Neuroprotection by anaesthetics in rodent models of traumatic brain injury. A systematic review and network meta-analysis

<table>
<thead>
<tr>
<th>Journal:</th>
<th>British Journal of Anaesthesia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>BJA-2018-00146-JH012.R2</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Review article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>n/a</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Archer, David; University of Calgary Cumming School of Medicine, Anesthesia McCann, Sarah; University of Edinburgh, Center for Clinical Brain Sciences Walker, Andrew; University of Calgary Cumming School of Medicine, Anesthesia Premji, Zahra; Haskayne School of Business University of Calgary, Libraries and Cultural Resources Rogan, Kyle; University of Calgary Cumming School of Medicine, Anesthesia Hutton, Meredith; University of Calgary Cumming School of Medicine, Anesthesia Gray, Laura; University of Leicester, Department of Health Sciences</td>
</tr>
</tbody>
</table>
Neuroprotection by anaesthetics in rodent models of traumatic brain injury. A systematic review and network meta-analysis

David P. Archer¹*, Sarah K. McCann², Andrew M. Walker¹, Zahra A. Premji³, Kyle J. Rogan¹,
Meredith J.H. Hutton¹, Laura J. Gray⁴
¹Cumming School of Medicine, University of Calgary, Calgary, Canada; ²Center for Clinical Brain Sciences, University of Edinburgh, Scotland, United Kingdom; ³Libraries and Cultural Resources, University of Calgary, Calgary, Canada; ⁴Biostatistics Research Group, Department of Health Sciences University of Leicester, United Kingdom

*Corresponding Author: David P. Archer, Department of Anesthesia, Foothills Medical Center, 1403 29 St. N.W., Calgary, AB, T2N 2T9, Canada; Phone: 1-(403) 944-2439, Email: darcher@ucalgary.ca

This manuscript is for literature review and analysis; it does not involve active investigation of humans or animals.

Running Title: Anaesthetic Neuroprotection in Preclinical Traumatic Brain Injury
Summary

Background: Anaesthetic neuroprotection in the setting of traumatic brain injury (TBI) remains unproven and is based upon results in the preclinical experiments. Here we sought to synthesize the results in rodent models of TBI and to evaluate the effects of publication bias, experimental manipulation and poor study quality on the effect estimates. Methods: Following a systematic review, we use pairwise meta-analysis to estimate the effect of anaesthetics, opioids, and sedative-hypnotics on neurological outcome and network meta-analysis to compare their relative efficacy. We sought evidence of bias related to selective publication, experimental manipulation, and study quality. Results: Sixteen studies, involving 32 comparisons were included (546 animals). Treatment improved neurological outcomes by 35%, 95% confidence interval (C.I.) 26 to 44%, P<0.001. Statistical heterogeneity was small (12%) but the 95% prediction interval for the estimate was wide (15 to 56%). Statistical power was low: 61% (90% C.I. 22-86%). Small sample size in the studies was a serious shortcoming reducing statistical heterogeneity and obscuring differences in outcome between drugs and between experimental conditions. Conclusions: Anaesthetics do provide neuroprotection in rodent models of TBI. The effect size estimates do not appear to be exaggerated by selective publication, experimental manipulation, or study design. The main shortcoming of the included studies was small sample sizes leading to low power and imprecision, which precluded the network meta-analysis from providing a meaningful ranking for efficacy among the drugs. Reliable preclinical investigations of neuroprotection by anaesthetics will require larger sample sizes.

KEYWORDS: anaesthetics, general (pharmacology); brain injuries, traumatic; mice; network meta-analysis; neuroprotection, drug effects; rats; review;
INTRODUCTION

Evidence to support drug choices for sedation of patients with severe traumatic brain injury remains elusive. In a systematic review of a wide variety of agents: ketamine, propofol, etomidate, and agents from diverse classes; benzodiazepines, α-2 agonists, and antipsychotics the authors found no convincing evidence to rank drug efficacy. The authors concluded that clinical trials were urgently needed. The ‘bench-to-beside’ research strategy begins with hypothesis testing in preclinical models in vitro and in vivo, leading ultimately to clinical trials. Using rodent models of ischaemic stroke and traumatic brain injury (TBI), successful development of novel treatments has been rare, leading to efforts to improve the quality of preclinical research. Explanations offered for the lack of success include the use of inappropriate animal models, exaggeration of effects by study design flaws, and bias. These features of preclinical research have the potential to generate false positive results, but we do not know to what extent these criticisms apply to preclinical research in TBI. Here we evaluated the results from experimental models of TBI in rats or mice with respect to overall effects of treatment, the comparative efficacy of the study drugs, and sources of bias. We used conventional pairwise meta-analysis with meta-regression and P-curve analysis to identify bias from study factors, selective publication and data manipulation. We performed a network meta-analysis to compare and rank the effects of individual drugs.

METHODS

The research approach and methods were performed according to guidelines for preclinical meta-analysis, pre-specified in a study protocol available online at Syrf.org.uk (search under “David Archer”). This systematic review is reported according to the PRISMA statement for network meta-analysis (PRISMA NMA checklist).

Search Strategy

The search strategy for the systematic review was developed by the team librarian (ZP) with one of the investigators (DA) and was independently reviewed by a second librarian. The database(s) searched included: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to 2017 May 19; Biosis all years, and Embase 1974- May 19, 2017 (Supplementary Data, Table S1, initial search January 3, 2017, final search...
May 19, 2017 – no change). We searched reference lists from included manuscripts and pertinent review articles. Titles and abstracts of manuscripts retrieved by the search were evaluated independently by two investigators (DA, AW). For these, the full article was reviewed to identify studies for inclusion.

**Inclusion and Exclusion Criteria**

We included abstracts, E-pub and published articles from 1946 to 2017 May 19, with no language restriction. For consistency of study design and participants, we limited the range of preclinical models to non-penetrating methods in rats or mice. We excluded studies of brain ischemia and blast injury because the experimental models are quite different from the non-penetrating models that we did include.

We included pharmacological interventions that are frequently administered during clinical management of TBI: sedative-hypnotics, intravenous or volatile anaesthetics, and opioids. It is not known that the mechanisms by which anaesthetics may exert neuroprotective effects coincide with those by which they exert their clinical effects (anaesthesia, sedation or analgesia). By including drugs of different classes, we have chosen to emphasize assessment by outcome without any mechanistic restrictions (for example see Cipriani and colleagues\(^\text{12}\)). For the meta-analysis, we excluded studies in which the intervention was delayed more than 1h after the injury. Studies were excluded if, after contact with the authors, the effect size of the outcome measures could not be expressed as a mean and standard deviation and when the number of experimental animals could not be determined.

We entered the following information into a spreadsheet: reference identification: authors, year of publication, source (journal, abstract), conflict of interest statement, regulatory approval, nature of subjects: species/strain, age, weight, sex; anaesthetic information: study drug (dose, timing of administration), and control drug or ‘awake’ controls. Characteristics of studies are summarized in Table S2 (Supplementary Data). Data were extracted independently in duplicate by DA and AW using a standardized *proforma* and verified by SM.

**Data Handling in Meta-Analysis**

For data handling in the analysis, we converted the neurological outcomes to an effect index: the pooled normalised mean difference (NMD, Equation 1).
We chose this outcome index (instead of standardized mean difference) for ease of interpretation – the units represent % neuroprotection. Neurological outcome scores may be either good or bad, depending on the scoring system - values were adjusted for the direction of effect. In the network meta-analysis, we compared all treatments to a joint comparator (no treatment or placebo or vehicle were lumped together).

Pairwise meta-analysis of NMD between control and treatment was performed using the inverse variance method for weighting, the DerSimonian-Laird estimator for $\tau^2$, and the Hartung-Knapp adjustment for the random effects model. The similarity of treatment effects among the included studies was quantified with the heterogeneity index, $I^2$, and tested using the Q statistic with a nominal significance value of $P < 0.1$. The expected range of true effects in similar studies was estimated with the prediction interval. To investigate potential sources of heterogeneity, we used meta-regression to identify effect size moderators among the pre-defined covariates.

The present analysis involves many small studies and is therefore at risk for overestimating the effect size because scientific journals may be more likely to publish studies with large effect sizes and studies that achieve nominal statistical significance ($P < 0.05$).

Experimental manipulation to achieve statistical significance (‘p-hacking’) includes data peeking (adding subjects until significance is achieved), selective reporting of data, and exclusion of outliers. We used two tools to identify and correct for these forms of publication bias. To correct for selective reporting of studies with large effect sizes we constructed a funnel plot and applied the trim-and-fill method. To correct for selective suppression of nonsignificant results and p-hacking, we applied P-curve analysis, which uses the distribution of statistically significant p values from the included studies to determine whether a true effect is likely to be present (evidential value of the results) and to estimate the size of the effect. The effect size estimated from the P-curve is the anticipated value that would be obtained if the included studies were rerun. Details of the P-curve analysis are provided in section A.4 of the Supplementary Data.

$$NMD = \frac{(\text{Mean control} - \text{Mean sham}) - (\text{Mean treatment} - \text{Mean sham})}{\text{Mean control} - \text{Mean sham}} \times 100\% \quad \text{Equation 1}$$
Pairwise meta-analysis was performed using the metafor package (version 4.8-4) with R (version 2.0.0), available online at the Cran Project. Meta-regression and subgroup analyses were performed with the software package ‘Comprehensive Meta-Analysis Version 3’ available online (www.meta-analysis.com).

We used the netmeta package (version 0.9-6) available online at Cran Project to perform network meta-analysis. Network geometry was assessed graphically: the netgraph command optimized the distance between nodes (treatments) for direct comparisons using a stress majorisation algorithm\textsuperscript{15}. Line width for edges connecting the nodes was proportional to the precision of the estimate of the comparison.

Certain design features of preclinical studies have been reported to bias, that is to systematically enhance or diminish the observed effect size. Here we used meta-regression to directly examine the impact of these factors on observed outcome. We used the risk of bias tool developed by the Collaborative Approach to Meta-Analysis and Review of Animal Data in Experimental Stroke (CAMARADES score)\textsuperscript{6} to evaluate the study risk of bias. In the first step, two investigators (KR, MH) independently extracted the components of a modified CAMARADES score [1] monitoring of blood pressure and blood gases, [2] statement of temperature control, [3] randomisation to treatment or control group, [4] allocation concealment, [5] blinded assessment of outcome, [6] sample-size calculation/power calculation/previous studies, [7] statement regarding regulatory compliance, [8] statement regarding possible conflict of interest. For each of these items compliance is scored “1”, otherwise “0” – a low score indicates increased risk of bias. Omitted from the original score were: ‘publication in a peer-reviewed journal’ (all studies were), and ‘avoidance of an anaesthetic with intrinsic neuroprotective activities’, resulting in a maximum score of 8. Since the CAMARADES tool has not been validated for TBI, we then used meta-regression of individual components against effect size to identify those categories that were associated with a significant exaggeration or reduction of effect in the present dataset. The studies identified as being at risk of bias were then managed in two ways. First, the meta-analysis was repeated excluding the suspect studies to determine to what extent the results were changed. Secondly, compliance in these categories was used to grade the study risk of bias. Using the Cochrane
Grade approach\textsuperscript{16}, a study that has characteristics that are associated with biased effect sizes was identified as having a moderate or high risk of bias. If a study was non-compliant in significant bias item, the risk of bias of that study was downgraded from “low” to “moderate”; if the study was non-compliant in two categories, the risk of bias was downgraded from “moderate” to “severe”. Using this approach, no results were excluded from the analysis.

Concerns about study bias are reflected in the “level of confidence” in the evidence, calculated using the CINeMA method (CINeMA)\textsuperscript{17}.

The level of agreement among the network comparisons (consistency) was evaluated both locally and globally. Local inconsistency was identified by disagreement between the direct and indirect evidence values for a specific comparison (e.g. isoflurane versus vehicle), generated with the netsplit command in netmeta and evaluated with a z test ($P<0.1$). Presence of disagreement (local inconsistency) resulted in downgrading the confidence in the evidence in CINeMA. Global consistency (also called incoherence) was assessed with Cochran’s Q for the whole network. For network meta-analysis the included studies must be, in principle, ‘jointly randomisable’ and that is it is possible to conceive of a large trial that includes all of the treatments (transitivity assumption). Transitivity was assessed by reviewing study designs to ensure that treatments were jointly randomisable. We assessed the level of confidence of the evidence using the Cochrane GRADE approach\textsuperscript{16} adapted for network meta-analysis using an online application by Salanti and colleagues\textsuperscript{17} (CINeMA). The number of comparisons in a meta-analysis is represented by $k$, the number of animals in a study or study arm is represented by $n$.

Results are reported as mean values with 95% confidence intervals (95% C.I.). The summary evidence grade for each comparison is indicated by color-coding (high = green; moderate = gold; low=red) on the forest plot.

\textbf{RESULTS}

The search strategy returned 16 studies for inclusion in the analysis (Figure 1, Table S2 Supplementary Data for characteristics of studies). The experimental models for TBI in rats or mice were weight drop (WD), controlled cortical impact (CCI), and fluid percussion injury (FPI). These techniques are considered to be useful for modelling concussion and diffuse brain injury\textsuperscript{34}. Sixteen treatments were compared: (chloral hydrate, diazepam (two doses),
etomidate, fentanyl, halothane (two doses), isoflurane, ketamine, morphine (two doses),
pentobarbital, propofol, sevoflurane, vehicle, and xenon). Most comparisons were two-armed
(12 of 16; 75%); three studies were three-armed and one study had eight arms. The 16 included
studies generated 32 pairwise comparisons for neurological outcome and 16 comparisons for
lesion volume. The neurological outcomes were used in the meta-analyses. The included
studies generated 56 direct comparisons for the network meta-analysis. Overall, the studies
involved 546 animals (345 (63%) rats). The studies were small: the average group size was 13.
Using the CAMARADES risk of bias scoring system (modified, maximum score = 8), a majority of
studies (11/16, 69%) were of moderate-to-high quality (moderate - modified quality score 4-6
(five studies); high quality – score 7-8 (six studies) (Table 1). Pre-reconciliation inter-rater
agreement on the study characteristics was ‘substantial’; $\kappa$ (Cohen) = 0.635; (0.508 to 0.762).
Compliance with study design features that mitigate investigator bias was good:
randomisation - 81%, blinded allocation – 63%, blinded assessment – 63%, and sample
size/power calculation/previous studies – 63%. Inter-rater agreement for the four bias-
mitigating study design characteristics was ‘high’; the range of $\kappa$ was 0.783 to 0.922 (probability
of random agreement, $P<0.001$).

**Effects of Treatment - Pairwise Meta-Analysis**

The treated animals had a 35% better neurological outcome than the control animals (95% C.I.:
26 to 44%; $P<0.001$; $k=32$; prediction interval: 15 to 56%). Among the predefined sources of
bias, only failure to randomise was identified by meta-regression as being associated with
increased effect. The timing of intervention was negatively correlated with effect size, so
including pre-treated studies would not inflate the estimate. The CAMARADES score did not
have a significant impact on the effect size ($P=0.2164$). Excluding the results from the six
comparisons that involved non-randomised studies reduced the estimated improvement to
28% (20 to 37%, $P<0.001$) which does not change the finding that treatments were associated
with improved outcome. The funnel plot (Figure 2), which is symmetrical with no studies added
by the trim and fill method, does not support a role for preferential publication of studies with
large positive effect sizes.
The P-curve analysis showed that the proportion of significant results that are \( p < 0.025 \) (82%) is greater than that which would be expected if there was no effect (50% of results), indicating that there is evidence of a true effect (\( Z = -2.53, p = 0.0057 \)). The statistical power (probability that they will reject the null hypothesis when the null hypothesis is false) of the tests included in the P-curve was low: 61% (90% C.I. 22-86%). The effect size estimated from the P-curve was 35% improvement (for details see Supplementary Data, A3).

There were many design differences (experimental/methodological heterogeneity) among the included studies: species (rats: 24 comparisons, mice: 8 comparisons), timing of intervention relative to TBI (before: 15 comparisons; after: 17 comparisons), conditions for the control group (anaesthetised or sedated: 16 comparisons; no treatment, vehicle, saline: 16), study conduct (e.g. allocation concealment, blinding) and the drugs being investigated. The contribution of these factors to the effect size was assessed with meta-regression. A selection of subgroup results is visualized in Figure 3. Overall, differences in study design had only a small impact on the variation in effect sizes of the comparisons. Statistical heterogeneity was low; \( I^2 = 12.3\% \) (0.0% to 42.9%), \( \tau^2 = 80.07, Q = 35.34 \) (df=31, \( P = 0.2706 \)). The estimates from the pairwise analysis are imprecise, as indicated by the wide prediction interval. The contribution of variation in effect among drugs was evaluated with the network meta-analysis.

We used backward step-wise meta-regression to evaluate the potential roles of the pre-defined study characteristics as effect modifiers and risk of bias: control conditions – anaesthetised vs vehicle, species, injury model, randomisation, blinded allocation, blinded assessment, sample size calculation and timing of intervention. Randomisation and timing of intervention were the only significant covariates (Table S5, Supplementary Data). Randomised comparisons (\( k = 26 \)) were associated with 28% (20% to 37%) improvement in neurological outcome compared to 63% (44% to 82%) improvement in non-randomised comparisons (\( k = 6; P < 0.001 \)). Based on the latter finding, lack of randomisation was identified as the only risk factor for bias in the study design category in the GRADE evidence evaluations. Drug administration before the injury (\( k = 15 \)) was associated with a 21% (9% to 33%) improvement compared to 47% (36% to 58%) improvement when treatment was concurrent with or followed injury (\( k = 17; P < 0.001 \)). Considering that a clinical setting such as neurosurgery may involve drug administration before
the potential injury and the finding that 'before' timing decreases rather than exaggerates the
effect size, we chose not to downgrade evidence based upon the timing of intervention.
Species, injury model, control conditions and CAMARADES score were not significant covariates
in meta-regression (P>0.05). There were no included studies that involved female animals. In
the CINeMA application for network GRADE analysis, we set the risk of publication bias for all
comparisons in the network at ‘undetected’.

**Network Meta-analysis**
The studies included in the analysis involved 16 treatments in 17 study ‘designs’; the network
provided 56 direct pairwise comparisons. Designs were defined by the subset of treatments
compared – isoflurane versus vehicle and halothane versus vehicle are two different designs.
The evidence graph (Figure 4) shows that the network is a sparsely-connected intersection of
two ‘star’ networks, the principal comparators being vehicle and isoflurane.
Most interventions were associated with an improvement in neurological outcome (Figure 5).
The exceptions were chloral hydrate, etomidate, morphine (1 mg.kg\(^{-1}\)), and propofol. The
confidence intervals of many of the drugs overlap, showing that the analysis could not
distinguish differences related to specific pharmacological characteristics. The ‘non-
randomised’ study design affected four direct treatment comparisons with vehicle: etomidate,
diazepam (5 and 10 mg.kg\(^{-1}\) intra-peritoneal, (i.p.) doses), and morphine (10 mg.kg\(^{-1}\) i.p.). We
identified the three non-randomised studies as having a moderate risk of bias - ‘some concerns’
in the study limitations section of the CINeMA evidence certainty assessment.

**Indirectness and Transitivity**
The included studies were all judged to be relevant to the research question (low indirectness).
Review of the designs of the included studies supported compliance with transitivity
assumption: we did not find any unique designs; the treatments were jointly randomisable.
There were not enough studies to permit direct evaluation of the distribution of potential effect
size moderators, randomisation and timing of intervention.

**Inconsistency (Heterogeneity and Incoherence)**
Overall, the network is coherent – \( Q_{\text{global}} (\chi^2) = 14.03; \text{df} = 17, P=0.6652; Q_{\text{between designs}} = 13.59; \text{df} =11, P = 0.2564; Q_{\text{within designs}} = 0.444; \text{df}=6, P=0.9985. \) However, there were seven instances
of local incoherence - differences between the direct and indirect estimates out of 40 possible comparisons (18%). The majority of local incoherence (6/7) involved diazepam or etomidate, which received major contributions from non-randomised studies.

DISCUSSION

Our synthesis of the evidence from rodent models of TBI suggests that many anaesthetics in common use improve neurological outcome in these settings by 30-40%. We did not find evidence that this result is influenced by bias due to study design, selective publication or experimental manipulation. It is therefore likely the estimates represent true effects and that many of these drugs may confound findings in experiments in which they are co-administered with specific target agents. The main shortcoming of the included studies, common to many studies in neuroscience, was their small size which resulted in imprecision and low statistical power. This is evident in the wide prediction interval in the estimate from the pairwise meta-analysis analysis (average observed effect, NMD = 35%, prediction interval = 15 to 56%).

Although the value of the heterogeneity index, $I^2 = 12\%$, would seem to suggest low heterogeneity, that is due to large sampling errors in the individual studies related, in turn, to their small size. $I^2$ is the proportion of the variance in observed effects that reflects variance in true effects rather than sampling error (Equation 2),

$$I^2 = \frac{T^2}{V_{obs}} \quad \text{Equation 2}$$

where $T$ is the standard deviation of the true effects and $V_{obs}$ is the variance of the observed effects. $I^2$ is, therefore, a relative, not an absolute measure of heterogeneity. When the sampling error is large, $I^2$ is small. The heterogeneity in the present studies is more reliably reflected by the prediction interval. The large sampling error and low precision also explain the counter-intuitive result that effect sizes were not sensitive to experimental conditions or drug selection. (When measured in units of meters, there is no difference in average height between grade school students in Year 1 and Year 3.)

Treatment ‘during’ TBI is unique to the experimental setting. The present data analysis is not adequate to address the role of treatment timing in outcome after experimental TBI and requires further experiments designed specifically to answer this question.
At a practical level, a research director for a pharmaceutical company, examining the forest plot (Figure 5) to identify a promising new agent will find it difficult to define a cut-off to proceed beyond rodent models. If a threshold of 70% improvement is chosen, none of the drugs makes the cut, but eleven of them have 95% confidence intervals that exceed the cut-off. Similar problems occur with most other thresholds. We are not criticizing the choice of small groups, simply demonstrating the practical consequences that ensue from the resulting lack of precision. Small sample sizes are common and appropriate in hypothesis-testing research designs that seek ‘proof of concept’ rather than precise estimation of effect size, thereby limiting the sample size to the minimum predicted by power analysis to achieve nominal statistical significance ($P<0.05$).

For the clinician, the relationship to patient care (translatability) is even more problematic for three reasons. First, in contrast with the standardized experimental models, human TBI is very diverse. Second, the lack of physiologic support to prevent secondary brain injury in most preclinical study designs is a major departure from clinical practice. Finally, clinical care often requires drug administration for prolonged periods after the original injury. When administered late ($\geq 2$ h after injury), propofol has been reported to have detrimental, rather than beneficial effects, suggesting that treatment timing may be critically important. We categorize the level of confidence in the evidence for translatability to human studies as “very low”.

It is likely that many of the drugs in the included studies can influence outcomes, potentially confounding the results of planned experimental interventions. The data is not precise enough to make recommendations about which agents to choose or avoid in preclinical experiments; even though chloral hydrate appears to have little effect, there was only one study informing the estimate. Although the improvement in outcome is probably not sufficient to merit investigation of these drugs as ‘treatments’ they do provide a degree of neuroprotection. Here the translatability of the models becomes an issue. To improve precision in future preclinical work in this field, adequately powered studies that are designed to more closely mimic clinical conditions may be financially and logistically feasible if collaboration among centres is designed a priori (e.g. multi-centre studies or pre-planned network meta-analysis).
Acknowledgments

The authors acknowledge the contribution of Helen Lee Robertson, Libraries and Cultural Resources, the University of Calgary in the preparation and review of the literature search strategy.

Conflict of Interest

The authors have no conflict of interest to declare.

Funding

Partial salary support provided by the University of Calgary, the University of Leicester and the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 702213

This manuscript is for literature review and analysis; it does not involve active investigation of humans or animals.

The protocol for this study was registered with SyRF.org.uk (search ‘David Archer’).

Authors’ Contributions

Study design, literature search strategy, and initial manuscript screening: D.P.A, S.K.M., A.M.W., Z.M.P.

Data extraction and verification: D.P.A., A.M.W., S.K.M.


Writing the first draft of the manuscript: D.P.A.

Manuscript revision: L.J.G., S.K.M., A.M.W. D.P.A

Data Availability Statement: All relevant data are included in the manuscript and the supplementary information.
References


(6) Sena ES, Van der Worp HB, Howells DW, Macleod MR. How can we improve the preclinical development of drugs for stroke? Trends Neurosci 2007; 30: 433-9


(9) Ioannidis JPA. Why most published research findings are false. PLOS Medicine 2005; 2; e124

(10) Simonsohn U, Nelson LD, Simmons JP. P-curve and effect size: correcting for publication bias using only significant results. Perspect Psychol Sci 2014; 9: 666-81


(13) IntHout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. BMJ Open 2016; 6; e010247
(14) Duval S, Tweedie R. A simple funnel-plot-based method of testing and adjusting for
publication bias in meta-analysis. *Biometrics* 2000; **56**: 455-63

204-5

assessing the certainty of evidence from preclinical animal studies PLoS one 2018; 13:
e0187271

and reduces secondary injury following trauma in an in vivo model of traumatic brain
injury. *Crit Care Med* 2015; **43**: 149-58

(18) Dixon CE, Ma X, Kline AE, et al. Acute etomidate treatment reduces cognitive deficits and

(19) Hayes RL, Lyeth BG, Jenkins LW, et al. Possible protective effect of endogenous opioids in

associated with inhibition of pro-inflammatory cytokines in adult rats after traumatic
brain injury. *Neuropeptides* 2016; **58**: 1-6

episode of anesthesia during the induction of experimental brain trauma on secondary
brain damage and inflammation. PloS one 2011; **6**: e19948

trauma through inhibition of nicotinamide adenine dinucleotide phosphate oxidase.
*Anesthesiology* 2013; **119**: 1370-88

(23) Lyeth BG, Liu S, Hamm RJ. Combined scopolamine and morphine treatment of traumatic
brain injury in the rat. *Brain Res* 1993; **617**: 69-75

(24) O'Connor CA, Cernak I, Vink R. Interaction between anesthesia, gender, and functional
outcome task following diffuse traumatic brain injury in rats. *J Neurotrauma* 2003; **20**: 533-41

(25) O'Dell DM, Gibson CJ, Wilson MS, DeFord SM, Hamm RJ. Positive and negative modulation
of the GABA<sub>A</sub> receptor and outcome after traumatic brain injury in rats. *Brain Res* 2000; 861: 325-32


Table 1. CAMARADES Risk of Bias Table for Included Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>Physiol</th>
<th>Temp</th>
<th>Random</th>
<th>Blinded Allocation</th>
<th>Blinded Assessment</th>
<th>Power Calculation</th>
<th>Regulatory Compliance</th>
<th>Conflict of Interest</th>
<th>Total Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Campos-Pires 2015</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Dixon 2003</td>
<td>0</td>
<td>1</td>
<td>0*</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Hayes 1990</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Liu 2016</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Luh 2011</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Luo 2013</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Lyeth 1993</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>O’Connor 2003</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>O’Dell 2000</td>
<td>1</td>
<td>1</td>
<td>0*</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Shapira 1992</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Smith 1993</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Statler 2000</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Statler 2006a</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Statler 2006b</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Tecoult 2000</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Zohar 2006</td>
<td>0</td>
<td>0</td>
<td>0*</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Compliance %</strong></td>
<td><strong>56</strong></td>
<td><strong>63</strong></td>
<td><strong>81</strong></td>
<td><strong>63</strong></td>
<td><strong>63</strong></td>
<td><strong>63</strong></td>
<td><strong>75</strong></td>
<td><strong>25</strong></td>
<td></td>
</tr>
</tbody>
</table>

* indicates studies with risk of bias due to lack of randomisation as identified by meta-regression (Table S3 Supplementary Data).
Figure 1. Results of the systematic review search strategy

169x130mm (150 x 150 DPI)
Figure 2. Contour-enhanced funnel plot (k=32 comparisons) does not suggest the presence of preferential publication of studies with large effect sizes. The vertical short- and long-dash lines indicating the summary estimate (35%) overlap because there were no studies imputed by trim and fill analysis and no correction of the estimate suggested. The shaded areas indicate the significance levels to be expected in the plot if there was no true difference between treated and control groups (light gray = p < 0.01, gray = 0.05 > p > 0.01, dark gray = p > 0.05).

256x171mm (72 x 72 DPI)
Figure 3. Outcome comparisons in subgroups showing that imprecision due to sampling error obscures methodological heterogeneity—(A) participants, mice vs rats; (B) timing of treatment, before vs during/after injury; (C) injury model, controlled concussive injury (CCI), fluid percussion (FP), weight drop (WD); (D) outcome assessment, blinded vs not blinded; (E) treatment allocation, randomised vs non-randomised; (F) control conditions, anaesthetised vs vehicle. NMD = normalised mean difference, percent improvement in outcome. * indicates a significant covariate in meta-regression (P<0.01). The meta-analysis estimate and 95% C.I. are indicated by the red line and the grey rectangle respectively.
Figure 4. Network graph for the TBI neurological outcome data. Nodes (black circles) indicate treatments; the interconnecting lines (edges) indicate direct comparison. The thickness of each edge is proportional to the precision of the estimate of the comparison. Numbers in treatment labels indicate mg.kg⁻¹ by intraperitoneal injection. Shading indicates multi-arm comparisons.

228x149mm (72 x 72 DPI)
Figure 5. Forest plot comparing neurological outcome estimates (95% C.I.) to vehicle/saline/no-treatment. The area of the box surrounding the estimate is proportional to the weighting in the meta-analysis. The color of the box indicates the GRADE level of confidence in the estimate (green: high, gold: moderate, red: low). Treatments are ranked according to P-score. NMD = normalised mean difference; C.I. = confidence interval. The route of administration for volatile anaesthetics was inhaled; the route of administration for other agents varied between intraperitoneal and intravenous.