

# Development and internal validation of prehospital prediction models for identifying intracerebral haemorrhage in suspected stroke patients

Mohammed Almubayyidh <sup>1,2</sup>, Adrian R Parry-Jones,<sup>1,3</sup> David A Jenkins<sup>4</sup>

**To cite:** Almubayyidh M, Parry-Jones AR, Jenkins DA. Development and internal validation of prehospital prediction models for identifying intracerebral haemorrhage in suspected stroke patients. *BMJ Neurology Open* 2024;6:e000878. doi:10.1136/bmjno-2024-000878

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/bmjno-2024-000878>).

Received 16 August 2024  
Accepted 29 September 2024



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Mohammed Almubayyidh;  
mohammed.almubayyidh@postgrad.manchester.ac.uk

## ABSTRACT

**Introduction** Distinguishing patients with intracerebral haemorrhage (ICH) from other suspected stroke cases in the prehospital setting is crucial for determining the appropriate level of care and minimising the onset-to-treatment time, thereby potentially improving outcomes. Therefore, we developed prehospital prediction models to identify patients with ICH among suspected stroke cases. **Methods** Data were obtained from the Field Administration of Stroke Therapy-Magnesium prehospital stroke trial, where paramedics evaluated multiple variables in suspected stroke cases within the first 2 hours from the last known well time. A total of 19 candidate predictors were included to minimise overfitting and were subsequently refined through the backward exclusion of non-significant predictors. We used logistic regression and eXtreme Gradient Boosting (XGBoost) models to evaluate the performance of the predictors. Model performance was assessed using the area under the receiver operating characteristic curve (AUC), confusion matrix metrics and calibration measures. Additionally, models were internally validated and corrected for optimism through bootstrapping. Furthermore, a nomogram was built to facilitate paramedics in estimating the probability of ICH. **Results** We analysed 1649 suspected stroke cases, of which 373 (23%) were finally diagnosed with ICH. From the 19 candidate predictors, 9 were identified as independently associated with ICH ( $p < 0.05$ ). Male sex, arm weakness, worsening neurological status and high systolic blood pressure were positively associated with ICH. Conversely, a history of hyperlipidaemia, atrial fibrillation, coronary artery disease, ischaemic stroke and improving neurological status were associated with other diagnoses. Both logistic regression and XGBoost demonstrated good calibration and predictive performance, with optimism-corrected sensitivities ranging from 47% to 49%, specificities from 89% to 90% and AUCs from 0.796 to 0.801. **Conclusions** Our models demonstrate good predictive performance in distinguishing patients with ICH from other diagnoses, making them potentially useful tools for prehospital ICH management.

## INTRODUCTION

Intracerebral haemorrhage (ICH), although comprising only 10%–15% of all stroke

## WHAT IS ALREADY KNOWN ON THIS TOPIC

- ⇒ Prehospital stroke scales currently in use are limited in their ability to identify intracerebral haemorrhage (ICH), as they were not originally designed for this purpose.
- ⇒ Differentiating ICH from other causes of stroke symptoms in the prehospital setting has become a priority, as these patients would benefit from pre-hospital blood pressure lowering and prompt triage to the closest, most appropriate stroke centre.

## WHAT THIS STUDY ADDS

- ⇒ In this secondary analysis of a prehospital clinical trial, we demonstrate that it is possible to distinguish patients with ICH from those experiencing stroke-like symptoms with acceptable predictive performance.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

- ⇒ Our findings could help prehospital clinicians identify ICH in suspected stroke cases. This, in turn, could facilitate correct triage, shorten transport times and potentially reduce onset-to-treatment time for patients with ICH.
- ⇒ Future research could explore the integration of the predictors identified in our study with newly developed point-of-care testing devices. This combined approach has the potential to improve significantly the accuracy of detecting ICH, enabling the initiation of treatment en route to hospital.

cases, is recognised as the most severe subtype of stroke, with the highest mortality rate and substantial disability among survivors.<sup>1 2</sup> Haematoma expansion, usually occurring within a few hours after symptom onset, plays a crucial role in the devastating outcomes after ICH, making it the key therapeutic target.<sup>3 4</sup> Implementing treatment strategies promptly, such as reversing anti-coagulation, intensively lowering blood pressure and providing immediate access to neurosurgical care, has been found to lower significantly the 30-day mortality rates of

patients with ICH.<sup>5</sup> The INTERACT4 study has further emphasised the importance of rapid intervention, demonstrating the clear benefits of administering intensive blood pressure-lowering treatments in the ambulance for patients with ICH.<sup>6</sup> However, this intervention was found to be harmful for those with a final diagnosis of ischaemic stroke (IS),<sup>6</sup> highlighting the need for diagnostic certainty in the prehospital setting. Mobile stroke units (MSUs)—specialised ambulances equipped with a CT scanner—can confirm diagnosis and expedite treatment.<sup>7</sup> Despite their potential, multiple barriers hinder their widespread implementation.<sup>8–9</sup> Given the time-sensitivity of ICH management, improving prehospital care could play a key role in enhancing patient outcomes; thus, it deserves particular attention.

Current prehospital stroke care pathways prioritise transferring suspected stroke patients to an endovascular therapy (EVT) capable centre, as patients with large vessel occlusion (LVO) can benefit from mechanical thrombectomy.<sup>10–11</sup> This treatment is limited by a narrow therapeutic time window to improve patient outcomes.<sup>12</sup> Recently, a prehospital trial investigated the impact of bypassing the nearest stroke centre and directly transferring patients with ICH to an EVT-capable centre.<sup>13</sup> The study found that this approach was associated with worse outcomes for ICH cases.<sup>13</sup> Patients with ICH could benefit from early stabilisation of the airway, blood pressure control and reversal of anticoagulation therapy at the nearest acute stroke centre.<sup>14</sup> Nevertheless, if transport time is not prohibitive, direct transfer to a hospital with neurosurgical capabilities could be favoured to ensure the early implementation of an ICH care bundle.<sup>5</sup> Hence, distinguishing between stroke subtypes in the prehospital setting becomes a priority, more than just recognising patients with LVO to determine the optimal transport destinations for suspected stroke cases.

Numerous prehospital stroke triage scales are in use to facilitate the rapid identification of stroke. However, these scales have been designed to identify undifferentiated stroke, rather than distinguishing between stroke subtypes.<sup>15</sup> In fact, previous studies have demonstrated that these prediction tools have limited accuracy in detecting ICH,<sup>16–17</sup> as well as in identifying patients with ICH who require neurosurgery.<sup>18</sup> This further underscores the necessity of improving the prehospital identification of ICH. The purpose of this study is to develop prehospital prediction models to identify patients with ICH among suspected stroke cases.

## METHODS

The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis+Artificial Intelligence checklist was followed when conducting this study (online supplemental appendix).<sup>19</sup>

## Data source and participant selection

The data were obtained from the Field Administration of Stroke Therapy-Magnesium (FAST-MAG) trial, which prospectively collected data from patients with suspected acute stroke in the prehospital setting.<sup>20</sup> The dataset was deidentified by the original researchers, thus eliminating the need for ethical approval for this secondary data analysis. This dataset is particularly valuable for the purposes of our study, as it contains 1700 suspected stroke cases assessed within 2 hours of the last known well time, along with a sufficient set of variables collected in the prehospital setting prior to the initiation of study drug infusion. For a detailed description of the trial design and methods, we refer the reader to the original publication.<sup>20</sup>

For this analysis, the outcomes were categorised into those with ICH and those with other diagnoses. Subarachnoid haemorrhage was excluded from the analysis due to the distinct clinical features that set it apart from other suspected stroke cases.<sup>21–22</sup>

## Sample size and selection of variables

Given a fixed sample size, we employed the sample size criteria from Riley *et al*, through the ‘pmsampsize’ package,<sup>23–24</sup> to calculate the maximum number of candidate predictor variables that could be considered during model development to minimise the risk of overfitting. To determine this, we made the following settings in the package. First, the primary outcome of our study is binary (ICH vs other diagnoses). Second, the prevalence of ICH in the studied population was 23%, corresponding to a  $\max(R^2_{cs})$  value of 0.66.<sup>25</sup> We conservatively assumed that the models would explain 15% of the variability, leading to an anticipated  $R^2_{cs}$  value of  $0.15 \times 0.66 = 0.10$ . Based on these parameters, the calculation results showed that we needed at least 1614 cases in total, including at least 372 positive cases of ICH, to examine 19 predictor variables for inclusion in the models.

Based on our literature search in this area,<sup>26</sup> the 19 selected variables were age, male sex, history of hypertension, diabetes mellitus, hyperlipidaemia, atrial fibrillation, coronary artery disease, IS, transient ischaemic attack (TIA), ICH, current alcohol use, facial droop, grip weakness or lack of strength, arm weakness (drift or falls down), unilateral weakness, decreased level of consciousness (LOC), improvement in neurological status, deterioration in neurological status and systolic blood pressure (SBP) levels. Diastolic blood pressure was not included in the analysis due to collinearity with SBP.

The neurological symptoms (weakness of the face, hand grip, arm or unilateral weakness) were assessed using the Los Angeles Prehospital Stroke Scale (LAPSS), a stroke recognition screening tool.<sup>27</sup> LOC was evaluated in this study using the prehospital Glasgow Coma Scale (GCS) score. Decreased LOC was defined as a prehospital GCS score of <15. To assess the degree of change in the patient’s neurological status from their first arrival on the scene to arrival at the emergency department, paramedics in the FAST-MAG trial used the Paramedic Global

**Table 1** Baseline characteristics, medical history, time intervals, symptoms and vital signs of the investigated cohort

Variables	All patients (n=1649)	ICH (n=373)	Other diagnoses (n=1276)	P value*
Demographic				
Age (years), median (IQR)	71 (59–81)	64 (55–76)	72 (60–82)	<0.001†
Number of male patients (relative number in %)	947 (57.4%)	250 (67.0%)	697 (54.6%)	<0.001‡
Medical history				
Hypertension	1285 (77.9%)	295 (79.1%)	990 (77.6%)	0.586‡
Diabetes mellitus	365 (22.1%)	70 (18.8%)	295 (23.1%)	0.087‡
Hyperlipidaemia	773 (46.9%)	133 (35.7%)	640 (50.2%)	<0.001‡
Atrial fibrillation	363 (22.0%)	30 (8.0%)	333 (26.1%)	<0.001‡
Coronary artery disease	346 (21.0%)	44 (11.8%)	302 (23.7%)	<0.001‡
Prior IS	108 (6.5%)	11 (2.9%)	97 (7.6%)	0.002‡
Prior TIA	151 (9.2%)	23 (6.2%)	128 (10.0%)	0.030‡
Prior ICH	23 (1.4%)	9 (2.4%)	14 (1.1%)	0.098‡
Current alcohol use	644 (39.1%)	169 (45.3%)	475 (37.2%)	0.006‡
Time intervals				
Onset-to-prehospital evaluation (mins), median (IQR)	23 (14–42)	23 (14–39)	24 (14–43)	0.232†
Prehospital-to-hospital arrival (mins), median (IQR)	32 (27–39)	32 (27–39)	33 (28–39)	0.569†
Onset-to-hospital arrival (mins), median (IQR)	58 (46–79)	58 (45–75)	59 (46–80)	0.154†
Symptoms				
Facial droop	1406 (85.3%)	320 (85.8%)	1086 (85.1%)	0.808‡
Grip weakness or lack of strength	1526 (92.5%)	363 (97.3%)	1163 (91.1%)	<0.001‡
Arm weakness (drift or falls down)	1499 (90.9%)	367 (98.4%)	1132 (88.7%)	<0.001‡
Unilateral weakness	1645 (99.8%)	373 (100.0%)	1272 (99.7%)	0.628‡
Decreased level of consciousness	429 (26.0%)	76 (20.4%)	353 (27.7%)	0.006‡
Improvement in neurological status	528 (32.0%)	62 (16.6%)	466 (36.5%)	<0.001‡
Deterioration in neurological status	92 (5.6%)	50 (13.4%)	42 (3.3%)	<0.001‡
Vital signs				
SBP (mm Hg), median (IQR)	160 (140–180)	176 (160–194)	155 (138–175)	<0.001†

\*P values were calculated using the Mann-Whitney U test for continuous variables or the  $\chi^2$  test for categorical variables to assess differences between ICH and other diagnoses; p values less than 0.05 indicated statistical significance.  
 †Mann-Whitney U test.  
 ‡ $\chi^2$  test  
 ICH, intracerebral haemorrhage; IS, ischaemic stroke; SBP, systolic blood pressure; TIA, transient ischaemic attack.

Impression of Change (PGIC) scale.<sup>28</sup> Improvement in neurological status was defined as a PGIC of 1–2, while deterioration in neurological status was defined as a PGIC of 4–5.

### Statistical analysis, models development and validation

The data were used to develop three different models using different modelling methods: one logistic regression model and two eXtreme Gradient Boosting (XGBoost) models. There were less than 5% missing data for variables of interest; therefore, a complete case analysis was carried out.<sup>29</sup>

Descriptive statistics were performed to outline the baseline characteristics of the cohort. Categorical data were expressed as frequencies and percentages, and numerical data were presented as medians (IQR).

Regardless of the univariable findings, all variables were used during the modelling and development process. The first XGBoost model was developed using all variables. Additionally, we employed backward elimination in multivariable logistic regression to select the strongest predictors for the logistic regression model and the second XGBoost model, retaining only those independent predictors that were statistically significant ( $p < 0.05$ ). ORs and 95% CIs for the included predictors were determined using the logistic regression model. We also used the XGBoost gain metric to determine the importance of these predictors on model performance throughout the boosting process, with a higher gain indicating greater importance in generating predictions in the XGBoost model.

**Table 2** ORs after backward selection and feature importance scores from XGBoost

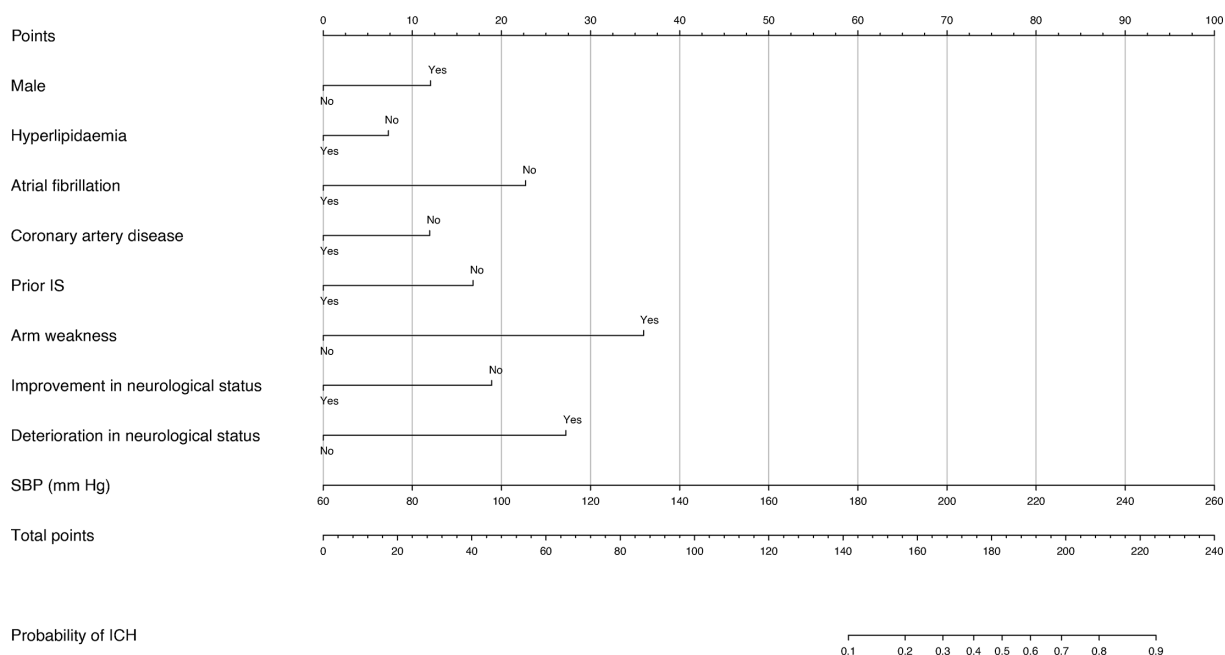
Variables	OR (95% CI)	Importance (gain)
Male	1.90 (1.44 to 2.49)***	0.06
Hyperlipidaemia	0.68 (0.51 to 0.89)**	0.04
Atrial fibrillation	0.30 (0.20 to 0.46)***	0.12
Coronary artery disease	0.53 (0.36 to 0.78)**	0.05
Prior IS	0.41 (0.21 to 0.81)*	0.02
Arm weakness (drift or falls down)	6.74 (2.86 to 15.89)***	0.08
Improvement in neurological status	0.37 (0.27 to 0.51)***	0.11
Deterioration in neurological status	4.24 (2.57 to 6.98)***	0.11
SBP	1.03 (1.02 to 1.03)***	0.41

\*p<0.05, \*\*p<0.01, \*\*\*p<0.001.

IS, ischaemic stroke; SBP, systolic blood pressure; XGBoost, eXtreme Gradient Boosting.

We used the full sample without splitting it into training and test sets, as recommended by Riley *et al.*<sup>25</sup> The models were internally validated with 1000 bootstrap resamples to correct for optimism in predictive measures, including sensitivity, specificity, positive predictive value and negative predictive value, as well as in discrimination and calibration performance. Optimism was calculated as the average difference in performance measures between the bootstrap samples and the original sample. The optimism-corrected estimates were obtained by subtracting the optimism from the performance measures of the original models. For all models, a probability cut-off of 0.4 was selected for classifications to enhance sensitivity while maintaining high specificity. Furthermore, sensitivity and specificity for each model were corrected for optimism and evaluated over the entire range of probability cut-offs.

The discrimination was assessed by the area under the receiver operating characteristic curve (AUC) to evaluate how effectively the models distinguished between the two groups of interest. An AUC of 1.0 indicates perfect discrimination, while an AUC of 0.5 suggests that classification is no better than random prediction. Calibration for the models was evaluated through calibration curves, slopes and intercepts to analyse the alignment between predicted and actual outcomes. Graphically, a 45° line signifies perfect calibration, whereas any deviation from this line reflects the extent of miscalibration. Numerically, the slope and intercept of the calibration curve describe the degree of calibration. A model is considered perfectly calibrated if the calibration curve has a slope of 1 and an intercept of 0.



**Figure 1** Logistic regression model nomogram for estimating the probability of ICH in suspected stroke cases. ICH, intracerebral haemorrhage; IS, ischaemic stroke; SBP, systolic blood pressure.

**Table 3** Performance measures of models for distinguishing patients with ICH

Statistical model used	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	AUC (95% CI)	Slope (95% CI)	Intercept (95% CI)
Model training							
XGBoost (19 variables)	53% (47.3% to 57.7%)	92% (90.0% to 93.1%)	65% (59.8% to 69.7%)	87% (84.9% to 88.6%)	0.849 (0.828 to 0.870)	0.929 (0.908 to 0.951)	0.030 (0.018 to 0.042)
Logistic regression	50% (44.7% to 55.1%)	90% (87.7% to 91.1%)	58% (53.0% to 63.2%)	86% (83.9% to 87.8%)	0.801 (0.777 to 0.826)	1.078 (1.057 to 1.099)	-0.023 (-0.033 to -0.012)
XGBoost (9 variables)	51% (45.7% to 56.1%)	91% (89.3% to 92.5%)	62% (57.1% to 67.1%)	86% (84.4% to 88.2%)	0.828 (0.805 to 0.850)	0.946 (0.926 to 0.965)	0.016 (0.006 to 0.027)
Optimism-corrected performance							
XGBoost (19 variables)	47% (42.0% to 51.1%)	90% (88.6% to 91.7%)	58% (53.6% to 62.9%)	85% (83.7% to 86.9%)	0.801 (0.783 to 0.819)	0.912 (0.894 to 0.931)	0.031 (0.020 to 0.041)
Logistic regression	49% (44.3% to 54.4%)	89% (87.6% to 91.0%)	57% (52.2% to 63.4%)	86% (83.8% to 87.8%)	0.796 (0.770 to 0.822)	1.061 (1.039 to 1.082)	-0.019 (-0.029 to -0.008)
XGBoost (9 variables)	48% (42.8% to 53.0%)	90% (88.6% to 91.8%)	59% (53.9% to 64.1%)	86% (83.7% to 87.4%)	0.799 (0.778 to 0.820)	0.934 (0.917 to 0.952)	0.018 (0.009 to 0.028)
AUC, area under the curve; ICH, intracerebral haemorrhage; NPV, negative predictive value; PPV, positive predictive value; XGBoost, eXtreme Gradient Boosting.							

Lastly, a nomogram was constructed based on the corrected logistic regression to facilitate the distinction between ICH and other diagnoses. In the nomogram, each predictor was assigned a score, which was summed up to a total point corresponding to the predicted probability of ICH. The variable with the highest predictive strength was assigned a maximum of 100 points, while other variables were assigned lower maximum values proportional to their effects. All statistical analyses were performed by using R Software (R Core Team, Vienna, Austria, V.4.2.2).

### Patient and public involvement

Neither patients nor the public were involved in the design of this study.

## RESULTS

### Clinical characteristics

Of the 1700 patients enrolled in the FAST-MAG trial,<sup>20</sup> 1649 met our eligibility criteria. Among these, 373 (23%) were diagnosed with ICH, while 1276 had other diagnoses, including 1013 (61%) with IS, 199 (12%) with TIA, 63 (4%) with stroke mimics (SM) and 1 (0.1%) with an unclassifiable diagnosis. Most of the included variables showed significant differences between ICH and other diagnoses ( $p < 0.05$ ). The baseline characteristics of the investigated cohort are summarised in [table 1](#).

### Selection of predictors

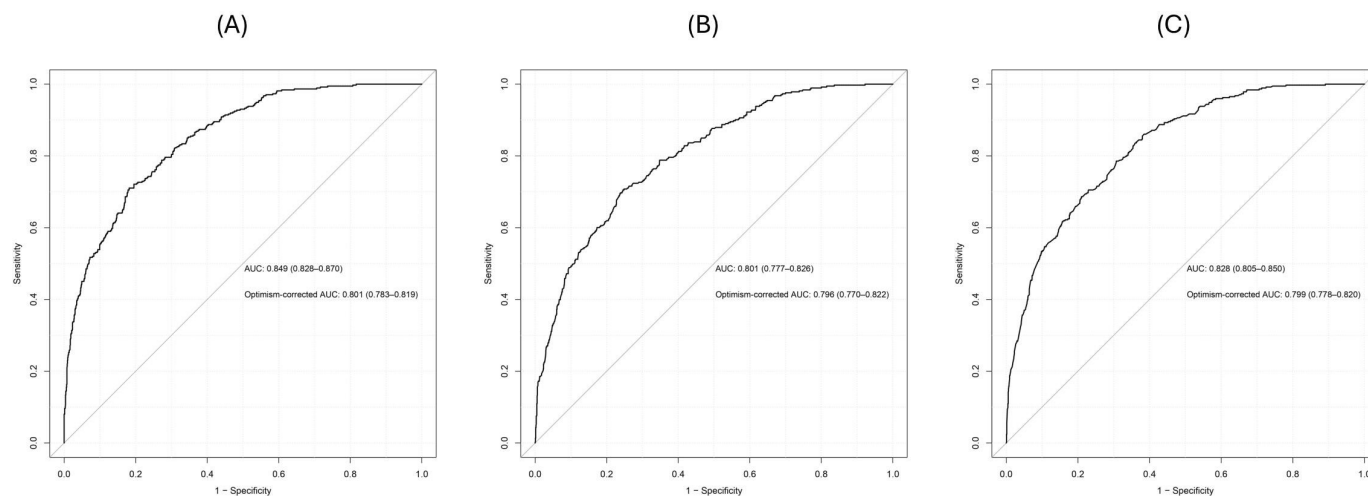
Following backward elimination, the variables selected for final inclusion in the logistic regression and second XGBoost models were male sex, history of hyperlipidaemia, atrial fibrillation, coronary artery disease, prior IS, arm weakness, improvement in neurological status, deterioration in neurological status and SBP levels. The ORs with 95% CIs for the logistic regression model and the feature importance scores for the second XGBoost model are outlined in [table 2](#).

The developed nomogram for estimating the likelihood of ICH in suspected stroke cases is shown in [figure 1](#). For instance, consider a male patient (12 points) without a history of atrial fibrillation (23 points), who presents with unilateral arm weakness (36 points) and experiences a deterioration in neurological status during transport (27 points+19 points for the lack of improvement) and has an SBP of 176mm Hg (58 points). The total score would be 175 points, corresponding to a 0.4 (40%) probability of ICH.

### Performance and internal validation of the models

The performance measures of the original and optimism-corrected models are summarised in [table 3](#). Overall, the models demonstrated good performance in distinguishing ICH from other diagnoses.

In the internal validation, the models showed comparable discrimination after optimism correction by bootstrapping ([figure 2](#)). The first XGBoost model with all



**Figure 2** ROC curves for the models with original and optimism-corrected AUCs and their 95% CIs. (A) XGBoost with 19 variables, (B) logistic regression, (C) XGBoost with nine variables. AUCs, area under the curves; ROC, receiver operating characteristic; XGBoost, eXtreme Gradient Boosting.

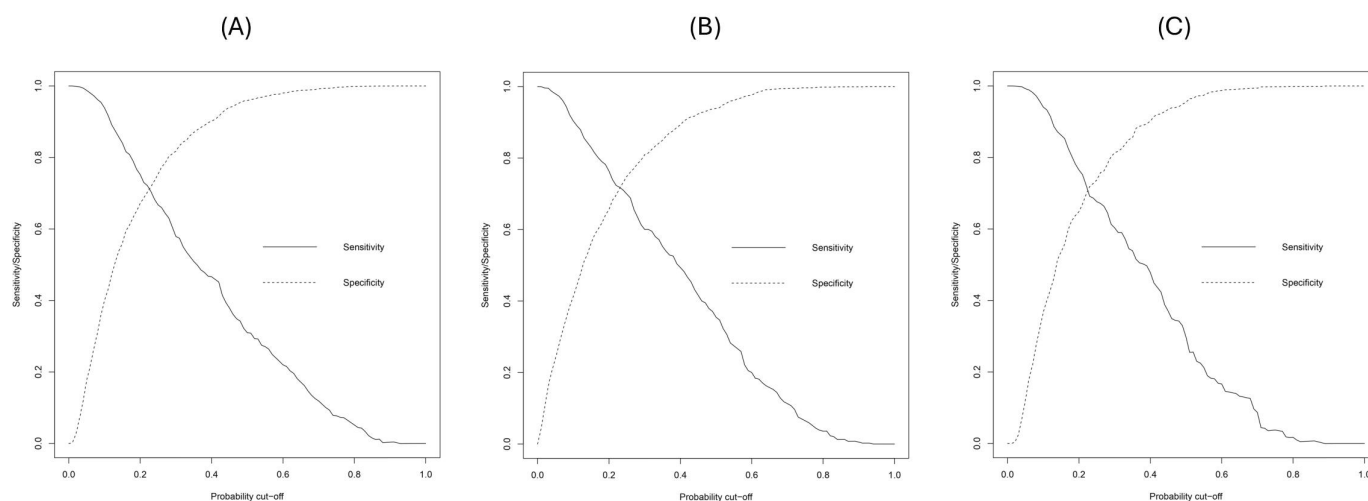
19 variables achieved an optimism-corrected sensitivity of 47% (95% CI 42.0% to 51.1%), a specificity of 90% (95% CI 88.6% to 91.7%) and an AUC of 0.801 (95% CI 0.783 to 0.819). Using the nine predictor variables from backward elimination, the logistic regression model yielded an optimism-corrected sensitivity of 49% (95% CI 44.3% to 54.4%), a specificity of 89% (95% CI 87.6% to 91.0%) and an AUC of 0.796 (95% CI 0.770 to 0.822). Meanwhile, the second XGBoost model had an optimism-corrected sensitivity of 48% (95% CI 42.8% to 53.0%), a specificity of 90% (95% CI 88.6% to 91.8%) and an AUC of 0.799 (95% CI 0.778 to 0.820). **Figure 3** summarises the corrected sensitivity and specificity of the three models at different probability cut-offs, which can be adjusted to improve either sensitivity or specificity as required.

Additionally, we found that the calibration curves for each prediction method were in good agreement with the classifications and actual observations after verification by bootstrap resampling (**figure 4**). All the calibration

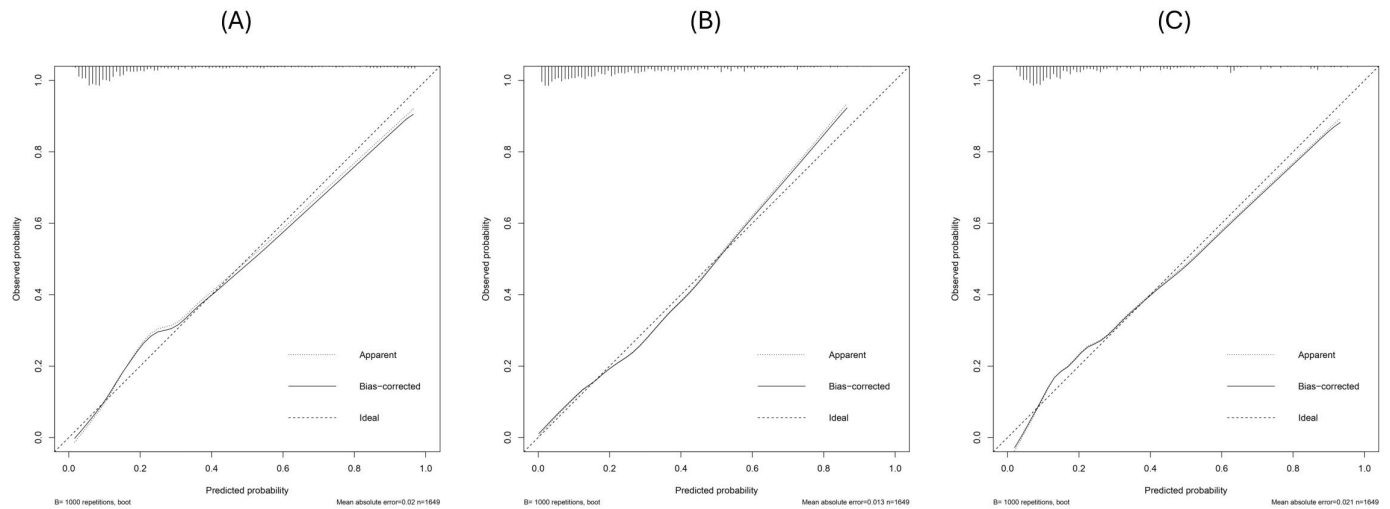
curves of the models were closely aligned with the 45° line, with the optimism-corrected slopes and intercepts ranging from 0.912 to 1.061 and  $-0.019$  to  $0.031$ , respectively (**table 3**).

## DISCUSSION

Patients with ICH often require interventions that are highly time-sensitive and are typically available only in specialised hospitals. Consequently, developing a prehospital prediction model to distinguish these patients from patients with other conditions that present with similar symptoms has become increasingly important. Such prediction methods could also facilitate the selection of patients for future prehospital intervention trials.<sup>30</sup> Recently, several prediction models have been proposed to distinguish patients with ICH in the prehospital setting.<sup>21 31 32</sup> However, these models were validated only to distinguish between stroke subtypes, without



**Figure 3** Corrected sensitivity and specificity at various probability cut-offs. (A) XGBoost with 19 variables, (B) logistic regression, (C) XGBoost with nine variables. XGBoost, eXtreme Gradient Boosting.



**Figure 4** Calibration curves of the models. (A) XGBoost with 19 variables, (B) logistic regression, (C) XGBoost with nine variables. XGBoost, eXtreme Gradient Boosting.

considering other SM conditions, which account for 26% of suspected stroke cases.<sup>33</sup> Hence, they may not capture the complexity of prehospital care adequately. Machine learning algorithms have also been employed to predict ICH and other types of strokes, but the specific predictors used to distinguish ICH from other conditions were not reported.<sup>34 35</sup> Here, we developed prehospital prediction models using a traditional method (logistic regression) and a machine learning approach (XGBoost) and provided a nomogram to distinguish patients with ICH from those with other conditions.

The AUCs of the models in this study ranged from 0.796 to 0.801 after bootstrap correction for optimism. Our models achieved higher AUCs compared with the study by Geisler *et al* (0.75),<sup>31</sup> comparable to those reported by Uchida *et al* (0.81–0.82),<sup>35</sup> but were lower than those reported by Hayashi *et al* (0.866).<sup>34</sup> Although Hayashi *et al*<sup>34</sup> achieved a higher AUC, their study included a wide range of 52 predictive features to improve performance. In contrast, we adhered to the guidelines proposed by Riley *et al*<sup>25</sup> to minimise overfitting and ensure a precise estimation of the number of variables that could be considered. Additionally, none of the existing prehospital models for ICH reported calibration measures, which are crucial since models with satisfactory discrimination may still have poor calibration,<sup>36</sup> limiting their prehospital utility. Our models, however, demonstrated good discrimination and calibration performance. The nine predictors identified in this study are easy to obtain in the prehospital setting and are represented in a nomogram, facilitating decision-making by prehospital clinicians and enhancing its practical application.

In terms of distinguishing features, previous prehospital prediction studies have found that younger age is associated with ICH,<sup>21 22</sup> while a history of diabetes,<sup>21</sup> atrial fibrillation,<sup>21 22</sup> hyperlipidaemia,<sup>22</sup> coronary artery disease<sup>22</sup> and IS<sup>32</sup> are associated with other diagnoses. These findings mostly agree with our results (table 2), although we found that male sex is significantly and

independently associated with ICH (OR 1.90,  $p < 0.001$ ). Regarding neurological symptoms, the paramedics in the FAST-MAG study used LAPSS, which consists of three items to assess unilateral deficits (face, hand grip and arm). Of these three items, only arm weakness was strongly associated with ICH in our study (OR 6.74,  $p < 0.001$ ). This finding partially aligns with that of Geisler *et al*,<sup>31</sup> where motor weakness of the arm or leg was included in their final model.

Interestingly, impaired LOC was not found to be associated with ICH at the time of the paramedics' arrival. This contradicts previous models, which reported an association between decreased LOC and an increased likelihood of ICH.<sup>31 32</sup> Part of the explanation for our finding can be attributed to the characteristics of the data, as we investigated patients within the first 2 hours of the time when they were last known to be well, whereas other prediction studies did not account for the duration between symptom onset and prehospital evaluation.<sup>31 32</sup> Notably, within a median of 32 min (IQR, 27–39) of prehospital care (table 1), a decline in neurological status was significantly associated with ICH (OR 4.24,  $p < 0.001$ ), while an improvement in neurological status was associated with other diagnoses (OR 0.37,  $p < 0.001$ ). This neurological deterioration may be due to haematoma expansion after ICH, which has been previously investigated in the prehospital setting using MSUs and found to occur frequently within the first 2 hours after symptom onset.<sup>3</sup> Hence, it is important to mitigate the risk of haematoma expansion within this critical period through necessary interventions, such as lowering SBP levels to 130–140 mm Hg.<sup>37</sup> This was proven to be effective in improving outcomes in the INTERACT4 trial for patients enrolled within 2 hours after symptom onset in the prehospital setting.<sup>6</sup> In this regard, prehospital SBP had the highest gain score in the XGBoost model and the highest point value in the nomogram (table 2 and figure 1), with higher SBP levels being associated with a higher probability of ICH, consistent with other prehospital models.<sup>21 31 32</sup>

Using a probability cut-off of 0.4, our models identified nearly half of the ICH cases and accurately ruled out 90% of other diagnoses. This could assist prehospital personnel in determining the appropriate level of hospital care, avoiding long transport times and thereby potentially reducing the onset-to-treatment time for patients with ICH. Additionally, it may help in prenotifying the receiving hospital of a suspected ICH, further expediting specialist assessment and treatment on the patient's arrival. However, a higher level of diagnostic certainty is necessary before considering initiating treatment in an ambulance. This ensures patient safety and avoids the risk of administering potentially harmful therapies to the wrong patient, such as treating a patient with IS with intensive blood pressure-lowering interventions.<sup>6</sup>

To improve diagnostic accuracy and reduce errors, future research might consider using the predictors identified in our study in combination with a simple point-of-care diagnostic test (POCT) to aid decision-making in the prehospital setting. While some POCT technologies are being developed,<sup>26</sup> measuring brain-related blood biomarkers using POCT devices is a promising solution.<sup>38</sup> This method is low-cost and can be evaluated easily in the field. Among the brain-related biomarkers, glial fibrillary acidic protein (GFAP) could serve as a diagnostic biomarker for distinguishing ICH from IS and SM conditions within the first few hours after the onset of symptoms.<sup>39</sup> It is hypothesised that GFAP is rapidly released into the bloodstream in cases of ICH due to blood-brain barrier disruption from the initial haematoma, but in patients with IS, the release is delayed.<sup>39</sup> A meta-analysis of 12 studies showed that plasma GFAP had a pooled sensitivity of 78% (95% CI 67% to 86%) and a specificity of 95% (95% CI 88% to 98%) in distinguishing patients with ICH from IS and SM conditions.<sup>40</sup> Given the high specificity, combining such a diagnostic method with a simple assessment of predictors could provide a cost-effective alternative to MSUs to expedite the treatment of patients with ICH in the field. However, this hypothesis has yet to be tested in a large prehospital trial of suspected stroke patients.

### Limitations

This study has several limitations. First, our analysis was confined to patients enrolled in the FAST-MAG trial, which was largely composed of white and Hispanic populations in the USA and excluded patients with SBP greater than 220 mm Hg or those in a comatose state. Consequently, the applicability of our findings to the general population is unclear. Second, we did not externally validate the models because a similar database is currently unavailable. Although bootstrap validation takes model overoptimism into account, and it is considered superior to data splitting,<sup>25</sup> further prehospital studies are needed to confirm the robustness and generalisability of our models. Additionally, other common ICH-related symptoms, such as headache and vomiting, were not collected in FAST-MAG study, yet they might have been helpful in

increasing the accuracy of distinguishing patients with ICH in our models; thus, they are worth investigating in future research.

### CONCLUSIONS

We have developed prehospital models that show promising results in distinguishing patients with ICH from other patients presenting with stroke-like symptoms, achieving moderate sensitivity and high specificity. This could assist in making informed decisions about patient destinations. However, external validation is necessary to confirm the performance and reliability of our models. Future research should also explore integrating POCT devices to further enhance the accuracy of identifying ICH, potentially enabling earlier initiation of appropriate treatments in the prehospital phase for these patients.

### Author affiliations

<sup>1</sup>Division of Cardiovascular Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

<sup>2</sup>Department of Aviation and Marine, Prince Sultan Bin Abdulaziz College for Emergency Medical Services, King Saud University, Riyadh, Saudi Arabia

<sup>3</sup>Geoffrey Jefferson Brain Research Centre, Manchester Academic Health Science Centre, Northern Care Alliance & University of Manchester, Manchester, UK

<sup>4</sup>Division of Informatics, Imaging and Data Science, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK

**Acknowledgements** This research is based on the National Institute of Neurologic Disease and Stroke's Archived Clinical Research data (Trial name: Field Administration of Stroke Therapy-Magnesium Trial; Principal Investigator: Jeffrey L. Saver, MD; grant number: U01NS044364) received from the Archived Clinical Research Dataset web site.

**Contributors** All the authors of this study fulfil the criteria of authorship. MA is the primary author and guarantor of this research, under the supervision of ARP-J and DAJ. MA conceptualised the idea and designed the methods, incorporating suggestions from ARP-J and DAJ. MA had full access to all study data and took responsibility for the integrity and accuracy of the analysis. MA drafted the manuscript with support from ARP-J and DAJ. The manuscript was critically reviewed and approved by all authors for final submission.

**Funding** The study is part of MA's PhD project, which is funded by King Saud University, Riyadh, Saudi Arabia (grant number: NA), through the Saudi Arabian Cultural Bureau in the UK (grant number: NA).

**Competing interests** None declared.

**Patient consent for publication** Not applicable.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data may be obtained from a third party and are not publicly available. The FAST-MAG database analysed in this study was obtained from the National Institute of Neurological Disorders and Stroke (<https://www.ninds.nih.gov/current-research/research-funded-ninds/clinical-research/archived-clinical-research-datasets>). Requests to access the database should be directed to the Clinical Research Liaison at [CRLiaison@ninds.nih.gov](mailto:CRLiaison@ninds.nih.gov).

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially,



and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

#### ORCID iD

Mohammed Almubayyidh <http://orcid.org/0000-0003-3266-1236>

#### REFERENCES

- Ziai WC, Carhuapoma JR. Intracerebral Hemorrhage. *Continuum (Mount Lawley)* 2018;24:1603–22.
- Lioutas V-A, Beiser AS, Aparicio HJ, et al. Assessment of Incidence and Risk Factors of Intracerebral Hemorrhage Among Participants in the Framingham Heart Study Between 1948 and 2016. *JAMA Neurol* 2020;77:1252–60.
- Bowry R, Parker SA, Bratina P, et al. Hemorrhage Enlargement Is More Frequent in the First 2 Hours: A Prehospital Mobile Stroke Unit Study. *Stroke* 2022;53:2352–60.
- Puy L, Parry-Jones AR, Sandset EC, et al. Intracerebral haemorrhage. *Nat Rev Dis Primers* 2023;9:14.
- Parry-Jones AR, Sammut-Powell C, Paroutoglou K, et al. An Intracerebral Hemorrhage Care Bundle Is Associated with Lower Case Fatality. *Ann Neurol* 2019;86:495–503.
- Li G, Lin Y, Yang J, et al. Intensive Ambulance-Delivered Blood-Pressure Reduction in Hyperacute Stroke. *N Engl J Med* 2024;390:1862–72.
- Cooley SR, Zhao H, Campbell BCV, et al. Mobile Stroke Units Facilitate Prehospital Management of Intracerebral Hemorrhage. *Stroke* 2021;52:3163–6.
- Walter S, Grunwald IQ, Helwig SA, et al. Mobile Stroke Units - Cost-Effective or Just an Expensive Hype? *Curr Atheroscler Rep* 2018;20:49.
- Gioia LC, Mendes GN, Poppe AY, et al. Advances in Prehospital Management of Intracerebral Hemorrhage. *Cerebrovasc Dis* 2024.
- Ramos A, Guerrero WR, Pérez de la Ossa N. Prehospital Stroke Triage. *Neurology (Ecranicon)* 2021;97:S25–s33.
- Ospel JM, Dmytriw AA, Regenhardt RW, et al. Recent developments in pre-hospital and in-hospital triage for endovascular stroke treatment. *J Neurointerv Surg* 2023;15:1065–71.
- Saver JL, Goyal M, van der Lugt A, et al. Time to Treatment With Endovascular Thrombectomy and Outcomes From Ischemic Stroke: A Meta-analysis. *JAMA* 2016;316:1279–88.
- Ramos-Pachón A, Rodríguez-Luna D, Martí-Fàbregas J, et al. Effect of Bypassing the Closest Stroke Center in Patients with Intracerebral Hemorrhage: A Secondary Analysis of the RACECAT Randomized Clinical Trial. *JAMA Neurol* 2023;80:1028–36.
- Richards CT, Oostema JA, Chapman SN, et al. Prehospital Stroke Care Part 2: On-Scene Evaluation and Management by Emergency Medical Services Practitioners. *Stroke* 2023;54:1416–25.
- Greenberg SM, Ziai WC, Cordonnier C, et al. 2022 Guideline for the Management of Patients With Spontaneous Intracerebral Hemorrhage: A Guideline From the American Heart Association/American Stroke Association. *Stroke* 2022;53:e282–361.
- Kleindorfer DO, Miller R, Moomaw CJ, et al. Designing a message for public education regarding stroke: does FAST capture enough stroke? *Stroke* 2007;38:2864–8.
- Oostema JA, Chassee T, Baer W, et al. Accuracy and Implications of Hemorrhagic Stroke Recognition by Emergency Medical Services. *Prehosp Emerg Care* 2021;25:796–801.
- Almqvist T, Falk Delgado A, Sjöstrand C, et al. Impact of prehospital stroke triage implementation on patients with intracerebral hemorrhage. *Ther Adv Neurol Disord* 2023;16:17562864231168278.
- Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378.
- Saver JL, Starkman S, Eckstein M, et al. Prehospital Use of Magnesium Sulfate as Neuroprotection in Acute Stroke. *N Engl J Med* 2015;372:528–36.
- Jin H-Q, Wang J-C, Sun Y-A, et al. Prehospital Identification of Stroke Subtypes in Chinese Rural Areas. *Chin Med J (Engl)* 2016;129:1041–6.
- Yamashita S, Kimura K, Iguchi Y, et al. Kurashiki Prehospital Stroke Subtyping Score (KP3S) as a means of distinguishing ischemic from hemorrhagic stroke in emergency medical services. *Eur Neurol* 2011;65:233–8.
- Riley RD, Snell KI, Ensor J, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat Med* 2019;38:1276–96.
- Ensor J, Martin EC, Riley RD. pmsampsize: calculates the minimum sample size required for developing a multivariable prediction model. R package version 1.1.2. 2022.
- Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020;368:m441.
- Almubayyidh M, Alghamdi I, Parry-Jones AR, et al. Prehospital identification of intracerebral haemorrhage: a scoping review of early clinical features and portable devices. *BMJ Open* 2024;14:e079316.
- Saver JL, Starkman S, Eckstein M, et al. Methodology of the Field Administration of Stroke Therapy – Magnesium (FAST-MAG) Phase 3 Trial: Part 2 – Prehospital Study Methods. *Int J Stroke* 2014;9:220–5.
- Shkirkova K, Schuberg S, Balouziyan E, et al. Paramedic Global Impression of Change During Prehospital Evaluation and Transport for Acute Stroke. *Stroke* 2020;51:784–91.
- Schafer JL. Multiple imputation: a primer. *Stat Methods Med Res* 1999;8:3–15.
- Sandset EC, Walter S, Song L. The challenges of large-scale prehospital stroke trials. *Lancet Neurol* 2022;21:948–9.
- Geisler F, Wesirow M, Ebinger M, et al. Probability assessment of intracerebral hemorrhage in prehospital emergency patients. *Neurol Res Pract* 2021;3:1.
- Uchida K, Yoshimura S, Hiyama N, et al. Clinical Prediction Rules to Classify Types of Stroke at Prehospital Stage. *Stroke* 2018;49:1820–7.
- Gibson LM, Whiteley W. The differential diagnosis of suspected stroke: a systematic review. *J R Coll Physicians Edinb* 2013;43:114–8.
- Hayashi Y, Shimada T, Hattori N, et al. A prehospital diagnostic algorithm for strokes using machine learning: a prospective observational study. *Sci Rep* 2021;11:20519.
- Uchida K, Kouno J, Yoshimura S, et al. Development of Machine Learning Models to Predict Probabilities and Types of Stroke at Prehospital Stage: the Japan Urgent Stroke Triage Score Using Machine Learning (JUST-ML). *Transl Stroke Res* 2022;13:370–81.
- Alba AC, Agoritsas T, Walsh M, et al. Discrimination and Calibration of Clinical Prediction Models: Users' Guides to the Medical Literature. *JAMA* 2017;318:1377–84.
- Arima H, Anderson CS, Wang JG, et al. Lower treatment blood pressure is associated with greatest reduction in hematoma growth after acute intracerebral hemorrhage. *Hypertension* 2010;56:852–8.
- Zylyftari S, Luger S, Blums K, et al. GFAP point-of-care measurement for prehospital diagnosis of intracranial hemorrhage in acute coma. *Crit Care* 2024;28:109.
- Mayer CA, Brunkhorst R, Niessner M, et al. Blood levels of glial fibrillary acidic protein (GFAP) in patients with neurological diseases. *PLoS One* 2013;8:e62101.
- Kumar A, Misra S, Yadav AK, et al. Role of glial fibrillary acidic protein as a biomarker in differentiating intracerebral haemorrhage from ischaemic stroke and stroke mimics: a meta-analysis. *Biomarkers* 2020;25:1–8.