ARE THE CURRENT OUTCOME PARAMETERS IN STROKE CLINICAL TRIALS CLINICALLY RELEVANT?

Kennedy R. Lees
Institute of Cardiovascular and Medical Sciences, University of Glasgow & Western Infirmary, Gardiner Institute, Glasgow, UK

Clinical trials in stroke fall into 3 categories: acute, rehabilitation and prevention trials. We can aslo consider these as either definitive or exploratory. The latter can be largely dismissed since for exploratory trials of any nature, clinical relevance of the outcome measure is of secondary importance: the aim is to answer a specific question. These outcomes can be as specialised and esoteric as the question. In practice, surrogate outcomes such as growth of infarct lesion on MRI, improvement in walking speed or reduction in growth of carotid wall intima-medial thickness may each have relevance to a clinical question. However, none would be adequate on its own to support a change in clinical practice. Thus, we need to consider which outcomes are used in large definitive trials, and to discuss their clinical relevance.

Stroke prevention trials have used various combinations of recurrent stroke, death, myocardial infarction and their combinations such as the "CAPRIE" endpoint - a composite outcome cluster of ischaemic stroke, myocardial infarction, or vascular death - but these combinations are often varied subtly. "MATCH" used the composite of ischaemic stroke, myocardial infarction, vascular death, or rehospitalisation for acute ischaemia. Some trials have chosen to concentrate on only recurrent ischaemic stroke or even on recurrent ischaemic stroke or TIA. Among these, TIA has limited clinical relevance; brief rehospitalisation has dubious importance in its own right; and if ischaemic stroke is reduced at the expense of an equal increase in haemorrhagic stroke it could be regarded as a Pyrrhic victory.

Rehabilitation trials have used diverse endpoints, with almost as many variants as there have been trials. Clinical relevance of many of the endpoints has been limited; however the majority of these trials were small. Larger rehabilitation trials that are intended to be definitive, such as AVERT, have sensibly chosen to use a validated functional measure in the form of the modified Rankin scale.

The principal outcome measure used in acute trials – but which applies equally well in rehabilitation trials – is the modified Rankin scale (mRS). This may even be included as an arbiter of relevance for recurrent stroke in prevention trials. The modified Rankin scale is often supported by a neurological scale – usually the NIHSS – and by a range of secondary measures. Change from baseline in NIHSS can have limited clinical relevence when arbitrary scores must be assigned to patients who die or are lost to follow up, since a bimodal distibution fo outcomes is generated that offers low statistical power.

The strengths of the mRS are its ordinal nature, ie the categories have a progressive order from best to worst, its favourable statistical properties since similar proportions of patients generally fall into each mRS category, its ease and reliability of measurement in all patients irrespective of language impairment, cognitive defict etc, its known inter-rater reliability and its relevance in every culture and region. Further, mRS is associated with healthcare resource use and costs, and so can be used to estimate cost-effectiveness. The principal criticism of mRS as a stroke outcome is that it emphasises motor recovery and physical functioning at the lower end of the scale, at the expense of cognitive features. Despite this, it has not so far been possible to demonstrate that the addition of cognitive measures such as COG-4 to mRS improves discrimination of treatment effects, perhaps because large strokes usually cause both cognitive and physical features that are inextricably linked.

Clincial relevance may also be considered in relation to the analytic approach applied to the outcome measure. Many of the thrombolysis trials have analysed mRS by dichotomisation, categorising patients according to excellent recovery (mRS 0 or 1) versus unfavourable (mRS 2-6) but sometimes a different cut-point is chosen, eg mRs 0-2 in some rtPA trials or even mRS 0-4 in trials of hemicraniectomy. The arbitrary nature of these splits, the inconsistency among trials and the chance of spurious conclusions about the success or failure of trials undermine the clinical relevance of this approach.

Ordinal analysis, in which all grades of mRS are considered together, is statistically more powerful, is more reliable, more informative and less dependent on case mix. Concerns about the need to satisfy statistical assumptions are unfounded if robust approaches are used such as the Wilcoxon-Mann-Whitney test and the Mann-Whitney measure, and these yield exact measures of probability and of effect size that can readily be converted to other measures of effect size such as the odds ratio or number needed to treat for dichotomous or ordinal outcomes.

Length of hospital stay is sometimes included as an outcome measure but suffers major disadvantages. Its relevance is undermined by the diverse care pathways and definitions of hospital versus rehabilitation or care home setting and by the spurious "improvement" that occurs with early mortality. A clinically relevant alternative is "home time" in which the number of nights that a patient spends in his/her own or relative's home among the first 90 days after stroke gives the obverse of length of stay. This measure tracks closely with healthcare costs and mRS score, and can reflect treatment effects, but still shows regional variation.

Thus, reliable and clinically relevant outcomes can be assessed in stroke trials using ordinal analysis of the mRS, and the main question for future is whether the addition of cognitive testing yields clinically relevant information or remains redundant.