RESEARCH COMMUNICATION

Consistency of Cytology Diagnosis for Cervical Cancer between Two Laboratories

Supannee Sriamporn[1], Onanong Kritpetcharat[2], Pekka Nieminen[3], Krittika Suwanrungraung[4], Supot Kamsa-ard[4], DM Parkin[5]

Abstract

The principal approach to the prevention of cancer of the cervix uteri has been through screening programmes, using the cervical smear (Pap test) to detect precursor lesions. The sensitivity and specificity of Pap smears depend on the skill of the observer in recognizing and classifying a variety of cellular abnormalities. We have studied the reproducibility of cytological diagnosis, according the Bethesada classification, made by cytologists in Khon Kaen, north-east Thailand, and in Helsinki, Finland, on smears taken from rural women undergoing screening during 1994-2001. A total of 313 slides were reviewed. The prevalence of abnormalities was relatively high, since the series included smears judged abnormal in Khon Kaen or from women who developed cancer during follow-up, as well as a group whose smears were negative.

In general, the reviewing cytologist in Finland evaluated more slides as abnormal than in the initial report. The level of agreement between the two observers was evaluated by calculating the coefficient of concordance (Kappa). The kappa score depended upon the degree of detail in the diagnosis; it was 0.43 for the presence or not of an epithelial abnormality (the General Categorization of the Bethesda system), and rather higher (0.5) for separating low grade from high grade (HSIL or worse) abnormalities or glandular lesions. Agreement was only fair (0.37) when the more detailed Bethesda categories (seven) were used.

The reproducibility of cervical cytology evaluations is critical to the success of screening programmes, and in this programme in a moderate-high risk population of women in rural Thailand, we found that agreement between skilled observers, at the level of tests requiring diagnostic follow-up or not, was only moderate. The women in this study are being traced to evaluate the true sensitivity of screening in terms of the lesions found on histology, during a prolonged follow up of 4 or more years.

Key Words: cytology - cervical cancer screening - reproducibility - Thailand

Asian Pacific J Cancer Prev, **6**, 208-212

Introduction

Cervical cancer is an important public health problem worldwide. It is the second most common cancer among women, with almost 80% of the cases occurring in developing countries, where, in many regions, it is the most common cancer of women (Parkin et al., 2001). Khon Kaen, in north-east Thailand, has an incidence that is moderately high, age-standardized rate (ASR) 16.8 per 100,000 in 1985-1999), and cervical cancer is the second most common cancer of women (after liver cancer) (Sriamporn et al., 2003, Sriplung et al., 2003) .

Until recently, the principal approach to the prevention of cancer of the cervix uteri has been through screening programmes, aimed at detection and treatment of precursor lesions. It is 50 years since Papanicolaou demonstrated that conventional cytology could be used to identify precancerous lesions of the cervix (Papanicolaou, 1954), and the cervical smear (Pap test) has been the mainstay of early detection ever since. It is generally accepted that it was the widespread introduction of organized cytological screening programmes that was responsible for the declines in incidence of (and

[1]Department of Epidemiology, Faculty of Public Health, [2]Department of Pathology, Faculty of Medicine, Khon Kaen University, Khon Kaen 40002, Thailand [3]Department of Obstetrics and Gynaecology, Helsinki Hospital, University of Helsinki, Finland [4]Cancer Unit, Srinagarind Hospital, Faculty of Medicine, Khon Kaen University, Thailand [5]Clinical Trial Service Unit and Epidemiological Studies Unit, Oxford University, UK
Corresponding Author: Dr Supannee Sriamporn, Thailand - Tel: +66 (0)43-347057; Fax: +66 (0)43-347058; e-mail: supannee@kku.ac.th

mortality from) cervical cancer in many countries, especially in North America, Europe, and Australia/New Zealand (Anttila et al., 2004, Hakama , 2005). However, experience in developing countries has not been as favourable, and, in most, screening activities have been relatively ineffective. There are several reasons, many of which relate to aspects of the logistics, or organization of screening activities. Thus, programmes often result in repeated screening of women at low risk, in inappropriate age groups, and may fail to ensure adequate diagnostic follow-up and appropriate treatment of those with abnormalities on cytology (Hakama, 2005). But, defects in the test itself may be partly responsible.

Screening tests are not intended to be diagnostic. They identify subjects with an increased probability of having the disease being sought, who must then undergo appropriate diagnostic investigations. Even with a highly reproducible test (based, for example, on biological measurement), screening tests yield a proportion of cases that prove to be false positive, or false negative, on diagnostic follow-up. The problem is compounded, however, when, as with exfoliative cytology, the test in question is subjective, and relies upon individual judgement concerning the grade of abnormality present, and how it may be categorized according to different classification schemes.

In this study, we compared the reproducibility of cytological diagnosis, according the Bethesada classification (Luff 1992), made by an examining cytologist in Khon Kaen, Thailand, and in Helsinki, Finland, on smears taken from women attending screening clinics in rural north-east Thailand during 1994-2001 .

Materials and Methods

The slides for study were those that could be traced in the Cancer Unit, and were judged to be of satisfactory quality for evaluation, taken from women who underwent a screening examination at the time of enrollment into a cohort study among villagers living in Khon Kaen Province, north-east Thailand.

There were three group of subjects; 1). Those who were considered to have had a positive Pap test (of any grade of severity). In this group there were 239 women. 2). Those subjects whose enrollment smear was evaluated as negative, but who were found later to have cancer (invasive or in situ) during follow up (57 cases). 3). Subjects selected as controls for these 57 cases (4 controls per each case), as part of a nested case-control study of cervical cancer and its antecedents (Sriamporn et al., 2004).

Out of the 3 groups, 313 slides could be traced. They were sent for review by an experienced cytopathologist in Finland (Cytologist II) in 2002. The results of the original cytological diagnosis in Khon Kaen (Cytologist I) were compared with those made on review. Comparisons were made at the level of presence/absence of epithelial cell abnormalities (with reactive, reparative, and inflammatory changes considered negative (no epithelial cell abnormality), whether the cytologists agreed on the presence or absence of glandular cells in the specimens, and on grade of abnormality diagnosed (when abnormal squamous cells were judged to be present).

The level of agreement was evaluated by calculating the coefficient of concordance (Kappa) (Cohen 1960, Landis and Koch 1977), whereby readings that are the same according to both observers are scored 1, and those that differ are scored zero. Kappa is calculated by summing all values and dividing by the total number.

The kappa-statistic measure of agreement is scaled to be 0 when the amount of agreement is what would be expected to be observed by chance and 1 when there is perfect agreement. For intermediate values, Landis and Koch (1977) suggest the following interpretations; Kappa below 0.0 means the agreement is poor, 0.00–0.20 = slight, 0.21–0.40 = fair, 0.41–0.60 = moderate, 0.61–0.80 = substantial, and 0.81–1.00 = almost perfect agreement respectively.

Results

The cytology was reported according to the Bethesda system (Luff 1992) in both laboratories. All of the slides selected for study had been reported as *"satisfactory for evaluation"* in Laboratory I (Khon Kaen) and this evaluation was confirmed by Laboratory II (Finland). The "general categorization" was the first level of reading, all specimens being classified as either *"negative for intraepithelial lesion or malignancy"* or as showing *"epithelial cell abnormalities"*.

In the original evaluation, 130 slides (41.5%) were judged to have an epithelial cell abnormality; the number was rather higher on review (157, 50.1%). Table 1 shows the level of agreement.

The category *"epithelial cell abnormalities"* includes both squamous and glandular cell abnormalities. This category is used whenever there are epithelial cell abnormalities, except for benign reactive or reparative changes.

Glandular cell abnormalities are categorized into four categories: *atypical glandular cells* (AGC), *atypical glandular cells - favor neoplasia*, *adenocarcinoma in-situ*, and *adenocarcinoma*. Whenever possible atypical glandular

Table 1. General Categorization of Findings Between both Cytologists

	Cytologist II		
	Negative for intraepithelial lesion or malignancy	Epithelial cell abnormality	Total
Cytologist I			
Negative for intraepithelial lesion or malignancy	125	58	183
Epithelial cell abnormality	31	99	130
Total	156	157	313

κ = 0.43, 95%CI = 0.32-0.54

Table 2. Presence of Glandular Cell Abnormalities

	Cytologist II			
	Negative	Presence of glandular cell abnormalities	Other epithelial cell abnormalities	Total
Cytologist I				
Negative	125	17	41	183
Presence of glandular cell abnormalities	2	20	4	26
Other epithelial cell abnormalities	29	7	68	104
Total	156	44	113	313

κ = 0.45, 95%CI = 0.36-0.53

cells are categorized as to whether they are endocervical or endometrial in origin.

Table 2 shows the evaluation of the two cytologists concerning the presence or absence of glandular cells among all subjects. Cytolgist II recognised glandular cell abnormalities in 44 cases (14%) and cytologist I in 26 (8.3%); there were 20 cases classified as such by both.

Squamous cell abnormalilities were reported in 105 slides at the original viewing (Cytologist I) and 125 on review of Cytologist II (Table 3).

Epithelial cell abnormalities are subdivided into four categories *"atypical squamous cells"* (ASC) is used when there are cytologic findings that are considered suggestive but not diagnostic of a squamous intraepithelial lesion. The "atypical squamous cell" category is formally subdivided into two subcategories: *"atypical squamous cells - of undetermined significance"* (ASC-US) and *"atypical squamous cells - cannot exclude a high-grade SIL"* (ASC-H). The *low-grade squamous intraepithelial lesion* (LSIL) category in the Bethesda System includes both HPV effects and mild dysplasia (CIN1). The *high-grade squamous intraepithelial lesion* (HSIL) category has a wide variation in cytological appearance ; it combines moderate and severe dysplasia together with carcinoma *in-situ*. The fourth category is *invasive squamous cell carcinoma* when frank malignant cells are observed in the smear.

Table 4 shows the evaluation of the two cytologists concerning the grade of squamous cell abnormality, amongst subjects in whom epithelial cell abnormalilies had been diagnosed among all subjects.

Table 5 shows the comparison of the two readers with respect to just 3 categories: normal or low-grade abnormality

Table 3. Presence of Squamous Cell Abnormalilities

	Cytologist II			
	Negative	Presence of squamous cell abnormalities	Other epithelial cell abnormalities	Total
Cytologist I				
Negative	125	45	13	183
Presence of squamous cell abnormalities	29	72	4	105
Other epithelial cell abnormalities	2	8	15	25
Total	156	125	32	313

κ = 0.43, 95%CI = 0.34-0.52

Table 4. Categorized Squamous Cell Abnormalities

	Cytologist II							
	Negative	ASC-US	ASC-H	LSIL	HSIL	Squamous cell carcinoma	Other epithelial cell	Total
Cytologist I								
Negative	125	25	4	7	8	1	13	183
ASC-US	25	19	3	3	9	0	4	63
ASC-H	0	0	0	0	0	0	0	0
LSIL	3	4	2	10	2	0	0	21
HSIL	1	0	2	0	11	0	0	14
Squamous cell carcinoma	0	0	0	0	3	4	0	7
Other epithelial cell	2	0	1	2	5	0	15	25
Total	156	48	12	22	38	5	32	313

κ = 0.37, 95%CI = 0.31-0.43

(ASC-US and LSIL), high -grade abnormality requiring diagnostic follow-up (HSIL (including ASC-H) and carcinoma), and glandular abnormality. The kappa score is moderate, because of good agreement on normal/low grade smears, and reasonable agreement between the observers for HSIL + (20 smears with a diagnosis in common, of 21 identified by cytologist I, and 55 by cytologist II).

Discussion

Many studies have suggested that reproducibility of cytological diagnoses, even between experienced cytologists, is quite low (Yobs et al., 1987, Klinkhamer et al., 1988, Woodhouse et al., 1999, Stoler & Schiffman, 2001). In general, cytopathologists appear to have more difficulty with recognizing and correctly interpreting glandular abnormalities than squamous abnormalities.

The Bethesda system was introduced in part to produce a more simplified and uniform classification scheme for reporting gynaecological cytology, thereby promoting interobserver agreement. It seems that it has not reduced this problem (Smith et al., 2000). In particular, the introduction of the "ASC-US" category seems to have provided a particularly non-reproducible receptacle for minor abnormalities. Currently approximately 4% to 5% of all cervical cytology specimens are classified as ASC in the United States (Jones & Davey, 2000). Agreement on ASC-US diagnosis in our study was low - of 63 slides labeled ASC-US in Khon Kaen, 25(39.7%) were judged to be normal on review, and 9 (14.3%) to show features of HSIL. Other studies have found the same problem (Sherman et al., 1994, Confortini et al., 2003).

ASC-US has a relatively low predictive value for high grade lesions - studies suggest that the prevalence of CIN 2,3 in women with ASC varies between 10% and 20% (Wright et al., 2002). As a result, management is a problem - recommendations are for a repeat smear, or for the use of HPV testing as a means of distinguishing abnormal cells of truly prognostic significance. Our study, like others, suggests that the ASC-US category has poor reproducibility, so that the value of this category for reporting cytology is doubtful. Previous studies have found that inter-observer agreement is better for the higher grades of abnormality, such as HSIL (Chhieng et al, 2001), although in the study reported here, there was still considerable disagreement, suggesting that, even for the better defined cytology categories, there are uncertainties in reporting criteria.

The kappa statistic is influenced by disease prevalence (Feinstein & Cicchetti, 1990). Normally, we might expect less than 5% of Pap tests to be abnormal, a circumstance in which concordance between reviewers will tend to be high, since the great majority of readings will be negative by both reviewers. In this study, the selection of subjects was such that a high proportion of cases was abnormal (almost half on the initial reading), which would result in rather lower concordance between observers. In addition, classification systems that utilize more diagnostic categories have inherently higher rates of variability than do classification systems with fewer diagnostic categories. Based on four categories (normal, ASC-US, LSIL and HSIL+) Stoler and Schiffman (2001) found only modest agreement (K = 0.46) between enrollment PAP smear report, and review by liquid-based cytology of 5000 specimens taken within the large US-based ALTS study, with a positivity rate of 37%. The largest source of disagreement was the ASC-US category, where, as in our study, 39.7% were judged to be negative on monolayer cytology (Yobs et al., 1987).

The reproducibility of cervical cytology evaluations is critical to the success of screening programmes. In our study, of smears taken from a moderate-high risk population of women in rural Thailand, we find that there is considerable interobserver variability between the original reading, and that during an independent review. In part, this was the result of a higher prevalence of smears being classified as having abnormalities on review. The discrepancies were greatest for the minor degrees of cytological abnormaility in the Bethesda classification, and the value of such a detailed categorization for screening purposes must be questionable. The women in this study are being traced to evaluate the true sensitivity of screening in terms of the lesions found on histology, during a prolonged follow up of 4 or more years.

Acknowledgements

This research is part of the project "a prospective study of cancer and other outcomes in a rural population in Thailand", which was approved by the research ethic committee, Faculty of Medicine, KKU. The study was

Table 5. Categorized Cell Abnormalities After Grouping into Three Categories: Normal or Low-grade Abnormality (ASC-US and LSIL), High-grade Abnormality (HSIL (including ASC-H) and Carcinoma), and Glandular Abnormality

	Cytologist II			
	LSIL or less	HSIL (including ASC-H) or worse	Glandular abnormality only	Total
Cytologist I				
LSIL or less	221	29	17	267
HSIL (including ASC-H) Or worse	1	20	0	21
Glandular abnormality only	4	6	15	25
Total	226	55	32	313

κ = 0.50, 95%CI = 0.42-0.58

supported in part by grants from Buddhist Aid Center of Japan, and the Association for International Cancer Research Scotland, UK, (grant number 98–55), whose contributions are gratefully acknowledged.

References

Anttila A, Ronco G, Clifford G, et al (2004). Weiderpass E. Cervical cancer screening programmes and policies in 18 European countries. *Br J Cancer*, **91**, 935-41.

Chhieng DC, Talley LI, Roberson J, et al (2002). Interobserver variability: comparison between liquid-based and conventional preparations in gynecologic cytology. *Cancer*, **96**, 67-73.

Cohen J (1960). A coefficient of aqreement for nominal scales. *Educational and psychological measurement*, **20**, 37-46.

Confortini M, Carozzi F, Dalla Palma P, et al (2003). Interlaboratory reproducibility of atypical squamous cells of undetermined significance report: a national survey. *Cytopathology*, **14**, 263-8.

Feinstein AR, Cicchetti DV (1990). High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol*, **43**, 543-9.

Hakama M (ed) (2005). Cervix Cancer Screening. IARC Handbooks of Cancer Prevention Volume 10. Lyon: International Agency for Research on Cancer .

Jones BA, Davey DD (2000). Quality management in gynecologic cytology using interlaboratory comparison. *Arch Pathol Lab Med*, **124**, 672-81.

Klinkhamer PJ, Vooijs GP, de Haan AF (1988). Intraobserver and interobserver variability in the diagnosis of epithelial abnormalities in cervical smears. *Acta Cytol*, **32**, 794-800.

Landis JR, Koch GG (1977). An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics*, **33**, 363-74.

Luff RD (1992). The Bethesda System for reporting cervical/vaginal cytologic diagnoses: report of the 1991 Bethesda workshop. The Bethesda System Editorial Committee. *Hum Pathol*, **23**, 719-21.

Papanicolaou GN (1954). Atlas of Exfoliative cytology, Boston, Massachusetts Commonweath Fund University press.

Parkin DM, Bray F, Ferlay J, Pisani P (2001). Estimating the world cancer burden: Globocan 2000. *Int J Cancer*, **94**, 153-6.

Sherman ME, Schiffman MH, Lorincz AT, et al (1994). Toward objective quality assurance in cervical cytopathology. Correlation of cytopathologic diagnoses with detection of high-risk human papillomavirus types. *Am J Clin Patho*, **102**, 182-7.

Smith AE, Sherman ME, Scott DR, et al (2000). Review of the Bethesda System atlas does not improve reproducibility or accuracy in the classification of atypical squamous cells of undetermined significance smears. *Cancer*, **90**, 201-6.

Sriamporn S, Pengsaa P, Hakama M, Suwanrungruang K, Parkin DM (2003). Cervix cancer in Khon Kaen, Northeast Thailand, 1985-1999. *Asian Pac J Cancer Prev*, **4**, 312-8.

Sriamporn S, Parkin DM, Pisani P, Suwanrungruang K, Pengsaa P (2004). Behavioural risk factors for cervical cancer from a prospective study in Khon Kaen, Northeast Thailand. *Cancer Detection and Prevention*, **28**, 334-9.

Sriplung H, Sontipong S, Martin N, et al (2003). Cancer in Thailand Vol. III 1995-1997. Bangkok Medical Publisher, Bangkok.

Stoler MH, Schiffman M (2001). Atypical Squamous Cells of Undetermined Significance-Low-grade Squamous Intraepithelial Lesion Triage Study (ALTS) Group. Interobserver reproducibility of cervical cytologic and histologic interpretations: realistic estimates from the ASCUS-LSIL Triage Study. *JAMA*, **285**, 1500-5.

Woodhouse SL, Stastny JF, Styer PE, et al (1999). Interobserver variability in subclassification of squamous intraepithelial lesions: Results of the College of American Pathologists Interlaboratory Comparison Program in Cervicovaginal Cytology. *Arch Pathol Lab Med*, **123**, 1079-84.

Wright TC, Cox JT, Massad LS, Twiggs LB, Wilkinson EJ (2002). 2001 Consensus Guidelines for the management of women with cervical cytological abnormalities. *JAMA*, **287**, 2120-9.

Yobs AR, Plott AE, Hicklin MD, et al (1987). Retrospective evaluation of gynecologic cytodiagnosis. II. Interlaboratory reproducibility as shown in rescreening large consecutive samples of reported cases. *Acta Cytol*, **31**, 900-10.