ABSTRACT

Objective Bleeding after endoscopic submucosal dissection (ESD) for early gastric cancer (EGC) is a frequent adverse event after ESD. We aimed to develop and externally validate a clinically useful prediction model (BEST-J score: Bleeding after ESD Trend from Japan) for bleeding after ESD for EGC.

Design This retrospective study enrolled patients who underwent ESD for EGC. Patients in the derivation cohort (n=8291) were recruited from 25 institutions, and patients in the external validation cohort (n=2029) were recruited from eight institutions in other areas. In the derivation cohort, weighted points were assigned to predictors of bleeding determined in the multivariate logistic regression analysis and a prediction model was established. External validation of the model was conducted to analyse discrimination and calibration.

Results A prediction model comprised 10 variables (warfarin, direct oral anticoagulant, chronic kidney disease with haemodialysis, P2Y12 receptor antagonist, aspirin, cilostazol, tumour size >30 mm, lower-third in tumour location, presence of multiple tumours and interruption of each kind of antithrombotic agent). The rates of bleeding after ESD at low-risk (0 to 1 points), intermediate-risk (2 points), high-risk (3 to 4 points) and very high-risk (≥5 points) were 2.8%, 6.1%, 11.4% and 29.7%, respectively. In the external validation cohort, the model showed moderately good discrimination, with a c-statistic of 0.70 (95% CI, 0.64 to 0.76), and good calibration (calibration-in-the-large, 0.05; calibration slope, 1.01).

Conclusions In this nationwide multicentre study, we derived and externally validated a prediction model for bleeding after ESD. This model may be a good clinical decision-making support tool for ESD in patients with EGC.

INTRODUCTION

Gastric cancer is the fifth most common cancer and 50% to 70% of gastric cancers in Eastern Asian
countries are detected at an early stage.\textsuperscript{1–4} Endoscopic submucosal dissection (ESD) is a minimally invasive treatment method for early gastric cancer (EGC) with almost no risk of lymph node metastasis. However, bleeding is one of the most frequent adverse events in ESD.\textsuperscript{3} Specifically, bleeding after ESD occurs in 4.1\% to 8.5\% of patients undergoing gastric ESD.\textsuperscript{6–11}

Thus far, many factors, including male sex, comorbidities, tumour characteristics and procedure time, have been identified as high-risk factors for bleeding after gastric ESD, and among them, antithrombotic (AT) agents are regarded as an important risk factor for bleeding.\textsuperscript{6, 7, 12, 13} AT agents include antiplatelet agents (APAs) (aspirin, P2Y12 receptor antagonist (P2Y12RA), cilostazol) and anticoagulants (ACs) (warfarin, direct oral AC (DOAC)), and because of their different mechanisms of action, these agents may carry different risks for bleeding after ESD. However, few studies have appropriately evaluated the risk of bleeding after ESD for each type of agent mainly because of the small number of cases. Furthermore, according to the recent Japanese guidelines for endoscopy with AT agents,\textsuperscript{14} aspirin and cilostazol can be continued or interrupted, P2Y12RA can be interrupted or replaced with aspirin or cilostazol and double antiplatelet therapy can be changed to monotherapy with aspirin or cilostazol. However, no studies have that evaluated the risk factors for bleeding have considered these factors.

Recently, clinical prediction models have been used in several fields to provide an estimate of the value of a therapy for an individual patient.\textsuperscript{15–17} However, no prediction models have been reported for bleeding after ESD for EGC. The haemostatic procedure for preventing bleeding in ESD sometimes differs among institutions with/without familiarity with ESD and among different geographical areas. Thus, it is desirable to develop a prediction model in a large-scale study that includes patients in various areas. Therefore, we aimed to derive and externally validate a clinically useful prediction model for bleeding after ESD for EGC based on a nationwide multicentre study in Japan.

**METHODS**

This study was conducted in accordance with the guidelines of the Declaration of Helsinki and was approved by the Ethics Committee of Tohoku University Graduate School of Medicine (2018-1-48) followed by the institutional review board of each institution before the recruitment of patients. Written informed consent for ESD was obtained from all patients before the procedure. The need for informed consent of this study was waived via the opt-out method on each participating hospital website.

**Study design and patient population**

To establish a reliable prediction model for bleeding after ESD for EGC, we conducted this retrospective study consisting of three steps: (1) identification of clinically significant predictors of bleeding after ESD for EGC using a large cohort, (2) development of a simple prediction model for predicting bleeding and (3) external validation of the model using an independent data set. This study followed the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guidelines.\textsuperscript{18}

The study population consisted of adult patients who underwent ESD for EGC. Patients in the derivation cohort were recruited from 25 institutions throughout Japan, except from the northern (Hokkaido and Tohoku) and southern (Kyushu and Shikoku) areas, between November 2013 and October 2016 (online supplementary figure 1). The patients were excluded if (1) ESD was not completed, (2) the follow-up duration after ESD was <28 days, (3) closure of the ulcer or the shielding method using polyglycolic acid sheets and fibrin glue was performed after ESD, (4) they refused use of their clinical data, (5) they underwent photodynamic therapy after ESD, (6) invasion of the muscularis propria or a deeper layer was pathologically confirmed or (7) ESD was performed on the remnant stomach. An external validation was conducted in an independent cohort of patients from eight institutions, which are located in the north or south of Japan (online supplementary figure 1) using these same inclusion and exclusion criteria.

**ESD procedure and pathological evaluation**

ESD was performed according to a standard ESD procedure\textsuperscript{19} at all institutions. Briefly, the procedure consisted of the following steps: (1) marking around the lesion, (2) submucosal injection of a solution, (3) mucosal incision outside of the marked region, (4) submucosal dissection and (5) retrieval of the specimen. Haemostasis of active bleeding and prophylactic coagulation of visible vessels on the ESD ulcer were performed using haemostatic forceps or clips during the procedure.

Scheduled second-look endoscopy (SLE) depended on the institution. During and after ESD, proton pump inhibitor, potassium-competitive acid blocker or H2 receptor antagonists were administered to all patients according to the selection of the doctors in charge. In those with intake of AT agents, the decision on continuing or interrupting the agents before ESD and, where present, the timing of interruption, and the use of heparin bridging were mainly based on the Japanese guidelines,\textsuperscript{14} which have been available in Japan since 2012.

**Candidates for predictive factors of bleeding after ESD for EGC**

On the basis of a comprehensive literature review, including a prior systematic review,\textsuperscript{20} and clinical relevance, we identified that the following variables were potentially associated with bleeding after ESD: sex, comorbidities (ischaemic heart disease, liver cirrhosis and chronic kidney disease (CKD) with haemodilysis), AT therapy (aspirin, P2Y12RA, cilostazol, warfarin and DOAC), the tumour characteristics (multiple tumours, lower-third in tumour location, tumour size >30 mm, undifferentiated type, submucosal invasion ≥500 mm from the muscularis mucosa (SM2) and ulceration) and procedure (ESD procedure time >120 min and SLE). We also selected characteristics that may influence bleeding, including age (≥75 years), piecemeal resection and mild/no endoscopic gastric atrophy according to the Kimura and Takemoto classification.\textsuperscript{21} With regard to the AT agents, the variables for the interruption of AT agents, heparin bridging and replacement of APAs were included in the model because of their possible influence on bleeding.

**Data collection and definition**

In each participating institution, demographic and clinical data were collected from the medical records and endoscopy and pathology reports.
In the subgroup analysis, the derived prediction model was evaluated for each centre-volume, which was divided into very high-volume (≥150 eligible patients per year), high-volume (≥100 and <150), intermediate-volume (≥50 and <100), and low-volume (<50) centres.

Statistical analysis
Categorical variables were summarised as n (%) and were compared using χ² tests. Continuous variables were summarised as medians and 25th to 75th percentiles, which were reported as P25 to P75, and were compared using the Wilcoxon rank-sum test.

In the derivation cohort, univariate and multivariate logistic regression analyses were used to test the associations of 23 candidate variables with bleeding after ESD. Multicollinearity among the variables was tested using the variance inflation factor (VIF). A cut-off value for inclusion of the assessed factors in the final prediction model was predefined as p<0.05. We assigned weighted points proportional to β regression coefficient values multiplied by 2 to the nearest integer for the factors determined in the multivariate analysis. Using these scores, we derived four risk categories, based on the risk of bleeding in previous reports.6–11,27 Low (<4.1% risk), intermediate (≥4.1% and ≤8.5% risk), high (≥8.5% and ≤15.9% risk) and very high (>15.9% risk). The trend in the risk for bleeding among the four risk groups was evaluated using the Cochran-Armitage trend test. The model’s overall performance was tested using Brier scores and Nagelkerke R², and its discrimination was examined using the c-statistic. The derived prediction model was internally validated by bootstrap resampling, which used 1000 random samples drawn with replacement from the original derivation data set.29,28 Decision curve analysis (DCA), as proposed by Vickers and Elkin,29 was performed to assess the net clinical benefit of the models for predicting bleeding after ESD.

External validation was conducted in the validation cohort to analyse the same performance metrics, including overall performance and discrimination. According to the proposal for better clinical prediction model,31 model calibration was tested using the calibration-in-the-large and calibration slope.

In the subgroup analysis, discrimination and calibration were tested separately in very high-volume, high-volume, intermediate-volume and low-volume centres, using the derivation cohort in order to take advantage of its larger sample size. Patients with missing data for variables included in the final model were excluded from the final derivation and validation cohorts. Data were analysed using SPSS V23.0 for Windows software (IBM Corp, Armonk, New York, USA) and R software V3.5.1 (The R Foundation for Statistical Computing, Vienna, Austria). A p value of <0.05 was considered statistically significant. All statistical analyses were performed by an independent statistician (TN). All authors had access to the study data and had reviewed and approved the final manuscript.

RESULTS
Study flow diagram and patient demographics
The flowchart of patient enrolment is shown in Figure 1. A total of 8291 patients were analysed as the derivation cohort. For the external validation cohort, independent 2029 patients met the eligibility criteria for this study. The patient demographics are shown in Table 1, and the number of enrolled patients at each institution is described in online supplementary table 1. Several baseline characteristics, such as age and cilostazol intake, were different between the derivation and validation cohorts. The
bleeding rates after ESD in the two cohorts were 4.7% (387 events) and 5.0% (102 events), respectively.

Model derivation
We included 23 candidate predictors in the univariate and multivariate models. Online supplementary table 2 shows the results of the univariate analyses for each candidate. There was no interaction between any two AT agents (online supplementary table 3). In the multivariate model, 10 variables were identified as predictors of bleeding (table 2). The VIF (1.00 to 4.58) was less than 5 to 10, indicating that the multicollinearity was not present.\(^{13}\) On the basis of the adjusted \(\beta\) regression coefficient, the scores of selected predictors were assigned an integer score: 4 points each for warfarin and DOAC; 3 points for CKD with haemodialysis; 2 points each for P2Y12RA and aspirin; 1 point each for cilostazol, a tumour size >30mm, lower-third in tumour location and presence of multiple tumours and −1 point for interruption of each kind of AT agents (table 2). Then, we established a prediction model (BEST-J score: Bleeding after ESD Trend from Japan; table 3). The prediction model had good overall performance, with a Nagelkerke \(R^2\) of 0.11 and a Brier score of 0.04, and showed moderately good discrimination, with a \(c\)-statistic of 0.71 (95% CI, 0.68 to 0.73; table 4). A cut-off of 1 point had high sensitivity for identifying patients with bleeding and a cut-off of 3 to 8 points had high specificity (online supplementary table 4). DCA revealed that this prediction model is useful for threshold probabilities between 3% and 50% (online supplementary figure 2). The model was then internally validated by bootstrap resampling procedure, which gave a mean \(c\)-statistic of 0.71 (95% CI, 0.68 to 0.73).

According to the definition, the total score was categorised as low-risk (0 to 1 points), intermediate-risk (2 points), high-risk (3 to 4 points) or very high-risk (≥5 points) for bleeding after ESD. As a result, the rates of bleeding for each risk category were 2.8%, 6.1%, 11.4% and 29.7% (table 3), and a significantly increasing trend of risk from low-risk to very high-risk groups was observed (p<0.001, Cochran-Armitage trend test). The \(c\)-statistic of the risk classification was 0.68 (95% CI, 0.65 to 0.71).

Because ACs and CKD with haemodialysis were the strongest risk factors for bleeding after ESD, we evaluated the prediction ability of a simple model by combining these two factors. However, this model had only modest discrimination ability (\(c\)-statistic, 0.61; 95%CI, 0.57 to 0.64; online supplementary table 5). In DCA, the net benefit of the simple model in predicting bleeding after ESD was lower than that of the risk score and risk classification (online supplementary figure 2).

Interruption of AT agents
The time of interrupting AT agents before ESD, the time of resuming them after ESD and the duration of interrupting them differed according to AT agent (online supplementary table 6). Thus, we evaluated the effect of interruption of AT agents on bleeding after ESD, based on the time of interrupting/resuming AT agents and the kind of them. Statistical significance was observed only for ≥6 days interruption of aspirin before ESD (OR, 0.39; p=0.020), although adjusted ORs of interruption for bleeding tended to differ across the AT agents (online supplementary table 7).

Example
A 65-year-old male, who takes warfarin and aspirin, has an EGC with a maximal diameter of 22mm, which is located in the middle-third of the stomach.

Total score =4 (for warfarin)+2 (for aspirin)=6.

Therefore, this patient falls in the very high-risk group with a total score of 6 (41.5% risk of bleeding after ESD). When warfarin and aspirin are interrupted (=−2 points), the total score decreases to 4 points (14.5% risk of bleeding).

Model validation
In the external validation cohort, the distribution of the prediction model and the bleeding rates are shown in table 3. The discriminative ability of the prediction model as measured by the \(c\)-statistic in the external validation cohort was 0.70 (95% CI, 0.64 to 0.76). This model was well calibrated with a calibration-in-the-large of 0.05 and a calibration slope of 1.01, and the overall performance was also good (table 4).
### Table 2  Multivariate logistic regression analysis of predictive factors for bleeding after ESD for EGC in the development cohort and the scoring system

<table>
<thead>
<tr>
<th>Predictive Factor</th>
<th>Adjusted OR</th>
<th>95% CI</th>
<th>P value</th>
<th>β regression coefficient</th>
<th>Points*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age ≥75 years</td>
<td>1.00</td>
<td>0.80 to 1.24</td>
<td>0.982</td>
<td>−0.002</td>
<td>–</td>
</tr>
<tr>
<td>Sex Male</td>
<td>1.21</td>
<td>0.92 to 1.59</td>
<td>0.170</td>
<td>0.191</td>
<td>–</td>
</tr>
<tr>
<td>Ischaemic heart disease Yes</td>
<td>1.41</td>
<td>0.97 to 2.03</td>
<td>0.069</td>
<td>0.341</td>
<td>–</td>
</tr>
<tr>
<td>Liver cirrhosis Yes</td>
<td>1.25</td>
<td>0.61 to 2.59</td>
<td>0.544</td>
<td>0.225</td>
<td>–</td>
</tr>
<tr>
<td>CKD with haemodialysis Yes</td>
<td>4.33</td>
<td>2.71 to 6.91</td>
<td>&lt;0.001</td>
<td>1.464</td>
<td>3</td>
</tr>
<tr>
<td>Endoscopic gastric atrophy Mild/no</td>
<td>0.84</td>
<td>0.54 to 1.30</td>
<td>0.436</td>
<td>−0.175</td>
<td>–</td>
</tr>
<tr>
<td>Aspirin Yes</td>
<td>2.24</td>
<td>1.55 to 3.24</td>
<td>&lt;0.001</td>
<td>0.807</td>
<td>2</td>
</tr>
<tr>
<td>P2Y12RA Yes</td>
<td>3.13</td>
<td>1.91 to 5.12</td>
<td>&lt;0.001</td>
<td>1.140</td>
<td>2</td>
</tr>
<tr>
<td>Cilostazol Yes</td>
<td>2.04</td>
<td>1.09 to 3.80</td>
<td>0.025</td>
<td>0.712</td>
<td>1</td>
</tr>
<tr>
<td>Warfarin Yes</td>
<td>8.74</td>
<td>4.92 to 15.54</td>
<td>&lt;0.001</td>
<td>2.168</td>
<td>4</td>
</tr>
<tr>
<td>DOAC Yes</td>
<td>8.16</td>
<td>4.74 to 14.04</td>
<td>&lt;0.001</td>
<td>2.099</td>
<td>4</td>
</tr>
<tr>
<td>Interruption of AT agents Each kind of agents</td>
<td>0.67</td>
<td>0.46 to 0.97</td>
<td>0.033</td>
<td>−0.403</td>
<td>–</td>
</tr>
<tr>
<td>Heparin bridging Yes</td>
<td>0.82</td>
<td>0.49 to 1.36</td>
<td>0.435</td>
<td>−0.205</td>
<td>–</td>
</tr>
<tr>
<td>Replacement of APAs Yes</td>
<td>1.24</td>
<td>0.57 to 2.66</td>
<td>0.587</td>
<td>0.212</td>
<td>–</td>
</tr>
<tr>
<td>The number of tumours Multiple</td>
<td>1.38</td>
<td>1.04 to 1.85</td>
<td>0.028</td>
<td>0.324</td>
<td>1</td>
</tr>
<tr>
<td>Tumour size &gt;30mm</td>
<td>1.72</td>
<td>1.28 to 2.31</td>
<td>&lt;0.001</td>
<td>0.545</td>
<td>1</td>
</tr>
<tr>
<td>Tumour location Lower third</td>
<td>1.68</td>
<td>1.35 to 2.10</td>
<td>&lt;0.001</td>
<td>0.520</td>
<td>1</td>
</tr>
<tr>
<td>Tumour differentiation Undifferentiated</td>
<td>1.40</td>
<td>0.87 to 2.24</td>
<td>0.167</td>
<td>0.334</td>
<td>–</td>
</tr>
<tr>
<td>Tumour depth SM2</td>
<td>1.22</td>
<td>0.82 to 1.82</td>
<td>0.328</td>
<td>0.199</td>
<td>–</td>
</tr>
<tr>
<td>Ulceration Positive</td>
<td>1.09</td>
<td>0.76 to 1.55</td>
<td>0.651</td>
<td>0.082</td>
<td>–</td>
</tr>
<tr>
<td>ESD procedure time &gt;120min</td>
<td>1.25</td>
<td>0.95 to 1.64</td>
<td>0.110</td>
<td>0.222</td>
<td>–</td>
</tr>
<tr>
<td>Resection type Piecemeal</td>
<td>0.76</td>
<td>0.17 to 3.48</td>
<td>0.721</td>
<td>−0.278</td>
<td>–</td>
</tr>
<tr>
<td>Second-look endoscopy Yes</td>
<td>1.01</td>
<td>0.80 to 1.27</td>
<td>0.950</td>
<td>0.007</td>
<td>–</td>
</tr>
</tbody>
</table>

*Weighted points were assigned proportional to β regression coefficient values multiplied by 2 to the nearest integer in the significant predictive factors.

APA, antiplatelet agent; AT, antithrombotic; CKD, chronic kidney disease; DOAC, direct oral anticoagulant; EGC, early gastric cancer; ESD, endoscopic submucosal dissection; P2Y12RA, P2Y12 receptor antagonist; SM2, submucosal invasion ≥500 μm from the muscularis mucosae.

### Table 3  Distribution of risk scores and risk classification for bleeding after ESD for EGC in the derivation cohort

<table>
<thead>
<tr>
<th>Total points</th>
<th>Patients (n=8288)*</th>
<th>Bleeding (n=387)</th>
<th>Rate of bleeding (95% CI) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2923</td>
<td>58</td>
<td>2.0 (1.5 to 2.6)</td>
</tr>
<tr>
<td>1</td>
<td>3344</td>
<td>117</td>
<td>3.5 (2.9 to 4.2)</td>
</tr>
<tr>
<td>2</td>
<td>1059</td>
<td>65</td>
<td>6.1 (4.8 to 7.8)</td>
</tr>
<tr>
<td>3</td>
<td>471</td>
<td>45</td>
<td>9.6 (7.1 to 12.6)</td>
</tr>
<tr>
<td>4</td>
<td>269</td>
<td>42</td>
<td>14.5 (10.7 to 19.1)</td>
</tr>
<tr>
<td>5</td>
<td>123</td>
<td>26</td>
<td>21.1 (14.3 to 29.4)</td>
</tr>
<tr>
<td>6</td>
<td>53</td>
<td>22</td>
<td>41.5 (28.1 to 55.9)</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>7</td>
<td>43.8 (19.8 to 70.1)</td>
</tr>
<tr>
<td>8</td>
<td>10</td>
<td>5</td>
<td>50.0 (18.7 to 81.3)</td>
</tr>
</tbody>
</table>

*There were missing data in three cases.

### Subgroup analysis

Model discrimination of the BEST-J score in very high-volume, high-volume and intermediate-volume centres was moderately good with c-statistics that ranged from 0.70 to 0.72; however, in low-volume centre, it was modest (c-statistic, 0.69) (table 6). The results pertaining to the model calibration are presented in table 6.

### Management after ESD

In the whole cohort, the periods of hospital stay after ESD for each risk category in those with and without bleeding after ESD are shown in online supplementary figure 3. The median periods of hospital stay in patients with bleeding after ESD in the low-risk, intermediate-risk, high-risk and very high-risk categories were 8, 8, 10 and 12 days, respectively, and those without...
We also evaluated the effect of SLE for bleeding after ESD in each risk category in the whole cohort. As a result, no significant difference in the bleeding rate after ESD was observed between patients with and without SLE in all the risk categories (online supplementary table 8).

### DISCUSSION

Given the relatively low incidence of bleeding after ESD for EGC, the availability of a prediction score would be of extreme importance. In the present study, we evaluated 23 candidate predictors and found 10 independent factors associated with bleeding. Our study highlights the strong impact of ACs, that is, warfarin and DOAC, on bleeding among AT agents. In addition, we first found the risk of P2Y12RA and cilostazol for bleeding after ESD. Furthermore, the impact of CKD with haemodialysis on bleeding was lower than that of ACs but higher than the bleeding risk to an extent depending on the number of kinds of AT agents. Especially, heparin bridging was not a risk factor for bleeding in the present study, which is contrary to the previous studies. Indeed, heparin was not initiated after ESD in 18.4% (36/196) of the patients with heparin bridging, and the patients who did receive heparin after ESD stopped heparin at a median of 2 days after the procedure, whereas dalteparin was initiated at 1 to 3 days after ESD with a mean dosing period of 2.5 days in a previous study. Therefore, caution is required when interpreting the risk of heparin bridging in the present study.

To date, two studies created the predictive models for bleeding after ESD. Toya et al reported the usefulness of combining the factors of ACs and resection size ≥35 mm for predicting bleeding after ESD for EGC. However, this study has the limitations of not evaluating the role of APA in detail and a small sample size from a single institution. A recent study established a risk-scoring model for predicting bleeding after colorectal ESD. In this model, points were assigned to use of AT agents except for aspirin alone and tumour size ≥30 mm, in addition to rectosigmoid area. However, because of low event rates and small sample size from a single institution, the variables of ACs and CKD were not adequately evaluated and the discrimination ability of this model (c-statistic, 0.634) was limited; a c-statistic of ≥0.70 is typically considered sufficient to make clinically useful individual predictions. In terms of existing bleeding scores, several scores for patients with atrial fibrillation are also available. However, these scores also have only modest discrimination (c-statistic ≤0.63) in the external validation cohort. Our prediction model (c-statistic, 0.70) satisfied the clinically useful discrimination ability in the external validation cohort. However, the c-statistic just reached its satisfactory value; thus, it should be noted that our model may not be very reliable. Moreover, since the lower limit of 95% CI was <0.70, further external validation using a larger cohort is required for elucidating whether this model has an acceptable discrimination ability or not. Nevertheless, since this model was established based on the analysis of almost all known key candidate predictors for many patients throughout Japan and achieved good results in the overall performance, calibration ability and net benefit, we believe that this model could be applied as a simple aid to clinical decision-making in routine practice.

### Table 4 Performance of the prediction model

<table>
<thead>
<tr>
<th>Derivation cohort</th>
<th>Validation cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nagelkerke R²</td>
<td>0.11</td>
</tr>
<tr>
<td>Brier score</td>
<td>0.04</td>
</tr>
<tr>
<td>c-statistic (95% CI)</td>
<td>0.71 (0.68 to 0.73)</td>
</tr>
<tr>
<td>Calibration-in-the-large</td>
<td>NA</td>
</tr>
<tr>
<td>Calibration slope</td>
<td>NA</td>
</tr>
</tbody>
</table>

Nagelkerke R² can range from 0 to 1, with 0 denoting that model does not explain any variation and 1 denoting that it perfectly explains the observed variation. The Brier score can range from 0 for a perfect model to 0.25 for a non-informative model with a 50% incidence of the outcome. c-statistic can range from 0.50 for a non-discriminative model to 1 for a perfect model with ≥0.70 as a reasonable discriminative ability. The calibration-in-the-large indicates whether predicted probabilities are systematically too low (value >0) or too high (value <0). The calibration slope indicates whether the model is overfitted (estimated risks too extreme, value <1) or underfitted (estimated risks too close to baseline risk, value >1). A model is perfectly calibrated if the calibration-in-the-large is 0 and calibration slope is 1.

*There were missing data in three cases.
NA, not applicable.

### Table 5 Distribution of the total risk score and risk classification for bleeding after ESD for EGC in the validation cohort

<table>
<thead>
<tr>
<th>Risk score</th>
<th>Patients (n=2029)</th>
<th>Bleeding (n=102)</th>
<th>Rate of bleeding (95% CI) (%)</th>
<th>Risk classification</th>
<th>Patients (n=2029)</th>
<th>Bleeding (n=102)</th>
<th>Rate of bleeding (95% CI) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>749</td>
<td>22</td>
<td>2.9 (1.8 to 4.4)</td>
<td>Low-risk</td>
<td>1481</td>
<td>41</td>
<td>2.8 (2.0 to 3.7)</td>
</tr>
<tr>
<td>1</td>
<td>732</td>
<td>19</td>
<td>2.6 (1.6 to 4.0)</td>
<td>Intermediate-risk</td>
<td>268</td>
<td>11</td>
<td>4.1 (2.7 to 7.2)</td>
</tr>
<tr>
<td>2</td>
<td>268</td>
<td>11</td>
<td>4.1 (2.1 to 7.2)</td>
<td>High-risk</td>
<td>224</td>
<td>33</td>
<td>14.7 (10.4 to 20.1)</td>
</tr>
<tr>
<td>3</td>
<td>146</td>
<td>21</td>
<td>14.4 (9.1 to 21.1)</td>
<td>Very high-risk</td>
<td>56</td>
<td>17</td>
<td>30.4 (18.8 to 44.1)</td>
</tr>
<tr>
<td>4</td>
<td>78</td>
<td>12</td>
<td>15.4 (8.2 to 25.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>41</td>
<td>11</td>
<td>26.8 (14.2 to 42.9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>5</td>
<td>50.0 (18.7 to 81.3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>1</td>
<td>33.3 (0.8 to 90.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0.0 (0.0 to 84.2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ESD, endoscopic submucosal dissection; EGC, early gastric cancer.
We here provide a new proposal for patient management after ESD, which is based on individual bleeding probabilities in the prediction model. Despite that patients at a low risk for bleeding can be safely managed, we found that the period of hospital stay of patients who underwent ESD without bleeding, which reflects the standard schedule of hospital stay for each risk category, did not differ among the risk categories (6 to 7 days was the median for the four categories). Thus, we propose a shortened hospital stay (e.g., 3 to 4 days after ESD) in patients with low risk, which may contribute to lowering the cost.

By contrast, some clinical approaches are recommended for patients at high risk. First, when such patients take AT agents, interruption of these can reduce the risk of bleeding and the effect of interruption depends on the number of kinds of AT agents interrupted, although the risk of thromboembolism by the interruption should also be considered. Second, caution would be given to endoscopists to look for possible bleeding sources, which would lead to more careful prophylactic coagulation of visible vessels on mucosal defect after resection. Furthermore, intensive monitoring of high-risk patients may prevent the onset of severe conditions even if bleeding occurs. On the other hand, SLE did not help prevent bleeding after gastric ESD, which is in accordance with a previous report in patients without AT agents. According to a recent study, the polyglycolic acid sheet with fibrin glue could not also decrease the risk of bleeding after ESD. However, several studies have reported the effect of mucosal defect closure for preventing bleeding after endoscopic resection. Thus, patients with a high risk for bleeding are candidates for this technique after resection. Although endoclips are sometimes insufficient to maintain closure of a large mucosal defect after ESD, recent advanced techniques such as endoloop, hand-suturing and OverStitch endoscopic suturing technique may overcome this issue.

Several strengths of the current study warrant mention. First, the designs of the derivation and the external validation are robust, according to the TRIPOD guidelines. The method also conformed with recent developments for improved prediction model assessment, including DCA. Furthermore, the sample size in this study was calculated based on the recommendations for deriving and externally validating a prognostic model, thereby increasing the reliability of the prediction model. Although we enrolled twice as many patients as the preplanned sample size, which was based on 10 EPV, in the derivation stage, this increase in sample size may have led to the better prediction model, because model performance was reported to be better as EPV increases from 10 to 50. Second, the data for deriving a prediction model were obtained from the largest cohort. Third, missing values were minimal, and actually, whole data are available for over 99% of the enrolled patients. Finally, the management of the AT agents during the perioperative period in the enrolled patients was mainly based on the recent Japanese guidelines, which are similar to the European and American guidelines.

This study has several potential limitations. First, this study is retrospective in nature. Second, the prediction model was developed and externally validated in Japan, which leads to two potential issues. One of them is that it is unclear whether this model is applicable to other geographical areas, such as Europe. The other issue is that points assigned to each variable may be different when the model is derived from other areas with less ESD expertise. Hence, fully independent validation in these areas is required and, if this prediction model has the poor performance, a new prediction model using the international cohort should be considered. Third, the timing of bleeding after ESD was not considered in this prediction model. Moreover, some reports showed that risk factors for early and late bleeding were different. Fourth, we should be careful about interpreting the risk of heparin bridging, as described previously. Lastly, our subgroup analysis in the low-volume centre did not reach a satisfactory value of discrimination; however, this analysis did not have enough power because of the small number of events.

In summary, we derived and externally validated a prediction model (BEST-J score) for bleeding after ESD for EGC, with good performance metrics. We, therefore, believe that this model can be used in routine practice to optimise the management of patients after ESD based on individual bleeding risk. However, further external validation in other geographical areas is required.

### Author affiliations

1. Division of Gastroenterology, Tohoku University Graduate School of Medicine, Sendai, Japan
2. Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
3. Department of Gastroenterology, Cancer Institute Hospital, Japanese Foundation for Cancer Research, Tokyo, Japan
4. Division of Endoscopy, Shizuoka Cancer Centre, Shizuoka, Japan
5. Department of Gastroenterology, Toranomon Hospital, Tokyo, Japan
6. Department of Gastroenterology, Ishikawa Prefectural Central Hospital, Kanazawa, Japan
7. Department of Gastroenterology, Osaka City University Graduate School of Medicine, Osaka, Japan
8. Department of Endoscopy, Fukushima Medical University Hospital, Fukushima, Japan
9. Department of Endoscopy, The Jikei University School of Medicine, Tokyo, Japan
10. Department of Gastroenterology, Kobe University Graduate School of Medicine, Kobe, Japan
11. Department of Gastroenterology, Tohoku University Hospital, Sendai, Japan
12. Department of Internal Medicine, Wakayama Medical University, Wakayama, Japan
13. Department of Gastroenterology and Metabolism, Ehime University Graduate School of Medicine, Toon, Japan
14. Division of Gastroenterology and Hepatology, Osaka General Medical Centre, Osaka, Japan
15. Division of Endoscopy, Hiroisaki University Hospital, Hiroisaki, Japan

---

**Table 6** Discrimination and calibration of the prediction model based on the centre-volume in the derivation cohort

<table>
<thead>
<tr>
<th>Centre-volume</th>
<th>Very high-volume</th>
<th>High-volume</th>
<th>Intermediate-volume</th>
<th>Low-volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>The number of patients, n</td>
<td>2358</td>
<td>2990</td>
<td>2278</td>
<td>665</td>
</tr>
<tr>
<td>The number of bleeding, n</td>
<td>120</td>
<td>123</td>
<td>117</td>
<td>27</td>
</tr>
<tr>
<td>c-statistic (95% CI)</td>
<td>0.70 (0.65 to 0.75)</td>
<td>0.72 (0.67 to 0.77)</td>
<td>0.76 (0.70 to 0.76)</td>
<td>0.69 (0.57 to 0.80)</td>
</tr>
<tr>
<td>Calibration-in-the-large</td>
<td>0.20</td>
<td>–0.21</td>
<td>0.09</td>
<td>–0.18</td>
</tr>
<tr>
<td>Calibration slope</td>
<td>1.12</td>
<td>1.01</td>
<td>1.01</td>
<td>0.88</td>
</tr>
</tbody>
</table>
Patient and public involvement Patients and/or the public were not involved in the design, conduct, or reporting or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval The study protocol was approved by the ethics committees of the respective institutions.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement The data that support the findings of this study are available from the first author, WH (ORCID: 0000-0001-9717-0281), upon reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Waku Hatta http://orcid.org/0000-0001-9717-0281
Atsushi Masamune http://orcid.org/0000-0001-7184-7282

REFERENCES


