Int. J. Cancer: **121,** 2218–2224 (2007)

Inappropriate gold standard bias in cervical cancer screening studies

Robert G. Pretorius[1]*, Yan-Ping Bao[2], Jerome L. Belinson[3], Raoul J. Burchette[4], Jennifer S. Smith[5] and You-Lin Qiao[2]*

[1]*Department of Obstetrics and Gynecology, S.C.P.M.G.-Fontana, Fontana, CA*
[2]*Cancer Institute/Hospital, Chinese Academy Medical Sciences, Beijing, China*
[3]*Department of Gynecology, The Cleveland Clinic Foundation, Cleveland, OH*
[4]*Department of Research and Evaluation, S.C.P.M.G., Los Robles, CA*
[5]*Department of Epidemiology, University of North Carolina, Chapel Hill, NC*

As acetic acid-aided visual inspection (VIA) and colposcopic-directed biopsy miss small ≥cervical intraepithelial neoplasia (CIN) 2, inflation of sensitivity of VIA may occur when colposcopic-directed biopsy is the gold standard for ≥CIN 2. To determine whether such inflation occurs, we reviewed 375 women with ≥CIN 2 from the Shanxi Province Cervical Cancer Screening Study II. These women had positive self or physician-collected tests for high-risk human papillomavirus or abnormal cervical cytology and had VIA followed by colposcopy with directed biopsy and endocervical curettage (ECC). If a cervical quadrant had no lesion, a random biopsy at the squamocolumnar junction within that quadrant was obtained. Sensitivity of colposcopic-directed biopsy was higher for ≥CIN 2 involving 3–4 cervical quadrants (81.3%) than for ≥CIN 2 involving 0–2 quadrants (49.0%, $p < 0.001$). Sensitivities of VIA, cytology of ≥ASC-US, ≥LSIL, and ≥HSIL were higher for ≥CIN 2 involving 3–4 quadrants than for ≥CIN 2 involving 0–2 quadrants. When a colposcopic-directed biopsy gold standard was compared with that of a 5-biopsy standard (which included ≥CIN 2 from colposcopic-directed biopsy, random biopsy, or ECC), the sensitivity for ≥CIN 2 of VIA was inflated by 20.0% (65.9% *vs.* 45.9%, $p < 0.001$). Sensitivities of other screening tests were not affected. Similar inflation of sensitivity of VIA was found with an endpoint of ≥CIN 3 (70.4% *vs.* 52.0%, $p = 0.0013$). Inflation of sensitivity of VIA depended upon agreement between colposcopic-directed biopsy and the screening tests as measured by kappa. Studies of VIA that used colposcopic-directed biopsy as the gold standard require reevaluation.

Key words: gold standard; acetic acid-aided visual inspection; colposcopy; cervical neoplasia

Though loop excision of the transformation zone is occasionally employed,[1] biopsy of colposcopic-detected lesions[2–4] is the most common gold standard for the diagnosis of cervical intraepithelial neoplasia (CIN) 2 or worse utilized in cervical cancer screening studies. In the Shanxi Province Cervical Cancer Screening Studies (SPOCCS) I and II,[5,6] we conducted population-based cervical cancer screening studies utilizing a 5-biopsy gold standard for CIN 2 or worse. In these studies, at colposcopy, the cervix was divided into 4 quadrants by 2 perpendicular lines drawn from 12 to 6 o'clock and from 9 to 3 o'clock. Each cervical quadrant was graded separately by colposcopic impression as negative (no lesion), low (suggestive of human papillomavirus (HPV) or CIN 1), high (suggestive of CIN 2 or 3), or cancer. Colposcopic-detected lesions were biopsied. Cervical quadrants without colposcopic-detected lesions had random biopsies at the squamocolumnar junction at 2, 4, 8 or 10 o'clock. An endocervical curettage (ECC) was then obtained. A bronchoscopic biopsy instrument was used to obtain virtually painless 2-mm biopsies. With this protocol, patients undergoing colposcopy had at least 4 cervical biopsies (some cervical quadrants with colposcopic-detected lesions had more than 1 biopsy) and an ECC.

In SPOCCS I, 1,997 women were screened with acetic acid-aided visual inspection (VIA), cervical cytology, self-test for high-risk HPV with specimen collected with Dacron swab and physician-collected test for high-risk HPV (specimen obtained from liquid-based cytology transport media). All 1,997 women underwent colposcopy with biopsy. Of 83 women with CIN 2 or worse diagnosed by the 5-biopsy gold standard in SPOCSS I, 62 (74.7%) were diagnosed by a colposcopic-directed biopsy and 21 (25.3%) were diagnosed by random biopsy in a colposcopically negative cervical quadrant and/or by a positive ECC.[7] In SPOCCS II, 8,497 women were screened with VIA, cervical cytology, self-test for high-risk HPV with specimen collected by a conical-shaped brush, and physician-collected high-risk HPV with specimen collected from the endocervix with a conical-shaped brush.[6] A total of 3,139 women had colposcopy with the 5-biopsy standard described above. Of 364 women in SPOCCS II with adequate colposcopy and CIN 2 or worse diagnosed by the 5-biopsy standard, 208 (57.1%) were diagnosed by a colposcopic-directed biopsy and 156 (42.9%) were diagnosed by a random biopsy in a colposcopically negative cervical quadrant and/or by a positive ECC.[8]

The use of colposcopic-directed biopsy as the gold standard in cervical cancer screening studies may underestimate the prevalence of CIN 2 or worse. However, it is reasonable to use colposcopic-directed biopsies if diagnosis is conditionally independent of the screening test results, given disease status (presence or absence of CIN 2 or worse).[9] In January of 2006, in a review of the SPOCCS I data, we reported a higher sensitivity of different diagnostic tests (*i.e.*, gold standard colposcopic-directed biopsy, VIA and cytology of high grade squamous intraepithelial lesion (HSIL) or greater) for the detection of CIN 2 or worse involving 3 or 4 quadrants of the cervix than for CIN 2 or worse involving 0–2 cervical quadrants (zero quadrants is diagnosis solely by ECC or by directed biopsy of vaginal vault lesion).[7] Given that the gold standard of colposcopic-directed biopsy and the screening tests of VIA and cytology of ≥HSIL appear to be interdependent, it was expected that the sensitivity of VIA and cytology of ≥HSIL for CIN 2 or worse would likely be inflated due to diagnostic inaccuracies when colposcopic-directed biopsy was used as the gold standard.[10]

In SPOCCS I, the sensitivity of VIA for the detection of 62 cases of CIN 2 or worse with a colposcopic-directed biopsy gold-standard (79.0%) was not significantly different from the sensitivity for detection of 83 cases of CIN 2 or worse with the 5-biopsy gold-standard (71.1%) ($p > 0.25$).[7] Similarly, the sensitivity of cytology of ≥HSIL for CIN 2 or worse with a colposcopic-directed biopsy gold-standard (85.5%) was not significantly different from that found with the 5-biopsy gold-standard (75.9%) ($p > 0.1$).[7]

Participants are from the Shanxi Province Cervical Cancer Screening Study II conducted in Xiangyuan and Yangcheng Counties, Shanxi Province, China.
Grant sponsors: Preventive Oncology International™; The Cancer Institute/Hospital, Chinese Academy of Medical Sciences, Beijing, China; Taussig Cancer Center of The Cleveland Clinic Foundation; Goldway Industrial INC, Shenzhen, China.
The first two authors contributed equally to this author.
***Correspondence to:** Department of Obstetrics and Gynecology, S.C.P.M.G.-Fontana, 9961 Sierra Ave., Fontana, CA 92335, USA. Fax: +909-427-4536. E-mail: Robert.G.Pretorius@kp.org (or) Cancer Institute/Hospital, Chinese Academy Medical Sciences, 17 S. Panjiayuan LN, Beijing 100021, China. Fax: +86-10-6771-3648.
E-mail: qiaoy@public.bta.net.cn
Received 11 September 2006; Accepted after revision 21 May 2007
DOI 10.1002/ijc.22991
Published online 26 July 2007 in Wiley InterScience (www.interscience.wiley.com).

This failure to achieve significance in difference of sensitivities for CIN 2 or worse was attributed to the low power associated with the small number of women with CIN 2 or worse ($N = 83$). Further, due to the limited number of women with CIN 3 or cancer ($N = 42$) in SPOCCS I, CIN 2 or worse was used as the study endpoint despite CIN 2 being a generally poor-defined entity, which may or may not progress to CIN 3 or invasive cervical cancer.[1]

To address the problems of low power and low numbers of CIN 3 or cancer in our prior publication, we review here the larger SPOCCS II study results to determine whether colposcopy or the screening tests missed CIN 2 or worse involving 0–2 quadrants of the cervix and whether any of the screening tests had significant inflation of sensitivity for CIN 2 or worse when a gold standard of colposcopic-directed biopsy was compared with that of a 5-biopsy standard. The analysis was repeated using CIN 3 or cancer as the endpoint.

Methods

The 375 women within the SPOCCS II that were diagnosed with CIN 2 or worse were reviewed. The study population and methods are described in detail in a previous publication.[6] In brief, 8,497 women residing in Shanxi Province, China were screened with 4 different screening tests for cervical neoplasia: self and physician-collected tests for high-risk HPV, cervical cytology and VIA. The self-tests for high-risk HPV were obtained in the local villages with a conical-shaped brush (Cervical Sampler®, Digene Corp., Gaithersburg, MD) that was placed high in the vagina and rotated 3–4 times. The brush was then placed in Special Transport Media (STM®, Digene Corp.) and tested for a mixture of 13 intermediate and high-risk types of HPV with the second-generation hybrid capture microplate-based HPV test (HC II test, Digene, Corp.). Women were recalled to a central clinic where they had physician-collected high-risk HPV, cervical cytology, VIA, and in selected cases, colposcopy and biopsy. Physician-collected cervical specimens were obtained from the endocervix with a conical-shaped brush like that used to obtain the self-tests for high-risk HPV and then tested for high-risk HPV with HC II. A cutoff value of 1.0 pg HPV DNA was defined as positive. Cervical specimens for liquid-based cytology were obtained with the Rovex brush and placed in CytoRich® transport media. Liquid-based cytology slides were prepared by the AutoCyte® Prep method (AutoCyte®, TriPath Imaging, Burlington, NC) and interpreted using the Bethesda Classification system by cytopathologists in Beijing who were blinded to the results of the other tests. Visual inspection was performed 1 min after 5% acetic acid was applied to the cervix. VIA was defined as positive when there was a white lesion that might or might not abut the squamocolumnar junction or a friable mass with an irregular surface.

Colposcopy and biopsy with the 5-biopsy gold standard (as described in the introduction) was performed on 3,139 women. Three thousand and sixty of the 3,139 (97.5%) women had positive self or physician-collected high-risk HPV tests or abnormal cervical cytology; 16 (0.5%) had negative self and physician-collected high-risk HPV tests and normal cervical cytology but had abnormal VIA; and 63 (2.0%) had negative self and physician-collected HPV tests, negative cervical cytology and normal VIA. Biopsies were processed and interpreted by pathologists without knowledge of the results of all other diagnostic tests. The final diagnosis was based on the worst biopsy (colposcopic-directed, random in a cervical quadrant without a lesion, or ECC) obtained. Although some subjects underwent cervical conization or hysterectomy as treatment for cervical neoplasia, results of these procedures were not considered in determining the final diagnosis. The Institutional Review Board for human subjects of The Cleveland Clinic Foundation approved this review of previously collected data.

The sensitivities of the self and physician-collected tests for high-risk HPV, VIA, cervical cytology with abnormal defined as $\geq$atypical squamous cells of uncertain significance (ASC-US), cervical cytology with abnormal defined as $\geq$low grade squamous intraepithelial lesion (LSIL), cervical cytology with abnormal defined as $\geq$HSIL, and colposcopy were calculated for CIN 2 or worse involving 0–2 quadrants of the cervix and for CIN 2 or worse involving 3–4 quadrants of the cervix. Differences in proportions were tested with the χ^2 statistic utilizing STAT 1 statistical software (Hewlett–Packard, Palo Alto, CA). Sensitivities of the screening tests were calculated and compared for 2 chosen gold standards: (i) colposcopic-directed biopsy and (ii) the 5-biopsy standard. The correlation between colposcopic-directed biopsy and the screening tests was tested with the kappa statistic utilizing SAS version 9.1.3 (SAS Institute, Cary, NC). The analysis was then repeated using CIN 3 or cancer as the endpoint.

Results

In the Shanxi Province Cervical Cancer Screening Study II, the mean age of the 8,497 screened women was 40.9 years (range 27–56). The mean number of pregnancies was 3.05 (range 0–13). Though the SPOCCS II protocol was to recall women to a central clinic within 3 months of their self-test for high-risk HPV, there were delays secondary to funding and transportation such that the time between the self-test and the other screening tests and colposcopy was between 4 and 16 months.

The distribution of colposcopy and biopsy and the rates of CIN 2 or worse within subgroups based on result of physician-collected high-risk HPV test and cervical cytology within SPOCCS II are shown in Table I. Three thousand five hundred and sixty-two of 8,497 screened women had positive self or physician-collected high-risk HPV tests or cervical cytology of $\geq$ASC-US and, by protocol should have had colposcopy and biopsy. Of these 3,562 women, 2,810 had positive physician-collected high-risk HPV tests or cervical cytology of $\geq$ASC-US and 752 had positive self-test for high-risk HPV but had negative physician-collected high-risk HPV tests and normal cervical cytology. Of the 3,562 that should have had colposcopy and biopsy, 3,060 had colposcopy with biopsy and 502 did not. Four hundred and sixty of the 502 women that failed to have colposcopy with biopsy were evaluated late in the study and had negative self and physician-collected high-risk HPV tests and cervical cytology of ASC-US. The remaining 42 women that failed to have colposcopy and biopsy

TABLE I – DISTRIBUTION OF COLPOSCOPY AND CIN 2 OR WORSE AMONG 8,497 WOMEN IN SHANXI PROVINCE CERVICAL CANCER SCREENING PROGRAM (SPOCCS) II

	Colposcopy		Pathology with 5-biopsy gold standard		Total
	# No colpo	# Colpo	# CIN 1 or less	# CIN 2 or worse	
HPV Pos, Cytology $\geq$ ASC-US					
VIA Neg	8	776	622	162	784
VIA Pos	4	271	116	159	275
Total	*12*	*1,047*	*738*	*321*	*1,059*
HPV Pos, Cytology Normal					
VIA Neg	25	848	840	33	873
VIA Pos	1	82	74	9	83
Total	*26*	*930*	*914*	*42*	*956*
HPV Neg, Cytology $\geq$ ASC-US					
VIA Neg	440	300	734	6	740
VIA Pos	24	31	51	4	55
Total	*464*	*331*	*785*	*10*	*795*
HPV Neg, Cytology Normal					
VIA Neg	4,532	759	5,289	2	5,291
VIA Pos	324	72	396	0	396
Total	*4,856*	*831*	*5,685*	*2*	*5,687*
Total	*5,358*	*3,139*	*8,122*	*375*	*8,497*

Colpo is colposcopy, CIN is cervical intraepithelial neoplasia, VIA is acetic acid-aided visual inspection, HPV is physician-collected high-risk human papillomavirus, ASC-US is atypical squamous cells of uncertain significance, Neg is negative, and Pos is positive.

TABLE II – TYPE OF BIOPSY AND NUMBER OF CERVICAL QUADRANTS INVOLVED AMONG 375 WOMEN WITH CIN 2 OR WORSE

Pathology diagnosis	Type	Number of cervical quadrants involved with CIN 2 or worse					
		0	1	2	3	4	Total
CIN 2	Colpo	1*	55	23	5	2	86
	Random/ECC	7	62	15	2	1	87
	Total	8	117	38	7	3	173
CIN 3	Colpo	0	25	40	23	18	106
	Random/ECC	11	34	20	9	1	75
	Total	11	59	60	32	19	181
SCC	Colpo	0	1	2	7	6	16
	Random/ECC	1	0	3	1	0	5
	Total	1	1	5	8	6	21
Total	*Colpo*	*1*[1]	*81*	*65*	*35*	*26*	*208*
	Random/ECC	*19*	*96*	*38*	*12*	*2*	*167*
	Total	*20*	*177*	*103*	*47*	*28*	*375*

[1]One woman had CIN 2 diagnosed on a colposcopic-directed vaginal vault biopsy. CIN is cervical intraepithelial neoplasia, ECC is endocervical curettage, Colpo is diagnosis by colposcopic-directed biopsy with or without positive ECC, Random/ECC is diagnosis by random biopsy with or without positive ECC or solely by positive ECC.

TABLE III – TYPE OF BIOPSY AND NUMBER OF CERVICAL QUADRANTS INVOLVED AMONG 202 WOMEN WITH CIN 3 OR CANCER

Pathology diagnosis	Biopsy type	Number of cervical quadrants involved with CIN 3 or worse					
		0	1	2	3	4	Total
CIN 3	Colpo	1*	50	34	7	7	99
	Random/ECC	23	44	11	4	0	82
	Total	24	94	45	11	7	181
SCC	Colpo	0	2	5	5	4	16
	Random/ECC	1	1	3	0	0	5
	Total	1	3	8	5	4	21
Total	*Colpo*	*1*[1]	*52*	*39*	*12*	*11*	*115*
	Random/ECC	*24*	*45*	*14*	*4*	*0*	*87*
	Total	*25*	*97*	*53*	*16*	*11*	*202*

[1]One woman had CIN 2 in a cervical quadrant with negative ECC and was diagnosed with CIN 3 solely on the basis of a vaginal vault biopsy. CIN is cervical intraepithelial neoplasia, ECC is endocervical curettage, Colpo is diagnosis by colposcopic-directed biopsy with or without positive ECC, Random/ECC is diagnosis by random biopsy with or without positive ECC or solely by positive ECC.

were self-test high-risk HPV negative with positive physician-collected high-risk HPV or abnormal cervical cytology that failed to return to the clinic for a second visit. Seventy-nine women with negative self and physician-collected high-risk HPV tests and normal cervical cytology, who by protocol should not have had colposcopy with biopsy, had the procedure. Sixteen of these 79 women had colposcopy solely because their VIA was positive. The remaining 63 women with all 4 negative screening tests were randomly selected for colposcopy and biopsy. None of these 79 women with negative self and physician-collected high-risk HPV tests and normal cytology had CIN 2 or worse.

As shown in Table I, the rate of CIN 2 or worse among women that had colposcopy and biopsy was 321/1047 (30.7%) for positive physician-collected high-risk HPV tests and positive cervical cytology, 42/930 (4.5%) for positive physician-collected high-risk HPV tests with negative cervical cytology, 10/331 (3.0%) for negative physician-collected high-risk HPV tests with cervical cytology of $\geq$ASC-US and 2/831 (0.2%) for negative physician-collected high-risk HPV tests and normal cervical cytology. Three hundred and fifty-three of 809 VIA positive women did not have colposcopy and biopsy. Three hundred and twenty-four of these 359 VIA positive women that did not have colposcopy and biopsy had negative physician-collected high-risk HPV and normal cervical cytology, 24 had negative physician-collected high-risk HPV and cervical cytology of $\geq$ASC-US, 1 had positive physician-collected high-risk HPV and normal cervical cytology and 4 had positive physician-collected high-risk HPV and cervical cytology of $\geq$ASC-US. By multiplying the number of VIA positive women that did not have colposcopy by the rate of CIN 2 or worse in each of the 4 subgroups and then adding, we estimate that about $(324)(0.002) + (24)(0.03) + (1)(0.045) + 4(0.307) = 2.06$ women with CIN 2 or worse and positive VIA were not detected.

Three hundred and seventy-five of the 8,497 (4.4%) women participating in SPOCCS II had CIN 2 or worse. The distribution of CIN 2 or worse, type of biopsy (colposcopic-directed with or without positive ECC *vs.* random with or without positive ECC or solely by positive ECC) and the number of cervical quadrants involved with CIN 2 or worse are presented in Table II. Of the 375 women with CIN 2 or worse, 173 had CIN 2, 181 had CIN 3 and 21 had invasive cervical cancer. Cervical intraepithelial neoplasia 2 or worse involved only the ECC in 19 women, the vaginal wall in 1 woman, 1 quadrant of the cervix in 177, 2 quadrants in 103, 3 quadrants in 47 and 4 quadrants in 28 women. Involvement of 3–4 cervical quadrants was found in 10 of 173 (5.8%) women with CIN 2, 51/181 (28.2%) women with CIN 3 and 14/21 (66.7%) women with invasive cancer (5.8% *vs.* 28.2% *vs.* 66.7%, $p < 0.001$). Colposcopic-directed biopsy with or without ECC showing CIN 2 or worse resulted in a diagnosis of CIN 2 in 86/173 (49.7%) women, CIN 3 in 106/181 (58.6%) women and invasive cancer in 16/21 (76.2%) women (49.7% *vs.* 58.6% *vs.* 76.2%, $p < 0.05$). Two hundred and eight of 375 (55.5%) women with CIN 2 or worse were diagnosed by a colposcopic-directed biopsy with or without an ECC showing CIN 2 or worse, 148 were diagnosed by a random biopsy from a cervical quadrant without a colposcopic-detected lesion with or without an ECC showing CIN 2 or worse and 19 were diagnosed solely by an ECC showing CIN 2 or worse. Of the 19 women with CIN 2 or worse diagnosed solely by a positive ECC, 7 had CIN 2, 11 had CIN 3 and 1 had invasive cancer.

Two hundred and two of the 8,497 (2.4%) women were diagnosed with CIN 3 or cancer. The type of biopsy and number of quadrants involved with CIN 3 or cancer are presented in Table III. When the endpoint is CIN 3 or cancer, cervical quadrants and ECCs that were interpreted as normal, CIN 1 or CIN 2 were coded as negative. Cervical intraepithelial neoplasia 3 or cancer was present only in the ECC in 24 women, only in a vaginal vault biopsy in 1 woman, involved 1 quadrant of the cervix in 97, 2 quadrants in 53, 3 quadrants in 16 and 4 quadrants in 11 women. One hundred and fifteen of 202 (56.9%) women with CIN 3 or cancer were diagnosed by a colposcopic-directed biopsy with or without an ECC showing CIN 3 or cancer, 63/202 (31.2%) were diagnosed by a random biopsy with or without an ECC showing CIN 3 or cancer, and 24/202 (11.9%) were diagnosed solely by an ECC showing CIN 3 or cancer. Among the 24 women with CIN 3 or cancer diagnosed solely by a positive ECC, there are 11 with CIN 3 and 1 with invasive cancer that appear in the 19 women with ECCs showing CIN 2 or worse (see first column of Table II). In addition, there are 7 women that had an ECC of CIN 3 with a random biopsy of CIN 2 and 5 women with an ECC of CIN 3 and a colposcopic-directed biopsy of CIN 2.

As shown in Table IV, colposcopic-directed biopsy is more likely to detect CIN 2 or worse when the CIN 2 or worse involves 3–4 quadrants of the cervix (81.3%) than when the CIN 2 or worse involves 0–2 quadrants of the cervix (49.0%) ($p < 0.001$); similar results were found for an endpoint of CIN 3 or cancer [85.2% for 3–4 quadrants *vs.* 52.6% for 0–2 quadrants, ($p = 0.0014$)]. With an endpoint of CIN 2 or worse, the sensitivities of VIA, cervical cytology of $\geq$ASCUS, cervical cytology of $\geq$LSIL and cervical cytology of $\geq$HSIL are higher for lesions involving 3–4 quadrants than for those involving 0–2 quadrants. In contrast, the sensitivities for CIN 2 or worse of self-test and physician-collected test for high-risk HPV are not functions of lesion size. With an endpoint of CIN 3 or cancer, the sensitivity of cervical cytology of $\geq$HSIL is higher for lesions involving 3–4 quadrants than for

TABLE IV – SENSITIVITY FOR HIGH-GRADE (CIN 2 OR WORSE OR CIN 3 OR CANCER) LESIONS BY DIFFERENT SCREENING METHODS, STRATIFIED BY NUMBER OF CERVICAL QUADRANTS INVOLVED

Screening test	CIN 2 or worse			CIN 3 or cancer		
	0–2 quadrants of cervix involved (N = 300) (%)	3–4 quadrants of cervix involved (N = 75) %	p	0–2 quadrants of cervix involved (N = 175) %	3–4 quadrants of cervix involved (N = 27) %	p
Colposcopic directed biopsy	49.0	81.3	<0.001	52.6	85.2	0.0014
Visual inspection positive	41.7	62.7	<0.001	52.0	51.9	0.99
Cytology						
≥ASC-US	85.3	96.0	0.013	93.7	96.3	0.60
≥LSIL	74.3	94.7	<0.001	86.9	96.3	0.16
≥HSIL	49.0	76.0	<0.001	64.6	88.9	0.01
HPV physician positive	96.3	98.7	0.30	97.1	96.3	0.81
HPV self-test positive	86.3	92.0	0.18	86.3	88.9	0.71

CIN is cervical intraepithelial neoplasia, ASC-US is atypical squamous cells of uncertain significance, LSIL is low-grade squamous intraepithelial lesion, HSIL is high-grade squamous intraepithelial lesion, HPV self-test is high-risk human papillomavirus obtained from the upper vagina, and HPV physician is high-risk human papillomavirus obtained from the endocervix.

TABLE V – COMPARISON OF SENSITIVITY OF SCREENING TESTS AND COLPOSCOPY WITH 5-BIOPSY AND COLPOSCOPIC-DIRECTED BIOPSY GOLD STANDARDS WITH ENDPOINTS OF CIN 2 OR WORSE AND CIN 3 OR CANCER

Screening test	CIN 2 or worse			CIN 3 or cancer		
	Sensitivity with 5-biopsy gold standard (≥CIN 2, N = 375) (%)	Sensitivity with colposcopic-directed biopsy standard (≥CIN 2, N = 208) (%)	p	Sensitivity with 5-biopsy gold standard (≥CIN 3, N = 202) (%)	Sensitivity with colposcopic-directed biopsy standard (≥CIN 3, N = 115) (%)	p
Colposcopic directed biopsy	55.5	100.0	<0.0001	56.9	100.0	<0.0001
Visual inspection positive	45.9	65.9	<0.0001	52.0	70.4	0.0013
Cytology						
≥ASC-US	88.2	90.4	0.26	94.1	95.7	0.55
≥LSIL	78.4	82.7	0.24	88.1	90.4	0.53
≥HSIL	54.4	58.7	0.31	67.8	72.2	0.42
HPV physician positive	96.8	97.6	0.57	97.0	97.4	0.85
HPV self-test positive	87.5	89.4	0.45	86.6	92.2	0.14

CIN is cervical intraepithelial neoplasia, ASC-US is atypical squamous cells of uncertain significance, LSIL is low-grade squamous intraepithelial lesion, HSIL is high-grade squamous intraepithelial lesion, HPV self-test is high-risk human papillomavirus obtained from the upper vagina, and HPV physician is high-risk human papillomavirus obtained from the endocervix.

TABLE VI – COMPARISON OF SPECIFICITY OF SCREENING TESTS WITH 5-BIOPSY AND COLPOSCOPIC-DIRECTED BIOPSY GOLD STANDARDS WITH ENDPOINTS OF CIN 2 OR WORSE AND CIN 3 OR CANCER

Screening test	CIN 2 or worse			CIN 3 or cancer		
	Specificity with 5-biopsy gold standard (≤CIN 1, N = 8,122) (%)	Specificity with colposcopic-directed biopsy standard (≤CIN 1, N = 8289) (%)	p	Specificity with 5-biopsy gold standard (≤CIN 2, N = 8295) (%)	Specificity with colposcopic-directed biopsy standard (≤CIN 2, N = 8382) (%)	p
Visual inspection positive	92.2	91.9	0.97	91.5	91.3	0.65
Cytology						
≥ASC-US	81.3	79.9	0.03	79.9	79.1	0.23
≥LSIL	93.2	91.8	<0.01	91.9	91.1	0.06
≥HSIL	98.7	97.7	<0.01	97.9	97.2	<0.01
HPV physician positive	79.7	78.1	0.02	78.1	77.3	0.23
HPV self-test positive	77.2	76.0	0.06	75.9	75.3	0.39

CIN is cervical intraepithelial neoplasia, ASC-US is atypical squamous cells of uncertain significance, LSIL is low-grade squamous intraepithelial lesion, HSIL is high-grade squamous intraepithelial lesion, HPV self-test is high-risk human papillomavirus obtained from the upper vagina, and HPV physician is high-risk human papillomavirus obtained from the endocervix.

lesions involving 0–2 quadrants while the sensitivities of the other screening tests are not functions of lesion size.

As shown in Table V, the sensitivity for CIN 2 or worse of VIA is 65.9% when determined by a gold standard of colposcopic-directed biopsy but only 45.9% when determined by the 5-biopsy gold standard. This 20.0% difference in sensitivity for CIN 2 or worse of VIA is significant ($p < 0.001$). Similarly, the sensitivity for CIN 3 or cancer of VIA is 70.4% when the gold standard is colposcopic-directed biopsy but only 52.0% with the 5-biopsy gold standard. This 18.4% difference is significant ($p = 0.001$). With neither endpoint (CIN 2 or worse or CIN 3 or cancer) was inflation of screening tests other than VIA found when the gold standard was changed from colposcopic-directed biopsy to the 5-biopsy standard. As shown in Table VI, though some of the differences in specificity of screening tests are statistically significant, even the greatest difference of 1.4% seen for cytology of ≥LSIL with an endpoint of CIN 2 or worse is not of clinical significance.

We examined the relationship between the concordances of the screening tests with diagnosis by colposcopic-directed biopsy (as determined by the kappa statistic). The 2-by-2 tables correlating the screening tests with colposcopic-directed biopsy using CIN 2 or worse as the endpoint are shown in Table VII; similar data using CIN 3 or cancer as the endpoint is shown in Table VIII. When the endpoint is CIN 2 or worse, VIA ($p < 0.0001$) and cytology of ≥LSIL ($p = 0.024$) have significant correlation with diagnosis by colposcopic-directed biopsy, while with an endpoint of CIN 3 or cancer, VIA ($p < 0.0001$) and self-test for high-risk HPV ($p = 0.008$) have significant correlations with diagnosis by colposcopic-directed biopsy.

TABLE VII – CORRELATION OF SCREENING TESTS AND DIAGNOSIS BY COLPOSCOPIC-DIRECTED BIOPSY USING CIN 2 OR WORSE AS ENDPOINT

Screening test	Test pos		Test neg		Kappa	*p*
	Colpo pos	Colpo neg	Colpo pos	Colpo neg		
Visual inspection	137	35	71	132	0.440	<0.0001
≥ASC-US	188	140	20	27	0.071	0.057
≥LSIL	172	122	36	45	0.102	0.024
≥HSIL	122	82	86	85	0.095	0.065
HPV physician-collected pos	203	160	5	7	0.020	0.329
HPV self-test pos	186	142	22	25	0.047	0.202

Acetic acid-aided visual inspection is visual inspection, colpo is diagnosis of CIN 2 or worse by colposcopic-directed biopsy with or without endocervical curettage of CIN 2 or worse, cervical cytology of ≥atypical squamous cells of uncertain significance is ≥ASC-US, cervical cytology of ≥low-grade squamous intraepithelial lesion is ≥LSIL, cervical cytology of ≥high-grade squamous intraepithelial lesion is ≥HSIL, positive is pos, negative is neg, and HPV is high-risk human papillomavirus.

TABLE VIII – CORRELATION OF SCREENING TESTS AND DIAGNOSIS BY COLPOSCOPIC-DIRECTED BIOPSY USING CIN 3 OR CANCER AS ENDPOINT

Screening test	Test pos		Test neg		Kappa	*p*
	Colpo pos	Colpo neg	Colpo pos	Colpo neg		
Visual inspection	81	24	34	63	0.423	<0.0001
≥ASC-US	110	80	5	7	0.041	0.271
≥LSIL	104	74	11	13	0.059	0.242
≥HSIL	83	54	32	33	0.104	0.128
HPV physician-collected pos	112	84	3	3	0.010	0.728
HPV self-test pos	106	69	9	18	0.140	0.008

Acetic acid-aided visual inspection is visual inspection, colpo is diagnosis of CIN 3 or worse by colposcopic-directed biopsy with or without endocervical curettage of CIN 3 or worse, cervical cytology of ≥atypical squamous cells of uncertain significance is ≥ASC-US, cervical cytology of ≥low-grade squamous intraepithelial lesion is ≥LSIL, cervical cytology of ≥high-grade squamous intraepithelial lesion is ≥HSIL, positive is pos, negative is neg, and HPV is high-risk human papillomavirus.

As the kappa statistic measuring correlation of screening tests and colposcopic-directed biopsy and inflation of sensitivity secondary to a correlated gold standard are both functions of the 4 numbers in the 2-by-2 tables seen in Tables VII and VIII, a mathematical relationship between the 2 exists. The equation for the ratio of inflation of sensitivity of screening tests secondary to a correlated gold standard divided by kappa for the correlation of the screening tests with diagnosis by colposcopic-directed biopsy is derived in the Appendix.

Discussion

In our previous report, we concluded that, as cytology of ≥HSIL and VIA missed CIN 2 or worse involving 0–2 quadrants of the cervix and colposcopic-directed biopsy missed CIN 2 or worse involving 0–2 quadrants of the cervix, inflation of sensitivity for CIN 2 or worse of these 2 screening tests was likely to occur when a gold standard of colposcopic-detected biopsy was employed.[7] We attributed our failure to detect significant inflation of sensitivity for CIN 2 or worse of these 2 screening tests to a lack of power. Our current review of roughly 4 times as many women confirms that significant inflation of sensitivity of correlated screening tests occurs when colposcopic-directed biopsy is the gold standard for CIN 2 or worse and provides more detail in explaining why this occurs. In contrast to our prior conclusion, lesion size is not the important factor in determining inflation of sensitivity of CIN 2 or worse (or CIN 3 or cancer) when colposcopic-directed biopsy is the gold standard. This is shown (Tables IV and V) for CIN 2 or worse in which, like colposcopy, 4 of the screening tests missed more CIN 2 or worse involving 0–2 cervical quadrants, yet only 1 of the 4 (VIA) had significant inflation of sensitivity when the gold standard of colposcopic-directed biopsy was compared to a 5-biopsy standard. It is even clearer for CIN 3 or cancer where colposcopy and cytology of ≥HSIL missed more CIN 3 or cancer involving 0–2 cervical quadrants, yet only VIA had significant inflation of sensitivity when the gold standard of colposcopic-directed biopsy was compared to that of a 5-biopsy standard.

As shown by the derivation of a mathematical relationship, agreement between the screening test and colposcopic-directed biopsy (as measured by kappa) is the significant factor in predicting inflation of sensitivity of CIN 2 or worse secondary to using colposcopic-directed biopsy as the gold standard. It is not enough that the gold standard of colposcopic-directed biopsy and the screening test both miss more lesions involving 0–2 quadrants of the cervix. For inflation of sensitivity to occur, the screening test and the gold standard must miss at least some of the exact same (whether small or large) lesions.

Repeating the analysis to determine the factors responsible for inflation of sensitivity of screening tests secondary to inaccurate gold standard with an endpoint of CIN 3 or cancer was done not only because CIN 2 is not a reliable indicator of neoplsia,[1] but also because lesion size increases with increasing grade of CIN. As shown in Table II, 3–4 quadrants of the cervix were involved with CIN 2 or worse in only 5.8% of women when the worst biopsy was CIN 2 compared with 28.1% when the worst biopsy was CIN 3 and 66.7% when the worst biopsy was invasive cancer. As our initial hypothesis was that inflation of sensitivity of screening tests secondary to the inaccurate gold standard of colposcopic-directed biopsy would be seen when the gold standard and the screening test both missed small lesions, testing the hypothesis with an endpoint of CIN 3 or cancer seemed prudent. When the analysis was repeated with an endpoint of CIN 3 or cancer, there was significant correlation of VIA with colposcopic-directed biopsy ($p < 0.001$) and there was significant inflation of sensitivity of VIA when a colposcopic-directed gold standard was compared with a 5-biopsy standard (18.4%, $p = 0.0013$).

As 3,139 of the 8,497 participants in SPOCCS II and 456 of the 809 women with positive VIA in SPOCCS II had colposcopy and biopsy to verify the diagnosis, the possibility of verification bias must be considered. The magnitude of verification bias in a cervical cancer screening study is low if the number of women with CIN 2 or worse among those that did not have colposcopy and biopsy is small. To limit cost and still minimize verification bias, in SPOCCS II, colposcopy and biopsy was planned for the 3,562 women that had positive self or physician-collected high-risk HPV tests or abnormal cervical cytology. Verification bias could have been minimized without referring self-test high-risk HPV positive yet physician-collected high-risk HPV negative and cytology negative women for colposcopy and biopsy as virtually all women with CIN 2 or worse have either a positive physician-collected high-risk HPV test or an abnormal cervical cytology. Data to sup-

port this assertion comes from the current trial in which there were only 2 cases of CIN 2 or worse among 831 women that had negative physician-collected high-risk HPV tests and normal cervical cytology, from SPOCCS I where there were no cases of CIN 2 or worse among 1,280 women with negative physician-collected high-risk HPV tests and normal cervical cytology,[5] and from the recent review conducted by Arbyn *et al*.[11] Verification bias in SPOCCS II is greater than originally planned as 502 of the 3,562 women that should have had colposcopy and biopsy did not have this evaluation. Though most (460) of the 502 women that did not have colposcopy came from the subgroup that had negative self and physician-collected high-risk HPV tests and had ASC-US cytology in which the risk of CIN 2 or worse is about 1%,[12] 42 of these women had other constellations of positive high-risk HPV tests and/or abnormal cervical cytology in which the risk of CIN 2 or worse is higher. For the analysis of inflation of sensitivity of VIA in particular, there must be few cases of CIN 2 or worse among the 353 women that were VIA positive but did not have colposcopy and biopsy. Using the rates of CIN 2 or worse in the 4 subgroups based on result of physician-collected high-risk HPV and cervical cytology, we estimated that 2 VIA positive women with CIN 2 or worse were not detected in SPOCCS II.

We conclude that, despite the verification bias inherent in SPOCCS II, inflation of sensitivity of VIA for CIN 2 or worse (and CIN 3 or cancer) is found when colposcopic-directed biopsy is used as the gold standard in cervical cancer screening studies. This inflation is secondary to correlation between VIA and colposcopic-directed biopsy and might have been expected as VIA and colposcopy both use acetic acid enhancement to detect CIN by visual means. The magnitude of the inflation of sensitivity of VIA in a given study cannot be determined without the results of a second, more accurate, gold standard.

Inflation of sensitivity of VIA associated with correlation of VIA with the gold standard of colposcopic-directed biopsy could be one explanation for the high values of apparent sensitivity of VIA reported by Sankaranarayanan *et al*.[13] We suspect that other screening tests that use optical methodology (*e.g*., cervicography[14] and speculoscopy[15]) are also likely correlated with colposcopy and hence also likely to have inflated sensitivity when colposcopic-directed biopsy is the gold standard for CIN 2, CIN 3, or cancer.

We advise a 5-biopsy gold standard for diagnosing CIN not only in cervical cancer screening studies but also in evaluation of women with abnormal cervical cytology. As noted previously, the multiple biopsies performed in SPOCCS I and II were obtained with a bronchoscopic biopsy instrument that takes 2-mm biopsies. In our experience, women have less discomfort with 4 cervical biopsies performed with the 2-mm instrument than they have with 1 biopsy performed with the usual biopsy instrument (*e.g*., baby Tischler). Though processing and interpretation of 2-mm biopsies may appear difficult, in neither the Cleveland Clinic where these biopsies have been the standard colposcopic biopsies for over 20 years nor in Beijing during SPOCCS I and II has this been a significant problem. SPOCCS I and II were performed in the same population and though the sensitivity of colposcopic-directed biopsy in SPOCCS II (56.3%) was lower than that in SPOCCS I (75.6%), utilizing the 5-biopsy gold standard resulted in virtually identical risks of CIN 2 or worse in the 2 studies.[5,6] We conclude that utilizing the 5-biopsy gold standard in colposcopy clinics would avoid missed diagnosis of CIN 2 or worse secondary to variation in expertise of colposcopy. As fewer cases of CIN 2 or worse would be missed at the initial colposcopic evaluation, adopting the 5-biopsy standard would also likely decrease the need for follow-up after a colposcopic diagnosis of CIN 1 or less.[16]

If the 5-biopsy standard is adopted for evaluation of women with abnormal cervical cytology, the rules for treatment of CIN may need to be changed. As shown in Table II, the CIN 2 or worse diagnosed by random biopsy or solely by positive ECC is smaller and of lower grade than that diagnosed by colposcopic-directed biopsy. Given that CIN 2 is a poorly defined entity and may resolve without treatment,[1] removal or ablation of the transformation zone in young women with small CIN 2 may not be indicated.[17] Ablation or excision of the transformation zone in women with CIN 2 is advised in older women that have large lesions, are unlikely to return for follow-up, or when the diagnosis of CIN 2 is uncertain (*e.g*., when the cervical cytology is HSIL, the ECC is positive for CIN 2, or the pathologic interpretation is uncertain).

References

1. ASCUS-LSIL Triage Study (ALTS) Group. Results of a randomized trial on the management of cytology interpretations of atypical squamous cells of undetermined significance. Am J Obstet Gynecol 2003;188:1383–92.
2. Sankaranarayanan R, Wesley R, Thara S, Dhakad N, Chandralekha B, Sebastian P, Chithrahara K, Parkin DM, Nair MK. Test characteristics of visual inspection with 4% acetic acid (VIA) and Lugol's iodine (VILI) in cervical cancer screening in Kerala, India. Int J Cancer 2003;106:404–8.
3. Wright TC, Jr, Denny L, Kuhn L, Pollack A, Lorincz A. HPV DNA testing of self-collected vaginal samples compared with cytologic screening to detect cervical cancer. JAMA 2000;283:81–6.
4. Herrero R, Hildesheim A, Bratti C, Sherman ME, Hutchinson M, Morales J, Balmaceda I, Greenberg MD, Alfara M, Burk RD, Wacholder S, Plummer M, et al. Population-based study of human papillomavirus infection and cervical neoplasia in rural Costa Rica. J Natl Cancer Inst 2000;92:464–74.
5. Belinson JL, Qiao YL, Pretorius R, Zhang WH, Elson P, Li L, Pan QJ, Fischer C, Lorincz A, Zahniser D. Shanxi province cervical cancer screening study. A cross sectional comparative trial of multiple techniques to detect cervical intraepithelial neoplasia. Gynecol Oncol 2001;83:439–44.
6. Belinson JL, Qiao YL, Pretorius RG, Zhang WH, Rong SD, Huang MN, Zhao FH, Wu LY, Ren SD, Huang RG, Washington MF, Pan QJ, Li, L, et al. Shanxi province cervical cancer screening study II: self-sampling for high-risk human papillomavirus compared to direct sampling for human papillomavirus and liquid based cervical cytology. Int J Gynecol Cancer 2003;13:819–26.
7. Pretorius RG, Kim RJ, Belinson JL, Elson P, Qiao YL. Inflation of sensitivity of cervical cancer screening tests secondary to correlated error in colposcopy. J Low Genital Tract Dis 2006;10:45–50.
8. Pretorius RG, Zhang WH, Belinson JL, Haung MN, Wu LY, Zhang X, Qiao YL. The relative importance of colposcopically directed biopsy, random cervical biopsy, and endocervical curettage in the diagnosis of cervical intraepithelial neoplasia II or worse. Am J Obstet Gynecol 2004;191:430–4.
9. Pepe MS. Incomplete data and imperfect reference tests. In: Pepe MS, ed. The statistical evaluation of medical tests for classification and prediction. Oxford: Oxford University Press, 2004.168–213.
10. Phelps CE, Hutson A. Estimating diagnostic test accuracy using a "fuzzy gold standard." Med Decis Making 1995;15:44–57.
11. Arbyn M, Buntinx F, Van Ranst M, Paraskevaidis E, Martin-Hirsch P, Dillner J. Virologic versus cytolgic triage of women with equivocal Pap smears: a meta-analysis of the accuracy to detect high-grade intraepithelial neoplasia. J Natl Cancer Inst 2004; 96:280–93.
12. Pretorius RG, Peterson P, Novak S, Azizi F, Sadeghi M, Lorincz AT. Comparison of two signal-amplification DNA tests for high-risk HPV as an aid to colposcopy. J Reprod Med 2002;47:290–6.
13. Sankaranarayanan R, Basu P, Wesley RS, Mahe C, Keita N, Mbalawa CC, Sharma R, Dolo A, Shastri SS, Nacoulma M, Nayama M, Somanathan T, et al.; IARC Multicentre Study Group on Cervical Cancer Early Detection. Accuracy of visual screening for cervical neoplasia: results from an IARC multicentre study in India and Africa. Int J Cancer 2004;110:907–13.
14. De Vuyst H, Clacys P, Njiru S, Muchiri L, Steyaert S, De Sutter P, Van Marck E, Bwayo J, Temmerman M. Comparison of pap smear, visual inspection with acetic acid, human papillomavirus DNA-PCR testing and cervicography. Int J Gynaecol Obstet 2005;89:120–6.
15. Lonky N, Mann WJ, Massad LS, Mutch DG, Blanco JS, Vasilev SA, Finan MA, Scotti RJ. Ability of visual tests to predict underlying cervical neoplasia. Colposcopy and speculoscopy. J Reprod Med 1995; 40:530–6.
16. Pretorius RG, Peterson P, Azizi F, Burchette RJ. Subsequent risk and presentation of cervical intraepithelial neoplasia (CIN) 3 or cancer after a colposcopic diagnosis of CIN 1 or less. Am J Obstet Gynecol 2006;195:1260–5.
17. American College of Obstetricians and Gynecologists. Evaluation and management of abnormal cervical cytology and histology in the adolescent: ACOG Committee Opinion No. 330. Obstet Gynecol 2006;107:963–8.
18. Bishop YMM, Fienberg SE, Holland PW, eds. Discrete multivariate analysis: theory and practice. Cambridge, MA: The MIT Press, 1975. 395–6.

Appendix

Figure A1 is a plot of inflation of sensitivity of the screening tests when a colposcopic-directed gold standard is compared with a 5-biopsy gold standard *versus* kappa for the correlation of the screening tests with diagnosis by colposcopic-directed biopsy. The points in Figure A1 appear to be on a straight line ($r = 0.998$) and the ratio of inflation of sensitivity secondary to the intermediate gold standard of colposcopic-directed biopsy to kappa appears to be 0.49. The r value is so close to 1.0 that we suspected a mathematical relationship between the 2 variables.

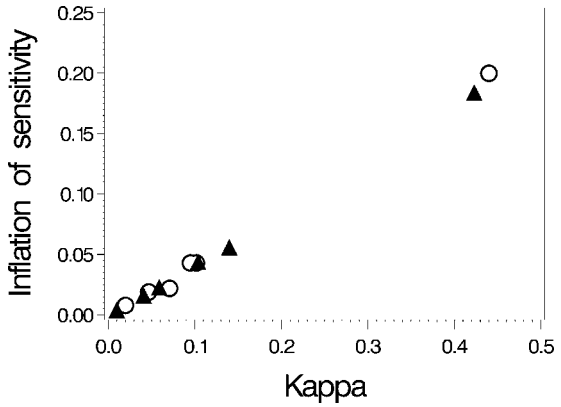

FIGURE A1 – Plot of inflation of sensitivity secondary to gold standard of colposcopic-directed biopsy *versus* kappa for correlation between screening tests and colposcopy with endpoints of CIN 2 or worse (open circles) and CIN 3 or cancer (solid triangles).

In the present case, the intermediate gold standard is colposcopic-directed biopsy. The cells in Table AI subdivide the positives of the universal "gold standard," in this case, the 5-biopsy standard. The sensitivity of a screening test determined by the 5-biopsy gold standard is $(A + B)/(A + B + C + D)$. The sensitivity of a screening test determined by the intermediate, colposcopic-directed biopsy gold standard is $A/(A + C)$. The difference between screening test sensitivity determined by an intermediate gold standard and the screening test sensitivity determined by the universal gold standard (defined in this article as inflation of sensitivity) is $[A/(A + C)] - [(A + B)/(A + B + C + D)]$. Kappa is $2(AD - BC)/[(A + B)(B + D) + (A + C)(C + D)]$.[18]

The ratio of inflation of sensitivity secondary to a correlated intermediate gold standard divided by kappa is $0.5 - [(A + B)/2][2/(A + B + C + D) - 1/(A + C)]$. In the situations described in this manuscript, the second term is always small compared to 0.5, so the ratio of inflation of sensitivity secondary to an inaccurate intermediate gold standard to kappa is always close to 0.5. The r value for the plot seen in Figure 1 is not exactly 1.0 and the ratio of inflation of sensitivity secondary to a correlated gold standard divided by kappa is not constant.

TABLE AI – MODEL OF A 2-BY-2 TABLE CORRELATING A SCREENING TEST AND AN INTERMEDIATE GOLD STANDARD WITHIN A POPULATION KNOWN TO BE POSITIVE BY A YET MORE UNIVERSAL "GOLD STANDARD"

	Intermediate gold standard positive	Intermediate gold standard negative	Universal gold standard positive
Screening test positive	A	B	A + B
Screening test negative	C	D	C + D
	A + C	B + D	A + B + C + D