

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

BMJ Open

Prognostic value of different lymph node staging methods for node positive cardia gastric cancer

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2021-050378
Article Type:	Original research
Date Submitted by the Author:	22-Feb-2021
Complete List of Authors:	Wang, Xiao-Qing; Anhui Medical College Bao, Min; Anhui Medical College Zhang, Cheng; First Affiliated Hospital of Anhui Medical University
Keywords:	Gastrointestinal tumours < ONCOLOGY, Gastrointestinal tumours < GASTROENTEROLOGY, SURGERY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **Title page**

2 **Article title:** Prognostic value of different lymph node staging methods for node positive cardia
3 gastric cancer

4 **Author information:** Xiao-Qing Wang 1, Min Bao 1, Cheng Zhang 2

5 1 Anhui Medical College, Hefei, Anhui, China PR

6 2 Anhui Provincial Cancer Institute, the First Affiliated Hospital of Anhui Medical University,
7 Hefei, Anhui, China PR

8 Corresponding to: Cheng Zhang, ahmuzc@sina.cn

9 **Disclaimer:** The views expressed in the submitted article are our own and not an official position
10 of the institution or funder.

11 **Word count:** 2052 (excluding its abstract, acknowledgments, tables, figure legends, and
12 references) abstract 274 .

13 **Number of figures and tables:** This draft included 4 figures and 2 tables.

14 **Disclosure of relationships and activities:** None.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

15 **Abstract**

16 Objective: To investigate the prognostic efficacy of lymph node ratio (LNR) and log odds of
17 positive lymph nodes (LODDS) in node positive cardia gastric adenocarcinoma (CGA).

18 Design: A SEER database review

19 Participants: A total of 1 038 patients with node positive CGA were enrolled from SEER database.

20 Seventy percent of the entire patients were randomly assigned to training set (N = 723) and the
21 rest was assigned to validating set (N = 315).

22 Interventions: The major endpoint was cancer specific survival (CSS). Optimal cut-off values
23 were determined by X-tile software. The prognostic power was evaluated using Akaike
24 Information Criterion (AIC) and Harrell concordance index (C-index). Cox stepwise regression
25 analysis was performed to construct nomogram for prediction of 1-, 2-, and 5-year CSS.

26 Results: The training set and validating set are similar in terms of clinical and demographic
27 features. The optimal cut-off values for LNR were 0.09 and 0.33, and for LODDS were -2.09 and
28 -0.65. CSS was significantly different by N, LNR and LODDS categories. The C-index of N stage
29 was lower than that of LNR or LODDS. The AIC of N stage was higher than that of LNR or
30 LODDS. Independent predictors included age, race, tumor grade, T stage, M stage and LNR (or
31 LODDS) and they were incorporated in nomograms for 1-, 2- and 5-year CSS prediction.
32 Calibration plots showed satisfied results of internal and external validity of the nomogram.

33 Conclusions: LNR and LODDS staging methods have better prognostic efficacy than traditional N
34 staging method in CGA patients with regional node metastasis. Besides, the two values are
35 promising substitute for N staging in nomogram development when other independent prognostic
36 factors are incorporated.

37
38 Key words: Cardia; Adenocarcinoma; Stomach Neoplasms; Lymph Node Ratio; Neoplasm
39 Staging

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49

Strengths and limitations of this study

- This study used national cancer registry data for cardia gastric adenocarcinoma research;
- Novel staging methods based on the number of positive lymph node was established for prognostic prediction;
- Nomograms based on the new staging methods were constructed and validated;
- This study needs to be confirmed by other populations.

Patient consent form: The SEER database review is granted exemption from obtaining patients' consents.

Word count: 2052 (excluding its abstract, acknowledgments, tables, figure legends, and references) abstract 274 .

For peer review only

50 Introduction

51 Gastric cancer (GC) generally includes 2 topographical categories: non-cardia GC that arises from
52 more distant sites and cardia GC that arises in gastroesophageal junction (GEJ). In contrast to
53 steady decline of non-cardia GC incidence, cardia GC occurs more frequently, particular in
54 high-income countries (1, 2). This trend is associated with obesity, gastroesophageal reflux
55 disease (GERD), and Barrett esophagus (2). In addition to different incidence trend,
56 clinicalpathological feature and long-term survival vary between the two GC subtypes (3). For
57 accurate prediction of survival, precise staging is required. The Tumor-Node-Metastasis (TNM)
58 classification 7th edition by the American Joint Committee on Cancer (AJCC) recommends at least
59 15 lymph nodes (LN) collection for N staging (4, 5). However inadequate LN harvest frequently
60 occurs due to many conditions, thus precise staging cannot be obtained sometimes. It has been
61 demonstrated that LN ratio (LNR) could better estimate survival of GC patients after curative
62 gastrectomy, regardless of the number of LN examined (6), and may be promising for aiding
63 TNM staging system (7). Apart from that, log odds of positive LN (LODDS) outperformed N and
64 LNR staging system when predicting survival of GC patients (8-10). Therefore the traditional N
65 staging classification may be substituted with different methods, with even improved performance.
66 Nevertheless little evidence evaluates the performance of the two LN staging systems
67 aforementioned in cardia GC, since it has distinct clinical characteristics and epidemiology from
68 overall GC.

69 Here we use nationwide cancer registry data to appraise the prognostic value of LNR and LODDS
70 in patients with node positive cardia gastric adenocarcinoma (CGA), and, if possible, construct
71 nomogram for survival prediction based on the new LN staging system.

72 Methods

73 Selection and Description of Participants

74 The inclusion criteria were as follows: 1) the International Classification of Disease for Oncology,
75 Third Edition (ICD-O-3) for primary tumor site was C16.0 (cardia); 2) broad histological recode
76 was 8140-8389: adenomas and adenocarcinomas; 3) diagnostic confirmation was positive
77 histology; 4) surgery was performed; 5) diagnosed during 2010-2015; 6) the definite number of
78 regional positive nodes was clear and not zero. We excluded cases with unknown race, T stage
79 information, tumor size and grade. As shown in Figure 1, the final cohort enrolled 1 038 patients
80 with node positive CGA, of whom 857 were male and 181 were female. Three hundred and thirty
81 eight (32.56%) were over 70 years old. Eight hundred and ninety six (86.32%) were white, 64
82 were black and 78 were other races. Next 70% of the entire patients were randomly assigned to
83 training set (N = 723) and the rest was assigned to validating set (N = 315).

84 Patient and Public Involvement

85 This study is a data review based on SEER program, so it was granted exemption from requiring
86 informed consent. All procedures performed in studies involving human participants were in
87 accordance with the ethical standards of the Ethics Committee of the Anhui Medical College. We
88 used SEER*Stat (version 8.3.8) to access to Incidence - SEER Research Data, 18 Registries, Nov
89 2019 Sub 2000-2017 (SEER 18 database) (11) for collection of node positive CGA patients
90 (username: 21268-Nov2019). Patients were not involved in the recruitment to and conduct of the

91 study. The findings of the study will be disseminated to all study participants by online article.

92 **Technical Information**

93 The main outcome was cancer specific survival (CSS), which was referred to as death specifically
 94 due to CGA and the period between first diagnosis and death. In addition, we extracted the
 95 following variables for analysis: sex, race, age, AJCC 7th TNM stage information, tumor size,
 96 grade, number of regional nodes examined and number of regional nodes positive. LNR and
 97 LODDS were calculated as previously reported (12). Briefly, LNR was defined as the ratio of the
 98 number of positive nodes divided by the total number of examined nodes. LODDS was calculated
 99 using the formula: $\log(\text{NPLN}+0.50)/(\text{NDLN}-\text{NPLN}+0.50)$, in which 0.50 was added to both the
 100 numerator and denominator to avoid an infinite number.

101 The optimal thresholds for cutting LNR and LODDS into trichotomous variables were determined
 102 by X-tile software (version 3.6.1) (13), which were based on the maximal log-rank chi-square
 103 value that represented the greatest group difference of CSS probability.

104 **Statistics**

105 The distributions of baseline features between training set and validating set were described and
 106 compared by chi-square test. Survival curves, median survival and CSS rates were generated using
 107 the Kaplan-Meier method. Outcome difference between groups was analyzed by the log-rank test.
 108 Multivariable Cox regression model was used to establish prognostic model for CSS. The
 109 prognostic power was evaluated using Akaike Information Criterion (AIC) and Harrell
 110 concordance index (C-index). A predictive model with lower AIC indicated better model fit, while
 111 with higher C-index indicated better discriminative ability. A value of C-index of 0.5 indicates no
 112 predictive power, and an index of 1.0 indicates complete differentiation. Cox stepwise regression
 113 analysis was also performed to construct nomogram for prediction of 1-, 2-, and 5-year CSS.
 114 Validation of nomogram was performed by internal and external calibration plots (14). Bootstraps
 115 with 1 000 resample were used for validation activities. All statistical analyses were performed
 116 using R software (version 3.5.3). A two-tailed P value of less than 0.05 was considered
 117 statistically significant.

118 **Results**

119 Table 1 summarized the demographic and clinical feature of the participants. Six hundred and
 120 twenty eight patients (60.50%) were diagnosed with a tumor less than 5cm. Six hundred and forty
 121 patients (61.66%) were with grade III or IV. The numbers of patients with T1, T2, T3 and T4
 122 respectively were 94, 125, 717 and 102. The numbers of patients with N1, N2 and N3 respectively
 123 were 488, 331 and 219. Seventy five patients (7.23%) were with distant metastasis at presentation.
 124 The median CSS was 27 months. The 1-, 2- and 5-year CSS rates were 76.8%, 53.0% and 29.2%,
 125 respectively. There was no statistical difference of baseline characteristics between training set
 126 and validating set. The detailed information of the two sets was also presented in Table 1.

Table 1. Baseline information of the included patients with node positive CGA,
 N(%).

Groups	Training set (N = 723)	Validating set (N = 315)	P-value
--------	---------------------------	-----------------------------	---------

Sex				
Male	596 (82.43)	261 (82.86)		0.939
Female	127 (17.57)	54 (17.14)		
Age				
<70	490 (67.77)	210 (66.67)		0.781
≥70	233 (32.23)	105 (33.33)		
Race				
White	628 (86.86)	268 (85.08)		
Black	40 (5.53)	24 (7.62)		0.437
Others	55 (7.61)	23 (7.30)		
Tumor size				
<5cm	442 (61.13)	186 (59.05)		0.573
≥5cm	281 (38.87)	129 (40.95)		
Grade				
I-II	279 (38.59)	119 (37.78)		0.859
III-IV	444 (61.41)	196 (62.22)		
T stage				
T1	70 (9.68)	24 (7.62)		
T2	83 (11.48)	42 (13.33)		0.600
T3	501 (69.30)	216 (68.57)		
T4	69 (9.54)	33 (10.48)		
N stage				
N1	348 (48.13)	140 (44.44)		0.545
N2	225 (31.12)	106 (33.65)		
N3	150 (20.75)	69 (21.91)		
M stage				
M0	678 (93.78)	285 (90.48)		0.079
M1	45 (6.22)	30 (9.52)		
Median survival (months)	28 (25, 32)	25 (21, 32)		0.361
CSS rate (%)				
1-year	77.0 (74.0, 80.2)	76.3 (71.6, 81.2)		
2-year	53.7 (50.1, 57.5)	51.4 (46.0, 57.5)		
5-year	30.3 (26.7, 34.5)	26.4 (20.9, 33.4)		

Abbreviation: CGA, cardia gastric adenocarcinoma; CSS, cancer-specific survival

127 According to X-tile software results, the optimal cut-off values for LNR were 0.09 and 0.33, and
 128 for LODDS were -2.09 and -0.65. Thus patients were separated into low (R1), medium (R2) or
 129 high LNR (R3) group, or low (L1), medium (L2) or high LODDS (L3) group. Next we illustrated
 130 the survival curves of the patients according to N, LNR or LODDS staging system. As shown in
 131 Figure 2 training set section, CSS was significantly different by all the three staging systems (all
 132 the log-rank P values < 0.0001); however the 95% CIs of N2 and N3 survival curve initially
 133 separated and partly overlapped afterwards. The inferior discriminative ability of N system was
 134 further supported by AIC and C-index. As shown in Table 2, the C-index of N stage was lower
 135 than that of LNR or LODDS. Similarly, the AIC of N stage was higher than that of LNR or

136 LODDS. The prognostic value of adjusted model was better than crude mode generally. In
 137 addition, the value of LNR system seemed to be worse than LODDS system; however the
 138 difference was not noticeable, so we considered both of the systems into nomogram construction.

Table 2. Prognostic values of variables for patients with node positive CGA (N = 1038).

Variables	Crude model			Adjusted model		
	HR (95% CI)	C-index	AIC	HR (95% CI)	C-index	AIC
Training set (N = 723)						
N stage		0.572	5412		0.633	5379
N1	1 (ref)			1 (ref)		
N2	1.44 (1.17, 1.79)			1.34 (1.07, 1.66)		
N3	1.98 (1.57, 2.51)			1.71 (1.34, 2.19)		
LNR*		0.607	5376		0.655	5343
R1	1 (ref)			1 (ref)		
R2	1.88 (1.44, 2.44)			1.83 (1.40, 2.39)		
R3	3.02 (2.30, 3.97)			2.74 (2.07, 3.63)		
LODDS*		0.609	5373		0.656	5339
L1	1 (ref)			1 (ref)		
L2	1.93 (1.48, 2.51)			1.86 (1.42, 2.44)		
L3	3.13 (2.38, 4.13)			2.87 (2.16, 3.81)		
Validating set (N = 315)						
N stage		0.603	1953		0.681	1926
N1	1 (ref)			1 (ref)		
N2	1.89 (1.36, 2.63)			1.88 (1.34, 2.64)		
N3	2.34 (1.62, 3.38)			2.18 (1.47, 3.24)		
LNR*		0.646	1927		0.702	1902
R1	1 (ref)			1 (ref)		
R2	2.20 (1.47, 3.30)			2.14 (1.41, 3.23)		
R3	4.16 (2.76, 6.28)			4.00 (2.59, 6.17)		
LODDS*		0.647	1927		0.703	1901
L1	1 (ref)			1 (ref)		
L2	2.07 (1.39, 3.09)			2.08 (1.38, 3.14)		
L3	4.22 (2.79, 6.39)			4.10 (2.65, 6.34)		

Abbreviations: CGA, cardia gastric adenocarcinoma; HR, hazard ratio; CI, confidence interval; AIC, Akaike information criterion; LNR, lymph node ratio; LODDS, log odds of positive lymph nodes.

Adjusted model considered age, sex, race, tumor size, grade, T stage and M stage.

* cut-off values for LNR were 0.09 and 0.33, and for LODDS were -2.09 and -0.65.

139 Stepwise Cox regression analysis showed age, race, tumor grade, T stage, M stage and LNR (or
 140 LODDS) were independent predictors, so these factors were included in nomograms. For both
 141 LNR and LODDS, the total score was 40, and higher score suggested lower survival (Figure 3A
 142 and 4A). Next calibration plot was used to assess the internal and external validity of the
 143 nomogram (Figure 3B, 3C, 4B and 4C). Since the cross-spot line was generally close to the grey

1
2
3 144 reference line, we concluded the predicted CSS was well correlated with the actual situation.
4

5 145 **Discussion**
6

7 146 The present study analyzes national cancer registry databases and demonstrates that survival of
8 147 patients with node-positive CGA is well predicted when the traditional N staging method is
9 148 substituted with LNR or LODDS system. This finding both exists in training and validating sets.
10 149 In training set, the survival curves separate clearly when patient grouping is implemented by LNR
11 150 or LODDS method, which is not achieved by traditional N staging system. Adjusted model that
12 151 simultaneously considers staging, clinical and demographic features outperforms crude model that
13 152 only takes staging into account. Therefore multiple independent survival factors are incorporated
14 153 in nomogram construction, which suggests older age at diagnosis, white, higher grade, greater
15 154 tumor infiltration, higher proportion of positive LN, and metastasis as risk factors. The
16 155 nomograms perform steadily in 1-, 2- and 5-year CSS prediction as the validation plots show.

17
18
19
20
21 156 Previous studies have demonstrated the superiority of LNR or LODDS for prognostic prediction
22 157 in GC after surgical resection (8-10, 15-17). However the GC patients are not further separated
23 158 and investigated according to primary tumor site, since there is much difference between cardia
24 159 and non-cardia GC in terms of tumor features, etiological factors, and biological behaviors (3). In
25 160 AJCC cancer staging 7th edition, tumors involving EGJ was categorized as esophagus cancer (5),
26 161 which was however argued by the viewpoint that GC staging system has a better ability to predict
27 162 survival of EGJ tumor (18, 19). In the latest 8th edition (20), a tumor that has its epicenter within 2
28 163 cm of EGJ and involves the EGJ (Siewert type I/II) is classified as esophageal cancer. Other
29 164 situation, including a tumor with epicenter more than 2 cm from EGJ or a tumor located with 2 cm
30 165 of EGJ but does not involve EGJ, is classified as stomach cancer. The superiority of the new
31 166 system is confirmed by a retrospective observational study from two high-volume institutions in
32 167 China, regardless of Siewert type (21). In terms of Siewert type II junctional adenocarcinoma, a
33 168 marginal advantage of the esophagus cancer system is found in discriminating survival rates after
34 169 3 and 5 years, however the advantage of GC system lies in division of the N3 category into N3a
35 170 and N3b, so the authors concludes neither the esophageal nor the stomach staging system is
36 171 flawless in predicting survival in Siewert type II junctional cancer (22). Above all, CGA is
37 172 probably a special entity that has a different biological property compared with genuine gastric
38 173 and genuine esophageal cancer. To the best of our knowledge, the present study first reveals a
39 174 superior performance of prognostic prediction based on LNR or LODDS in node positive CGA
40 175 patients. Unfortunately we are unable to consider Siewert type due to unavailable information
41 176 from SEER database; therefore we encourage further studies to pay special attention on tumor
42 177 location.

43
44
45
46
47
48
49
50 178 LNR and LODDS have been proved to be the strongest indicators of survival in gastric
51 179 adenocarcinoma when LN harvest is inadequate (16, 17). It is demonstrated that in general, more
52 180 LN resected is associated with better survival, which may be the result of either improved N
53 181 classification or a therapeutic effect of lymphadenectomy. For esophageal cancer, worldwide data
54 182 shows that yielding 10 nodes for pT1, 20 for pT2, and 30 or more for pT3/T4 is recommended for
55 183 maximum 5-year survival (23). For GC, greater LN harvest also shows improved survival (24). It
56 184 is suggested that at least 16 nodes be assessed pathologically and evaluation of more than 30
57 185 nodes is desirable (25). Overall it is encouraged to harvest as many LN as possible, balancing the

186 extent of LN resection necessary for accurate N staging and maximum survival without
187 unnecessarily increasing the morbidity of radical lymphadenectomy. Nevertheless, many
188 conditions would lead to insufficient LN harvest. It is estimated that only one fifth GC patients
189 have sufficient LN examined in Iran (26), while more than 15 LNs are examined in 64% of
190 patients in the US (25). The LNR and LODDS staging methods do not require adequate number of
191 LN assessment. In fact, the new N category method is stable when nodal assessment is insufficient
192 during surgery not only for GC (8, 15-17) but also for colorectal cancer (27), esophageal cancer
193 (28), oral squamous cell carcinoma (29), gallbladder cancer (30), etc.

194 One limitation of this study is that the recruited patients were diagnosed during 2010-2015 and
195 staged based on TNM 7th edition that defined 3 N categories. In 8th edition, N category of GC
196 includes N1, N2, N3a and N3b, which improves survival prediction in patients with junctional
197 cancer (22). So whether LNR or LODDS based staging system outperforms TNM 8th edition
198 needs to be further investigated. Another limitation is that our results are based on training set and
199 confirmed by validating set; however the features of the two groups are similar. So this finding
200 needs to be proved among populations with distinct features.

201 In conclusion, LNR and LODDS staging methods have better prognostic efficacy than traditional
202 N staging method in CGA patients with regional node metastasis. Besides, the two values are
203 promising substitute for N staging in nomogram development when other independent prognostic
204 factors are incorporated.

205 **a. Contributorship statement:** WXQ work conception, data interpretation, critical review for
206 important content, final approval of the manuscript and agreement to be accountable for all
207 aspects of the work; BM administrative work, funding, critical review for important content,
208 final approval of the manuscript and agreement to be accountable for all aspects of the work;
209 ZC acquisition, analysis of data, drafting the work, final approval of the manuscript and
210 agreement to be accountable for all aspects of the work.

211 **b. Competing interests:** None.

212 **c. Funding:** This work was supported by [Major Project of Natural Science of Anhui Provincial
213 Department of Education] grant number [KJ2019ZD73].

214 **d. Data sharing statement:** Dataset available from the SEER website
215 (<https://seer.cancer.gov/data-software/>).

216
217 **Acknowledgement:** We thank all patient advisers for their special contribution.

218 219 **References:**

- 220 1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018:
221 GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA*
222 *Cancer J Clin.* [Journal Article]. 2018 2018-11-01;68(6):394-424.
- 223 2. Abdi E, Latifi-Navid S, Zahri S, Yazdanbod A, Pourfarzi F. Risk factors predisposing to cardia
224 gastric adenocarcinoma: Insights and new perspectives. *Cancer Med.* [Journal Article; Research
225 Support, Non-U.S. Gov't; Review]. 2019 2019-10-01;8(13):6114-26.
- 226 3. Amini N, Spolverato G, Kim Y, Squires MH, Poultsides GA, Fields R, et al. Clinicopathological
227 features and prognosis of gastric cardia adenocarcinoma: a multi-institutional US study. *J SURG*
228 *ONCOL.* [Comparative Study; Journal Article; Multicenter Study]. 2015 2015-03-01;111(3):285-92.

- 229 4. Sobin L, Gospodarowicz M, Wittekind C. TNM classification of malignant tumours. 7th edition.:
230 Wiley-Blackwell; 2009.
- 231 5. Edge S, Byrd D, Compton C, Fritz A, Greene F, Trotti A. AJCC cancer staging manual. 7th
232 edition.: Springer; 2009.
- 233 6. Hou Y, Wang X, Chen J. Prognostic significance of metastatic lymph node ratio: the lymph node
234 ratio could be a prognostic indicator for patients with gastric cancer. *WORLD J SURG ONCOL.*
235 [Journal Article]. 2018 2018-10-04;16(1):198.
- 236 7. Zhu J, Xue Z, Zhang S, Guo X, Zhai L, Shang S, et al. Integrated analysis of the prognostic role
237 of the lymph node ratio in node-positive gastric cancer: A meta-analysis. *INT J SURG.* [Journal Article;
238 Meta-Analysis; Review]. 2018 2018-09-01;57:76-83.
- 239 8. Sun Z, Xu Y, Li DM, Wang ZN, Zhu GL, Huang BJ, et al. Log odds of positive lymph nodes: a
240 novel prognostic indicator superior to the number-based and the ratio-based N category for gastric
241 cancer patients with R0 resection. *CANCER-AM CANCER SOC.* [Journal Article; Research Support,
242 Non-U.S. Gov't]. 2010 2010-06-01;116(11):2571-80.
- 243 9. Zhao E, Zhou C, Chen S. Prognostic nomogram based on log odds of positive lymph nodes for
244 gastric carcinoma patients after surgical resection. *FUTURE ONCOL.* [Journal Article]. 2019
245 2019-12-01;15(36):4207-22.
- 246 10. Gu P, Deng J, Sun Z, Wang Z, Wang W, Liang H, et al. Superiority of log odds of positive lymph
247 nodes (LODDS) for prognostic prediction after gastric cancer surgery: a multi-institutional analysis of
248 7620 patients in China. *SURG TODAY.* [Journal Article]. 2021 2021-01-01;51(1):101-10.
- 249 11. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov)
250 SEER*Stat Database: Incidence - SEER Research Data, 18 Registries, Nov 2019 Sub (2000-2017) -
251 Linked To County Attributes - Time Dependent (1990-2017) Income/Rurality, 1969-2018 Counties,
252 National Cancer Institute, DCCPS, Surveillance Research Program, released April 2020, based on the
253 November 2019 submission..
- 254 12. Yu Y, Zhang P, Yao R, Wang J, Wang P, Xue X, et al. Prognostic value of log odds of positive
255 lymph nodes in node-positive lung squamous cell carcinoma patients after surgery: a SEER
256 population-based study. *Transl Lung Cancer Res.* [Journal Article]. 2020 2020-08-01;9(4):1285-301.
- 257 13. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker
258 assessment and outcome-based cut-point optimization. *CLIN CANCER RES.* [Journal Article;
259 Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, Non-P.H.S.; Research Support, U.S.
260 Gov't, P.H.S.]. 2004 2004-11-01;10(21):7252-9.
- 261 14. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer
262 prognosis. *J CLIN ONCOL.* [Journal Article; Review]. 2008 2008-03-10;26(8):1364-70.
- 263 15. Jian-Hui C, Shi-Rong C, Hui W, Si-le C, Jian-Bo X, Er-Tao Z, et al. Prognostic value of three
264 different lymph node staging systems in the survival of patients with gastric cancer following D2
265 lymphadenectomy. *Tumour Biol.* [Journal Article]. 2016 2016-08-01;37(8):11105-13.
- 266 16. Aurello P, Petrucciani N, Nigri GR, La Torre M, Magistri P, Tierno S, et al. Log odds of positive
267 lymph nodes (LODDS): what are their role in the prognostic assessment of gastric adenocarcinoma? *J*
268 *GASTROINTEST SURG.* [Comparative Study; Journal Article]. 2014 2014-07-01;18(7):1254-60.
- 269 17. Kilic MO, Gundogdu SB, Ozden S, Saylam B, Tez M. The prognostic value of different node
270 staging systems in patients with ≤ 15 lymph nodes following surgery for gastric adenocarcinoma.
271 *ACTA CHIR BELG.* [Journal Article]. 2018 2018-02-01;118(1):1-6.
- 272 18. Suh YS, Han DS, Kong SH, Lee HJ, Kim YT, Kim WH, et al. Should adenocarcinoma of the

- 1
2
3 273 esophagogastric junction be classified as esophageal cancer? A comparative analysis according to the
4 274 seventh AJCC TNM classification. *ANN SURG*. [Comparative Study; Journal Article; Research
5 275 Support, Non-U.S. Gov't]. 2012 2012-05-01;255(5):908-15.
- 6 276 19. Huang Q, Shi J, Feng A, Fan X, Zhang L, Mashimo H, et al. Gastric cardiac carcinomas involving
7 277 the esophagus are more adequately staged as gastric cancers by the 7th edition of the American Joint
8 278 Commission on Cancer Staging System. *Mod Pathol*. [Journal Article; Research Support, Non-U.S.
9 279 Gov't; Validation Study]. 2011 2011-01-01;24(1):138-46.
- 10 280 20. Amin M, Edge S, Greene F, Byrd D, Brookland R. *AJCC cancer staging manual*. 8th edition.:
11 281 Springer; 2016.
- 12 282 21. Liu K, Feng F, Chen XZ, Zhou XY, Zhang JY, Chen XL, et al. Comparison between gastric and
13 283 esophageal classification system among adenocarcinomas of esophagogastric junction according to
14 284 AJCC 8th edition: a retrospective observational study from two high-volume institutions in China.
15 285 *GASTRIC CANCER*. [Comparative Study; Journal Article; Observational Study]. 2019
16 286 2019-05-01;22(3):506-17.
- 17 287 22. Karstens KF, Ghadban T, Sawez S, Konczalla L, Woestemeier A, Bachmann K, et al. Comparison
18 288 of the 8th UICC staging system for esophageal and gastric cancers in Siewert type II junctional
19 289 adenocarcinomas. *Eur J Surg Oncol*. [Comparative Study; Journal Article]. 2020 2020-04-01;46(4 Pt
20 290 A):638-43.
- 21 291 23. Rizk NP, Ishwaran H, Rice TW, Chen LQ, Schipper PH, Kesler KA, et al. Optimum
22 292 lymphadenectomy for esophageal cancer. *ANN SURG*. [Journal Article; Research Support, Non-U.S.
23 293 Gov't]. 2010 2010-01-01;251(1):46-50.
- 24 294 24. Macalindong SS, Kim KH, Nam BH, Ryu KW, Kubo N, Kim JY, et al. Effect of total number of
25 295 harvested lymph nodes on survival outcomes after curative resection for gastric adenocarcinoma:
26 296 findings from an eastern high-volume gastric cancer center. *BMC CANCER*. [Journal Article;
27 297 Research Support, Non-U.S. Gov't]. 2018 2018-01-12;18(1):73.
- 28 298 25. Schoenleber SJ, Schnelldorfer T, Wood CM, Qin R, Sarr MG, Donohue JH. Factors influencing
29 299 lymph node recovery from the operative specimen after gastrectomy for gastric adenocarcinoma. *J*
30 300 *GASTROINTEST SURG*. [Journal Article]. 2009 2009-07-01;13(7):1233-7.
- 31 301 26. Khanjani N, Mirzaei S, Nasrolahi H, Hamed SH, Mosalaei A, Omidvari S, et al. Insufficient
32 302 lymph node assessment in gastric adenocarcinoma. *J Egypt Natl Canc Inst*. [Journal Article]. 2019
33 303 2019-10-22;31(1):2.
- 34 304 27. Pei JP, Zhang CD, Fan YC, Dai DQ. Comparison of Different Lymph Node Staging Systems in
35 305 Patients With Resectable Colorectal Cancer. *FRONT ONCOL*. [Journal Article]. 2018
36 306 2018-01-20;8:671.
- 37 307 28. Cao J, Yuan P, Ma H, Ye P, Wang Y, Yuan X, et al. Log Odds of Positive Lymph Nodes Predicts
38 308 Survival in Patients After Resection for Esophageal Cancer. *ANN THORAC SURG*. [Journal Article;
39 309 Multicenter Study]. 2016 2016-08-01;102(2):424-32.
- 40 310 29. Jin W, Zhu Z, Wu Y, Ding X, Wu H, Song X, et al. Prognostic value of log odds of positive
41 311 lymph nodes in patients with resectable oral squamous cell carcinoma. *ORAL ONCOL*. [Journal
42 312 Article; Research Support, Non-U.S. Gov't]. 2020 2020-09-01;108:104709.
- 43 313 30. Li J, Sun Y, Zhao B, Tang C, Fan D, Jiang W, et al. Lymph Node Ratio-Based Staging System for
44 314 Gallbladder Cancer With Fewer Than Six Lymph Nodes Examined. *FRONT ONCOL*. [Journal
45 315 Article]. 2020 2020-01-20;10:542005.
- 46
47
48
49
50
51
52
53
54
55
56
57
58
59 316
60

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

317 **Illustrations**

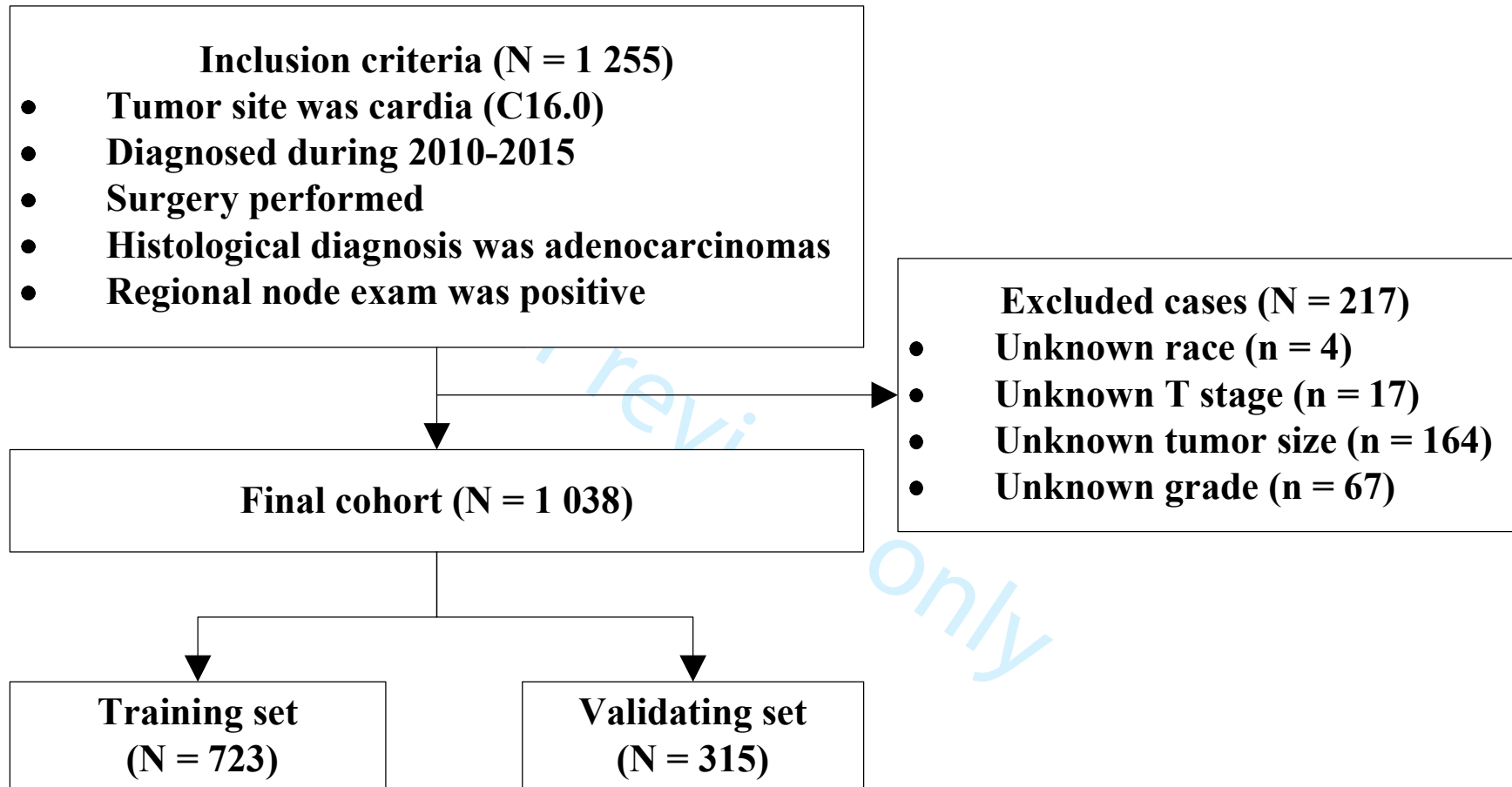
318 Figure 1. Flow diagram of patient selection and grouping.

319 Figure 2. Survival curves of training and validating sets by different staging systems.

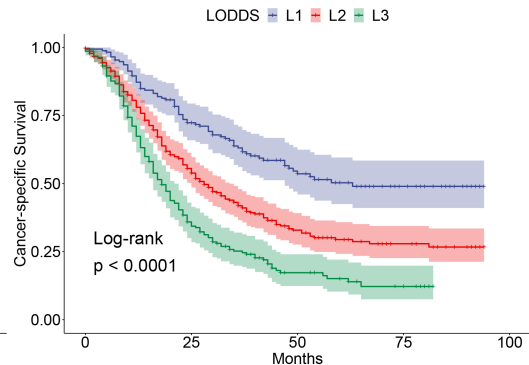
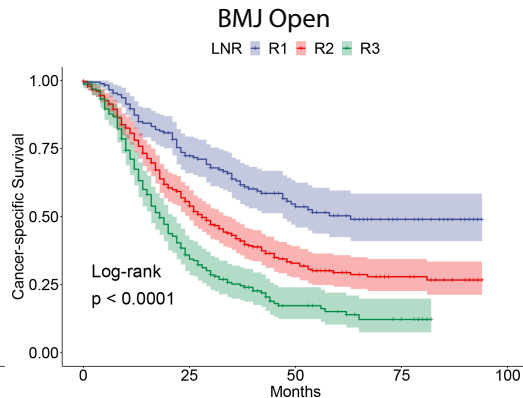
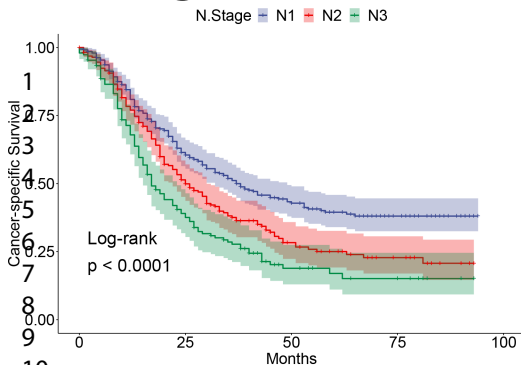
320 Figure 3. Construction and validation of nomogram based on Tumor-Lymph node ratio-Metastasis
321 staging system. B) Internal validation; C) External validation.

322 Figure 4. Construction and validation of nomogram based on Tumor-Log odds of positive lymph
323 node-Metastasis staging system. B) Internal validation; C) External validation.

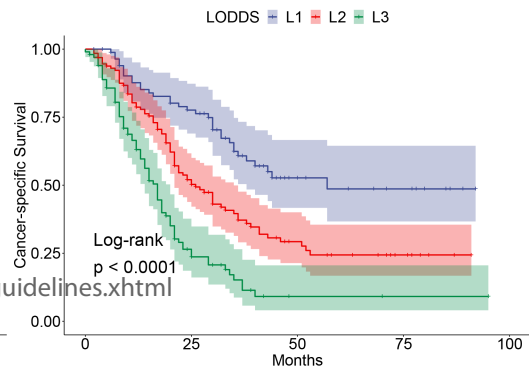
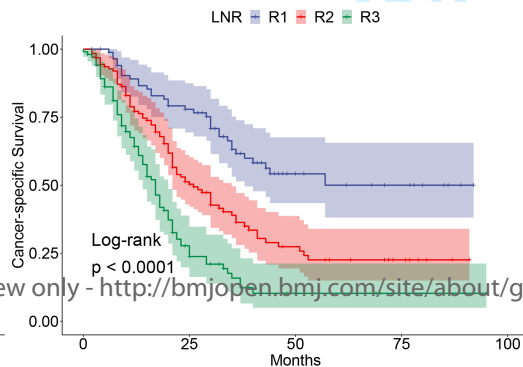
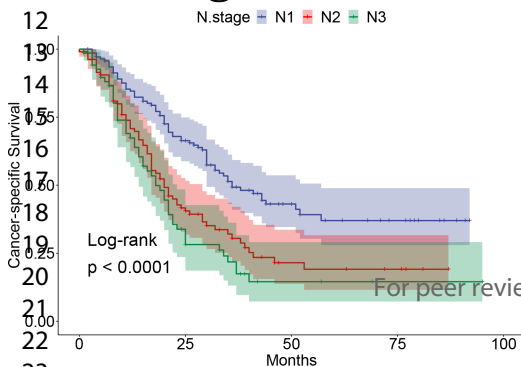
For peer review only

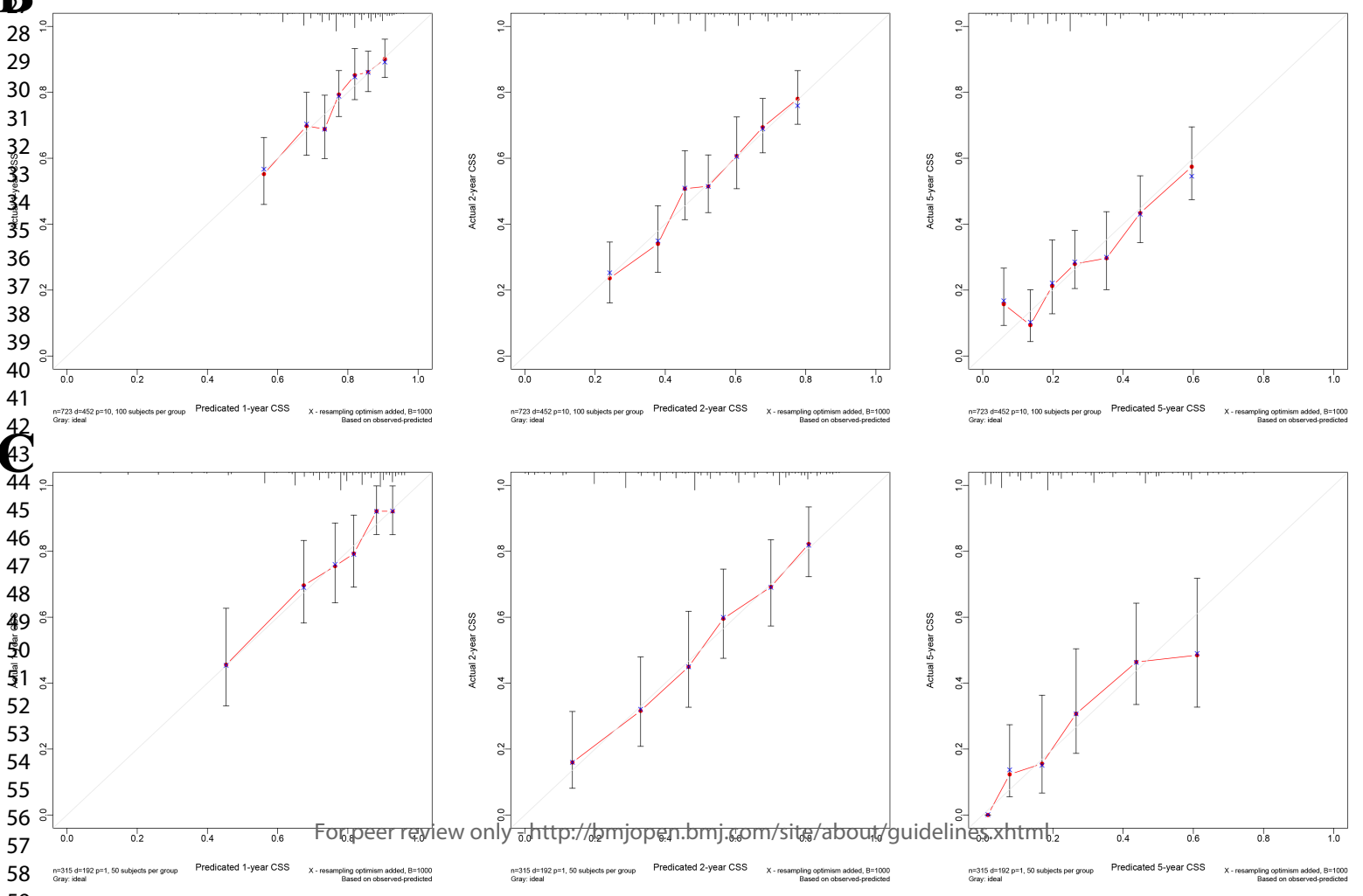
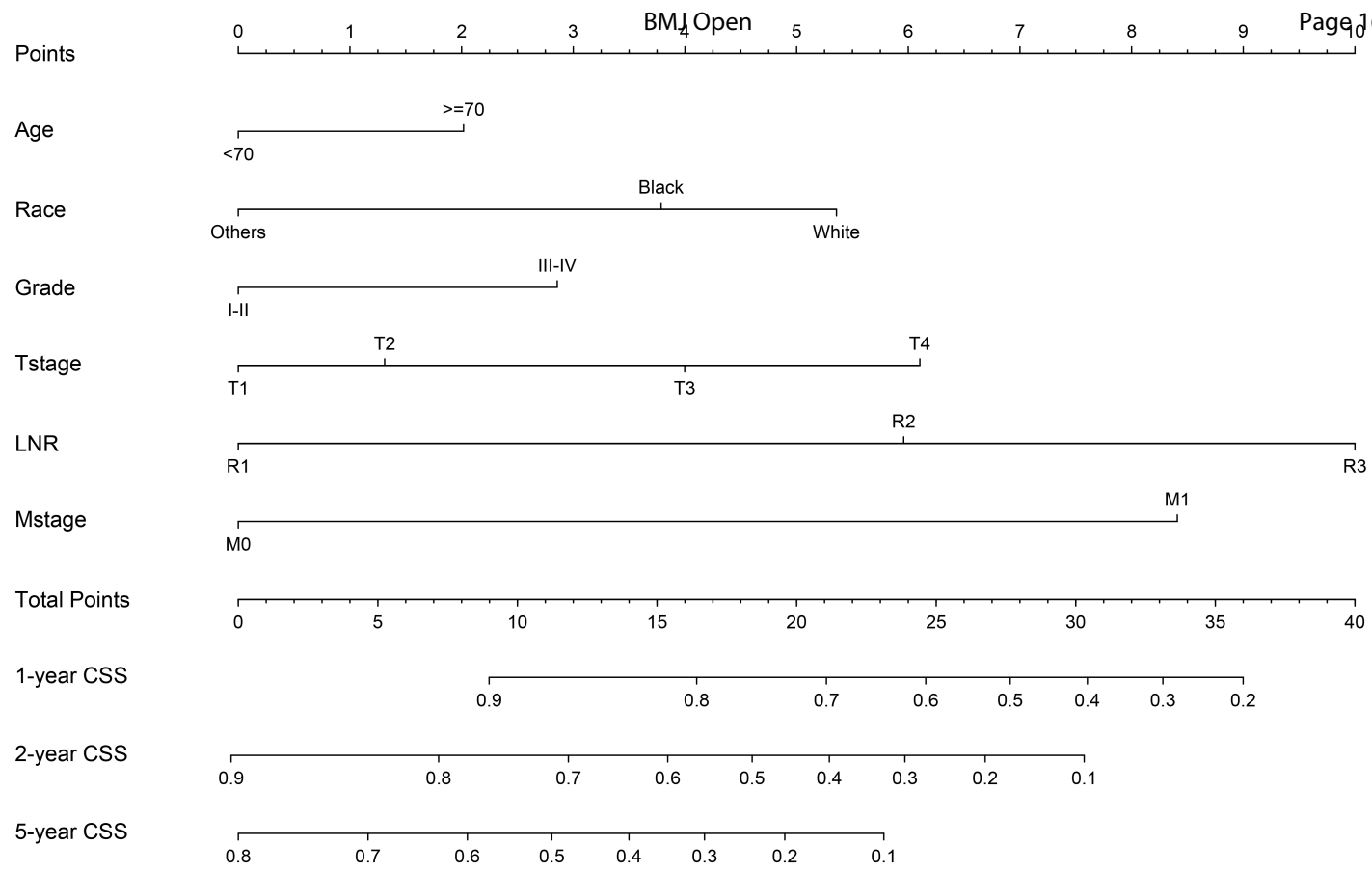


Training Set

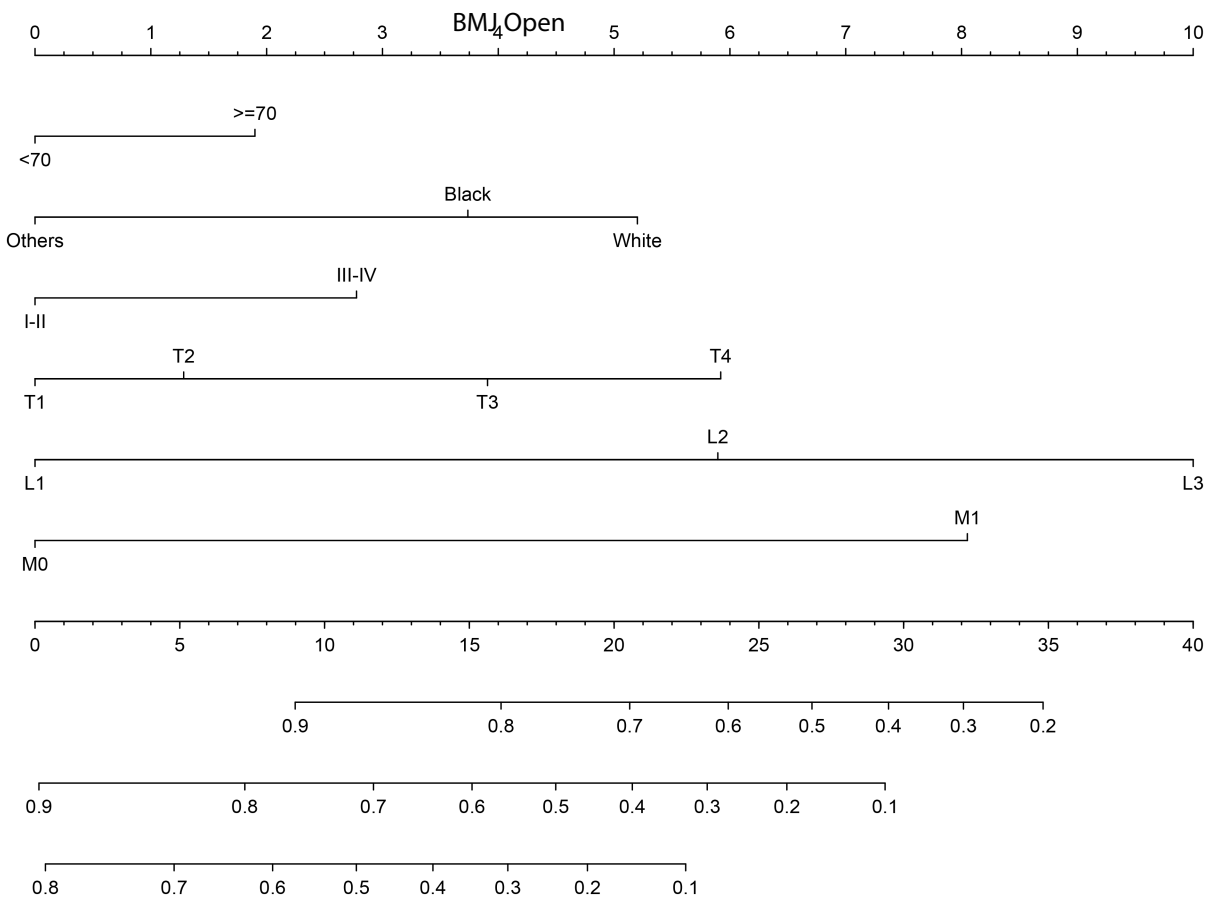


Validating Set



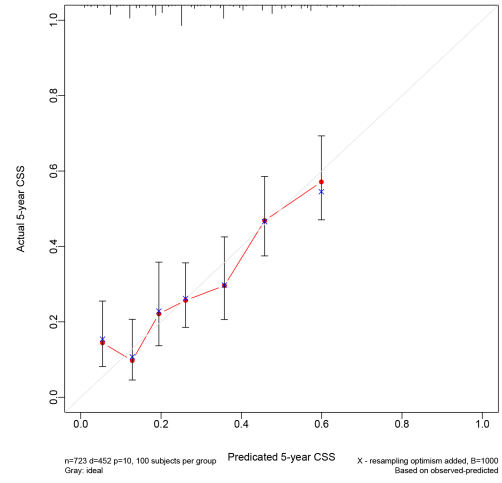
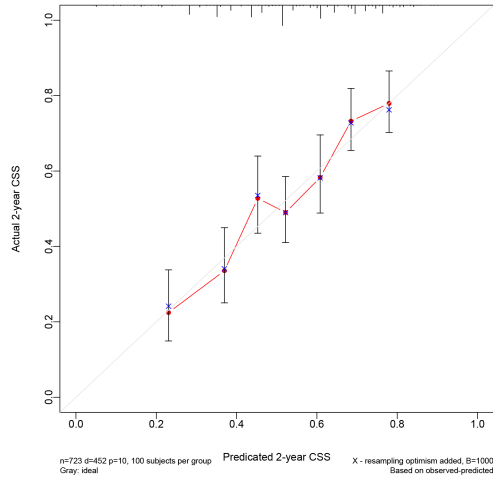
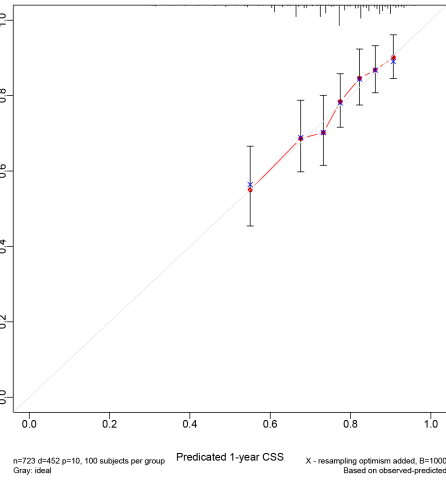


- 1 Age
- 2
- 3
- 4 Race
- 5
- 6 Grade
- 7
- 8
- 9 Tstage
- 10
- 11 LODDS
- 12
- 13
- 14 Mstage
- 15
- 16 Total Points
- 17
- 18
- 19 1-year CSS
- 20
- 21
- 22 2-year CSS
- 23
- 24 5-year CSS
- 25



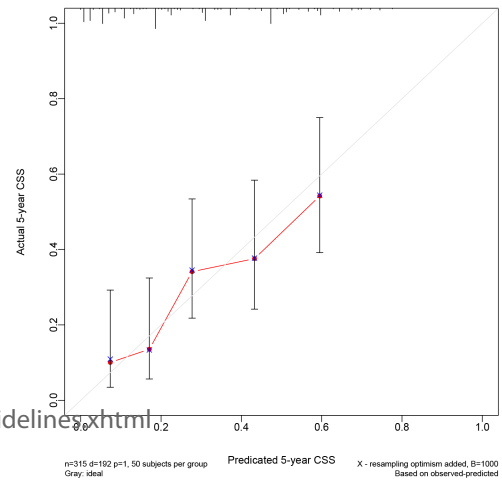
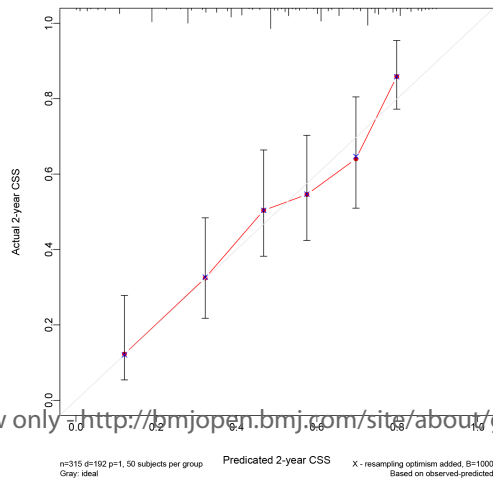
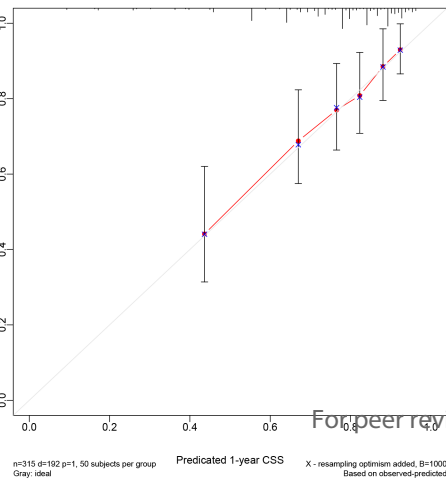
B

- 28
- 29
- 30
- 31
- 32
- 33
- 34
- 35
- 36
- 37
- 38
- 39
- 40
- 41
- 42



C

- 43
- 44
- 45
- 46
- 47
- 48
- 49
- 50
- 51
- 52
- 53
- 54
- 55
- 56
- 57
- 58
- 59



BMJ Open

Prognostic value of different lymph node staging methods for node positive cardia gastric cancer: a register-based retrospective cohort study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2021-050378.R1
Article Type:	Original research
Date Submitted by the Author:	18-Jun-2021
Complete List of Authors:	Wang, Xiao-Qing; Anhui Medical College Bao, Min; Anhui Medical College Zhang, Cheng; First Affiliated Hospital of Anhui Medical University
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	Gastrointestinal tumours < ONCOLOGY, Gastrointestinal tumours < GASTROENTEROLOGY, SURGERY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **Title page**

2 **Article title:** Prognostic value of different lymph node staging methods for node positive cardia
3 gastric cancer: a register-based retrospective cohort study

4 **Author information:** Xiao-Qing Wang 1, Min Bao 1, Cheng Zhang 2

5 1 Anhui Medical College, Hefei, Anhui, China PR

6 2 Anhui Provincial Cancer Institute, the First Affiliated Hospital of Anhui Medical University,
7 Hefei, Anhui, China PR

8 Corresponding to: Cheng Zhang, ahmuzc@sina.cn

9 **Disclaimer:** The views expressed in the submitted article are our own and not an official position
10 of the institution or funder.

11 **Word count:** 3191 (excluding its abstract, acknowledgments, tables, figure legends, and
12 references) abstract 281 .

13 **Number of figures and tables:** This draft included 3 figures and 2 tables.

14 **Disclosure of relationships and activities:** None.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

15 **Abstract**

16 Objective: To investigate the prognostic efficacy of lymph node ratio (LNR) and log odds of
17 positive lymph nodes (LODDS) in node positive cardia gastric adenocarcinoma (CGA).

18 Design: A register-based retrospective cohort study.

19 Participants: A total of 1 038 patients with node positive CGA were enrolled from SEER database,
20 and randomly assigned (7:3) in training set (N = 723) or validating set (N = 315).

21 Interventions: The major endpoint was cancer specific survival (CSS). Optimal cut-off values
22 were determined by X-tile software. The prognostic power was evaluated using Akaike
23 Information Criterion (AIC) and Harrell concordance index (C-index). Cox stepwise regression
24 analysis was performed to construct nomogram for prediction of 1-, 2-, and 5-year CSS. The
25 prediction model was further evaluated by calibration curve, receiver operator characteristic (ROC)
26 curve and decision curve analysis (DCA) plot.

27 Results: The training set and validating set are similar in terms of clinical and demographic
28 features. The optimal cut-off values for LNR were 0.09 and 0.33, and for LODDS were -2.09 and
29 -0.65. CSS was significantly different by N, LNR and LODDS categories. The C-index of N stage
30 was lower than that of LNR or LODDS. The AIC of N stage was higher than that of LNR or
31 LODDS. Independent predictors included race, T stage, M stage and LNR (or LODDS) and they
32 were incorporated in nomograms for 1-, 2- and 5-year CSS prediction. Calibration plots showed
33 satisfied results of internal and external validity of the nomogram.

34 Conclusions: LNR and LODDS staging methods have better prognostic efficacy than traditional N
35 staging method in CGA patients with node metastasis. Besides, the two values are promising
36 substitute for N staging in nomogram development when other independent prognostic factors are
37 incorporated.

38
39 Key words: Cardia; Adenocarcinoma; Stomach Neoplasms; Lymph Node Ratio; Neoplasm
40 Staging

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

41 **Strengths and limitations of this study**

- 42 ● This study used national cancer registry data for cardia gastric adenocarcinoma research;
- 43 ● Novel staging methods based on the number of positive lymph node was established for
44 prognostic prediction;
- 45 ● Nomograms based on the new staging methods were constructed and validated;
- 46 ● This study needs to be confirmed by other populations.

47 **Patient consent form:** The SEER database review is granted exemption from obtaining patients'
48 consents.

49 **Word count:** 3191 (excluding its abstract, acknowledgments, tables, figure legends, and
50 references) abstract 281.

For peer review only

51 Introduction

52 Gastric cancer (GC) generally includes 2 topographical categories: non-cardia GC that arises from
53 more distant sites and cardia GC that arises in gastroesophageal junction (GEJ). In contrast to
54 steady decline of non-cardia GC incidence, cardia GC occurs more frequently, particular in
55 high-income countries (1, 2). This trend is associated with obesity, gastroesophageal reflux
56 disease (GERD), and Barrett esophagus (2). In addition to different incidence trend,
57 clinicalpathological feature and long-term survival vary between the two GC subtypes (3). For
58 accurate prediction of survival, precise staging is required. The Tumor-Node-Metastasis (TNM)
59 classification 7th edition by the American Joint Committee on Cancer (AJCC) recommends at least
60 15 lymph nodes (LN) collection for N staging (4, 5). However inadequate LN harvest frequently
61 occurs due to many conditions, thus precise staging cannot be obtained sometimes. It has been
62 demonstrated that LN ratio (LNR) could better estimate survival of GC patients after curative
63 gastrectomy, regardless of the number of LN examined (6), and may be promising for aiding
64 TNM staging system (7). Apart from that, log odds of positive LN (LODDS) outperformed N and
65 LNR staging system when predicting survival of GC patients (8-10). Therefore the traditional N
66 staging classification may be substituted with different methods, with even improved performance.
67 Nevertheless little evidence evaluates the performance of the two LN staging systems
68 aforementioned in cardia GC, since it has distinct clinical characteristics and epidemiology from
69 overall GC.

70 Here we use nationwide cancer registry data to appraise the prognostic value of LNR and LODDS
71 in patients with node positive cardia gastric adenocarcinoma (CGA), and, if possible, construct
72 nomogram for survival prediction based on the new LN staging system.

73 Methods

74 *Study design and Participants Selection*

75 This study is a SEER register-based retrospective cohort study, which aimed to enroll patients
76 with node positive cardia gastric adenocarcinoma (CGA), review crucial clinical characteristics
77 and observe survival of this population. The source of SEER data is registered cancer cases from
78 various locations throughout the United States. The permission of data access was obtained by
79 sending application form and receiving confirmation mail with valid username (21268-Nov2019)
80 and password.

81 We used SEER*Stat (version 8.3.8) to access to Incidence - SEER Research Data, 18 Registries,
82 Nov 2019 Sub 2000-2017 (SEER 18 database) (11) for collection of node positive CGA patients.
83 The inclusion criteria were as follows: 1) the International Classification of Disease for Oncology,
84 Third Edition (ICD-O-3) for primary tumor site was C16.0 (cardia); 2) broad histological recode
85 was 8140-8389: adenomas and adenocarcinomas; 3) diagnostic confirmation was positive
86 histology; 4) surgery was performed; 5) diagnosed during 2010-2015; 6) the definite number of
87 regional positive nodes was clear and not zero. We excluded cases with unknown race, T stage
88 information, tumor size and grade. As shown in Figure 1, the final cohort enrolled 1 038 patients
89 with node positive CGA, of whom 857 were male and 181 were female. Three hundred and thirty
90 eight (32.56%) were over 70 years old. Eight hundred and ninety six (86.32%) were white, 64
91 were black and 78 were other races. Next 70% of the entire patients were randomly assigned to

1
2
3 92 training set (N = 723) and the rest was assigned to validating set (N = 315).
4

5 93 ***Technical Information***

6
7 94 The main outcome was cancer specific survival (CSS), which was referred to as death specifically
8 95 due to CGA and the period between first diagnosis and death. In addition, we extracted the
9 96 following variables for analysis: sex, race, age, AJCC 7th TNM stage information, tumor size,
10 97 grade, number of regional nodes examined and number of regional nodes positive. The stage
11 98 information was further corrected according to AJCC 8th criteria. LNR and LODDS were
12 99 calculated as previously reported (12). Briefly, LNR was defined as the ratio of the number of
13 100 positive nodes divided by the total number of examined nodes. LODDS was calculated using the
14 101 formula: $\log(\text{NPLN}+0.50)/(\text{NDLN}-\text{NPLN}+0.50)$, in which 0.50 was added to both the numerator
15 102 and denominator to avoid an infinite number.

16
17
18
19 103 The optimal thresholds for cutting LNR and LODDS into trichotomous variables were determined
20 104 by X-tile software (version 3.6.1) (13), which were based on the maximal log-rank chi-square
21 105 value that represented the greatest group difference of CSS probability. LNR and LODDS were
22 106 cut into 3 levels because they are proposed as the alternative indicators for N stage in node
23 107 positive GC that included N1, N2 and N3.

24 108 ***Statistics***

25
26
27 109 The distributions of baseline features between training set and validating set were described and
28 110 compared by chi-square test. Survival curves, median survival and CSS rates were generated using
29 111 the Kaplan-Meier method. Outcome difference between groups was analyzed by the log-rank test.
30 112 After testing proportional hazard assumption, multivariable Cox regression model was used to
31 113 establish prognostic model for CSS. The prognostic power was evaluated using Akaike
32 114 Information Criterion (AIC) and Harrell concordance index (C-index). A predictive model with
33 115 lower AIC indicated better model fit, while with higher C-index indicated better discriminative
34 116 ability. A value of C-index of 0.5 indicates no predictive power, and an index of 1.0 indicates
35 117 complete differentiation. Cox stepwise regression analysis was also performed to construct
36 118 nomogram for prediction of 1-, 2-, and 5-year CSS. Validation of nomogram was performed by
37 119 internal and external calibration plots (14). Bootstraps with 1 000 resample were used for
38 120 validation activities. Receiver operator characteristic (ROC) curves and areas under the ROC
39 121 curves (AUCs) were calculated to evaluate how accurately the CSS was predicted by different
40 122 models. Decision curve analysis (DCA) was performed to determine the clinical application of
41 123 different model: the proportion of true positive results minus the proportion of false positive
42 124 results, and then, the relative risks of false positive and false negative results were weighted to
43 125 obtain the net benefits of decision-making. All statistical analyses were performed using R
44 126 software (version 3.5.3). A two-tailed P value of less than 0.05 was considered statistically
45 127 significant.

46 128 ***Patient and Public Involvement***

47
48 129 The development of the research question and outcome measures were not informed by patients'
49 130 priorities, experience, and preferences. The patients were involved during the retrospective review
50 131 of public database where cases were diagnosed during 2010-2015. Patients were not involved in
51 132 the recruitment to and conduct of the study. The findings of the study will be disseminated by

133 online article to all study participants whose identity kept confidential during the whole research.

134 ***Ethics approval statement***

135 The Ethics committee(s) and IRB name: the Ethics Committee of Anhui Medical College; Reason
136 for exemption: The observational nature of the study

137 **Results**

138 Table 1 summarized the demographic and clinical feature of the participants. Six hundred and
139 twenty eight patients (60.50%) were diagnosed with a tumor less than 5cm. Six hundred and forty
140 patients (61.66%) were with grade III or IV. The numbers of patients with T1, T2, T3 and T4
141 respectively were 94, 125, 717 and 102. The numbers of patients with N1, N2 and N3 respectively
142 were 479, 330 and 229. Seventy five patients (7.23%) were with distant metastasis at presentation.
143 The median CSS was 27 months. The 1-, 2- and 5-year CSS rates were 76.8%, 53.0% and 29.2%,
144 respectively. There was no statistical difference of baseline characteristics between training set
145 and validating set. The detailed information of the two sets was also presented in Table 1.

Table 1. Baseline information of the included patients with node positive CGA, N(%).

Groups	Training set (N = 723)	Validating set (N = 315)	P-value
Sex			
Male	596 (82.43)	261 (82.86)	0.939
Female	127 (17.57)	54 (17.14)	
Age			
<70	490 (67.77)	210 (66.67)	0.781
≥70	233 (32.23)	105 (33.33)	
Race			
White	628 (86.86)	268 (85.08)	0.437
Black	40 (5.53)	24 (7.62)	
Others	55 (7.61)	23 (7.30)	
Tumor size			
<5cm	442 (61.13)	186 (59.05)	0.573
≥5cm	281 (38.87)	129 (40.95)	
Grade			
I-II	279 (38.59)	119 (37.78)	0.859
III-IV	444 (61.41)	196 (62.22)	
T stage			
T1a	17 (2.35)	4 (1.27)	0.224
T1b	53 (7.33)	20 (6.35)	
T2	83 (11.48)	42 (13.33)	
T3	501 (69.29)	216 (68.57)	
T4a	49 (6.78)	28 (8.89)	
T4b	20 (2.77)	5 (1.59)	
N stage			
N1	332 (45.92)	147 (46.67)	0.921

N2	229 (31.67)	101 (32.06)	
N3	162 (22.41)	67 (21.27)	
M stage			
M0	678 (93.78)	285 (90.48)	0.079
M1	45 (6.22)	30 (9.52)	
Low nodes yield			
Yes	532 (73.58)	243 (77.14)	0.300
No	191 (26.42)	72 (22.86)	
No. of nodes harvest	17 (12, 25)	16 (11, 24)	0.400
No. of positive nodes	3 (1, 6)	3 (1, 6)	1.000
Median survival (months)	28 (25, 32)	25 (21, 32)	0.361
CSS rate (%)			
1-year	77.0 (74.0, 80.2)	76.3 (71.6, 81.2)	
2-year	53.7 (50.1, 57.5)	51.4 (46.0, 57.5)	
5-year	30.3 (26.7, 34.5)	26.4 (20.9, 33.4)	

Abbreviation: CGA, cardia gastric adenocarcinoma; CSS, cancer-specific survival

146 According to X-tile software results, the optimal cut-off values for LNR were 0.09 and 0.33, and
 147 for LODDS were -2.09 and -0.65. Thus patients were separated into low (R1), medium (R2) or
 148 high LNR (R3) group, or low (L1), medium (L2) or high LODDS (L3) group. For model
 149 optimization, LNR and LODDS were also categorized into trichotomous factors using cut-off
 150 values of P_{25} and P_{75} . The discrimination ability of the model based on interquartile was lower
 151 (Suppl. Table 1), so this model was not further analyzed. Next we illustrated the survival curves of
 152 the patients according to N, LNR or LODDS staging system. As shown in Figure 2 training set
 153 section, CSS was significantly different by all the three staging systems (all the log-rank P values
 154 < 0.0001); however the 95% CIs of N2 and N3 survival curve initially separated and partly
 155 overlapped afterwards. The inferior discriminative ability of N system was further supported by
 156 AIC and C-index. As shown in Table 2, the C-index of N stage was lower than that of LNR or
 157 LODDS. Similarly, the AIC of N stage was higher than that of LNR or LODDS. The clinical
 158 characteristics with statistical significance for CSS were further incorporated in the Cox regression
 159 model as potential confounders (Suppl. Table 2), and all the variables met proportional hazard
 160 assumption (Suppl. Figure 1). The prognostic value of adjusted model was better than crude mode
 161 generally. In addition, the value of LNR system seemed to be worse than LODDS system;
 162 however the difference was not noticeable, so we considered both of the systems into nomogram
 163 construction.

Table 2. Prognostic values of variables for patients with node positive CGA (N = 1038).

Variables	Crude model			Adjusted model		
	HR (95% CI)	C-index	AIC	HR (95% CI)	C-index	AIC
Training set (N = 723)						
N stage		0.582	5403		0.632	5365
N1	1 (ref)			1 (ref)		
N2	1.53 (1.24, 1.91)			1.42 (1.14, 1.77)		

N3	2.15 (1.70, 2.71)			2.03 (1.60, 2.59)		
LNR*		0.607	5376		0.643	5350
R1	1 (ref)			1 (ref)		
R2	1.88 (1.44, 2.44)			1.74 (1.33, 2.29)		
R3	3.02 (2.30, 3.97)			2.63 (1.97, 3.50)		
LODDS*		0.609	5373		0.644	5346
L1	1 (ref)			1 (ref)		
L2	1.93 (1.48, 2.51)			1.80 (1.36, 2.37)		
L3	3.13 (2.38, 4.13)			2.77 (2.07, 3.70)		
Validating set (N = 315)						
N stage		0.596	1957		0.675	1931
N1	1 (ref)			1 (ref)		
N2	1.81 (1.31, 2.51)			1.75 (1.25, 2.46)		
N3	2.18 (1.51, 3.15)			2.23 (1.50, 3.30)		
LNR*		0.646	1927		0.691	1913
R1	1 (ref)			1 (ref)		
R2	2.20 (1.47, 3.30)			1.91 (1.26, 2.90)		
R3	4.16 (2.76, 6.28)			3.58 (2.30, 5.56)		
LODDS*		0.647	1927		0.789	1914
L1	1 (ref)			1 (ref)		
L2	2.07 (1.39, 3.09)			2.08 (1.38, 3.14)		
L3	4.22 (2.79, 6.39)			4.10 (2.65, 6.34)		

Abbreviations: CGA, cardia gastric adenocarcinoma; HR, hazard ratio; CI, confidence interval; AIC, Akaike information criterion; LNR, lymph node ratio; LODDS, log odds of positive lymph nodes.

Adjusted model considered race, tumor size, grade, T stage and M stage.

* cut-off values for LNR were 0.09 and 0.33, and for LODDS were -2.09 and -0.65.

164 Stepwise Cox regression analysis showed race, tumor grade, low nodes yield, T stage, M stage
 165 and LNR (or LODDS) were independent predictors, so these factors were included in nomograms.
 166 For both LNR and LODDS, the total score was 40, and higher score suggested lower survival
 167 (Figure 3 and Suppl. Figure 2). Next calibration plot was used to assess the internal and external
 168 validity of the nomogram (Figure 3 and Suppl. Figure 2). Since the cross-spot line was generally
 169 close to the grey reference line, we concluded the predicted CSS was well correlated with the
 170 actual situation. In addition, ROC curves indicated that the AUC of the model based on N stage
 171 was lower than that of the model based on the nomogram of LNR or LODDS (Suppl. Figure 3).
 172 DCA plot also showed that the nomogram model was superior to traditional model (Suppl. Figure
 173 3).

174 Discussion

175 The present study analyzes national cancer registry databases and demonstrates that survival of
 176 patients with node-positive CGA is well predicted when the traditional N staging method is
 177 substituted with LNR or LODDS system. This finding both exists in training and validating sets.
 178 In training set, the survival curves separate clearly when patient grouping is implemented by LNR
 179 or LODDS method, which is not achieved by traditional N staging system. Adjusted model that

1
2
3 180 simultaneously considers staging, clinical and demographic features outperforms crude model that
4 181 only takes staging into account. Therefore multiple independent survival factors are incorporated
5 182 in nomogram construction, which suggests older age at diagnosis, white, higher grade, greater
6 183 tumor infiltration, higher proportion of positive LN, and metastasis as risk factors. The
7 184 nomograms perform steadily in 1-, 2- and 5-year CSS prediction as the validation plots show.

10 185 Previous studies have demonstrated the superiority of LNR or LODDS for prognostic prediction
11 186 in GC after surgical resection (8-10, 15-17). However the GC patients are not further separated
12 187 and investigated according to primary tumor site, since there is much difference between cardia
13 188 and non-cardia GC in terms of tumor features, etiological factors, and biological behaviors (3). In
14 189 AJCC cancer staging 7th edition, tumors involving EGJ was categorized as esophagus cancer (5),
15 190 which was however argued by the viewpoint that GC staging system has a better ability to predict
16 191 survival of EGJ tumor (18, 19). In the latest 8th edition (20), a tumor that has its epicenter within 2
17 192 cm of EGJ and involves the EGJ (Siewert type I/II) is classified as esophageal cancer. Other
18 193 situation, including a tumor with epicenter more than 2 cm from EGJ or a tumor located with 2 cm
19 194 of EGJ but does not involve EGJ, is classified as stomach cancer. The superiority of the new
20 195 system is confirmed by a retrospective observational study from two high-volume institutions in
21 196 China, regardless of Siewert type (21). In terms of Siewert type II junctional adenocarcinoma, a
22 197 marginal advantage of the esophagus cancer system is found in discriminating survival rates after
23 198 3 and 5 years, however the advantage of GC system lies in division of the N3 category into N3a
24 199 and N3b, so the authors concludes neither the esophageal nor the stomach staging system is
25 200 flawless in predicting survival in Siewert type II junctional cancer (22). Above all, CGA is
26 201 probably a special entity that has a different biological property compared with genuine gastric
27 202 and genuine esophageal cancer. To the best of our knowledge, the present study first reveals a
28 203 superior performance of prognostic prediction based on LNR or LODDS in node positive CGA
29 204 patients. Unfortunately we are unable to consider Siewert type due to unavailable information
30 205 from SEER database; therefore we encourage further studies to pay special attention on tumor
31 206 location.

32
33
34
35
36
37
38
39
40 207 LNR and LODDS have been proved to be the strongest indicators of survival in gastric
41 208 adenocarcinoma when LN harvest is inadequate (16, 17). It is demonstrated that in general, more
42 209 LN resection is associated with better survival, which may be the result of either improved N
43 210 classification or a therapeutic effect of lymphadenectomy. For esophageal cancer, worldwide data
44 211 shows that yielding 10 nodes for pT1, 20 for pT2, and 30 or more for pT3/T4 is recommended for
45 212 maximum 5-year survival (23). For GC, greater LN harvest also shows improved survival (24). It
46 213 is suggested that at least 16 nodes be assessed pathologically and evaluation of more than 30
47 214 nodes is desirable (25). Overall it is encouraged to harvest as many LN as possible, balancing the
48 215 extent of LN resection necessary for accurate N staging and maximum survival without
49 216 unnecessarily increasing the morbidity of radical lymphadenectomy. Nevertheless, many
50 217 conditions would lead to insufficient LN harvest. It is estimated that only one fifth GC patients
51 218 have sufficient LN examined in Iran (26), while more than 15 LNs are examined in 64% of
52 219 patients in the US (25). The LNR and LODDS staging methods do not require adequate number of
53 220 LN assessment. In the present study, low nodes yield is a risk factor for poor survival in univariate
54 221 analysis; however it loses significance in LNR or LODDS based multivariate model, which
55 222 indicates that it probably exerts little impact with consideration of LNR or LODDS. In fact, the

223 new node category method is stable when nodal assessment is insufficient during surgery not only
224 for GC (8, 15-17) but also for colorectal cancer (27), esophageal cancer (28), oral squamous cell
225 carcinoma (29), gallbladder cancer (30), etc.

226 The association between LNR and survival is an exciting aspect of cardia GC that is currently
227 emerging and may be clinically meaningful. The higher ratio of positive LN indicates worse
228 outcome in cardia GC. Patients are at 2-3 folds higher risk of cancer specific death if the ratio is
229 over 33%. The ratio of 9-33% also indicates a double risk. This effect is independent of other
230 crucial clinical characteristics, thus providing a useful tool for surgeons to predict the prognosis,
231 and to be taken as evidence for the surgeon to tend towards truly radical, i.e., complete lymph
232 node clearance rather than limited clearance (31). In addition, LNR minimizes the “stage
233 migration” phenomenon that can be observed using the current N staging system (32).

234 One limitation of this study is that some important factors that are associated with survival are not
235 considered in the model due to unavailable data source. For example, ECOG/KPS score is
236 commonly taken into account in survival analysis due to its remarkable relationship with general
237 status and prognosis. Unfortunately the SEER 18 database does not record the score at diagnosis,
238 so the impact of it is not considered in this analysis. Treatment mode is also associated with
239 clinical outcome. This study enrolled patients who received gastric resection; however other
240 information about chemo- or radiotherapy is not available in SEER 18 database. Randomized
241 clinical trial demonstrates that compared with surgery alone, preoperative administration of
242 carboplatin and paclitaxel with concurrent radiotherapy significantly improved overall survival
243 among patients with esophageal or GEJ cancer (HR = 0.657) (33). The NCCN clinical practice
244 guidelines for GEJ cancer recommend preoperative chemoradiation or perioperative
245 chemotherapy due to substantial survival benefit compared with surgery alone (34). To overcome
246 this limitation, a database that provides with fully detailed medical records is needed for analysis.
247 In addition, consideration of the potential factors aforementioned would greatly improve
248 prognostic power of survival prediction model. Another limitation is that our results are based on
249 training set and confirmed by validating set; however the features of the two groups are similar.
250 So this finding needs to be proved among populations with distinct features.

251 In conclusion, LNR and LODDS staging methods have better prognostic efficacy than traditional
252 N staging method in CGA patients with regional node metastasis. Besides, the two values are
253 promising substitute for N staging in nomogram development when other independent prognostic
254 factors are incorporated.

- 255 **a. Contributorship statement:** WXQ work conception, data interpretation, critical review for
256 important content, final approval of the manuscript and agreement to be accountable for all
257 aspects of the work; BM administrative work, funding, critical review for important content,
258 final approval of the manuscript and agreement to be accountable for all aspects of the work;
259 ZC acquisition, analysis of data, drafting the work, final approval of the manuscript and
260 agreement to be accountable for all aspects of the work.
- 261 **b. Competing interests:** None.
- 262 **c. Funding:** This work was supported by [Major Project of Natural Science of Anhui Provincial
263 Department of Education] grant number [KJ2019ZD73] and [Domestic Visiting Study Project
264 for Outstanding Young Talents in Colleges and Universities in Anhui Province] grant number

265 [2021].
 266 **d. Data sharing statement:** Dataset available from the SEER website
 267 (<https://seer.cancer.gov/data-software/>).
 268

269 **Acknowledgement:** We thank all patient advisers for their special contribution.
 270

271 References:

- 272 1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018:
 273 GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA*
 274 *Cancer J Clin.* [Journal Article]. 2018 2018-11-01;68(6):394-424.
 275 2. Abdi E, Latifi-Navid S, Zahri S, Yazdanbod A, Pourfarzi F. Risk factors predisposing to cardia
 276 gastric adenocarcinoma: Insights and new perspectives. *Cancer Med.* [Journal Article; Research
 277 Support, Non-U.S. Gov't; Review]. 2019 2019-10-01;8(13):6114-26.
 278 3. Amini N, Spolverato G, Kim Y, Squires MH, Poultsides GA, Fields R, et al. Clinicopathological
 279 features and prognosis of gastric cardia adenocarcinoma: a multi-institutional US study. *J SURG*
 280 *ONCOL.* [Comparative Study; Journal Article; Multicenter Study]. 2015 2015-03-01;111(3):285-92.
 281 4. Sobin L, Gospodarowicz M, Wittekind C. TNM classification of malignant tumours. 7th edition.:
 282 Wiley-Blackwell; 2009.
 283 5. Edge S, Byrd D, Compton C, Fritz A, Greene F, Trotti A. AJCC cancer staging manual. 7th
 284 edition.: Springer; 2009.
 285 6. Hou Y, Wang X, Chen J. Prognostic significance of metastatic lymph node ratio: the lymph node
 286 ratio could be a prognostic indicator for patients with gastric cancer. *WORLD J SURG ONCOL.*
 287 [Journal Article]. 2018 2018-10-04;16(1):198.
 288 7. Zhu J, Xue Z, Zhang S, Guo X, Zhai L, Shang S, et al. Integrated analysis of the prognostic role
 289 of the lymph node ratio in node-positive gastric cancer: A meta-analysis. *INT J SURG.* [Journal Article;
 290 Meta-Analysis; Review]. 2018 2018-09-01;57:76-83.
 291 8. Sun Z, Xu Y, Li DM, Wang ZN, Zhu GL, Huang BJ, et al. Log odds of positive lymph nodes: a
 292 novel prognostic indicator superior to the number-based and the ratio-based N category for gastric
 293 cancer patients with R0 resection. *CANCER-AM CANCER SOC.* [Journal Article; Research Support,
 294 Non-U.S. Gov't]. 2010 2010-06-01;116(11):2571-80.
 295 9. Zhao E, Zhou C, Chen S. Prognostic nomogram based on log odds of positive lymph nodes for
 296 gastric carcinoma patients after surgical resection. *FUTURE ONCOL.* [Journal Article]. 2019
 297 2019-12-01;15(36):4207-22.
 298 10. Gu P, Deng J, Sun Z, Wang Z, Wang W, Liang H, et al. Superiority of log odds of positive lymph
 299 nodes (LODDS) for prognostic prediction after gastric cancer surgery: a multi-institutional analysis of
 300 7620 patients in China. *SURG TODAY.* [Journal Article]. 2021 2021-01-01;51(1):101-10.
 301 11. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov)
 302 SEER*Stat Database: Incidence - SEER Research Data, 18 Registries, Nov 2019 Sub (2000-2017) -
 303 Linked To County Attributes - Time Dependent (1990-2017) Income/Rurality, 1969-2018 Counties,
 304 National Cancer Institute, DCCPS, Surveillance Research Program, released April 2020, based on the
 305 November 2019 submission..
 306 12. Yu Y, Zhang P, Yao R, Wang J, Wang P, Xue X, et al. Prognostic value of log odds of positive
 307 lymph nodes in node-positive lung squamous cell carcinoma patients after surgery: a SEER

- 1
2
3 308 population-based study. *Transl Lung Cancer Res.* [Journal Article]. 2020 2020-08-01;9(4):1285-301.
- 4 309 13. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker
5 310 assessment and outcome-based cut-point optimization. *CLIN CANCER RES.* [Journal Article;
6 311 Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, Non-P.H.S.; Research Support, U.S.
7 312 Gov't, P.H.S.]. 2004 2004-11-01;10(21):7252-9.
- 8 313 14. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer
9 314 prognosis. *J CLIN ONCOL.* [Journal Article; Review]. 2008 2008-03-10;26(8):1364-70.
- 10 315 15. Jian-Hui C, Shi-Rong C, Hui W, Si-le C, Jian-Bo X, Er-Tao Z, et al. Prognostic value of three
11 316 different lymph node staging systems in the survival of patients with gastric cancer following D2
12 317 lymphadenectomy. *Tumour Biol.* [Journal Article]. 2016 2016-08-01;37(8):11105-13.
- 13 318 16. Aurello P, Petrucciani N, Nigri GR, La Torre M, Magistri P, Tierno S, et al. Log odds of positive
14 319 lymph nodes (LODDS): what are their role in the prognostic assessment of gastric adenocarcinoma? *J*
15 320 *GASTROINTEST SURG.* [Comparative Study; Journal Article]. 2014 2014-07-01;18(7):1254-60.
- 16 321 17. Kilic MO, Gundogdu SB, Ozden S, Saylam B, Tez M. The prognostic value of different node
17 322 staging systems in patients with ≤ 15 lymph nodes following surgery for gastric adenocarcinoma.
18 323 *ACTA CHIR BELG.* [Journal Article]. 2018 2018-02-01;118(1):1-6.
- 19 324 18. Suh YS, Han DS, Kong SH, Lee HJ, Kim YT, Kim WH, et al. Should adenocarcinoma of the
20 325 esophagogastric junction be classified as esophageal cancer? A comparative analysis according to the
21 326 seventh AJCC TNM classification. *ANN SURG.* [Comparative Study; Journal Article; Research
22 327 Support, Non-U.S. Gov't]. 2012 2012-05-01;255(5):908-15.
- 23 328 19. Huang Q, Shi J, Feng A, Fan X, Zhang L, Mashimo H, et al. Gastric cardiac carcinomas involving
24 329 the esophagus are more adequately staged as gastric cancers by the 7th edition of the American Joint
25 330 Commission on Cancer Staging System. *Mod Pathol.* [Journal Article; Research Support, Non-U.S.
26 331 Gov't; Validation Study]. 2011 2011-01-01;24(1):138-46.
- 27 332 20. Amin M, Edge S, Greene F, Byrd D, Brookland R. *AJCC cancer staging manual.* 8th edition.:
28 333 Springer; 2016.
- 29 334 21. Liu K, Feng F, Chen XZ, Zhou XY, Zhang JY, Chen XL, et al. Comparison between gastric and
30 335 esophageal classification system among adenocarcinomas of esophagogastric junction according to
31 336 AJCC 8th edition: a retrospective observational study from two high-volume institutions in China.
32 337 *GASTRIC CANCER.* [Comparative Study; Journal Article; Observational Study]. 2019
33 338 2019-05-01;22(3):506-17.
- 34 339 22. Karstens KF, Ghadban T, Sawez S, Konczalla L, Woestemeier A, Bachmann K, et al. Comparison
35 340 of the 8th UICC staging system for esophageal and gastric cancers in Siewert type II junctional
36 341 adenocarcinomas. *Eur J Surg Oncol.* [Comparative Study; Journal Article]. 2020 2020-04-01;46(4 Pt
37 342 A):638-43.
- 38 343 23. Rizk NP, Ishwaran H, Rice TW, Chen LQ, Schipper PH, Kesler KA, et al. Optimum
39 344 lymphadenectomy for esophageal cancer. *ANN SURG.* [Journal Article; Research Support, Non-U.S.
40 345 Gov't]. 2010 2010-01-01;251(1):46-50.
- 41 346 24. Macalindong SS, Kim KH, Nam BH, Ryu KW, Kubo N, Kim JY, et al. Effect of total number of
42 347 harvested lymph nodes on survival outcomes after curative resection for gastric adenocarcinoma:
43 348 findings from an eastern high-volume gastric cancer center. *BMC CANCER.* [Journal Article;
44 349 Research Support, Non-U.S. Gov't]. 2018 2018-01-12;18(1):73.
- 45 350 25. Schoenleber SJ, Schnelldorfer T, Wood CM, Qin R, Sarr MG, Donohue JH. Factors influencing
46 351 lymph node recovery from the operative specimen after gastrectomy for gastric adenocarcinoma. *J*

- 1
2
3 352 GASTROINTEST SURG. [Journal Article]. 2009 2009-07-01;13(7):1233-7.
4 353 26. Khanjani N, Mirzaei S, Nasrolahi H, Hamed SH, Mosalaei A, Omidvari S, et al. Insufficient
5 354 lymph node assessment in gastric adenocarcinoma. J Egypt Natl Canc Inst. [Journal Article]. 2019
6 355 2019-10-22;31(1):2.
7 356 27. Pei JP, Zhang CD, Fan YC, Dai DQ. Comparison of Different Lymph Node Staging Systems in
8 357 Patients With Resectable Colorectal Cancer. FRONT ONCOL. [Journal Article]. 2018
9 358 2018-01-20;8:671.
10 359 28. Cao J, Yuan P, Ma H, Ye P, Wang Y, Yuan X, et al. Log Odds of Positive Lymph Nodes Predicts
11 360 Survival in Patients After Resection for Esophageal Cancer. ANN THORAC SURG. [Journal Article;
12 361 Multicenter Study]. 2016 2016-08-01;102(2):424-32.
13 362 29. Jin W, Zhu Z, Wu Y, Ding X, Wu H, Song X, et al. Prognostic value of log odds of positive
14 363 lymph nodes in patients with resectable oral squamous cell carcinoma. ORAL ONCOL. [Journal
15 364 Article; Research Support, Non-U.S. Gov't]. 2020 2020-09-01;108:104709.
16 365 30. Li J, Sun Y, Zhao B, Tang C, Fan D, Jiang W, et al. Lymph Node Ratio-Based Staging System for
17 366 Gallbladder Cancer With Fewer Than Six Lymph Nodes Examined. FRONT ONCOL. [Journal
18 367 Article]. 2020 2020-01-20;10:542005.
19 368 31. Spillane AJ, Cheung BL, Winstanley J, Thompson JF. Lymph node ratio provides prognostic
20 369 information in addition to american joint committee on cancer N stage in patients with melanoma, even
21 370 if quality of surgery is standardized. ANN SURG. [Journal Article]. 2011 2011-01-01;253(1):109-15.
22 371 32. Rosa F, Tortorelli AP, Alfieri S, Doglietto GB. Lymph node ratio for gastric cancer: useful
23 372 instrument or just an expedient to retrieve fewer lymph nodes? ANN SURG. [Letter; Comment]. 2014
24 373 2014-04-01;259(4):e65.
25 374 33. van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge HM, Wijnhoven BP, et al.
26 375 Preoperative chemoradiotherapy for esophageal or junctional cancer. N Engl J Med. [Journal Article;
27 376 Multicenter Study; Randomized Controlled Trial; Research Support, Non-U.S. Gov't]. 2012
28 377 2012-05-31;366(22):2074-84.
29 378 34. Ajani JA, D'Amico TA, Bentrem DJ, Chao J, Corvera C, Das P, et al. Esophageal and
30 379 Esophagogastric Junction Cancers, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. J
31 380 Natl Compr Canc Netw. [Journal Article]. 2019 2019-07-01;17(7):855-83.

381 382 **Illustrations**

383 Figure 1. Flow diagram of patient selection and grouping.

384 Figure 2. Survival curves of training and validating sets by different staging systems.

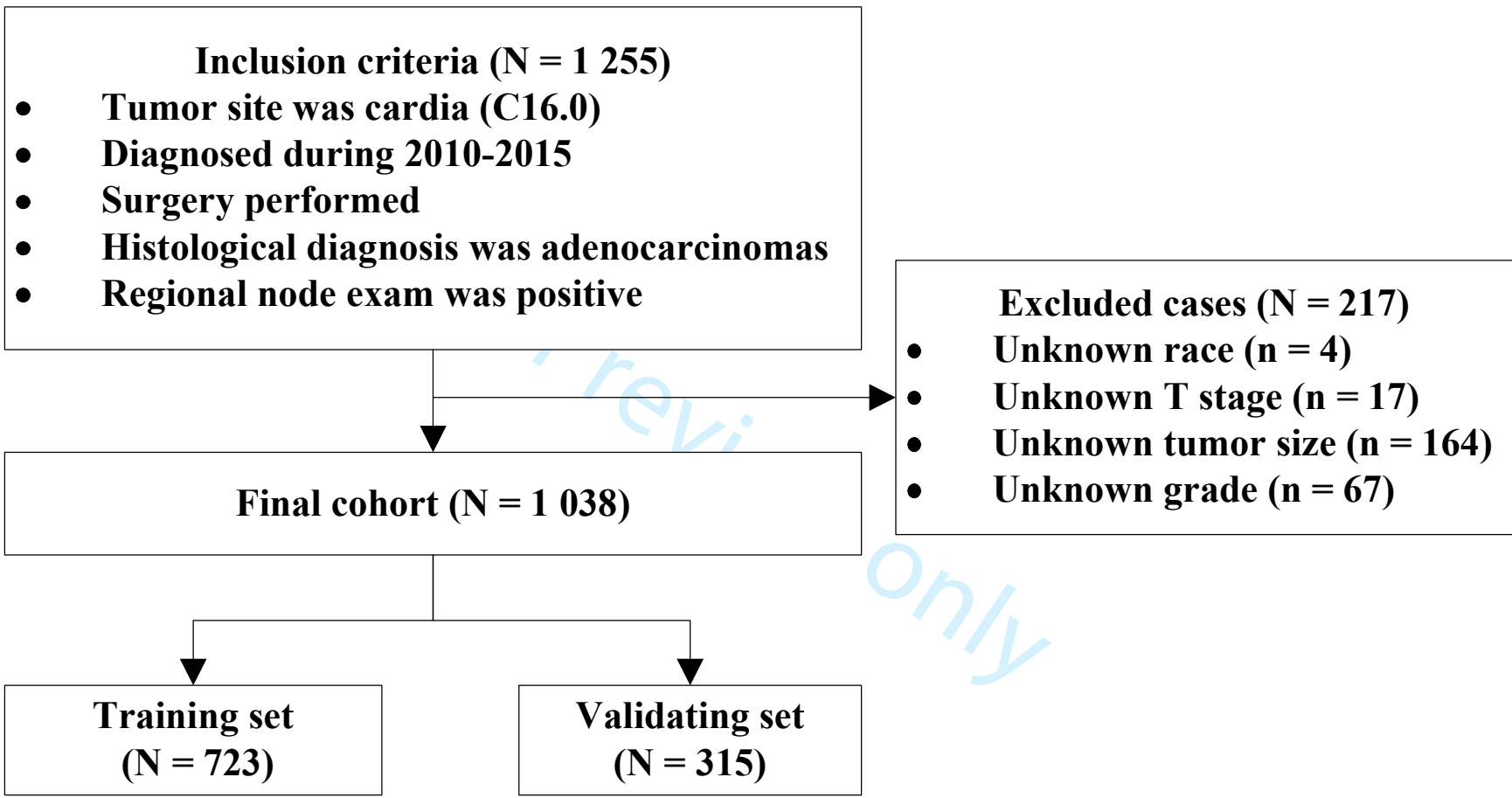
385 Figure 3. Construction of nomogram based on Tumor-Lymph node ratio-Metastasis staging system
386 and calibration plots for the nomogram.

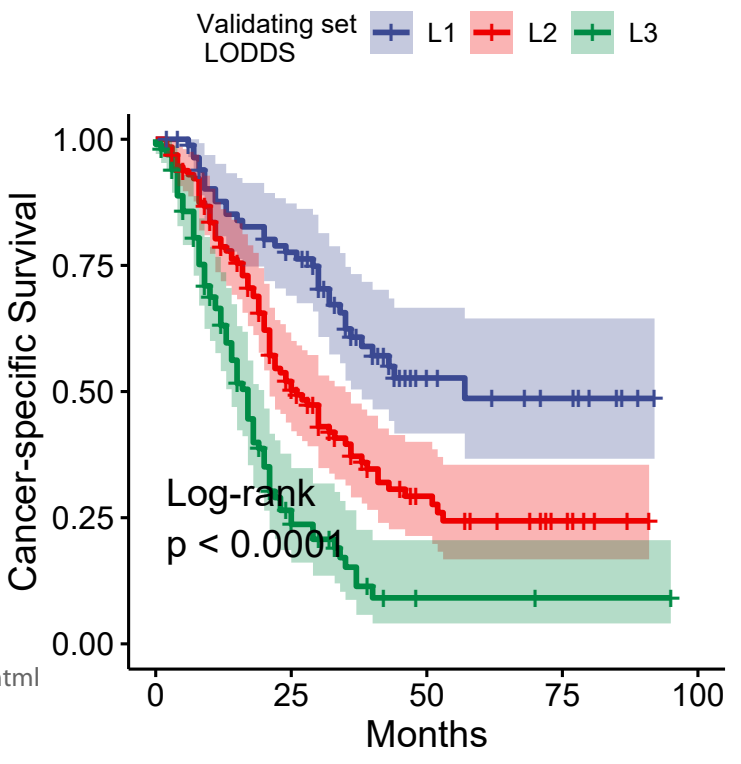
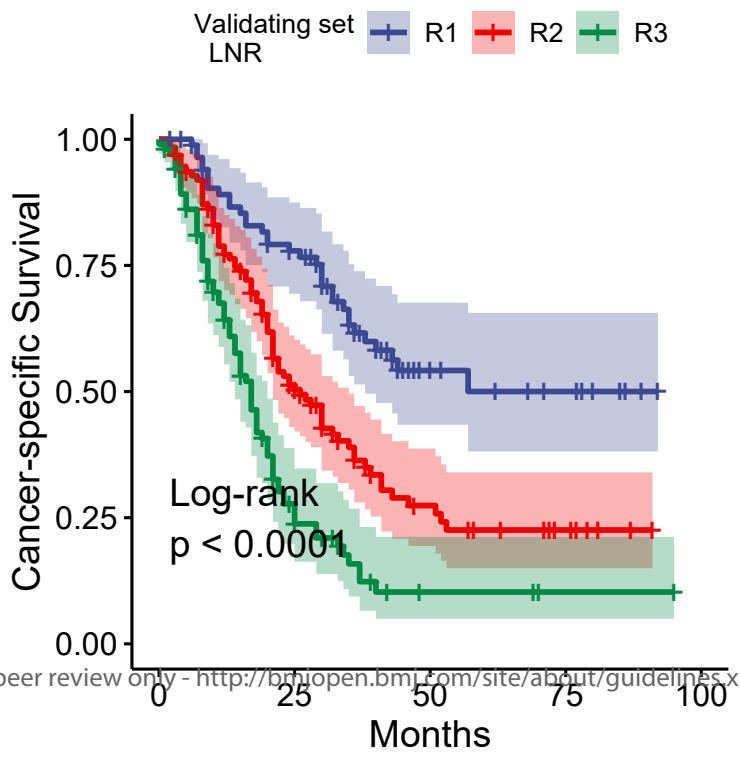
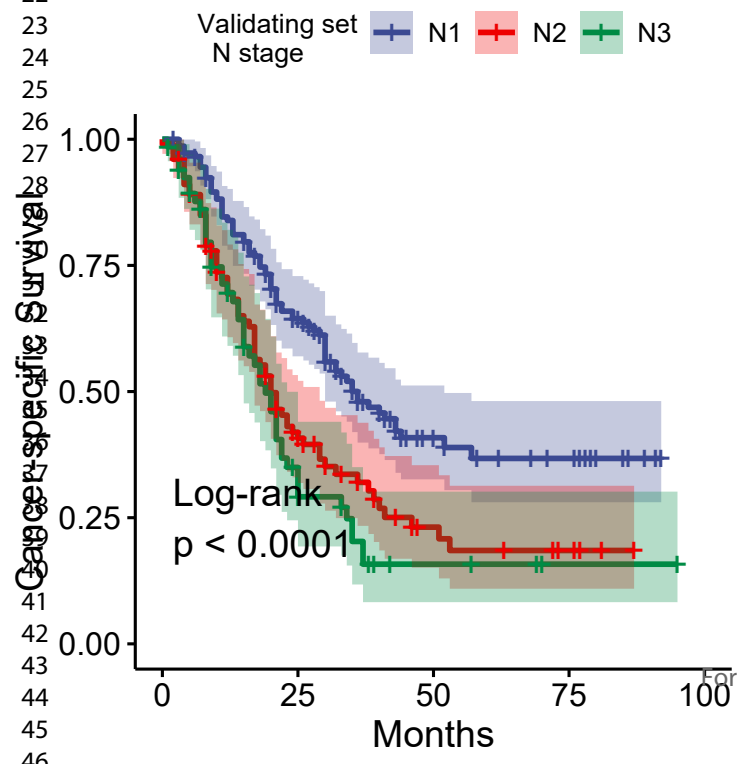
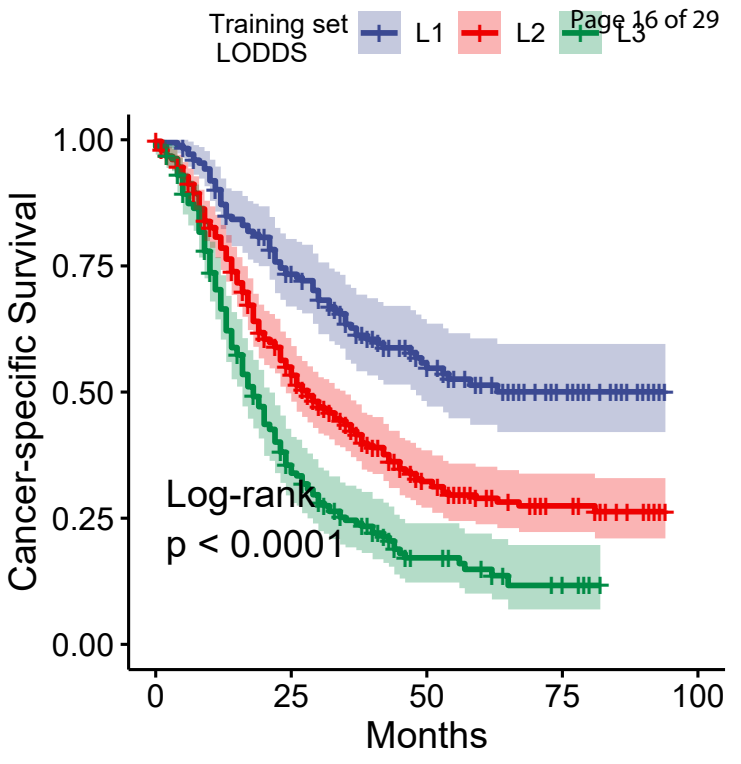
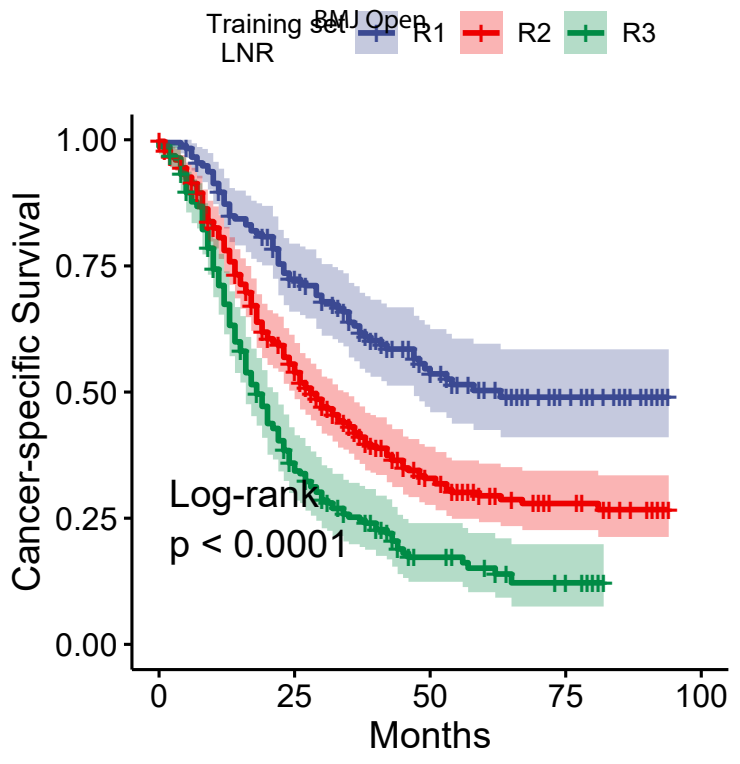
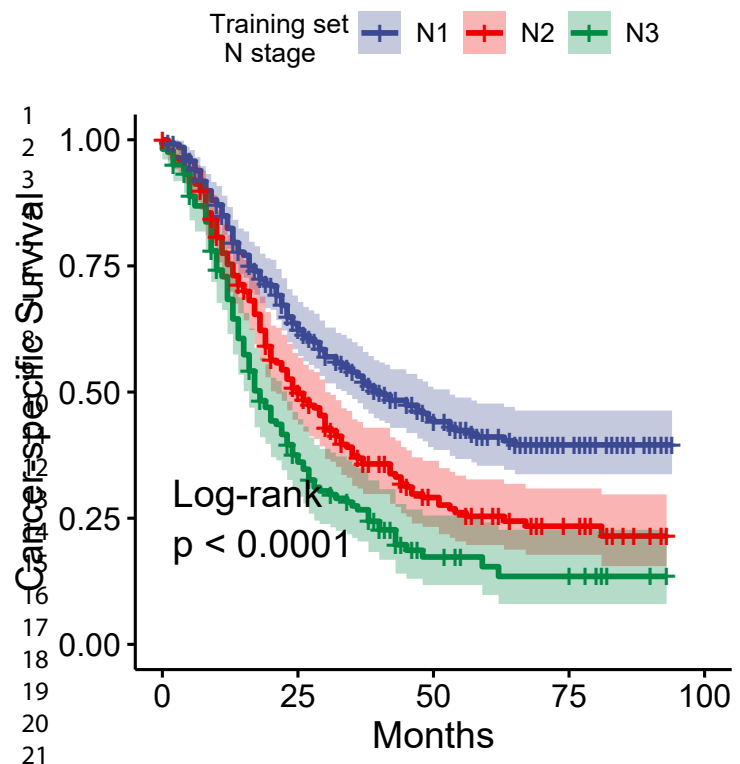
387
388 Suppl. Figure 1. Plots of Schoenfeld, Martingale, and Deviance residuals for proportional hazard
389 assumption test in models that incorporate N stage, lymph node ratio and log odds of positive
390 lymph nodes.

391 Suppl. Figure 2. Construction of nomogram based on Tumor-Log odds of positive lymph nodes
392 -Metastasis staging system and calibration plots for the nomogram.

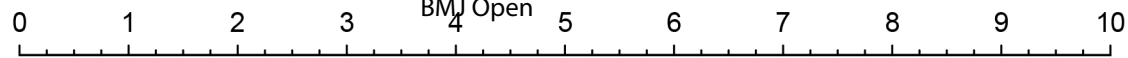
393 Suppl. Figure 3. Receiver operator characteristic (ROC) curves and decision curve analysis (DCA)
394 plots for comparison of the prediction powers of the different models.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46

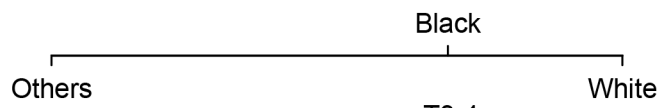




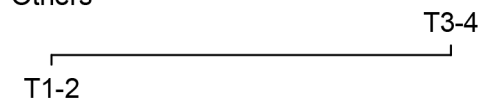
Points



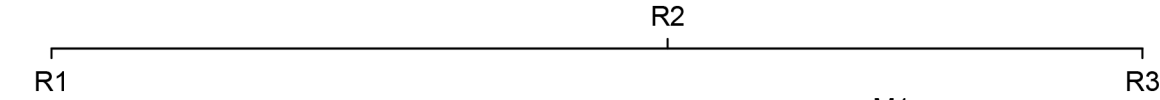
Race



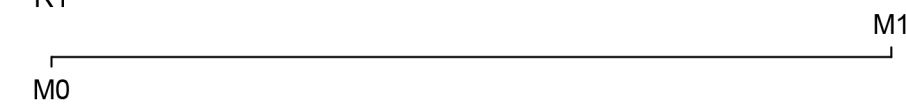
T.stage



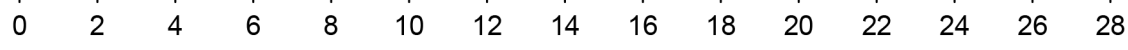
LNR



M.stage



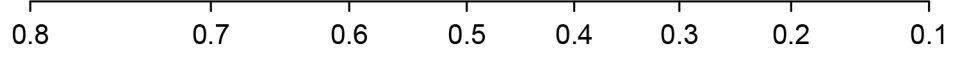
Total Points



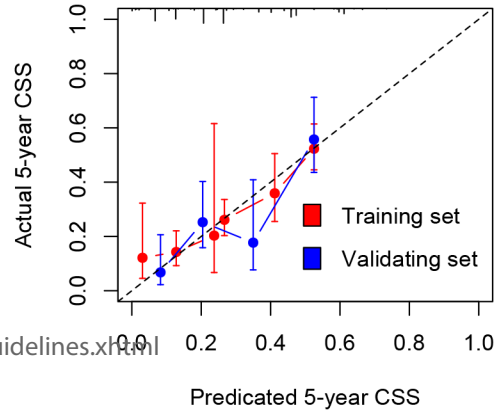
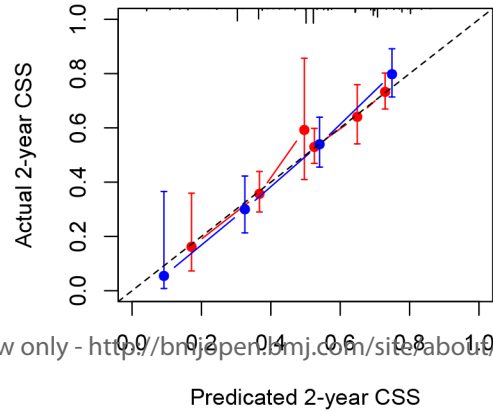
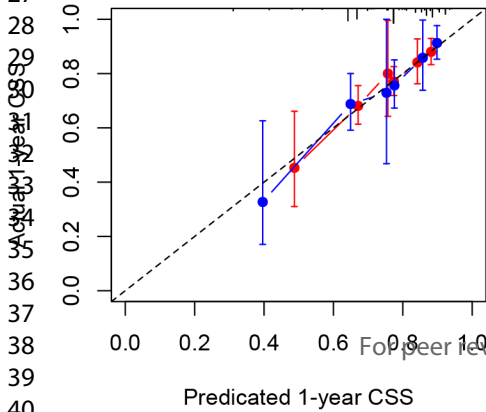
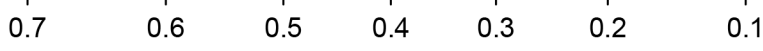
1-year CSS

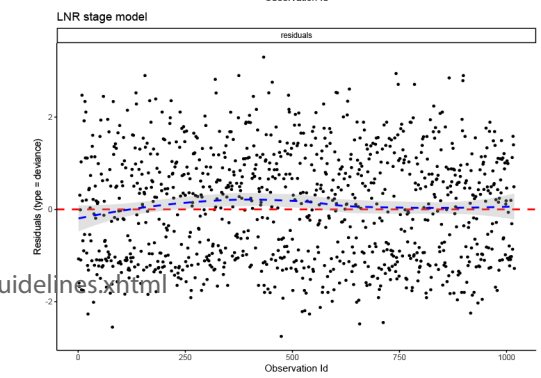
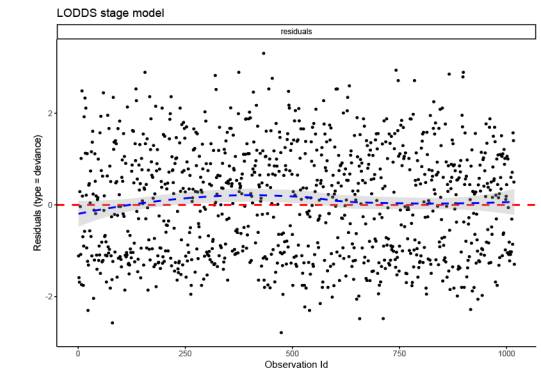
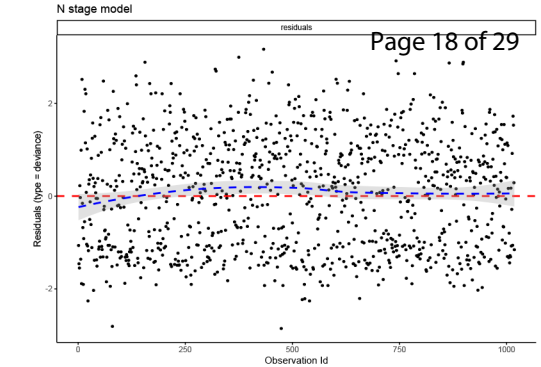
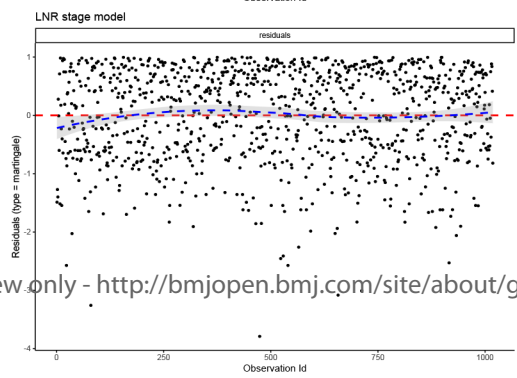
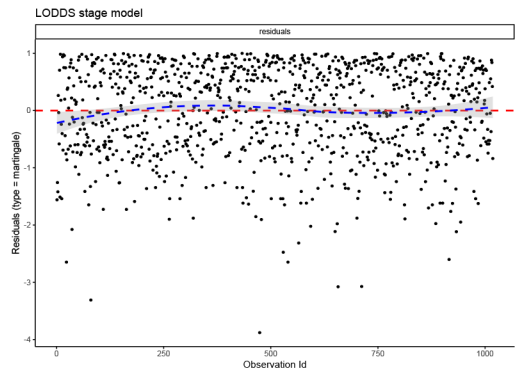
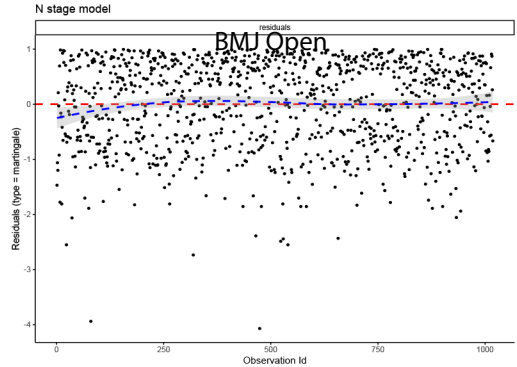
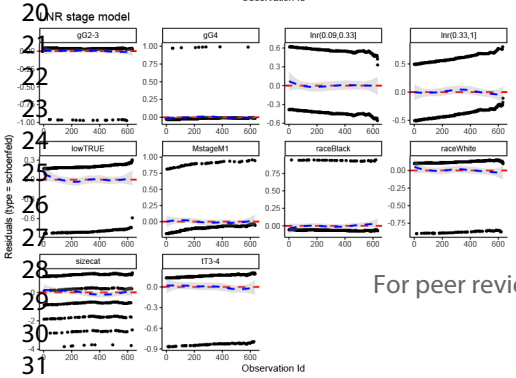
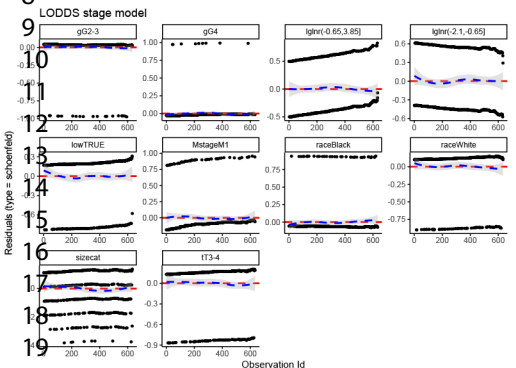
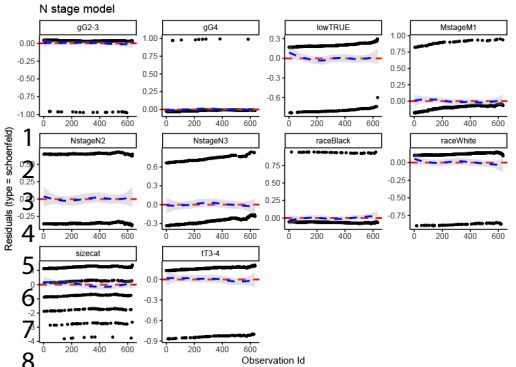


2-year CSS

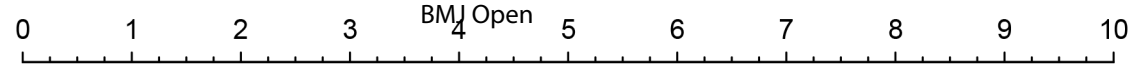


5-year CSS

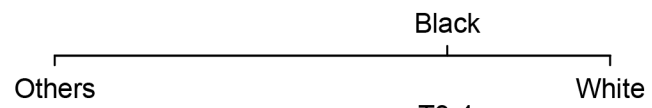




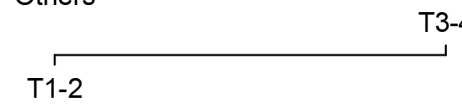
Points



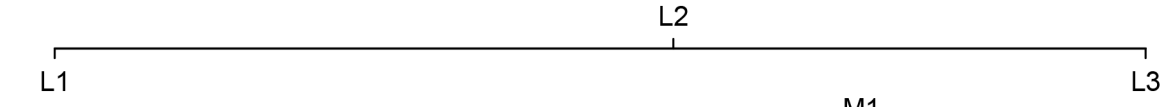
Race



T.stage



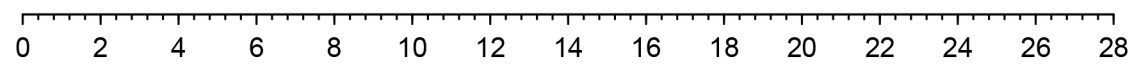
LODDS



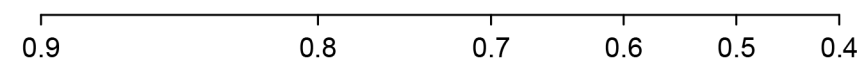
M.stage

M0

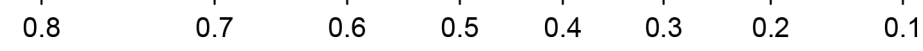
Total Points



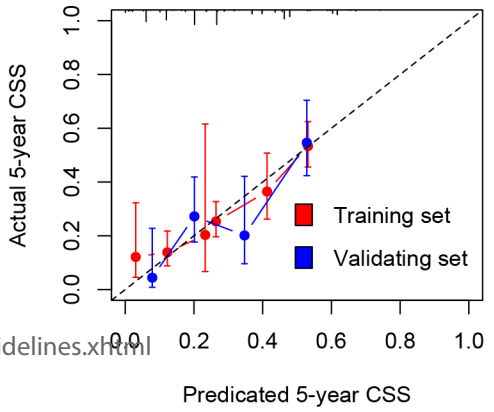
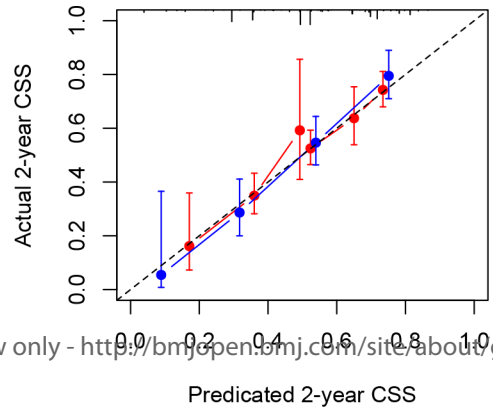
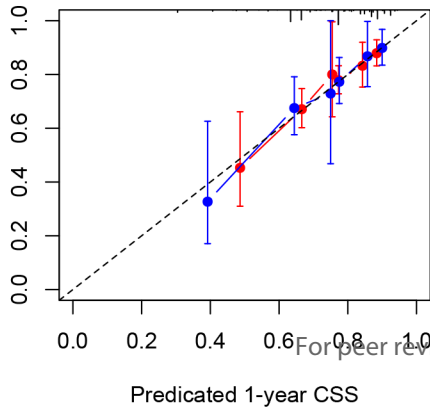
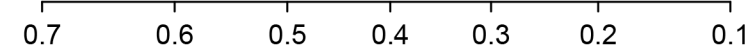
1-year CSS



2-year CSS

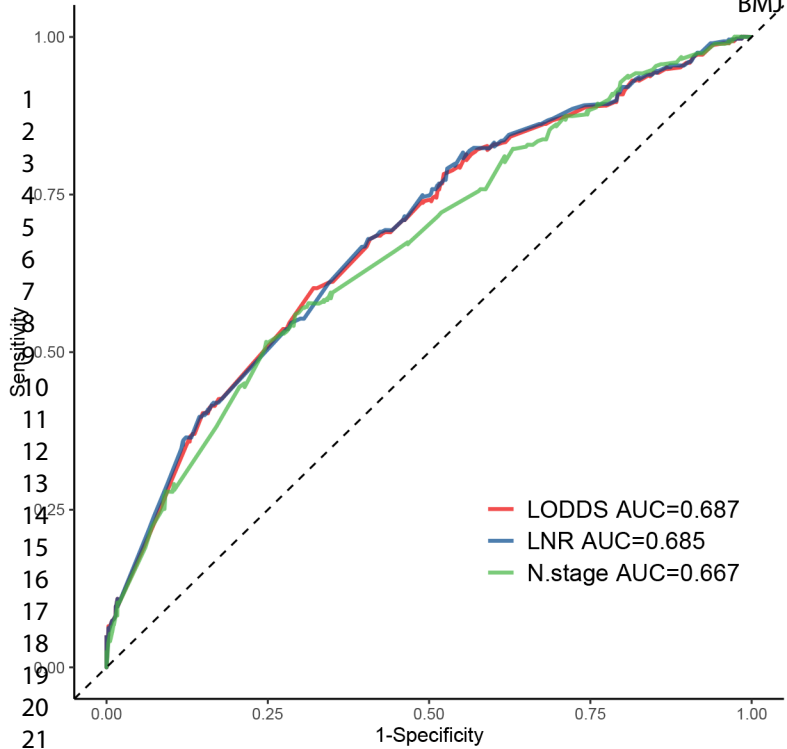


5-year CSS

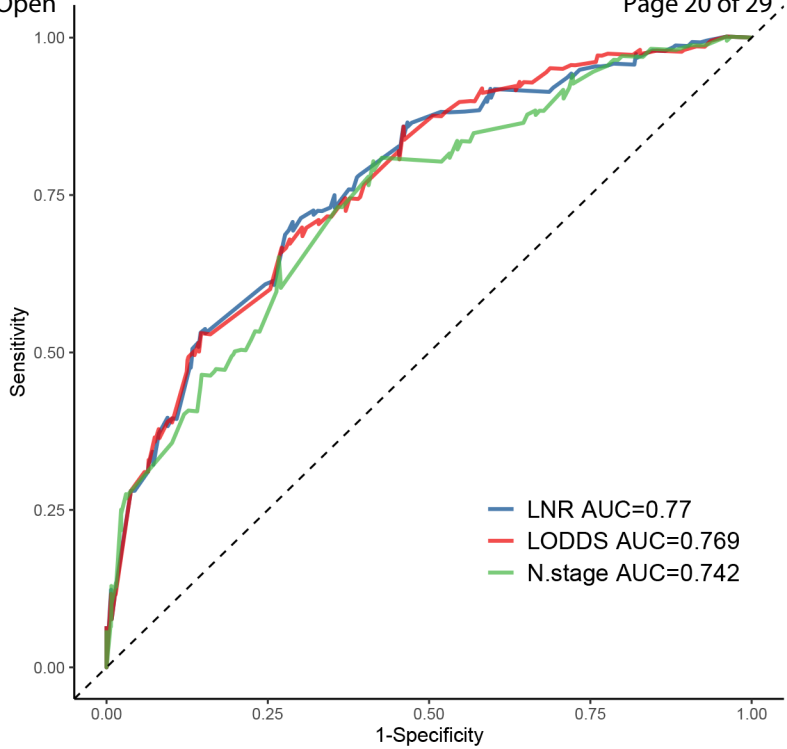


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

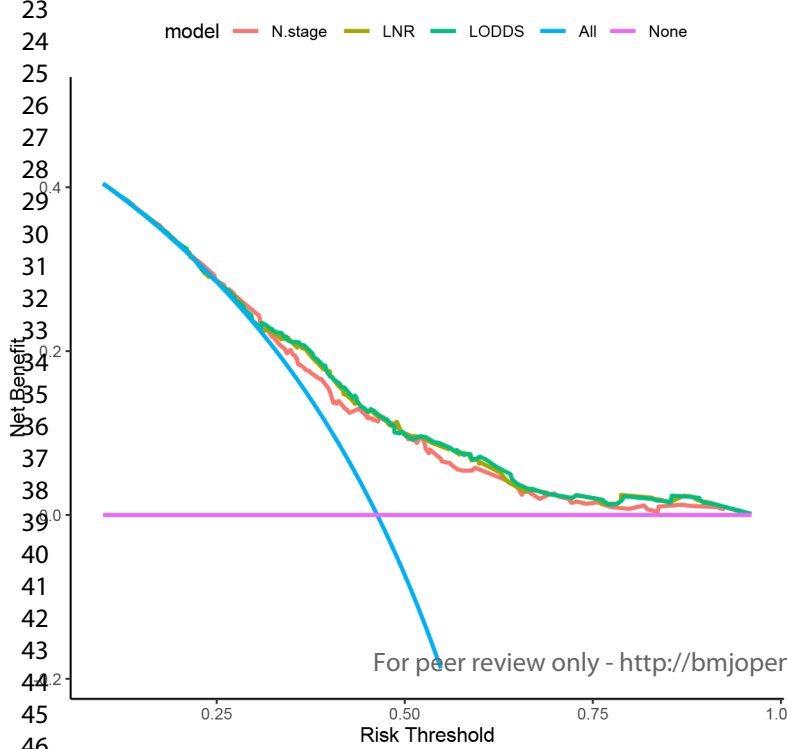
Training set



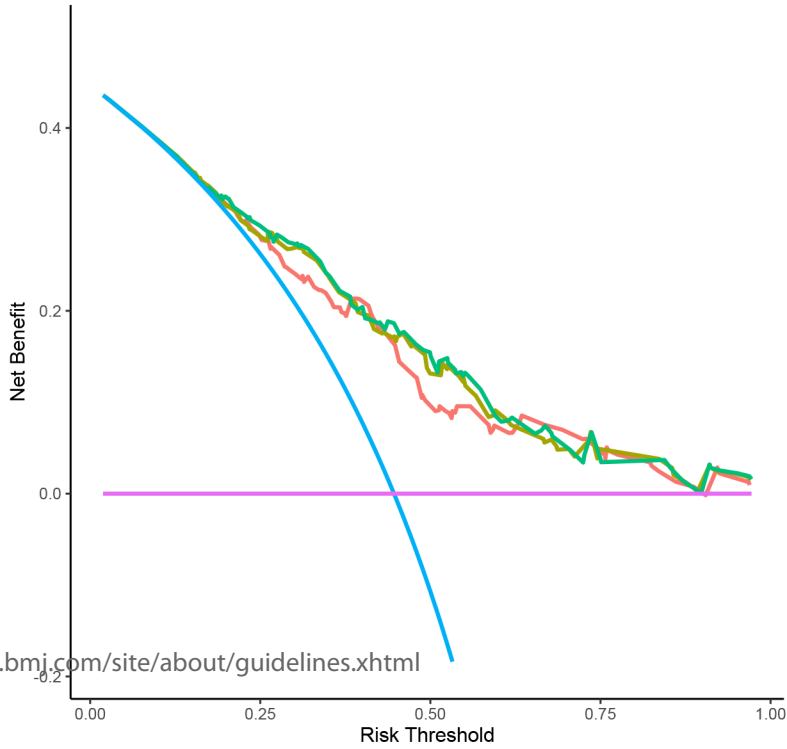
Validating set



Training set



Validating set



Supplementary Table 1. Comparison of the discrimination ability of different models based on different cutoff values.

	Crude model			Adjusted model	
	HR (95% CI)	C-index	AIC	HR (95% CI)	C-index
LNR					
Cutoff_1		0.607	5376	0.643	5350
<0.09	1 (ref)			1 (ref)	
0.09~0.33	1.88 (1.44, 2.44)			1.74 (1.33, 2.29)	
>0.33	3.02 (2.30, 3.97)			2.63 (1.97, 3.50)	
Cutoff_2		0.605	5378	0.641	5355
<0.09	1 (ref)			1 (ref)	
0.09~0.40	1.97 (1.52, 2.54)			1.85 (1.38, 2.54)	
>0.40	3.16 (2.38, 4.21)			2.72 (2.02, 3.67)	
LODDS					
Cutoff_1		0.609	5373	0.644	5346
<-2.09	1 (ref)			1 (ref)	
-2.09~-0.65	1.93 (1.48, 2.51)			1.80 (1.36, 2.37)	
>-0.65	3.13 (2.38, 4.13)			2.77 (2.07, 3.70)	
Cutoff_2		0.605	5378	0.640	5352
<-2.10	1 (ref)			1 (ref)	
-2.09~-0.37	2.00 (1.54, 2.59)			1.86 (1.42, 2.44)	
>-0.37	3.26 (2.45, 4.33)			2.83 (2.10, 3.81)	

Cutoff_1 was generated by minimal p-value method via X-tile software, cutoff_2 was generated using P25 and P75.

Supplementary Table 2. Univariate analysis of the potential confounders.

Variables	HR (95% CI)	P-value
Age (increased by 10ys)	1.06 (0.98 – 1.14)	0.065
Tumor size (increased by 1cm)	1.13 (1.06 – 1.19)	<0.001
Sex		
Male	1	
Female	1.09 (0.89 – 1.34)	0.386
Race		
Others	1	
Black	1.42 (0.91 – 2.21)	0.118
White	1.57 (1.13 – 2.19)	0.007
Grade		
G1	1	
G2-3	2.08 (1.20 – 3.60)	0.009
G4	3.08 (1.32 – 7.22)	0.009
T		
T1-2	1	
T3-4	1.70 (1.38 – 2.10)	<0.001
N		
N1	1	
N2	1.61 (1.34 – 1.93)	<0.001
N3	2.15 (1.77 – 2.62)	<0.001
M		
M0	1	
M1	2.61 (2.00 – 3.39)	<0.001
Low nodes yield		
No	1	
Yes	1.64 (1.35 – 1.99)	<0.001

1
2
3 Editor's Comments to Author (if any):
4
5

6 - Please revise your title so that it includes your study design. This is the preferred format for the
7 journal.

8 **Response:** The study design is included in the title. It is a register-based retrospective cohort
9 study.

10 - Please revise the abstract >> design section. A "database review" is not an appropriate
11 description of your study.

12 **Response:** This section has been revised as "a register-based retrospective cohort study".

13 - Please revise the Patient and Public Involvement statement. This section should be included as a
14 sub-heading in the methods section of all manuscripts. It should provide a brief description of any
15 patient involvement in study design or conduct of the study, as well as any plans to disseminate
16 the results to study participants. If patients and or public were not involved please state this.

17 The Patient and Public Involvement statement should NOT contain details of participant
18 recruitment, patient consent or ethics approval. This information should be included elsewhere in
19 your methods section. Please see our blog for further information regarding PPI:
20 [http://blogs.bmj.com/bmjopen/2018/03/23/new-requirements-for-patient-and-public-involvement-](http://blogs.bmj.com/bmjopen/2018/03/23/new-requirements-for-patient-and-public-involvement-statements-in-bmj-open/)
21 [statements-in-bmj-open/](http://blogs.bmj.com/bmjopen/2018/03/23/new-requirements-for-patient-and-public-involvement-statements-in-bmj-open/)

22 **Response:** The PPI statement has been revised according to the information from blog.

23 *"The development of the research question and outcome measures were not informed by patients'*
24 *priorities, experience, and preferences. The patients were involved during the retrospective review of*
25 *public database where cases were diagnosed during 2010-2015. Patients were not involved in the*
26 *recruitment to and conduct of the study. The findings of the study will be disseminated by online article*
27 *to all study participants whose identity kept confidential during the whole research."*

28 - Please work on improving the reporting of the methods. For example, what was the study's
29 design? What are the settings? More information is needed on the data source used. Was it an
30 anonymised dataset? What permissions were obtained? Did this study require approval from your
31 ethics committee? If not then please explain why not.

32 **Response:** The reporting of the methods has been revised at the beginning of the Methods section.

33 *"This study is a SEER register-based retrospective cohort study, which aimed to enroll patients with*
34 *node positive cardia gastric adenocarcinoma (CGA), review crucial clinical characteristics and*
35 *observe survival of this population. The source of SEER data is registered cancer cases from various*
36 *locations throughout the United States. The permission of data access was obtained by sending*
37 *application form and receiving confirmation mail with valid username (21268-Nov2019) and password.*
38 *This study was granted exemption from requiring informed consent. All procedures performed in*
39 *studies involving human participants were in accordance with the ethical standards of the Ethics*
40 *Committee of the Anhui Medical College."*

41 - Along with your revised manuscript, please provide a completed copy of the STROBE checklist
42 (<http://www.strobe-statement.org/>).

43 **Response:** We provide a STROBE checklist and each item is linked to the line number of the
44 revised manuscript.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ##Reviewer: 1

4 Dr. Rui Zhong, Southwest Medical University
5
6

7 Comments to the Author:

8 1.The continuous factors evaluated, LNR, age, tumour size, LODDS were all categorized and then
9 included in the model. Categorization results in substantial loss of statistical power and reduced
10 interpretability. For instance, by selecting the cutoff of age at 65, the interpretation is that a 64
11 year old is the same as a 30 year old, but the 64 year old is different than a 65 year old. This makes
12 no biological sense. Further, although the authors state that the 'best optimal cutoff' was selected
13 (which is a form of data dredging), it is questioned that the best optimal cutoff would result in
14 cutoffs as they did (i.e. age=65, tumour size=25, etc). The best option would be to leave the
15 continuous variables as continuous for modeling purposes.
16

17
18 **Response:** Thank you for your suggestion. According to the coding rule of SEER database, age
19 and tumor size are not always in numeric format (e.g. year of 85+, or tumor size of >990 mm),
20 nevertheless you kindly remind us that stratification by cutoff values will result in substantial loss
21 of statistical power. Therefore the revised manuscript transforms age and tumor size into
22 categorized variables that contain much more strata. Age is transformed as a variable of 9 levels,
23 with the lowest of 0-10 and the highest of 80+ (interval of 10 years). However univariate analysis
24 shows that age is not associated with survival (see Suppl. Table 1.), so it is not included in the
25 final model. Likewise tumor size is transformed as a variable of 6 levels, with the lowest of less
26 than 1cm and the highest of over 5cm (interval of 1cm).
27

28 LNR and LODDS are still kept as trichotomous factors because they are used as alternative
29 indicators for AJCC N stage.
30
31

32
33
34 2.What selection process was used to select factors for inclusion in the multivariable model?
35 Please explain whether a single factor regression analysis was performed before the variables
36 entered the multivariate COX regression. In addition, please explain how this process addresses
37 the potential effects of confounding or collinearity.
38

39 **Response:** In previous manuscript, we selected factors due to clinical significance (for example,
40 higher stage or large tumor size indicates unfavorable outcome). Inspired by your question, we
41 consider both univariate model results and clinical significance in the revised manuscript, and
42 finally include tumor size, race, grade, T stage, M stage and low nodes yield as adjusted variables
43 for stepwise Cox regression model. The results of univariate analyses are listed in the
44 supplementary table 1. So the confounding effect is addressed by multivariable analysis that
45 incorporates with potential confounders.
46
47
48

49
50 3.Please consider augmenting the discussion of the findings concerning LNR. LNR association
51 with survival is an exciting aspect of cardia gastric cancer that is currently emerging and may be
52 clinically meaningful.
53

54 **Response:** Thank you for your suggestion. In the previous paper, we focused on the advantage of
55 LNR when LN harvest is inadequate. In the revised paper, we further discuss the clinical meaning
56 of LNR. We are very grateful for the comment that helps us improve this study.
57
58

59 4.Why did the author use the x-titile software for the cutoff value to choose the third quartile
60

1
2
3 instead of the binary or interquartile range?

4 **Response:** We transformed LNR and LODDS into trichotomous variables and selected two cut-off
5 points that represented the greatest group difference of CSS probability according to the minimum
6 p-value method. Thus LNR was re-coded as R1, R2 and R3; LODDS was re-coded as L1, L2 and
7 L3, which was similar with the trichotomous AJCC N staging (N1, N2 and N3). Since only
8 node-positive patients were enrolled, no N0 patients existed here. We address this issue in the
9 revised manuscript in order to make it clear and understandable for readers. Thank you for your
10 question.
11
12
13

14
15 5.As the author said whether LNR or LODDS based staging system outperforms TNM 8th edition
16 needs to be further investigated. We want to know whether all the stagings can be corrected to the
17 eighth edition based on the existing fields of the seer databases.
18

19 **Response:** Inspired by your suggestion, all the stagings are now corrected to the 8th edition.
20 Accordingly, the results have been modified. Thank you.
21
22

23 6.Given the importance of the Cox PH model for the development of the nomogram, it would
24 appropriate to include validation that the assumptions of a Cox PH model are met. Please include
25 plots (in the main text or a supplemental figure) of the Schoenfeld, Martingale, and Deviance
26 residuals.
27

28 **Response:** The plots for PH model validation are included in the supplementary figure 1. Plots of
29 Schoenfeld, Martingale, and Deviance residuals for models that incorporate N stage, LNR and
30 LODDS are all presented. The tests show that PH assumption is met in all models ($P>0.05$). Thank
31 you for the advice.
32
33

34 7.ECOG/Karnofsky performance scores not utilized. Please comment on why these were not
35 utilized as they serve as significant reference points for PC treatment. If possible, this would be a
36 great thing to include in this analysis or the analyses suggested above.
37

38 **Response:** Unfortunately, variables that reflect general status are not available in SEER database,
39 so ECOG or KPS was not utilized. We address this issue as a limitation in the Discussion Section
40 of the revised paper. Thank you for the comment.
41

42 *"...One limitation of this study is that some important factors that are associated with survival are*
43 *not considered in the model due to unavailable data source. For example, ECOG/KPS score is*
44 *commonly taken into account in survival analysis due to its remarkable relationship with general*
45 *status and prognosis. Unfortunately the SEER 18 database does not record the score at diagnosis,*
46 *so the impact of it is not considered in this analysis.....To overcome this limitation, a database*
47 *that provides with fully detailed medical records is needed for analysis."*
48
49
50

51 8.altered cutoff values would be an essential factor to analyze during the further optimization of
52 this model. This needs to be further explained in the discussion.
53

54 **Response:** Thank you for your suggestion. In order to cut LNR and LODDS into trichotomous
55 factor, the previous study selected two points by using the minimal p-value method via X-tile
56 software. Unlike one cutoff point selection, ROC and maximally selected rank statistics cannot be
57 applied in two-point selection. Therefore the revised manuscript attempts to generate trichotomous
58 factors using P25/P75, construct regression model and compare discrimination ability with X-tile
59
60

1
2
3 based cutoff values. As a result, the X-tile based values have higher power and are finally included
4 in further analysis. We address this issue in the revised paper and present the process in Suppl.
5 Table 2. Although the previous results remain, this step is very crucial for optimizing the models
6 and improving the study reliability. We deeply appreciate this comment.
7

8 *"...For model optimization, LNR and LODDS were also categorized into trichotomous factors*
9 *using cut-off values of P25 and P75. The discrimination ability of the model based on interquartile*
10 *was lower (Suppl. Table 2), so this model was not further analyzed...."*
11
12

13
14 9. Please explain why the N stages do not appear in this nomogram. For example, they do not show
15 significance in the multivariate COX regression?

16 **Response:** In the present study, we attempt to construct a new alternative indicator for N stage,
17 because the current N stage classification may not perform well in cardia gastric cancer, so we
18 presented the nomograms that cooperated with LNR or LODDS, other than N staging. In other
19 words, N stages DO appear in the nomogram, but in the form of LNR or LODDS. In addition, the
20 other reviewer suggested us to pick one plot to avoid confusion and unclear message; we only
21 show one nomogram that cooperates with LNR in the main text, and show the other plot in the
22 supplementary file. We hope the revision is acceptable.
23
24
25

26
27 10. We believe that the nomogram established by the author should be compared with traditional
28 models, such as ROC curve and Decision Curve Analysis.

29 **Response:** In the revised paper, ROC and DCA curve are made to compare the prognostic powers
30 between the new nomogram models based on LNR and LODDS and the traditional model based
31 on N stage. The results indicate that the new nomogram models are better. For details, please see
32 the supplementary figure 3. We really thank you for the suggestion that further confirms our
33 results.
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 ##Reviewer: 2

4 Dr. Rasa Zarnegar , Weill Cornell Medical College

6
7 Comments to the Author:

8 I think this is a nice paper with 2 nomograms for the determination for CGA.

9
10
11 I think there are some revisions that would make this paper better. The SEER DB has access to
12 total number of nodes analyzed and the number of positive nodes. It is important to use the current
13 guidelines 8th Edition for this analysis even though the data was from prior to implementation of
14 the 8th. The concept should still hold and validate the rigor of the study based on current
15 guidelines.
16

17 **Response:** All the 7th stagings are now corrected to the 8th edition. Thank you for your brilliant
18 suggestion.
19

20
21 2. Raw data on the patient population No of nodes harvested and total positive is required to
22 determine the frequency of low yield in the study design.

23 **Response:** In the revised paper, we describe the number of nodes harvested and total positive, and
24 the frequency of low nodes yield (please see Table 1). The frequency of low yield is also in the
25 multivariable model because it is associated with survival in univariate analysis (Suppl. Table 2).
26
27

28
29 3. There should be discussion and data on neoadjuvant therapy as this likely impact survival and
30 the number of patient that received therapy.

31 **Response:** Neoadjuvant therapy is likely to influence survival; unfortunately the database,
32 Incidence - SEER Research Data, 18 Registries, Nov 2019 Sub 2000-2017 (SEER 18 database),
33 does not provides with information about chemotherapy, therefore it is unavailable for this study.
34 We treat it as a limitation and discuss the impact of neoadjuvant therapy in the Discussion Section.
35 *"...Treatment mode is also associated with clinical outcome. This study enrolled patients who
36 received gastric resection; however other information about chemo- or radiotherapy is not
37 available in SEER 18 database. Randomized clinical trial demonstrates that compared with
38 surgery alone, preoperative administration of carboplatin and paclitaxel with concurrent
39 radiotherapy significantly improved overall survival among patients with esophageal or GEJ
40 cancer (HR = 0.657) (31). The NCCN clinical practice guidelines for GEJ cancer recommend
41 preoperative chemoradiation or perioperative chemotherapy due to substantial survival benefit
42 compared with surgery alone (32). To overcome this limitation, a database that provides with fully
43 detailed medical records is needed for analysis...."*
44
45
46
47
48
49

50 4. I suggest picking one nomogram. I think its important to send a clear message and by
51 presenting 2 nomograms the authors are creating confusion and an unclear message. I think by
52 being more focused on one approach would allow for improved implementation. The authors may
53 want to compare whichever they select with conventional lymphadenectomy is so desired.

54 **Response:** Thanks for the advice. We pick the nomogram that incorporates with LNR because the
55 calculation of LNR is easier than LODDS, which is more convenient in clinical practice. The
56 nomogram and calibration curves based on LODDS are shown in supplementary figure 2. In
57 addition, for more clear presentation, we put the calibration curves of training (red) and validating
58
59
60

1
2
3 set (blue) into one plot.
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

STROBE Statement—checklist of items that should be included in reports of observational studies

All line numbers are based on the manuscript with tracked change.

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract title page (b) Provide in the abstract an informative and balanced summary of what was done and what was found page2
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported page 4
Objectives	3	State specific objectives, including any prespecified hypotheses page 4
Methods		
Study design	4	Present key elements of study design early in the paper page 4
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection page 4-5
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up page 4 <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls NA <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants NA (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed NA <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable page 5
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group page 5
Bias	9	Describe any efforts to address potential sources of bias page 5
Study size	10	Explain how the study size was arrived at NA
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why page 5-6
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding page 5-6 (b) Describe any methods used to examine subgroups and interactions page 4-6 (c) Explain how missing data were addressed page 4 (d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed NA <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed NA <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy NA (e) Describe any sensitivity analyses NA

Results

Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed page 4 (b) Give reasons for non-participation at each stage page 4 (c) Consider use of a flow diagram figure 1, page 4
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders page 6 (b) Indicate number of participants with missing data for each variable of interest figure 1, page 4 (c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount) Table 1 page 6-7
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time page 6 <i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure NA <i>Cross-sectional study</i> —Report numbers of outcome events or summary measures NA
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included table 1, page 7-9 (b) Report category boundaries when continuous variables were categorized page 7 (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses NA
Discussion		
Key results	18	Summarise key results with reference to study objectives page 9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias page 10-11
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence page 9-11
Generalisability	21	Discuss the generalisability (external validity) of the study results page 11
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based page 11

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Prognostic value of different lymph node staging methods for node-positive cardia gastric cancer: a register-based retrospective cohort study

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2021-050378.R2
Article Type:	Original research
Date Submitted by the Author:	29-Jul-2021
Complete List of Authors:	Wang, Xiao-Qing; Anhui Medical College Bao, Min; Anhui Medical College Zhang, Cheng; First Affiliated Hospital of Anhui Medical University
Primary Subject Heading:	Oncology
Secondary Subject Heading:	Gastroenterology and hepatology
Keywords:	Gastrointestinal tumours < ONCOLOGY, Gastrointestinal tumours < GASTROENTEROLOGY, SURGERY

SCHOLARONE™
Manuscripts



I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our [licence](#).

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which [Creative Commons](#) licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 **Title page**

2 **Article title:** Prognostic value of different lymph node staging methods for node-positive cardia
3 gastric cancer: a register-based retrospective cohort study

4 **Author information:** Xiao-Qing Wang 1, Min Bao 1, Cheng Zhang 2

5 1 Anhui Medical College, Hefei, Anhui, China PR

6 2 Anhui Provincial Cancer Institute, the First Affiliated Hospital of Anhui Medical University,
7 Hefei, Anhui, China PR

8 Corresponding to: Cheng Zhang, ahmuzc@sina.cn

9 **Disclaimer:** The views expressed in the submitted article are our own and not an official position
10 of the institution or funder.

11 **Word count:** 2962 (excluding its abstract, acknowledgments, tables, figure legends, and
12 references)

13 **Number of figures and tables:** This draft included 3 figures and 2 tables.

14 **Disclosure of relationships and activities:** None.

15 ABSTRACT

16 **Objective:** To investigate the prognostic efficacy of lymph node ratio (LNR) and log odds of
17 positive lymph nodes (LODDS) in node-positive cardia gastric adenocarcinoma (CGA).

18 **Design:** A registry-based retrospective cohort study.

19 **Setting:** Patients diagnosed with node-positive CGA in the Surveillance, Epidemiology, and End
20 Results (SEER) database from 2010 to 2015.

21 **Participants:** A total of 1038 patients were enrolled and randomly assigned (7:3) to the training
22 set (N = 723) or validating set (N = 315).

23 **Primary outcome measure:** Cancer-specific survival (CSS).

24 **Results:** The baseline characteristics of the training and validation sets were similar. Based on the
25 optimal cut-off values, LNR was classified into low (<0.09), medium (0.09~0.33), and high (>0.33)
26 groups; LODDS was also classified into low (<-2.09), medium (-2.09~-0.65), and high (>-0.65)
27 groups. CSS was significantly different across LNR and LODDS subgroups. The Harrell
28 concordance index of the N stage was lower than that of the LNR or LODDS. The Akaike
29 information criterion of the N stage was higher than that of the LNR or LODDS. Independent
30 predictors included race, T stage, M stage, and LNR (or LODDS), and they were incorporated into
31 nomograms for 1-, 2-, and 5-year CSS prediction. Calibration plots showed satisfactory results for
32 internal and external validity of the nomogram.

33 **Conclusions:** LNR and LODDS staging methods have better prognostic efficacy than the
34 traditional N staging method in CGA with node metastasis. Moreover, the two values are
35 promising substitutes for N staging in nomogram development when other independent prognostic
36 factors are incorporated.

37 KEY WORDS

38 Cardia; Adenocarcinoma; Stomach Neoplasms; Lymph Node Ratio; Neoplasm Staging

39 STRENGTHS AND LIMITATIONS OF THIS STUDY

- 40 ● This study used the national cancer registry data for cardia gastric adenocarcinoma research;
- 41 ● Novel staging methods based on the number of positive lymph nodes have been established
42 for prognostic prediction.
- 43 ● Nomograms based on the new staging methods were constructed and validated;
- 44 ● The validity of the outcomes of the study needs to be confirmed in other populations.

45 **Patient consent form:** The SEER database review is granted exemption from obtaining patients'
46 consents.

47 **Word count:** 2962 (excluding its abstract, acknowledgments, tables, figure legends, and
48 references)

49 INTRODUCTION

50 Gastric cancer (GC) generally includes two topographical categories: non-cardia GC that occurs at
51 a more distal part of the stomach and GC of the cardia that occurs at the gastroesophageal junction
52 (GEJ). In contrast to the steady decline in the incidence of non-cardia GC, GC of the cardia occurs
53 more frequently, particularly in high-income countries (1, 2). This trend is associated with obesity,
54 gastroesophageal reflux disease, and Barrett esophagus (2). In addition to the difference in the
55 incidence trend, the clinic pathological features and long-term survival vary between the two GC
56 subtypes (3). Precise staging is necessary for the accurate prediction of survival. The
57 Tumor-Node-Metastasis (TNM) classification 7th edition of the American Joint Committee on
58 Cancer (AJCC) recommends harvesting of at least 15 lymph nodes (LN) for N staging (4, 5).
59 However, inadequate LN harvest is frequent because of various reasons; thus, precise staging is
60 difficult. It has been demonstrated that the LN ratio (LNR) could provide a better estimate of the
61 survival of patients with GC after curative gastrectomy, regardless of the number of LNs
62 examined (6), and might be a promising aid along with the TNM staging system (7). Furthermore,
63 in previous reports, the log odds of positive LN (LODDS) outperformed the N and LNR staging
64 systems in predicting the survival of patients with GC (8-10). Therefore, the traditional N staging
65 classification might be substituted with different methods with improved performance.
66 Nevertheless, few studies have evaluated the performance of the two LN staging systems in GC of
67 the cardia, which has distinct clinical characteristics and epidemiology than other types of GC.

68 Here, we used the data of a nationwide cancer registry to evaluate the prognostic value of LNR
69 and LODDS in patients with node-positive cardia gastric adenocarcinoma (CGA), and, if possible,
70 construct a nomogram for the prediction of survival based on the new LN staging system.

71 METHODS

72 Study design and Participant Selection

73 This study was a Surveillance, Epidemiology, and End Results (SEER) registry-based
74 retrospective cohort study, which aimed to enroll patients with node-positive CGA, review their
75 critical clinical characteristics, and observe the survival of this population. The source of the
76 SEER data is registered cases of cancer from various locations throughout the United States.
77 Permission for data access was obtained by sending an application form and receiving
78 confirmation mail with a valid username (21268-Nov2019) and password.

79 A SEER*Stat (version 8.3.8) was used to access the Incidence-SEER Research Data, 18 Registries,
80 Nov 2019 Sub 2000-2017 (SEER 18 database) (11) and to obtain data of node-positive patients
81 with CGA. The inclusion criteria were as follows: 1) the International Classification of Disease for
82 Oncology, Third Edition (ICD-O-3) code for the primary tumor site was C16.0 (cardia); 2) broad
83 histological recode was 8140-8389 adenomas and adenocarcinomas; 3) diagnostic confirmation
84 was by positive histology; 4) surgery was performed; 5) diagnosis was during 2010-2015; and 6)
85 the definite number of positive regional nodes was known and was not zero. Cases with unknown
86 race, T stage information, tumor size, or tumor grade were excluded. As shown in Figure 1, the
87 final cohort comprised 1038 patients with node-positive CGA, of whom 857 were male, and 181
88 were female. A total of 338 (32.56%) patients were above 70 years of age. Eight hundred and
89 ninety-six (86.32%) patients were White, 64 were Black, and 78 were of other races. Of the total

90 cohort, 70% of the patients were randomly assigned to the training set (N = 723), and the
91 remaining were assigned to the validation set (N = 315).

92 **Technical Information**

93 The main outcome was cancer-specific survival (CSS), which was defined as the period between
94 the first diagnosis and death specifically due to CGA. In addition, we extracted the following
95 variables for analysis: sex, race, age, AJCC 7th TNM stage information, tumor size, tumor grade,
96 number of regional nodes examined, and number of regional nodes that were positive. The
97 information about the stage of cancer was further corrected according to the AJCC 8th criteria.
98 LNR and LODDS were calculated as previously reported (12). Briefly, LNR was defined as the
99 ratio of the number of positive nodes divided by the total number of examined nodes. LODDS was
100 calculated using the formula: $\log(\text{NPLN}+0.50)/(\text{NDLN}-\text{NPLN}+0.50)$, in which 0.50 was added to
101 both the numerator and denominator to avoid an infinite number.

102 The optimal thresholds for dividing LNR and LODDS into trichotomous variables were
103 determined using the X-tile software (version 3.6.1) (13), which were based on the maximal
104 log-rank chi-square value that represented the greatest group difference in CSS probability. LNR
105 and LODDS were classified into three levels because they are proposed as alternative indicators
106 for N stage in node-positive GC, including N1, N2, and N3.

107 **Statistics**

108 The distributions of baseline characteristics between the training and validation sets were
109 described and compared using chi-square test. Survival curves, median survival, and CSS rates
110 were generated using Kaplan-Meier method. Outcome differences between the groups were
111 analyzed using log-rank test. After testing proportional hazard assumption, a multivariable Cox
112 regression model was used to establish a CSS prognostic model. The prognostic power was
113 evaluated using Akaike information criterion (AIC) and Harrell concordance index (C-index). A
114 predictive model with a lower AIC indicated a better model fit, while a higher C-index indicated a
115 better discriminative ability. A C-index value of 0.5 indicated no predictive power, and an index
116 of 1.0 indicated complete differentiation. Cox stepwise regression analysis was also performed to
117 construct a nomogram for the prediction of 1-, 2-, and 5-year CSS. Validation of the nomogram
118 was performed using internal and external calibration plots (14). Bootstraps with 1000 resamples
119 were used for the validation activities. Receiver operator characteristic (ROC) curves and areas
120 under the ROC curves (AUCs) were calculated to evaluate the accuracy of CSS prediction using
121 different models. Decision curve analysis (DCA) was performed to determine the clinical
122 application of different models: the proportion of true positive results minus the proportion of
123 false positive results, and the relative risks of false-positive and false-negative results were
124 weighted to obtain the net benefits of decision-making. All statistical analyses were performed
125 using R software (version 3.5.3). A two-tailed P value of less than 0.05 was considered
126 statistically significant.

127 **Patient and Public Involvement**

128 Due to the retrospective and observational nature of the study, the research question and outcome
129 measures were not developed and influenced by patients' priorities, experiences, and preferences.
130 Patients were not involved in the design, recruitment, and conduct of this study. Patients were not

131 asked to assess the burden of the intervention and time required to participate in the research. The
 132 findings of the study will be disseminated online and are freely available for public.

133 **Ethics approval statement**

134 The Ethics Committee of Anhui Medical College exempted the requirement for ethics approval
 135 because of the observational nature of the study.

136 **RESULTS**

137 Table 1 summarizes the demographic and clinical features of the participants. In all, 628 patients
 138 (60.50%) were diagnosed with a tumor less than 5cm. Six hundred and forty patients (61.66%)
 139 were diagnosed with grade III or IV cancer. The number of patients with T1, T2, T3, and T4 stage
 140 was 94, 125, 717, and 102, respectively. The number of patients with N1, N2, and N3 stage was
 141 479, 330, and 229, respectively. Seventy-five patients (7.23%) had distant metastasis at
 142 presentation. The median CSS was 27 months. The rate of 1, 2, and 5-year CSS was 76.8%, 53.0%,
 143 and 29.2%, respectively. There was no statistical difference in the baseline characteristics between
 144 the training and validating set. The detailed information about the two sets is also presented in
 145 Table 1.

Table 1. Baseline information of the included patients with node positive CGA, N(%).

Groups	Training set (N = 723)	Validating set (N = 315)	P-value
Sex			
Male	596 (82.43)	261 (82.86)	0.939
Female	127 (17.57)	54 (17.14)	
Age			
<70	490 (67.77)	210 (66.67)	0.781
≥70	233 (32.23)	105 (33.33)	
Race			
White	628 (86.86)	268 (85.08)	0.437
Black	40 (5.53)	24 (7.62)	
Others	55 (7.61)	23 (7.30)	
Tumor size			
<5cm	442 (61.13)	186 (59.05)	0.573
≥5cm	281 (38.87)	129 (40.95)	
Grade			
I-II	279 (38.59)	119 (37.78)	0.859
III-IV	444 (61.41)	196 (62.22)	
T stage			
T1a	17 (2.35)	4 (1.27)	0.224
T1b	53 (7.33)	20 (6.35)	
T2	83 (11.48)	42 (13.33)	
T3	501 (69.29)	216 (68.57)	
T4a	49 (6.78)	28 (8.89)	
T4b	20 (2.77)	5 (1.59)	

N stage			
N1	332 (45.92)	147 (46.67)	
N2	229 (31.67)	101 (32.06)	0.921
N3	162 (22.41)	67 (21.27)	
M stage			
M0	678 (93.78)	285 (90.48)	
M1	45 (6.22)	30 (9.52)	0.079
Low nodes yield			
Yes	532 (73.58)	243 (77.14)	0.300
No	191 (26.42)	72 (22.86)	
No. of nodes harvest	17 (12, 25)	16 (11, 24)	0.400
No. of positive nodes	3 (1, 6)	3 (1, 6)	1.000
Median survival (months)	28 (25, 32)	25 (21, 32)	0.361
CSS rate (%)			
1-year	77.0 (74.0, 80.2)	76.3 (71.6, 81.2)	
2-year	53.7 (50.1, 57.5)	51.4 (46.0, 57.5)	
5-year	30.3 (26.7, 34.5)	26.4 (20.9, 33.4)	

Abbreviation: CGA, cardia gastric adenocarcinoma; CSS, cancer-specific survival

146 According to X-tile software results, the optimal cut-off values for LNR were 0.09 and 0.33, and
 147 for LODDS were -2.09 and -0.65. Thus, patients were classified into the low (<0.09, R1), medium
 148 (0.09~0.33, R2), or high LNR (>0.33, R3) groups; or low (<-2.09, L1), medium (-2.09~-0.65, L2),
 149 or high LODDS (>-0.65, L3) groups. For model optimization, LNR and LODDS were also
 150 categorized into trichotomous factors using the cut-off values of P_{25} and P_{75} . The discrimination
 151 ability of the model based on the interquartiles was poor (Suppl. Table 1); hence, this model was
 152 not analyzed further. Next, we created the survival curves of the patients according to the N
 153 staging, LNR, or LODDS staging system. As shown in Figure 2 in the training set, CSS was
 154 significantly different between all the three staging systems (all the log-rank P values < 0.0001);
 155 however the 95% confidence intervals (CIs) of N2 and N3 survival curve were initially divergent
 156 and partly overlapped afterward. The inferior discriminative ability of the N system was further
 157 reinforced by the AIC and C-index. As shown in Table 2, the C-index of the N stage was lower
 158 than that of LNR or LODDS. Similarly, the AIC of the N stage was higher than that of the LNR or
 159 LODDS. The clinical characteristics with statistical significance for CSS were further
 160 incorporated in the Cox regression model as potential confounders (Suppl. Table 2), and all the
 161 variables met the proportional hazard assumption (Suppl. Figure 1, all the P values > 0.05). The
 162 prognostic value of the adjusted model was generally better than that of the crude model. In
 163 addition, the prognostic value of the LNR system seemed to be poorer than that of the LODDS
 164 system; however, the difference was not significant; hence, we incorporated both the systems for
 165 nomogram construction.

Table 2. Prognostic values of variables for patients with node positive CGA (N = 1 038).

Variables	Crude model			Adjusted model		
	HR (95% CI)	C-index	AIC	HR (95% CI)	C-index	AIC

Training set (N = 723)					
N stage		0.582	5403		0.632 5365
N1	1 (ref)			1 (ref)	
N2	1.53(1.24, 1.91)			1.42(1.14, 1.77)	
N3	2.15 (1.70, 2.71)			2.03(1.60, 2.59)	
LNR		0.607	5376		0.643 5350
R1	1 (ref)			1 (ref)	
R2	1.88 (1.44, 2.44)			1.74 (1.33, 2.29)	
R3	3.02 (2.30, 3.97)			2.63 (1.97, 3.50)	
LODDS		0.609	5373		0.644 5346
L1	1 (ref)			1 (ref)	
L2	1.93 (1.48, 2.51)			1.80 (1.36, 2.37)	
L3	3.13 (2.38, 4.13)			2.77 (2.07, 3.70)	
Validating set (N = 315)					
N stage		0.596	1957		0.675 1931
N1	1 (ref)			1 (ref)	
N2	1.81 (1.31, 2.51)			1.75 (1.25, 2.46)	
N3	2.18 (1.51, 3.15)			2.23 (1.50, 3.30)	
LNR		0.646	1927		0.691 1913
R1	1 (ref)			1 (ref)	
R2	2.20 (1.47, 3.30)			1.91(1.26, 2.90)	
R3	4.16 (2.76, 6.28)			3.58 (2.30, 5.56)	
LODDS		0.647	1927		0.789 1914
L1	1 (ref)			1 (ref)	
L2	2.07 (1.39, 3.09)			2.08 (1.38, 3.14)	
L3	4.22 (2.79, 6.39)			4.10 (2.65, 6.34)	

Abbreviations: CGA, cardia gastric adenocarcinoma; HR, hazard ratio; CI, confidence interval; AIC, Akaike information criterion; LNR, lymph node ratio; LODDS, log odds of positive lymph nodes.

Adjusted model considered race, tumor size, grade, T stage and M stage.

Stepwise Cox regression analysis showed race, T stage, M stage, and LNR (or LODDS) were independent predictors; hence, these factors were included in the nomograms. For both LNR and LODDS, the total score was 40, and a higher score suggested lower survival (Figure 3 and Suppl. Figure 2). Next, the calibration plot was used to assess the internal and external validity of the nomogram (Figure 3 and Suppl. Figure 2). Since the cross-spot line was generally close to the grey reference line, we concluded that the predicted CSS was well correlated with the actual state. In addition, ROC curves indicated that the AUC of the model based on N stage was lower than that of the model based on the nomogram of LNR or LODDS (Suppl. Figure 3). However the DCA plot does not show advantage of the nomogram (Suppl. Figure 3).

DISCUSSION

The present study analyzed the databases of the national cancer registry and demonstrated that survival of patients with node-positive CGA could be well predicted when the traditional N staging method is substituted with an LNR or LODDS system. This outcome was seen both in

1
2
3 179 training and validation set. In the training set, the survival curves clearly separated when the
4 180 patient grouping was implemented following the LNR or LODDS method, which was not
5 181 achieved by the traditional N staging system. An adjusted model that simultaneously considered
6 182 the staging, clinical, and demographic features, outperformed the crude model that only considers
7 183 staging. Therefore, multiple independent survival factors were incorporated in the nomogram
8 184 construction, which suggested White, deeper infiltration of the tumor, higher proportion of
9 185 positive LN, and metastasis as risk factors. The nomograms performed consistently across the 1, 2,
10 186 and 5-year prediction of the CSS as seen in the validation plots.

11 187 Previous studies have demonstrated the superiority of LNR or LODDS for prognostic prediction
12 188 in GC after surgical resection (8-10, 15-17). However, the patients were not further classified
13 189 according to the primary tumor site; this is a critical limitation since there is a significant
14 190 difference between cardia and non-cardia GC in terms of tumor features, etiological factors, and
15 191 biological behaviors (3). In the AJCC cancer staging 7th edition, tumors involving GEJ were
16 192 categorized as esophageal cancer (5). This was debatable because the GC staging system has a
17 193 better ability to predict survival of a GEJ tumor (18, 19). In the latest 8th edition (20), a tumor that
18 194 has its epicenter within 2 cm of the GEJ and involves the GEJ (Siewert type I/II) is classified as
19 195 esophageal cancer. Other types of GCs, including a tumor with an epicenter more than 2 cm from
20 196 the GEJ or a tumor located with 2 cm of the GEJ but not involving the GEJ, are classified as GC.
21 197 The superiority of the new system was confirmed by a retrospective observational study from two
22 198 institutions in China that have a high volume of cases of GC, regardless of the Siewert type (21).
23 199 In terms of the Siewert type II junctional adenocarcinoma, a marginal superiority of the
24 200 esophageal cancer was found in discriminating survival rates after three and five years. However
25 201 the advantage of the GC system lies in the division of the N3 category into N3a and N3b. Hence,
26 202 the authors concluded that neither the esophageal nor the stomach staging system is accurate in
27 203 predicting survival in Siewert type II junctional cancer (22). Moreover, CGA is probably a special
28 204 entity that has different biological characteristics compared to distinct gastric or esophageal cancer.
29 205 To the best of our knowledge, the present study is the first to demonstrate a superior prognostic
30 206 prediction based on LNR or LODDS in patients with node-positive CGA. Unfortunately, we were
31 207 unable to consider the Siewert type due to a lack of information in the SEER database; hence,
32 208 further studies are necessary with a special focus on tumor location.

33 209 LNR and LODDS have been proven to be the strongest indicators of survival in gastric
34 210 adenocarcinoma when LN harvest is inadequate (16, 17). It has been demonstrated that, in general,
35 211 more extensive LN resection is associated with better survival, which might be due to either
36 212 improved N classification or a therapeutic effect of lymphadenectomy. For esophageal cancer, the
37 213 worldwide data shows that harvesting 10 nodes for pT1, 20 for pT2, and 30 or more for pT3/T4 is
38 214 desirable for reaching maximum 5-year survival (23). For GC, a higher LN harvest also shows
39 215 improved survival (24). It is suggested that at least 16 nodes be evaluated pathologically and
40 216 evaluation of more than 30 nodes is desirable (25). Overall, it is encouraged to harvest as many
41 217 LNs as possible; balancing the extent of LN resection necessary for accurate N staging and
42 218 maximum survival without unnecessarily increasing the morbidity caused by radical
43 219 lymphadenectomy. Nevertheless, many conditions can lead to inadequate LN harvest. It is
44 220 estimated that only one-fifth of the patients with GC have an adequate number of LN examined in
45 221 Iran (26), while more than 15 LNs are examined in 64% of the patients in the US (25). The LNR

222 and LODDS staging methods do not require an adequate number of LNs to be evaluated. In the
223 present study, a low LN yield was found to be a risk factor for poor survival in univariate analysis;
224 however, it was not significant in the LNR or LODDS based multivariate model, which indicates
225 that LN harvest has little impact on prediction of survival based on LNR or LODDS. In fact, the
226 new node category method is consistent when nodal assessment is inadequate during surgery not
227 only for GC (8, 15-17) but also for colorectal cancer (27), esophageal cancer (28), oral squamous
228 cell carcinoma (29), gallbladder cancer (30), and others.

229 The association between LNR and survival is a promising aspect of cardia GC that is currently
230 emerging and might be clinically relevant. A higher ratio of positive LN indicates a worse
231 outcome in cardia GC. Patients are at 2-3 times higher risk of cancer-specific death if the ratio is
232 over 33%. The ratio of 9-33% also indicates a twofold risk. This effect is independent of other
233 crucial clinical characteristics; thus, it is a useful tool for surgeons to predict the prognosis. This is
234 also evidence supporting truly radical surgery, i.e., complete lymph node resection rather than
235 limited resection (31). In addition, LNR minimizes the “stage migration” phenomenon that occurs
236 with the current N staging system (32).

237 One limitation of this study is that some important factors associated with survival have not been
238 considered in the model due to unavailable data. For example, the Eastern Cooperative Oncology
239 Group or Karnofsky Performance Status score is commonly considered in the survival analysis
240 due to its remarkable relationship with the general status and prognosis. Unfortunately, the SEER
241 18 database does not record the score at diagnosis; hence, its impact is not considered in this
242 analysis. The treatment modality is also associated with clinical outcomes. This study enrolled
243 patients who underwent gastric resection; however other information about chemo- or
244 radiotherapy is not available in the SEER 18 database. A previous randomized clinical trial
245 demonstrated that compared with surgery alone, preoperative administration of carboplatin and
246 paclitaxel with concurrent radiotherapy significantly improved the overall survival among patients
247 with esophageal or GEJ cancer (HR = 0.657) (33). The National Comprehensive Cancer Network
248 clinical practice guidelines for GEJ cancer recommend preoperative chemoradiation or
249 perioperative chemotherapy due to substantial survival benefits compared with surgery alone (34).
250 To overcome this limitation, a database that provides fully detailed medical records is necessary
251 for analysis. Moreover, the inclusion of these factors would greatly improve the prognostic power
252 of the survival prediction model. Another limitation is that our results are based on the training set
253 and confirmed by the validation set; however, the baseline characteristics of the two groups are
254 similar. Hence, these results need to be validated among populations with different characteristics.
255 The third limitation is clinical usability. The DCA result is proposed for assessing the potential
256 clinical impact of risk models for recommending treatment or intervention, and the suggested
257 clinical usability of the nomogram may be poorer than that of other models. In this regard,
258 although this model may have some merits regarding outcome prediction, its use for guiding
259 clinical decisions should be further studied.

260 In conclusion, staging methods based on LNR and LODDS have better prognostic ability than the
261 traditional N staging method in patients of CGA with regional lymph node metastasis. Moreover,
262 the two values are promising substitutes for N staging in nomogram development when other
263 independent prognostic factors are incorporated.

264 **Contributorship statement:** WXQ work conception, data interpretation, critical review for
265 important content, final approval of the manuscript and agreement to be accountable for all
266 aspects of the work; BM administrative work, funding, critical review for important content, final
267 approval of the manuscript and agreement to be accountable for all aspects of the work; ZC
268 acquisition, analysis of data, drafting the work, final approval of the manuscript and agreement to
269 be accountable for all aspects of the work.

270 **Competing interests:** None.

271 **Funding:** This work was supported by [Domestic Visiting Study Project for Outstanding Young
272 Talents in Colleges and Universities in Anhui Province] grant number [gxnfx2021183] and
273 [Major Project of Natural Science of Anhui Provincial Department of Education] grant number
274 [KJ2019ZD73].

275 **Data sharing statement:** Dataset available from the SEER website
276 (<https://seer.cancer.gov/data-software/>).

277 **Acknowledgement:** We thank all patient advisers for their special contribution.
278

References:

- 280 1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018:
281 GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA*
282 *Cancer J Clin.* [Journal Article]. 2018 2018-11-01;68(6):394-424.
- 283 2. Abdi E, Latifi-Navid S, Zahri S, Yazdanbod A, Pourfarzi F. Risk factors predisposing to cardia
284 gastric adenocarcinoma: Insights and new perspectives. *Cancer Med.* [Journal Article; Research
285 Support, Non-U.S. Gov't; Review]. 2019 2019-10-01;8(13):6114-26.
- 286 3. Amini N, Spolverato G, Kim Y, Squires MH, Poultsides GA, Fields R, et al. Clinicopathological
287 features and prognosis of gastric cardia adenocarcinoma: a multi-institutional US study. *J SURG*
288 *ONCOL.* [Comparative Study; Journal Article; Multicenter Study]. 2015 2015-03-01;111(3):285-92.
- 289 4. Sobin L, Gospodarowicz M, Wittekind C. *TNM classification of malignant tumours. 7th edition.*:
290 Wiley-Blackwell; 2009.
- 291 5. Edge S, Byrd D, Compton C, Fritz A, Greene F, Trotti A. *AJCC cancer staging manual. 7th*
292 *edition.*: Springer; 2009.
- 293 6. Hou Y, Wang X, Chen J. Prognostic significance of metastatic lymph node ratio: the lymph node
294 ratio could be a prognostic indicator for patients with gastric cancer. *WORLD J SURG ONCOL.*
295 [Journal Article]. 2018 2018-10-04;16(1):198.
- 296 7. Zhu J, Xue Z, Zhang S, Guo X, Zhai L, Shang S, et al. Integrated analysis of the prognostic role
297 of the lymph node ratio in node-positive gastric cancer: A meta-analysis. *INT J SURG.* [Journal Article;
298 Meta-Analysis; Review]. 2018 2018-09-01;57:76-83.
- 299 8. Sun Z, Xu Y, Li DM, Wang ZN, Zhu GL, Huang BJ, et al. Log odds of positive lymph nodes: a
300 novel prognostic indicator superior to the number-based and the ratio-based N category for gastric
301 cancer patients with R0 resection. *CANCER-AM CANCER SOC.* [Journal Article; Research Support,
302 Non-U.S. Gov't]. 2010 2010-06-01;116(11):2571-80.
- 303 9. Zhao E, Zhou C, Chen S. Prognostic nomogram based on log odds of positive lymph nodes for
304 gastric carcinoma patients after surgical resection. *FUTURE ONCOL.* [Journal Article]. 2019
305 2019-12-01;15(36):4207-22.
- 306 10. Gu P, Deng J, Sun Z, Wang Z, Wang W, Liang H, et al. Superiority of log odds of positive lymph

- 1
2
3 307 nodes (LODDS) for prognostic prediction after gastric cancer surgery: a multi-institutional analysis of
4 308 7620 patients in China. *SURG TODAY*. [Journal Article]. 2021 2021-01-01;51(1):101-10.
- 5
6 309 11. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov)
7 310 SEER*Stat Database: Incidence - SEER Research Data, 18 Registries, Nov 2019 Sub (2000-2017) -
8 311 Linked To County Attributes - Time Dependent (1990-2017) Income/Rurality, 1969-2018 Counties,
9 312 National Cancer Institute, DCCPS, Surveillance Research Program, released April 2020, based on the
10 313 November 2019 submission..
- 11
12 314 12. Yu Y, Zhang P, Yao R, Wang J, Wang P, Xue X, et al. Prognostic value of log odds of positive
13 315 lymph nodes in node-positive lung squamous cell carcinoma patients after surgery: a SEER
14 316 population-based study. *Transl Lung Cancer Res*. [Journal Article]. 2020 2020-08-01;9(4):1285-301.
- 15
16 317 13. Camp RL, Dolled-Filhart M, Rimm DL. X-tile: a new bio-informatics tool for biomarker
17 318 assessment and outcome-based cut-point optimization. *CLIN CANCER RES*. [Journal Article;
18 319 Research Support, Non-U.S. Gov't; Research Support, U.S. Gov't, Non-P.H.S.; Research Support, U.S.
19 320 Gov't, P.H.S.]. 2004 2004-11-01;10(21):7252-9.
- 20
21 321 14. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer
22 322 prognosis. *J CLIN ONCOL*. [Journal Article; Review]. 2008 2008-03-10;26(8):1364-70.
- 23
24 323 15. Jian-Hui C, Shi-Rong C, Hui W, Si-le C, Jian-Bo X, Er-Tao Z, et al. Prognostic value of three
25 324 different lymph node staging systems in the survival of patients with gastric cancer following D2
26 325 lymphadenectomy. *Tumour Biol*. [Journal Article]. 2016 2016-08-01;37(8):11105-13.
- 27
28 326 16. Aurello P, Petrucciani N, Nigri GR, La Torre M, Magistri P, Tierno S, et al. Log odds of positive
29 327 lymph nodes (LODDS): what are their role in the prognostic assessment of gastric adenocarcinoma? *J*
30 328 *GASTROINTEST SURG*. [Comparative Study; Journal Article]. 2014 2014-07-01;18(7):1254-60.
- 31
32 329 17. Kilic MO, Gundogdu SB, Ozden S, Saylam B, Tez M. The prognostic value of different node
33 330 staging systems in patients with ≤ 15 lymph nodes following surgery for gastric adenocarcinoma.
34 331 *ACTA CHIR BELG*. [Journal Article]. 2018 2018-02-01;118(1):1-6.
- 35
36 332 18. Suh YS, Han DS, Kong SH, Lee HJ, Kim YT, Kim WH, et al. Should adenocarcinoma of the
37 333 esophagogastric junction be classified as esophageal cancer? A comparative analysis according to the
38 334 seventh AJCC TNM classification. *ANN SURG*. [Comparative Study; Journal Article; Research
39 335 Support, Non-U.S. Gov't]. 2012 2012-05-01;255(5):908-15.
- 40
41 336 19. Huang Q, Shi J, Feng A, Fan X, Zhang L, Mashimo H, et al. Gastric cardiac carcinomas involving
42 337 the esophagus are more adequately staged as gastric cancers by the 7th edition of the American Joint
43 338 Commission on Cancer Staging System. *Mod Pathol*. [Journal Article; Research Support, Non-U.S.
44 339 Gov't; Validation Study]. 2011 2011-01-01;24(1):138-46.
- 45
46 340 20. Amin M, Edge S, Greene F, Byrd D, Brookland R. *AJCC cancer staging manual*. 8th edition.:
47 341 Springer; 2016.
- 48
49 342 21. Liu K, Feng F, Chen XZ, Zhou XY, Zhang JY, Chen XL, et al. Comparison between gastric and
50 343 esophageal classification system among adenocarcinomas of esophagogastric junction according to
51 344 AJCC 8th edition: a retrospective observational study from two high-volume institutions in China.
52 345 *GASTRIC CANCER*. [Comparative Study; Journal Article; Observational Study]. 2019
53 346 2019-05-01;22(3):506-17.
- 54
55 347 22. Karstens KF, Ghadban T, Sawez S, Konczalla L, Woestemeier A, Bachmann K, et al. Comparison
56 348 of the 8th UICC staging system for esophageal and gastric cancers in Siewert type II junctional
57 349 adenocarcinomas. *Eur J Surg Oncol*. [Comparative Study; Journal Article]. 2020 2020-04-01;46(4 Pt
58 350 A):638-43.

- 1
2
3 351 23. Rizk NP, Ishwaran H, Rice TW, Chen LQ, Schipper PH, Kesler KA, et al. Optimum
4 352 lymphadenectomy for esophageal cancer. *ANN SURG*. [Journal Article; Research Support, Non-U.S.
5 353 Gov't]. 2010 2010-01-01;251(1):46-50.
- 7 354 24. Macalindong SS, Kim KH, Nam BH, Ryu KW, Kubo N, Kim JY, et al. Effect of total number of
8 355 harvested lymph nodes on survival outcomes after curative resection for gastric adenocarcinoma:
9 356 findings from an eastern high-volume gastric cancer center. *BMC CANCER*. [Journal Article;
11 357 Research Support, Non-U.S. Gov't]. 2018 2018-01-12;18(1):73.
- 12 358 25. Schoenleber SJ, Schnelldorfer T, Wood CM, Qin R, Sarr MG, Donohue JH. Factors influencing
13 359 lymph node recovery from the operative specimen after gastrectomy for gastric adenocarcinoma. *J*
14 360 *GASTROINTEST SURG*. [Journal Article]. 2009 2009-07-01;13(7):1233-7.
- 16 361 26. Khanjani N, Mirzaei S, Nasrolahi H, Hamed SH, Mosalaei A, Omidvari S, et al. Insufficient
17 362 lymph node assessment in gastric adenocarcinoma. *J Egypt Natl Canc Inst*. [Journal Article]. 2019
18 363 2019-10-22;31(1):2.
- 20 364 27. Pei JP, Zhang CD, Fan YC, Dai DQ. Comparison of Different Lymph Node Staging Systems in
21 365 Patients With Resectable Colorectal Cancer. *FRONT ONCOL*. [Journal Article]. 2018
22 366 2018-01-20;8:671.
- 24 367 28. Cao J, Yuan P, Ma H, Ye P, Wang Y, Yuan X, et al. Log Odds of Positive Lymph Nodes Predicts
25 368 Survival in Patients After Resection for Esophageal Cancer. *ANN THORAC SURG*. [Journal Article;
26 369 Multicenter Study]. 2016 2016-08-01;102(2):424-32.
- 28 370 29. Jin W, Zhu Z, Wu Y, Ding X, Wu H, Song X, et al. Prognostic value of log odds of positive
29 371 lymph nodes in patients with resectable oral squamous cell carcinoma. *ORAL ONCOL*. [Journal
30 372 Article; Research Support, Non-U.S. Gov't]. 2020 2020-09-01;108:104709.
- 32 373 30. Li J, Sun Y, Zhao B, Tang C, Fan D, Jiang W, et al. Lymph Node Ratio-Based Staging System for
33 374 Gallbladder Cancer With Fewer Than Six Lymph Nodes Examined. *FRONT ONCOL*. [Journal
34 375 Article]. 2020 2020-01-20;10:542005.
- 36 376 31. Spillane AJ, Cheung BL, Winstanley J, Thompson JF. Lymph node ratio provides prognostic
37 377 information in addition to american joint committee on cancer N stage in patients with melanoma, even
38 378 if quality of surgery is standardized. *ANN SURG*. [Journal Article]. 2011 2011-01-01;253(1):109-15.
- 39 379 32. Rosa F, Tortorelli AP, Alfieri S, Doglietto GB. Lymph node ratio for gastric cancer: useful
40 380 instrument or just an expedient to retrieve fewer lymph nodes? *ANN SURG*. [Letter; Comment]. 2014
41 381 2014-04-01;259(4):e65.
- 43 382 33. van Hagen P, Hulshof MC, van Lanschot JJ, Steyerberg EW, van Berge HM, Wijnhoven BP, et al.
44 383 Preoperative chemoradiotherapy for esophageal or junctional cancer. *N Engl J Med*. [Journal Article;
45 384 Multicenter Study; Randomized Controlled Trial; Research Support, Non-U.S. Gov't]. 2012
46 385 2012-05-31;366(22):2074-84.
- 48 386 34. Ajani JA, D'Amico TA, Bentrem DJ, Chao J, Corvera C, Das P, et al. Esophageal and
49 387 Esophagogastric Junction Cancers, Version 2.2019, NCCN Clinical Practice Guidelines in Oncology. *J*
50 388 *Natl Compr Canc Netw*. [Journal Article]. 2019 2019-07-01;17(7):855-83.

389

FIGURE LEGENDS

390 Figure 1. Flow diagram of the patient selection and grouping.

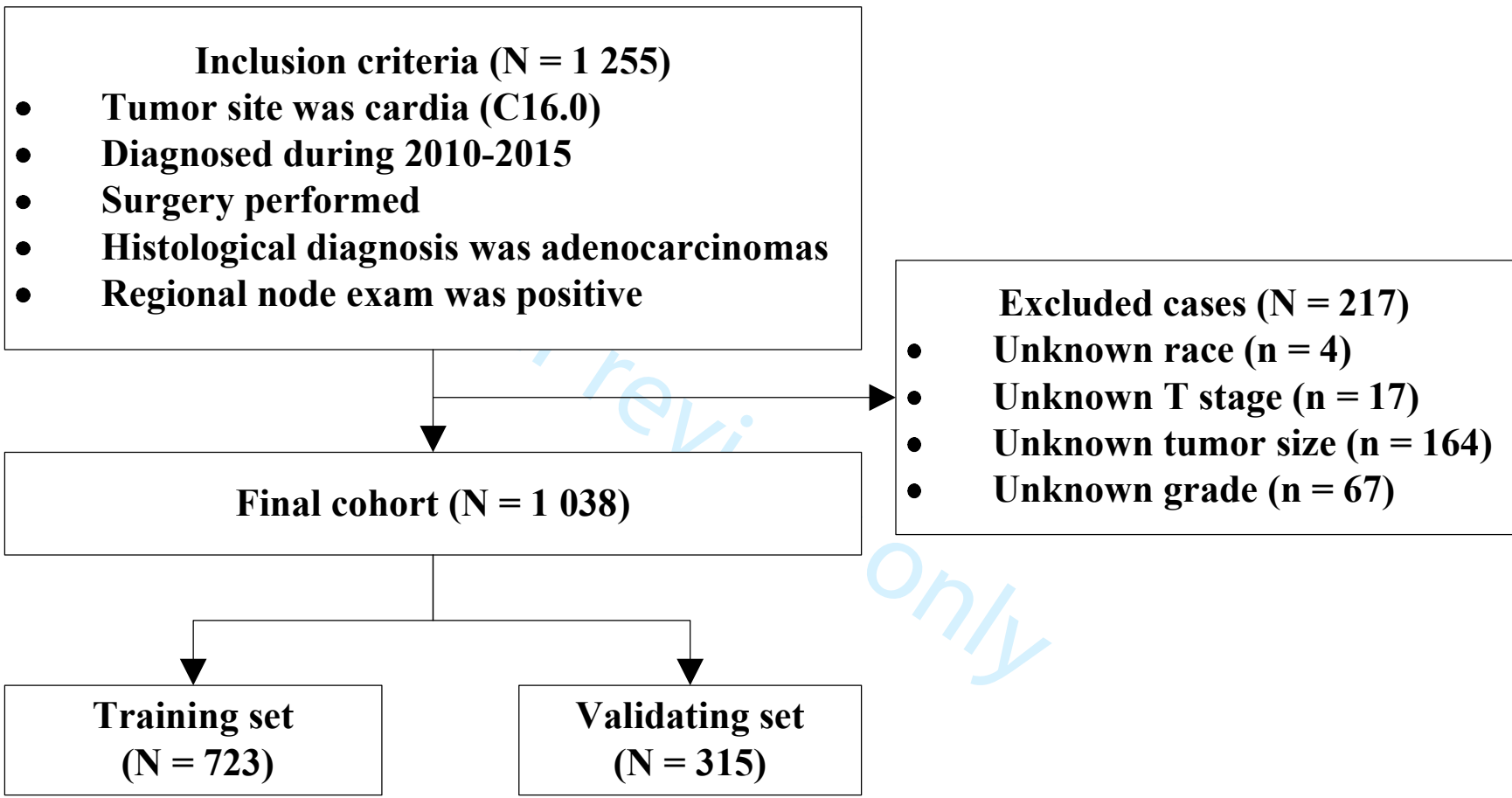
391 Figure 2. Survival curves of the training and validating sets by different staging systems.

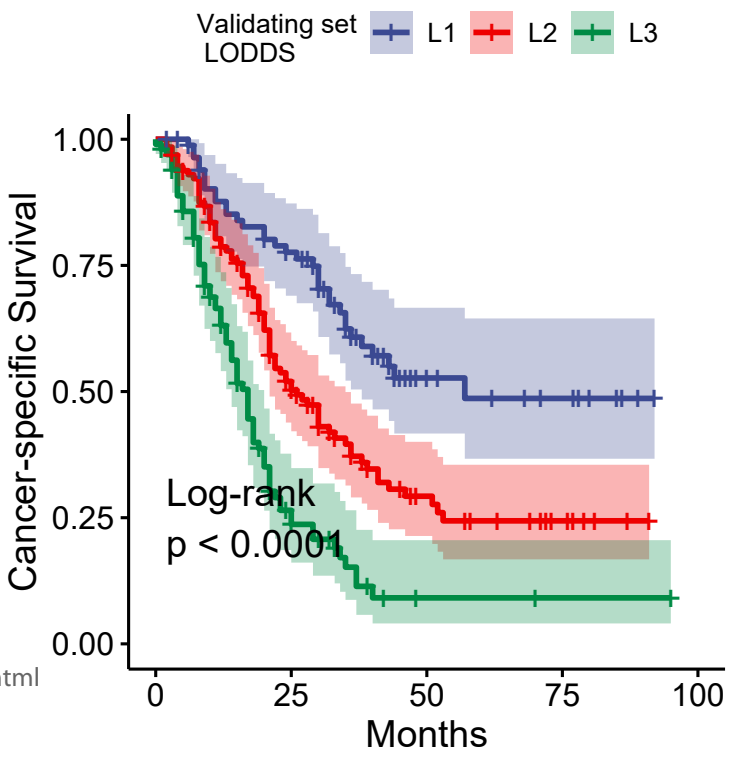
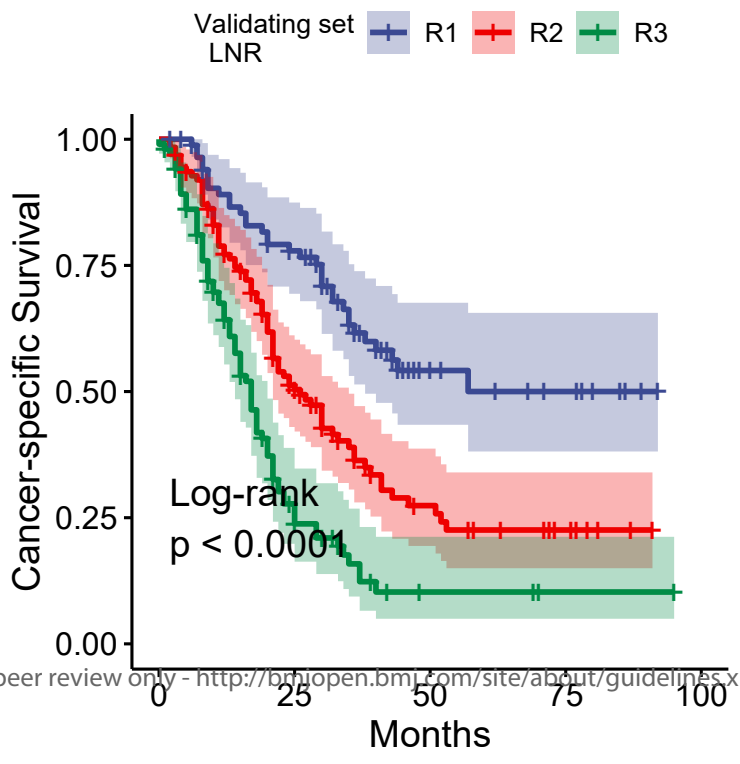
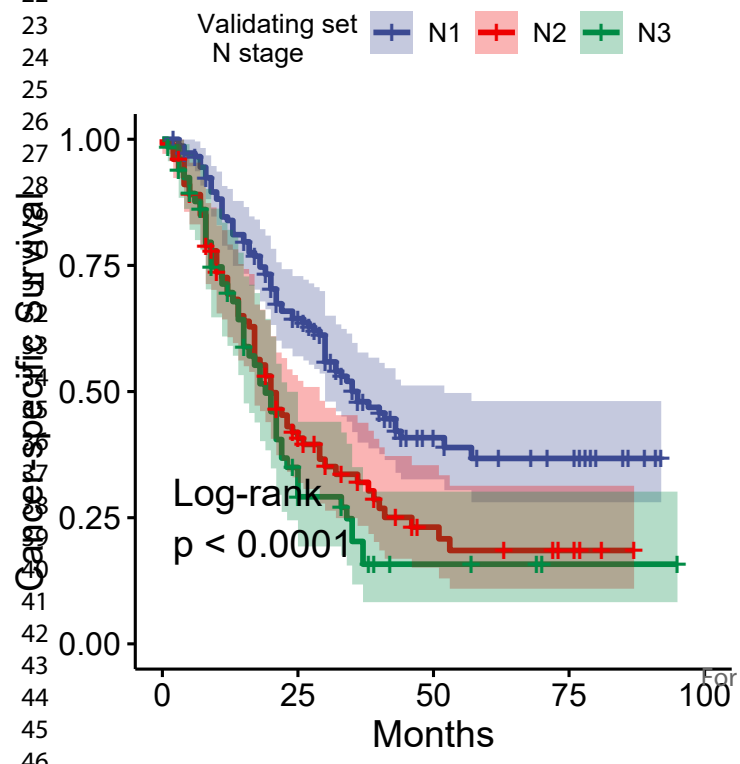
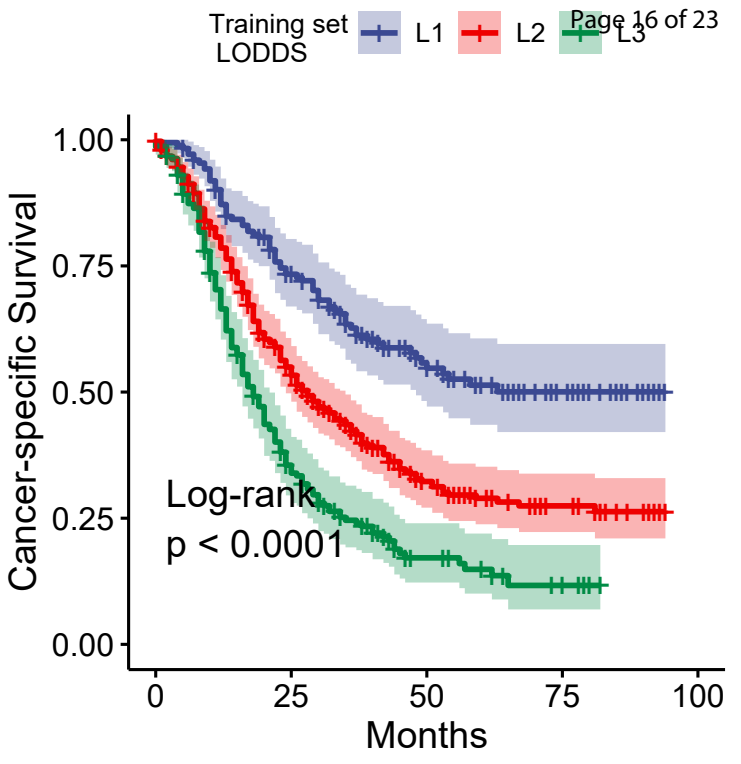
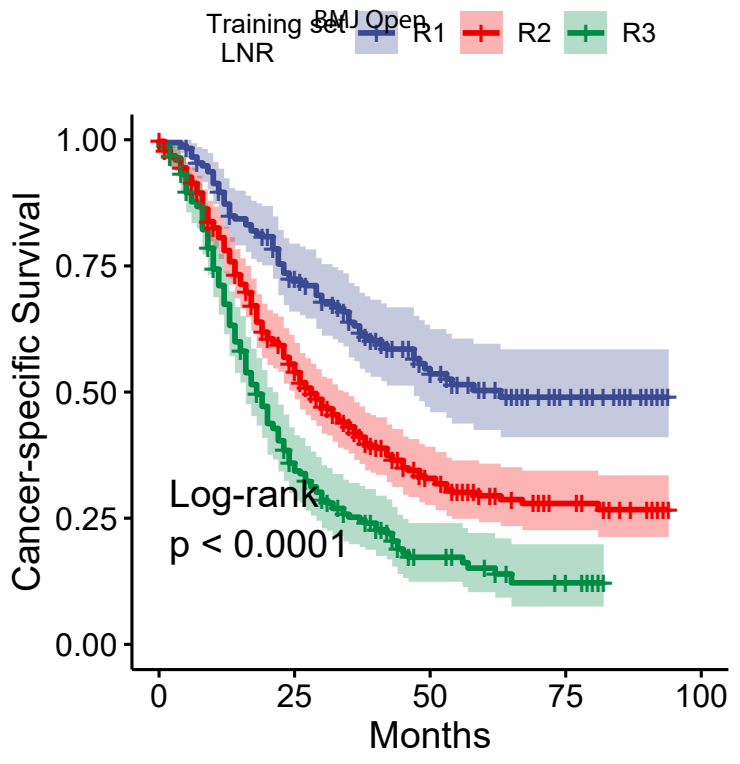
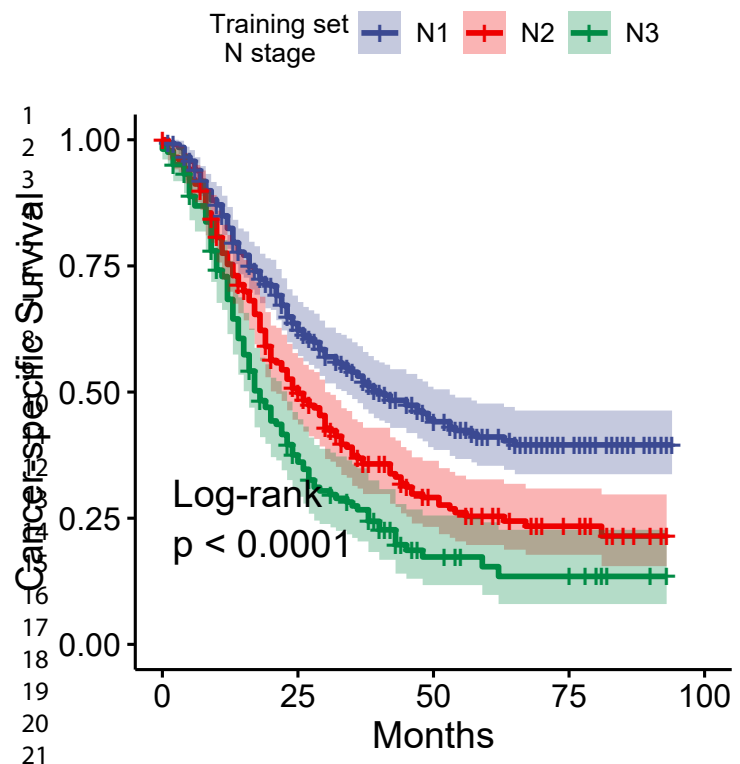
392 Figure 3. Construction of nomogram based on Tumor-Lymph node ratio-Metastasis staging system
393 and calibration plots for the nomogram.
394

1
2
3 395 Suppl. Figure 1. Kaplan Meier plots for proportional hazard assumption test in models that were
4 396 based on N stage, lymph node ratio, and log odds of positive lymph nodes.
5
6 397 Suppl. Figure 2. Construction of nomogram based on Tumor-Log odds of positive lymph nodes
7 398 -Metastasis staging system and calibration plots for the nomogram.
8 399 Suppl. Figure 3. Receiver operator characteristic curves and decision curve analysis plots for
9 400 comparison of the prediction powers of the different models.
10
11 401
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

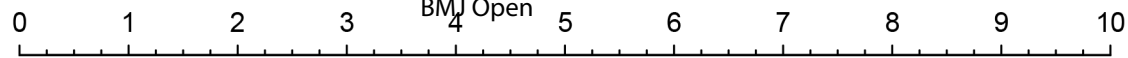
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46



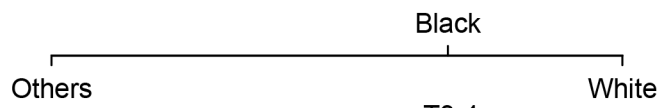


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

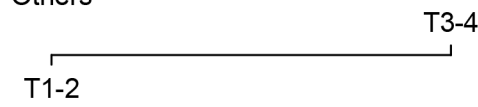
Points



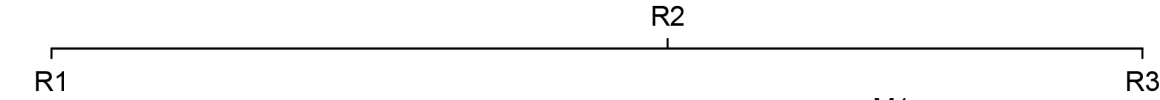
Race



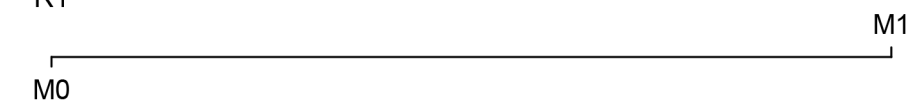
T.stage



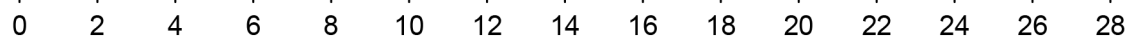
LNR



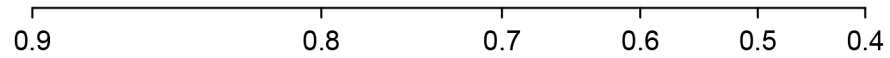
M.stage



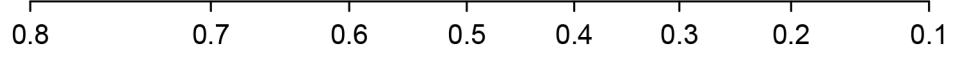
Total Points



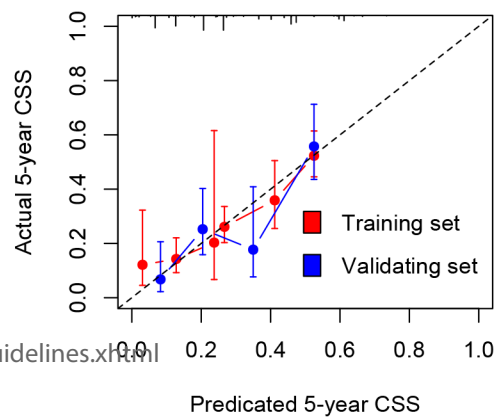
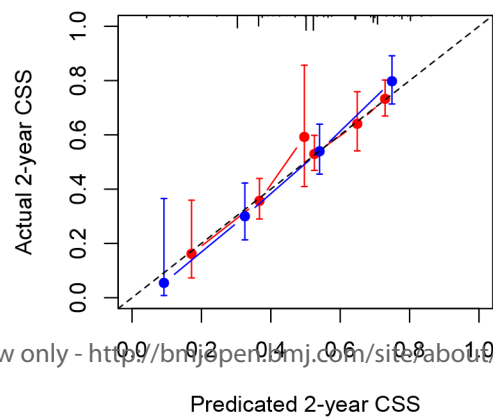
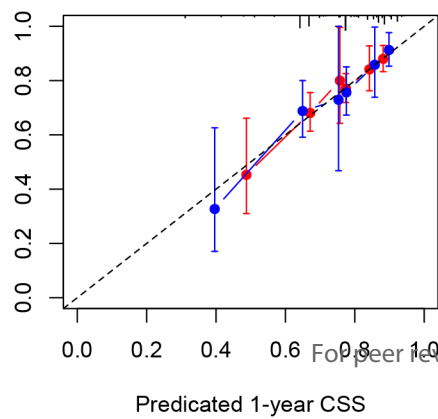
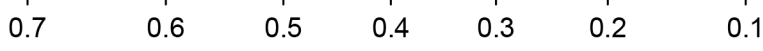
1-year CSS

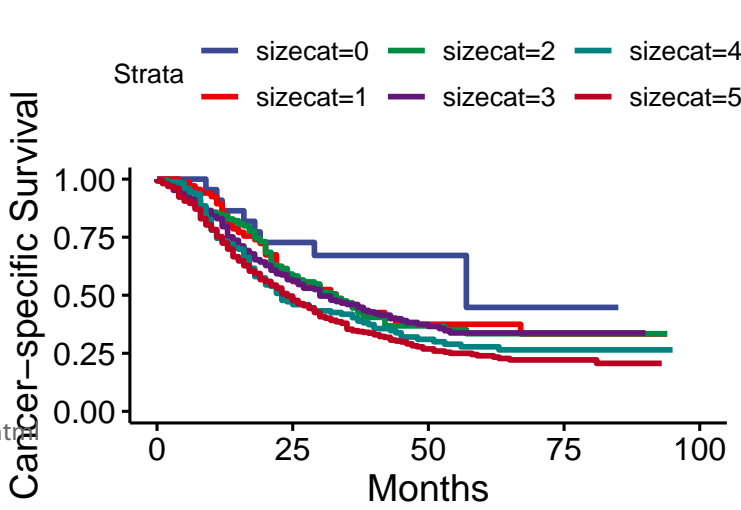
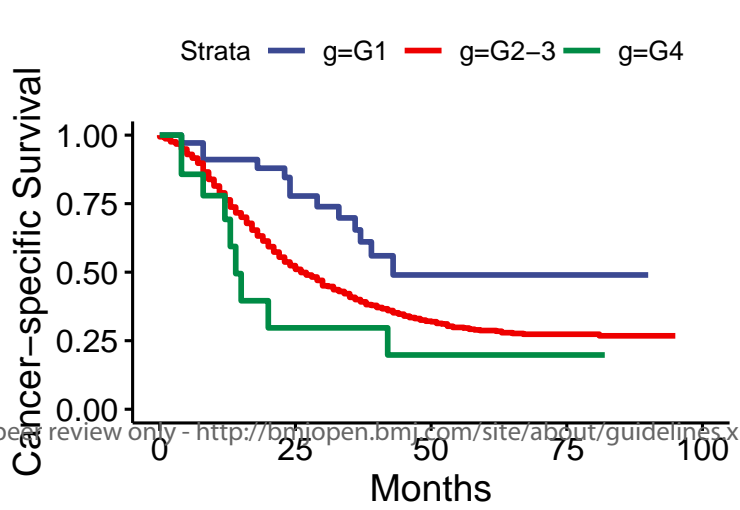
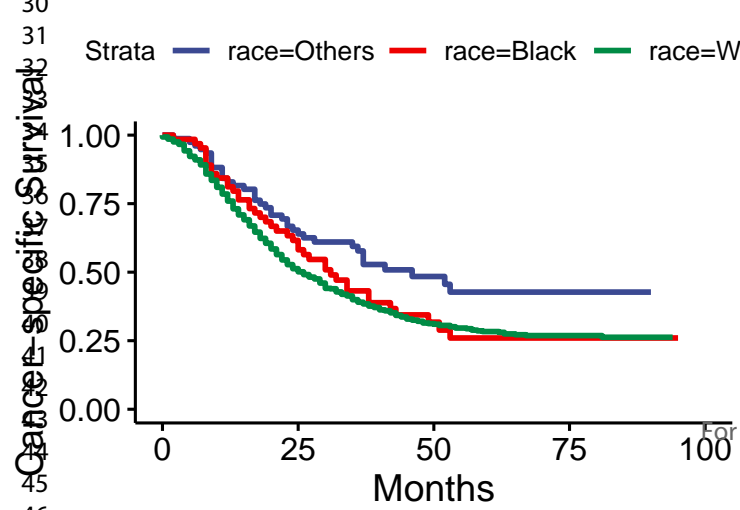
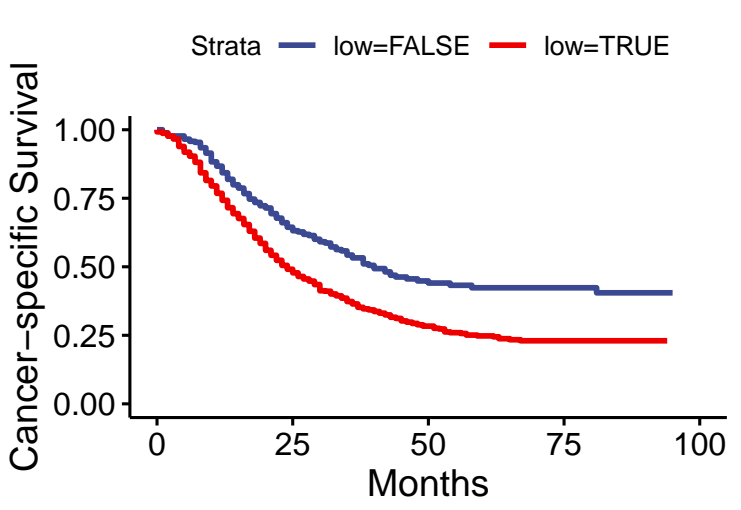
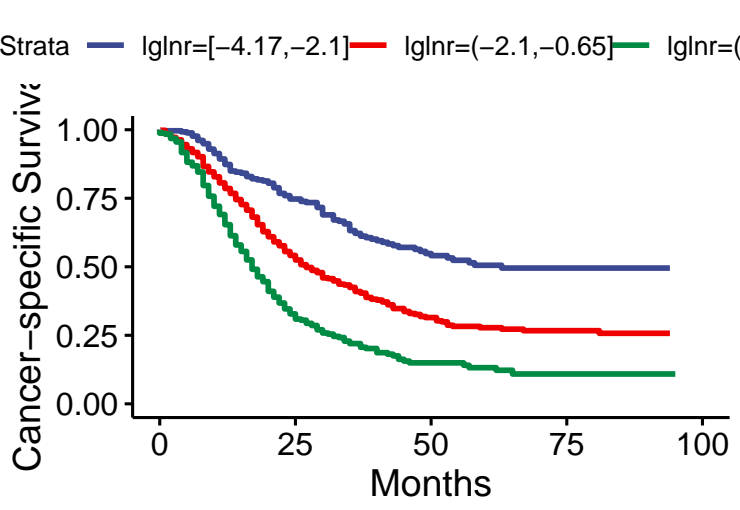
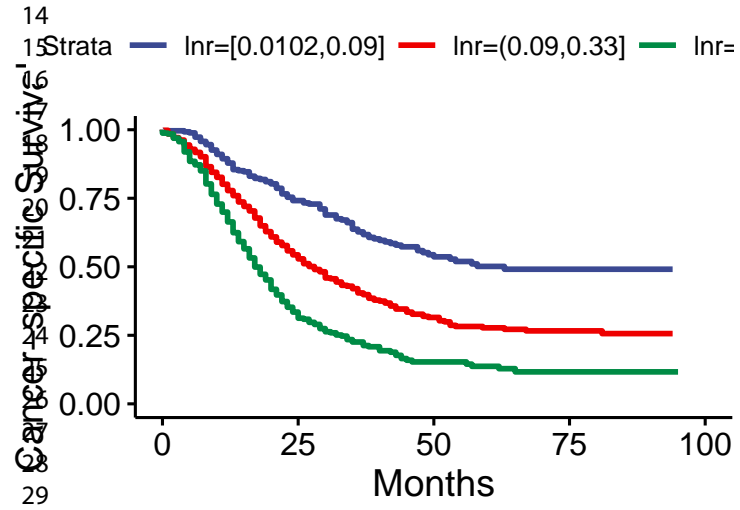
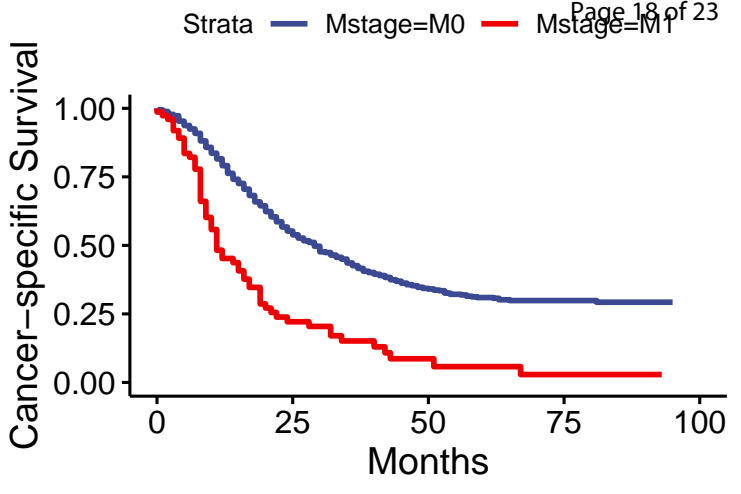
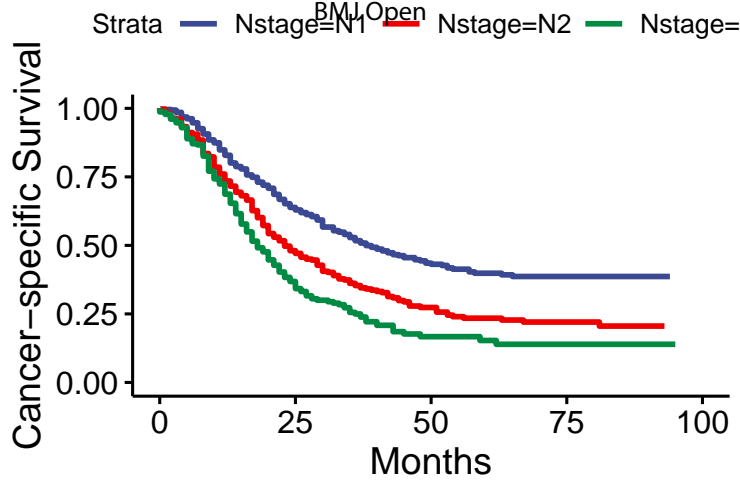
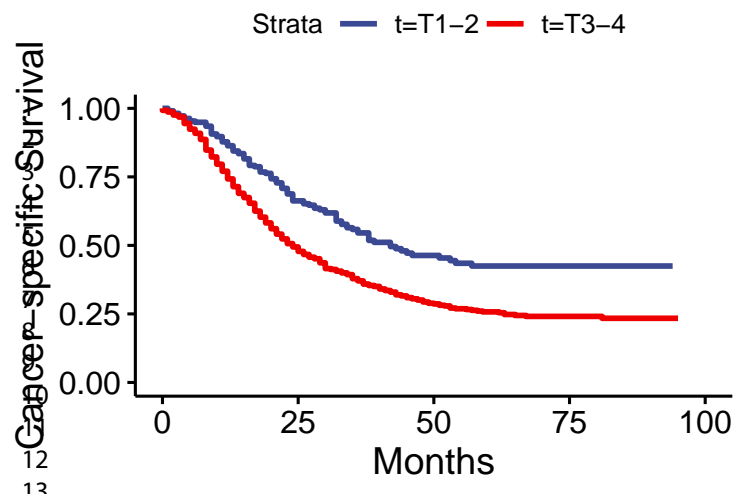


2-year CSS

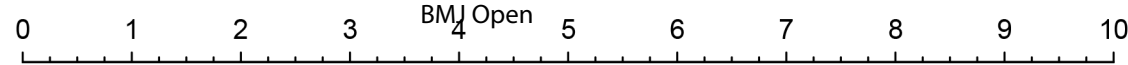


5-year CSS

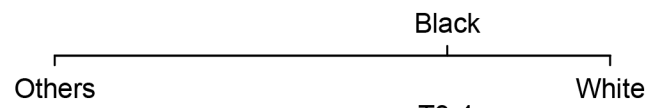




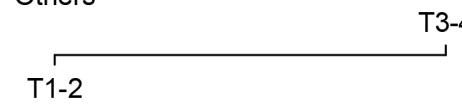
Points



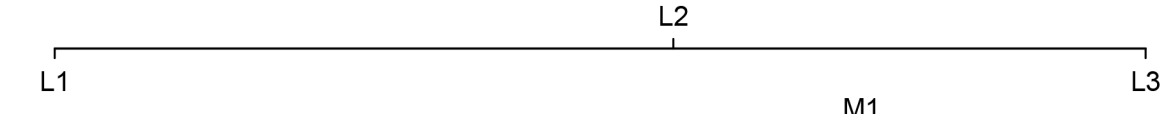
Race



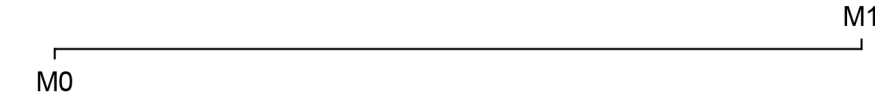
T.stage



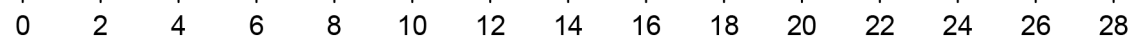
LODDS



M.stage



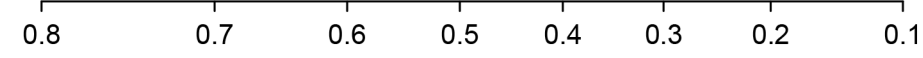
Total Points



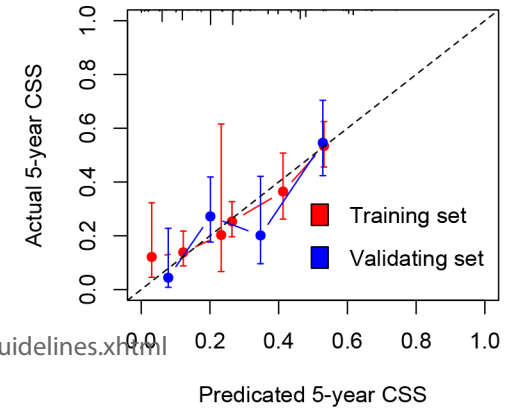
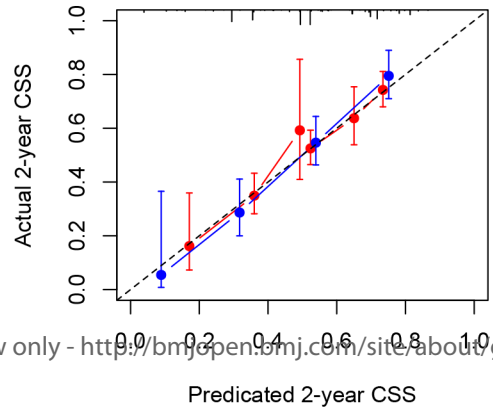
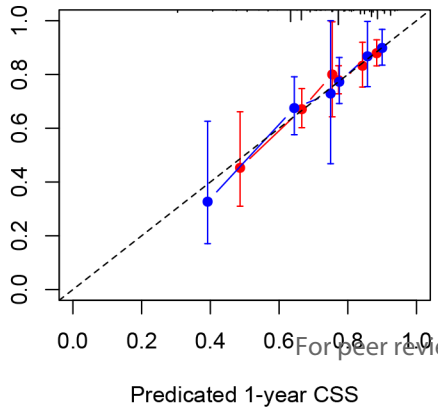
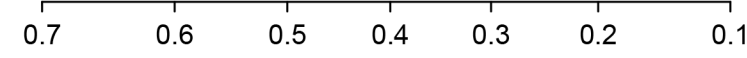
1-year CSS



2-year CSS

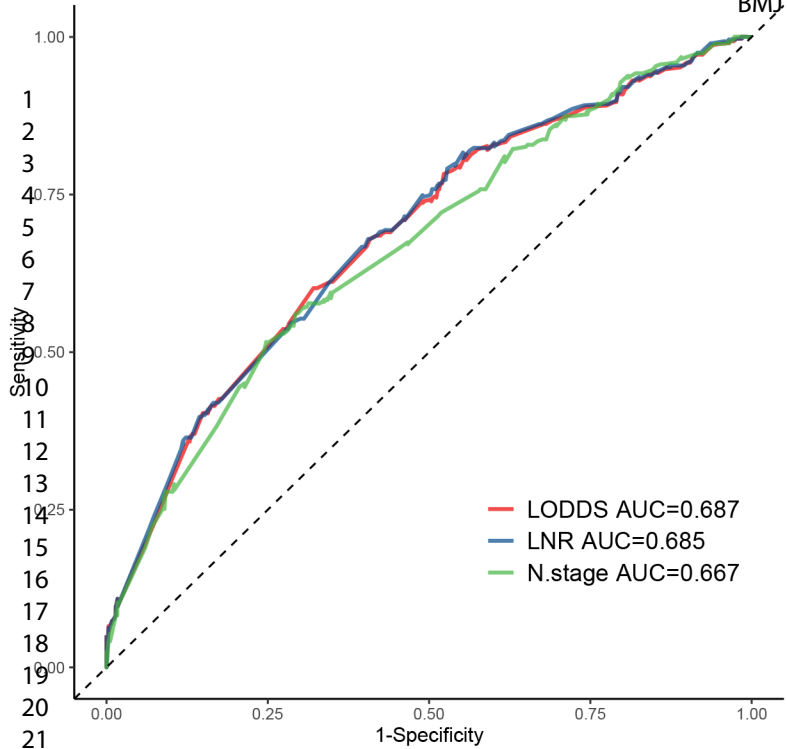


5-year CSS

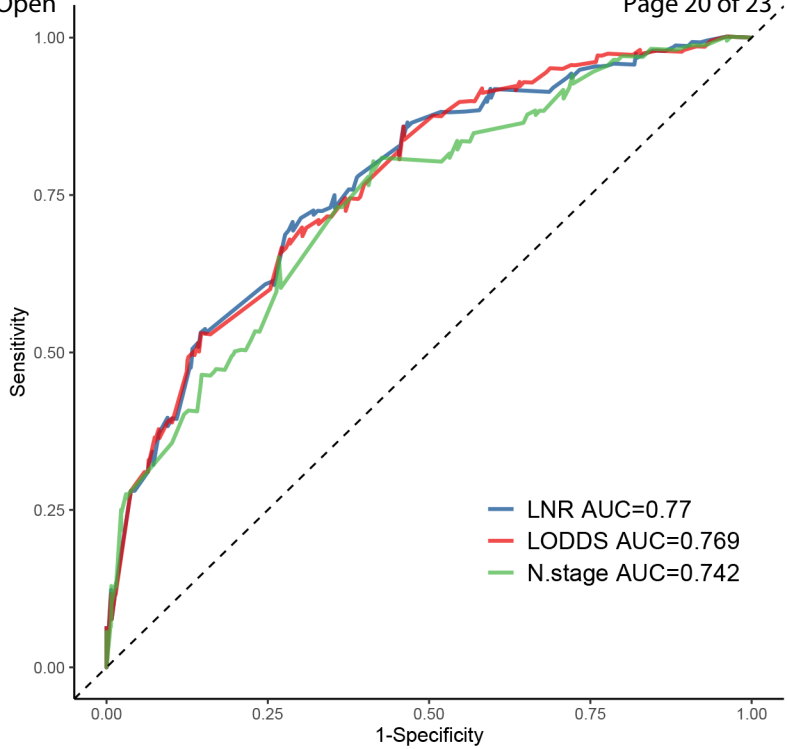


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41

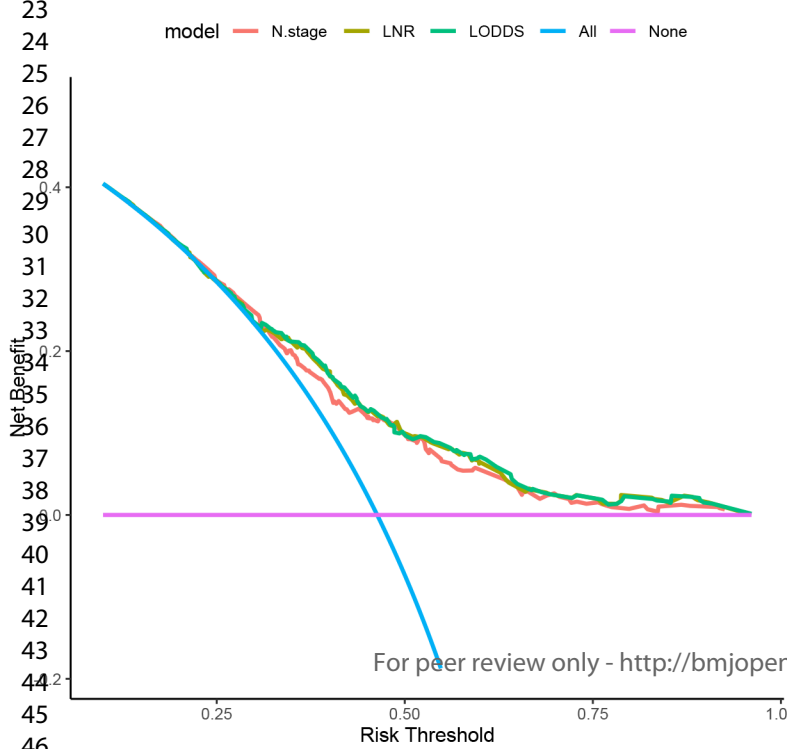
Training set



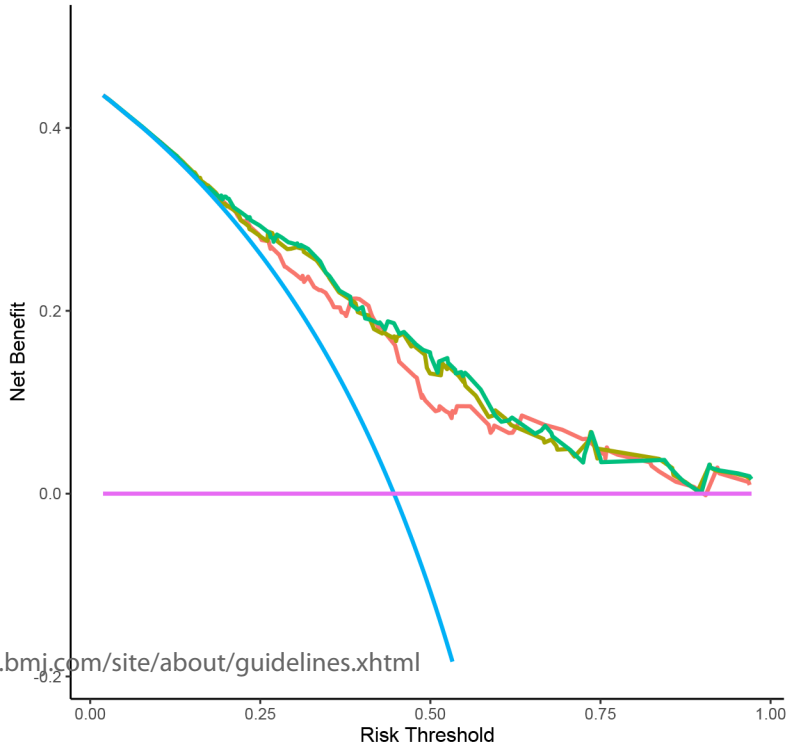
Validating set



Training set



Validating set



Supplementary Table 1. Comparison of the discrimination ability of different models based on different cutoff values.

	Crude model			Adjusted model	
	HR (95% CI)	C-index	AIC	HR (95% CI)	C-index
LNR					
Cutoff_1		0.607	5376	0.643	5350
<0.09	1 (ref)			1 (ref)	
0.09~0.33	1.88 (1.44, 2.44)			1.74 (1.33, 2.29)	
>0.33	3.02 (2.30, 3.97)			2.63 (1.97, 3.50)	
Cutoff_2		0.605	5378	0.641	5355
<0.09	1 (ref)			1 (ref)	
0.09~0.40	1.97 (1.52, 2.54)			1.85 (1.38, 2.54)	
>0.40	3.16 (2.38, 4.21)			2.72 (2.02, 3.67)	
LODDS					
Cutoff_1		0.609	5373	0.644	5346
<-2.09	1 (ref)			1 (ref)	
-2.09~-0.65	1.93 (1.48, 2.51)			1.80 (1.36, 2.37)	
>-0.65	3.13 (2.38, 4.13)			2.77 (2.07, 3.70)	
Cutoff_2		0.605	5378	0.640	5352
<-2.10	1 (ref)			1 (ref)	
-2.09~-0.37	2.00 (1.54, 2.59)			1.86 (1.42, 2.44)	
>-0.37	3.26 (2.45, 4.33)			2.83 (2.10, 3.81)	

Cutoff_1 was generated by minimal p-value method via X-tile software, cutoff_2 was generated using P25 and P75.

Supplementary Table 2. Univariate analysis of the potential confounders.

Variables	HR (95% CI)	P-value
Age (increased by 10ys)	1.06 (0.98 – 1.14)	0.065
Tumor size (increased by 1cm)	1.13 (1.06 – 1.19)	<0.001
Sex		
Male	1	
Female	1.09 (0.89 – 1.34)	0.386
Race		
Others	1	
Black	1.42 (0.91 – 2.21)	0.118
White	1.57 (1.13 – 2.19)	0.007
Grade		
G1	1	
G2-3	2.08 (1.20 – 3.60)	0.009
G4	3.08 (1.32 – 7.22)	0.009
T		
T1-2	1	
T3-4	1.70 (1.38 – 2.10)	<0.001
N		
N1	1	
N2	1.61 (1.34 – 1.93)	<0.001
N3	2.15 (1.77 – 2.62)	<0.001
M		
M0	1	
M1	2.61 (2.00 – 3.39)	<0.001
Low nodes yield		
No	1	
Yes	1.64 (1.35 – 1.99)	<0.001

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the abstract title page (b) Provide in the abstract an informative and balanced summary of what was done and what was found page2
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported page 3
Objectives	3	State specific objectives, including any prespecified hypotheses page3
Methods		
Study design	4	Present key elements of study design early in the paper page 3
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection page 3-4
Participants	6	(a) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up page 4 <i>Case-control study</i> —Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls NA <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants NA (b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed NA <i>Case-control study</i> —For matched studies, give matching criteria and the number of controls per case NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable page 4
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group page 3-4
Bias	9	Describe any efforts to address potential sources of bias page 4
Study size	10	Explain how the study size was arrived at NA
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why page 4
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding page 4 (b) Describe any methods used to examine subgroups and interactions page 4 (c) Explain how missing data were addressed page 4 (d) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed NA <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed NA <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy NA (e) Describe any sensitivity analyses NA
Results		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible,

		examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed page 5
		(b) Give reasons for non-participation at each stage page 5
		(c) Consider use of a flow diagram figure 1, page 3
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders page 5-6
		(b) Indicate number of participants with missing data for each variable of interest figure 1, page 3
		(c) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount) Table 1 page 5-6
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time page 5
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure NA
		<i>Cross-sectional study</i> —Report numbers of outcome events or summary measures NA
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included table 1, page 5-6
		(b) Report category boundaries when continuous variables were categorized page 6
		(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses NA
Discussion		
Key results	18	Summarise key results with reference to study objectives page 7-8
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias page 9
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence page 8-9
Generalisability	21	Discuss the generalisability (external validity) of the study results page 9
Other information		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based page 10

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at www.strobe-statement.org.