Pooling of Animal Experimental Data Reveals Influence of Study Design and Publication Bias

Malcolm R. Macleod, PhD; Tori O'Collins, BSci; David W. Howells, PhD; Geoffrey A. Donnan, MD

Background and Purpose—The extensive neuroprotective literature describing the efficacy of candidate drugs in focal ischemia has yet to lead to the development of effective stroke treatments. Ideally, the choice of drugs taken forward to clinical trial should be based on an unbiased assessment of all available data. Such an assessment might include not only the efficacy of a drug but also the in vivo characteristics and limits—in terms of time window, dose, species, and model of ischemia used—to that efficacy. To our knowledge, such assessments have not been made. Nicotinamide is a candidate neuroprotective drug with efficacy in experimental stroke, but the limits to and characteristics of that efficacy have not been fully described.

Methods—Systematic review and modified meta-analysis of studies of experimental stroke describing the efficacy of nicotinamide. The search strategy ensured ascertainment of studies published in full and those published in abstract only. DerSimonian and Laird random effects meta-analysis was used to account for heterogeneity between studies.

Results—Nicotinamide improved outcome by 0.287 (95% confidence interval 0.227 to 0.347); it was more effective in temporary ischemia models, after intravenous administration, in animals without comorbidities, and in studies published in full rather than in abstract. Studies scoring highly on a quality measure gave more precise estimates of the global effect.

Conclusions—Meta-analysis provides an effective technique for the aggregation of data from experimental stroke studies. We propose new standards for reporting such studies and a systematic approach to aggregating data from the neuroprotective literature. (Stroke. 2004;35:1203-1208.)

Key Words: meta-analysis ■ stroke ■ animal models ■ neuroprotection ■ nicotinamide

The failure of neuroprotective drugs in clinical trials represents a major challenge to the doctrine that animals provide a scientifically valid model for human stroke. This failure has provided the impetus for the creation of the Stroke Academic Industry Roundtable (STAIR) in an attempt to overcome the difficulties in taking animal neuroprotectants to successful clinical trial in humans.1–3 It has been argued that for many drugs so tested, the animal data were not sufficiently robust to warrant the expectations placed on them.4 Furthermore, the sheer volume of the published neuroprotective literature, with >4000 publications describing the neuroprotective efficacy of >700 drugs (our unpublished observations), renders it virtually impossible for any individual to maintain an overview of the field.

Systematic review and meta-analysis have contributed greatly to the interpretation and aggregation of data in the clinical sciences. Systematic review uses a methodical approach to minimize the risk of bias in the selection of studies for inclusion, whereas meta-analysis combines results from individual studies to produce a better estimate of treatment effect. Stratified meta-analysis can then be used to explore the impact of particular study characteristics.5

Nicotinamide (Vitamin B3) is a precursor of nicotinic adenine dinucleotide, an important cofactor in energy metabolism.6 Nicotinamide also inhibits poly (ADP-ribose) polymerase, an enzyme activated after DNA strand breaks; animals with targeted deletions in PARP have reduced infarct volume after experimental ischemia,7 suggesting that drugs that inhibit PARP activity may be effective in stroke.

We have investigated the characteristics of nicotinamide in experimental stroke using the techniques of systematic review, meta-analysis, and stratified meta-analysis. Specifically, we have calculated a global estimate of the efficacy of nicotinamide, and we have examined the impact of study quality and various study characteristics on the estimate of effect size.

Materials and Methods

Studies of nicotinamide in animal models of stroke were identified from Pubmed (1974 to June 2003), Embase (1980 to June 2003), and...
BIOSIS (1969 to June 2003). Our search strategy used the words [nicotinamide OR Vitamin B3] AND [stroke OR ischemia]. We also performed hand-searching of abstracts of scientific meetings, reference lists of identified publications, and requests to senior authors of identified publications for references to other studies. We included all controlled studies of the effect of nicotinamide in animal models of focal cerebral ischemia in which the outcome was expressed as a volume of infarction or a neurological score.

We extracted data for mean outcome, standard deviation (SD), and number of animals per group for individual comparisons (defined as the assessment of outcome in treatment and control groups after treatment with a given dose of drug or vehicle, with treatment starting a given time before or after the induction of cerebral ischemia). Values for data expressed graphically were requested from authors. When nicotinamide was administered in multiple doses, the comparison was grouped according to the first dose at the first time it was administered.

When neurological tests were performed at different times, only the final test was included. When 1 group of animals were scored in >1 neurological domain (for instance, motor and sensory scores), or when both neurological score and infarct volume were measured, data were combined using meta-analysis (see later) to give an overall estimate of effect size and its standard error. We defined effect size as the proportional reduction in outcome (infarct volume, neurological score, or combined score) in treated animals relative to untreated ischemic controls.

Methodological quality of individual studies was assessed according to published criteria refined in discussion between basic and clinical scientists. These criteria were: peer-reviewed publication; statement of control of temperature; random allocation to treatment or control; blinded induction of ischemia; blinded assessment of outcome; use of anesthetic without significant intrinsic neuroprotective activity; animal model (aged, diabetic, or hypertensive); sample size calculation; compliance with animal welfare regulations; and statement of potential conflict of interests. Each study was given a quality score out of a possible total of 10 points, and the group median was calculated.

For each comparison, the mean outcome for the treatment group and the SDs in treatment and control groups were expressed as a proportion of the outcome in the control group, and the effect size (the difference between the treatment and control groups) and its standard error were calculated. Data were aggregated using a weighted mean difference method; because of anticipated heterogeneity between studies, we used the random effects model of DerSimonian and Laird, in which the weighting given to individual comparisons depends on the variance within those comparisons and on overall heterogeneity (see supplementary material, available online at http://stroke.ahajournals.org). This is a generally more conservative technique than fixed-effects meta-analysis.

When a control group served >1 experimental group, the number of observations in that control group was, for the purpose of the meta-analysis, divided by the number of experimental groups served.

To explore the impact of study characteristics on estimates of effect size, we then performed a stratified meta-analysis with experiments grouped according to: methodological score; use of aged, diabetic, or hypertensive experimental animals; anesthetic used; whether the data had been published in full or in abstract; permanent or temporary ischemia; outcome measure; route of drug delivery; and species and gender of animal used. The significance of differences between n groups was assessed by partitioning heterogeneity and by using the distribution with n degrees of freedom (df).

**Results**

Electronic searching identified 157 publications, of which 14 described experiments reporting the effect of nicotinamide in focal cerebral ischemia in which the outcome was expressed as a volume of infarction or a neurological score, and hand-searching identified a further 4 publications. There were 10 full articles and 8 abstracts. Four abstracts described work that was also described in full articles, and 1 further abstract has been published in full since the search was performed. This meta-analysis is therefore based on data from 11 full articles and 3 abstracts.

Study characteristics are shown in the Table. No study described a sample size calculation or disclosed a potential conflict of interest, even though some studies included in
authorship individual(s) holding a patent for the use of nicotinamide in ischemic stroke. Random allocation to treatment group and blinded assessment of outcome were described in 3 studies each, and 2 studies reported that ischemia was induced by an investigator blinded to treatment allocation. The median quality score (see Methods) was 3.5 (range 0 to 7).

The global estimate of the effect of nicotinamide was 0.287 (95% confidence interval 0.227 to 0.347, \( P<0.00001 \)), an improvement in outcome of \( \approx 30\% \) (Figure 1). There was substantial statistical heterogeneity (\( \chi^2=207, \text{df}=70, P<0.00001 \)), consistent with substantial biological heterogeneity between the studies. Doses of nicotinamide from 100 mg/kg to 750 mg/kg were significantly protective, as were treatments beginning 90 minutes to 6 hours after the onset of ischemia (Figure 2). Although the lack of efficacy beyond 6 hours was expected, an early limit to the time window for efficacy was not. This was observed even when drug was administered before middle cerebral artery occlusion and therefore cannot be explained by a failure of drug to reach brain. This suggests that the concept that neuroprotective drugs have maximum efficacy when administered as soon as possible after stroke onset may need revised, and that some drugs may be most effective if administered later in the course of stroke.

Nicotinamide was less effective in animals with diabetes or hypertension (0.218, 0.131, to 0.304) than in healthy animals (0.300, 0.232 to 0.367; \( \chi^2=7.4, \text{df}=1, P<0.01 \)). There was a nonsignificant increase in efficacy in studies using ketamine anesthesia (0.414, 0.276 to 0.553 versus 0.265, 0.200 to 0.331; \( \chi^2=2.5, \text{df}=1, P=0.11 \)). Comparisons published in abstract only gave a significantly lower estimate of effect size (0.162, 0.066 to 0.258) than those subjected to peer review and published in full (0.306, 0.241 to 0.371; \( \chi^2=12.2, \text{df}=1, P<0.001 \)). More generally, comparisons from studies scoring highly for methodological quality tended to give a more precise estimate of the global estimate than those from low-quality studies (Figure 3).

**Figure 1.** Point estimate and 95% confidence intervals for global estimate and each of 71 comparisons ranked by effect size. Effect size is the improvement in treated animals expressed as a proportion of the outcome in control animals. The diamond indicates the global estimate and its 95% confidence interval. The solid vertical line marks where treatment and control are equal. The size of each point reflects the weight of that comparison in the meta-analysis. For a more detailed version of this figure, see supplementary material (available online at http://stroke.ahajournals.org).

**Figure 2.** Point estimates and 95% confidence intervals of effect size for (a) dose and (b) time to treatment. Where the 95% confidence interval does not reach the solid vertical line, outcome is significantly different from control \( (P<0.05) \). The text indicates effect size and 95% confidence interval; number of comparisons and of animals contributing to each point; \( \chi^2 \) test, and probability of observed heterogeneity. The size of each point reflects the weight of that comparison in the meta-analysis.
Nicotinamide was more effective when cerebral ischemia was temporary (0.367, 0.262 to 0.476) than when permanent (0.238, 0.171 to 0.306; $\chi^2 = 11.0$, df = 1, $P<0.001$), and after intravenous (0.313, 0.234 to 0.391) rather than intraperitoneal (0.262, 0.169 to 0.356) administration ($\chi^2 = 4.8$, df = 1, $P<0.05$). Comparisons reporting both infarct volume and neurological score gave a higher estimate of effect size (0.392, 0.287 to 0.478) than those reporting infarct volume (0.250, 0.173 to 0.327) or neurological score (0.074, 0.039 to 0.186) alone ($\chi^2 = 30.1$, df = 2, $P<0.0001$). Nicotinamide was effective in rats (0.301, 0.245 to 0.372) but not in mice (0.074, 0.039 to 0.186) ($\chi^2 = 20.3$, df = 1, $P<0.0001$), and although outcome in male animals (0.297, 0.232 to 0.362) was not significantly higher than that in females (0.204, 0.048 to 0.360; $\chi^2 = 2.4$, df = 1, $P=0.12$), a biologically important effect cannot be excluded. In a post hoc analysis, stratification by animal model suggested that efficacy was greater in surgical occlusion than in photothrombotic or filament occlusion models ($\chi^2 = 37.3$, df = 3, $P<0.0001$).

**Discussion**

We have shown a robust neuroprotective effect of nicotinamide in experimental stroke. Although the effect is modest compared with that reported for many other drugs in single studies, this is the first systematic meta-analysis of any neuroprotective drug in experimental stroke. No true comparison with other drugs is possible, as yet, because such rigorous assessments have not been available.

Nava-Ocampo et al.\(^{30}\) describe a meta-analysis of glutamate release inhibitors in experimental stroke. However, their search strategy was limited to MEDLINE. Although they, too, extracted data for individual comparisons, they excluded 17 of 47 comparisons to attain statistical heterogeneity, and they do not describe the meta-analysis technique used. The systematic review of nimodipine in experimental stroke by Horn et al.\(^{8}\) used a limited search strategy that would have missed 4 of our 14 studies. Because we disaggregated data into individual comparisons, we have been able to explore the effect of drug dose and time to treatment rather than simply a global effect of drug. Finally, grouping according to study characteristics has allowed, for the first time, a systematic exploration of the impact of study design on effect size.

Although the stratified meta-analyses were prespecified, results should be interpreted with caution because this is a form of subgroup analysis. Increased efficacy with temporary ischemia is biologically plausible and provides further support for the combination of neuroprotection with thrombolysis.\(^{31,32}\) Experiments using healthy animals, ketamine anesthesia, or male animals may overstate effect size. Ischemic stroke generally occurs in elderly patients with associated medical problems, and our data are consistent with the work of Davis et al showing reduced efficacy for the NMDA receptor antagonist D-CPPene in aged rats.\(^{33}\) Financial and ethical considerations discourage the use of aged rats, but despite the availability of spontaneously hypertensive rats, and of streptozotocin to render animals diabetic, this approach is not widespread. Most models of experimental stroke require anesthesia for the induction of ischemia; some anesthetic agents, including ketamine, have marked intrinsic neuroprotective activity, particularly when administered in...
combination with other drugs, and if it is necessary to use a stroke model requiring anesthesia, these anesthetics should be avoided.

The smaller effect size for studies published only in abstract demonstrates publication bias. Meta-analysis can only consider available data, and groups who have found no effect of nicotinamide may not have published their results at all, even in abstract. It is therefore likely that the impact of publication bias is even larger than we estimate here. Although journals may not favor negative studies, a medium for their publication should be found to avoid further distortion of the literature.

Using random effects meta-analysis, the weighting given to individual comparisons is derived from the variance of data within that comparison and the heterogeneity between comparisons. Alternative weighting systems based on study quality have some attractions, but these are often subjective judgements, and a weighting based on variance reflects sample size and, to a degree, the quality of the data. Studies varied in their methodological quality score, and low-quality score was associated with less precise estimates of the overall observed effect size. We believe that the components of the score have an important bearing on study quality. However, some components are more important than others, and some important components may have been omitted; the development of more sophisticated quality scores, perhaps with weighting of different components, is an important area for future research.

Some components such as publication after peer review, randomization to treatment group, and blinded assessment of outcome are widely accepted. For the rest, we believe that it is important for ischemia to be induced blinded to treatment allocation (or for randomization to occur after the induction of ischemia) to prevent a bias in the severity of the induced infarct. Although it would be best to avoid all anesthesia, some anesthetics have much higher intrinsic neuroprotective activity, and their use is, we believe, relevant to study quality. Although there is no evidence that experiments using aged, diabetic, or hypertensive animals provide a better model of human stroke, given the prevalence of these comorbidities in human stroke, it seems likely that such models may be more relevant. Sample size calculations are uncommon in the animal literature, and most studies are underpowered. We believe that such calculations should be routinely reported. Finally, because the financial interests of authors or sponsors may lead to biased data interpretation, many journals now require a statement of any potential conflicts of interest. However, publications may predate the requirement for such disclosure and, even now, not all journals require such disclosure. This remains an important quality issue.

Despite such concerns, these studies compare favorably with others in the animal literature, although less favorably with clinical studies. There is no fundamental reason why such standards cannot be achieved in basic science. In our laboratory, we have adopted the principles of randomization to treatment group; performance of surgery blinded to treatment allocation; blinded assessment of outcome; minimization of use of anesthetics with intrinsic neuroprotective activity; increased use of hypertensive and diabetic animals; and full reporting of potential conflicts of interest. We recommend that others do the same. Journals publishing studies in experimental stroke should consider the development of guidelines similar to the CONSORT guidelines for clinical studies to act as a force for quality improvement.

We have demonstrated the use of systematic review and meta-analysis in the preclinical assessment of candidate stroke drugs. Extending this approach to other putative neuroprotectants will allow a more systematic assessment of relative efficacy, will generate hypotheses for testing in further animal experiments, will provide robust information about the characteristics of individual drugs, and may provide the basis for a new classification of neuroprotective drugs based on their in vivo characteristics rather than their putative mode of action.

Given the huge number of putative neuroprotective agents, such a classification represents a considerable challenge. We propose a collaborative approach, modeled on the Cochrane Collaboration, to develop a rigorous evidence-based summary of animal experimental stroke data. This would inform the choice of drugs for clinical trial and therefore protect trial participants from exposure to potentially dangerous drugs with limited, and often overestimated, efficacy.

Acknowledgments
We are grateful to Kenneth Maynard and James Adams for the provision of data for inclusion in the meta-analysis. This work was supported by an intramural grant from the National Health and Medical Research Council of Australia. M.R.M. acknowledges the support of the Brain and Spine Foundation and the Wellcome Trust.

References


Pooling of Animal Experimental Data Reveals Influence of Study Design and Publication Bias
Malcolm R. Macleod, Tori O'Collins, David W. Howells and Geoffrey A. Donnan

*Stroke*. 2004;35:1203-1208; originally published online April 1, 2004;
doi: 10.1161/01.STR.0000125719.25853.20

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2004 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/35/5/1203

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in *Stroke* can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to *Stroke* is online at:
http://stroke.ahajournals.org//subscriptions/