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% (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. It ranks 13th in cancer deaths among women, accounting for more deaths than any other cancer of the female reproductive system. 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. 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.

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
Because the Cohen κ was calculated using the Fleiss κ, 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 Koch 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 overdiagnosed 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 was 0.54 (0.40–0.65), implying only slight overall agreement (Figure 1). The gynecologic pathologists had a higher overall κ of 0.40 (0.26–0.55) than the general surgical pathologists’ of 0.32 (0.19–0.45), 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 epithelial tumors.
pathologists A (gynecologic) and D (gynecologic) of 0.80 (0.60–1.00). The gynecologic pathologists had a slightly higher overall $\kappa$ of 0.47 (0.35–0.58) than the general surgical pathologists’ of 0.45 (0.29–0.62), but their 95% CIs overlapped (Figure 2).

A root cause analysis of the major discrepancy cases was performed (Table 4). Sampling error accounted for roughly half of the cases, 43% (14 of 33). The erroneous FSD was attributed to interpretation error in 21% of the cases (7 of 33), where 36% of the cases (12 of 33) were discrepant for other nonidentifiable causes.

Finally, the clinical outcomes of interest were evaluated in the 33 major discrepant cases. Direct impact on surgical management was seen only in 3 cases, 0.5% (3 of 618) of the overall cohort. Performance of an unnecessary procedure was identified in one overdiagnosed patient (Figure 3, A and B). Two patients were underdiagnosed on the FSD and return for a second surgery was indicated. The first case was a patient who declined the second surgery and received chemotherapy instead (Figure 3, C and D). In the second case, the cyst ruptured intraoperatively; hence, the patient was upstaged to pT1c and no second surgery was performed (Figure 3, E and F). However, no patient had 30-day postoperative mortality.

**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

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 literature9 (8%). We propose that our higher rates reflect a more stringent application of

---

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

| Discrepancy                      | All Cases (N = 618) | Epithelial Tumors (n = 366) | Sex Cord Tumors (n = 61) | Germ Cell Tumors (n = 58) | Metastasis (n = 36) | Other Benign Nonneoplastic Entities (n = 97) |
|----------------------------------|---------------------|------------------------------|--------------------------|---------------------------|-------------------|-----------------------------------------|
| No discrepancy (82.4%, n = 509) | 292                 | 45                          | 54                       | 28                        | 90                |
| Minor discrepancies (12.3%, n = 76) | 51                  | 12                          | 2                        | 4                         | 7                 |
| Major discrepancies (5.3%, n = 33) | 23                  | 4                           | 2                        | 4                         | 0                 |
| Major discrepancies overdiagnosed (n = 7) | 5                   | 2                           | 0                        | 0                         | 0                 |
| Major discrepancies undiagnosed (n = 26) | 18              | 2                           | 2                        | 4                         | 0                 |

---

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

| Discrepancy | General Surgical Pathologists, No. (%) | Gynecologic Pathologists, No. (%) | $\chi^2$ test |
|-------------|---------------------------------------|----------------------------------|---------------|
| None (n = 509) | 344 (83) | 165 (81) | .65 |
| Minor (n = 76) | 50 (12) | 26 (13) |   |
| Major (n = 33) | 21 (5) | 12 (6) |   |

---

**Figure 1.** The $\kappa$ agreement between each pathologist's frozen section diagnosis and the gold standard final diagnosis (interater). $\kappa$ statistics were generated by Cohen $\kappa$ and tested against the null hypothesis of the true $\kappa$ of 0. Gynecologic pathologists (GYN) are pathologists A, B, C and D, and general surgical pathologists (non-GYN) are pathologists E, F and G.
criteria to define discrepancies and the inclusion of all major- and minor-discrepancy cases (the discrepancy rate was 5.3% for major and 12.3% for minor). We believe that a lower threshold allows pathologists to identify the factors that lead to major and minor discrepancies that may ultimately contribute to more significant events. This will also shed light on potential areas for improvement that will ultimately reduce the potential for discrepancy overall.

Indeed, supervising training pathology residents and pathology assistants for adequate examining and sampling of ovarian lesions; however, some invasion might be appreciated only on permanent sections, so changing the diagnosis line to definitive paraffin section diagnosis.

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 “at least borderline tumor” and discussing the case with surgeons intraoperatively may give a better insight about clinical implication. Continuous quality assurance should be conducted in regard to discrepancies between FSD and final paraffin section diagnosis.

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 “at least borderline tumor” and discussing the case with surgeons intraoperatively may give a better insight about clinical implication. Continuous quality assurance should be conducted in regard to discrepancies between FSD and final paraffin section diagnosis.

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 $\kappa$ 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.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 studies22 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 “at least borderline tumor” and discussing the case with surgeons intraoperatively may give a better insight about clinical implication. Continuous quality assurance should be conducted in regard to discrepancies between FSD and final paraffin section diagnosis.

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

| Cause of Discrepancy | Epithelial Tumors (n = 23) | Sex Cord Tumors (n = 4) | Germ Cell Tumors (n = 2) | Metastasis (n = 4) | Other Benign Nonneoplastic Entities (n = 0) |
|----------------------|---------------------------|------------------------|--------------------------|------------------|------------------------------------------|
| Interpretation error (n = 7) | 5 | 2 | 0 | 0 | 0 |
| Sampling error (n = 14) | 8 | 1 | 2 | 3 | 0 |
| Other nonidentifiable error (n = 12) | 10 | 1 | 0 | 1 | 0 |
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).

Figure 3.
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
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68(6):394–424.
2. American Cancer Society. Cancer statistics center. http://cancerstatisticscenter.cancer.org. Accessed October 1, 2020.
3. Kurman RJ, Carcangiu ML, Herrington CS, Young RH. WHO Classification of Tumours of Female Reproductive Organs. 4th ed. Lyon, France: International Agency for Research on Cancer; 2014. WHO Classification of Tumours; vol 6.
4. Prat J; FIGO Committee on Gynecologic Oncology. Staging classification for cancer of the ovary, fallopian tube, and peritoneum. Int J Gynaecol Obstet. 2014;124(1):1–5.
5. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas. 1960;20:37–46.
6. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol Bull. 1971;76(5):378–382.
7. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–174.
8. Berek JS, Cram C, Friedlander M. Cancer of the ovary, fallopian tube, and peritoneum. Int J Gynaecol Obstet. 2015;131(suppl 2):S111–S122.
9. Sukumaran R, Somanathan T, Mathews A, et al. Role of frozen section in intraoperative assessment of ovarian masses: a tertiary oncology center experience. Indian J Surg Oncol. 2014;5(2):99–103.
10. Winther C, Graem N. Accuracy of frozen section diagnosis: a retrospective analysis of 4785 cases. AFMS. 2011;119(4–5):259–262.
11. Geomini P, Bremer G, Kruitwagen R, Mol BW. Diagnostic accuracy of frozen section diagnosis of the adnexal mass: a metaanalysis. Gynecol Oncol. 2003;96(1):1–9.
12. Suprasert P, Khunamornpong S, Phueng A, Settakorn J, Siriaunkglu S. Accuracy of intra-operative frozen sections in the diagnosis of ovarian masses. Asian Pac J Cancer Prev. 2008;9(4):737–740.
13. Subbion A, Devi UK, Bafna UD. Accuracy rate of frozen section studies in ovarian cancers: a regional cancer institute experience. Indian J Cancer. 2013;50(4):302–305.
14. Ponguvuverayakul T, Khunamornpong S, Settakorn J, Sukpan K, Suprasert P, Siriaunkglu S. Accuracy of frozen-section diagnosis of ovarian mucinous tumors. Int J Gynecol Cancer. 2012;22(3):400–406.
15. Hashmi AA, Naz S, Edhi MM, et al. Accuracy of intraoperative frozen section for the evaluation of ovarian neoplasms: an institutional experience. World J Surg Oncol. 2016;31:14:91.
16. Houack N, Nikrui N, Duska L, et al. Borderline tumors of the ovary: correlation of frozen and permanent histopathologic diagnosis. Obstet Gynecol. 2000;95(6, pt 1):839–843.
17. Medeiros LR, Rosa DD, Edelweiss MI, et al. Accuracy of frozen-section analysis in the diagnosis of ovarian tumors: a systematic quantitative review. Int J Gynecol Cancer. 2005;15(2):192–202.
18. Stewart CJ, Brennan BA, Hammond IG, Leung YC, McCartney AJ. Intraoperative assessment of ovarian tumors: a 5-year review with assessment of discrepant diagnostic cases. Int J Gynecol Pathol. 2006;25(3):216–222.
19. Menzin AW, Rubin SC, Noundoff JS, LiVolisi VA. The accuracy of a frozen section diagnosis of borderline ovarian malignancy. Gynecol Oncol. 1995;59(2):183–185.
20. Ratnavelu NDG, Brown AP, Mallett S, et al. Intraoperative frozen section analysis for the diagnosis of early stage ovarian cancer in suspicious pelvic masses. Cochrane Database Syst Rev. 2016;3(3):CD010360.
21. Kim K, Chung HH, Kim JW, Park NH, Song YS, Kang SB. Clinical impact of under-diagnosis by frozen section examination is minimal in borderline ovarian tumors. Eur J Surg Oncol. 2009;35(9):969–973.
22. Camattea S, Moricea P, Thoury A, et al. Impact of surgical staging in patients with microscopic “stage I” ovarian borderline tumors: analysis of a continuous series of 101 cases. Eur J Cancer. 2004;40(12):1842–1849.