EARLY ONLINE RELEASE

Note: This article was posted on the Archives Web site as an Early Online Release. Early Online Release articles have been peer reviewed, copyedited, and reviewed by the authors. Additional changes or corrections may appear in these articles when they appear in a future print issue of the Archives. Early Online Release articles are citable by using the Digital Object Identifier (DOI), a unique number given to every article. The DOI will typically appear at the end of the abstract.

The DOI for this manuscript is doi: 10.5858/arpa.2020-0686-OA

The final published version of this manuscript will replace the Early Online Release version at the above DOI once it is available.
Accuracy and Reproducibility of Frozen Section Diagnosis in Ovarian Tumors

A 10-Year Experience at a Tertiary Cancer Center

Feras Zaiem, MD; Hany Deirawan, MD; Raghad Kherallah, MD; Omar Fehmi; Hyejeong Jang, MS; Seongho Kim, PhD; Sudeshna Bandyopadhyay, MD; Rouba Ali-Fehmi, MD

Context.—Intraoperative consultation—frozen section diagnosis (FSD)—determines tumor pathology and guides the optimal surgical management of ovarian neoplasms intraoperatively.

Objective.—To evaluate the diagnostic accuracy of the FSD and analyze the discrepancy between the FSD and final diagnosis.

Design.—This is a retrospective study of 618 ovarian neoplasm FSDs from 2009 to 2018 at a tertiary health care center. The discrepant cases were reviewed and reevaluated by gynecologic and general surgical pathologists. The outcomes of interest were performing unnecessary procedure, returning for a second surgery, and 30-day postoperative mortality.

Results.—The sensitivity and the positive predictive value of the FSD were lower in borderline tumors than in benign and malignant epithelial ovarian tumors. Major and minor discrepancies were identified in 5.3% (33 of 618) and 12.3% of (76 of 618) cases, respectively. A root cause analysis of the major discrepant cases showed that sampling error accounted for 43% (14 of 33). The discrepancy distributions of gynecologic and general surgical pathologists were statistically similar in the overall cohort (\( P = .65 \)). The overall \( k \) for diagnostic agreement among gynecologic pathologists, general surgical pathologists, and final diagnosis was 0.18 (0.10–0.26, \( P < .001 \)), implying only a slight overall agreement. Of the major discrepant cases, only 3 had a clinical implication. One overdiagnosed patient underwent unnecessary procedure and 2 underdiagnosed patients were recommended to return for a second surgery. No patient had 30-day postoperative mortality.

Conclusions.—Frozen section diagnosis remains a definitive diagnostic tool in ovarian neoplasms and plays a crucial role in guiding intraoperative surgical management.

(Arch Pathol Lab Med. doi: 10.5858/arpa.2020-0686-OA)

Ovarian cancer is the third most common gynecologic tumor, after cervical and uterine cancer.\(^1\) It ranks 13th in cancer deaths among women, accounting for more deaths than any other cancer of the female reproductive system.\(^2\) Ovarian neoplasms are classified histologically into epithelial, germ cell, sex cord–stromal, and miscellaneous, with epithelial tumors being the most common. Epithelial tumors are stratified according to clinical behavior into benign, borderline, and malignant.\(^3\) The clinical diagnosis of ovarian cancer is problematic given its acute and subacute presentation, limited sensitivity of laboratory and radiologic techniques, and the risk of cancer cells seeding during biopsy. Surgical exploration remains the diagnostic procedure of choice in ovarian cancer.\(^4\)

Intraoperative consultation, also known as intraoperative frozen section diagnosis (FSD) is particularly useful in guiding surgical management when evidence of malignancy is not definitive. Frozen section diagnosis provides real-time objective clinical information that can guide the surgeon on the extent of the needed surgery. This is achieved by determining if the lesion is neoplastic in nature and its malignant potential.

The aims of this study are to report our experience in ovarian neoplasms intraoperative consultation at a tertiary health care center, to evaluate the impact of pathologists’ gynecologic experience on the FSD, and to shed light on the clinical implication of intraoperative consultation.

MATERIALS AND METHODS

Institutional review board approval was obtained for medical record and pathology reports of patients with ovarian neoplasms seen at Detroit Medical Center/Wayne State University, Detroit, Michigan, between January 1, 2009, and July 31, 2018. We included patients with ovarian neoplasms who underwent surgery in which an intraoperative consultation (FSD) was requested. The electronic medical records of the included patients were reviewed, including...
preoperative clinical impression, pathology reports, and postoperative clinical course. Ovarian tumors were classified according to World Health Organization criteria.5

Cases with an FSD of defer to permanent (n = 23) were excluded from our study. Cases with a discrepancy between FSD and final diagnosis were identified and classified into major or minor discrepancy. Major discrepancies were defined as those that needed reclassification to a different category of benign, borderline, or malignant tumors with either overdiagnosis or underdiagnosis and hence had a major impact on the surgical outcome. Minor discrepancies were defined as those that needed reclassification within the same category of benign, borderline, or malignant tumor without major impact on the clinical outcome or clinical decision-making.

Seven experienced pathologists participated in the study. They were categorized into 4 gynecologic pathologists (pathologists A, B, C, and D) and 3 general surgical pathologists (pathologists E, F, and G) according to their fellowship training and area of focused sign-out. The frozen section glass slides of the major discrepancy cases (range, 1–4 frozen section slides) were unlabeled and distributed among the 7 pathologists. They were blinded to the FSD and the final pathologic diagnosis and independently reviewed and recorded their interpretation using a standardized form. This form provided basic patient information, including age, tumor size, gross description, tumor rupture status, and a classification scheme following the World Health Organization classification. The diagnostic accuracy (sensitivity, specificity, positive predictive value, and negative predictive value) of the FSD is calculated using the final pathologic diagnosis as a gold standard test. The agreement was calculated first between each pathologist’s FSD diagnosis and the final diagnosis (intra-rater reliability), then between pathologists (inter-rater reliability).

A root cause analysis of the major discrepancy cases was performed. To accomplish that, we stratified those cases into interpretation error, sampling error, or other nonidentifiable error. Interpretation error is an error due to an interpretation mistake in the FSD. It was inferred when the majority of reviewing pathologists (≥6 of 7) agreed with the final diagnosis and disagreed with the erroneous FSD interpretation. Sampling error is an error due to undersampling of the lesion in the FSD setting. This was inferred when the majority of reviewing pathologists (≥6 of 7) agreed with the erroneous FSD interpretation and disagreed with the final diagnosis. The cases that did not fall into the previous 2 categories were assigned as nonidentifiable error.

Finally, we evaluated the clinical end points of those major discrepancy cases. The outcomes of interest were performance of unnecessary procedure, return for a second surgery, and 30-day postoperative mortality.

Statistical Methods

Associations between 2 categorical were was assessed using a χ² test. Exact agreement percentages and Cohen κ were calculated to measure the degree of agreement (interrater agreement) for each pair (between 2 raters or between a rater and gold standard).6 Because the Cohen κ is an agreement measure only for a pair, an overall κ for multiple raters (ie, among 3 or more raters) was calculated using the Fleiss κ.6 and overall agreement with gold standard and between gynecologic pathologists and general surgical pathologists was calculated using Cohen κ with combined occasions. The κ coefficients can be interpreted using the guidelines outlined by Landis and Koch5 as follows: less than 0.00, poor; 0.00 to 0.20, slight; 0.21 to 0.40, fair; 0.41 to 0.60, moderate; 0.61 to 0.80, substantial; and 0.81 to 1.00, almost perfect agreement.

Results

A total of 618 patients with ovarian neoplasms were included in the study. Epithelial tumors accounted for most of the ovarian neoplasms, 59.2% (366 of 618), followed by sex cord tumors, 9.8% (61 of 618); germ cell tumors, 9.4% (58 of 618); metastasis, 5.8% (36 of 618); and other benign nonneoplastic entities, 15.7% (97 of 618). Of the epithelial tumors, 74.5% (273 of 366) were benign, 8.4% (31 of 366) borderline, and 17.1% (62 of 366) malignant.

The diagnostic accuracy of the FSD in the total cohort is detailed in Table 1. The sensitivity of FSD in epithelial tumors was lower in borderline cases at 79.2% compared with benign (98.7%) and malignant cases (87%). Similarly, the positive predictive value of FSD epithelial tumors was lower in borderline cases (80.8%) than in benign (96.5%) and malignant cases (95.3%). The sensitivity of FSD in germ cell tumors was the highest (98.3%), followed by sex cord tumors (96.7%). The specificity and negative predictive value of FSD were high in all nonepithelial tumors. The positive predictive value of FSD was the highest in germ cell tumors (99.3%), followed by nonneoplastic entities.

Of the overall cohort cases, discrepancy between FSD and final diagnosis was seen in 17.6% (109 of 618). Those cases were categorized as major in 5.3% (33 of 618) and minor in 12.3% (76 of 618). Mucinous epithelial tumors accounted for the most major discrepancies, 27.3% (9 of 33). Among the major discrepancies, 7 were overdiaognosed and 26 were underdiagnosed cases. Detailed characteristics of the discrepant cases are seen in Table 2. The discrepancy distributions of the overall cohort between gynecologic and general surgical pathologists were statistically similar to each other (P = .65) (Table 3).

Seven experienced pathologists were blinded to FSD and final diagnosis and independently reviewed the 33 major discrepant cases. The overall κ for diagnostic agreement between all 7 gynecologic and general surgical pathologists and final diagnosis was 0.18 (0.10–0.26, P < .001), implying only slight overall agreement (Figure 1). The gynecologic pathologists had a higher overall κ of 0.24 (0.13–0.35) than the general surgical pathologists’ 0.10 (0.01–0.21), but their 95% CIs overlapped (Figure 1).

Regarding agreements between pathologists, the overall κ was 0.46 (0.40–0.52), implying a moderate overall agreement among raters. The highest κ was found between

Table 1. Patient Characteristics and Diagnostic Accuracy of Frozen Section Diagnosis in Ovarian Neoplasms

<table>
<thead>
<tr>
<th>Category</th>
<th>Sensitivity, %</th>
<th>Specificity, %</th>
<th>Positive Predictive Value, %</th>
<th>Negative Predictive Value, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benign epithelial tumor (n = 273)</td>
<td>98.7</td>
<td>90.5</td>
<td>96.5</td>
<td>96.8</td>
</tr>
<tr>
<td>Borderline epithelial tumor (n = 31)</td>
<td>79.2</td>
<td>98.2</td>
<td>80.8</td>
<td>98.1</td>
</tr>
<tr>
<td>Malignant epithelial tumor (n = 62)</td>
<td>87</td>
<td>99</td>
<td>95.3</td>
<td>97.1</td>
</tr>
<tr>
<td>Sex cord tumors (n = 61)</td>
<td>96.7</td>
<td>99.6</td>
<td>96.7</td>
<td>99.6</td>
</tr>
<tr>
<td>Germ cell tumors (n = 58)</td>
<td>98.3</td>
<td>99.8</td>
<td>99.3</td>
<td>99.8</td>
</tr>
<tr>
<td>Metastatic tumors (n = 36)</td>
<td>94.4</td>
<td>99.7</td>
<td>94.4</td>
<td>99.7</td>
</tr>
<tr>
<td>Other nonneoplastic entities (n = 97)</td>
<td>95</td>
<td>99.6</td>
<td>97.9</td>
<td>99</td>
</tr>
</tbody>
</table>

Frozen Section Diagnosis in Ovarian Tumors—Zaim et al
Intraoperative consultation (FSD) plays a crucial role in determining the appropriate surgical management of ovarian neoplasms. However, the surgeon and the pathologist should be aware of the limitation of FSD. One of the main limitations is sampling error, especially in tumors larger than 10 cm. Other limitations include freezing artifacts, time limitation, technical problems, lack of ancillary studies, and the pathologist’s subspecialty and experience.9

Frozen section diagnosis of ovarian tumors continues to pose a significant diagnostic challenge for practicing pathologists. A recent study of 4785 frozen section diagnoses ranked ovarian tumors as the third most frequently discrepant.10 A large body of literature has addressed the reliability of intraoperative FSD in ovarian tumors with variable conclusions. In a meta-analysis11 of 18 studies, the accuracy of FSD in ovarian tumors diagnosis was good, with sensitivity ranging from 65% to 97% and specificity from 97% to 100%. Borderline tumors have been notoriously problematic on FSD.12–14 Our results also showed that epithelial borderline tumors have lower sensitivity and positive predictive value than both benign and malignant tumors. The diagnostic criteria of ovarian borderline and malignant tumors require adequate extensive sampling to establish 10% atypical proliferation features with absence of invasion for borderline tumors and at least a single focus of frank invasion for malignant tumors; this might explain their FSD difficulty and discrepancy.15,16 Our study showed only a slightly higher discrepancy rate between FSD and final diagnosis in ovarian tumors (17.6%) than reported in the literature7 (8%). We propose that our higher rates reflect a more stringent application of

### DISCUSSION

Surgical management of ovarian tumors depends on the nature of the neoplasm. Benign tumors are treated conservatively with cystectomy or oophorectomy. Borderline tumors require pelvic lymph node dissection with limited omental sampling. However, fertility and ovarian function–sparing surgery is still preferred in young women. In malignant tumors, staging and surgical reduction including total abdominal hysterectomy with bilateral salpingo-oophorectomy, lymph node sampling, peritoneal sampling, and omentectomy are recommended.8

Preoperative clinical impression, radiology, and CA125 serum levels are not a definitive diagnostic method in ovarian tumors. Intraoperative consultation (FSD) plays a crucial role in determining the appropriate surgical management of ovarian neoplasms. However, the surgeon and the pathologist should be aware of the limitation of FSD. One of the main limitations is sampling error, especially in tumors larger than 10 cm. Other limitations include freezing artifacts, time limitation, technical problems, lack of ancillary studies, and the pathologist’s subspecialty and experience.9

### Table 2. Characteristics of the Cases Based on Discrepancy Between Frozen Section Diagnosis and Final Diagnosis

<table>
<thead>
<tr>
<th>All Cases (N = 618)</th>
<th>Epithelial Tumors (n = 366)</th>
<th>Sex Cord Tumors (n = 61)</th>
<th>Germ Cell Tumors (n = 58)</th>
<th>Metastasis (n = 36)</th>
<th>Other Benign Nonneoplastic Entities (n = 97)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No discrepancy (82.4%, n = 509)</td>
<td>292</td>
<td>45</td>
<td>54</td>
<td>28</td>
<td>90</td>
</tr>
<tr>
<td>Minor discrepancies (12.3%, n = 76)</td>
<td>51</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>Major discrepancies (5.3%, n = 33)</td>
<td>23</td>
<td>4</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Major discrepancies overdiagnosed (n = 7)</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Major discrepancies underdiagnosed (n = 26)</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 3. Correlations Between Pathologists and Overall Cohort Cases

<table>
<thead>
<tr>
<th>Discrepancy</th>
<th>General Surgical Pathologists, No. (%) (n = 415)</th>
<th>Gynecologic Pathologists, No. (%) (n = 203)</th>
<th>P Valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>None (n = 509)</td>
<td>344 (83)</td>
<td>165 (81)</td>
<td>.65</td>
</tr>
<tr>
<td>Minor (n = 76)</td>
<td>50 (12)</td>
<td>26 (13)</td>
<td></td>
</tr>
<tr>
<td>Major (n = 33)</td>
<td>21 (5)</td>
<td>12 (6)</td>
<td></td>
</tr>
</tbody>
</table>

a χ² test.
Our study was consistent with the latter, showing an overall agreement between gynecologic pathologists and general surgical pathologists in the overall cohort and roughly a similar discrepancy rate between them. However, the fact that sampling error accounts roughly for half of the major discrepant cases (43%) might explain the overall low $k$ of 0.18 (0.10–0.26, $P < .001$).

The immediate clinical ramifications of FSD error are critical and may result in harm to the patient. Overdiagnosis during FSD can result in unnecessary widening of the surgical field with increased morbidity and mortality. Another concern with overdiagnosis is the loss of fertility and ovarian function in younger patients from radical resections.\textsuperscript{21,22} Our cohort showed only one overdiagnosed patient on FSD who had unnecessary surgery. However, she was 75 years old and high morbidity and mortality was the main concern rather than loss of fertility. Underdiagnosis can result in suboptimal surgical treatment and lead to a second surgery and possible tumor spread. Few studies\textsuperscript{23} have shown that the restaging procedure has no impact on survival in borderline ovarian tumors. Our study identified 2 underdiagnosed patients; however, one patient had already had the cyst ruptured intraoperatively and was upstaged for pT1c and the other patient declined the option for a secondary surgery and instead received chemotherapy.

This study is limited by its retrospective nature and by the fact that it does not account for advances in imaging modalities to guide surgical management. Another limitation is that although this was a large cohort study ($N = 618$), the actual number of problematic cases on frozen sections was limited, which may have affected the relevant outcomes. Nevertheless, the strengths of this article are derived from the evaluation of the gynecologic training of pathologists on the intraoperative diagnosis, the study of the root cause analysis of the major discrepant cases, and the assessment of the clinical implication of the discrepant cases on patient outcome.

In conclusion, the diagnostic accuracy rate for FSD remains high for benign and malignant ovarian tumors but is relatively low for borderline ovarian tumors. Pathologists should take into consideration a few factors that might help to mitigate FSD discrepancies, such as learning and appreciating the histologic limitations of FSD and appropriate deferral to permanent section diagnosis when needed. We suggest more thoughtful examination and sampling of ovarian lesions; however, some invasion features of the tumor might be appreciated only on permanent sections, so changing the diagnosis line to permanent sections might help to mitigate FSD discrepancies, such as learning and appreciating the histologic limitations of FSD and appropriate deferral to permanent section diagnosis when needed.

Table 4. Characteristics of Major Discrepant Cases ($n = 33$)

<table>
<thead>
<tr>
<th>Cause of Discrepancy</th>
<th>epithelial tumors ($n = 23$)</th>
<th>sex cord tumors ($n = 4$)</th>
<th>germ cell tumors ($n = 2$)</th>
<th>metastasis ($n = 4$)</th>
<th>other benign nonneoplastic entities ($n = 0$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretation error ($n = 7$)</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sampling error ($n = 14$)</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Other nonidentifiable error ($n = 12$)</td>
<td>10</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 2. The $k$ agreements between pathologists (interrater). The $k$ statistics were generated by Cohen $k$ and tested against the null hypothesis of the true $k$ of 0, except for overall $k$ statistics by Fleiss $k$. Gynecologic pathologists (Gyn) are pathologists A, B, C, and D, and general surgical pathologists (Non-Gyn) are pathologists E, F, and G.
Figure 3. Patient 1 was overdiagnosed with a serous borderline tumor on frozen section (A, frozen section slide); the final diagnosis was serous cystadenofibroma (B, histology slide). Patient 2 was underdiagnosed on frozen section as mucinous borderline tumor (C, frozen section slide); the final diagnosis was mucinous carcinoma (D, histology slide). Patient 3 was underdiagnosed with benign luteinized cyst (E, frozen section slide); the final diagnosis was consistent with serous borderline tumor (F, histology slide) (hematoxylin-eosin, original magnification ×10 objective).
We thank Suzanne Jacques, MD; Faisal Qureshi, MD; Rafic Beydoun, MD; Fulvio Lonardo, MD; Dongping Shi, MD; and Amy Harper, MD, for their contributions to this paper.

References