HPV TESTING AND UNDERSTANDING VALIDITY:
A tough row to hoe

Mark H. Stoler, MD
ASC Companion Meeting
USCAP 2008

OBJECTIVES:

1. Describe the concept of marker validation in the context of HPV tests.
2. Present why the currently available HPV test is highly valid for its existing indications.
3. Emphasize the difference between clinical and analytic validation.
4. Discuss the issues regarding validating competitive HPV tests.

Introduction

All talks about cervical neoplasia testing start with the declaration: the Pap smear is the best example of a successful screening test! Most talks then follow with a reminder that the Pap, while of obvious clinical validity, was widely implemented without the benefit of a randomized clinical trial, one of the current standards of proof or best evidence that a test should be part of general practice. Well, now we have an abundance of data regarding the clinical performance of the Pap alone and in relation to other potential screening tests. Because of this application of the benefits of evidence-based medicine, HPV testing is integral to all aspects of cytopathology practice in the realm of cervical neoplasia. This evolution of molecular testing has created an expanded market for HPV testing of women in the cervical cancer screening system. Currently, only one HPV test is FDA approved. This same test has also been the most highly validated. But new competitive tests will bring competition to this market and it is incumbent upon us to understand the relative performance and function of these tests. Indeed, I believe we need ongoing national dialogue regarding the continued evolution of cervical screening paradigms as regards the interplay between HPV testing, vaccination and cytology.

Marker Validation: a process that is often incomplete

Markers can be diagnostic, prognostic or predictive in nature. A diagnostic marker or (test) can be defined as an item of information concerning the pathology of interest that can increase the accuracy of diagnosis. If the test is positive, the patient has CIN3. Prognostic markers allow for the prediction of outcomes, such as the risk that the patient may develop CIN3, if no CIN 3 is found today. Predictive markers are usually distinguished from prognostic as the former is defined by some as being of utility in predicting therapeutic response, which of course may blur together with prognosis. A Pap interpretation can be both diagnostic and prognostic. The same applies to HPV testing, perhaps even more strongly.
Ideally markers are highly correlated with biologic processes. The literature is full of candidate markers that fail to become part of routine practice. They in a sense, fail the process. The process of determining marker utility is a much more daunting problem than much of the literature suggests and explains why many of the claims in the literature are confusing, conflicting or just plain poor science. In partial response to the inappropriate or premature use of markers in many settings, guidelines for the evaluation of prognostic factors have been developed and refined. The American Joint Committee on Cancer (A.J.C.C.), C.A.P., Working Group on Ethical, Legal and Social Implications of the human genome research of the National Institutes of Health (N.I.H./E.L.S.I.) all have published position papers on this topic. By analogy, many similar concerns can be applied to diagnostic markers as well. All proposals have certain common threads. To be useful a proposed new factor should be significant; meaning it is not a chance occurrence, is independent and most importantly is clinically relevant. Clinical relevance is the most stringent standard since it requires that the additional information provided by the factor will affect patient management or outcome in a useful manner. Does the factor, lead to more accurate diagnosis, better prediction of outcome, better therapeutic triage or perhaps more cost-effective care? A factor may be highly correlated with biologic processes, but if management options are limited it may be of little clinical relevance and hence should probably not be in routine use. Only markers having high clinical outcome scores (validity) supported by high quality evidence are recommended for standard practice. All these systems recognize that determination of analytic sensitivity and specificity is only a first step and factors like technical feasibility, assay codification, methodological accuracy, precision, quality assurance, cost and turn around time are all prerequisites before well designed and statistically powered trials of clinical sensitivity, specificity, predictive value and validity/utility can be fully established and a new marker can be considered ready for “prime time” general practice. The row is long and the work is hard and expensive. But this is because we now know what is truly necessary.

The Current Context

In 2001, The American Society of Colposcopy and Cervical Pathology (ASCCP)-sponsored consensus conference established a series of evidence-based guidelines to help guide clinical practice and management of women with cervical abnormalities. The National Cancer Institute-sponsored clinical trial called ALTS firmly established the clinical value of HPV testing in the management of patients with equivocal cytologic abnormality. To a large extent, the wealth of data generated by ALTS “guided the guidelines” developed by the ASCCP-sponsored consensus conference. In my commentary accompanying the 2002 JAMA publication of the ASCCP guidelines it was noted:

“Human papillomavirus (HPV) testing has matured, appears clinically validated and should become integral to both screening and clinical management….. [The] bar has been raised for bringing forward newer HPV diagnostics. Having well-established positive and negative predictive values for the current [FDA-approved] test, which are applicable to most
populations, allows for clear probabilistic reporting of the results in direct correlation with those of the source cervical cytology. Any new test must document its performance relative to this standard, because many of the proposed management guidelines are based on the performance data…”.

In the ensuing time, multiple independent studies have validated the use of HPV testing in the management of women with abnormal cytology. Studies have demonstrated that HPV testing in a primary screening setting is more sensitive, more reproducible and of better predictive value compared to cytology alone. When used in combination with cytology, HPV compensates for the relative insensitivity of a single Pap test. Long awaited randomized trials have now validated these concepts.

Last year, ASCCP and the community updated and revised the guidelines as published in Oct 2007. This edition further expanded the indications for HPV testing. Yet circa the Spring of 2008 there are still no other FDA-approved HPV tests, although several non-approved tests are being sold under ASR status or in Europe and major clinical trials are underway by several commercial entities. The lack of multiple, competitive well-validated tests is still a problem, as noted in the new guidelines:

“These Guidelines expand clinical indications for HPV testing based on studies using validated HPV assays. One cannot assume that management decisions that are based on results of HPV tests that have not been similarly validated will result in the outcomes that are intended by these guidelines. Furthermore the application of these guidelines using such [unvalidated] tests may increase the potential for patient harm. Therefore, the appropriate use of these guidelines requires that laboratories utilize only HPV tests that have been analytically and clinically validated with proven acceptable reproducibility, clinical sensitivity, specificity and positive and negative predictive values for cervical cancer and verified precancer (CIN2,3), as documented by U.S. Food and Drug Administration (FDA) approval and/or publication in peer-reviewed scientific literature.”

A companion piece to the guidelines was independently published (extensively paraphrased here) to clarify the meaning of the phrase “analytically and clinically validated with proven acceptable reproducibility, clinical sensitivity, specificity and positive and negative predictive values for cervical cancer and precancer (CIN2,3)”, and in addition to more accurately reflect the discourse during the consensus process. As alluded to in the section on bringing forward validated tests, a candidate test must be reliable and reproducible so that a specimen sent to one laboratory would yield the same result at other laboratories. It must have its technical and analytic performance parameters codified. But most importantly, the tests must also be accurate in judging whether a clinically relevant HPV infection is present or not. Clinical relevance or validity means detecting precancer or reassuring patients they do not have precancer. It has little to with detecting virus.
Accuracy is not simple to achieve in the realm of HPV diagnostics. In general, most tests are analytically, but may not be clinically validated. In infectious disease testing it may be clinically important to have maximal analytical sensitivity for very small numbers of infectious viruses like HIV or hepatitis C. In contrast, absolute analytic sensitivity for the smallest possible number of molecules of HPV is not a desirable result. We know that many patients have HPV detectable by molecular analyses in which no clinical evidence of disease can immediately be demonstrated. Excessive analytic sensitivity of HPV molecular diagnostics can cause clinically non-specific outcomes i.e., referral to colposcopy and possible biopsy or surgery in the absence of CIN2,3. Furthermore, colposcopy, the main tool we have for detecting clinical lesions, is pretty insensitive. If viral testing is too sensitive, the clinicians caring for the patient may come to consider the tests as false positive in that there may be no demonstrable evidence of disease by cytology, colposcopy or histology. But if the tools of disease ascertainment are flawed the conclusion may well be wrong.

HPV testing is an especially complex kind of molecular diagnostic assay, which is certainly one of the reasons that there are not more FDA-approved tests. Approximately 15 HPV types can cause CIN3, AIS, and carcinoma, but with varying carcinogenic risk. The numbers of HPV genotypes assayed, and the positive cut-off for the number of viral copies of each, are related analytic as well as clinical issues. It is the interplay between analytic sensitivity and clinical sensitivity that is critical to the issue of clinical validation. Hence the current FDA-approved test cutpoint has been optimized and clinically validated to trade off sensitivity vs. specificity for a clinically important end point rather than some arbitrary analytic cut-off of numbers of HPV molecules. **The critical cut point is sensitivity for predicting the presence of, or future detection of CIN3 within the screening interval.** While the guidelines use the safer and broader target of CIN2,3 as a management threshold for safety, the more rigorous and reproducible standard of adjudicated CIN3 was preferred in ALTS and is recommended as the better endpoint for validation testing.

It is also critical to recognize that no single test or even combination of tests will attain perfect clinical sensitivity. That is, all tests have an inherent false negative rate. Moreover, aside from test failures, the inherent variability of cervical sampling is intrinsic to any screening test and is prone to error leading to false negative rates. Practically, the upper limit of clinical sensitivity is on the order of 95-97% rather than 100% at any analytic sensitivity when testing for carcinogenic HPV, as a marker of risk for cervical cancer. Even in the theoretical scenario in which an ultra-analytically sensitive test achieved near 100% sensitivity, resulting in extreme over-referral to colposcopy (or if we simply sent everyone to colposcopy) we now understand that the diagnostic standard of colposcopy is imperfectly sensitive. Thus, we must recognize the true limits for achieving maximum programmatic sensitivity for detection of precancer as well as the huge financial and undeniable iatrogenic costs of trying in vain to achieve it.

Finally, most clinicians and their patients have no desire to understand the nuances between clinical and analytic validation, nor should they really have to in an evidence-based practice. Most assume as a matter of course that any test offered by a clinical
laboratory has been clinically validated for the indications for which they are using the test. We should not betray their trust.

**Take Home Points:**

- **HPV Testing Is Integral To Contemporary Clinical Management.**
- **Clinical Validation of HPV Tests Requires Clinical Trials Using a CIN3 Endpoint.**
- **Such Trials are Beyond the Routine for Clinical Laboratories.**
- **Given the Excellent Performance of the Current Test, There is Little Room for Improvement in New Tests in Terms of Clinical Performance.**

**Selected Reading:**

Holtzman, NA and Watson, MS Promoting safe and effective genetic testing in the United States http://www.nhgrl.nih.gov/ELSI/TFGT_final/


Bulkmans N, Berkhof J, Rozendaal L et al. Human papillomavirus DNA testing for the detection of cervical intraepithelial neoplasia grade 3 and cancer: 5-year follow-up


Stoler MH. HPV testing: If it's not clinically valid, it's dangerous. *HPV Today*. 2006;9:4-5.


