,		
Number of risk factors	Patients with PC/total patients	Proportion of PC
0	18/289	6.2%
1	30/157	19.1%
2	28/52	53.8%
3	5/5	100%

Table 1. Proportion of patients with PC in different groups according to model A (BorrmannIV, tumor diameter \ge 5cm, CA125 \ge 35U/ml)

Table 2. Negative predictive value of the model A with different cutoff values

Cutoff value	Negative predictive value
1*	93.8%
2	89.2%
3	84.7%

*Patients with one or more risk factors requiring staging laparoscopy





ROC curve of model A, AUC = 0.763

Number of risk	Patients with PC/total patients	Proportion of PC		
factors				
0	3/150	2.0%		
1	23/180	12.8%		
2	24/120	20.0%		
3	26/48	54.2%		
4	5/5	100%		

Table3. Proportion of patients with PC in different groups according to model B (BorrmannIV, tumor diameter \ge 5cm, CA125 \ge 35U/ml, T4)

Table4. Negative predictive value of the model B with different cutoff values

Cutoff value	Negative predictive value	
1*	98.0%	
2	92.1%	
3	88.9%	
4	84.7%	

*Patients with one or more risk factors requiring staging laparoscopy

Figure 2. Model B



ROC curve of model B, AUC = 0.780

Number of risk factors	Patients with PC/total patients	Proportion of PC
0	8/198	4.0%
1	28/178	15.7%
2	19/73	26.0%
3	17/24	70.8%
4	3/3	100%

Table 5. Proportion of patients with PC in different groups according to model C (BorrmannIV, tumor diameter≥5cm, CA125≥35U/ml, Lauren diffused type)

Table 6. Negative predictive value of the model C with different cutoff values

Cutoff value	Negative predictive value
1*	96.0
2	90.4%
3	87.8%
4	84.8%

*Patients with one or more risk factors requiring staging laparoscopy Figure 3. Model C



ROC curve of model C, AUC = 0.769

Number of risk factors	Patients with PC/total patients	Proportion of PC
0	1/117	0.9%
1	13/142	9.2%
2	20/100	20.0%
3	20/100	20.0%
4	23/40	57.5%
5	4/4	100%

Table 7. Proportion of patients with PC in different groups according to model D (BorrmannIV, tumor diameter≥5cm, CA125≥35U/ml, T4, N2/3)

	Table 8. Negative 1	oredictive y	value of the	model D with	different	cutoff values
--	---------------------	--------------	--------------	--------------	-----------	---------------

Cutoff value	Negative predictive value
1*	99.1%
2	94.6%
3	90.5%
4	88.2%
5	84.6%





ROC curve of model D, AUC = 0.786





Comparison of different models

In order to construct a PC prediction model that takes into account the requirements of conciseness, predictive accuracy and coverage of more patients, we compared each model. We find that model A, which is constructed with three independent risk factors, can spare 57.5% of patients from staging laparoscopy (SL) with cutoff value of 1, which is significantly higher than model B. However, the maximum negative predictive value (NPV) is 93.8%, which means that 6.2% of PC is missed in patients who are exempted from SL. Model C and B both consist of 4-factors, equally easy to use. And when the cutoff value of model C is 1, the false negative rate (FNR) of PC among patients exempted from SL is 4%, higher than that of model B. Model D has more factors and is more complex than model B. The FNR of PC among patients exempted from SL is 0.9% with cutoff value of 1 in model D, which is better than that of model B but with limited improvement, and the percentage of patients exempted from SL reduce to 23%, less than 29.8% of model B. The comparison of the remaining models with model B is similar and will not be listed. Although reducing the factors of model B can simplify the model, it cannot achieve satisfactory NPV; replacing the factors of model B will reduce the NPV and increase FNR of PC in the population exempted from SL; adding factors will not significantly improve the NPV, but will make the model more complicated and make the model more difficult to apply. A stratified prediction model consisting of Borrmann IV, tumor diameter ≥5 cm, CA125 ≥35 U/ml manand cT4 meets the requirements of conciseness, predictive accuracy and making more people exempt from SL.

The code of 10-fold cross validation over 1000 iterations using the "caret" package in R (version 4.1.1) is as follows.

```
library (pROC)
library (ggplot2)
library(dplyr)
library (caret)
ggm <- read.csv('ggm2.csv')
ggm$outcome1 <- ifelse(ggm$outcome==0, 'no', 'yes')</pre>
ggm$outcome1 <- factor(ggm$outcome1, levels = c('no', 'yes'))</pre>
# Define training control,设置参数10-折,重复1000次
set. seed (1234)
train.control <- trainControl(method = "repeatedcv", summaryFunction=twoClassSummary,</pre>
                              number = 10, repeats = 1000, savePredictions = T, classProbs=T, )
# logistic模型
model <- train(outcomel<sup>~</sup>Borrmann+size+CA125+pT, data = ggm, method = "glm",
               preProc=c("center", "scale"), trControl = train.control)
#> Warning in train.default(x, y, weights = w, ...): The metric "Accuracy" was not
#> in the result set. ROC will be used instead.
# Summarize the results
print(model)
#> Generalized Linear Model
#>
#> 503 samples
#>
    4 predictor
     2 classes: 'no', 'ves'
#>
#>
#> Pre-processing: centered (4), scaled (4)
#> Resampling: Cross-Validated (10 fold, repeated 1000 times)
#> Summary of sample sizes: 453, 453, 452, 453, 452, 452, ...
#> Resampling results:
#>
#>
     ROC
                Sens
                           Spec
\#>
     0.7841435 0.9488126 0.3458083
##10折+1000次重复 总共应该有10*1000=10000个AUC值
length(model$resample$ROC)
#> [1] 10000
##AUC极差
range(model$resample$ROC)
#> [1] 0.4211310 0.9895833
##AUC均值
mean(model$resample$ROC)
#> [1] 0.7841435
##95%置信区间
quantile(model$resample$ROC, c(0.025, 0.975))
        2.5%
                 97.5%
#>
#> 0.6160714 0.9241071
View(model$results)
library (plotROC)
g <- ggplot(model$pred, aes(m=yes, d=factor(obs, levels = c("no", 'yes')))) +
    geom roc(n.cuts=0) +
   coord equal() +
   style roc()
g + annotate("text", x=0.75, y=0.25, label=paste("AUC =", round(mean(model$resample$ROC), 4),
'[',
                                                  round(quantile(model$resample$ROC, 0.025), 4),
' –' ,
```

ggm

```
']'))
#> Warning in verify_d(data$d): D not labeled 0/1, assuming no = 0 and yes = 1!
```



ggm

```
#g + annotate("text", x=0.75, y=0.25, label="AUC = 0.768[0.590-0.913]")
#g + annotate("text", x=0.75, y=0.25, label=paste("AUC =", round((calc_auc(g))$AUC, 4)))
##单纯用总人群的logistic建模结果AUC及输出
glml <- glm(outcome~Borrmann+size+CA125+pT,family = "binomial", data = ggm)
exp(coef(g1m1))
#> (Intercept)
                 Borrmann
                                 size
                                             CA125
                                                            рT
    0.0429079
               4.8869917
                             3.3156355
                                         3.9513712
                                                     1.9830790
#>
exp(confint(glm1))
#>
                    2.5 %
                              97.5 %
#> (Intercept) 0.02161591 0.07747542
#> Borrmann
              2.37407465 10.24874954
#> size
              1.82552301 6.16593550
#> CA125
              1.70386066 9.09195904
              1.02610380 4.04291035
#> pT
summary(g1m1)
\#>
#> Call:
#> glm(formula = outcome ~ Borrmann + size + CA125 + pT, family = "binomial",
#>
      data = ggm)
#>
#> Deviance Residuals:
#>
      Min
                10
                     Median
                                   3Q
                                           Max
#> -1.3165 -0.5158 -0.4041 -0.2899
                                        2.5261
#>
#> Coefficients:
#>
              Estimate Std. Error z value Pr(>|z|)
#> (Intercept) -3.1487
                           0.3238 -9.725 < 2e-16 ***
#> Borrmann
                1.5866
                           0.3716 4.270 1.96e-05 ***
#> size
                 1.1986
                           0.3091 3.878 0.000105 ***
#> CA125
                1.3741
                           0.4248 3.235 0.001218 **
                           0.3471 1.972 0.048569 *
#> pT
                0.6847
#> ----
#> Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> (Dispersion parameter for binomial family taken to be 1)
#>
#>
      Null deviance: 444.03 on 502 degrees of freedom
#> Residual deviance: 357.51 on 498 degrees of freedom
#> AIC: 367.51
#>
#> Number of Fisher Scoring iterations: 5
ggm$prob <- predict(glm1, ggm, type="response")</pre>
roc1 <- roc(ggm$outcome, ggm$prob)</pre>
roc1
#>
#> Call:
#> roc.default(response = ggm$outcome, predictor = ggm$prob)
#>
#> Data: ggm$prob in 422 controls (ggm$outcome 0) < 81 cases (ggm$outcome 1).
#> Area under the curve: 0.7895
plot(roc1)
```