**C-reactive protein in outcome prediction after subarachnoid haemorrhage and the role of machine learning.**

*Gaastra, Ben1\*; MRCS*

*Barron, Peter2;*

*Newitt, Laura2; BMBS*

*Chhugani, Simran2;*

*Turner, Carole3; MSc*

*Kirkpatrick, Peter3; FRCS (SN)*

*MacArthur, Ben4; PhD*

*Galea, Ian5; PhD*

*Bulters, Diederik1; FRCS (SN)*

1 Department of Neurosurgery, Wessex Neurological Centre, University Hospital Southampton, Southampton, SO16 6YD, UK

2 University of Southampton Faculty of Medicine, University Road, Southampton SO17 1BJ, UK

3 Department of Neurosurgery, Cambridge University Hospital, Box 166, Hills Road, Cambridge, CB2 0QQ, UK.

4 Mathematical Sciences, University of Southampton, Southampton, SO16 6YD, UK

5 Clinical Neurosciences, Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, SO16 6YD, UK

\* corresponding author. [bgaastra@nhs.net](mailto:bgaastra@nhs.net). Tel: 07791536702

Cover title: Outcome prediction after subarachnoid haemorrhage

Tables: 2

Figures: 4

Key words: subarachnoid hemorrhage, outcome studies, C-reactive protein

Word count: 5995

**Abstract**

**Background and purpose:** Outcome prediction after aneurysmal subarachnoid haemorrhage (aSAH) is challenging. C-reactive protein (CRP) has been reported to be associated with outcome, but it is unclear if this is independent of other predictors and applies to aSAH of all grades. Therefore, the role of CRP in aSAH outcome prediction models is unknown. The purpose of this study is to assess if CRP is an independent predictor of outcome after aSAH, develop new prognostic models incorporating CRP, and test whether these can be improved by application of machine learning.

**Methods:** This was an individual patient level analysis of data from patients within 72 hours of aSAH from two prior studies. A panel of statistical learning methods including logistic regression, random forest and support vector machines (SVM) were used to assess the relationship between CRP and modified Rankin Score (mRS). Models were compared to the full SAHIT prediction tool of outcome after aSAH and internally validated using cross-validation.

**Results:** 1017 patients were included for analysis. CRP on the first day after ictus was an independent predictor of outcome. The full SAHIT model achieved an AUC of 0.831. Addition of CRP to the predictors of the full SAHIT model improved model performance (AUC = 0.846, p = 0.01). This improvement was not enhanced when learning was performed using a random forest (AUC = 0.807), but was with an SVM (AUC of 0.960, p <0.001).

**Conclusions:** CRP is an independent predictor of outcome after aSAH. Its inclusion in prognostic models improves performance, although the magnitude of improvement is probably insufficient to be relevant clinically on an individual patient level, and of more relevance in research. Greater improvements in model performance are seen with SVMs but these models have the highest classification error rate on internal validation and require external validation and calibration.

**Introduction**

Aneurysmal subarachnoid haemorrhage (aSAH) is a devastating form of stroke. Up to 50% of patients are estimated to die(1), and of the survivors 50% have cognitive impairment and 40% are unable to return to their previous work(2). Predicting which patients will have poor outcome after aSAH would be valuable in order to prognosticate and guide treatment. Better models would also improve our ability to control for covariates and reduce sample sizes for research studies.

Several outcome prediction tools following aSAH have been developed(3). The best available is the SAHIT prediction tool, developed from over 10 000 patients and validated in a separate cohort of over 3000 patients(4). Three predictive models were generated: a core, neuroimaging and full model. The models performed well, with good discrimination (AUC 0.76-0.81). The authors report R2 statistic values of 22-31%, meaning that up to 78% of variance in outcome after aSAH is not explained by the predictors included in the SAHIT models. To improve on this, it will be necessary to either develop better statistical methods, or include additional predictors, not incorporated in SAHIT.

C-reactive protein (CRP) is an acute phase reactant made in the liver and released into the blood in response to inflammation. It is routinely available making it an ideal addition to predictive models. Both serum and CSF CRP levels rise following aSAH and peak around day 3-4 post ictus(5-7). Elevated levels were associated with poor outcome following aSAH in two small series of less than 100 patients(6, 7). A further study of 178 patients highlighted the problem that although CRP levels on admission were associated with outcome, with their sample size they were not independently predictive (8). Another study of 803 patients reported CRP was only an independent predictor in good grade (World Federation of Neurological Surgeons (WFNS) grade 1 or 2) but not poor grade (WFNS 3-5) aSAH patients(9). In summary, although there is evidence CRP is associated with outcome after aSAH, it is unclear if this is independent from other known predictors or across all aSAH grades and therefore if it should be added to prediction models such as SAHIT.

The SAHIT predictive tool, along with the majority of other outcome prediction tools for aSAH, was developed using the classical statistical learning method of logistic regression(3, 4), which is a generalised linear model. More modern machine learning methods offer advantages over classical statistics as they are better able to analyse complex high dimensional data, account for nonlinear associations between predictors, and thereby generate models that generalize more successfully to unseen data(10, 11). Although machine learning techniques have shown promise for the prediction of cerebral vasospasm and delayed cerebral ischaemia(12, 13), only simple decision trees have been applied to outcome prediction. These have been shown to have comparable performance to logistic regression methods (14-16). More powerful machine learning methods, including ensemble methods such as random forests and nonlinear methods such as support vector machines, have not been explored in outcome prediction after aSAH and use of these methods may allow improve predictive models(17).

The aims of this study are therefore to: (1) assess if CRP is an independent predictor of outcome after aSAH; (2) develop improved predictive models of outcome after aSAH by including CRP in machine learning models.

**Methods**

The data that support the findings of this study are available from the corresponding author upon reasonable request. Predictive models are reported in keeping with the TRIPOD statement (checklist in the Supplemental material).

Patient data

Patients with aSAH who had at least one recorded CRP and WFNS on day 0, 1 or 2 after ictus were eligible for inclusion. De-identified individual patient level data was identified from two sources:

1. The Wessex Neurological Centre (WNC, University Hospital Southampton, UK) neurovascular database. Following National Research Ethics Service (NRES) and Health Research Authority (HRA) approval (17/LO/0964), prospectively collected data from aSAH patients aged over 18 years old admitted to the WNC between August 2009 and September 2017 was selected. Additional clinical data for serial CRP and WFNS values were collected from the electronic and paper notes retrospectively. Outcomes were collected prospectively by a specialist nurse trained in assessment of the modified Rankin Scale, and confirmed retrospectively in the paper notes.
2. The randomized multicentre controlled trial Simvastatin in Aneurysmal Subarachnoid Haemorrhage (STASH)(18)

Variables/predictors

The following variables/predictors were available:

1. Age
2. CRP (mg/L) on day 0, 1 and 2
3. WFNS grade on day 0, 1 and 2
4. Treatment status (clip, coil or no treatment)
5. Fisher grade on admission
6. Past medical history of hypertension
7. Aneurysm size (≤ 12 mm, 13-24 mm or > 24 mm)
8. Aneurysm location (anterior cerebral artery, middle cerebral artery, internal carotid artery, posterior circulation)

All data was treated as categorical apart from age and CRP which were considered as continuous variables. Aneurysm size categories were defined as per the SAHIT predictive tool(4).

The variables/predictors available are the same as the complete SAHIT predictive tool with the addition of CRP.

Outcome

Outcome was defined as best modified Rankin Scale (mRS) 3-6 months post ictus. All patients from the STASH trial had outcome at 6 months, patients from the WNC had outcome at 3-6 months.

Outcome was dichotomised into good (mRS 0-3) and poor outcome (mRS 4-6). This dichotomisation was chosen to match the SAHIT predictive tool(4).

Missing data

Missing CRP and WFNS values were imputed using multiple imputation with a predictive mean matching method(19) to generate 5 imputations. Imputation was only employed if data was missing at random. Other missing data was coded with a dummy variable to allow for inclusion of patients with missing data in the predictive models.

CRP and WFNS inclusion

Binary logistic regression and area under the receiver operator characteristics (AUC) curve analysis was used to compare the predictive values of CRP on day 0, 1 and 2. Reported AUC values were pooled across the 5 imputed datasets using Rubin’s Rules(20). To identify best predictive day the DeLong test(21) was used to assess for difference in AUC, as this cannot be implemented for pooled AUCs it is reported for the first imputed dataset. If no difference was identified the earlier time point was used in order to minimize confounding of CRP rise with secondary illness such as aspiration pneumonia. The best performing day was used for the rest of the analysis. The process was repeated to identify the best predictive performing day of WFNS recording.

Model development

A variety of statistical learning methods were used to develop the outcome prediction models.

Binary logistic regression was used at first incorporating the same predictors as in the SAHIT predictive tool (age, WFNS, past medical history of hypertension, Fisher grade, size and location of aneurysm, treatment status). The model was further developed by the addition of CRP as a predictor.

The supervised classification machine learning methods random forest and support and support vector machines (SVMs) were employed and trained on all predictors available. Both these methods were optimized to improve predictive performance. In the random forest the number of variables considered at each split of the tree and the number of trees was manually optimized. Both linear and nonlinear SVMs (using a radial basis function kernel) were trained, and all model parameters were optimised with 10-fold cross validation. Nominal categorical predictors were converted to dummy variables to allow inclusion in the SVM model.

Model performance

Models were compared to the full SAHIT predictive tool. Individual model performance was assessed using the area under the receiver operating characteristic (ROC) curve (AUC). Reported AUC values were pooled across the 5 imputed datasets using Rubin’s Rules and reported with 95% confidence intervals(20). The DeLong test was used to compare AUCs(21) and is reported for the first imputed dataset. The net reclassification improvement (NRI) and integrated discrimination improvement (IDI) were used to assess the improvement in model performance after addition of predictors.

To identify the proportion of variance explained by the logistic regression models a variance-function-based R2 statistic was used(22). In order to control for collinearity, dominance analysis was used to analyse the importance of individual predictors to a multivariable regression model(23). Variable importance from random forests were reported by Gini index and mean decrease in accuracy and for the SVM by ranking the coefficients of the support vectors.

Model validation

Logistic regression models and SVMs were trained using 10-fold cross validation to avoid over fitting. Random forests were optimised using out of bag (OOB) error, a natural measure that approaches the leave one out cross validation error as the number of trees gets large. Throughout training a data split into train and test data sets in proportions 2/3 and 1/3 respectively was used. Average error is reported across the 5 imputed datasets.

Model calibration

Internal model calibration was assessed using a calibration test based on a likelihood ratio statistic method described by Nattino et al.(24) designed to assess prediction tools based on dichotomous outcomes. When required the estimates produced by the classifiers were calibrated using isotonic regression and Platt scaling (25).

Sensitivity analyses

Analysis was repeated using mRS dichotomised into 0-2 as good outcome and 3-6 as poor outcome. This is the same dichotomisation used by Turner et al. in the study reporting CRP as an independent predictor of outcome in good grade but not poor grade aSAH patients(9).

Further sensitivity analyses were also completed using study site, alternative aneurysm size categories, follow up time as a covariate, and imputation rather than dummy variables for missing Fisher grade, hypertension and aneurysm size.

Statistical analysis

All analyses were conducted using R software (version 3.6.2, R Foundation for Statistical Computing).

**Results**

Patient inclusion, missing data and demographics

A total of 1017 patients were included in the study: 552 from the WNC neurovascular database and 465 from the STASH trial (Figure 1 details patient inclusion flow chart).

All patients had both WFNS and CRP data on at least one of day 0, 1 or 2. CRP data was missing on day 0, 1 and 2 in 760 (75%), 450 (44%) and 373 (37%) patients respectively. WFNS data was missing on day 0, 1 and 2 in 499 (49%), 332 (33%) and 281 (28%) patients respectively. Any missing CRP and WFNS data for day 0, 1 and 2 were imputed.

Table 1 details demographic data on patients included in the study and details other missing data. WFNS is presented on day 1 (see below).

CRP and WFNS

Supplementary Figure 1 shows histograms of day 0, 1 and 2 CRP values. Univariate logistic regression showed that CRP on all three days was significantly associated with outcome after aSAH (p< 0.002). The pooled AUCs for day 0, 1 and 2 CRP were 0.557 (0.472-0.638), 0.685 (0.630-0.734) and 0.709 (0.673-0.757) respectively. The AUCs for day 1 and 2 were both significantly improved compared to day 0 CRP (p< 0.001). There was no significant difference between the AUCs for day 1 and 2 CRP (p = 0.192). Day 1 CRP was therefore used for the remainder of the analysis. The R2 for day 1 CRP was 8%. Given the distribution of CRP seen in supplementary Figure 1, log CRP was also assessed as a predictor of outcome, but it did not improve predictive performance (p = 0.998) and was therefore not used in the analysis.

Univariate logistic regression showed that WFNS on day 0, 1 and 2 was significantly associated with outcome after aSAH (p< 0.001). The pooled AUCs for day 0, 1 and 2 WFNS were 0.690 (0.638-0.736), 0.763 (0.723-0.799) and 0.757 (0.710-0.799) respectively. The AUCs for day 1 and 2 were both significantly improved compared to day 0 WFNS (p< 0.003). There was no significant difference between the AUCs for day 1 and 2 WFNS (p = 0.48). Day 1 WFNS was therefore used for the remainder of the analysis. The R2 for day 1 WFNS was 15%, similar to that reported in other studies(4). Day 1 and 2 WFNS is likely to be post-resuscitation, compared to day 0, which is keeping with the finding that post-resuscitation WFNS is more predictive of outcome that pre-resuscitation(26).

If both CRP and WFNS are used as predictors in a multivariate logistic regression model they both remain significantly associated with outcome (CRP: p< 0.001; WFNS: p< 0.001), suggesting that the prognostic effect of CRP is independent of WFNS.

Predictive models of outcome

*Logistic regression model*

Inputting the same predictors as the full SAHIT predictive tool, hereby referred to as the “SAHIT model”, in this dataset using a logistic regression model generates a pooled AUC of 0.831 (0.797-0.860) (R2 = 24%), results comparable to those reported by Jaja *et al.*(4).

When CRP is added as a predictor to the full SAHIT model it remains an independent predictor of outcome (p< 0.001). The pooled AUC for this model is 0.846 (0.814-0.873) (R2 = 27%). The AUC is significantly improved compared to the SAHIT model alone in this dataset (p = 0.01). The NRI and IDI were calculated for the addition of CRP to the SAHIT model and demonstrated the addition of CRP significantly improved model performance (p = 0.03 and p< 0.001 respectively).

The general dominance of each individual predictor in the multivariable regression model is reported in Figure 2. WFNS, Fisher grade, treatment status and CRP were highlighted as the strongest predictors in the model.

*Random forest*

A random forest was trained using 500 trees on the SAHIT predictors and CRP and manually optimised to consider two predictors at each split of the tree. Out of bag error was 17% with a pooled model AUC of 0.807 (0.766-0.843). The AUC is significantly worse when compared to the regression model trained on the same data (p <0.001). A random forest was also trained on the SAHIT predictors and similarly optimised, achieving a pooled AUC of 0.782 (0.739-0.820) and performing significantly worse than the regression model trained on the SAHIT predictors (p <0.001).

CRP, age and WFNS were highlighted as the three most important variables to the predictive model as assessed by the mean decrease in Gini index; treatment status, WFNS and CRP were highlighted as the three most important variables to the predictive model as assessed by the mean decrease in accuracy (Figure 2).

*Support vector machine (SVM)*

A SVM was trained on the SAHIT predictors and CRP and optimised using 10-fold cross validation. The best performing model was achieved using a radial basis function kernel with model parameters: cost (related to model regularisation) of 4 and a γ (related to the radius of influence of the support vectors) of 2. This model achieved 1% misclassification with a pooled AUC of 0.960 (0.932-0.977), performing significantly better than both the regression model (p <0.001) and the random forest (p <0.001) trained on the same data.

WFNS, CRP and age were also highlighted as the three most important predictive variables in the SVM (Figure 2).

When an SVM is trained using the same predictors as the SAHIT model and optimised using 10-fold cross validation the best performing model was achieved using a radial basis function kernel with a cost of 4 and a γ of 2. The model achieved 5% misclassification with an AUC of 0.895 (0.850-0.927). This AUC was improved compared to the full SAHIT model (p = 0.004).

Table 2 and Supplementary Table I (please see https://www.ahajournals.org/journal/str) summarise the performance of the predictive models.

Model validation

All logistic regression models were internally validated using 10-fold cross validation. For all models the average classification error rate on the test data was 12% for the model using the same predictors as SAHIT and 11% for the model using CRP in addition. The SVM was also internally validated using 10-fold cross validation. The average classification error rate in the test data was 20% for the model using the same predictors as SAHIT and 19% for the model using CRP in addition. As described above, the random forest had an out of bag error rate of 17%.

Model calibration

Internal model calibration was assessed for the logistic regression and SVM models. The random forest was not assessed as it did not improve on the performance of the logistic regression.

Using the calibration test described by Nattino et al. the logistic regression models trained on the SAHIT predictors both with and without CRP did not demonstrate evidence of miscalibration (p = 0.237 and p = 0.413, respectively). Both the SVM trained on the SAHIT predictors alone and with CRP demonstrated evidence of miscalibration (p <0.001, Figure 3). In order to calibrate both these models, isotonic regression was applied to the estimates produced by the SVMs. The calibrated models were reassessed with the calibrated SVM trained on the SAHIT predictors alone showing no evidence of miscalibration (p = 0.859) and the SVM trained with the addition of CRP demonstrating improved calibration (Figure 3) but with ongoing evidence of miscalibration (p <0.001). Platt scaling failed to improve on isotonic regression (Figure 3).

The calibrated SVM without CRP achieved an AUC of 0.857 (0.824-0.890) and with CRP 0.920 (0.894-0.947).

Sensitivity analysis

None of the sensitivity analyses significantly altered the results of the models (please see supplemental material <https://www.ahajournals.org/journal/str> ).

**Discussion**

A number of models have been developed to predict outcome after aSAH(3). The SAHIT predictive tool is based on the largest population and has been rigorously validated(4). It is therefore the most widely accepted. However, it only explains up to 31% of the variation in outcome, meaning other predictors not in the model may play a significant role in outcome prediction. In this study we aimed to assess whether CRP is an additional independent predictor of outcome after aSAH and whether incorporating it into a range of outcome prediction models improved performance.

We have shown in a cohort of 1017 patients that day 1 CRP is an independent predictor of outcome controlling for all other known key predictors.

Figure 4 displays the effect on probability of poor outcome for each predictor included in the multivariable logistic regression models. In terms of relative importance, this study shows that CRP is a key predictor of outcome after aSAH with an R2 = 8%. In the regression analysis controlling for collinearity it is the fourth most important predictor as assessed by general dominance after WFNS, Fisher grade and treatment status (see Figure 2). It should be noted that in this nomenclature, all the effect of treatment status was due to poor performance of patients not treated, and that CRP had a much greater influence on outcome than clipping or coiling, which had almost identical outcomes (Figure 4). Both the SVM and Random Forest models identified WFNS, CRP and age as most important. This emphasises the importance of including CRP in future prediction models.

However, logistic regression models including CRP only achieved an R2 of 27%. This suggests that despite the inclusion of CRP there still remain other unknown but significant predictors of outcome after aSAH. Further studies need to be designed to look for these unknown predictors if we are to improve prognostication. These studies require unbiased analysis of a wide range of predictors including the consideration of predictors beyond the standard demographic, clinical and radiological variables, such as proteomic and genetic data.

We have gone on to consider how CRP can best be incorporated into prediction models and demonstrate that incorporation of day 1 CRP to traditional multivariate logistic regression models significantly improves predictive performance above that of the full SAHIT predictive tool.

Random forests did not improve upon logistic regression. This may be inherent to our study design. Random forests are well suited to categorical classification problems with multiple classes and may perform better using Rankin class rather than dichotimisation into good and poor outcome.

The SVM was superior to logistic regression with a marked improvement in discrimination (AUC=0.960). However, it should be noted that the SVM models had the greatest error on internal validation suggesting overfitting. External validation is necessary to validate the improvements achieved. However, given machine learning methods are of greatest performance benefit in large datasets with high dimensional data and future research is likely to involve larger datasets with more predictors, this observation is relevant and these methods are likely to become increasingly important(10, 11).

None of the logistic regression models showed evidence of miscalibration, but the SVM models did. This could be improved with isotonic regression and Platt scaling. Following calibration, the SVM trained on the SAHIT predictors alone was well calibrated, but despite clear improvements, the SVM using CRP in addition still showed some evidence of miscalibration (Figure 3). Calibration reduced the AUC of both models but both still markedly outperformed the logistic regression models trained on the same predictors. The fact the model cannot be fully calibrated suggests evidence of overfitting and future studies should ensure SVM models are calibrated on external datasets to minimise this.

Generalisability

In the development of these models, we considered if simplified models, not requiring intensive processing, could be derived. Details of these are available in the online supplement. They were unable to capture the improvements seen from utilising CRP and SVMs. Therefore, use of any of the models in this manuscript would require use of an online calculator.

In the logistic regression models the addition of CRP to the SAHIT predictors improved outcome prediction (pooled AUC 0.846 compared to 0.831), although statistically significant this is unlikely to influence clinical management on an individual patient basis. Implementation of an online calculator is a significant barrier to adoption and it is notable that even SAHIT itself is mainly used as a research tool, and only routinely used clinically in a minority of institutions. Therefore, given the improvements from addition of the SAHIT predictors to WFNS are larger than that afforded by CRP this seems unlikely to be widely adopted outside of research.

In order to generate predictive tools, which influence clinical management, they will need to show much greater benefit over WFNS alone. If the magnitude of improvement seen with SVMs can be externally validated, an online calculator with the addition of SVMs as well as CRP may be able to achieve this.

This analysis was limited to patients within 72 hours of ictus and cannot be applied to patients outside of this window. We have shown that CRP (and WFNS) is less predictive on day 0 than either day 1 or 2. It might be expected that CRP becomes confounded by other intervening treatments and complications over time. We have not observed any differences in CRP between coiled and clipped patients on day 1 in this study suggesting this is too early for treatment to be significant. We have however seen such differences at later timepoints, and that the predictive value of CRP drops after day 2 (data not presented). This may explain why previous studies which included patients whose CRP was obtained within 120 hours of ictus have demonstrated that CRP is an independent predictor of outcome in good grade, but not poor grade, aSAH patients(9). This distinction is important, as it is poor grade patients where prognostication is of particular importance when guiding potentially invasive and costly treatment strategies, and it would not be possible to use CRP in prognostic models if these only applied to a subgroup of patients.

Limitations

In this study all patients had CRP and WFNS available on at least one of day 0, 1 or 2. Day 1 CRP and WFNS was used in the analysis and was missing in 450 (44%) and 332 (33%) patients respectively. Data was missing at random for both day 1 CRP and WFNS with respect to outcome (CRP: p = 0.134 (Kruskal Wallis test); WFNS: p = 0.665 (chi-squared test)). All patients with missing data had day 1 CRP/WFNS imputed using the CRP/WFNS data available from the day before or after (or both). Multiple imputation has been used with pooled AUCs reported. Consequently, although there is missing CRP and WFNS data, robust methodology has been used to account for this. It was not possible to report a DeLong’s test result to compare pooled AUCs from different predictive models and we therefore have reported the result for the first imputed dataset.

A limitation of this study is that the predictive models have not been validated in an external dataset. The models have, however, been internally validated using cross validation and out of bag error. In previous outcome prediction models after aSAH using smaller datasets a validation set approach dividing data into train and test data sets has commonly been presented. For smaller data sets this method is superior to cross validation to avoid overfitting of data, however as sample sizes increase the utility of cross validation improves(27). A sensitivity analysis using a validation set approach with data divided into train and test datasets with proportions 70% and 30% showed error rates consistent with those identified by cross validation.

Like many studies of outcome in aSAH, this study is limited by its outcome measure, mRS, which was not developed for aSAH and may not be sufficiently sensitive to identify the nuances of aSAH. An SAH specific outcome score has been developed but is awaiting external validation(28). The use of scores like this may improve prognostication in this unique condition. This study also uses dichotomised outcome data and future efforts using the full ordinal nature of the mRS may also improve models.

Outcome was recorded at 3-6 months in this study in keeping with SAHIT(4). This range means that these predictive models do not predict outcome at a specific timepoint. In order to achieve this, larger datasets with specific outcome times would be required. However, the sensitivity analysis including time of follow up as a covariate showed no change in the significance of the results and model performance was not affected.

Some covariates, such as hypertension and aneurysm size, were not available in the STASH data set, which may have influenced the prediction models. This is a limitation of using retrospective data, but may be addressed in external validation and future prospective studies.

Finally, several radiological scales have been shown to have greater predictive value for outcome than the Fisher scale including the modified Fisher Score, Claassen score, Barrow Neurological Institute score and Hijdra score. Unfortunately, these were not available in this dataset. This has historically been the case for many datasets including SAHIT. However, future studies should use newer scores with greater prognostic value.

**Conclusions**

This study demonstrates that day 1 CRP is an independent predictor of outcome after aSAH. Its incorporation into predictive models of outcome after aSAH improves model performance over that seen using the SAHIT predictors alone. The machine learning method SVM generates further significant improvement in prognostic modelling, but requires validation and calibration in an external cohort.

**Sources of Funding:** The STASH study was funded by the British Heart Foundation (SP/08/003/24065).

**Disclosures:** none

**Supplemental Materials:**

Expanded methods, results and discussion  
Online: Tables I-V, Figures I-III

Reference (29)

References

1. Stegmayr B, Eriksson M, Asplund K. Declining mortality from subarachnoid hemorrhage: changes in incidence and case fatality from 1985 through 2000. Stroke. 2004;35(9):2059-63.

2. Al-Khindi T, Macdonald RL, Schweizer TA. Cognitive and functional outcome after aneurysmal subarachnoid hemorrhage. Stroke. 2010;41(8):e519-36.

3. Jaja BN, Cusimano MD, Etminan N, Hanggi D, Hasan D, Ilodigwe D, et al. Clinical prediction models for aneurysmal subarachnoid hemorrhage: a systematic review. Neurocrit Care. 2013;18(1):143-53.

4. Jaja BNR, Saposnik G, Lingsma HF, Macdonald E, Thorpe KE, Mamdani M, et al. Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: the SAHIT multinational cohort study. BMJ. 2018;360:j5745.

5. Höllig A, Stoffel-Wagner B, Clusmann H, Veldeman M, Schubert GA, Coburn M. Time Courses of Inflammatory Markers after Aneurysmal Subarachnoid Hemorrhage and Their Possible Relevance for Future Studies. Front Neurol. 2017;8:694.

6. Fountas KN, Tasiou A, Kapsalaki EZ, Paterakis KN, Grigorian AA, Lee GP, et al. Serum and cerebrospinal fluid C-reactive protein levels as predictors of vasospasm in aneurysmal subarachnoid hemorrhage. Clinical article. Neurosurg Focus. 2009;26(5):E22.

7. Romero FR, Cataneo DC, Cataneo AJ. C-reactive protein and vasospasm after aneurysmal subarachnoid hemorrhage. Acta Cir Bras. 2014;29(5):340-5.

8. Juvela S, Kuhmonen J, Siironen J. C-reactive protein as predictor for poor outcome after aneurysmal subarachnoid haemorrhage. Acta Neurochir (Wien). 2012;154(3):397-404.

9. Turner CL, Budohoski K, Smith C, Hutchinson PJ, Kirkpatrick PJ, Murray GD, et al. Elevated Baseline C-Reactive Protein as a Predictor of Outcome After Aneurysmal Subarachnoid Hemorrhage: Data From the Simvastatin in Aneurysmal Subarachnoid Hemorrhage (STASH) Trial. Neurosurgery. 2015;77(5):786-92; discussion 92-3.

10. Bzdok D, Krzywinski M, Altman N. Points of Significance: Machine learning: a primer. Nat Methods. 2017;14(12):1119-20.

11. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. Nat Methods. 2018;15(4):233-4.

12. Dumont TM. Prospective Assessment of a Symptomatic Cerebral Vasospasm Predictive Neural Network Model. World Neurosurg. 2016;94:126-30.

13. Ramos LA, van der Steen WE, Sales Barros R, Majoie CBLM, van den Berg R, Verbaan D, et al. Machine learning improves prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage. J Neurointerv Surg. 2019;11(5):497-502.

14. Hostettler IC, Muroi C, Richter JK, Schmid J, Neidert MC, Seule M, et al. Decision tree analysis in subarachnoid hemorrhage: prediction of outcome parameters during the course of aneurysmal subarachnoid hemorrhage using decision tree analysis. J Neurosurg. 2018;129(6):1499-510.

15. Liu J, Xiong Y, Zhong M, Yang Y, Guo X, Tan X, et al. Predicting Long-Term Outcomes After Poor-Grade Aneurysmal Subarachnoid Hemorrhage Using Decision Tree Modeling. Neurosurgery. 2020.

16. de Toledo P, Rios PM, Ledezma A, Sanchis A, Alen JF, Lagares A. Predicting the outcome of patients with subarachnoid hemorrhage using machine learning techniques. IEEE Trans Inf Technol Biomed. 2009;13(5):794-801.

17. Shi Z, Hu B, Schoepf UJ, Savage RH, Dargis DM, Pan CW, et al. Artificial Intelligence in the Management of Intracranial Aneurysms: Current Status and Future Perspectives. AJNR Am J Neuroradiol. 2020;41(3):373-9.

18. Kirkpatrick PJ, Turner CL, Smith C, Hutchinson PJ, Murray GD, Collaborators S. Simvastatin in aneurysmal subarachnoid haemorrhage (STASH): a multicentre randomised phase 3 trial. Lancet Neurol. 2014;13(7):666-75.

19. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. 2011. 2011;45(3):67.

20. DB R. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley and Sons; 2004.

21. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. Biometrics. 1988;44(3):837-45.

22. Zhang D. A Coefficient of Determination for Generalized Linear Models. The American Statistician. 2017;71(4):310-6.

23. Kraha A, Turner H, Nimon K, Zientek LR, Henson RK. Tools to support interpreting multiple regression in the face of multicollinearity. Front Psychol. 2012;3:44.

24. Nattino G, Finazzi S, Bertolini G. A new calibration test and a reappraisal of the calibration belt for the assessment of prediction models based on dichotomous outcomes. Stat Med. 2014;33(14):2390-407.

25. Huang Y, Li W, Macheret F, Gabriel RA, Ohno-Machado L. A tutorial on calibration measurements and calibration models for clinical prediction models. J Am Med Inform Assoc. 2020;27(4):621-33.

26. van Donkelaar CE, Bakker NA, Veeger NJ, Uyttenboogaart M, Metzemaekers JD, Eshghi O, et al. Prediction of outcome after subarachnoid hemorrhage: timing of clinical assessment. J Neurosurg. 2017;126(1):52-9.

27. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One. 2019;14(11):e0224365.

28. Pace A, Mitchell S, Casselden E, Zolnourian A, Glazier J, Foulkes L, et al. A subarachnoid haemorrhage-specific outcome tool. Brain. 2018.

29. Zhang Z, Zhang H, Khanal MK. Development of scoring system for risk stratification in clinical medicine: a step-by-step tutorial. Ann Transl Med. 2017;5(21):436.

**Figure legends**

Figure 1. Flow chart for patients included in study.

Diagram

Description automatically generated

Figure 2. Relative importance of predictors for the first imputed dataset. A: mean decrease in accuracy (random forest). B: mean decrease in Gini index (random forest). C: coefficients of support vectors (SVM). D: General dominance of predictors reported as R2 statistic (logistic regression)

Chart

Description automatically generated

Figure 3. Calibration belt plots for SVMs trained on SAHIT predictors before and after calibration (A & B) and trained on SAHIT predictors + CRP before and after calibration (C, D &E). B & D calibrated with isotonic regression (IR) and E with Platt scaling (PS).

Diagram, engineering drawing

Description automatically generated

Figure 4. Effect plots for predictors included in multivariable logistic regression model for the first imputed dataset.

Diagram, engineering drawing

Description automatically generated

Table 1 – Demographic data from patients included in study. WFNS reported for the first imputed dataset.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Wessex Neurological Centre | STASH | Total |
| Total (n) | 552 | 465 | 1017 |
| Age | 56 (21-79) | 50 (20-69) | 54 (20-79) |
| Sex | M: 160  F: 392 | M: 157  F: 308 | M: 318  F: 700 |
| WFNS (day 1)  I  II  III  IV  V | 236  83  36  93  104 | 185  122  36  70  52 | 421  205  72  163  156 |
| Fisher  1  2  3  4  Missing | 8  31  127  191  195 | 5  62  162  235  1 | 13  93  289  426  196 |
| Treatment status  Clip  Coil  No treatment | 131  405  16 | 168  287  10 | 299  692  26 |
| Location  Middle cerebral artery  Anterior cerebral artery  Posterior circulation  Internal cerebral artery | 101  240  77  134 | 109  134  187  35 | 210  374  264  169 |
| Hypertension  Yes  No  Missing | 141  411  0 | 0  0  465 | 141  411  465 |
| Size  ≤ 12 mm  13-24 mm  > 24 mm  Missing | 413  19  0  120 | 0  0  0  465 | 413  19  0  585 |
| Outcome  Good  Poor | 451  101 | 389  76 | 840  177 |

Table 2 – Performance of predictive models comparing SAHIT predictors and SAHIT + CRP. AUCs reported with 95% confidence intervals.

|  |  |  |
| --- | --- | --- |
|  | AUC for SAHIT predictors | AUC for SAHIT predictors + CRP |
| Logistic regression | 0.831 (0.797-0.860) | 0.846 (0.814-0.873) |
| Random forest | 0.782 (0.739-0.820) | 0.807 (0.766-0.843) |
| Support vector machine | 0.895 (0.850-0.927) | 0.960 (0.932-0.977) |

SUPPLEMENTAL MATERIAL

**C-reactive protein in outcome prediction after subarachnoid haemorrhage and the role of machine learning.**

*Gaastra, B; Barron, P; Newitt, L; Chhugani, S; Turner, C; Kirkpatrick, P; MacArthur, B; Galea, I; Bulters, D*

Supplemental methods

*Simplified predictive tools*

Advanced statistical learning methods such as SVMs are challenging for clinicians to interpret and require inputting a range of data usually into an online tool. This is a barrier to the use of prognostic tools at the bedside to aid clinical decision making and brings into question the clinical use of advanced predictive modelling outside the research setting. We therefore explored generating simple predictive tools that could be easily applied at the bedside and, if validated, aid decision making in clinical practice.

In order to generate simplified predictive tools we applied two methods using the key features identified as most relevant in the predictive models:

1. Classification decision trees
2. A scoring system based on the methods described by Zhang et al.(29) where continuous variables are converted to categorical levels, using a Locally Weighted Scatterplot Smoothing (LOESS) function.

The first imputed dataset was used for this analysis. Treatment status was not included in these simplified models as one of the aims of a bedside predictive model is to help guide management including the decision whether to treat or not. Simplified predictive tool performance was compared to that of WFNS alone, rather than the SAHIT prediction tool, as WFNS is the best individual predictor of outcome and most easily applied at the bedside

Supplemental results

Three SAHIT models were developed by Jaja et al. (4): core, neuroimaging and full. The core model includes patient age, WFNS grade and past medical history of hypertension. The neuroimaging model consist of the core model plus Fisher grade and size and location of ruptured aneurysm. The AUCs for the core SAHIT and neuroimaging model in this dataset are 0.775 (0.739-0.812) and 0.815 (0.782-0.848) respectively.

In addition to the logistic regression models trained on the SAHIT predictors and the SAHIT predictors plus CRP further logistic regression models were generated incorporating polynomial and interaction terms. The specific non-linear and interaction terms were chosen using an intelligence-based method based on clinical experience and review of the data. The additional models were trained using all predictors, considering age and CRP as third order polynomials in view of their potentially non-linear relationship with outcome and allowing for interaction of WFNS and CRP. This interaction was considered as good and poor grade SAH may have a different relationship with CRP.

The model trained using CRP and age as third order polynomials achieved an AUC of 0.850 (0.817-0.877). Considering interaction terms alone the model achieved an AUC of 0.846 (0.814-0.873). The final additional model incorporated both polynomial and interaction terms and achieved an AUC 0.850 (0.817-0.878) (R2 = 29%). CRP remained an independent predictor of outcome (p < 0.001). This more complex model did not perform significantly better than the regression model developed on the same predictors without interaction or polynomial terms (p = 0.16). See Table III for summary of additional models.

*Simplified predictive tools*

In order to generate simple predictive tools that do not require inputting of data by the user into analytical software, and could theoretically, once validated be used as clinical bedside predictive tools, two methods were employed. Treatment status was not included so that these models could be developed to aid management decisions, such as decision to treat.

*Decision tree*

A classification tree was trained using all available variables and manually optimised, including exclusion of the predictors hypertension and size. The minimum number of observations that must exist at each node and the minimum number of observations in any terminal ranged between 1-2 and the complexity parameter ranged between 0.008-0.01. This decision tree achieved an AUC of 0.766 (0.727-0.806). This tree had a depth of 8 and was felt no more practical to implement than an online tool. The tree was pruned based on variable importance to limit to a maximum depth of 2, as trees deeper than this are felt to be impractical to apply at the bedside. Only two variables WFNS (1-3 vs 4-5) and CRP (<42, ≥45) remained subsequent to this pruning and this tree performed with an AUC of 0.747 (0.708-0.786) (Figure II). In comparison to WFNS alone this model did not improve predictive performance (p = 0.15).

*Scoring system*

A scoring system predictive model was developed based on the methods described by Zhang *et al*.(29). The predictors age, CRP and WFNS were included in the scoring model. These three predictors were selected as they featured at least twice as the top three most important predictors in the random forest, SVM and logistic regression models (Figure 3).

The continuous predictors age and CRP were converted to categorical levels using a LOESS smoothening function with the cut offs ascertained visually at points of inflexion on the graph (Figure III). This methodological approach using a LOESS smoothing function to convert continuous predictors to categorical levels has not been applied in aSAH before and offers significant benefits by allowing the cut offs for levels to be assigned more objectively. Scores were assigned to the categorical variables WFNS and treatment status and the newly generated age and CRP levels based on the coefficients of a logistic regression model rounded to the nearest integer. Table IV presents the scores assigned. This resulted in a scoring system ranging from 0-4 points.

The newly developed score was used as the input in a univariate logistic regression model to predict outcome and performed with an AUC of 0.772 (0.735-0.810) (R2 = 17%). This model was internally validated with 10-fold cross validation with an average classification error rate of 12%. The predicted probability of poor outcome for each score is presented in Table V. In comparison to WFNS alone this model did not improve predictive performance (p = 0.44).

Supplemental discussion

In order to develop clinically applicable, easy to apply and interpret prognostic models both decision trees and scoring systems have previously been generated. We considered two approaches for developing similar models incorporating CRP. The decision tree presented is especially simple but did not offer any significant benefit over WFNS alone. The scoring system likewise did not offer any significant benefit over WFNS alone. It therefore remains that WFNS is likely to remain the mainstay of bedside prognostication.

Supplemental sensitivity analyses

All sensitivity analyses were performed in the first imputed dataset.

All analyses in the manuscript were repeated with outcome dichotomised mRS 0-2 as good and mRS 3-6 as bad. There were no changes to the significance of the reported results. Of note CRP remained an independent predictor of outcome for both good and poor grade SAH patients when added to the SAHIT predictors.

Study site was not significant as a predictor of outcome in any of the predictive models, as may be expected given the data distribution was similar in both datasets. The dummy variable for missing data was also not a significant predictor of outcome in any of the models.

When follow up time was included as a covariate in the logistic regression models it was not a significant predictor of outcome and there were no changes to the significance of the reported results.

The use of dummy variables to allow for inclusion of missing data in the analyses has limitations and we have consequently imputed the remaining missing data as a further sensitivity analysis. Missing Fisher grade, hypertension and aneurysm size data was imputed. All of these predictors were missing at random with respect to outcome (Fisher grade: p = 0.239 (Kruskal Wallis test); hypertension: p = 0.462 (chi-squared test); aneurysm size: p = 0.198 (chi-squared test)). There were no changes to the significance of reported results.

As the majority of aneurysms (96%) within this dataset fall into the ≤12mm size category a sensitivity analysis dividing this category into two further groups (0-10mm and 11-12mm) was performed. This cut off was chosen in our dataset using a LOESS smoothening function, as described above, to identify the point where aneurysmal size was most likely to influence outcome after aSAH. The use of this additional size category made no change to the significance of the results.

Patients included in this study were treated between 2007 and 2017. In order to control for changing healthcare over this decade a sensitivity analysis including year of treatment as a covariate was performed. Due to data availability this was only performed in the STASH cohort. Year of treatment was not a significant predictor of outcome and there were no changes to the significance of the reported results.

Finally, log CRP was also used in all analyses to replace CRP and did not alter the significance of the results.

Supplemental tables

|  |  |
| --- | --- |
| Method | AUC (95% confidence intervals) |
| Logistic regression (see table II for details of models)  WFNS alone  SAHIT full model  *Predictors: age, WFNS, hypertension, Fisher, size & location of aneurysm, treatment status*  SAHIT full model + CRP  *Predictors: age, WFNS, hypertension, Fisher, size & location of aneurysm, treatment status, CRP* | 0.763 (0.723-0.799)  0.831 (0.797-0.860)  0.846 (0.814-0.873) |
| Random forest  SAHIT full model  *Predictors: age, WFNS, sex, hypertension, Fisher, size & location of aneurysm, treatment status, CRP*  SAHIT full model + CRP  *Predictors: age, WFNS, hypertension, Fisher, size & location of aneurysm, treatment status, CRP* | 0.782 (0.739-0.820)  0.807 (0.766-0.843) |
| Support vector machine  SAHIT full model  *Predictors: age, WFNS, sex, hypertension, Fisher, size & location of aneurysm, treatment status, CRP*  SAHIT full model + CRP  *Predictors: age, WFNS, hypertension, Fisher, size & location of aneurysm, treatment status, CRP* | 0.895 (0.850-0.927)  0.960 (0.932-0.977) |
| Simple predictive tools  Decision tree  *Predictors: age, WFNS, sex, Fisher, size & location of aneurysm*  Pruned decision tree  *Predictors: WFNS, CRP*  Scoring system  *Predictors: WFNS, CRP, age* | 0.766 (0.727-0.806)  0.747 (0.708-0.786)  0.772 (0.735-0.810) |

Table I. Performance of predictive models

|  |  |  |  |
| --- | --- | --- | --- |
| Intercept and Predictors | β coefficient | Odds ratio | 95% CI |
| *Logistic regression model: WFNS alone* | | | |
| Intercept | -2.95 |  |  |
| WFNS grade  1  2  3  4  5 | Reference  0.88  1.23  2.33  2.61 | 2.41  3.43  10.24  13.61 | 1.30-4.46  1.58-7.48  5.94-17.65  7.91-23.4 |
| *Logistic regression model: SAHIT full model* | | | |
| Intercept | -4.22 |  |  |
| Hypertension  Missing  Yes  No | Reference  0.16  -0.01 | 1.18  0.99 | 0.57-2.45  0.53-1.85 |
| Age | 0.02 | 1.02 | 1.00-1.04 |
| WFNS grade  1  2  3  4  5 | Reference  0.91  1.05  2.24  2.39 | 2.49  2.85  9.40  10.87 | 1.28-4.84  1.19-6.82  5.17-17.08  5.87-20.13 |
| Fisher  Missing  1  2  3  4 | Reference  -13.84  -0.73  -0.18  0.94 | 0  0.48  0.83  2.56 | 0  0.14-1.71  0.41-1.70  1.38-4.75 |
| Size  Missing  ≤ 12mm  12-24mm | Reference  -0.16  1.02 | 0.85  2.78 | 0.47-1.53  0.76-10.22 |
| Location  MCA  ACA  Posterior  ICA | Reference  -0.22  -0.26  -0.57 | 0.80  0.77  0.56 | 0.48-1.34  0.44-1.36  0.29-1.09 |
| Treatment  Clip  Coil  Conservative | Reference  -0.09  3.45 | 0.91  31.42 | 0.59-1.41  8.85-111.57 |
| *Logistic regression model: SAHIT full model + CRP* | | | |
| Intercept | -4.60 |  |  |
| Hypertension  Missing  Yes  No | Reference  0.17  0.09 | 1.19  1.09 | 0.57-2.47  0.58-2.07 |
| Age | 0.02 | 1.02 | 1.00-1.04 |
| WFNS grade  1  2  3  4  5 | Reference  0.87  0.95  2.02  2.13 | 2.4  2.57  7.51  8.41 | 1.23-4.68  1.05-6.29  4.08-13.82  4.49-15.74 |
| Fisher  Missing  1  2  3  4 | Reference  -13.61  -0.60  -0.04  1.03 | 0  0.55  0.96  2.82 | 0  0.15-1.97  0.46-2.00  1.50-5.33 |
| Size  Missing  ≤ 12mm  12-24mm | Reference  -0.18  0.99 | 0.84  2.69 | 0.46-1.51  0.73-9.93 |
| Location  MCA  ACA  Posterior  ICA | Reference  -0.21  -0.30  -0.57 | 0.81  0.74  0.57 | 0.48-1.37  0.42-1.33  0.29-1.11 |
| Treatment  Clip  Coil  Conservative | Reference  -0.08  3.59 | 0.92  36.32 | 0.59-1.44  9.98-132.16 |
| CRP | 0.01 | 1.01 | 1.01-1.02 |

Table II. Details on logistic regression models including intercept. The predicted probability of poor compared to good outcome is P(poor outcome) = 1/(1 + exp(-(intercept + β1X1 + β2X2 + … + βi­Xi))), where βi­ is the regression coefficient for predictor Xi. Predictor value is one when present and zero when absent apart from age and CRP which are considered as continuous variables.

|  |  |
| --- | --- |
| Logistic regression model | AUC (95% confidence intervals) |
| Polynomial terms  *Predictors: age, WFNS, sex, hypertension, Fisher, size & location of aneurysm, treatment status, CRP*  *Polynomial (third order): age, CRP*  Interaction terms  *Predictors: age, WFNS, sex, hypertension, Fisher, size & location of aneurysm, treatment status, CRP*  *Interaction: CRP & WFNS*  Polynomial and interaction terms  *Predictors: age, WFNS, sex, hypertension, Fisher, size & location of aneurysm, treatment status, CRP*  *Polynomial (third order): age, CRP*  *Interaction: CRP & WFNS* | 0.850 (0.817-0.877)  0.846 (0.8135-0.873)  0.850 (0.817-0.878) |

Table III. Performance of additional logistic regression models.

|  |  |
| --- | --- |
| Variable | Score |
| WFNS  1-3  4  5 | 0  1  2 |
| CRP  < 42  ⩾ 42 | 0  1 |
| Age  < 71  ⩾ 71 | 0  1 |

Table IV. Scores applied to each variable. The total score is generated by summing all variables.

|  |  |
| --- | --- |
| **Score** | **Predicted probability of poor outcome** |
| 0 | 0.07 |
| 1 | 0.17 |
| 2 | 0.36 |
| 3 | 0.60 |
| 4 | 0.80 |

Table V. Predicted probability of poor outcome for each score.

Supplemental figures

Chart, histogram

Description automatically generated

Figure I. Histograms showing CRP on day 0 (A), day 1 (B) and day 2 (C) for the first imputed dataset.

Diagram

Description automatically generated

Figure II. Pruned decision tree. Probability of outcome is reported.

Chart

Description automatically generated

Figure III. LOESS curves for CRP and age. Cut offs, visually assigned at inflexions of the graph, for scoring system represented by black lines.

A picture containing table

Description automatically generated