

## A multicentre development and validation study of a novel Lower Gastrointestinal Bleeding score-The Birmingham Score

Smith, Samuel; Bazarova, Alina; Ejenavi, Efe; Qurashi, Maria; Shivaji, Uday; Harvey, Phil R; Slaney, Emma; McFarlane, Michael; Baker, Graham; Elnagar, Mohamed; Yuzari, Sarah; Gkoutos, George V.; Ghosh, Subrata; Iacucci, Marietta

DOI:

[10.1007/s00384-019-03459-z](https://doi.org/10.1007/s00384-019-03459-z)

License:

Creative Commons: Attribution (CC BY)

*Document Version*

Publisher's PDF, also known as Version of record

*Citation for published version (Harvard):*

Smith, S, Bazarova, A, Ejenavi, E, Qurashi, M, Shivaji, U, Harvey, PR, Slaney, E, McFarlane, M, Baker, G, Elnagar, M, Yuzari, S, Gkoutos, GV, Ghosh, S & Iacucci, M 2020, 'A multicentre development and validation study of a novel Lower Gastrointestinal Bleeding score-The Birmingham Score', *International journal of colorectal disease*, vol. 35, no. 2, pp. 285–293. <https://doi.org/10.1007/s00384-019-03459-z>

[Link to publication on Research at Birmingham portal](#)

### General rights

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### Take down policy

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.



# A multicentre development and validation study of a novel lower gastrointestinal bleeding score—The Birmingham Score

Samuel C. L. Smith<sup>1</sup> · Alina Bazarova<sup>1</sup> · Efe Ejenavi<sup>2</sup> · Maria Qurashi<sup>2</sup> · Uday N. Shivaji<sup>1,3</sup> · Phil R. Harvey<sup>4</sup> · Emma Slaney<sup>4</sup> · Michael McFarlane<sup>5</sup> · Graham Baker<sup>6</sup> · Mohamed Elnagar<sup>6</sup> · Sarah Yuzari<sup>6</sup> · Georgios Gkoutos<sup>1</sup> · Subrata Ghosh<sup>1,2,3</sup> · Marietta Iacucci<sup>1,2,3,7</sup>

Accepted: 12 November 2019 / Published online: 16 December 2019

© The Author(s) 2019

## Abstract

**Purpose** Lower gastrointestinal bleeding (LGIB) is common and risk stratification scores can guide clinical decision-making. There is no robust risk stratification tool specific for LGIB, with existing tools not routinely adopted. We aimed to develop and validate a risk stratification tool for LGIB.

**Methods** Retrospective review of LGIB admissions to three centres between 2010 and 2018 formed the derivation cohort. Using regression analysis within a machine learning technique, risk factors for adverse outcomes were identified, forming a simple risk stratification score—The Birmingham Score. Retrospective review of an additional centre, not included in the derivation cohort, was performed to validate the score.

**Results** Data from 469 patients were included in the derivation cohort and 180 in the validation cohort. Admission haemoglobin OR 1.07(95% CI 1.06–1.08) and male gender OR 2.29(95% CI 1.40–3.77) predicted adverse outcomes in the derivation cohort AUC 0.86(95% CI 0.82–0.90) which outperformed the Blatchford 0.81(95% CI 0.77–0.85), Rockall 0.60(95% CI 0.55–0.65) and AIM65 0.55(0.50–0.60) scores and in the validation cohort AUC 0.80(95% CI 0.73–0.87) which outperformed the Blatchford 0.77(95% CI 0.70–0.85), Rockall 0.67(95% CI 0.59–0.75) and AIM 65 scores 0.61(95% CI 0.53–0.69). The Birmingham Score also performs well at predicting adverse outcomes from diverticular bleeding AUC 0.87 (95% CI 0.75–0.98). A score of 7 predicts a 94% probability of adverse outcome.

**Conclusion** The Birmingham Score represents a simple risk stratification score that can be used promptly on patients admitted with LGIB.

**Keywords** Lower gastrointestinal bleeding · Risk stratification · Risk factors · Triaging tool

## Introduction

Lower gastrointestinal bleeding (LGIB) has an incidence of approximately 33 per 100,000 population and a frequency approaching that of upper gastrointestinal causes [1]. Common causes include diverticular disease, haemorrhoids, colitis and colorectal cancer/polyps [2–4]. LGIB can range from benign to potentially life threatening, with the majority of cases resolving spontaneously [2].

Evaluation with a risk stratification score may be useful in aiding clinical decision-making. There are few validated risk stratification scoring systems designed specifically for LGIB [5], unlike upper gastrointestinal bleeding, for which there are well-validated and robust predictive scoring models [6–8]. The current scoring systems specific for LGIB [9–11] have limitations including requirement for inpatient colonoscopy and prolonged patient observation, meaning patients require

✉ Marietta Iacucci  
m.iacucci@bham.ac.uk

<sup>1</sup> Institute of Translational Medicine and Institute of Immunology and Immunotherapy, University of Birmingham, Birmingham, UK

<sup>2</sup> University Hospitals Birmingham NHS Foundation Trust, Queen Elizabeth Hospital, Birmingham, UK

<sup>3</sup> National Institute for Health Research (NIHR) Birmingham Biomedical Research Centre, Birmingham, UK

<sup>4</sup> Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, UK

<sup>5</sup> University Hospital Coventry and Warwickshire NHS Trust, Birmingham, UK

<sup>6</sup> The Royal Wolverhampton NHS Trust, Birmingham, UK

<sup>7</sup> University of Calgary, Calgary, Canada

admission and the scores cannot be used as an emergency triage tool [9, 11–13]. Therefore, they are not widely used in clinical practice [14]. More recently, Oakland and colleagues developed and validated a LGIB score predicting suitability for safe early discharge [15]. The criteria for safe discharge were the absence of certain interventions such as blood transfusion, which may not truly reflect severity of cases [15]. It also requires the knowledge of previous admissions with LGIB, which is information that may not consistently be available to admitting clinicians.

Developing a robust and reproducible scoring system would allow clinicians to potentially facilitate early discharge for low risk patients and identify high-risk patients, who may require more intensive management such as close monitoring or critical care. The aim of this study was to identify risk factors for adverse outcomes from LGIB and subsequently develop and validate a risk stratification tool that would allow early decision making at the point of admission without the need for prolonged observation and before adverse outcomes such as blood transfusion or haemostatic intervention have taken place.

## Patients and methods

### Setting and participants

We conducted a multicentre retrospective review of all admissions with LGIB (age  $\geq 16$  years) presenting to University Hospital Birmingham NHS Trust (UHB), University Hospital Coventry and Warwickshire (UHCW), Sandwell and West Birmingham NHS Trust (SWBH) and the Royal Wolverhampton NHS Trust (RWH) between 2010 and 2018. We used ICD coding to identify acute admissions with haematochezia and rectal bleeding. Retrospective electronic medical records were reviewed to confirm LGIB at the point of admission; therefore, it was not possible to gather data on the number of patients undergoing upper GI endoscopy for exclusion of upper GI bleeding. Patients attending for outpatient appointments for LGIB and patients attending the emergency department and subsequently discharged without admission were excluded. Case notes were reviewed to collect demographic data, co-morbidities (grouped as per Glasgow-Blatchford score [7]), admission laboratory testing, admission vital signs, presence of syncope and altered mental status. Altered mental status was recorded if confusion, disorientation or reduced conscious level was noted on admission hospital records, which is recorded as a standard of practice. Adverse outcomes collected included blood transfusion, endoscopic intervention (APC, injection therapy, clips), CT angiography (positive or negative), surgical intervention, re-bleeding (re-admission within 30 days) and mortality. Sub-group analysis was performed on those patients with a proven endoscopic diagnosis of LGIB secondary to diverticular disease.

## Outcomes and endpoints

The primary outcome of the study was adverse outcomes. The adverse outcomes of blood transfusion, endoscopic intervention, CT angiography, surgical intervention, re-bleeding and mortality were pooled to allow statistical analysis given the low incidence of mortality from LGIB (3.4%) [3]. These outcomes were selected as they reflect a requirement for an inpatient admission.

### Statistical analysis

#### Risk score derivation

For the derivation cohort, data from UHB, UHCW and SWBH were analysed, using regression analysis within a machine learning technique (namely R-package CARROT-Bazarova and Raseta <https://CRAN.R-project.org/package=CARROT>, 2018), which has been previously used for prediction in clinical set-ups [16] (25, 26) and combines principles of good practice from machine learning such as cross-validation [17] and subset regression [18] restricted by the rule of ten events per variable ('one in ten rule') [19]. The latter means that only the regression models where the number of variables multiplied by ten does not exceed the number of patients in the less likely category of the training set (adverse outcomes). Therefore, regression models consisting of all possible combinations of factors satisfying this condition were used to predict adverse outcomes. The variables considered were gender, age, co-morbidities, admission haemoglobin, albumin, INR, urea, blood pressure, heart rate, melena, syncope and altered mental state. Within the package CARROT, the number of cross-validations [17] was set to 1000. For each cross-validation, the dataset was randomly partitioned into the training set (90% of the data) and the test set (10% of the data), and the corresponding regression models were fit to the training set and their predictive power was assessed on the test set. Given 159 patients experienced adverse outcomes, according to the 'rule of 10' on average of up to 14 variables could be included in the model during each cross-validation (around 16,000 models). The best models were identified via maximizing average AUROCs computed over all cross-validations on the test sets. Then, the best model fitted to the whole dataset and the corresponding ORs were computed. The factors with 95% CI for ORs containing value 1 were excluded from the set of variables. Then, machine-learning algorithm was applied to the modified set of variables. This procedure was repeated until the final model consisted of factors with ORs not containing 1 when fitted to the whole dataset. The score was developed by fitting the corresponding regression to the whole dataset; the coefficients of the regression were rounded. Then, the curve was produced and its most efficient breakpoints determined by Youden's J-statistic [20] were used to create

an integer valued discrete score. The AUROC corresponding to the newly developed score were calculated using a bootstrapping method to assess the predictive power of the score available through an R-package pROC [21].

### Risk score validation

Validation of The Birmingham Score took place by using the risk algorithm on the dataset from RWH. Estimates of the area under the receiver operating curve (AUROC) were calculated by applying the score directly to the whole dataset and its confidence intervals were obtained by using a bootstrapping method repeated 2000 times for this dataset against The Birmingham Score using R-package pROC.

For comparisons of the derivation and validation groups, we used independent *t*-test statistics in order to compare means of the continuous variables (Table 1). For categorical outcomes, we used chi-squared and Fisher's exact test. Analysis was performed using statistical software R (R Core Team 2017). A *p* value of < 0.05 was considered to be statistically significant.

### Ethics

Formal local audit approval was obtained at each participating centre (CARMS-13587).

## Results

### Derivation cohort

#### Patient characteristics

A total of 473 patients were identified in the derivation cohort, 4 were removed due to a lack of clinical data (Table 1).

#### Risk score derivation

Logistic regression modelling demonstrated the performance of each variable in predicting adverse outcomes (Table 2). The output of R-package CARRoT was a model consisting of four variables, gender OR 2.31 (95% CI 1.40–3.82), admission Haemoglobin OR 1.07 (95% CI 1.05–1.08), admission urea OR 1.03 (95% CI 0.97–1.09) and syncope OR 0.56 (95% CI 0.23–1.39) demonstrating an average AUROC over all cross-validations of 0.86. Having observed that 95% CI for syncope contains the value 1 and suggests that absence of syncope is associated with adverse outcomes, we excluded syncope from the set of variables and re-ran the analysis on the modified set of variables. The best-selected model consisted of three variables: gender, admission haemoglobin and admission urea. We excluded urea from the set of variables given that its 95% CI for OR contained 1 and repeated the procedure again. The final output was admission haemoglobin OR 1.07 (95%

**Table 1** Patient characteristics of the derivation and validation groups

	Derivation cohort ( <i>n</i> = 469)	Validation cohort ( <i>n</i> = 180)	<i>P</i> value	Diverticular cohort ( <i>n</i> = 57)	<i>P</i> value
Median age (range)	71 (16–98)	71 (19–100)	0.876	77 (29–97)	0.0003
Age ≥ 80 years	147 (31.3%)	59 (32.8%)	0.797	21 (36.8%)	0.49
Co-morbidities (cardiac failure, ischaemic heart disease, major co-morbidity, renal/liver failure, metastatic cancer)	230 (49.3%)	92 (51.1%)	0.701	21 (36.8%)	0.110
Median admission Haemoglobin (range)	120 g/L (29–181)	121 g/L (54–171)	0.157	114 g/L (52–162)	0.472
Haemoglobin ≤ 100 g/L	139 (29.6%)	47 (26.1%)	0.428	20 (35.1%)	0.323
Median admission systolic BP (range)	132 mmHg (85–218)	130 mmHg (66–218)	0.887	135 mmHg (95–210)	0.323
Admission systolic BP ≤ 90 mmHg	6 (1.3%)	1 (0.56%)	0.680	0 (0.0%)	
Mean admission heart rate (range)	82 bpm (48–131)	86 bpm (30–240)	0.005	81 bpm (60–114)	0.853
Admission heart rate ≥ 100 bpm	51 (11%)	41 (22.8%)	≤ 0.001	4 (7.0%)	0.503
Adverse outcomes					
Blood transfusion	118 (25.2%)	45 (25.0%)	1.0	18 (31.6%)	0.376
Endoscopic intervention	24 (5.1%)	4 (2.3%)	0.162	1 (1.8%)	
CT angiography	38 (8.1%)	0 (0.0%)		2 (3.5%)	
Surgery	11 (2.3%)	6 (3.3%)	0.583	0 (0.0%)	
Re-bleeding	29 (6.2%)	17 (9.4%)	0.201	3 (5.2%)	
Mortality in-hospital	10 (2.1%)	8 (4.4%)	0.115	0 (0.0%)	
Mortality < 30 days	15 (3.2%)	6 (3.3%)	1.0	0 (0.0%)	
Mean length of hospital stay (range)	5.8 days (1–107)	7.1 (1–87)	0.179	6.5 (0–35)	0.486

**Table 2** Univariate analysis of variables collected

Variable	OR	OR 95% CI	P value
Gender	1.63	(1.11, 2.4)	0.01
Age	1.01	(1.0, 1.02)	0.02
Co-morbidities (cardiac failure, ischaemic heart disease, major co-morbidity, renal/liver failure, metastatic cancer)	1.28	(1.09, 1.52)	0.003
Haemoglobin	1.07	(1.05–1.07)	$< 2 \times 10^{-16}$
Urea	1.12	(1.07, 1.18)	$3.3 \times 10^{-6}$
Albumin	3.78	(1.83, 8.13)	0.00041
INR	1.97	(1.15, 3.34)	0.01
BP	0.98	(0.97, 0.99)	$5.29 \times 10^{-5}$
HR	1.01	(0.99, 1.02)	0.22
Syncope	0.99	(0.51, 1.77)	0.97
Altered mental state	4.71	(1.29, 22.09)	0.03

CI 0.06–1.08) and gender OR 2.29 (95% CI 1.4–3.77) demonstrating an average AUROC over all cross-validations of 0.86—The Birmingham Score. The corresponding AUROC was 0.86 (95% CI 0.82–0.90) for this score.

### Comparison with other gastrointestinal bleeding scores

The Birmingham Score outperforms the GBS 0.81 (95% CI 0.77–0.85), modified Oakland score 0.84 (95% CI 0.80–0.88), Rockall 0.60 (95% CI 0.55–0.65) and AIM65 0.55 (0.50–0.60) (Fig. 1). Note that for Oakland score, we did not take into account previous history of re-bleeding and DRE findings, as the information was not collected in this study. The cut-offs for The Birmingham Score are shown in Table 3. Using The Birmingham Score, the probabilities of a patient experiencing an adverse outcome were calculated (Table 4).

### Validation Cohort

#### Patient characteristics

In the validation cohort, a total of 203 patients were identified; however, 23 were removed due to a lack of clinical data (Table 1). There was a statistically significant difference between the derivation vs. validation groups in the mean admission heart rate (82 vs. 87  $p = 0.032$ ) and number of patients admitted with a heart rate  $\geq 100$  (11% vs. 22.7%  $p = 0.001$ ).

#### Risk score validation and comparison with other gastrointestinal bleeding scores

Using the validation dataset, the risk stratification scores were applied. The Birmingham Score AUROC 0.80 (95% CI 0.73–0.87) outperformed the GBS AUROC 0.77 (95% CI 0.70–

0.85), modified Oakland Score 0.77 (95% CI 0.70–0.85), Rockall Score AUROC 0.67 (95% CI 0.59–0.75) and AIM65 score AUROC 0.61 (95% CI 0.53–0.69) (Fig. 2).

The distribution of scores and proportion of patients having an adverse outcome in the combined cohort (derivation and validation) are shown in Fig. 3. Overall, 145 patients (22.3%) were admitted with a Birmingham Score of  $< 2$  with 10 (6.9%) noted to have an adverse outcome. A total of 201 (31.0%) patients scored  $\geq 5$  and 148 of those had an adverse outcome (73.6%). An example of how The Birmingham Score might be used in clinical practice is shown in Fig. 4.

### Prediction of adverse outcomes associated with diverticular bleeding

A total of 57 patients (8.7%) had an endoscopic diagnosis of LGIB secondary to diverticular disease. There was a significant difference between this cohort and the derivation cohort in terms of age, with the diverticular group have a higher median age of 77 (range 29–97) vs. 71 (range 16–98)  $p = 0.0003$ . The Birmingham Score predicted adverse outcomes in this cohort of patients with an AUROC 0.87 (95% CI 0.75–0.98), Blatchford Score 0.83 (0.71–0.95), modified Oakland score 0.81 (0.67–0.96), AIM65 0.55 (0.50–0.68) and Rockall score 0.50 (0.50–0.67).

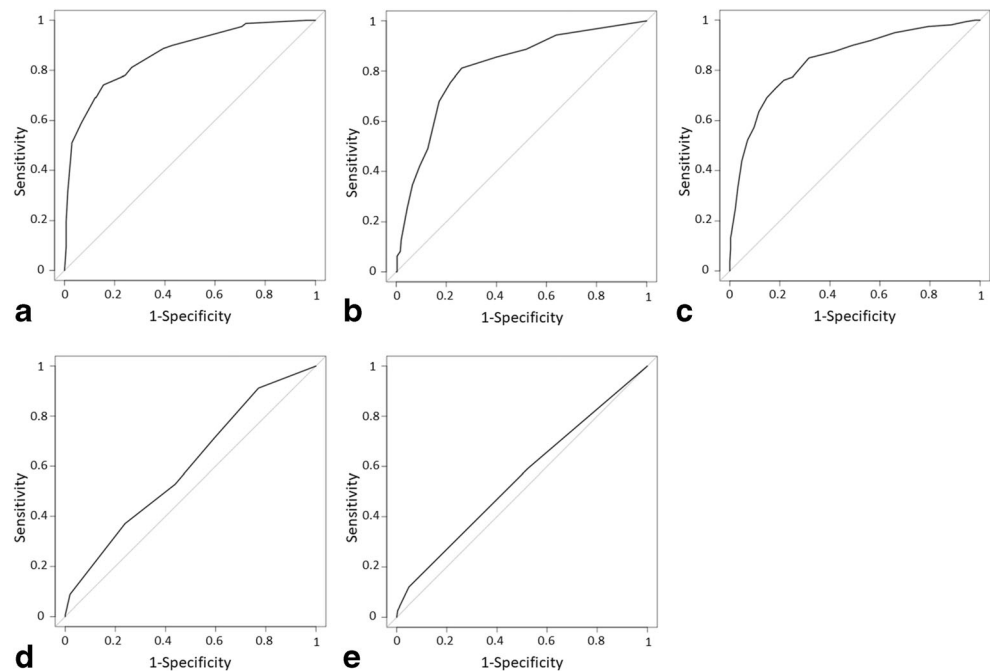
### Discussion

We have developed and validated a triaging tool that can guide decision-making on hospitalization, by using a multicentre database, which outperforms validated upper GI bleeding scores including Glasgow Blatchford Score, Rockall and AIM65. The risk factors included are admission haemoglobin and male gender. A Birmingham score of 6–7 points equates to a probability of an adverse outcome of 90% and a score of  $< 2$  gives a probability of an adverse outcome of 4%.

The strength of The Birmingham Score is that it was developed from a large database from multiple acute care hospitals. These centres include academic centres (UHB and UHCW) as well as community centres (SWBH and RWH) making the results more generalizable. As individual case notes were reviewed, accurate data on adverse outcomes was collated, which may not have been possible from administrative databases or population statistics that had been used in other LGIB studies. The score can be used at the point of admission with one single blood test allowing prompt decisions and is more simple than other scores including the Oakland score but predicts adverse outcome in a similar manner.

Risk stratification tools are integral part of the management of upper gastrointestinal bleeding [6–8], which allows prompt recognition of high-risk patients requiring intensive treatment

**Fig. 1** ROC curves for the derivation dataset. (a) Birmingham Score AUROC 0.86 (95% CI 0.82–0.90), (b) GBS AUROC 0.81 (95% CI 0.77–0.85), (c) modified Oakland score Rockall score AUROC 0.84 (95% CI 0.80–0.88), (d) Rockall score AUROC 0.60 (95% CI 0.55–0.65), and (e) AIM65 AUROC 0.55 (95% CI 0.50–0.60)



and low risk patients facilitating early discharge. In LGIB, existing stratification tools are not widely used clinically [14]. One potential difficulty in developing a LGIB risk stratification score is the heterogeneous nature of LGIB. There is a wide range of aetiologies including diverticular disease, colitis, malignancy and haemorrhoids [3] and as a result acuity and prognosis will vary. Tapaskar et al. ( $n = 170$  patients) in a single-centre retrospective study demonstrated no single risk stratification score has the best predictive ability to predict adverse outcomes [22]. Limitations of previous studies include inclusion criteria of patients having a colonoscopy on the index admission [13], require prolonged observation of patients reducing its use as a triaging tool [9, 11, 12] and poor performance at differentiating low and high-risk patients [10, 23].

More recently, attempts have been made to identify patients who are at low-risk and not requiring in-hospital intervention. Hreinsson et al. ( $n = 581$  patients) developed the SHA<sub>2</sub>PE

score which predicts low-risk patients not requiring in-hospital intervention with an AUC of 0.83 and NPV of 96% [24]. This study was limited by its retrospective, single-centre design, including patients discharged from ER and only including patients who had an endoscopy. Oakland et al. (external validation  $n = 288$ ) also identified risk factors associated with a low likelihood of adverse outcome, with a score of  $< 8$  predicting a 95% probability of safe discharge [15]. Similar to the present study, the Oakland score excluded patients already admitted to hospital, and included blood transfusion, therapeutic intervention to control bleeding, in-hospital mortality, surgical intervention and re-admission within 28 days as adverse outcomes, the absence of which reflected safe discharge after presentation [15]. The most common adverse outcome in the Oakland study was blood transfusion (25%) which is similar to the present study. This score is heavily weighted by admission haemoglobin (up to 22 points from a total of 35), similar to The Birmingham Score.

The Birmingham Score represents a simple risk stratification score, which is objective and does not require observation of patients nor endoscopic information. Therefore, it can be useful clinically for timely decision-making at the time of presentation, although as with all

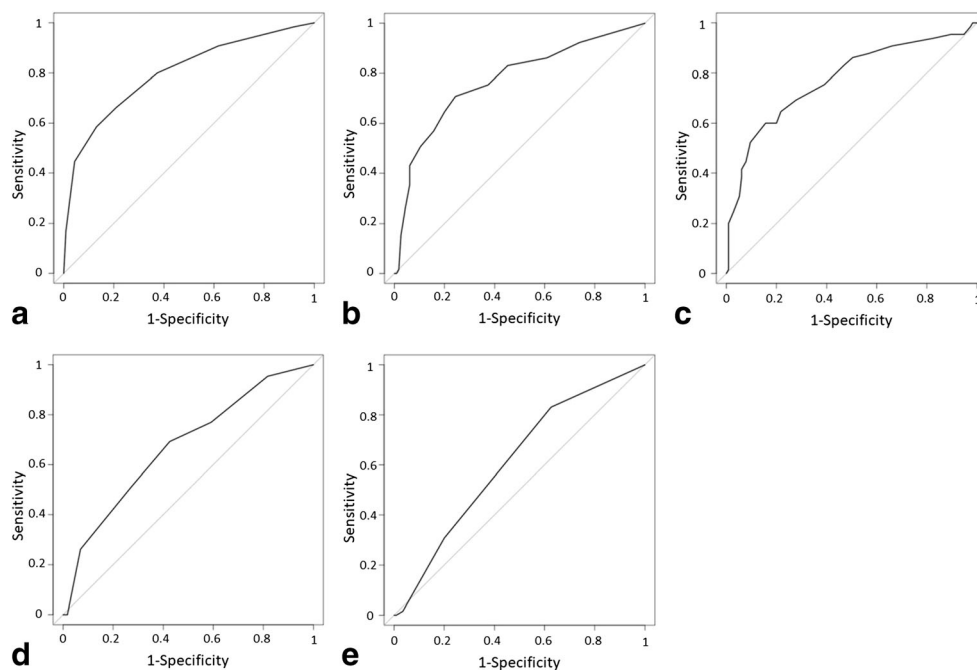
**Table 3** The Birmingham Score

Variable	Points
Male gender	1
Admission haemoglobin	
< 83	6 points
83–< 97	5 points
97–< 112	4 points
112–< 116	3 points
116–< 133	2 points
133–< 147	1 points
> 147	0 points

**Table 4** Probabilities of experiencing adverse outcomes using The Birmingham Score

Birmingham score (points)	Probability % (95% CI)
6–7	90 (82–95)
4–5	38 (30–47)
2–3	18 (13–24)
< 2	4 (1–11)

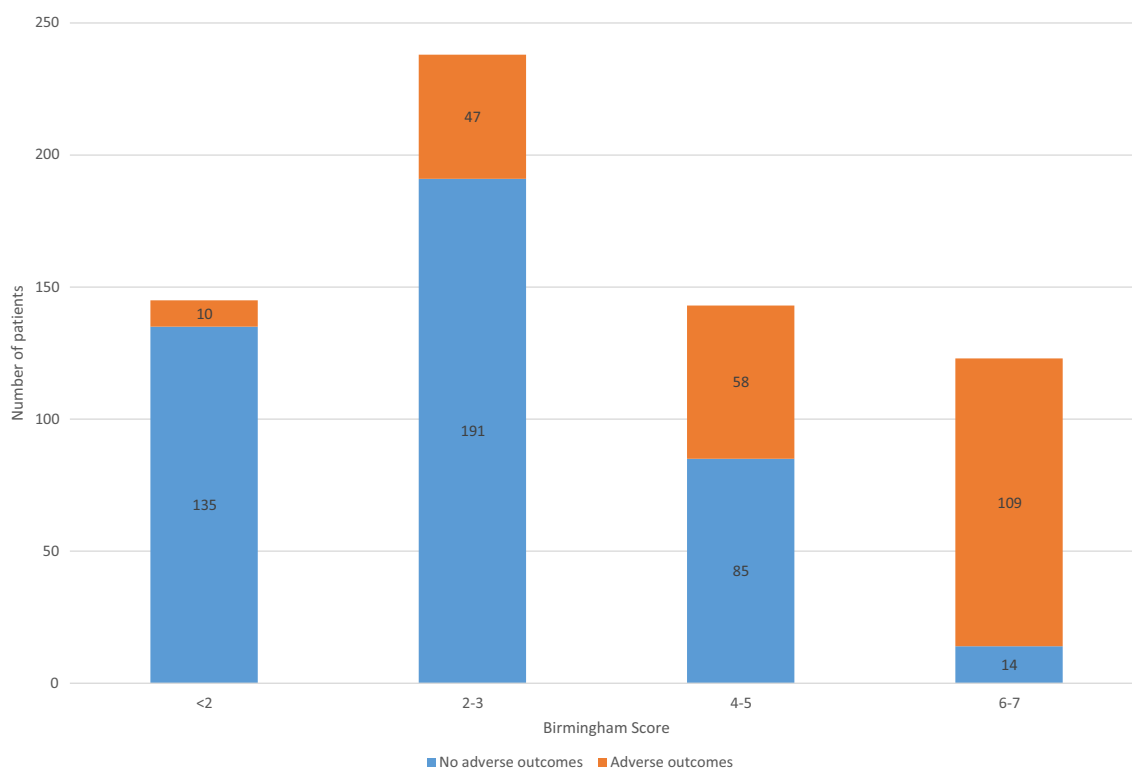
**Fig. 2** ROC curves for the validation dataset. (a) Birmingham score AUROC 0.80 (95% CI 0.73–0.87), (b) GBS AUROC 0.77 (95% CI 0.70–0.85), (c) modified Oakland score AUROC 0.77 (95% CI 0.70–0.85), (d) Rockall score AUROC 0.67 (95% CI 0.59–0.75), and (e) AIM65 AUROC 0.61 (95% CI 0.53–0.69)



risk predictive scores, it has to be used in clinical context and clearly if there are significant co-morbidities or frailty for example then a clinical decision has to be made at the individual patient level. Surprisingly, factors such as age and co-morbidities did not predict adverse outcomes from LGIB; however, this is consistent with the Oakland Score [15] which does not feature co-morbidities in its risk

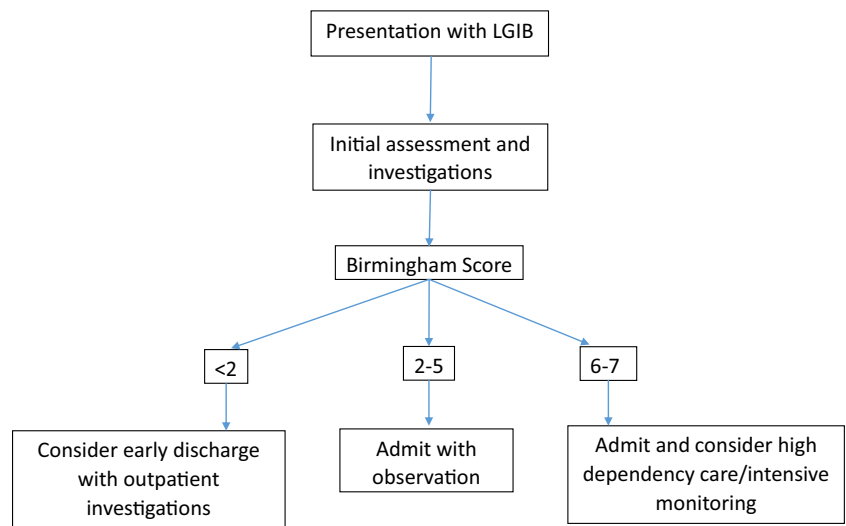
model and age has a maximum weighting of 2 points (out of 35).

The adverse outcomes chosen were those that reflect a requirement for inpatient management and were pooled together in order to power a meaningful risk stratification score and since the most common adverse outcome was blood transfusion, The Birmingham Score is therefore heavily weighted



**Fig. 3** Distribution of Birmingham scores and proportion of patients with adverse outcomes

**Fig. 4** Flowchart of The Birmingham Score in use in the Emergency Department



by admission haemoglobin. Tapaskar et al. also demonstrated the importance of admission haemoglobin in predicting severe bleeding (OR 1.28 (1.10–1.49)) [22]. This is similar to the Oakland Score [15]; however, The Birmingham Score is simpler with less variables required and does not require knowledge of previous LGIB admissions, which may not be immediately available to admitting clinicians. Male gender was found to increase risk of an adverse outcome, which is consistent with the Oakland score [9, 15]. When comparing validated upper GI bleeding scores, we found that the Glasgow Blatchford Score outperforms the Rockall and AIM65 scores, which is also consistent with another study [15]. We made comparisons with upper GI bleeding scores as within the UK, these scores are in common use and therefore the data available reflects this. The Birmingham Score performed well in predicting adverse outcomes from diverticular disease (AUROC 0.87 (95% CI 0.75–0.98)), which given this is most common cause of LGIB means The Birmingham Score is likely to be more widely applicable in clinical practice [3]. Using The Birmingham Score, 22.3% of patients had a low probability of an adverse outcome (Birmingham Score < 2) and therefore could be considered for early discharge, potentially allowing significant healthcare savings.

One of the strengths of the present study is the size of the derivation cohort used ( $n = 469$ ), which is more than other LGIB studies [9–11, 13, 22]. The machine learning technique is a strength to the present study as it allows internal validation at the stage of training which is achieved by multiple cross-validations within the derivation cohort. This way, the best predictive variables are selected for the final score which in turn leads to the highest predictive power measured via AUROC. Variable selection based on internal cross-validations also yields a score which is more likely to stay robust upon external validation, which we demonstrate on the validation cohort. Internal and external validations are

important components in predictive clinical research as they guarantee the best predictive power, robustness of the selected model and allow to avoid overfitting [25]. The Birmingham Score performs favourably to the Oakland score, however is less complex with fewer components.

There are limitations to this study. The retrospective nature of the study mean the conclusions cannot be generalized until prospective validation takes place. The patients included were those whom were admitted with LGIB excluding patients attending to emergency departments and subsequently discharged. However, this study suggests that low-risk patients are admitted without clear criteria to use in triage. There is bias towards more significant LGIB necessitating admission, given patients presenting to emergency departments and subsequently discharged were not included. Together with this, the inclusion of severe presentations of patients unfit for inpatient colonoscopy means there is an element of selection bias. The centres included covers a wide geographical area; however, it was not possible to establish whether individual patients re-presented at other hospitals in the region, which means some episodes of re-admission may not have been captured. The patients included are those whose primary reason for admission was LGIB; therefore, The Birmingham Score cannot be applied to patients having a LGIB during an inpatient stay for other reasons or those patients presenting with multiple presenting complaints in addition to LGIB. The majority of adverse outcomes included in this study could be described as soft outcomes (blood transfusion, endoscopic intervention, CT angiography) as practice may vary between hospitals. Given hard outcomes such as mortality is low in LGIB, a large multicentre study would be needed to give sufficient sample size to power the study. This is a similar limitation in other LGIB studies [15].

In summary, The Birmingham Score represents a simple risk stratification tool that is easily calculated at the point of

admission and does not require endoscopic data nor patient observation for a prolonged period. The Birmingham Score outperforms the Glasgow Blatchford Score, Rockall, modified Oakland and AIM65 in predicting adverse outcomes from LGIB. This can potentially guide early clinical decision making including identifying patients that may require more intensive treatment. Further prospective validation is required before The Birmingham Score can be utilized in routine clinical practice.

**Author contribution** Study conception and design: MI, SG  
Data acquisition: SS, EE, MQ, PH, ES, MM, GB, ME, SY  
Analysis and interpretation of data: SS, AB, GG, SG, MI  
Drafting of the article: SS, AB, US, MI  
Critical revision of manuscript: SG, MI  
Final approval of the article: MI

**Funding information** MI and SG are funded by the NIHR Birmingham Biomedical Research Centre at the University Hospitals Birmingham NHS Foundation Trust and the University of Birmingham.

## Compliance with ethical standards

**Ethics** Formal local audit approval was obtained at each participating centre (CARMS-13587).

**Disclaimer** The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

**Conflict of interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Lanas A, Garcia-Rodriguez LA, Polo-Tomas M, Ponce M, Alonso-Abreu I, Perez-Aisa MA et al (2009) Time trends and impact of upper and lower gastrointestinal bleeding and perforation in clinical practice. *Am J Gastroenterol* 104(7):1633–1641
- Strate LL, Ayanian JZ, Kotler G, Syngal S (2008) Risk factors for mortality in lower intestinal bleeding. *Clin Gastroenterol Hepatol* 6(9):1004–1010 quiz 955
- Oakland K, Guy R, Uberoi R, Hogg R, Mortensen N, Murphy MF, Jairath V, UK Lower GI Bleeding Collaborative (2018) Acute lower GI bleeding in the UK: patient characteristics, interventions and outcomes in the first nationwide audit. *Gut* 67(4):654–662
- Ghassemi KA, Jensen DM (2013) Lower GI bleeding: epidemiology and management. *Curr Gastroenterol Rep* 15(7):333
- Strate LL, Gralnek IM (2016) ACG clinical guideline: management of patients with acute lower gastrointestinal bleeding. *Am J Gastroenterol* 111(4):459–474
- Rockall TA, Logan RF, Devlin HB, Northfield TC (1996) Risk assessment after acute upper gastrointestinal haemorrhage. *Gut* 38(3):316–321
- Blatchford O, Murray WR, Blatchford M (2000) A risk score to predict need for treatment for upper-gastrointestinal haemorrhage. *Lancet (London, England)* 356(9238):1318–1321
- Saltzman JR, Tabak YP, Hyett BH, Sun X, Travis AC, Johannes RS (2011) A simple risk score accurately predicts in-hospital mortality, length of stay, and cost in acute upper GI bleeding. *Gastrointest Endosc* 74(6):1215–1224
- Strate LL, Orav EJ, Syngal S (2003) Early predictors of severity in acute lower intestinal tract bleeding. *Arch Intern Med* 163(7):838–843
- Kollef MH, O'Brien JD, Zuckerman GR, Shannon W (1997) BLEED: a classification tool to predict outcomes in patients with acute upper and lower gastrointestinal hemorrhage. *Crit Care Med* 25(7):1125–1132
- Velayos FS, Williamson A, Sousa KH, Lung E, Bostrom A, Weber EJ, Ostroff JW, Terdiman JP (2004) Early predictors of severe lower gastrointestinal bleeding and adverse outcomes: a prospective study. *Clin Gastroenterol Hepatol* 2(6):485–490
- Newman J, Fitzgerald JE, Gupta S, von Roon AC, Sigurdsson HH, Allen-Mersh TG (2012) Outcome predictors in acute surgical admissions for lower gastrointestinal bleeding. *Color Dis* 14(8):1020–1026
- Aoki T, Nagata N, Shimbo T, Niikura R, Sakurai T, Moriyasu S, Okubo H, Sekine K, Watanabe K, Yokoi C, Yanase M, Akiyama J, Mizokami M, Uemura N (2016) Development and validation of a risk scoring system for severe acute lower gastrointestinal bleeding. *Clin Gastroenterol Hepatol* 14(11):1562–70 e2
- Committee ASoP, Pasha SF, Shergill A, Acosta RD, Chandrasekhara V, Chathadi KV et al (2014) The role of endoscopy in the patient with lower GI bleeding. *Gastrointest Endosc* 79(6):875–885
- Oakland K, Jairath V, Uberoi R, Guy R, Ayaru L, Mortensen N, Murphy MF, Collins GS (2017) Derivation and validation of a novel risk score for safe discharge after acute lower gastrointestinal bleeding: a modelling study. *Lancet Gastroenterol Hepatol* 2(9):635–643
- Rutter AV, Crees J, Wright H, Raseta M, van Pittius DG, Roach P et al (2019) Identification of a glass substrate to study cells using Fourier transform infrared spectroscopy: are we closer to spectral pathology? *Appl Spectrosc* 3702819875828
- Geisser S (1993) Predictive inference. Chapman & Hall
- Zhang Z (2016) Variable selection with stepwise and best subset approaches. *Ann Transl Med* 4(7):136
- Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR (1996) A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 49(12):1373–1379
- Youden WJ (1950) Index for rating diagnostic tests. *Cancer* 3(1):32–35
- Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC et al (2011) pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12:77
- Tapaskar N, Jones B, Mei S, Sengupta N (2018) Comparison of clinical prediction tools and identification of risk factors for adverse outcomes in acute lower gi bleeding. *Gastrointest Endosc*

23. Xavier SA, Machado FJ, Magalhaes JT, Cotter JB (2018) Acute lower gastrointestinal bleeding: are STRATE and BLEED scores valid in clinical practice? *Colorectal Dis*
24. Hreinsson JP, Sigurdardottir R, Lund SH, Bjornsson ES (2018) The SHA2PE score: a new score for lower gastrointestinal bleeding that predicts low-risk of hospital-based intervention. *Scand J Gastroenterol*:1–6
25. Steyerberg EW, Vergouwe Y (2014) Towards better clinical prediction models: seven steps for development and an ABCD for validation. *Eur Heart J* 35(29):1925–31

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.