Predictive modeling for determination of microscopic residual disease at primary cytoreduction: An NRG Oncology/Gynecologic Oncology Group 182 Study☆,☆☆,★

Neil S. Horowitz [a,1], G. Larry Maxwell [h,*,1], Austin Miller [b], Chad A. Hamilton [c], Bunja Rungruang [d], Noah Rodriguez [e], Scott D. Richard [f], Thomas C. Krivak [g], Jeffrey M. Fowler [i], David G. Mutch [j], Linda Van Le [k], Roger B. Lee [l], Peter Argenta [m], David Bender [n], Krishnansu S. Tewari [o], David Gershenson [p], James J. Java [b], Michael A. Bookman [q]

[a] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Brigham & Women's Hospital, Boston, MA 02115, United States
[b] NRG Oncology/Gynecologic Oncology Group, Statistical and Data Center, Roswell Park Cancer Institute, Buffalo, NY 14263, United States
[c] Gynecologic Cancer Center of Excellence, Walter Reed National Military Medical Center and Uniformed Services University of the Health Sciences, Bethesda, MD 20889, United States
[d] Division of Gynecologic Oncology, Medical College of Georgia of Augusta University, Augusta, GA 30912, United States
[e] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Kaiser Permanente Irvine Medical Center, Irvine, CA 92868, United States
[f] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Hahnemann University Hospital, Philadelphia, PA 19102, United States
[g] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Western Pennsylvania Hospital, Pittsburgh, PA 15222, United States
[h] Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, and the Inova Schar Cancer Institute, Inova Fairfax Women's Hospital, Falls Church, VA 22042, United States
[i] Dept. of Gynecologic Oncology, Ohio State University Comprehensive Cancer Center, Columbus, OH 43210, United States
[j] Department of Obstetrics and Gynecology, Washington University School of Medicine, Saint Louis, MO 63110, United States
[k] Department of OB/GYN, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, United States
[l] Department of GYN/ONC, Tacoma General Hospital, Tacoma, WA 98405, United States
[m] Division of Gynecologic Oncology, Department of Obstetrics, Gynecology and Women's Health, University of Minnesota, Minneapolis, MN 55455, United States
[n] Gyn/Onc Division, University of Iowa, Iowa, IA 52242, United States
[o] Department of Gynecologic Oncology, University of California at Irvine, Orange, CA 92868, United States
[p] Department of GYN/ONC, The University of Texas MD Anderson Cancer Center, Houston, TX 77230, United States
[q] US Oncology Research and Arizona Oncology, Tucson, AZ 85711, United States

HIGHLIGHTS

- A predictive model for R0 was developed and validated.
- Variables included disease score, stage, CA125, ascites, and stage-age interaction.
- Model alone should not determine management of advanced stage ovarian cancer.

A B S T R A C T

Objective. Microscopic residual disease following complete cytoreduction (R0) is associated with a significant survival benefit for patients with advanced epithelial ovarian cancer (EOC). Our objective was to develop a prediction model for R0 to support surgeons in their clinical care decisions.

Methods. Demographic, pathologic, surgical, and CA125 data were collected from GOG 182 records. Patients enrolled prior to September 1, 2003 were used for the training model while those enrolled after constituted the validation data set. Univariate analysis was performed to identify significant regression predictors of R0 and these variables were subsequently analyzed using multivariable regression. The regression model was reduced using backward selection and predictive accuracy was quantified using area under the receiver operating characteristic area under the curve (AUC) in both the training and the validation data sets.

Results. Of the 3882 patients enrolled in GOG 182, 1480 had complete clinical data available for the analysis. The training data set consisted of 1007 patients (234 with R0) while the validation set was comprised of 473 patients (122 with R0). The reduced multivariable regression model demonstrated several variables predictive of R0 at cytoreduction: Disease Score (DS) ($p < 0.001$), stage ($p = 0.009$), CA125 ($p < 0.001$), ascites ($p < 0.001$), and stage-age interaction ($p = 0.01$). Applying the prediction model to the validation data resulted in an AUC of 0.73 (0.67 to 0.78, 95% CI). Inclusion of DS enhanced the model performance to an AUC of 0.83 (0.79 to 0.88, 95% CI).

Conclusions. We developed and validated a prediction model for R0 that offers improved performance over previously reported models for prediction of residual disease. The performance of the prediction model suggests additional factors (i.e. imaging, molecular profiling, etc.) should be explored in the future for a more clinically actionable tool.

1. Introduction

Ovarian cancer is the fourth most common gynecologic malignancy among women in the United States. By American Cancer Society estimates, 22,280 women will be newly diagnosed in 2016 and 14,240 women will die of their disease [1]. Patients with ovarian cancer are typically treated with a combination of surgical cytoreduction and adjuvant chemotherapy. Although outcomes are adversely affected by the grade, histology, and stage of disease, anatomical distribution of disease and the volume of residual disease following surgical debulking may be the strongest determinants of survival [2]. Upper abdominal surgery is often needed to achieve maximal cytoreduction of disease and institutional improvements in patient outcome have coincided from programmatic transition to more aggressive operative intervention [3]. The Gynecologic Oncology Group (GOG) has previously defined "optimal" residual as <1 cm, but multiple reports now support maximum cytoreduction to microscopic residual (R0) as being the new standard [4–7].

The management of patients in whom R0 cannot be obtained surgically remains controversial. Some authors advocate for Neoadjuvant Chemotherapy (NACT) with Interval Debulking Surgery (IDS) in this scenario [8,9] while others support aggressive Primary Cytoreductive Surgery (PDS) as long as the patients can be debulked to <1 cm of disease [10]. Some institutional studies show improved outcomes from cytoreduction to minimal gross residual disease (>R0 < 1 cm) [10]. However, these improvements have not been reproduced using multi-institutional data [6,7]. In a study by Horowitz et al. [11], multivariable modeling of 2655 GOG patients with <1 cm of disease affirmed that disease distribution and disease residual, not the surgical procedure, appear to be the most important determinants of patient prognosis. In addition, a prospective clinical trial performed by the European Organization for Research and Treatment of Cancer, investigators found that R0 patients with NACT and IDS had a median survival of 38 months and 28% 5-year survival compared to patients primarily debulked to minimal gross residual disease (>R0 < 1 cm), who had a median survival of 32 months and a 23% 5-year survival [12]. Although no statistical analysis was provided, these results do not demonstrate worse outcome for patients that receive NACT and have a higher chance for R0 at IDS compared to patients that undergo PDS with >R0 < 1 cm. Based on these studies [9–12], many feel that radical surgery should be reserved

for those patients that can be debulked to R0, but this is not easy to predict prior to surgery. In cases in which R0 is feasible and safe for the patient, the full range of upper abdominal and conventional cytoreductive procedures should be deployed [13,14].

Accurate pre-operative identification of patients who are likely to achieve R0 status at surgery could allow for more careful selection of patients for PDS while patients with a low likelihood of R0 could be considered for NACT and IDS, or a minimally invasive diagnostic assessment with laparoscopy to determine R0 feasibility. Previous attempts to develop prediction models of postoperative residual disease have predominantly used <1 cm (instead of R0) as an endpoint for measuring accuracy. In addition, the models developed thus far have either not been validated in a set of patients independent from which the model was developed or failed external validation. We sought to develop a more robust validated model for prediction of R0 in women with ovarian cancer using data from a multi-institutional cooperative group randomized trial.

2. Patients and methods

All patient data were abstracted from GOG patients enrolled on the international collaborative trial GOG-182-ICON5 [15] using existing case report forms. Patients were enrolled with FIGO stage III or IV histologically confirmed epithelial ovarian cancer (EOC) or primary peritoneal carcinoma (PPC). All patients underwent PDS to either ≤1 cm or >1 cm residual disease before being randomized to one of the five platinum and paclitaxel-based chemotherapy regimens. No statistically significant treatment effects on progression-free survival (PFS) or overall survival (OS) were found among the five chemotherapy regimens. Further details of eligibility criteria and results from the original study have been published [15].

Patient demographics and tumor characteristics, including age, race, performance status, tumor grade, presence of ascites (any versus none), and histology were extracted from the GOG database. Information describing surgical procedures and preoperative extent of disease based on 56 anatomic locations was abstracted from GOG surgical reporting forms and diagrams. All operative and pathology reports were evaluated to obtain accurate descriptions of disease (distribution and residual) and surgical procedures. Preoperative CA125 was obtained from medical records submitted to the GOG as part of enrollment in GOG-182.

Initial sites of disease identified during primary debulking surgery were used to develop the preoperative disease score (DS). The DS classes were defined as follows: (1) DS-low, with pelvic and retroperitoneal spread; (2) DS-moderate (DS-mod), with additional spread to the abdomen but sparing the upper abdomen; or (3) DS-high, with the presence of upper abdominal disease affecting the diaphragm, spleen, liver or pancreas. The current study included all patients with FIGO stage III or IV EOC or PPC. The residual disease outcome was scored as completely resected disease (R0), minimal gross residual disease ($>R0 < 1$ cm), or suboptimal cytoreduction (>1 cm) following PDS. The raw (unadjusted) proportion of patients with R0 outcomes was estimated for each of Institution in the dataset. Exploratory analysis showed considerable variability in the R0 estimates among Institutions supported by fewer patients. The 3-level R0 stratification factor was defined at the tertiles of R0 distribution. The R0 factor levels were low (0 to 16%), middle (16.5 to 31%) and high (31.5 to 100%).

Model training and validation subsets were defined based on the date of patient enrollment on the study. The patients enrolled between February 7, 2001 and August 31, 2003 were assigned to the training data set while the remaining patients enrolled from September 1, 2003 to September 1, 2004 were used for model validation. While the training and validation patients were similar by virtue of enrollment on the same study, this temporal allocation to training and validation sets provides a prospective assessment of model performance [16].

Using the training data, exploratory univariate analysis was performed to identify predictors of R0 following PDS. Covariates for the full multivariable logistic regression model included patient characteristics believed to be both available to inform the surgery decision and related to the R0 outcome. The model specification included main effects and second-order interaction terms for age, stage, presence of ascites, CA125 (log transformed), and GOG performance score (similar to Karnofsky). This full model was then reduced using backward selection constrained to respect interaction term hierarchy, with a covariate retention threshold of $p < 0.20$. The reduction did not significantly degrade the model fit ($p = 0.7$ by full and reduced model Chi square test). This process was repeated with DS included as a covariate. To explore institutional effects on predictive accuracy, the R0 model including DS was respecified to adjust for Institute (parent) or level R0 strata. This factor was included as a fixed effect in the multivariable model, with the middle tertile as the reference group.

R0 prediction scores for the patients in the validation data were obtained using coefficient estimates from the reduced model fit to the training data. The ability of the models to accurately distinguish between R0 vs > R0 was graphically presented with Receiver Operating Characteristic (ROC) curves. ROC curves display the classification accuracy of a predictive model score as the classification threshold is varied. For different cut-off points of the model score, the true positive fraction is plotted in function of the false positive fraction. The classification accuracy is summarized by the area under the ROC (AUC) and its 95% confidence limits. AUC values >0.80 were considered to have some utility in predicting the response of individual patients [17]. The Youden Index [18] estimated from the training data was used to classify patients as having either a high or low propensity for R0. This index is the basis of other indicators of classification accuracy, including estimates for sensitivity, specificity, positive and negative predictive values.

All data analyses were generated using SAS/STAT software, Version 9.4. Copyright 2012, SAS Institute Inc. SAS is a registered trademark of SAS Institute Inc., Cary, NC, USA.

3. Results

Of the 3882 patients in the original dataset of GOG-182, 1480 had complete preoperative data elements and were used to generate the training cohort ($n = 1007$) and validation ($n = 473$) cohort for this study. Although a significant number of patients were excluded because they did not have documentation of CA125 and ascites data ($n = 2215$), the proportion of R0 and $>R0$ was similar between the included and excluded cases (Supplementary Fig. 1). A total of 356 patients (24%) had R0, 23% of the training cohort ($n = 234$) and 26% of the validation cohort ($n = 122$).

There were 73 parent hospitals and 240 affiliates that contributed patients to this study. The Institutes had a median of 15 patients in this dataset and 75% of Institutes had 30 or fewer patients. The median R0 for the parent institution was 23% while the median R0 for affiliates was 17%. The distribution of R0 by anonymized institution and number of patients enrolled is shown in Supplementary Figs. 2 and 3 and Supplementary Tables 1 and 2. Approximately 75% of affiliate hospitals in our sample were represented by <8 patients. Our sample was not strong enough to support hospital-level R0 estimates and so patients from affiliates were combined with patients from the parent program in providing an institution variable for modeling.

The demographics and disease characteristics for patients with R0 vs > R0 disease for the training cohort are presented in Tables 1 and 2. Many of the clinical characteristics were similar between those with R0 and >R0 residual. However, the probability of having R0 disease was significantly associated with age, performance status, stage, extent of disease (DS) at exploration, and presence of ascites (Table 3). Not surprisingly, older, symptomatic patients with advanced stage, large volume disease, and ascites were less likely to have R0 residual. In addition, patients with >R0 were more likely to be older with symptoms and be less ambulatory compared to patients with R0 (Table 1).

CA125 values were significantly lower for R0 (median 345 u/mL) compared to $>R0 < 1$ cm (median 839 u/mL) or >1 cm (median 1241 u/mL), while the values between $>R0 < 1$ cm and >1 cm did not differ significantly. Indeed, patients in all three residual disease groups had the full range of CA125 values, making classification on the basis of CA125 alone problematic. An exploratory analysis considered log-transformed CA125 as a continuous predictor of the residual disease status (R0, $>R0 < 1$ cm, and >1 cm). Increasing CA125 was related to increasing the predicted probability of $>R0 < 1$ cm and >1 cm, and decreasing probability of R0. The pattern is intuitively satisfying, and the predicted probabilities become more compelling for CA125 value >1000 u/mL. At this point, the predicted probably of R0 dropped to 0.18 (0.16 to 0.21), vs 0.51 (0.47 to 0.54) for $>R0 < 1$ cm and 0.31 (0.28 to 0.34) for >1 cm. Unfortunately, the CA125 model is least effective in lower ranges. For example, at CA125 = 35 u/mL, the predicted probably of R0 is 0.55 (0.47 to 0.63), providing no more information than a coin flip.

A multivariable logistic regression model to predict the probability of R0 disease (reference: $>R0$) was specified to include factors believed to be both available to inform the surgery decision and related to the R0 outcome. In preliminary analysis, non-linear effects of age and CA125 were not statistically significant, and were thus excluded. We elected to control for institutions as a variable in our modeling which was inclusive of affiliates. Since 75% of affiliates were represented by <8 patients, we did not use hospital as a variable in the adjusted model. The full regression model was reduced by backward selection leaving age (linear) ($p < 0.001$), log-transformed preoperative CA125 (linear) ($p < 0.001$), DS ($p < 0.001$), stage ($p = 0.009$), presence of ascites ($p < 0.001$), and the age-stage interaction term ($p = 0.01$) as statistically significant factors. Final model specification and coefficient estimates are shown in Table 4.

We reproduced the full and reduced multivariable regression models including and excluding DS from the modeling. The performance of the model exclusive of DS revealed an AUC for the model in the training set of 0.78 (95% CI 0.74–0.81) (Fig. 1A) and in the validation data set of 0.73 (95% CI 0.67–0.78) (Fig. 1B). Inclusion of DS in the prediction model resulted in an AUC 0.87 (95% CI 0.84–0.90) in the training set (Fig. 2A) and an AUC of 0.83 (95% CI 0.79 to 0.88) in the validation data set (Fig. 2B). An AUC of 0.87 was obtained in the validation set when both DS and institution were included in the final model (Supplementary Table 2 and Supplementary Fig. 4). With the DS model scoring

Table 1
Demographics of the Training Cohort by Residual Disease Group.

		>R0	R0	Overall	p value
Overall	N	773 (76.8)	234 (23.2)	1007 (100%)	
Age (Y)	Mean/StdErr	59.5/0.4	56.2/0.7	58.7/0.3	<0.001
	Median/min/max	59.4/29.5/87.5	55.7/24.5/84.7	58.9/24.5/87.5	
Age (Cat)	0–54	274 (35.4%)	107 (45.7%)	381 (37.8%)	0.005
	55–64	248 (32.1%)	74 (31.6%)	322 (32.0%)	
	65+	251 (32.5%)	53 (22.6%)	304 (30.2%)	
Race	Unknown	13 (1.7%)	4 (1.7%)	17 (1.7%)	0.708
	Asian	15 (1.9%)	3 (1.3%)	18 (1.8%)	
	Black	34 (4.4%)	7 (3.0%)	41 (4.1%)	
	White	711 (92.0%)	220 (94.0%)	931 (92.5%)	
Ethnicity	Hispanic	13 (1.7%)	4 (1.7%)	17 (1.7%)	0.682
	Non-Hispanic	704 (91.1%)	209 (89.3%)	913 (90.7%)	
	Not known	56 (7.2%)	21 (9.0%)	77 (7.6%)	
BMI	Under	27 (3.5%)	8 (3.4%)	35 (3.5%)	0.412
	Normal	321 (41.5%)	89 (38.0%)	410 (40.7%)	
	Over	213 (27.6%)	69 (29.5%)	282 (28.0%)	
	Obese	186 (24.1%)	54 (23.1%)	240 (23.8%)	
	Missing	26 (3.4%)	14 (6.0%)	40 (4.0%)	
Performance	Asymptomatic	320 (41.4%)	127 (54.3%)	447 (44.4%)	0.001
	Ambulatory	397 (51.4%)	98 (41.9%)	495 (49.2%)	
	In bed <50%	56 (7.2%)	9 (3.8%)	65 (6.5%)	
Performance	Symptomatic	453 (58.6%)	107 (45.7%)	560 (55.6%)	<0.001
	Asymptomatic	320 (41.4%)	127 (54.3%)	447 (44.4%)	

the validation dataset, the optimal Youden Index = 0.28 [18] had corresponding sensitivity and specificity of 0.80 and 0.76 respectively. Additional indicators of classification accuracy are provided in Table 5.

4. Discussion

Large numbers of patients with clinical information are required to develop prediction models that control for multiple confounding influences. Typically, one predictor for every 10 events is recommended in order to avoid creation of an exploratory model that does not validate in subsequent independent testing [19,20]. Although prior models using small numbers of patients and a regression model with multiple variables demonstrated impressive internal prediction accuracy [21], these initial findings were not substantiated in subsequent validation studies in independent cohorts of patients [22,23].

Many surgeons have now adopted R0 as the most appropriate goal for primary or interval surgery, even though clinical outcomes are

Table 2
Pathologic and surgical characteristics of the Training Cohort by Residual Disease Group.

		>R0	R0	Overall	p value
Overall	N	773 (76.8)	234 (23.2)	1007 (100%)	
Site	Ovary	662 (85.6%)	220 (94.0%)	882 (87.6%)	<0.001
	Other	111 (14.4%)	14 (6.0%)	125 (12.4%)	
Histology	Adeno unsp	5 (0.6%)	2 (0.9%)	7 (0.7%)	<0.001
	Clear cell	9 (1.2%)	19 (8.1%)	28 (2.8%)	
	Endomet	41 (5.3%)	28 (12.0%)	69 (6.9%)	
	Mucinous	7 (0.9%)	1 (0.4%)	8 (0.8%)	
	Mixed epi	41 (5.3%)	13 (5.6%)	54 (5.4%)	
	Undiffer	7 (0.9%)	4 (1.7%)	11 (1.1%)	
	Transition	3 (0.4%)	2 (0.9%)	5 (0.5%)	
	Serous	660 (85.4%)	165 (70.5%)	825 (81.9%)	
Grade	1	20 (3.1%)	11 (5.6%)	31 (3.7%)	0.267
	2	130 (20.1%)	40 (20.2%)	170 (20.1%)	
	3	497 (76.8%)	147 (74.2%)	644 (76.2%)	
Stage	3	638 (82.5%)	214 (91.5%)	852 (84.6%)	<0.001
	4	135 (17.5%)	20 (8.5%)	155 (15.4%)	
Ascites	Yes	681 (88.1%)	117 (50.0%)	798 (79.2%)	<0.001
	No	92 (11.9%)	117 (50.0%)	209 (20.8%)	
Ext of diseas	DS-low	14 (1.8%)	37 (15.8%)	51 (5.1%)	<0.001
	DS-mod	100 (12.9%)	131 (56.0%)	231 (22.9%)	
	DS-high	659 (85.3%)	66 (28.2%)	725 (72.0%)	

Adeno unsp - Adenocarcinoma unspecified; Endomet - Endometrioid, Epi – Epithelial; Undiff – Undifferentiated; Transition – Transitional; Ext of Diseas – Extent of disease; DS - Disease score; Mod - moderate.

influenced by other factors, including the initial extent of disease. Our prediction model incorporating the clinical parameters of preoperative CA125 (linear), disease score, stage, and age-stage interaction is the first non-surgical, validated, and reproducible predictor for disease residual and is applicable to the contemporary endpoint of R0. Such a model and/or further refinements could complement clinical judgement when considering a patient for PDS versus NACT. This model was built on an extensive training set derived from multiple institutions avoiding the problem of overfitting while enhancing generalizability. Our current model included 1007 patients, 237 of which had the residual disease endpoint (R0) thereby mitigating the effect of overfitting. Further validation of our prediction model has shown consistency in performance. Prior prediction models of residual disease have used <1 cm as an endpoint for defining "optimal" residual disease [22,23, 25–28] and we have chosen R0 as the more contemporary definition for "optimal" for the current modeling. Development of an accurate predictive model for achieving R0 status would have significant clinical utility for many surgeons in making decisions about primary cytoreduction versus NACT.

Establishment of an exploratory model and subsequent testing in an independent validation set also should be approached with caution in order to establish generalizability while avoiding selection bias [24]. If one has 2 populations of patients (i.e. A and B), it is ideal to select patients from both A & B to create the prediction model and then test the model in an independent set of patients from A & B. Development of the model in one population (group A) and then validation in another population (group B) can lead to poor performance of the model in independent cohorts. In a study by Dowdy et al. [25], a prediction model incorporating ascites and diffuse peritoneal thickening was developed to identify 23 of 87 patients with suboptimal (>1 cm) cytoreduction achieving 79% accuracy. Although this prediction model showed promising performance characteristics, two subsequent studies used patients from other institutions to externally validate the model and noted that accuracy dropped to 48% [21]–65% [23]. Similarly, the accuracy of two prediction models for patients with <1 cm of residual disease developed by investigators from Italy [26] demonstrated a drop in AUC from 0.78 and 0.81 to 0.56–0.59 and 0.55–0.60 respectively when assessed in an independent patient cohort by investigators in The Netherlands [27]. In the current study, patients from 73 institutions participating in GOG-182 reflected a diversity of surgical skill levels while providing standardization of adjuvant therapy. Our training set and validation set of patients were similar in terms of the institutions from which

A.

B.

Fig. 1. Receiver operating characteristic curves with area under the curve (AUC) and 95% confidence interval estimates from model excluding disease distribution score, fit on the training (A) and validation (B) data. No adjustment for Institution.
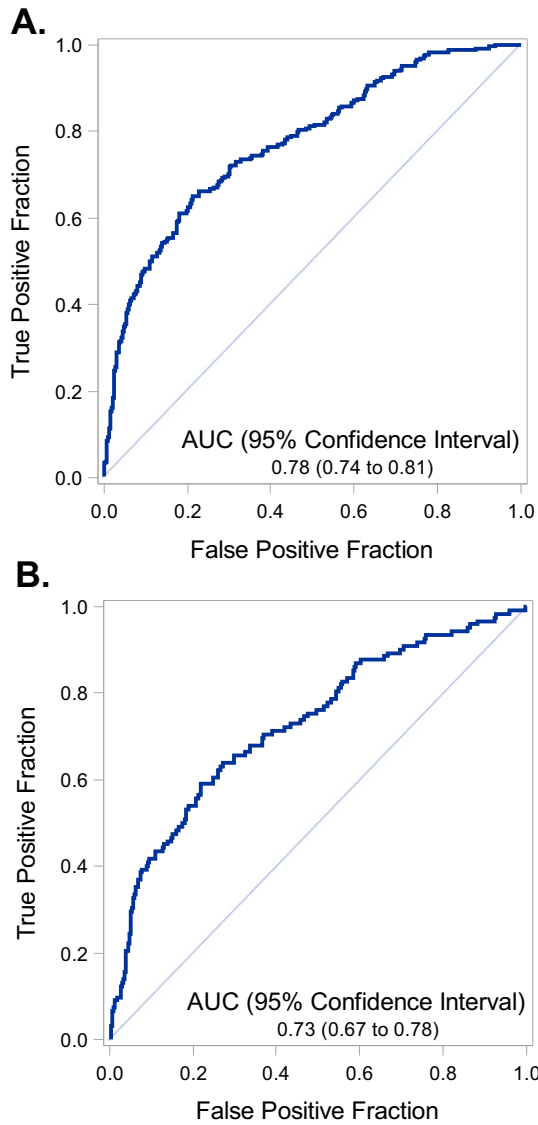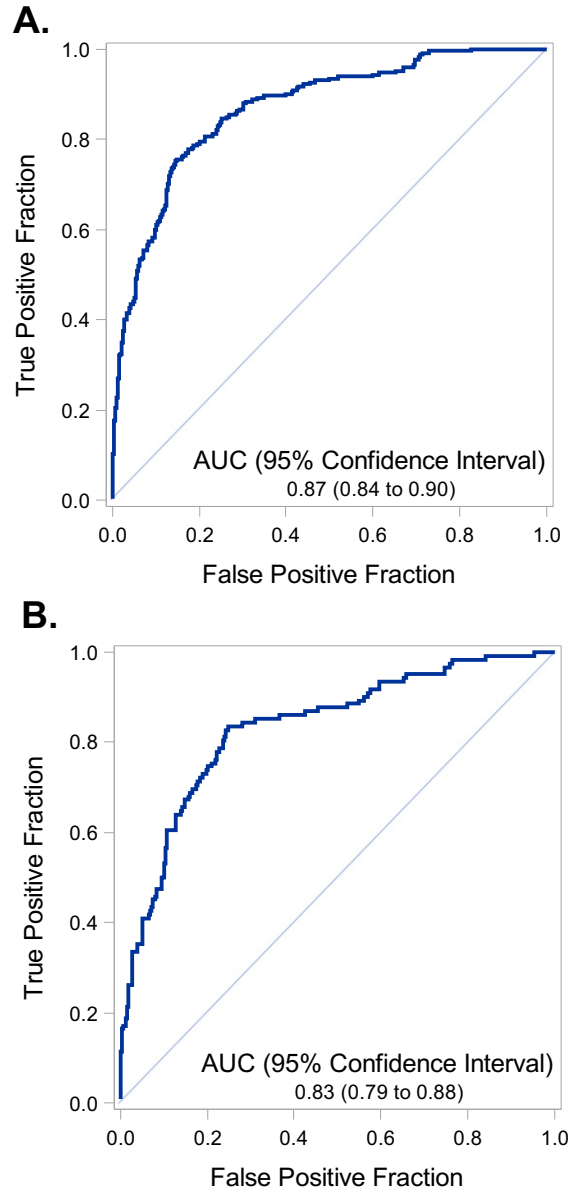
A.

B.

Fig. 2. Receiver operating characteristic curves with area under the curve (AUC) and 95% confidence interval estimates from model including disease distribution score, fit on the training (A) and validation (B) data. No adjustments for Institution.

they were enrolled avoiding effects of selection bias on the prediction model performance in validation while enhancing the generalizability of a model that might result from single institution reporting. Consecutive patients were used to create the training model and the subsequent validation test set so that the temporal relation allowed for prospective testing of the model thereby further enhancing its reliability [16].

Investigators from the Mayo Clinic recently published a prediction model for R0 [29]. Although the model (including ascites, omental thickening, and diffuse peritoneal thickening) was not externally validated using an independent sample set, investigators noted an AUC of 0.748 when internally assessing the model using a bootstrap approach [29]. Conversely, our model for R0 was developed using a large training set reflective of patients from multiple institutions and represents the only externally validated model reported to date. In the validation data, our model inclusive of DS had a AUC of 0.83 (0.79 to 0.88). Inclusion of institution as a variable in the model improved the AUC further to 0.87 in the validation dataset.

Despite its strengths, our current study has limitations. Distribution of disease (DS) was obtained from operative notes and not pre-operative CT scans. Previous reports have highlighted the limitations of pre-operative CT scans for identification of intraperitoneal disease. In a study by Ko et al. [30], 68 of 148 patients with stage III and IV EOC had false negative metastasis that were not detected on initial CT but identified during surgery. Re-examination of the CT scans a second time identified 63% of lesions that were missed on the initial examination. Most importantly, no case was found to be false positive for metastatic disease [30]. Although we recognize that limitations in CT detection of metastasis might affect negative predictive value, patients unexpectedly found to have widespread disease intraoperatively could have aggressive cytoreduction postponed if radical debulking cannot render the patient R0. When we assessed the performance of the model for R0 prediction with and without inclusion of DS in the validation dataset, we identified a substantial improvement in the AUC from 0.73 in the model without DS to 0.83 in the model inclusive of DS in the validation cohort. One could assume that the accuracy of the model could fall somewhere in that range depending on the sensitivity of radiologic imaging in detection of disease distribution. Finally, although inclusion of institutions improved the accuracy of the prediction model to an AUC

Table 3
Treatment and outcomes of the Training Cohort by Residual Disease Group.

		>R0	R0	Overall	p value
Overall	N	773 (76.8)	234 (23.2)	1007 (100%)	
Treatment	Pac + Car	168 (21.7%)	37 (15.8%)	205 (20.4%)	0.174
	Pac + Car + Gem	148 (19.1%)	46 (19.7%)	194 (19.3%)	
	Pac + Dox + Car	156 (20.2%)	52 (22.2%)	208 (20.7%)	
	Car + Top + Pac + Car	138 (17.9%)	54 (23.1%)	192 (19.1%)	
	Car + Gem + Pac + Car	163 (21.1%)	45 (19.2%)	208 (20.7%)	
Alive/dead	Alive	113 (14.6%)	91 (38.9%)	204 (20.3%)	<0.001
	Dead	660 (85.4%)	143 (61.1%)	803 (79.7%)	
Surg aggr scr	Mean/StdErr	5.0/0.1	5.2/0.2	5.0/0.1	0.262
	Median/min/max	4.0/0.0/15.0	4.0/1.0/15.0	4.0/0.0/15.0	
Score group	CS-Low	180 (23.3%)	36 (15.4%)	216 (21.4%)	0.014
	CS-Mod	465 (60.2%)	164 (70.1%)	629 (62.5%)	
	CS-High	128 (16.6%)	34 (14.5%)	162 (16.1%)	
Surgery grp	NUAP	643 (83.2%)	190 (81.2%)	833 (82.7%)	0.481
	UAP	130 (16.8%)	44 (18.8%)	174 (17.3%)	
Resid abstrtd	Microscopic		234 (100.0%)	234 (23.2%)	<0.001
	Opt (<=1)	481 (62.2%)		481 (47.8%)	
	Sub Opt	292 (37.8%)		292 (29.0%)	

Pac – Paclitaxel; Car - Carboplatin; Dox - Liposomal doxorubicin; Top - Topotecan; Gem - Gemcitabine; Dis - Disease; Surg aggr scr - Surgical aggressiveness score; CS - surgical complexity; Grp - Group; NUAP – no upper abdominal procedures; UAP - upper abdominal surgery; Resid abstrtd – Residual disease abstracted; Opt – optimal.

of 0.87, there are limitations in the accuracy of the institution rate of R0 which was limited to small numbers of patients enrolled in GOG 182 at each site. R0 rates of 100% or 0% from a few of the institutions enrolling low numbers of patients highlight this limitation.

Given the limitations of imaging and clinical prediction of R0 at cytoreduction, some surgeons have advocated use of diagnostic laparoscopy to evaluate disease distribution and assess resectability. Using intraoperative findings at diagnostic laparoscopy to develop a risk score, Fagotti et al. [31] developed a model to increase their rates of achieving R0 at PDS up to 57% of patients. At a threshold score of 10, the likelihood of R0 was 0 and the risk of laparotomy was 32%. Using a predictive index value (PIV) ≥ 8 the probability of achieving R0 at cytoreduction was zero. The discriminating performance of this predictor for R0 revealed

an AUC of 0.885. Several other prospective trials using laparoscopic disease assessment to triage patients are ongoing potentially allowing personalized surgical management [14,32].

Alternative to clinical or surgical parameters to predict R0 cytoreduction, molecular markers and gene expression signatures are likely to be informative for surgical outcome. Using data from 13 publicly available databases Reister et al. developed and validated a gene expression signature to predict suboptimal cytoreduction [33]. Based on protein expression of POSTN, CXCL14, and pSmad 2/3 these authors report a 93% accuracy of determining a high risk and low risk group for suboptimal cytoreduction. Perhaps combining our model with this genomic tool will provide surgeons the opportunity to triage patients appropriately for primary cytoreduction or NACT.

In conclusion, our model is an independently reproducible model for prediction of disease residual and is applicable to the more contemporary endpoint of R0 as the goal of surgery. However, it is not meant to be used as the sole determinant of whether a surgeon should offer a patient PDS versus NACT and IDS. Nonetheless, we hope that model will provide surgeons with additional data to use in conjunction with the medical condition of the patient, clinical intuition, and their patient's preferences in deciding on a course of action.

Table 4
Maximum likelihood coefficient estimates to predict R0 (ref: >R0) from the predictive models using the training dataset.

Covariate	Log odds estimate	Standard error	p value
Model excluding DS (Fig. 1A and B)			
Intercept	2.55	0.64	<0.001
Log(Pre-op CA125) (continuous)	−0.35	0.06	<0.001
Age (continuous)	−0.03	0.01	<0.001
Ascites			
Yes	Reference		
No	1.65	0.19	<0.001
Stage			
3	Reference		
4	−3.92	1.56	0.01
Age ∗ stage interaction	0.06	0.03	0.03
Model including DS (Fig. 2A and B)			
Intercept	1.54	0.71	0.03
Log(Pre-op CA125) (continuous)	−0.27	0.07	<0.001
Age (continuous)	−0.04	0.01	<0.001
Ascites			
Yes	Reference		
No	1.19	0.22	<0.001
Stage			
3	Reference		
4	−4.28	1.64	0.01
Age ∗ stage interaction	0.07	0.03	0.01
Disease score			
DS-high	Reference		
DS-mod	2.32	0.20	<0.001
DS-low	2.53	0.37	<0.001

DS – Disease score.
Log odds estimates greater than zero indicate a greater propensity for microscopic residual disease after surgery, as compared to patients in the reference group, given the other covariates in the model.

Table 5
Classification accuracy of the predictive models.

Model	Accuracy	Training Data	Validation Data
Excluding DS (Youden = 0.23)	AUC	0.78 (0.74 to 0.81)	0.73 (0.67 to 0.78)
	Sensitivity	0.65 (0.58 to 0.70)	0.61 (0.52 to 0.69)
	Specificity	0.79 (0.76 to 0.82)	0.74 (0.70 to 0.79)
	PPV	0.48 (0.42 to 0.53)	0.45 (0.37 to 0.53)
	NPV	0.88 (0.85 to 0.90)	0.84 (0.80 to 0.88)
Including DS (Youden = 0.28)	AUC	0.87 (0.84 to 0.90)	0.83 (0.79 to 0.88)
	Sensitivity	0.76 (0.70 to 0.81)	0.80 (0.72 to 0.86)
	Specificity	0.85 (0.82 to 0.87)	0.76 (0.72 to 0.81)
	PPV	0.60 (0.55 to 0.66)	0.54 (0.47 to 0.61)
	NPV	0.92 (0.90 to 0.94)	0.91 (0.88 to 0.94)

DS - Disease score, AUC – area under curve, PPV - positive predictive value, NPV - negative predictive value.
Measures of discrimination accuracy of the R0 prediction scores. Patients were classified as testing positive or negative using the Youden index estimated from the Training data. Accuracy indicators from the Training and Validation data are shown separately. Patients with predicted R0 probability greater or equal to the Youden index were classified as testing positive, or having a high likelihood of R0 after surgery. PPV is the proportion of patients with a positive test who actually attained R0. NPV is the proportion of patients with a negative test who actually failed to attain R0. Two-sided, α = 0.05 Jeffrey's confidence intervals indicate reliability of the estimates.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygyno.2017.10.011.

Conflicts of interest

References

[1] http://www.cancer.org/research/cancerfactsstatistics/cancerfactsfigures2016/index.

[2] S.J. Chang, R.E. Bristow, D.S. Chi, W.A. Cliby, Role of aggressive surgical cytoreduction in advanced ovarian cancer, J. Gynecol. Oncol. 26 (2015) 336–342.

[3] D.S. Chi, E.L. Eisenhauer, O. Zivanovic, Y. Sonoda, N.R. Abu-Rustum, D.A. Levine, et al., Improved progression free and overall survival in advanced ovarian cancer as a result of a change in surgical paradigm, Gynecol. Oncol. 114 (2009) 26–31.

[4] K.K. Shih, D.S. Chi, Maximal cytoreductive effort in epithelial ovarian cancer surgery, J. Gynecol. Oncol. 21 (2010) 75–80.

[5] D.S. Chi, E.L. Eisenhauer, J. Lang, J. Huh, L. Haddad, N.R. Abu-Rustum, et al., What is the optimal goal of primary cytoreductive surgery for bulky stage IIIC epithelial ovarian carcinoma (EOC)? Gynecol. Oncol. 03 (2006) 559–564.

[6] W.E. Winter III, G.L. Maxwell, C. Tian, J.W. Carlson, R.F. Ozols, P.G. Rose, et al., Prognostic factors for advanced epithelial ovarian cancer: a gynecologic oncology group study, J. Clin. Oncol. 25 (2007) 3621–3627.

[7] N. Rodriguez, A. Miller, S.D. Richard, B. Rungruang, C.A. Hamilton, M.A. Bookman, et al., Upper abdominal procedures in advanced stage ovarian or primary peritoneal carcinoma patients with minimal or no gross residual disease: an analysis of GOG 182, Gynecol. Oncol. 130 (2013) 487–492.

[8] S.J. Chang, M. Hodeib, J. Chang, R.E. Bristow, Survival impact of complete cytoreduction to no gross residual disease for advanced-stage ovarian cancer: a meta-analysis, Gynecol. Oncol. 130 (2013) 493–498.

[9] S.J. Chang, R.E. Bristow, Evolution of surgical treatment paradigms for advanced stage ovarian cancer: redefining "optimal" residual disease, Gynecol. Oncol. 125 (2012) 483–492.

[10] J.N. Barlin, K.C. Long, E.J. Tanner, G.J. Gardner, M.M. Leitao Jr., D.A. Levine, et al., Optimal (≤1 cm) but visible residual disease: is extensive debulking warranted? Gynecol. Oncol. 130 (2013) 284–288.

[11] N.S. Horowitz, A. Miller, B. Rungruang, D.S. Richard, N. Rodriguez, M.A. Bookman, et al., Does aggressive surgery improve outcomes? Interaction between preoperative disease burden and complex surgery in patients with advanced stage ovarian cancer: an analysis of GOG 182, J. Clin. Oncol. 33 (2015) 937–943.

[12] I. Vergote, C.G. Tropé, F. Amant, G.B. Kristensen, T. Ehlen, N. Johnson, et al., Neoadjuvant chemotherapy or primary surgery in stage IIIC or IV ovarian cancer, N. Engl. J. Med. 363 (2010) 943–953.

[13] N.S. Horowitz, A. Miller, B. Rungruang, S. Richard, N. Rodriguez, M.A. Bookman, et al., Miles of agreement and millimeters of controversy, J. Clin. Oncol. 33 (2015) 3521–3522.

[14] A.M. Nick, R.L. Coleman, P.T. Ramirez, A.K. Sood, A framework for a personalized surgical approach to ovarian cancer, Nat. Rev. Clin. Oncol. 12 (2015) 239–245.

[15] M.A. Bookman, M.F. Brady, W.P. McGuire, P.G. Harper, D.S. Alberts, M. Friedlander, et al., Evaluation of new platinum-based treatment regimens in advanced-stage ovarian cancer: a phase III trial of the gynecologic cancer intergroup, J. Clin. Oncol. 27 (2009) 1419–1425.

[16] D.G. Altman, Y. Vergouwe, P. Royston, K.G. Moons, Prognosis and prognostic research: validating a prognostic model, BMJ 338 (2009) b605.

[17] Harrell FE. Springer Series in Statistics: Regression Modeling Strategies (Corrected ed.) 2010; New York, NY: Springer New York. Pg 247.

[18] W.J. Youden, Index for rating diagnostic tests, Cancer 13 (1950) 32–35.

[19] F.E. Harrell Jr., K.L. Lee, R.M. Califf, D.B. Pryor, R.A. Rosati, Regression modeling strategies for improved prognostic prediction, Stat. Med. 3 (1984) 143–152.

[20] F.E. Harrell Jr., K.L. Lee, D.B. Mark, Multivariable prognostic models: issues in developing models, evaluating assumptions, and adequacy, and measuring and reducing errors, Stat. Med. 15 (1996) 361–387.

[21] R.E. Bristow, L.R. Duska, N.C. Lambrou, E.K. Fishman, M.J. O'Neill, E.L. Trimble, et al., A model for predicting surgical outcome in patients with advanced epithelial ovarian carcinoma using computed tomography, Cancer 89 (2000) 1532–1540.

[22] A.E. Axtell, M.H. Lee, R.E. Bristow, S.C. Dowdy, W.A. Cliby, S. Raman, et al., Multi-institutional reciprocal validation study of computed tomography predictors of suboptimal primary cytoreduction in patients with advanced ovarian cancer, J. Clin. Oncol. 25 (2007) 384–389.

[23] O. Gemer, M. Gdalevich, M. Ravid, B. Piura, A. Rabinovich, T. Gasper, et al., A multicenter validation of computerized tomography models as predictors of non-optimal primary cytoreduction of advanced epithelial ovarian cancer, Eur. J. Surg. Oncol. 35 (2009) 1109–1112.

[24] S. Kang, S.Y. Park, To predict or not to predict? The dilemma of predicting the risk of suboptimal cytoreduction in ovarian cancer, Ann. Oncol. 22 (Suppl. 8) (2011) viii23–viii28 Dec.

[25] S.C. Dowdy, S.A. Mullany, K.R. Brandt, B.J. Huppert, W.A. Cliby, The utility of computed tomography scans in predicting suboptimal cytoreductive surgery in women with advanced ovarian carcinoma, Cancer 101 (2004) 346–352.

[26] G. Ferrandina, G. Sallustio, A. Fagotti, G. Vizzielli, E. Cucci, et al., Role of CT scan-based and clinical evaluation in the preoperative prediction of optimal cytoreduction in advanced ovarian cancer: a prospective trial, Br. J. Cancer 101 (2009) 1066–1073.

[27] I.J. Rutten, R. van de Laar, R.F. Kruitwagen, F.C. Bakers, M.J. Ploegmakers, T.W. Pappot, et al., Prediction of incomplete primary debulking surgery in patients with advanced ovarian cancer: an external validation study of three models using computed tomography, Gynecol. Oncol. 140 (2016) 22–28.

[28] R.S. Suidan, P.T. Ramirez, D.M. Sarasin, J.B. Teicher, S. Milonov, R.B. Iyer, et al., A multi-center prospective trial evaluating the ability of preoperative computed tomography scan and serum CA125 to predict suboptimal cytoreduction at primary debulking surgery for advanced ovarian, fallopian tube, and peritoneal cancer, Gynecol. Oncol. 134 (2014) 455–461.

[29] J.M. Janco, G. Glaser, B. Kim, M.E. McGree, A.L. Weaver, W.A. Cliby, et al., Development of a prediction model for residual disease in newly diagnosed advanced ovarian cancer, Gynecol. Oncol. 138 (2015) 70–77.

[30] S.M. Ko, D.C. Jung, Oh. YT, S.H. Kim, S. Kang, Discrepancy between preoperative imaging and surgical findings in patients with advanced ovarian cancer: a study of initial false-negative lesions, J. Comput. Assist. Tomogr. 37 (2013) 568–571.

[31] A. Fagotti, G. Vizzielli, F. Fanfani, et al., Definition of a dynamic laparoscopic model for the prediction of incomplete cytoreduction in advanced epithelial ovarian cancer: proof of a concept, Gynecol. Oncol. 139 (2015) 5–9.

[32] M.J. Rutten, K.N. Garenstroom, T. Van Gorp, et al., Laparoscopy to predict the result of primary cytoreductive surgery in advanced ovarian cancer patients (LapOvCa-trial): a multicenter randomized controlled study, BMC Cancer 12 (2012) 31.

[33] M. Riester, W. Wei, L. Waldron, A.C. Culhane, L. Trippa, E. Oliva, et al., Risk prediction for late-stage ovarian cancer by meta-analysis of 1525 patient samples, J. Natl. Cancer Inst. 106 (5) (2014).