Patient recruitment forecasting in clinical trials using time-dependent Poisson-gamma model and homogeneity testing criteria

Volodymyr Anisimov*
Data Science
Center for Design & Analysis
Amgen, London, UK

Lucas Oliver
Data Science
Center for Design & Analysis
Amgen, London, UK

Abstract

Clinical trials in the modern era are characterized by their complexity and high costs and usually involve hundreds/thousands of patients to be recruited across multiple clinical centres in many countries, as typically a rather large sample size is required in order to prove the efficiency of a particular drug.

As the imperative to recruit vast numbers of patients across multiple clinical centres has become a major challenge, an accurate forecasting of patient recruitment is one of key factors for the operational success of clinical trials. A classic Poisson-gamma (PG) recruitment model assumes time-homogeneous recruitment rates. However, there can be potential time-trends in the recruitment driven by various factors, e.g. seasonal changes, exhaustion of patients on particular treatments in some centres, etc. Recently a few authors considered some extensions of the PG model to time-dependent rates under some particular assumptions. In this paper, a natural generalization of the original PG model to a PG model with non-homogeneous time-dependent rates is introduced. It is also proposed a new analytic methodology for modelling/forecasting patient recruitment using a Poisson-gamma approximation of recruitment processes in different countries and globally. The properties of some tests on homogeneity of the rates (non-parametric one using a Poisson model and two parametric tests using Poisson and PG model) are investigated. The techniques for modeling and simulation of the recruitment using time-dependent model are discussed. For re-projection of the remaining recruitment it is proposed to use a moving window and re-estimating parameters at every interim time. The results are supported by simulation of some artificial data sets.

1 Introduction

Contemporary late-phase clinical trials involve hundreds or even thousands of patients to be recruited across multiple clinical centres in many countries as typically it is required to achieve a rather large sample size to prove the efficiency of a particular drug. Citing Ken Getz, director, Tufts Center for Study of Drug Development, USA, patient recruitment and retention are among the greatest challenges that the clinical research enterprise faces today, and they are a major cause of drug development delays.

To address these challenges, novel approaches and predictive analytic techniques are required for efficient data analysis, monitoring, and decision-making.

Consider typical multicentre clinical trials used in pharmaceutical industry where the patients are recruited by many clinical centres in different countries and then randomized to different treatments.

---

*E-mail: vanisimo@amgen.com

Using Poisson models for modelling patient recruitment in clinical trials is a well-accepted approach. Several papers use Poisson processes in different clinical centres with fixed recruitment rates (Carter et al., [19]; Senn [27, 28].

However, in real trials different centres typically have different capacity and productivity, and the recruitment rates vary. To model this variation, Anisimov and Fedorov [1, 2] introduced a Poisson-gamma (PG) model, where the patient arrival at clinical centres is modeled using Poisson processes with some rates and the variation in rates among different centres is modeled using a gamma distribution.

This model can be described in the framework of the empirical Bayesian approach where the prior distribution of the rates is a gamma distribution with parameters that are evaluated either using historical data or data provided by study investigators. In [1] a maximum likelihood technique was proposed for estimating the parameters of the rates and the Bayesian technique for adjusting the posterior distribution of the rates at any interim time using recruitment data in the individual centres. Various applications to real trials are considered in [1, 3, 4, 5].

This technique was developed further to account for random delays and closure of clinical centres (Anisimov [3, 4, 5, 8]) and for modelling events in event-driven trials (oncology) (Anisimov [6, 8] and Anisimov et al. [11]).

Note that independently Gajewski et al. [20] considered later a similar approach to modelling recruitment as a Poisson process with gamma distributed rate, however their model is applicable to only one clinical centre.

The PG model was also used in Mijoule et al. [24] to consider some extensions related to other distributions of the recruitment rates and sensitivity analysis, in Minois et al. [25] to evaluate the duration of recruitment process when historical trials are available, and in Bakhshi et al. [15] for the evaluation of parameters of a PG model using meta-analytic techniques for historic trials. A survey on using mixed Poisson models is provided in a discussion paper [7].

It is also worth mentioning the recent investigations devoted to centralized statistical monitoring and forecasting recruitment performance and also forecasting patient recruitment under various restrictions and creating an optimal cost-efficient recruitment design, Anisimov and Austin [9, 10, 12, 13].

There are also other approaches to recruitment modelling described in the literature, however, they are mainly dealing with the analysis of the global recruitment and therefore have some limitations. These approaches typically require rather large number of centres and patients and cannot be applied to predicting recruitment on centre/country level (see survey papers by Barnard et al. [16]; Heitjan et al. [22], Gkioni et al. [21].

Currently, the originally developed PG model, [1, 5], gained world-wide recognition and in some recent papers is now called "one of the most popular techniques" and an "industry-standard" model.

This model assumes that patient recruitment rates do not change over time. However, in real trials there can be some time-trends, seasonal changes, and recruitment can be changing in some countries over time due to different reasons, e.g. a slowdown in recruitment near the end of a trial, etc.

Therefore, to capture these situations, in some recent papers [18, 23, 26, 29, 30, 31], the standard PG model was extended to the cases where the recruitment rates can be time-dependent.

Lan et al. [23] proposed a non-homogeneous Poisson process model that allows for staggered centre activation and heterogeneity within centres is modeled by a gamma distribution and assumes that after a period of steady recruitment, the centre mean recruitment rate gradually declines as a negative exponential with some coefficient. They calculate the posterior distribution and consider some applications to real trials. Urbas et al. [31] proposed a more general time-dependent PG model that allows for a wider range of recruitment rate functions. The model is fitted using a maximum likelihood approach and is used to select the best model among a set of candidate models. Perperoglou et al. [26] compared the performance of the standard PG model to two time-dependent models: the models proposed in [23] and [31]. The study found that the time-dependent models outperformed the PG model in terms of prediction accuracy, especially in trials with time-varying recruitment rates. Best et al. [18] considered a standard PG model as the starting point and developed a more flexible version that introduces variation in rates over time as a function of COVID-19-dependent covariates which is implemented in a fully Bayesian probabilistic framework. Turchetta et al. [29] proposed a time-dependent PG model allowing recruitment rates to vary over time through B-splines that is fitted using a Bayesian approach. The model is evaluated in a simulation study and found to perform well in a variety of scenarios. In another paper, Turchetta et al. [30] consider predictive performance of the proposed technique in forecasting the recruitment process of two HIV vaccine trials.

All six papers above contribute to the advancement of recruitment modelling in multicentre clinical trials by extending the standard time-homogeneous PG model to accommodate time-dependent recruitment rates.

The flexibility of the original PG model allows for rather natural extension to model time-dependent rates on a centre level. However, the main question here is: what type of time-dependence to use? Also, how to detect the time-dependent behavior of the rates, and how to test for this dependence in real trials. Thus, the choice of a suitable time-dependent model is trial specific.

In this paper, the authors consider a natural generalization of a PG model to a time-dependent model using a general framework of a standard PG model [1, 5] which is similar to the way considered in [31]. The basic methodology of using a PG model with time-dependent rates is developed by the first author in [14]. This paper extends some results of [14], in particular, we investigate two new parametric criteria for testing the recruitment rates for homogeneity, a Poisson criterion and a PG one in addition to a non-parametric Poisson criterion [14]. In [14] it is also proposed a novel analytic methodology for modelling/forecasting patient recruitment using the recent results in [9] on a PG approximation of the sums of PG processes in different centres that can be applied to modelling recruitment even in not so large countries/regions, which is also described in this paper for completeness of presentation.

Note that this analytic methodology is not reflected in the six papers mentioned above but is crucial as the country process under general assumptions is not a PG process and also for a small number of centres a normal approximation for the total recruitment process in general cannot be applied. Therefore, for using analytic tools for modelling/forecasting patient recruitment on different levels instead of Monte Carlo simulation, we need to develop some analytic approximations of country processes for time-dependent models similar to [9], as well.

The first author also proposed two new criteria, a non-parametric criterion and the parametric one using Poisson assumptions for testing the recruitment rates for time-dependence, Section 4.1, and the fast algorithm of simulation of a time-dependent model in R using discrete event simulation, Section 3.2, that is used in implementations.

The new parametric criterion for testing rates for time-dependence using assumptions of a PG model, Section 4.5, is analyzed by the second author who also ran many simulated scenarios using Monte Carlo technique to test the applicability of non-homogeneous PG model for prediction and interim re-projection of recruitment and compare with using a classic PG model. Some examples are provided in Section 5. These results are briefly discussed in Section 6.

2 Poisson-gamma model with homogeneous rates

Consider a multicentre clinical trial which is designed to recruit n patients by N clinical centres. A standard PG model, [1, 5], assumes that the recruitment rates in different centres are modeled as independent gamma distributed random variables with some parameters that do not depend on time. Thus, the recruitment process in any interval where the centre is active follows a time-homogeneous mixed Poisson process with gamma distributed rate (PG process).

First consider some definitions. Denote by $\Pi(a)$ a Poisson random variable with parameter a, and by $\Pi_a(t)$ an ordinary homogeneous Poisson process with rate a. This means, for any $t > 0$,

$$\mathbf{P}(\Pi_a(t) = k) = e^{-at}\frac{(at)^k}{k!},\ k = 0, 1, \ldots$$

Also denote by $\mathrm{Ga}(\alpha, \beta)$ a gamma distributed variable with parameters (α, β) (shape and rate) and probability density function

$$f(x, \alpha, \beta) = \frac{e^{-\beta x}\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)},\ x > 0, \tag{1}$$

where $\Gamma(\alpha) = \int_0^\infty e^{-x}x^{\alpha-1}\mathrm{d}x$ is a gamma function.

2.1 The case of homogeneous rates

Consider first the basic definitions used in a standard PG model as some similar definitions will be used in PG model with time-dependent rates. Let $\Pi_\lambda(t)$ be a doubly stochastic Poisson process where $\lambda = \mathrm{Ga}(\alpha, \beta)$. According to terminology in [17], this is a PG process with parameters (t, α, β). Denote it as $\mathrm{PG}(t, \alpha, \beta)$ to reflect the dependence on (α, β). Then for any $k = 0, 1, \ldots$,

$$\mathbf{P}(\mathrm{PG}(t, \alpha, \beta) = k) = \frac{\Gamma(\alpha + k)}{k!\,\Gamma(\alpha)}\,\frac{t^k\beta^\alpha}{(\beta + t)^{\alpha+k}}. \tag{2}$$

For $t = 1$, $\Pi_\lambda(1)$ has the same distribution as $\Pi(\lambda)$, which is a PG variable with parameters (α, β). For this case we omit $t = 1$ and use the notation $\mathrm{PG}(\alpha, \beta)$ instead of $\mathrm{PG}(1, \alpha, \beta)$.

Note that $\Pi_\lambda(t)$ has a negative binomial distribution and for any $\alpha > 0, \beta > 0, t > 0$, the following relation is true: for any $k = 0, 1, \ldots,$

$$\mathbf{P}(\text{PG}(t, \alpha, \beta) = k) = \mathbf{P}(\text{Nb}(\alpha, \beta/(\beta + t)) = k) \tag{3}$$

where $\text{Nb}(\alpha, p)$ denotes a random variable which has a negative binomial distribution with size α and probability p:

$$\mathbf{P}(\text{Nb}(\alpha, p) = k) = \frac{\Gamma(\alpha + k)}{k!\,\Gamma(\alpha)} p^\alpha (1 - p)^k, \; k = 0, 1, \ldots$$

This relation is useful in calculations of a PG distribution using R, as there are standard functions pnbinom(), dnbinom().

Now consider modelling recruitment. Denote by $n_i(t)$ the recruitment process in centre i (the number of patients recruited in time interval $[0, t]$), and by $n(I_j, t)$ the recruitment process in some region with the set I_j of N_j clinical centres. Also denote by u_i the date of initiation of centre i.

Assume that the recruitment process in centre i during an active recruitment stage follows a homogeneous PG process with rate λ_i which has a gamma distribution with parameters (α_i, β_i). Then the mean rate m_i and the variance s_i^2 are calculated as $m_i = \alpha_i/\beta_i$, $s_i^2 = \alpha_i/\beta_i^2$. Correspondingly, the instantaneous rate at time t is $\lambda_i(t) = \lambda_i \chi(u_i < t)$, where $\chi(A)$ is the indicator function of the event A.

Consider a more convenient representation via a cumulative rate. Denote $x(t, u) = \max(0, t - u)$ (the duration of active recruitment at time t for a centre activated at time u). So, if centre i is active at time $t > u_i$, then $x(t, u_i) = t - u_i$.

Then the cumulative rate of the process $n_i(t)$ is $\Lambda_i(t, u_i) = \lambda_i x(t, u_i)$. This means, if $\lambda_i = \text{Ga}(\alpha_i, \beta_i)$, $n_i(t)$ is a PG process with parameters $(x(t, u_i), \alpha_i, \beta_i)$, and the distribution of $n_i(t)$ can be calculated using (2) where in the right-hand side we should use $x(t, u_i)$ instead of t, and parameters (α_i, β_i).

Correspondingly, for any region with a subset of centres I_j, the recruitment process $n(I_j, t)$ in this region as a sum of the recruitment processes in the centres from this region is a nonhomogeneous Poisson process with the instantaneous rate $\lambda(I_j, t) = \sum_{i \in I_j} \lambda_i(t)$ and, in distribution, $n(I_j, t) = \Pi(\Lambda(I_j, t))$, where the region cumulative rate has the form

$$\Lambda(I_j, t) = \sum_{i \in I_j} \Lambda_i(t, u_i).$$

Note that in general a sum of gamma distributed variables with different parameters doesn't have a gamma distribution, therefore, in general, the process $n(I_j, t)$ is not a PG process. The technique for the approximation of $n(I_j, t)$ by a PG process was developed in [9].

3 Non-homogeneous PG model

Consider now an extension of the classic PG model to time-dependent rates. Assume that the recruitment rates can proportionally change over time according to some non-negative function $r(t), t \geq 0$.

A realistic type of dependence can be a piecewise linear dependence: linear increase in some initial interval, then a constant rate in some larger interval in the middle of recruitment, and then linear decrease in the last interval closer to the end of recruitment. Another opportunity for function $r(t)$ can be an exponential decay starting from 1 until some proportion, say, 50%, at the end of recruitment.

Of course the main question can arise - how to define the form of function $r(t)$. However, this question is for separate investigation, and potentially the form of function $r(t)$ can be chosen using similar historic trials, or monitoring the data for the current study and use some extrapolations of the rate dependence up to the interim point for the remaining interval.

Assume also for simplicity that the function $r(t)$ is the same for all centres, that means, time-dependence is defined only by the duration of the recruitment in the trial. However, in general, this assumption can be relaxed and we can assume different dependence in different regions, though regional type dependence would be harder to extract from historic trials. Therefore, a non-homogeneous PG model can be defined according to [14] as follows.

Assume that if a centre i is active at time t, then the recruitment rate has the form

$$\lambda_i r(t) \tag{4}$$

where λ_i is some baseline rate which has a gamma distribution with parameters (α_i, β_i), and $r(t)$ is some non-negative function. Then, if a centre i is initiated at time u_i, the instantaneous rate at time t is $\lambda_i(t) = \lambda_i r(t) \chi(u_i < t)$.

Define the cumulative proportional change in the rate for centre i in the interval $[a, b]$ as

$$R_i(a, b, u_i) = \int_a^b r(x)\chi(u_i < x)\mathrm{d}x$$

In particular, if no change in time, $r(x) \equiv 1$, and $R_i(a, b, u_i)$ is just a duration of recruitment window in the interval $[a, b]$.

The cumulative rate for centre i in the interval $[a, b]$ is $\Lambda_i(a, b, u_i) = \lambda_i R_i(a, b, u_i)$. Correspondingly, the recruitment process $n_i(t)$ in interval $[0, t]$ is again a doubly stochastic PG process with cumulative rate $\Lambda_i(0, t, u_i)$. Then,

$$\mathbf{E}[\Lambda_i(0, t, u_i)] = m_i R_i(0, t, u_i), \ \mathbf{Var}[\Lambda_i(0, t, u_i)] = s_i^2 R_i^2(0, t, u_i)$$

Consider now the recruitment process in some country s with the set of centres I_s. Then a country process

$$n(I_s, t) = \sum_{i \in I_s} n_i(t)$$

has a cumulative rate

$$\Lambda(I_s, t) = \sum_{i \in I_s} \Lambda_i(0, t, u_i)$$

Note that $n(I_s, t)$ is a mixed Poisson process but in general not a PG process.

Denote the mean and variance of $\Lambda(I_s, t)$ as

$$\begin{aligned} E(I_s, t) &= \sum_{i \in I_s} m_i R_i(0, t, u_i), \\ S^2(I_s, t) &= \sum_{i \in I_s} s_i^2 R_i^2(0, t, u_i) \end{aligned} \tag{5}$$

Assuming that $S^2(I_s, t) > 0$, let us introduce the variables

$$A(I_s, t) = E^2(I_s, t)/S^2(I_s, t), B(I_s, t) = E(I_s, t)/S^2(I_s, t) \tag{6}$$

The following statement is an extension of the result in [9] to the case where the rates are gamma distributed with different parameters.

Lemma 3.1 *The distribution of $n(I_s, t)$ can be well approximated by the distribution of a PG random variable $PG(A(I_s, t), B(I_s, t))$.*

In [9] it is shown using numerical calculations that this approximation provides a very good fit even for a small number of centres, $N = 2, 3$, and with a larger number of centres the difference between the exact and approximative distributions is decreasing to a negligible value around 10^{-4} or less.

This result can be used to create the approximate values of the mean and predictive bounds for the recruitment process $n(I_s, t)$ at any time point t.

Indeed, by definition,

$$\mathbf{E}[n(I_s, t)] = E(I_s, t) \tag{7}$$

and P-confidence predictive bounds can be calculated using a quantile function of the negative binomial distribution

```
qnbinom(P,size=M^2/S2,prob=M/(M+S2))
```

where at time t we set $M = E(I_s, t)$, $S2 = S^2(I_s, t)$.

To create predictions for the global recruitment we just need to use in relation (5) the sum by all active centres.

This technique allows for creation of analytic predictions of the recruitment process $n(I_s, t)$ over time with mean and predictive bounds assuming that the type of time dependence, function $r(t)$, is known.

3.1 Estimation at the interim stage

The technique for estimating parameters using interim data may depend on the form of time-dependence.

Using an assumption of a standard PG model [1, 5], where it is assumed that the parameters (α_i, β_i) of the rates are the same for all centres, and also assuming that the function $r(t)$ is known, for estimating parameters (α, β) we can use the same maximum likelihood technique developed in [1, 5] where in calculations of the likelihood we need to use the values $R_i(0, t, u_i)$ instead of recruitment windows v_i. The posterior distribution of the rates can be calculated similarly as in [1, 5].

If the function $r(t)$ depends on some additional parameters, then the technique of estimation will depend on the form of the function $r(t)$. In this case we can write a likelihood expression similarly as in [1, 5] and use $2 + d$ optimisation, where d is the number of parameters used in the definition of the function $r(t)$.

Note that for time-dependent models, to closely capture the dependence on time, it can be recommended to use a moving window including data only for the last several months. The choice of the length of the window is trial specific, but as a general recommendation, the number of centres and the number of patients recruited during this window should be large enough to apply a maximum likelihood technique in the framework of a PG model. In applications to late-stage clinical trials, the length of the window can be chosen as 2,3,4 months.

3.2 Simulation of non-homogeneous PG-model

In general, it is preferable to use an analytic technique described in Section 3, Lemma 3.1. However, there can be different restrictions on the recruitment, or it can be required to evaluate different metrics related to recruitment performance. For example, to evaluate the probability that the recruitment process will fall into some region, say, within 20% bounds from the predicted mean during a particular number of months, and it's hard to develop the analytic relations for these situations. Therefore, in these cases, Monte Carlo simulation can be the most appropriate way. In addition, simulation can serve also as a useful tool to verify different analytic approaches.

Here we describe the technique of simulation which is based on a discrete event simulation approach and used directly in applications.

3.2.1 Algorithm of simulation using R

Consider a trial with N centres. Assume that for a centre i the time of initiation is u_i. Denote $vecu = (u_1, \ldots, u_N)$.

Consider also a given interval of simulation, $[0, T]$, which should be large enough to ensure that for nearly all simulated trajectories the stopping point will be reached. T can be evaluated as the upper predictive bound of the recruitment time using some analytic expressions, similar to [5].

Define vectors of parameters:
$veca = (\alpha_1, \ldots, \alpha_N), vecb = (\beta_1, \ldots, \beta_N)$.

1. Create a sequence $vecr = r(1 : T)$ using function $r(t)$

2. Generate a vector of gamma distributed rates in centres:
   $vecla$ = rgamma(N, shape = $veca$, rate = $vecb$)

3. Define an output matrix for one run:
   $outmat$ = matrix(nrow = N, ncol = T)

4. Use a loop by the number of centres N:

   – For a centre i, create the vector
   $veclaik = vecla[i]*vecr$
   of the rate in centre i for each day

   – Update this vector by setting
   $veclaik[k] = 0$ for $k < vecu[i]$

   – Simulate the numbers of recruited patients in each day $k = 1, \ldots, T$ using vector of rates and Poisson distribution,
   $vecptsday$ = rpois(T, $veclaik$)

   – Use cumsum($vecptsday$) to generate a cumulative number of patients in centre i over time

   – Save $outmat[i,]$ = cumsum($vecptsday$)

Now as the output in one simulation run, we get a matrix *outmat* of the simulated trajectories of the recruitment in each centre.

Depending on the purpose, we can use matrix calculations to calculate country or global trajectories and calculate different characteristics, e.g. time to reach a total sample size, time to reach country targets, etc.

By repeating this step many times, say, 10^4 simulation runs, we can calculate different summary statistics.

A computing time is quite small. To calculate the trajectories of the mean, median and predictive bounds for the global recruitment over time using 10^4 runs, it takes seconds for $N \leq 100$. For these cases, the predictions using analytic expressions or simulation practically coincide.

4   Testing the recruitment rates for homogeneity

Consider a trial with N centres initiated at different times $\{u_i, i = 1, \ldots, N\}$ and during the recruitment period consider two disjoint intervals $[a, b]$ and $[c, d]$ (in months) where $a < b < c < d$. Define the observed number of patients in each interval $n[a, b]$ and $n[c, d]$ and denote the total number of patients $n[a, b, c, d] = n[a, b] + n[c, d]$.

For a given centre, define the duration of the active recruitment window in the interval $[y, z]$ if the centre is initiated at time u as the following function:

$$u(y, z, u) = \begin{cases} z - y & \text{if} \quad u < y \\ u - y & \text{if} \quad y \leq u < z \\ 0 & \text{if} \quad u \geq z \end{cases} \tag{8}$$

Denote by $U[a, b]$ and $U[c, d]$ the total durations of recruitment windows in each interval, that means,

$$U[a, b] = \sum_i u(a, b, u_i); \; U[c, d] = \sum_i u(c, d, u_i) \tag{9}$$

and put $U[a, b, c, d] = U[a, b] + U[c, d]$.

4.1   Poisson-type test

Let us propose a non-parametric test using Poisson assumptions. Assume here a Poisson recruitment model.

Consider testing the hypothesis:

$\mathbf{H_0}$: The mean recruitment rates in both intervals are the same.

Lemma 4.1 *Given the total number of patients $n([a, b, c, d])$ and hypothesis $\mathbf{H_0}$, the number of patients $n([a, b])$ has a binomial distribution $Bin(n([a, b, c, d]), p(a, b))$ where*

$$p(a, b) = \frac{U[a, b]}{U[a, b, c, d]}$$

Proof. Under our assumptions, the numbers of patients $n([a, b])$ and $n([c, d])$ recruited in intervals $[a, b]$ and $[c, d]$, respectively, are independent Poisson random variables with rates $mU[a, b]$ and $mU[c, d]$. Therefore, the proof follows from the known fact for Poisson variables. Indeed, consider two independent Poisson variables $\Pi(a_1)$ and $\Pi(a_2)$ with parameters a_1 and a_2. Then the conditional distribution of $\Pi(a_1)$, given $\Pi(a_1) + \Pi(a_2) = n$, is a binomial distribution $\text{Bin}(n, p)$ with $p = a_1/(a_1 + a_2)$.

4.2   Criterion for testing hypothesis H_0

Define two P-values:
Upper P-value

$$P_{Upp}(a, b, c, d) = \mathbf{P}(\text{Bin}(n([a, b, c, d]), p(a, b)) \geq n([a, b])) \tag{10}$$

Lower P-value

$$P_{Low}(a, b, c, d) = \mathbf{P}(\text{Bin}(n([a, b, c, d]), p(a, b)) \leq n([a, b])) \tag{11}$$

Consider some critical level, say, $\delta = 0.1$. Then:
1) If $P_{Upp}(a, b, c, d) \leq \delta$, with high probability the number of patients $n([a, b])$ is unusually large and it is likely that

the mean recruitment rate in interval $[a, b]$ is larger compared to interval $[c, d]$.
2) Otherwise, the mean rate in interval $[a, b]$ is smaller.

Correspondingly, we can detect intervals where the hypothesis H_0 is not true. This allows us to detect intervals with rather high or rather low recruitment rates.

Note that this test is non-parametric as it does not include the rate m which is beneficial.

The proposed approach is exact, it does not require estimation of the parameters, and can be used for any number of centres and any number of recruited patients in any intervals.

As the next stage, we are coming to the following question. How many centres do we need to detect a particular proportional difference in rates with a given confidence?

4.2.1 Dependence of P-values on the number of centres and duration of the intervals.

Let us analyze the dependence of the P-values defined in (10),(11) on the durations of active recruitment windows and the number of centres for a time-dependent model.

Suppose that the rates for each centre in each interval $[a, b]$ and $[c, d]$ are the same and are equal to m_1 in interval $[a, b]$ and m_2 in interval $[c, d]$, respectively. Let us analyze the dependence of P_{Upp} on the values of rates and $U[a, b]$ and $U[c, d]$.

This will allow to evaluate the number of centres needed for testing a particular difference in the rates m_1, m_2.

For simplicity of notation, let us use index j to define a particular interval ($j = 1$ corresponds to $[a, b]$, $j = 2$ corresponds to $[c, d]$). Correspondingly, denote by U_j the total duration of recruitment windows in interval j defined in (9).

Then the number of recruited patients n_j in interval j is a Poisson random variable, $\pi_j = \Pi(m_j U_j)$, and the expression for upper P-value can be written as

$$P_{Upp} = \mathbf{P}(X \geq 0) \tag{12}$$

where $X = \text{Bin}(\pi_1 + \pi_2, p) - \pi_1$, $p = U_1/(U_1 + U_2)$.

Then the mean of X is calculated as:

$$
\begin{aligned}
\mathbf{E}[X] = \ & \mathbf{E}[\text{Bin}(\pi_1 + \pi_2, p)] - \mathbf{E}[\pi_1] = \mathbf{E}[\pi_1 + \pi_2]p - \mathbf{E}[\pi_1] \\
= \ & m_1 U_1(p - 1) + m_2 U_2 p = \frac{U_1 U_2}{U_1 + U_2}(m_2 - m_1)
\end{aligned}
\tag{13}
$$

To calculate the variance, note that π_1 and π_2 are independent. Then, using the formula of total variance, we get

$$
\begin{aligned}
\mathbf{Var}[X] &= \mathbf{E}[\mathbf{Var}[X \mid n_1, n_2]] + \mathbf{Var}[\mathbf{E}[X \mid n_1, n_2]] \\
&= \mathbf{E}[(n_1 + n_2)p(1 - p)] + \mathbf{Var}[n_1(p - 1) + n_2 p] \\
&= (m_1 U_1 + m_2 U_2)p(1 - p) + m_1 U_1(p - 1)^2 + m_2 U_2 p^2 \\
&= \frac{U_1 U_2}{U_1 + U_2}(m_1 + m_2)
\end{aligned}
\tag{14}
$$

Note that in the real trials, the expressions for the mean and the variance of X are typically rather large as U_1 and U_2 are large. Therefore, we can consider a normal approximation for X in the form

$$X \approx \mathbf{E}[X] + \sqrt{\mathbf{Var}[X]}Z \tag{15}$$

where $Z = \mathcal{N}(0, 1)$ is a standard normal random variable. Then

$$
\begin{aligned}
P_{Upp} = \mathbf{P}(X \geq 0) &\approx \mathbf{P}\left(Z \geq \frac{\frac{U_1 U_2}{U_1 + U_2}(m_1 - m_2)}{\sqrt{\frac{U_1 U_2}{U_1 + U_2}(m_1 + m_2)}}\right) \\
&= \mathbf{P}\left(Z \leq \sqrt{\frac{U_1 U_2}{U_1 + U_2}}\frac{m_2 - m_1}{\sqrt{m_1 + m_2}}\right)
\end{aligned}
\tag{16}
$$

Note that if $m_1 > m_2$, then $P_{Upp} \to 0$ as $U_1, U_2 \to \infty$. Likewise, if $m_1 < m_2$, $P_{Upp} \to 1$ as $U_1, U_2 \to \infty$.

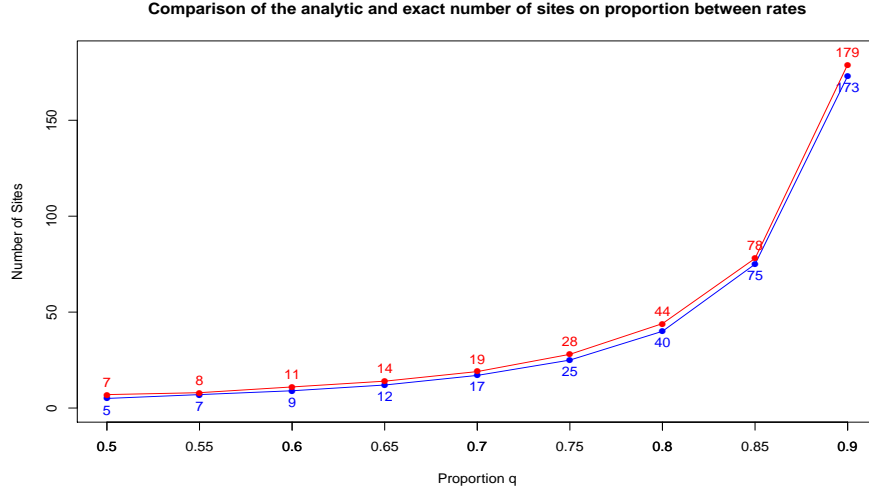We investigate the properties of this test when $m_1 > m_2$ (the rate is decreasing).

Figure 1: Dependence of the number of centres/sites N on the proportion q in the rates. Analytic results are shown by a blue line and simulation results by a red one.

Note that the relation $P_{Upp} = \delta$ can be approximated by the relation

$$\sqrt{\frac{U_1 U_2}{U_1 + U_2}} \frac{m_2 - m_1}{\sqrt{m_1 + m_2}} = z_\delta \tag{17}$$

where z_δ is a δ-quantile of a standard normal distribution.

This relation has a general form and includes the total durations of recruitment windows U_1, U_2. To analyze the dependence on the number of centres and durations of intervals, consider a specific case when all centres are active in each interval, and there are N_j active centres in interval j.

In this case $U_j = L_j N_j, j = 1, 2$. Correspondingly,

$$\mathbf{E}[X] = \frac{N_1 L_1 N_2 L_2}{N_1 L_1 + N_2 L_2}(m_2 - m_1), \mathbf{Var}[X] = \frac{N_1 L_1 N_2 L_2}{N_1 L_1 + N_2 L_2}(m_1 + m_2) \tag{18}$$

and the relation (17) has the form

$$\sqrt{\frac{N_1 L_1 N_2 L_2}{N_1 L_1 + N_2 L_2}} \frac{m_2 - m_1}{\sqrt{m_1 + m_2}} = z_\delta \tag{19}$$

In particular, if $N_1 = N_2 = N$, $L_1 = L_2 = L$ and $m_2 = qm_1$ for $q \in (0, 1)$ (the rate has decreased from interval 1 to interval 2), then relation (19) implies

$$N = \frac{2z_\delta^2}{m_1 L} \cdot \frac{1 + q}{(1 - q)^2} \tag{20}$$

Thus, $N = O(1/(1 - q)^2)$, and detecting small differences in rates requires many clinical centres, or we need to increase the duration of the interval L.

Consider a numerical example. Let's put $\delta = 0.1$, $m_1 = 0.04$ (patients per centre per day), $L = 90$ (days) and $q \in (0.5, 0.9]$. Figure 1 shows how many centres are required to detect the proportion in the rates q with confidence level 90% ($P_{Upp} \leq 0.1$). If q is closer to 1, the number of centres is growing as $1/(1 - q)^2$.

In Figure 1, analytic results based on the normal approximation are shown by a blue line. The curve for the number of centres obtained by using Monte Carlo simulation which is performed according to Section 4.3.1 is shown by a red line. Note that both curves nearly coincide. That means, for evaluating the performance of Poisson test we can use analytic results based on simpler relation (20).

4.3 The power of a Poisson test

The power of a Poisson test can be evaluated using Monte Carlo simulation. We consider here a special case where $N_1 = N_2 = N$, $L_1 = L_2 = L$ and $m_2 = qm_1$ for $q \in (0, 1)$.

P-values are defined as in (10), (11) and for our case have the form

$$\begin{aligned} P_{Upp} &= \mathbf{P}(\mathrm{Bin}(n,p) \geq n_1) \\ P_{Low} &= \mathbf{P}(\mathrm{Bin}(n,p) \leq n_1) \end{aligned} \tag{21}$$

where $n = n_1 + n_2$, n_j is the number of recruited patients in interval j, and $p = N_1 L_1/(N_1 L_1 + N_2 L_2)$.

Criterion using P-values:
For a given confidence level δ,
if $P_{Upp} \leq \delta$, we assume that $m_1 > m_2$;
if $P_{Low} \leq \delta$, we assume that $m_1 < m_2$.

Let us investigate the dependence of these quantities as random variables depending on the observed data (n_1, n_2).

Consider the approach for testing both hypotheses and analyzing the properties of this test and its power.

Consider first the upper P-value and given some hypothesis H define a statistic

$$T_H = T_H(n_1, n_2) = \mathbf{P}_H(\mathrm{Bin}(n_1 + n_2, p) \geq n_1 \mid (n_1, n_2)) \tag{22}$$

where $\mathbf{P}_H()$ is P-value but now it is considered as a random function depending on (n_1, n_2).

Consider testing two hypotheses
1. H_0: $m_1 = m_2 = m$
2. H_1: $m_2 = qm_1$ where $q \in (0, 1)$

Define the probability of a Type I error δ (usually it is denoted as α, but α is already reserved for the shape parameter of a gamma distribution).

Then, using the hypothesis H_0, we need to define a significance level $a(\delta)$ (critical region) such that for a given δ,

$$\mathbf{P}(T_{H_0}(n_1, n_2) \leq a(\delta)) = \delta \tag{23}$$

It is known that under hypothesis H_0, for a continuous distribution, P-value has a uniform distribution, and this is also confirmed by simulations for our model. Therefore, as the initial approximation we can choose $a(\delta) = \delta$. However, in our case we have a discrete distribution, thus, a more precise evaluation of the value $a(\delta)$ can be done using Monte Carlo simulation with up to 10^6 runs.

Consider now the hypothesis H_1 and define the power of the test as

$$\mathbf{P}(T_{H_1}(n_1, n_2) \leq a(\delta)) = 1 - \varepsilon \tag{24}$$

where ε is a probability of a Type II error, known as a "false negative". The value $1 - \varepsilon$ is a probability of correctly rejecting the null hypothesis H_0 given H_1.

Consider now the behavior of the power for some cases of m_i, N_i and L_i using Monte Carlo simulation. Note that $\mathbf{E}[T_{H_1}(n_1, n_2)] = P_{Upp}$ as defined in (21). Therefore, when we use the criterion

$$P_{Upp} \leq a(\delta) \tag{25}$$

to detect a difference in rates, if the relation (25) holds, this does not mean that with high probability

$$T_{H_1}(n_1, n_2) \leq a(\delta) \tag{26}$$

This is confirmed later based on a large number of Monte Carlo simulations. Thus, in general the test in the form (25) does not guarantee a high power of this test and this depends on the variance of $T_{H_1}(n_1, n_2)$.

4.3.1 Simulation results

In Figure 1, the calculation of the number of centres is based on the normal approximation using (20). However, the required number of centres can be calculated nearly exactly using Monte Carlo simulation of 10^6 runs.

For this purpose we need to evaluate by simulation for every
$q = 0.5, 0.55, \ldots, 0.85, 0.9$, the minimal N such that in (26)

$$\mathbf{E}[T_{H_1}(n_1, n_2)] \leq \delta$$

Table 1: P-values and Power for different proportions of mean rates m_1, m_2

Proportion-q	Sites-Number	Bound_Type-I-error	Pvalue_H0	Pvalue_H1	Power
0.5	7	0.093	0.540	0.088	0.729
0.55	8	0.096	0.537	0.101	0.697
0.6	11	0.098	0.532	0.092	0.722
0.65	14	0.092	0.528	0.097	0.695
0.7	19	0.094	0.524	0.099	0.695
0.75	28	0.097	0.519	0.098	0.704
0.8	44	0.099	0.516	0.099	0.705
0.85	78	0.092	0.512	0.101	0.685
0.9	179	0.096	0.508	0.1	0.694

Then the calculated nearly exact numbers of centres needed to detect the difference using criterion (25) are (7, 8, 11, 14, 19, 28, 44, 78, 179), as reflected in Figure 1. It's interesting to note that when we use these numbers, the power of test defined in (24) varies in the range 0.7-0.73, so, not so large.

Correspondingly, to get a larger power, say, 80%, the number of centres should be increased. This will be explored further in the next part.

4.3.2 Analysis of the power of the test $P_{Upp} \leq \delta$

Consider now the analysis of the power of the test $P_{Upp} \leq \delta$ using relation (24). Consider the same scenario as above,
$m_1 = 0.04$ (patients per centre per day);
$m_2 = qm_1$; $L_1 = L_2 = L$ and $L = 90$ (days); $\delta = 0.1$.

Let us evaluate the power of the test $P_{Upp} \leq \delta$ for any given $q = 0.5, 0.55, \ldots, 0.85, 0.9$, using a corresponding number of centres $N(q)$ from the vector (7, 8, ..., 78, 179) calculated as shown above using Monte Carlo simulation.

Consider Monte Carlo simulation using 10^6 independent runs of three different vectors:
vector $vec.n_1$ of the number of patients n_1 in the 1st interval with mean rate m_1,
vector $vec.n_{2,H_0}$ of the number of patients n_2 in the 2nd interval for hypothesis H_0 (simulation when $m_2 = m_1$),
and vector $vec.n_{2,H_1}$ for hypothesis H_1, when $m_2 = qm_1$.

As the result, we get three independently simulated vectors, $vec.n_1, vec.n_{2,H_0}, vec.n_{2,H_1}$.

First, we evaluate the value $a(\delta)$ in the relation (23) using data set $vec.n_1, vec.n_{2,H_0}$ related to H_0.

Then, using the part of simulated data set related to H_1, $vec.n_1, vec.n_{2,H_1}$, we evaluate the power of the test

$$\mathbf{P}(T_{H_1}(n_1, n_2) \leq a(\delta)) \tag{27}$$

Table 1 shows that for every q the evaluated value $a(\delta)$ in the column "Bound_Type-I-error" is very close to $\delta = 0.1$, as expected. The P-value for hypothesis H_0 is about 0.5 and the P-value for hypothesis H_1 is about 0.1, which is also expected.

Nevertheless, the power of this test is about 0.7 for all values q. It is also expected that the power can be less than $1 - \delta = 0.9$.

Therefore, this raises another question of interest. What should be the minimal number of centres required to reach a particular level of the power? Let us consider a numeric solution to this problem for the case when required power is 80%.

Then, for any given $q = 0.5, 0.55, \ldots, 0.85, 0.9$, according to (24), using Monte Carlo simulation with 10^6 runs, we need to find a minimal $N(q, 2)$ such that the probability

$$\mathbf{P}(T_{H_1}(n_1, n_2) \leq a(\delta)) \geq 80\% \tag{28}$$

where the values (n_1, n_2) are simulated for a given scenario under H_1, so we use vectors $vec.n_1$ and $vec.n_{2,H_1}$.

Table 2 shows the results, where the column "Sites-Number" provides the required number of centres/sites to get power 80%.

Here we can see again that the P-value for hypothesis H_0 is about 0.5 as expected, however, the P-value for hypothesis H_1 is now lower and is about 0.06. This is also expected as the number of centres is larger compared to the previous case, thus, given hypothesis H_1, the P-value should be less.

Table 2: The optimal number of centres/sites to reach power 0.8 for different proportions of mean rates m_1, m_2

Proportion-q	Sites-Number	Bound_Type-I-error	Pvalue_H0	Pvalue_H1	Power
0.5	9	0.097	0.535	0.059	0.821
0.55	11	0.099	0.531	0.063	0.810
0.6	15	0.094	0.527	0.057	0.821
0.65	19	0.095	0.525	0.062	0.806
0.7	26	0.097	0.521	0.064	0.804
0.75	38	0.100	0.517	0.064	0.808
0.8	62	0.092	0.513	0.061	0.803
0.85	111	0.094	0.510	0.063	0.803
0.9	249	0.096	0.507	0.065	0.801

Table 3: Comparing the number of centres/sites to get Upper Pvalue ≤ 0.1 and to reach Power=0.8 for different proportions of mean rates m_1, m_2

Proportion-q	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90
N-sites-Pvalue test	7	8	11	14	19	28	44	78	179
N-sites-Power-0.8	9	11	15	19	26	38	62	111	249

Table 3 shows the comparison of the number of centres for using P-value test and for using power of the test 80% using columns "Sites-Number" from tables 2 and 3. This is also shown visually in Figure 2.

Note that for a not so high proportion q, say $q \leq 0.8$, these values don't differ much, though of course the numbers of centres required to reach a higher power are visibly larger.

Nevertheless, when using test (21), in practice if the test does not show a difference in rates, which means, P-value is not a small value, we should not expect non-homogeneity in the rates.

However, if P-value is small, say ≤ 0.1, according to power calculations, with a reasonable probability which can be even around 70%, we may expect that the mean rates in these two intervals are different.

In this case it can be recommended to consider a Poisson-gamma test described in Section 4.5 which is based on calculating P-values using a Poisson-gamma recruitment model.
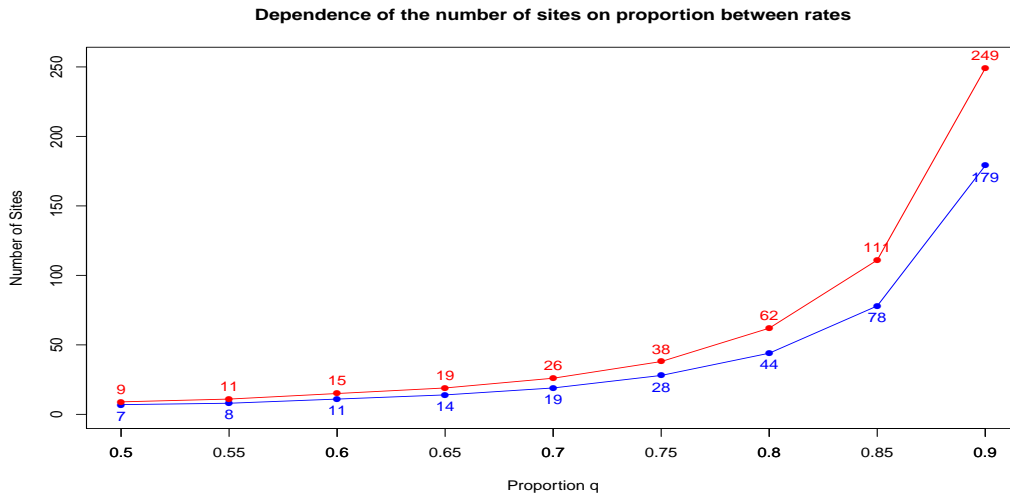

Figure 2: Comparing the number of centres/sites required to get $P_{Upp} \leq 0.1$ (blue line) with the number of centres/sites required to reach Power=0.8 (red line) for different proportion of mean rates m_1, m_2.

4.4 Poisson parametric test

Let us consider another approach which is based on using estimated rates for a Poisson model in the total interval of observations for calculating P-values for any given sub-interval.

Consider again two disjoint intervals $J_1 = [a, b]$ and $J_2 = [c, d]$ of length L_1 and L_2 with the number of active centres in these intervals N_1 and N_2. Assume for simplicity that the centres are active during each interval. Then the duration of the total recruitment window in interval J_j is $U_j = N_j L_j, j = 1, 2$.

Assume Poisson model assumptions. Suppose the rates for each centre in these intervals are m_1 and m_2 (per centre per day), respectively. That means, the number of recruited patients n_j in interval J_j is a Poisson random variable, $\pi_j = \Pi(m_j U_j)$.

Consider testing a hypothesis H_0 that $m_1 = m_2 = m$.

Denote the observed data (number of recruited patients) in each interval by n_1 and n_2 and put $n = n_1 + n_2$.

Then for a Poisson model, given H_0, the estimated rate $\widehat{m}$ in the union of the intervals $J_1 \cup J_2$ is

$$\widehat{m} = \frac{n_1 + n_2}{U_1 + U_2} = \frac{n}{U} \tag{29}$$

where $U = U_1 + U_2$. Correspondingly, define for interval J_1 the lower and upper P-values as

$$P_{Upp} = \mathbf{P}(\Pi(\widehat{m}U_1) \geq n_1)$$
$$P_{Low} = \mathbf{P}(\Pi(\widehat{m}U_1) \leq n_1) \tag{30}$$

These relations can be re-written in the form

$$P_{Upp} = \mathbf{P}(\Pi(np) \geq n_1)$$
$$P_{Low} = \mathbf{P}(\Pi(np) \leq n_1) \tag{31}$$

where

$$p = \frac{U_1}{U_1 + U_2}$$

It's interesting that for the previous non-parametric test, P-values have a similar form given in (21) where a binomial distribution is used instead of a Poisson one. Note also that for a given n,

$$\mathbf{Var}[\text{Bin}(n, p)] = np(1 - p) < \mathbf{Var}[\Pi(np)] = np$$

Therefore, it is expected that the first non-parametric test is more powerful (requires lower number of centres to detect a difference in means).

Note also that the approach in this section uses the estimated mean rate, and this is another reason to expect that the first non-parametric test is more powerful. This is supported by calculations below.

4.4.1 Analysis of P-values

Let us now investigate the dependence of P-values in (31) on m_j, N_j and L_j assuming the general case of possibly different m_1, m_2. This will allow us to evaluate the number of centres needed for testing a particular difference in the rates m_1, m_2 and compare with the first test.

Note that

$$P_{Upp} = \mathbf{P}(\Pi(np) \geq n_1) = \mathbf{P}(\Pi(np) - n_1 \geq 0)$$

Consider now the expression $X_2 = \Pi((\pi_1 + \pi_2)p) - \pi_1$. Then the calculations below follow similar steps as in Section 4.2.1. As $\mathbf{E}[\pi_j] = m_j U_j$,

$$\mathbf{E}[X_2] = \mathbf{E}[\Pi((\pi_1 + \pi_2)p)] - \mathbf{E}[\pi_1] = \mathbf{E}[\pi_1 + \pi_2]p - \mathbf{E}[\pi_1]$$
$$= \mathbf{E}[\pi_1](p - 1) + \mathbf{E}[\pi_2]p = \frac{U_1 U_2}{U_1 + U_2}(m_2 - m_1) \tag{32}$$

Note that $\mathbf{Var}[\Pi((\pi_1 + \pi_2)p) \mid \pi_1, \pi_2] = (\pi_1 + \pi_2)p$, $\mathbf{Var}[\pi_j] = m_j U_j$. Therefore, the formula of total variance gives

$$\mathbf{Var}[X_2] = \mathbf{E}[\mathbf{Var}[X_2 \mid \pi_1, \pi_2]] + \mathbf{Var}[\mathbf{E}[X_2 \mid \pi_1, \pi_2]]$$
$$= \mathbf{E}[(\pi_1 + \pi_2)p] + \mathbf{Var}[\pi_1(p - 1) + \pi_2 p]$$
$$= (m_1 U_1 + m_2 U_2)p + m_1 U_1(p - 1)^2 + m_2 U_2 p^2 \tag{33}$$
$$= \frac{1}{U^2}\Big(m_1 U_1(U_1(U_1 + U_2) + U_2^2) + m_2 U_1 U_2(2U_1 + U_2)\Big)$$

If to compare the expressions above for $\mathbf{E}[X_2]$ and $\mathbf{Var}[X_2]$ with the expressions (13) and (14) for $\mathbf{E}[X]$ and $\mathbf{Var}[X]$ in the first Poisson test, one can see that the mean is the same, however, it's easy to calculate that

$$\mathbf{Var}[X_2] - \mathbf{Var}[X] = \frac{1}{U^2}(U_1^3 m_1 + U_1^2 U_2 m_2) > 0$$

Denote

$$D = \sqrt{m_1 U_1(U_1(U_1 + U_2) + U_2^2) + m_2 U_1 U_2(2U_1 + U_2)} \tag{34}$$

Now consider an approximation for X_2 by a normal distribution similar to (15):

$$X_2 \approx \mathbf{E}[X_2] + \sqrt{\mathbf{Var}[X_2]} \cdot Z$$
$$= \frac{U_1 U_2}{U_1 + U_2}(m_2 - m_1) + \frac{D}{U_1 + U_2} \cdot Z$$

Then

$$P_{Upp} \approx \mathbf{P}(X_2 \geq 0) = \mathbf{P}\left(Z \geq \frac{\frac{U_1 U_2}{U_1 + U_2}(m_1 - m_2)}{\sqrt{\mathbf{Var}[X_2]}}\right) \tag{35}$$

and similarly for P_{Low}.

Note that for $m_1 > m_2$, as $\mathbf{Var}[X_2] > \mathbf{Var}[X]$, the value P_{Upp} in (35) above is greater than P_{Upp} in the first Poisson test, see (16). That means, using a parametric test we need to increase the values U_1 and U_2, or correspondingly, increase the number of centres N, to reach the same level δ compared to a non-parametric test.

Consider now a particular case when $L_1 = L_2 = L$ and $N_1 = N_2 = N$. Then using similar calculations we get the relation for N required to reach a particular confidence level δ:

$$N = \frac{3z_\delta^2}{m_1 L} \cdot \frac{1 + q}{(1 - q)^2} \tag{36}$$

Here it is clearly visible that N required for a parametric test is greater compared to (20) for a non-parametric one.

4.5 Poisson-gamma test

Consider an approach for testing the hypothesis H_0 that parameters (α, β) of the PG model are the same for two disjoint intervals $[a, b]$ and $[c, d]$. Assume there is possibly a different number of centres in these intervals.

Consider combining data in both intervals. Then, estimate parameters (α, β) of a PG model using this data and maximum likelihood technique described in [1, 5]. In each interval, using estimated parameters $(\hat{\alpha}, \hat{\beta})$ and active recruitment windows, we can create the predictive PG distribution.

Suppose we have some schedule of centre initiations for N centres given by $(u_1, \ldots, u_N)$. Then we have N_1 centres active in interval $[a, b]$, and N_2 centres active in interval $[c, d]$ (not necessarily $N_1 = N_2$). For each interval, say, $[a, b]$, define the duration of the active recruitment window for centre i using (8), as $v_i = u(a, b, u_i)$, and correspondingly for interval $[c, d]$.

Denote the pools of active centres as I_1 and I_2 for intervals $j = 1, 2$, $[a, b]$ and $[c, d]$, respectively. Then in interval j we have the data (k_i, v_i), the number of patients recruited and the centre's active recruitment window.

Now, for every centre combine data from intervals $j = 1, 2$ to get data $(k_i, v_i), i \in I_1 \cup I_2$, the total number of patients and total active recruitment window which is the sum of windows on both intervals (for simplicity keep the same notation k_i, v_i for the total patients and sum of windows).

We can use this data to estimate parameters of a PG model, e.g. using maximum likelihood technique and numerical optimisation (assuming we have a suitable amount of data). For data of N centres, $\{(k_i, v_i), i = 1, \ldots, N\}$, the log-likelihood $\mathcal{L}(\alpha, \beta)$ has the form, [1, 5]:

$$\sum_{i=1}^{N} [\ln \Gamma(\alpha + k_i) + \alpha \ln \beta - \ln \Gamma(\alpha) - (\alpha + k_i) \ln(\beta + v_i) + k_i \ln v_i - \ln(k_i!)]$$

Now, given estimated parameters $(\hat{\alpha}, \hat{\beta})$ using the union of the intervals, denote $m = \hat{\alpha}/\hat{\beta}$, $s^2 = \hat{\alpha}/\hat{\beta}^2$.

In interval j with pool of active centres I_j, consider the country mean and variance of the cumulative rate. So we have

$$E(I_j) = m \sum_{i \in I_j} v_i, \quad S^2(I_j) = s^2 \sum_{i \in I_j} v_i^2$$

Introduce the variables

$$A(I_j) = E^2(I_j)/S^2(I_j), \quad B(I_j) = E(I_j)/S^2(I_j)$$

In [9] it was proved that the number of recruited patients n_j in interval j can be approximated by a PG variable $PG(A(I_j), B(I_j))$ with parameters $(A(I_j), B(I_j))$, where $PG(A, B)$ is a mixed Poisson variable with the rate $Ga(A, B)$. Now we can define the upper and lower P-values in interval j as

$$\begin{aligned}
P_{Upp}(I_j) &= \mathbf{P}(PG(A(I_j), B(I_j)) \geq n_j) \\
P_{Low}(I_j) &= \mathbf{P}(PG(A(I_j), B(I_j)) \leq n_j)
\end{aligned} \tag{37}$$

where $n_j = \sum_{i \in I_j} k_i$ is the total number of patients recruited in the pool of centres I_j in corresponding interval j.

To calculate these P-values, we can use the cumulative probability function of a negative binomial distribution in R via the relation:

$$\mathbf{P}(PG(A(I_j), B(I_j)) \leq k) = \text{pnbinom}\left(k, \text{size} = \frac{E^2(I_j)}{S^2(I_j)}, \text{prob} = \frac{E(I_j)}{E(I_j) + S^2(I_j)}\right)$$

4.5.1 Some analytic considerations

To analyze the dependence on the number of centres consider a particular case when all centres are active during the whole intervals. Define by L_1 and L_2 the durations of intervals. Then $E(I_j) = mN_jL_j$, $S^2(I_j) = s^2 N_j L_j^2$, so $A(I_j) = \frac{m^2}{s^2} N_j$, $B(I_j) = \frac{m}{s^2 L_j}$, and

$$\begin{aligned}
P_{Upp}(I_j) &= \mathbf{P}\left(PG\left(\frac{m^2}{s^2} N_j, \frac{m}{s^2 L_j}\right) \geq n_j\right) \\
P_{Low}(I_j) &= \mathbf{P}\left(PG\left(\frac{m^2}{s^2} N_j, \frac{m}{s^2 L_j}\right) \leq n_j\right)
\end{aligned} \tag{38}$$

Now consider the behavior of values in (38) when hypothesis H_0 is not true. That means we may have different parameters of a PG model (m_j, s_j^2) in different intervals. Note that the parameters (m, s^2) in the relations above are estimated using data in the union of intervals assuming H_0. Now, as $\mathbf{E}[\Pi_\lambda(t)] = \mathbf{E}[\lambda]t$, using the method of moments estimates, approximately,

$$m \approx \frac{n_1 + n_2}{N_1 L_1 + N_2 L_2} = \frac{n_1}{N_1 L_1} \frac{N_1 L_1}{N_1 L_1 + N_2 L_2} + \frac{n_2}{N_2 L_2} \frac{N_2 L_2}{N_1 L_1 + N_2 L_2} \tag{39}$$

Note that n_j is a PG variable with parameters (A_j, B_j), where $A_j = \frac{m_j^2}{s_j^2} N_j$, $B_j = \frac{m_j}{s_j^2 L_j}$ and $\mathbf{E}[n_j] = m_j N_j L_j$. Then, $m_j \approx \frac{n_j}{N_j L_j}$, and from (39),

$$m \approx m_1 \frac{N_1 L_1}{N_1 L_1 + N_2 L_2} + m_2 \frac{N_2 L_2}{N_1 L_1 + N_2 L_2} \tag{40}$$

So, if $m_1 > m_2$, then the estimated value m using a union of intervals satisfies the relation $m_1 > m > m_2$.

However, under H_0, $\mathbf{E}[PG(A(I_j), B(I_j))] = mN_jL_j$. Thus, if $j = 1$ in (38) for P_{Upp}, using the means of the left and right-hand side parts under the sign of probability we have

$$\mathbf{E}[PG(A(I_1), B(I_1))] = mN_1L_1 < \mathbf{E}[n_1] = m_1 N_1 L_1.$$

Thus, as $N_1 \to \infty$, $N_2 \to \infty$, the right-hand side has a higher order of tendency to ∞, thus $P_{Upp} \to 0$, and we can use this test to detect the case when $m_1 > m_2$.
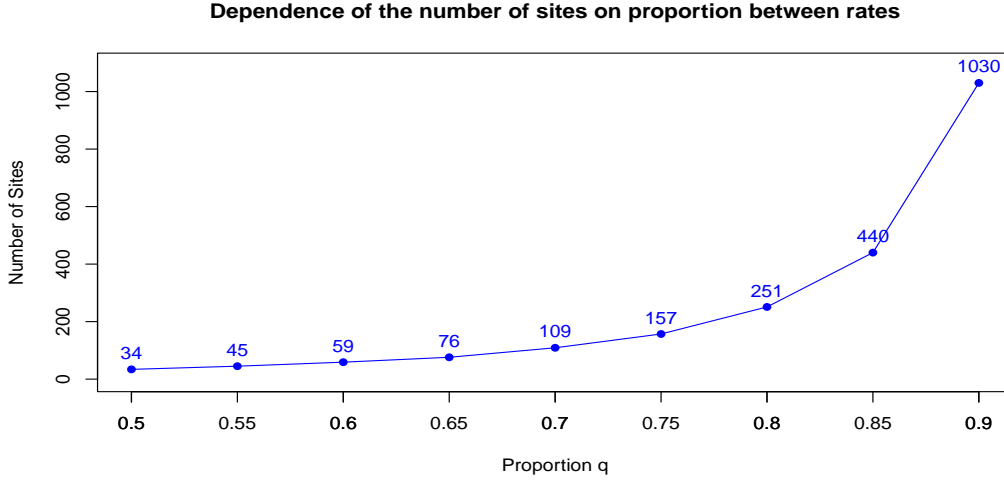
Figure 3: Dependence of the number of centres needed to detect the difference in rates by Monte Carlo simulation.

4.6 PG criterion power testing

We follow the same methodology and scenario as given in Section 4.3. Again, consider a scenario where:
$m_1 = 0.04$ (patients per centre per day);
$m_2 = qm_1$; $L_1 = L_2 = L$ and $L = 90$ (days); $\delta = 0.1$.
Also, consider the shape parameter $\alpha = 1/1.2^2$ for all centres in both intervals (recall that we are assuming $N_1 = N_2 = N$).

Now, using Monte Carlo simulation we need to find the minimal N such that $\mathbf{E}[T_{H_1}(n_1, n_2)] \leq \delta$. For a given value of q, this is calculated numerically by using a sequence of N and selecting the smallest N so that this inequality is satisfied. In calculation of P-values for the PG test, we use maximum likelihood approach for estimation of parameters over the joint intervals.

For each run of Monte Carlo simulation, interval data is generated using a Poisson-gamma process where N gamma variables $vecla = (\lambda_1, \ldots, \lambda_N)$ are simulated for N different centres. Then, in the first interval, we generate N Poisson variables using recruitment rates $vecla$. In the second interval, for hypothesis H_0, we generate data using the same rates. For hypothesis H_1, in the second interval we use rates $q \times vecla$ to account for changing rate parameters.

For values $q = 0.5, 0.55, \ldots, 0.85, 0.9$, Figure 3 demonstrates the dependence of N on q for this particular scenario. Note that these values are close to exact values and will be even closer given a larger simulation number $> 10^5$.

4.6.1 Analysis of the power of the test $P_{Upp} \leq \delta$

Consider the same scenario as above. Let us evaluate the power of the test $P_{Upp} \leq \delta$ for the vector of q values $(0.5, 0.55, \ldots, 0.85, 0.9)$, using a corresponding number of centres $N(q)$ from the vector $(34, 45, \ldots, 440, 1030)$ calculated as shown above by simulation.

Now we use Monte Carlo simulation with 10^5 runs by simulating the number of patients in the first interval using rates $vecla$. In the second interval, use rates $vecla$ and $q \times vecla$, corresponding to hypotheses H_0 and H_1 respectively. As the result we get three simulated vectors, $vec.n_1, vec.n_{2,H_0}, vec.n_{2,H_1}$, which are the vectors of the number of patients recruited in each site.

First we evaluate the value $a(\delta)$ in the relation $\mathbf{P}(T_{H_0}(n_1, n_2) \leq a(\delta)) = \delta$ using data set $vec.n_1, vec.n_{2,H_0}$ related to H_0.

Then, using the part of simulated data set related to H_1, $vec.n_1, vec.n_{2,H_1}$, we evaluate the power of the test $\mathbf{P}(T_{H_1}(n_1, n_2) \leq a(\delta))$. Table 4 shows the results.

We can also provide a solution of this problem for the case of power equal to 80%. For this case, we must find minimal N such that $\mathbf{P}(T_{H_1}(n_1, n_2) \leq a(\delta)) \geq 1 - \varepsilon$ with $\varepsilon = 0.2$. Table 5 shows the results, where the column "Sites-Number" provides the required number of centres/sites to get 80% power.

Table 4: P-values and Power for different proportions of mean rates m_1, m_2

Proportion-q	Sites-Number	Bound_Type-I-error	Pvalue_H0	Pvalue_H1	Power
0.5	34	0.066	0.517	0.098	0.657
0.55	45	0.061	0.517	0.098	0.647
0.6	59	0.063	0.514	0.099	0.655
0.65	76	0.066	0.513	0.101	0.653
0.7	109	0.066	0.512	0.099	0.663
0.75	157	0.068	0.509	0.100	0.665
0.8	251	0.070	0.507	0.100	0.669
0.85	440	0.072	0.504	0.102	0.664
0.9	1030	0.074	0.502	0.100	0.679

Table 5: The optimal number of centres/sites to reach power 0.8 for different proportions of mean rates m_1, m_2

Proportion-q	Sites-Number	Bound_Type-I-error	Pvalue_H0	Pvalue_H1	Power
0.5	85	0.069	0.512	0.055	0.810
0.55	107	0.066	0.511	0.057	0.799
0.6	138	0.068	0.507	0.059	0.800
0.65	179	0.067	0.510	0.060	0.793
0.7	246	0.068	0.509	0.061	0.794
0.75	352	0.072	0.506	0.060	0.801
0.8	570	0.074	0.504	0.060	0.805
0.85	1002	0.073	0.504	0.061	0.799
0.9	2000	0.073	0.503	0.067	0.782

Here we can see that the number of centres is much larger in order to obtain a power of 80%, and the P-value for hypothesis H_1 is now much lower and is about 0.06. Table 6 shows the comparison of the number of centres for both approaches. Note that the difference in values differs quite a bit, with more than 50 extra centres needed for power of 80%. This is also shown visually in Figure 4.

As this test involves estimating parameters, it can be oriented to trials with quite large number of centres and patients recruited. The properties of this test can be evaluated using Monte Carlo simulation. Correspondingly, to detect the proportional difference in the mean rates, it is required more centres compared to Poisson test.

Note that the Poisson test is more conservative as the random variation in the number of recruited patients is larger for a PG model. Therefore, it can be the case when Poisson test shows time-dependence, but in reality this is just explained by random fluctuations. In this case it is recommended to apply a PG test as the next step. However, if Poisson test doesn't show time-dependence, there is no need to additionally use a PG test.

5 Simulation example using non-homogeneous rates

Consider simulating some artificial data in R using time-dependent rates as outlined in Section 3.2.1. We compare a moving window approach with the classic approach using all data, where rate estimation is performed using a maximum likelihood technique. We also apply the Poisson and Poisson-gamma rate tests to the artificial data.

Table 6: Comparing the number of centres/sites to get upper P-value ≤ 0.1 and to reach power 0.8 for different proportions of mean rates m_1, m_2

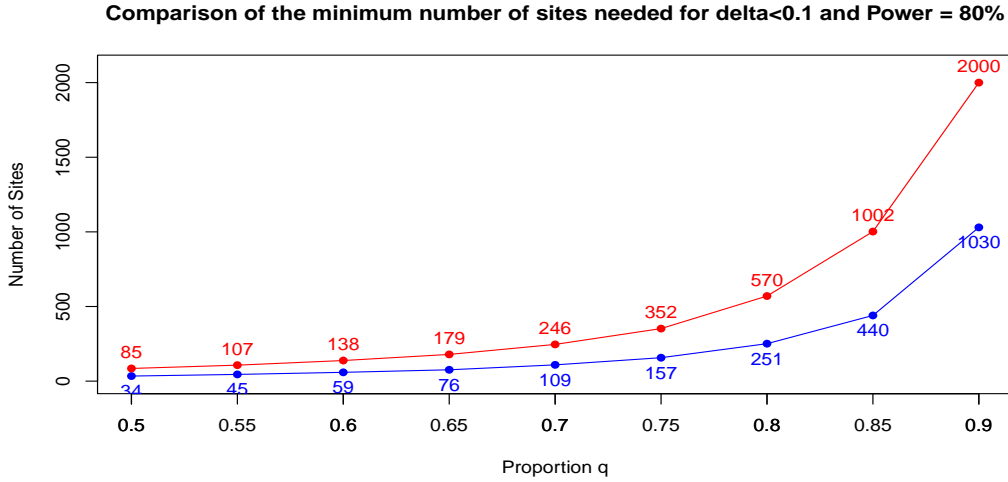| Proportion-q	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90
N-sites-Pvalue test	34	45	59	76	109	157	251	440	1030
N-sites-Power-0.8	85	107	138	179	246	352	570	1002	2000

Figure 4: Comparing the number of centres/sites required to get Upper P-value ≤ 0.1 (blue line) with the number of centres/sites required to reach Power of 0.8 (red line) for different proportion of mean rates m_1, m_2.

Consider the following parameters of simulation:

- 200 centres where each centre has mean rate 0.02 and coefficient of variation of 1.2. So, for each centre i, $\alpha_i = 1/1.2^2$ and $\beta_i = \alpha_i/0.02$
- Centres activated within 4 months using uniform grid
- In each centre i, the rate $\lambda_i r(t)$ has the following form:
  Exponential decrease over 400 days from $2.5\times$ (the mean rate) to $0.2\times$ (the mean rate) (shown in Figure 5)
- Recruitment target of 1000 patients to be recruited in 400 days

Then, a time-dependent PG process can be simulated using technique described in Section 3.2 and we can perform rate testing at some interim time, specifying some intervals and interval width to test on.

Consider simulating data for 400 days, ensuring recruitment target is reached, and then consider an interim time of 200 days.

We can look at the average global rate over time. If each centre i has simulation rate parameters (α_i, β_i) and rate function $r_i(t)$, then the mean rate at time t is $\frac{\alpha_i}{\beta_i} r_i(t)$. Then we can average over all centres at time t, as shown in Figure 5. In the following simulation, $r_i(t)$ and parameters (α_i, β_i) are the same across all centres.

The simulated trajectory of recruitment is shown in Figure 6. Note that the global target of 1000 was reached on day 384.

Given an interim time of 200 days, we can perform rate testing on two adjacent, disjoint intervals of 60 days. Both Poisson and Poisson-gamma tests were able to detect a declining rate between days 80-140 and 140-200 using the simulated data.

Now, given a detected rate change, we can use a moving window approach, where we use only the latest 60 days of data and maximum likelihood technique for estimating rate parameters for future prediction.

Consider two approaches for prediction: using all data from the start and using only the latest 60 days of data.

The results using all data were an estimated mean completion date on day 269 with lower and upper bounds of 263 and 274. This gave a probability of success of 1. The results using only the most recent 60 days of data gave an estimated mean completion date on day 287, with lower and upper bounds of 279 and 295, and also a probability of success of 1. Lower and upper bounds are calculated with 80% confidence interval.

The results are shown visually in Figure 7 using all data, and Figure 8 using the moving window approach.

It can be seen that using a moving window approach after rate-testing detects a change can give slightly more accurate prediction and is better suited to recent data. This approach is an improvement upon the standard approach when rates
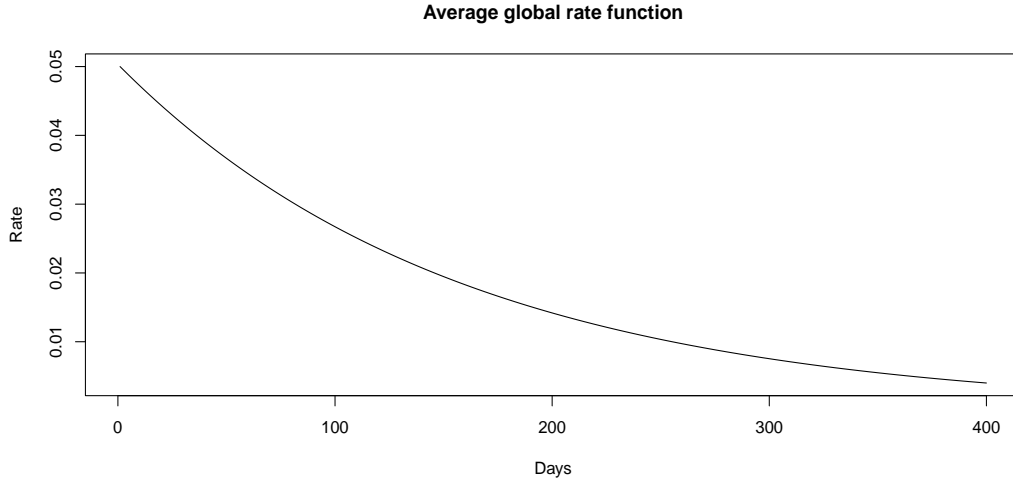
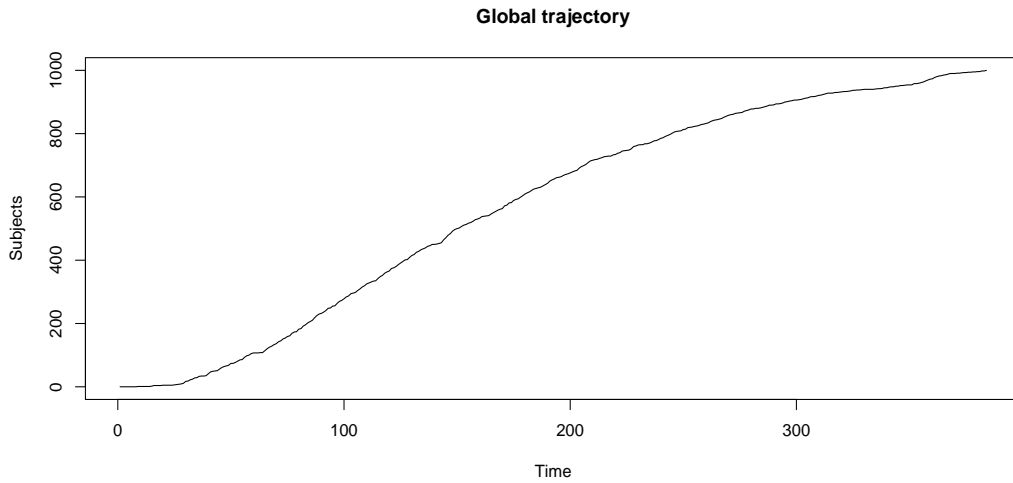Figure 5: Average global rate function with exponential decay.

Figure 6: Simulated trajectory of recruitment in all centres with time in days.

are declining steadily. However, if rates continue to decline then predictions can be poor as future time-dependence and rate decline is not considered.

Note that using a smaller window of data may improve prediction, but rate tests may not be able to successfully detect a changing rate using small intervals for testing. Also, rate estimation in a small window may give unreliable estimates when the amount of data is small.

A more advanced approach can be to use the methodology outlined in Section 3. First, assuming the rate function $r(t)$ is known, we can apply a maximum likelihood approach as in Section 3.1. For our simulated data, this gave a estimated mean completion data on day 391 with lower and upper bounds of 363 and 425. Here, the probability of success to complete recruitment in 400 days was 0.645.

The results are shown visually in Figure 9. It is clear that the approach when $r(t)$ is known yields superior predictions over the previous approaches. However, in reality the function $r(t)$ is not known, so the main task is to suitably approximate the rate function and estimate the form of the function through the use of previous data.

Similar scenarios can also be considered, where recruitment is coming from centres in countries with different rate parameters and forms of rate function $r(t)$. The conclusions will be similar to the above.
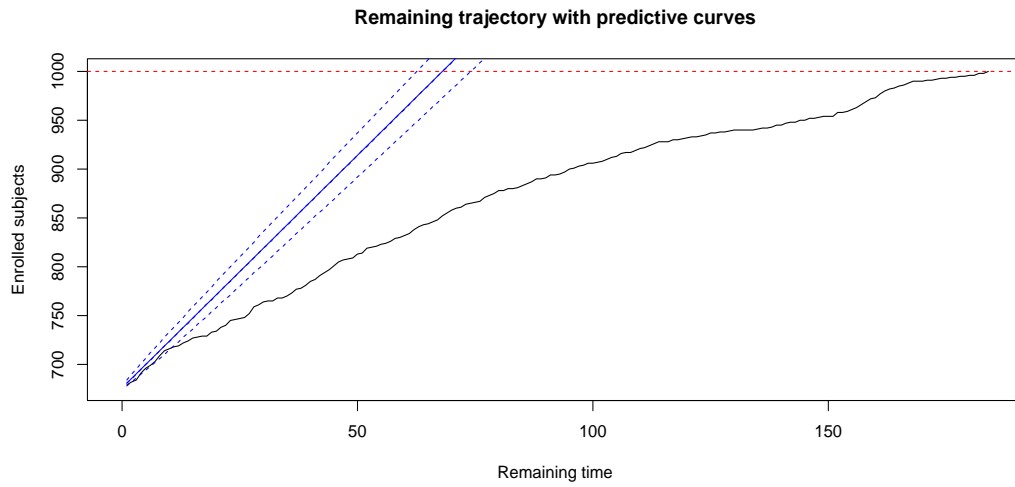
19

Figure 7: Reprojection at 200 days using all data and maximum likelihood technique. Reprojected mean and bounds in blue, global target in red.
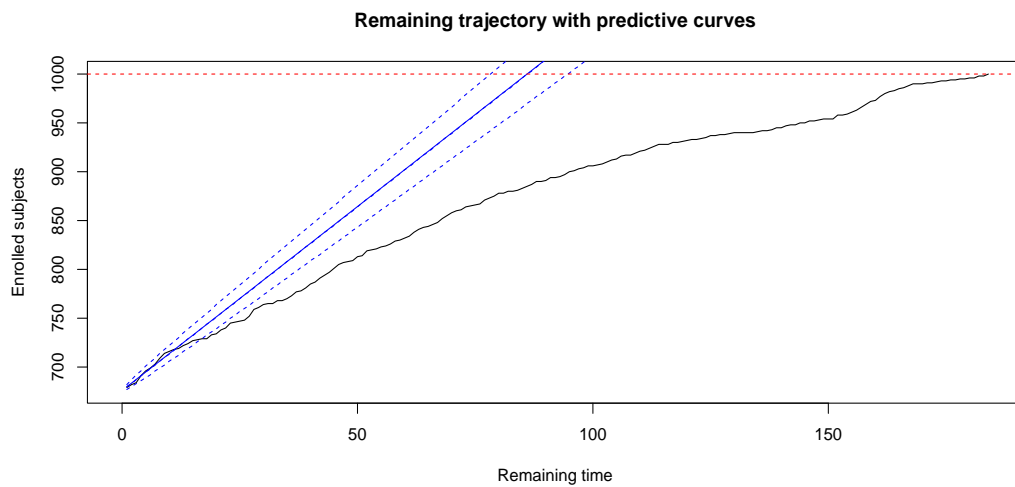

Figure 8: Reprojection at 200 days using most recent 60 days of data and standard maximum likelihood technique. Reprojected mean and bounds in blue, global target in red.
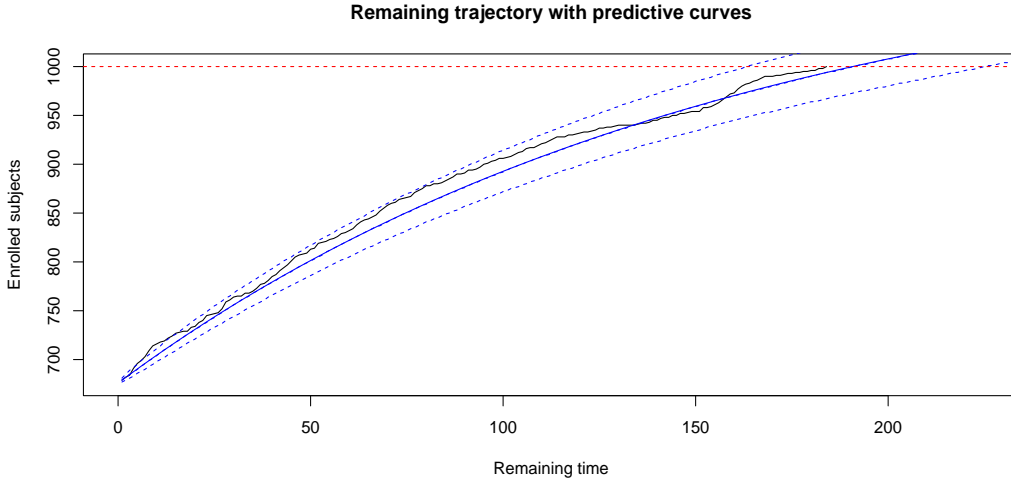
Figure 9: Reprojection at 200 days using time-dependent maximum likelihood technique, assuming $r(t)$ is known. Reprojected mean and bounds in blue, global target in red.

6 Implementations

To verify how the technique for modelling and forecasting non-homogeneous patient recruitment in clinical trials is working, there were considered different case studies using artificial data and exhaustive Monte Carlo simulation.

At the initial stage when we only have planned rates and some schedule of centre initiation, a few specific forms of function $r(t)$ including piece-wise linear dependence and exponential decay over time were considered. Recruitment was simulated using technique proposed in Section 3.2 and prediction of the recruitment was created using the analytic technique proposed in Section 3. The results shows a perfect coincidence of the analytically calculated mean and predictive bounds with the results of simulation even for not so large number of clinical centres.

At the interim stage, first it is recommended to test the recruitment rates on time-dependence using two previous time-intervals and the developed Poisson and PG criteria. If the tests do not show any time-dependence, for predicting future recruitment we can use a standard PG model and the techniques developed in [5].

However, if the tests show time-dependence, there can be different options. One way can be to use a window of the most recent data, e.g. 3 months, and create predictions using the same technique as in [5]. In this case, the recruitment rates will use only the latest 3-months of data and will be adjusted to the time-trend by repeating this procedure at every interim time in the future. Results from simulation show that when rates are changing monotonically, this approach is an improvement over the techniques in [5] .

Another more advanced opportunity can be to assume some particular time-dependence, say, linear or exponential function $r(t)$. Then, we can estimate parameters of the rates and of the function $r(t)$ using data in moving window and maximum likelihood technique, and then use these parameters for predicting future recruitment using the analytic technique proposed in Section 3.

The results of simulation of several scenarios show that if the rate is steadily changing over time (linear or exponential increase/decrease), both approaches show improved results compared to using a standard PG model, where the second approach is also outperforming the first one.

Acknowledgment

References

[1] V. Anisimov and V. Fedorov, Modelling, prediction and adaptive adjustment of recruitment in multicentre trials, *Statistics in Medicine*, 26, 27, 2007, pp. 4958–4975.

[2] V. Anisimov and V. Fedorov, Design of multicentre clinical trials with random enrolment, Chapter 25, in book *Advances in Statistical Methods for the Health Sciences*; Series: Statistics for Industry and Technology, Balakrishnan, N.; Auget, J.-L.; Mesbah, M.; Molenberghs, G., Birkhäuser Boston, 2007, pp. 387–400.

[3] V. Anisimov, D. Downing and V. Fedorov, Recruitment in multicentre trials: prediction and adjustment, in *mODa 8 - Advances in Model-Oriented Design and Analysis*, 2007, pp. 1–8.

[4] V. Anisimov, Predictive modelling of recruitment and drug supply in multicenter clinical trials. In: *Proc. of the Joint Statistical Meeting*, Biopharmaceutical Section, Washington, DC, American Statistical Association, 2009, pp. 1248–1259.

[5] V. Anisimov, Statistical modeling of clinical trials (recruitment and randomization), *Communications in Statistics - Theory and Methods*, 40, 19-20, 2011, pp. 3684–3699.

[6] V. Anisimov, Predictive event modelling in multicentre clinical trials with waiting time to response, *Pharmaceutical Statistics*, v. 10, iss. 6, 2011, pp. 517-522.

[7] V. Anisimov, Discussion on the paper "Real-time prediction of clinical trial enrollment and event counts: a review" by D.F. Heitjan et al., *Contemporary Clinical Trials*, 40, 2016, pp. 7–10.

[8] V. Anisimov, Modern analytic techniques for predictive modelling of clinical trial operations, Chapter 8, in book *Quantitative Methods in Pharmaceutical Research and Development: Concepts and Applications*, O. Marchenko, N. Katenka (Eds), Springer International Publ., 2020, pp. 361–408.

[9] V. Anisimov and M. Austin, Centralized statistical monitoring of clinical trial enrollment performance, *Communications in Statistics - Case Studies and Data Analysis*, 6, 4, 2020, pp. 392–410.

[10] V. Anisimov and M. Austin, Modeling restricted enrollment and optimal cost-efficient design in multicenter clinical trials, *arXiv:2212.12930*, 2022, 22 pages.

[11] V. Anisimov, S. Gormley, R. Baverstock and C. Kineza, Predicting event counts in event-driven clinical trials accounting for cure and ongoing recruitment, Chapter 9, in book "Data Analysis and Related Applications 2. Multivariate, Health and Demographic Data Analysis", V. 10 - Big Data, Artificial Intelligence and Data Analysis, Edited by K. N. Zafeiris, C. H. Skiadas, Y. Dimotikalis, A. Karagrigoriou, C. Karagrigoriou-Vonta, Wiley & ISTE, Aug 2022, pp. 121-142.

[12] V. Anisimov and M. Austin, Forecasting and optimizing patient enrolment in clinical trials under various restrictions, chapter 23, in book: *Stochastic Processes, Statistical Methods, and Engineering Mathematics*, Springer Proceedings in Mathematics & Statistics 408, 2023, pp. 511–540.

[13] V. Anisimov, An analytic methodology for forecasting patient enrolment performance in multicentre clinical trials, ICoMS '23: Proceedings of the 2023 6th International Conference on Mathematics and Statistics, July 14-16, 2023, Leipzig, Germany. ACM, New York, NY, USA, 2023, pp. 91-97.

[14] V. Anisimov, Modelling/forecasting patient recruitment in clinical trials using Poisson-gamma model with time-dependent rates, ICoMS '24: Proceedings of the 2024 7th International Conference on Mathematics and Statistics", June 23-25, 2024, Amarante, Portugal, ACM, New York, NY, USA, 2024, (to appear).

[15] A. Bakhshi, S. Senn and A. Phillips, Some issues in predicting patient recruitment in multi-centre clinical trials. *Statistics in Medicine*, 32(30), 2013, pp. 5458–5468.

[16] K.D. Barnard, L. Dent and A. Cook, A systematic review of models to predict recruitment to multicentre clinical trials, *BMC Medical Research Methodology*, 10, 63, 2010.

[17] J.M. Bernardo and A.F.M. Smith, *Bayesian Theory*, John Wiley & Sons: Hoboken, NJ, USA, 2004.

[18] N. Best, I. Perevozskaya, D. Lunn, G. Archer, J. Euesden, C. Fillmore, V. Sherina, D. Thompson and M. Zwierzyna, 'Predicting the COVID-19 pandemic impact on clinical trial recruitment', Statistics in Biopharmaceutical Research, 14(1), 2022, pp. 67–79.

[19] R.E. Carter, S.C. Sonne and K.T. Brady, Practical considerations for estimating clinical trial accrual periods: Application to a multi-center effectiveness study, *BMC Medical Research Methodology*, 2005, 5, pp. 11–15.

[20] B.J. Gajewski, S.D. Simon and S.E. Carlson, Predicting accrual in clinical trials with Bayesian posterior predictive distributions. *Statistics in Medicine*, 27, 2008, pp. 2328–2340.

[21] E. Gkioni, R. Riusd, S. Dodda and C. Gamblea, A systematic review describes models for recruitment prediction at the design stage of a clinical trial, *Journal of Clinical Epidemiology*, 115, 2019, pp. 141–149.

[22] D.F. Heitjan, Z. Ge and G.S. Ying, Real-time prediction of clinical trial enrollment and event counts: a review, *Contemporary Clinical Trials*, 45, part A, 2015, pp. 26–33.

[23] Y. Lan, G. Tang and D.F. Heitjan, Statistical modeling and prediction of clinical trial recruitment. *Statistics in Medicine*, 38(6), 2019, pp. 945-955.

[24] G. Mijoule, S. Savy and N. Savy, Models for patients' recruitment in clinical trials and sensitivity analysis, *Statistics in Medicine*, 31(16), 2012, pp. 1655–1674.

[25] M.N. Minois, V. Lauwers-Cances, S. Savy, M. Attal, S. Andrieua, V. Anisimov and N. Savy, Using Poisson-gamma model to evaluate the duration of recruitment process when historical trials are available. *Statistics in Medicine*, 36(23), 2017, pp. 3605–3620.

[26] A. Perperoglou, Y. Zhang, and D-K. Kipourou, Modeling time-varying recruitment rates in multicenter clinical trials, Biometrical J., v. 65, iss. 6, Aug 2023, pp. 1–12.

[27] S. Senn, *Statistical Issues in Drug Development*. Wiley: Chichester, 1997.

[28] S. Senn, Some controversies in planning and analysis multi-center trials. *Statistics in Medicine*, 17, 1998, pp. 1753–1756.

[29] A. Turchetta, N. Savy, D. A. Stephens, E. E.M. Moodie, and M. B. Klein, A time-dependent Poisson-Gamma model for recruitment forecasting in multicenter studies, Statistics in Medicine, 42, 2023, pp. 4193–4206.

[30] A. Turchetta, E. E.M. Moodie, D. A. Stephens, N. Savy, and Z. Moodie, The time-dependent Poisson-gamma model in practice: Recruitment forecasting in HIV trials, Contemporary Clinical Trials, June, 2024.

[31] S. Urbas, C. Sherlock, and P. Metcalfe, Interim recruitment prediction for multi-center clinical trials, Biostatistics, 2022, pp. 85–506.