INTEROBSERVER VARIABILITY IN THE INTERPRETATION OF ROUTINE GYNECOLOGICAL SPECIMENS WITHIN A LARGE PRACTICE

F.A. MEIER[1], R. VARNEY[1], S.R. KINI[1], A.R. GABA[1], C.H. STONE[1], D.S. SCHULTZ[1], T. KLIMOWICZ[1],

M. BANSAL[1], C. VRBIN[2], S.S. RAAB[2], R.J. ZARBO[1]

1. **Department of Pathology and Laboratory Medicine, Henry Ford Hospital, Detroit, Michigan**
2. **Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania**

BACKGROUND: The interpretation of gynecological specimens (pap smear cytologies and the histopathology of cervical biopsies) is one of the most frequent examples of morphological diagnosis in clinical medicine. In our large pathology practice we examined 63,678 pap smears and approximately 3000 correlating Cx Bx's in 2004.

Because of their high volume in our practice, such specimens are interpreted by many observers. These observers make their interpretations against different diagnostic horizons, depending on whether the interpreter is a cytotechnologist (a medical technologist with special professional competence in cytological screening and diagnosis), a cytopathologist (a physician with specialty qualification in pathology and subspecialty training and experience in morphologic diagnosis from exfoliated cells), or a surgical pathologist (a physician with specialty qualifications in pathology and subspecialty focus in the diagnosis of histological sections but usually not exfoliated cytological smears).

Cytotechnologists spend their working days examining smears, both screening for and re-screening (confirming) the presence or (much more often) the absence of diagnostic findings in cytological preparations. Cytopathologists confirm or revise findings turned up by screening and re-screening and place diagnostic findings in the appropriate classification. Surgical pathologists examine, as cytopathologists usually also examine, histological material from biopsies which cytological specimens either have triggered or accompany.

Over time, an individual patient's cytological and histopathological findings tend to be reviewed by several different cytotechnologists, cytopathologists, and surgical pathologists.

The context of this study is the variation which such multiple review introduces into the diagnostic process.

METHODS: To quantify interobserver (percent) agreement and variability (by the Kappa statistic), we presented 30 liquid-based cytology slides and 24 cervical biopsy slides for interpretation by: 10 cytotechnologists, who, as in usual practice, reviewed the pap smears only; 4 cytopathologists, who, also as in routine practice, reviewed both pap smears and cervical biopsy slides; and 9 surgical pathologists who, as in the usual routine, examined just cervical biopsy slides. Both smears and sections were de-identified as to case number, etc. Among the biopsy slides, only 'diagnostic slides' (those with interpretable findings) were included. Agreement was measured by the fraction of interpreters giving the same result from the same material (percent agreement). Variation in diagnostic concordance was assessed using the kappa statistic (kappa ranges): very high concordance was assessed as a kappa of 0.8, high concordance a kappa of 0.6-0.79, moderate concordance 0.4-0.59, and fair-to-slight concordance of 0.39 or less.

For cytological slides, the agreement and concordance among cytotechnologists, among cytopathologists, and between these two groups were all three determined. For histological slides, agreement and concordance among cytopathologists, among surgical pathologists, as well as also between these two groups, were all three assessed.

For cytological smears, classifiers placed diagnoses in one of five categories: negative (neg), atypical squamous cells of undetermined significance (ASCUS), atypical squamous cells of undetermined significance, probably high grade (ASC-H), low grade squamous intraepithelial lesion (LSIL) and high grade squamous intraepithelial lesion (HSIL). For histological slides, classifiers placed diagnoses in one of four categories: negative (neg), cervical intraepithelial neoplasia-I (CIN I), cervical intraepithelial neoplasia – II (CIN II), and cervical interepithelial neoplasia – III / carcinoma in-situ (CIN III/CIS).

Table 1

% Agreement By Cytology Smear Among CTOTECHNOLOGISTS

Cytotechs	NIL (%)	ASCUS (%)	ASC-H (%)	LSIL (%)	HSIL (%)
1	91	9	0	0	0
2	0	0	0	27	73
3	9	36	0	55	0
4	9	36	0	55	0
5	100	0	0	0	0
6	0	0	0	55	45
7	18	9	73	0	0
8	0	18	27	55	0
9	0	0	0	100	0
10	100	0	0	0	0
11	27	55	0	18	0
12	45	9	45	0	0
13	0	0	0	0	100
14	0	18	9	73	0
15	27	45	9	18	0
16	9	0	64	0	27
17	91	0	9	0	0
18	0	0	0	100	0
19	18	0	55	9	18
20	0	0	0	9	91
21	0	0	9	91	0
22	9	27	55	0	9
23	100	0	0	0	0
24	0	0	0	100	0
25	0	0	0	0	100
26	18	18	64	0	0
27	91	9	0	0	0
28	18	45	27	0	9
29	45	27	18	0	9
30	0	0	18	9	73

Table 3 - % Agreement By Cervical Biopsy Slide Among Cytopathologists

Cytopathologists	Negative (%)	CIN I (%)	CIN II (%)	CIN III (%)
1	0	100	0	0
2	50	40	0	10
3	0	90	10	0
4	0	80	20	0
5	0	20	0	80
6	50	50	0	0
7	30	70	0	0
8	70	20	10	0
9	10	70	20	0
10	0	0	50	50
11	20	50	30	0
12	100	0	0	0
13	60	40	0	0
14	0	20	60	20
15	0	0	30	70
16	90	10	0	0
17	0	90	10	0
18	60	40	0	0
19	0	0	10	90
20	0	50	50	0
21	20	60	20	0
22	20	50	30	0
23	60	40	0	0
24	20	10	30	40

Table 2

% Agreement By Cytology Smear Among Cytopathologists

Cytopathologists	NIL (%)	ASCUS (%)	ASC-H (%)	LSIL (%)	HSIL (%)
1	60	40	0	0	0
2	0	0	0	60	40
3	0	40	0	60	0
4	0	60	0	40	0
5	80	20	0	0	0
6	0	0	0	60	40
7	0	20	40	0	40
8	0	20	20	40	20
9	0	0	0	100	0
10	80	20	0	0	0
11	0	100	0	0	0
12	60	20	20	0	0
13	0	0	0	0	100
14	0	40	0	60	0
15	0	40	20	20	20
16	0	0	60	0	40
17	40	20	20	0	20
18	0	20	0	80	0
19	0	20	40	20	20
20	0	0	0	20	80
21	0	0	0	100	0
22	20	20	40	0	20
23	80	0	20	0	0
24	0	0	0	60	40
25	0	0	20	0	80
26	0	20	80	0	0
27	40	60	0	0	0
28	0	60	0	0	40
29	0	40	20	0	40
30	0	0	0	0	100

Table 4 – % Agreement by Cervical Biopsy Slide Among Surgical Pathologists

Surg.Pathologists	Negative (%)	CIN I (%)	CIN II (%)	CIN III (%)
1	0	100	0	0
2	80	0	0	20
3	0	60	40	0
4	0	80	20	0
5	0	20	40	40
6	100	0	0	0
7	20	60	20	0
8	80	20	0	0
9	20	20	60	0
10	0	0	0	100
11	0	20	60	20
12	100	0	0	0
13	60	40	0	0
14	0	0	100	0
15	0	0	0	100
16	80	20	0	0
17	0	80	20	0
18	80	20	0	0
19	0	0	0	100
20	0	20	40	40
21	20	60	0	20
22	40	40	20	0
23	60	40	0	0
24	0	0	20	80

Group and Inter-Group

Percent Agreement and Degree of Concordance

DESIGNATIONS	GROUP/INTER-GROUP AGREEMENT (%)	KAPPA VALUE (RANGES)
Cytotechnologist (smears)	79	0.398-0.801
Cytopathologist (smears)	66.6	0.348-0.681
Cytotechnologist/cytopathologist (smears)	75.2	0.22-0.74
Cytopathologist (Cx Bx slides)	58.3	0.390-0.598
Surgical pathologist (Cx Bx slides)	51.2	0.066-0.709
Cyto/surgical pathologist (Cx Bx slides)	50.5	0.066-0709

DISCUSSION: Agreement among cytotechnologists was moderately high, almost 80%, but the kappa range was relatively wide, from almost .40('moderate') to .80 ('very high'). Agreement among cytopathologists was, interestingly, less (66.6%), but with a narrower kappa range over a lower spectrum of concordance from about .35 ('fair') to .68 ('high'). In contrast, the *inter-group* agreement between the aggregate of cytotechnologists and aggregate of cytopathologists was higher (75%), almost as high as the cytotechnologists *'intra-group'* agreement 79% ('high'), but the - *'inter-group'* kappa range was substantially wider: from 0.22 ('poor') – 0.74 ('high').

Agreement on cervical biopsy slides was less, in general, than the agreement on pap smears. The agreement among cytopathologists was only 58% with a low-to-moderate kappa but a narrower kappa range (.29-.60). Agreement among surgical pathologists was even worse, 51%, with the widest kappa range measured, one extending from the 'slight' agreement range to the 'high' agreement range (0.066 – 0.709). This level of disagreement and range of discord are further reflected in the intergroup comparison between cyto- and surgical pathologists, where the same 51% agreement and the same wide 0.066-0.709 kappa range are observerved.

Historically, the combination of pap smear cytology and cervical biopsy pathology has proven successful in improving patient outcomes by decreasing the incidence, morbidity, and mortality from cervical cancer, however, their agreement and 'precision' as diagnostic methods appears severely limited by interobserver variation.

CONCLUSIONS: Agreement about pap smear diagnoses among cytotechnologists and between this group of screening diagnosticians and cytopathologists was moderately high (in the 75-80% range), but kappas within the two groups of cyto-diagnosticians were only moderate. Among cytopathologists interobserver concordance varied over a wide range of kappas (from 'poor' to 'high').

Agreement about cervical biopsies' histological diagnoses were lower (in the 50-60% range), with spectacularly wide kappas, (from 'poor' to 'high' depending on diagnoses and specific pairs of diagnosticians).

This study appears indeed to suggest not only that interobserver variability in the interpretation of routine gynecological specimens is wide but also that rates of (dis)agreement vary substantially among different groups of observers: cytotechnologists' agreement > cytopathologists' and surgical pathologists' agreement).

It further argues that agreement is better for the 'screening' pap smears than for the putatively 'definitive' cervical biopsies. Regarding the latter, poor agreement, the wide variation in diagnostic concordance is a cause for concern. Interobserver variability among cytopathologists and surgical pathologists produces non-biological variation in diagnoses that, over time, can produce variation in therapeutic patient management that would not be due to changes in the patient's condition, or to the quality or type of specimen, but, rather, to who interprets it.