

# Human papillomavirus testing as a cytology gold standard: comparing Surinam with the Netherlands

Mitchell S Wachtel<sup>1</sup>, Mathilde E Boon<sup>2</sup>, Hans Korporaal<sup>2</sup> and Lambrecht P Kok<sup>3</sup>

<sup>1</sup>St Joseph Regional Health Center, Bryan, TX, USA; <sup>2</sup>Leiden Cytology and Pathology Laboratory, Leiden, The Netherlands and <sup>3</sup>University of Groningen, Groningen, The Netherlands

**Polymerase chain reaction to detect high-risk human papillomavirus has been suggested as a gold standard for cytology. The Netherlands and Surinam were prospectively compared in regard to the proportions of Negative, Atypical Squamous Cells of Undetermined Significance, and Squamous Intraepithelial Lesion smears that had detectable high-risk human papillomavirus. For the Netherlands, 14 600 negative, 270 Atypical Squamous Cells of Undetermined Significance and 120 Squamous Intraepithelial Lesion smears were evaluated by polymerase chain reaction. For Surinam, 150 negative, 50 Atypical Squamous Cells of Undetermined Significance, and 150 Squamous Intraepithelial Lesion smears were evaluated by polymerase chain reaction. In all, 4% of Dutch and 80% of Surinamese negative smears had detectable high-risk human papillomavirus ( $\chi^2 = 1313$ ,  $P < 0.00001$ ). In total, 41.9% of Dutch and 84% of Surinamese Atypical Squamous Cells of Undetermined Significance smears had detectable high-risk human papillomavirus ( $\chi^2 = 28$ ,  $P < 0.00001$ ). Totally, 67.5% of Dutch and 94% of Surinamese SIL smears had detectable high-risk human papillomavirus ( $\chi^2 = 30$ ,  $P < 0.00001$ ). The Negative: Atypical Squamous Cells of Undetermined Significance odds ratio was 0.058 for the Netherlands and 0.762 for Surinam ( $\chi^2_{\text{homog}} = 31$ ,  $P < 0.00001$ ). The Negative: Squamous Intraepithelial Lesion odds ratio was 0.020 for the Netherlands and 0.255 for Surinam ( $\chi^2_{\text{homog}} = 31$ ,  $P < 0.00001$ ). The Atypical Squamous Cells of Undetermined Significance: Squamous Intraepithelial Lesion odds ratio was 0.347 for the Netherlands and 0.335 for Surinam ( $\chi^2_{\text{homog}} = 0.005$ ,  $P > 0.75$ ). Human papillomavirus DNA testing may not be a suitable gold standard in general because its use would make specificity and sensitivity prevalence-dependent. A new statistic, the percent of Negative pap smears with detectable high-risk human papillomavirus, is posited, which may be important if human papillomavirus DNA testing is used clinically.**

*Modern Pathology* (2005) 18, 349–353, advance online publication, 23 July 2004; doi:10.1038/modpathol.3800242

**Keywords:** squamous intraepithelial lesion; human papillomavirus; polymerase chain reaction; prevalence; odds ratio

Recent studies in the United States have shown great promise in the use of the polymerase chain reaction (PCR) to detect high-risk human papillomavirus DNA (HPV<sup>hr</sup>) in order to triage specimens with diagnoses of atypical squamous cells of undetermined significance (ASCUS).<sup>1,2</sup> The technique has been suggested as a quality assurance device.<sup>3–5</sup> The utility of this technique for developing nations has not been fully assessed. Dutch groups have shown that persistent infection with HPV<sup>hr</sup> is necessary for the development and maintenance of high-grade squamous intraepithelial

lesion (SIL).<sup>6</sup> PCR testing of Surinamese cervical scrapings by Dutch groups has been performed and reported as a viable technique.<sup>7</sup> Surinam has five times the incidence of cervical carcinoma that the Netherlands has; a study of paraffin-embedded specimens detected 13 different HPV genotypes in the Surinamese group vs nine in the Dutch group.<sup>8</sup>

This study shows that, for the diagnostic categories Negative, ASCUS, and SIL, a greater fraction of cervicovaginal smears will have detectable HPV<sup>hr</sup> in Surinam than in the Netherlands. This study also shows that, whereas Dutch and Surinamese Negative: ASCUS and Negative:SIL odds ratios differ, Dutch and Surinamese ASCUS:SIL odds ratios are quite similar. The results question the use of HPV DNA testing as a gold standard in general because the calculated sensitivity and specificity then become prevalence-dependent. The use of HPV DNA testing for creation of an ASCUS:SIL ratio, as previously

Correspondence: Dr ME Boon, MD, Leiden Cytology and Pathology Laboratory, PO Box 16084, NL-2301 GB Leiden, The Netherlands.

E-mail: m.e.boon@lcp.nl

Received 26 April 2004; revised and accepted 2 June 2004; published online 23 July 2004

suggested, is, however, supported.<sup>5</sup> In addition, a new statistic, the percent of negative smears with HPV<sup>hr</sup>, is posited, which may be very important in evaluating the clinical utility of HPV DNA testing.

## Materials and methods

### Case Selection

From a Dutch private cytology laboratory were obtained 300 000 consecutive pap smears; from a Surinamese screening program were obtained 10 000 pap smears. Both sets of smears received KOPAC system diagnoses,<sup>9</sup> which translate into Bethesda system classes: Negative (KOPAC S1), ASCUS (KOPAC S2-3), and SIL (KOPAC S4-9). The frequency distribution of diagnoses is displayed in Table 1. One of us (MEB), known internationally for her skills as a cytopathologist, evaluated all of the Dutch and Surinamese smears used in this study. From the Dutch laboratory, Negative smears were obtained by systematic randomization; if the smears were in her opinion Negative, they were analyzed by PCR. This process was continued until 14 600 Negative Dutch smears were obtained for PCR testing. Similarly, 270 Dutch ASCUS smears, 120 Dutch SIL smears, 100 Surinamese Negative smears, 50 Surinamese ASCUS smears, and 150 Surinamese

**Table 1** Frequency distribution (and percents) of smear diagnoses by country

	<i>The Netherlands</i>	<i>Surinam</i>
Negative	292 200 (97.4%)	7450 (74.5%)
ASCUS	5400 (1.8%)	1800 (18%)
SIL	2400 (0.8%)	750 (7.5%)
Total	300 000 (100%)	10 000 (100%)

**Table 2** Frequency distribution (and percents) of smears with and without detectable HPV<sup>hr</sup> by cytological diagnosis and country, and the  $\chi^2$  statistics with Yates correction for testing hypotheses that the Netherlands and Surinam had the same proportion of smears with detectable HPV<sup>hr</sup>

	<i>With detectable HPV<sup>hr</sup></i>	<i>Without detectable HPV<sup>hr</sup></i>	<i>Total</i>	$\chi^2$
<i>Negative</i>				
<i>The Netherlands</i>	584 (4%)	14 016 (96%)	14 600 (100%)	1313*
<i>Surinam</i>	80 (80%)	20 (20%)	100 (100%)	
<i>ASCUS</i>				
<i>The Netherlands</i>	113 (41.9%)	157 (58.1%)	270 (100%)	28*
<i>Surinam</i>	42 (84%)	8 (16%)	50 (100%)	
<i>SIL</i>				
<i>The Netherlands</i>	81 (67.5%)	39 (32.5%)	120 (100%)	30*
<i>Surinam</i>	141 (94%)	9 (6%)	150 (100%)	

\* $P < 0.00001$ .

SIL smears were obtained for this prospective study. The number of cases rejected was not recorded.

### Polymerase Chain Reaction

PCR was performed as previously described.<sup>7,10,11</sup> In brief, using standard methods and commercially available reagents (QIAamp DNA Mini Kit, Qiagen, the Netherlands), target DNA was amplified using commercially available consensus primers for the conserved region of HPV<sup>hr</sup> for types 16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58; MY09/MY11 primer/probes set (Digene/Abbot, USA) and 45 amplification cycles.

### Statistics

To test hypotheses that proportions did not differ,  $\chi^2$  statistics with Yates correction were calculated.<sup>12</sup> To test hypotheses that odds ratios did not differ,  $\chi^2_{\text{homog}}$  statistics were calculated.<sup>12</sup> A statistic, the percent of Negative smears with detectable HPV<sup>hr</sup>, was posited.

## Results

Table 2 displays the frequency distribution of PCR test results stratified by diagnostic group and country. For smears diagnosed as Negative, 80% of Surinamese and 4% of Dutch smears had detectable HPV<sup>hr</sup> ( $\chi^2 = 1313$ ,  $P < 0.00001$ ). For smears diagnosed as ASCUS, 84% of Surinamese and 41.9% of Dutch smears had detectable HPV<sup>hr</sup> ( $\chi^2 = 28$ ,  $P < 0.00001$ ). For smears diagnosed as SIL, 94% of Surinamese and 67.5% of Dutch smears had detectable HPV<sup>hr</sup> ( $\chi^2 = 30$ ,  $P < 0.00001$ ). For each diagnostic group, a greater percent of Surinamese than Dutch smears had detectable HPV<sup>hr</sup>.

Table 3 displays odds ratios stratified by country. The Negative:ASCUS odds ratio was 0.762 for

**Table 3** Odds ratios by country and the  $\chi^2_{\text{homog}}$  statistics for testing hypotheses that the Netherlands and Surinam had equal odds ratios

	<i>The Netherlands</i>	<i>Surinam</i>	$\chi^2_{\text{homog}}$
Negative:ASCUS	0.058	0.762	31*
Negative:SIL	0.020	0.255	31*
ASCUS:SIL	0.347	0.335	0.005**

\* $P < 0.00001$ .

\*\* $P > 0.75$ .

Surinam and 0.058 for the Netherlands ( $\chi^2_{\text{homog}} = 31$ ,  $P < 0.00001$ ). The Negative:SIL odds ratio was 0.255 for Surinam and 0.020 for the Netherlands ( $\chi^2_{\text{homog}} = 31$ ,  $P < 0.00001$ ). The ASCUS:SIL odds ratio was 0.335 for Surinam and 0.347 for the Netherlands ( $\chi^2_{\text{homog}} = 0.005$ ,  $P > 0.75$ ). When Negative and abnormal (either ASCUS or SIL) diagnoses were compared, the Dutch ratio of the odds of detectable HPV<sup>hr</sup> was less than 1/10 the Surinamese ratio. For both Surinam and the Netherlands, SIL diagnoses conferred about three times the odds of detectable HPV<sup>hr</sup> as did ASCUS diagnoses.

## Discussion

The fractions of Surinamese Negative smears, ASCUS smears, and SIL smears that had detectable HPV<sup>hr</sup> were greater than those of Dutch Negative smears, ASCUS smears, and SIL smears. The Negative:ASCUS and Negative:SIL odds ratios for the Netherlands differed from those in Surinam. The ASCUS:SIL odds ratio for the Netherlands was similar to that seen in Surinam.

Because Negative:ASCUS and Negative:SIL odds ratios in Surinam differed from those in the Netherlands, the sensitivity and the specificity of an abnormal cytologic examination cannot remain constant with respect to prevalence if HPV testing is used as the gold standard. This is also true because HPV infection and dysplastic/cancerous lesions do not bear a simple linear relationship. HPV infection is often transient in sexually active, young women with normal cervicovaginal cytology: in 93% of initially infected women, the same viral type was not detected upon re-examination four menstrual cycles later.<sup>13</sup> Heterogeneity is also supported by a meta-analysis of many studies from around the world that showed HPV types associated with cervical cancer varied by region.<sup>14</sup> Although HPV has been divided into high- and low-risk categories, studies evidence a wider array than previously thought of carcinogenic HPV serotypes.<sup>15</sup> Additionally, other cofactors in the carcinogenesis of this infection, such as smoking and bacterial flora,<sup>16,17</sup> have been posited.

In all, 80% of Surinamese smears without cytologic aberrations had detectable HPV<sup>hr</sup>; 4% of Dutch smears without cytologic aberrations had detectable

HPV<sup>hr</sup>. The bulk of the smears in both countries lack cytologic aberrations; were HPV DNA testing to be used as a gold standard, the sensitivity of an abnormal cytologic diagnosis would be considerably lower in Surinam than in the Netherlands. Similarly, because there is such a high background rate of HPV<sup>hr</sup> positivity in Surinam, the specificity of an abnormal cytologic diagnosis would of necessity be lower. Given that the pap smear is known to be an excellent screening tool in developing countries,<sup>18</sup> it follows HPV DNA testing is a questionable gold standard in general for cytology because if it were used the sensitivity and specificity would show variability independent of the judgment and ability of the person or laboratory who evaluates the smears.

Surinamese smears with detectable HPV<sup>hr</sup> have 3/4 the odds of receiving a Negative diagnosis from the person who reviewed the smears for this study as they would have of receiving an ASCUS diagnosis; the corresponding ratio for Dutch smears is 1/20. Over 80% of Surinamese ASCUS smears had detectable HPV<sup>hr</sup>. It follows that using HPV DNA testing to triage Surinamese patients with ASCUS diagnoses may expose a large fraction of these patients to unneeded colposcopy. The latter assertion should be empirically evaluated; if it is shown to be the case, the percent of Negative pap smears with detectable HPV<sup>hr</sup> should be determined before HPV DNA testing is used for purposes of ASCUS triage. When the percent of Negative pap smears with detectable HPV<sup>hr</sup> is very high, 80% for Surinam in this study, HPV testing is of questionable value in general.

Because the ASCUS:SIL odds ratios were the same for the Netherlands and for Surinam, it follows that the ASCUS:SIL odds ratio, in this study, was relatively independent of prevalence. This independence raises the possibility that the similarity of the ASCUS:SIL ratios of the two countries reflects differences in the assessment of abnormal smears.<sup>5</sup> The cause of such differences can be varied; no definite statements can be made without examining the histologic follow-up, which was not available for the Surinamese study. Because the ASCUS:SIL odds ratio was in this study prevalence-independent with respect to detectable HPV<sup>hr</sup>, the results support the ASCUS:SIL odds ratio as a quality control device, as previously described.<sup>5</sup> As noted, the ASCUS:SIL odds ratio must be accompanied by another measure or set of measures if it is to be interpretable.<sup>5</sup> By contrast, the fraction of ASCUS cases that have detectable HPV<sup>hr</sup>, which had been previously recommended as a quality control measure,<sup>3,4</sup> is probably not a good measure of diagnostic judgment in and of itself because, as this study shows, that fraction will vary for reasons that have nothing to do with the judgment of the cytologist. The conclusion that individual diagnostic categories show prevalence-dependence in regard to detectable HPV<sup>hr</sup> is supported by other studies.<sup>5,19,20</sup>

The finding that the Netherlands and Surinam differ is well-supported by prior comparisons of the two countries. Krul *et al*,<sup>21</sup> in a study of cervical carcinoma, found differences in tumor-associated somatic genetic alterations, possibly due to differences in the genetic pathways that lead to squamous cancer and also, as previously noted, found differences in HPV genotypes between the countries.<sup>8</sup>

In summary, fractions of Negative, ASCUS, and SIL smears with detectable HPV<sup>hr</sup> are greater in Surinam than in the Netherlands. Relative to a Negative diagnosis, an abnormal diagnosis confers a much greater risk of a smear's having detectable HPV<sup>hr</sup> in the Netherlands than in Surinam. For both the Netherlands and Surinam, an ASCUS diagnosis confers about one-third the odds of having detectable HPV<sup>hr</sup> that an SIL diagnosis confers. Taken together, the findings reported in this study indicate that using HPV DNA testing for cytology as a gold standard in general is questionable, although it may be of use to create an ASCUS:SIL odds ratio. If HPV DNA testing is to be used, it may be vital to assess the percent of Negative pap smears with detectable HPV<sup>hr</sup>.

## References

- Solomon D, Schiffman M, Tarone R. ALTS Study Group. Comparison of three management strategies for patients with atypical squamous cells of undetermined significance: baseline results from a randomized trial. *J Natl Cancer Inst* 2001;93:293–299.
- Sherman M, Schiffman M, Cox JT. 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 Group. *J Natl Cancer Inst* 2002;94:102–107.
- Zuna RE, Moore W, Dunn ST. HPV DNA testing of the residual sample of liquid-based Pap test: utility as a quality assurance monitor. *Mod Pathol* 2001;14:147–151.
- Stoler MH. Toward objective quality assurance: the eyes don't have it. *Am J Clin Pathol* 2002;117:520–522.
- Wachtel MS, Dahm PF. The ASCUS:SIL ratio and the reference laboratory pathologist. *Cytopathology* 2003;14:249–256.
- Nobbenhuis MA, Walboomers JM, Helmerhorst TJ, *et al*. Relation of human papillomavirus status to cervical lesions and consequences for cervical-cancer screening: a prospective study. *Lancet* 1999;354:20–25.
- Wisman GB, Hollema H, de Jong S, *et al*. Telomerase activity as a biomarker for (pre)neoplastic cervical disease in scrapings and frozen sections from patients with abnormal cervical smear. *J Clin Oncol* 1998;16:2238–2245.
- Krul EJ, Van De Vijver MJ, Schuurin E, *et al*. Human papillomavirus in malignant cervical lesions in Surinam, a high-risk country, compared to the Netherlands, a low-risk country. *Int J Gynecol Cancer* 1999;9:206–211.
- Boon ME, Suurmeijer AJH (eds). The precursor lesions. In: *The Pap Smear* 2nd edn. Coulomb Press Leyden: Leiden, The Netherlands, 1993, pp 105–149.
- Reesink-Peters N, Helder MN, Wisman GB, *et al*. Detection of telomerase, its components, and human papillomavirus in cervical scrapings as a tool for triage in women with cervical dysplasia. *J Clin Pathol* 2003;56:31–35.
- Ngan HY, Cheung AN, Liu SS, *et al*. Telomerase assay and HPV 16/18 typing as an adjunct to conventional cytological cervical cancer screening. *Tumour Biol* 2002;23:87–92.
- Fleiss JL. *Statistical Methods for Rates and Proportions* 2nd edn. John Wiley & Sons, Inc.: New York, 1981.
- Hinchliffe SA, van Velzen D, Korporaal H, *et al*. Transience of cervical HPV infection in sexually active, young women with normal cervicovaginal cytology. *Br J Cancer* 1995;72:943–945.
- Clifford GM, Smith JS, Plummer M, *et al*. Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *Br J Cancer* 2003;88:63–73.
- Munoz N, Bosch FX, de Sanjose S, *et al*. International Agency for Research on Cancer Multicenter Cervical Cancer Study Group. Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003;348:518–527.
- Boon ME, van Ravenswaay Claasen HH, Kok LP. Urbanization and baseline prevalence of genital infections including *Candida*, *Trichomonas*, and human papillomavirus and of a disturbed vaginal ecology as established in the Dutch Cervical Screening Program. *Am J Obstet Gynecol* 2002;187:365–369.
- Castle PE, Wacholder S, Lorincz AT, *et al*. A prospective study of high grade cervical neoplasia risk among human papillomavirus-infected women. *J Natl Cancer Inst* 2002;94:1406–1414.
- Guidozzi F. Screening for cervical cancer. *Obstet Gynecol Surv* 1996;51:247–252.
- Ho GY, Bierman R, Beardsley L, *et al*. Natural history of cervicovaginal papillomavirus infection in young women. *N Engl J Med* 1998;338:423–428.
- Herrero R, Hildesheim A, Bratti C, *et al*. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. *J Natl Cancer Inst* 2000;92:464–474.
- Krul EJ, Kersemaekers AM, Zomerdijk-Nooyen YA, *et al*. Different profiles of allelic losses in cervical carcinoma cases in Surinam and the Netherlands. *Cancer* 1999;86:997–1004.

## Appendix A: Explanation of relevant statistical concepts

*Sensitivity* is the fraction of gold standard-positive patients who are test-positive. *Specificity* is the fraction of gold standard-negative patients who are test-negative. *Positive predictive value* is the fraction of test-positive patients who are gold standard-positive. *Negative predictive value* is the fraction of test-negative patients who are gold standard-negative. *Test level* is the fraction of patients who are test-positive. *Prevalence* is the fraction of patients who are gold standard-positive. The *odds ratio*, derivable from any four-fold table, is the product of

the left upper and right lower corners divided by the product of the right upper and left lower corners.

In a *retrospective study*, two groups, one gold standard-positive, the other gold standard-negative, are evaluated by the test. Calculated directly are the sensitivity, specificity, and odds ratio. Knowing the test level permits estimates of predictive values. **If the odds ratio changes when prevalence changes, sensitivity and specificity cannot be constant if prevalence changes.**

In a *prospective study*, two groups, one test-positive, the other test-negative, are evaluated by the gold standard. Calculated directly are the predictive values and the odds ratio. Knowing prevalence permits estimates of sensitivity and specificity.

In a *naturalistic study*, one group is evaluated both by the gold standard and the test. Calculated directly are the fractions *true positive* (test-positive and gold standard-positive), *false positive* (test-positive and

gold standard-negative), *false negative* (test-negative and gold standard-positive), and *true negative* (test-negative and gold standard-negative), as well as the odds ratio. Test level is: true positive + false positive. Prevalence is: true positive + false negative. True positive is: sensitivity  $\times$  prevalence. False positive is:  $(1 - \text{specificity}) \times (1 - \text{prevalence})$ . Hence:

$$\text{Test level} = \text{true positive} + \text{false positive}$$

$$\text{Test level} = (\text{sensitivity} \times \text{prevalence}) + [(1 - \text{specificity}) \times (1 - \text{prevalence})]$$

$$\text{Test level} = \text{prevalence} \times (\text{sensitivity} + \text{specificity} - 1) + (1 - \text{specificity})$$

If sensitivity and specificity do not vary when prevalence changes, the last equation is a straight line. **If prevalence and test level lack a simple linear relationship, sensitivity and specificity cannot be constant if prevalence changes.**