Quality Assurance in Gynecologic Cytology

The Value of Cytotechnologist-Cytopathologist Discrepancy Logs

Edmund S. Cibas, MD,[1,4,6] *Barbara Dean, CT(ASCP),*[1] *Nicholas Maffeo, MBA,*[2] *and Elizabeth N. Allred, MS*[3-6]

Key Words: Quality control; Cytotechnologist; Cytopathologist; Cytology; Papanicolaou smears; Performance evaluation

Abstract

We describe a simple method for displaying and evaluating the concordance or discordance between cytotechnologists (CTs) and cytopathologists (CPs) on gynecologic cases. The provisional diagnoses made by the CTs and the final diagnoses of the CPs are captured by the laboratory information system; data generated for specified periods are displayed as a 10 × 10 matrix that classifies each possible diagnosis made by the CT and CP into 1 of 10 major categories. Matrices are generated for the entire laboratory and for individual CTs; individual CTs are evaluated based on their deviation from the laboratory average. Three statistical measures are generated: percentage of discordant diagnoses, a kappa statistic, and a weighted measure.

During a 2.5-year period, approximately 4,200 cases were referred to a CP for review every 6 months. The median discordance in diagnoses increased during 2 years from 21% to 34%, and the kappa value fell from 0.69 to 0.38. This was attributed primarily to 1 CT, whose performance, as well as that of the entire laboratory, improved after remedial action. Measures of CT-CP diagnostic concordance are a useful and efficient measure of CT performance and can be incorporated into mandatory semiannual performance evaluations.

The review of Papanicolaou (Pap) tests by a cytotechnologist (CT) involves screening (detection) and interpretation. The screening ability of the CT governs the sensitivity of the Pap test and is understandably the focus of much of the quality assurance activity of the laboratory. Efforts to minimize and quantify false-negative results are paramount. At least 10% of cases interpreted as negative are rescreened by another CT. This rescreening identifies false-negative results and permits the calculation of a false-negative fraction, which can be used to evaluate and compare the screening performance of individual CTs.[1,2]

The responsibility for the final interpretation of a nonnegative Pap slide rests with the cytopathologist (CP), who is required to review all Pap tests that contain reactive, atypical, dysplastic, or malignant cells. When a case is referred to a CP for review, it is accompanied by the provisional interpretation of the CT. The work of the CP is made more or less difficult depending on how closely the CT's provisional diagnosis matches the final interpretation by the CP. A less-than-vigilant CP may even misclassify a specimen because he or she was swayed unduly by the CT's interpretation. For these reasons, the Clinical Laboratories Improvement Amendments of 1988 Final Rule requires that the mandatory semiannual evaluation of a CT be based, in part, on interpretation of cases submitted to the CP for review.[3]

We describe a software program that tabulates the degree of concordance between CTs and CPs, thus providing objective measures for performance evaluations.

Materials and Methods

In an effort to streamline our quality control activities and provide more objective measures of CT and CP performance, we met once or twice a month from July 1997 to

August 1998 to develop and test add-on menu items for the existing laboratory information system. These included calculations of false-negative fractions, case volume statistics, turnaround time reports, productivity reports, and CT-CP diagnostic discrepancy logs. The latter are the focus of this article.

The Brigham and Women's Hospital (Boston, MA) Cytology System is one of many component systems that make up the Brigham Integrated Computer System, which is a system of networked PCs. The cytology system was designed and developed at Brigham and Women's Hospital using the MUMPS (Massachusetts General Utility Multi-Programming System, InterSystems Corp, Cambridge, MA) programming language (now called M).

As a database system designed to capture events related to patient cytology specimens, it is necessary that the computer be a good fit into the workplace. The cytology system was designed to capture events occurring from initial accessioning to final reporting of the specimen. All computer entries are audited, and audit information includes a time stamp, the user's identity, and the information entered or changed. By using this trail of information, feedback can be provided relating to the processing of the laboratory specimens.

The system tracks the entry and edits of the provisional diagnosis by the CT and the final interpretation made by the CP to qualify and quantify the degree of concordance between the CT and the CP. A grid displays concordance and discordance of CTs and CPs on the evaluation of gynecologic cytology cases, with all possible diagnoses grouped into 10 major categories: (1) unsatisfactory; (2) within normal limits; (3) benign cellular changes (BCC); (4) endometrial cells, cytologically benign, in a postmenopausal woman; (5) atypical squamous cells of undetermined significance (ASCUS); (6) atypical glandular cells of undetermined significance; (7) low-grade squamous intraepithelial lesion; (8) squamous intraepithelial lesion, difficult to grade; (9) high-grade squamous intraepithelial lesion; and (10) carcinoma. A 10×10 matrix can be generated for the entire laboratory and for individual CTs. All discordant cases are displayed in list form by individual CT. Thus, for example, one could retrieve and review all cases called ASCUS by CT 1 but changed to BCC by the CPs.

The data in these matrices are summarized by several measures of agreement. First, the program calculates the percentage of discordant cases. Second, a kappa statistic is calculated according to standard methods.[4] This is a statistical measure of the degree of concordance between observers. The kappa values range from 0 to 1.0. Values above 0.4 represent good agreement, and those above 0.75 represent excellent agreement. Finally, a weighted score is derived by assigning a greater penalty for larger discordances than smaller ones and a greater penalty for "undercalls" than for overcalls. The penalty assigned for each possible discordance was based on grids developed at Thomas Jefferson University (Philadelphia, PA) for grading cytotechnology students **Table 1**. The weighted score is obtained by taking the number of discordant diagnoses in each cell, multiplying it by the penalty value assigned to that cell, totaling the products for each cell, and dividing this value by the total number of discordant diagnoses.

Results

Between July 1997 and December 1999, 14 CTs were employed at the Brigham and Women's Hospital. During this period, 3 CTs left the laboratory, 3 were hired, and 8 CTs

Table 1
Weights (Penalty Values) for Each Type of Discrepancy

	Cytopathologist Diagnosis									
	UNS	**WNL**	**BCC**	**EPM**	**ASCUS**	**AGUS**	**LSIL**	**SILUNC**	**HSIL**	**CAR**
Cytotechnologist diagnosis										
UNS	0	10	15	15	15	15	20	21	22	25
WNL	25	0	2	3	4	7	6	7	8	10
BCC	15	2	0	1	2	6	5	6	7	9
EPM	10	3	2	0	3	5	6	7	8	9
ASCUS	10	4	2	1	0	2	4	5	6	8
AGUS	10	6	4	3	2	0	3	4	5	8
LSIL	10	7	6	5	3	1	0	1	5	7
SILUNC	10	9	7	6	4	3	2	0	1	6
HSIL	10	10	9	7	6	2	4	1	0	6
CAR	10	10	9	9	8	8	7	5	4	0

AGUS, atypical glandular cells of undetermined significance; ASCUS, atypical squamous cells of undetermined significance; BCC, benign cellular changes; CAR, carcinoma; EPM, endometrial cells, cytologically benign, in a postmenopausal woman; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; SILUNC, squamous intraepithelial lesion, difficult to grade; UNS, unsatisfactory; WNL, within normal limits.

were employed continuously. Eleven CPs reviewed Pap specimens during this period. During any given 6-month period of the study, from 6 to 9 CPs reviewed Pap specimens.

Each CT was evaluated twice a year, based in part on the accuracy of the diagnoses they made on nonnegative Pap specimens that they referred for CP review. The number of cases referred to a CP for review, the percentage of discordant diagnoses **Table 2**, a kappa statistic **Table 3**, and a weighted score **Table 4** were calculated for each CT and for the entire laboratory. **Table 5** summarizes these measures for the entire laboratory during this period. The computer program was not designed to measure discrepancies between a CT and any individual CP, only for the group of CPs as a whole.

During this time, the number of cases referred for CP review ranged from 3,925 to 4,469 per 6-month period. An example of the matrix for the entire laboratory for a 6-month period is given in **Table 6**. The median percentage of discordant diagnoses rose from 21% in the second half of 1997 to 34% in the first half of 1999 (Table 5). This was mirrored by a steep decline in the kappa statistic from 0.69 to 0.38. The weighted score decreased from 3.55 to 3.04, indicating that, although there was an increasing discordance between CTs and CPs during this period, the discordances, on average, were less severe.

A rise in discordance rates during the first four 6-month periods of study was seen for 7 of the 8 CTs continuously employed during this period; only CT 1 showed a decrease in the percentage of discordant cases. The steepest increase was seen with CT 7, for whom the percentage of discordant cases more than doubled.

In August 1999, evaluations were based on measures of performance from January through June 1999. The high

Table 2
Percentage of Discordant Cases for Entire Laboratory and for Individual Cytotechnologists*

	July-December 1997	January-June 1998	July-December 1998	January-June 1999	July-December 1999
Median (entire laboratory)	20.9 (3,925)	28.3 (4,469)	28.4 (4,307)	34.0 (4,287)	18.3 (4,154)
Cytotechnologist No.					
1	21.6 (148)	27.0 (71)	20.0 (55)	19.0 (47)	22.0 (64)
2	14.8 (488)	25.8 (523)	21.6 (399)	29.0 (372)	15.6 (301)
3	13.0 (61)	35.0 (57)	37.0 (62)	34.0 (68)	30.0 (43)
4	17.8 (231)	32.0 (56)	—	—	—
5	20.0 (485)	28.1 (452)	25.0 (272)	40.5 (131)	18.3 (273)
6	25.3 (178)	18.6 (161)	28.0 (78)	45.0 (29)	16.0 (95)
7	23.4 (1,127)	43.3 (1,472)	44.5 (1,516)	55.8 (1,664)	16.2 (973)
8	23.0 (517)	35.5 (575)	33.2 (467)	33.6 (494)	21.6 (486)
9	20.9 (375)	28.3 (474)	28.1 (342)	37.1 (283)	23.3 (223)
10	19.2 (297)	—	—	—	—
11	50.0 (18)	40.2 (487)	—	—	—
12	—	18.0 (57)	35.7 (291)	42.9 (371)	23.2 (539)
13	—	—	31.5 (302)	28.9 (439)	12.4 (539)
14	—	—	28.4 (282)	28.8 (389)	14.9 (617)

* The number of cases referred by the cytotechnologists for cytopathologist review is in parentheses.

Table 3
kappa Statistic for the Entire Laboratory and for Individual Cytotechnologists

	July-December 1997	January-June 1998	July-December 1998	January-June 1999	July-December 1999
Entire laboratory	0.69	0.48	0.48	0.38	0.75
Cytotechnologist No.					
1	0.69	0.60	0.72	0.74	0.71
2	0.77	0.61	0.68	0.58	0.76
3	0.81	0.45	0.46	0.51	0.58
4	0.73	0.57	—	—	—
5	0.72	0.61	0.66	0.44	0.75
6	0.53	0.65	0.55	0.37	0.66
7	0.64	0.35	0.31	0.17	0.77
8	0.66	0.46	0.50	0.50	0.69
9	0.69	0.59	0.58	0.47	0.68
10	0.74	—	—	—	—
11	0.27	0.39	—	—	—
12	—	0.72	0.48	0.32	0.68
13	—	—	0.52	0.60	0.82
14	—	—	0.57	0.58	0.79

Table 4
Weighted Score for the Entire Laboratory and for Individual Cytotechnologists

	July-December 1997	January-June 1998	July-December 1998	January-June 1999	July-December 1999
Entire laboratory	3.55	3.24	3.10	3.04	3.56
Cytotechnologist No.					
1	4.19	3.68	2.64	3.33	2.43
2	4.60	3.11	3.35	2.80	3.49
3	3.63	3.75	3.00	3.30	3.46
4	3.41	2.28	—	—	—
5	3.51	3.07	2.96	3.11	3.02
6	4.51	3.67	3.77	3.31	4.87
7	3.31	3.09	2.82	2.92	3.72
8	3.22	3.42	3.52	2.99	3.68
9	3.29	3.43	3.25	3.14	3.21
10	3.51	—	—	—	—
11	2.67	3.59	—	—	—
12	—	3.40	3.63	3.39	3.33
13	—	—	3.07	3.11	4.09
14	—	—	3.29	3.53	3.55

Table 5
Summary of Discordance Measures for the Entire Laboratory by 6-Month Period

	July-December 1997	January-June 1998	July-December 1998	January-June 1999	July-December 1999
Median percentage of discordant cases (range)	21 (13-50)	28 (18-43)	28 (20-45)	34 (19-56)	18 (12-30)
Mean kappa (range)	0.69 (0.27-0.81)	0.48 (0.39-0.72)	0.48 (0.31-0.72)	0.38 (0.17-0.74)	0.75 (0.58-0.82)
Mean weighted score (range)	3.55 (2.67-4.60)	3.24 (2.28-3.75)	3.10 (2.64-3.77)	3.04 (2.80-3.53)	3.56 (2.43-4.87)

Table 6
Cytotechnologist-Cytopathologist Discrepancy Log: January-June 1999 (n = 4,287)*

	Cytopathologist Diagnosis									
	UNS	WNL	BCC	EPM	ASCUS	AGUS	LSIL	SILUNC	HSIL	CAR
Cytotechnologist diagnosis										
UNS	**7**	2	2	—	1	—	—	—	—	—
WNL	7	**28**	25	3	38	1	—	—	—	—
BCC	7	30	**447**	4	141	10	—	1	2	—
EPM	—	11	2	**16**	—	1	—	—	—	—
ASCUS	15	160	808	4	**1,440**	6	71	7	19	—
AGUS	2	8	74	6	10	**16**	—	—	2	3
LSIL	—	1	7	—	151	—	**351**	15	17	—
SILUNC	1	—	—	—	9	—	7	**7**	4	—
HSIL	—	—	1	—	54	2	33	23	**160**	1
CAR	—	—	1	—	—	—	—	—	—	**11**

AGUS, atypical glandular cells of undetermined significance; ASCUS, atypical squamous cells of undetermined significance; BCC, benign cellular changes; CAR, carcinoma; EPM, endometrial cells, cytologically benign, in a postmenopausal woman; HSIL, high-grade squamous intraepithelial lesion; LSIL, low-grade squamous intraepithelial lesion; SILUNC, squamous intraepithelial lesion, difficult to grade; UNS, unsatisfactory; WNL, within normal limits.
* Numbers in bold type indicate concordance.

percentage of diagnostic disagreements between CT 7 and the CPs (55.8%), accompanied by the low kappa statistic (0.17), were noted, and remedial action was instituted. CT 7 was asked to review all cases judged to be nonnegative with the on-service CP at a double-headed microscope for the next 6 months. (Laboratory policy at that time specified that only cases in which a discrepancy of 2 major diagnostic categories occurred needed to be reviewed by the CT and CP at a double-headed microscope).

At the next semiannual performance evaluation, a marked improvement was seen in the work of CT 7 (the discordance rate fell from 55.8% to 16.2%) and in the work of the other 6 CTs, whose rates followed the trend of CT 7. The median percentage of discordant cases for the entire

laboratory decreased from 34.0% to 18.3%, and the kappa value rose from 0.38 to 0.75.

Discussion

Aside from the valuable daily interactions between CTs and CPs, which provide subjective impressions on the quality of a CT's performance, objective measures of diagnostic precision can be useful. In our laboratory, the objective measures of CT performance confirmed the subjective impression of the CPs during the study period, namely, that assessments by CT 7 of the degree of an abnormality were frequently 1 diagnostic category higher than that of the CP. Remedial action was instituted, and the CPs noticed an improvement in the provisional diagnoses recorded by CT 7. This was reflected in the data generated for the next semiannual performance evaluation.

It is not clear why the performance of the entire laboratory improved when remedial action was applied to only 1 CT. Granted, the data for CT 7 skewed the laboratory values because of the higher productivity of CT 7; during the first 6-month period, the cases for CT 7 accounted for 28.7% of the work submitted to the CPs for review; in subsequent periods, this percentage was even higher. It is possible, for example, that criteria for diagnosis passed on from the CPs to 1 CT filtered across to other CTs. It is also possible that word of remedial action leaked out and prompted other CTs to improve their performance.

Although the discordance rate increased and the kappa values declined during the first 4 study periods (Table 5), the weighted score revealed that the average discordance decreased in severity. Thus, although there was an increase in the number of discordances, these were comparatively minor disagreements. This was not surprising and again confirmed the subjective impression of the CPs, who, although they sensed an increase in discordant diagnoses, also sensed that the disagreements were mostly minor.

We are not aware of published data that tabulate and analyze CT-CP diagnostic concordance. For this reason, there are no readily accessible benchmarks, and we used our judgment in determining thresholds for remedial action. In this early outing, a discordance rate that rose above 50% and a kappa value below 0.30 triggered action. These benchmarks very likely will be modified as more experience in this laboratory and others is accrued. For example, a relative increase in the percentage of discordance may prove to be a more meaningful threshold for action compared with an absolute percentage of discordances. Benchmarks also may vary from laboratory to laboratory because of variables that may be impossible to control. For example, the concordance between CTs and CPs may be a function of the time the CTs have worked in the laboratory.

The program we designed also is useful because it identifies, in list form, all the discordant cases by accession number. These are sorted by diagnostic heading, which makes it possible to retrieve slides for focused review. For example, the supervisor may choose to review all serious discordances, such as cases called BCC by the CT that were revised to high-grade squamous intraepithelial lesion by the CP. Alternatively, if a CT has a tendency to overcall, the supervisor can audit cases called ASCUS by the CT and changed to within normal limits or BCC by the CP.

Although the design and testing of this program took many hours, now that it is in place, the reports can be queued to print in a matter of seconds. This is critical because, although efforts to improve the quality of cytologic evaluation deserve consideration, if they are laborious, time-consuming, or expensive, they are doomed to failure and may instead prove to be encumbrances to cancer prevention.[5] We believe that the program we describe can be a simple tool, one of many, for use in evaluating the complex tasks of the CT.

From the Departments of [1]Pathology and [2]Information Systems and the [3]Neuroepidemiology Unit, Brigham and Women's Hospital; [4]Children's Hospital Medical Center; [5]Harvard School of Public Health; and [6]Harvard Medical School, Boston, MA.

Address reprint requests to Dr Cibas: Dept of Pathology, Brigham and Women's Hospital, 75 Francis St, Boston, MA.

Acknowledgment: We thank Shirley E. Greening, MS, JD, CFIAC, for providing the grids used at Thomas Jefferson University, Philadelphia, PA, for grading cytotechnology students.

References

1. Krieger P, Naryshkin S. Random rescreening of cytologic smears: a practical and effective component of quality assurance programs in both large and small laboratories. *Acta Cytol*. 1994;38:291-298.
2. Krieger PA, Cohen T, Naryshkin S. A practical guide to Papanicolaou smear rescreens: how many slides must be evaluated to make a statistically valid assessment of screening performance? *Cancer*. 1998;84:130-137.
3. US Department of Labor, Occupational Health and Safety Administration. Laboratory Requirements. Condition: Cytology, 58 *Federal Register* 7169 (1992).
4. Fleiss JL. *Statistical Methods for Rates and Proportions*. 2nd ed. New York, NY: John Wiley; 1988:212-236.
5. Krieger PA, McGoogan E, Vooijs GP, et al. Quality assurance/control issues: IAC Task Force summary. *Acta Cytol*. 1998;42:133-140.

HOLOGIC®

First and Only FDA Cleared
Digital Cytology System

Genius™ Cervical AI

Genius™ Review Station

Genius™ Digital Imager

Empower Your Genius With Ours

Make a Greater Impact on Cervical Cancer
with the Advanced Technology of the
Genius™ Digital Diagnostics System

Click or Scan
to discover more

genius™
DIGITAL DIAGNOSTICS