Treatment of Intracerebral Hemorrhage in Animal Models: Meta-Analysis

Joseph Frantzias, BSc, Emily S. Sena, PhD, Malcolm R. Macleod, PhD, and Rustam Al-Shahi Salman, MA, PhD

Objective: Interventions that improve functional outcome after acute intracerebral hemorrhage (ICH) in animals might benefit humans. Therefore, we systematically reviewed the literature to find studies of nonsurgical treatments tested in animal models of ICH.

Methods: In July 2009 we searched Ovid Medline (from 1950), Embase (from 1980), and ISI Web of Knowledge (from 1969) for controlled animal studies of nonsurgical interventions given after the induction of ICH that reported neurobehavioral outcome. We assessed study quality and performed meta-analysis using a weighted mean difference random effects model.

Results: Of 13,343 publications, 88 controlled studies described the effects of 64 different medical interventions (given a median of 2 hours after ICH induction) on 38 different neurobehavioral scales in 2,616 treated or control animals (median 14 rodents per study). Twenty-seven (31%) studies randomized treatment allocation, and 7 (8%) reported allocation concealment; these studies had significantly smaller effect sizes than those without these attributes (p < 0.001). Of 64 interventions stem cells, calcium channel blockers, anti-inflammatory drugs, iron chelators, and estrogens improved both structural outcomes and neurobehavioral scores in >1 study. Meta-regression revealed that together, structural outcome and the intervention used accounted for 65% of the observed heterogeneity in neurobehavioral score (p < 0.001, adjusted r² = 0.65).

Interpretation: Further animal studies of the interventions that we found to improve both functional and structural outcomes in animals, using better experimental designs, could target efforts to translate effective treatments for ICH in animals into randomized controlled trials in humans.

The global burden of acute spontaneous (nontraumatic) intracerebral hemorrhage (ICH) and its outcome appear unchanged over the past quarter century,1,2 despite the improvements in outcome that can be achieved by organized stroke unit care and neurosurgical hematoma evacuation.3–6 The quest for other effective therapies has been fuelled by the recent failures of recombinant activated factor VII and the neuroprotectant drug NXY-059 to improve outcome after acute ICH in humans.7,8 Randomized controlled trials of medical therapies such as blood pressure lowering are ongoing,9 but the search for other candidate interventions may best start with methodologically robust laboratory research in animal models that accurately mimic human ICH.10 Therefore, we aimed to undertake a systematic review of nonsurgical interventions in controlled studies of animal models of ICH reporting neurobehavioral outcomes, to explore their methodological quality, and to perform a meta-analysis of the effects of each class of intervention.

Subjects and Methods

Eligibility Criteria

We sought controlled studies, regardless of their language of publication, of nonsurgical interventions given to wild-type (nontransgenic) animals after the induction of ICH using autologous blood or collagenase injection11 that reported neurobehavioral outcome.

Information Sources

In July 2009, we searched Ovid Medline (from 1950), Ovid Embase (from 1980), and ISI Web of Knowledge (from 1969) using comprehensive electronic search strategies (Supporting
Information Table 1). We also searched the proceedings of the Society for Neuroscience and the first and second International Symposia on Cerebral Hemorrhage using the ISI Web of Knowledge Conference Proceedings Citation Index. We screened the bibliographies of eligible studies for other eligible studies. We contacted authors to clarify study eligibility, where necessary.

**Study Selection**

One investigator (J.F.) screened all titles and available abstracts for eligibility, and removed duplicates. Studies that appeared to be eligible were read in full by 2 investigators (J.F., and E.S.S. or R.A.-S.S.), and disagreements were resolved by either discussion or arbitration by a third investigator (M.R.M.). We obtained translations of studies that were not written in English.

**Data Collection**

Two investigators independently extracted data on experimental design, study quality attributes, intervention characteristics, functional outcome (neurobehavioral score, measured on any scale), and structural outcomes (brain water content, or hematoma size). We recorded these data in the Collaborative Approach to Meta-Analysis and Review of Animal Data from Experimental Stroke (CAMARADES) Microsoft Access 2003 data manager application. For every treatment comparison (a given dose of an intervention at a given time of administration after ICH), we extracted the number of animals in each treatment group, the mean outcome score, and the standard deviation or standard error of the mean. We extracted all neurobehavioral outcomes at the latest time of assessment, as well as at 7 ± 2 days after the induction of ICH (if reported) and structural outcome data at the latest time when animals were culled. We extracted outcome data on untreated sham groups (in which ICH had not been induced) from experiments where these were included; in experiments that did not, we inferred sham neurobehavioral scores for functionally unimpaired animals, sham hematoma volumes of zero, and sham brain water content values corresponding to the contralateral brain region of the control/vehicle group. Unless outcomes were quantified at the relevant time points, we measured them from publications' figures (using Adobe measuring tools). We contacted authors to obtain unpublished or missing data.

**Quality Assessment**

We assessed each study’s quality according to the CAMARADES 10-item checklist,12 which consists of reporting of a sample size calculation; use of animals with comorbidities (eg, hypertension or diabetes); control of animals’ temperature; use of anesthetics other than ketamine (because of its marked intrinsic neuroprotectant activity13); randomized treatment allocation; treatment allocation concealment; blinded assessment of outcome; publication in a peer-reviewed journal; statement of compliance with regulatory requirements; and statement of potential conflicts of interest.

**Data Analysis**

**META-ANALYSIS.** The prespecified primary outcome was functional outcome as measured by neurobehavioral score at the latest time point after ICH induction. Prespecified secondary outcomes were structural measures of brain injury: either brain water content or hematoma volume. We quantified effect sizes with the normalized (weighted) mean difference summary statistic, using the summary measures of outcome provided in the individual comparisons in studies where sufficient data were available to allow this analysis (outcome reported for [1] animals with ICH receiving the intervention, [2] control animals with ICH receiving vehicle, and [3] sham animals neither undergoing ICH nor receiving treatment). If the same group of animals was assessed using several neurobehavioral scores or structural outcomes in 1 study, we combined these using fixed effects meta-analysis and used this pooled measure for further analysis. We used the DerSimonian and Laird weighted mean difference random effects model to aggregate the weighted summary statistic for each individual comparison into a pooled estimate of effect size,14 grouping interventions by their main putative mechanism of action as attributed by the authors of individual studies.

**SENSITIVITY ANALYSES.** We assessed the effect of key methodological study attributes by stratifying the pooled estimate of effect in all studies by the use of randomization, allocation concealment, and blinded assessment of outcome. For each class of intervention, we assessed whether the pooled estimate of primary (neurobehavioral) outcome at the last time point of assessment was modified if we restricted analyses to outcome data provided at 7 ± 2 days after ICH induction. In a post hoc sensitivity analysis, we explored whether the addition of studies that administered interventions prior to the induction of ICH affected the pooled estimate of effect on the primary outcome.

**ASSESSMENT OF HETEROGENEITY.** We used the chi-square statistic to assess the significance of differences between studies (with n − 1 degrees of freedom), and used the Bonferroni correction to calculate significance, taking into account the number of comparisons performed.

**META-REGRESSION.** We identified studies describing functional and structural outcomes in the same group of animals. We used meta-regression to explore the relationship between functional and structural outcome and: aspects of study quality; the intervention tested; the time of treatment; and the time of outcome assessment. Meta-regression extends the random effects meta-analysis model by taking into account 1 or more study-level covariates and determines how much heterogeneity can be explained by taking into account both within- and between-study variance.15 We performed meta-regression using STATA 10 with the linear function metareg. We built the regression model in a hierarchical manner, by running the regression analysis 1 variable (or group of dummy groups) at a time. The variable with the most significant change in the F ratio was the first variable to enter the model. The second variable was then chosen by the largest change in the F ratio in the model that already contained the first variable. This process was iterated until there were no significant changes in the F ratio. The adjusted $R^2$ value provided indicates how much residual heterogeneity is accounted for by the covariates.
Results
Our searches identified 98 potentially eligible studies of the effect of nonsurgical interventions on neurobehavioral outcome (Fig 1). We excluded 10 studies because ICH was not induced using collagenase or autologous blood injection,\textsuperscript{16–18} outcome data could not be extracted or obtained,\textsuperscript{19} variance values of zero precluded meta-analysis,\textsuperscript{20,21} or interventions were administered before the induction of ICH\textsuperscript{22–25} (although we included 3 of these studies in a post hoc sensitivity analysis\textsuperscript{24–26}). After corresponding authors clarified the data in 2 studies,\textsuperscript{27,28} we included 88 studies reporting 151 different treatment comparisons (87 of which had sham groups) that described the effects of nonsurgical interventions on neurobehavioral outcome in 2,616 treated or control rodents with ICH (Supporting Information Table 1).\textsuperscript{26–113}

Study Characteristics
The most frequently used animal model of ICH was collagenase injection (53 [60%] studies; see Supporting Information Table 1), and ICH was induced in the striatum in all but 1 study (99%).\textsuperscript{51} The median number of treated or control animals used in each study was 14 (interquartile range, 12–21; full range, 6–100). Only 2 studies reported deaths prior to the planned time of culling: 7 of 63 animals died during ICH induction in 1 study, and 5 of 18 died during another (representing a case fatality rate of 15%).\textsuperscript{86,94} Sixty-four different interventions were given at a median interval after ICH induction of 2 hours (interquartile range, 20 minutes to 6 hours). The included studies reported outcome using 38 different neurobehavioral scales (see Supporting Information Table 1), most often the forelimb placing test (10%), neurological severity score (10%),\textsuperscript{114} corner turn test (9%), and modified limb-placing test (9%). Brain water content was the most frequently reported structural outcome (38 [43%] studies, most of which [89%] used wet and dry brain weights to quantify it), and hematoma volume was reported in 30 (34%) studies (and was measured by several techniques including a variety of histological methods, spectrophotometry, or magnetic resonance imaging).

Risk of Bias of Included Studies
The median study quality score was 4/10 (interquartile range, 3–5). Of 88 publications, 27 (31%) reported random allocation to group, 7 (8%) reported allocation concealment, and 43 (49%) reported the blinded assessment of outcome. No study reported a sample size calculation, only 1 study reported the use of animals with comorbidities, and 28 (32%) used ketamine, which is an anesthetic agent with intrinsic neuroprotective properties.\textsuperscript{13} For neurobehavioral score, studies that did not report random allocation to group were associated with larger effect sizes (32%; 95% confidence interval [CI], 26–39; n = 107) compared to those that did (21%; 95% CI, 13–29; n = 42; chi-square = 437; df = 1; p < 0.0001).
Similarly, studies that did not report allocation concealment were associated with larger estimates of effect (31%; 95% CI, 26–37; n = 132) compared to studies that did (20%; 95% CI, 6–33; n = 17; chi-square = 33, df = 1, p < 0.0001). The reporting of blinded assessment of outcome accounts for a significant proportion of between-study heterogeneity (chi-square = 43, df = 1, p < 0.001), but there was no difference in effect sizes between studies that did blind outcome assessment and those that did not (29%; 95% CI, 22–36 vs 30%; 95% CI, 21–38).

**Neurobehavioral Outcomes**

Stem cells, calcium channel blockers, anti-inflammatory drugs, iron chelators, growth factors, thrombin inhibitors, and peroxisome proliferator-activated receptor gamma agonists improved neurobehavioral scores at the last time point of assessment, although there was significant between-study heterogeneity in several of these classes of intervention (Fig 2). In a sensitivity analysis restricted to studies reporting neurobehavioral outcome at 7 ± 2 days after interventions delivered within 24 hours of ICH (using 44% of the data used to evaluate the primary outcome), the improvement of neurobehavioral outcome became statistically significant for some interventions (anti-oxidants, glutamate receptor antagonists, heme oxygenase inhibitors, and minocycline) and insignificant for others (calcium channel blockers, thrombin inhibitors, and tumor necrosis factor (TNF)-α inhibitor antisense oligonucleotide). In a post hoc sensitivity analysis of the primary outcome in studies of interventions already included in the meta-analysis, also including studies that treated animals prior to ICH induction had no significant impact on the pooled estimate for the 3 intervention groups (antioxidants, estrogens, and TNF-α inhibitor antisense oligonucleotides) reporting both pre- and post-ICH delivery of intervention (see Fig 2). We did not further analyze data from 2 studies that treated animals prior to the induction ICH with agents that had not also been administered after ICH induction.

**Structural Outcomes**

In the studies that reported brain water content in addition to neurobehavioral scores (Fig 3), the pooled reduction of brain water content was 34% (95% CI, 25–43) in 78 comparisons involving 867 animals, with substantial heterogeneity between studies (chi-square = 956, p < 0.0001). Among 19 classes of intervention, 12 (63%) significantly reduced brain water content (see Fig 3), and 8 of these also significantly improved neurobehavioral scores (see Fig 2): stem cells, calcium channel blockers, anti-inflammatory drugs, iron chelators, estrogens, microglial inhibitors, propofol, and an angiotensin II receptor blocker. Of these 8 interventions, the angiotensin II receptor blocker (81%; 95% CI, 68–94) and anti-inflammatory drugs (19%; 95% CI, 0.5–38) also reduced hematoma volume. There was neither an improvement in neurobehavioral outcome nor a reduction in brain water content for antioxidant drugs, glutamate receptor antagonists, hypothermia, and gap junction inhibitors. The effects of the remaining 11 interventions on functional and structural outcomes were discordant.

**Meta-Regression of Structural and Neurobehavioral Outcomes**

Fifty-three comparisons reported both a functional and a structural outcome in the same group of animals. Structural outcome explained 38% of the observed heterogeneity in functional outcome (t² = 816, adjusted t² = 0.38, Fig 4). In a multivariate model, structural outcome and the intervention used accounted for 65% of the observed heterogeneity in neurobehavioral score (F(15, 37) = 5.31, p < 0.001, df = 52, t² = 286, adjusted t² = 0.65). Relative to structural benefit, additional functional benefit was observed when animals were treated with anti-inflammatory drugs (30%; 95% CI, 8–51), antipoptotic drugs (38%; 95% CI, 11–65), and albumin (67%; 95% CI, 22–112) compared to antioxidants as the reference group (chosen because this group contained the largest quantity of data). Gap junction inhibitors reduced functional benefit (−30%; 95% CI, −54 to −7). These findings were not affected by whether the structural outcome reported was hematoma volume or brain water content. Study quality had no effect on the relationship between structural and functional outcome.

**Discussion**

In a systematic review and meta-analysis of 88 controlled studies describing the effects of 64 different nonsurgical interventions on 38 different neurobehavioral scales in 2,616 rodents, interventions that improved both neurobehavioral score and structural outcome(s) in 2 or more animal studies (albeit with some between-study heterogeneity) included stem cells, calcium channel blockers, anti-inflammatory drugs, estrogens, and iron chelators.

We benefited from comprehensive search strategies without language bias, prespecified outcomes and analytic approaches, sensitivity analyses, and correction of statistical tests for multiple comparisons. Other reviews have not meta-analyzed existing data. Nevertheless, publication bias is known to result in an overestimation of effect sizes in animal models, and may have affected...
FIGURE 2: Weighted mean difference meta-analysis of the effects of nonsurgical interventions on neurobehavioral outcome in controlled animal studies. Diamonds represent grouped estimates of effect (and their associated 95% confidence intervals) for either classes of intervention or several studies of the same intervention. Effect estimates are organized by classes of intervention followed by individual interventions, in descending order of sample size. Heterogeneity between studies within each class of intervention is indicated in parentheses. Squares represent point estimates of effect, and horizontal lines are their 95% confidence intervals. Details of interventions and experimental design in the individual studies are provided in Supplementary Table 1. ATSC = adipose tissue-derived stromal cells; BMSC = bone marrow stem cells; NSC = nonspecific suppressor cells; VEGF = vascular endothelial growth factor; MSC = mesenchymal stem cells; UCBC = umbilical cord blood culture; GCSF = granulocyte colony-stimulating factor; GABA = gamma-aminobutyric acid; TNF = tumor necrosis factor.
our findings. Although the use of weighted mean difference meta-analysis allows the combination and comparison of outcomes across a large number of neurobehavioral scales, these scales measure different functions, and the findings presented here represent a summary of available data.

FIGURE 3: Weighted mean difference meta-analysis of the effects of nonsurgical interventions on brain water content in controlled animal studies. Diamonds represent grouped estimates of effect (and their associated 95% confidence intervals) for either classes of intervention or several studies of the same intervention. Effect estimates are organized in the same order as Figure 2. Heterogeneity between studies within each class of intervention is indicated in parentheses. Squares represent point estimates of effect, and horizontal lines are their 95% confidence intervals. Details of interventions and experimental design in the individual studies are provided in Supplementary Table 1.

FIGURE 4: Meta-regression of functional (neurobehavioral score) and structural (brain water content or hematoma size) outcomes that were measured in the same groups of animals. The size of each point reflects the precision of each comparison.
Methodological problems known to affect animal studies were evident in the studies included in our analyses.\textsuperscript{117} Their methodological quality was generally poor (median score, 4/10), but variation among studies enabled us to confirm the potent influences of randomization and allocation concealment on effect sizes. No study reported a sample size calculation, the use of ketamine anesthesia was frequent despite its known intrinsic neuroprotective properties,\textsuperscript{13} and every study bar 1 used only healthy animals, which is known to bias animal studies of stroke.\textsuperscript{118} Furthermore, young animals without comorbidities are unrepresentative of humans who suffer ICH,\textsuperscript{9} and the predominantly striatal location of ICH induction in the included studies did not fully represent ICH in humans, which occurs in lobar regions and the posterior fossa, may extend into other brain compartments, and often causes hydrocephalus. Studies used a diverse array of neurobehavioral scales, and several used only 1 scale when a battery of tests might have given a more complete description of neurobehavioral outcome.\textsuperscript{119} Many of the scales have not been validated and rely on primarily motor tasks. Moreover, the neurobehavioral scale used should relate to whether an ICH is cortical or striatal, because the neurological impairments arising from ICH vary by the anatomical location of the ICH.\textsuperscript{119} Furthermore, because rodents have proportionately less white matter than humans,\textsuperscript{11} and may have greater neuroplasticity,\textsuperscript{115} there may be problems with the relevance of the animal models to human ICH. There are differences between the 2 main rodent models of ICH,\textsuperscript{120} but we found no evidence that 1 was any more relevant to the human condition than the other. There was a general shortage of external validation of interventions that appeared to improve outcome.

We have used meta-analysis to derive summary estimates of efficacy from a collection of relatively small studies, many of which were not sufficiently powered to detect modest treatment effects. Where possible, we have used weighted rather than standardized mean differences, because this is a more powerful statistical approach when the size of contributing studies is small.\textsuperscript{121} The use of meta-analysis to aggregate data for neurobehavioral outcomes described using ordinal scales is well established (and indeed many of the contributing studies use parametric statistics to analyze these data), and is justified because parametric analyses of nonparametric data becomes more valid when large numbers of studies are aggregated in this way.\textsuperscript{122}

Importantly, only 40% of the variation in treatment effect on neurobehavioral outcome was attributable to the observed effect on structural outcome, increasing to 65% when drug-specific effects were also taken into account. Although this provides some basis for the use of structural outcomes in clinical trial programs, it does suggest first that such structural outcomes capture only a proportion of efficacy, and second that the relationship between structural and functional outcome—and therefore the utility of such structural outcome measures—may vary substantially between different drug classes.

Discordance has previously been reported between animal and human studies in other diseases,\textsuperscript{123} and we have shown that this is also the case for ICH. Of the interventions that improved both structural outcomes and neurobehavioral scores in animal models (see Figs 2 and 3), anti-inflammatory drugs have not improved outcome in humans.\textsuperscript{124} Similarly, recombinant activated factor VII did not improve outcome in humans, and indeed the first evidence of benefit on structural outcomes in animals was reported after the start of human trials.\textsuperscript{7,125}

For progress to be made in translational ICH research, the quality of animal studies must improve.\textsuperscript{126} We have reasonable understanding of some of the pathophysiological processes underlying ICH in animals,\textsuperscript{3,127} but future research might usefully focus on understanding to what extent these mechanisms are important in humans, allowing drug development to be targeted at the key processes. Furthermore, we need animal models of ICH that better mimic important pathophysiological processes in humans, including hematoma expansion and recurrence,\textsuperscript{3,10,115} and we need these experiments to be conducted in such a way as to minimize the risk of study quality bias. Only then may the translational paradigm be reliable enough for effective treatments in animal models to be tested in randomized controlled trials in humans.

Acknowledgment

R.A.-S.S. was funded by a clinician scientist fellowship from the UK Medical Research Council. M.R.M. acknowledges the support of the MRC Trials Methodology Hub.

We thank E. Lebedeva, A. Wong, and Dr C. Fournaris for their generous assistance with translating studies for this review.

Potential Conflicts of Interest

Nothing to report.

References


