Pooled analysis suggests benefit of catheter-based hematoma removal for intracerebral hemorrhage

Pitchaiah Mandava, MD, PhD, MSEE, Santosh B. Murthy, MD, MPH, Neel Shah, MD, Yves Samson, MD, Marek Kimmel, PhD, and Thomas A. Kent, MD

Neurology® 2019;92:e1688-e1697. doi:10.1212/WNL.0000000000007269

Correspondence

Dr. Mandava
pmandava@bcm.edu

RELATED ARTICLE

Editorial

Winners, losers, and controls

Page 689

MORE ONLINE

CME Course

NPub.org/cmelist

Abstract

Objective

To develop models of outcome for intracerebral hemorrhage (ICH) to identify promising and futile interventions based on their early phase results without need for correction for baseline imbalances.

Methods

We developed a pooled outcome model from the control arms of randomized control trials and tested different interventions against the model at comparable baseline conditions. Eligible clinical trials and large case series were identified from multiple library databases. Models based on baseline factors reported in the control arms were tested for the ability to predict functional outcome (modified Rankin Scale score) and mortality. Interventions were grouped into blood pressure control, fibrinolytic-assisted hematoma evacuation, hemostatic medications, and neuroprotective agents. Statistical intervals around the model were generated at the $p = 0.1$ level to screen how each trial's outcome compared to expected outcome.

Results

Fourteen control arms with 3,386 patients were used to develop 7 alternate models for functional outcome. The model incorporating baseline NIH Stroke Scale, age, and hematoma volume yielded the best fit (adjusted $R^2 = 0.89$). All early phase treatments that eventually resulted in negative late phase trials were identified as negative by this method. Early phase fibrinolytic-assisted hematoma evacuation studies showed the most promise trending toward improved functional outcome with no suggestion of an increase in mortality, supporting its further study.

Conclusions

We successfully developed an outcome model for ICH that identified interventions destined to be negative while identifying a promising one. Such an approach may assist in prioritizing resources prior to multicenter trial.

From the Michael E. DeBakey VA Medical Center Stroke Program (P.M., N.S.) and Analytical Software and Engineering Research Laboratory, Department of Neurology (P.M., N.S., T.A.K.), Baylor College of Medicine, Houston, TX; Department of Neurology (S.B.M.) and Clinical and Translational Neuroscience Unit (S.B.M.), Feil Family Brain and Mind Research Institute, Weill Cornell Medicine, New York, NY; APHP (Y.S.), Urgences Cerebro-Vasculaire, Pitié-Salpêtrière, and UPMC Paris Sorbonne Universités (Y.S.), Paris, France; Departments of Statistics and Bioengineering (M.K.) and Chemistry (T.A.K.), Rice University, Houston; Institute of Biosciences and Technology (IBT) (T.A.K.), Texas A&M Health Science Center–Houston Campus; and Department of Neurology (T.A.K.), Houston Methodist Hospital and Research Institute, TX.

Go to Neurology.org/N for full disclosures. Funding information and disclosures deemed relevant by the authors, if any, are provided at the end of the article.

Glossary

AIC = Akaike Information Criterion; **ATACH** = Antihypertensive Treatment of Acute Cerebral Hemorrhage; **BP** = blood pressure; **CI** = confidence interval; **ENOS** = Efficacy of Nitric Oxide in Stroke; **FAST** = Recombinant Factor VIIa in Acute Intracerebral Haemorrhage; **GCS** = Glasgow Coma Scale; **ICH** = intracerebral hemorrhage; **INTERACT** = Intensive Blood Pressure Reduction in Acute Cerebral Hemorrhage Trial; **IVH** = intraventricular hemorrhage; **MISTIE** = Minimally Invasive Surgery plus rt-PA for Intracerebral Hemorrhage Evacuation; **mRS** = modified Rankin Scale; **NIHSS** = NIH Stroke Scale; **PATCH** = Platelet Transfusion versus Standard Care After Acute Stroke due to Spontaneous Cerebral Haemorrhage Associated With Antiplatelet Therapy; **RCT** = randomized clinical trial; **rFVIIa** = recombinant factor VIIa; **SSS** = Scandinavian Stroke Scale.

Intracerebral hemorrhage (ICH) is a potentially devastating disease with no proven treatment,¹ despite focused translational² and clinical research.³ A number of therapeutic strategies have been advanced and tested in both early and late phase clinical trials, the latter thus far without success.^{4–17} We recently reviewed the failures in phase 3 randomized clinical trials (RCTs) in ischemic stroke and identified that many of these trials were destined to fail because their early phase counterparts were erroneously interpreted as promising.¹⁸ These false-positive results were largely due to imbalances in important baseline factors and noise generated by errors in subjective measures.¹⁹ We previously demonstrated that a complex, nonlinear relationship between baseline factors and outcome violates assumptions necessary to consider these statistical correction methods valid.¹⁸

These issues highlight the difficulties in identifying a clear pathway from early phase to successful late phase trials. While no method can be expected to be perfectly predictive, we found that comparison to a large pooled sample of trial controls at their own baseline conditions made it possible to identify the likely outcome of phase 3 trials from the results of earlier trials in ischemic stroke.^{18,20} As part of this approach, we generate statistical intervals around this outcome model to test whether a treatment's outcomes exceed the variance expected in a large population.¹⁸ This approach requires no additional statistical correction for imbalances since trials are compared to the pooled model at their own baseline parameters.

In this study, we applied a similar approach to ICH by pooling outcome results from the control arms of ICH trials and comparing different baseline variables appropriate for ICH for their predictive value, generating an outcome model in the process. Subsequently, we superimposed the results of ICH trials and case series against these models at their own reported baseline conditions to compare their outcomes to a pooled control population.

Methods

We first built models to predict mortality and functional outcomes, using the control arms of all available RCTs in ICH identified using prespecified criteria. The treatment arms of

different therapeutic interventions (RCTs and phase I/II single arm studies that did not have control arms) were then superimposed on these models to assess benefit or harm and to answer the question of whether to proceed to phase III. Details of model development are outlined below.

Control arm model development

Literature search

A literature search was performed by 2 authors (P.M. and S.B.M.) using key words “intracerebral hemorrhage” or “intracerebral haemorrhage.” The search methodology is outlined in detail in figure e-1 (doi.org/10.5061/dryad.vj2dj82). Filters of “Clinical trial” (all phases), “Human,” and “English” were also applied. To further identify control arms used to generate the model, the following selection criteria were applied to studies published up to March 1, 2018: (1) baseline initial severity of deficit, either NIH Stroke Scale (NIHSS) or Scandinavian Stroke Scale (SSS), was provided in terms of median or mean of the group; SSS was converted to NIHSS using a published conversion factor²¹; (2) additional baseline factors were provided, such as baseline hematoma volume and age; (3) 30–180 days mortality or functional outcome was provided.

Model development from control arms of RCTs

Candidate models were developed for each outcome using a combination of clinical and radiologic parameters known to influence ICH outcome such as age, NIHSS, Glasgow Coma Scale (GCS), baseline hematoma volume, and presence of intraventricular hemorrhage (IVH). The most common functional outcome measure reported was modified Rankin Scale (mRS). Favorable outcome in most ICH trials was defined as proportion achieving mRS 0–3. For trials that considered mRS 0–2 as the only measure of favorable outcome, we developed separate models for this measure.

Not all trials presented all variables of interest. The predictive value of baseline variables or combination of variables cannot be compared across models if there are dissimilar groups with different number of trials in each model. To address this issue, the favorable outcome (mRS 0–3) and mortality models were subdivided in 3 groups based on baseline variables provided by the trials to identify the predictive variable or set of variables that best determine outcome in each group. Trials were

divided into 3 nested groups with the requirement that all trials in a group have all variables under consideration. Group 1 included all trials that reported baseline NIHSS, age, and hematoma volumes; group 2 had trials that reported NIHSS, GCS, age, and hematoma volume; and group 3 comprised trials that reported NIHSS, age, hematoma volume, and IVH percentage.

The process of model development is detailed in earlier publications,^{22,23} summarized here along with changes in minimization routine. Outcomes reported as proportions for the control arms of the RCTs were transformed by means of Freeman-Tukey transformation.^{24,25}

$$\frac{1}{2} \left[\arcsin \sqrt{\frac{X}{n+1}} + \arcsin \sqrt{\frac{X+1}{n+1}} \right]$$

In the equation above, n is the number of patients in the control arm of an RCT and X/n is the proportion of patients with an outcome (mRS 0–3 or mortality). The Freeman-Tukey transformation stabilizes the variance and is suggested when there is wide variation in n^{25} as it has been suggested superior to arcsin-sqrt transformation.²⁶ The variance of each transformed proportion was shown by Freeman and Tukey²⁴ to be $\frac{821}{n+0.5}$ and the reciprocal of this variance was used as the weight in a weighted least squares method to determine the best fit model for each outcome.

Depending on the number of predictors in each group, several candidate models are possible. For example, group 1 had 7 models for mRS 0–3 and mortality accounting for all 3 predictor variables of NIHSS, age, and hematoma volume, in various combinations ($2^3 - 1$). Of the several candidate models, optimal models were chosen based on the lowest Akaike Information Criterion (AIC), an information theoretic approach first suggested by Akaike²⁷ and exhaustively characterized by Burnham and Anderson²⁸ (section 5, doi.org/10.5061/dryad.vj2dj82).

AIC²⁷ is a measure of the deviation of a model function from the data that generated the model.²⁸ If the ratio of number of observations (n) that generated the model to the number of estimated parameters of the model (k) used is below 40 then a corrected AIC_c is calculated by adding a penalty term (data available from section 5: doi.org/10.5061/dryad.vj2dj82 and equation below).

$$AIC_c = AIC + \frac{2k(k+1)}{n-k-1}$$

The model with the smallest AIC_c value is considered the best candidate model and ranked 1. To see how other models perform in comparison to the model ranked 1, a difference value Δ_i is calculated by subtracting the AIC_c of a candidate model from the AIC_c of the one ranked 1. Burnham and Anderson²⁸ suggested that models with $\Delta_i \leq 2$ of the model ranked 1 have strong support to be possible alternate models

and those with values between 4 and 7 have less support and those beyond 10 have no support to be alternate models.

Adjusted R^2 , which takes into account the number of variables that were considered for each model, is reported. Models were developed in R-language with figures drawn in Matlab.

Sampling and simulation

To identify the influence of number of trials on model selection, 14 trials that presented mRS 0–3 were sampled randomly to generate subsets from 7 to 13 trials, and models generated and AIC_c calculated. This sampling of all subsets of trials was simulated 10,000 times and the AIC_c of the alternate models averaged and ranked. The lower bound of a population of 7 trials is necessary since for any value below 7, the denominator ($n-k-1$) in equation 2 will be zero or negative, resulting in an implausible negative or infinity penalty term in AIC_c.

Comparison of different therapeutic modalities for ICH

Literature search to identify various therapeutic regimens for ICH

A separate literature search was performed (P.M. and S.B.M.) to identify all phases of clinical trials and case series that dealt with treatment of acute ICH. Using search inclusion and exclusion search strategies detailed above, 4 treatment regimens were selected for analysis here: (1) antihypertensive agents, (2) clot aspiration augmented by fibrinolysis, (3) hemostatic agents, and (4) neuroprotective agents.

Comparison of the treatment arms against the pooled outcome model

Treatment arms were visualized in 2 ways. Three-dimensional figures demonstrating outcome were employed when using 2 baseline variables. The model function is shown as a surface after back transformation of derived model from angles to proportions.²⁵ Prediction intervals corresponding to $p = 0.1$ bounding the model function surface²⁹ were selected for this demonstration as a screening threshold for success or harm. To compare treatment arms against the models, the individual studies are visually inspected with respect to their location relative to the upper and lower $p < 0.1$ prediction interval bounds to assess whether their outcomes exceeded or are less than the $p = 0.1$ variance threshold in this pooled population. The viewing angles of the figures are rotated for each treatment group to better visualize the studies and their different baseline factors.

In the case of 3 baseline variables, 4-dimensional figures cannot be visualized. Here, we show point estimates and confidence intervals (CIs) of those trials that had early and late phase follow-ups to assess whether the early trials were predictive of future outcomes for antihypertensive, catheter-based interventions, and hemostatic agent. The CIs were generated using the Agresti-Coull method.³⁰ To test if a particular model among 2 alternate models better explains

a dataset, the mean square error of the 2 models was compared using F test.³¹

Data availability

All supplementary data are available at doi.org/10.5061/dryad.vj2dj82. Further anonymized data can be made available to qualified investigators upon request to the corresponding author.

Results

Twenty-seven control arms of RCTs from 1999 to 2017 were identified that met criteria and therefore used to specify the model. The baseline characteristics (NIHSS, GCS, age, hematoma volume, percentage of patients with IVH) and long-term outcomes (mortality and mRS 0–3) are listed in table e-1 (doi.org/10.5061/dryad.vj2dj82). Age was reported in all trials. In addition to age, baseline factors reported in decreasing frequency were NIHSS, GCS, hematoma volume, and percentage of IVH. Volume of IVH was reported in very few trials ($n = 4$).

Functional independence (mRS 0–3) model

Fourteen trials (group 1) reported baseline NIHSS, age, and hematoma volume and long-term mRS 0–3 proportions. These 14 control arms with 3,386 patients had a mean NIHSS of 15.5, mean age of 65.7, and mean hematoma volume was 21.6 mL. Models were developed for 7 combinations of the baseline factors and ranked according their AIC_c (table e-2; doi.org/10.5061/dryad.vj2dj82). The mRS 0–3 model function generated from NIHSS, age, and hematoma volume had the lowest AIC_c and ranked the best model (adjusted $R^2 = 0.89$; table e-2). The mRS 0–3 model based on NIHSS and hematoma volume shown in figure 1 (adjusted $R^2 = 0.81$) was ranked the second best with Δ_i of 3.35. The other 5 models had $\Delta_i > 4$ and thus considerably less support to be alternate models.

From 11 trials (group 2) that reported GCS in addition to NIHSS, age, and hematoma volume, 15 models were developed. The models based on NIHSS and age (adjusted $R^2 = 0.76$), NIHSS and hematoma volume (adjusted $R^2 = 0.75$), and NIHSS, age, and hematoma volume (adjusted $R^2 = 0.85$) were ranked 1–3 and Δ_i within 2 of each other. Models based on GCS alone and combined with other variables crossed the Δ_i threshold of 4, which translates to considerably less support to be acceptable alternate models.

From 10 trials (group 3) that reported percentage IVH in addition to NIHSS, age, and hematoma volume, 15 models for mRS 0–3 were developed. The model based on NIHSS and age was ranked the best model in this group (adjusted $R^2 = 0.87$). The model with percentage IVH alone had a low rank compared to other models (rank = 11/15, adjusted $R^2 = 0.11$) (table e-1; doi.org/10.5061/dryad.vj2dj82).

Models based on NIHSS and hematoma volume for groups 1–3 were superimposed as shown in figure e-2 (doi.org/10.5061/dryad.vj2dj82).

Figure 1 Modified Rankin Scale (mRS) 0–3 model function based on baseline NIH Stroke Scale (NIHSS) and hematoma volume of control arms of 14 randomized clinical trials

Middle surface is the model function (adjusted $R^2 = 0.81$). Surfaces on either side represent $\pm p = 0.1$ prediction interval surfaces. Color bar indicates the percentage ($\times 100$) achieving an mRS score 0–3.

5061/dryad.vj2dj82). The 3 surfaces (red: group 1, green: group 2, and black: group 3) are indistinguishable and confirmed by F tests (groups 1 and 2: $p = 0.49$ and group 1 and 3: $p = 0.50$), supporting that the models are not different in statistical terms.

Sampling and simulation

When the number of RCTs is below 9, only the model based on NIHSS alone can be supported and all other models including those based on 2 variables at a time or all 3 variables of NIHSS, age, and volume are not contenders (section 2 and table e-3; doi.org/10.5061/dryad.vj2dj82). Models based on all 3 variables require at least 13 RCTs.

Functional mRS 0–2 model

A functional outcome model for mRS 0–2 was developed (adjusted $R^2 = 0.82$). This model was restricted to testing early phase treatment trials (Antihypertensive Treatment of Acute Cerebral Hemorrhage [ATACH]¹⁵ and Intensive Blood Pressure Reduction in Acute Cerebral Hemorrhage Trial [INTERACT]¹³) that only reported mRS 0–2.

Comparison of 4 different treatment regimens against the mRS 0–3 model

We tested antihypertensive agents, fibrinolytic-assisted approaches, hemostatic agents, and neuroprotective agents against the models. In figures 2–5, uppercase letters identify a treatment arm plotted at the baseline NIHSS and hematoma volume (table e-4, A–D; doi.org/10.5061/dryad.vj2dj82 for details of trials tested and letter assignments). The non-intervention arm of an RCT or prospective cohort study was

Figure 2 Effect of treatment with antihypertensives compared with model of modified Rankin Scale (mRS) 0–3

Treatment arms (upper case) and control arm (lower case) plotted onto the surfaces. Trials specifically discussed are in black letters and trials tested against the models but not discussed are in red letters (see table e-2A, doi.org/10.5061/dryad.vj2dj82, for details of trials tested and letter assignments). Intensive Blood Pressure Reduction in Acute Cerebral Hemorrhage Trial (INTERACT)-2: E and e; Antihypertensive Treatment of Acute Cerebral Hemorrhage (ATACH)-II: F and f; Efficacy of Nitric Oxide in Stroke-intracerebral hemorrhage (ENOS-ICH) 6 hours: I and i. Confidence intervals for the treatment arms are not shown to avoid overlapping confidence intervals. See figure 6 for confidence intervals of selected trials. Color bar indicates the percentage (×100) achieving an mRS score 0–3. NIHSS = NIH Stroke Scale.

assigned a lower-case letter. Missing letters indicate that the trials did not report mRS 0–3. Trials discussed are highlighted by large black font and other trials tested against the models but not specifically discussed are in red. Selected treatment arms above or below the statistical interval are also boxed for aid in visualization.

Blood pressure (BP) intervention

Several antihypertensive medications in different time windows and different BP ranges were tested in RCTs, single arm series, and early phase trials with no control arms (figure 2 and table e-4A; doi.org/10.5061/dryad.vj2dj82). A multinational early phase INTERACT¹³ followed by a later phase INTERACT-2¹⁴ (figure 2, treatment arm: E; control arm: e) tested several antihypertensive agents with the choice of agents left to the investigator. INTERACT-2¹⁴ mRS 0–3 outcomes were neutral by our analysis as results are very close to the predicted model surface. We speculate that a suggestion of benefit in the original interpretation of INTERACT-2¹⁴ was likely due to baseline imbalances since the treatment arm had a 1-point lower NIHSS than the control arm, which could explain 2.4% of the 3.6% reported difference in outcome.¹⁴ Nicardipine was tested in an early phase trial, ATACH,¹⁵ and a late phase trial, ATACH-II,¹⁶ halted for futility (treatment arm: F; control arm: f). Again, note that there is no treatment benefit with respect to the

model. Efficacy of Nitric Oxide in Stroke (ENOS)-ICH¹⁷ tested glyceryl trinitrate, an antihypertensive, and mRS 0–3 outcome of the treatment arm of the 6-hour subgroup (I) is well within the ±90% intervals indicating no potential benefit. The control arm of the ENOS-ICH 6-hour group (i) had functional outcome below the –90% prediction interval, suggesting that the outcomes in this arm are not representative of the broad group of control arms that generated the model.

Clot aspiration augmented by thrombolytic use trials

Four thrombolytic-augmented aspiration trials provided^{32–35} NIHSS, age, and hematoma volume along with mRS 0–3 outcomes (figure 3 and table e-4B; doi.org/10.5061/dryad.vj2dj82). Both the early phase trial of urokinase-assisted aspiration (boxed K in figure 3) and phase 2 Minimally Invasive Surgery plus rt-PA for Intracerebral Hemorrhage Evacuation (MISTIE)-II³⁴ intervention arm (boxed M) had mRS 0–3 proportions above the +90% surface, suggesting that this is a promising strategy. Using a slightly different albeit larger catheter-based study, the Intraoperative Stereotactic Computed Tomography-Guided Endoscopic Surgery (ICES)³⁵ treatment arm (L) is also above +90% surface for mRS 0–3 outcome. Stereotactic Treatment of Intracerebral Hematoma by Means of a Plasminogen Activator (SICHPA), a study that used stereotaxic evacuation³³ (N), lies on the +90% interval.

Figure 3 Clot aspiration treatment compared with model of modified Rankin Scale (mRS) 0–3

Treatment arms (upper case) and control arm (lower case) plotted onto the surfaces. All fibrinolytic augmented surgical trials including Minimally Invasive Surgery plus rt-PA for Intracerebral Hemorrhage Evacuation (MISTIE)-II and an early phase urokinase-augmented aspiration trial (UK-ASP) showed better than expected outcomes. Letter assignments: UK-ASP: K and k; Intraoperative Stereotactic Computed Tomography-Guided Endoscopic Surgery (ICES): L; MISTIE-II: M and m; Stereotactic Treatment of Intracerebral Hematoma by Means of a Plasminogen Activator (SICHPA): N and n. See confidence intervals of selected trials in figure 6. Color bar indicates the percentage (×100) achieving an mRS score 0–3. NIHSS = NIH Stroke Scale.

Trials of hemostatic agents

An early phase trial⁴ tested 3 doses of recombinant factor VIIa (rFVIIa: O, P, Q) against a control arm (o) as a precursor to the Recombinant Factor VIIa in Acute Intracerebral Haemorrhage (FAST) trial.⁵ All 3 treatment arms of the early phase trial⁴ had lower NIHSS than the control arm, thus favoring better outcome in the treatment arms, and indeed this trial was suggested to be positive in the original interpretation (figure 4, table e-4C; doi.org/10.5061/dryad.vj2dj82). However, there was no positive signal for the treatment arms when compared to the pooled model in our analysis, as all are on the middle surface or within the $\pm 90\%$ intervals. The follow-on FAST trial tested low-dose and high-dose regimens of rFVIIa⁴ (R and S) against a control arm (r). The imbalances in terms of NIHSS did not persist in the larger FAST⁵ and no benefit was seen in either model as shown (compare R, S, and r).

In the Platelet Transfusion versus Standard Care After Acute Stroke due to Spontaneous Cerebral Haemorrhage Associated With Antiplatelet Therapy (PATCH) trial,⁷ the functional outcome (mRS 0–3) was much below the -90% prediction surface consistent with harm (figure 4: treatment arm U). The control arm of the PATCH trial was below the -90% interval, suggesting that the control arm is not representative of the other control arms. PATCH trial was interpreted as negative in the original publication,⁷ comparable to our results.

Figure 4 Effect of hemostatic agents

Letter assignments: early phase trial with recombinant factor VIIa (rFVIIa): O, P, Q, o; Recombinant Factor VIIa in Acute Intracerebral Haemorrhage (FAST): R, S, r; Tranexamic Acid for Hyperacute Primary Intracerebral Haemorrhage (TICH): T and t; Platelet Transfusion versus Standard Care After Acute Stroke due to Spontaneous Cerebral Haemorrhage Associated With Antiplatelet Therapy (PATCH): U and u; The Spot Sign for Predicting and Treating ICH Growth Study (STOP-IT): V and v. PATCH treatment arm outcome was worse than expected (boxed U). Control arms of PATCH (u) and early phase trial of rFVIIa (o) were not representative of other control arm outcomes. Color bar indicates the percentage ($\times 100$) achieving a modified Rankin Scale (mRS) score 0–3. NIHSS = NIH Stroke Scale.

Neuroprotective agents

Clinical trials of citicoline,¹⁰ NXY-059,¹¹ and deferoxamine⁸ were tested against the model (figure 5 and table e-4D; doi.org/10.5061/dryad.vj2dj82). There were baseline imbalances between the citicoline¹⁰ trial arms (figure 5, treatment arm W and control arm w) favoring better outcomes in treatment arm with lower baseline severity. When compared to our model, no benefit was seen. Deferoxamine⁸ was tested (X) as a single arm treatment trial showed no benefit. Cerebral Hemorrhage and NXY-059 Treatment (CHANT)¹¹ treatment arm (NXY-059: Z) showed harm compared to the model. Overall, no neuroprotectant trial analyzed here showed promise.

Comparing early and late phase trials against the 3 variable (NIHSS, age, and hematoma volume) models

One of the goals of our method is to assess whether it provides a useful signal to predict benefit or harm in an early phase. We selected trials that either followed up on an early phase result (e.g., BP intervention) or is still in the early phase with a larger phase trial underway (e.g., the MISTIE series) using the model based on 3 predictor variables of NIHSS, age, and hematoma volume. Since it is difficult to visualize a 4D model (outcome and 3 baseline variables), we show point estimates and CIs for control arm outcome (black, figure 6) at the same treatment arm baseline NIHSS, age, and hematoma volume. The reported outcome of a treatment arm and the 90% CI are shown in red. While we did not detect a true separation (intervals not overlapping), some trends were evident.

ATACH-A¹⁵ (tier 3: aggressive BP management of 110–140), ATACH-B¹⁵ (tier 2: BP target of 140–170), and

Figure 5 Neuroprotectant treatments compared against the modified Rankin Scale (mRS) 0–3 model

Letter assignments: citicoline: W, w; deferoxamine-US: X; NXY-059: Z, z. Color bar indicates the percentage ($\times 100$) achieving an mRS score 0–3. NIHSS = NIH Stroke Scale.

Figure 6 Point estimate comparison of control and treatment outcomes for trials providing baseline NIH Stroke Scale (NIHSS), age, and hematoma volume

INTERACT¹³ only presented mRS 0–2 outcomes, while the follow-on trials ATACH-II¹⁶ and INTERACT-2¹⁴ presented the complete mRS distribution. Therefore, ATACH-A¹⁵ and ATACH-B¹⁵ and INTERACT¹³ were compared against our mRS 0–2 model (* in figure 6). There was no suggestion of positive results for either early phase trial, consistent with negative later phase outcomes. Early phase trial utilizing rFVIIa⁴ did not show any benefit with respect to mRS 0–3 and would have suggested that undertaking the larger FAST trial⁵ would be negative, as indeed was the case.

The point estimate for the intervention arm of urokinase-augmented aspiration³² (n = 9) suggested benefit, albeit with large overlapping CIs. The recently completed MISTIE-II³⁴ (n = 52) outcome also appears near the high end of predicted outcomes although again with overlapping error bars and less impressive difference than with the 2-variable model (compare with letter M in figure 3). Late phase MISTIE-III is completed and results pending.

Mortality

Twenty trials reported NIHSS, age, and hematoma volume. These 20 control arms had 3,784 patients, with the mean NIHSS of the control arms being 14.1, mean age 65.1, and

mean hematoma volume 20.1 mL. Models for mortality were developed for different combinations of the baseline factors for groups 1, 2, and 3 defined in the Methods. All the models in each group were ranked according to their AIC_c (summarized here and in table e-5; doi.org/10.5061/dryad.vj2dj82). In all 3 groups, models based on NIHSS and hematoma volume were ranked the best in terms of AIC_c (group 1 adjusted R² = 0.67; group 2 adjusted R² = 0.78; group 3 adjusted R² = 0.69). Details of the various models of mortality and the result of comparisons of the 4 treatment regimens against the model are detailed in data available from Dryad (section 2). The vast majority of treatment regimens fell within the $\pm 90\%$ intervals (figures e-3–e-6). The exceptions were the 2 treatment arms from early phase ATACH¹⁵ and the PATCH⁷ trial. These 3 trial arms had mortality above the +90% interval, suggesting potential harm because of greater than expected mortality.

Discussion

The main goal of our approach is to generate models with predictive value for study outcomes in ICH based on baseline factors and to test whether we could identify potentially

beneficial treatments at an early phase. This effort is meant to help investigators in early phase trials to test outcomes against a model generated from the pooled outcomes of control arms of a larger and potentially more representative of a general population likely to be encountered with the transition to phase 3. The models generated here differ from individual patient-based models that generally have the intent of prognosticating outcomes or for selection into an appropriate treatment modality.^{36–40} Gregson et al.⁴¹ employed meta-analysis derived from individual patient-level data to answer a specific question whether interventional surgery for ICH is superior to nonsurgical treatment. The overall goal of this and our method is similar: to detect a treatment effect in published trials, ours differs in providing a broader control population, although addition of representative individual datasets could add to its flexibility.

We tested several interventions that had completed early and late phase trials, and found that all those trials proved to be negative in phase 3 were identified as negative at their early phase by this method. It is noteworthy that even small differences in baseline factors accounted for differences in outcome that could confound even large phase 3 trials. Given that there are no established efficacious therapies tested in a phase 3 trial, we cannot estimate the positive predictive value of our method, but the consistent appearance of fibrinolytic-augmented aspiration therapies at the upper end of beneficial threshold without increasing mortality was in contrast to the consistent appearance of other interventions near or worse than their control outcomes. Although recognizing that our screening criteria for potential efficacy was set at a conservative significance level, our results agree with the performance of the recently concluded MISTIE-III trial whose results are pending, and the ongoing Early Minimally Invasive Removal of Intracerebral Hemorrhage (ENRICH) trial. Note also that our method assumes the later phase trial is performed identical to the early phase, which is often not the case as inclusion or exclusion criteria and other factors are sometimes changed to accommodate a larger population sample.

Our review of the ICH literature to generate our models revealed several findings that have implications for the outcome variance we encountered. For example, trials did not use a consistent/common scale for measurement of baseline severity, employing NIHSS, SSS, and GCS, requiring using a published conversion factor to convert from SSS to NIHSS. Converting scales likely introduces noise in the measure. Also, we are not aware of a conversion factor to translate GCS to NIHSS. While there is a high correlation (Spearman $\rho = 0.87$) reported between GCS and NIHSS, there is no validated conversion factor reported.⁴²

Because we are testing the potential therapeutic effect on the outcome of a trial population rather than individual patient outcomes, we employed mean or median measures of baseline factors to generate the model and compare to

treatment interventions as one might do with the results of their early phase trials. These measures of central tendency may not represent the true distribution of baseline measures especially for non-Gaussian distributions. Availability of the individual subject data that go into the measures could provide a check on whether the distributions are skewed.

The number of factors available to generate the model is limited by the lack of consistency of reporting. For example, not all studies employed the same baseline severity index. Here we compared treatment arms against models that incorporated NIHSS and volume even when the treatment arms presented both NIHSS and GCS in addition to volume. We used the models based on NIHSS and not GCS since we showed that models of functional outcome incorporating NIHSS are more predictive vis-à-vis GCS based models. Therefore, the trials performed with GCS^{43–45} as the only measure of severity of ICH cannot be tested against the model.

The time window for recruitment also varied considerably, from under 3 hours–72 hours, which may influence outcomes such as with respect to hematoma expansion. Follow-up time periods were also not consistent (30/90/180 days). While in a study pertaining to non-posterior-fossa ischemic strokes, Duncan et al.⁴⁶ showed there is plateau effect in functional recovery at 90 days, the time frame of the plateau effect is unknown for ICH.

Another noteworthy limitation of our study is lack of availability of data on IVH volumes. Presence of IVH, when available, was used as a variable in our current trial-based models, but these models ranked rather low in the ranking (mRS 0–3: 12/15 in table e-3; mortality: 12/15 in table e-5; doi.org/10.5061/dryad.vj2dj82). Reporting IVH volume, measured using techniques such as the modified Graeb score for example,⁴⁷ would likely have helped study the influence of IVH better; but lack of availability of IVH volume, reported in only 4 trials, precluded further exploration. Other factors specific to ICH such as active bleeding (e.g., spot sign) may influence the effect of therapy. However, a recent secondary analysis of ATACH-II trial incorporating the spot sign did not show any benefit of considering this factor.⁴⁸

While the influence of IVH and active bleeding is somewhat subsumed under the contribution to initial stroke severity, having these and more variables to consider could improve the sensitivity of the model. However, simply increasing the number of variables, while attractive, is not advisable without good justification, since a penalty is assessed in model selection both in the setting of AIC and the closely related Bayesian Information Criterion.²⁸ This penalty increases as the ratio of number of predictive variables to number of data points increases.²⁴ If reporting of ICH trials becomes more consistent, additional clinical or neuroimaging factors can be considered for the model.

While our method may not be able to detect small treatment effects in ICH until reporting is more consistent, in the interim, comparing treatment results to a pooled sample may be especially useful to detect approaches that are likely to be futile. While our results remain speculative without confirmatory follow-up studies, we nevertheless tentatively identified fibrinolytic-augmented aspiration therapies as the most promising of current approaches we were able to analyze. We suggest that considering a pooled outcome model could be useful for prioritizing resources to pursue therapies for ICH with implications for other conditions in which baseline factors influence outcome.

Note added in proof: The MISTIE III trial was reported following acceptance of the final version of this manuscript. While the primary outcome modified intention to treat group did not achieve significant improvement, the per protocol treatment group (in whom residual blood achieved the 15ml threshold) did demonstrate improvement (absolute risk reduction: 10.9%, $p = 0.02$).⁴⁹

Author contributions

P. Mandava: study concept, design, acquisition and analysis of data, drafting of the manuscript including critical revisions for intellectual content, study supervision, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. S. Murthy: study design, acquisition of data, drafting of the manuscript including critical revisions. N. Shah: critical revision of the manuscript for intellectual content. Y. Samson: critical revision of the manuscript for intellectual content. M. Kimmel: statistical advice, drafting of the manuscript including critical revisions. T. Kent: study concept, design, acquisition and analysis of data, drafting of the manuscript including critical revisions for intellectual content, study supervision.

Acknowledgment

Drs. Nikola Sprigg and Julio José Secades Ruiz provided additional data regarding the Tranexamic acid in Intracerebral Hemorrhage (TICH) and the Citicoline trials. Dr. Barbara Gregson confirmed nonavailability of NIHSS for STICH-I and STICH-II trials.

Study funding

No targeted funding reported.

Disclosure

P. Mandava jointly holds the copyright for the analytical methodology used in the study: pPREDICTS. S. Murthy is supported by the National Institutes of Health through the grant K23NS105948 and the Leon Levy Foundation. N. Shah, Y. Samson, and M. Kimmel report no disclosures relevant to the manuscript. T. Kent jointly holds the copyright for the analytical methodology used in the study: pPREDICTS. Supported in part by R01 NS094535 and grant no. BE-0048 from the Welch Foundation. Go to Neurology.org/N for full disclosures.

Publication history

Received by *Neurology* March 19, 2018. Accepted in final form December 6, 2018.

References

- Qureshi AI, Tuhrim S, Broderick JP, Batjer HH, Hondo H, Hanley DF. Spontaneous intracerebral hemorrhage. *New Engl J Med* 2001;344:1450–1460.
- Chang CF, Cai L, Wang J. Translational intracerebral hemorrhage: a need for transparent descriptions of fresh tissue sampling and preclinical model quality. *Transl Stroke Res* 2015;6:384–389.
- Gonzales NR. Ongoing clinical trials in intracerebral hemorrhage. *Stroke* 2013;44(6 suppl 1):S70–S73.
- Mayer SA, Brun NC, Begtrup K, et al. Recombinant activated factor VII for acute intracerebral hemorrhage. *New Engl J Med* 2005;352:777–785.
- Mayer SA, Brun NC, Begtrup K, et al. Efficacy and safety of recombinant activated factor VII for acute intracerebral hemorrhage. *New Engl J Med* 2008;358:2127–2137.
- Sprigg N, Renton CJ, Dineen RA, Kwong Y, Bath PM. Tranexamic acid for spontaneous intracerebral hemorrhage: a randomized controlled pilot trial (ISRCTN50867461). *J Stroke Cerebrovasc Dis* 2014;23:1312–1318.
- Baharoglu MI, Cordonnier C, Al-Shahi Salman R, et al. Platelet transfusion versus standard care after acute stroke due to spontaneous cerebral haemorrhage associated with antiplatelet therapy (PATCH): a randomised, open-label, phase 3 trial. *Lancet* 2016;387:2605–2613.
- Selim M, Yeatts S, Goldstein JN, et al. Safety and tolerability of deferoxamine mesylate in patients with acute intracerebral hemorrhage. *Stroke* 2011;42:3067–3074.
- Haley EC, Jr, Thompson JL, Levin B, et al. Gavestinel does not improve outcome after acute intracerebral hemorrhage: an analysis from the GAIN International and GAIN Americas studies. *Stroke* 2005;36:1006–1010.
- Secades JJ, Alvarez-Sabin J, Rubio F, Lozano R, Dávalos A, Castillo J. Citicoline in intracerebral haemorrhage: a double-blind, randomized, placebo-controlled, multicentre pilot study. *Cerebrovasc Dis* 2006;21:380–385.
- Lyden PD, Shuaib A, Lees KR, et al. Safety and tolerability of NXY-059 for acute intracerebral hemorrhage: the CHANT Trial. *Stroke* 2007;38:2262–2269.
- Lee SH, Park HK, Ryu WS, et al. Effects of celecoxib on hematoma and edema volumes in primary intracerebral hemorrhage: a multicenter randomized controlled trial. *Eur J Neurol* 2013;20:1161–1169.
- Anderson CS, Huang Y, Wang JG, et al. Intensive blood pressure reduction in acute cerebral haemorrhage trial (INTERACT): a randomised pilot trial. *Lancet Neurol* 2008;7:391–399.
- Anderson CS, Heeley E, Huang Y, et al. Rapid blood-pressure lowering in patients with acute intracerebral hemorrhage. *New Engl J Med* 2013;368:2355–2365.
- Antihypertensive Treatment of Acute Cerebral Hemorrhage Investigators. Antihypertensive treatment of acute cerebral hemorrhage. *Crit Care Med* 2010;38:637–648.
- Qureshi AI, Palesch YY, Barsan WG, et al. Intensive blood-pressure lowering in patients with acute cerebral hemorrhage. *New Engl J Med* 2016;375:1033–1043.
- Krishnan K, Scutt P, Woodhouse L, et al. Glyceryl trinitrate for acute intracerebral hemorrhage: results from the Efficacy of Nitric Oxide in Stroke (ENOS) trial, a subgroup analysis. *Stroke* 2016;47:44–52.
- Kent TA, Shah SD, Mandava P. Improving early clinical trial phase identification of promising therapeutics *Neurology* 2015;85:274–283.
- Mandava P, Krumpelman CS, Shah JN, White DL, Kent TA. Quantification of errors in ordinal outcome scales using shannon entropy: effect on sample size calculations. *PLoS One* 2013;8:e67754.
- Mandava P, Kent TA. Intra-arterial therapies for acute ischemic stroke. *Neurology* 2007;68:2132–2139.
- Gray LJ, Ali M, Lyden PD, Bath PM. Virtual International Stroke Trials Archive Collaboration. Interconversion of the National Institutes of Health Stroke Scale and Scandinavian Stroke Scale in acute stroke. *J Stroke Cerebrovasc Dis* 2009;18:466–468.
- Mandava P, Kent TA. A method to determine stroke trial success using multidimensional pooled control functions. *Stroke* 2009;40:1803–1810.
- Mandava P, Shah SD, Sarma AK, Kent TA. An outcome model for intravenous rt-PA in acute ischemic stroke. *Transl Stroke Res* 2015;6:451–457.
- Freeman MF, Tukey JW. Transformations related to the angular and the square root. *Ann Math Stat* 1950;21:607–611.
- Zar J. *Biostatistical Analysis*. 3rd ed. Upper Saddle River, NJ: Prentice-Hall; 1996: 282–283.
- Uchino K, Billheimer D, Cramer SC. Entry criteria and baseline characteristics predict outcome in acute stroke trials. *Stroke* 2001;32:909–916.
- Akaike H. Prediction and entropy. In: Atkinson AC, ed. *A Celebration of Statistics*. New York: Springer-Verlag; 1985:1–24.
- Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Approach Information-Theoretic Approach*. New York: Springer; 2002.
- Process Modeling. NIST/SEMATECH e-Handbook of Statistical Methods. 2012. Available at: itl.nist.gov/div898/handbook/tool aids/pdf/pmd.pdf. Accessed March 30, 2014.
- Agresti A, Coull BA. Approximate is better than “exact” for interval estimation of binomial proportions. *Am Stat* 1988;52:119–126.
- Using the F-test to compare two models: non-linear regression. *Stat Topic Files*. 2013. Available at: sakai.duke.edu/access/content/group/25e08a3d-9fc4-41b0-a7e9-815732c14ba/New%20folder/Stat%20Topic%20Files/Non-Linear%20Regression/FTestTutorial.pdf. Accessed March 30, 2014.

32. Zuccarello M, Brott T, Derex L, et al. Early surgical treatment for supratentorial intracerebral hemorrhage: a randomized feasibility study. *Stroke* 1999;30:1833–1839.
33. Teernstra OP, Evers SM, Lodder J, et al. Stereotactic treatment of intracerebral hematoma by means of a plasminogen activator: a multicenter randomized controlled trial (SICHPA). *Stroke* 2003;34:968–974.
34. Hanley DF, Thompson RE, Muschelli J, et al. Safety and efficacy of minimally invasive surgery plus alteplase in intracerebral haemorrhage evacuation (MISTIE): a randomized, controlled, open-label, phase 2 trial. *Lancet Neurol* 2016;15:1228–1237.
35. Vespa P, Hanley D, Betz J, et al. ICES (intraoperative stereotactic computed tomography-guided endoscopic surgery) for brain hemorrhage: a multicenter randomized controlled trial. *Stroke* 2016;47:2749–2755.
36. Ariesen MJ, Algra A, van der Worp HB, Rinkel GJ. Applicability and relevance of models that predict short term outcome after intracerebral haemorrhage. *J Neurol Neurosurg Psychiatry* 2005;76:839–844.
37. Satopaa J, Mustanoja S, Meretoja A, et al. Comparison of all 19 published prognostic scores for intracerebral hemorrhage. *J Neurol Sci* 2017;379:103–108.
38. Weimar C, Benemann J, Diener HC; German Stroke Study Collaboration. Development and validation of the Essen Intracerebral Haemorrhage Score. *J Neurol Neurosurg Psychiatry* 2006;77:601–605.
39. Hemphill JC III, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke* 2001;32:891–897.
40. Rost NS, Smith EE, Chang Y, et al. Prediction of functional outcome in patients with primary intracerebral hemorrhage: the FUNC score. *Stroke* 2008;39:2304–2309.
41. Gregson BA, Broderick JP, Auer LM, et al. Individual patient data subgroup meta-analysis of surgery for spontaneous supratentorial intracerebral hemorrhage. *Stroke* 2012;43:1496–1504.
42. Prasad K, Dash D, Kumar A. Validation of the Hindi version of National Institute of Health Stroke Scale. *Neurol India* 2012;60:40–44.
43. Morgenstern LB, Frankowski RF, Shedden P, Pasteur W, Grotta JC. Surgical treatment for intracerebral hemorrhage (STICH): a single-center, randomized clinical trial. *Neurology* 1998;51:1359–1363.
44. Mendelow AD, Gregson BA, Fernandes HM, et al. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial intracerebral haematomas in the International Surgical Trial in Intracerebral Haemorrhage (STICH): a randomised trial. *Lancet* 2005;365:387–397.
45. Mendelow AD, Gregson BA, Rowan EN, et al. Early surgery versus initial conservative treatment in patients with spontaneous supratentorial lobar intracerebral haematomas (STICH II): a randomised trial. *Lancet* 2013;382:397–408.
46. Duncan PW, Goldstein LB, Matchar D, Divine GW, Feussner J. Measurement of motor recovery after stroke: outcome assessment and sample size requirements. *Stroke* 1992;23:1084–1089.
47. Morgan TC, Dawson J, Spengler D, et al. The Modified Graeb Score: an enhanced tool for intraventricular hemorrhage measurement and prediction of functional outcome. *Stroke* 2013;44:635–641.
48. Morotti A, Brouwers HB, Romero JM, et al. Intensive blood pressure reduction and spot sign in intracerebral hemorrhage: a secondary analysis of a randomized clinical trial. *JAMA Neurol* 2017;74:950–960.
49. Hanley DF, Thompson RE, Rosenblum, et al. Efficacy and safety of minimally invasive surgery with thrombolysis in intracerebral haemorrhage evacuation (MISTIE III): a randomized, controlled, open-label blinded endpoint phase 3 trial.