

Table 1 Demographics and clinical characteristics of eligible patients

Characteristics	Total cohort N(%)	Development cohort N (%)	Validation cohort N (%)	P-value
Number of patients	3348	2344(70%)	1004(30%)	
Age				0.732
<45	330(9.9%)	236(10.1%)	94(9.3%)	
45-64	1613(48.1%)	1128(48.1%)	485(48.5%)	
65-84	1234(36.9)	866(36.9)	368(36.7%)	
≥85	171(5.1%)	114(5.9%)	57(5.7%)	
Sex				0.396
Female	1480(44.2%)	1025(43.7%)	455(45.3%)	
Male	1868(55.8%)	1319(56.3%)	549(54.7%)	
Race				0.343
White	2623(78.3%)	1852(79.0%)	771(76.8%)	
Black	292(8.7%)	202(8.6%)	90(9.0%)	
Asian or Pacific Islander	412(12.3%)	274(11.8%)	138(13.7%)	
American Indian/Alaska Native	21(0.7%)	16(0.7%)	5(0.5%)	
Marital				0.513
Single	651(19.5%)	471(20.1%)	180(17.9%)	
Married	1923(57.4%)	1330(56.7%)	593(59.1%)	
Divorced or Separated	355(10.6%)	255(10.9%)	100(10%)	
Widowed	404(12.1%)	278(11.9%)	126(12.5%)	
Unmarried or Domestic Partner	15(0.4%)	10(0.4%)	5(0.5%)	
Radiation				0.559
No	2586(77.2%)	1817(77.5%)	769(76.6%)	
Yes	762(22.8%)	527(22.5%)	235(23.4%)	
Chemotherapy				0.204
No	968(28.9%)	693(29.36%)	275(27.4%)	
Yes	2380(71.1)	1651(70.4%)	729(72.6%)	
Summary_stage				0.595
Regional	2551(76.2%)	1780(75.9%)	771(76.8%)	
Distant	797(23.8%)	564(24.1%)	233(23.2%)	
AJCC_Stage				0.086

Characteristics	Total cohort N(%)	Development cohort N (%)	Validation cohort N (%)	P-value
IIIA	330(9.9%)	228(9.7%)	102(10.2)	0.918
IIIB	1653(49.3%)	1168(49.8%)	485(48.3%)	
IIIC	625(18.7%)	417(17.8%)	208(20.7%)	
IVA	489(14.6%)	340(14.5%)	149(14.8%)	
IVB	251(7.5%)	191(8.2%)	60(6.0%)	
T				
T1	66(2.0%)	47(2.0%)	19(1.9%)	
T2	351(10.5%)	244(10.4%)	107(10.7%)	
T3	2210(66.0%)	1555(66.3%)	655(65.2%)	
T4a	497(14.8%)	347(14.8%)	150(15.0%)	
T4b	224(6.7%)	151(6.5%)	73(7.2%)	
N				0.726
N1a	921(27.5%)	657(28.0%)	264(26.3%)	0.924
N1b	986(29.5%)	695(29.7%)	291(29.0%)	
N1c	5(0.1%)	3(0.1%)	2(0.2%)	
N2a	726(21.7%)	498(21.3%)	228(22.7%)	
N2b	710(21.2%)	491(20.9%)	219(21.8%)	
Tumor_size				
≤3.7cm	849(25.4%)	590(25.2%)	259(25.8%)	
3.8-5.6cm	1579(47.1%)	1107(47.2%)	472(47.0%)	
≥5.7cm	920(27.5%)	647(27.6%)	273(27.2%)	
Regional_nodes_examined				0.640
1-12	686(20.5%)	489(21.9%)	197(19.6%)	0.345
13-16	842(25.1%)	592(25.3%)	250(24.9%)	
≥17	1820(56.4%)	1263(53.8%)	557(55.5%)	
Regional_nodes_positive				
1-3	1913(57.1%)	1358(57.9%)	555(55.3%)	
4-8	976(29.2%)	668(28.5%)	308(30.7%)	
≤9	459(13.7%)	318(13.6%)	141(14.0%)	