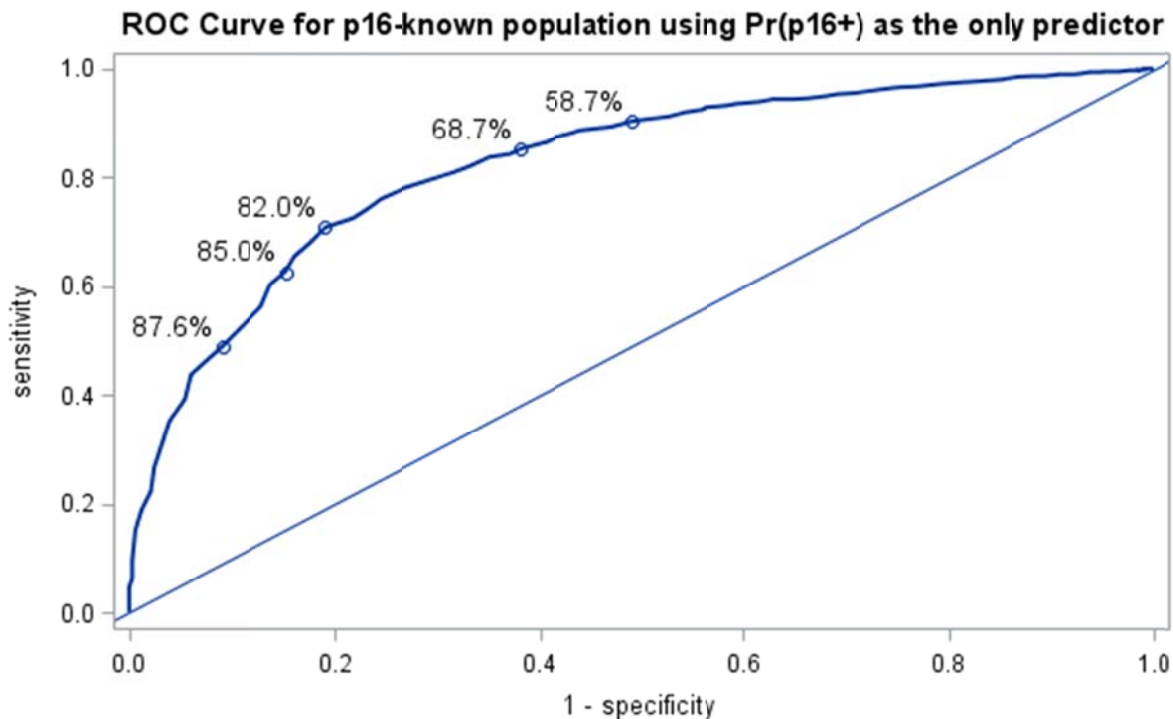


Appendix 4 (as supplied by the authors): Leave-one-out cross validation methods



Cut-off of Pr(p16+)	Sensitivity	Specificity
58.7%	90%	51%
68.7%	85%	62%
82.0% (optimal)	71%	81%
85.0%	63%	85%
87.6%	49%	90%

Caption: The OPC population with known p16 status was used as the validation subset (leave-one-out validation). For each patient in the validation subset (n=1282), p16 status was iteratively assumed missing and estimated using 100 imputations (as described in the methods) using the remaining validation subset population (n=1281). Thus, 1282 sets of 100 imputations was performed and Pr(p16+) was recorded for each patient with known p16 status. Using Pr(p16+) as the only predictor of p16 status in the validation subset, Receiver Operator Characteristic (ROC) curve analysis was conducted to evaluate the predictive ability of different cut-points of Pr(p16+). The optimal cut-point was defined as the value of Pr(p16+) that maximized the sum of specificity and sensitivity of predicting the true p16 status. Other cut-points (85%, 90% sensitivity, and 85%, 90% specificity) were also evaluated. The points on the graph (○) represent cut-points of Pr(p16+).