External Validation of a Laparoscopic-based Score to Evaluate Resectability for Patients with Advanced Ovarian Cancer Undergoing Interval Debulking Surgery

E. CHEREAU1, V. LAVOUE4, M. BALLESTER1, C. COUTANT1, F. SELLÉ2, A. CORTEZ3, E. DARAÏ1, J. LEVEQUE4 and R. ROUZIER1

Departments of 1Gynecology-Obstetrics, 2Medical Oncology and 3Pathology, Tenon Hospital, Paris, France; 4Department of Gynecology-Obstetrics, CHU Anne de Bretagne, Rennes, France

Abstract. Aim: To evaluate the relevance of laparoscopic index of Fagotti et al during staging laparoscopy (S-LPS) to predict optimal cytoreduction during interval debulking surgery (IDS) after neoadjuvant chemotherapy for ovarian cancer. Patients and Methods: Fifty-two patients with stage III–IV ovarian cancer were retrospectively analyzed. We evaluated discrimination with a receiver operating characteristic (ROC) curve analysis and calibration of Fagotti et al’s model among our population and compared this performance with their data. Results: A score >4 was associated with optimal resection with sensitivity and positive predictive value (PPV) of 95% and 82% respectively. The ROC curve analysis gave an area under the curve (AUC) of 0.72 (95% confidence interval (CI) 0.65-0.80) for our population compared to 0.88 (95% CI 0.84-0.91) in Fagotti et al’s population. Percentages predicted in our population were unsatisfactory (p<0.01), illustrating the different rates of optimal cytoreduction between the centers (average error of 25%). Conclusion: The laparoscopic index of Fagotti et al is relevant in prediction of optimal cytoreduction among women undergoing IDS.

The largest randomized screening trial comparing primary cytoreductive surgery to interval debulking surgery (IDS) for ovarian cancer after neoadjuvant chemotherapy (NACT) has reported no difference in overall survival between these two strategies (1). Optimal surgery appears to be a major prognosis factor for survival also in the case of IDS (2-4). However, there is a general consensus to triage only responsive patients (complete/partial response) to interval debulking surgery (IDS).

A selection as correct as possible of patients in which optimal cytoreduction is actually achievable after NACT would avoid unnecessary explorative laparotomies, complications, and delay in starting new drug regimens, if necessary.

In this context, only one study based on CA125 has been published but none based on imaging. Laparoscopy could help to evaluate resectability with limited morbidity when compared to laparotomic evaluation and to select the right candidates and the right time for performing optimal IDS.

While several pre-operative scoring systems have been developed to predict likelihood optimal cytoreduction at the time of primary surgery (5, 6), only one score, reported by Fagotti et al (7), considered IDS after NACT. Fagotti et al proposed a laparoscopic index based on four significant parameters (mesenteric retraction, bowel infiltration, stomach infiltration and liver metastasis) to identify patients who would be candidates for successful IDS.

However, as no external validation of this laparoscopic index has been published to date, we were prompted to evaluate the usefulness of this laparoscopy-based score to identify good candidates with advanced ovarian cancer for optimal IDS.

Patients and Methods

From 2002 to 2010, 70 patients were candidates for IDS after NACT for advanced ovarian cancer. All data were entered prospectively into a database and analyzed retrospectively for this study. All patients underwent initial laparoscopy at the time of ovarian cancer diagnosis. Among these patients, 52 underwent staging laparoscopy (S-LPS) at the time of IDS and were thus eligible for this study. Eighteen patients were excluded for IDS because of the absence of a response based on CA125 serum level and computerized tomography (CT) scan results. Individual records of all patients were reviewed and analyzed. Patient and tumor characteristics, number of NACT courses, S-LPS and intraoperative findings were analyzed. Resection was defined as optimal (residual disease <1 cm) or non-optimal according to Fagotti et al (7).

All women gave informed written consent to therapeutic procedures and to the data analysis related to their malignancy in...
accordance with the Institutional Review Board’s institutional guidelines and the Declaration of Helsinki.

We evaluated the score described in a recent article published by Fagotti et al (7). For each value of described predictive index value (PIV) we compared the percentage of patients with optimal resection in each group. Moreover, in accordance with this study, we compared sensitivity, specificity positive predictive value (PPV) and negative predictive value (NPV) for the threshold reported (PIV>4).

For each predictive parameter described by Fagotti et al (7) (namely omental cake, peritoneal carcinosis, diaphragmatic carcinosis, mesenteric retraction, bowel infiltration, stomach infiltration and liver metastases) and for one additional parameter, ascites, we calculated sensitivity, specificity, PPV and NPV to predict complete resection using IDS in our population. As described elsewhere (7), each predictive index parameter was given a value from 0 to 2 (0 when absent; 1 when specificity ≥5%, PPV ≥50% and NPV ≥50% and 2 when the summed overall accuracy >60%).

The model performance was quantified with respect to discrimination and calibration. Discrimination (i.e., whether the relative ranking of individual predictions was in the correct order) was quantified with the area under (AUC) the receiver operating characteristic (ROC) curve. Calibration (i.e., agreement between observed outcome frequencies and predicted probabilities) was studied with graphical representations of the relationship between the observed outcome frequencies and the predicted probabilities (calibration curves). We also evaluated average error (E ave) between predictions and observations obtained from the calibration curves. The clinical significance of calibration is high; it reflects the accuracy of individual predictions. In this study, we evaluated discrimination of the PIV for complete resection among our population and Fagotti et al’s patients. These criteria were not evaluated in Fagotti et al’s paper. We calculated these performance indicators from individual data in Fagotti’s series to determine overfitting and exportability of their score.

We compared discrimination of the model of Fagotti et al’s applied to their population and to our population to an optimal model. This optimal logistic regression model was developed from our patient cohort. The performance of the optimal model can be regarded as the maximum that can be expected for a model based on external data. The aim of this model was not to identify independent predictive factors of resectability but to determine the maximal informativeness of the combination of parameters.

We calculated the number of patients to include to validate the correlation between scores and resectability. To demonstrate that there is a linear correlation with an r>0.4 between prediction and outcome, with alpha=0.05, and beta=0.2, the inclusion of 47 patients was necessary.

All analyses were performed using the R software package with the Design, Hmisc, Rpart and Verification libraries (http://lib.stat.cmu.edu/R/CRAN/).

Results

Epidemiological and surgical characteristics of the population.
The clinicopathological and surgical characteristics of the 52 patients are reported in Table I.

Validation of laparoscopic score for IDS. Significant S-LPS predictive factors for non successful optimal resection were omental cake (p=0.008) peritoneal carcinosis (p=0.02) and
mesenteric retraction ($p=0.007$) in univariate analysis. Other laparoscopic parameters were non-significant (Table II).

When we compared the distribution of PIV in our population and in Fagotti et al’s population (Figure 1), we highlighted a very similar distribution of the two populations. No differences were observed for PIV scores. For the rate of optimal resection, except for PIV scores 0 and 2, there was a significant difference between the populations with more optimal resection in our population ($p=0.002$ for PIV 4 and $p=0.03$ for PIV 6) (Figure 2).
Sensitivity, specificity, PPV, NPV and accuracy for each predictive parameter calculated in our population are given in Table II.

With a threshold PIV of 4 for our patients, sensitivity and PPV were 95% and 82%, respectively. These results were in accordance with the results of Fagotti et al’s. The discriminating performance of the S-LPS score by ROC curve analysis is given in Figure 3. The AUC calculated with the patient data from the original (7) paper was 0.88 (95% confidence interval (CI) =0.84-0.91). In our population, the AUC was 0.72 (95% CI=0.65-0.80).

In multivariate analysis, none of these parameters were significant, probably due to the size of the sample, but the area under the ROC curve of the logistic regression model was satisfactory (AUC=0.88, 95% CI=0.83-0.94) (Figure 3). This model represents the optimal logistic regression model. The performance of this model can be regarded as the maximum that can be expected for a model based on external data. The lack of difference between the AUC of Fagotti et al’s model compared to the optimal logistic regression model (overlapping of confidence intervals) supported the equivalence of these models.

Calibration plots are given in Figure 4. The PIV of Fagotti et al’s had an excellent calibration (p=0.9, E ave=0.001) whereas for our population, percentages predicted were unsatisfactory with a higher proportion of optimal resection than expected (p<0.001, E ave=0.25). The E ave of 25% could be explained by the different rates of optimal cytoreduction between the centers (83% in our population and 60% in Fagotti et al’s population, p=0.03).

**Discussion**

The present study confirms the relevance of Fagotti et al’s laparoscopic index value (7) to the prediction of resectability for IDS of patients with advanced ovarian cancer.

Because complete cytoreduction is the cornerstone of treatment for patients with advanced ovarian cancer and has a direct relationship with survival, maximal cytoreductive surgery is the standard for these patients (2). When the extent of the disease was considered non-resectable during primary staging laparoscopy, patients underwent NACT and were eligible for IDS (8). Complete cytoreduction is still mandatory during IDS, and staging laparoscopy could be helpful to select patients eligible for IDS.

To date, IDS after NACT represents a valuable alternative to primary cytoreduction in the case of an unresectable disease (3). However, recent debated data suggested a similar overall survival and a decreased morbidity for patients who underwent IDS instead of primary cytoreduction (1), which raises the alternative of selection optimal cytoreductive surgery at the time of IDS more often.

**Preoperative evaluation of resectability.** Preoperative evaluation of resectability remains a challenge for clinicians in advanced ovarian cancer. Predicting the possibility of performing successful surgery has been studied for primary cytoreduction using CA125 levels (9-13), radiological imaging modalities (14-17) and clinical findings (14, 15). To date a universally applicable model that can predict in which patients optimal primary cytoreduction will be achieved remains elusive. Laparoscopic-based models were the most promising tools in predicting complete resection (12). For IDS, only one study has correlated the variation of the CA125 level with optimal cytoreduction (18), and no study has been published on using radiological or clinical findings related to cytoreduction. Therefore, staging laparoscopy and the laparoscopic index described by Fagotti et al (7) seems to be promising to discriminate patients applicable for successful IDS.

**Justification of evaluation laparoscopy at the time of IDS.** Bristow et al (4) reported the absence of benefits to survival by adding further chemotherapy cycles compared to surgical effort. In this context, forward laparoscopic evaluation is mandatory to reduce morbidity compared to explorative

### Table II. Predictive index parameters included in the score of Fagotti et al (7) that were applied to our population and the univariate analysis of the laparoscopic predictive factors for optimal resection for interval debulking surgery.

<table>
<thead>
<tr>
<th>Index Parameter</th>
<th>p-Value</th>
<th>Sensitivity (%) (range)</th>
<th>Specificity (%) (range)</th>
<th>PPV (%) (range)</th>
<th>NPV (%) (range)</th>
<th>Accuracy (%) (range)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Omental cake</td>
<td>0.008</td>
<td>100% (84-100)</td>
<td>28% (18-28)</td>
<td>46% (39-46)</td>
<td>100% (66-100)</td>
<td>56% (44-56)</td>
</tr>
<tr>
<td>Peritoneal carcinosis</td>
<td>0.02</td>
<td>100% (82-100)</td>
<td>26% (17-26)</td>
<td>40% (32-40)</td>
<td>100% (66-100)</td>
<td>50% (38-50)</td>
</tr>
<tr>
<td>Diaphragmatic carcinosis</td>
<td>0.27</td>
<td>89% (78-97)</td>
<td>25% (12-34)</td>
<td>58% (51-63)</td>
<td>67% (33-90)</td>
<td>60% (48-68)</td>
</tr>
<tr>
<td>Mesenteric retraction</td>
<td>0.007</td>
<td>92% (84-98)</td>
<td>43% (21-58)</td>
<td>81% (74-86)</td>
<td>67% (33-90)</td>
<td>79% (67-87)</td>
</tr>
<tr>
<td>Bowel infiltration</td>
<td>0.71</td>
<td>84% (76-93)</td>
<td>20% (7-34)</td>
<td>63% (57-70)</td>
<td>45% (16-76)</td>
<td>60% (50-70)</td>
</tr>
<tr>
<td>Stomach infiltration</td>
<td>0.32</td>
<td>84% (82-86)</td>
<td>50% (3-97)</td>
<td>98% (95-99)</td>
<td>11% (1-22)</td>
<td>83% (79-86)</td>
</tr>
<tr>
<td>Liver metastases</td>
<td>1</td>
<td>82% (82-85)</td>
<td>0% (0-79)</td>
<td>95% (95-99)</td>
<td>0% (0-18)</td>
<td>79% (79-85)</td>
</tr>
</tbody>
</table>

PPV: Positive predictive value; NPV: negative predictive value.
laparotomy. Because the relationship between CA125 level, ascites and peritoneal carcinoma has been demonstrated (19), the timing of S-LPS could be decided upon a decrease of CA125 rate, and the laparoscopic index value could help surgeons to proceed with IDS or to postpone IDS until after one or two more cycles of chemotherapy.

Although Fagotti et al’s laparoscopic index is relevant to our population with an AUC of 0.77, in this population only two parameters were sufficiently relevant to be attributed a score of 2 points: stomach infiltration and liver metastases. The lack of difference between the AUC of Fagotti et al’s model (AUC=0.88, 95% CI=0.84-0.91) compared to the optimal logistic regression model (AUC=0.88, 95% CI=0.83-0.94) supports the equivalence of these models and therefore the pertinence of the score, at least in terms of discrimination. However, this score needs to evolve because optimal resection was defined as <1 cm residual disease by Fagotti et al while the true definition is no visible disease.

Expert centers. Another major point when debating optimal or complete cytoreduction for advanced ovarian cancer is the management of patients in expert centers. Indeed, publications increasingly highlight the importance not only of surgical skills and experience, but also of the high quality of surgical care delivered by a multidisciplinary team with oncology nurses, anesthesiologists, pathologists, dieticians and physiotherapists for the postoperative management of patients with advanced ovarian cancer (20). Many retrospectives studies have pointed out that debulking operations performed by gynecologic oncologists result in the best outcomes (including survival) compared to surgery performed by general gynecologists or general surgeons (21-23). Centralization of these patients in expert centers is therefore recommended. The calibration results need to be correlated to the rate of optimal resection among the patients of the studied population. Indeed, the rate of complete or optimal cytoreduction may vary between two centers or in one center between two periods. Chi et al reported a rate of optimal cytoreduction from 45% to 80% in 2000 and 2009, respectively (24, 25). In our study, the difference between the two calibration curves was explained by the different rates of optimal cytoreduction between the centers. This difference was somewhat well illustrated with the value of E ave of 25%. This finding is relevant because discrimination of the score was satisfactory: patients with the worst prediction indeed had the worst outcome.

Conclusion

In conclusion, the laparoscopic index of Fagotti et al is relevant in the prediction of optimal cytoreduction among women undergoing IDS for advanced ovarian cancer in our population. The remaining challenge for clinicians is to evaluate the timing of IDS by weighing a limited number of chemotherapy cycles with optimal cytoreduction. In their article, Fagotti et al suggested a new algorithm for the management of patients with advanced ovarian cancer based upon the PIV evaluated during S-LPS at a patient’s initial surgery and before IDS. This algorithm needs to be evaluated ideally with a prospective multicenter study. However, other tools such as the CA125 level, CT scan or PET/CT could also play a part in the decision to perform IDS, especially for the timing of IDS.

References


Received October 11, 2011
Revised November 15, 2011
Accepted November 16, 2011