



Published in final edited form as:

Gynecol Oncol. 2018 September ; 150(3): 521–526. doi:10.1016/j.ygyno.2018.07.003.

Inter-pathologist and pathology report agreement for ovarian tumor characteristics in the Nurses' Health Studies

Mollie E. Barnard^a, Alexander Pyden^b, Megan S. Rice^{c,d}, Miguel Linares^c, Shelley S. Tworoger^{a,e}, Brooke E. Howitt^{f,g}, Emily E. Meserve^{g,h}, and Jonathan L. Hecht^b

^aDepartment of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, MA 02115

^bDepartment of Pathology, Beth Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02215

^cClinical and Translational Epidemiology Unit, Department of Medicine, Massachusetts General Hospital, Bartlett Ext., 9th Floor 55 Fruit St. Boston, MA 02114

^dChanning Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115

^eDepartment of Cancer Epidemiology, Moffitt Cancer Center, 12902 USF Magnolia Drive, Tampa, FL 33612

^fDepartment of Pathology, Stanford University School of Medicine, 300 Pasteur Drive, Lane 235, Stanford, CA 94305-5324

^gDepartment of Pathology, Brigham and Women's Hospital, 75 Francis St, Boston, MA 02115

^hDepartment of Anatomic & Clinical Pathology, Maine Medical Center, Tufts University School of Medicine, 22 Bramhall Street, Portland, Maine 04102-3175

Abstract

Background—Grade and histotype of ovarian carcinomas are often used as surrogates of molecular subtypes. We examined factors affecting pathologists' reproducibility in two prospective studies.

Contact information (corresponding author): Jonathan L. Hecht, MD PhD, Department of Pathology, Beth-Israel Deaconess Medical Center, 330 Brookline Ave, Boston, MA 02215, 617-667-3817 (voice), 617-667-7120 (fax), jlhecht@BIDMC.harvard.edu.

DISCLOSURE/CONFLICT OF INTEREST:

All authors have approved of the submitted version of the manuscript; the manuscript is not under consideration elsewhere; none of the authors have any business relationships that might lead to a conflict of interest.

AUTHOR CONTRIBUTION:

All authors listed below confirm that they have contributed to the intellectual content of this submitted manuscript and have met the following three requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the manuscript for intellectual content; and (c) final approval of the article. Authors: Mollie E. Barnard, Alexander Pyden, Megan S. Rice, Miguel Linares, Shelley S. Tworoger, Brooke E. Howitt, Emily E. Meserve, Jonathan L. Hecht.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Methods—Two pathologists independently reviewed slides from 459 incident ovarian cancers in the Nurses' Health Study (NHS) and NHSII. We described agreement on tumor characteristics using percent agreement and Cohen's standard kappa (κ) coefficients. We used logistic regression, with disagreement as the outcome, to evaluate the contribution of case and tumor characteristics to agreement.

Results—Inter-rater agreement was 95% ($\kappa=0.81$) for carcinoma versus borderline, 89% ($\kappa=0.58$) for grade and 85% ($\kappa=0.71$) for histotype. Inter-rater grading disagreement was higher for non-serous histotypes (OR=4.66, 95% CI 2.09–10.36) and lower for cancers with bizarre atypia (OR=0.13, 95% CI 0.04–0.38). Agreement with original pathology reports was 94% ($\kappa=0.73$) for carcinoma versus borderline, 78% ($\kappa=0.60$) for histotype, and 79% ($\kappa=0.24$) for grade. Grading disagreement was significantly lower for tumors with 'solid, pseudoendometrioid or transitional' (SET) architecture (OR=0.08, 95% CI 0.01 – 0.84). Date of original diagnosis, hospital type, number of slides available for review, tumor stage, and slide quality were not related to agreement.

Conclusion—Overall, inter-rater agreement for tumor type and grade for archival tissue specimens was good. Agreement between the consensus review and original pathology reports was lower. Factors contributing to grading disagreement included non-serous histotype, absence of bizarre atypia, and absence of SET architecture.

Keywords

Ovarian cancer; grade; histotype; agreement

INTRODUCTION

Ovarian cancer is a heterogeneous disease. [1] Although sub-classification using molecular diagnostics is an emerging trend, [2–4] existing data repositories used in clinical and epidemiologic research largely rely on the morphologic classifications of histotype (e.g., serous, endometrioid, clear cell, mucinous) and grade reported in pathology reports or tumor registries. [5, 6] Morphologic classification of ovarian cancer, however, is subjective.

Reproducibility studies evaluating inter-observer agreement among expert pathologists have shown excellent agreement for histotype ($\kappa = 0.77 - 0.97$) [5, 7], and Köbel *et al.* reported that adding a panel of immunohistochemical stains can further improve inter-observer agreement. [8] In contrast, agreement for grade is only fair. For example, reported agreement on International Federation of Gynecology and Obstetrics (FIGO) grade ranges from $\kappa = 0.25 - 0.26$. [5, 6].

Inter-observer agreement and agreement with the original pathology report in the setting of an unselected population or central review during clinical trials or epidemiologic studies have not been extensively reported. Kommos *et al.* included a central pathology review during a phase 3 drug trial and cited 65% agreement on histotype with the greatest discrepancy for serous misclassified as endometrioid type. Agreement was not significantly better for cases diagnosed at academic versus private hospitals. [9] Similar results were observed by Lopez-Guerrero *et al.* for a central pathology review of early-stage ovarian

carcinoma in the Spanish Group for Ovarian Cancer Research (GEICO), with concordance of 76% ($\kappa = 0.50$; $p < 0.0001$). [10] In contrast, Young *et al.* and Ozols *et al.* reported little to no misclassification of histotype between central pathology review and the original report. [11, 12] In general, borderline tumors were rarely misclassified as carcinoma versus borderline. [9, 11] In the only study to assess grade, an evaluation in the Surveillance Epidemiology and End Results (SEER) Residual Tissue Repository, Matsuno *et al.* found only fair grading agreement between expert review and original pathologist diagnosis (49% agreement, $\kappa = 0.25$). [6]

In addition, relatively few studies have evaluated predictors of grading agreement. Matsuno *et al.* commented that grading agreement between expert pathologist review and SEER data (based on the original diagnosis) was better for tumors identified as carcinoma (versus borderline) by the pathology review (agreement in 62% of cases and $\kappa = 0.32$), but was not improved when restricted to cases where the reviewer agreed with SEER on tumor histotype. [6] There have been no reports on the effects of variables related to data collection, such as time since original diagnosis, hospital type, slide quality and number of slides available. Further, while high-grade tumors tend to have greater mitotic activity and multinucleated cells, [13] the effects of tumor characteristics on grading agreement, such as extensive necrosis, bizarre atypia or with 'solid, pseudoendometrioid or transitional' (SET) architecture, have not been analyzed. This may be especially relevant for studies relying on pathology report data because cancers with these features were formerly classified as high-grade endometrioid carcinoma, but are now being diagnosed as serous. [14]

The Nurses' Health Studies (NHS/NHSII) are large, prospective cohort studies that request tumor tissue and pathology reports from reported cases of epithelial ovarian cancer during up to 40 years of follow-up. The NHS/NHSII have obtained tumors slides or blocks from over 450 ovarian cancer cases, allowing for assessment of agreement in the evaluation of tumor characteristics. Using data from NHS/NHSII, we examined agreement on ovarian carcinoma versus borderline, histotype and grade [15] independently assigned by two expert gynecologic pathologists with access to the same archival tissue slides. We also quantified tumor histotype and grading agreement between the consensus of expert gynecologic pathologists and information abstracted from the original pathology reports.

To better understand potential predictors of poor inter-rater agreement, we examined factors that may affect grading agreement in an epidemiologic setting, including variables related to data collection (i.e., time since original diagnosis, hospital type, slide quality, number of slides available), and tumor characteristics (i.e., necrosis, bizarre atypia, SET architecture, tumor stage). Further, we anticipated greater inter-pathologist disagreement among cases with fewer or older slides. We also expected that cases with SET morphology or necrosis would be graded higher on the grading scale by the consensus review of pathologists relative to data abstracted from pathology reports. [16]

MATERIALS AND METHODS

Study Population

As of August 2010, the latest cancer diagnosis date for cases in this study, the NHS/NHSII included 1,311 confirmed cases of incident epithelial ovarian cancer with pathology report data. The NHS/NHSII requested tumor tissue from all of these cases, and received specimens from 459 of them. Common reasons that tumor tissue was not available included, death of the patient, destruction of the tissue block, or inability of the hospital to send a tissue sample. [17] Of the 459 cases with slides available for review by two expert gynecologic pathologists, 41 did not have record-based information on histopathologic features, due to confirmation via linkage to the relevant tumor registry or death certificates.

Slide Review

Between October 2014 and May 2016, two expert gynecologic pathologists (JH and EM) reviewed slides from all 459 cases with tumor slides. Both were blinded to morphologic assessment on original pathology reports. The reviewing pathologists assigned values for carcinoma versus borderline, histotype, grade, and grading-related features: gland formation, nuclear atypia and mitotic rate. They also recorded the number of slides available for review, commented on slide quality, and assessed tumor architectural features such as SET (percent solid, pseudoendometrioid or transitional architecture), geographic necrosis, and bizarre atypia. Cases for which the two reviewers disagreed on carcinoma versus borderline, histotype, or grade, were adjudicated by a third gynecologic pathologist (BH). Extreme disagreements on tumor architecture, as defined by >30% difference in percent SET, were also adjudicated by the third pathologist. The agreed upon or arbitrated values were treated as the consensus values of the three reviewing pathologists. For the 418 cases from whom we also obtained pathology reports, we abstracted values for date of original diagnosis, carcinoma versus borderline, histotype, grade, and stage, when reported. The grading system was generally not specified in the original reports, but was likely the 3-tiered FIGO system given the prevalence of its usage in the United States.

The pathologist review of archival tumor specimens included assignment of histopathologic features. Tumors were categorized as either carcinoma or borderline. Tumor histotypes were described as serous, endometrioid, clear cell, mucinous, or other. Grade was reported for all epithelial ovarian carcinomas using a three-tier grading system in which serous carcinomas were assigned high-grade or low-grade (grades 1 or 3), clear cell carcinomas were assigned grade 3, and endometrioid and mucinous carcinoma were graded based on architecture according to the FIGO system (Grade 1 showed < 5% of solid tumor growth; Grade 2 with 5%–50%, and Grade 3 with >50%). [18] Reviewing pathologists were blinded to morphologic assessment on original pathology reports when reviewing the slides.

Our grading scheme is an modification of the WHO 2014 grading system [19] to allow for comparison to the original pathology reports, the majority of which reported on a the 3-tiered grading scale. In WHO 2014, low-grade and high-grade serous are distinct diseases, so a numerical grade is not assigned; clear cell carcinomas are all considered high-grade for which a numerical grade is not assigned, and endometrioid and mucinous carcinomas are

graded on a 3-tiered system. [19] In contrast, a numerical grade was assigned in the majority of original reports regardless of histotype. The grading system in the original reports was generally not specified, but was likely the 3-tiered FIGO system (including grade 2 serous carcinoma) in older cases, or a modification similar to ours in recent cases. Assignment of a numerical grade was often included to accommodate the data format of existing tumor registries such as NHS or SEER [20], older cancer reporting protocols, [21] or clinical practice guidelines. [22]

SET architecture was assessed as a binary variable for which tumors were defined as SET if greater than 50% of the tissue had solid, pseudoendometrioid, or transitional architecture. [14] Bizarre atypia and geographic necrosis were also evaluated as binary with categories of either present or absent. Bizarre atypia and geographic necrosis were not adjudicated when the two primary reviewers disagreed, so the consensus review of these variables describes each feature as present if observed by at least one pathologist. Comparisons between the consensus pathology opinion and the original pathology report were restricted to the following variables: carcinoma versus borderline, histotype and grade as these were the only factors available in the majority of pathology reports. In the majority of cases, tissue blocks were not available for immunohistochemistry so histotype was assigned based on morphology alone.

Statistical Analysis

We described agreement on ovarian cancer tumor characteristics using percent agreement and Cohen's standard kappa (κ) coefficients, [23] only including cases for which both pathologists or the consensus review and pathology report had known values. Consistent with prior literature on grading agreement, κ coefficients were interpreted as poor (0.01–0.20), fair (0.21–0.40), moderate (0.41–0.60), substantial (0.61–0.80) and near perfect (0.81–1.00). [5, 6, 23] Analyses of agreement on grade were restricted to carcinoma, and agreement on SET was restricted to serous tumors, per definition. [14, 16]

We used logistic regression, with disagreement as the outcome of interest, to assess disagreement for each feature. We evaluated the contribution of histologic features and slide characteristics to inter-reviewer disagreement and to disagreement between the consensus of the pathologists and the original pathology report. The exposures of interest included year of diagnosis, hospital type, the number of slides available for review, slide quality as assessed by the reviewing pathologist, histotype, SET architecture, geographic necrosis, and bizarre atypia. For agreement between expert pathologists and the pathology report, we additionally considered stage of disease as abstracted from the pathology report. We report odds ratios and 95% confidence intervals to quantify the odds of disagreement for a particular feature comparing those with and without each exposure listed above. Odds ratios are presented from univariate analyses, from “adjusted” models that adjusted for predictors with $p < 0.10$ in the univariate analyses, and from “fully adjusted” models that were mutually adjusted for all other exposure variables, except SET architecture, which was evaluated in a model restricted to serous tumors.

The study protocols for NHS/NHSII were approved by the institutional review board of Brigham and Women's Hospital. All analyses were conducted using SAS 9.4 (Cary, NC,

USA). Statistical significance was assessed assuming a two-sided test with an alpha level of 0.05.

RESULTS

NHS/NHSII ovarian cancer cases with tumor tissue were representative of all ovarian cancer cases within the cohorts with respect to carcinoma versus borderline, histotype, grade and stage (Table 1). The mean year of diagnosis was 1997 for all cases and 1998 for cases with both pathology reports and tumor slides. The majority of cases were carcinoma, serous, stage 3, and grade 3.

Inter-pathologist agreement was excellent (95%; $\kappa=0.81$) for carcinoma versus borderline, good for grade (89%; $\kappa=0.58$) and histotype (85%; $\kappa=0.71$), and moderate for SET architecture (81%; $\kappa=0.52$). Agreement between the pathologist consensus review and data abstracted from pathology reports was excellent for carcinoma versus borderline (94%; $\kappa=0.73$), and moderate for histotype (78%; $\kappa=0.60$) and grade (79%; $\kappa=0.24$; Table 2).

We modeled the odds of disagreement between pathologists for the grade to evaluate what factors impacted inter-reviewer agreement. The reviewing pathologists were significantly more likely to disagree on grade for non-serous compared to serous ovarian cancer (OR=4.66, 95% CI 2.09–10.36), and less likely to disagree when bizarre atypia was present (OR=0.13, 95% CI 0.04–0.38). None of the other variables were significantly associated with inter-reviewer disagreement on grade (Table 3).

We also evaluated the odds of disagreement on grade between the consensus of the pathologists and data abstracted from pathology reports. Restricting to cases with values from both contemporary pathology review and pathology reports, we observed that the consensus opinion was significantly less likely to disagree with the original pathology reports on grade for cases in which SET architecture was present (OR=0.08, 95% CI 0.01 – 0.84). Disagreement on grade was suggestively higher when histotype was non-serous compared to serous (OR=2.49, 95% CI 0.87 – 7.09). In total, there were 16 cases classified as endometrioid grade 1 or 2 in the pathology reports. Of these, 4 were classified as serous by contemporary slide review. There were also 15 cases classified as endometrioid grade 3 in the pathology reports, and 13 of these were classified as serous by contemporary slide review. Other variables we evaluated that were not associated with disagreement included, date of diagnosis, hospital type, number of slides, slide quality, pathologic stage, and geographic necrosis (Table 4).

DISCUSSION

The goals of this study were two-fold: (1) to measure reproducibility of grade assignment and histotype for ovarian cancer in the setting of an epidemiologic study and (2) to identify factors that affect grading disagreement. We observed excellent reproducibility among expert pathologists in assignment of carcinoma versus borderline, good reproducibility for histotype, grade and SET architecture. When comparing contemporary slide review by expert pathologists to original pathology reports, percent agreement was excellent for carcinoma versus borderline, and moderate for histotype and grade.

We observed greater inter-pathologist agreement for grade than previously reported. This may be explained by differences in tumor sub-type distribution compared to populations previously studied. For example, the SEER database study, which found only 49% agreement with $\kappa = 0.25$, included a mix of non-Hispanic white women and Asian women from Hawaii, [6] while ours included mostly non-Hispanic, white women living in the mainland United States. In addition, earlier reproducibility studies used selected cases, possibly representing a difficult subset. For example, 50 cases from a clinical case series of 575 were reviewed by three pathologists to determine inter-observer variation in histotype and grade assignment. Inter-observer variation for histotype was very good ($\kappa = 0.77$) with minimal reproducibility in assignment of FIGO grade ($\kappa = 0.27$). [5] Tumor cell type can be reproducibly diagnosed and is of independent prognostic significance in patients with maximally debulked ovarian carcinoma. The difference may be due to a bias in selecting difficult cases. Another explanation for the higher inter-pathologist agreement observed in our study is that the training of our pathologists was more consistent; all three of our expert pathologists (JH, EM and BH) shared a hospital affiliation and on-going training opportunities.

Agreement for grade differed by histotype and other tumor morphologic features. For example, grade (inter-pathologist grade and consensus vs. pathology report abstraction) had worse agreement among non-serous versus serous cases. This likely reflects a reduction of the number of categories for cancers of serous histology (high-grade=3 versus low-grade=1) in the grading scheme, and serous was the most common histotype.

Agreement between contemporary pathologist review and original pathology reports was good, suggesting that pathology report abstraction is a good approach for epidemiologic studies. Grade discordance was lower both between pathologists and with original pathology reports for serous tumors with SET architecture. This result is unexpected since we had assumed that a proportion of serous cancers with SET architecture would have been misclassified as low-grade endometrioid based on the presence of glandular architecture, particularly in the original pathology reports. However, tumors with SET architecture often retain some areas with classic serous histology, and show geographic necrosis and a high mitotic rate as clues to their histotype. [16] Interestingly, our study suggested that case and slide characteristics (i.e., year of diagnosis, hospital type, slide quality, and number of slides) were not strong predictors of grading agreement, supporting similar epidemiologic evaluation of cases spanning multiple hospital types and a broad range in time. However, the majority of carcinoma cases (78.3%) had more than one tumor slide available for review and only 12.7% were flagged as poor-quality thus reducing power to assess these factors.

To our knowledge, this is one of only a few large studies on grading agreement among unselected cases from an epidemiologic database. [5, 6] Further, this is the first study to assess factors that may affect grading agreement. Key strengths of this study include the large sample size, evaluation of all tumors by at least two pathologists, and adjudication review of tumors by a third pathologist for which the two primary pathologists disagreed on key tumor characteristics. Another strength of this study is that all tumor slides collected by the Nurses' Health Studies were made available for review, rather than a random subset of slides as is done in some clinical studies. [6]

Limitations of this study affected both study design and analysis. First, many diagnoses recorded in the original pathology reports were confirmed by expert consultation, but we do not have information as to which cases. This may have elevated measured agreement between the contemporary slide review by expert pathologists, and tumor characteristics recorded on pathology reports. However, this is often done in clinical practice and should therefore reflect ovarian cancer cases recruited into most population-based studies. Further, since grading agreement was better than anticipated, we had less power to look at predictors of disagreement. As expected, in our study, percent agreement values were noticeably greater than kappa values. While kappa is generally treated as a more valid measure of agreement because it accounts for random chance, the imbalance among categories may lead to under-estimation of agreement when kappa is used. [24] A similar discrepancy between percent agreement and kappa was seen in prior studies. [6] Finally, immunohistochemical stains, known to aid in correctly assigning histotype, were not available in the study cases. [8]

It is increasingly evident that ovarian cancer etiology, risk, and prognosis differ by tumor subtype. [25] While subtype classification using molecular diagnostics is an emerging trend, [2–4] existing data repositories used in clinical and epidemiologic research largely rely on the morphologic classifications of histotype (e.g., serous, endometrioid, clear cell, mucinous) and grade reported in pathology reports or tumor registries (which also abstract this information from original pathology reports). [5, 6] Consistent with the few other available studies, the results of this study suggest reasonable, yet imperfect, agreement on important tumor characteristics (i.e. carcinoma versus borderline, histotype, and grade). This suggests that epidemiologic analyses examining associations by different tumor characteristics can provide valid, if somewhat attenuated results, due to modest misclassification. It is worth exploring if our results are generalizable to other populations and, if so, how they could be leveraged to estimate or eliminate the potential effect of misclassification when estimating subtype-specific associations.

Acknowledgments

FINANCIAL SUPPORT:

M. Barnard was supported by the National Cancer Institute of the National Institutes of Health under Award Number T32 CA09001 and Award Number F99 CA212222.

The Nurses' Health Study is supported by grants UM1 CA186107 and P01 CA87969, while the Nurses' Health Study II is supported by grant UM1 CA17672 by the National Institute of Health.

We would like to thank the study participants and staff of the NHS and NHSII and the following state cancer registries: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The authors assume full responsibility for analyses and interpretation of these data.

References

1. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. *Nature*. 2011; 474:609–15. [PubMed: 21720365]
2. Kobel M, Kalloger SE, Lee S, Duggan MA, Kelemen LE, Prentice L, et al. Biomarker-based ovarian carcinoma typing: a histologic investigation in the ovarian tumor tissue analysis consortium. *Cancer Epidemiol Biomarkers Prev*. 2013; 22:1677–86. [PubMed: 23880734]

3. Leong HS, Galletta L, Etemadmoghadam D, George J, Kobel M, et al. Australian Ovarian Cancer S. Efficient molecular subtype classification of high-grade serous ovarian cancer. *J Pathol*. 2015; 236:272–7. [PubMed: 25810134]
4. Winterhoff B, Hamidi H, Wang C, Kalli KR, Fridley BL, Dering J, et al. Molecular classification of high grade endometrioid and clear cell ovarian cancer using TCGA gene expression signatures. *Gynecol Oncol*. 2016; 141:95–100. [PubMed: 27016234]
5. Gilks CB, Ionescu DN, Kalloger SE, Kobel M, Irving J, Clarke B, et al. Tumor cell type can be reproducibly diagnosed and is of independent prognostic significance in patients with maximally debulked ovarian carcinoma. *Hum Pathol*. 2008; 39:1239–51. [PubMed: 18602670]
6. Matsuno RK, Sherman ME, Visvanathan K, Goodman MT, Hernandez BY, Lynch CF, et al. Agreement for tumor grade of ovarian carcinoma: analysis of archival tissues from the surveillance, epidemiology, and end results residual tissue repository. *Cancer Causes Control*. 2013; 24:749–57. [PubMed: 23378140]
7. Kobel M, Kalloger SE, Baker PM, Ewanowich CA, Arseneau J, Zherebitskiy V, et al. Diagnosis of ovarian carcinoma cell type is highly reproducible: a transcanadian study. *Am J Surg Pathol*. 2010; 34:984–93. [PubMed: 20505499]
8. Kobel M, Bak J, Bertelsen BI, Carpen O, Grove A, Hansen ES, et al. Ovarian carcinoma histotype determination is highly reproducible, and is improved through the use of immunohistochemistry. *Histopathology*. 2014; 64:1004–13. [PubMed: 24329781]
9. Kommoss S, Pfisterer J, Reuss A, Diebold J, Hauptmann S, Schmidt C, et al. Specialized pathology review in patients with ovarian cancer: results from a prospective study. *Int J Gynecol Cancer*. 2013; 23:1376–82. [PubMed: 24257551]
10. Lopez-Guerrero JA, Pecharroman AG, Palacios J, Romero I, Lana EMC, Hardisson D, et al. Central pathology review of early-stage ovarian carcinoma: Description and correlation with follow-up—A study by the Spanish Group for Ovarian Cancer Research (GEICO). *Journal of Clinical Oncology*. 2014; 32:5583.
11. Ozols RF, Bundy BN, Greer BE, Fowler JM, Clarke-Pearson D, Burger RA, et al. Phase III trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage III ovarian cancer: a Gynecologic Oncology Group study. *J Clin Oncol*. 2003; 21:3194–200. [PubMed: 12860964]
12. Young RC, Walton LA, Ellenberg SS, Homesley HD, Wilbanks GD, Decker DG, et al. Adjuvant therapy in stage I and stage II epithelial ovarian cancer. Results of two prospective randomized trials. *N Engl J Med*. 1990; 322:1021–7. [PubMed: 2181310]
13. Malpica A, Deavers MT, Lu K, Bodurka DC, Atkinson EN, Gershenson DM, et al. Grading ovarian serous carcinoma using a two-tier system. *Am J Surg Pathol*. 2004; 28:496–504. [PubMed: 15087669]
14. Howitt BE, Hanamornroongruang S, Lin DI, Conner JE, Schulte S, Horowitz N, et al. Evidence for a dualistic model of high-grade serous carcinoma: BRCA mutation status, histology, and tubal intraepithelial carcinoma. *Am J Surg Pathol*. 2015; 39:287–93. [PubMed: 25581732]
15. Shimizu Y, Kamoi S, Amada S, Akiyama F, Silverberg SG. Toward the development of a universal grading system for ovarian epithelial carcinoma: testing of a proposed system in a series of 461 patients with uniform treatment and follow-up. *Cancer*. 1998; 82:893–901. [PubMed: 9486579]
16. Soslow RA, Han G, Park KJ, Garg K, Olvera N, Spriggs DR, et al. Morphologic patterns associated with BRCA1 and BRCA2 genotype in ovarian carcinoma. *Mod Pathol*. 2012; 25:625–36. [PubMed: 22193042]
17. Hecht JL, Kotsopoulos J, Gates MA, Hankinson SE, Tworoger SS. Validation of tissue microarray technology in ovarian cancer: results from the Nurses' Health Study. *Cancer Epidemiol Biomarkers Prev*. 2008; 17:3043–50. [PubMed: 18990746]
18. Kurman R, Carcangiu M, Herrington C, Young R. WHO classification of tumours of the female reproductive organs. Lyon: IARC press; 2014. [4]
19. Kurman RJ. International Agency for Research on Cancer., World Health Organization. WHO classification of tumours of female reproductive organs. 4. Lyon: International Agency for Research on Cancer; 2014.
20. p. SEER database.

21. Fleming ID. American Joint Committee on Cancer., American Cancer Society., American College of Surgeons. AJCC cancer staging manual. 5. Philadelphia: Lippincott-Raven; 1997.
22. Morgan RJ Jr, Armstrong DK, Alvarez RD, Bakkum-Gamez JN, Behbakht K, Chen LM, et al. Ovarian Cancer, Version 1.2016, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2016; 14:1134–63. [PubMed: 27587625]
23. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb). 2012; 22:276–82. [PubMed: 23092060]
24. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. Journal of clinical epidemiology. 1990; 43:543–9. [PubMed: 2348207]
25. National Academies of Sciences Engineering and Medicine. Ovarian Cancers: Evolving Paradigms in Research and Care. Washington, DC: The National Academies Press; 2016.

Highlights

Ovarian cancer registries often rely on morphologic classification of histotype/
grade from local pathology reports

Central review pathologists' reproducibility and agreement with original pathology
reports is good

Agreement with original reports varied by histologic features, but not by slide
quality

Table 1

Characteristics of ovarian cancer cases in the Nurses' Health Studies and the subset of cases with tumor slides *

| | All cases [†] (n=1311) | Cases with tumor slides [‡] (n=418) [§] |
|---------------------------------|---------------------------------|---|
| Mean year of diagnosis (SD) | 1997 (9) | 1998 (7) |
| Mean years since diagnosis (SD) | 18.5 (8.6) | 18.2 (7.1) |
| Invasion | | |
| Invasive (%) | 87.5 | 87.1 |
| Borderline (%) | 11.1 | 12.7 |
| Missing (%) | 1.4 | 0.2 |
| Histotype | | |
| Serous (%) | 57.8 | 55.3 |
| Mucinous (%) | 8.0 | 8.9 |
| Endometrioid (%) | 13.2 | 15.8 |
| Clear cell (%) | 5.3 | 6.5 |
| Other (%) | 11.8 | 13.2 |
| Missing (%) | 3.9 | 0.5 |
| Stage | | |
| Stage 1 (%) | 26.1 | 31.6 |
| Stage 2 (%) | 7.0 | 6.5 |
| Stage 3 (%) | 53.9 | 56.2 |
| Stage 4 (%) | 8.2 | 4.8 |
| Missing (%) | 4.8 | 1.0 |
| Grade | | |
| 1 (%) | 7.2 | 7.1 |
| 2 (%) | 15.3 | 16.5 |
| 3 (%) | 51.8 | 53.3 |
| Missing (%) | 25.8 | 23.1 |

* Tumor characteristics as recorded on pathology reports.

[†] All cases with tumor information abstracted from pathology reports.[‡] Cases with both tumor slides and tumor information abstracted from pathology reports.[§] A total of 459 NHS cases had tumor slides and were reviewed by two pathologists, but only 418/459 also had tumor information abstracted from pathology reports.

Table 2

Agreement in evaluation of ovarian cancer tumor characteristics

| | Number of cases* | Percent agreement | Cohen's Kappa (κ) |
|---|------------------|-------------------|----------------------------|
| Inter-rater | | | |
| Invasion | 459 | 0.948 | 0.81 |
| Histotype | 459 | 0.852 | 0.71 |
| WHO grade [†] | 369 | 0.894 | 0.58 |
| SET [‡] | 216 | 0.810 | 0.52 |
| Pathologist Consensus vs. Pathology Report | | | |
| Invasion | 417 | 0.935 | 0.73 |
| Histotype | 364 | 0.783 | 0.60 |
| Grade [†] | 236 | 0.788 | 0.24 |

* Number of cases with tumor characteristics reviewed by two pathologists (inter-rater) or with tumor characteristics reviewed by two pathologists and tumor information abstracted from pathology reports (pathologist consensus vs. pathology report).

[†] Among invasive tumors

[‡] Among serous tumors

Table 3

Predictors of inter-rater WHO grading discordance for invasive cases

| | Discordant cases/total cases | Unadjusted odds ratio (95% CI) | Adjusted odds ratio* (95% CI) | Fully adjusted odds ratio[†] (95% CI) |
|---|-------------------------------------|---|--|---|
| Date of diagnosis (5- year increments) | 39/369 | 0.93 (0.75, 1.16) | 0.99 (0.79, 1.25) | 0.99 (0.79, 1.26) |
| Number of slides | 39/369 | 0.99 (0.95, 1.02) | 0.99 (0.95, 1.02) | 0.99 (0.95, 1.02) |
| Slide Quality | | | | |
| Good | 32/322 | (ref) | (ref) | (ref) |
| Poor | 7/47 | 1.59 (0.66, 3.83) | 1.55 (0.59, 4.10) | 1.51 (0.55, 4.11) |
| Histotype | | | | |
| Serous | 14/248 | (ref) | (ref) | (ref) |
| Non-serous | 20/72 | 6.43 (3.05, 13.56) | 4.26 (1.96, 9.23) | 4.66 (2.09, 10.36) |
| Missing | 5/49 | 1.90 (0.65, 5.54) | 2.00 (0.66, 6.07) | 1.94 (0.63, 5.93) |
| Geographic necrosis | | | | |
| Absent | 21/160 | (ref) | (ref) | (ref) |
| Present | 18/209 | 0.62 (0.32, 1.22) | 0.98 (0.47, 2.01) | 1.08 (0.51, 2.27) |
| Bizarre atypia | | | | |
| Absent | 35/184 | (ref) | (ref) | (ref) |
| Present | 4/185 | 0.09 (0.03, 0.27) | 0.12 (0.04, 0.37) | 0.13 (0.04, 0.38) |

* Adjusted for variables that were predictors of discordance in univariate models ($p < 0.10$). These included histotype and bizarre atypia.

[†] Mutually adjusted for all other predictors in the table.

Table 4

Predictors of grading discordance between pathologists and pathology reports

| | Discordant cases/total cases | Unadjusted odds ratio (95% CI) | Adjusted odds ratio* (95% CI) | Fully adjusted odds ratio [†] (95% CI) |
|--|------------------------------|-----------------------------------|----------------------------------|--|
| Date of diagnosis (5-year increments) | 49/232 | 1.07 (0.83, 1.37) | 1.02 (0.78, 1.34) | 1.01 (0.77, 1.33) |
| Number of slides | 49/232 | 0.99 (0.96, 1.02) | 1.00 (0.97, 1.03) | 1.00 (0.97, 1.03) |
| Slide Quality | | | | |
| Good | 42/201 | (ref) | (ref) | (ref) |
| Poor | 7/31 | 1.10 (0.45, 2.74) | 0.93 (0.35, 2.51) | 0.90 (0.33, 2.45) |
| Hospital | | | | |
| Hospital with fellowship program | 4/22 | (ref) | (ref) | (ref) |
| NCI center or academic hospital without fellowship program | 16/54 | 1.90 (0.55, 6.49) | 1.81 (0.50, 6.49) | 1.80 (0.50, 6.47) |
| Community hospital | 28/147 | 1.06 (0.33, 3.37) | 0.97 (0.29, 3.27) | 0.96 (0.28, 3.24) |
| Unknown | 1/9 | 0.56 (0.05, 5.86) | 0.65 (0.06, 7.19) | 0.65 (0.06, 7.23) |
| Stage | | | | |
| 1/2 | 19/53 | (ref) | (ref) | (ref) |
| 3/4 | 30/179 | 0.36 (0.18, 0.72) | 0.58 (0.25, 1.35) | 0.55 (0.23, 1.29) |
| Histotype | | | | |
| Serous | 29/170 | (ref) | (ref) | (ref) |
| Non-serous | 14/29 | 4.54 (1.98, 10.42) | 2.52 (0.90, 7.10) | 2.49 (0.87, 7.09) |
| Missing | 6/33 | 1.08 (0.41, 2.85) | 0.97 (0.36, 2.65) | 0.97 (0.34, 2.72) |
| SET | | | | |
| 50% | 25/124 | (ref) | (ref) | (ref) |
| >50% | 1/31 | 0.13 (0.02, 1.02) | 0.08 (0.01, 0.71) | 0.08 (0.01, 0.84) |
| Geographic necrosis | | | | |
| Absent | 25/94 | (ref) | (ref) | (ref) |
| Present | 24/138 | 0.58 (0.31, 1.10) | 0.72 (0.36, 1.41) | 0.71 (0.35, 1.46) |
| Bizarre atypia | | | | |
| Absent | 28/98 | (ref) | (ref) | (ref) |
| Present | 21/134 | 0.47 (0.25, 0.88) | 0.61 (0.31, 1.21) | 0.63 (0.31, 1.27) |

* Adjusted for variables that were predictors of discordance in univariate models ($p < 0.10$). These included stage, histotype, geographic necrosis and bizarre atypia. The SET odds ratio is adjusted for stage, geographic necrosis and bizarre atypia only.

[†] Mutually adjusted for all other predictors in the table. The model used to estimate the odds ratio for SET was not adjusted for histotype.