

Table S1: Baseline characteristics of the training set and the validation set.

Variables	All patients n (%)	Training set n (%)	Validation set n (%)	P-value
Age				0.591
< 45	2338 (53.0)	1624 (52.5)	714 (54.0)	
45-47	1083 (24.5)	760 (24.6)	323 (24.4)	
> 47	994 (22.5)	708 (22.9)	286 (21.6)	
Race				0.388
White	3310 (75.0)	2323 (75.1)	987 (74.6)	
Black	333 (7.5)	241 (7.8)	92 (7.0)	
Other	772 (17.5)	528 (17.1)	244 (18.4)	
Grade				0.765
I	2466 (55.9)	1735 (56.1)	731 (55.3)	
II	1211 (27.4)	836 (27.0)	375 (28.3)	
III	583 (13.2)	411 (13.3)	172 (13.0)	
IV	155 (3.5)	110 (3.6)	45 (3.4)	
T stage				0.469
T1	3600 (81.5)	2510 (81.2)	1090 (82.4)	
T2	324 (7.3)	248 (8.0)	107 (8.1)	
T3	441 (10.1)	296 (9.6)	114 (8.6)	
T4	50 (1.1)	38 (1.2)	12 (0.9)	
N stage				0.533
N0	4023 (91.1)	2812 (90.9)	1211 (91.5)	
N1	237 (5.4)	170 (5.5)	67 (5.1)	
N2	155 (3.5)	110 (3.6)	45 (3.4)	
M stage				0.762
M0	4232 (95.9)	2962 (95.8)	1270 (96.0)	
M1	183 (4.1)	130 (4.2)	53 (4.0)	
Tumor size (cm)				0.834
< 3.6	1116 (25.3)	778 (25.2)	338 (25.5)	
3.6-7.8	3027 (68.5)	2124 (68.7)	903 (68.3)	
> 7.8	272 (6.2)	190 (6.1)	82 (6.2)	
SEER stage				0.716
Localized	3232 (73.2)	2259 (73.1)	973 (73.5)	
Regional	978 (22.2)	687 (22.2)	291 (22.0)	
Distant	205 (4.6)	146 (4.7)	59 (4.5)	
Surgery				0.595
No	70 (1.6)	47 (1.5)	23 (1.7)	
Yes	4345 (98.4)	3045 (98.5)	1300 (98.3)	
Lymphadenectomy				0.687
No	1879 (42.6)	1322 (42.8)	557 (42.1)	
Yes	2536 (57.4)	1770 (57.2)	766 (57.9)	
Radiotherapy				0.597
No	3509 (79.5)	2451 (79.3)	1058 (80.0)	
Yes	906 (20.5)	641 (20.7)	265 (20.0)	
Chemotherapy				0.644
No	3546 (80.3)	2489 (80.5)	1057 (79.9)	
Yes	869 (19.7)	603 (19.5)	266 (20.1)	