Original Article

Telecytology

Diagnostic Accuracy in Cervical-Vaginal Smears

STEPHEN S. RAAB, MD, M. SUE ZALESKI, SCT(ASCP), PATRICIA A. THOMAS, MD, THEODORE H. NIEMANN, MD, CHRISTINA ISACSON, MD, AND CHRIS S. JENSEN, MD

Although cervical-vaginal telecytology is a promising tool, diagnostic accuracy has not been extensively evaluated. The authors examined the accuracy of five cytotechnologists who retrospectively reviewed 50 cervical-vaginal smears using the video monitor, and 2 months later, using the light microscope. Accuracy was expressed in terms of crude agreement with the original diagnosis and number of false positives (FPs) and false negatives (FNs). With a greater than one step difference as discrepant, the group crude agreement using the video monitor and the light microscope was 85.6% and 95.6%, respectively. The group number of FNs and FPs for the light microscope was 8 and 7, respectively, and for the video monitor was 34 and 7, respectively. There was a wide range of individual performance. We conclude that accuracy of telecytology is high, but less than that of light microscopy. The major reason for lower telecytologic accuracy was undercalling dysplasia. (Key words: Telecytology; Cervical-vaginal smear; Automated screening; Cytology; Dysplasia; Diagnostic accuracy; Reproducibility) Am J Clin Pathol 1996;105:599–603.

Two important applications of cervical-vaginal telecytology are in off-site diagnosis and in automated screening.[1–17] The utility and ultimately the use of cervical-vaginal telecytology depends on the evaluation of several factors, one of which is diagnostic accuracy.[14,18–20]

Studies of cervical-vaginal telecytologic accuracy have mainly focused on automated screening.[10,11,13–17] In the PAPNET screening system, Koss and colleagues[11] reported that the interpretation of telecytologic images resulted in few missed "dysplasias." However, these studies examined the accuracy of the screening system and not the observer. This is an important distinction, because the telecytologic competency of the observer will determine the number of cases to be manually rescreened and if manual rescreening is even necessary.

Comparison of individual observer diagnostic accuracy using the video monitor and the light microscope is the next step in evaluation. Recently, Ziol and colleagues[21] reported that crude agreement for pathologists using the light microscope and the video monitor was similar. Additional research is needed to confirm these results and to characterize the accuracy of other observer types, such as cytotechnologists.

From the Department of Pathology, University of Iowa, Iowa City, Iowa.

Manuscript received July 24, 1995; revision accepted November 8, 1995.

Address reprint requests to Dr. Raab: University of Iowa Hospitals and Clinics, Department of Pathology, 200 Hawkins Drive, 5216 RCP, Iowa City, IA 52242-1009.

In our study, we examined telecytologic diagnostic accuracy for cytotechnologists in terms of:

1. Group and individual number of discordant diagnoses (crude agreement);
2. Group and individual number of false-negative diagnoses; and
3. Group and individual number of false-positive diagnoses.

Other accuracy issues that we evaluated were the effect of observer experience and observer proficiency using the video monitor.

MATERIALS AND METHODS

Fifty cervical-vaginal smears were retrospectively selected from the University of Iowa cytology files from the years 1992–1993. Each of the cervical-vaginal smears consisted of one slide. All cases had histologic follow-up that confirmed the original cytologic diagnosis. The diagnoses of the cases consisted of 16 benign (including 6 repair), 17 low grade squamous intraepithelial lesions (LGSILs) and 17 high grade squamous intraepithelial lesions (HGSILs).[22–25] The 16 benign cases also had 6 to 15 months (mean 10 months) of additional benign cervical-vaginal smear follow-up. Cases with diagnoses of atypical squamous cells of undetermined significance (ASCUS), atypical glandular cells of undetermined significance (AGUS), and carcinoma were excluded.[22–25]

Each case was randomly relabeled with a number from 1 to 50. The slides were then screened by an experienced cytotechnologist who was unaware of the previous cytologic and histologic diagnoses. The cytotechnologist placed five dots per slide, marking the areas which were diagnostic or most worrisome for a squamous intraepithelial lesion. In the benign cases, these areas often exhibited reactive or reparative changes.

Each observer reviewed all 50 cervical-vaginal smears twice, first by video monitor, and again 2 months later by light microscopy. Viewing time for each dot was 12 seconds for each method. Video microscopy was performed using an Olympus BH-2 microscope (Tokyo, Japan) equipped with a three chip RBG Sony camera (Tokyo, Japan). The images were shown on a Sony monitor model 1943 with 750 lines of resolution. Conventional light microscopy was performed individually on an Olympus BH-2 microscope. Objective lenses were identical for each method.

Standardized forms with specific instructions were given to each observer. Clinical histories that accompanied the original cytologic requisition forms were provided. Each case included a patient age, menstrual history, exogenous hormonal status, and previous history of clinically significant cervical-vaginal lesions. The observers classified each case into one of the following categories: benign, ASCUS, AGUS, LGSIL, HGSIL, and carcinoma.[22-25] The observers did not consult each other.

All observers were cytotechnologists who had different levels of cytologic experience. Observers 3 and 5 had more than 10 years of experience. Observers 1, 2, and 4 had 18 months of experience. The observers had little or no prior experience with telecytology.

Diagnostic accuracy of telecytology and light microscopy was determined by measuring:

1. Group and individual number of discordant diagnoses (crude agreement);
2. Group and individual number of false-negative diagnoses; and
3. Group and individual number of false-positive diagnoses.

Crude agreement was calculated by comparing the telecytologic and light microscopic diagnoses to the original cytologic diagnosis and was expressed as a percentage.[18-20,26] The number of discordant diagnoses using each technique was determined. Discordant diagnoses were calculated in two ways. In the first way, telecytologic and light microscopic diagnoses were considered discordant if there was not exact agreement with the original diagnosis. The second way of calculating discordant diagnoses depended on considering cytologic diagnoses to be semiquantitative.[19,20] Each diagnosis corresponded to a step from benign to malignant, and the ordering of the diagnoses was: benign, ASCUS and AGUS, LGSIL, HGSIL, and carcinoma.[19] Using this method, telecytologic and light microscopic diagnoses that were identical or one step different from the original diagnosis were considered to be concordant. Diagnoses that were more than one step different from the original cytologic diagnosis were considered to be discordant. For example, if the telecytologic diagnosis was ASCUS and the original diagnosis was benign or LGSIL, there was concordance. If the original diagnosis was HGSIL, there was discordance.

TABLE 1. CRUDE AGREEMENT OF FIVE OBSERVERS BETWEEN THE LIGHT MICROSCOPE OR THE VIDEO MONITOR AND THE ORIGINAL DIAGNOSIS

	All Differences as Discordant		*>1 Step Difference as Discordant*	
Observer	*Microscope*	*Monitor*	*Microscope*	*Monitor*
1	70	62	96	90
2	64	64	98	86
3	70	56	100	86
4	68	48	94	70
5	64	70	90	96
Average	67.2	60.0	95.6	85.6

The number of false-negative diagnoses using the video monitor and the light microscope was determined by calculating the number of "missed" dysplasias. Determination of this number depended on considering the diagnoses as semiquantitative.[19,20] A telecytologic or light microscopic diagnosis was a false negative if that diagnosis was more than one step lower than the original cytologic diagnosis. For example, if the light microscopic or telecytologic diagnosis was ASCUS or benign, and the original diagnosis was HGSIL, it was considered a false negative. If the original diagnosis was LGSIL, and the light microscopic or telecytologic diagnosis was benign, it was considered a false negative.

The number of false-positive diagnoses was determined by calculating the number of overcalls of dysplasia or ASCUS. A telecytologic or light microscopic diagnosis was an overcall if that diagnosis was dysplasia (carcinoma, HGSIL, or LGSIL) or ASCUS and the original diagnosis was benign.

RESULTS

The crude agreement between the video monitor or the light microscope and the original diagnosis is shown in Table 1. Using both methods of calculating discordant

TABLE 2. NUMBER OF FALSE-NEGATIVE DIAGNOSES OF FIVE OBSERVERS USING THE LIGHT MICROSCOPE AND THE VIDEO MONITOR

	No. of Missed HGSILs		*No. of Missed LGSILs*	
Observer	*Microscope*	*Monitor*	*Microscope*	*Monitor*
1	1	3	1	2
2	1	4	0	2
3	0	4	0	3
4	0	9	1	6
5	3	0	1	1
Total	5	20	3	14

HGSIL = high grade intraepithelial neoplasia; LGSIL = low grade intraepithelial neoplasia.

diagnoses, as a group, the cytotechnologists had a higher crude agreement using the light microscope ($P = .11$). Individual cytotechnologist performance was variable. Observer 5 had a higher crude agreement using the video monitor. The other observers generally had higher accuracy using the light microscope. Observer 4 performed the poorest using the video monitor; crude agreements were more than 20 percentage points lower on the video monitor than on the light microscope. We did not demonstrate a difference in crude agreement using the video monitor and the light microscope based on observer experience.

The number of false-negative diagnoses using the video monitor and the light microscope is shown in Table 2. As a group, the cytotechnologists missed fewer HGSILs and LGSILs using the light microscope than using the video monitor. A range of diagnostic ability was observed. Observer 5 performed better using the video monitor. Observer 4 performed poorest using the video monitor and missed 9 of 17 (53%) HGSILs. Across all observers, the total number of LGSILs and HGSILs that were called benign using the light microscope and the video monitor were 4 and 18, respectively ($P = .11$).

The number of false-positive diagnoses is shown in Table 3. Overcalls of dysplasia and ASCUS are listed separately. As a group, the cytotechnologists made an equal number of overcalls with the video monitor and the light microscope. There were few overcalls of dysplasia and a slightly higher number of overcalls of ASCUS. Observer 5 had a tendency to overcall ASCUS using the video monitor, whereas observer 2 had a tendency to overcall ASCUS using the light microscope.

DISCUSSION

We examined individual observer diagnostic accuracy using the video monitor and the light microscope and conclude:

1. Diagnostic accuracy of telecytology was high.
2. For the group of cytotechnologists, there were fewer discordant diagnoses using the light microscope than the video monitor.
3. For the group of cytotechnologists, a higher false-negative rate was observed for the video monitor than for the light microscope. This may indicate that many observers are conservative when using the video monitor and reluctant to make a dysplastic diagnosis.
4. For the group of cytotechnologists, an equal number of false-positive diagnoses (dysplastic and ASCUS overcalls) were made using the video monitor and the light microscope.
5. Individuals perform with varying ability using the video monitor and some individuals perform better using the video monitor.
6. Performance using the video monitor and the light microscope did not depend on general cytotechnology experience.

Our data of group crude agreement were similar to the data of Ziol and colleagues.[21] We reported group crude agreements of 67.2% and 60.0% for the light microscope and the video monitor, respectively, whereas Ziol and colleagues[21] reported group crude agreements of 64% and 62%, respectively. In actuality, these data reflect relatively poor concordance with the original diagnosis for both the light microscope and the video monitor. For our data, the kappa statistic for the light microscope and the video monitor was 0.34 and 0.20, respectively; a kappa statistic less than 0.5 is considered poor.[19,20,26] Despite differences in study design, our data are similar to results on cervical-vaginal interobserver variation reported previously.[27-35] In grading squamous abnormalities, Klinkhamer and colleagues[29] reported that only 44.1% of observer diagnoses were exactly concordant with original diagnoses.

Such low concordance values are acceptable, because many of the differences in diagnosis do not affect patient

TABLE 3. NUMBER OF FALSE-POSITIVE DIAGNOSES OF FIVE OBSERVERS USING THE LIGHT MICROSCOPE AND THE VIDEO MONITOR

	No. of Overcalls of Dysplasia		*No. of Overcalls of ASCUS*	
Observer	*Microscope*	*Monitor*	*Microscope*	*Monitor*
1	0	0	0	0
2	0	1	3	1
3	0	0	0	0
4	1	0	1	0
5	1	1	1	4
Total	2	2	5	5

ASCUS = atypical squamous cells of undetermined significance.

care. Only when the observer diagnosis and the original diagnosis differ by more than one step is the discordance important. This is because a greater than one step difference in cytologic diagnosis often results in a different treatment protocol.[28] Using a greater than one step difference as discordant, we observed group crude agreements of 95.6% and 85.6% for the light microscope and video monitor, respectively. Ziol and colleagues[21] did not report these data. Our findings indicate that the group diagnostic accuracy of the video monitor was high, but still lower than the diagnostic accuracy of the light microscope. The kappa statistic for the light microscope and the video monitor was 0.91 and 0.71, respectively. Values greater than 0.9 are considered excellent and values between 0.5 and 0.9 are considered good.[26]

One reason that telecytology had a lower accuracy was that most of the observers tended to undercall dysplasia. Low grade squamous intraepithelial lesions were classified as benign smears and HGSILs were classified as benign or ASCUS smears. This indicates that for most observers, there may be a certain reluctance using the video monitor to make a dysplastic diagnosis. This conclusion differs from that of Ziol and colleagues[21] who reported that teletransmission is as reliable as light microscopy. In our study, sources of undercall include unfamiliarity with telecytology or the study design and difficulties in recognizing dysplastic changes on a video monitor. Chance variation could not be excluded.

In contrast to the disproportionate number of false-negative diagnoses, observers made few false-positive diagnoses with both the video monitor and the light microscope. Using the video monitor, the entire group made only two diagnoses that were more than one step higher than the original diagnoses.

Our findings have different import depending if cervical-vaginal telecytology is used for long distance diagnosis or for automated screening. For long distance diagnosis, high accuracy across all diagnostic categories is desirable.[1-9,36-38] Long distance diagnoses presumably will be made by pathologists, who were not evaluated in our study. We must determine if pathologists will exhibit the same accuracy tendencies as cytotechnologists.

Telecytologic diagnostic accuracy depends on several factors, one of which is the quality of the image displayed on the video monitor. Image quality is affected by image capture, transmission, and video display. In our study, there was no digitalization step, which further may compromise image quality. Some commercial systems use digitalization. The comparison of diagnostic accuracy between these systems and our model has not been made.

In cervical-vaginal screening, the key is not definitive diagnosis, but classification of smears into two categories: (1) those needing review, and (2) those not needing review.[11,14] Presumably, this classification will be made by cytotechnologists, and it is desirable to have few false negatives in the "not review" category. We determined that across all observers, the total number of LGSILs and HGSILs that were called benign using the light microscope and the video monitor were 4 and 18, respectively. This false-negative rate is higher than reported by Koss and associates,[11] Slagel and coworkers,[14] and others.[10] The discrepancy between our and other's data may be due to differences in study design, but needs further investigation.

The variability in cytotechnology performance using the video monitor may indicate that telecytologic proficiency has an innate component. In our study, one observer performed better with the video monitor. In using the video monitor, a select field is presented and the observer does not have to try to find the cells of concern. Thus, the observer's ability to focus on the whole slide is of less importance than on the observer's ability to make a decision on select areas.[39] Telecytology may be beneficial to some cytotechnologists who have difficulty in maintaining concentration when examining an entire smear.

We did not find that cytotechnology experience played a role in telecytologic accuracy. Zaleski and colleagues[39] reported similar results in the bronchial brush specimen. The ability to improve telecytologic accuracy with increased telecytologic use has yet to be studied.

We conclude that telecytology has high diagnostic accuracy, although, the finding from this study indicate it is not yet as accurate as light microscopy. Accuracy studies are only the first step in determining the utility of telecytology, which also must be evaluated in other areas, such as cost effectiveness and effect on patient management.

Acknowledgments. The authors thank Amber Donnelly, Gary Hendricks, Sara Wiemerslage, Susan Martens, and Cindy Strovers for their participation in this study and Katie Dasse for assistance in the manuscript preparation.

REFERENCES

1. Weinstein RS. Prospects for telepathology. *Hum Pathol* 1986;17:433-434.
2. Weinstein RS, Bloom KJ, Rozek LS. Telepathology: Long distance diagnosis. *Am J Clin Pathol* 1989;91(suppl):S39-S42.
3. Erickson D. Do you see what I see? *Sci Am* 1990;262:88-89.
4. Weinstein RS. Telepathology: Practicing pathology in two places at once. *Clinical Laboratory Management Review* 1992;171-173.
5. Weinstein RS, Bloom KJ, Rozek LS. Telepathology and the networking of pathology diagnostic services. *Arch Pathol Lab Med* 1987;111:646-652.

6. Weinstein RS. Telepathology comes of age in Norway. *Hum Pathol* 1991;22:511–513.
7. Nordrum I, Engum B, Rinde, et al. Remote frozen section service: A telepathology project in Norway. *Hum Pathol* 1991;22:514–518.
8. Wold LE, Weiland LH. Telepathology at the Mayo Clinic. *Clinical Laboratory Management Review* 1992;174–175.
9. Eide TJ, Nordrum I, Engum B, et al. Use of telecommunications in pathology and anatomic services. *Tidsskr Nor Laegeforen* 1991;111:17–19.
10. Ouwerkerk-Noordam E, Boon ME, Beck S. Computer-assisted primary screening of cervical smears using the PAPNET method: Comparison with conventional screening and the evaluation of the role of the cytologist. *Cytopathology* 1994;5:211–218.
11. Koss LG, Lin E, Schrieber K. Elger P, Mango L. Evaluation of the PAPNET cytologic screening system for quality control of cervical smears. *Am J Clin Pathol* 1994;101:220–229.
12. Williams C, Rosenthal DL. Cytopathology in the 21st century. *Am J Clin Pathol* 1993;99(suppl):S31–S33.
13. Ploem JS, van Driel-Kulker AMJ, Verwoerd NP. LEYTAS: A cytology screening system using the new modular image analysis computer (MIAC) from Leitz. In: Burger G, Ploem JS, Goerttler K, eds. Clinical Cytometry and Histometry. London: Academic Press, 1987, pp 24–35.
14. Slagel DD, Zaleski S, Cohen MB. The efficacy of automated cervical cytology screening. *Diagn Cytopathol* 1995;13:26–30.
15. Hutchinson ML, Cassin BM, Bull HG III. The efficacy of an automated preparation device for cervical cytology. *Am J Clin Pathol* 1991;96:300–305.
16. Bahr GF, Bibbo M, Oehme M, Puls JH, Reale FR, Wied GL. An automated device for the production of cell preparations suitable for automatic assessment. *Acta Cytol* 1978;22:243–249.
17. Ashfaq R, Liang Y, Saboorian MH. Evaluation of PAPNET™ system for rescreening of negative cervical smears. *Diagn Cytopathol* 1995;13:31–36.
18. Raab SS. Diagnostic accuracy in cytopathology. *Diagn Cytopathol* 1994;10:68–75.
19. Raab SS, Bottles K, Cohen MB. Technology assessment in anatomic pathology: An illustration of technology assessment techniques in fine-needle aspiration biopsy. *Arch Pathol Lab Med* 1994;18:1173—1180.
20. Valenstein P. Technology assessment for the diagnostic laboratory. American Society of Clinical Pathology Special Topics Council Commission on Continuing Education: Should this test be done? Lecture notes. American Society of Clinical Pathology Meeting, 1992.
21. Ziol M, Vacher-Lavenu M, Ferrand J, et al. Telepathology and cervical Papanicolaou smears: Is teletransmission of selected fields available for interpretation of squamous cell abnormalities? *Acta Cytol* 1995;39:376.
22. The Bethesda Committee. The Bethesda system for reporting cervical/vaginal cytologic diagnoses. *Acta Cytol* 1993;37:115–124.
23. The Bethesda Committee. The Bethesda System for Reporting Cervical/Vaginal Diagnoses. New York: Springer Verlag, 1994.
24. The Bethesda Committee. Current issues: The 1991 Bethesda system for reporting cervical/vaginal diagnoses. *Diagn Cytopathol* 1993;9:235–293.
25. Koss LG. The new Bethesda system for reporting results of smears of the uterine cervix. *J Natl Cancer Inst* 1990;82:988–991.
26. Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical Epidemiology: A Basic Science for Clinical Medicine. Boston: Little, Brown, 1991.
27. Koss LG. Epidermoid carcinoma of the uterine cervix and related precancerous lesions. Part II: Cytologic diagnosis and its consequences. In: Koss LG, ed. Diagnostic Cytology and Its Histologic Bases. Philadelphia: JB Lippincott, 1992, pp 470–471.
28. Vooijs GP. Benign proliferative reactions, intraepithelial neoplasia, and invasive cancer of the uterine cervix. In: Bibbo M, ed. Comprehensive Cytopathology. Philadelphia: WB Saunders, 1991, pp 221–225.
29. Klinkhamer PJM, Vooijs GP, de Haan AFJ. Intraobserver and interobserver variability in the diagnosis of epithelial abnormalities in cervical smears. *Acta Cytol* 1988;32:794–800.
30. Klinkhamer PJM, Vooijs GP, de Haan AFJ. Intraobserver and interobserver variability in the quality assessment of cervical smears. *Acta Cytol* 1989;33:215–218.
31. Evans DMD, Shelley G, Cleary B. Baldwin Y. Observer variation and quality control of cytodiagnosis. *J Clin Pathol* 1974;27:945–950.
32. 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* 1987;31:900–910.
33. Yobs AR, Swanson RA, Lamotte LC, Jr. Laboratory reliability of the Papanicolaou smear. *Obstet Gynecol* 1985;65:235–244.
34. Seybolt JF, Johnson WD. Cervical cytodiagnostic problems: A survey. *Am J Obstet Gyencol* 1971;109:1089–1103.
35. Koss LG. Cytologic evaluation of the uterine cervix: Factors influencing its accuracy. *Pathologist* 1982;36:401–407.
36. Linder J, Allen KA, Hansen S. Remote cytologic diagnosis by videotelephone (Abstr). *Acta Cytol* 1988;32:772.
37. Linder J, Masada CT. Frozen section diagnosis by videotelephone (Abstr). *Lab Invest* 1989;60:54A.
38. Becker FL, Specht CS, Jones R, Rueda-Pedraza ME, O'Leary TJ. Use of remote video microscopy (telepathology) as an adjunct to neuro-surgical frozen section consultation. *Hum Pathol* 1993;24:909–911.
39. Zaleski MS, Thomas PA, Ferguson DJ, Niemann TH, Raab SS. Telecytology: Diagnostic accuracy with comparison to light microscopy in the bronchial brush specimen. *J Surg Pathol* 1996 (in press).