Assessment of Prognostic Scores for Emergency Department Patients With Upper Gastrointestinal Bleeding

Pierre-Clément Thiebaud, MSc*; Eliana Wassermann, PhD; Mathilde de Caluwe; Clément Prebin; Florent Noel, MD; Agnès Dechartres, PhD; Pierre-Alexis Raynal, MD; Judith Leblanc, PhD; Youri Yordanov, PhD

*Corresponding Author. E-mail: pierre-clement.thiebaud@aphp.fr.

Study objective: Early prognostic stratification could optimize the management of patients with upper gastrointestinal bleeding and reduce unnecessary hospitalizations. The aim of this study was to assess and compare the performance of existing prognostic scores in predicting therapeutic intervention and death.

Methods: A systematic search of the literature identified existing prognostic scores. A multicenter retrospective cohort study included adult patients hospitalized for upper gastrointestinal bleeding from January 1, 2019, to December 31, 2020. The primary outcome was a composite including therapeutic intervention within 7 days (blood transfusion, endoscopic, surgical, or interventional radiology hemostasis) and/or 30-day death. Discrimination performance was estimated by the area under the curve (AUC). The ability to identify low-risk patients was analyzed using sensitivity and negative predictive value (NPV) for defined thresholds.

Results: The systematic search identified 39 prognostic scores, 12 of which could be analyzed. Among the 990 patients included, therapeutic intervention and/or death occurred in 755 (76.4%) patients. Scores with the highest discriminative performance to predict the primary composite outcome were Glasgow-Blatchford score (GBS) (AUC 0.869 [0.842 to 0.895]), modified GBS (AUC 0.872 [0.847 to 0.898]) and modified GBS 2 (AUC 0.855 [0.827 to 0.884]). The best performance to identify low-risk patients was for GBS$\leq$1 (sensitivity 0.99 [0.99 to 1.00], NPV 0.89 [0.75 to 0.97]) and modified GBS=0 (sensitivity 0.99 [0.98 to 1.00], NPV 0.84 [0.71 to 0.94]).

Conclusions: The GBS and the modified GBS are the 2 best performing scores because they achieve both key objectives: stratifying patients based on their risk of therapeutic intervention and/or death and identifying low-risk patients who may qualify for outpatient management. [Ann Emerg Med. 2025;85:31-42.]

Please see page 32 for the Editor's Capsule Summary of this article.

Readers: click on the link to go directly to a survey in which you can provide **feedback** to *Annals* on this particular article.
A **podcast** for this article is available at www.annemergmed.com.

0196-0644/$-see front matter

https://doi.org/10.1016/j.annemergmed.2024.06.024

INTRODUCTION

Background

Upper gastrointestinal bleeding is a frequent and potentially fatal medical emergency, with an incidence of 60 to 120 per 100,000 adults per year and a mortality rate of 3% to 11%.[1-3] Upper gastrointestinal endoscopy is the standard diagnostic and therapeutic modality for upper gastrointestinal bleeding, but some patients do not require any specific therapeutic intervention during their hospital stay. Effective patient stratification based on the risk of therapeutic intervention or death could optimize the management of patients with upper gastrointestinal bleeding and reduce unnecessary hospitalizations. Several prognostic scores have been developed for this purpose, among which the Glasgow-Blatchford score (GBS) seems to be the most accurate to predict therapeutic intervention and/or death.[4,5] It can also be used to identify low-risk patients who could be discharged from the emergency department (ED) and managed as outpatients. European and international guidelines for nonvariceal upper gastrointestinal bleeding strongly and moderately recommend using the GBS with a threshold of less than or equal to 1 to identify these low-risk patients based on a moderate and low level of evidence, respectively.[6,7] Two other scores have been widely studied and externally validated, the preendoscopic Rockall and AIMS65 (albumin, international normalized ratio, mental status, systolic blood pressure, age over 65), but their use is not recommended (misclassification of 4% to 7% and 20% of high-risk patients

Editor's Capsule Summary

What is already known on this topic

Risk stratification scores may identify high-risk emergency department patients presenting with upper gastrointestinal bleeding.

What question this study addressed

Which scores best identify the patients with upper gastrointestinal bleeding who will undergo therapeutic interventions within 7 days or die within 30 days?

What this study adds to our knowledge

This retrospective study in 990 patients examined 12 existing upper gastrointestinal bleeding prediction tools. Three versions of the Glasgow-Blatchford Score performed best in predicting high-risk subjects, but 1 of these had limited utility in predicting low risk.

How this is relevant to clinical practice

No clear high-performing upper gastrointestinal bleeding complication prediction tool currently exists.

as low risk, respectively).[7-9] Many others have been proposed more recently but they have limited or no external validation. Some of them pertain only to specific populations, such as nonvariceal upper gastrointestinal bleeding. All these scores have been developed or validated in different settings (ED, endoscopy, or gastroenterology department) and with various outcomes, not all of which are clinically relevant. To identify the most relevant prognostic scores for current practice, it is necessary to assess and compare the performance of the various existing scores on the same population and with the same primary outcome. This approach is particularly important for the most recent scores that have not yet been validated.

Goals of This Investigation

The aim of this study was to systematically identify existing prognostic scores, assess and compare their performance in predicting therapeutic intervention or death, and identifying low-risk patients using a large multicenter cohort from the Greater Paris University Hospitals.

METHODS

Study Design and Settings

This is a retrospective cohort study using data from the Greater Paris University Hospitals health data warehouse, which collects data from patients in 38 hospitals spread across the Greater Paris area (more than 22,000 beds, 1.5 million hospitalizations each year), including 15 adult EDs. The overall database contains routine data of patients cared for in the Greater Paris University Hospitals. Data collected in the electronic health record software (ORBIS, Dedalus Healthcare) and in the claim database (French *Programme de Médicalisation des Systèmes d'Information*, coded following the International Classification of Diseases 10th revision [ICD-10]) are integrated in the Greater Paris University Hospitals health data warehouse on a daily and monthly basis, respectively. The research database follows the Informatics for Integrating Biology & the Bedside (i2b2) standard.[10] However, the health data warehouse does not monitor patients who are discharged after their ED visit or following their hospitalization. The health data warehouse data was used to select patients and access their electronic medical reports. The inclusion period was from January 1, 2019, to December 31, 2020, with a follow-up period of 30 days. This study was approved by the Greater Paris University Hospitals health data warehouse scientific and ethics committee (IRB00011591) and registered on clinicaltrials.gov (NCT05927493). The Reporting of Studies Conducted Using Observational Routinely-collected Data (RECORD) and Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statements have been followed for the reporting of this study (Appendix E1, available at http://www.annemergmed.com).[11,12]

Selection of Participants

Eligible patients were all adults (≥18 years of age) who were admitted to the hospital (length of stay ≥24 hours) after visiting the ED during the inclusion period, diagnosed with upper gastrointestinal bleeding based on the ICD-10 codes and for whom both ED and hospitalization reports were available in the health data warehouse (Appendix E1). Patients were not included if the diagnosis was not a suspected upper gastrointestinal bleeding based on the ED report. Patients could be included more than once after a 30-day latency period between 2 visits.

Score Identification and Selection

A systematic search of the literature was performed using Medline, Embase, and the Cochrane Library from inception to May 10, 2023. The search strategy, using the Ingui filter with Geersing update, is available in Appendix E1.[13,14] All studies presenting a new prognostic score in upper gastrointestinal bleeding patients have been included. No limitations to language, date, or design have been applied. The assessed studies included patients from various settings

with various inclusion criteria and different outcomes. The reference lists of included articles were reviewed for any relevant study not identified by our search. Studies selection and management were conducted using Rayyan© software.[15] Following the removal of duplicates, one investigator (PCT) screened the titles and abstracts. Two independent reviewers (PCT, PAR) then screened the full-text of eligible studies based on the specified inclusion criteria. An independent third investigator (YY) was consulted to resolve any discrepancies. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram was used to document the selection process.[16] However, the PRISMA checklist did not apply, as no data extraction was carried out apart from the score components.

Among identified studies, scores were eligible if their components were adequately reported, they used only clinical or biological variables (eg, no endoscopic or therapeutic data), and did not require gastric tube placement or computed tomography (both are not recommended in international guidelines). Scores were excluded if they included data that was not available in the health data warehouse, components with an expected high number of missing data as these variables were not routinely collected in these situations, or if the published information did not allow their precise calculation.

Data Collection and Management

Data extraction was performed using the Informatics for Integrating Biology and the Bedside (i2b2) platform (Appendix E1). Then a first patient selection was electronically performed using age, ICD-10 codes, period of interest, visit type, and availability of ED and hospitalization records. Final patient inclusion was manually performed by 3 independent physician extractors (PCT, MdC, and CP) to confirm inclusion and exclusion criteria. They were trained to collect data using a standardized electronic case report form within the REDCap® electronic data capture tools.[17] The extractors were not blinded to the study hypothesis. However, prognostic scores were calculated only retrospectively during statistical analysis. A data set of 10% of the included patients was randomly selected and double-checked by an independent assessor (FN) to assess the data's reliability.

The explanatory variables collected were those used in the selected scores or in at least 3 of the identified scores in the literature review. These variables were collected from the ED and/or biological reports, not from hospitalization reports. The vital signs and biological results collected were the first available values. The definitions of explanatory variables are specified in the Appendix E1. All these variables were used to calculate the selected prognostic scores. The type of hospitalization (conventional or intensive care), final diagnosis, timing to endoscopy, and length of stay were also collected.

Outcomes

The primary outcome was a predefined composite including a therapeutic intervention within 7 days of admission (defined as blood transfusion, endoscopic, surgical, or interventional radiology hemostasis) or 30-day death. Rebleeding was not included in the composite outcome because its definition varies between studies and it is not a relevant criterion if it does not result in a therapeutic intervention. The components of the primary composite outcome were those suggested to standardize the outcome of studies on gastrointestinal bleeding.[18] We decided to use this primary composite outcome, which appeared to be the most relevant in clinical practice, even if some of these scores have been developed with other outcomes. Moreover, the most popular scores have been studied with various outcomes over time. For example, the GBS initially included rebleeding in its outcome, but it was later removed from the main composite outcome in the study conducted by Stanley et al,[5] which serves as a reference for international guidelines. The secondary outcomes were blood transfusion within 7 days, endoscopic hemostasis within 7 days, and death within 30 days. Surgical and interventional radiology hemostasis were not studied as secondary outcomes due to the expected low number of events. Outcome measures were collected from ED or hospitalization reports. Patients discharged from the hospital within 7 days of their admission (before the end of the outcome measure period) were considered as not having a therapeutic intervention after discharge. Data from the national mortality registry are integrated into the health data warehouse and allows to track deaths occurring after hospital discharge, which made it a highly reliable measure.

Analysis

The extreme or outlier data points for vital signs and biological tests were double-checked. The baseline characteristics and outcomes of the included patients were described. Continuous variables were reported as medians with first and third quartiles. Categoric variables were reported as number and percentages.

The different investigated scores were calculated and their discrimination assessed, first to predict the primary composite outcome and then each secondary outcome. The areas under the curve (AUCs) were calculated with their corresponding 95% confidence intervals (95% CI) using the DeLong method. For each outcome, the AUC of the different scores were compared using the DeLong test.[19] These analyses were

first conducted on all upper gastrointestinal bleeding patients and then on a subgroup of patients not suspected of having portal hypertension (no liver disease). Receiver operating characteristic (ROC) curves were plotted for scores with good discriminative performance (AUC>0.80).

For thresholds found in the literature review of scores identifying low-risk patients and provided that its discriminative performance was good, the sensitivity, specificity, positive predictive value (PPV), and NPV were estimated. Confidence intervals for these estimates were calculated using the exact binomial method. The rate of patients identified as low risk was also reported.

The main analyses were performed on complete case data. Sensitivity analyses were performed for the primary composite outcome after imputation of missing data by multiple imputations with chained equations under the missing-at-random hypothesis. Inter-rater reliability regarding data collection was analyzed for 10% of the included patients using the intraclass correlation coefficient for continuous variables and the Cohen's kappa for categoric variables. All tests were 2 sided, and a P value $<.05$ was considered significant. Analyses were performed using the R software version 4.2.2 (packages readr, tableone, ggplot2, pROC, epiR, mice, and irr).

RESULTS

Score Selection

The literature review identified 5,929 records. Out of 3,981 records that were not duplicates, 3,850 were excluded based on title or abstract screening by one author. Two assessors reviewed the full-texts of the remaining 131 articles. This process identified 80 prognostic scores for upper gastrointestinal bleeding, of which 39 met the eligibility criteria (Figure E1, available at http://www.annemergmed.com). Finally, we included 12 scores in the study: pre-endoscopic Rockall, GBS, modified GBS, modified GBS 2, age-extended GBS, Canada-United Kingdom-Adelaide (CANUKA), Cologne-WATCH (C-WATCH), Horibe gastrointestinal bleeding (HARBINGER), hematemesis, heart rate, hemoglobin, blood pressure, blood urea nitrogen (H3B2), Kalula, N score, and modified N score (Table E1, available at http://www.annemergmed.com).[4,8,20-29] Noneligible scores, as well as eligible but excluded scores, and the reason for their noninclusion are presented in Tables E2 and E3 (available at http://www.annemergmed.com).

Patient Characteristics

Out of 1,814 eligible patients (corresponding to 1,814 ED visits in 1,594 unique patients), 824 were excluded (Figure 1). The characteristics of the 990 included patients are shown in Table 1. The median age was 70.7 years [57.6 to 81.4], and 653 (66%) were men. Among them, 832 patients (84.3%) had an upper gastrointestinal endoscopy, of which 36.8% (n=306/832) required endoscopic hemostasis. Overall, 350 patients (35.4%) were hospitalized in the intensive care unit, and 671 (68.0%) received a blood transfusion. The mortality rate at day 30 was 6.7% (n=66). Therapeutic intervention and/or death occurred in 755 patients (76.4%). The primary causes of bleeding were gastroduodenal ulcer or erosion in 324 patients (32.7%) and portal hypertension in 122 patients (12.3%).

Primary Composite Outcome

The discriminative performance of each score to predict therapeutic intervention and/or death is summarized in Table 2. In all upper gastrointestinal bleeding patients, the scores with good discriminative performance were the GBS (AUC 0.869 [0.842 to 0.895]), modified GBS (AUC 0.872 [0.847 to 0.898]), modified GBS 2 (AUC 0.855 [0.827 to 0.884]), age-extended GBS (AUC 0.854 [0.826 to 0.881]), and CANUKA (AUC 0.829 [0.799-0.859]). The corresponding ROC curves are shown in Figure 2. There was no significant difference between the AUC of the GBS, modified GBS, and modified GBS 2 (Table E4, available at http://www.annemergmed.com). The AUCs of the other scores were significantly lower. Results were similar in patients not suspected of having portal hypertension (n=778), with good discriminative performance for the same 5 scores and significant superiority of the GBS, modified GBS, and modified GBS 2 (Table E5, available at http://www.annemergmed.com).

Table 3 provides sensitivity, specificity, and PPV and NPV for scores with good discriminative performance to predict therapeutic intervention or death. The sensitivity was excellent (>0.97) and the NPV was moderate (<0.90) for the 5 scores and 8 thresholds analyzed. The best performance was noted for GBS≤1 (sensitivity 0.99 [0.99 to 1.00], NPV 0.89 [0.75 to 0.97]) and modified GBS=0 (sensitivity 0.99 [0.98 to 1.00], NPV 0.84 [0.71 to 0.94]). CANUKA ≤2 performs similarly but identifies around half as many low-risk patients. The contingency tables are available in Table E6 (available at http://www.annemergmed.com).

Sensitivity analysis with imputed data yielded similar results (Tables E7 and E8, available at http://www.annemergmed.com). Due to the high rate of missing data for 4 biological variables used only in the C-Watch score (C-reactive protein 45%, alanine transaminase 22%, leukocytes 8%, and platelets 7%), these variables were not imputed. The C-Watch score was therefore not included in the sensitivity analysis.

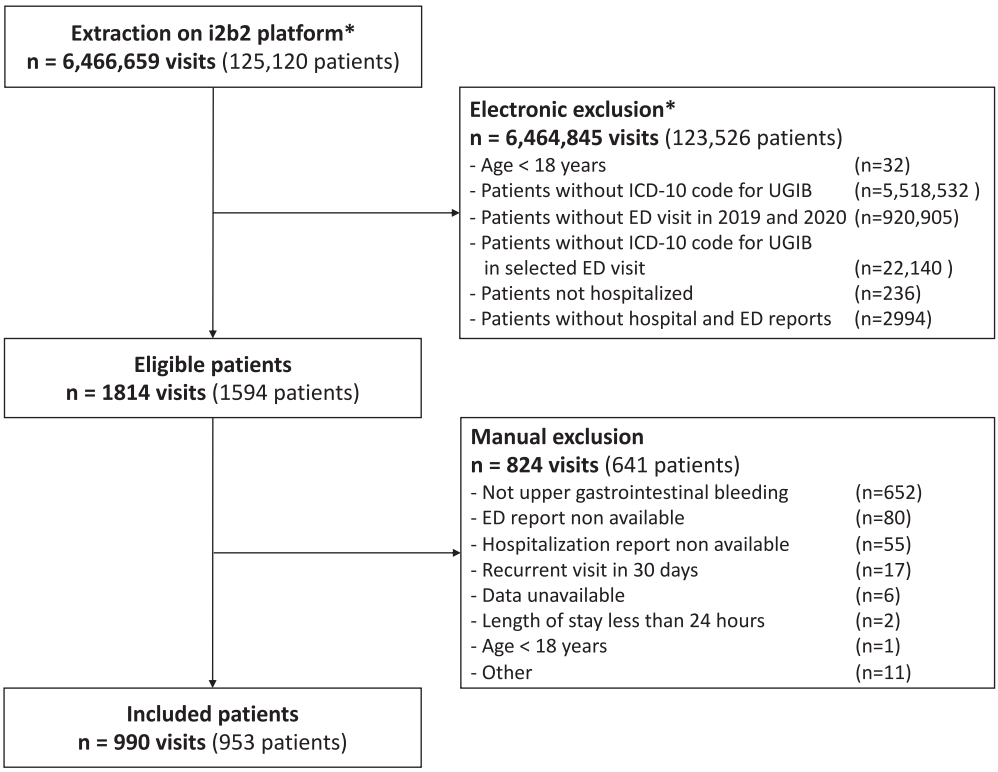

Figure 1. Flow-chart of participants. *Additional information is provided in Appendix E1.

Secondary Outcomes

The discriminative performance of each score to predict secondary outcomes is reported in Tables E9-E11 (available at http://www.annemergmed.com). The GBS, modified GBS, modified GBS 2, age-extended GBS and CANUKA showed good discriminative performance (AUC>0.80) in predicting blood transfusion but had poor discriminative performance in predicting death (AUC≤0.70) or endoscopic hemostasis (AUC≤0.60).

Inter-rater Reliability

Inter-rater reliability regarding data collection was substantial for presence of heart failure (κ=0.666 [0.572 to 0.760]), active malignancy (κ=0.648 [0.553 to 0.743]), melena (κ=0.764 [0.679 to 0.848]), syncope (κ=0.752 [0.666 to 0.838]), type of hospitalization (κ=0.777 [0.691 to 0.863]), and final diagnosis (κ=0.733 [0.645 to 0.822]) and very good or perfect for all other variables (Table E12, available at http://www.annemergmed.com).

LIMITATIONS

This study has several limitations, particularly due to its retrospective design. First, a selection bias could be associated with patient identification using ICD-10 coding, as some patients may be incorrectly or not referenced. Second, this study only included hospitalized patients. These patients are more severe than the overall population of patients visiting the ED with upper gastrointestinal bleeding, resulting in spectrum bias and an underestimation of the NPVs. This is linked to a logistical problem, as the medical records of patients discharged after their ED visit were not available in the health data warehouse. Third, the PRISMA statement was used to guide our literature search and the resulting findings. This systematic search was a necessary preliminary step to identify existing scores, but no data were extracted other than the scores themselves. Consequently, the PRISMA checklist did not apply. Fourth, numerous identified scores could not be evaluated as they contained variables that could not be retrospectively retrieved or with too many missing data. Furthermore, the excessive missing data for C-reactive protein (45%) precluded sensitivity analysis for the C-Watch score. Fifth, a statistically significant difference in AUC may not translate to clinical relevance. Similarly, differences in the performance of scores using a threshold to identify low-risk patients may not be clinically relevant. Sixth, the significance threshold was not adjusted for multiple comparisons between AUCs of different scores.

Table 1. Characteristics of patients hospitalized for upper gastrointestinal bleeding.*

Characteristics	Overall N=990	Missing Data, n (%)
Age, median (Q1-Q3)	70.7 (57.6-81.4)	
Men sex, n (%)	653 (66.0)	
Comorbidities, n (%)		
Heart failure	100 (10.1)	
Coronary disease	188 (19.0)	
Kidney failure	152 (15.4)	
Liver disease	212 (21.4)	
Malignancy	167 (16.9)	
Treatments, n (%)		
Nonsteroidal antiinflammatory	28 (2.8)	
Antiagregant	323 (32.6)	
Anticoagulant	242 (24.4)	
Proton pump inhibitor	323 (32.6)	
Symptoms, n (%)		
Hematemesis	326 (32.9)	
Melena	693 (70.0)	
Syncope	168 (17.0)	
Vital signs, n (%)		
Pulse rate (/min)	90 (76-104)	13 (1.3)
Systolic blood pressure (mm Hg)	122 (105-139)	8 (0.8)
Glasgow score	15 (15-15)	14 (1.4)
Laboratory results, median (Q1-Q3)		
Hemoglobinemia (g/dL)	8.5 (6.7-10.9)	6 (0.6)
Urea (mg/dL)	27.5 (17.7-44.3)	41 (4.1)
Creatininemia (mg/dL)	0.98 (0.77-1.40)	26 (2.6)
C-reactive protein (mg/dL)	0.8 (0.5-2.8)	450 (45.5)
Leukocytes (/μL)	8,950 (6,440-11,940)	81 (8.2)
Platelets (10^3/μL)	226 (158-308)	72 (7.3)
Alanine aminotransferase (U/L)	21 (14-36.5)	215 (21.7)
Hospitalization, n (%)		
Conventional	640 (64.6)	
Intensive unit	350 (35.4)	
Length of stay (days), median (Q1-Q3)	8 (5-13)	
Endoscopy, n (%)	832 (84.3)	3 (0.3)
Time to endoscopy (days), median (Q1-Q3) (n=832)	1 (0-2)	34 (3.4)
Diagnosis, n (%)		
Gastroduodenal ulcer or erosion	324 (32.7)	
Portal hypertension	122 (12.3)	
Esophagitis and Mallory-Weiss	88 (8.9)	
Malignancy	57 (5.8)	
Angiodysplasia	41 (4.1)	
Other upper gastrointestinal bleeding	47 (4.7)	
Lower gastrointestinal bleeding	60 (6.1)	
Nongastrointestinal bleeding	2 (0.2)	
Unknown	249 (25.2)	

Table 1. Continued.

Characteristics	Overall N = 990	Missing Data, n (%)
Therapeutic interventions at day 7, n (%)		
Blood transfusion	671 (68.0)	3 (0.3)
Endoscopic hemostasis	306 (31.0)	4 (0.4)
Interventional radiology hemostasis	22 (2.2)	3 (0.3)
Surgery	10 (1.0)	3 (0.3)
Death at day 30, n (%)	66 (6.7)	
Therapeutic intervention at day 7 or death at day 30, n (%)	755 (76.4)	
Scores, median (Q1-Q3)		
GBS	10 (7-13)	59 (6.0)
Modified GBS	9 (6-11)	59 (6.0)
Modified GBS 2	7 (4-9)	20 (2.0)
Extended GBS	12 (8-15)	59 (6.0)
Preendoscopic Rockall	3 (2-5)	13 (1.3)
CANUKA	8 (6-10)	59 (6.0)
H3B2	3 (2-4)	59 (6.0)
C-Watch	3 (3-4)	522 (52.7)
HARBINGER	1 (1-2)	57 (5.8)
N	2 (1-4)	43 (4.3)
Modified N	1 (0-3)	43 (4.3)
Kalula	2 (1-3)	45 (4.5)

Q, quartile.
*SI conversion factors: to convert hemoglobin to g/L, multiply by 10; to convert urea to mmol/L, multiply by 0.357; to convert creatinine to μmol/L, multiply by 88.4; to convert C-reactive protein to mg/L, multiply by 10; to convert leukocytes to 10^9/L, multiply by 0.001; to convert platelets to 10^9/L, multiply by 1; to convert alanine aminotransferase to μkat/L, multiply by 0.0167.

Table 2. Discriminative performance of prognostic scores to predict therapeutic intervention and/or death in patients with upper gastrointestinal bleeding (case complete data).

Scores	All UGIB Patients N = 990 AUC (95% CI)	UGIB Patients Not Suspected of Having Portal Hypertension N = 778 AUC (95% CI)
All UGIB scores		
Glasgow-Blatchford	**0.87*** (0.84-0.89)	**0.87** (0.84-0.90)
Modified Glasgow-Blatchford	**0.87** (0.85-0.90)	**0.88** (0.85-0.91)
Age-extended Glasgow-Blatchford	**0.85** (0.83-0.88)	**0.86** (0.83-0.89)
CANUKA	**0.83** (0.80-0.86)	**0.84** (0.80-0.87)
C-Watch	0.75 (0.69-0.80)	0.75 (0.70-0.80)
Pre-endoscopic Rockall	0.66 (0.62-0.70)	0.67 (0.62-0.71)
HARBINGER	0.58 (0.55-0.62)	0.59 (0.54-0.63)
Nonvariceal UGIB scores		
Modified Glasgow-Blatchford 2	**0.86** (0.83-0.88)	**0.86** (0.83-0.89)
Kalula	0.79 (0.76-0.82)	0.80 (0.76-0.83)
H3B2	0.75 (0.72-0.79)	0.75 (0.71-0.79)
Modified N score	0.65 (0.61-0.69)	0.64 (0.59-0.68)
N score	0.56 (0.52-0.60)	0.55 (0.50-0.60)

CI, confidence interval; *UGIB*, upper gastrointestinal bleeding.
*Values in bold indicate AUC>0.80.

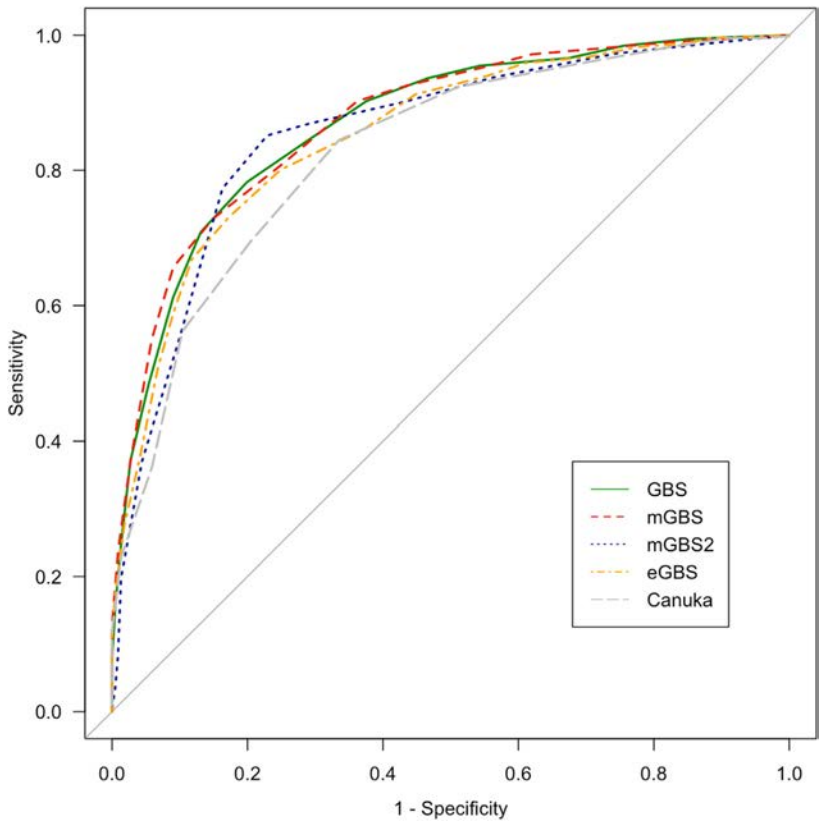

Figure 2. ROC curves for scores with AUC>0.80 to predict therapeutic intervention or death in patients with upper gastrointestinal bleeding. *mGBS*, modified Glasgow-Blatchford score; *eGBS*, extended Glasgow-Blatchford score.

Seventh, the study is limited to a French region, which may limit the generalizability of the results. However, it was conducted in a large number of EDs in Paris and Île-de-France, the most populous region in France, thus ensuring a diverse population. Eighth, inter-rater reliability regarding data collection was only substantial for heart disease, active malignancy, melena, and syncope. Although this may modify the GBS, it has no effect on the modified GBS. Ninth, the composite primary outcome is not the one used during the derivation of the different scores, but it seemed the most clinically relevant to meet the study's objective. Finally, outcomes were

Table 3. Sensitivity, specificity, PPV, and NPV of different scores and cut-offs to stratify patients with upper gastrointestinal bleeding according their risk of therapeutic intervention and/or death (case complete data).

	Low-risk Patients		High-risk Patients		Sensitivity	Specificity	PPV	NPV
Score	**Cutoff**	**n (%)**	**Cutoff**	**n (%)**	**% (95% CI)**	**% (95% CI)**	**% (95% CI)**	**% (95% CI)**
GBS	≤1	37 (4.0)	>1	892 (96.0)	0.99 (0.99-1.00)	0.15 (0.11-0.20)	0.79 (0.76-0.82)	0.89 (0.75-0.97)
	≤2	65 (7.0)	>2	864 (93.0)	0.98 (0.97-0.99)	0.24 (0.19-0.31)	0.81 (0.78-0.83)	0.83 (0.72-0.91)
mGBS	0	45 (4.8)	>0	884 (95.2)	0.99 (0.98-1.00)	0.17 (0.12-0.23)	0.79 (0.76-0.82)	0.84 (0.71-0.94)
	≤1	81 (8.7)	>1	848 (91.3)	0.98 (0.97-0.99)	0.30 (0.24-0.36)	0.82 (0.79-0.84)	0.81 (0.71-0.89)
mGBS2	≤1	78 (8.1)	>1	890 (91.9)	0.97 (0.96-0.98)	0.26 (0.20-0.32)	0.81 (0.78-0.84)	0.74 (0.63-0.84)
eGBS	≤2	43 (4.6)	>2	886 (95.4)	0.99 (0.98-0.99)	0.15 (0.11-0.21)	0.79 (0.76-0.82)	0.79 (0.64-0.90)
CANUKA	≤1	11 (1.2)	>1	918 (98.8)	0.99 (0.98-0.99)	0.04 (0.02-0.07)	0.77 (0.74-0.79)	0.73 (0.39-0.94)
	≤2	23 (2.5)	>2	906 (97.5)	1.00 (0.99-1.00)	0.09 (0.06-0.14)	0.78 (0.75-0.80)	0.87 (0.66-0.97)

eGBS, extended Glasgow-Blatchford score; *mGBS*, modified Glasgow-Blatchford score.

analyzed on day 7 for interventions and day 30 for death, without considering the precise timing of each intervention. However, the timing of interventions could be relevant for stratifying patients and guiding management decisions during initial phase of care.

DISCUSSION

In this study, we assessed and compared the performance of multiple scores in identifying patients at risk of therapeutic intervention and/or death in an upper gastrointestinal bleeding population. The GBS and its modified versions (modified GBS, modified GBS 2) have the highest discriminative performance in this indication. These scores could contribute to optimize patient management according to their individual level of risk. More than global risk stratification, the identification of low-risk patients for therapeutic intervention or death is a significant issue in EDs because these patients could be managed as outpatients, thereby reducing unnecessary hospitalizations and health care costs. For that purpose, GBS$\leq$1 and modified GBS=0 had the best results with excellent sensitivity, but no score showed good NPV. The advantage of the modified GBS is that it excludes subjective items, such as comorbidities (cardiac failure, liver disease) and symptoms (melena, syncope).

Regarding discrimination, the results for secondary outcomes indicate that the good performance of these different scores is mainly driven by the prediction of blood transfusion. None of the analyzed scores appeared satisfactory for death prediction. According to the literature, AIMS65 often outperformed the GBS and pre-endoscopic Rockall in predicting death.[30,31] The more recent Age, Blood tests and Comorbidities (ABC) score may perform even better.[32,33] However, both AIMS65 and ABC include data that are not routinely collected in EDs, such as albumin or the international normalized ratio for AIMS65, and albumin or the American Society of Anesthesiologists (ASA) score for ABC. The ASA score is primarily used by anesthesiologists and is based on multiple factors that cannot be reliably retrieved retrospectively. Consequently, AIMS65 and ABC scores could not be analyzed in our study. No score has yet been validated in the literature to accurately predict the occurrence of hemostatic procedure.[30,32,34] Several additional scoring systems have been proposed in recent years, with disappointing results or awaiting validation.[23,24,28,35-37] The better performance of the assessed scores in predicting blood transfusion than other outcomes has already been reported. As reported by Oakland, the GBS performed well to predict blood transfusion (AUCs 0.77 to 0.93) and was moderate to predict rebleeding (AUCs 0.63 to 0.75), endoscopic hemostasis (AUCs 0.58 to 0.78), and need for surgery or interventional radiology (AUCs 0.61 to 0.71).[31] For the CANUKA score, Oakland et al[27] reported an AUC of 0.81 for blood transfusion, 0.70 for death, and 0.61 for adverse outcome excluding transfusion. This can be explained, on one hand, by the fact that hemoglobin is the variable that contributes the most points in these scores, with other variables being also correlated with the severity of bleeding. On the other hand, blood transfusion is the most frequently occurring event in the composite criteria. It is possible that this was also the case in the derivation study of the GBS, although these data are not available.[4] Death and need for endoscopic hemostasis are more complex outcomes. Death is influenced by various factors beyond initial presentation particularly because it was assessed at 30 days. Endoscopic hemostasis decisions may depend on the timing of the endoscopy or expertise of the medical team, which cannot be captured by these scores. The superiority of the GBS compared with other scores in predicting blood transfusion has already been reported.[32] However, transfusion could probably be predicted by hemoglobin and vital parameters, without the need for a score, as already suggested by Horibe et al.[21] Moreover, the decision to transfuse is subjective and may not be associated with the need for a hemostatic procedure or occurrence of death. The choice of a criterion excluding transfusion should therefore be considered when studying prognostic scores for the initial stratification of patients in EDs.

Regarding the identification of low-risk patients, our results should be interpreted with caution because the performance of the scores has been estimated on hospitalized patients only. Specificity and NPV are likely underestimated compared with their calculation in all ED patients, including discharged ones with presumed favorable outcomes. Previous studies including all ED patients reported good results for the GBS and CANUKA, supporting the use of these scores to identify low-risk patients eligible for safe discharge.[5,27,38] The GBS, with a threshold of less than or equal to 1, is considered the best performing score for this purpose, as confirmed by our results. Its use is recommended in the latest international guidelines.[6,7,30,32,39,40] It would identify approximately 20% of low-risk patients.[5,27,38] In our study including only hospitalized patients, the use of the GBS could have avoided a maximum of 4% of hospitalizations. This number may have been affected by the noninclusion of patients hospitalized with suspected upper gastrointestinal bleeding but not identified by the selected ICD-10 codes, which could result in a higher risk study sample. However, this rate was 7.5% in another recent French study of patients identified by their reason for consultation.[38] In the

present study, the majority of low-risk patients were probably already discharged from the ED even if a score is rarely used. The emergency physicians' clinical judgment could therefore be as efficient as prognostic scores, as already shown for endoscopists but this remains to be studied.[41]

In addition, there may be a distinction between patients who had no therapeutic intervention or death and those who could have actually been discharged. A patient may require hospitalization for another reason, whether related to upper gastrointestinal bleeding or not. It is also possible to transfuse a patient in the ED and then manage them as an outpatient. Ultimately, the clinical judgment of the emergency physician should prevail in deciding the patient's disposition.

In the future, it will be necessary to conduct high-quality interventional prospective studies to analyze the clinical effect, the safety and the economic influence of the implementation of a prognostic score using a relevant clinical outcome and involving the follow-up of all patients, including those discharged from the ED. Alternatively, it might be useful to derive a new score with the sole objective of identifying patients eligible for a safe discharge. Another approach could be to accelerate the integration of artificial intelligence tools in clinical practice. For example, a recent study showed that a machine learning model outperformed the GBS in predicting therapeutic interventions and death (AUC 0.90 versus 0.87, P=.004) and identifying low-risk patients.[42] Although machine learning models can potentially offer improved predictive performance, they often lack interpretability, which can cause trust issues for physicians who may prefer models with transparent decisionmaking processes that they can understand, such as the GBS. Integrating machine learning models into electronic health record systems also presents challenges, as it requires significant modifications to existing hospital computer systems and workflows. The use of machine learning for risk prediction in the context of upper gastrointestinal bleeding is still in its early stages, and further research is needed to refine these models, validate their performance in diverse clinical settings, and address concerns related to interpretability, trust, and seamless integration into clinical decision support systems within electronic health records.[43]

In summary, the GBS and the modified GBS are the 2 best performing scores because they may be able to achieve both key objectives: stratifying patients based on their risk of therapeutic intervention or death and identifying low-risk patients who may be likely to qualify for outpatient management. Although the modified GBS2 performs equally well for overall stratification, only the GBS and modified GBS appear to perform somewhat better in identifying low-risk patients.

The authors would like to thank David Hajage (Sorbonne Université, INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique, UMR-S 1136) for his methodological support and Vanessa Lemaître (Sorbonne Université, Improving Emergency Care University Hospital Federation) for her logistical support.

Supervising editor: Allan B. Wolfson, MD. Specific detailed information about possible conflict of interest for individual editors is available at https://www.annemergmed.com/editors.

Author affiliations: From the Université de Paris Cité (Thiebaud), INSERM, UMR-S 942, Improving Emergency Care University Hospital Federation, Paris, France; Sorbonne Université (Thiebaud, Yordanov), INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique, UMR-S 1136, AP-HP, Hôpital Saint-Antoine, Service d'Accueil des Urgences, Paris, France; AP-HP.Sorbonne Université (Wassermann), Hôpital Saint-Antoine, Unité de recherche clinique de l'Est Parisien, Paris, France; AP-HP.Sorbonne Université (De Caluwe, Prebin, Noel, Raynal), Hôpital Saint-Antoine, Service d'Accueil des Urgences, Paris, France; Sorbonne Université (Dechartres), INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique, UMR-S 1136, AP-HP, Hôpital Pitié-Salpêtrière, Département de Santé Publique, Paris, France; Sorbonne Université (Leblanc), INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique, AP-HP, Hôpital Saint-Antoine, Unité de recherche clinique Est Parisien, Paris, France.

Author contributions: PCT, JL, and YY conceived the study, designed the trial, and obtained legal permissions. PCT supervised the conduct of the trial. AD was involved in the systematic review of the literature. PCT, PAR, and YY reviewed the literature. EW and JL performed the selection of patient data. PCT, MDC, CP, and FN collected data. PCT managed and analyzed the data. PCT and YY were involved in data interpretation and drafted the initial manuscript. All authors contributed substantially to its revision. PCT takes responsibility for the manuscript as a whole.

Data sharing statement: Data supporting this study can be made available on reasonable request to the corresponding author, on condition that the research project is accepted by the scientific and ethics committee of the Greater Paris University Hospital health data warehouse.

All authors attest to meeting the four ICMJE.org authorship criteria: (1) Substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work; AND (2) Drafting the work or revising it critically for important intellectual content; AND (3) Final approval of the version to be published; AND (4) Agreement to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding and support: By *Annals'* policy, all authors are required to disclose any and all commercial, financial, and other relationships in any way related to the subject of this article as per ICMJE conflict

of interest guidelines (see www.icmje.org). This study was conducted without financial support. The authors have no conflict of interest relevant to this article to disclose.

Publication dates: Received for publication May 8, 2024. Revisions received June 8, 2024, and June 17, 2024. Accepted for publication June 20, 2024.

Trial registration number: This study was registered on clinicaltrials.gov (NCT05927493).

REFERENCES

1. Laine L, Yang H, Chang SC, et al. Trends for incidence of hospitalization and death due to GI complications in the United States from 2001 to 2009. *Am J Gastroenterol*. 2012;107:1190-1196.
2. Thiebaud P-C, Yordanov Y, Galimard J-E, et al. Management of upper gastrointestinal bleeding in emergency departments, from bleeding symptoms to diagnosis: a prospective, multicenter, observational study. *Scand J Trauma Resusc Emerg Med*. 2017;25:78.
3. Nahon S, Hagège H, Latrive JP, et al. Epidemiological and prognostic factors involved in upper gastrointestinal bleeding: results of a French prospective multicenter study. *Endoscopy*. 2012;44:998-1008.
4. Blatchford O, Murray WR, Blatchford M. A risk score to predict need for treatment for upper-gastrointestinal haemorrhage. *Lancet*. 2000;356:1318-1321.
5. Stanley AJ, Laine L, Dalton HR, et al. Comparison of risk scoring systems for patients presenting with upper gastrointestinal bleeding: international multicentre prospective study. *BMJ*. 2017;356:i6432.
6. Gralnek IM, Stanley AJ, Morris AJ, et al. Endoscopic diagnosis and management of nonvariceal upper gastrointestinal hemorrhage (NVUGIH): European Society of Gastrointestinal Endoscopy (ESGE) Guideline - Update 2021. *Endoscopy*. 2021;53:300-332.
7. Barkun AN, Almadi M, Kuipers EJ, et al. Management of nonvariceal upper gastrointestinal bleeding: guideline recommendations from the international consensus Group. *Ann Intern Med*. 2019;171:805-822.
8. Rockall TA, Logan RF, Devlin HB, et al. Risk assessment after acute upper gastrointestinal haemorrhage. *Gut*. 1996;38:316-321.
9. Saltzman JR, Tabak YP, Hyett BH, et al. A simple risk score accurately predicts in-hospital mortality, length of stay, and cost in acute upper GI bleeding. *Gastrointest Endosc*. 2011;74:1215-1224.
10. Murphy SN, Weber G, Mendis M, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc*. 2010;17(2):124-130.
11. Snell KIE, Levis B, Damen JAA, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). *BMJ*. 2023;381:e073538.
12. Benchimol EI, Smeeth L, Guttmann A, et al. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) statement. *PLoS Med*. 2015;12:e1001885.
13. Ingui BJ, Rogers MA. Searching for clinical prediction rules in MEDLINE. *J Am Med Inform Assoc*. 2001;8:391-397.
14. Geersing GJ, Bouwmeester W, Zuithoff P, et al. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS One*. 2012;7:e32844.
15. Ouzzani M, Hammady H, Fedorowicz Z, et al. Rayyan-a web and mobile app for systematic reviews. *Syst Rev*. 2016;5:210.
16. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
17. Harris PA, Taylor R, Thielke R, et al. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42:377-381.
18. Jensen DM, Barkun A, Cave D, et al. Acute gastrointestinal bleeding: proposed study outcomes for new randomised controlled trials. *Aliment Pharmacol Ther*. 2021;54:616-626.
19. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44:837-845.
20. Cheng DW, Lu YW, Teller T, et al. A modified Glasgow Blatchford Score improves risk stratification in upper gastrointestinal bleed: a prospective comparison of scoring systems. *Aliment Pharmacol Ther*. 2012;36:782-789.
21. Horibe M, Kaneko T, Yokogawa N, et al. A simple scoring system to assess the need for an endoscopic intervention in suspected upper gastrointestinal bleeding: A prospective cohort study. *Dig Liver Dis*. 2016;48:1180-1186.
22. Hoffmann V, Neubauer H, Heinzler J, et al. A novel easy-to-use prediction scheme for upper gastrointestinal bleeding: cologne-WATCH (C-WATCH) Risk Score. *Medicine (Baltimore)*. 2015;94:e1614.
23. Ito N, Funasaka K, Furukawa K, et al. A novel scoring system to predict therapeutic intervention for non-variceal upper gastrointestinal bleeding. *Intern Emerg Med*. 2022;17:423-430.
24. Ito N, Funasaka K, Fujiyoshi T, et al. Modified N score is helpful for identifying patients who need endoscopic intervention among those with black stools without hematemesis. *Dig Endosc*. 2022;34:1157-1165.
25. Laursen SB, Hansen JM, Schaffalitzky de Muckadell OB. The Glasgow Blatchford score is the most accurate assessment of patients with upper gastrointestinal hemorrhage. *Clin Gastroenterol Hepatol*. 2012;10:1130-1135.e1.
26. Kalula SZ, Swingler GH, Louw JA. Clinical predictors of outcome in acute upper gastrointestinal bleeding. *S Afr Med J*. 2003;93:286-290.
27. Oakland K, Kahan BC, Guizzetti L, et al. Development, validation, and comparative assessment of an international scoring system to determine risk of upper gastrointestinal bleeding. *Clin Gastroenterol Hepatol*. 2019;17:1121-1129.e2.
28. Sasaki Y, Abe T, Kawamura N, et al. Prediction of the need for emergency endoscopic treatment for upper gastrointestinal bleeding and new score model: a retrospective study. *BMC Gastroenterol*. 2022;22:337.
29. Romagnuolo J, Barkun AN, Enns R, et al. Simple clinical predictors may obviate urgent endoscopy in selected patients with nonvariceal upper gastrointestinal tract bleeding. *Arch Intern Med*. 2007;167:265-270.
30. Tham J, Stanley A. Clinical utility of pre-endoscopy risk scores in upper gastrointestinal bleeding. *Expert Rev Gastroenterol Hepatol*. 2019;13:1161-1167.
31. Oakland K. Risk stratification in upper and upper and lower GI bleeding: which scores should we use? *Best Pract Res Clin Gastroenterol*. 2019;42-43:101613.
32. Orpen-Palmer J, Stanley AJ. A review of risk scores within upper gastrointestinal bleeding. *J Clin Med*. 2023;12:3678.
33. Laursen SB, Oakland K, Laine L, et al. ABC score: a new risk score that accurately predicts mortality in acute upper and lower gastrointestinal bleeding: an international multicentre study. *Gut*. 2021;70:707-716.
34. Yang L, Sun R, Wei N, et al. Systematic review and meta-analysis of risk scores in prediction for the clinical outcomes in patients with acute variceal bleeding. *Ann Med*. 2021;53:1806-1815.
35. Jin X, Wang X, Mao P. Early identification of high-risk factors for upper gastrointestinal bleeding. *Evid Based Complement Alternat Med*. 2022;2022:5641394.
36. Iino C, Mikami T, Igarashi T, et al. Evaluation of scoring models for identifying the need for therapeutic intervention of upper gastrointestinal bleeding: A new prediction score model for Japanese patients. *Dig Endosc*. 2016;28:714-721.
37. Marks I, Janmohamed IK, Malas S, et al. Derivation and validation of a novel risk score to predict need for haemostatic intervention in acute

upper gastrointestinal bleeding (London Haemostat Score). *BMJ Open Gastroenterol*. 2023;10:e001008.
38. Pognonec C, Dirhoussi Z, Cury N, et al. External validation of Glasgow-Blatchford, modified Glasgow-Blatchford and CANUKA scores to identify low-risk patients with upper gastrointestinal bleeding in emergency departments: a retrospective cohort study. *Emerg Med J*. 2023;40:451-457.
39. Siau K, Hearnshaw S, Stanley AJ, et al. British Society of Gastroenterology (BSG)-led multisociety consensus care bundle for the early clinical management of acute upper gastrointestinal bleeding. *Frontline Gastroenterol*. 2020;11:311-323.
40. Laine L, Barkun AN, Saltzman JR, et al. ACG clinical guideline: upper gastrointestinal and ulcer bleeding. *Am J Gastroenterol*. 2021;116:899-917.
41. Brullet E, Garcia-Iglesias P, Calvet X, et al. Endoscopist's judgment is as useful as risk scores for predicting outcome in peptic ulcer bleeding: a multicenter study. *J Clin Med*. 2020;9:408.
42. Shung DL, Au B, Taylor RA, et al. Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. *Gastroenterology*. 2020;158:160-167.
43. Nigam GB, Murphy MF, Travis SPL, et al. Machine learning in the assessment and management of acute gastrointestinal bleeding. *BMJ Med*. 2024;3:e000699.