Gynecologic Oncology 140 (2016) 22–28

Contents lists available at ScienceDirect

Gynecologic Oncology

journal homepage: www.elsevier.com/locate/ygyno

# Prediction of incomplete primary debulking surgery in patients with advanced ovarian cancer: An external validation study of three models using computed tomography

Iris J.G. Rutten [a,b,*], Rafli van de Laar [a,b], Roy F.P.M. Kruitwagen [a,b], Frans C.H. Bakers [c], Marieke J.M. Ploegmakers [d], Teun W.F. Pappot [e], Regina G.H. Beets-Tan [b,c,f], Leon F.A.G. Massuger [g], Petra L.M. Zusterzeel [g], Toon Van Gorp [a,b]

[a] Department of Obstetrics and Gynecology, Maastricht University Medical Center, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands
[b] GROW School for Oncology and Developmental Biology, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands
[c] Department of Radiology, Maastricht University Medical Center, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands
[d] Department of Radiology and Nuclear Medicine, Radboud University Nijmegen Medical Center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands
[e] Department of Radiology, Rijnstate Hospital, P.O. Box 9555, 6800 TA Arnhem, The Netherlands
[f] Department of Radiology, Netherlands Cancer Institute, P.O. Box 90203, 1006 BE Amsterdam, The Netherlands
[g] Department of Obstetrics and Gynecology, Radboud University Nijmegen Medical Center, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands

## HIGHLIGHTS

- Ovarian cancer models that predict incomplete primary debulking have limited accuracy
- A radiologist's subjective assessment seems as successful as using a prediction model
- Prediction models should be interpreted with caution in clinical decision-making

## ARTICLE INFO

*Article history:*
Received 3 November 2015
Received in revised form 13 November 2015
Accepted 18 November 2015
Available online 24 November 2015

*Keywords:*
External validation
Prediction models
Epithelial ovarian cancer
Complete primary cytoreductive surgery
Computed tomography

## ABSTRACT

*Objective.* To test the ability of three prospectively developed computed tomography (CT) models to predict incomplete primary debulking surgery in patients with advanced (International Federation of Gynecology and Obstetrics stages III–IV) ovarian cancer.

*Methods.* Three prediction models to predict incomplete surgery (any tumor residual > 1 cm in diameter) previously published by Ferrandina (models A and B) and by Gerestein were applied to a validation cohort consisting of 151 patients with advanced epithelial ovarian cancer. All patients were treated with primary debulking surgery in the Eastern part of the Netherlands between 2000 and 2009 and data were retrospectively collected. Three individual readers evaluated the radiographic parameters and gave a subjective assessment. Using the predicted probabilities from the models, the area under the curve (AUC) was calculated which represents the discriminative ability of the model.

*Results.* The AUC of the Ferrandina models was 0.56, 0.59 and 0.59 in model A, and 0.55, 0.60 and 0.59 in model B for readers 1, 2 and 3, respectively. The AUC of Gerestein's model was 0.69, 0.61 and 0.69 for readers 1, 2 and 3, respectively. AUC values of 0.69 and 0.63 for reader 1 and 3 were found for subjective assessment.

*Conclusions.* Models to predict incomplete surgery in advanced ovarian cancer have limited predictive ability and their reproducibility is questionable. Subjective assessment seems as successful as applying predictive models. Present prediction models are not reliable enough to be used in clinical decision-making and should be interpreted with caution.



## 1. Introduction

Ovarian cancer has a high mortality which translates to a 5-year survival of 38–46% in Europe and the United States [1,2]. These survival rates drop from 73-90% in early stage epithelial ovarian cancer to 17–39% in advanced stages (International Federation of Gynecology and

* Corresponding author at: Maastricht University Medical Centre, P.O. Box 5800, 6202 AZ Maastricht, The Netherlands.
*E-mail address:* iris.rutten@mumc.nl (I.J.G. Rutten).

http://dx.doi.org/10.1016/j.ygyno.2015.11.022
0090-8258/

Obstetrics, (FIGO) stage III and IV) [2]. The most important predictor of survival in advanced stage disease is the amount of residual tumor after cytoreductive surgery [3–5]. Maximal cytoreduction has been the mainstay of advanced ovarian cancer treatment for many years now since Griffiths et al. established the inverse relationship between the amount of residual disease and overall survival in 1975 [6]. Multiple studies have consistently confirmed these findings and more recent data demonstrate a significant survival gain for complete resection to no macroscopic residual disease in comparison to optimal resection to a tumor residual of ≤1 cm [4,5]. Incomplete resection (tumor residual >1 cm) has little beneficial effects for survival while it can cause substantial perioperative morbidity. It is widely agreed that surgery should be avoided when incomplete resection is expected. Neoadjuvant chemotherapy followed by interval debulking surgery provides a good treatment alternative with similar survival rates when primary surgery is deemed not feasible or impossible [3,7].

Whether or not complete tumor resection can be achieved depends on patient related factors such as age and comorbidity, the extent and location of disease and the skill and experience of the operating surgeon [8]. Numerous prediction models have been developed in order to select patients with advanced ovarian cancer who are unlikely to benefit from primary debulking surgery [9–24]. They have incorporated clinical features and computed tomography (CT) features in different combinations. Frequently used CT predictors are peritoneal thickening, ascites, suprarenal lymphadenopathy, involvement of the bowel mesentery, diaphragmatic involvement and liver involvement [25]. Recurring clinical features included in the prediction models are serum Ca-125 levels and World Health Organization Performance Status (WHO-PS) [25].

Agreement between the developed models is limited and the great variety of combinations of predictors found to be associated with ovarian cancer resectability makes the reproducibility and clinical applicability of the models questionable. Many of the earlier models defined optimal debulking as <2 cm residual disease instead of ≤1 cm [9–12] or included early stage ovarian cancer [10,11,13,14]. Furthermore most studies, including many of the newer models, were developed with retrospective data from small populations and were only validated internally or not validated at all. External validation in a population other than the study population is mandatory before a model can be implemented in daily practice. Two studies tested the accuracy of the models developed by Bristow et al. [15] and Dowdy et al. [16] and found a substantial decline in predictive performance when the prediction models were tested on a different population [26,27]. Since then more models have been developed but few have been tested for external validity [17–24].

The goal of this study is to externally validate three prospectively developed CT models published by Ferrandina et al. [17] and by Gerestein et al. [18] to predict incomplete debulking surgery in advanced stage ovarian cancer patients undergoing primary debulking surgery. We aim to apply the three models on our validation cohort and test their accuracy when interpreted by different readers.

## 2. Methods

### 2.1. Selection of patients

A Medical Ethical Committee approved this study and waived the requirement for obtaining informed consent. All patients diagnosed with primary advanced (FIGO stage III and IV) epithelial ovarian cancer that were treated in the Eastern part of the Netherlands (one specialized academic centre and six regional hospitals) between 1 January 2000 and 1 January 2009 were identified. Patients were included if they fulfilled the following criteria: 1) patients were treated with primary debulking surgery and 2) a preoperative CT scan was performed within three months before surgical treatment. Data on patient characteristics, diagnosis and therapy were retrospectively collected from local hospital records. Collection of imaging data was also done retrospectively. Patients were excluded in case of missing data regarding surgical outcome or when the preoperative CT scan could not be retrieved. None of the included patients were used in the development of any of the models under evaluation and as such, this population represents an independent external validation cohort..

### 2.2. Prediction models

The prediction models under evaluation in this study were constructed by Ferrandina et al. in 2009 and Gerestein et al. in 2011 [17,18].

Ferrandina described two approaches to predict incomplete primary debulking in advanced ovarian cancer: models A and B. In model A, each radiographic or clinical parameter that showed a specificity ≥75%, a positive predictive value (PPV) ≥50%, and a negative predictive value (NPV) ≥50% in predicting surgical outcome was assigned a score of 1 point. In addition, if a parameter showed an overall accuracy ≥60%, it was assigned a second point. In total, 4 radiographic parameters were assigned 1 or 2 points and the clinical parameter WHO-PS was assigned 2 points in the final model as well (see Table 1). A cut-off of more than 5 points out of a possible 8 point score was used to predict incomplete debulking. In model B, the relationship between each possible predictor with surgical outcome was tested in univariable logistic regression analysis at a significance level of 5%. Significant predictors were included in a multivariable logistic regression model using a stepwise elimination method. All predictors achieving a p-value <0.10 were assigned a score of 1 point. The final model included 4 radiographic parameters and 1 clinical parameter. The parameters are specified in Table 1. A cut-off of >3 points out of a possible 5 point score was used to predict incomplete debulking. Reported Area Under the Curve (AUC) values were 0.81 for model A and 0.82 for model B.

Gerestein performed univariable regression analysis at a significance level of 30% to select parameters that were associated with surgical outcome in advanced ovarian cancer patients. The selected parameters were included in a multivariable Cox regression model using a backward elimination method. Two radiographic predictors and one clinical predictor were included in the final prediction model (see Table 1). The model was internally validated by a bootstrap method which resulted in a c-index of 0.67.

### 2.3. Data analysis

CT scans from patients in our external validation cohort were assessed in a retrospective manner by three radiologists from different hospitals (two academic centers and one regional hospital). All readers had a special interest in gynecologic oncology imaging and had eight years (FB) or four years (MP and TP) of experience in this field. For their assessment the readers individually scored the radiographic parameters included in the original prediction models by Ferrandina and Gerestein blinded to any clinical information and surgical outcome. Two of the three radiologists (FB, TP) were also asked to give their own subjective judgment whether incomplete surgery was expected. Clinical parameters were retrieved from hospital records. All patients underwent laparotomy aimed at achieving maximal cytoreduction according to prevailing guidelines. Incomplete tumor resection was defined as any tumor residual greater than 1 cm in diameter. Surgical findings and histopathological confirmation were used as reference standard.

### 2.4. Statistical analysis

First, Ferrandina's prediction models A and B were applied to the validation cohort. Using the described cut-off values of 5 in model A and 3 in model B, the sensitivity, specificity, PPV, NPV, pre-test probability and post-test probability were calculated. A true positive (TP) was defined as a patient in whom an incomplete debulking was correctly predicted (score > 5 or >3). A true negative (TN) was defined as a patient in