MULTIPLE SCLEROSIS JOURNAL | MSJ

Multiple Sclerosis Journal
17(1) 81–88

DOI: 10.1177/1352458510382129
msj.sagepub.com
SAGE

A novel adaptive design strategy increases the efficiency of clinical trials in secondary progressive multiple sclerosis

Jeremy Chataway[1,2], Richard Nicholas[2], Susan Todd[3], David H Miller[1,4], Nicholas Parsons[5], Elsa Valdés-Márquez[3], Nigel Stallard[5] and Tim Friede[5]

Abstract

Background: Adaptive seamless designs (ASDs) have been proposed to test multiple candidate compounds using an interim decision point which allows potentially effective therapies to be taken into the next design stage and to be assessed using a phase III outcome.

Objective: To determine whether ASDs are feasible in secondary progressive multiple sclerosis (SPMS) and to compare them with conventional trial designs.

Methods: We develop an innovative adaptive trial design for SPMS, which builds on recent developments in statistical methodology. A literature search and individual clinical datasets were used to inform a framework to run simulations to evaluate the proposed design.

Results: ASDs are feasible in SPMS with MRI informing an interim decision point and Expanded Disability Status Scale (EDSS) as the final disability endpoint. Furthermore ASDs are more efficient than conventional designs with sample size savings of up to 40%. Sample sizes of 1000–1250 patients are sufficient to test up to four experimental treatments. Controlled recruitment is important to realize the full benefits of ASDs.

Conclusions: Although more complex in design, ASDs have the potential to be more efficient and more powerful than conventional designs.

Keywords

adaptive seamless design, secondary progressive multiple sclerosis, interim analysis, biomarker, computer simulations

Date received: 28th March 2010; revised: 8th July 2010; accepted: 25th July 2010

Introduction

Despite the success of increasingly sophisticated compounds in modulating the early phases of multiple sclerosis (MS), both delaying the conversion of clinically isolated syndromes to the full disease state[1] and reducing relapse frequency,[2] the pivotal problem of altering an established gradient of progression (primary or secondary), as the disease evolves from an inflammatory to an axonal destructive state,[3] remains. There is no shortage of candidate therapies and the clear challenge is the implementation of an efficient trial design which can relatively quickly determine whether progression can be modulated or not. Classical development programmes take in excess of 10 years from phase I inception to completion of phase III,[4] and ultimately will only determine whether the single drug being studied is of value.

Typically disease-modifying trials in MS are conducted by comparing a single novel treatment with a control treatment, e.g. placebo. As is well established in relapsing–remitting MS, in phase II this comparison is done using MRI outcomes with relatively short

[1]National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Foundation Trust, Queen Square, London, UK.
[2]Imperial College Healthcare NHS Trust, London, UK.
[3]Applied Statistics, University of Reading, Reading, UK.
[4]Department of Neuroinflammation, Institute of Neurology, University College, London, UK.
[5]Warwick Medical School, The University of Warwick, Coventry, UK.

Corresponding author:
Professor Tim Friede, Georg-August-University Göttingen, Department of Medical Statistics, Humboldtallee 32, D-37073 Göttingen, Germany
Email: tim.friede@med.uni-goettingen.de

follow-up, whereas in phase III long-term disability outcomes are used. The efficiency of such programmes can potentially be increased by allowing a number of treatments to be tried concurrently, with poorly performing treatments dropped quickly as the study proceeds. In order to maintain the scientific integrity of such an approach, modern statistical techniques are required. These so-called flexible or adaptive two-stage trial designs with treatment selection[5–7] commence with a multi-arm stage, dropping treatments for futility at an interim analysis, whilst taking promising compound(s) through to the final stage (Figure 1). A confirmatory phase III analysis can then be conducted at the end of the study using combined information from the first and second stages. Such an approach is more efficient than the traditional phase II followed by phase III approach.[6] Previously, such designs have largely been described in settings where the interim treatment selection is based on the final outcome, which is of course unrealistic in progressive MS.

The key requirement for adaptive designs to work in this setting is the existence of a short-term outcome measure for interim decision making and a long-term outcome sensitive to treatment changes. The short-term outcome must be *biologically plausible* in that it gives some indication as to whether the mechanism of action of a test treatment is working as anticipated at the interim analysis, and thereby allows futility stopping. However, it would not necessarily predict either the size of the treatment effects on the disability outcome or the disease progression of an individual patient as a *surrogate marker* would.[8]

Here we present and then demonstrate the feasibility of an adaptive seamless design (ASD) in secondary progressive MS (SPMS) that uses currently available phase II MRI outcomes for futility stopping. Furthermore, we show that the use of such a trial based on currently available disability measures and the known relationship between late disability and early MRI assessment can lead to substantially increased efficiency in terms of number of subjects, time and resources required.

Methods

Literature review and individual patient datasets

To inform the study parameters, both a comprehensive literature review (up to 28 April 2008) and individual clinical dataset appeals through the UK MS Society Clinical Trials Network were undertaken. As some studies contain a mix of populations the literature search included all publications with reference to progressive disease. The following databases were searched: Medline, Pubmed/Premedline, Embase, Cumulative Index to Nursing and Allied Health Literature, Cochrane Central Register of Controlled Trials and Database of Systematic Reviews, Database of Abstracts of Reviews of Effectiveness, Health Technology Assessment and Google Scholar. We used the following inclusion criteria.

1. Design: randomized controlled trials (RCTs), systematic reviews, natural history cohorts, epidemiological or observational studies.

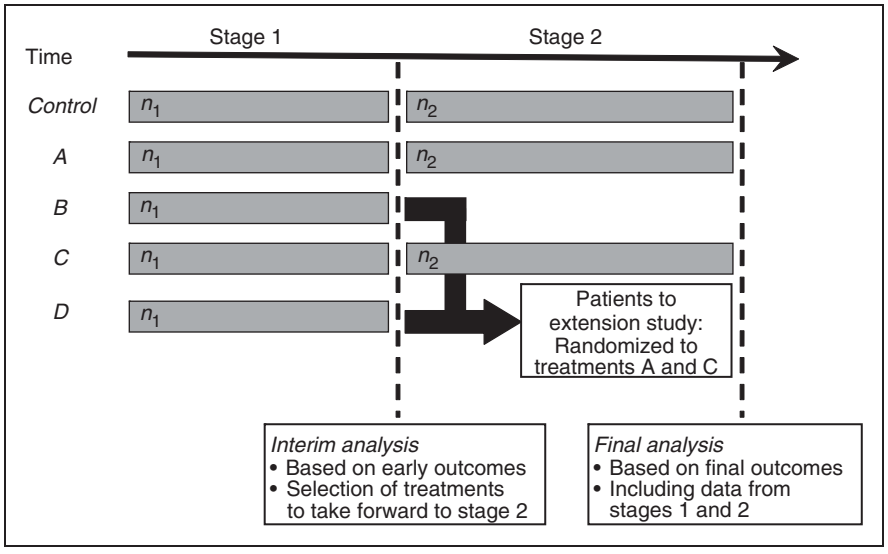

Figure 1. An adaptive seamless design (ASD) for four test treatments arms (and a single control) in stage 1 with stopping for futility at an interim analysis, based on an early outcome, and a final analysis including combined information from stage 1 and 2 based on a final outcome measure.

2. Clinical outcome: Expanded Disability Status Scale (EDSS),[9] Multiple Sclerosis Impact Scale 29 (MSIS-29)[10] or Multiple Sclerosis Functional Composite (MSFC).[11,12]
3. Any MRI outcome (e.g. T2 burden, number of enhanced lesions, T1 black holes, atrophy rate).
4. Size: $\geq$150 subjects.

Studies were then excluded based on: duration <1 year, non-progressive disease phase, inappropriate outcome (e.g. health economics), cross-sectional studies, review/abstract only. Abstracts were reviewed by two neurologists (JC, RN) independently. In any studies where there was disagreement, the full paper was obtained and discussed until consensus was achieved. Individual patient datasets were obtained through appeals for clinical datasets.

Statistical methods

To apply ASDs in MS, two components are required: first, methodology is needed to develop a design for the trial; and, second, a framework is necessary to facilitate the conduct of simulation studies that evaluate how the trial might run in practice. The design was based on the 'combination test approach' to ASDs.[6,7] The combination function was used to combine stagewise *p*-values from the two stages of the trial and application of the closed test principle was used to control the overall type I error rate. The method was extended, as part of this work in MS, to incorporate early outcomes for interim treatment selection.[13] The framework for the simulation studies was based on clinical scenario evaluation[14] adapted for MS.[15]

Key to the evaluation process is the specification of a range of disease-specific features (Figure 2).

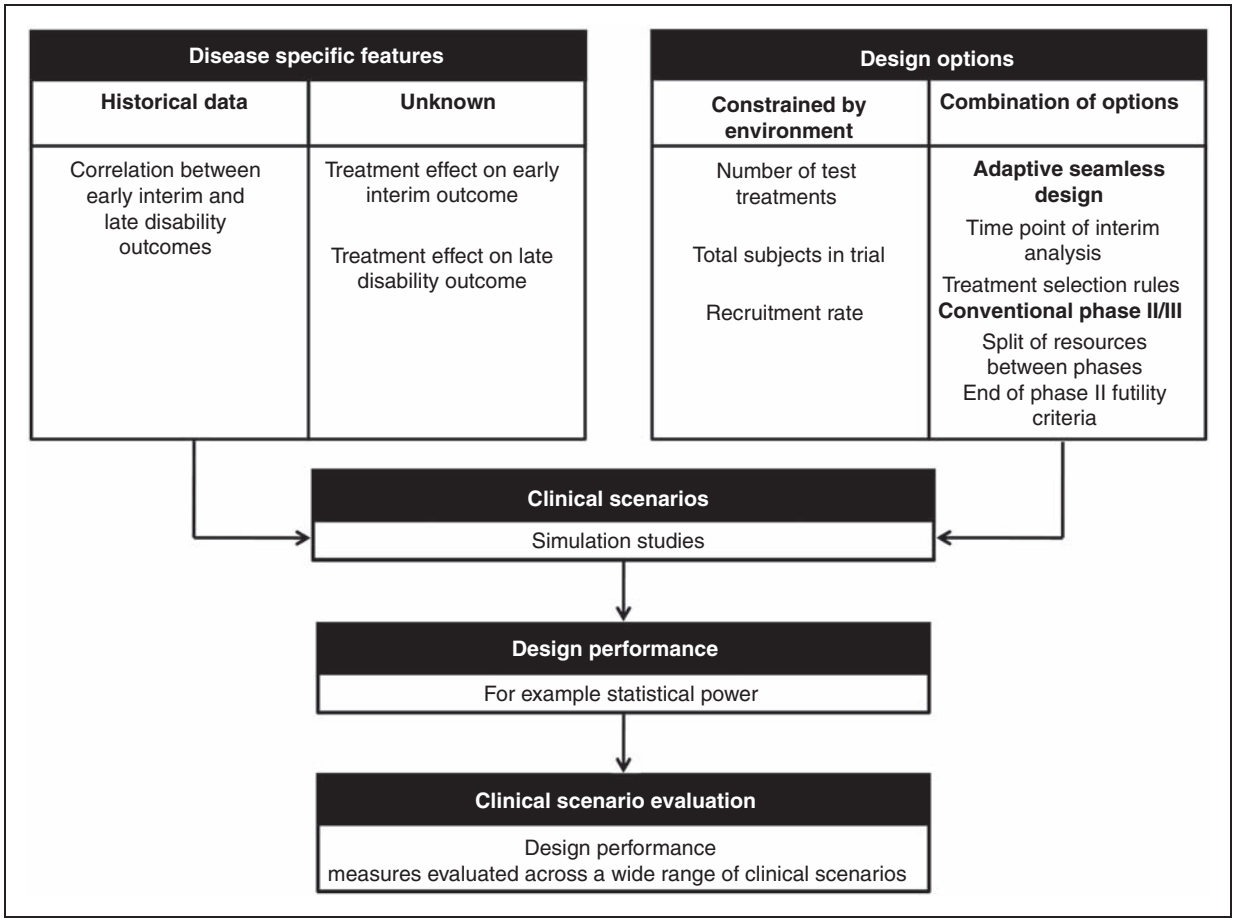

Figure 2. A clinical scenario evaluation paradigm identified the disease-specific features that relate to progressive multiple sclerosis (MS), with historical data informing the model where available, and unknown features being assumed with the support of expert opinion. Design options either constrained by the healthcare environment or statistical decisions arising from optimization. Combining these elements allowed the development of clinical scenarios which were then simulated. The performance of each clinical scenario is judged allowing coherent comparison with traditional designs, leading to the development of pragmatic recommendations.

These include likely treatment effects and correlations between early and late outcomes in the trial. Design options such as the number of treatments, total subjects and recruitment rate are, in reality, constrained by the healthcare environment. However, the optimal combination of 'statistical' options arises from the application of specialist methodology and evaluation under different scenarios. These include the timing of the interim analysis, rules for treatment selection and the allocation of resources between the trial stages. Combining disease-specific features and the design options generates clinical scenarios. Running simulations allows a comparative evaluation of the design performance of each scenario by determining the statistical power. Simulations were undertaken using functions developed by the project team and now freely available as package *asd* at the comprehensive R archive network (CRAN: http://www.cran.r-project.org/). Briefly, early and final outcome data are randomly generated for known stage 1 group sizes, correlations and effects for assumed normally distributed test statistics, and treatment selection at interim analysis made using one of the available decision rules (eg. select best treatment only). Stage 2 final outcome data are also randomly generated for known stage 2 group sizes for the selected treatment(s) and stagewise *p*-values combined and the closed test principle applied to control the overall type I error rate.[6,7] The simulation process is repeated many times using different randomizations and counts made of the number of times individual elementary hypotheses are rejected; overall trial power follows from counting how often false null hypotheses are correctly rejected. Full details of the statistical methods used are given elsewhere[13] or are available from the corresponding author upon request.

Results

Can adaptive designs be applied to secondary progressive MS?

The primary outcome of any phase III trial in SPMS will have to reflect changes in disability, likely measured by EDSS.[16] Our review of the literature identified 46 papers covering 29 studies which satisfied the specified inclusion criteria, including (sometimes multiple) publications from RCTs ($n = 14$), uncontrolled trials ($n = 1$), observational studies ($n = 2$), observational longitudinal studies ($n = 9$), placebo database ($n = 1$) and meta-analyses ($n = 2$) with a total of 16,107 patients, 11,146 of which had progressive disease (both primary and secondary) (see Supplementary Table 1). All studies identified reported (E)DSS. Studies which used either entirely or partly SPMS patients were then considered, of which 10 RCTs were identified, with patient follow-up of up to 1 year ($n = 2$), of 2 years ($n = 4$) and of 3 years ($n = 4$) with a total of 4427 patients. Individual patient data were also obtained from seven neuroscience centres: longitudinal natural history studies ($n = 6$) and RCT control groups ($n = 1$) (see Supplementary Table 2), all of which reported EDSS, confirming again that this is clearly the most common endpoint used and would be appropriate for an ASD. Measuring EDSS at 3 years therefore appears to be a reasonable and feasible choice.

Recruitment rates into a trial are central to implementing an ASD. For example, recruiting at a constant rate over a period of 2 years and using a 12-month early endpoint, an interim analysis conducted at 18 months would yield data on 25% of the total number of patients (solid line in Figure 3). At the 18-month time point, however, another 50% of the total number of patients will have been recruited into the trial. The optimal recruitment strategy is not to recruit at a constant rate, but have fast recruitment initially, followed by slower recruitment up to the interim analysis, after which time recruitment is accelerated towards the end of the recruitment period (dashed line in Figure 3).

Imaging endpoints exist that make small-scale phase II clinical trials in progressive MS possible.[17,18] These include non-enhancing T1 hypointense lesions, changes in whole-brain volume (atrophy) and grey matter fraction, as well as more established measures such as T2 lesion volume (T2LV). Recent sample size calculations for whole brain atrophy suggest minimum 2-year sample sizes/arm of 32 (50% treatment effect at 80% power); although realistically total group sizes will be closer to 100.[18] However, despite these promising new measures, our search demonstrated that, currently, robust longitudinal (3-year) data is only available for T2LV making this the prime candidate for use in ASDs. Ultimately, it would be anticipated that this would be combined with other measures such as atrophy.

Disease-specific features

The literature search confirmed that there was no effective treatment for progressive MS.

Based on data from 41 patients, the recent sample size calculations for whole brain atrophy referred to above, considered a total of 24 scenarios leading to sample sizes which correspond to standardized effect sizes ranging from 0.10 to 0.76 with the centre 50% lying between 0.27 and 0.53.[18] Furthermore, effect sizes from 0.1 to 0.4 are generally considered as small to medium effect sizes[19] and thus this range was chosen to assess power for alternative ASDs.

From the literature search, just three studies in SPMS were found which presented both detailed

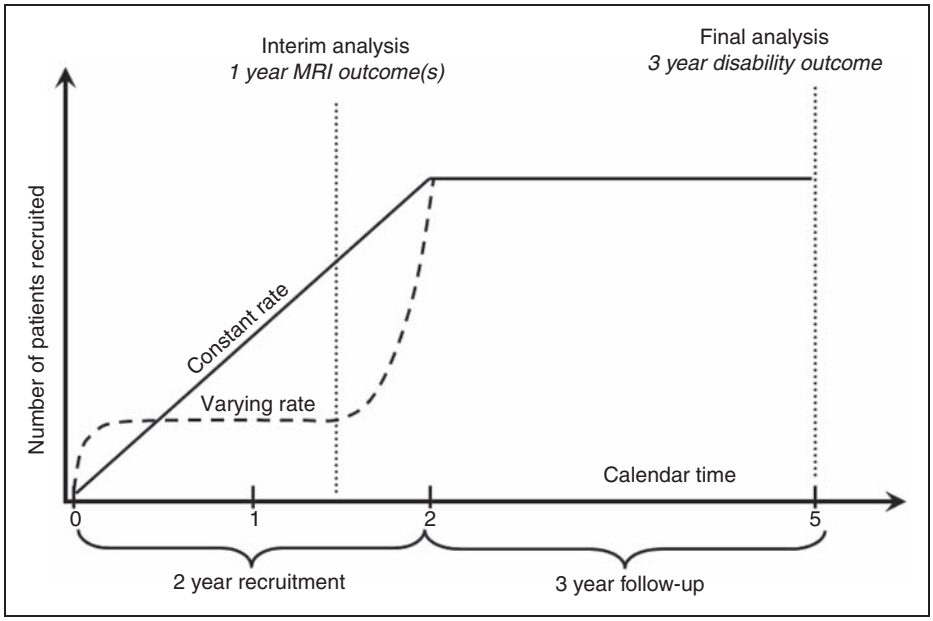

Figure 3. Recruitment in adaptive seamless design (ASD) takes place over 2 years, with an interim analysis at 18 months based on MRI data, where one or more experimental treatments are selected. The final analysis is based on a 3-year disability outcome (Expanded Disability Status Scale).

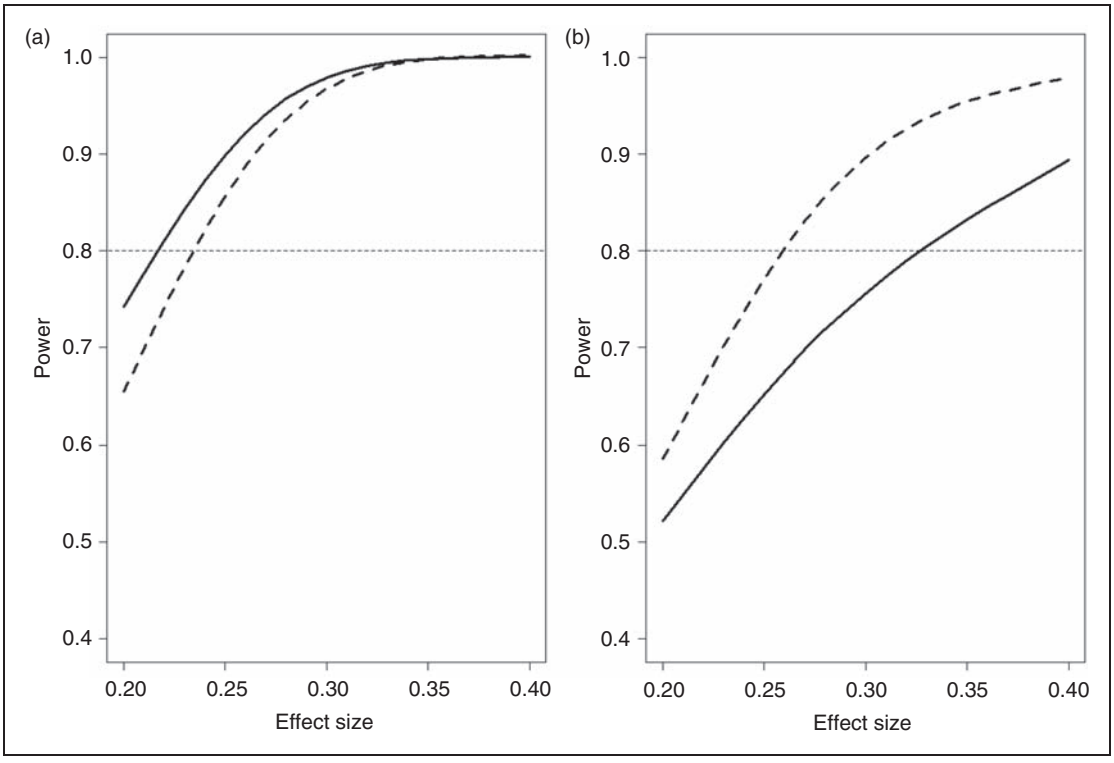

Figure 4. Power to reject to the null hypothesis for three test treatments and a single control treatment, for a range of effect sizes ($n = 1200$ patients; 600 in stage 1 and 600 in stage 2 of the adaptive seamless design [ASD]), for (a) a single effective treatment with the given effect size for both the early and the final outcome measures and (b) two effective treatments for the early outcome and one effective treatment and one partially effective treatment for the final outcome. Plots show power for a fixed (—) and a flexible selection rule (– – –).

MRI data and 3 year EDSS scores: (1) European Study Group on interferon beta-1b in secondary progressive MS;[20] (2) SPECTRIMS study group;[21] (3) North American Study group on Interferon beta-1b in Secondary Progressive MS.[22] However the only interim (in fact, 1 year) MRI 3-year EDSS correlation that was reported from these studies was from the European Study Group, which gave a Spearman correlation between change in EDSS at 3 years and percentage change in T2LV at 1 year of 0.13 with 95% CI (0.05, 0.21).[20] Correlations between 1-year MRI measurements and 3-year change in EDSS were not reported for any other MRI parameters. A number of other correlations were described, but they were, for example, between baseline MRI and 3-year EDSS scores which are not appropriate for our purposes.

Choosing design options

From the perspective of the UK health system and infrastructure, it seems feasible to recruit a total of 1200 patients to an ASD within 2 years, although of course this technique can model any sample size. Given these constraints and the framework of an ASD, the questions are (i) when to conduct the interim analysis and (ii) how to select treatments at interim. Simulations established the power to reject the null hypothesis for three test treatments and a single control treatment, for a range of effect sizes from small (0.2) to moderate (0.4) and for two selection rules: a fixed rule that simply selects the best performing treatment at interim and a flexible rule that allows more treatments to be selected depending on the spread of observed treatment means. Two scenarios were modelled: a single effective treatment with the given effect size for both the early and the final outcome measures (Figure 4(a)) and two effective treatments for the early outcome and one effective treatment and one partially effective treatment for the final outcome (Figure 4(b)).

For a moderate effect size (0.25) for a single effective treatment for the early and late outcomes (Figure 4(a)), both rules provided at least 80% power with a marginal preference for the fixed selection rule. However, the situation differs when two treatments are equally effective in terms of the early outcome, but one of these treatments is less effective for the final outcome measure (Figure 4(b)). Here the fixed, select-one-treatment-only, rule performs much more poorly than the flexible selection rule; the flexible selection rule attains approximately 80% power for an effect size of 0.25. The flexible rule, although not optimal for the scenario of a single fixed effect treatment through the trial (Figure 4(a)), is thus more robust than the fixed selection rule and is the preferred choice for the clinical scenarios described here.

Adaptive designs are more powerful than conventional designs in SPMS

ASDs were compared with conventional designs. A conventional design might consist of a multi-arm phase II trial comparing a number of test compounds with a common control, using n_1 patients per treatment arm. The best-performing treatment from this trial is then selected, based on an early outcome measure. A phase III trial then tests this compound against a control treatment, using n_2 patients per treatment arm. In three scenarios patients were considered allocated to treatment groups in stages 1 and 2 (phase II and III) in a ratio of 1:1 (– – –), 1:2 (—) and 1:3 (···) for n_1 and n_2 (Figure 5). The ASD is always more powerful than the conventional design and the time point of the interim analysis does not affect the power of the ASD as much as it does the power of the conventional design. This means that for the ASD we can afford a later interim analysis, and this gives a higher chance of picking the best treatment. As the number of treatments increases, then the advantage of the ASD becomes smaller.[6] For a small number of treatments the sample size saving can be large for the ASD approach. For instance, for two test compounds, for equal patient numbers in stage 1 and stage 2 treatment

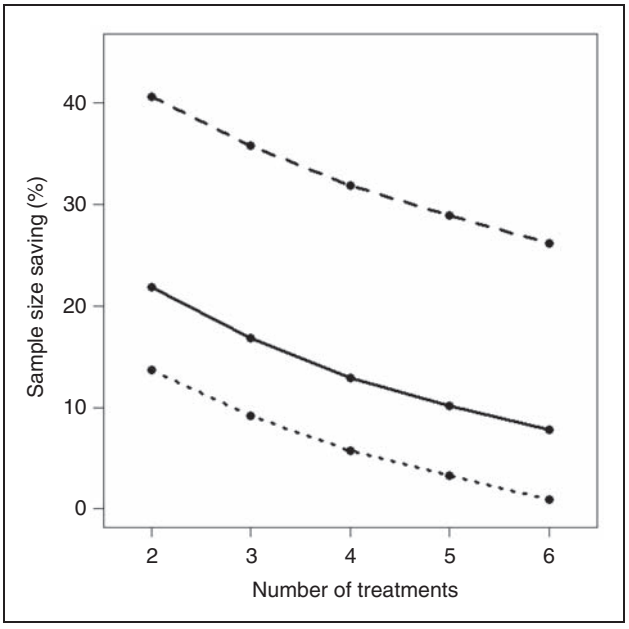

Figure 5. Comparison of sample size savings in adaptive seamless design (ASD) compared with conventional phase II/phase III clinical designs for two, three, four, five and six test treatments (and a single control treatment). Plots show saving for ASDs where patients are allocated to treatment groups in stages 1 and 2 (phase II and III) in a ratio of 1:1 (– – –), 1:2 (—) and 1:3 (. . .).

groups (1:1), for the ASD 80% power is obtained for a trial size of 750 patients and for the conventional design 80% power is obtained for a total phase II and phase III trial size of 1250 patients, a saving in sample size of 40% for the ASD design in total and in patients on placebo specifically.

Discussion

This work has developed the framework to take forward realistic adaptive seamless trial designs in SPMS constructed from analyses of data from previous trials, relevant research literature and individual clinical datasets.

In progressive MS the interim measure would be informed predominantly although not necessarily exclusively by MRI, of which currently the measure with the most longitudinal evidence is T2LV. The final outcome based on current recommendations is disability based and is the EDSS.[16] For the ASD, a key decision is which disability and MRI endpoints to use, and at which time points to record them. There is a current lack of robust longitudinal data linking the interim to the final measures, and whilst the ASD requires a positive effect on T2LV at interim analysis for a potential therapy to move forward, outcome measures such as atrophy[17,18] or clinical function scales, may provide additional information regarding efficacy and/or safety to further inform that decision.

Much of the challenge of implementation of a treatment selection design of the type described here arises exactly because of its flexibility and therefore it is important to consider carefully the range of options and assumptions prior to commencing a real trial. This is done by evaluating a large number of plausible clinical scenarios using simulation. A framework comprehensible to clinicians is essential to allow full engagement and rational decision making in trial design. In SPMS disease-specific issues including the lack of an available therapy mean that treatment effects need to be assumed. ASDs are generally based on 'futility' where test treatments are tested against a 'gold' standard, this absence means that we have tested the robustness of ASDs to variable treatment effects.

It can be seen that sample sizes of 1000–1250 can reasonably be expected to be sufficient to test up to four active compounds, with conservative imputation of values such as interim–final outcome correlations, effect sizes and selection rules. Moreover, this can be achieved despite relatively crude interim (T2LV) and final (EDSS) measurements. Clearly, as these choices become more sophisticated, the sample size will reduce.[18] As demonstrated in the simulation study, ASDs can produce sample size savings of up to 40%. Additional benefits of the ASD include reducing the number of subjects on placebo, reducing development time and ultimately reducing costs.

One finding is the importance of recruitment patterns. Whereas in conventional designs it is always optimal to recruit as rapidly as possible, ASDs benefit from a staggered recruitment. Ideally recruitment should start rapidly to recruit sufficient numbers for interim decision making and would only be restarted once the treatments have been selected. From a practical point of view a temporary recruitment stop might not be desirable. Therefore, one would continue to recruit, but at a lower rate.

Adaptive trial designs have been suggested for neurological diseases such as stroke[23] and neuropathic pain.[24] In terms of its efficiency and flexibility we believe it is now right to apply this technique to SPMS therapeutics, underscored by the recent failure of the large phase III trial of dirucotide (myelin basic protein),[25] to move the area on from single-agent, one-by-one, parallel design models.

Acknowledgements

The authors are grateful to the MS Society of Great Britain and Northern Ireland and to the following people for contributing data: Professor CP Hawkins (Institutes for Science and Technology in Medicine, Keele University), Dr J Palace, Ms A Cavey and Ms J Carey (Department of Neurology, John Radcliffe Hospital, Oxford), Professor AJ Thompson (Department of Brain Repair and Rehabilitation, Institute of Neurology, UCL), Dr N Tubridy (St Vincent's University Hospital, Dublin). JC and RN are grateful for support from the NIHR Biomedical Research Centre.

Funding

This study was funded by grants from the MS Society of Great Britain and Northern Ireland (grant reference numbers 878/08 and 913/09). Work undertaken at Imperial College Healthcare NHS Trust and University College London Hospitals NHS Foundation Trust received a proportion of funding from the Department of Health's NIHR Biomedical Research Centers funding scheme.

Conflict of interest statement

The authors have no commercial interests that might pose a conflict of interest in connection with the submission of this manuscript.

References

1. Clerico M, Faggiano F, Palace J, et al. Recombinant interferon beta or glatiramer acetate for delaying conversion of the first demyelinating event to multiple sclerosis. *Cochrane Database Systematic Rev* 2008; 2: CD005278.
2. Menge T, Weber MS, Hemmer B, et al. Disease-modifying agents for multiple sclerosis: recent advances and future prospects. *Drugs* 2008; 68: 2445–2468.

3. Compston A and Coles A. Multiple sclerosis. *Lancet* 2008; 372: 1502–1517.
4. DiMasi JA, Hansen RW and Grabowski HG. The price of innovation: New estimates of drug development costs. *J Health Econ* 2003; 22: 151–185.
5. Bauer P and Kieser M. Combining different phases in the development of medical treatments within a single trial. *Statist Med* 1999; 18: 1833–1848.
6. Bretz F, Schmidli H, Koenig F, et al. Confirmatory seamless phase II/III clinical trials with hypo-theses selection at interim: general concepts. *Biometrical J* 2006; 48: 623–634.
7. Friede T and Stallard N. A comparison of methods for adaptive treatment selection. *Biometrical J* 2008; 50: 767–781.
8. Burzykowski T, Molenberghs G and Buyse ME. *The Evaluation of Surrogate Endpoints*. New York: Springer, 2005.
9. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an Expanded Disability Status Scale (EDSS). *Neurology* 1983; 33: 1444–1452.
10. Hobart J, Lamping D, Fitzpatrick R, et al. The Multiple Sclerosis Impact Scale (MSIS-29): a new patient-based outcome measure. *Brain* 2001; 124: 962–973.
11. Whitaker JN, McFarland HF, Rudge P, et al. Outcomes assessment in multiple sclerosis clinical trials: a critical analysis. *Mult Scler* 1995; 1: 37–47.
12. Rudick R, Fischer J, Antel J, et al. Clinical outcomes assessment in multiple sclerosis. *Ann Neurol* 1996; 40: 469–479.
13. Friede T, Parsons N, Stallard N, et al. Designing a seamless phase II/III clinical trial using early outcomes for treatment selection: an application in multiple sclerosis. *Statist Med* 2010; (submitted).
14. Benda N, Branson M, Maurer W, et al. Aspects of modernizing drug development using clinical scenario planning and evaluation. *Drug Inform J* 2010; 44: 299–315.
15. Friede T, Nicholas R, Stallard N, et al. Refinement of the clinical scenario evaluation framework for assessment of competing development strategies with an application to multiple sclerosis. *Drug Inform J* 2010; (in press).
16. EMEA. Guidelines on clinical investigation of medicinal products for the treatment of multiple sclerosis. European Medicines Agency, 2005. Available at http://www.emea.europa.eu/pdfs/human/ewp/056198en.pdf (accessed 24 November 2009).
17. Barkhof F, Calabresi PA, Miller DH, et al. Imaging outcomes for neuroprotection and repair in multiple sclerosis trials. *Nat Rev Neurol* 2009; 5: 256–266.
18. Altmann DR, Jasperse B, Barkhof F, et al. Sample sizes for brain atrophy outcomes in trials for secondary progressive multiple sclerosis. *Neurology* 2009; 72: 595–601.
19. Cohen JA. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. New York: Lawrence Erlbaum Associates, 1988.
20. Sormani MP, Bruzzi P, Beckmann K, et al. MRI metrics as surrogate endpoints for EDSS progression in SPMS patients treated with IFN beta-1b. *Neurology* 2003; 60: 1462–1466.
21. Li DK, Zhao GJ and Paty DW, University of British Columbia MS/MRI Analysis Research Group: The SPECTRIMS Study Group. Randomized controlled trial of iterferon-beta-1a in secondary progressive MS: MRI results. *Neurology* 2001; 56: 1505–1513.
22. North American Study group on Interferon beta-1b in Secondary Progressive MS. Interferon beta-1b in secondary progressive MS: results from 3-year controlled study. *Neurology* 2004; 63: 1788–1795.
23. Krams M, Lees KR and Berry DA. The past is the future: innovative designs in acute stroke therapy trials. *Stroke* 2005; 36: 1341–1347.
24. Maca J, Bhattacharya S, Dragalin V, et al. Adaptive seamless phase II/III designs - Background, operational aspects, and examples. *Drug Inform J* 2006; 40: 463–473.
25. Freedman MS, Verco T, Ossanna M, et al. A double-blind, placebo-controlled, multi-center study to evaluate the efficacy and safety of dirucotide in patients with secondary progressive multiple sclerosis. *Mult Scler* 2009; 15: S272–S273.