Colposcopists' agreement on cervical biopsy site

L. Elit[1], *M.D.*; **J.A. Julian**[2, 3], *M.Sc.*; **J.W. Sellors**[4], *M.D.*; **M. Levine**[2, 3, 5], *M.D.*

[1]*Department of Obstetrics and Gynecology McMaster University, Hamilton, Ontario*
[2]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario*
[3]*Henderson Research Centre, Hamilton, Ontario*
[4]*Reproductive Health Program, PATH, Seattle, WA (USA)*
[5]*Department of Internal Medicine, McMaster University, Hamilton, Ontario (Canada)*

Summary

Objective: To determine the inter-observer agreement among colposcopists on the most abnormal area of the cervix from which a biopsy would be obtained and whether any attributes predict agreement.

Material and Methods: Fifty cervigrams were reviewed and 72 colposcopists from five countries indicated the site to biopsy and whether an ECC should be obtained. Prior to the study, six Canadian colposcopists met to achieve consensus on the most diseased area for biopsy. Consensus was also reached on whether an ECC was indicated. For each cervigram, percent agreement was determined between each study colposcopist and the consensus. Data were analyzed to determine the attributes associated with the consensus response.

Results: The percent overall agreement of the colposcopists with the consensus diagnoses had a mean of 0.70 (95% CI, 0.65-0.75). The use of ECC was most common in Canada (15% of cases). The following factors were assessed by multivariate analysis to determine their influence on individual agreement with the consensus recommendation for the site to biopsy: country, duration of practice (less than or greater than 1 year), professional group (nurse, family doctor, pathologist, gynecologist, gynecologic oncologist), expert status (recognized national/international expert vs colposcopist), and gender. No factor was significantly associated.

Conclusion: This international study was feasible and the level of inter-observer agreement among colposcopists on the location of the most severe lesions in cervical images is good.

Key words: Colposcopy reliability.

Introduction

Colposcopy still has a pivotal role as a diagnostic tool in the clinical management of women with abnormal cervical cytology. Colposcopy involves the assessment of the lower genital tract using 5-15 fold magnification after application of saline, 3-5% acetic acid and/or Lugol's solution. A biopsy of the most severely abnormal area influences subsequent management. The process of colposcopy was accepted as a standard of care in the 1970s with little assessment of its test performance. In the last decade, work has been conducted to document the accuracy and reproducibility of colposcopy [1, 2].

In the process of conducting an international randomized controlled trial assessing the best management strategy for biopsy proven CIN 1 [3], potential sources of measurement error needed to be assessed. Given that colposcopy is a diagnostic test with a significant component of observer interpretation, reproducibility needed to be evaluated. Prior work by Sellors suggests that agreement among experienced colposcopists is excellent (kappa > 0.90) for the site of the worst lesion [4]. Given that this randomized trial would involve many international centers, we undertook to measure observer variability among colposcopists of the most abnormal area of the cervix from which a biopsy should be obtained. We also evaluated those factors that could affect performance.

Material and Methods

For this study we used the cervigrams that had been obtained by Sellors, (for a prior work) in a standard manner after the application of 5% acetic acid [2]. Patients' cervical cytology and biopsy results were known. This information allowed the selection of a slide set (50 individual cases) that included a spectrum of normal, CIN 1, CIN 2, CIN 3 and cancer (Figure 1 and 2). However, this information was not used to define the site of the worst lesion. The selection of colpophotographs for our study over-represented the CIN 1 group, as this was the focus of the randomized trial (Table 1).

A group of six colposcopists from Ontario and Quebec assessed the slide set individually. They were designated as leaders by their peers for any one of the following reasons: they had authored chapters in textbooks or edited books on colposcopy, spoken at major colposcopy meetings, or were team leaders in their respective colposcopy units. They were asked to individually complete an assessment of cervigrams as if the patient presented with her first CIN 1 cervical cytology results. They were then asked to locate the most severe area of abnormality, if any, and indicate the o'clock position as if a clock face had been superimposed on the cervix. If no punch biopsy was recommended, they were asked to indicate whether an endocervical curetting should be obtained. They then met as a group to review the cervigrams. They were asked to come to consensus on where the most severe lesion was located. In other words, if a biopsy were taken anywhere else, the colposcopist would have missed the disease.

This project was made possible through the financial support of the Canadian Institute for Health Research
This paper was presented at the American Society for Colposcopy and Cervical Pathology March 2006.

Revised manuscript accepted for publication December 20, 2006

Table 1. — *Information on the cervigrams.*

Classification	Cytology	Colposcopy impression	Biopsy~
Benign	11	14	16
ASCUS	10		
HPV	18*	23	11
CIN 1			
CIN 2	9	10	10
CIN 3			
Cancer	1	1	1
Missing	1	2	1
Not done	0	0	11
Total	50	50	50

*At the time of this study, CIN 1 was used as both a cytology and histologic term in Ontario, Canada. Currently LSIL would be the cytologic term to describe the previous cytologic category of HPV and CIN 1.
~ biopsy could be from punch or excisional approach.

During 2002 and 2003, colposcopists from Canada, Brazil, Mexico, Scotland, and Australia reviewed the cervical photographs. The photographs were reviewed using the same sequence. In the majority of cases the images were viewed on a screen to produce an image 1 m wide. The observers sat 1.5 m from the image in a darkened room. In a small number of cases the slides were projected by CD-ROM onto a 15" computer screen (2 cases), or as 5"x7" photographs (2 cases). The colposcopist was asked to complete the assessment as if the patient presented with her first CIN 1 cervical cytology results. Again the colposcopists were asked to define the worst area for biopsy. They were asked to indicate one point on the clock face for biopsy. This could be one clock face number or an ECC or no biopsy. When this process occurred in a group context, no discussion was permitted until the end of the session.

The data were analyzed using SAS version 9.1. Percent agreement in relation to other colposcopists' choice, by consensus definition and by slide was determined. The location of the worst lesion was defined as a range (± 1 the o'clock reading) rather than the exact reading on the clock face given by an observer. Multivariable analysis using linear regression modeling was used to determine the factors: country, gender of the physician, duration of independent colposcopy practice (less than or greater than 1 year), profession (nurse, family practice, gynecologist, gynecologic oncologist, pathologist), and expertise (international expert vs clinic colposcopist) that affect agreement.

Results

Seventy-two experience colposcopists from 13 different cities in five countries (Canada, Brazil, Mexico, Scotland, Australia) reviewed the cervigrams and their results were compared to the consensus. The percent agreement of the colposcopists with the consensus diagnosis ranged from 0.42 to 0.94 with a mean of 0.70 (95% CI, 0.65-0.75) and a median at 0.68. There was no change in mean or median when the data from the four observers using photographs or CD-ROM were removed. The mean agreement with the 72 colposcopists ranged from 0.19 to 0.40 with a mean of 0.31 (95% CI, 0.26-0.36). There was no change in mean when the four cases using photographs or CD-ROM were removed. Twelve o'clock was the most common site to recommend for biopsy (22% of cases) (Table 2).

Table 2. — *Frequency of site for biopsy.*

Clock face hourly site	Frequency %	Number
1	7.4	268
2	7.9	283
3	4.3	156
4	1.5	53
5	2.8	101
6	8.0	287
7	4.9	177
8	2.4	85
9	2.4	85
10	3.2	115
11	9.2	331
12	22.0	791
ECC	11.0	397
No biopsy	10.1	362
Multiple sites	3.0	109

There were regional differences for recommending endocervical (ECC) curettage. In Canada, an ECC was recommended in 15% of cases compared to a rate of 2-8% in other countries.

In the multivariate linear modeling of the data from the 72 colposcopists (Table 3), none of the five factors was found to predict agreement with the consensus site for biopsy.

Table 3. — *Factors that may influence agreement with the consensus site for biopsying the cervix.*

Parameter	Number of colposcopists	Mean agreement with consensus	Standard deviation	p value
Country:				
Brazil	14	0.68	0.08	
Canada	46	0.71	0.10	0.73
Mexico	4	0.70	0.15	
Other	8	0.71	0.08	
Gender of clinician:				
Male	41	0.72	0.09	0.15
Female	31	0.69	0.10	
Duration of practice:				
< 1 yr	9	0.70	0.07	0.60
≥ 1 yr	63	0.71	0.10	
Profession:				
Pathologist	2	0.79	0.10	
Nurse	2	0.64	0.03	
Family doctor	3	0.77	0.03	0.42
Gynecologist	51	0.70	0.10	
Gynecologic Oncologist	14	0.71	0.09	
Expert:				
Yes	4	0.77	0.17	0.24
No	68	0.70	0.09	

Discussion

The premise under which colposcopy operates is that there are colposcopic criteria [5-7] which when applied give an impression of the diagnosis and define the area with the most abnormal changes on the cervix. A biopsy of the worst area will reflect the most significant disease in the cervix. Sellors has published a review on the accuracy and reproducibility of colposcopy as a diagnostic test [2]. Our paper adds to the understanding of colposcopic validity and reproducibility.

In our study, we evaluated the inter-observer agreement of the worst lesion from which a biopsy would be recommended. When assessing colposcopists who deliver care in large colposcopy clinics around the world concerning their level of agreement about where a biopsy should be taken, their level of agreement with a consensus diagnosis was average (0.70). The most common site for biopsy was 12 o'clock in keeping with the findings from the ALTS trial [8]. In our work, gender, experience, international/national recognition, professional discipline and country do not appear to influence performance.

The limitations of our study include the use of a convenience sample of images. These cervigrams were used as they were easily accessible, they had been obtained in a standardized manner and there was full information concerning referral cervical cytology, biopsy and disposition of each patient. A superior approach would have been to use a randomly selected sample of colposcopic images of the appropriate grades of lesions with and without acetic acid, at different magnifications and using the green filter. This would have better simulated the colposcopy scenario. Another limitation is that our study addresses internal consistency but accuracy is not clear. Although we had biopsy results from most of the cases, whether that histology correlates with the consensus diagnosis is not clear. Another limitation is that we used mixed methods. By and large, the cervigrams were usually reviewed by projection on a large screen however there were a few cases where this was not feasible so web-based or 5" x 7" photographs were used (< 8% of colposcopists). Another limitation is the application of the clock face to locate lesions on the cervix. The difficulty with the use of the clock face is that the lesion could be near the os or far away. Since the use of the clock face cannot distinguish between these two sites within a designated location, this could lead to over-estimation of agreement. A superior method of assessment would be to use computer simulation with an "x" to mark the site. Use of a kappa statistic would have been preferable to account for the rate of agreement by chance alone. However, this was not feasible. The consensus was either a point biopsy or a range of sites for biopsy as dictated by the worst appearing lesion on the cervigrams. As, well, the colposcopists were asked to provide only one biopsy site however in 3% of the cases, they chose multiple sites. In our study, colposcopists were exposed to 50 cervigrams. These could be reviewed in one hour. A larger sample of cervigrams would narrow the confidence intervals and increase our precision.

The purpose of this evaluation was to understand the level of agreement and the associated factors for colposcopic assessment of lesion location. The plan was that if one jurisdiction or individual had extremely poor results then that site/individual would not participate in the trial. If the problem of poor agreement on where to biopsy was generalized then we would have discussed and redefined criteria for selecting the most abnormal site. An additional approach would have been retraining and selection of colposcopists that met a criterion standard of agreement with the expert consensus. As a result of this process, where agreement was less than 0.65 (outside of the 95% confidence interval), that colposcopist did not recruit patients to the trial either because that center or those individuals choose not to participate. (This accounted for 2 centers and 2 individuals.)

Conclusions

We have found that it is feasible to have 72 colposcopists from various countries participate in an assessment of colposcopic skills. We have shown good internal consistency with a consensus standard for identifying the most abnormal area on the cervix, which should be biopsied in order to determine further management.

Acknowledgements

The authors which to acknowledge the support of the following physicians and colposcopy units: S. Bates and Brantford General Hospital, Brantford, Canada; J. Bentley and Queen Elizabeth II, Halifax, Canada; F. Coelho and Hospital General, Sao Paulo, Brazil; M. Cruickshank and Aberdeen Royal Infirmary, Scotland; J. Kamel and Mexico City, Mexico; T. Le and Royal University Hospital, Saskatoon, Canada; M. Fortier and St. Sacrement Hospital, Quebec City, Canada; A. Lytwyn and Women's College and Sunnybrook Hospitals, Toronto, Canada; R. Robbins and London Health Sciences Centre, London, Canada; P. Roth and Hamilton Health Sciences Centre, Hamilton, Canada; F. Russomano and Instituto Fernandes Figueira, Rio de Janeiro, Brazil; J. Tan and Royal Women's Hospital, Carlton, Australia; C. Zeferino and University of Campinas, Campinas, Brazil.

References

[1] Van Niekerk W.A., Dunton C.J., Richart R.M., Hilgarth M., Kato H., Kaufamn R.H. *et al.*: "Colposcopy, cervicography, speculoscopy and endoscopy". *IAC Task Force Summary Acta Cytol.*, 1998, *32*, 33.

[2] Sellors J.: "Screening Technology. Handbook on Cervical Cancer Screening". France, IARC/WHO. 2005.

[3] Elit L., Sellors J., Julian J., Lytwyn A., McLachlin M., Mahony J., *et al.*: "Randomized trial of immediate treatment versus colposcopic follow-up for biopsy proven CIN 1". Proceedings of the 22nd International Papillomavirus Conference, Vancouver, BC, Canada, May 2-6, 2005.

[4] Sellors J.W., Nieminen P., Vesterinen E., Paavonen J.: "Observer variability in the scoring of colpophotographs". *Obstet. Gynecol.*, 1990, *76*, 1006.

[5] Reid R., Scalzi P.: "Genital warts and cervical cancer: an improved colposcopic index for differentiating benign papillomaviral infections from high-grade cervical intraepithelial neoplasia". *J. Obstet. Gynecol.*, 1985, *153*, 611.

[6] Carreiro C., Di Gesu A., Conte R., Ferreri R., Loizzi P.: "Grading colposcopic appearance: paired comparison between two methods for differentiating benign papillomaviral infection from high-grade dysplasia of the uterine cervix". *Int. J. Gynecol. Obstet.* 1990, 34, 139.

[7] Coppleson M., Pixley E.: "Colposcopy of cervix". In: Coppleson M. (ed.). Gynecologic Oncology. Edinburgh: Churchill Livingstone, 1992, 297.

[8] Guido R.S., Jeronimo J., Schiffman M., Solomon D.: "The distribution of neoplasia arising on the cervix: Results from the ALTS trial". *Am. J. Obstet. Gynecol.*, 2005, *193,* 1331.

Address reprint requests to:
L. ELIT, M.D.
Juravinski Cancer Centre
699 Concession Street
L8V 5C2 Hamilton, Ontario (Canada)