COMMENTARY

Stroke research at a road block: the streets from adversity should be paved with meta-analysis and good laboratory practice

Ulrich Dirnagl[1] and Malcolm R Macleod[2]

[1]*Departments of Neurology and Experimental Neurology, and Center for Stroke Research Berlin, Charite University Medicine, Berlin, Germany, and* [2]*Department of Neurology, NHS Forth Valley, Stirling, and Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK*

In this issue, Bath *et al.* use state of the art meta-analytical tools to address the pressing question of why NXY-059, a compound at the time considered to fulfil all the recommendations for the evaluation of preclinical data regarding neuroprotective drugs, has failed clinically. They demonstrate quantitatively that a negative publication bias existed, that the compound was indeed neuroprotective in experimental stroke, but that bias may have resulted in an overestimation of efficacy, and that efficacy in healthy, male, adolescent animals is a poor predictor of success in clinical trial. The study contains important messages for researchers, journal editors, the pharmaceutical industry and science policy makers. Bias is both prevalent and relevant in experiments modelling human stroke. Simple measures can reduce, perhaps substantially, the impact of such bias. The decision to proceed to clinical trial should be based on a thorough and systematic review of the animal data.
British Journal of Pharmacology (2009) **157**, 1154–1156; doi:10.1111/j.1476-5381.2009.00211.x

This is a Commentary on the Research Paper in this issue of *BJP* entitled *Effects of NXY-059 in experimental stroke: an individual animal meta-analysis* (Bath *et al.*, pp. 1157–1171). To view this article visit http://www3.interscience.wiley.com/journal/121548564/issueyear?year=2009

Translating the numerous promising therapeutic agents emerging from basic and preclinical research into effective therapy for patients is a major challenge for modern medicine. Among many other biomedical fields, this 'translational roadblock' is plaguing the search for an effective treatment for acute stroke (Endres *et al.*, 2008). Hundreds of neuroprotectants have made the transition from bench to randomized clinical trials, only to demonstrate futility. Despite billions of dollars spent, not one pharmacological agent, improving outcome in stroke patients by acting on brain cells, has been approved by the Food and Drug Administration (FDA). The only pharmacological stroke therapy with level I evidence is the recanalization of the ischaemic territory by i.v. thrombolysis within 4.5 h after symptom onset (Hacke *et al.*, 2008). After the recent negative outcome of one of the largest neuroprotection trials ever performed in stroke patients, testing

the efficacy of the free radical scavenger NXY-059 (SAINT II, Shuaib *et al.*, 2007), a 'nuclear winter' for translational stroke research has been proclaimed.

A number of reasons may account for the dismal failures of bench-to-bedside translation in stroke. Understanding and eliminating these is of utmost importance if we are to improve the survival and quality of life of patients with this common and devastating disorder. A number of articles have speculated why there is such a striking difference between apparent efficacy in preclinical as opposed to clinical studies: among other explanations, the doses effective in animals may not have been achieved in humans due to treatment-limiting side effects; the delays to treatment in clinical trials may have been substantially longer than those used when efficacy was observed in animals; the outcomes measured in animals (mostly reduction in infarct volumes) may not be directly relevant to outcome measures in human trials (usually death or disability); or animal experiments may simply not model efficacy of stroke drugs with sufficient fidelity to be useful. Recently it has been suggested that bias in animal experiments may have led to inflated estimates of the efficacy of such drugs in animals. Such bias could come from a number

Correspondence: Departments of Neurology and Experimental Neurology, and Center for Stroke Research Berlin, Charite University Medicine, Berlin, Germany.
E-mail: ulrich.dirnagl@charite.de

of sources; the internal validity of such studies may be compromised by a lack of randomization (selection bias) or blinding (performance or ascertainment bias), by a failure to report excluded animals (attrition bias) or by small sample sizes, low statistical power and therefore poor positive predictive value. The external validity of such studies may be compromised by the lack of inclusion of animals with co-morbidities, or old animals, or female animals; or by publication bias. Indeed, several meta-analyses have provided quantitative evidence that study quality has confounded evidence of efficacy in preclinical stroke research (Crossley *et al.*, 2008; Macleod *et al.*, 2008).

In this issue of the *BJP*, the NXY-059 Efficacy Meta-analysis in individual Animals with Stroke (NEMAS) investigators (Bath *et al.*, 2009) use state of the art meta-analytical tools to address the pressing question of why NXY-059, a compound at the time considered to fulfil all the Stroke Therapy Academic Industry Roundtable (STAIR) recommendations for the evaluation of preclinical data regarding neuroprotective drugs (Stroke therapy academic industry roundtable, 1999), has failed clinically (Shuaib *et al.*, 2007). Two aspects of the study make it outstanding: it is the first meta-analysis in the field using individual animal data, improving its power and allowing subgroup analysis as well as investigation of interaction between variables. Perhaps more importantly, the study was able to include large sets of unpublished data, in particular from AstraZeneca, the company developing the compound. The importance of this approach was recently demonstrated by Kirsch *et al.* in an analysis of the clinical efficacy of new-generation antidepressants. Using all data submitted to the FDA (including unpublished data), they showed that these widely prescribed medications are only superior to placebo in the most severely depressed patients (Kirsch *et al.*, 2008). A previous report had contended that 'the preclinical evaluation of NXY-059 lacked strenuous testing and was not shown to reproducibly lead to robust protection in extended time windows in clinically relevant stroke models, at several different academic research laboratories' (Savitz, 2007). In addition, a meta-analysis of published preclinical studies on NXY-059 (Macleod *et al.*, 2008) had demonstrated that the evidence for efficacy in experimental stroke was confounded by study quality, in particular by a lack of randomization and concealment of treatment group allocation. The analysis of the NEMAS investigators, using unpublished data from individual animals, is now able to quantitatively demonstrate that a negative publication bias existed, that NXY-059 was indeed neuroprotective in experimental stroke, but that bias may have resulted in an overestimation of efficacy. Like previous analyses, they clearly demonstrate that efficacy in healthy, male, adolescent animals is a poor predictor of success in clinical trial.

Of course, measures to avoid bias reported in publications may not reflect all of the measures that were taken. For instance, the Marshall papers cited are held to 'clearly state' that temperature was controlled during the procedure; in fact they report that 'after surgery, animals were placed in incubators to maintain body temperature'. No doubt the NEMAS investigators know what actually happened, as some of them co-authored the Marshall papers, but the scoring of *reported* study quality can and should be based only on what is reported. Quality scoring can be even more complex; while Yoshimoto *et al.* (reference see Bath *et al.*, 2009) did indeed not state explicitly that they blinded the induction of ischaemia, they randomized animals to treatment group after the induction of ischaemia, and therefore the induction of ischaemia must have been blinded to treatment group allocation. Because of the central importance of these measures to avoid bias, Bath *et al.* are absolutely correct to assert that each publication should clearly report all measures taken to avoid bias.

Besides exposing methodological weaknesses and raising the issue of overestimated effect sizes, the meta-analysis by Bath *et al.* suggests that NXY-059 was neuroprotective in experimental stroke in various species, including non-human primates. Why then did the clinical trial fail? It may be overly simplistic to argue that the combination of quality bias, publication bias and an inappropriate generalization from healthy young animals to unhealthy old animals is sufficient to explain the lack of efficacy in the human trial. Other factors that may be involved include: (i) the concentration of NXY-059 that was achieved at the presumed site of action in human brain is not known; (ii) the median delay to initiation of treatment in the SAINT II study was almost 4 h, compared with less than 2.5 h in animal studies; and (iii) the clinical trial findings come from a few large adequately powered studies conducted in a highly regulated and audited setting, whereas the animal data come from a larger number of much smaller underpowered studies conducted in a less regulated, less audited environment. There is no direct evidence supporting the most troubling explanation – that animal models currently available do not model human stroke with sufficient fidelity to be useful – but the field is desperately waiting for at least one example of successful translation to demonstrate 'proof of concept'.

The study by Bath *et al.* (2009) contains important messages for researchers, journal editors, the pharmaceutical industry and science policy makers. Bias is both prevalent and relevant in experiments modelling human stroke (Dirnagl, 2006). Simple measures can reduce, perhaps substantially, the impact of such bias. Two important journals in this field (*Stroke*, *Journal of Cerebral Blood Flow and Metabolism*) have recently published 'good laboratory practice' (GLP) requirements for the publication of preclinical stroke studies (Macleod *et al.*, 2009), which include proper reporting (e.g. of animal characteristics, or inclusion and exclusion criteria), sample size calculations, true randomization, allocation concealment, blinded assessment of outcome and reporting of potential conflicts of interest. Meanwhile, the STAIR criteria of 1999 have been updated to include GLP issues (Fisher *et al.*, 2009). Scientific societies, funding bodies and journals must find ways to minimize publication bias (e.g. internet searchable repositories of unpublished high quality neutral datasets) and to ensure that published data is made available for quantitative meta-analysis.

In the meantime, we should consider this not to be a 'nuclear winter' but a time to improve the way we conduct and report animal experiments and the way we design clinical trials. The decision to proceed to trial should be based on a thorough and systematic review of the animal data, considering not only efficacy but also the quality and range of

supporting evidence, the presence of any publication bias and the conditions of maximum efficacy that a clinical trial might seek to recapitulate.

We speculate that the weaknesses in preclinical stroke studies exposed by Bath *et al.* (2009), as well as several other recent papers, are not specific to this field. Indeed, a translational roadblock has been diagnosed in many other fields of modern medicine. By openly discussing these issues, and through the systematic application of meta-analytical tools in preclinical research, the stroke field might serve as a role model for other areas of medicine in which tremendous advances at the bench still wait for successful translation into benefit for patients.

Acknowledgements

Conflict of interest

None.

References

Bath PMW, Gray LJ, Bath AJG, Buchan A, Miyata T, Green AR, on behalf of the NXY-059 Efficacy Meta-analysis in individual Animals with Stroke (NEMAS) investigators (2009). Effects of NXY-059 in experimental stroke: an individual animal meta-analysis *Br J Pharmacol* **157**: 1157–1171.

Crossley NA, Sena E, Goehler J, Horn J, van der Worp B, Bath PM *et al.* (2008). Empirical evidence of bias in the design of experimental stroke studies: a meta-epidemiological approach. *Stroke* **39**: 929–934.

Dirnagl U (2006). Bench to bedside: the quest for quality in experimental stroke research. *J Cereb Blood Flow Metab* **26**: 1465–1478.

Endres M, Engelhardt B, Koistinaho J, Lindvall O, Meairs S, Mohr JP *et al.* (2008). Improving outcome after stroke: overcoming the translational roadblock. *Cerebrovasc Dis* **25**: 268–278.

Fisher M, Feuerstein G, Howells DG, Hurn PD, Kent TA, Savitz SI *et al.* (2009) Update of the Stroke Therapy Academic Industry Roundtable (STAIR) preclinical recommendations. *Stroke* [epub ahead of print].

Hacke W, Kaste M, Bluhmki E, Brozman M, Dávalos A, Guidetti D *et al.* (2008). Thrombolysis with alteplase 3 to 4.5 h after acute ischemic stroke. *N Engl J Med* **359**: 1317–1329.

Kirsch I, Deacon BJ, Huedo-Medina TB, Scoboria A, Moore TJ, Johnson BT (2008). Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* **5** (2): e45.

Macleod MR, van der Worp HB, Sena ES, Howells DW, Dirnagl U, Donnan GA (2008). Evidence for the efficacy of NXY-059 in experimental focal cerebral ischaemia is confounded by study quality. *Stroke* **39**: 2824–2829.

Macleod MR, Fisher M, O'Collins V, Sena ES, Dirnagl U, Bath PM *et al.* (2009). Preventing Introduction of Bias at the Bench. *J Cereb Blood Flow Metab* **29**: 221–223.

Savitz SI (2007). A critical appraisal of the NXY-059 neuroprotection studies for acute stroke: a need for more rigorous testing of neuroprotective agents in animal models of stroke. *Exp Neurol* **205**: 20–25.

Shuaib A, Lees KR, Lyden P, Grotta J, Davalos A, Davis SM *et al.* (2007). NXY-059 for the treatment of acute ischemic stroke. *N Engl J Med* **357**: 562–571.

Stroke therapy academic industry roundtable (Fisher M, Chair) (1999). Recommendations for standards regarding preclinical neuroprotective and restorative drug development. *Stroke* **30**: 2752–2758.