Comparison of risk scoring systems for patients presenting with upper gastrointestinal bleeding
international multicentre prospective study

Stanley, Adrian J; Laine, Loren; Dalton, Harry R; Ngu, Jing H.; Schultz, Michael; Abazi, Rozeta; Zakko, Liam; Thornton, Susan; Wilkinson, Kelly; Khor, Cristopher J L; Murray, Iain A; Laursen, Stig Borbjerg; International Gastrointestinal Bleeding Consortium

Published in:
B M J

DOI:
10.1136/bmj.i6432

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Document license
CC BY-NC

Citation for published version (APA):
Comparison of risk scoring systems for patients presenting with upper gastrointestinal bleeding: international multicentre prospective study

Adrian J Stanley,1 Loren Laine,2 Harry R Dalton,3 Jing H Ngu,4 Michael Schultz,5,7 Roseta Abazi,6 Liam Zakko,2 Susan Thornton,1 Kelly Wilkinson,3 Cristopher J L Khor,4 Iain A Murray,1 Stig B Laursen6 on behalf of the International Gastrointestinal Bleeding Consortium

ABSTRACT

OBJECTIVE
To compare the predictive accuracy and clinical utility of five risk scoring systems in the assessment of patients with upper gastrointestinal bleeding.

DESIGN
International multicentre prospective study.

SETTING
Six large hospitals in Europe, North America, Asia, and Oceania.

PARTICIPANTS
3012 consecutive patients presenting over 12 months with upper gastrointestinal bleeding.

MAIN OUTCOME MEASURES
Comparison of pre-endoscopy scores (admission Rockall, AIMS65, and Glasgow Blatchford) and post-endoscopy scores (full Rockall and PNED) for their ability to predict predefined clinical endpoints: a composite endpoint (transfusion, endoscopic treatment, surgery, or 30 day mortality), endoscopic treatment, 30 day mortality, rebleeding, and length of hospital stay. Optimum score thresholds to identify low risk and high risk patients were determined.

RESULTS
The Glasgow Blatchford score was best (area under the receiver operating characteristic curve (AUROC) 0.86) at predicting intervention or death compared with the full Rockall score (0.70), PNED score (0.69), admission Rockall score (0.66, and AIMS65 score (0.68) (all P<0.001). A Glasgow Blatchford score of ≤1 was the optimum threshold to predict survival without intervention (sensitivity 98.6%, specificity 34.6%). The Glasgow Blatchford score was better at predicting endoscopic treatment (AUROC 0.75) than the AIMS65 (0.62) and admission Rockall scores (0.61) (both P<0.001). A Glasgow Blatchford score of ≥7 was the optimum threshold to predict endoscopic treatment (sensitivity 80%, specificity 57%). The PNED (AUROC 0.77) and AIMS65 scores (0.77) were best at predicting mortality, with both superior to admission Rockall score (0.72) and Glasgow Blatchford score (0.64; P<0.001). Score thresholds of ≥2 for PNED, ≥2 for AIMS65, ≥4 for admission Rockall, and ≥5 for full Rockall were optimal at predicting death, with sensitivities of 65.8-78.6% and specificities of 65.0-65.3%. No score was helpful at predicting rebleeding or length of stay.

CONCLUSIONS
The Glasgow Blatchford score has high accuracy at predicting need for hospital based intervention or death. Scores of ≤1 appear the optimum threshold for directing patients to outpatient management. AUROCs of scores for the other endpoints are less than 0.80, therefore their clinical utility for these outcomes seems to be limited.

TRIAL REGISTRATION
Current Controlled Trials ISRCTN16235737.

Introduction
Upper gastrointestinal bleeding is a common cause of admission to hospital worldwide, with a UK incidence of 103-172 per 100 000 adults each year and mortality of 8-14%.1-3 Many risk assessment scores have been developed to predict clinically relevant outcomes, including mortality, need for hospital based intervention, rebleeding, and length of hospital stay.4-9 Several of these, including the Rockall score and recently described progetto nazionale emorragia digestive (PENED) score require endoscopy before calculation.4,5 The Rockall score has been widely reported and it has been suggested that the PNED score is superior to it, although it has not been externally validated.7 Other endoscopy based scores have been described but are not appropriate for unselected patients with upper gastrointestinal bleeding, have not been externally validated, or have been shown to be inferior to the Rockall score or PNED score.10 However, requiring endoscopy to calculate a score might delay risk assessment in...
some healthcare settings, as there can be considerable delays in performing endoscopy out of hours or on weekends.\(^\text{11}\)

Recently, much interest has been shown in pre-endoscopic risk scores for upper gastrointestinal bleeding, which can be calculated shortly after presentation to hospital. The most widely studied scores are the abbreviated “admission” Rockall score, Glasgow Blatchford score, and recently described AIMS65 (albumin level <30 g/L (A), international normalised ratio >1.5 (I), altered mental status (M), systolic blood pressure <90 mm Hg (S), and age >65 years (65)) score.\(^\text{1,2}\) These scoring systems use clinical, haemodynamic, and (for Glasgow Blatchford and AIMS65 scores) readily available laboratory variables. Some studies have suggested that these scores could be used to identify patients at very low risk who could be managed as outpatients.\(^\text{12-14}\) Studies have also suggested that these scores could identify patients at higher risk who might require urgent endoscopy or management in high dependency or intensive care units.\(^\text{15,16}\)

Several studies have compared scores in their ability to predict various outcomes.\(^\text{12-16,39}\) Although the Glasgow Blatchford score seems useful in identifying very low risk patients in the UK, many studies have been small and single centre and no international study has compared all the commonly used scores. Recent guidelines have suggested use of risk scores for patients with upper gastrointestinal bleeding, but uncertainty remains about their exact role in clinical practice.\(^\text{20-23}\) In this prospective international study we compared five endoscopic and pre-endoscopic risk assessment scores for their ability to predict clinically relevant endpoints. In addition we assessed the clinical utility of these scores, by determining optimal thresholds for identifying patients at very low risk who could be managed as outpatients, and higher risk patients who might require specific management strategies aimed at improving outcome.

**Methods**

The study was undertaken in Yale-New Haven Hospital (USA), Glasgow Royal Infirmary (Scotland), Royal Cornwall Hospital Truro (England), Odense University Hospital (Denmark), Singapore General Hospital (Singapore), and Dunedin Hospital (New Zealand).

The predetermined clinical endpoints were the composite endpoint of need for hospital based intervention (red blood cell transfusion, endoscopic treatment, interventional radiology, or surgery) or death, endoscopic treatment, 30 day mortality, rebleeding within seven days, and length of hospital stay.

**Patient management**

Patients were included if they presented to the hospital with evidence of upper gastrointestinal bleeding defined by haematemesis, coffee-ground vomiting, or melaena. We excluded patients who developed upper gastrointestinal bleeding while an inpatient for another reason.

All patients presenting to each hospital with upper gastrointestinal bleeding were initially assessed in the emergency department or acute assessment unit. Two centres (Glasgow and Odense) had a policy of non-admission for those with Glasgow Blatchford scores of 1 or less, and one centre (Truro) for those with Glasgow Blatchford scores of 2 or less and aged less than 70 years, unless required for other reasons. Outpatient endoscopy was arranged in all three centres. Proton pump inhibitors (PPIs) were not routinely given to all admitted patients before endoscopy. The practice in all centres was to perform endoscopy within 24 hours in all admitted patients where possible. After endoscopy, the policy in all centres was to administer high dose PPVs by intravenous bolus followed by infusion to patients with high risk ulcer stigmata who required endoscopic treatment, and to other selected patients depending on clinical judgment.\(^\text{20-23}\) For patients with suspected variceal bleeding, the policy in all centres was to give intravenous vasopressors and antibiotics before endoscopy.\(^\text{24}\)

The endoscopic practice in all centres for patients with high risk stigmata of non-variceal bleeding was by injection of dilute adrenaline (epinephrine) into and around the bleeding point, thermal contact or clips, or both, but not adrenaline alone.\(^\text{20-23}\) Band ligation or injection of tissue glue with or without transjugular intrahepatic portosystemic stent shunt was performed in cases of oesophageal or gastric variceal bleeding, respectively.\(^\text{24}\) In line with recent evidence and guidelines, the policy was to administer red blood cells at a haemoglobin threshold of 70-80 g/L (7-8 g/dL), or as guided by the clinician in patients with severe haemorrhage.\(^\text{24,25}\)

**Data collection**

At each centre, data on consecutive, unselected patients presenting to the hospital with upper gastrointestinal bleeding were collected over a 12 month period between March 2014 and March 2015. Data on patients managed as outpatients in Glasgow, Truro, and Odense were also included. A dedicated research nurse, doctor, or medical student collected data at each site. The data collected included patient characteristics and haemodynamic and laboratory variables at presentation necessary to calculate the full Rockall, admission Rockall, Glasgow Blatchford, AIMS65, and PNED scores (see supplementary table 1). Endoscopic findings were recorded.

Blood transfusion, endoscopic treatment, and interventional radiology or surgery were recorded, as was rebleeding within seven days, as previously defined (see supplementary table 2),\(^\text{30}\) 30 day mortality, and length of hospital stay. Supplementary table 3 shows the collected data.

**Statistical analysis**

We compared each scores’ ability to predict the predetermined outcomes, using calculation of area under the receiver operating characteristic curves (AUROCs) and 95% confidence intervals. All comparisons were based on patients in whom all the compared scores could be calculated. The AUROCs were considered dependent.
undertaken using STATA 11.0 (StataCorp, College Station, TX). We calculated rates of transfusion, endoscopic treatment, and mortality for those classified as low risk. Proportions were compared using Pearson’s χ² test and Fischer’s exact test. Medians were compared using the Mann-Whitney U test and the Kruskal-Wallis equality of populations rank test.

We aimed for an adequate sample size to compare all the predetermined outcomes. Existing data indicated that the highest power was needed for comparative analyses on mortality. Therefore assuming AUROCs of 0.86 for mortality, 0.78 for AIMS65 score⁸⁻²⁰ and full Rockall score,¹⁰ ²¹ with an α of 5%, power of 90%, rank correlation of 0.5, and mortality of 5%,¹⁹ we determined that we required a sample size of 2814 patients. We used a two tailed significance level of 5%. Data analysis was undertaken using STATA 11.0 (StataCorp, College Station, TX).

Only the local principal investigator and dedicated nurse, junior doctor, or medical student had access to the identifiable data from their centre, with the anonymised data sent to SBL and AJS for central analysis. Our report follows STROBE guidelines.

Patient involvement
No patients were involved in setting the research question or the outcome measures, nor were they involved in developing plans for recruitment, design, or implementation of the study. No patients were asked to advise on interpretation or writing up of results. There are no plans to disseminate the results of the research to study participants or the relevant patient community.

Results
Patient characteristics and baseline scores
A total of 3012 patients were included in the study, with 2868 (95%) followed up for 30 days. The median age of patients was 65 years, and 58% were men. Table 1 shows the patients’ characteristics, endoscopic findings, interventions, and outcomes. A total of 1348 (45%) patients needed hospital based intervention or died within 30 days. Endoscopic treatment was performed in 574 (19%), 37 (1.2%) required interventional radiology or surgery, and 144 (5%) rebled within seven days. The median length of stay was 3 days (95% confidence interval 0 to 16 days), and 30 day mortality was 7%. Overall, the mean Glasgow Blatchford score was 6.6 (95% confidence interval 0 to 14), mean AIMS65 score was 1.0 (0 to 3), mean admission Rockall score was 2.7 (0 to 5), mean full Rockall score was 3.9 (1 to 7), and mean PNED score was 3.0 (0 to 8). Supplementary table 4 show these data by centre.

Comparison of scores’ ability to predict outcomes

**Intervention or mortality**

The Glasgow Blatchford score had the highest discriminative ability (AUROC 0.86) at predicting need for intervention or death compared with the PNED score (0.69; P < 0.001), full Rockall score (0.70; P < 0.001), admission Rockall score (0.66; P < 0.001), and AIMS65 score (0.68; P < 0.001) (table 2 and fig 1). The Glasgow Blatchford score performed consistently well in all centres, with AUROCs ranging between 0.85 and 0.91 (see supplementary table 5). However, performance of AIMS65 and the full Rockall scoring systems varied by centre, with AUROCs ranging from 0.54-0.75 and 0.57-0.79, respectively. AUROCs among centres were 0.59-0.72 for admission Rockall score and 0.70-0.76 for PNED score (see supplementary table 5).

Table 3 shows the ability of the three pre-endoscopic scores to predict patients at very low risk who did not require intervention and survived at optimal cutoffs. A Glasgow Blatchford score of 1 or less had a sensitivity of 98.6%, specificity of 34.6%, positive predictive value of 96.6%, and negative predictive value of 56.0% for this combined endpoint. Intervention or death was recorded in 3.4% of the 564 patients with Glasgow Blatchford scores of 1 or less, compared with 16% of the

### Table 1 | Characteristics, treatment, and outcome of patients (n=3012). Values are numbers (percentages) unless stated otherwise

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Men</strong></td>
<td>1750 (58)</td>
</tr>
<tr>
<td><strong>Comorbidity</strong></td>
<td></td>
</tr>
<tr>
<td>Ischaemic heart disease</td>
<td>580 (19)</td>
</tr>
<tr>
<td>Liver disease</td>
<td>453 (15)</td>
</tr>
<tr>
<td>Renal failure</td>
<td>266 (9)</td>
</tr>
<tr>
<td>Any malignancy</td>
<td>430 (14)</td>
</tr>
<tr>
<td>Median (95% CI) systolic blood pressure (mm Hg)</td>
<td>125 (90 to 170)</td>
</tr>
<tr>
<td>Median (95% CI) pulse</td>
<td>89 (61 to 126)</td>
</tr>
<tr>
<td>Median (95% CI) haemoglobin level (g/L)</td>
<td>112 (58 to 162)</td>
</tr>
<tr>
<td><strong>Findings at endoscopy</strong></td>
<td></td>
</tr>
<tr>
<td>Nothing abnormal</td>
<td>297 (14)</td>
</tr>
<tr>
<td>Erosive disease*</td>
<td>583 (28)</td>
</tr>
<tr>
<td>Gastric/duodenal ulcer</td>
<td>572 (28)</td>
</tr>
<tr>
<td>Variceal bleeding</td>
<td>143 (7)</td>
</tr>
<tr>
<td>Upper gastrointestinal cancer</td>
<td>70 (3)</td>
</tr>
<tr>
<td><strong>No endoscopy</strong></td>
<td>937 (31)</td>
</tr>
<tr>
<td>Median (95% CI) time to endoscopy (hours)</td>
<td>19 (3 to 72)</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
</tr>
<tr>
<td>Median (95% CI) No of transfusions</td>
<td>1.3 (0 to 6)</td>
</tr>
<tr>
<td>Endoscopic treatment</td>
<td>574 (19)</td>
</tr>
<tr>
<td>Surgery/interventional radiology</td>
<td>37 (1.2)</td>
</tr>
<tr>
<td><strong>Outcome</strong></td>
<td></td>
</tr>
<tr>
<td>No need for intervention, or death</td>
<td>1636 (55)</td>
</tr>
<tr>
<td>Rebleeding</td>
<td>144 (5)</td>
</tr>
<tr>
<td>Mortality</td>
<td>208 (7)</td>
</tr>
<tr>
<td>Mean (95% CI) score</td>
<td>6.6 (0 to 14)</td>
</tr>
<tr>
<td><strong>Glasgow Blatchford</strong></td>
<td>1.0 (0 to 3)</td>
</tr>
<tr>
<td>AIMS65</td>
<td>2.7 (0 to 5)</td>
</tr>
<tr>
<td>Admission Rockall</td>
<td>3.9 (1 to 7)</td>
</tr>
<tr>
<td>Full Rockall</td>
<td>3.0 (0 to 8)</td>
</tr>
<tr>
<td>PNED</td>
<td></td>
</tr>
</tbody>
</table>

*Oesophagitis, gastritis, or duodenitis.*
and overall mortality 0.4%. The proportion of low risk patients defined by Glasgow Blatchford scores of 1 or less varied by site, ranging from 9% in Dunedin to 31% in Glasgow.

Endoscopic treatment
The Glasgow Blatchford score (AUROC 0.75) performed better in predicting need for endoscopic treatment than the other pre-endoscopic scores: AIMS65 (0.62) and admission Rockall (0.61; P=0.001 for both; fig 2). The Glasgow Blatchford scoring system performed equally well in predicting this endpoint in all centres (see supplementary table 5). A threshold Glasgow Blatchford score of 7 or more was best at predicting endoscopic treatment, with a sensitivity of 80.4%, specificity of 57.4%, positive predictive value of 31.3%, and negative predictive value of 92.4% (table 4).

Mortality
The AIMS65 score (AUROC 0.77) and PNED score (0.77) had similar discriminative abilities for predicting 30 day mortality (table 2 and fig 3). The PNED score was better at predicting mortality than the admission Rockall score (0.72; P=0.05), full Rockall score (0.72; P=0.05), and Glasgow Blatchford score (0.64; P<0.001). AIMS65 was better at predicting mortality than the Glasgow Blatchford score (P=0.001) and admission Rockall score (P=0.05). Furthermore, the AIMS65 score had a near statistically significantly higher AUROC compared with the full Rockall score (P=0.06). The best score thresholds at predicting 30 day mortality were 4 or more for PNED, 2 or more for AIMS65, 3 or more for admission Rockall, 5 or more for full Rockall, and 5 or more for Glasgow Blatchford. Table 5 shows the predictive ability of all five scores to identify high risk patients.

Rebleeding and length of stay
PNED, which includes rebleeding in the score, was best (AUROC 0.85) at predicting rebleeding compared with the Glasgow Blatchford score (0.66; P<0.001), full Rockall score (0.64; P<0.001), admission Rockall score (0.62; P<0.001), and AIMS65 (0.60; P<0.001). The Glasgow Blatchford score was better at predicting rebleeding compared with the AIMS65 score (P=0.04). All risk scores had poor discriminative abilities for predicting length of stay for more than three days. The AUROCs were 0.68 for PNED score, 0.65 for full Rockall score, 0.64 for admission Rockall score, 0.63 for Glasgow Blatchford score, and 0.61 for AIMS65 score.

Missing values
Values were missing for the Glasgow Blatchford score (n=80), AIMS65 score (n=511), admission Rockall score (n=43), full Rockall score (n=1000), PNED score (n=178), comorbidity (n=1), systolic blood pressure (n=41), pulse (n=38), haemoglobin concentration (n=28), findings at endoscopy (n=2), true low risk status (n=28), total units of blood transfused (n=23), performance of endoscopic treatment (n=20), performance of surgery or interventional radiology (n=5), rebleeding (n=51), and mortality (n=1). Comparisons of AUROCs were undertaken using
Table 3 | Outcomes at optimal thresholds of Glasgow Blatchford score, AIMS65, and admission Rockall score in prediction of need for intervention, or death. Values are numbers (percentages) unless stated otherwise

<table>
<thead>
<tr>
<th>Scoring system</th>
<th>Cut-off</th>
<th>Low risk patients*</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
<th>Received transfusion</th>
<th>Endoscopic treatment</th>
<th>Surgery or interventional radiology†</th>
<th>Mortality‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glasgow Blatchford</td>
<td>≤1</td>
<td>564 (19.2)</td>
<td>98.6</td>
<td>94.6</td>
<td>56.0</td>
<td>10 (1.8)</td>
<td>8 (1.4)</td>
<td>2 (0.4)</td>
<td>2 (0.4)</td>
<td>6 (0.7)</td>
</tr>
<tr>
<td>AIMS65</td>
<td>0</td>
<td>865 (34.6)</td>
<td>81.5</td>
<td>49.9</td>
<td>74.7</td>
<td>59.9</td>
<td>165 (19)</td>
<td>107 (12)</td>
<td>9 (1.0)</td>
<td>6 (0.7)</td>
</tr>
<tr>
<td>Admission Rockall</td>
<td>0</td>
<td>436 (16.7)</td>
<td>95.6</td>
<td>23.4</td>
<td>86.5</td>
<td>50.9</td>
<td>41 (4.9)</td>
<td>29 (6.7)</td>
<td>0 (0)</td>
<td>1 (0.2)</td>
</tr>
</tbody>
</table>

PPV=positive predictive value; NPV=negative predictive value.

Several patients fulfilled more than one endpoint.

*Classified as low risk according to risk scoring system.
†Number (%) of patients needing surgery, or arterial embolisation among patients classified as low risk.
‡Number (%) of patients dying within 30 days from presentation among patients classified as low risk.

Discussion

This international multicentre study shows that the Glasgow Blatchford score is an accurate risk score at all sites for predicting need for clinical intervention, or death, after upper gastrointestinal bleeding. Although scores for AIMS65 and PNED are the best for predicting mortality and the Glasgow Blatchford score is best for predicting endoscopic treatment, accuracy and clinical utility is relatively low for these endpoints.

Strengths and limitations of this study

Strengths of our study include its international multicentre design and large number of consecutive patients studied. By assessing the five pre-endoscopy and post-endoscopy scores that appeared most promising for clinical use, we could investigate the optimum way to risk assess patients with upper gastrointestinal bleeding, both early after presentation and after endoscopic diagnosis and treatment. We believe our choice of several clinically relevant endpoints in addition to a composite endpoint to compare the scores provides clinicians with useful information. Clinicians may be keen to identify those at high risk for need of endoscopic...

Table 4 | Need for endoscopic treatment at optimal score thresholds of each of the pre-endoscopic scores: Glasgow Blatchford score, AIMS65, and admission Rockall score

<table>
<thead>
<tr>
<th>Scoring system</th>
<th>Cut-off</th>
<th>No (%) of patients classified as high risk</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glasgow Blatchford</td>
<td>≥7</td>
<td>1456 (50)</td>
<td>84.4</td>
<td>57.4</td>
<td>31.3</td>
<td>92.4</td>
</tr>
<tr>
<td>AIMS65</td>
<td>≥3</td>
<td>1619 (65)</td>
<td>74.7</td>
<td>38.7</td>
<td>25.9</td>
<td>87.6</td>
</tr>
<tr>
<td>Admission Rockall</td>
<td>≥3</td>
<td>1686 (57)</td>
<td>69.8</td>
<td>45.9</td>
<td>23.5</td>
<td>86.5</td>
</tr>
</tbody>
</table>

PPV=positive predictive value; NPV=negative predictive value.
RESEARCH

Fig 3 | Comparisons of scores in prediction of 30 day mortality (n=1707). AUROC=area under the receiver operating characteristic curve

A weakness of our study was that we excluded patients who were already inpatients when they developed upper gastrointestinal bleeding. These patients are already known to be at high risk of poor outcome. In addition, 31% of patients did not undergo endoscopy. In three centres, there was a guideline to not admit patients with Glasgow Blatchford scores of either 1 or 2 or less and aged less than 70 years, and many did not attend for outpatient endoscopy. In addition, some clinicians decided endoscopy was not required after an apparent trivial bleed, or was deemed inappropriate for some ill patients with multiple comorbidities. Interestingly, the UK national audit on upper gastrointestinal bleeding reported that 26% of patients notified to gastroenterologists did not have inpatient endoscopy.13 We do not believe this had a major effect on our study findings, as 95% of patients were followed up for 30 days after presentation. Importantly, the usefulness of risk scores that require endoscopy is limited because endoscopy (if undertaken) is often delayed for up to 24 hours or more, whereas clinicians generally want risk stratification early after presentation.

A fairly large number of patients had missing values for full Rockall score (largely as a result of not undergoing endoscopy) or AIMS65 score (largely due to lack of measurement of albumin). Although 30 day mortality was similar in patients with and without missing values for both scoring systems, those with missing values had lower values of the other scores and less often underwent intervention, which could lead to an underestimation of the ability of the AIMS65 score to identify low risk patients. However, in the centre with the lowest missing values for AIMS65 score (Odense at 1.9%), subgroup analysis confirmed our main finding that the Glasgow Blatchford score had a higher discriminative ability to predict need for intervention or death compared with other risk scores, including AIMS65.

All data on patients managed as outpatients were included. In these outpatients, apart from performance of endoscopy they were otherwise assessed, investigated, and managed in the same way as all others and were followed up at 30 days using patient records. None of those managed as outpatients died and none required transfusion, surgery, or interventional radiology; endoscopic treatment was undertaken in one (a 28 year old with a minor Mallory-Weiss tear.). Therefore we do not believe that these policies had a major impact on our conclusions.

The power calculation for mortality was based on AIMS65 and full Rockall scores. Using our baseline assumptions, the number of patients in our study with full data on AIMS65 and full Rockall scores still provided 80% power to predict mortality. For our primary endpoint of intervention or death, using previous AUROC data for AIMS65 score of 0.7 and for the Glasgow Blatchford score of 0.8 (worst case value) and power of 90%, we needed only approximately 500 patients.10

Table 5 | 30 day mortality at optimal score thresholds of Glasgow Blatchford score, AIMS65, PNED, and full Rockall score

<table>
<thead>
<tr>
<th>Scoring system</th>
<th>Cut-off</th>
<th>No (%) of high risk patients*</th>
<th>Sensitivity (%)</th>
<th>Specificity (%)</th>
<th>PPV (%)</th>
<th>NPV (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIMS65</td>
<td>≥2</td>
<td>672 (26.9)</td>
<td>65.8</td>
<td>76.2</td>
<td>18.0</td>
<td>96.6</td>
</tr>
<tr>
<td>PNED</td>
<td>≥4</td>
<td>1065 (37.6)</td>
<td>77.3</td>
<td>65.3</td>
<td>14.1</td>
<td>97.5</td>
</tr>
<tr>
<td>Admission Rockall</td>
<td>≥24</td>
<td>1130 (38.1)</td>
<td>78.6</td>
<td>65.0</td>
<td>14.3</td>
<td>97.6</td>
</tr>
<tr>
<td>Full Rockall</td>
<td>≥25</td>
<td>815 (40.5)</td>
<td>74.0</td>
<td>61.7</td>
<td>11.1</td>
<td>97.3</td>
</tr>
<tr>
<td>Glasgow Blatchford</td>
<td>≥25</td>
<td>1812 (61.8)</td>
<td>84.7</td>
<td>40.2</td>
<td>9.9</td>
<td>97.9</td>
</tr>
</tbody>
</table>

*PPV=positive predictive value; NPV=negative predictive value.
*Classified as high risk according to risk scoring system.
Therefore we believe we achieved sufficient power in this study, which is the largest prospective study on this topic to date.

Comparison with other studies
A recent single centre retrospective study from Australia of 424 patients with upper gastrointestinal bleeding compared all these scores apart from PNED.18 The researchers reported that the AIMS65 score was best at predicting mortality with a similar AUROC to ours, although lower than other smaller studies.16 19 29 They also found that AIMS65, Glasgow Blatchford, and full Rockall scores were similar at predicting their composite endpoint, which unlike ours did not include blood transfusion. However, we believe this is an important endpoint to assess, particularly for identification of low risk patients for possible outpatient management. A variety of low and high risk score thresholds have been suggested for clinical use by these and other small studies.

Published guidelines suggest outpatient management for patients with a Glasgow Blatchford score of 0.20 21 23 Some authors, however, have suggested that this low risk threshold could be extended to 1 or less or to 2 or less.13 14 22 We identified a Glasgow Blatchford score of 1 or less as the optimum score threshold to identify patients who would not require intervention or die, with a sensitivity of 98.6%. The accuracy of this threshold was consistently high across international sites, which comprised 19% of all patients in our study, compared with the 8.6% patients who had a Glasgow Blatchford score of 0. Of the patients with a Glasgow Blatchford score of 1 or less, all cause mortality was only 0.4%. This compares favourably with accepted low risk thresholds of commonly used scores for other medical conditions, including CURB65 for chest infection and the pulmonary embolism severity index.32 33 Other studies have reported more accurate prediction of death, rebleeding, or length of stay using these scores, but our study has the advantages of a multicentre design and large numbers of participants and suggests these outcomes are not well predicted by the studied scores.

The median length of hospital admission for a patient presenting with upper gastrointestinal bleeding in the UK is four or five days,19 20 and a recent study showed the mean in-hospital cost to be £2458 (€2909) per patient.34 Use of an accurate score within emergency departments or acute assessment units, to identify very low risk patients with upper gastrointestinal bleeding not requiring admission has clear benefits. Interestingly, sites varied in the proportion of very low risk patients defined by a Glasgow Blatchford score of 1 or less. The differences are probably due to different models of healthcare provision, including extent of primary care services. At the other end of the spectrum of severity of upper gastrointestinal bleeding, early identification of high risk patients who might benefit from urgent endoscopy or higher level care could help guide management. An observational study suggested that patients with a Glasgow Blatchford score greater than 12 have decreased mortality if endoscopy is undertaken less than 13 hours after presentation.29 Our data suggest that a Glasgow Blatchford score of 7 or more and AIMS65 score of 2 or more have the highest combination of sensitivity and specificity for predicting endoscopic treatment or 30 day mortality, respectively, but the positive predictive value for both is low. Therefore the clinical utility of these scores to direct care in high risk patients seems limited.

Conclusions and policy implications
The Glasgow Blatchford score is accurate at predicting need for clinical intervention, or death in patients with upper gastrointestinal bleeding in all countries studied. A score of 1 or less is the optimum threshold for identifying very low risk patients suitable for outpatient management. A Glasgow Blatchford score of 7 or more is best at predicting need for endoscopic treatment, and a PNED score of 4 or more and AIMS65 score of 2 or more are best at predicting mortality, although accuracy in predicting these endpoints is relatively low. No score seems accurate at predicting rebleeding or length of hospital stay. This information can help direct management of very low risk patients with upper gastrointestinal bleeding, but further studies using these, or new scores, are required to clarify their role in directing management of higher risk patients.

International Gastrointestinal Bleeding Consortium: P Dunne and J Fleming (Glasgow Royal Infirmary, Glasgow, UK), A Holloway, O Proud, J Sutton, N Michell, and E Pinchin (Royal Cornwall Hospital, Cornwall, UK), H S Ong, Y K Chin, and D S Koay (Singapore General Hospital, Singapore), T Steiner (Southern District Health Board, Dunedin Hospital, Dunedin, New Zealand).

Contributors: RA, LZ, ST, KW, CILK, SBL, PN, JF, AH, AT, JS, NM, EP, HSO, YKC, DSK and TS collected data. AJS and SBL wrote the paper with considerable input from LL, HRD, JHN, MS and IAM. All coauthors approved the final manuscript. AJS is the study guarantor.

Funding: This study was sponsored by NHS Greater Glasgow and Clyde Health Board, Scotland, UK. Data collection in Glasgow, Singapore, and Dunedin was funded by a health board endowment, hospital research, and guthealthnetwork grant, respectively. The funders had no input to the study design, data collection, or interpretation, writing of report, or submission for publication.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author) and declare that: (1) no authors have support from any company for the submitted work; (2) no authors have relationships with any company that might have an interest in the submitted work; (3) no author, their spouses, partners, or children have financial relationships that may be relevant to the submitted work and (4) no authors have any non-financial interests that may be relevant to the submitted work.

Ethical approval: This study was approved by the west of Scotland ethics committee (reference 14/WS/0012; project No 145837) and each centre obtained approval from their local research committee or review board. Owing to the observational, non-interventional nature of the study, the ethical committee (and local committees and boards) agreed that individual patient consent was not required.

Data sharing: Patient level data and/or full dataset available from the corresponding author AJS. Consent was not obtained but the presented data are anonymised and risk of identification is low.

Transparency: The study guarantor (AJS) affirms that the manuscript is an honest, accurate and transparent account of the study being submitted. The funders had no input to the study design, data collection, or interpretation, writing of report, or submission for publication. All authors agree that individual patient consent was not required.


Supplementary information: Supplementary tables 1 to 5.