A framework for prospectively defining progression rules for internal pilot studies monitoring recruitment

Journal Title
XX(X):2–22

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Lisa V. Hampson [1], Paula R. Williamson[2], Martin J. Wilby [3], Thomas Jaki [1]

Abstract

Just over half of publicly funded trials recruit their target sample size within the planned study duration. When recruitment targets are missed, the funder of a trial is faced with the decision of either committing further resources to the study or risk that a worthwhile treatment effect may be missed by an underpowered final analysis. To avoid this challenging situation, when there is insufficient prior evidence to support predicted recruitment rates, funders now require feasibility assessments to be performed in the early stages of trials. Progression criteria are usually specified and agreed with the funder ahead of time. To date, however, the progression rules used are typically *ad hoc*. In addition, rules routinely permit adaptations to recruitment strategies but do not stipulate criteria for evaluating their effectiveness. In this paper we develop a framework for planning and designing internal pilot studies which permit a trial to be stopped early if recruitment is disappointing or to continue to full recruitment if enrolment during the feasibility phase is adequate. This framework enables a progression rule to be pre-specified and agreed upon prior to starting a trial. The novel two-stage designs stipulate that if neither of these situations arises, adaptations to recruitment should be made and subsequently evaluated to establish whether they have been successful. We derive optimal progression rules for internal pilot studies which minimize the expected trial overrun and maintain a high probability of completing the study when the recruitment rate is adequate. The advantages of this procedure are illustrated using a real trial example.

Keywords

Internal Pilot; Operational Feasibility; Operational Power; Progression Rule; Randomised Controlled Trial; Recruitment

Prepared using sagej.cls *[Version: 2015/06/09 v1.01]*

1 Introduction

Many randomised controlled trials (RCTs) struggle to recruit their target sample size within the planned study duration [1;2]. Delays to patient recruitment increase costs and hold up clinical research, with the risk that new research questions will arise before a study can report its findings. Furthermore, trials missing their recruitment targets will be underpowered to detect clinically relevant treatment effects, meaning efficacious treatments may be wrongly abandoned for futility. A review of a cohort of UK publicly-funded multi-centre clinical trials recruiting between 1994-2002 found that only 31% of trials met their original enrolment target and 45% failed to reach even 80% of this figure [3]. Recruitment difficulties persisted in a similar cohort of trials recruiting between 2002-2008, with 45% of trials missing their target sample size. Almost half of the cohort (45%) were awarded either a time extension, additional funding, or both, although only 55% of trials receiving this additional support reached their sample size target [2].

Since there is a limited amount of funding available for research, competing projects must be prioritised. Public funding bodies and commercial enterprises will want to maximise the value of the research portfolio they support, diverting funding away from trials unlikely to meet their objectives on budget and on time, to projects that are. Public funders are now paying closer attention to recruitment and require feasibility assessments in cases where there is insufficient evidence before a trial begins to support predicted recruitment rates. Such assessments can be reliably made in the early stages of a trial: recruitment rates in the first two months of a trial have been found to be correlated with the proportion of the recruitment target eventually accrued [4]. The UK National Institute for Health Research (NIHR) expects all trials funded by its Health Technology Assessment Programme which incorporate an internal pilot study to pre-specify an enrolment target for the pilot phase [5]. If this target is not met, investigators must propose remedial steps to boost accrual and if no suitable modifications can be identified, funding may be withdrawn. The pilot phase of these trials can be viewed as a 'probationary period' which enables funders to screen out those studies for which the recruitment rate is so low that allowing enrolment to continue to completion would result in a lengthy overrun.

Designs for internal pilot studies used to refine sample size calculations are well established [6;7]. In contrast, the design of internal pilot studies to demonstrate the operational feasibility of a trial has received little attention. Lee [8] derives a target recruitment rate and interim minimum recruitment goals which ensure that the final accrual target is met with high probability under the target recruitment rate.

[1] Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK
[2] MRC North-West Hub for Trials Methodology Research, Department of Biostatistics, University of Liverpool, UK
[3] Walton Centre NHS Foundation Trust, Liverpool, UK

Corresponding author:
Lisa Hampson, Medical and Pharmaceutical Statistics Research Unit, Department of Mathematics and Statistics, Lancaster University, Lancaster LA1 4YF, UK
Email: l.v.hampson@lancaster.ac.uk

Prepared using sagej.cls

Gajewski et al.[9] suggest abandoning an on-going trial if there is a high posterior predictive probability that the final analysis will be severely underpowered, but do not propose any concrete designs. Jiang et al.[10] propose a similar criterion, abandoning a trial if the 95% posterior predictive interval for the time to complete recruitment excludes 1.25 times the planned trial duration assuming recruitment follows a homogeneous Poisson process. In our experience, *ad hoc* designs tailored to trials on a case-by-case basis are typically used to monitor recruitment in practice: see, for example, the NERVES trial[11], a multi-centre RCT in sciatica, about which more detail is given in Section 5. In addition, progression rules often permit adaptations to recruitment strategies without imposing criteria for determining their effectiveness.

To address these shortcomings, we formulate recruitment progression rules for internal pilot studies incorporating up to two feasibility assessments. The framework proposes that these rules should be agreed upon prior to the start of recruitment since this will reduce uncertainty for both funders and trialists. At the first analysis, decisions are based on the total number of patients enrolled since the trial opened. Rules recommend progression to the main trial if this total is promisingly high, and recommend abandoning the trial for operational futility if it is inadequate. For intermediate outcomes, investigators must make adaptations to boost recruitment and the effectiveness of these is evaluated at a second analysis based on the number of patients enrolled since the first assessment. At this second analysis, a final decision is taken on whether to progress with recruitment or abandon the trial. Predicting recruitment into a clinical trial is a complex task. As a means of deriving enrolment targets that can be specified and discussed with the funder ahead of time, we formulate optimal progression rules which minimise the average expected overrun of a trial past its planned duration under a pragmatic model for patient and centre accrual. We acknowledge that departures from this model may occur when the trial is conducted which will imply that properties of optimal rules are not attained exactly. Nevertheless, the proposed framework represents a general statistical approach to formulating progression criteria; devising these ahead of time enables the funder to decide whether or not they agree with the proposed criteria before the trial begins. Furthermore, in contrast to *ad hoc* rules, the proposed progression criteria are certain to be optimal in at least one instance. We will investigate the impact of deviations from assumptions on the properties of our proposed progression rules and show that designs remain efficient if departures are small. Pre-specified progression criteria will still prove useful even when unforeseen events external to the trial occur since the planned rule will provide a sensible and efficient starting point which can be adapted.

In Section 3, we define our new two-stage designs for internal pilot studies, before moving on to derive optimal versions in Section 4. Applications to the NERVES trial are considered in Section 5. The optimal timings of feasibility assessments is considered in Section 6, and the article concludes with a brief discussion.

2 Illustrative example

The NERVES (NErve Root block VErsus Surgery) trial (ISRCTN number: ISRCTN04820368) funded by the NIHR is a multi-centre RCT intended to compare the effectiveness of transforaminal epidural steroid injection with surgery for the treatment of sciatica[11]. The primary clinical outcome is the Oswestry Disability Questionnaire score at 18 weeks, assumed to be normally distributed. The trial is currently open to recruitment. A total of 200 patients are required to have 90% power to detect a standardised

effect of 0.5 at the two-sided 5% significance level allowing for a 10% loss to follow-up.

Based on early feasibility work it is estimated that six centres will open to enrolment and 30 patients can be recruited per centre per year. These figures were derived from clinicians reviewing the number of eligible patients and positing potential consent rates. Prior uncertainty about recruitment parameters was enough to prompt the trial investigators to design the first phase of the NERVES trial as an internal pilot study but not so much as to require an external pilot. The pilot phase will target two centres. This number was chosen on pragmatic grounds: targeting two centres means that a specialist unit and a non-specialist referral unit can be opened, but this number avoids committing too many resources to a trial that could be abandoned for operational futility. While the pilot phase is underway, the remaining four centres will be initialised so that they can open immediately upon progression to the main trial. Progression will depend upon several criteria such as satisfactory rates of accrual, consent and treatment switching, although for present purposes we will focus only on the recruitment criterion. Let N represent the total number recruited across both sites during the 6-month pilot phase. The recruitment progression rule states that:

If $N \geq 30$	Progress to complete recruitment using current enrolment strategies.	
Otherwise	Develop strategies to boost recruitment and progress incorporating these adaptations.	(*NERVES*)

Note that while the NERVES internal pilot does not stipulate a stopping rule for recruitment, it does incorporate stopping rules for other criteria. Under the assumed recruitment rate and centre opening times, the expected time needed to recruit 200 patients is 17.3 months. In what follows we will use the NERVES trial and its set-up to illustrate how optimal progression rules can be applied and their operating characteristics.

3 Progression rules for internal pilot studies monitoring recruitment

3.1 Two-stage designs for internal pilot studies

Consider a multi-centre RCT designed to compare the relative efficacy of two treatments labelled A and B. Let θ measure the advantage of A over B. Suppose that the trial's primary objective is to test $H_0 : \theta = 0$ against $\theta \neq 0$ with type I error rate α. Using recruitment as a surrogate for feasibility tacitly assumes that power depends on sample size or, more formally, that Fisher's information for θ is determined by sample size. Indeed, this does hold for many different data types, for example, when the primary endpoint is a continuous, binary, ordinal or count variable [12]. Let $n_{\max}$ denote the total sample size needed to satisfy the power criterion.

Suppose early feasibility work identifies $\mathcal{C}$ centres from which eligible patients can be recruited. Let t represent the time (in months) since commencement of recruitment, so that recruitment begins at calendar time $t = 0$. Furthermore, let t_p denote the initial estimate of the time needed to complete recruitment used to inform funding requests where t_p will be chosen to ensure that the trial's findings will be relevant upon its conclusion. The first stage of the RCT will be conducted as an internal pilot study, where progression to the main trial is assured if early recruitment is sufficient. We consider two-stage designs for the internal

pilot, scheduling feasibility assessments at calendar times t_1 and t_2. Let N_1 denote the total number of patients recruited over the interval $[0, t_1]$ and let N_2 denote the total number of patients recruited over the interval $[t_1, t_2]$. We propose designs with progression rules of the form:

At assessment 1:

If $N_1 \geq u_1$	Progress to complete recruitment using current enrolment strategies.
If $N_1 \leq l_1$	Stop recruitment for operational futility.
Otherwise	Modify enrolment strategies and proceed to feasibility assessment 2.

(1)

At assessment 2:

If $N_2 \geq u_2$	Progress to complete recruitment using modified enrolment strategies.
Otherwise	Stop recruitment for operational futility.

Rule (1) is defined tacitly assuming that recruitment will stop immediately once n_{max} patients have been enrolled. If this target is reached during the pilot phase, 'progression' is unnecessary and the final efficacy analysis can be conducted at once. Setting $l_1 = -1$ in (1) implies that early stopping for operational futility is not permitted at the first interim analysis. Meanwhile, setting $u_1 \geq n_{max}$ implies that progression to the main trial at time t_1 is not permitted, since progression is allowed only when it is unnecessary. Similarly, all choices of u_2 with $u_2 \geq n_{max} - u_1$ imply that progression is not permitted at time t_2. Therefore, we can restrict attention to rules with $-1 \leq l_1 < u_1 \leq n_{max}$ and $0 \leq u_2 \leq n_{max} - u_1$.

Adaptations to enrolment strategies introduced after the first feasibility assessment can take any form. Potential changes include widening the trial eligibility criteria; altering the means of communication with patients; offering incentives; and staff training [13;14]. One could also open additional centres, although we defer consideration of this adaptation until the Discussion. In view of the variety of possible adaptations, strong assumptions would be needed to relate the recruitment rate incorporating adaptations to the recruitment rate without them. Therefore, if the internal pilot study continues to a second stage, the effectiveness of the modifications is assessed on the basis of the number of patients recruited between times t_1 and t_2 only.

The timings of interim analyses also require particular attention. A careful choice of t_1 will ensure that this assessment time is long enough to allow recruitment to get up to speed before the first assessment is conducted. If adaptations are recommended, it will take some time for changes to be communicated to centres and become embedded in practice. Assessment time t_2 must be carefully chosen to be long enough to allow adaptations to take effect but not so long as to lose efficiency if they are ineffective.

3.2 A proposal for centre opening times

For ease of explanation, we propose designs for internal pilot studies which stipulate that centres should open to recruitment in up to three 'waves', as illustrated in Figure 1. However, in Appendix B of the

Supplementary Materials accompanying this manuscript, we describe how progression rules can be found under more complex models for centre opening times. Whilst in principle our designs could be formulated for the case that all $\mathcal{C}$ centres participate in the internal pilot, following the NERVES trial we prefer to stipulate that $c_1 < \mathcal{C}$ centres should participate in the first phase of the pilot. We assume that all c_1 centres will be ready to begin recruiting at time $t = 0$. During the period $[0, t_1]$, the remaining $(\mathcal{C} - c_1)$ centres are initialised so that they can open at time t_1 should the pilot study progress to the main trial at this earliest opportunity. If, however, the pilot phase is extended at time t_1, we stipulate that an additional $c_2 - c_1 < \mathcal{C} - c_1$ centres must open and adaptations to recruitment are rolled out simultaneously across the c_2 centres recruiting from time t_1 onwards. If the pilot study progresses to the main trial at the second feasibility assessment, the remaining $(\mathcal{C} - c_2)$ centres must open at time t_2. The impact on design properties of departures from the recruitment pattern in Figure 1 is assessed in Section 5.4.

Several aspects of our proposed designs merit further discussion. Our preference to involve only a subgroup of centres in the pilot phase reflects current standard practice when there is uncertainty about a trial's operational feasibility: see, for example, the protocols of the NERVES [11], CASPER [15] and FACT [16] trials. Our preference arises because opening centres is not without cost. Staggering opening times reflects a compromise between our wish to open all sites as quickly as possible and our desire to minimise our regret at wasting resources on a trial that is later abandoned for inadequate accrual. Alternatively, if adaptations to recruitment are recommended at time t_1, the decision to open only a subset of the $(\mathcal{C} - c_1)$ newly initialised centres at time t_1 may simply reflect the fact that it will be impractical to roll out adaptations across a large number of centres simultaneously. Having said that, as noted above, our designs could in principle be derived setting $c_1 = c_2 = \mathcal{C}$.

3.3 Choice of pilot centres

So that it is possible to extrapolate from early recruitment trends to make reliable judgements about the feasibility of the main trial, the first groups of c_1 and c_2 centres must be carefully chosen to be representative of the $\mathcal{C}$ centres that will eventually participate in the trial. Here representative means similar in terms of the overall average monthly recruitment rate. We do not have in mind a formal quantitative approach for selecting centres to participate in the internal pilot. Instead, we suggest that pilot centres should be selected on the basis of factors thought *a priori* to be predictive of centre-specific recruitment rates. For example, if accrual rates are expected to be higher in specialist centres than mixed care settings and if the former are to comprise around half of all $\mathcal{C}$ centres, then specialist centres should also comprise 50% of the sites chosen to participate in each stage of the internal pilot. Deliberately choosing all pilot centres to be specialist centres would, in this example, bias upwards interim estimates of the average monthly recruitment rate, with the risk that an infeasible trial would be allowed to progress. For this reason, the NERVES pilot centres are a mixture of specialist and non-specialist referral units.

3.4 Modelling individual patient accrual

Properties of progression rule (1) are evaluated under a pre-specified model for patient accrual in a multi-centre trial. Below we review models for individual patient accrual that have been suggested in

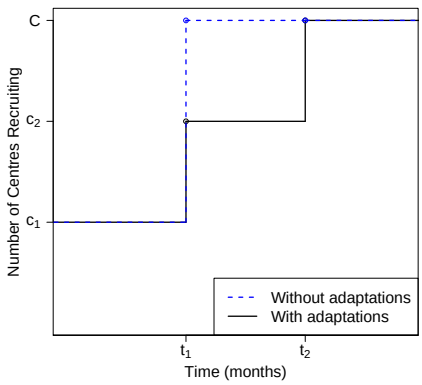


Figure 1. Opening times of $\mathcal{C}$ centres participating in a RCT incorporating progression rule (1).

the literature, and our own pragmatic proposal.

3.4.1 A review of existing models for patient accrual Systematic reviews of the literature [17–19] have found that stochastic recruitment models can take the form of a Poisson process; a two-state Markov model; a first-order autoregressive time-series model; or assume differences between the observed and expected cumulative numbers recruited follow a Brownian motion. Lee [8] and Senn [20] were amongst the first authors to propose a Poisson model for recruitment, where this will be reasonable if one can assume that patients enter a trial independently of one another and there is a negligible probability of two patients arriving simultaneously. Variations proposed for modelling the total number of patients recruited include a homogeneous Poisson process [9;10]; a contagious Poisson model where the recruitment rate for the interval $(t_i,\, t_{i+1})$ is taken to be the total number of patients recruited up to time t_i divided by t_i [8]; a Poisson process with piecewise constant rate over the time interval of interest [21]; and a non-homogeneous Poisson process with time-varying rate modelled using cubic B-splines [22]. Methods [9;10;21;22] model the total number recruited by aggregating across centres and do not incorporate information on centre-specific recruitment rates. However, hierarchical models accommodating centre-specific information have been proposed which assume that recruitment into each centre, once open, follows a homogeneous Poisson process with rate sampled from a Gamma [23;24] or decentred-Pareto [25] distribution.

To derive prior predictions from recruitment models (for example, forecasts of the time needed to complete recruitment), initial guesses at model parameters must be deduced either from relevant completed trials or expert opinion [26], although the latter may be challenging unless model parameters have a direct interpretation. Once a trial is underway, interim predictions can be derived from recruitment models fitted using maximum likelihood estimation [21]; using Bayes theorem to update prior distributions elicited from expert opinion [9;10]; or using an empirical Bayes analysis [23], whereby a Bayesian approach is

taken to predict the remaining recruitment time by substituting prior hyperparameters with maximum likelihood estimates derived from the observed trial data. Assuming a Poisson-Gamma hierarchical recruitment model, Anisimov[24] proposes using accumulating trial data to forecast whether poor performing centres will recruit further patients in a future time interval and adjusts the number of trial centres to ensure accrual can be completed by a specified deadline with high probability.

3.4.2 A proposal for a pragmatic model for patient accrual We formulate progression criteria based on relatively simple, pragmatic, statistical recruitment models that can be specified in advance, rather than rely on complex models whose goodness-of-fit may be difficult to verify even in the early stages of the trial. Specifically we evaluate properties of our proposed progression rules assuming all centres recruit to capacity and there is no competition between centres for patients. Furthermore, we assume that recruitment into each centre follows a Poisson process so that total recruitment into the trial, aggregating across all centres, also follows a Poisson process with monthly rate $\Lambda(t)$ at time t. In the absence of adaptations, recruitment into centre i occurs at a constant monthly rate λ_i, for $i = 1, \ldots, \mathcal{C}$. Let $\lambda = (1/\mathcal{C}) \sum_{i=1}^{\mathcal{C}} \lambda_i$ represent the average monthly recruitment rate per centre. Since a representative sample of centres is selected to participate in the pilot phase, it is reasonable to assume that $\Lambda(t) = c_1\lambda$ for $0 \leq t \leq t_1$, and if the study progresses to the main trial at the first feasibility assessment $\Lambda(t) = \mathcal{C}\lambda$, for $t > t_1$. However, if the pilot study continues to a second stage, adapting enrolment strategies will lead to a shift in the average monthly recruitment rate per centre. Let η denote the fraction increase in λ effected by the adaptations. Then, $\Lambda(t) = c_2\lambda(1+\eta)$ for $t_1 < t \leq t_2$, and this increases to $\Lambda(t) = \mathcal{C}\lambda(1+\eta)$ for $t > t_2$ if progression is recommended at the second feasibility assessment. In what follows, a Bayesian approach will be adopted to accommodate prior uncertainty about λ and η.

3.5 Properties of trials incorporating internal pilots

Several properties are of interest when evaluating recruitment progression rules of the form (1). Define a trial's operational power as the probability of completing recruitment. Assuming that recruitment is always completed once progression has been recommended, this can be said to occur if either: a) n_{max} patients are enrolled during the internal pilot study; or b) the criterion for progression defined in (1) is met at an interim analysis. Therefore, for fixed λ and η, operational power is

$$\begin{aligned} &\text{pr}\{N_1 \geq u_1\} + \text{pr}\{l_1 < N_1 < u_1, N_2 \geq \min\{u_2, n_{\max} - N_1\}\} \\ &= \text{pr}\{X \geq u_1\} + \sum_{n_1=l_1+1}^{u_1-1} \text{pr}\{X = n_1\}\text{pr}\{Y \geq \min\{u_2, n_{\max} - n_1\}\} \end{aligned} \tag{2}$$

where X and Y are Poisson distributed random variables with parameters $c_1\lambda t_1$ and $c_2\lambda(1+\eta)(t_2 - t_1)$, respectively. Similarly, the probability of progression to the main trial at t_2 is

$$\sum_{n_1=l_1+1}^{u_1-1} \text{pr}\{X = n_1\}\text{pr}\{u_2 \leq Y < n_{\max} - n_1\},$$

assuming that progression is unnecessary if $n_{\max}$ or more patients have been recruited by time t_2.

Let T represent the trial duration (in months), where the trial ends when recruitment has been completed or stopped early for operational futility. In Supplementary Appendix A, we show that for fixed values of λ and η the expected overrun of a trial beyond time t_p is

$$\mathbb{E}\left[\max\{0, T-t_p\}; \lambda, \eta\right] = \int_{t_p}^{\infty} (t-t_p) \sum_{n_1=u_1}^{n_{\max}-1} \text{pr}\{X=n_1\} h(t-t_1; n_{\max}-n_1, \mathcal{C}\lambda)\, \mathrm{d}t$$
$$+ \int_{t_p}^{\infty} (t-t_p) \sum_{n_1=l_1+1}^{u_1-1} \sum_{n_2=u_2}^{n_{\max}-n_1-1} \text{pr}\{X=n_1\}\, \text{pr}\{Y=n_2\}\, h(t-t_2; n_{\max}-n_1-n_2, \mathcal{C}\lambda(1+\eta))\, \mathrm{d}t\,, \tag{3}$$

where $h(z; a, b)$ is the probability density function (pdf) of a Gamma(a, b) random variable given by

$$h(z; a,\, b) = \frac{b^a z^{a-1} \exp\{-bz\}}{\Gamma(a)}, \quad z \geq 0.$$

4 Optimising recruitment progression rules

4.1 Optimal design criteria

When choosing the boundaries (l_1, u_1, u_2) of an internal pilot study, we stipulate two properties designs must possess:

(D1) *Designs should recommend adaptations to recruitment with probability at most κ when λ is promisingly large.*

(D2) *Designs should achieve operational power not less than $1-\rho$ under the lowest promising values of λ and η, denoted by λ^{min} and η^{min}.*

Through desideratum (D1), we calibrate designs to limit the probability of requiring adaptations when the trial would otherwise come to a timely conclusion; here we take a promisingly large value of λ to be $\tilde{\lambda}$, the prior best estimate. Power criterion (D2) reflects a wish to complete recruitment and conduct a high powered test of efficacy whenever feasible. We offer some general guidance on setting κ and ρ. For reasons of science and ethics, the probability of completing a feasible trial should be high, so that setting $\rho = 0.1$ or 0.2 will, in general, be appropriate. Meanwhile, the most suitable value for κ will vary from trial to trial depending on the anticipated cost and difficulty of adapting recruitment processes. However, a value for κ between 0.1 and 0.4 will be appropriate in many cases; larger than conventional risks of incorrect decision making will be acceptable since making unnecessary adaptations may be an inconvenience but will not jeopardise patient safety or trial integrity. In all future evaluations of methods, we set $\kappa = 0.15$ and $\rho = 0.1$.

Letting $\tilde{\eta}$ represent the prior best guess at η, one approach to setting operational power would be to specify it under $\tilde{\lambda}$ and $\tilde{\eta}$. However, to guard against using excessively optimistic prior guesses at parameters, we consider whether smaller values are consistent with a tolerable risk of a significant

overrun. To this end, let T_1 represent the time (in months) to completion of recruitment under the design setting $l_1 = -1$ and $u_1 = 0$, so that the internal pilot study progresses to the main trial at t_1 unless recruitment has already been completed. Furthermore, let T_2 denote the time (in months) to completion of recruitment under the design setting $l_1 = -1$, $u_1 = n_{\max}$ and $u_2 = 0$. Then, define $\lambda_L(t_1)$ and $\eta_L(t_1, t_2)$ as solutions to

$$\text{pr}\{T_1 \geq \nu t_p; \lambda = \lambda_L(t_1)\} = \zeta_1 \tag{4}$$

$$\text{pr}\{T_2 \geq \nu t_p; \lambda = \lambda_L(t_1), \eta = \eta_L(t_1, t_2)\} = \zeta_2. \tag{5}$$

The calculations needed to evaluate the probabilities in Equations (4)-(5) are described in Supplementary Appendix A. Since ζ_1 and ζ_2 should represent tolerable risks of a significant overrun, we suggest setting these parameters to values between 0.05 - 0.2. Meanwhile, previous authors have taken a 25% overrun to be significant [10]. Therefore, in future evaluations of methods we set $\nu = 1.25$, $\zeta_1 = 0.05$ and $\zeta_2 = 0.1$. We set $\zeta_2 > \zeta_1$ because trials always progressing at time t_2 rather than t_1 take longer to complete due to the staggered opening times of centres. We write $\lambda_L(t_1)$ and $\eta_L(t_1, t_2)$ to emphasise the dependence of these parameters on t_1 and t_2. This arises because of the impact of t_1 and t_2 on the opening times of centres: earlier scheduling of assessments means recruitment can potentially be rolled out across all $\mathcal{C}$ centres sooner so that smaller values of λ and η are associated with acceptable risks of a significant overrun. We refer to $\lambda^{min} = \min\{\tilde{\lambda}, \lambda_L(t_1)\}$ and $\eta^{min} = \min\{\tilde{\eta}, \eta_L(t_1, t_2)\}$ as the lowest promising values of λ and η, respectively, taking minimums to ensure we consider values both promising and realistic. We specify a design's operational power at $\lambda = \lambda^{min}$ and $\eta = \eta^{min}$.

In the class of designs with desiderata (D1) and (D2), we seek the progression rule which we expect to minimise the overrun of the trial past time t_p. Rather than minimise $\mathbb{E}[\max\{0, T - t_p\}; \lambda, \eta]$ for particular values of λ and η, we minimise a Bayesian version of this criterion, integrating $\mathbb{E}[\max\{0, T - t_p\}; \lambda, \eta]$ with respect to a joint prior density for (λ, η) representing uncertainty about these parameters. Rules minimising this Bayesian criterion can be thought of as minimising the *average* expected overrun of the trial. We characterise uncertainty about λ and η as independent priors derived from (possibly discounted) expert opinion. Strategies for eliciting expert prior opinion on λ and η can be found in Supplementary Appendix C. We acknowledge that there will be an element of subjectivity to individuals' answers to the proposed elicitation questions. However, these opinions will often be based on relevant data. In the UK public sector, almost all funding proposals will be informed by early feasibility work: centres set to participate in the trial will complete screening logs to record the number of patients meeting trial eligibility criteria over a number of weeks. These figures are then discounted to take into account likely consent rates. Experts will be able to draw on these data, as well as their own experiences of similar trials, when stating their opinions. Alternatively, if there was substantial experience of conducting studies in a disease, priors for λ and η could be derived from historical recruitment data by performing a meta-analysis of completed trials [26].

Gajewski et al. [9] model opinion on the expected interarrival time of a Poisson recruitment process as an inverse-gamma distribution. We adopt this approach and model prior opinion on λ as a Gamma(α_1, β_1) distribution. Considering the elicitation of opinion on η, we note that investigators are likely to have some idea *a priori* of the nature of the adaptations to recruitment that would be feasible in their trial. For example, the nature of the intervention may preclude widening the eligibility criteria but it may

be deemed possible to refine the screening process. Experts' prior experiences with previous trials will inform their opinions on the likely impact of such modifications. Of course there is a risk that, once implemented, adaptations would prove ineffective at increasing the average monthly recruitment rate per centre (indeed, modifications investigators are certain would be effective should be incorporated from the outset of the trial if within resource constraints). To capture opinion on this risk we model prior beliefs for η as a mixture distribution, placing weight ω on the scenario that $\eta = 0$ and weight $(1-\omega)$ on the scenario that η is sampled from a Gamma($\alpha_2,\ \beta_2$) distribution. Given these priors, we seek the internal pilot design with desiderata (D1) and (D2) minimising the average expected overrun

$$F = \omega \int_0^\infty \mathbb{E}\left[\max\{0, T - t_p\}; \lambda, \eta = 0\right] h(\lambda; \alpha_1, \beta_1) \mathrm{d}\lambda$$
$$+ (1-\omega) \int_0^\infty \int_0^\infty \mathbb{E}\left[\max\{0, T - t_p\}; \lambda, \eta\right] h(\eta; \alpha_2, \beta_2)\, h(\lambda; \alpha_1, \beta_1)\, \mathrm{d}\eta\, \mathrm{d}\lambda.$$

4.2 Finding optimal progression rules

We use a direct search to find the triple $(l_1^\star, u_1^\star, u_2^\star)$ minimising F in the class of rules with desiderata (D1) and (D2). The computational cost of such a search is manageable because we only need to evaluate the average expected overrun for a subset of designs. To see this, first note that we can restrict our search to the subset of triples with $-1 \leq l_1 < n_{\max}$ and $l_1 < u_1 \leq n_{\max}$ satisfying $\mathrm{pr}\{l_1 < N_1 < u_1; \tilde{\lambda}\} \leq \kappa$. Suppose l_1^a and u_1^a are boundary values satisfying this constraint. Then, fixing $l_1 = l_1^a$ and $u_1 = u_1^a$, equation (2) implies that as u_2 increases, operational power at $\lambda = \lambda^{min}$ and $\eta = \eta^{min}$ decreases due to the more stringent criterion for progression imposed at the second interim. Turning our attention to the expected trial overrun, we see that the integrand of the RHS integral in (3) is the sum of non-negative terms. Since fewer terms are included in this summation as u_2 increases, it follows that for any configuration of λ and η, $\mathbb{E}[\max\{0,\ T - t_p\}; \lambda, \eta]$ is a decreasing function of u_2, which in turn implies that F must be a decreasing function of u_2. Thus, setting u_2 equal to the largest integer $\tau \leq n_{max}$ for which operational power at $\lambda = \lambda^{min}$ and $\eta = \eta^{min}$ is not less than $1 - \rho$, it follows that the design defined by (l_1^a, u_1^a, τ) minimises F in the class of rules satisfying constraints (D1) and (D2) with $l_1 = l_1^a$ and $u_1 = u_1^a$. Repeating this step across all pairs $(l_1,\ u_1)$ satisfying constraint (D1) yields the globally optimal design we seek.

Finding an optimal design requires repeated evaluations of the average expected overrun of a trial, which itself requires evaluation of two- and three-fold integrals. We evaluate integrals numerically in R [27] using the `cubature` package [28]. Setting $c_1 = 2$, $c_2 = 4$, $\mathcal{C} = 6$, $t_1 = 6$, $t_2 = 12$, $n_{\max} = 200$, $\lambda^{min} = 2.1$, $\eta^{min} = 0.115$, it takes 130 minutes on a 2.9 GHz Intel Core i7 processor to find the progression rule minimising F with desiderata (D1) and (D2). Computing time is reduced to 10.1 minutes if the optimal rule is found evaluating F using 10 million simulations. In the examples we have considered, values of F obtained using numerical integration and simulation have been consistent to 2 decimal places. Furthermore, the impact of Monte Carlo error on comparisons of rules can be reduced by basing simulations for different designs on the same sequence of pseudo-random numbers. This less accurate (but quicker) computational approach will likely suffice when the aim is to consider many different testing scenarios in order to develop a broader understanding of the impact of settings on

properties of rules.

4.3 Specifying an internal pilot in the trial protocol

To draw together the ideas presented thus far, we make some recommendations on how to specify the design of a two-stage internal pilot study in a trial protocol. First, the protocol should name the c_1 and c_2 centres that will be targeted in each pilot phase and give a brief justification for why these sites are considered to be representative of the $\mathcal{C}$ centres that will eventually participate in the trial. The protocol should also list the timings of the interim analyses, t_1 and t_2, and the decision criteria that will be applied at each. A complete description of the progression criteria would also define the following parameters: a) the constraints (D1) and (D2), including the values of λ and η at which operational power and the maximum probability of recruitment adaptations are specified; b) the Bayesian optimal design criterion F and the design priors for λ and η under which F is derived. This level of detail will ensure that the progression rule is reproducible. Properties of the design should also be characterised, both under the anticipated centre opening times and under other realistic accrual patterns. Key operating characteristics include the probability of recommending recruitment adaptations, the probability of progressing to full recruitment, expected trial overrun, and the risk of a significant overrun.

5 Application to the NERVES trial

5.1 Properties of the original NERVES progression rule

We begin by considering the single-stage *ad hoc* progression rule implemented in the NERVES trial. Properties of Rule *NERVES* are evaluated given that two centres will participate in the pilot phase and the four remaining centres will open immediately at 6 months. This rule satisfies design criterion (D2) because it does not permit stopping for operational futility. However, it violates criterion (D1) since it recommends adaptations with probability 0.476 when $\lambda = 2.5$.

To illustrate how prior distributions for λ and η might be derived for the NERVES trial, recall that the prior best guess at λ was 2.5: this value will be used as $\tilde{\lambda}$ for the purposes of specifying design constraints (D1) and (D2). Furthermore, suppose that expert opinion states there is a 75% chance that $\lambda \leq 3$. However, since investigators may be overly optimistic about recruitment, it may be preferable to optimise progression rule (1) under a design prior formed by modifying elicited distributions so that λ has a Gamma prior with mode shifted to the left of 2.5 (to 2) and variance preserved at α_1/β_1^2, which gives $\lambda \sim \text{Gamma}(13.519,\ 6.260)$. A mixture distribution for η with weight 0.4 on $\eta = 0$ and 0.6 on $\eta \sim \text{Gamma}(2.900,\ 12.664)$ is implied by the opinion that there is an 60% chance that planned adaptations to recruitment will be effective and the conditional beliefs that given $\eta > 0$, the most likely value is $\tilde{\eta} = 0.15$ and there is a 75% chance that $\eta \leq 0.3$. These hypothetical opinions appear feasible given the range of results reported by studies comparing strategies to improve recruitment in RCTs [13]. Assuming these design priors for λ and η, the average expected overrun of the NERVES trial is $F = 2.807$ months. We compare properties of the original NERVES progression rule with those of optimal two-stage designs below.

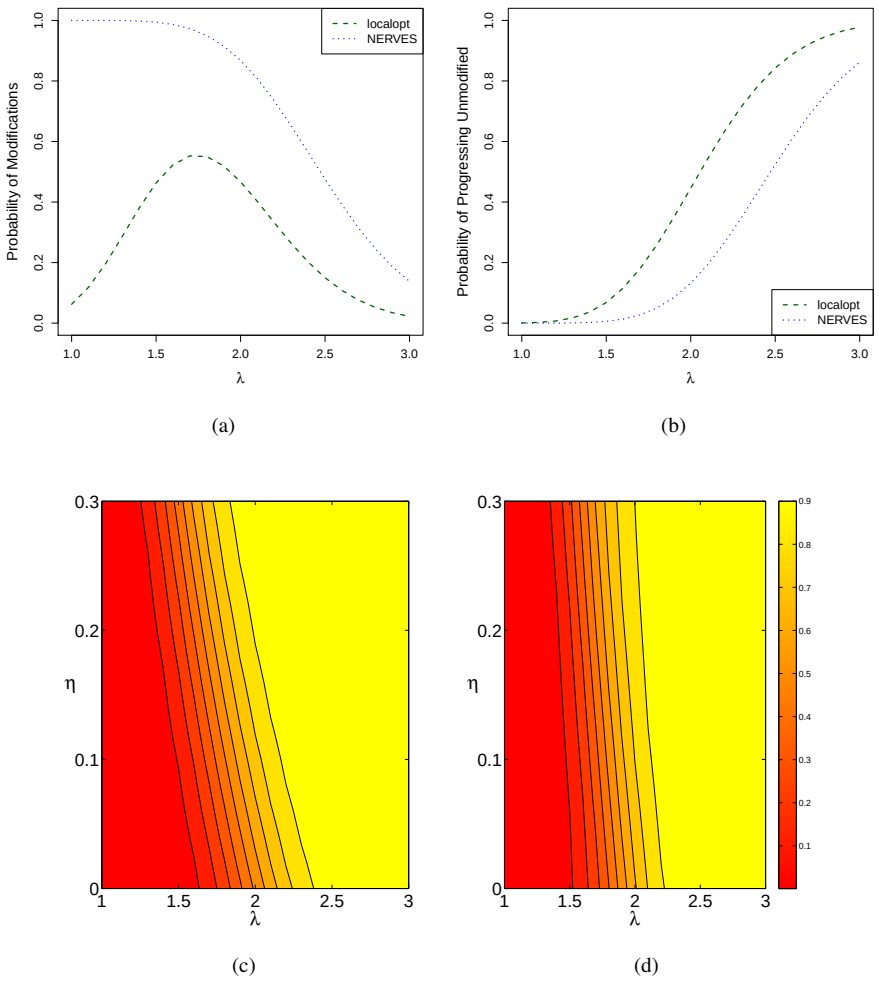
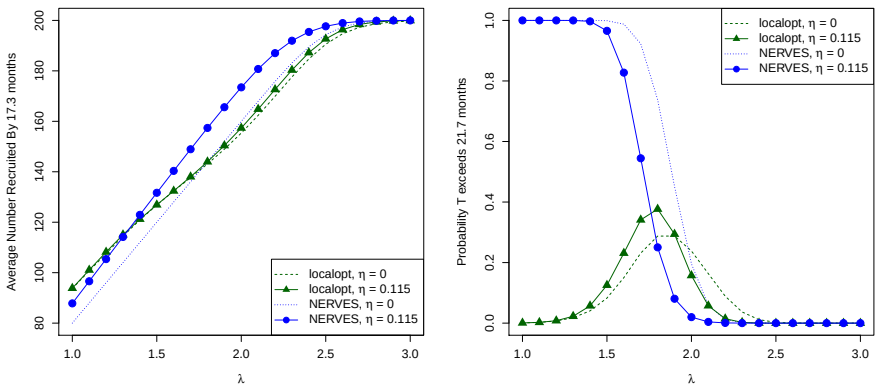


Figure 2. a) Probability of recommending adaptations to recruitment for Rules *NERVES* and *localopt*; b) Probability of progressing to the main trial at time t_1 to complete recruitment for Rules *NERVES* and *localopt*; c) Conditional probability under Rule *localopt* of progressing to complete recruitment at time t_2 given progress to Stage 2 of the pilot study; d) Probability under Rule *localopt* of completing recruitment. All evaluations of rules set $c_1 = 2$, $c_2 = 4$, $\mathcal{C} = 6$, $t_1 = 6$ and $t_2 = 12$. Rule *localopt* is defined with (l_1 = 17, u_1 = 25, u_2 = 48). Contour plots have contours drawn at probability intervals of 0.1.

5.2 Optimal progression rules for the NERVES trial

We derive two-stage rules of the form (1) for the NERVES trial, preserving the time of the first feasibility assessment as $t_1 = 6$ months and fixing $t_2 = 12$ months. We consider rules controlling the



(a) (b)

Figure 3. Properties of Rules *NERVES* and *localopt*: a) Expected number of patients recruited by time t_p for studies progressing to the main trial to complete recruitment; b) Probability that the trial duration exceeds t_p by more than 25%. All evaluations set $c_1 = 2$, $c_2 = 4$, $\mathcal{C} = 6$, $t_1 = 6$, $t_2 = 12$, $n_{\max} = 200$ and $t_p = 17.3$. Expected numbers recruited are based on 10 million simulations. Rule *localopt* is defined with $(l_1 = 17, u_1 = 25, u_2 = 48)$.

maximum probability of recruitment adaptations at 0.1 when $\lambda = 2.5$. Operational power is stipulated at $\lambda^{min} = \min\{2.5, \lambda_L(t_1)\} = 2.112$ and $\eta^{min} = \min\{\tilde{\eta}, \eta_L(t_1, t_2)\} = 0.115$, where $\lambda_L(t_1 = 6)$ and $\eta_L(t_1 = 6, t_2 = 12)$ are found as solutions to Equations (4) - (5). Then, setting $c_1 = 2$, $c_2 = 4$, $\mathcal{C} = 6$, $n_{\max} = 200$, we find the design minimising F under the Gamma$(13.519,\ 6.260)$ prior for λ and mixture prior for η defined above. This locally optimal rule, labelled 'Rule *localopt*', stipulates that if 17 or fewer patients are recruited in the first 6 months of the trial, accrual must stop. If 25 or more patients are recruited, the trial should progress to complete recruitment without modifications or further feasibility assessments. Otherwise, if between 17 and 25 patients are recruited, adaptations are required and a second feasibility assessment must be conducted at 12 months to evaluate their effectiveness. At the time of this second assessment, if fewer than 48 patients have been recruited in the last 6 months, the trial must stop, otherwise it must progress to complete recruitment incorporating changes to recruitment. The average expected overrun of the trial under this design is $F = 1.235$ months.

Recall that T represents trial duration and assume a significant overrun occurs if $T \geq 1.25\, t_p$. Figures 2 and 3 compare properties of Rule *NERVES* with those of Rule *localopt* as the true average monthly recruitment rate per centre, λ, varies. We see that the two-stage rule recommends modifications to recruitment with lower probability than the single-stage Rule *NERVES*. Figure 3(b) highlights further advantages of permitting early stopping for operational futility and of evaluating adaptations to recruitment strategies. Specifically, for trials incorporating two-stage progression rules, the risk of a significant overrun approaches 0 for extreme values of λ since in these scenarios the trial will almost surely either stop early for operational futility, incurring no overrun, or progress to complete recruitment

within t_p months. However, since Rule *NERVES* does not permit early stopping, the risk of a significant overrun under this rule quickly approaches 1 as λ falls below 2. Additionally, not evaluating the effectiveness of recruitment adaptations means that if modifications are ineffective, under Rule *NERVES* the trial would incur a large penalty in terms of an increased risk of a significant overrun. This penalty is smaller for the two-stage design. Studying the risk curve for Rule *localopt*, we see that for $\lambda < 1.906$, the risk of a significant overrun is greater when $\eta = 0.115$ than when $\eta = 0$ because when $\eta = 0.115$ there is a greater probability of progressing at Stage 2 to complete recruitment rather than abandoning the trial. The ordering of the risk curves is reversed for $\lambda > 1.906$ as trials progressing at Stage 2 become more likely to complete accrual within $1.25\,t_p$ months. Comparing Figures 3(a) and 3(b), we see that for certain pairs of λ and η, the risk of a significant overrun under Rule *NERVES* is greater than under Rule *localopt* but we expect to recruit more patients by time t_p. In these cases, the distribution of T has heavier tails under Rule *NERVES* but the probability that $T \leq t_p$ is greater because this single-stage rule recommends adaptations with higher probability and rolls out accrual across all $\mathcal{C}$ centres after six months, whereas this may be postponed until 12 months under the two-stage design.

Figure 4 evaluates the robustness of the efficiency of Rule *localopt* to a prior-data conflict: in this scenario we may regret optimising the progression rule under priors for λ and η which place large probability mass on parameter values inconsistent with the observed data. It is clear that the relative efficiency of Rule *localopt* is poor when the true average monthly recruitment rate per centre, λ, lies in the tails of the Gamma$(13.519,\ 6.260)$ prior distribution used for optimisation. However, for values of λ within the 50% central interval of $(1.740,\ 2.521)$, the efficiency of Rule *localopt* is within 67% of that of the locally optimal design for all values of η considered; for many pairs of parameter values, the relative efficiency is within 20%. The robustness of the efficiency of Rule *localopt* to deviations from the assumption that the average monthly recruitment rate per centre remains constant across the pilot and main phases of the trial is examined in Supplementary Appendix D.

5.3 Impact of progression rules on efficacy testing

The attained error rates of the test of H_0 will deviate from their nominal values if early stopping for operational futility is permitted. For example, if we are permitted to test H_0 only if the progression criteria are met and accrual is completed, the overall type I error rate will be less than α because a trial may be stopped early for operational futility without rejecting H_0 in cases where it otherwise would have gone on to commit a type I error. While we do not advocate adjusting the final efficacy analysis for the possibility of early stopping for operational futility, we do recommend exploring the impact of monitoring recruitment on a trial's power ahead of time.

Table 1 quantifies what the impact of Rule *localopt* would be on the power and type I error rate of the NERVES trial. Values of λ are indexed by the expected trial duration calculated as $t_1 + (n_{\max} - c_1 t_1 \lambda)/(\mathcal{C}\lambda)$. For present purposes we assume that all recruited patients are followed-up for their primary endpoint (that is, there are no missing data), meaning that in the absence of feasibility testing NERVES has 94.2% power to detect a standardised effect of 0.5 at the two-sided 5% significance level. When Rule *localopt* is incorporated into the NERVES trial, power decreases when there is a non-negligible probability of stopping early. If we permit efficacy testing only when the target sample size is met,

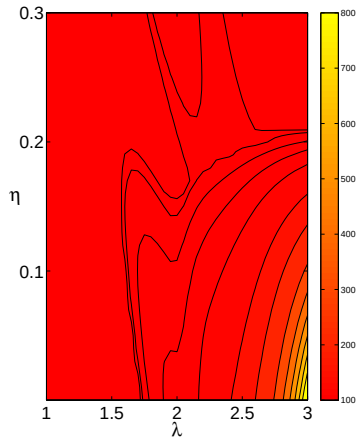


Figure 4. Value of $\mathbb{E}[\max\{T - t_p, 0\}; \lambda, \eta]$ attained by Rule *localopt* expressed as a percentage of the minima attained by the locally optimal design. All locally optimal designs are derived setting $c_1 = 2$, $c_2 = 4$, $\mathcal{C} = 6$, $t_1 = 6$, $t_2 = 12$, $n_{\max} = 200$ and $t_p = 17.3$, and specifying operational power at $\lambda = 2.1$ and $\eta = 0.115$. Contours are drawn at 101, 102, 105, 110, 120, 150 and 200% and at every 100 percentage points thereafter.

Table 1. Operating characteristics of an RCT incorporating Rule *localopt* when efficacy testing is: a) always permitted on termination of the trial; and b) only permitted if $n_{\max} = 200$ subjects are recruited. Properties relate to a two-sided test of $H_0 : \theta = 0$ conducted setting $\alpha = 0.05$, assuming data are normally distributed and θ is a difference in means. Listed are attained two-sided type I error rates and power for detecting a standardised effect of 0.5. The probability of early stopping is the probability of stopping for operational futility at any feasibility assessment. All properties are calculated under $\eta = 0.115$ setting $c_1 = 2$, $c_2 = 4$, $\mathcal{C} = 6$, $t_1 = 6$ and $t_2 = 12$.

True average accrual length (months)	Testing scenario a) Power	Type I Error	Testing scenario b) Power	Type I Error	Probability stop early	E(N) when stop early
15.0	0.942	0.050	0.942	0.050	2.791×10^{-4}	16.351
17.3	0.936	0.050	0.935	0.050	0.008	22.075
20.0	0.873	0.050	0.834	0.044	0.115	39.394
22.0	0.729	0.050	0.595	0.032	0.369	43.087

deviations in power and type I error rate increase as λ decreases. Losses in power are ameliorated and the overall type I error rate preserved at its nominal level when efficacy testing is always permitted on trial termination. However, losses in power still persist because efficacy tests are underpowered when recruitment is abandoned before completion. Table 1 lists results for $\eta = 0.115$; increasing η leads to a reduced probability of early stopping and thus smaller deviations in power.

5.4 Robustness of optimal progression rules to deviations from recruitment assumptions

In practice, centres must obtain ethics and R&D approvals before they can open to recruitment. We refer to centres in receipt of these approvals as 'initialised'. Unforeseen delays in obtaining approvals may mean that centre opening times cannot follow the pattern envisaged in Figure 1 [29]. Such departures will imply that properties of progression rules deviate from their pre-specified values. Let $N_I(t)$ denote the total number of centres initialised after the trial has been open for t months. Supplementary Figure S1 illustrates centre opening times for the NERVES trial under progression rule (1) when

$$N_I(t) = \min\{\mathcal{C},\, c_1 + \lfloor \gamma t \rfloor\} \qquad \text{for } t \geq 0, \tag{6}$$

setting $t_1 = 6$ and $t_2 = 12$, and where γ is a constant reflecting the rate at which new centres are initialised. Supplementary Appendix B describes how optimal versions of rule (1) are derived when initialisation times follow model (6) with $\gamma = 0.3$ and $\gamma = 0.5$.

Supplementary Figure S2 assesses the robustness of properties of Rule *localopt* to delays in centre openings. Small departures from the anticipated pattern imply deviations from design properties are also small. If $\gamma = 0.5$ in model (6), the initialisation of centres is only marginally slower than was anticipated when the internal pilot was designed: Rule *localopt* (l_1 = 17, $u_1 = 25$, u_2 = 48) remains the optimal rule for this modified pattern, with an average expected overrun of $F = 1.330$ months. If, however, $\gamma = 0.3$, centres are much slower to open than was initially expected. If we had optimised the progression rule for this accrual pattern, Rule *localopt* would no longer satisfy power criterion (D1) and instead the optimal progression rule with desiderata (D1) and (D2) has boundaries (l_1 = 17, u_1 = 25, u_2 = 35) and achieves $F = 2.242$ months.

6 Impact of varying assessment schedules

So far we have followed the NERVES trial to assume that feasibility assessments will be scheduled at six-monthly intervals. To explore the impact of varying the feasibility assessment schedule, we optimise the interim assessment timings t_1 and t_2. In practice, these timings would typically be rounded to the nearest month. With this in mind, we restrict our search for optimal times to integer pairs satisfying the constraints $\pi_1 \leq t_1 \leq t_p - \pi_1$ and $t_1 + \pi_1 \leq t_2 \leq t_p$. The interval π_1 must be long enough for our assumption that all $\mathcal{C}$ centres will be ready to open by $t_1 = \pi_1$ to be plausible. More generally, we believe that designs with $\pi_1 \approx 0$ would be impractical because: a) it would be difficult to justify stopping a trial at the first interim analysis on the basis of very few data; and b) if adaptations are made to recruitment processes at an interim analysis, π_1 must be long enough to allow centres to become familiar with new trial procedures and begin recruiting at their optimal rate before the next feasibility assessment.

Considering the NERVES trial, we set $\pi_1 = 4$ along with $c_1 = 2$, $c_2 = 4$, $\mathcal{C} = 6$, $n_{\max} = 200$ and $t_p = 17.3$, and perform a direct search over pairs (t_1, t_2) satisfying our constraints. For each pair, we find the progression rule with desiderata (D1) - (D2) minimising F under the design priors for λ and η defined in Section 5.1. Operational power is specified at $\lambda = \min\{2.5, \lambda_L(t_1)\}$ and $\eta = \min\{0.15, \eta_L(t_1, t_2)\}$, where $\lambda_L(t_1)$ and $\eta_L(t_1, t_2)$ are found as solutions to Equations (4) - (5). We find that the optimal

assessment schedule is $t_1 = 4$ and $t_2 = 12$ months. All evaluations of F were performed using numerical integration. The optimal rule for this analysis schedule, labelled 'Rule *globalopt*' is defined by (l_1 = 7, u_1 = 15, u_2 = 66) and achieves $F = 1.100$ months. Setting $\lambda^{min} = \min\{2.5, \lambda_L(t_1)\} = 1.964$ and $\eta^{min} = \min\{0.15, \eta_L(t_1, t_2)\} = 0.147$, this rule has operational power 0.903 under $\lambda = \lambda^{min}$ and $\eta = \eta^{min}$, and recommends recruitment adaptations with probability 0.104 when $\lambda = 2.5$. Comparing Rules *localopt* and *globalopt* in Figure 5 reveals the impact of conducting the second feasibility assessment with more data. Figure 5(a) shows that as the true average monthly recruitment rate per centre tends towards 1, the conditional probability of progression at the second assessment under Rule *localopt* is slower to decrease than this probability under Rule *globalopt*. To explain this, recall that we model $N_2 \sim \text{Pois}(c_2\lambda(1 + \eta)(t_2 - t_1))$. Therefore reducing the length of the interval $[t_1, t_2]$ means smaller shifts in the location of the distribution of N_2 result from varying λ or η. Thus Rule *localopt* is less able than Rule *globalopt* to distinguish between promising and inadequate recruitment scenarios. Since Rule *globalopt* is powered to complete recruitment under a lower value of λ, probabilities of completing recruitment are larger than under Rule *localopt*. Thus, it follows that the risk of a significant overrun is greater for Rule *globalopt* when the true value of λ is small.

7 Discussion

In this paper we have developed two-stage recruitment progression rules for internal pilot studies that can be specified ahead of time and which permit early stopping for operational futility and the evaluation of adaptations to recruitment procedures. These designs stipulate that centres should enter the trial in up to three waves at times $t = 0$, t_1 and t_2. However, we have shown that our methods can be extended to incorporate more complex models for centre opening times, for example, to accommodate non-constant total recruitment rates during the intervals (t_1, T) and (t_2, T) which may arise if it is unreasonable to stipulate that $\mathcal{C}$ centres should be initialised by the time of the first feasibility assessment.

Our proposed methods could be extended in several ways. For instance, it would be straightforward to derive optimal progression rules minimising a different criterion to the average expected trial overrun; suitable alternatives include the average risk of a significant overrun or a large percentile of the marginal distribution of $\max\{T - t_p, 0\}$. Alternatively, our two-stage designs could be applied using different (non-optimal but intuitive) criteria to set (l_1, u_1, u_2). In particular, one alternative seems worthy of further consideration, motivated by the fact that the most controversial outcome of an internal pilot study would be to recommend early stopping at the first interim analysis. An intuitive approach to specifying l_1 would be to set $(l_1 + 1)$ as the smallest N_1 such that if adaptations to recruitment were made at time t_1, the value of η needed to ensure that the posterior predictive probability $\xi = \text{pr}\{T \leq 1.25\, t_p \mid \eta, N_1\}$ exceeded 0.8 is considered feasible; here ξ would be calculated assuming the trial would always progress to completion at time t_2. The remaining boundaries u_1 and u_2 would then be fixed by constraints (D1) and (D2). Fixing l_1 in this manner implies that if the trial reaches the first interim analysis with $N_1 = l_1$, we would conclude that an infeasibly large boost to recruitment is needed to ensure the risk of a significant overrun remains acceptable. Of course, the trialist and funder would need to pre-specify what boost to the average monthly recruitment rate would be considered infeasible. However, the direct interpretation of l_1 may prove helpful for trialists seeking to understand why their trial has been abandoned at the earliest

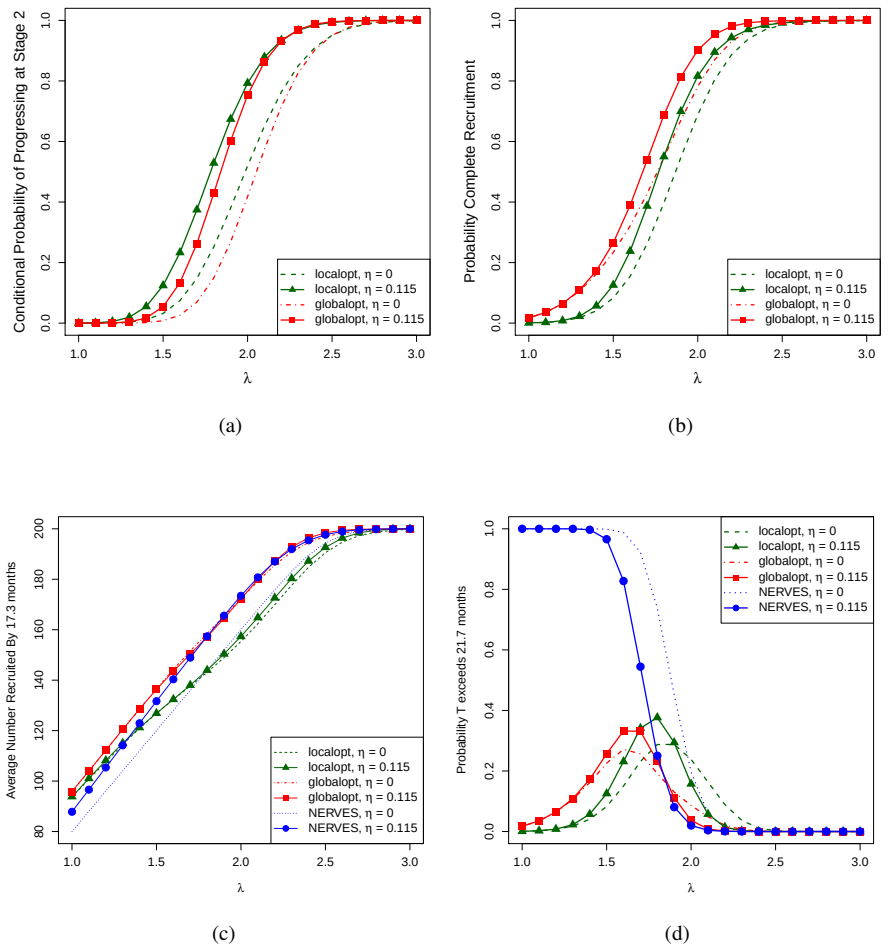


Figure 5. Properties of Rules *NERVES*, *localopt* and *globalopt*: a) Conditional probability of progressing to the main trial at time t_2 to complete recruitment given we proceed to Stage 2 of the pilot study; b) Probability of completing recruitment; c) Expected number of patients recruited by time t_p for studies progressing to the main trial to complete recruitment; d) Probability that the trial duration exceeds t_p by more than 25%. All evaluations set $c_1 = 2$, $c_2 = 4$, $\mathcal{C} = 6$, $n_{\max} = 200$ and $t_p = 17.3$. Expected numbers recruited are based on 10 million simulations. Rule *localopt* is defined with (l_1 = 17, u_1 = 25, u_2 = 48) while Rule *globalopt* is defined with (l_1 = 7, u_1 = 15, u_2 = 66).

opportunity.

Throughout we have restricted attention to two-stage designs, assuming that the timeframe of most trials will not allow for the implementation and assessment of multiple adaptations to recruitment. However, this may be feasible and desirable in long-term trials with a recruitment period lasting several years. To extend our approach to this setting, additional constraints would be needed to ensure unique optimal designs exist. On another note, the pattern of centre opening times envisaged in Figure 1 tacitly assumes that the option of boosting recruitment by opening extra centres is unavailable. However, it would be straightforward to extend our framework to accommodate this case, the only change being that we would anticipate recruiting across different numbers of centres should the trial progress with adaptations and without them.

Declaration of conflicting interests

The Authors declare that there is no conflict of interest

Funding

Lisa Hampson was supported by Medical Research Council Grant MR/J014079/1. Paula Williamson was funded by the MRC North-West Hub for Trials Methodology Research (MR/K025635/1). Thomas Jaki was supported by the National Institute for Health Research Grant NIHR-SRF-2015-008-001. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the National Institute for Health Research or the Department of Health.

Supplemental material

Supplementary figures and derivations supporting results stated in the main text are presented in an additional document, which is distributed with the paper as web-based supporting materials.

References

1. Bower P, Wilson S and Mathers N. How often do UK primary care trials face recruitment delays? Family Practice 2007; 24: 601–603.
2. Sully BGO, Julious SA and Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. Trials 2013; 14: 166–174.
3. Campbell MK, Snowdon C, Francis D et al. Recruitment to randomised trials: strategies for trial enrolment and participation study. The STEPS study. Health Technology Assessment 2007; 11(48).
4. Haidich AB and Ioannidis JPA. Patterns of patient enrollment in randomized controlled trials. Journal of Clinical Epidemiology 2001; 54: 877–883.
5. National Institute for Health Research. Recruitment rate progression rules for internal pilot studies for HTA trials. Available at http://www.themedcalls.nihr.ac.uk_data/assets/pdf_file/0003/75198/Progression-rules-for-internal-pilot-studies.pdf.
6. Wittes J, Schabenberger O, Zucker D et al. Internal pilot studies I: type I error rate of the naive t-test. Statistics in Medicine 1999; 18: 3481–3491.
7. Zucker DM, Wittes JT, Schabenberger O et al. Internal pilot studies II: comparison of various procedures. Statistics in Medicine 1999; 18: 3493–3509.
8. Lee Y. Interim recruitment goals in clinical trials. Journal of Chronic Diseases 1983; 36: 379–389.

9. Gajewski BJ, Simon SD and Carlson SE. Predicting accrual in clinical trials with Bayesian posterior predictive distributions. Statistics in Medicine 2008; 27: 2328–2340.
10. Jiang Y, Simon S, Mayo MS et al. Modeling and validating Bayesian accrual models on clinical data and simulations using adaptive priors. Statistics in Medicine 2015; 34: 613–629.
11. National Institute for Health Research. HTA- 12/201/10: Multi-centre randomised control trial comparing the clinical and cost effectiveness of trans-foraminal epidural steroid injection to surgical microdiscectomy for the treatment of chronic radicular pain secondary to prolapsed intervertebral disc herniation. Available at http://www.nets.nihr.ac.uk/projects/hta/1220110.
12. Whitehead J. The Design and Analysis of Sequential Clinical Trials (Revised 2nd Edition). Chichester: Wiley, 1997.
13. Treweek S, Mitchell E, Pitkethly M et al. Strategies to improve recruitment to randomised controlled trials (Review). The Cochrane Library 2010; .
14. Bower P, Brueton V, Gamble C et al. Interventions to improve recruitment and retention in clinical trials: a survey and workshop to assess current practice and future priorities. Trials 2014; 15: 399–407.
15. National Institute for Health Research. HTA- 08/19/04: Collaborative care and active surveillance for screen-positive elders with sub-clinical depression: a pilot study and definitive and randomised evaluation - the CASPER trial. Available at http://www.nets.nihr.ac.uk/projects/hta/081904.
16. National Institute for Health Research. HTA- 13/04/46: A randomised comparison of femtosecond laser assisted vs manual phacoemulsification cataract surgery for adults with visually significant cataract: the FACT trial. Available at http://www.nets.nihr.ac.uk/projects/hta/130446.
17. Barnard K, Dent L and Cook A. A systematic review of models to predict recruitment to multicentre clinical trials. BMC Medical Research Methodology 2010; 10: 63–70.
18. Zhang X and Long Q. Modelling and prediction of subject accrual and event times in clinical trials: a systematic review. Clinical Trials 2012; 9: 681–688.
19. Heitjan D, Ge Z and Ying G. Real-time prediction of clinical trial enrollment and event counts: A review. Contemporary Clinical Trials 2015; : online.
20. Senn S. Some controversies in planning and analysing multi-centre trials. Statistics in Medicine 1998; 17: 1753–1765.
21. Tang G, Kong Y, Chang C et al. Prediction of accrual closure date in multi-center clinical trials with discrete-time Poisson process models. Pharmaceutical Statistics 2012; 11: 351–356.
22. Zhang X and Long Q. Stochastic modeling and prediction for accrual in clinical trials. Statistics in Medicine 2010; 29: 649–658.
23. Anisimov VV and Fedorov VV. Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. Statistics in Medicine 2007; 26: 4958–4975.
24. Anisimov V. Statistical modelling of clinical trials (recruitment and randomization). Communications in Statistics - Theory and Methods 2011; 40: 3684–3699.
25. Mijoule G, Savy S and Savy N. Models for patients' recruitment in clinical trials and sensitivity analysis. Statistics in Medicine 2012; 31: 1655–1674.
26. Bakhshi A, Senn S and Phillips A. Some issues in predicting patient recruitment in multi-centre clinical trials. Statistics in Medicine 2013; 32: 5458–5468.
27. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, 2013. URL `http://www.R-project.org/`.

28. Johnson SG and Narasimhan B. cubature: Adaptive multivariate integration over hypercubes, 2013. URL http://CRAN.R-project.org/package=cubature. R package version 1.1-2.
29. Kearney A, McKay A, Hickey H et al. Opening research sites in multicentre clinical trials within the UK: a detailed analysis of delays. BMJ Open 2014; 4: e005874.