"Inadequate outcome measures are the biggest impediment to successful clinical trials in progressive MS". **YES**

Jeremy Chataway

Queen Square Multiple Sclerosis Centre, Department of Neuro-inflammation, UCL Institute of Neurology, University College London, London, UK

jeremy.chataway@uclh.nhs.uk

The closer progressive multiple sclerosis (MS) is approached, the more awkward it is to define and the more slippery it is to measure. Of course on one level it is easy – if 5 years previously a person affected by MS can walk or see or balance normally, but now they walk with a stick or have a visual acuity of 6/60 or are falling over, then it is clear (to anyone) that progression has occurred, whichever yardstick is used. No, the problem is that progression is (generally) slow, it is multi-dimensional and the time frequency of such very hard clinical end-points is too low to be of value in the 2-3 year life span of trials in progressive MS.

Returning to the title statement above, there are in particular two areas which must be thought about, since they are fundamental to the issues of outcome measures in progressive MS and a brief revision of these starts to clarify the issues at hand.

The first is the science of clinical measurement. Turn to any medical statistics book[1] and the key concepts include: biological variation, the skill of the observer, the interaction between the observer and the subject, and the precision with which the data are recorded. To take an everyday example: systolic blood pressure could be recorded to the nearest 10 mm Hg, or the nearest 5 mm Hg or the nearest 1 mm Hg. Some investigators would use Korotkov sound four, others five. Imprecision, or non-standardisation, has the capacity to greatly alter the final outcome. Continuing with blood pressure, it varies from day to day, and season to season. Replication and quantification of these rhythms is necessary standard statistical practice.

In the measurement of progressive MS, by whatever means, the opportunities for error abound. Examples that easily come to mind would be the effect on fatigue on a walking distance; depression on a test of cognition; whether the functional electrical stimulator (FES) is consistently on or off when the 25 foot walk is recorded; the blinding of the subject, the assessor and all other trial personnel.

The second is the properties of the outcome measures themselves, in particular the rating scales. The psychometric issues in neurology have been well reviewed in detail by experts in the field and only a few areas are used for illustration here, without going into the mathematical theories needed to enhance rating scale analysis, such as latent trait theories.[2] In the Ashworth spasticity scale there are six categories ordered in increasing spasticity: from 'no increase in muscle tone' to 'affected part rigid in flexion or extension'. They are ordinal (ranked) assignments and the absolute interval differences are unknown and are likely to have different meanings at different portions of the scale. Another major concern is that of construct validity, whether the scales actually measure the health concerns that they purport to measure. To spare the reader, I thought I'd leave out the Expanded Disability Status Scale (EDSS), which has been raked over many times before. Indeed the properties required were discussed in the *MSJ* 20 years ago and included: sensitivity, reliability, validity, independence of dimensions, ceiling effects.[3] As we know, the EDSS comes up short.

Yet the paradox is, that even when a concerted attempt is made to overcome these issues it may not succeed. The Multiple Sclerosis Functional Scale (MSFC), which is reliable, covers three major domains of interest, yields a single score which can be compared across studies, and has good correlations with indices such as MRI and quality of life. Despite all of these properties and their promise, when it comes down to it, as an outcome scale, in a trial, the MSFC was rejected by the regulators. Reasons given were that the summary Z score was seen as dimensionless and abstract.[4] Let down again by the outcome scale.

Outcome problems are not confined to clinician or patient observed scales. As we were prompted recently in the *MSJ*, there are numerous traps for the unwary, with so called objective measures such as MRI in progressive MS.[5] Brain atrophy is of course very current in phase 2 trials in progressive MS, but there are a number of issues which will add error to the result and therefore impede the outcome measurement. We are reminded of the effects of age, hydration status and drug-induced pseudoatrophy. Moreover as a worked example, the tool that is used, MRI, in the context of a scanner upgrade can change the atrophy rate by about 1.5%.[5] Frightening, when considering a background rate of whole brain atrophy in progressive MS of about 0.5%/year. The confounding effects on more advanced MRI parameters, particularly in the heterogeneous multi-site environments of clinical trials will of course magnify the problem.

Moreover the difficulties are not just confined to measuring human beings with progressive MS. In the parallel world of animal models, a similar suite of concerns and worries exist with outcomes and their measurement. In one survey of 2600 reports, which included Experimental Autoimmune Encephalomyelitis (EAE), a blinded assessment of outcome was reported in about 30% of studies, with the rate for EAE being only 20%.[6]

I think it is clear from what has been said above, that from the basic to the complex outcome, from animal to human, in trying to solve a problem such as progressive MS, there is plenty that can and does go wrong. The numerous other hurdles, recruitment, drug choice and trial design, though problematic, are of a lower order of magnitude.

References

1.Bland M. An introduction to medical statistics. Oxford Medical Publications. 2nd edition.

2.Hobart J, Cano S, Zajicek J, Thompson A. Rating scales as outcome measures for clinical trials in neurology: problems, solutions and recommendations. *Lancet Neurol* 2007;**6**:1094-1105.

3.Whitaker J, McFarland H, Rudge P, Reingold S. Outcomes assessment in multiple sclerosis trials: a critical analysis. *MSJ* 1995;**1**:37-47.

4.Cohen J, Reingold S, Polman C, et al. Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects. *Lancet Neurology* 2012;**11**:467-76.

5.Barkhof F. Brain atrophy measurements should be used to guide therapy monitoring in MS – NO. *MSJ* 2016;1-2.

6. Macleod M, McLean A, Kyriakopoulou, et al. Risk of bias in reports of in vivo research: a focus for improvement. *PLOS Biol.* 2015;**13** (10):e1002273.

Conflicts of interest