Is It Possible to Conduct a Multi-Arm Multi-Stage Platform Trial in Parkinson’s Disease: Lessons Learned from Other Neurodegenerative Disorders and Cancer

Marie-Louise Zeissler\(^a\), Vivien Li\(^{b,c,d}\), Mahesh K.B. Parmar\(^d\) and Camille Buchholz Carroll\(^a,\ast\)

\(^a\)Applied Parkinson’s Research Group, University of Plymouth, Faculty of Health: Medicine, Dentistry and Human Sciences, Plymouth, United Kingdom
\(^b\)Queen Square MS Centre, Department of Neuroinflammation, UCL Institute of Neurology, London, UK
\(^c\)Department of Uro-Neurology, National Hospital for Neurology and Neurosurgery, UCL Institute of Neurology, Queen Square, London, United Kingdom
\(^d\)MRC Clinical Trials Unit at UCL, University College London, London, United Kingdom

Accepted 7 February 2020

Abstract. Many potential disease modifying therapies have been identified as suitable for clinical evaluation in Parkinson’s disease (PD). Currently, the evaluation of compounds in phase II and phase III clinical trials in PD are set up in isolation, a process that is lengthy, costly and lacks efficiency. This review will introduce the concept of a multi-arm, multi-stage (MAMS) trial platform which allows for the assessment of several potential therapies at once, transitioning seamlessly from a phase II safety and efficacy study to a phase III trial by means of an interim analysis. At the interim checkpoint, ineffective arms are dropped and replaced by new treatment arms, thereby allowing for the continuous evaluation of interventions. MAMS trial platforms already exist for prostate, renal and oropharyngeal cancer and are currently being developed for progressive multiple sclerosis (PMS) and motor neuron disease (MND) within the UK. As a MAMS trial will evaluate many potential treatments it is of critical importance that a widely endorsed core protocol is developed which will investigate outcomes and objectives meaningful to patients. This review will discuss the challenges of drug selection, trial design, stratification and outcome measures and will share strategies implemented in the planned MAMS trials for MND and PMS that may be of interest to the PD field.

Keywords: Parkinson’s disease, adaptive clinical trial, outcome measure, clinical trial protocol, multiple sclerosis, motor neuron disease

INTRODUCTION

Parkinson’s disease (PD) is a neurodegenerative disease for which there are currently no therapies that delay or stop progression. The disease encompasses an array of motor and non-motor symptoms which
have an increasing impact on patients’ and carers’ quality of life.

An estimated 0.85% of the population over 60 years of age is affected by PD in the UK [1, 2]. This amounted to an estimated annual care cost (direct and indirect) of €3.6 billion in 2018 [3, 4]. As both prevalence and incidence are projected to increase, it is imperative to develop disease-modifying therapies.

With the ever increasing pace of research, including in vitro and in vivo screening systems, advances in virtual drug modelling and bioinformatics approaches, the number of suitable drug candidates for clinical evaluation as potential disease modifying therapies is on the rise [5–8].

The development, financing and conducting of a clinical trial can take many years, creating a bottleneck for the clinical evaluation of potential therapies, particularly in the phase III setting. There is therefore a need to adopt an innovative and adaptive approach that allows for the seamless streamlining of trials and the testing of multiple hypotheses at once. One such example is the multi-arm, multi-stage (MAMS) platform trial design. Most MAMS trials are phase III trials that allow the evaluation of multiple therapeutic interventions against a common placebo arm (multi-arm), with pre-determined interim assessments that will evaluate safety and efficacy of all interventions (multi-stage). This allows for the early detection of ineffective treatment arms, which can then be replaced in favour of other interventions (Fig. 1). Most importantly, the multi-stage approach allows for the incorporation of phase II equivalent findings to be carried forward to the final phase III results without requiring the initiation of a separate trial. A MAMS trial is not only quicker and cheaper than a conventional approach, but also more efficient as fewer patients are required to come to a phase III powered conclusion regarding a particular intervention.

There are many challenges inherent to the implementation of a platform trial. In PD, eight phase II trials have provided sufficient evidence of efficacy for the initiation of eight phase III trials within the last 10 years. Whilst all phase II studies exhibited at least a trend towards activity, only one phase III trial (the ADAGIO study) successfully showed evidence of benefit on its efficacy endpoints (Table 1). Although it is possible that these drugs were not neuroprotective, current outcome measures, trial design features and patient population heterogeneity may play a significant role in the ability of trials to show efficacy of a new treatment.

This review will focus on the challenges ahead for the implementation of a MAMS platform in PD including drug selection, trial design, stratification and outcome measures. Many of these challenges are not unique to PD and therefore insights from other diseases that have embarked on this process will be shared. Here we will particularly focus on current efforts made in the development for MAMS trial platforms in motor-neuron disease (MND) and progressive multiple sclerosis (PMS) as well as one well established MAMS trial for the investigation of metastatic prostate cancer (STAMPEDE). Furthermore, we will discuss why common design decisions such as restriction of study populations to early disease and lack of stratification should be reconsidered.

MAMS PLATFORM TRIALS: STAMPEDE AND NEURODEGENERATIVE DISEASES

The STAMPEDE trial is perhaps one of the most successful examples of a MAMS platform. The protocol was initiated in October 2005 for men with poor prognosis prostate cancer who were starting long-term hormone therapy for the first time. Long-term hormone therapy was discovered in the 1950s and since then there had been no new treatments for men with this stage of disease. In the first version of the protocol, 5 new treatments (added to long term hormone therapy) were evaluated against a control arm of long-term hormone therapy alone. At the time of writing (October 2019), more than 11,000 patients had been recruited to the study, and version number 20 of the protocol had been launched. Over the last 14 years, the control (standard of care) arm of STAMPEDE has been adapted and improved 4 times – 3 times due to results emerging from research arms within STAMPEDE and once from results from another trial. Over a period of 20 years from the start of the protocol, STAMPEDE will have evaluated 10 different new treatments. The aim is to evaluate further treatments within STAMPEDE until at least 2030 and continue to improve the survival of men with advanced prostate cancer.

Within the UK, MAMS adaptive platform trials are being developed for MND and PMS. As with PD, these initiatives have arisen to address the unmet need for disease modifying therapies with strong backing from patients, carers and charitable groups. A MAMS platform design was selected in both conditions to enable more efficient testing of potential therapies.
Fig. 1. MAMS trial schematic. In this example, a five year phase III set up is shown that assumes one out of four tested interventions (D) show a positive signal at an interim analysis after 18 months. By the time one phase III conclusion has been reached, 12 study drugs will have been initiated into the trial, eight will have been terminated at the interim analysis stage, two will have shown sufficient evidence to be carried forward to phase III and two will be pending interim results in year six. (--- analysis).

<table>
<thead>
<tr>
<th>Phase III studies</th>
<th>Clinical Trial Identifier</th>
<th>Endpoint met</th>
<th>Was there evidence of disease improvement in phase II</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isradipine</td>
<td>NCT02168842</td>
<td>Unpublished</td>
<td>No, but trend towards efficacy</td>
</tr>
<tr>
<td>Inosine</td>
<td>NCT02642393</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Creatine</td>
<td>NCT00449865</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Coenzyme Q10</td>
<td>NCT00740714</td>
<td>No</td>
<td>Yes (n = 2) Yes in advanced disease (n = 1)</td>
</tr>
<tr>
<td>Rasagiline</td>
<td>NCT00256204</td>
<td>Yes</td>
<td>No, but trend towards efficacy</td>
</tr>
<tr>
<td>Ganoderma</td>
<td>NCT03594656</td>
<td>Unpublished</td>
<td>Unpublished</td>
</tr>
<tr>
<td>Pramipexole</td>
<td>NCT00321854</td>
<td>No</td>
<td>No, but trend towards efficacy</td>
</tr>
<tr>
<td>Levodopa+Carbidopa</td>
<td>ISRCTN30518857</td>
<td>No</td>
<td>Unpublished</td>
</tr>
</tbody>
</table>

The MND Systematic Multi-Arm Adaptive Randomised Trial (MND SMART) is being sponsored by the University of Edinburgh with funding from a number of charitable organisations and has recently gained regulatory approval. It takes advantage of Scotland’s CARE-MND integrated care and research platform.

For PMS, the UK MS Society Clinical Trials Network was initiated in 2007 to develop an efficient, multi-arm trial of repurposed drugs. This led to the development of MS-Secondary Progressive Multi-Arm Randomisation Trial (MS SMART), a multi-arm trial evaluating three candidate neuroprotective drugs with distinct mechanistic actions [11]. The trial was completed in 2018 with 445 patients recruited. Although none of the research arms showed a benefit on the primary phase II outcome (MRI brain atrophy) over placebo, this trial did demonstrate the feasibility of conducting a multi-arm trial of this scale.

The UK MS Society’s 2018-2022 Research Strategy has focused on designing and delivering a clinical MAMS trial platform to efficiently identify and evaluate potential treatments for PMS. It will incorporate both phase II (interim) and phase III (final) evaluations, with the possibility of dropping arms at interim stages where there is insufficient signal for efficacy. Building on the work of treatment selection, trial...
design, outcomes and infrastructure strategic working groups, a programme grant has been awarded and the trial will commence setup from early 2020 with recruitment beginning in 2021.

CHALLENGES TO INITIATING A MAMS PLATFORM IN PD

The continued failure to discover new medicines for neurodegenerative diseases reflects bottlenecks at two pivotal stages of drug development (i) discovery/identification of promising candidate interventions and (ii) translation of these agents into treatments that are available to patients through cost and time-efficient testing in clinical trials. To maximise the potential of a MAMS platform trial which will run over many years and interrogate many research questions, it is crucial that there is a pipeline in place that will continuously identify and evaluate suitable drug candidates. Furthermore, outcome measures have to be chosen that are sensitive enough to changes in disease progression over interim stages as well as the full duration of the trial. Decisions on outcome measures, trial population and design need to be widely endorsed not only in the scientific community to encourage widespread collaboration on the programme, but also by patients as their support is crucial for success.

Challenge 1: Drug selection

Many diseases have common etiopathological processes that are echoed in PD such as inflammation, protein trafficking and accumulation, oxidative stress and mitochondrial dysfunction. This is particularly true for neurodegenerative diseases (Supplementary Table 1) but can be found across many medical conditions including cancer [9] and diabetes [10]. It is therefore conceivable that agents targeting these mechanisms in other diseases may also be of value in PD.

Drug repurposing is an effective way to identify new therapeutic strategies as it uses existing clinical efficacy, safety and regulatory data and if off-patent, drugs are likely to be quicker and cheaper to get to licensed approval status.

For the successful implementation of a MAMS trial platform, a process for the selection of suitable drugs needs to be in place. In PD, a process has been devised through the linked clinical trials initiative (LCT) [11, 12]. Similar to its predecessor, the National Institute of Neurological Disorders and Stroke (NINDS) Committee to Identify Neuroprotective Agents for Parkinson’s (CINAPS) [13, 14], the identification and prioritisation of candidates to be entered into pilot clinical trials for PD is the main objective of the LCT committee. In short, the process involves the screening and selection of candidates on the basis of their potential efficacy in PD, evidence for safety, ability to penetrate the blood-brain-barrier, evidence of efficacy in PD animal models, the possibility for measuring target engagement and their commercial/patent status. This results in the compilation of detailed dossiers for a subset of drugs that meet basic pharmacological and safety criteria which are then scored and presented at an annual meeting of international experts where prioritisation for clinical drug testing is finalised [11, 12].

In the MS SMART study, a similar process of systematic review, commissioned by the MS-Clinical Trials Network was undertaken of animal and human trials of putative neuroprotective drugs. This process included a three-part systematic review and meta-analysis assessing the neurodegenerative disease literature including clinical and pre-clinical (animal and human induced pluripotent stem cell) publications to identify drugs ranked by a product score that takes into account efficacy, safety, study size and quality. The human analysis was extended to other neurodegenerative diseases including amyotrophic lateral sclerosis (ALS), Huntington’s, Alzheimer’s and PD, because of the existence of shared pivotal pathways in neurodegeneration. The focus of the upcoming PMS MAMS adaptive platform trial is on repurposed (already approved drugs) and rescued drugs (drugs at advanced stage of development but abandoned before approval) that act on mechanistic pathways in PMS [15].

MND SMART is utilising the infrastructure established by MS SMART for drug selection. Additionally, late-breaking drugs that may not have been tested previously in neurodegenerative diseases but show promise from preclinical data or from clinical trials in other diseases with similar underlying pathological processes are being considered.

Beyond the selection and prioritisation of candidate drugs there are other practical considerations that pose unique challenges to a MAMS trial protocol, such as the use of a common placebo. Traditionally, the MAMS trial design allows for the comparison of multiple therapies against a common control arm. Whilst this is a useful approach to reduce trial cost, it makes it difficult to incorporate and compare drugs that have different routes of administration. Out of the
Challenge 2: Outcome measures

Motor progression

Currently, the Movement Disorder Society - Unified PD rating scale (MDS-UPDRS) is used worldwide for the assessment of the severity of PD symptoms and is endorsed by regulators such as the FDA. The scale consists of four parts: Parts I and II - Non-Motor and Motor Experience of Daily Living; Part III – Motor Examination; Part IV – Motor Complications. This revision of the UPDRS rating scale was introduced in 2008 and aimed to allow finer differentiation at earlier stages of the disease [17].

Clinical trials have generally utilised either the clinically assessed part III motor score as their outcome measure or a composite score of several parts of the scale. A search of clinicaltrials.gov looking at neuroprotective trials within the last 10 years shows that out of 31 phase II studies with a motor outcome, 42% used the part I-III or I-IV total score and 55% assessed part III only. In contrast, a composite score is more likely to be used in the phase III setting (out of eight phase III trials, six utilised a change in I-III or I-IV composite score as their primary endpoint, one assessed part III only and one trial did not use the UPDRS as an outcome measure) [18–25].

However, changes in the MDS-UPDRS total score (within the first five years of disease) are largely driven by changes in the MDS-UPDRS Part III score; as part I and II score changes are not as pronounced over time [26] their addition into a total score may reduce sensitivity to change [27]. This is in line with findings examining the old UPDRS scale [28]. Although several studies have suggested that composite scores are able to detect clinically relevant differences over time [26, 27, 29], the authors of the new scale clearly recommended against the use of I-III composite scores as their unequal factorial loading reduces sensitivity to change [17].

Assessing part III only has its own limitations: it provides only a snapshot of the patients’ condition and is subject to significant inter- and intra-rater variability, which can undermine score accuracy. This is especially relevant for large multi-centre trials where many different raters are assessing patients throughout the study. Part III results are also dependent on the medication state of the patient [26, 28, 30] with OFF state results being more reliable compared with the ON state [30].

Longitudinal observations of clinical progression on the rating scale are important in determining the sensitivity of the scale in detecting differences at various stages of disease. Whilst some data indicate that overall disease progression up to at least five years from diagnosis is linear [26, 31, 32], it is generally accepted that the rate of motor symptom decline decelerates beyond five years [28, 32–34]. Progression studies using the revised scale are only just emerging [26, 27, 35]; to date these studies do not report data of patients beyond five years of disease, representing an important knowledge gap.

Cognition

Mild cognitive impairment (MCI) affects 20%–33% of patients at diagnosis and 80% of PD patients will go on to develop PD associated dementia within 12 years of disease, making cognitive dysfunction one of the most prevalent non-motor symptoms of PD [36, 37]. Cognitive function is thus often monitored in neuroprotective trials. Progression of cognitive impairment could be described as inverse to that of motor symptoms, with slow initial progression for up to 10 years followed by a rapid decline thereafter [32, 38, 39].

Recent evidence suggests that progression on the MoCA, MMSE, and Scales for Outcomes in Parkinson’s disease-Cognition (SCOPA-Cog) in non-demented PD patients is not reliably detectable over the typical timeframe of a clinical trial (1 year) [36, 40] with the MoCA being potentially more sensitive than the MMSE at detecting MCI [37].

Time to event measures

Historically, the most common time to event measure employed in neuroprotective trials of PD has been time to initiation of levodopa therapy. A study published in 2016 showed that 50% of neuroprotective trials identified at the time investigated time to initiation of dopaminergic treatment [41]. However, this measure is highly dependent on the judgement of the site investigator. In the case of studies where study drug is withdrawn after commencement of symptomatic therapy (ST), differing motivation of patients...
to continue the study may also play a role. Time to commencement of ST is therefore prone to bias. Other events that signify milestones for disease progression are onset of dementia, falls, nursing home placement and death.

Time to falls is of particular interest here as a recent study suggests that approximately 70% of PD patients fall within the first 5 years of diagnosis [42] and thus recording time to first fall may be a relatively rapidly evolving event in PD in comparison to events that occur much later such as dementia, nursing home placement or death.

The choice and feasibility of time to event measures employed in a clinical trial largely depend on the patient population being investigated. For example, time to death is dependent on the demographic of patients included in the study [43] with younger onset patients generally having a lower life expectancy [44–48], although some studies have found no difference [49].

Delay of these disease progression milestones are highly meaningful to patients and may therefore represent an alternative to conventional primary outcomes used in phase III trials.

**Non-motor symptoms**

Progression of non-motor symptoms (NMS) such as sleep, anxiety, apathy, depression, autonomic dysfunction and cognition are less influenced by dopamine therapy and significantly progress over short periods of 1-2 years even in an early disease population [50]. This makes measurement of NMS an attractive tool for evaluating disease progression in patients that require symptomatic treatment. NMS in PD can be measured by part I of the MDS-UPDRS scale or the NMS scale for PD. Both are suitable for detecting prevalence of NMS within PD populations [51] and have been used as secondary outcomes in PD trials. The MDS-NMS has recently been published [52].

**Quality of life measures**

Quality of life (QoL) measures are patient reported outcomes that attempt to measure physical, mental and social wellbeing that can either be related to general health or disease specific depending on the instrument. There are many generic and PD specific tools available although the most common QoL scale used in clinical trials of PD is the Parkinson’s Disease Questionnaire (PDQ-39) which captures the impact of both motor and non-motor symptoms on quality of life [53]. The impact of PD on QoL has been reviewed by others. In summary, PD has a profound impact on quality of life, has been shown to correlate with disease severity and has been useful at determining the impact of symptomatic therapy on patients in early as well as late disease [53, 54].

**Biomarkers**

A biomarker is an objectively measured characteristic that serves as an indicator of normal biological processes, pathogenic processes or pharmacological responses to a therapeutic intervention [55]. There is currently no biomarker that can reliably and sensitively track PD severity [32, 56, 57]. Consequently, biomarkers are not always measured and if so, they are generally ancillary exploratory outcomes in PD trials. Only 11 of 29 published phase II studies included at least one biomarker measurement in their protocol. These included 11 studies with brain imaging assessments [24, 58–64], one assessment of 8-hydroxy-2’‐deoxyguanosine measurement in plasma and one in urine [65], as well as two studies with cerebrospinal fluid (CSF) analysis of protein aggregation [66, 67]. Progress on the usefulness of imaging techniques and fluid or tissue biomarkers has been reviewed extensively by others [56, 57, 68]. Importantly, evidence regarding their relevance to clinical worsening of the disease is often contradictory.

**Digital health technologies (DHT)**

DHTs can be broadly categorised into passive (such as wearables) and active (apps that require data input or task completion) data capture. Where active measures are used, compliance may create problems with accuracy and reliability of data sets over long study periods [69–72]. Uploading of data collected by apps and sensors represents an additional challenge, making studies liable to data gaps due to patients living in poor network areas or outdated hospital IT systems. Mobilise-D is an EU funded project that has set out to improve the accurate assessment of daily life mobility by developing standardised tools [73]. A MAMS trial could provide a platform for integrated assessment of different digital measures that can be tied in to routinely collected clinical data of disease progression thereby validating their use.

**Shaping future outcomes**

The challenge of assessing the neuroprotective effect of drugs is not unique to PD. Measurement of worsening neurological function in a slowly
progressive condition with relatively short (2–3 years) trials is inherently difficult. In MS there are also significant variations in progression between patients and changes in rate over time across several domains of neurological function. MRI brain atrophy measurements can be used to quantify aspects of neurodegeneration [74] and have been employed as the primary outcome in the phase II MS-STAT [75] and MS-SMART trials, as well as a secondary outcome in phase III trials of PMS. However, clinical outcome measures of relevance to patients are critically important to determine the functional impact of treatments on patients with PMS and expected by regulators in phase III trials.

The Expanded Disability Status Scale (EDSS) is currently the most widely used scale in phase III PMS clinical trials. Similar to the UPDRS scale in PD, where motor function is the main driver of progression [26], the EDSS overemphasises walking ability, whilst minimising the contribution of cognition and arm function [76]. Considering the range of disability typical of PMS patients, this is a major shortcoming.

To overcome this limitation, the Multiple Sclerosis Functional Composite (MSFC) was designed and includes measures of lower limb function (timed 25-foot walk), upper limb function and hand dexterity (9-hole peg test) and cognitive function (Paced Auditory Serial Addition Test) [77]. Subsequent amendments replaced the Paced Auditory Serial Addition Test by the Symbol Digit Modalities Test, a test of processing speed, and low-contrast letter acuity as a measurement of visual function [78]. A composite outcome measure has been successfully used in the INFORMS trial [79] and was an exploratory outcome in the ORATORIO trial (phase III trial of fingolimod and Ocrelizumab respectively) [80].

In the design of the UK MS Society’s PMS adaptive trial platform, a composite clinical endpoint, based on EDSS, 9-hole peg test and timed 25-foot walk, was decided as the primary phase III outcome based on expert consensus from the outcomes working group. As a composite endpoint is reached by confirmed disability progression on just one of the scales included in the composite measure, more progression events are recorded than if just assessing one scale alone, thereby reducing the time required to reach a progression endpoint as well as the number of participants required. It was also favoured by patient and public interest groups as it encompassed a broader range of meaningful measurements, particularly in non-ambulant patients in whom walking tests may not be possible or relevant.

Composite outcomes have also been successfully developed for MND. Given the typical trajectory of the disease, mortality is an important outcome for MND clinical trials. However, the large sample sizes and duration required to power such a study makes this practically challenging.

The most commonly used and best validated tool is the Amyotrophic Lateral Sclerosis-Functional Rating Score ALS-FRS(R). The ALS-FRS(R) has the advantage of assessing a wide range of neurological functions including communication, activities of daily living and mobility [81, 82]. Thus, the combined assessment of function and survival (CAFS) ranks outcomes based on survival time and change in ALS-FRS(R) score was developed [83].

Combining endpoints might be a useful strategy in PD to enable the simultaneous analysis of multiple equally important endpoints and may serve to give aspects of the disease, such as cognition, that become more pronounced at later stages of disease, a greater weight.

To date only one phase III trial (The NET-PD LS-1 trial) has attempted to measure a one year delay in cumulative disability using a global statistic measure combining the Modified Rankin Scale, Symbol Digit Modalities Test, PDQ-39 Summary Index, Schwab and England Activities of Daily Living scale, and ambulatory capacity over a five year period as primary efficacy endpoint. In addition to the necessity for long term monitoring of these measures, which has implications on cost and attrition rates, the tests and scales employed exhibited slower progression and higher variability across the study population than originally anticipated [23].

**Challenge 3: Trial design**

**Inclusion criteria**

It is important that the trial design for a MAMS platform has strategies in place to minimise or account for patient heterogeneity in terms of disease severity and rate of progression as this will affect variance of efficacy outcomes. Almost all phase III clinical trials conducted within the last 10 years have restricted recruitment to early, untreated disease for this reason (Supplementary Table 2). This patient group is more likely to follow a linear progression trajectory which makes it easier and more reliable to compute efficacy endpoints [31] and drug naive PD can be modelled more accurately [35]. It is also considered that this restrictive recruitment strategy will capture a population where pathological dete-
Corruption within affected brain regions may still be at a redeemable stage [84]. It should be noted here that the assessment of a drug naïve PD population requires adjustments in order to account for patients that initiate dopamine therapy during the course of the trial. A common strategy is to define the endpoint as change from baseline to the end of the study duration or until initiation of dopaminergic therapy. However, this method treats differential progression as equal and does not take into account the potential of these groups responding differently to the study drug.

Limitation of recruitment to early disease has other disadvantages. Most importantly, unless the treatment effect of a neuroprotective drug is comparable to symptomatic benefit of levodopa, patients are likely to require and receive adjunctive dopamine therapy. Therefore, if no benefit can be detected whilst patients receive symptomatic therapy, the agent is not going to have an impact on the patient’s quality of life. Furthermore, early disease trials generally preclude data collection on efficacy of treatments in later disease stages and therefore there is no data to support efficacy or safety as disease advances, which will require further costly trials. Any beneficial effect at later disease stages, where different neurological systems are starting to be involved, would be missed by such a design and exploration of these drugs for PD are most likely going to be stopped after a negative result in an early disease group. In fact, one phase II trial of co-enzyme Q10 stratified patients into early (not treated with levodopa) and more advanced (PD with wearing off) arms and found a treatment benefit (change in total UPDRS score) only in more advanced disease [85]. Although this result came from a very small study (average of 12 patients per arm) and may represent the differential sensitivity of the scale at these two disease stages, a potential beneficial effect in this patient group could not have been (and was not) captured in the later phase III trial that focused on early, untreated disease [19]. These are important considerations that are particularly relevant in the context of establishing a continuous platform trial where there is a high throughput of interventions.

Stratification

A trial that allows for the recruitment of de novo and more advanced PD will require stratification of patients based on disease severity. However, it might be possible to further categorise patients into subtypes that have differential predicted disease progression. Patients with baseline characteristics of increased age, male sex, mild cognitive impairment, REM sleep behaviour disorder and orthostatic hypotension are more likely to progress faster, whereas tremor-dominant patients follow a milder disease course [86–90]. The differences in disease progression between these clinical subtypes can be very significant [86]. Progress in practically distinguishing between PD subtypes has been made by Fereshtehnejad et al and Velseboer et al, who have developed subtype calculators for newly diagnosed patients [87, 91]. Unless these groups are treated as clinically distinct in trial design, differences in their treatment response will never be detected and thus their importance as etiologically distinct groups cannot be verified. However, whether predictions of these subtype calculators hold true in longitudinal data sets remains to be investigated and previous attempts to verify clinical subtypes in additional cohorts have pointed towards a lack of reproducibility as well as instability of identified phenotypes in early disease populations [92, 93].

Although the reproducibility of clinical subtypes in additional cohorts is often inconsistent [92], evidence suggests that tremor dominance in particular, is highly predictive of slower progression on the UPDRS scale [86–90, 94–96]. This may be due to the fact that the UPDRS and MDS-UPDRS scale is weighted towards tremor and an early presentation of high tremor scores may reduce the probability of tremor dominant individuals to progress.

PD is complex and multi-factorial and therefore it might be wise to step back from the single endpoint paradigm of trial design that is based on the assumption that a population as a whole will respond and show improvement homogeneously. One way to acknowledge the complex multifactorial nature of PD in a MAMS trial without tightening inclusion criteria, could be to measure and power interim progression on individuals with low tremor scores only, whilst ultimate progression for phase III analysis may be a non-UPDRS based measure. Whether this is a feasible approach remains to be determined. In the MND-SMART trial, interim analysis excludes individuals classified as long survivors in order to reduce heterogeneity at that point without excluding long survivors from the trial. The planned inclusion EDSS range of 3.0–8.0 for the PMS MAMS trial is also broader than in previous phase III trials (typically 3.5 or 4.0–6.5). Making clinical trials more inclusive has been emphasised by patient and public involvement groups. Baseline EDSS will be one of the minimisation factors in randomisation to ensure comparability of study arms.
Fig. 2. Methodology to investigate disease modification. The delayed start design is a two period trial design. In period one, patients are initially randomised into placebo (delayed start) and treatment arms (early start) at the end of which, the placebo group is switched to the study drug (period two). Both groups will receive the study drug for the remaining duration of the trial. Should a treatment be neuroprotective rather than symptomatic, the early start group should have a reduced progression rate as compared to placebo in period one, a significantly improved MDS-UPDRS score from baseline at the end of period two and an equal or reduced rate of disease progression in the early start group compared to the delayed start group in period two. Thus, such a design has three endpoints [31]. The wash-out design is a two period design that evaluates a global change in the outcome measure of choice from baseline over a drug administration period and the maintenance of this change after the study drug is withdrawn (washed out).

**Methodology**

There are two distinct trial designs that aim to distinguish between symptomatic and disease modifying effects of a study drug. These are the delayed start and wash-out design (Fig. 2). Within the past 10 years, four phase III trials have used a delayed start design, one has employed a wash out design and three chose to assess beneficial effects without attempting to distinguish between symptomatic and neuroprotective benefits (Table 2). The delayed start design has been successfully employed by the ADAGIO study that investigated the MAO inhibitor rasagiline, which is currently used as a symptomatic treatment of PD. However, the design has been criticised as it assumes linearity of disease progression over the trial period and whilst it may be reasonable to make such an assumption in a *de novo* PD patient group, its feasibility in more advanced disease is questionable. Furthermore, differential dropout rates between the two arms have to be imputed. This is particularly relevant in trials that examine therapies with known symptomatic effect on a drug naïve patient population (such as the ADAGIO study), as the need for symptomatic therapy will occur earlier in the delayed start group resulting in increased drop out.

The delayed start design is an interesting consideration for a MAMS trial design as it already allows for an interim analysis after period one which could serve as an interim checkpoint in a multi-arm trial. A primary outcome measure that is sensitive enough to detect changes within the relatively short time scales of period one and two would have to be chosen.

The wash-out design requires knowledge of the pharmacology of the drug and thus evaluation of blood and plasma levels would have to be incorporated into the initial phase II component of the MAMS trial. This ensures that the wash-out period is long enough to allow for clearance of the drug from the body so that a persisting benefit thereafter, indicative of its influence on underlying disease pathology, can be confirmed. The advantage of this design is that period one is generally longer than period one of a delayed start design allowing for the evaluation of potentially slower evolving outcome measures. However, long duration effects are often poorly understood and therefore it is hard to conclude whether a given wash-out period will be sufficient to preclude long lasting symptomatic benefits. For example, despite the short half-life of levodopa in plasma [97, 98] clinical benefits persist much longer [99, 100]. Thus the results of the ELLDOPA study, where levodopa treatment led to sustained improvement of the total UPDRS score over a two week wash out period (indicative of slowed disease progression) remained inconclusive and were further confounded by its imaging substudy that showed a decline in dopamine transporter (DAT) activity ([123I] β-CIT uptake) [99, 101].
Table 2  
Trial methodology and endpoints of phase III neuroprotective trials and their preceding phase II trials within the last 10 years; ST, symptomatic therapy

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Endpoint</th>
<th>Intervention</th>
<th>Phase</th>
<th>Clinical trial identifier</th>
<th>Treatment duration</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delayed start</td>
<td>1) Difference in rate of progression (period 1)</td>
<td>Rasagiline</td>
<td>3</td>
<td>NCT00256204</td>
<td>Period 1: 9 months; Period 2: 9 months</td>
<td>Allowed 3 months wash in for slope analyses</td>
</tr>
<tr>
<td></td>
<td>2) non-inferiority of slopes (period 2)</td>
<td>Ganoderma</td>
<td>3</td>
<td>NCT03594656</td>
<td>Period 1: 6 months; Period 2: 12 months</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3) difference in total UPDRS score between early and delayed (period 1+2)</td>
<td>Pramipexole</td>
<td>3</td>
<td>NCT00321854</td>
<td>Period 1: 6–9 months; Period 2: 2 months</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Levodopa+Carbidopa</td>
<td>3</td>
<td>ISRCTN40518857</td>
<td>Period 1: 9 months; Period 2: 9 months</td>
<td></td>
</tr>
<tr>
<td>Wash out</td>
<td>Difference in rate of change from shared baseline</td>
<td>Inosine</td>
<td>3</td>
<td>NCT02642393</td>
<td>Period 1: 24 months; Period 2: 3 months</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change from baseline vs each arm</td>
<td>Isradipine</td>
<td>2</td>
<td>NCT00909545</td>
<td>Period 1: 12 months; Period 2: 0.5 months</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change from baseline</td>
<td>Ubiquinol-10</td>
<td>2</td>
<td>N/A</td>
<td>Period 1: 12 months (advanced PD arm), 24 months (early PD arm); Period 2: 2 months</td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>Change from baseline</td>
<td>Creatine</td>
<td>3</td>
<td>NCT00449865</td>
<td>60 months</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Change from baseline for those receiving ST vs placebo</td>
<td>Coenzyme Q10 with Vitamin E</td>
<td>3</td>
<td>NCT00740714</td>
<td>16 months or initiation of ST</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Futility boundary 70% of placebo group progression</td>
<td>Isradipine</td>
<td>3</td>
<td>NCT02168842</td>
<td>36</td>
<td>DATATOP placebo + trial placebo arm</td>
</tr>
<tr>
<td></td>
<td>Change from baseline</td>
<td>Coenzyme Q10</td>
<td>2</td>
<td>NCT00063193</td>
<td>12 months or initiation of ST</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Futility boundary 70% of placebo group progression</td>
<td>Coenzyme Q10</td>
<td>2</td>
<td>NCT00076492</td>
<td>12 months or initiation of ST</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Futility boundary 70% of placebo group progression</td>
<td>Inosine</td>
<td>2</td>
<td>NCT00833690</td>
<td>23 months</td>
<td></td>
</tr>
</tbody>
</table>
Distinguishing between a symptomatic and a disease modifying therapy at the interim check point may be challenging. A MAMS trial is only feasible if an interim analysis is conducted before the recruitment target for the phase III element is reached. Introducing a wash out or delayed start design prior to the interim analysis may increase the trial duration beyond the feasibility threshold for a MAMS trial. Therefore, a MAMS trial design that postpones distinguishing between disease modification and symptomatic effects until the phase III analysis may be more practical.

PROPELING MAMS INTO ACTION

A MAMS trial is a substantial logistical undertaking and thus requires significant investment to ensure trial delivery. STAMPEDE, which has been running for more than 10 years currently employs tens of core staff that are responsible for continued data collection, management, analysis and biobanking, amongst a range of other roles as well as facilitating the seamless introduction of new study arms. Beyond the core staff required for the trial, which is generally funded through disease specific charities, funding for individual research arms is sourced separately through standard clinical research grants or industry sponsorship. MAMS trial arms should

1) investigate drugs that are as divergent as possible (a practical consideration that also retains the possibility of combining effective treatments at a later stage) and
2) be powered to compare active arms to placebo only

thereby avoiding direct competition between drugs making it more appealing to industry and has attracted industry to the STAMPEDE trial.

Gaining regulatory approval for a seamless phase II/III design may be perceived as a barrier to the implementation of an adaptive trial platform. Although this approach is novel in PD, the Medical Research Council Clinical Trials Unit for the STAMPEDE trial has a proven track record for the implementation of adaptive trials within the UK and have experienced regulators to be open minded and helpful in the development of adaptive trial protocols. The MND SMART trial has obtained regulatory approval to commence recruitment in early 2020. This represents a growing recognition that in situations where there is an unmet need for effective therapies, and/or patient populations are limited, such designs have the potential to facilitate ongoing evaluation of drugs over time and are thus worthy of the considerable effort required for their implementation and continued adaptation. Selection of the primary outcomes would need to be made with regulatory agencies in mind and any adaptations made during the course of the trial would also need to be aligned with regulatory requirements.

In order to establish a MAMS trial platform consensus will have to be reached on 1) drug selection, 2) an appropriate patient population for study, 3) methodology for identifying disease modification and whether this is necessary as well as 4) effective and relevant outcome measures. Current phase III studies have been focused on an early PD patient population and whilst practical considerations have played a role in dictating this, recent development in identification of progression subgroups as well as the development of composite measures may open the gate to more inclusive recruitment strategies.

To reach consensus on these issues in PD, a Delphi process is currently being developed to inform the design of a MAMS platform. This method is an iterative approach whereby experts from different backgrounds such as clinicians, funders, industry, academics, regulators and patients will be invited to complete a series of questionnaires reaching consensus over multiple survey rounds.

We propose that the Delphi process should inform the remit for focused working groups to develop trial design, outcome measures and trial infrastructure, thereby ensuring that the wider research community, as well as patients have the opportunity to engage in the development of the MAMS platform.

CONCLUSION

A MAMS trial is an excellent platform that gives the opportunity to simultaneously test the benefit of multiple agents with the added benefit of integrated interim analyses allowing for the removal of arms that do not show sufficient evidence of efficacy. To date, clinical investigation of neuroprotective therapies is hampered by a lack of availability and consensus around methodology and outcome measures. By including a broad field of clinical research stakeholders, as well as patients, it is hoped that a strategy for implementing a MAMS trial will be formulated that is internationally endorsed and of
relevance to patients, thereby giving it the best chance for success.

Dedicated effort over two years in the MS field, which has faced similarly complex issues, especially with regards to outcome measures, has led to a research grant award for the implementation of their MAMS trial. This is encouraging and inspires confidence that a similar approach will make a MAMS trial for PD possible in the near future.

ACKNOWLEDGMENTS

This work is supported by the Cure Parkinson’s Trust

CONFLICT OF INTEREST

The authors have no conflict of interest to report.

SUPPLEMENTARY MATERIAL

The supplementary material is available in the electronic version of this article: https://dx.doi.org/10.3233/JPD-191856.

REFERENCES


Kordower JH, Olanow CW, Dodiya HB, Chu Y, Beach TG, Adler CH, Halliday GM, Bartus RT (2013) Disease duration and the integrity of the nigrostriatal system in PD. *Arch Neurol* 70, 141.


Mollenhauer B, Zimmermann J, Sixel-Döring F, Focke NK, Wicke T, Ebentheuer J, Schaumburg M, Lang E, Friede T, Trenkwalder C (2019) Baseline predictors for progression 4 years after Parkinson’s disease diagnosis in...
the De Novo Parkinson Cohort (DeNoPa). Mov Disord 34, 67–77.


