A prognostic model of immunohistochemistry biomarkers for high-grade serous ovarian cancer

by

Zhuxuan Fu

BS, Shandong University at Weihai, 2014

MPH, Tufts University, 2016

Submitted to the Graduate Faculty of the

Department of Biostatistics

Graduate School of Public Health partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2020

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

By

Zhuxuan Fu

It was defended on

October 19, 2020

and approved by

Francesmary Modugno, MS, PhD, MPH,
Womens Cancer Research Center, Magee-Womens Research Institute and Hillman Cancer Center.
Division of Gynecologic Oncology, Department of Obstetrics, Gynecology and Reproductive Sciences. In. University of Pittsburgh School of Medicine.

Ada O Youk, Ph.D.,
Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh.

Thesis Advisor or Dissertation Director: Lu Tang, Ph.D.,
Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh.

A prognostic model of immunohistochemistry biomarkers for high-grade serous ovarian cancer

Zhuxuan Fu, MS

University of Pittsburgh, 2020

Abstract

Background: Ovarian cancer is the most lethal gynecologic cancer in the United States. High-grade serous ovarian cancer (HGSOC) accounts for 70%-90% of all ovarian cancer death. It is crucial to identify efficient prognostic biomarkers to inform treatment decision making.

Method: Tissue microarrays and clinical data were obtained from patients diagnosed with invasive HGSOC enrolled in studies participating in the Ovarian Tumor Tissue Analysis consortium. Cox proportional hazard regression analysis (CoxPHR) with *lasso* penalty was performed to select the most important variables related to overall survival (OS) from clinical prognostic data and 9 immunohistochemistry (IHC) biomarkers of interest, MyD88, TLR4, FOLT1, CD8+ tumor-infiltrating lymphocytes (CD8+ TILs), p16, PTEN, progesterone receptor (PR), estrogen receptor (ER) and androgen receptor (AR) using a training set of 254 patients with all 9 IHC data. The external validation was conducted using the test set of 1563 patients with data of the selected IHC biomarker by *lasso*. Hazard ratios (HRs) and 95% confidence intervals (CIs) of the selected variables were estimated from the CoxPHR. Kaplan-Meier curves were used to visually compare survival across the selected variables. A nomogram was generated to estimate the 2-year and 3-year survival.

Results: The median OS time of the training set was 5.04 years (95% CI 4.36- 5.99 years). The selected variables from CoxPHR with *lasso* penalty include age at diagnosis, stage,

debulking status, AR, TLR4, CD8+ TILs, and p16. The median OS of the test set is 3.41 years (95% CI 3.21-3.63 years). The cases in the test set are at a more advanced stage. C-index from the prediction model fitting in the test set is 0.63. In the prediction model, CD8+ is inversely associated with the hazard of death (P for trend = 0.0011).

Conclusion: The CoxPHR model with *lasso* penalty identifies four IHC biomarkers, AR, TLR4, CD8+, and p16, along with age at diagnosis, tumor stage, and debulking status, as prognostic factors for HGSOC survival. Further study containing more IHC candidates and clinical variables, such as chemotherapy response, and using continuous IHC scores should be performed to increase the accuracy of the prediction model for HGSOC survival.

Table of Contents

List of Tables

List of Figures

Preface

The thesis is submitted for the degree of Master of Science at Biostatic Department, Graduate School of Public Health, University of Pittsburgh.

The research was conducted under the supervision of Dr. Tang, Lu. Without his kindly assistance, I would not have been able to conduct this analysis. He raised many excellent points in our discussion and considerably improve the quality of my thesis.

I also wish to thank Dr. Modugno Francesmary. I enjoyed the time in the past three years working with her. Dr. Modugno helps me improve my knowledge of ovarian cancer and the skills of statistical analysis and commutation. She also provided me the opportunity of working with other research investigators in a multi-center study. The enthusiasm of her made a strong impression on me and motived me to explore in the field of ovarian cancer. She also supported me emotionally. Every time I felt stressed or I was ready to quit, she talked with me rationally and gave me some great advice.

I would like to thank Dr. Youk, Ada, my advisor in the Biostatistics Department, Dr. Songer, Thomas, my advisor in the Epidemiology Department, and Lori Smith, the student affairs manager in Epidemiology Department. It is not easy to pursue an MS degree and a Ph.D. degree at the same time. With their guidance in every step throughout the process, I can stratify the requirements for both degrees.

Most importantly, I would like to thank my family and friends for the unfailing support. My best friend, Lanting Yang, took care of me carefully and patiently and encouraged me when I broke my leg. I felt not alone when studying and living in this city. This accomplishment would not have been possible without her.

1.0 Background

1.1 High-grade serous ovarian cancer

1.1.1 Statistics of HGSOC

Ovarian cancer is a heavy burden in the United States. About 13,940 women will die from ovarian cancer in 2020 in the U.S, according to the American Cancer Society, which is more than any other gynecologic cancer (American Cancer Society, 2020; Olsen et al., 2013). Epithelial ovary cancer (EOC) accounts for nearly 90% of malignant ovarian cancer (American Cancer Society, 2020). High-grade serous ovarian cancer (HGSOC), accounting for more than 60% of EOC and 70%-90% of all ovarian cancer deaths, is the most deadly histologic subtype of EOC (Bowtell et al., 2015; Gurung, Hung, Morin, & Gilks, 2013; Lauren C. Peres et al., 2019).

There are two systems for staging ovarian cancer, the International Federation of Gynecology and Obstetrics system and the American Joint Committee on Cancer TNM staging system. These systems are the same in that they both use three factors to stage cancer, including the extent of the tumor, the spread to nearby lymph nodes, the spread to distant sites. Higher numbers indicate more advanced cancer. The Surveillance, Epidemiology, and End Results registry (SEER) program of the National Cancer Institute (NCI) use a simplified approach to stage ovarian cancer as localized (roughly Stage I disease), regional (encompassing Stage II and III disease) and distant (Stage IV disease) (National Cancer Institute, March 9, 2015). HGSOC primarily arises from the fallopian tube and then spread to the ovary (J. Kim et al., 2018; Kurman & Shih Ie, 2010; Salvador et al., 2009). More than 75% of HGSOC cases are diagnosed at an

advanced stage (stage III/IV) (L. C. Peres et al., 2019; Seidman, Zhao, & Yemelyanova, 2011). According to the SEER registry data, the prognosis for patients diagnosed in the advanced stages is very poor (Baldwin et al., 2012; L. C. Peres et al., 2019). The data from 40,692 EOC patients in SEER from 1995-2007 indicates that the 5-year survival of EOC is 17% for stage IV, and the 10-year survival is only 8% (Baldwin et al., 2012). Another study including 17,837 HGSOC patients in SEER from 2004-2014 estimates the 5-year survival of HGSOC to be 32% for the distant stage, and the 10-year survival to be further reduced by half (L. C. Peres et al., 2019).

1.1.2 Clinical factors related to survival

Many clinical factors affect the survival of HGSOC patients in addition to the tumor stage. First, optimal surgical debulking, which is defined as no visible cancer or no tumor larger than 1 cm after debulking surgery (American Cancer Society), is associated with better survival for EOC patients and HGSOC patients (Aluloski et al., 2017; Chang, Bristow, & Ryu, 2012; Chang, Hodeib, Chang, & Bristow, 2013; Dao et al., 2016). Second, different types of treatment might be associated with HGSOC survival, but the effect is unclear. There are two typical treatments for EOC patients. One is neoadjuvant chemotherapy, which involves 3 cycles of chemotherapy treatment before debulking surgery and then an additional 3-5 cycles of chemotherapy after debulking surgery (Inciura et al., 2006). The other is adjuvant chemotherapy, which involves 6-8 cycles of chemotherapy after surgery (Inciura et al., 2006). Although there is no study estimating the effect of different treatments on HGSOC survival, several studies estimated the effect on EOC survival. One study, including 925 patients with early-stage EOC, reveals that adjuvant chemotherapy is associated with better overall survival (Carlson, 2003). Nevertheless, there is no difference found between two types of treatments on survival in another study of 574 patients with

advanced OC (Inciura et al., 2006). Compared to the types of treatment, the response to chemotherapy is more important to HGSOC survival. Platinum sensitivity, which is defined as the time between the last dose of platinum-based chemotherapy and the evidence of cancer progression equal to or larger than 6 months (Matulonis et al., 2016), produces a better survival for HGSOC patients (Aluloski et al., 2017; Dao et al., 2016). Studies also reveal that BRCA mutation is not only a risk factor of HGSOC but also a factor related to HGSOC survival (Bell et al., 2011; Dao et al., 2016; S. I. Kim et al., 2019). Patients with germline BRCA mutations had better prognosis (Dao et al., 2016; S. I. Kim et al., 2019).

Cancer antigen 125 (CA125) is the gold standard of biomarkers for diagnosis and prognosis for HGSOC (Akeson et al., 2009; Rein et al., 2011; Salminen et al., 2020). Xu et al. indicate that the nadir CA-125 level, defined as the CA-125 values at two weeks after the evaluation at the start of diagnosis, is an independent predictor for HGSOC overall survival (Xu et al., 2013). In a recent study, Salminen et al. indicate the nadir CA125-STn, instead of CA-125 can predict HGSOC survival (Salminen et al., 2020). The authors in a literature review investigated the association of CA-125 level at different time-point, including prechemotherapy, post-chemotherapy, during chemotherapy, preoperative, postoperative, half-life, with nadir, and ovarian cancer survival (Gupta & Lis, 2009). The authors conclude that the CA-125 level is a significant prognostic factor for ovarian cancer overall survival, although the results were controversial (Gupta & Lis, 2009).

1.1.3 Immunohistochemistry biomarkers of interest

The Ovarian Tumor Tissue Analysis (OTTA) consortium was formed in 2010 to validate prognostic markers by histologic subtypes (Susan J. Ramus et al., 2013). There were 70 study sites from North America, South American, Europe, and Asia participating in OTTA. OTTA have

conducted two types of large-scale research projects, including immunohistochemistry (IHC) and RNA expression. Nine IHC biomarkers associated with EOC survival have been published or under review by OTTA, including myeloid differentiation primary response gene 88 (MyD88) (Block et al., 2018), Toll-like receptor 4 (TLR4) (Block et al., 2018), Folate receptor 1 (FOLR1) (Köbel et al., 2014), CD8+ tumor-infiltrating lymphocytes (CD8+ TILs) (Ovarian Tumor Tissue Analysis et al., 2017), p16 (Rambau et al., 2018), phosphatase and tensin homolog (PTEN), progesterone receptor (PR) (Sieh et al., 2013), estrogen receptor (ER) (Sieh et al., 2013) and androgen receptor (AR). Positive MyD88 is associated with poor HGSOC survival compared to negative MyD88 (Block et al., 2018). Positive CD8+ TILs and strong PR expression are associated with better HGSOC survival (Ovarian Tumor Tissue Analysis et al., 2017; Sieh et al., 2013). CD8+ TILs also have interacted with BRCA mutation on HGSOC survival (Ovarian Tumor Tissue Analysis et al., 2017). Although FOLR1 is not associated with HGSOC survival, there is a significant interaction between FOLR1 and stage on HGSOC survival (Köbel et al., 2014).

There is no published result on AR and PTEN by OTTA. However, a study combining 87 HGSOC and 31 endometrioid, another histology subtype of EOC, finds positive PR and AR are associated with overall survival (Jönsson et al., 2015). Lower expression levels of PTEN are associated with poor prognosis in HGSOC (Bakkar et al., 2015; Filipe C. Martins et al., 2014; Shen, Li, Liu, & Cheng, 2017) because the inactivation of PTEN causes chemotherapy resistance (Nero, Ciccarone, Pietragalla, & Scambia, 2019).

The current study ***aims*** to build a prognostic model for HGSOC overall survival by identifying the most important prognostic biomarkers from 9 candidate IHC biomarkers in OTTA. The model would provide potential opportunities to inform treatment decision making and extend the survival of HGSOC patients.

1.2 Statistical learning algorithms of variable selection for survival

1.2.1 Cox proportional hazard regression

Cox proportional hazard regression analysis (CoxPHR) is usually used to predict survival time from a set of potential predictors. The hazard function is written as

$$h(t) = h_0(t)e^{x_i^t\beta},$$

where $h_0(t)$ is the baseline hazard when all the predictors, x_i, are equal to 0; $h(t)$ is the expected hazard at time t, which represent the instantaneous death rate at time t; β is the vector of regression coefficients.

The CoxPHR model is a semi-parametric model (Cox, 1972) that can be written as

$$\frac{h(t)}{h_0(t)} = e^{x_i^t\beta}.$$

By taking the natural logarithm (ln) on both sides of the hazard function, the formula can be rewritten as

$$ln\, h(t) = ln\, h_0(t) + x_i^t\beta.$$

Although it is not a generalized linear model (Nelder & Wedderburn, 1972), inference on the coefficients β can be achieved via partial likelihood without specifying $h_0(t)$, the baseline hazard, proposed by Cox, 1975 (Cox, 1975).

The observed dataset is denoted as (y_i, δ_i, x_i), where y_i is the observed time, δ_i is the failure or right-censoring indicator (1 = failure, 0 = censored), x_i is a set of predictors. For individual with $\delta_i = 0$, the likelihood is

$$L_i(\beta) = S_i(y_i).$$

For individual with $\delta_i = 1$, the likelihood is

$$L_i(\beta) = S_i(y_i)h_i(y_i).$$

In both formulas above, $S_i(y_i)$ is the survival function, which gives the probability that a patient will survived beyond time y.

The full likelihood is

$$L(\beta) = \prod_{i=1}^{n} h_i(y_i)^{\delta_i} S_i(y_i)$$

$$= \prod_{i=1}^{n} \left[\frac{h_i(y_i)}{\sum_{j\in\mathcal{R}(y_i)} h_j(y_j)}\right]^{\delta_i} \left[\sum_{j\in\mathcal{R}(y_i)} h_j(y_i)\right]^{\delta_i} S_i(y_i).$$

The seconded and third terms of the full likelihood formula are not relevant to inference β. β is derived from the first term of the formula by Cox, which is

$$L_{partial}(\beta) = \prod_{i=1}^{n} \left[\frac{h_i(y_i)}{\sum_{j\in\mathcal{R}(y_i)} h_j(y_j)}\right]^{\delta_i}$$

$$= \prod_{i=1}^{n} \left[\frac{e^{x_i^t\beta}}{\sum_{j\in\mathcal{R}(y_i)} e^{x_j^t\beta}}\right]^{\delta_i}.$$

The likelihood function is called the partial likelihood function. The maximum likelihood estimation (MLE) method can be applied to the partial likelihood to calculate β (Lai, Hayashida, & Akutsu, 2013). The estimated β, denoted as $\hat{\beta}$, maximizes the log partial likelihood function and therefore, it is named as the maximum partial likelihood estimator (MPLE).

1.2.2 CoxPHR model with *lasso* penalty for variable selection (*glmnet* package)

The sample size, n, normally is larger than the number of predictors, p. However, if $n \gg p$ is not satisfied, prediction performance could suffer if too many nonimportant predictors are

included in the model. Additionally, high correlation between predictors may occur due to the inclusion of redundant variables. Hence, a subset of predictors should be selected. A regularization term, which is imposing a cost on the optimization function to prevent overfitting, can be added to the partial likelihood function to reach this goal. The current study uses the least absolute shrinkage and selection operator (*lasso*) estimator, which is an L_1- penalized estimator (Tibshirani, 1996). The L_1 norm, as known as Manhattan Distance norm, is the sum of the magnitudes of the vectors in a space.

The L_1 norm of β is defined as

$$\| \beta \|_1 = \sum_j |\beta_j|.$$

To estimate β with variable selection, the question is now to search for β which maximize the partial likelihood

$$\mathrm{L}_{partial}(\beta) = \prod_{i=1}^{n} \left[\frac{e^{x_i^t \beta}}{\sum_{j \in \mathcal{R}(y_i)} e^{x_j^t \beta}} \right]^{\delta_i},$$

subject to $\sum_j |\beta_j| \leq s$.

Friedman et al., 2010 provided the *glmnet* packaged in R to fit *lasso* and other regularization paths for the CoxPHR model (Friedman, Hastie, & Tibshirani, 2010). In *glmnet*, the equivalent question is written as

$$\arg\max_{\beta} l(\beta) - \sum_{j=1}^{p} \lambda[(1-\alpha) \| \beta \|_2^2/2 + \alpha \| \beta \|_1],$$

where $l(\beta)$ is the log partial likelihood, λ is the tuning parameter to controls the penalty; $\sum_{j=1}^{p} \lambda[(1-\alpha) \| \beta \|_2^2/2 + \alpha \| \beta \|_1]$ is the elastic net penalty, including the L_1 norm (*lasso*) and L_2 norm (ridge) penalties. When α is set to 1, which is the default in *glmnet* package, the *lasso*

penalty is applied. The current study uses the L_1 norm penalty because the L_1 norm penalty can force some of the coefficient estimates to be exactly equal to zero when the tuning parameter, λ, is sufficiently large. The *lasso* performers the best subset selection.

CoxPHR model with *lasso* penalty has strengths over stepwise regression, another widely used method to selection predictive variables. There are three main approaches for stepwise regression, forward selection, backward elimination, and bidirectional procedure. The algorithm used with all three approaches proposed by Efroymson in 1960 is an automated procedure to select variables based on some prespecified criteria, such as t statistics, AIC, BIC and adjusted R^2 (Efroymson, 1960). However, using different prespecified criteria could produce different sets of selected variables. Forward stepwise and backward stepwise selection approaches generally cannot produce identical models (Gareth James, 2013). Moreover, stepwise regression does not work well when 1) the predictors are correlated, or 2) the number of parameters is larger than that of the number of observations. The CoxPHR model with *lasso* penalty can still conduct variable selection in these two situations, and it has a more systematic method for generating a sequence of candidate models through the tuning of shrinkage parameter λ. For example, $\lambda = 0$ corresponds to the saturated model with no variable selection, and the model becomes sparse and more parsimonious as λ increases. Another strength of the CoxPHR model with *lasso* penalty is computational efficiency to explore the best combinations from many parameters.

1.2.3 Tuning parameter selection (*glmnet* package)

Cross-validation to select the tuning parameter, λ, is built in the *glmnet* package. The methods of measurement include deviance and Harrell's C-index for the CoxPHR model. The deviance is defined using the partial-likelihood for CoxPHR model in *cv.glmnet* function.

Harrell's C-index, also known as the concordance index, proposed by Harrell et al., 1982, is used to assess the goodness of fit of the CoxPHR model (F. E. Harrell, Jr., Califf, Pryor, Lee, & Rosati, 1982). A prognostic score for each patient, denoted as η, is defined as $x_i^t\beta$. A higher prognostic score should be related to a longer survival time. Harrell's C-index quantifies the correlation between the prognostic sore, η, and survival time, T.

For two patients i and j ($i \neq j$), three scenarios could happen.

- In scenario 1, both had an event during the follow-up time. A concordant pair is identified when $\eta_i > \eta_j$ and $T_i > T_j$. A discordant pair is identified when $\eta_i > \eta_j$ and $T_i < T_j$.
- In scenario 2, both were censored. Since it is unknown which one had a longer survival time, the pair of patients is not be counted.
- In scenario 3, the patient i was censored, and the patient j had an event. If $T_i > T_j$, a concordant pair is identified when $\eta_i > \eta_j$, and a discordant pair is identified when $\eta_i < \eta_j$. If $T_i < T_j$, the pair of patients is not be counted.

All possible pairs of patients must be investigated. Harrell's C-index is calculated as

$$C - index = \frac{\#\ concordant\ pairs}{\#\ concordant\ pairs + \#\ discordant\ pairs}$$

C-index ranges from 0 to 1. The value of C-index less than 0.5 means that the model's prediction ability is no better than a random chance. A value of C-index close to 1 means that the model has an excellent goodness-of-fit.

The *cv.glmnet* function in the *glmnet* package in returns the mean cross-validated C-index and the estimate of the standard error of the mean cross-validated C-index. The *cv.glmnet* function returns all values of the tuning parameter, λ, used in the fit, and also specifies the value that gives

the maximum mean cross-validated C-index, $lambda.min$, and that gives the C-index within one standard error of the maximum, $lambda.1se$. The later, $lambda.1se$, is more often to use to select the best model (Friedman et al., 2010).

1.2.4 Additional validation and calibration (*rms* package)

The *rmc* packaged developed by Frank Harrell can be used to evaluate the prediction ability of a fitted CoxPHR model (F. E. Harrell, 2006). In the current study, the functions, *validate.cph* and *calibrate*, are used to validate the fitted Cox prognostic model in the test set. The resampling methods built in the two functions include cross-validation, bootstrap, 0.632 resampling, and randomization. Resampling approaches involves repeatedly drawing samples from a training set and fitting the same statistical method multiple times. The function, *validate.cph*, returns a bias-corrected D_{xy} index, which equal to $2 \times (C - index - 0.5)$. The function, *calibrate*, gives a curve comparing the predicted probabilities to the observed probabilities at a specified time point.

2.0 Method

2.1 Patients and clinical variables

A total of 5,944 patients diagnosed with invasive HGSOC and enrolled in the collaborative studies participating in the OTTA consortium were considered eligible in the analysis. The site-specific Institutional Review Boards approved study protocols. Inclusion criteria included the availability of tissue microarrays (TMAs) for IHC analysis, clinical follow-up data, age at diagnosis, tumor stage, and debulking status. Clinical data were obtained from medical records, cancer registries, death certificates, and pathology reports. There were 129 cases with missing data for stage there were excluded from the analysis. There were 254 cases with complete data for 9 IHC biomarkers and 5,561 cases with at least one missing IHC biomarker. Data from the participating study sites were centrally harmonized.

2.2 Immunohistochemistry (IHC)

Tumors were obtained at initial debulking surgery and arrayed on TMAs. Stained slides of TMAs were centralized and performed at the Mayo Clinic (Rochester, Minnesota) for MyD88, TLR4 and CD8 TILs, at Leica Microsystems (Wetzlar, Germany) for FOLR1, and at Genetic Pathology Evaluation Centre (Vancouver, BC, Canada) for ER and PR. The information on PTEN and AR is not available. Details are given in the published studies (Block et al., 2018; Köbel et al., 2014; Ovarian Tumor Tissue Analysis et al., 2017; Rambau et al., 2018; Sieh et al., 2013).

Two observers score the staining intensity for each IHC. For AR, a 2-tiered system (negative and positive) was used. ER and PR were scored using a 3-tiered system (<1%, 1 to 50%, and >50% of tumor cell nuclei positive). For p16 expression, a 3- tiered system was used (<1%, 1 to 75%, and >75% of tumor cell nuclei positive). MyD88, TLR4 and PTEN were scored using a 4-tiered system (negative, week, moderate and strong expression). CD8 TILs were scored into a 4-tiered system (none, 1-2 IEL/40 x HPF, 3-19 IEL/40 x HPF, and >=20 IEL/40 x HPF). FOLR1 was scored into a 6-tiered system (absent, weak, 1-50% irrespective of subcellular localization, >50% with membranous localization, 50-95% with cytoplasmic staining and >95% with cytoplasmic staining). A preliminary analysis was conducted to compare the association between IHC biomarkers and HGSOC survival in published OTTA paper using CoxPHR model adjusting for study site, age (continuous), stage (I/II and III/IV) and debulking status (optimal, suboptimal, and unknown) in STATA SE 16.1.

2.3 Analysis

The goal of this thesis was to build a prognostic model for HGSOC overall survival. To achieve this, the following steps were used: step 1) split the dataset into the training set and test set; step 2) fit a CoxPHR model with *lasso* penalty in the training set to identify the most important individual predictors from clinical variables and IHC biomarkers; step 3) select the best model from models with age, stage and building status (model 0), the dummy variables selected by *lasso* (model 1), the dummy variables selected by *lasso* with other dummy variables grouped in the categorical variables (model 2), and further adjusting for study site (model 3); step 4) conduct external validation in the test set; and step 5) visualize the final prognostic model.

2.3.1 Training set and test set

There were total 5944 HGSOC cases from OTTA. Cases with missing stage were excluded from the current study, leaving 5815 cases. The training set contained 254 cases with complete data for 9 IHC biomarkers from HOP (Lo-Ciganic et al., 2012), MAL (Glud et al., 2004), NOT (Williams, Martin, Moss, Durrant, & Deen, 2012), TUE, UKO (Balogun et al., 2011) and VAN (Köbel et al., 2010; Prentice et al., 2007). After the step of variable selection, 1563 cases with data in the selected IHC biomarkers compose the test set from AOV(Kelemen, Köbel, Chan, Taghaddos, & Dinu, 2013; Köbel et al., 2014), BAV (Song et al., 2009), BRZ, CAL (Bromley et al., 2012), CNI (Kamieniak et al., 2015), GER (Royar, Becher, & Chang-Claude, 2001), HAW (Goodman, Lurie, Thompson, McDuffie, & Carney, 2008; Lurie et al., 2008), HOP (Lo-Ciganic et al., 2012), LAX (S. J. Ramus et al., 2012), MAL (Glud et al., 2004), MAY (Goode et al., 2010; Kelemen et al., 2008), NOT (Williams et al., 2012), POC (Garcia-Closas et al., 2007), TUE, UKO (Balogun et al., 2011), VAN (Köbel et al., 2010; Prentice et al., 2007) and WMH (Emmanuel et al., 2014) (**Appendix Table 1**). The left 5561 cases were in the candidates' pool. After variables selection by *Lasso*, 1563 cases with selected IHC biomarkers from the candidates' pool were in the test set (**Figure 1)**. Student's t-test was used to compare age at diagnosis (continuous), and Chi-square test was used to compare stage (I/II and III/IV), debulking status (optimal, suboptimal, and unknown) and 9 IHC biomarkers between cases in the training set and in the candidates' pool. The statistical tests were two-sided and conducted using STATA SE 16.1.

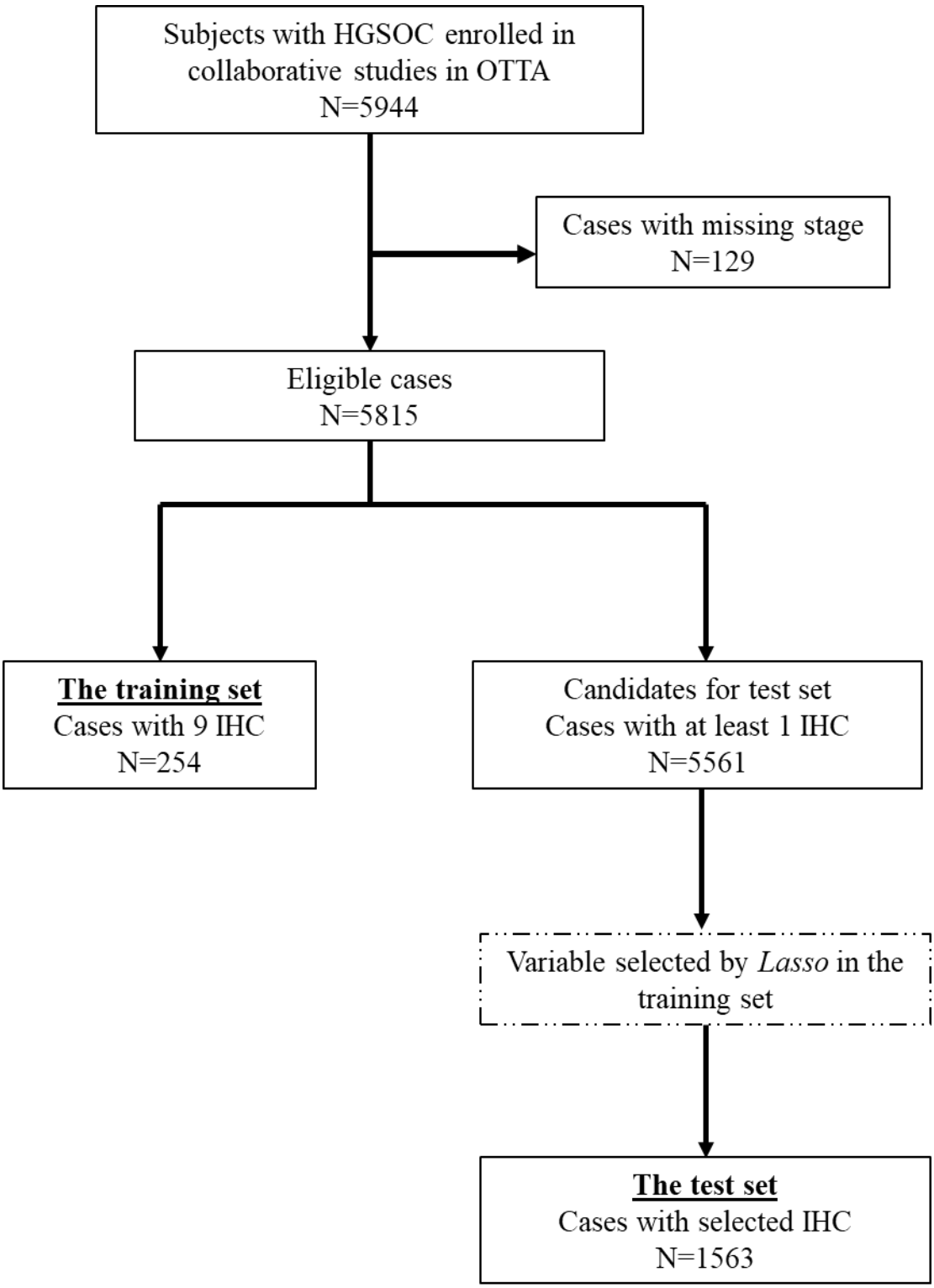

Figure 1 Cases election for taining set and test set

2.3.2 Variable selection

Data from the training set was used to fit in the CoxPHR model with *lasso* penalty using the *glmnet* package in R software. Because the *cv.glmnet* function cannot perform the CoxPHR model with grouped *lasso* penalty (i.e. does not accommodate categorical variables with more than two levels directly), the *model.matrix* function was used to create the predictor matrix with dummy variables. Thus, each level of the original categorical variables was treated as an individual predictor and could potentially be selected in the prediction models. From the original training set, the predictor matrix contained 29 predictors not including intercept, because a CoxPHR model does not contain an intercept. The default 10-fold cross-validation was used to select the tuning parameter based on the C-index. The C-index corresponding to the tuning parameter that gives error within one standard error of the maximum was chosen.

2.3.3 Model selection and internal validation

The *rms* package in R software is used to refit and evaluate four CoxPHR models: **model 0** as a baseline model using age and the clinical variables, such as stage and debulking status, **model 1** using the dummy variables selected by *lasso*, **model 2** using the dummy variables selected by *lasso* with other dummy variables grouped in the categorical variables, and **model 3** further adjusting for study sites. Because *Surv* function in the package cannot adjust the bias induced by left truncation, the follow-up duration is corrected by the date of death or the date of last known to be alive since the date of diagnosis minus the date of interviews since the date of diagnosis. The computation of corrected follow-up duration was checked in STATA SE 16.1. The estimated

parameters using the corrected follow-up duration are close to those from *stset* function dealing with left-truncation data in STATA SE 16.1.

The goodness of fit for a model was determined by C-index as well as calibrations of 1-year, 2- year, 3- year and 5-year survival using 500 bootstrap samples. The internal validation was implemented by *validate* and *calibrate* functions built in the *rms* package.

2.3.4 External validation

Student's t-test was used to compare age at diagnosis (continuous), and Chi-square test was used to compare stage (I/II and III/IV), debulking status (optimal, suboptimal, and unknown) and 9 IHC biomarkers among the test set, training set and excluded cases. Kaplan-Meier curves were visually compared to assess differences in survival. The statistical tests were two-sided and conducted using STATA SE 16.1.

2.3.5 Final model and nomogram

The final CoxPHR model estimated the hazard ratios (HRs) and 95% confidence intervals (CIs) of the selected clinical variables and IHC biomarkers using the test set. Kaplan-Meier curves visually compared survival across each categorical variable adjusted for other covariates to estimate the proportional hazards assumption. Wald test were performed to test the linear trend for selected IHC biomarkers in the final model. The tests were implemented in STATA SE 16.1. The goodness of fit for the refitted model in the test set was determined by C-index as well as calibrations of 1-year, 2- year, 3- year and 5-year survival using 500 bootstrap samples. Receiver operating characteristic (ROC) curves with area under the curve (AUC) corresponding to the

predictions on 1-year, 2- year, 3- year and 5-year survival were presented using the *timeROC* package in R software.

The prognostic model was also fitted using the *rms* package in R software. Based on the calibration of 1-year, 2- year, 3- year and 5-year survival, a nomogram to predict 3-year and 5-year survival of patients with HGSOC is obtained by *nomogram* function. The nomogram gives a prognostic index calculated by summing the number of the risk points corresponding to each weighted covariate in the prognostic model. The possibility of the 3-year and 5-year survival for each patient can be subsequently calculated based on the prognostic index.

3.0 Results

3.1 Preliminary analysis for IHC biomarkers

The associations between IHC biomarkers and HGSOC survival in the current study are similar to the results from published OTTA study (**Table 1**). Strong expression of MyD88 in the tumor cells is associated with higher hazard of HGSOC death (HR 1.13, 95% CI 1.02- 1.26). Expression of CD8 is inversely associated with hazard of HGSOC death. However, strong expression of PR is not associated with HGSOC survival in the current data (HR 0.85, 95% CI 0.68-1.05), although Sieh et al. indicated a significant association between strong expression of PR and HGSOC survival (Sieh et al., 2013). All published OTTA studies adjusted for study site, age and stage in their models. Studies estimating the association between FOLR1 and p16 also adjusted for residual disease after surgery (Köbel et al., 2014; Rambau et al., 2018). The discrepancies on the effect of PR and ER on HGSOC survival between the current study and Sieh et al.'s study could be due to that Sieh et al. further adjusted for age-squared in the Cox model (Sieh et al., 2013).

Table 1 Comparing the association between IHC biomarkers and HGSOC survival to the published OTTA study

IHC biomarkers	OTTA				Current study		
	Level	**N (%)**	**HR (95% CI)**	**Confounders**	**Level**	**N (%)**	**HR (95% CI)[1]**
MyD88 (Block et al., 2018)	Weak	712 (26)	ref	Adjusted for study, age (continuous) and stage (I/II, III/IV. Unknown)	Negative/ weak	769 (25.41)	ref
	Strong	2064 (74)	1.13 (1.01-1.26)		Moderate/strong	2257 (74.59)	1.13 (1.02, 1.26)
TLR4 (Block et al., 2018)	Weak	734 (29)	ref	Adjusted for study, age (continuous) and stage (I/II, III/IV. Unknown)	Negative/ weak	768 (28.80)	ref
	Strong	1788 (71)	1.06 (0.94, 1.18)		Moderate/strong	1899 (71.20)	1.13 (1.01, 1.26)
FOLR1 (Köbel et al., 2014)	Negative (Abstract/weak)	358 (23.8)	ref	Stratified by study and adjusted for age at diagnosis, residual disease (not macroscopic, macroscopic or missing) and FIGO stage (I/II, III/IV or missing)	Negative (Abstract/weak)	375 (23.67)	ref
	Positive (1-50%, membranous >50%, cytoplasmic50-95%, and cytoplasmic >95%)	1149 (76.2)	0.99 (0.84, 1.18)		Positive (1-50%, membranous >50%, cytoplasmic50-95%, and cytoplasmic >95%)	1209 (76.33)	0.99 (0.86, 1.13)
CD8+ (Ovarian Tumor Tissue Analysis et al., 2017)	Negative (none)	546 (17.1)	ref	Adjusted for study, age (continuous), and stage (I/II, III/IV, unknown)	None	588 (17.36)	ref
	Low (1-2 IEL/40 x HPF)	546 (17.1)	0.86 (0.75, 0.99)		1-2 IEL/40 x HPF	577 (17.04)	0.91 (0.80, 1.04)
	Moderate (3-19 IEL/40 x HPF)	1394 (43.6)	0.77 (0.69, 0.87)		3-19 IEL/40 x HPF	1476 (43.58)	0.80 (0.72, 0.90)
	High (20+ IEL/40 x HPF)	710 (22.2)	0.57 (0.49, 0.65)		20+ IEL/40 x HPF	746 (22.03)	0.60 (0.53, 0.69)
p16 (Rambau et al., 2018)	Heterogeneous	1550 (37.9)	ref	Adjusted for study, age, time interval, stage and residual tumor	Negative (<1%)	269 (6.61)	ref

	Absent	244 (6.0)	1.06 (0.90, 1.25)		1 -75%	1534 (37.68)	1.00 (0.86, 1.17)
	Block	2292 (56.1)	1.03 (0.95, 1.11)		> 75%	2268 (55.71)	1.03 (0.89, 1.20)
Progesterone receptor (Sieh et al., 2013)	Negative	1144 (68.9)	ref	Stratified by site, and adjusted for age, age-squared, and stage (localized, regional, advanced) at diagnosis	Negative	1107 (66.49)	ref
	Weak	393 (23.7)	1.02 (0.89, 1.18)		Weak	421 (25.29)	1.05 (0.92, 1.19)
	Strong	124 (7.5)	0.71 (0.55, 0.91)		Strong	137 (8.23)	0.85 (0.68, 1.05)
Estrogen receptor (Sieh et al., 2013)	Negative	326 (19.3)	ref	Stratified by site, and adjusted for age, age-squared, and stage (localized, regional, advanced) at diagnosis	Negative	350 (20.73)	ref
	Weak	347 (20.5)	1.08 (0.89, 1.31)		Weak	360 (21.33)	0.93 (0.78, 1.11)
	Strong	1018 (60.2)	1.05 (0.89, 1.24)		Strong	978 (57.94)	0.94 (0.81, 1.08)

CI, confidence interval; HR, hazard ratio; IEL, intraepithelial lymphocytes; IHC, Immunohistochemistry; ref, reference.

[1] Adjusting for study site, age as continuous, stage (I/II and III/IV) and debulking status (optimal, suboptimal, and unknown)

s

3.2 Clinical characteristics of patients

Out of 5815 cases, 254 patients with HGSOC from 6 sites in OTTA are included in the variable selection process. The characteristics of these patients are presented in **Table 2**. The participating studies and case ascertainment are summarized in **Appendix Table 1**. The median OS is 5.04 years (95% CI 4.36-5.99), which is longer than the median OS among cases in the candidates' pool, 3.50 years (95% CI 3.36-3.63). Of the 254 cases, 124 (48.82%) have more advanced stage. While 82.84% of cases in candidates' pool are in stage III/ IV ($P<0.001$). There are 56.69% of cases and 34.11% cases with an optimal debulking status in the training set and in the candidates' pool, respectively.

AR expression is seen in 43.31% HGSC cases in the training set. Nearly half cases (48.03%) have ER expression in >50% tumor cells in the training set. More than 70% of cases have a moderate or strong intensity of TLR4 and MyD88 expression. FOLR1 expression is not presented in 36 cases and is weakly presented in 40 cases, while 88 cases have FOLR1 expression in >50% of tumor cells with cytoplasmic staining. 90 cases (35.43%) had 3-19 IEL/40 * HPF of CD8+ and 70 (27.56%) cases have more than 20 IEL/40 * HPF. More than 60% of cases have PR expression in <1% tumor cells, which is close to the percentage in the candidates' pool ($P = 0.169$). The distribution of p16 expressed in the training set and in the candidates' pool are not statistically different ($P=0.525$).

Table 2 Demographic and immunohistochemistry characteristics of patients with invasive high grade serous ovarian cancer cases

Variables	Total (N=5815), n (%)	Cases in training set (N=254), n (%)	Cases in candidates' pool (N=5561), n (%)	P value
Overall survival, years				
median (95% CI)	3.58 (3.43, 3.70)	5.04 (4.36, 5.99)	3.50 (3.36, 3.63)	<0.001
Age (year), mean (SD)				
mean (SD)	60.22 (10.71)	60.28 (11.26)	60.22 (10.68)	0.9345
Stage				
stage I/II	1084 (18.64)	130 (51.18)	954 (17.16)	<0.001
stage III/IV	4731 (81.36)	124 (48.82)	4607 (82.84)	
Debulking Status				
optimal	2041 (35.10)	144 (56.69)	1897 (34.11)	<0.001
suboptimal	526 (9.05)	7 (2.76)	519 (9.33)	
unknown	3248 (55.86)	103 (40.55)	3145 (56.55)	
Immunohistochemistry				
Androgen Receptor				
negative	2308 (65.57)	144 (56.69)	2164 (66.26)	0.002
positive	1212 (34.43)	110 (43.31)	1102 (33.74)	
missing	3739	-	3739	
Estrogen Receptor				
<1% of tumor cells	350 (20.73)	68 (26.77)	282 (19.67)	0.002
1 to 50% of tumor cell nuclei positive	360 (21.33)	64 (25.20)	296 (20.64)	
> 50% of tumor cells positive	978 (57.94)	122 (48.03)	8856 (59.69)	
missing	4127	-	4127	
Progesterone Receptor				
<1% of tumor cells	1694 (61.40)	158 (62.20)	1536 (61.32)	0.169
1 to 50% of tumor cell nuclei positive	815 (29.54)	66 (25.98)	749 (29.90)	
> 50% of tumor cells positive	250 (9.06)	30 (11.81)	220 (8.78)	
missing	3306	-	3306	
Toll-like receptor 4, TLR4				
negative	306 (11.47)	22 (8.66)	284 (11.77)	0.001
weak intensity	462 (17.32)	38 (14.96)	424 (17.57)	
moderate intensity	1735 (65.05)	190 (74.80)	1545 (64.03)	
strong intensity	164 (6.15)	4 (1.57)	160 (6.63)	
missing	3148	-	3148	

Myeloid differentiation primary response 88, MyD88				
negative	361 (11.93)	43 (16.93)	318 (11.47)	0.004
weak intensity	408 (13.48)	24 (9.45)	384 (13.85)	
moderate intensity	1409 (46.56)	130 (51.18)	1279 (46.14)	
strong intensity	848 (28.02)	57 (22.44)	791 (28.54)	
missing	2789	-		
Folate receptor alpha, FOLR1				
absent	175 (11.05)	36 (14.17)	139 (10.45)	0.017
weak staining	200 (12.63)	40 (15.75)	160 (12.03)	
1–50% of tumor cells irrespective of subcellular localization	400 (25.25)	51 (20.08)	349 (26.24)	
>50% of tumor cells with membranous localization	299 (18.88)	39 (15.35)	260 (19.55)	
50–95% of tumor cells with cytoplasmic staining	375 (23.67)	71 (27.95)	340 (22.86)	
>95% of tumor cells with cytoplasmic staining	135 (8.52)	17 (6.69)	118 (8.87)	
missing	4231	-	4231	
Phosphatase and tensin homolog, PTEN				
negative	587 (19.05)	32 (12.60)	555 (19.63)	0.049
weak intensity	1544 (50.10)	137 (53.94)	1407 (49.75)	
moderate intensity	773 (25.08)	67 (26.38)	706 (24.96)	
strong intensity	178 (5.78)	18 (7.09)	160 (5.66)	
missing	2733	-	2733	
Cluster of differentiation 8, CD8				
no IEL	588 (17.36)	53 (20.87)	535 (17.08)	0.018
1-2 IEL/40 x HPF	577 (17.04)	41 (16.14)	536 (17.11)	
3-19 IEL/40 x HPF	1476 (43.58)	90 (35.43)	1386 (44.24)	
20 or more IEL/40 x HPF	746 (22.03)	70 (27.56)	676 (21.58)	
missing	2428	-	2428	
p16				
<1% of tumor cells	269 (6.61)	15 (5.91)	254 (6.65)	0.525
1 to 75% of tumor cell nuclei positive	1534 (37.68)	104 (40.94)	1430 (37.46)	
> 75% of tumor cells positive	2268 (55.71)	135 (53.15)	2133 (55.88)	
missing	1744	-	1744	

IEL, intraepithelial lymphocytes

3.3 Variable selection for the prognostic model

The tuning parameter was selected based on C-index using 10-fold cross-validation (**Figure 2**). The tuning parameter that gave the C-index within one standard error of the maximum is 0.083 corresponding to the C-index of 0.6471 (standard error 0.02103).

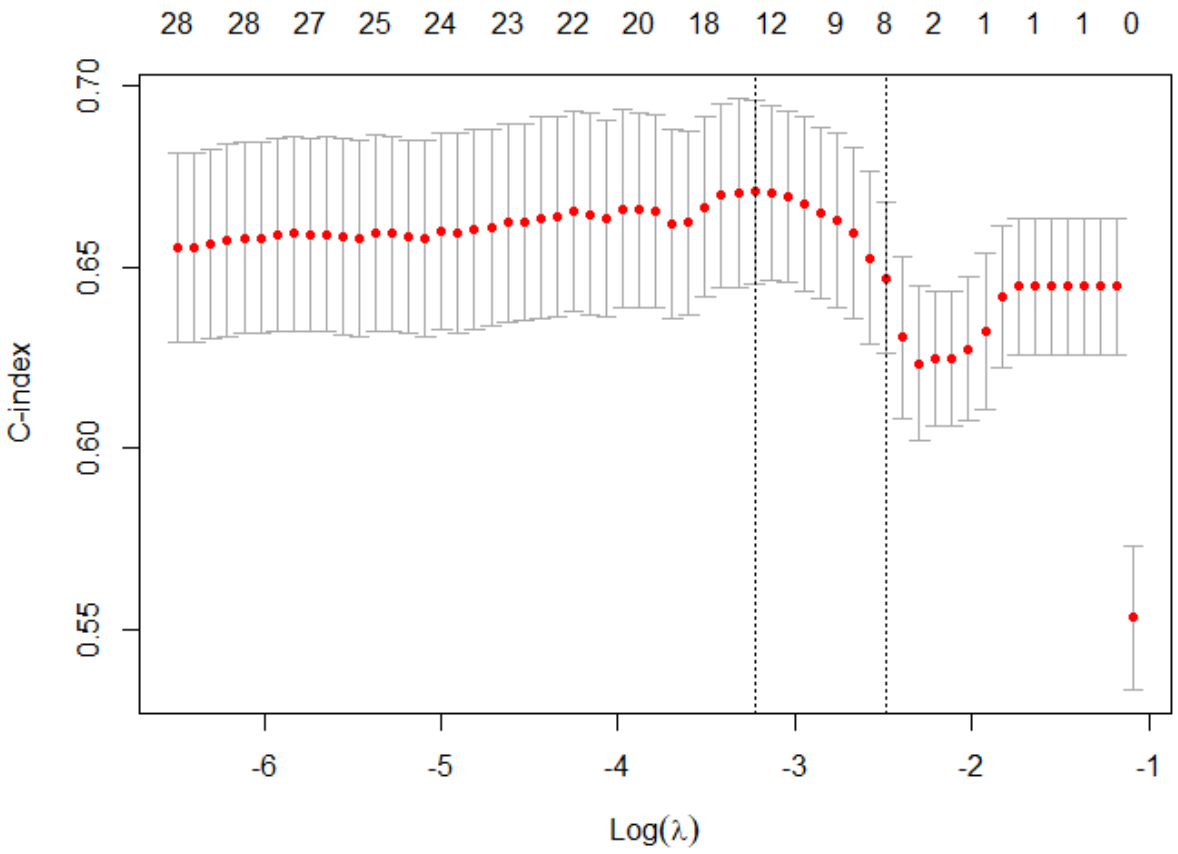

Figure 2 Trend of C-index with the different tuning parameter, *lambda*

Of 28 dummy variables, 8 variables with non-zero coefficients are selected (**Table 3**) according to the tuning parameter that gives the C-index within one standard error of the

maximum. The third level of TLR4 expression, the third and fourth level of CD8+ TILs, and the second level of p16 expression out of the dummy IHC variables are selected.

Table 3 Aossiciaton of overall survival with clinical and immunohistochemistrical variables of patients with HGSOC selected from the CoxPHR model with Lasso penalty

Predictors	β
Age (year), mean (SD)	0.0018
Stage	
Stage I/II	ref
Stage III/IV	0.8125
Debulking Status	
Optimal	ref
Suboptimal	0.3204
Unknown	Not selected
Immunohistochemistry	
Androgen Receptor	
Negative	ref
Positive	-0.0901
Toll-like receptor 4, TLR4	
negative	ref
weak intensity	Not selected
moderate intensity	-0.0418
strong intensity	Not selected
Cluster of differentiation 8, CD8	
no IEL	ref
1-2 IEL/40 x HPF	Not selected
3-19 IEL/40 x HPF	-0.0464
20 or more IEL/40 x HPF	-0.0750
p16	
<1% of tumor cells	ref
1 to 75% of tumor cell nuclei positive	-0.0540
> 75% of tumor cells positive	Not selected

IEL, intraepithelial lymphocytes.

Table 4 indicates the predictors in four refitted Cox models using age, stage and debulking status (Model 0), using dummy variables selected by *lasso* (Model 1), using the dummy variables

selected by *lasso* with other dummy variables grouped in the categorical variables (Model 2), and further adjusting for study site (Model 3), respectively. **Table 5** presents the parameters from four refitted Cox models. C-index from 500 bootstrap validation samples are 0.6611, 0.6922, 0.6863 and 0.6877 for Model 0, Model 1, Model 2 and Model 3, respectively. Further adjusting for study site does not improve C-index. The estimated βs from Model 1 (**Table 5**) are smaller than the estimates from the CoxPHR model with *lasso* penalty (**Table 3**) because *lasso* shrinks βs towards to zero.

Table 4 Predictors in three models

Variables	Model 0	Model 1	Model 2	Model 3
Age (year) as continuous variable	Yes	Yes	Yes	Yes
Stage				
Stage I/II	ref	ref	ref	ref
Stage III/IV	Yes	Yes	Yes	Yes
Debulking Status				
Optimal	ref	ref	ref	ref
Suboptimal	Yes	Yes	Yes	Yes
Unknown	Yes	No	Yes	Yes
Immunohistochemistry				
Androgen Receptor				
Negative	No	ref	ref	ref
Positive	No	Yes	Yes	Yes
Estrogen Receptor				
<1% of tumor cells	No	No	No	No
1 to 50% of tumor cell nuclei positive	No	No	No	No
> 50% of tumor cells positive	No	No	No	No
Progesterone Receptor				
<1% of tumor cells	No	No	No	No
1 to 50% of tumor cell nuclei positive	No	No	No	No
> 50% of tumor cells positive	No	No	No	No
Toll-like receptor 4, TLR4				
negative	No	ref	ref	ref
weak intensity	No	No	Yes	Yes
moderate intensity	No	Yes	Yes	Yes
strong intensity	No	No	Yes	Yes
Myeloid differentiation primary response 88, MyD88				
negative	No	No	No	No
weak intensity	No	No	No	No
moderate intensity	No	No	No	No

strong intensity	No	No	No	No
Folate receptor alpha, FOLR1				
absent	No	No	No	No
weak staining	No	No	No	No
1–50% of tumor cells irrespective of subcellular localization	No	No	No	No
>50% of tumor cells with membranous localization	No	No	No	No
50–95% of tumor cells with cytoplasmic staining	No	No	No	No
>95% of tumor cells with cytoplasmic staining	No	No	No	No
Phosphatase and tensin homolog, PTEN				
negative	No	No	No	No
weak intensity	No	No	No	No
moderate intensity	No	No	No	No
strong intensity	No	No	No	No
Cluster of differentiation 8, CD8				
no IEL	No	ref	ref	ref
1-2 IEL/40 x HPF	No	No	Yes	Yes
3-19 IEL/40 x HPF	No	Yes	Yes	Yes
20 or more IEL/40 x HPF	No	Yes	Yes	Yes
p16				
<1% of tumor cells	No	ref	ref	ref
1 to 75% of tumor cell nuclei positive	No	Yes	Yes	Yes
> 75% of tumor cells positive	No	No	Yes	Yes
Study Site as categorical variable	No	No	No	Yes

Table 5 Comparison of parameters from CoxPHR model with Lasso penalty and refitted CoxPHR models

Variables	**β from refitted Cox model 0**	**β from refitted Cox model 1**	**β from refitted Cox model 2**	**β from refitted Cox model 3[1]**
Age (year), mean (SD)	0.0138	0.0098	0.0104	0.0090
Stage				
Stage I/II	ref	ref	ref	ref
Stage III/IV	1.0599	1.0613	0.9899	0.9778
Debulking Status				
Optimal	ref	ref	ref	ref
Suboptimal	1.3696	1.2421	1.2711	1.7723
Unknown	0.1999	NA	0.2811	1.1230
Immunohistochemistry				
Androgen Receptor				
Negative	NA	ref	ref	ref
Positive	NA	-0.3739	-0.2938	-0.2445
Toll-like receptor 4, TLR4				
negative	NA	ref	ref	ref
weak intensity	NA	NA	-0.0631	0.0076
moderate intensity	NA	-0.0882	-0.1775	-0.0293

strong intensity	NA	NA	-0.0367	0.2429
Cluster of differentiation 8, CD8				
no IEL	NA	ref	ref	ref
1-2 IEL/40 x HPF	NA	NA	-0.4566	-0.5428
3-19 IEL/40 x HPF	NA	-0.5788	-0.8490	-0.9284
20 or more IEL/40 x HPF	NA	-0.7084	-0.9223	-0.9792
p16				
<1% of tumor cells	NA	ref	ref	ref
1 to 75% of tumor cell nuclei positive	NA	-0.2959	-0.1360	0.2084
> 75% of tumor cells positive	NA	NA	0.2902	0.5640
D_{xy}	0.3221	0.3844	0.3726	0.3753
C-index	0.6611	0.6922	0.6863	0.6877

IEL, intraepithelial lymphocytes; NA, not applicable

[1] adjust for study site

The calibration of 1-year, 2-year, 3-year and 5-year survival predicted by Model 1 and Model 2 are presented in **Figure 3**. The mean errors of model 1 was 0.017 for 3-year survival and 0.033 for 5-year survival. The mean errors of model 1 was 0.026 for 3-year survival and 0.035 for 5-year survival. Compared to Model 2, Model 1 has a slightly better performance to predict the survival of HGSOC for 3-year and 5-year survival, but Model 2 is more interpretable. Thus, Model 2, as the final prognostic model, was used for external validation.

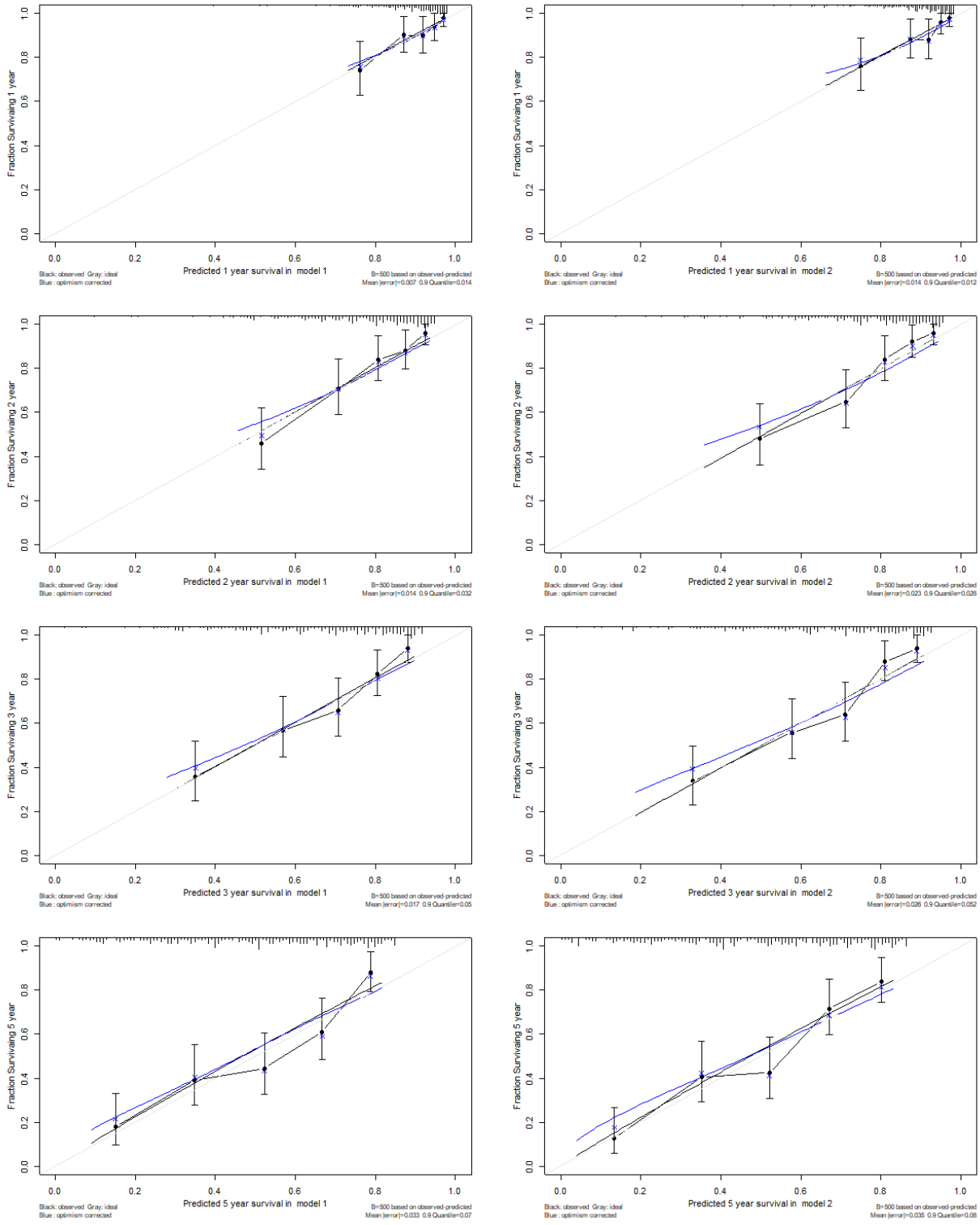

Figure 3 Calibration of 1-year, 2-year, 3-year and 5-year survival for Model 1 and Model 2 in the training set

3.4 External validation

A total of 1,563 patients with data in AR, TLR4, CD8 and p16 expression in 17 study sites from OTTA (**Appendix Table 1**) were used as a test set to perform the external validation. The median OS in the test set is 3.41 years (95 % CI 3.21-3.63 years), which is significantly shorter than the OS in the training set (5.41 years, 95% CI 4.36-5.99) but closes to the excluded cases (3.55 years, 95% CI 3.34-3.73) (**Figure 4**).

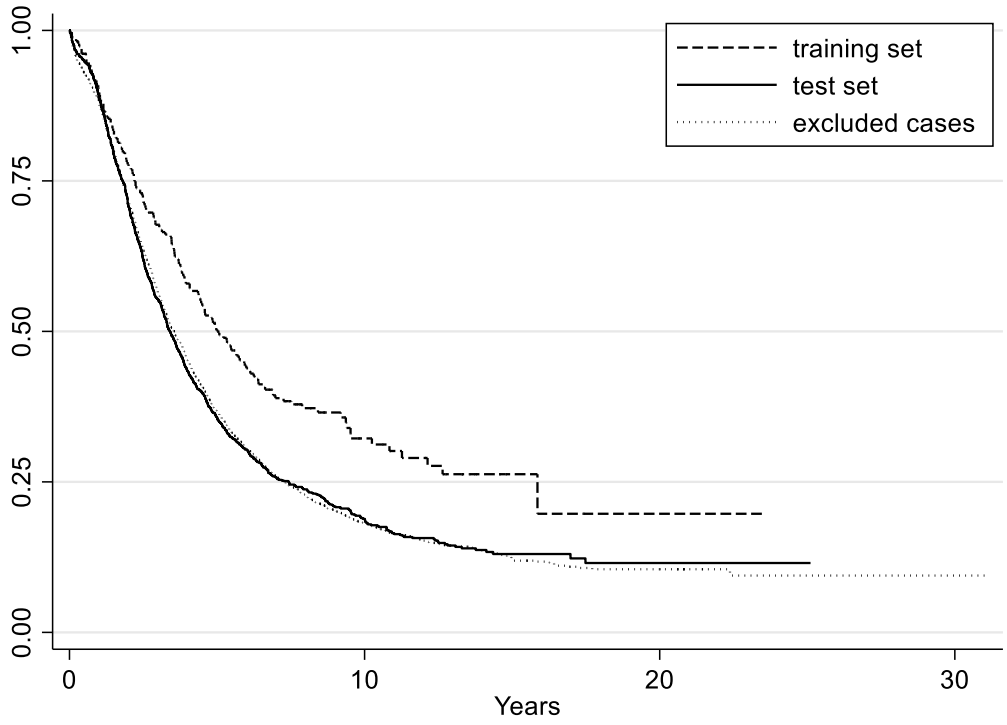

Figure 4 Kaplan–Meier curves for the test set, training set and excluded cases

The test set shares similar distributions of age, AR and p16 expression as the training set. Patients in the test set are likely to be in an advanced stage and had suboptimal debulking status (**Table 6**). TLR4 is not expressed in 14.78% cases and weakly expressed in 18.04% cases in the test set, compared to 8.66% and 14.96% in the training set. Cases in test set have less absent CD8+ TILs (14.84%), but also less 20 or more IEL/40 x HPF (23.42%) compared to cases in the training set,

20.87% and 27.56%, respectively. Compared to the excluded cases, cases in the test set are more like to be older (mean 61.21 vs. 59.83 of excluded cases, P <0.0001) and have optimal debulking status (45.55% vs. 29.64 of excluded cases, P < 0.001). However, the percentages of cases in stage I/II are close to each other (16.44% in the test set vs. 17.43% of excluded cases, P = 0.378).

Table 6 Comparison of selected demographic and immunohistochemistry characteristics of patients between the test set, training set and exlcuded cases

Variables	Test set (N=1563), n (%)	Training Set (N=254), n (%)	P-value comparing test set to training set	Excluded cases (N=3998), n (%)	P-value comparing test set to excluded cases
Overall survival, years					
Median (95% CI)	3.41 (3.21, 3.63)	5.04 (4.36, 5.99)	<0.001	3.55 (3.34, 3.73)	0.778
Age (year), mean (SD)					
mean (SD)	61.21 (11.25)	60.28 (11.26)	0.2211	59.83 (10.43)	<0.0001
Stage					
Stage I/II	257 (16.44)	130 (51.18)	<0.001	697 (17.43)	0.378
Stage III/IV	1306 (83.56)	124 (48.82)		3301 (82.57)	
Debulking Status					
Optimal	712 (45.55)	144 (56.69)	<0.001	1185 (29.64)	<0.001
Suboptimal	189 (12.09)	7 (2.76)		330 (8.25)	
Unknown	662 (42.35)	103 (40.55)		2483 (62.11)	
Immunohistochemistry					
Androgen Receptor					
negative	959 (61.36)	144 (56.69)	0.158	1205 (70.76)	<0.001
positive	604 (38.64)	110 (43.31)		498 (29.24)	
missing	-	-		2295	
Toll-like receptor 4, TLR4					
negative	231 (14.78)	22 (8.66)	<0.001	53 (6.24)	<0.001
weak intensity	282 (18.04)	38 (14.96)		142 (16.71)	
moderate intensity	934 (59.76)	190 (74.80)		611 (71.88)	
strong intensity	116(7.42)	4 (1.57)		44 (5.18)	
missing				3148	
Cluster of differentiation 8, CD8					
no IEL	232 (14.84)	53 (20.87)	0.01	303 (19.30)	0.002
1-2 IEL/40 x HPF	260 (16.63)	41 (16.14)		276 (17.58)	
3-19 IEL/40 x HPF	705 (45.11)	90 (35.43)		681 (43.38)	
20 or more IEL/40 x HPF	366 (23.42)	70 (27.56)		310 (19.75)	

missing	-	-		2428	
p16					
<1% of tumor cells	77 (4.93)	15 (5.91)	0.631	177 (7385)	0.001
1 to 75% of tumor cell nuclei positive	611 (39.09)	104 (40.94)		819 (36.34)	
> 75% of tumor cells nuclei positive	875 (55.98)	135 (53.15)		1258 (55.81)	
missing	-	-		1744	

IEL, intraepithelial lymphocytes

3.5 Final model and visualization

Due to a high discrepancy on survival time between the test set and training set, the hazard ratio of OS with age, stage, debulking status, AR, TLR4, CD8+ TILs and p16 expression from the prognostic model refitting in the test set are obtained (**Table 7)**. The hazard is 2.40 times (95 CI% 1.96-2.93) higher for patients in stage III/IV compared to patients in I/II. The suboptimal debulking status increases the 1.47 times of hazard of death (95% CI 1.21-1.78) compared to the optimal debulking status. Higher density of CD8+ TILs is associated with lower hazard of death (0.92, 0.88, 0.72 for 1-2 IEL/40 x HPF, 3-19 IEL/40 x HPF and 20 or more IEL/40 x HPF compared to no IEL, P for trend =0.0011). Although TLR4 and p16 expression status are also selected to build the prognostic model, no significant association of TLR4 or p16 expression levels to the OS was found. Positive AR expression is associated with a 18% lower hazard of death compared to negative AR expression, but the association is not statistically significant (95% CI 0.81-1.04, P = 0.166).

Table 7 Association of overall survival with selected clinical and immunohistochemistry variables of the predict Cox model

Variables	Hazard Ratio (95% CI)
Age as continuous	1.02 (1.02, 1.03)
Stage	
Stage I/II	ref
Stage III/IV	2.40 (1.96, 2.93)
Debulking Status	
Optimal	ref
Suboptimal	1.47 (1.21, 1.78)
Unknown	1.48 (1.30, 1.68)
Immunohistochemistry	
Androgen Receptor	
Negative	ref
Positive	0.92 (0.81, 1.04)
Toll-like receptor 4, TLR4	
negative	ref
weak intensity	1.07 (0.86, 1.32)
moderate intensity	1.04 (0.87, 1.24)
strong intensity	1.03 (0.79, 1.35)
Cluster of differentiation 8, CD8	
no IEL	ref
1-2 IEL/40 x HPF	0.92 (0.75, 1.13)
3-19 IEL/40 x HPF	0.88 (0.74, 1.04)
20 or more IEL/40 x HPF	0.72 (.59, 0.88)
p16	
<1% of tumor cells	ref
1 to 75% of tumor cell nuclei positive	0.91 (0.68, 1.21)
> 75% of tumor cells nuclei positive	0.91 (0.68, 1.20)
C-index	0.6309

IEL, intraepithelial lymphocytes

Kaplan-Meier curves comparing survival across categories of each variable from the final model adjusting for age and the other variables are presented in **Figure 5**. In each panel, curves in each panel, except for AR, do not cross each other, which indicates that the proportional hazards assumption is not violated. Curves in the panel C for AR crosses at very early times and comes

close to each other at later time. P for the proportional hazard assumption test based on Schoenfeld residuals is 0.0045 for AR. The slight violation of the proportional hazard assumption by AR is acceptable. No further stratified analysis by time is conduced.

C-index for the prognostic model refitting in the test set using 500 bootstrap samples is 0.6309 (**Table 7**). The calibration of 1-year, 2-year, 3-year and 5-year survival are presented in **Figure 6**. The mean error for 3- year, 5-year survival is 0.008 and 0.009, respectively. The line corrected by 500 bootstrap samples (blue) is close to the diagonal line (grey) in the plot for 3-year and 5-year survival with 67.45% and 69.57% of AUC, respectively.

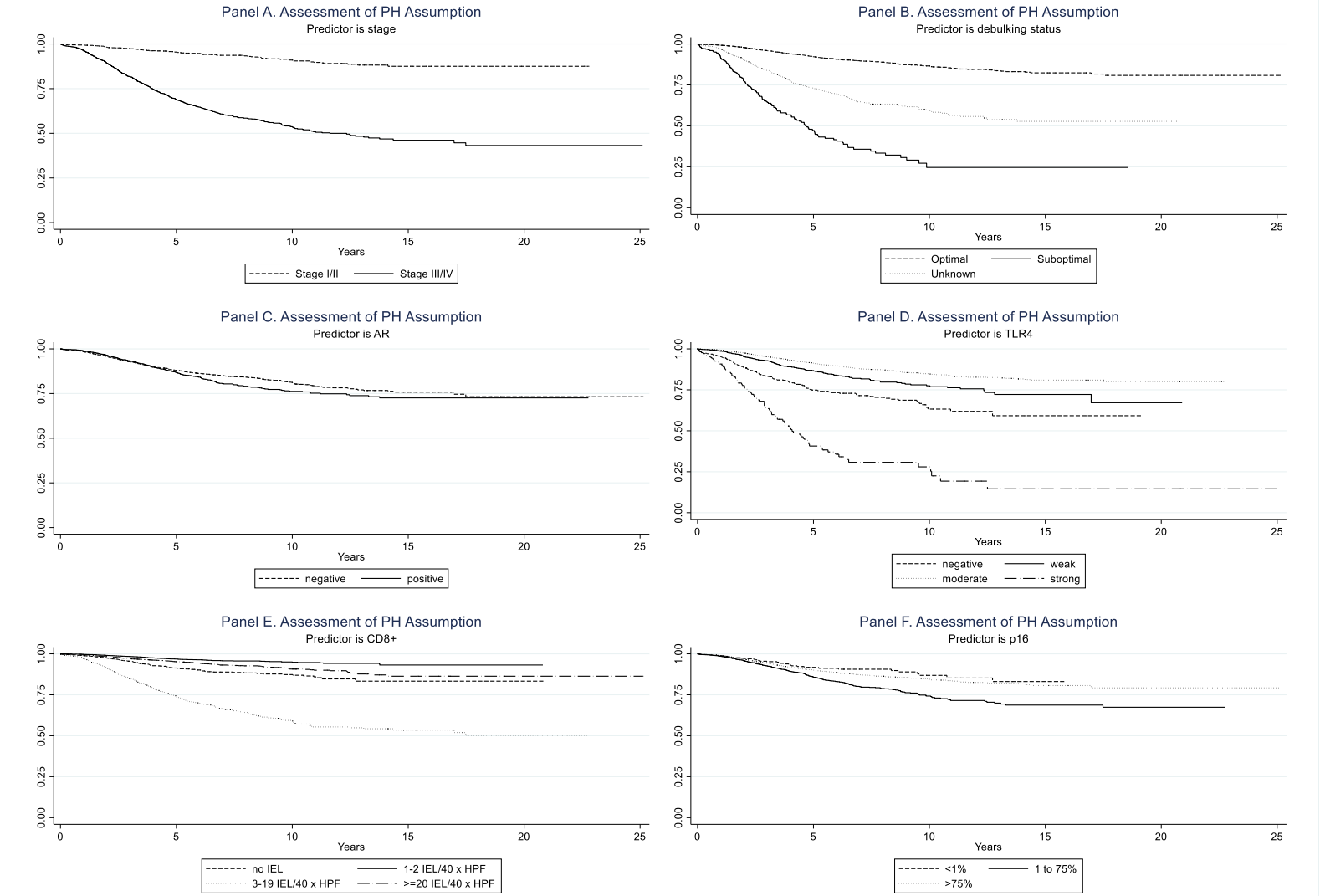

Figure 5 Kaplan-Meier curves comparing survival across categories of each variable from the final model adjusting for age and the other variables

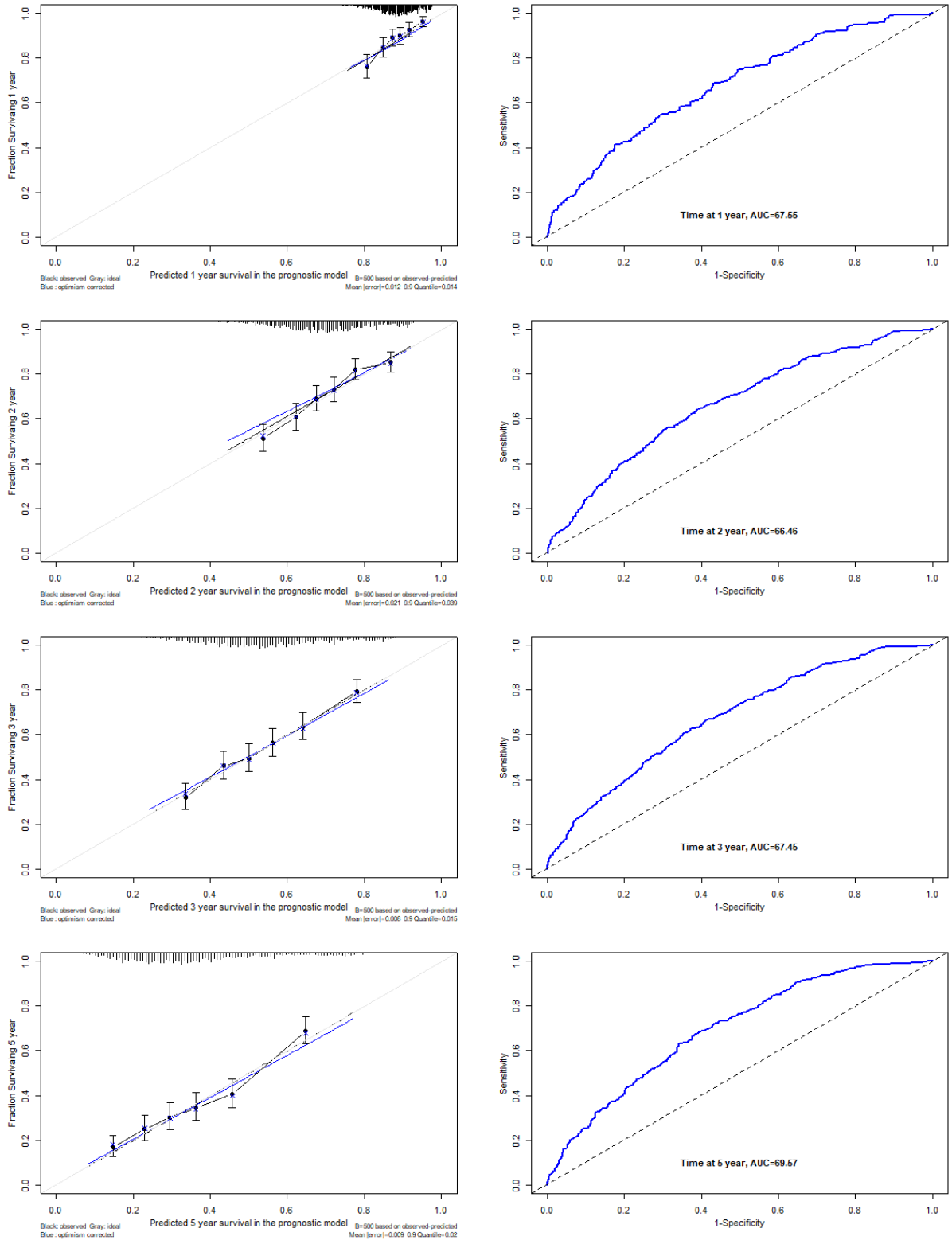

Figure 6 Calibration and ROC curves of 1-year, 2-year, 3-year and 5-year survival for the prognostic model in the test set

Based on the association of OS with the selected clinical and IHC variables, a nomogram was generated (**Figure 7**). To interpret the nomogram, for a HGSOC patient at stage I/II aged 60 with suboptimal debulking status, positive AR expression, moderate TLR4 expression, no IEL CD8 and p16 expression of 1-75% tumor cells, the patient got total task points of 107.5 (0+50+31+0+2.5+24+0). The estimated possibilities of 3 -year and 5-year survival of this patient was 0.7 and 0.54, respectively.

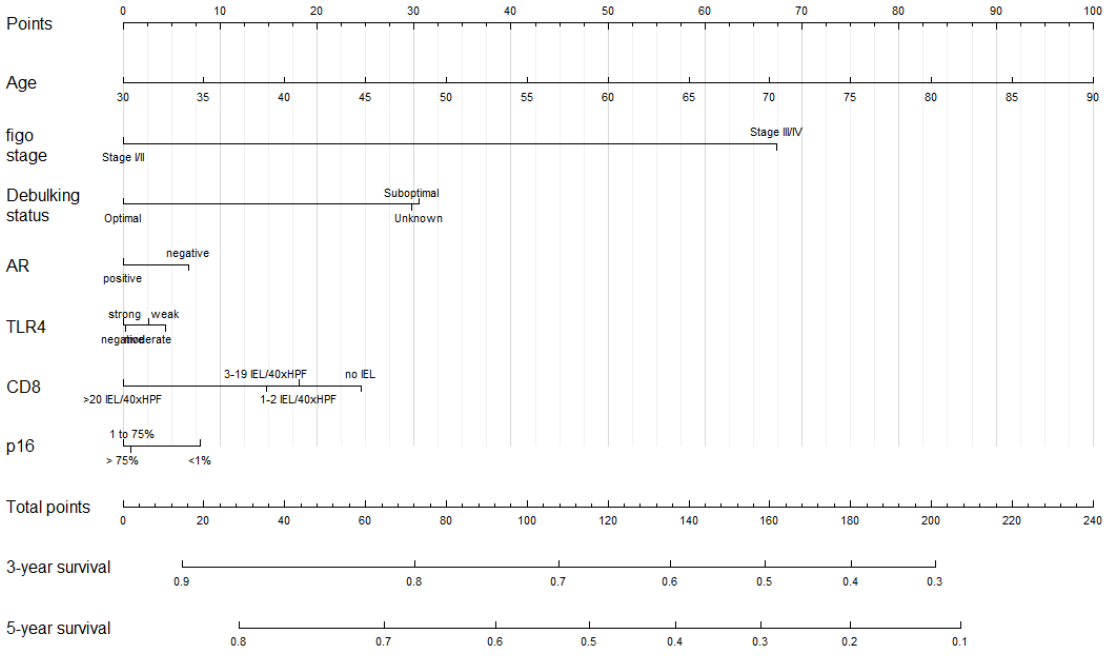

Figure 7 Nomogram of high-grade serous ovarian cancer

4.0 Discussion

4.1 Summary

The prognostic model includes age at diagnosis, stage, debulking status, AR, TLR4, CD8+ and p16. CD+8 shows a strong dose-response relationship with the survival of HGSOC patients. The C-index of the final CoxPHR model fit in the test set is 0.6309, which indicates a fair model. Even with the fair predictive power of the final model, the study proposes a novel approach to identify the important prognostic factors, which provides a potential opportunity to help clinical decision-making and extend the survival of HGSOC patients.

The nomogram containing selected IHC biomarkers helps HGSCO patients and the physicians calculate the possibility of survival and understand the disease visually. The accuracy of the prognostic model for 3-year and 5-year survival is 67.45% and 69.57%. The results should be discussed with the physicians to help patients understand the context. The nomogram is based on numbers of patients with HGSOC, and it cannot predict how long a particular patient will live.

4.2 Comparison with previous studies and potential biologic mechanism

Previous studies are published to estimate the association of individual IHC biomarker and HGSOC survival. However, there is no study building a prediction model for HGSOC survival by selecting prognostic factors from clinical variables and 9 IHC biomarkers.

AR activation plays a role in ovarian carcinogenesis via causing down-regulation of THF-beta pathway, upregulation of telomerase and epidermal growth factor receptor, increasing of IL-6, and IL-8 levels and other changes of downstream and upstream regulation (Mizushima & Miyamoto, 2019; Zhu et al., 2017). At the same time, AR expression shows a prognostic significance of ovarian cancer, especially in serous subtypes (Zhu et al., 2017). One study examining the association of AR expression and EOC survival in 154 EOC cases indicates that higher AR expression is a favorable prognostic factor (Nodin et al., 2010). Martins et al. using TCGA RNA-sequencing data also revealed that low AR expression is associated with shorter OS (F. C. Martins et al., 2014). The conclusion is consistent with the result of HGSOC survival from the current study, although the result does not reach statistical significance (AR positive vs AR negative HR 0.92, 95% CI 0.81-1.04, P=0.166). It is possible that AR expression affects HGSOC survival via involving in response to chemotherapy, but the mechanism is still unclear (Elattar et al., 2012; Sun, Huang, Lu, Chang, & Chao, 2015).

Although in the current analysis, there is no linear trend of TLR4 on HGSOC survival, moderate/strong TLR4 expression is associated with longer survival compared to negative/weak expression TLR4 expression (**Figure 7**). The OTTA study reveals that higher TLR4 expression is associated with longer survival in low-grade serous ovarian cancer, but no associated with HGSOC (Block et al., 2018). Even though the association of TLR4 expression and HGSOC does not reach statistical significance, TLR4 expression is still an important factor in the prognostic model. Studies show that positive TLR4 is correlated with platinum-sensitive patients (d'Adhemar et al., 2014; Luo, He, & Wang, 2015), thereby associating with longer OS of ovarian cancer. TLRs plays a crucial role in establishing innate immunity (Takeda & Akira, 2005). The TLR4 signaling pathway can be MyD88-dependent and MyD88-independent. The TIR domain-containing

adaptors, TRIF and TRAM are essential in the MyD88- independent TLR4 pathway. Activation of IRF-3 and NF-κB also participates in the pathway (Takeda & Akira, 2005). Some studies link MyD88-dependent TLR4 signaling pathways to chemoresistance in EOC (Kelly et al., 2006), and MyD88 expression to HGSOC survival (Block et al., 2018). While in the current study, only TLR4 is selected to build the prognostic model for HGSOC survival. It suggests that MyD88-independent TLR4 signaling pathways might play a crucial role in chemoresistance and affect HGSOC survival.

The biomarker of CD8+ TILs is selected in the prognostic model. CD8+ TILs density shows a highly significant dose-response association with HGSCOS survival. The result confirms the conclusion indicated by previous studies that CD8+ TILs density is a favorable prognostic factor for HGSCOS patients (Ovarian Tumor Tissue Analysis et al., 2017; Pinto et al., 2018; Sato et al., 2005). CD8+ TILs produce cytokines, such as tumor necrosis factor (TNF) and interferon-γ (IFN-γ), and enzymes, such as granzyme-B and perforin, to eliminate the targeted cells during the process of tumorigenesis. The immune response of CD8+ TILs interacts with other immune molecules, such as CD27 and MHC-II, affects clinical outcomes in ovarian cancer (Turner, Buchsbaum, Straughn, Randall, & Arend, 2016).

The biomarker p16 is considered a tumor suppressor (Lukas et al., 1995; Romagosa et al., 2011). Absence and overexpression of p16 are both abnormal. In the nomogram (**Figure 7**), the risk point of >75% of tumor cells nuclei positive weights over the risk point of <1% of tumor cells, but both are higher than the risk point of 1-75% of tumor cells nuclei positive. The eukaryotic cell cycle composes of four discrete phases, M, G_1, S and G_2. The G_1 phase is the interval between mitosis and initiation of DNA replication (Cooper, 2000). p16, as a negative cell cycle regulator, acts in the late G_1 phase to regulate the transition to the S phase (Kamb & McCormack, 2001).

Many studies show that p16 is downregulated in cancer (Fukushima et al., 2002; Guida et al., 2009; Moore et al., 2001). While p16 overexpression is also observed in some types of cancer, especially human papillomavirus (HPV) related cancer (Klaes et al., 2001; Mulvany, Allen, & Wilson, 2008; Volgareva et al., 2004). The overexpressing level of p16, as a prognostic marker, is various for different types of cancer. In breast cancer, p16 in 25% of the tumors is associated with unfavorable prognosis (Milde-Langosch, Bamberger, Rieck, Kelp, & Löning, 2001). In the current study, the overexpression of p16 is defined as over 75% of tumor cells nuclei positive. It is one of the predictors in the prognostic model but does not reach statistical significance on HGSOC survival compared to <1% and 1%-75% of tumor cells nuclei positive.

4.3 Limitations

Ideally, the quantitative IHC score should be used in the prognostic model (Galon et al., 2012). However, such data cannot be obtained. For the different IHC biomarkers, different score systems are used to keep the ordinal variables with as many levels as possible. AR is the only dichotomous variable. ER, PR and p16 use a 3-tiered system. TLR4, MyD88, PTEN and CD8+ use a 4-tiered system. FOLR1 uses a 6-tiered system. The other issue related to IHC scores is that the inter-observer agreement for HGSOC subset from OTTA is not obtained. The studies using available IHC data of all histologic subtypes of ovarian cancer from OTTA indicate that 2-score system has a higher inter-observer agreement for MyD88, TLR4 and CD8+ (Block et al., 2018; Ovarian Tumor Tissue Analysis et al., 2017). The OTTA study estimating the association of FOLR1 and ovarian cancer survival indicates a 5-tiered system, combing “absent” and “weak” together, is as high as 0.9 in a subset of 183 patients (Köbel et al., 2014).

The other limitation of the current study is that *glmnet* package cannot select ordinal variables with all levels together. However, the internal validation, C-index, of the model fitted by the selected levels only is not different from that of the model fitted by the selected predictors with all levels (0.6922 vs 0.6863). Due to the shortcoming of the package itself, no interaction term is added in the CoxPHR model with *lasso* penalty to improve the interpretability of variable selection. The *gglasso* package was designed by Yang, et al to solve group-lass learning problems (Yang & Zou, 2015). Categorical variables with more than two levels can be selected in the model using this package. In the current study using *glmnet* package, the dummy variables were selected first and then the dummy variables with other dummy variables grouped in the categorical variable were refitted in the model. However, the *gglasso* package cannot by applied to CoxPHR.

The missing pattern of IHC biomarkers causes the discrepancy in the sample sizes of the training set and the test set. The IHC biomarkers were selected using the training set with all 9 biomarkers. Compared to the training set, the test set has more HGSOC in stage III/IV and the median OS is much shorter. To explore whether the availability of markers affecting the overall survival through stage, the relationships between availability of each IHC biomarkers and stage are estimated (Table 8). The absolute Cramér's V values ranged from 0.025 to 0.13, indicating weak relationship of availability of each IHC biomarkers and stage. Thus, the conflicting prediction power of the prognostic model in the training set and in the test set is not due to the availability of each IHC biomarkers. The missing pattern of IHC biomarkers affects the variable selection slightly.

Table 8 Cramér's V of availiablity of each IHC biomarkers and stage

	AR	ER	PR	TLR4	MyD88	FOLR1	PTEN	CD8	p16
Stage	-0.026	0.13	0.039	0.037	0.067	0.075	0.046	0.025	0.042

Other covariates that might be related to some IHC biomarkers and HGSOC survival are not included in the variable selection due to the inaccessibility of the data. As mentioned before, the CA-125 level is a prognostic factor for ovarian cancer survival. The types of treatment might be associated with ovarian cancer survival. Patients with stage III/IV ovarian cancer who received neoadjuvant chemotherapy (NACT) but not followed by debulking surgery have worse OS compared to women with NACT and debulking surgery (Y. L. Liu et al., 2020). A meta-analysis including 15 cohort studies indicates that early initiation of chemotherapy is associated with longer OS of ovarian cancer patients (Y. Liu et al., 2017). A complete response to chemotherapy after NACT improves OS in HGSOC patients (Cohen et al., 2019; Santoro et al., 2019). Previous studies also reveal that BRCA1/2 mutations are associated with long-term survival among all ovarian cancer cases and HGSOC cases (Chetrit et al., 2008; Huang, 2018; S. I. Kim et al., 2019). However, in the training set, only 2 cases have data on BRACA1/2 mutation status. Thus, BRACA1/2 mutation status is also not included in the current analysis.

Other potential IHC candidates for HGSOC prognosis need to be evaluated. CD3+ and CD4+ TILs indicating the immune response against cancer cells are found to associate with OS in many tumor types, such as breast cancer, colon cancer and cervical cancer (Ancuta et al., 2009; Hadrup, Donia, & Thor Straten, 2013; Nedergaard, Ladekarl, Thomsen, Nyengaard, & Nielsen, 2007; Rathore et al., 2014). CD20+ B lymphocytes is associated with a favourable prognosis in lung cancer and gastric cancer (Al-Shibli et al., 2008; Hennequin et al., 2016). Adding IHC variables related to ovarian cancer and other cancer in the step of variable selection might increase the accuracy of the final prognostic model and help identify new prognostic biomarkers for HGSOC.

4.4 Conclusion and future study

The final prognostic model indicates that the four IHC biomarkers, AR, TLR4, CD8+ and p16, play more crucial roles in the prognosis of HGSOC than the other IHC biomarkers, ER, PR, MyD88, FOLR1 and PTEN based on a statistical algorithm. The potential biologic mechanism needs to be investigated to explain why AR, TLR4, CD8+ and p16 weights over ER, PR, MyD88, FOLR1 and PTEN to predict HGSOC survival. Based on the limitations future studies are recommended treating the IHC biomarkers as continuous values and considering other IHC candidates and clinical variables, such as adjuvant chemotherapy or not, chemotherapy response, pre-treatment and post-treatment CA125 level, BRCA1 and BRCA2 status, as well as other.

The *glmnet* package with its vignette is an easy-to-use and efficient package when conducting lasso and elastic-net regularized regression. However, it cannot achieve a model with *grouped lasso* penalty, and cannot deal with ordinal predictors. Although cross-validation is built in the package to select the tuning parameters, external validation still needs to be conducted by refitting the model and using other packages. A comprehensive R package fitting different types of survival and predictor data in regularized CoxPHR analysis should be developed.

Appendix A Characteristics of ovarian cancer studies included in the current analysis

Appendix Table 1 Characteristics of 17 Ovarian Cancer Studies from the Ovarian Tumor Tissue Analysis (OTTA) consortium

Site	Study Name	Region	Study Period	Case ascertainment	Training set			Test set		
					Total Cases, n	Alived, n (%)	Dead, n (%)	Total Cases, n	Alived, n (%)	Dead, n (%)
AOV	Alberta Ovarian Tumor Types Study	Canada	1978-2010	Populaiton-based prospective Alberta Cancer Registry	-	-	-	69	30 (43.48)	39 (56.52)
BAV	Bavarian Ovarian Cancer Cases and Controls	Germany	2002-2006	Gynecologic Oncology Center at the Comprehensive Cancer Center Erlangen-Nuremberg	-	-	-	141	35 (24.82)	106 (75.18)
BRZ	Brazil Gynecologic Tumor Bank Study	Brazil	1987-2010	University Hospital of Ribeirao Preto School of Medicine (HCRP), case series with prospective follow up	-	-	-	55	24 (43.64)	31 (56.36)
CAL	Calgary Serous Carcinoma Study	Canada	2003–2007	Hospital-based retrospective observational study	-	-	-	71	15 (21.13)	56 (78.87)
CNI	CNIO Ovarian Cancer Study	Spain	2006-2013	Hospitals in Medical Oncology Divisions	-	-	-	51	34 (66.67)	17 (33.33)
GER	German Ovarian Cancer Study	Germany	1993-1996	Hospital admission in the study areas	-	-	-	56	7 (12.50)	49 (87.50)
HAW	Hawaii Ovarian Cancer Case-Control Study	USA	1993-2008	Hawaii Tumor Registry	-	-	-	68	19 (27.94)	49 (72.06)
HOP	Hormones and Ovarian cancer PrEdiction	USA	2003-2009	Hospital registries and active surveillance of medical practices in three catchment areas	23	8 (34.78)	15 (65.22)	3	2 (66.67)	1 (33.33)
LAX	Women's Cancer Program at the Samuel Oschin	USA	1989 - present	Women's Cancer Program Biorepository	-	-	-	189	55 (29.10)	134 (70.90)

	Comprehensive Cancer Institute									
MAL	MALignant OVArian cancer	Denmark	1994-1999	Gynecological departments in Copenhagen, Frederiksberg and seven surrounding counties	3	0 (0.00)	3 (100.00)	3	0 (0.00)	3 (100.00)
MAY	Mayo Clinic Ovarian Cancer Study	USA	2000–2009	Mayo Clinic medical records and State death certificates	-	-	-	163	28 (17.18)	135 (82.82)
NOT	Nottingham Study	UK	1991-2008	Hospital records and Trent cancer registry	41	8 (19.51)	33 (80.49)	73	11 (15.07)	62 (84.93)
POC	Polish Ovarian Cancer Study	Poland	1998-2006	Pathology departments of Szczecin, and main oncology hospitals of Poznan, Opole and Rzeszów	-	-	-	70	29 (41.43)	41 (58.57)
TUE	Tuebingen University Women's Hospital study	Germany	1999-2008	Department of Obstetrics and Gynaecology, Eberhard Karls Universitats Tübingen, Tübingen Germany	-	-	-	143	46 (32.17)	97 (67.83)
UKO	United Kingdom Ovarian Cancer Population Study	UK	2006–2010	Ten major Gynaecological Oncology NHS centers in England, Wales and Northern Ireland	59	22 (37.29)	37 (62.71)	8	1 (12.50)	7 (87.50)
VAN	OVCARE	Canada	2003-present	VGH and/or BC Cancer Agency Division of Gynecologic Oncology	128	52 (40.63)	76 (59.38)	261	72 (27.59)	189 (72.41)
WMH	Westmead Hospital: Molecular Biology of Gynaecologic Disease	Australia	1992-2012	The Crown Princess Mary Cancer Centre and affiliated hospitals	-	-	-	139	20 (14.39)	119 (85.61)
Total	-	-	-	-	254	90 (35.43)	164 (64.57)	1563	428 (27.38)	1135 (72.62)

Bibliography

Akeson, M., Jakobsen, A. M., Zetterqvist, B. M., Holmberg, E., Brännström, M., & Horvath, G. (2009). A population-based 5-year cohort study including all cases of epithelial ovarian cancer in western Sweden: 10-year survival and prognostic factors. *Int J Gynecol Cancer, 19*(1), 116-123. doi:10.1111/IGC.0b013e3181991b13

Al-Shibli, K. I., Donnem, T., Al-Saad, S., Persson, M., Bremnes, R. M., & Busund, L. T. (2008). Prognostic effect of epithelial and stromal lymphocyte infiltration in non-small cell lung cancer. *Clin Cancer Res, 14*(16), 5220-5227. doi:10.1158/1078-0432.Ccr-08-0133

Aluloski, I., Tanturovski, M., Jovanovic, R., Kostadinova-Kunovska, S., Petrusevska, G., Stojkovski, I., & Petreska, B. (2017). Survival of Advanced Stage High-Grade Serous Ovarian Cancer Patients in the Republic of Macedonia. *Open access Macedonian journal of medical sciences, 5*(7), 904-908. doi:10.3889/oamjms.2017.215

American Cancer Society. (April 11, 2018). Surgery for Ovarian Cancer. Retrieved from https://www.cancer.org/cancer/ovarian-cancer/treating/surgery.html#:~:text=Debulking%20is%20very%20important%20when,This%20is%20called%20optimally%20debulked

American Cancer Society. (2020). Cancer Facts & Figures 2020. Retrieved from https://www.cancer.org/cancer/ovarian-cancer/about/key-statistics.html

Ancuta, E., Ancuţa, C., Zugun-Eloae, F., Iordache, C., Chirieac, R., & Carasevici, E. (2009). Predictive value of cellular immune response in cervical cancer. *Rom J Morphol Embryol, 50*(4), 651-655.

Bakkar, R. M., Xie, S. S., Urbauer, D. L., Djordjevic, B., Vu, K., & Broaddus, R. R. (2015). Intact PTEN Expression by Immunohistochemistry is Associated With Decreased Survival in Advanced Stage Ovarian/Primary Peritoneal High-grade Serous Carcinoma. *Int J Gynecol Pathol, 34*(6), 497-506. doi:10.1097/pgp.0000000000000205

Baldwin, L. A., Huang, B., Miller, R. W., Tucker, T., Goodrich, S. T., Podzielinski, I., . . . Seamon, L. G. (2012). Ten-year relative survival for epithelial ovarian cancer. *Obstet Gynecol, 120*(3), 612-618. doi:10.1097/AOG.0b013e318264f794

Balogun, N., Gentry-Maharaj, A., Wozniak, E. L., Lim, A., Ryan, A., Ramus, S. J., . . . Menon, U. (2011). Recruitment of newly diagnosed ovarian cancer patients proved challenging in a multicentre biobanking study. *J Clin Epidemiol, 64*(5), 525-530. doi:10.1016/j.jclinepi.2010.07.008

Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D. W., Dao, F., . . . Data coordination, c. (2011). Integrated genomic analyses of ovarian carcinoma. *Nature, 474*(7353), 609-615. doi:10.1038/nature10166

Block, M. S., Vierkant, R. A., Rambau, P. F., Winham, S. J., Wagner, P., Traficante, N., . . . Goode, E. L. (2018). MyD88 and TLR4 Expression in Epithelial Ovarian Cancer. *Mayo Clinic proceedings, 93*(3), 307-320. doi:10.1016/j.mayocp.2017.10.023

Bowtell, D. D., Böhm, S., Ahmed, A. A., Aspuria, P. J., Bast, R. C., Jr., Beral, V., . . . Balkwill, F. R. (2015). Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nat Rev Cancer, 15*(11), 668-679. doi:10.1038/nrc4019

Bromley, A. B., Altman, A. D., Chu, P., Nation, J. G., Nelson, G. S., Ghatage, P., . . . Köbel, M. (2012). Architectural patterns of ovarian/pelvic high-grade serous carcinoma. *Int J Gynecol Pathol, 31*(5), 397-404. doi:10.1097/PGP.0b013e31824c2372

Carlson, R. H. (2003). Ovarian Cancer: Adjuvant Chemo Reduces Recurrence, But So Does Adequate Staging. *Oncology Times, 25*(6). Retrieved from https://journals.lww.com/oncology-times/Fulltext/2003/03250/Ovarian_Cancer__Adjuvant_Chemo_Reduces_Recurrence,.2.aspx

Chang, S. J., Bristow, R. E., & Ryu, H. S. (2012). Impact of complete cytoreduction leaving no gross residual disease associated with radical cytoreductive surgical procedures on survival in advanced ovarian cancer. *Ann Surg Oncol, 19*(13), 4059-4067. doi:10.1245/s10434-012-2446-8

Chang, S. J., Hodeib, M., Chang, J., & Bristow, R. E. (2013). Survival impact of complete cytoreduction to no gross residual disease for advanced-stage ovarian cancer: a meta-analysis. *Gynecol Oncol, 130*(3), 493-498. doi:10.1016/j.ygyno.2013.05.040

Chetrit, A., Hirsh-Yechezkel, G., Ben-David, Y., Lubin, F., Friedman, E., & Sadetzki, S. (2008). Effect of BRCA1/2 mutations on long-term survival of patients with invasive ovarian cancer: the national Israeli study of ovarian cancer. *J Clin Oncol, 26*(1), 20-25. doi:10.1200/jco.2007.11.6905

Cohen, P. A., Powell, A., Böhm, S., Gilks, C. B., Stewart, C. J. R., Meniawy, T. M., . . . Singh, N. (2019). Pathological chemotherapy response score is prognostic in tubo-ovarian high-grade serous carcinoma: A systematic review and meta-analysis of individual patient data. *Gynecol Oncol, 154*(2), 441-448. doi:10.1016/j.ygyno.2019.04.679

Cooper, G. M. (2000). *The cell : a molecular approach*. Washington, D.C.; Sunderland, Mass.: ASM Press ; Sinauer Associates.

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society: Series B (Methodological), 34*(2), 187-202. doi:10.1111/j.2517-6161.1972.tb00899.x

Cox, D. R. (1975). Partial likelihood. *Biometrika, 62*(2), 269-276. doi:10.1093/biomet/62.2.269

d'Adhemar, C. J., Spillane, C. D., Gallagher, M. F., Bates, M., Costello, K. M., Barry-O'Crowley, J., . . . O'Leary, J. J. (2014). The MyD88+ Phenotype Is an Adverse Prognostic Factor in Epithelial Ovarian Cancer. *PLOS ONE, 9*(6), e100816. doi:10.1371/journal.pone.0100816

Dao, F., Schlappe, B. A., Tseng, J., Lester, J., Nick, A. M., Lutgendorf, S. K., . . . Levine, D. A. (2016). Characteristics of 10-year survivors of high-grade serous ovarian carcinoma. *Gynecologic Oncology, 141*(2), 260-263. doi:https://doi.org/10.1016/j.ygyno.2016.03.010

Efroymson, M. A. (1960). Multiple regression analysis. *Mathematical Methods for Digital Computers, Ralston A. and Wilf,H. S., (eds.), Wiley, New York.*

Elattar, A., Warburton, K. G., Mukhopadhyay, A., Freer, R. M., Shaheen, F., Cross, P., . . . Edmondson, R. J. (2012). Androgen receptor expression is a biological marker for androgen sensitivity in high grade serous epithelial ovarian cancer. *Gynecol Oncol, 124*(1), 142-147. doi:10.1016/j.ygyno.2011.09.004

Emmanuel, C., Chiew, Y. E., George, J., Etemadmoghadam, D., Anglesio, M. S., Sharma, R., . . . deFazio, A. (2014). Genomic classification of serous ovarian cancer with adjacent borderline differentiates RAS pathway and TP53-mutant tumors and identifies NRAS as an oncogenic driver. *Clin Cancer Res, 20*(24), 6618-6630. doi:10.1158/1078-0432.Ccr-14-1292

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *2010, 33*(1), 22. doi:10.18637/jss.v033.i01

Fukushima, N., Sato, N., Ueki, T., Rosty, C., Walter, K. M., Wilentz, R. E., . . . Goggins, M. (2002). Aberrant methylation of preproenkephalin and p16 genes in pancreatic intraepithelial neoplasia and pancreatic ductal adenocarcinoma. *Am J Pathol, 160*(5), 1573-1581. doi:10.1016/s0002-9440(10)61104-2

Galon, J., Pagès, F., Marincola, F. M., Angell, H. K., Thurin, M., Lugli, A., . . . Fox, B. A. (2012). Cancer classification using the Immunoscore: a worldwide task force. *J Transl Med, 10*, 205. doi:10.1186/1479-5876-10-205

Garcia-Closas, M., Brinton, L. A., Lissowska, J., Richesson, D., Sherman, M. E., Szeszenia-Dabrowska, N., . . . Chanock, S. J. (2007). Ovarian cancer risk and common variation in the sex hormone-binding globulin gene: a population-based case-control study. *BMC cancer, 7*, 60. doi:10.1186/1471-2407-7-60

Gareth James, D. W. T. H. R. T. (2013). *An introduction to statistical learning : with applications in R*: New York : Springer, [2013] ©2013.

Glud, E., Kjaer, S. K., Thomsen, B. L., Høgdall, C., Christensen, L., Høgdall, E., . . . Blaakaer, J. (2004). Hormone therapy and the impact of estrogen intake on the risk of ovarian cancer. *Arch Intern Med, 164*(20), 2253-2259. doi:10.1001/archinte.164.20.2253

Goode, E. L., Maurer, M. J., Sellers, T. A., Phelan, C. M., Kalli, K. R., Fridley, B. L., . . . Hartmann, L. C. (2010). Inherited determinants of ovarian cancer survival. *Clin Cancer Res, 16*(3), 995-1007. doi:10.1158/1078-0432.Ccr-09-2553

Goodman, M. T., Lurie, G., Thompson, P. J., McDuffie, K. E., & Carney, M. E. (2008). Association of two common single-nucleotide polymorphisms in the CYP19A1 locus and ovarian cancer risk. *Endocr Relat Cancer, 15*(4), 1055-1060. doi:10.1677/erc-08-0104

Guida, M., Sanguedolce, F., Bufo, P., Di Spiezio Sardo, A., Bifulco, G., Nappi, C., & Pannone, G. (2009). Aberrant DNA hypermethylation of hMLH-1 and CDKN2A/p16 genes in benign, premalignant and malignant endometrial lesions. *Eur J Gynaecol Oncol, 30*(3), 267-270.

Gupta, D., & Lis, C. G. (2009). Role of CA125 in predicting ovarian cancer survival - a review of the epidemiological literature. *Journal of ovarian research, 2*, 13-13. doi:10.1186/1757-2215-2-13

Gurung, A., Hung, T., Morin, J., & Gilks, C. B. (2013). Molecular abnormalities in ovarian carcinoma: clinical, morphological and therapeutic correlates. *Histopathology, 62*(1), 59-70. doi:10.1111/his.12033

Hadrup, S., Donia, M., & Thor Straten, P. (2013). Effector CD4 and CD8 T cells and their role in the tumor microenvironment. *Cancer Microenviron, 6*(2), 123-133. doi:10.1007/s12307-012-0127-6

Harrell, F. E. (2006). *Regression Modeling Strategies*: Springer-Verlag.

Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *Jama, 247*(18), 2543-2546.

Hennequin, A., Derangère, V., Boidot, R., Apetoh, L., Vincent, J., Orry, D., . . . Ladoire, S. (2016). Tumor infiltration by Tbet+ effector T cells and CD20+ B cells is associated with survival in gastric cancer patients. *Oncoimmunology, 5*(2), e1054598. doi:10.1080/2162402x.2015.1054598

Huang, Y. W. (2018). Association of BRCA1/2 mutations with ovarian cancer prognosis: An updated meta-analysis. *Medicine (Baltimore), 97*(2), e9380. doi:10.1097/md.0000000000009380

Inciura, A., Simavicius, A., Juozaityte, E., Kurtinaitis, J., Nadisauskiene, R., Svedas, E., & Kajenas, S. (2006). Comparison of adjuvant and neoadjuvant chemotherapy in the management of advanced ovarian cancer: a retrospective study of 574 patients. *BMC cancer, 6*, 153-153. doi:10.1186/1471-2407-6-153

Jönsson, J.-M., Arildsen, N. S., Malander, S., Måsbäck, A., Hartman, L., Nilbert, M., & Hedenfalk, I. (2015). Sex Steroid Hormone Receptor Expression Affects Ovarian Cancer Survival. *Translational oncology, 8*(5), 424-433. doi:10.1016/j.tranon.2015.09.002

Kamb, A., & McCormack, K. (2001). p16 Tumor Suppressor. In D. E. Fisher (Ed.), *Tumor Suppressor Genes in Human Cancer* (pp. 183-195). Totowa, NJ: Humana Press.

Kamieniak, M. M., Rico, D., Milne, R. L., Muñoz-Repeto, I., Ibáñez, K., Grillo, M. A., . . . García, M. J. (2015). Deletion at 6q24.2-26 predicts longer survival of high-grade serous epithelial ovarian cancer patients. *Mol Oncol, 9*(2), 422-436. doi:10.1016/j.molonc.2014.09.010

Kelemen, L. E., Köbel, M., Chan, A., Taghaddos, S., & Dinu, I. (2013). Differentially methylated loci distinguish ovarian carcinoma histological types: evaluation of a DNA methylation assay in FFPE tissue. *Biomed Res Int, 2013*, 815894. doi:10.1155/2013/815894

Kelemen, L. E., Sellers, T. A., Schildkraut, J. M., Cunningham, J. M., Vierkant, R. A., Pankratz, V. S., . . . Goode, E. L. (2008). Genetic variation in the one-carbon transfer pathway and ovarian cancer risk. *Cancer Res, 68*(7), 2498-2506. doi:10.1158/0008-5472.Can-07-5165

Kelly, M. G., Alvero, A. B., Chen, R., Silasi, D.-A., Abrahams, V. M., Chan, S., . . . Mor, G. (2006). TLR-4 Signaling Promotes Tumor Growth and Paclitaxel Chemoresistance in Ovarian Cancer. *Cancer Research, 66*(7), 3859-3868. doi:10.1158/0008-5472.Can-05-3948

Kim, J., Park, E. Y., Kim, O., Schilder, J. M., Coffey, D. M., Cho, C.-H., & Bast, R. C., Jr. (2018). Cell Origins of High-Grade Serous Ovarian Cancer. *Cancers, 10*(11), 433. doi:10.3390/cancers10110433

Kim, S. I., Lee, M., Kim, H. S., Chung, H. H., Kim, J. W., Park, N. H., & Song, Y. S. (2019). Effect of BRCA mutational status on survival outcome in advanced-stage high-grade serous ovarian cancer. *Journal of ovarian research, 12*(1), 40. doi:10.1186/s13048-019-0511-7

Klaes, R., Friedrich, T., Spitkovsky, D., Ridder, R., Rudy, W., Petry, U., . . . von Knebel Doeberitz, M. (2001). Overexpression of p16(INK4A) as a specific marker for dysplastic and neoplastic epithelial cells of the cervix uteri. *Int J Cancer, 92*(2), 276-284. doi:10.1002/ijc.1174

Köbel, M., Madore, J., Ramus, S. J., Clarke, B. A., Pharoah, P. D. P., Deen, S., . . . Kelemen, L. E. (2014). Evidence for a time-dependent association between FOLR1 expression and survival from ovarian carcinoma: implications for clinical testing. An Ovarian Tumour Tissue Analysis consortium study. *British journal of cancer, 111*(12), 2297-2307. doi:10.1038/bjc.2014.567

Köbel, M., Reuss, A., du Bois, A., Kommoss, S., Kommoss, F., Gao, D., . . . Gilks, C. B. (2010). The biological and clinical value of p53 expression in pelvic high-grade serous carcinomas. *J Pathol, 222*(2), 191-198. doi:10.1002/path.2744

Kurman, R. J., & Shih Ie, M. (2010). The origin and pathogenesis of epithelial ovarian cancer: a proposed unifying theory. *Am J Surg Pathol, 34*(3), 433-443. doi:10.1097/PAS.0b013e3181cf3d79

Lai, Y., Hayashida, M., & Akutsu, T. (2013). Survival analysis by penalized regression and matrix factorization. *ScientificWorldJournal, 2013*, 632030. doi:10.1155/2013/632030

Liu, Y., Zhang, T., Wu, Q., Jiao, Y., Gong, T., Ma, X., & Li, D. (2017). Relationship between initiation time of adjuvant chemotherapy and survival in ovarian cancer patients: a dose-response meta-analysis of cohort studies. *Scientific reports, 7*(1), 9461-9461. doi:10.1038/s41598-017-10197-1

Liu, Y. L., Filippova, O. T., Zhou, Q., Iasonos, A., Chi, D. S., Zivanovic, O., . . . Tew, W. P. (2020). Characteristics and survival of ovarian cancer patients treated with neoadjuvant chemotherapy but not undergoing interval debulking surgery. *Journal of gynecologic oncology, 31*(1), e17-e17. doi:10.3802/jgo.2020.31.e17

Lo-Ciganic, W. H., Zgibor, J. C., Bunker, C. H., Moysich, K. B., Edwards, R. P., & Ness, R. B. (2012). Aspirin, nonaspirin nonsteroidal anti-inflammatory drugs, or acetaminophen and risk of ovarian cancer. *Epidemiology, 23*(2), 311-319. doi:10.1097/EDE.0b013e3182456ad3

Lukas, J., Parry, D., Aagaard, L., Mann, D. J., Bartkova, J., Strauss, M., . . . Bartek, J. (1995). Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature, 375*(6531), 503-506. doi:10.1038/375503a0

Luo, X. Z., He, Q. Z., & Wang, K. (2015). Expression of Toll-like receptor 4 in ovarian serous adenocarcinoma and correlation with clinical stage and pathological grade. *Int J Clin Exp Med, 8*(8), 14323-14327.

Lurie, G., Wilkens, L. R., Thompson, P. J., McDuffie, K. E., Carney, M. E., Terada, K. Y., & Goodman, M. T. (2008). Combined oral contraceptive use and epithelial ovarian cancer risk: time-related effects. *Epidemiology, 19*(2), 237-243. doi:10.1097/EDE.0b013e31816334c5

Martins, F. C., Santiago, I., Trinh, A., Xian, J., Guo, A., Sayal, K., . . . Brenton, J. D. (2014). Combined image and genomic analysis of high-grade serous ovarian cancer reveals PTEN loss as a common driver event and prognostic classifier. *Genome biology, 15*(12), 526. doi:10.1186/s13059-014-0526-8

Martins, F. C., Santiago, I. d., Trinh, A., Xian, J., Guo, A., Sayal, K., . . . Brenton, J. D. (2014). Combined image and genomic analysis of high-grade serous ovarian cancer reveals PTEN loss as a common driver event and prognostic classifier. *Genome biology, 15*(12), 526-526. doi:10.1186/s13059-014-0526-8

Matulonis, U. A., Sood, A. K., Fallowfield, L., Howitt, B. E., Sehouli, J., & Karlan, B. Y. (2016). Ovarian cancer. *Nature reviews. Disease primers, 2*, 16061-16061. doi:10.1038/nrdp.2016.61

Milde-Langosch, K., Bamberger, A. M., Rieck, G., Kelp, B., & Löning, T. (2001). Overexpression of the p16 cell cycle inhibitor in breast cancer is associated with a more malignant phenotype. *Breast Cancer Res Treat, 67*(1), 61-70. doi:10.1023/a:1010623308275

Moore, P. S., Sipos, B., Orlandini, S., Sorio, C., Real, F. X., Lemoine, N. R., . . . Scarpa, A. (2001). Genetic profile of 22 pancreatic carcinoma cell lines. Analysis of K-ras, p53, p16 and DPC4/Smad4. *Virchows Arch, 439*(6), 798-802. doi:10.1007/s004280100474

Mulvany, N. J., Allen, D. G., & Wilson, S. M. (2008). Diagnostic utility of p16INK4a: a reappraisal of its use in cervical biopsies. *Pathology, 40*(4), 335-344. doi:10.1080/00313020802035907

National Cancer Institute. (March 9, 2015). Cancer Staging. Retrieved from https://www.cancer.gov/about-cancer/diagnosis-staging/staging

Nedergaard, B. S., Ladekarl, M., Thomsen, H. F., Nyengaard, J. R., & Nielsen, K. (2007). Low density of CD3+, CD4+ and CD8+ cells is associated with increased risk of relapse in squamous cell cervical cancer. *British journal of cancer, 97*(8), 1135-1138. doi:10.1038/sj.bjc.6604001

Nelder, J. A., & Wedderburn, R. W. M. (1972). Generalized Linear Models. *Journal of the Royal Statistical Society: Series A (General), 135*(3), 370-384. doi:10.2307/2344614

Nero, C., Ciccarone, F., Pietragalla, A., & Scambia, G. (2019). PTEN and Gynecological Cancers. *Cancers, 11*(10), 1458. doi:10.3390/cancers11101458

Nodin, B., Zendehrokh, N., Brändstedt, J., Nilsson, E., Manjer, J., Brennan, D. J., & Jirström, K. (2010). Increased androgen receptor expression in serous carcinoma of the ovary is associated with an improved survival. *Journal of ovarian research, 3*(1), 14. doi:10.1186/1757-2215-3-14

Olsen, C. M., Nagle, C. M., Whiteman, D. C., Ness, R., Pearce, C. L., Pike, M. C., . . . Webb, P. M. (2013). Obesity and risk of ovarian cancer subtypes: evidence from the Ovarian Cancer Association Consortium. *Endocr Relat Cancer, 20*(2), 251-262. doi:10.1530/erc-12-0395

Ovarian Tumor Tissue Analysis, C., Goode, E. L., Block, M. S., Kalli, K. R., Vierkant, R. A., Chen, W., . . . Ramus, S. J. (2017). Dose-Response Association of CD8+ Tumor-Infiltrating Lymphocytes and Survival Time in High-Grade Serous Ovarian Cancer. *JAMA oncology, 3*(12), e173290-e173290. doi:10.1001/jamaoncol.2017.3290

Peres, L. C., Cushing-Haugen, K. L., Köbel, M., Harris, H. R., Berchuck, A., Rossing, M. A., . . . Doherty, J. A. (2019). Invasive Epithelial Ovarian Cancer Survival by Histotype and Disease Stage. *Journal of the National Cancer Institute, 111*(1), 60-68. doi:10.1093/jnci/djy071

Peres, L. C., Cushing-Haugen, K. L., Köbel, M., Harris, H. R., Berchuck, A., Rossing, M. A., . . . Doherty, J. A. (2019). Invasive Epithelial Ovarian Cancer Survival by Histotype and Disease Stage. *Journal of the National Cancer Institute, 111*(1), 60-68. doi:10.1093/jnci/djy071

Pinto, M. P., Balmaceda, C., Bravo, M. L., Kato, S., Villarroel, A., Owen, G. I., . . . Ibañez, C. (2018). Patient inflammatory status and CD4+/CD8+ intraepithelial tumor lymphocyte

infiltration are predictors of outcomes in high-grade serous ovarian cancer. *Gynecologic Oncology, 151*(1), 10-17. doi:https://doi.org/10.1016/j.ygyno.2018.07.025

Prentice, L. M., Klausen, C., Kalloger, S., Köbel, M., McKinney, S., Santos, J. L., . . . Aparicio, S. A. (2007). Kisspeptin and GPR54 immunoreactivity in a cohort of 518 patients defines favourable prognosis and clear cell subtype in ovarian carcinoma. *BMC Med, 5*, 33. doi:10.1186/1741-7015-5-33

Rambau, P. F., Vierkant, R. A., Intermaggio, M. P., Kelemen, L. E., Goodman, M. T., Herpel, E., . . . Köbel, M. (2018). Association of p16 expression with prognosis varies across ovarian carcinoma histotypes: an Ovarian Tumor Tissue Analysis consortium study. *The journal of pathology. Clinical research, 4*(4), 250-261. doi:10.1002/cjp2.109

Ramus, S. J., Antoniou, A. C., Kuchenbaecker, K. B., Soucy, P., Beesley, J., Chen, X., . . . Chenevix-Trench, G. (2012). Ovarian cancer susceptibility alleles and risk of ovarian cancer in BRCA1 and BRCA2 mutation carriers. *Hum Mutat, 33*(4), 690-702. doi:10.1002/humu.22025

Ramus, S. J., Köbel, M., Sieh, W., Anglesio, M. S., Millstein, J., Bowtell, D. D., . . . Huntsman, D. G. (2013). Abstract A27: The ovarian tumor tissue analysis (OTTA) consortium. *Clinical Cancer Research, 19*(19 Supplement), A27. doi:10.1158/1078-0432.OVCA13-A27

Rathore, A. S., Kumar, S., Konwar, R., Makker, A., Negi, M. P., & Goel, M. M. (2014). CD3+, CD4+ & CD8+ tumour infiltrating lymphocytes (TILs) are predictors of favourable survival outcome in infiltrating ductal carcinoma of breast. *Indian J Med Res, 140*(3), 361-369.

Rein, B. J. D., Gupta, S., Dada, R., Safi, J., Michener, C., & Agarwal, A. (2011). Potential Markers for Detection and Monitoring of Ovarian Cancer. *Journal of Oncology, 2011*, 475983. doi:10.1155/2011/475983

Romagosa, C., Simonetti, S., López-Vicente, L., Mazo, A., Lleonart, M. E., Castellvi, J., & Ramon y Cajal, S. (2011). p16(Ink4a) overexpression in cancer: a tumor suppressor gene associated with senescence and high-grade tumors. *Oncogene, 30*(18), 2087-2097. doi:10.1038/onc.2010.614

Royar, J., Becher, H., & Chang-Claude, J. (2001). Low-dose oral contraceptives: protective effect on ovarian cancer risk. *Int J Cancer, 95*(6), 370-374. doi:10.1002/1097-0215(20011120)95:6<370::aid-ijc1065>3.0.co;2-t

Salminen, L., Nadeem, N., Jain, S., Grènman, S., Carpén, O., Hietanen, S., . . . Hynninen, J. (2020). A longitudinal analysis of CA125 glycoforms in the monitoring and follow up of high grade serous ovarian cancer. *Gynecol Oncol, 156*(3), 689-694. doi:10.1016/j.ygyno.2019.12.025

Salvador, S., Gilks, B., Köbel, M., Huntsman, D., Rosen, B., & Miller, D. (2009). The fallopian tube: primary site of most pelvic high-grade serous carcinomas. *Int J Gynecol Cancer, 19*(1), 58-64. doi:10.1111/IGC.0b013e318199009c

Santoro, A., Angelico, G., Piermattei, A., Inzani, F., Valente, M., Arciuolo, D., . . . Zannoni, G. F. (2019). Pathological Chemotherapy Response Score in Patients Affected by High Grade Serous Ovarian Carcinoma: The Prognostic Role of Omental and Ovarian Residual Disease. *Frontiers in oncology, 9*, 778-778. doi:10.3389/fonc.2019.00778

Sato, E., Olson, S. H., Ahn, J., Bundy, B., Nishikawa, H., Qian, F., . . . Odunsi, K. (2005). Intraepithelial CD8⁺ tumor-infiltrating lymphocytes and a high CD8⁺/regulatory T cell ratio are associated with favorable prognosis in ovarian cancer. *Proceedings of the National Academy of Sciences of the United States of America, 102*(51), 18538-18543. doi:10.1073/pnas.0509182102

Seidman, J. D., Zhao, P., & Yemelyanova, A. (2011). "Primary peritoneal" high-grade serous carcinoma is very likely metastatic from serous tubal intraepithelial carcinoma: assessing the new paradigm of ovarian and pelvic serous carcinogenesis and its implications for screening for ovarian cancer. *Gynecol Oncol, 120*(3), 470-473. doi:10.1016/j.ygyno.2010.11.020

Shen, W., Li, H. L., Liu, L., & Cheng, J. X. (2017). Expression levels of PTEN, HIF-1α, and VEGF as prognostic factors in ovarian cancer. *Eur Rev Med Pharmacol Sci, 21*(11), 2596-2603.

Sieh, W., Köbel, M., Longacre, T. A., Bowtell, D. D., deFazio, A., Goodman, M. T., . . . Ramus, S. J. (2013). Hormone-receptor expression and ovarian cancer survival: an Ovarian Tumor Tissue Analysis consortium study. *The Lancet. Oncology, 14*(9), 853-862. doi:10.1016/S1470-2045(13)70253-5

Song, H., Ramus, S. J., Tyrer, J., Bolton, K. L., Gentry-Maharaj, A., Wozniak, E., . . . Gayther, S. A. (2009). A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. *Nat Genet, 41*(9), 996-1000. doi:10.1038/ng.424

Sun, N. K., Huang, S. L., Lu, H. P., Chang, T. C., & Chao, C. C. (2015). Integrative transcriptomics-based identification of cryptic drivers of taxol-resistance genes in ovarian carcinoma cells: Analysis of the androgen receptor. *Oncotarget, 6*(29), 27065-27082. doi:10.18632/oncotarget.4824

Takeda, K., & Akira, S. (2005). Toll-like receptors in innate immunity. *International Immunology, 17*(1), 1-14. doi:10.1093/intimm/dxh186

Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267-288. Retrieved from www.jstor.org/stable/2346178

Turner, T. B., Buchsbaum, D. J., Straughn, J. M., Randall, T. D., & Arend, R. C. (2016). Ovarian cancer and the immune system — The role of targeted therapies. *Gynecologic Oncology, 142*(2), 349-356. doi:https://doi.org/10.1016/j.ygyno.2016.05.007

Volgareva, G., Zavalishina, L., Andreeva, Y., Frank, G., Krutikova, E., Golovina, D., . . . Kisseljov, F. (2004). Protein p16 as a marker of dysplastic and neoplastic alterations in cervical epithelial cells. *BMC cancer, 4*, 58. doi:10.1186/1471-2407-4-58

Williams, E., Martin, S., Moss, R., Durrant, L., & Deen, S. (2012). Co-expression of VEGF and CA9 in ovarian high-grade serous carcinoma and relationship to survival. *Virchows Arch, 461*(1), 33-39. doi:10.1007/s00428-012-1252-9

Xu, X., Wang, Y., Wang, F., Jia, L., Zhou, Y., Deng, F., . . . Wang, J. (2013). Nadir CA-125 level as prognosis indicator of high-grade serous ovarian cancer. *Journal of ovarian research, 6*, 31-31. doi:10.1186/1757-2215-6-31

Yang, Y., & Zou, H. (2015). A fast unified algorithm for solving group-lasso penalize learning problems. *Statistics and Computing, 25*(6), 1129-1141. doi:10.1007/s11222-014-9498-5

Zhu, H., Zhu, X., Zheng, L., Hu, X., Sun, L., & Zhu, X. (2017). The role of the androgen receptor in ovarian cancer carcinogenesis and its clinical implications. *Oncotarget, 8*(17), 29395-29405. doi:10.18632/oncotarget.12561