

**Table S1** Characteristics of the patients in the Chinese dataset

<b>Variable</b>	<b>Median (IQR)/N (%)</b>
<b>Year of surgery</b>	
2001–2003	348 (21.8%)
2004–2006	351 (22.0%)
2007–2010	896 (56.2%)
<b>Age, years</b>	58 (50–65)
<b>Gender</b>	
Female	527 (33.0%)
Male	1068 (67.0%)
<b>Tumour location</b>	
Cardia	97 (6.1%)
Upper one-third	444 (27.8%)
Middle one-third	329 (20.6%)
Lower one-third	638 (40.0%)
Unknown	87 (5.5%)
<b>Tumour size, cm</b>	4.0 (3.0–6.0)
Unknown	35 (2.2%)
<b>Tumour differentiation</b>	
Poorly or undifferentiated	983 (61.6%)
Well or moderately differentiated	434 (27.2%)
Unknown	178 (11.2%)
<b>T stage</b>	
T1	120 (7.5%)
T2	199 (12.5%)
T3	170 (10.7%)
T4a	912 (57.2%)
T4b	194 (12.2%)
<b>MLNs</b>	2 (0–6)
<b>NLNs</b>	13 (6–21)
<b>LNR</b>	0.14 (0.00–0.40)
<b>LODDS</b>	-1.61 (-2.94–-0.35)
<b>Adequate nodal evaluation</b>	
No ( $\leq 15$ )	613 (38.4%)
Yes ( $> 15$ )	982 (61.6%)
<b>Institution</b>	
The Sun Yat-Sen University Cancer Centre	1093 (68.5%)
The Nanfang Hospital	502 (31.5%)

IQR, interquartile range; MLNs, metastatic lymph nodes; THNs, total harvested

lymph nodes; NLNs, negative lymph nodes; LNR, lymph node ratio; LODDS, log ratio of metastatic lymph nodes.

**Table S2** Patient characteristics after multiple imputations for missing data in the SEER and Chinese sets

<b>Variable</b>	<b>Median (IQR)/N (%)</b>
<b>The SEER set (N = 15,320)</b>	
<b>Race</b>	
White	9899 (64.6%)
Black	1887 (12.3%)
Other	3534 (23.1%)
<b>Tumour location</b>	
Cardia	5043 (32.9%)
Upper one-third	1950 (12.7%)
Middle one-third	7677 (50.1%)
Lower one-third	650 (4.2%)
<b>Tumour size, mm</b>	
	4.5 (2.7–6.2)
<b>Tumour differentiation</b>	
Poorly or undifferentiated	9940 (64.9%)
Well or moderately differentiated	5380 (35.1%)
<b>The Chinese set (N = 1,595)</b>	
<b>Tumour location</b>	
Cardia	99 (6.2%)
Upper one-third	467 (29.3%)
Middle one-third	359 (22.5%)
Lower one-third	670 (42.0%)
<b>Tumour size, cm</b>	
	4.0 (3.0–6.0)
<b>Tumour differentiation</b>	
Poorly or undifferentiated	1116 (70.0%)
Well or moderately differentiated	479 (30.0%)

SEER, Surveillance, Epidemiology, and End Results database; IQR, interquartile range.

**Table S3** Impact of total harvested node counts on the prognostic performances of MLN-, LNR- and LODDS-based nomograms among the SEER set

Model	Number of total harvested lymph nodes		
	1–10	11–15	> 15
	C-index <sup>a</sup>	C-index <sup>b</sup>	C-index <sup>c</sup>
<b>MLNs</b>	0.733	0.745	0.759
<b>LNR</b>	0.731	0.745	0.756
<b>LODDS</b>	0.733	0.745	0.759

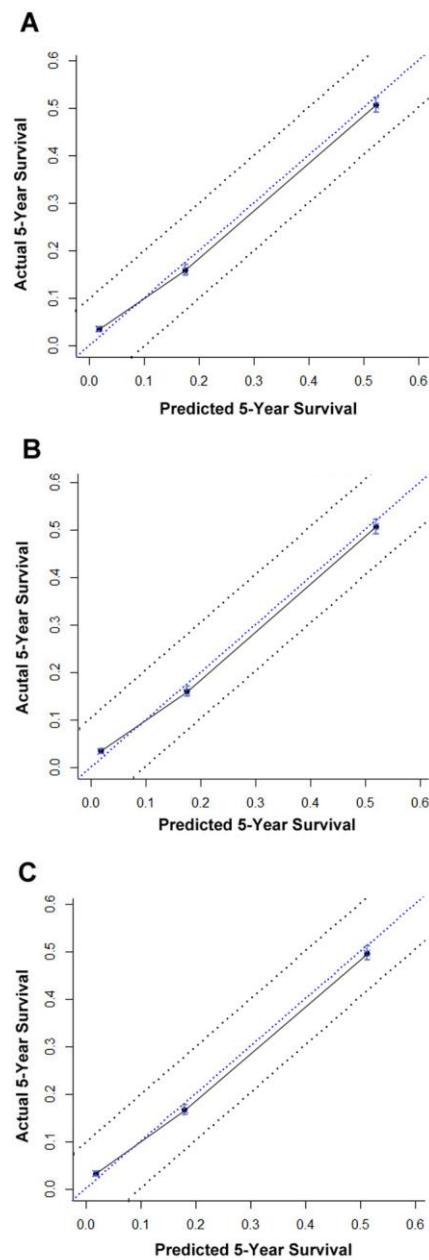
C-index, concordance index; MLNs, metastatic lymph nodes; LNR, lymph node ratio; LODDS, log ratio of metastatic lymph node.

a.  $P > 0.373$  for all pairwise comparisons among the C-indices of the three nomograms.

b.  $P > 0.480$  for all pairwise comparisons among the C-indices of the three nomograms.

c.  $P > 0.278$  for all pairwise comparisons among the C-indices of the three nomograms.

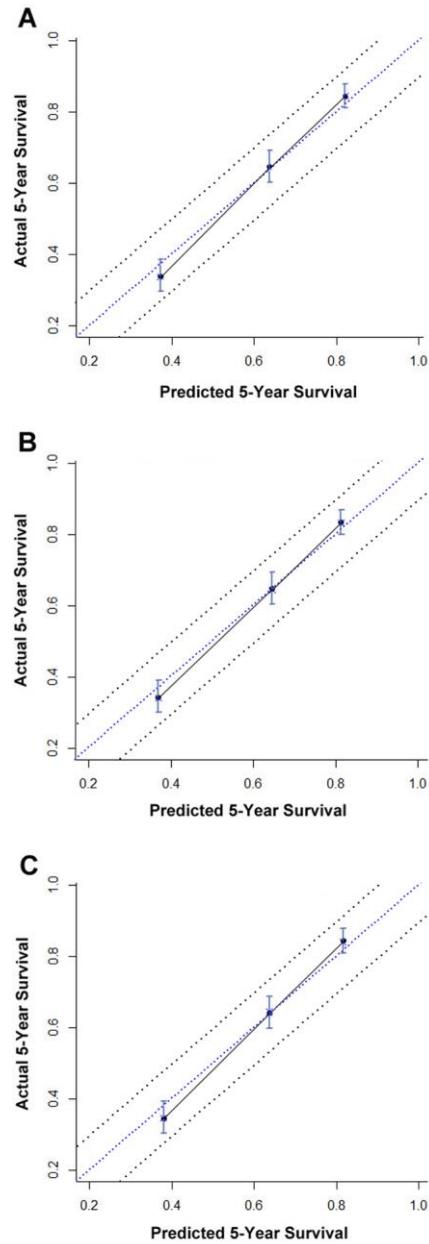
**Fig. S1** Calibration plots of the prognostic nomograms for the SEER set.



In the calibration plots of the (A) MLN-, (B) LNR-, and (C) LODDS-based nomograms, the nomogram-predicted 5-year OS rates showed a close correlation with the ideal 45-degree reference line (blue dotted line). All predictions lie within a 10% margin of error (within the black dotted lines). The mean deviations between the actual survival probabilities and those predicted with the MLN-, LNR-, and

LODDS-based nomograms were 0.016, 0.014, and 0.014, respectively

**Fig. S2** Calibration plots of the prognostic nomograms for the Chinese set.



In the calibration plots of the (A) MLN-, (B) LNR-, and (C) LODDS-based nomograms, the nomogram-predicted 5-year OS rates showed a close correlation with the ideal 45-degree reference line (blue dotted line). All predictions lie within a 10% margin of error (within the black dotted lines). The mean deviations between the actual survival probabilities and those predicted with the MLN-, LNR-, and

LODDS-based nomograms were 0.022, 0.017, and 0.023, respectively