44.
- (28) Ratnam S, Franco EL, Ferenczy A. Human papillomavirus testing for primary screening of cervical cancer precursors. *Cancer Epidemiol Biomarkers Prev* 2000; 9:945-51.
- (29) Begg CB, Greenes RA. Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* 1983;39:207-15.
- (30) Walter SD. Estimation of test sensitivity and specificity when disease confirmation is limited to positive results. *Epidemiology* 1999;10:67-72.
- (31) The ALTS Group. A randomized trial on the management of atypical cells of undetermined significance. *Am J Obstet Gynecol*. In press.
- (32) Franco EL, Ferenczy A. Assessing gains in diagnostic utility when human papillomavirus testing is used as an adjunct to papanicolaou smear in the triage of women with cervical cytologic abnormalities. *Am J Obstet Gynecol* 1999;181:382-6.
- (33) Biggerstaff BJ. Comparing diagnostic tests: a simple graphic using likelihood ratios. *Stat Med*. 2000;19:649-63.
- (34) Ferreccio C, Bratti C, Sherman ME, Herrero R, Hildesheim A, et al. A comparison of single and combined visual, cytologic, and virologic tests as cervical cancer screening methods in a region at high risk of cervical cancer. *Cancer Epidemiol Biomarkers Prevent*. 2003. In press.
- (35) Macaskill P, Walter SD, Irwig L, Franco EL. Assessing the gain in diagnostic performance when combining two diagnostic tests. *Stat Med* 2002; 21:2527-46.

## NOTES

<sup>1</sup>*Editor's note:* SEER is a set of geographically defined, population-based, central cancer registries in the United States, operated by local nonprofit organizations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes the data available to the public for scientific research.

The author thanks Dr. Ruth Pfeiffer, Biostatistics Branch, DCEG, NCI for her constructive suggestions on a draft of the manuscript.