

1 **Original Research Article**

2 **Prediction Model for Severe Thrombocytopenia induced by Gemcitabine plus Cisplatin**
3 **Combination Therapy in Patients with Urothelial Cancer**

4

5 **Short running title:** Prediction model for severe thrombocytopenia

6

7 Noriaki Matsumoto^{1,2†}, Tomohiro Mizuno^{1*†}, Yosuke Ando¹, Koki Kato¹, Masanori Nakanishi³, Tsuyoshi
8 Nakai¹, Jeannie K Lee⁴, Yoshitaka Kameya⁵, Wataru Nakamura⁶, Kiyoshi Takahara⁶, Ryoichi Shiroki⁶, and
9 Shigeki Yamada¹

10

11 ¹Department of Pharmacotherapeutics and informatics, Fujita Health University School of Medicine,
12 Toyoake, Aichi, Japan

13 ²Sugi Pharmacy Co., Ltd., Obu, Aichi, Japan

14 ³Department of Nephrology, Fujita Health University School of Medicine, Toyoake, Aichi, Japan

15 ⁴Department of Pharmacy Practice & Science, University of Arizona R. Ken Coit College of Pharmacy,
16 Tucson, Arizona, USA

17 ⁵Department of Information Engineering, Meijo University, Japan

18 ⁶Department of Urology, Fujita Health University School of Medicine, Toyoake, Aichi, Japan

19

20 †These authors contributed equally to this work.

21

22 ***Corresponding author**

23 Tomohiro Mizuno, Ph.D.

24 Department of Clinical Pharmacy, Fujita Health University

25 1-98 Dengakugakubo, Kutsukake-cho, Toyoake, 470-1192, Japan

26 Tel: (+81)562-93-2493;

27 Fax: (+81)562-93-4593

28 E-mail: tomohiro.mizuno@fujita-hu.ac.jp

29 **Acknowledgements**

30 This work was supported by the JSPS KAKENHI [grant number 21K06696] and the Research Center
31 for Pathogenesis of Intractable Diseases, Research Institute of Meijo University, Japan. We would like to
32 thank Editage (<https://www.editage.com/>) for editing and reviewing this manuscript for English
33 language.

34

35 **Conflict of interest statement**

36 The authors declare no conflicts of interest regarding the subject of this study. The interests from the past
37 three years are as follows: NM is an employee of Sugi Pharmacy Co., Ltd. SY received honoraria from
38 AstraZeneca K.K., Daiichi Sankyo Co., Ltd., EA Pharma Co., Ltd., Eisai Co., Ltd., Meiji Seika Pharma
39 Co., Ltd., Nippon Shinyaku Co., Ltd., Nipro Corporation, Otsuka Pharmaceutical Factory, Pfizer Japan
40 Inc., Taisho Pharmaceutical Co., Ltd., Takeda Pharmaceutical Co., Ltd., Toa Eiyo Ltd., and Towa
41 Pharmaceutical Co., Ltd. and funding from EA Pharma Co., Ltd., Kissei Pharmaceutical Co., Ltd.,
42 Mochida Pharmaceutical Co., Ltd., SUGI Holdings Co., Ltd., and Otsuka Pharmaceutical Factory. TM
43 received honoraria from Kyowa Kirin Co., Ltd., Nanzando Co., Ltd., and WELCIA Holdings Co., Ltd.

44

45 **Abstract**

46 **Study Objective:** Severe thrombocytopenia is often induced by combination therapy using gemcitabine
47 and cisplatin (GC). Thrombocytopenia reduces therapeutic intensity and requires platelet transfusion in
48 some cases. This study aimed to develop a predictive model for the incidence of severe
49 thrombocytopenia using machine learning to increase the safety and efficacy of GC therapy.

50 **Design:** A retrospective cohort study was conducted in patients with bladder cancer at the initiation of
51 GC therapy at Fujita Health University Hospital between January 2006 and December 2021. We
52 performed receiver operating characteristic analysis to determine the cutoff values of the associated
53 factors. Univariate and multivariate analyses were conducted to identify risk factors associated with the
54 occurrence of severe thrombocytopenia. The prediction model was constructed from an ensemble model
55 and gradient-boosted decision trees to estimate the risk of an outcome using the risk factors associated
56 with the occurrence of severe thrombocytopenia.

57 **Main Results:** Of 192 patients included in this study, 48 patients (25%) experienced severe
58 thrombocytopenia induced by GC therapy. Pretreatment platelet count, hemoglobin, hematocrit,
59 lymphocyte count, prognostic nutritional index, and gemcitabine dose on day 1 administration were
60 determined as risk factors using univariate analysis. Multivariate analyses revealed that platelet count \leq
61 $21.4 (\times 10^4/\mu\text{L})$ [odds ratio=7.54, $p < 0.01$], hemoglobin ≤ 11.9 (g/dL) [odds ratio=2.78, $p=0.01$],
62 lymphocyte count $\leq 1.458 (\times 10^3/\mu\text{L})$ [odds ratio=2.20, $p=0.04$], and dose of gemcitabine ≥ 775.245
63 (mg/m^2) [odds ratio=3.49, $p < 0.01$] were risk factors of severe thrombocytopenia. The performance of
64 the predictive model using these associated factors was high (area under the curve: 0.75, accuracy=0.77,
65 precision=0.53, recall=0.69, and F-measure=0.60).

66 **Conclusions:** Platelet count, hemoglobin level, lymphocyte count, and gemcitabine dose contributed to
67 the development of a prediction model to identify the incidence of GC-induced thrombocytopenia.

68

69 **KEYWORDS:** Prediction model, Thrombocytopenia, Gemcitabine, Cisplatin, Multivariate analysis

70

71 1. INTRODUCTION

72 Severe thrombocytopenia is often induced by combination therapy using gemcitabine and cisplatin
73 (GC). Grade 3 or 4 thrombocytopenia was more frequent in patients receiving GC therapy than in those
74 receiving methotrexate, vinblastine, doxorubicin, and cisplatin (MVAC) therapy in the randomized
75 controlled phase III study (GC: 57%, MVAC: 21%)¹. Severe thrombocytopenia results in reduced
76 therapeutic efficacy and requires platelet transfusion in some cases¹. Poor performance status, low platelet
77 count, low lymphocyte count, low polymorphonuclear leukocyte count, and types of chemotherapy were
78 associated with severe thrombocytopenia². Although the previous study defined regimens containing high
79 doses of doxorubicin, epirubicin, cisplatin, ifosfamide, cyclophosphamide, etoposide, and cytarabine as
80 high-risk chemotherapy regimens associated with thrombocytopenia, the risk factors for
81 thrombocytopenia induced by GC therapy had been unclear. Recently, chemotherapy cycles and cancer
82 stage³, platelet number, and potassium concentration⁴ have been reported as risk factors for
83 thrombocytopenia induced by GC therapy. The potassium level was within the standard values in the
84 previous study, and the association between potassium levels and thrombocytopenia remains unclear. The
85 neutrophil-lymphocyte ratio (NLR), platelet count, and prognostic nutritional index (PNI) are predictive
86 markers of severe thrombocytopenia induced by carboplatin and pemetrexed combination therapy⁵. Since
87 the NLR and PNI are markers of the systemic inflammatory response, systemic inflammation may be
88 associated with the incidence of chemotherapy-induced thrombocytopenia.

89 Although Zhou *et al.* reported the effectiveness of a predictive model for the risk of chemotherapy-
90 induced thrombocytopenia, which defined thrombocytopenia as a platelet count $< 100,000/\text{ml}$ ⁶, the
91 prediction ability of severe thrombocytopenia inducing the cancellation of chemotherapy has not been
92 evaluated. In addition, this study did not use common machine learning methods (e.g., neural networks
93 and gradient-boosted decision trees). We previously reported that a machine learning model developed
94 using neural networks and gradient-boosted decision trees was effective in predicting cisplatin-induced

95 acute kidney injury (AKI)⁷. This previously used algorithm evaluates not only the prediction ability of the
96 model containing various factors, but also calculates the rank of contribution as a predictive factor based
97 on sensitivity analysis. Hence, this algorithm clarifies the construction of a prediction model.

98 We believe that a predictive model for the incidence of severe thrombocytopenia may increase the
99 safety and efficacy of GC therapy. This study aimed to develop a predictive model for the incidence of
100 severe thrombocytopenia.

101

102 **2. METHODS**

103 **2.1. Data collection**

104 A retrospective cohort study was conducted on patients with bladder cancer at the initiation of GC
105 therapy at Fujita Health University Hospital from January 2006 to December 2021. The exclusion criteria
106 were as follows: 1) patients without essential laboratory values within 30 days before the initiation of GC
107 therapy and 2) patients without day 1 and/or 2 administration of GC. The following data were collected
108 from the medical records in Fujita Health University Hospital; gender, age, body weight, height, body
109 surface area (BSA), dose of GC, white blood cell (WBC) count, hemogram, platelet count, hemoglobin,
110 hematocrit, total protein, albumin, serum creatinine, aspartate aminotransferase (AST), alanine
111 aminotransferase (ALT), estimated glomerular filtration rate (eGFR) creatinine, blood urea nitrogen
112 (BUN), and concentrations of sodium and potassium.

113

114 **2.2. Evaluation of the association between inflammatory markers with severe thrombocytopenia**

115 To identify the risk factors associated with the occurrence of severe thrombocytopenia, we used C-
116 reactive protein (CRP), neutrophil count, lymphocyte count, neutrophil-to-lymphocyte ratio (NLR),
117 platelet-to-lymphocyte ratio (PLR), prognostic nutritional index (PNI), lymphocyte-to-monocyte ratio
118 (LMR), and Glasgow Prognostic Score (GPS).

119 The neutrophil count, lymphocyte count, NLR, PLR, PNI, and LMR were calculated as follows:

120 neutrophil count, WBC count \times segmented cell ratio; lymphocyte count, WBC count \times lymphocyte ratio;
121 NLR, neutrophil count/lymphocyte count; PLR, platelet count/lymphocyte count; PNI, $10 \times$ albumin +
122 $0.005 \times$ lymphocyte count; and LMR, lymphocyte count/monocyte count. GPS was assigned as follows;
123 GPS=2: CRP \leq 1.0 mg/dL and albumin \geq 3.5 g/dL, GPS=1: CRP $>$ 1.0 mg/dL or albumin $<$ 3.5 g/dL,
124 GPS=0: CRP $>$ 1.0 mg/dL and albumin $<$ 3.5 g/dL.

125

126

127 **2.3. Definition of severe thrombocytopenia**

128 The severity of thrombocytopenia was assessed from day 1 of the first course to day 0 of the second
129 course of GC therapy with Common Terminology Criteria for Adverse Events (CTCAE) version 5.0, as
130 follows: $7.5 \times 10^4/\mu\text{L} \leq$ platelet count $<$ lower limit of normal ($15.8 \times 10^4/\mu\text{L}$) was considered as grade 1,
131 $5.0 \times 10^4/\mu\text{L} \leq$ platelet count $<$ $7.5 \times 10^4/\mu\text{L}$ as grade 2, $2.5 \times 10^4/\mu\text{L} \leq$ platelet count $<$ $5 \times 10^4/\mu\text{L}$ as grade
132 3, and platelet count $<$ $2.5 \times 10^4/\mu\text{L}$ as grade 4.

133 Severe thrombocytopenia was defined as grade 3 or higher based on the above criteria, and the
134 subjects were divided into non-severe and severe thrombocytopenia groups (Figure 1).

135

136 **2.4. Statistical analysis**

137 Fisher's exact test was used to evaluate categorical variables, and the Mann-Whitney U-test was used
138 for continuous variables. A *p*-value of less than 0.05 was considered statistically significant. To evaluate
139 the association between thrombocytopenia and baseline patient characteristics, laboratory values, and
140 inflammatory markers, these factors were compared between the two groups. For values found to be
141 significantly associated with severe thrombocytopenia, receiver operating characteristic (ROC) curve
142 analysis was performed to determine the cutoff value using the Yoden index. The DeLong's test was also
143 performed for the two correlated ROC curves. Univariate and multivariate analyses were performed to
144 evaluate whether these cut-off values were associated with the incidence of severe thrombocytopenia. All

145 statistical analyses were performed using EZR Version 1.55, (Saitama Medical Center, Jichi Medical
146 University, Saitama, Japan), which is a graphical user interface for “R” (The R Foundation for Statistical
147 Computing, Vienna, Austria). A modified version of the R commander was designed by adding statistical
148 functions widely used in biostatistics⁸.

149

150 **2.5. Machine learning**

151 We constructed a prediction model based on a previous report⁷. Briefly, Prediction One version
152 3.2.0.3 (Sony Corporation, Tokyo, Japan) was used to construct the prediction model. Prediction One
153 created an ensemble model consisting of neural networks, which learn representations of both key factors
154 and their interactions from the dataset, and gradient-boosted decision trees estimate the risk of an outcome
155 using all predictors. Hyperparameters for selecting the best model were used to build the model. The
156 datasets used in the prediction model were shuffled before machine learning. A prediction model was
157 constructed using variables that were statistically significant in the multivariate analysis. To prevent
158 overfitting and improve the generalization performance, we conducted cross-validation. The area under
159 the curve (AUC), accuracy, precision, recall, and F-value, which were used to evaluate the prediction
160 model, were calculated using Prediction One. The rank of contribution as a predictive factor was
161 determined using sensitivity analysis.

162

163 **2.6. Ethics approval**

164 This study was approved by the ethics board of Fujita Health University Hospital (ethics committee
165 approval number: HM22-335) and was conducted in accordance with the appropriate guidelines. An opt-
166 out approach was used to obtain informed consent according to the guidelines of the ethics board.

167 **3. RESULTS**

168 **3.1. Baseline characteristics**

169 The study design is illustrated in Figure 1. This study included 282 patients who received first-line
170 GC therapy at Fujita Medical University Hospital between January 2006 and December 2021. Seventy
171 patients were excluded due to lack of essential laboratory values within 30 days before the initiation of
172 GC therapy. Twenty patients were excluded due to cancellation of chemotherapy on day 1 and/or day 2 of
173 administration. A total of 192 patients were included in this study. Forty-eight patients (25%) experienced
174 severe thrombocytopenia induced by GC therapy. The patients were divided into non-severe (n=144) and
175 severe thrombocytopenia (n = 48) groups.

176 Baseline characteristics, laboratory values, and inflammatory markers are shown in Table 1.
177 Pretreatment platelet count ($p \leq 0.01$), hemoglobin ($p=0.04$), hematocrit ($p=0.02$), lymphocyte count (p
178 ≤ 0.01), and PNI ($p=0.03$) were significantly lower in the severe thrombocytopenia group than in the
179 non-severe thrombocytopenia group. Dose of gemcitabine on day1 was significantly higher in the severe
180 thrombocytopenia group than in the non-severe thrombocytopenia group ($p=0.03$).

181

182 **3.2. ROC curve analysis**

183 ROC curve analysis was performed to determine the cutoff values for platelet count, hematocrit,
184 hemoglobin, lymphocyte count, PNI before GC therapy, and dose of gemcitabine (day1) (Figure 2), which
185 were 21.4 (specificity: 0.750, sensitivity: 0.667), 36.4 (specificity: 0.660 sensitivity: 0.542), 11.9
186 (specificity: 0.681, sensitivity: 0.500), 1.458 (specificity: 0.611 sensitivity: 0.625), 44.580 (specificity:
187 0.632, sensitivity: 0.500), and 775.245 (specificity: 0.382, sensitivity: 0.854), respectively. AUC and 95%
188 CI of platelet count, hematocrit, hemoglobin, lymphocyte count, PNI before GC therapy, and dose of day1
189 gemcitabine (day1) were 0.718 (95% CI: 0.63–0.806), 0.612 (95% CI: 0.521–0.702), 0.601 (95% CI:
190 0.508–0.694), 0.649 (95% CI: 0.561–0.737), 0.605 (95% CI: 0.509–0.702), and 0.604 (95% CI: 0.517–
191 0.691), respectively.

192

193 **3.3. Risk factors of severe thrombocytopenia**

194 Univariate and multivariate analyses were conducted to determine risk factors for severe
195 thrombocytopenia (Table 2). Cutoff values were used for ROC curve analysis. There were no significant
196 differences between the AUC of hemoglobin and hematocrit ($p=0.31$), PNI, and lymphocyte count
197 ($p=0.30$) in the ROC curve analysis (Supplementary Figure 1). Since hemoglobin and hematocrit, PNI,
198 and lymphocyte count were correlated, we used hemoglobin and lymphocyte counts in the multivariate
199 analysis. Hemoglobin was used in the multivariate analysis as it is closely associated with hemolytic
200 anemia that occurs in gemcitabine-associated thrombotic thrombocytopenic purpura (TTP)⁹. The
201 lymphocyte count was also used because it is reportedly associated with thrombocytopenia in patients
202 undergoing various chemotherapy². The univariable analysis revealed that platelet count ≤ 21.4 ($\times 10^4/\mu\text{L}$)
203 (odds ratio=5.93, $p < 0.01$), hemoglobin ≤ 11.9 (g/dL) (odds ratio=2.12, $p=0.04$), lymphocyte count \leq
204 1.458 ($\times 10^3/\mu\text{L}$) (odds ratio= 2.39, $p=0.01$), and dose of gemcitabine ≥ 775.245 (mg/m²) (odds ratio=3.07,
205 $p < 0.01$) were the risk factors of severe thrombocytopenia. The multivariate analysis followed the results
206 of the univariate analysis. The patients with platelet count ≤ 21.4 ($\times 10^4/\mu\text{L}$) (odds ratio=7.54, 95% CI:
207 3.43–16.60, $p < 0.01$), hemoglobin ≤ 11.9 (g/dL) (odds ratio=2.78, 95% CI: 1.27–6.09, $p=0.01$),
208 lymphocyte count ≤ 1.458 ($\times 10^3/\mu\text{L}$) (odds ratio=2.20, 95% CI: 1.03–4.69, $p=0.04$), and dose of
209 gemcitabine ≥ 775.245 (mg/m²) (odds ratio=3.49, 95% CI: 1.41–8.65, $p < 0.01$) had a high risk of severe
210 thrombocytopenia.

211

212 **3.4. Performance of the prediction model for severe thrombocytopenia**

213 The prediction model for severe thrombocytopenia was established by machine learning using the
214 dataset for platelet count, hemoglobin, lymphocyte count, and dose of gemcitabine, which were
215 determined as risk factors for severe thrombocytopenia in multivariate analysis. These data were converted
216 to binary values using the cutoff values. The performance of the prediction model was as follows AUC:

217 0.75, accuracy: 0.77; precision: 0.53; recall: 0.69; F-measure: 0.60 (Table 3). When platelet count alone
218 was used as a predictive factor, F-measure decreased by 5% compared to the current prediction model
219 (Supplementary Table 1; AUC: 0.69, accuracy: 0.73, precision: 0.47, recall: 0.67, F-measure: 0.55).
220 Contribution as a predictive factor of severe thrombocytopenia was ranked in the following order; platelet
221 counts ≤ 21.4 ($\times 10^4/\mu\text{L}$), dose of gemcitabine ≥ 775.245 (mg/m^2), hemoglobin ≤ 11.9 (g/dL), and
222 lymphocyte count ≤ 1.458 ($\times 10^3/\mu\text{L}$) (Table 4).

223

224 4. DISCUSSION

225 Our results suggest that platelet count, hemoglobin level, lymphocyte count, and dose of gemcitabine
226 are risk factors for severe thrombocytopenia, and the performance of the prediction model using these
227 factors showed a high prediction ability.

228 In the present study, the percentage of patients who experienced grade 3 or 4 thrombocytopenia
229 induced by GC therapy was 25%, which is lower than that in previous reports¹. The reason for this
230 difference may be that the observation period was shorter in the present study than in the previous phase
231 III study. Dose-limiting toxicity of gemcitabine against severe thrombocytopenia was observed in our
232 study. In a previous report, peak blood concentration of gemcitabine ($14.3 \mu\text{g}/\text{mL}$) was associated with a
233 50% reduction in platelet count¹⁰. The dose of gemcitabine was higher ($> 1,500 \text{ mg}/\text{m}^2$) in the previous
234 study than in the current study, and the median dose of gemcitabine was $800 \text{ mg}/\text{m}^2$. Therefore, in this
235 study, the peak blood concentration was predicted to be $< 10 \text{ g}/\text{mL}$. Hence, to prevent severe
236 thrombocytopenia, our results suggest that patients with risk factors (e.g., low platelet count, low
237 hemoglobin level, and low lymphocyte count) require a dose adjustment of gemcitabine to $< 775 \text{ mg}/\text{m}^2$.
238 To the best of our knowledge, no reports recommend dose adjustment in patients with a high risk of
239 thrombocytopenia undergoing GC therapy. Our findings may contribute to decreasing the incidence of
240 severe thrombocytopenia.

241 Pre-chemotherapy lymphocyte, hemoglobin, and platelet counts were associated with the incidence

242 of severe thrombocytopenia following GC therapy in the present study. Commonly, the pre-chemotherapy
243 platelet count is associated with the incidence of chemotherapy-induced thrombocytopenia^{2,4,5,11}, whereas
244 there are differences among studies regarding whether lymphocyte counts and hemoglobin are markers of
245 thrombocytopenia. Lymphocyte counts² and hemoglobin¹¹ are associated with the incidence of
246 chemotherapy-induced thrombocytopenia, but there have been no reports showing an association between
247 lymphocyte count or hemoglobin and the incidence of GC therapy-induced thrombocytopenia. Although
248 potassium concentration is a risk factor for GC therapy-induced thrombocytopenia⁴, this factor was not
249 associated with the incidence of GC therapy-induced thrombocytopenia in our study. The NLR and LMR,
250 which are markers of systemic inflammatory response, were higher in the severe thrombocytopenia group
251 than in the non-severe group. As a previous report suggested that NLR is associated with severe
252 thrombocytopenia induced by carboplatin and pemetrexed⁵, there may be an underlying mechanism
253 similar to that induced by GC therapy.

254 Thrombotic thrombocytopenic purpura (TTP) is life-threatening complication of GC therapy^{9, 12, 13}.
255 TTP has also been reported in patients receiving a single dose of pemetrexed¹⁴. TTP is accompanied by
256 fever, hemolytic anemia, thrombocytopenia, neurological complications and renal impairment⁹. Therefore,
257 the mechanism underlying severe thrombocytopenia may be similar to that underlying TTP. Endothelial
258 cell injury is recognized as a trigger for TTP, and endothelial dysfunction increases the release of tissue
259 plasminogen activators, thrombomodulin, and plasminogen activator inhibitor¹⁵. Treatment with cisplatin,
260 gemcitabine, and a tyrosine kinase inhibitor targeting the vascular endothelial growth factor receptor
261 promotes endothelial damage and activates the coagulation cascade and platelets¹⁶. Gemcitabine-induced
262 TTP is associated with the cumulative dose of gemcitabine¹⁷. These features of TTP are consistent with
263 our results, in which a low platelet count, low hemoglobin level, and gemcitabine dose were associated
264 with severe thrombocytopenia. Low hemoglobin levels may be associated with hemolytic anemia.
265 Although no patients had gemcitabine-associated TTP in our study, these predictors of severe
266 thrombocytopenia have potential as markers of gemcitabine-associated TTP.

267 A predictive model for the risk of chemotherapy-induced thrombocytopenia was presented in a
268 previous report⁶. This predictive model was developed using the bootstrapping technique, but not the
269 common methods of machine learning (e.g., neural networks and gradient-boosted decision trees). Our
270 study is the first to develop a predictive model for chemotherapy-induced thrombocytopenia using
271 machine learning methods. We could not compare the prediction ability of the current and previous models
272 because no studies have used a similar method of machine learning. A prediction model for acute kidney
273 injury induced by cisplatin was reported in our previous report⁷. The performance of this prediction model
274 was high in older patients (AUC: 0.76; accuracy: 0.71; precision: 0.31; recall: 1.00; F-measure: 0.47),
275 while that of our current prediction model was of the same degree as that of the previous model (AUC:
276 0.75; accuracy: 0.77; precision: 0.53; recall: 0.69; F-measure: 0.60). The F-measure reflects the values for
277 both recall and precision, which is important for assessing the risks of overestimation and underestimation.
278 Since the development datasets differed between the two studies, the comparison of F-measure is not
279 appropriate. However, our results suggest that the current predictive model can contribute to increasing
280 the safety and efficacy of GC therapy. It is a reasonable result that platelet count was determined as the
281 most contributing predictive factor, but when platelet count alone was used as a predictive factor, the F-
282 measure decreased by 5% compared with the current prediction model. Therefore, a prediction model that
283 uses all four predictors identified in this study is recommended.

284 Our study has several limitations. First, we could not obtain validation data as our cohort was small.
285 Although cross-validation was conducted to prevent overfitting and improve the generalization
286 performance, further multicenter studies are required. Second, the clinical importance of the cutoff values
287 was unclear because we evaluated the dataset from patients who received first-line GC therapy. To validate
288 the importance of the cutoff values, a prospective study should be conducted in the future. Third, we could
289 not evaluate the improvement in clinical outcomes and economic utility of this prediction model. To
290 evaluate the impact of our prediction model on clinical benefits, a prospective study is required.

291 In conclusion, platelet count, hemoglobin level, lymphocyte count, and gemcitabine dose contributed

292 to the development of a prediction model for the incidence of GC-induced thrombocytopenia.

293

294

295

296 **References**

- 297 1. von der Maase H, Hansen SW, Roberts JT, et al. Gemcitabine and cisplatin versus methotrexate,
298 vinblastine, doxorubicin, and cisplatin in advanced or metastatic bladder cancer: results of a large,
299 randomized, multinational, multicenter, phase III study. *J Clin Oncol* 2000;17:3068-77.
- 300 2. Blay JY, Le Cesne A, Mermet C, et al. A risk model for thrombocytopenia requiring platelet
301 transfusion after cytotoxic chemotherapy. *Blood* 1998;2:405-10.
- 302 3. Wang Z, Cai XJ, Chen LY, et al. Factors potentially associated with gemcitabine-based
303 chemotherapy-induced thrombocytopenia in Chinese patients with nonsmall cell lung cancer. *J Cancer*
304 *Res Ther* 2018;Supplement:S656-s60.
- 305 4. Takahashi N, Sunaga T, Fujimiya T, et al. Risk Associated with Severe Hematological Toxicity in
306 Patients with Urothelial Cancer Receiving Combination Chemotherapy of Gemcitabine and Cisplatin.
307 *Chemotherapy* 2020;1-2:29-34.
- 308 5. Takahashi M, Takahashi K, Kaneda H, Kawaguchi T, Nagayama K. Pretreatment Platelet Count
309 and Neutrophil/Lymphocyte Ratio Are Predictive Markers for Carboplatin Plus Pemetrexed Therapy-
310 induced Thrombocytopenia. *Anticancer Res* 2021;11:5729-37.
- 311 6. Zhou S, Song B, Li C, et al. The predictive model for risk of chemotherapy-induced
312 thrombocytopenia based on antineoplastic drugs for solid tumors in eastern China. *Sci Rep*
313 2023;1:3185.
- 314 7. Okawa T, Mizuno T, Hanabusa S, et al. Prediction model of acute kidney injury induced by
315 cisplatin in older adults using a machine learning algorithm. *PLoS One* 2022;1:e0262021.
- 316 8. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics.
317 *Bone Marrow Transplant* 2013;3:452-8.
- 318 9. Kasi PM. Thrombotic thrombocytopenic purpura and gemcitabine. *Case Rep Oncol* 2011;1:143-
319 8.
- 320 10. Fogli S, Danesi R, De Braud F, et al. Drug distribution and pharmacokinetic/pharmacodynamic
321 relationship of paclitaxel and gemcitabine in patients with non-small-cell lung cancer. *Ann Oncol*
322 2001;11:1553-9.
- 323 11. Song B, Zhou S, Li C, et al. A Prediction Model for Chemotherapy-Induced Thrombocytopenia
324 Based on Real-World Data and a Close Relationship Between AST/ALT Ratio and Platelet Count in
325 Patients with Solid Tumors. *Int J Gen Med* 2022;8003-15.
- 326 12. Suzuki Y, Kanazawa K, Kanai R, et al. A case of primary lung squamous cell carcinoma mimicking
327 malignant mesothelioma producing granulocyte colony stimulating factor with chemotherapy
328 (cisplatin and gemcitabine)-associated thrombotic thrombocytopenic purpura (TTP); An autopsy
329 case report. *Lung Cancer* 2019;105-08.
- 330 13. Nishijima Y, Hirata H, Himeno A, et al. Drug-induced thrombotic thrombocytopenic purpura
331 successfully treated with recombinant human soluble thrombomodulin. *Intern Med* 2013;10:1111-4.
- 332 14. Alabiso I, Baratelli C, Brizzi MP, Bitossi R, Ottone A, Tampellini M. Acute and fatal
333 thrombocytopenic thrombotic purpura after a single dose of pemetrexed. *Int J Clin Pharm*
334 2014;6:1141-3.
- 335 15. Thachil J. Causes of thrombotic thrombocytopenic purpura. *Lancet Oncol* 2007;9:757-8.
- 336 16. Kuenen BC, Levi M, Meijers JC, et al. Potential role of platelets in endothelial damage observed

337 during treatment with cisplatin, gemcitabine, and the angiogenesis inhibitor SU5416. *J Clin Oncol*
338 2003;11:2192-8.
339 17. Zupancic M, Shah PC, Shah-Khan F. Gemcitabine-associated thrombotic thrombocytopenic
340 purpura. *Lancet Oncol* 2007;7:634-41.
341
342

343 **Table 1. Baseline characteristics, laboratory values and inflammation markers**

Factors		Not-severe thrombocytopenia (n=144)	Severe thrombocytopenia (n=48)	p-value
Baseline characteristics				
Gender				
Male	(%)	106 (73.6)	39 (81.2)	0.34 ^a
Female	(%)	38 (26.4)	9 (18.8)	
Age (years)	median [range]	71 [38 90]	73 [48 83]	0.29
Body weight (kg)	median [range]	60.5 [38.8 96.9]	60.3 [36.5 87.7]	0.75
Height (cm)	median [range]	163.8 [139.0 184.0]	163.1 [138.0 175.5]	0.82
BSA (m ²)	median [range]	1.7 [1.2 2.1]	1.7 [1.2 2.0]	0.76
Dose of cisplatin (mg/m ²)	median [range]	55.5 [0 71.0]	56.5 [34.0 73.6]	0.23
Dose of gemcitabine (day 1) (mg/m ²)	median [range]	800.7 [500.5 1032.2]	858.6 [592.8 1018.5]	0.03
Laboratory values				
White blood cell count (×10 ³ /μL)	median [range]	6.1 [3.2 65.6]	5.7 [3.1 14.3]	0.30
Hemogram				
Segmented cell (%)	median [range]	62.4 [37.1 84.0]	66.3 [38.0 87.0]	0.10
Basophil (%)	median [range]	1.0 [0 3.0]	0.9 [0 2.0]	0.66
Eosinophil (%)	median [range]	3.0 [0.3 21.0]	3.0 [0.8 21.5]	0.62
Monocyte (%)	median [range]	6.0 [2.9 11.0]	6.0 [3.2 10.0]	0.86
Lymphocyte (%)	median [range]	26.0 [7.0 46.0]	23.9 [7.0 47.0]	0.08
Platelet count (×10 ⁴ /μL)	median [range]	25.5 [12.8 45.6]	20.1 [6.1 34.9]	< 0.01
Hemoglobin (g/dL)	median [range]	12.8 [6.7 17.1]	12.0 [8.8 15.1]	0.04
Hematocrit (%)	median [range]	38.5 [25.4 50.5]	36.3 [26.0 45.1]	0.02
Total protein (g/dL)	median [range]	6.8 [5.6 8.1]	6.7 [5.5 7.8]	0.44
Albumin (g/dL)	median [range]	3.9 [2.5 4.8]	3.9 [2.6 4.7]	0.32
Serum creatinine (mg/dL)	median [range]	0.9 [0.4 1.6]	0.9 [0.5 1.9]	0.11
eGFR creatinine (mL/min/1.73m ²)	median [range]	61.3 [32.1 130.0]	58.2 [27.4 93.5]	0.17
BUN (mg/dL)	median [range]	14.8 [7.7 29.5]	15.2 [8.9 24.0]	0.35
AST (U/L)	median [range]	19 [10 73]	18 [10 78]	0.31
ALT (U/L)	median [range]	15 [3 78]	13 [6 58]	0.15
Na (mmol/L)	median [range]	141 [128 146]	141 [130 146]	0.93
K (mmol/L)	median [range]	4.2 [3.0 5.1]	4.2 [3.6 5.2]	0.27
Inflammation markers				
CRP > 1.0 (mg/dL)	(%)	20 (13.9)	8 (16.7)	0.64 ^a
CRP ≤1.0 (mg/dL)	(%)	124 (86.1)	40 (83.3)	
Neutrophil count (×10 ³ /μL)	median [range]	3.8 [1.5 55.1]	3.6 [1.8 12.4]	0.53
Lymphocyte count (×10 ³ /μL)	median [range]	1.6 [0.5 4.6]	1.4 [0.4 2.4]	< 0.01
NLR	median [range]	2.4 [0.9 12.0]	2.9 [0.8 12.4]	0.07
PLR	median [range]	15.8 [5.5 39.9]	15.0 [5.8 40.2]	0.76
PNI	median [range]	47.1 [8.4 65.4]	44.6 [5.7 55.5]	0.03
LMR	median [range]	4.3 [1.3 9.3]	3.8 [1.4 11.0]	0.08
GPS	median [range]	0 [0 2]	0 [0 2]	0.22

344 345 AST: aspartate aminotransferase, ALT: alanine aminotransferase, BSA: body surface area, BUN: blood

346 urea nitrogen, CRP: c-reactive protein, eGFR: estimated glomerular filtration rate, GPS: Glasgow

347 Prognostic Score, K: concentration of potassium, LMR: lymphocyte-monocyte ratio, Na: concentration of
348 sodium, NLR: neutrophil-lymphocyte ratio, PLR: platelet-lymphocyte ratio, PNI: prognostic nutritional
349 index, ^aFisher's exact test, the rest: Mann-Whitney U-test,
350
351

352 **Table 2. The risk factors of severe thrombocytopenia**

Factors	Thrombocytopenia		Univariable analysis		Multivariable analysis	
	Not-severe (n = 144)	Severe (n = 48)	Odds ratio (95%CI)	p-value	Odds ratio (95%CI)	p-value
Platelet count > 21.4 ($\times 10^4/\mu\text{L}$)	108	16	5.93	< 0.01	7.54	< 0.01
Platelet count \leq 21.4 ($\times 10^4/\mu\text{L}$)	36	32	(2.80 13.05)		(3.43 16.60)	
Hemoglobin > 11.9 (g/dL)	98	24	2.12	0.04	2.78	0.01
Hemoglobin \leq 11.9 (g/dL)	46	24	(1.03 4.37)		(1.27 6.09)	
Lymphocyte count > 1.458 ($\times 10^3/\mu\text{L}$)	88	19	2.39	0.01	2.20	0.04
Lymphocyte count \leq 1.458 ($\times 10^3/\mu\text{L}$)	56	29	(1.17 4.97)		(1.03 4.69)	
Dose of gemcitabine (day1) < 775.245 (mg/m^2)	55	8	3.07	< 0.01	3.49	< 0.01
Dose of gemcitabine (day1) \geq 775.245 (mg/m^2)	89	40	(1.30 8.17)		(1.41 8.65)	

353

354 CI: confidence interval

355

356

357 **Table 3. Model performance for predicting severe thrombocytopenia**

Evaluation value	Evaluation score
AUC	0.75
Accuracy	0.77
Precision	0.53
Recall	0.69
F-measure	0.60

358

359 AUC: area under the curve

360

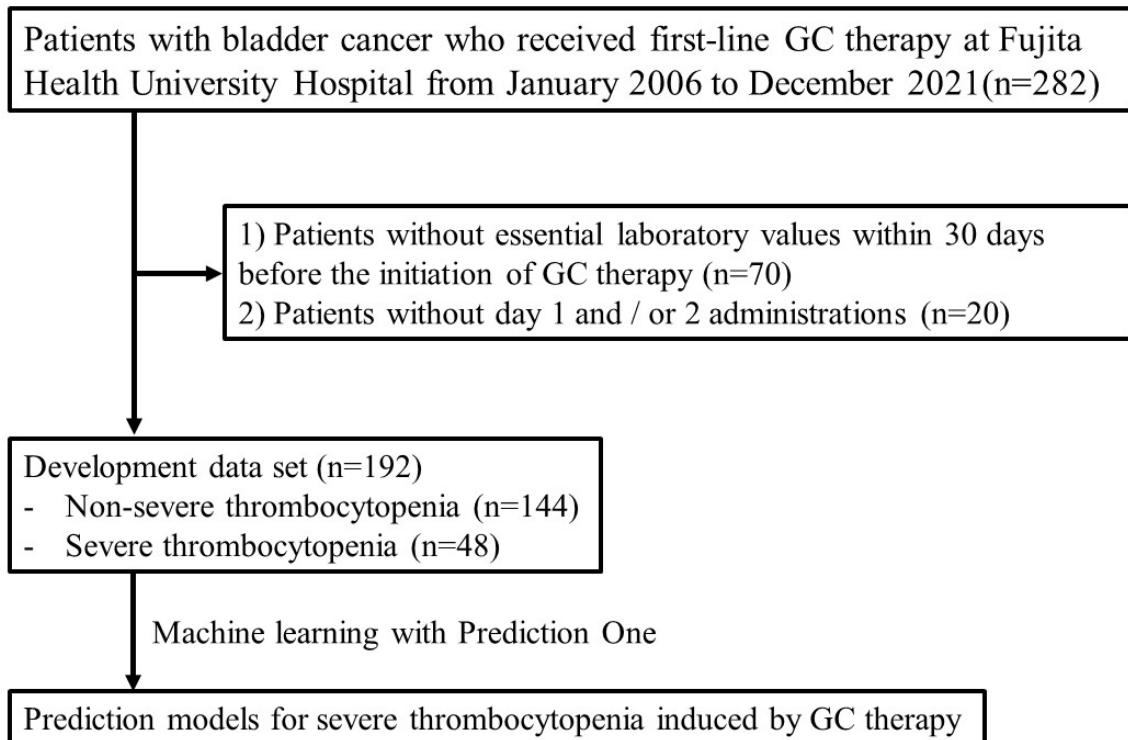
361

362 **Table 4. Model performance for predicting severe thrombocytopenia**

Predictive factor	Rank of contribution
Platelet count ≤ 21.4 ($\times 10^4/\mu\text{L}$)	1
Dose of gemcitabine ≥ 775.245 (mg/m^2)	2
Hemoglobin ≤ 11.9 (g/dL)	3
Lymphocyte count ≤ 1.458 ($\times 10^3/\mu\text{L}$)	4

363

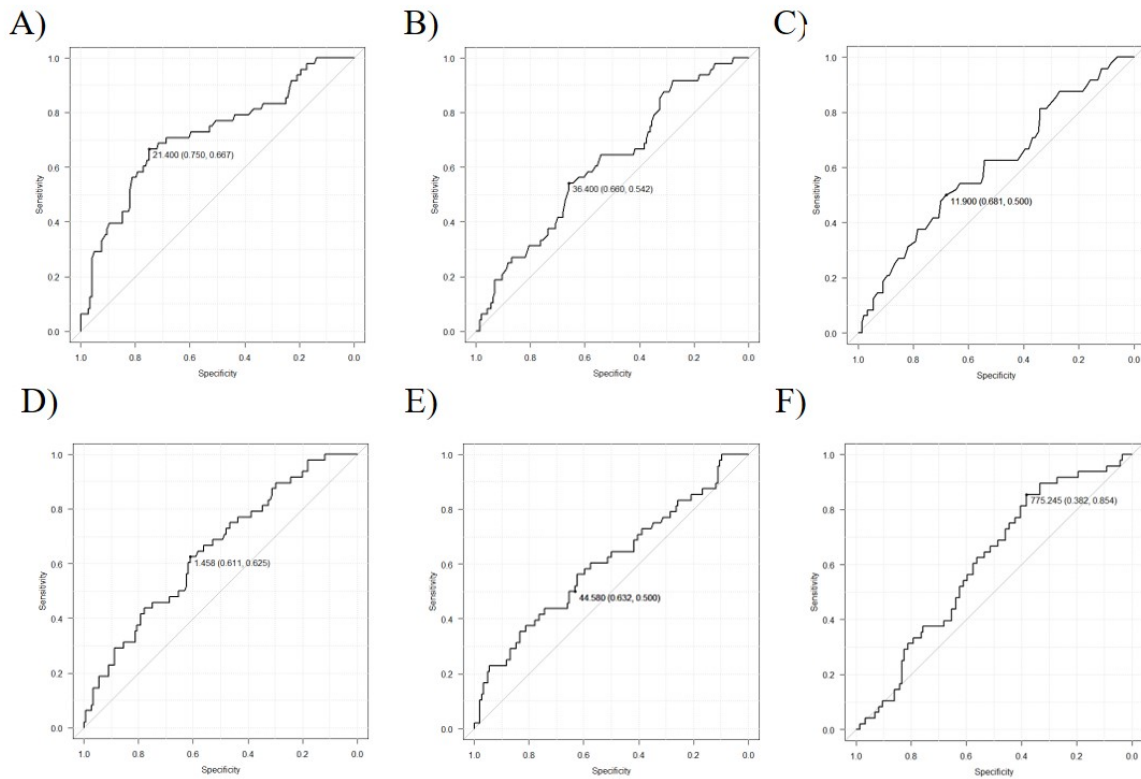
364



365 **Figure 1. Study design**

366 GC: gemcitabine plus cisplatin combination

367



368 **Figure 2. Receiver operating characteristic curves to calculate sensitivity and specificity at cut-off**
 369 **values.**

370 A) platelet count, B) hematocrit C) hemoglobin, D) lymphocyte count, E) PNI, F) dose of gemcitabine
 371 (day1)

372 Cut-off value for platelet count (A), hematocrit (B), hemoglobin (C), lymphocyte count (D), PNI (E) and
 373 dose of gemcitabine (F) was 21.4 (specificity: 0.750, sensitivity: 0.667), 36.4 (specificity: 0.660
 374 sensitivity: 0.542), 11.9 (specificity: 0.681, sensitivity: 0.500), 1.458 (specificity: 0.611 sensitivity 0.625),
 375 44.580 (specificity: 0.632, sensitivity: 0.500), and 775.245 (specificity: 0.382, sensitivity: 0.854),
 376 respectively. AUC and 95%CI of platelet count, hematocrit, hemoglobin, lymphocyte count, PNI before
 377 GC therapy and dose of gemcitabine was 0.718 (95%CI: 0.63 - 0.806), 0.612 (95%CI: 0.521 - 0.702),
 378 0.601 (95%CI: 0.508 - 0.694), 0.649 (95%CI: 0.561 - 0.737), 0.605 (95%CI: 0.509 - 0.702) and 0.604
 379 (95%CI: 0.517 - 0.691), respectively.

380 AUC: Area under the receiver operating characteristic curve, CI: confidence interval, PNI: prognostic
381 nutritional index
382