human immunodeficiency virus infection. Antimicrob Agents Chemother 35:2544–2550, 1991

(48) Ingber D, Fujita T, Kishimoto S, et al: Synthetic analogues of fumagillin that inhibit angiogenesis and suppress tumour growth. Nature 348:555–557, 1990

Notes

[1]In autocrine function, the cells producing a growth factor are the target cells for that growth factor. In paracrine function, cells other than those actually producing the growth factor are the target cells.

Affiliations of authors: J. M. Pluda, L. E. Shay, A. Foli, D. Adamo, R. Yarchoan (Medicine Branch, Clinical Oncology Program, Division of Cancer Treatment), P. J. Cohen (Dermatology Branch, Intramural Research Program, Division of Cancer Biology, Diagnosis, and Centers), M. R. Cooper (Clinical Pharmacology Branch, Clinical Oncology Program, Division of Cancer Treatment), S. Broder (Office of the Director), National Cancer Institute, National Institutes of Health, Bethesda, Md.

S. Tannenbaum (Clinical Pathology Department), B. R. Goldspiel (Department of Pharmacy), Warren Grant Magnuson Clinical Center, National Institutes of Health, Bethesda, Md.

Present address: J. M. Pluda, Investigational Drug Branch, Cancer Therapy Evaluation Program, Division of Cancer Treatment, National Cancer Institute.

Present address: P. J. Cohen, Departments of Pathology and Dermatology, University of Medicine and Dentistry, New Jersey Medical School, Newark, N.J.

Correspondence to: James M. Pluda, M.D., National Cancer Institute, 6130 Executive Blvd., Executive Plaza North, Rm. 715, Rockville, MD 20852.

We thank the AIDS clinical research nurses, the Medicine Branch medical staff, the nursing staff of the National Cancer Institute, and the pharmacy staff of the Warren Grant Magnuson Clinical Center for their help in the performance of this trial.

Manuscript received February 11, 1993; revised June 15, 1993; accepted July 6, 1993.

Improving Consistency in Cervical Cytology Reporting

*Heather Mitchell**

Background: **During the 1970s, the Papanicolaou method of classifying cervical cytology specimens and reporting diagnoses was replaced by more descriptive reporting systems. The plethora of reporting terms caused much confusion and a lack of standardization. To improve this situation, "The Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses" was approved at a National Cancer Institute Workshop in 1988. In Australia, the Victorian Cervical Cytology Registry (VCCR) was established in 1989. Because of the absence of a standard format for reporting cervical cytology in that country, a coding schedule was developed by local cytopathologists. While the pattern of reporting smear diagnoses was found to be reasonably consistent within individual laboratories, substantial variation in reporting abnormal cervical smear diagnoses by 29 laboratories in Victoria, Australia, was observed. In 1992, a working party of the National Health and Medical Research Council of Australia proposed that a modified Bethesda System be adopted by Australian laboratories.** ***Purpose:*** **The aim of this study was to promote more uniform reporting of cervical/vaginal cytologic diagnoses by cytopathology laboratories in Victoria, Australia.** ***Methods:*** **From the computer database, VCCR staff identified 80 slides that had been registered during the first half of 1991 and that covered the range of low-grade reports and negative reports. Each slide was identified by research number only. Two sets of 40 slides were compiled. Of the 29 laboratories that had worked with the VCCR during 1991, 22 agreed to participate in this study in 1992. One slide set was sent to each laboratory. An evaluation of the intralaboratory and interlaboratory consistency in reporting a set of 40 slides was undertaken. Analysis of the results compared the degree of consistency using current descriptive terminology that operates locally in Victoria with that which would pertain if the proposed Australian modification to the Bethesda System were adopted.** ***Results:*** **Intralaboratory agreement with previously reported slides was low on the squamous descriptor (49% agreement with original report) but higher on the human papillomavirus descriptor (76% agreement with original report) when the results were analyzed using the current terminology. Wide variation in reporting was apparent between laboratories; only 5% of the slides had agreement by all laboratories. Both intralaboratory and interlaboratory agreement improved substantially when results were grouped into the categories of the proposed Australian modification of the Bethesda Reporting System.** ***Conclusion and Implication:*** **Substantial improvement in the consistency of reporting cervical cytology specimens would be likely if terminology incorporating the broad categories of the Bethesda System were adopted. [J Natl Cancer Inst 85:1592–1596, 1993]**

During the 1950s and 1960s, diagnoses of cervical cytology specimens were generally reported according to the five classes of the Papanicolaou classification (*1*). The simple coding system was replaced during the 1970s by more descriptive reporting that used such terms as benign atypia, koilocytic atypia, dyskaryosis, dysplasia, or cervical intraepithelial neoplasia. This plethora of reporting terms resulted in what has been called "diagnostic chaos" (*2*). In an attempt to improve this situation, the Bethesda System for Reporting Cervical/Vaginal Cytologic Diagnoses was approved at a National Cancer Institute Workshop in 1988 (*3*).

Use of the Bethesda System has increased substantially year by year

*See "Notes" section following "References."

within the United States (*4*) since its inception. Although formal surveys in other countries have not been undertaken, the impression is that the Bethesda System is used relatively infrequently outside of the United States. Experience with cytology registries in Australia is that most laboratories continue to issue cytology reports as a descriptive diagnosis of the cellular features. In 1992, a working party of the National Health and Medical Research Council of Australia proposed that a modified Bethesda System be adopted by Australian laboratories (*5*). There are two areas that differ from the Bethesda System. First, while it is recommended that endocervical status be reported, the section of the Bethesda System entitled "Adequacy of the Specimen" is not included in the Australian terminology. Second, the proposed Australian modification of the Bethesda System provides six general categories of reports rather than three. These six categories are as follows: 1) negative (which would include benign cellular changes), 2) low-grade epithelial abnormalities, 3) high-grade epithelial abnormalities, 4) high-grade nonepithelial lesions, 5) inconclusive, and 6) technically unsatisfactory. By contrast, the three categories of the Bethesda System are 1) within normal limits, 2) benign cellular changes, and 3) epithelial cell abnormality.

Because of the absence of a standard format for reporting cervical cytology, a coding schedule was developed by Victorian cytopathologists when the Victorian Cervical Cytology Registry (VCCR) began in 1989. This coding schedule has been used subsequently for recording the results of tests of more than 500000 cervical cytology specimens each year. The VCCR coding schedule has five sections: 1) squamous cell code, 2) human papillomavirus code, 3) endocervical cell code, 4) noncervical cell code, and 5) recommendation code. The human papillomavirus code was allocated a separate section because of local interest in monitoring the prevalence of women with cytologic features of the virus.

The pattern of reporting smear diagnoses has been shown to be reasonably consistent over time within individual laboratories. For example, laboratories that report a high proportion of their smears as abnormal are generally consistent year by year in this reporting pattern (*6*). However, substantial variation has been apparent between laboratories. Thus, during the first 6 months of 1992, when the results of 267000 smears were registered, the probability of a woman receiving a diagnosis of abnormal squamous cells varied from 7% (approximately one chance in 13) to more than 26% (approximately one chance in four), depending upon which laboratory reported her smear diagnosis. Given the large number of tests involved, this variation is more likely to be due to different reporting practices by different laboratories rather than to differences in the client base. Most of the variation between laboratories has been in the reporting of minor reactive and inflammatory change, mild atypia, and mild dysplasia. Where laboratories report a relatively high proportion of smears in these categories, a concomitant reduction in the proportion reported as "no abnormal cells" is apparent.

We report here the findings of a study undertaken with 22 cytopathology laboratories during 1991 and 1992. Previously reported slides were circulated for reporting to determine which cytology appearances within the range of negative to low-grade abnormalities were associated with the greatest variation in reporting. The aim of the study was to ascertain if a greater interlaboratory consistency of cervical cytology diagnoses would be achieved if Australian laboratories were to adopt a Bethesda-style reporting approach rather than to continue to use a predominantly descriptive cytology reporting approach.

Methods

From the computer database, the VCCR staff identified 80 slides that had been registered during the first half of 1991 and that covered the range of low-grade reports and negative reports. All selected slides had either a histology report (n = 73) or a further cytology (n = 7) report within the next 8 months; the results of all of these later investigations were from the same range of normal- and low-grade abnormalities but were not necessarily directly concordant with the original cytology reports. Identifying information was removed from the slides so that each slide was identified by a research number only. Two sets of 40 slides each were compiled. Forty slides was considered the maximum number that an individual laboratory could reasonably be asked to review and represented up to 1 day's work for an average cytotechnologist working in an Australian laboratory. The two slide sets appeared to be acceptably comparable as judged by analysis of the range of original cytology reports as well as the results of later investigations (see Table 1 for details).

Each of the 29 laboratories that had worked with the VCCR during 1991 received a written invitation to participate in the study. Of the 29 laboratories, 22 agreed to participate. One slide set was sent to each laboratory for reporting and coding into the VCCR protocol. The accompanying letter to each laboratory indicated the range of reports that had been originally assigned to the slides. Each laboratory was asked to replicate, as much as possible, usual working conditions (i.e., screen each slide in 5-8 minutes, with only one cytotechnologist to screen all 40 slides). It was requested that a pathologist view the slide and grade the abnormality only if that would be the usual practice in the laboratory. Before sending the slide sets to each laboratory, all slides were rechecked to ensure that no identifying information, other than the research number, was on the slide.

The degree of intralaboratory agreement for the VCCR squamous cell and human papillomavirus cell codes was ascertained for slides that were reviewed by the original reporting laboratory. Table 2 shows the available coding options for squamous cells and human papillomavirus cells Intralaboratory and interlaboratory agreement under the proposed Australian modification of the Bethesda System was determined (Tables 3 and 4). Interlaboratory agreement on this coding schedule was profiled (Table 4) Under this new classification, the VCCR codes of "no abnormal cells" and of minor reactive and inflammatory changes would be grouped together to form the negative category, and the VCCR codes of mild atypia, human papillomavirus effect (possible and present), and mild dysplasia would comprise the low-grade epithelial abnormality category.

Results

Twenty-two of the 29 laboratories participated in the study; these 22 laboratories reported 97.4% of the 542000 smear diagnoses registered during 1992. The seven cytopathology laboratories that did not participate in this study tended to be low-volume laboratories reporting less than 15000 smears per year, to not be registered with the National Association of Testing Authorities, and to have pathologists who did not have membership in or a fellowship with the International Academy of Cytology.

Intralaboratory Agreement on Cytology Report

Of the 40 slides that were presented to each laboratory for review, some

Table 1. Comparison of two slide sets

Original report	Results of later investigations*				
	No abnormal cells	Minor reactive and inflammatory change	Mild atypia	Mild dysplasia	Total
Slide set 1					
No abnormal cells	7	2	0	0	9
Minor reactive and inflammatory change	1	6	2	0	9
Mild atypia	2	3	5	1	11
Mild dysplasia	1	2	2	6	11
Total	11	13	9	7	40
Slide set 2					
No abnormal cells	7	2	0	0	9
Minor reactive and inflammatory change	1	6	1	1	9
Mild atypia	2	4	4	1	11
Mild dysplasia	0	3	2	6	11
Total	10	15	7	8	40

*Values = number of slides.

Table 2. VCCR coding options for the squamous cell code and human papillomavirus cell code

Cell	Code
Squamous	0 = unsatisfactory 1 = no abnormal cells 2 = minor reactive and inflammatory changes 3 = mild atypia including changes of human papillomavirus infection without dysplasia 4 = cell changes of uncertain significance 5 = mild dysplasia (CIN 1) including equivocal or possible mild dysplasia* 6 = moderate dysplasia (CIN 2)* 7 = severe dysplasia and carcinoma in situ (CIN 3)* 8 = suspicious of microinvasion or invasion 9 = invasive squamous cell carcinoma
Human papillomavirus	1 = absent 2 = possible 3 = present

*CIN = cervical intraepithelial neoplasia (*16*).

Table 3. Degree of intralaboratory agreement for 18 laboratories

Original report	Review report*				
	No abnormal cells	Minor reactive and inflammatory change	Mild atypia	Mild dysplasia	Moderate dysplasia
Squamous cell code					
No abnormal cells	7	6	0	0	0
Minor reactive and inflammatory change	4	3	0	1	0
Mild atypia	2	1	6	0	0
Mild dysplasia	3	0	4	6	2

Original report	Review report*		
	Absent	Possible	Present
Human papillomavirus cell code			
Absent	23	1	2
Possible	1	0	0
Present	4	3	11

Original report	Review report*	
	Negative	Other
Agreement according to proposed Australian modification of the Bethesda System		
Negative	20	1
Other	8	16

*Values = number of slides.

Table 4. Degree of interlaboratory agreement for two sets of 40 slides, each assessed by 11 laboratories

	No. of slides with agreement among				
	11 laboratories	≥10 laboratories	≥9 laboratories	≥8 laboratories	≥7 laboratories
Classification according to current VCCR coding schedule					
Squamous cell code					
No abnormal cells	3	7	18	21	28
Minor reactive and inflammatory change	0	0	0	1	1
Mild atypia	0	1	3	3	6
Mild dysplasia	1	1	1	4	7
Total (% of slides)	4 (5)	9 (11)	22 (28)	29 (36)	42 (53)
Human papillomavirus cell code					
Absent	26	39	50	52	54
Possible	0	0	0	0	1
Present	2	3	6	8	10
Total (% of slides)	28 (35)	42 (53)	56 (70)	60 (75)	65 (81)
Classification according to the two relevant categories of the proposed Australian modification of the Bethesda System					
Negative	9	21	31	37	45
Low-grade epithelial abnormality	2	6	13	21	23
Total (% of slides)	11 (14)	27 (34)	44 (55)	58 (73)	68 (85)

may have been originally screened by the laboratory concerned—possibly two, three, or four of the slides may have been originally reported by the laboratory. Eighteen laboratories assessed slides that they had originally reported. Forty-five of the 80 slides in total were reassessed by the original laboratory that had reported the cytology. Table 3 shows the original and review reports for the squamous cell and human papillomavirus cell codes. There was agreement on the squamous cell code for 49% (22 of 45) of the slides, giving a kappa statistic of 0.008, which indicated poor agreement beyond chance (*7*). The level of agreement on the human papillomavirus cell code was higher, with 76% (34 of 45) of the slides showing a kappa statistic of 0.63, which indicated fair to good agreement beyond chance.

If these results are grouped according to the proposed Australian modification of the Bethesda System, there was intralaboratory agreement on 80% (36 of 45) of the slides. The kappa statistic for this level of agreement is 0.60, which again indicated fair to good agreement beyond chance (Table 3).

Interlaboratory Agreement on Cytology Report

The 80 slides in this study were each reviewed by 11 laboratories (11 laboratories reviewed one set of 40 slides; the other 11 laboratories reviewed the other set of 40 slides). There was unanimous agreement between the 11 laboratories about the squamous cell code for four slides (no abnormal cells for three slides and mild dysplasia for one slide). All laboratories agreed about human papillomavirus status for 28 slides (no evidence of human papillomavirus for 26 slides and human papillomavirus changes present for two slides). (*See* Table 4 *for details.*)

If the categories of the proposed Australian modification of the Bethesda System were used, there was complete agreement between the 11 laboratories for 11 slides (negative for nine slides and low-grade epithelial abnormality for two slides). At least seven of the 11 laboratories agreed on a modified Bethesda categorization for 68 (85%) of the 80 slides. See Table 4 for details. The proportion of slides with agreement by seven or more laboratories improved significantly with the use of this smaller number of categories (chi-square = 16.67, $P<.001$).

Discussion

This study demonstrates that there was substantial intralaboratory and interlaboratory variation in the reporting of diagnoses from cervical cytology specimens, even under somewhat artificial conditions. Despite the introductory letter indicating the range of reports that had originally been assigned to the slides, one in 12 of the reports provided in this review was outside this range. The current VCCR coding schedule for squamous cells (10 options) resulted in low rates of agreement both within and between laboratories. The kappa statistic for intralaboratory agreement on the squamous cell code indicated poor agreement beyond chance. Only 53% of the slides had agreement by seven or more of the 11 reviewing laboratories on the squamous cell code. When the modified Bethesda coding system was used, the improvement in the number that had agreement is impressive.

It is difficult to compare our results with the results of previous studies. The lack of a common terminology and the variation in study methodologies make it difficult to be certain that "like is being compared with like." Nevertheless, it appears that the least agreement among VCCR cytologists occurs in the middle range of the spectrum from normal to malignant. Published studies have found the most variation in the following categories: moderate and mild dysplasia (*8-10*), atypical squamous metaplasia and mild dysplasia (*11*), and moderate dysplasia to borderline carcinoma in situ (*12*). None of these studies attempted to classify, separately, the changes of human papillomavirus from the

changes of mild dysplasia or from what was variously described as benign reaction, benign atypia, or inflammatory atypia.

This study found that four of 80 slides had complete agreement on the squamous cell diagnosis by all 11 laboratories. Thomas et al. (*13*) documented complete agreement among seven laboratories on 22 of 140 slides, which ranged in diagnosis from benign changes to invasive malignancy; seven coding options were allowed. Derman et al. (*10*) found that none of 146 cases that were graded as "suspicious" (indicating mild or moderate dysplasia) received 100% agreement when evaluated by 180 laboratories; this study allowed four coding options. Yobs et al. (*12*) compared agreement between two laboratories on a series of 19474 cases. Ten coding options were allowed. Interlaboratory agreement was highest for slides with malignant features (70%-78%) and for negative slides (67%), but agreement was substantially less for all other categories. Agreement on categories that are pertinent to the range evaluated in this study was as follows: 37% agreement on 2836 cases of benign reaction, 9% agreement on 327 cases of minimal dysplasia, and 10% agreement on 92 cases of mild to moderate dysplasia.

Realistic expectations for the desirable degree of agreement among laboratories may be difficult to formulate. Factors that are likely to influence the degree of agreement include the number of observers, the number of slides evaluated, and the range and number of reporting options that are allowed. In relation to the latter option, statistical modeling has concluded that optimal classification of gynecologic cytology occurs when three categories of diagnosis are allowed (*14*). The modeling evaluated discrimination and divergence when two, three, four, five, and eight reporting categories were allowed. Of practical relevance to the cervical screening program is that use of a small number of reporting categories corresponds well to "patient action." After screening, a woman is essentially directed toward one of three options: 1) referral for further investigation, 2) early repeat cytology, or 3) repeat cytology at the usual recommended interval.

The ad hoc development of cytology reporting styles over the last two decades may have suited the individual preferences of cytologists but has seriously hindered comparative analyses among laboratories or countries. It is likely that patient management has been complicated by differing sequential smear diagnoses from different laboratories. Thus the adoption of a common terminology such as the Bethesda System would represent a major advance. It is perhaps regrettable that the Bethesda System, in its pure form, has not been adopted by Australian cytologists. The divergence arose from two main concerns. First, there was concern about the implications of the heading "Adequacy of the specimen" and about the validity of the information that would be used to define this aspect of a specimen. Second, in Australia, there has been a tradition of using an inconclusive report for up to 1% of smears in which the cytologic features raise the possibility of a high-grade lesion but a specific diagnosis is not possible. Local cytologists believed that the Bethesda System did not have an appropriate section for these reports.

As a screening modality, cervical cytology has produced a very worthwhile benefit in reducing mortality (*15*). Nevertheless, its shortcomings are frequently debated. The status of the test would be substantially improved if greater consistency in reporting were evident. Cytologists may need to choose between a desire for individual style in reporting and the increase in credibility that would result from greater consistency in reporting. The Bethesda System deserves further serious consideration. An international approach to reporting terminology would be most beneficial.

References

(*1*) Papanicolaou GN: Atlas of Exfoliative Cytology. Cambridge, Mass: Harvard Univ Press, 1954, pp 20–21

(*2*) Kurman RJ, Malkasian GD Jr, Sedlis A, et al: From Papanicolaou to Bethesda: the rationale for a new cervical cytologic classification. Obstet Gynecol 77:779–782, 1991

(*3*) The 1988 Bethesda System for reporting cervical/vaginal cytologic diagnoses. Developed and approved at the National Cancer Institute Workshop, Bethesda, Maryland, USA, December 12-13, 1988. Acta Cytol 33:567–574, 1989

(*4*) Davey DD, Nielsen ML, Rosenstock W, et al: Terminology and specimen adequacy in cervicovaginal cytology. The College of American Pathologists Interlaboratory Comparison Program experience. Arch Pathol Lab Med 116:903–907, 1992

(*5*) Working Party on the Management of Women with Abnormalities Detected in Cervical Cancer Screening: Screening for the prevention of cervical cancer: management of women with an abnormal smear. National Health and Medical Research Council, Draft Report, August 1992

(*6*) Victorian Cervical Cytology Registry: Participating Laboratories' Statistics and VCCR Protocols. Melbourne: 1990, 1991, 1992

(*7*) Fleiss JL: Statistical Methods for Rates and Proportions. New York: John Wiley & Sons, 1981, pp 212–225

(*8*) Curry H, Thompson DW, Dietrich M, et al: Proficiency testing in cytology laboratories in Ontario, Canada: a decade of experience. Acta Cytol 31:203–225, 1987

(*9*) Penner DW: An overview of the College of American Pathologists' programs in surgical pathology and cytopathology. Data summary of diagnostic performance in cervical cytopathology. Acta Cytol 33:439–442, 1989

(*10*) Derman H, Koss L, Hyman MP, et al: Cervical cytopathology 1—peers compare performance. Pathol 35:317–325, 1981

(*11*) Rombach JJ, Cranendonk R, Velthuis FJJ: Monitoring laboratory performance by statistical analysis of rescreening cervical smears. Acta Cytol 31:887–894, 1987

(*12*) Yobs AR, Plott AE, Hicklin MD, et al: Retrospective evaluation of gynecologic cytodiagnosis. II. Interlaboratory reproducibility as shown in rescreening large consecutive samples of reported cases. Acta Cytol 31:900–910, 1987

(*13*) Thomas GD, Head C, Thorogood J: Quality assessment in cervical cytology. a pilot study. J Clin Pathol 41:215–219, 1988

(*14*) Okagaki T, Fujimura M, Zelterman D: Information, discrimination and divergence in cytology. III. Optimization of classification of Papanicolaou smears. Acta Cytol 35:30–34, 1991

(*15*) Tomatis L: Cancer: causes, occurrence and control. Lyon: IARC, 1990, pp 276–279

(*16*) Koss LG: Diagnostic Cytology and its Histopathologic Bases, 4th ed. Philadelphia: Lippincott, 1992, pp 444–466

Notes

Affiliation of author: Victorian Cervical Cytology Registry, Carlton South, Australia.

Correspondence to: Heather Mitchell, M.D., F.R.A.C.P., F.A.F.P.H.M., Medical Director, P.O. Box 161, Carlton South, Australia 3053.

Supported by the Department of Health, Housing and Community Services of the Australian government.

I thank the staff of each laboratory that participated in this study.

Manuscript received March 3, 1993; revised June 28, 1993; accepted July 6, 1993.