STATISTICS IN MEDICINE
Statist. Med. 2007; **26**:4958–4975
Published online 18 July 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/sim.2956

Modelling, prediction and adaptive adjustment of recruitment in multicentre trials

Vladimir V. Anisimov[1,*,†] and Valerii V. Fedorov[2]

[1]*GlaxoSmithKline, Research Statistics Unit, NFSP-South, Third Avenue, Harlow, Essex CM19 5AW, U.K.*
[2]*GlaxoSmithKline, Research Statistics Unit, 1250 South Collegeville Road, P.O. Box 5089, Collegeville, PA 19426-0989, U.S.A.*

SUMMARY

This paper is focused on statistical modelling, prediction and adaptive adjustment of patient recruitment in multicentre clinical trials. We consider a recruitment model, where patients arrive at different centres according to Poisson processes, with recruitment rates viewed as a sample from a gamma distribution. A statistical analysis of completed studies is provided and properties of a few types of parameter estimators are investigated analytically and using simulation. The model has been validated using many real completed trials. A statistical technique for predictive recruitment modelling for ongoing trials is developed. It allows the prediction of the remaining recruitment time together with confidence intervals using current enrolment information, and also provision of an adaptive adjustment of recruitment by calculating the number of additional centres required to accomplish a study up to a certain deadline with a pre-specified probability. Results are illustrated for different recruitment scenarios. Copyright © 2007 John Wiley & Sons, Ltd.

KEY WORDS: multicentre trial; random recruitment; Poisson-gamma model; estimation; prediction; adaptive adjustment

1. INTRODUCTION

The recruitment time is one of the key decision variables in the design stage of a clinical trial. Existing techniques for recruitment planning are mainly deterministic and do not account for the various uncertainties in input information and stochastic fluctuations of the recruitment in time. Therefore, a large proportion of trials fail to complete by the enrolment deadline.

We consider a recruitment model, where the patients arrive at different centres according to Poisson processes with time-constant rates. This is a well-accepted assumption in the literature.

*Correspondence to: Vladimir V. Anisimov, GlaxoSmithKline, Research Statistics Unit, NFSP-South, Third Avenue, Harlow, Essex CM19 5AW, U.K.
†E-mail: vladimir.v.anisimov@gsk.com

Received 1 December 2005
Accepted 27 April 2006

For example, Senn [1] considers a model where patients arrive according to Poisson processes with fixed and equal recruitment rates. The properties of a balanced randomization scheme with several treatments for a model with Poisson arrivals are studied by Rukhin [2]. Anisimov *et al.* [3] focus on modelling the recruitment time for deterministic rates and on the properties of estimators for the combined response to treatment in a 'random effect' setting (see also [4]). In real trials, the recruitment rates vary across the centres and cannot be predicted with certainty. We suggest to model this variability assuming that the rates are viewed as a sample from a gamma-distributed population [4] and call this model a Poisson-gamma model.

A Poisson-gamma model and an associated negative binomial distribution have been widely used in various applications [5–9]. Doubly stochastic Poisson processes are considered by Cox [10] and Cox and Isham [11]. Snyder and Miller [12] study random point processes and, in particular, doubly stochastic Poisson processes. The recruitment model with random rates can be also viewed in the empirical Bayesian setting as the recruitment rates are actually not known in advance and can be considered as random variables with some prior distribution.

There are two major aspects in the recruitment analysis. The first is to explore the statistical properties of a Poisson-gamma recruitment model. For this case, we consider repeated trials where for different trials the rates are considered as independent samples from a gamma distribution, and investigate the probabilistic properties of specific distributions describing the recruitment. We also propose estimators of the parameters and verify their properties using Monte Carlo simulation. Analysis of completed GSK trials in different therapeutic areas shows that for a sufficiently large number of centres (>20) a Poisson-gamma model is in good agreement with real data and can serve as a basic recruitment model.

The second aspect is the prediction of the recruitment for ongoing trials. The main problem here is the construction of adequate estimators of recruitment rates. One way would be to assume that the rates are some unknown constants and then construct for each centre the maximum likelihood (ML) estimators using the number of recruited patients. However, statistical analysis of real trials shows that the models with fixed rates do not match well with real data, as they have a different shape of a special statistic which we introduced as a measure of centre occupancy (Section 3 [13]). Moreover, for trials with many centres, at early stages usually many centres are still empty and ML estimators of rates in these centres are zero, which does not reflect the real situation.

Therefore, we consider rates as random variables having a prior gamma distribution with unknown parameters. Using current recruitment information, we estimate these parameters and then calculate for each centre a posterior distribution of the rate which we use for prediction of the recruitment (empirical Bayesian setting). This approach is also compared with the straightforward ML approach, where the rates are considered to be fixed.

In addition, we consider an adaptive adjustment of the recruitment and derive tools for calculating the number of additional centres required to complete a trial up to a certain deadline with a prespecified probability.

The paper is organized as follows. Section 2 deals with the analysis of the recruitment model and investigates distributions of the number of recruited patients and centres with particular number of patients. Section 3 compares several statistical methods for estimating parameters of a Poisson-gamma model for completed trials and validates the model using real data. Section 4 develops a statistical technique for predicting the remaining recruitment time. Section 5 considers adaptive adjustment of recruitment.

DOI: 10.1002/sim

2. MODELLING OF RECRUITMENT

Consider a multicentre clinical trial designed to recruit n patients by N clinical centres. We use a model where the patients enter different centres according to independent Poisson processes [1, 3, 4, 13, 14]. Denote by λ_i the recruitment rate in centre i. We focus our attention on the case when the rates λ_i are viewed as a sample from a gamma-distributed population and consider a competitive recruitment policy [3] with no restrictions on the number of patients recruited by centres. The recruitment is stopped when the total number of recruited patients reaches a prescribed level n.

2.1. Analysis of the recruitment time

Denote by $T(n, N)$ the recruitment time (time needed to recruit n patients by N centres). Let $n_i(t)$ be the number of patients recruited by centre i up to time t and $n(t) = \sum_{i=1}^{N} n_i(t)$ be the total number of patients recruited up to time t by all N centres. Suppose that centre i is initiated at some (possibly random) time u_i, $i = 1, 2, \ldots, N$. Then for any i, $n_i(t) = 0$, $t < u_i$, and $n_i(t) = \Pi_{\lambda_i}(t - u_i)$ for $t \geqslant u_i$, where we denote by $\Pi_\lambda(\cdot)$ an ordinary Poisson process with rate λ. The overall rate is given by

$$\Lambda(t) = \sum_{i=1}^{N} \lambda_i \chi(u_i \leqslant t)$$

where $\chi(A) = 1$ if A is true, and $\chi(A) = 0$ otherwise. The process $n(t)$ is a nonhomogeneous Poisson process with instantaneous (possibly random) rate $\Lambda(t)$, and each newly initiated centre j will add an additional rate λ_j.

Assume first that all centres are initiated simultaneously at time zero. If the rates λ_i are constants, then $n_i(t)$ is a Poisson process with rate λ_i and $n(t)$ is a Poisson process with rate $\Lambda = \sum_{i=1}^{N} \lambda_i$. According to the properties of a Poisson process, $T(n, N)$, as the time of the nth event for the process $n(t)$, is gamma distributed with parameters (n, Λ).

Here, and in what follows, we denote by $\mathrm{Ga}(\alpha, \beta)$ a gamma-distributed random variable with parameters (α, β) and probability density function (p.d.f.) $p(x, \alpha, \beta) = \mathrm{e}^{-\beta x} \beta^\alpha x^{\alpha-1} \Gamma(\alpha)^{-1}$, where α and β are the shape and rate parameters, respectively, and $\Gamma(\alpha)$ is a gamma function. Note that $\mathbf{E}[\mathrm{Ga}(\alpha, \beta)] = \alpha/\beta$ and $\mathrm{Var}[\mathrm{Ga}(\alpha, \beta)] = \alpha/\beta^2$.

Consider now the case when the rates λ_i are viewed as a sample from a gamma distribution with parameters (α, β). In this case Λ, as their sum, is also gamma distributed with parameters $(\alpha N, \beta)$. Thus, the process $n(t)$ is a Poisson process with random rate Λ (doubly stochastic or Cox process [10, 11]), and for any fixed t the variable $n(t)$ has a negative binomial distribution with parameters $(\alpha N, t/\beta)$ [8]. Time $T(n, N)$ is a superposition of two independent gamma random variables and can be represented in the form $\mathrm{Ga}(n, \mathrm{Ga}(\alpha N, \beta))$. Observing that $\mathrm{Ga}(\alpha, \beta) = \mathrm{Ga}(\alpha, 1)/\beta$, we can represent $T(n, N)$ in the equivalent form: $T(n, N) = \beta\, \mathrm{Ga}(n, 1)/\mathrm{Ga}(\alpha N, 1)$. Random variable $\eta = \beta\, \mathrm{Ga}(n, 1)/\mathrm{Ga}(A, 1)$, where $\mathrm{Ga}(n, 1)$ and $\mathrm{Ga}(A, 1)$ are independent, has a Pearson type VI distribution [15, p. 381] with p.d.f.

$$p(x, n, A, \beta) = \frac{1}{\mathscr{B}(n, A)} \frac{x^{n-1} \beta^A}{(x + \beta)^{n+A}}, \quad x \geqslant 0 \tag{1}$$

DOI: 10.1002/sim

where $\mathscr{B}(i, a)$ is a beta function, and for $A>2$,

$$M(n, A, \beta) = \mathbf{E}[\eta] = \frac{\beta n}{A-1}, \quad S^2(n, A, \beta) = \mathrm{Var}[\eta] = \frac{\beta^2 n(n+A-1)}{(A-1)^2(A-2)} \tag{2}$$

Consequently, $T(n, N)$ has p.d.f. (1) and moments (2) with $A = \alpha N$.

Using the asymptotic properties of a gamma distribution, one can verify that, as $n \to \infty$ and $N \to \infty$, the ratio $T(n, N)/M(n, \alpha N, \beta)$ converges in probability to 1, and the random variable $(T(n, N) - M(n, \alpha N, \beta))/S(n, \alpha N, \beta)$ converges in distribution to $\mathscr{N}(0, 1)$, where $\mathscr{N}(0, 1)$ is a standard normal random variable.

The condition that both $N \to \infty$ and $n \to \infty$ is necessary and sufficient. The intuitive explanation is that if the values λ_i are random, asymptotically many centres are required to get averaging in the number of recruited patients over the centres as in the Law of Large Numbers. Therefore, if only one of the variables (n or N) tends to ∞, we do not have the consistency and asymptotic normality of $T(n, N)$. Note that in the case of deterministic rates we need only the condition $n \to \infty$, which simply follows from the representation $T(n, N) = \mathrm{Ga}(n, 1)/\Lambda$.

Consider now the case where centre i is initiated at some (possibly random) time u_i and the rates λ_i are viewed as a sample from a gamma-distributed population. Denote $\Sigma(t) = \sum_{i=1}^{N}(t - u_i)\lambda_i \chi(t \geqslant u_i)$, where $\Sigma(t)$ stands for the cumulative rate of the process $n(t)$ on the interval $[0, t]$. Using the properties of nonhomogeneous Poisson process and gamma distribution, we observe that

$$\mathbf{P}(T(n, N) \leqslant t) = \mathbf{P}(\Pi_{\Sigma(t)} \geqslant n) = \mathbf{P}(\mathrm{Ga}(n, \Sigma(t)) \leqslant 1) = \mathbf{P}(\mathrm{Ga}(n, 1) \leqslant \Sigma(t)) \tag{3}$$

where Π_a stands for a Poisson variable with rate a. Relation (3) is used in Section 5. The right-hand side in (3) allows us to compute the cumulative distribution function of $T(n, N)$ using Monte Carlo simulation.

2.2. *Centre occupancy models*

To apply the proposed model for the analysis of completed studies and to validate it, we consider some distributions related to the centre occupation properties.

In this section, for the sake of simplicity, we assume that all centres are initiated simultaneously. Denote by n_i the number of patients recruited by centre i. Suppose first that the recruitment rates λ_i are given. Let

$$p_i = \lambda_i/\Lambda, \quad i = 1, \ldots, N \tag{4}$$

Then, the vector $(n_1, \ldots, n_N)$ has a multinomial distribution with parameters $(p_1, \ldots, p_N)$:

$$\mathbf{P}(n_1 = k_1, \ldots, n_N = k_N) = \frac{n!}{k_1! \ldots k_N!} \prod_{i=1}^{N} p_i^{k_i}$$

where $\sum_{i=1}^{N} k_i = n$ [16, p. 31].

Consider the following characterization of the vector $(n_1, \ldots, n_N)$. For given N and n, denote by $\nu(n, N, j)$ the number of centres that recruited exactly j patients, $j = 0, 1, \ldots, n$. Random variables $\nu(n, N, j)$ are dependent and $\sum_{j=0}^{n} \nu(n, N, j) = N$.

When $p_i \equiv 1/N$, the investigation of $\nu(n, N, j)$ is related to the classical occupancy problem (random allocation of n balls into N cells, [17, Chapter II, 5]). In particular, it is possible to obtain

DOI: 10.1002/sim

a rather simple expression for the distribution of $v(n, N, 0)$ (the number of empty centres) (see [17, p. 60] and [8, Chapter 10.4, p. 414, (10.41)]).

Consider now the case when λ_i are viewed as independent gamma-distributed variables with parameters (α, β). Using (4), we can represent p_i in the form

$$p_i = \mathrm{Ga}(\alpha, \beta)(\mathrm{Ga}(\alpha, \beta) + \mathrm{Ga}(\alpha(N-1), \beta))^{-1} = \mathrm{Ga}(\alpha, 1)(\mathrm{Ga}(\alpha, 1) + \mathrm{Ga}(\alpha(N-1), 1))^{-1}$$

where the variables $\mathrm{Ga}(\alpha, 1)$ and $\mathrm{Ga}(\alpha(N-1), 1)$ are independent.

Thus, p_i does not depend on β and has a beta distribution with parameters $(\alpha, \alpha(N-1))$ [18, p. 212]. Consequently, the vector $(p_1, \ldots, p_N)$ has a Dirichlet distribution with parameters $\alpha_i = \alpha, i = 1, \ldots, N$, and the vector $(n_1, \ldots, n_N)$ has a Dirichlet (or beta) compound multinomial distribution [16, p. 80], which is a mixed Dirichlet multinomial distribution:

$$\mathbf{P}(n_1 = k_1, \ldots, n_N = k_N) = \frac{n!}{k_1! \ldots k_N!} \frac{\Gamma(\alpha N)}{\Gamma(\alpha)^N} \frac{\prod_{i=1}^N \Gamma(\alpha + k_i)}{\Gamma(\alpha N + n)} \tag{5}$$

where $\sum_{i=1}^N k_i = n$. Note that this distribution also does not depend on parameter β.

Let $m(n, N, \alpha, j) = \mathbf{E}[v(n, N, j)]$. Later it will be shown that the array $\{m(n, N, \alpha, j), j = 0, \ldots, n\}$ provides an adequate characterization of the recruitment.

To investigate the properties of $m(n, N, \alpha, j)$, let us use the representation $v(n, N, j) = \sum_{i=1}^N \chi_i(n, N, j)$, where $\chi_i(n, N, j) = 1$, if the ith centre has recruited exactly j patients, and $\chi_i(n, N, j) = 0$ otherwise. Note that $\mathbf{P}(\chi_i(n, N, j) = 1) = \mathbf{P}(n_i = j)$. Therefore,

$$\mathbf{E}[v(n, N, j)] = \sum_{i=1}^N \mathbf{P}(n_i = j) \tag{6}$$

If $\lambda_i \equiv \lambda$, then for any i, n_i has a binomial distribution with parameters $(n, 1/N)$ and (6) implies

$$\mathbf{E}[v(n, N, j)] = \binom{n}{j} \left(\frac{1}{N}\right)^{j-1} \left(1 - \frac{1}{N}\right)^{n-j}, \quad j = 0, 1, \ldots, n \tag{7}$$

where $\binom{n}{j}$ denotes binomial coefficients.

Consider now the case of gamma-distributed rates.

Lemma 2.1

If λ_i are sampled from a gamma-distributed population with parameters (α, β), then

$$m(n, N, \alpha, j) = N \binom{n}{j} \frac{\mathscr{B}(\alpha + j, \alpha(N-1) + n - j)}{\mathscr{B}(\alpha, \alpha(N-1))}, \quad j = 0, 1, 2, \ldots, n \tag{8}$$

The proof uses representation (6). In this case for any i, p_i is beta distributed and n_i has a beta-binomial distribution. Thus,

$$\mathbf{P}(n_i = j) = \mathbf{E}[\mathbf{P}(n_i = j \mid p_i)] = \binom{n}{j} \mathbf{E}[p_i^j (1 - p_i)^{n-j}]$$

Applying formula (35.154) from [16] or directly using the formula for the beta-binomial distribution [19, p. 428], we prove (8).

DOI: 10.1002/sim

Table I. Mean number of empty centres.

$N\backslash s^2$	0	0.25	0.5	0.75	1	1.25	1.5
20	0	0.01	0.15	0.47	0.9	1.41	1.94
40	0	0.25	1.07	2.25	3.55	4.88	6.18
60	0.07	1.16	3.13	5.41	7.71	9.91	11.98
80	0.53	3.07	6.44	9.9	13.19	16.24	19.03
100	1.81	6.17	11	15.61	19.84	23.67	27.13

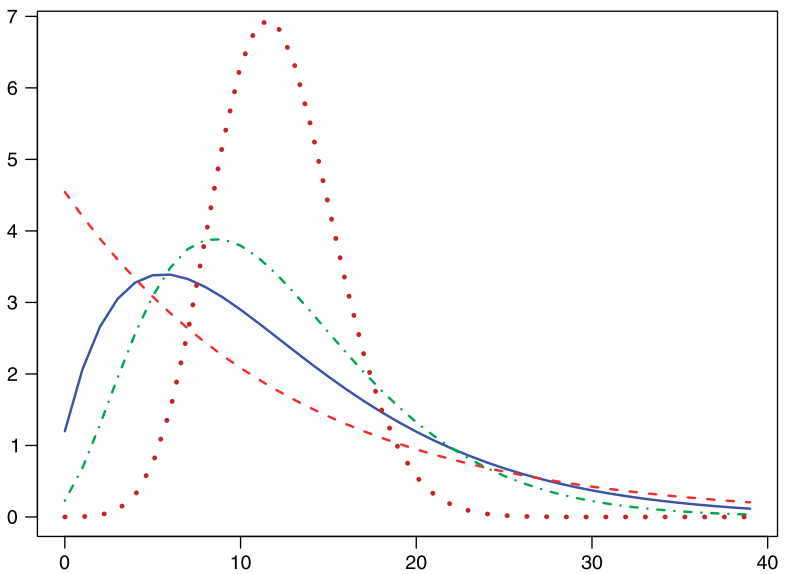

Figure 1. Mean number of centres with a given number of patients for $n = 720$, $N = 60$. A dotted line corresponds to the case when λ_i are deterministic and equal, a dash-dotted line—λ_i are gamma distributed with $\mathrm{Var}[\lambda] = 0.25$, a solid line—$\mathrm{Var}[\lambda] = 0.5$, a dashed line—$\mathrm{Var}[\lambda] = 1$. Here $\mathbf{E}[\lambda] = 1$ for all 4 cases and therefore $\alpha = 1/\mathrm{Var}[\lambda]$.

It is also possible to calculate the closed-form expression for the variance of $v(n, N, j)$ (see [13]), which will be used in Section 3 for validation of the model.

If $j = 0$, then the value $m(n, N, \alpha, 0)$ is the mean number of empty centres. It is intuitively clear that the variation in λ_i increases the variation in the number of patients over the centres and increases the mean number of empty centres.

In Table I, the mean number of empty centres is calculated for the case $n = 400$ and different values of the variance $s^2 = \mathrm{Var}[\lambda]$, where N takes values $20, \ldots, 100$. Note that $s^2 = 0$ corresponds to the case when the rates are deterministic and equal (7). These calculations give a clear evidence why in trials with large number of initiated centres the number of empty centres is often substantial.

A typical behaviour of the mean number of centres $m(n, N, \alpha, j)$ (8) as a function of j at fixed n and N and different α is shown in Figure 1.

DOI: 10.1002/sim

3. PARAMETER ESTIMATION

In this section, we compare different methods of the parameter estimation for the Poisson-gamma recruitment model: ML method, method of least squares (MLS) and method of moments (MM). For simplicity, assume throughout the section that all centres are initiated at the same time. The properties of estimators of parameter α are verified using an exhaustive simulation and applied to a number of completed GSK studies. Note that observing only $\{n_i\}$ we cannot estimate β, as formulae (5), (8) do not include it. But, at the same time, all conclusions about the type of recruitment model are based only on our knowledge of n, N and α. Parameter β can be estimated knowing in addition the duration of recruitment in different centres (see Section 4).

3.1. Maximum likelihood method

We consider a completed trial and suppose that n patients are recruited by N centres. Given $n_i, i = 1, \ldots, N$, and using (5), a log-likelihood function $\mathscr{L}(\alpha)$ can be written in the form:

$$\mathscr{L}(\alpha) = C + \ln\Gamma(\alpha N) + \sum_{j=0}^{n} \nu(j) \ln\Gamma(\alpha + j) - N \ln\Gamma(\alpha) - \ln\Gamma(\alpha N + n) \tag{9}$$

where $\nu(j) = \nu(n, N, j)$ is the number of centres that recruited exactly j patients, and we used the fact that $\sum_{i=1}^{N} \ln\Gamma(\alpha + n_i) = \sum_{j=0}^{n} \nu(j) \ln\Gamma(\alpha + j)$.

The asymptotic variance of ML estimator $\widehat{\alpha}$ is

$$\mathrm{Var}[\widehat{\alpha}] \approx V_{\mathrm{as}}^2(\alpha) = -1 \Big/ \mathbf{E}\left[\frac{\partial^2}{\partial\alpha^2}\mathscr{L}(\alpha)\right]$$

(see for instance, [20, p. 60]). Using (8) and (9), one can calculate that

$$\mathbf{E}\left[\frac{\partial^2}{\partial\alpha^2}\mathscr{L}(\alpha)\right] = N^2(\psi(\alpha N) - \psi(\alpha N + n)) - N\psi(\alpha) + \sum_{j=0}^{n} m(n, N, \alpha, j)\psi(\alpha + j)$$

where $\psi(\alpha)$ is a trigamma function, $\psi(\alpha) = (\partial^2/\partial\alpha^2) \ln\Gamma(\alpha)$ [21]. Thus, $V_{\mathrm{as}}^2(\alpha)$ can be calculated numerically. Note that, in agreement with Rao–Cramer inequality, $V_{\mathrm{as}}(\alpha)$ is a low bound of the actual variance of the estimator.

Figure 2 shows the behaviour of $V_{\mathrm{as}}(\alpha)$ as a function of N for $\alpha = 2$ and different values of n. As we can see, after some point, an increase in the number of centres N does not lead to any substantial decrease of $V_{\mathrm{as}}(\alpha)$.

3.2. Method of least squares

The MLS estimator can be considered as the point of minimum of the function

$$S^2(\alpha) = \sum_{j=0}^{n} (\nu(j) - m(n, N, \alpha, j))^2 \tag{10}$$

where the values $\nu(j)$ are defined as in (9). Note that there is no sense in constructing the sum of residuals using the values n_i as $\mathbf{E}[n_i] = n/N$ and does not depend on α.

DOI: 10.1002/sim

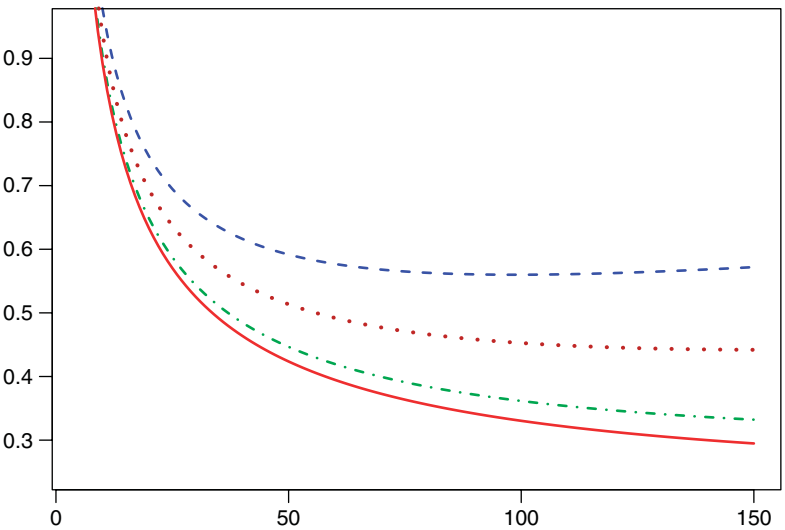

Figure 2. Standard deviation of the ML estimator as the function of N for $\alpha = 2$ and different n. Solid line—$n = 1200$, dashed-dotted line—$n = 800$, dotted line—$n = 400$, dashed line—$n = 250$.

3.3. Method of moments

To construct the MM estimator, we consider the values $\Sigma_n = \sum_{i=1}^{N} \mathbf{E}_\alpha[n_i^2]$, where $\mathbf{E}_\alpha[\ldots]$ is the expectation given that the parameter α is known, and $\overline{\Sigma}_n = \sum_{i=1}^{N} n_i^2$. Taking into account that n_i is a beta-binomial random variable and using formulae for the first two moments of a beta distribution [18, p. 217], we can conclude that $\Sigma_n = n + (\alpha + 1)n(n-1)/(\alpha N + 1)$. Equating Σ_n and $\overline{\Sigma}_n$, one may find that

$$\widehat{\alpha}_{\mathrm{MM}} = (n^2 - \Sigma_n)(N(\Sigma_n - n) - n(n-1))^{-1} \tag{11}$$

Note that $n^2 > \Sigma_n$ as $N > 1$; however, the solution has sense only when the denominator is positive. Note that $\widehat{\alpha}_{\mathrm{MM}}$ is a biased estimator. However, if $n \to \infty$ and $N \to \infty$ in such a way that $\lim N/n = c < \infty$, then the bias vanishes and the estimator is consistent.

3.4. Simulation results

We were not able to analyse analytically the asymptotic properties of ML and MLS estimators and use Monte Carlo simulations instead. Different sample realizations of the vector $(n_1, \ldots, n_N)$ were simulated for particular n, N and α, and then the parameter α was estimated using ML method, MLS and MM.

Figure 3 shows the results of simulation for the scenario: $\alpha = 2$, $n = 1200$, $N = 100$, and illustrates the fact that the mean values $\{m(n, N, \alpha, j),\ j = 0, 1, \ldots\}$ can serve as an adequate measure of centre occupancy. For different simulation runs, the graphs have a similar form.

Figure 4 shows the graphs of the empirical density function of the estimator $\widehat{\alpha}$ for the same scenario and 1.5×10^4 simulation runs using all three estimation methods according to the algorithms suggested in Sections 3.1, 3.2 and (11). As we see, all estimators are similarly distributed. They are practically unbiased and the empirical density function is close to a normal density. The

DOI: 10.1002/sim

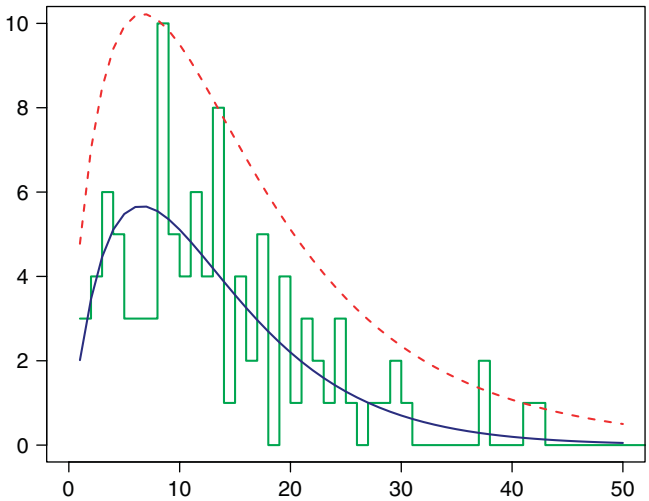

Figure 3. Mean number of centres and simulated data for $\alpha = 2$, $n = 1200$, $N = 100$. A step-wise line is a sample realization of the vector $\mathbf{v} = (v(0), v(1), \ldots)$ for one particular run, a continuous line shows the graph of the mean $(m(n, N, \alpha, j), j = 0, 1, \ldots)$ (see (8)). The behaviour of $m(n, N, \alpha, j)$ plus two standard deviations $\mathrm{sd}(n, N, \alpha, j)$ is indicated by a dashed line, where $\mathrm{sd}(n, N, \alpha, j) = \sqrt{\mathrm{Var}[v(n, N, j)]}$, and is calculated in a closed form.

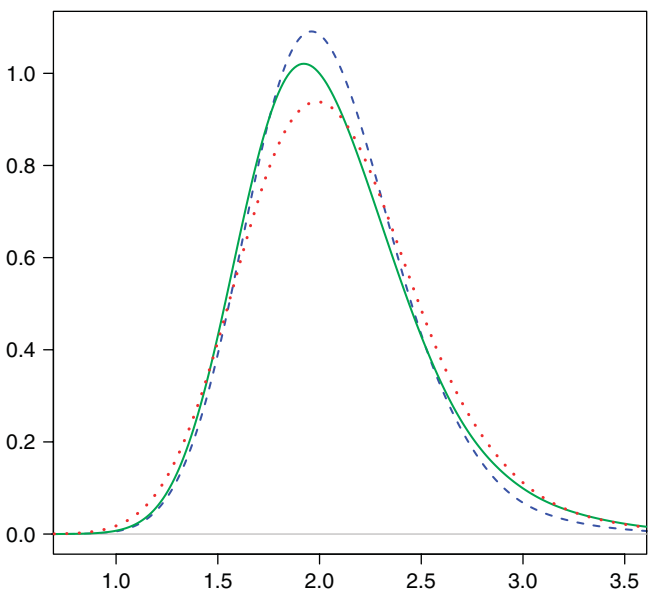

Figure 4. Comparison of the empirical density function of ML, MLS and MM estimators for the scenario $n = 1200$, $N = 100$, $\alpha = 2$ using 15 000 simulation runs. A solid line corresponds to MLS, dashed line—ML, dotted line—MM.

DOI: 10.1002/sim

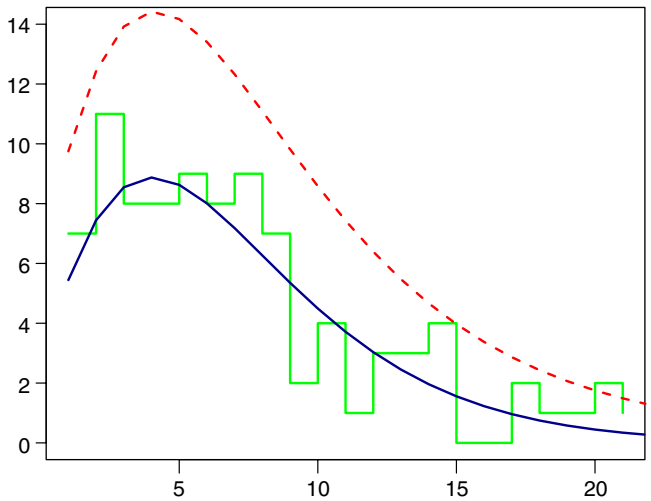

Figure 5. Empirical data and mean number of centres. Study A.

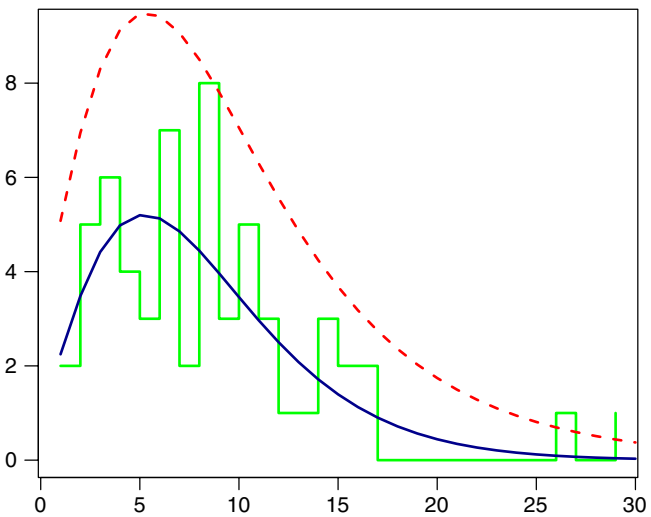

Figure 6. Empirical data and mean number of centres. Study B.

results of exhaustive simulation show that the proposed technique works well for trials with the number of centres $N \geqslant 20$.

3.5. Validation of the model

To check how a Poisson-gamma recruitment model fits the real recruitment data, several tens of completed GSK studies were analysed with $N > 20$. Practically for all studies ML estimator $\widehat{\alpha}$ was in the interval $(1.2, 4.0)$. The results of the analysis of the following two studies A and B, where

DOI: 10.1002/sim

$\bar{v} = (v(1), v(2), \ldots)$, are typical examples, and for each study all three types of estimators of α are calculated. In this case, the number of empty centres has not been reported and we use 'truncated' estimators (the item corresponding to $j = 0$ in corresponding relations (9)–(11) is omitted). At large N, it adds the additional error in estimation of the order $O(1/N)$ and for these studies the effect is practically negligible.

Study A: $n = 629$, $N = 91$, $\bar{v} = (7, 11, 8, 8, 9, 8, 9, 7, 2, 4, 1, 3, 3, 4, 0, 0, 2, 1, 1, 2, 1, 0, 0, 0, \ldots)$, (MLE, MLE, MME) = (2.846, 2.112, 2.684).

Study B: $n = 475$, $N = 59$, $\bar{v} = (2, 5, 6, 4, 3, 7, 2, 8, 3, 5, 3, 1, 1, 3, 2, 2, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 0, \ldots)$, (MLE, MLE, MME) = (3.372, 3.08, 2.933).

Figures 5 (study A) and 6 (study B) show the graph of data (vector $\bar{v}$)—step-wise line, function $m(n, N, \widehat{\alpha}_{\mathrm{ML}}, j)$—solid line, and the graph of $m(n, N, \widehat{\alpha}_{\mathrm{ML}}, j) + 2\mathrm{sd}$—dashed line, where sd is also a function of $n, N, \widehat{\alpha}_{\mathrm{ML}}$, and j. As we see, practically all data are within two sd near the theoretical mean, where $\widehat{\alpha}_{\mathrm{ML}}$ is a ML estimator of the parameter α.

For all analysed studies, a Poisson-gamma model fits real data well when many centres (~ 20 or more) are initiated [13]. When only a few centres are initiated, the rates should be estimated individually in each centre.

4. PREDICTION OF RECRUITMENT

Consider a multicentre study where n patients are planned to be recruited by N centres. Assume that we have reached some intermediate time point t_1 and denote by k_i the number of patients recruited by centre i up to this time. Let τ_i be the actual duration of recruitment by centre i in this period, $i = 1, \ldots, N$ (τ_i is the difference between t_1 and the actual date of ith centre initiation, as centres usually are initiated with delays). Our aim is to construct the prediction of the remaining recruitment time using the interim data $\{k_i, \tau_i, i = 1, \ldots, N\}$.

4.1. Estimation of parameters

We consider a Poisson-gamma recruitment model and assume that the rates λ_i have a prior gamma distribution with unknown parameters (α, β). In this case, k_i as a random variable has a negative binomial distribution with parameters $(\alpha, \tau_i/\beta)$ [8, p. 204]. Using the parameterization in terms of mean rate $m = \alpha/\beta$, we get:

$$\mathbf{P}(k_i = k) = P(k, \tau_i, \alpha, m) = \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)} \left(\frac{m\tau_i}{\alpha + m\tau_i}\right)^k \left(\frac{\alpha}{\alpha + m\tau_i}\right)^\alpha, \quad k = 0, 1, \ldots$$

The numbers of recruited patients in different centres are independent. Therefore, given the interim data $\{k_i, \tau_i, i = 1, \ldots, N\}$, the log-likelihood function is

$$\mathscr{L}(\alpha, m) = \sum_{i=1}^{N} \mathscr{L}(\alpha, m, k_i, \tau_i)$$

where $\mathscr{L}(k_i, \tau_i, \alpha, m) = \ln P(k_i, \tau_i, \alpha, m)$. A Fisher information matrix in one centre is $I_i(k_i, \tau_i, \alpha, m) = -\mathbf{E}[\nabla^2 \mathscr{L}(k_i, \tau_i, \alpha, m)]$, where ∇^2 is a matrix of second derivatives. In total, a Fisher information matrix is

$$I(\alpha, m) = \sum_{i=1}^{N} I_i(k_i, \tau_i, \alpha, m)$$

DOI: 10.1002/sim

Assume that t_1 is fixed, $N \to \infty$ and that one of the two conditions is true: either τ_i are fixed and $\min_i \tau_i \geqslant a > 0$, or τ_i are independent random variables and, for all i, $\mathbf{P}(\tau_i > 0) \geqslant \delta > 0$, and in addition that there exists the limit in probability $I_0(\alpha_{\text{true}}, m_{\text{true}}) = \lim_{N\to\infty} I(\alpha_{\text{true}}, m_{\text{true}})/N$, where $I_0(\alpha_{\text{true}}, m_{\text{true}})$ is the asymptotic information matrix. Then, from the ML estimation theory, it follows that at large N, under standard conditions of regularity, the distribution of the ML estimator $(\widehat{\alpha}, \widehat{m})$ is approximated by a bivariate normal distribution with mean $(\alpha_{\text{true}}, m_{\text{true}})$ and variance–covariance matrix $I_0^{-1}(\alpha_{\text{true}}, m_{\text{true}})/N$.

4.2. Prediction of the remaining time

Consider a prediction of the remaining recruitment time. Let $K_1 = \sum_{i=1}^{N} k_i$ be the total number of patients recruited up to time t_1, $K_2 = n - K_1$ be the remaining number of patients left to recruit and $T(K_2, N)$ be the remaining recruitment time. Put $\Lambda = \sum_{i=1}^{N} \lambda_i$, where λ_i is a recruitment rate in centre i. Assume that after time point t_1 all N centres involved in the trial continue to recruit without interruption. Thus, given rates $\{\lambda_i\}$, the patients in centre i are recruited according to a Poisson process with rate λ_i and the overall recruitment rate is Λ.

If Λ is known, then, according to Section 2.1, $T(K_2, N)$ can be represented as

$$T(K_2, N) = \text{Ga}(K_2, 1)/\Lambda \tag{12}$$

However, as Λ is unknown, we can construct the predictors of $T(K_2, N)$ using the predictors of Λ. We focus on a Bayesian approach that provides a parsimonious way of using a prior information. Assume that $\{\lambda_i\}$ are random and have a prior gamma distribution which is a conjugate prior to a Poisson distribution. Suppose first that parameters (α, β) are known. Then $\Lambda = \text{Ga}(\alpha N, \beta)$ (has a prior gamma distribution). Given the observed values $(k_1, \ldots, k_n)$, let us calculate the posterior distribution of Λ.

Let $\Pi_\lambda(t)$ stand for a Poisson process with rate λ, where λ has a prior gamma distribution with parameters (α, β). Assume that we observe k events on the interval $[0, t]$, which means, $\Pi_\lambda(t) = k$. Then, the posterior distribution of λ is also gamma, with parameters $(\alpha + k, \beta + t)$ (compare with [19, p. 429]).

Thus, given the number of events k_i and durations τ_i, the predicted total rate Λ has the same posterior distribution as the variable

$$\widetilde{\Lambda}_1 = \sum_{i=1}^{N} \text{Ga}(\alpha + k_i, \beta + \tau_i) \tag{13}$$

In a particular case when $\tau_i \equiv \tau$ (the same duration of recruitment in all centres), (13) can be replaced by $\widetilde{\Lambda}_1 = \text{Ga}(\alpha N + K_1, \beta + \tau)$.

Consider also for comparison another approach using ML estimation. In centre i, an ML estimator of the rate is k_i/τ_i (given that $\tau_i > 0$), and the ML estimator of the overall rate Λ is

$$\widetilde{\Lambda}_2 = \sum_{i=1}^{N} k_i/\tau_i \tag{14}$$

In particular, if $\tau_i \equiv \tau$, (14) has the form $\widetilde{\Lambda}_2 = K_1/\tau$.

Consider the prediction of the remaining recruitment time. Using the Bayesian approach, we may conclude that the predicted remaining time $\widetilde{T}_1$ has the same posterior distribution as the

DOI: 10.1002/sim

random variable

$$\widetilde{T}_1 = \mathrm{Ga}(K_2, 1)/\widetilde{\Lambda}_1 \tag{15}$$

Using the ML method, we can represent the predicted remaining time as

$$\widetilde{T}_2 = \mathrm{Ga}(K_2, 1)/\widetilde{\Lambda}_2 \tag{16}$$

Compare the quality of both predictors in the case when $\tau_i \equiv \tau$. If the rates $\{\lambda_i\}$ were known, (12) implies that given the data the mean remaining time is $\mathbf{E}[T(K_2, N) \mid \{\lambda_i, k_i\}] = K_2/\Lambda$. As $K_2 = n - K_1 = n - \Pi_\Lambda(\tau)$, let us consider the following quantities:

$$\Delta_{T_1} = \frac{\mathrm{Ga}([n - \Pi_\Lambda(\tau)]_+, 1)}{\mathrm{Ga}(\alpha N + \Pi_\Lambda(\tau), 1)}(\beta + \tau) - \frac{n - \Pi_\Lambda(\tau)}{\Lambda} \tag{17}$$

$$\Delta_{T_2} = \frac{\mathrm{Ga}([n - \Pi_\Lambda(\tau)]_+, 1)}{\Pi_\Lambda(\tau)}\tau - \frac{n - \Pi_\Lambda(\tau)}{\Lambda} \tag{18}$$

where $[a]_+ = \max(a, 1)$. These quantities show the variation of the predicted time near its mean calculated at given rates and can serve as the measures of the quality of prediction.

We investigate the population variability of these quantities by assuming that $\Lambda = \mathrm{Ga}(\alpha N, \beta)$ and is independent of other variables. In this case, it is not possible to find closed-form expressions for the mean and the variance of Δ_{T_1} and Δ_{T_2}. Thus, we provide the analysis using the approximations at large N and simulation, as well.

Using a normal approximation for $\mathrm{Ga}(V, 1)$ and for $\Pi_V(t)$ at large V, it is possible to construct the asymptotic expansions for the values Δ_{T_i} and prove that as $N \to \infty$ in distribution $\Delta_{T_1} \to 0$, $\Delta_{T_2} \to 0$, and also that $\mathbf{E}[\Delta_{T_1}] = O(1/N)$, $\mathbf{E}[\Delta_{T_2}] = O(1/N)$, $\mathrm{Var}[\Delta_{T_1}] = O(1/N)$, and $\mathrm{Var}[\Delta_{T_2}] = O(1/N)$. Using Taylor expansion, it is also possible to find the approximate expressions for $\mathrm{Var}[\Delta_{T_1}]$ and $\mathrm{Var}[\Delta_{T_2}]$. However, these expressions are rather complicated. Thus, numerical calculations are provided and the results are compared with direct Monte Carlo simulation of the expressions Δ_{T_1} and Δ_{T_2} in (17), (18).

Figure 7 shows a numerically calculated difference of the approximated expressions for the variances of Δ_{T_2} and Δ_{T_1} for the scenario: $n = 600$, $N = 60$, $m = 2/3$, and for three different values of α. For this scenario, the mean predicted time $T = n/(mN) = 15$. For example, if $\alpha = 3$, then at $\tau < 4$ the variance of the Bayesian predictor (15) is less compared to with the ML predictor (16). These results are supported by Monte Carlo simulation of expressions Δ_{T_1} and Δ_{T_2}. Consider the same scenario and the cases $\tau = 0.6$, 1, 4 and 10, where 10^6 simulation runs were provided. Denote by $\mathrm{Apsd}(\Delta_{T_j})$ the approximated value of sd for Δ_{T_j}, $j = 1, 2$ (these values are calculated numerically up to the error of the order $O(1/N)$). The results of calculations of the empirical mean, sd, and $\mathrm{Apsd}(\Delta_{T_j})$ are given in Table II.

As we see, at smaller τ, Bayesian approach gives better prediction. When τ is larger, the quality of both predictors is practically the same.

If the durations of recruitment $\{\tau_i\}$ are different, then the results may depend on the values $\{\tau_i\}$. However, for any particular scenario, the mean and sd of the expressions similar to Δ_{T_j} can be evaluated using relations (13) or (14) and Monte Carlo simulation.

DOI: 10.1002/sim

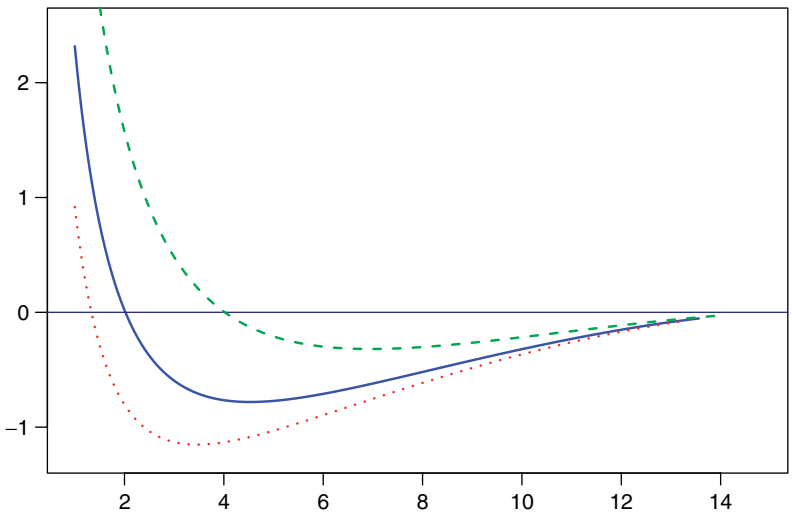

Figure 7. Plot of the approximated expression $\mathrm{Var}[\Delta_{T_2}] - \mathrm{Var}[\Delta_{T_1}]$ for the scenario: $n = 600$, $N = 60$, $m = 2/3$. A dashed line corresponds to $\alpha = 3$, a continuous line—$\alpha = 1.5$, and a dotted line—$\alpha = 1$.

Table II. Comparison of empirical mean and sd for both approaches.

τ	$\mathrm{Mean}(\Delta_{T_1})$	$\mathrm{Mean}(\Delta_{T_2})$	$\mathrm{sd}(\Delta_{T_1})$	$\mathrm{sd}(\Delta_{T_2})$	$\mathrm{Apsd}(\Delta_{T_1})$	$\mathrm{Apsd}(\Delta_{T_2})$
0.6	0.002	0.707	2.01	3.55	1.83	2.49
1	−0.002	0.41	1.83	2.52	1.68	1.98
4	−0.005	0.094	1.11	1.02	1.04	0.94
10	0.0001	0.038	0.51	0.47	0.61	0.54

4.3. Accounting for errors in estimation of parameters

The conclusions above are related to the case when the parameters α and β are known. As (α, β) are not known, we can use $(\widehat{\alpha}, \widehat{\beta})$ (or $(\widehat{\alpha}, \widehat{m})$) estimated on the interval $[0, t_1]$ instead. Consider again the case $\tau_i \equiv \tau$. Using Bayesian re-estimation, the predicted time is represented as

$$\widetilde{T}_1 = \frac{\mathrm{Ga}(K_2, 1)}{\mathrm{Ga}(\hat{\alpha}N + K_1, 1)}(\widehat{\alpha}/\widehat{m} + \tau)$$

At given $(\widehat{\alpha}, \widehat{\beta})$, $\widetilde{T}_1$ has a Pearson type VI distribution (1) with p.d.f. $p(x) = p(x, K_2, \hat{\alpha}N + K_1, \widehat{\alpha}/\widehat{m} + \tau)$. The impact of additional errors in estimating parameters on the precision of prediction cannot be precisely evaluated, as the estimators cannot be written in a closed form.

One way is to use the approximation by a bivariate normal distribution as in Section 4.1. Denote by (η_1, η_2) the bivariate normal vector with covariance matrix $B_0(\alpha, m)$. Then at large N, $(\widehat{\alpha}, \widehat{m}) \approx (\alpha, m) + (\eta_1, \eta_2)/\sqrt{N}$. Using (2), it is possible to write the asymptotic expansions for $\mathbf{E}[\widetilde{T}_1]$ and $\mathrm{Var}[\widetilde{T}_1]$, and show that the errors in estimating parameters add the additional errors in the mean and the variance of the predicted time of the order $O(1/N)$. However, for any given n

DOI: 10.1002/sim

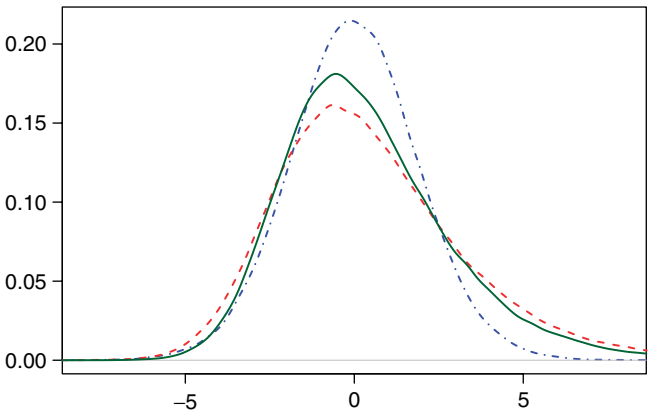

Figure 8. Empirical density function of the difference in predicted time and average remaining time. Mean duration of study $T = 15$, initial interval for prediction $\tau = 1$.

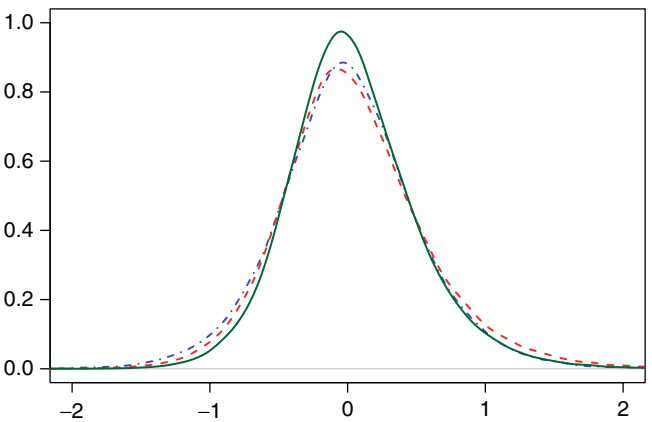

Figure 9. Empirical density function of the difference in predicted time and average remaining time. Mean duration of study $T = 15$, initial interval for prediction $\tau = 10$.

and N, there are additional errors related to asymptotic approximation. Thus, both predictors are also compared using simulations.

A typical situation is illustrated in Figures 8 and 9 for the scenario $n = 600$, $N = 60$, $\alpha = 1.5$, and $m = \frac{2}{3}$ and two cases: $\tau = 1$ (Figure 8) and $\tau = 10$ (Figure 9) (2×10^5 simulation runs). A solid line corresponds to the empirical density function of Δ_{T_2}, a dashed line to the empirical density of Δ_{T_1} with parameters $(\widehat{\alpha}, \widehat{m})$ estimated on $[0, \tau]$, and a dashed-dotted line to the empirical density of Δ_{T_1} with known parameters (Δ_{T_j} defined in (17) and (18)). Note that to construct the predicted time we use only data simulated on $[0, \tau]$ (numbers of recruited patients). A dashed-dotted line is drawn only for comparison with the case when the parameters are known.

Simulation results show that, if the parameters were known, at small τ a Bayesian approach gives better results, which is in agreement with results of Section 4.2. However, when parameters

DOI: 10.1002/sim

(α, β) are also estimated, both methods give close results, as at smaller τ the variation in the estimation of parameters as well as in the estimation of the overall rate using the ML method is larger. When τ is larger (Figure 9), there is practically no difference between both methods and cases when parameters are either known or estimated. It is intuitively clear as at larger τ the quality of estimating parameters and overall rate are better.

However, using Bayesian re-estimation is preferable from several points of view. First, it preserves the properties of a Poisson-gamma model after re-estimating which allows the use of closed-form expressions for predicted characteristics and avoids Monte Carlo simulation. When there are centres with different durations of recruitment (long term and relatively short term), Bayesian approach provides better estimators of the rates in short-term centres. In addition, when there are centres with small numbers of patients (low-performing centres), in particular empty centres, the parameters (α, β) can be estimated using long-term centres and then can be used to re-estimate the recruitment rates in low-performing centres.

The results of simulation of different scenarios show that when τ is more than $2T/3$, the difference in p.d.f. of the predicted time for the cases when parameters (α, β) are either known or estimated is practically negligible.

In the general case, when the durations of recruitment τ_i can be different in different centres, the Bayesian predicted time has the form

$$\widetilde{T}_1 = \frac{\mathrm{Ga}(K_2, 1)}{\sum_{i=1}^{N} \mathrm{Ga}(\hat{\alpha}N + k_i, 1)/(\widehat{\alpha}/\widehat{m} + \tau_i)} \tag{19}$$

and its distribution cannot be calculated in the closed form. However, the empirical density function and other characteristics of $\widetilde{T}_1$ can be calculated with high precision using Monte Carlo simulation. Actually, for given values $(\widehat{\alpha}, \widehat{m})$, it takes only minutes to generate 10^6 simulations of $\widetilde{T}_1$ for particular scenarios considered above.

5. ADAPTIVE ADJUSTMENT OF RECRUITMENT

Consider a multicentre study aimed to recruit n patients by N centres. Suppose that, at some intermediate time point t_1 the parameters of recruitment are re-estimated using current information and denote by $\widetilde{T}_1$ the predicted remaining time estimated as in Section 4 (see (15) or (19)). Assume that the targeted deadline for recruitment time is T and the predicted time is rather long. That means, on average either $\mathbf{E}[\widetilde{T}_1] > T - t_1$ or $\mathbf{P}(\widetilde{T}_1 > T - t_1) \geqslant p$, where p is some pre-specified probability, say, $p = 0.8$.

In this case with high probability, the trial may fail to complete the recruitment before deadline. The question is: how many additional centres should be added with the purpose of reaching the deadline with probability $1 - \delta$?

To get a closed-form solution, let us again assume for simplicity that, for all i, $\tau_i \equiv \tau$ and all new centres will be initiated with the same delay d $(d < T - t_1)$. Also, let the parameters (α, m) be known. In this case, to complete the recruitment before the deadline with probability $1 - \delta$, we need to add M new centres, where

$$M \geqslant \frac{A + Bz_{1-\delta}^2/2 + z_{1-\delta}\sqrt{AB + Q + B^2 z_{1-\delta}^2/4}}{\alpha B} \tag{20}$$

DOI: 10.1002/sim

$A = K_2 - (\alpha N + K_1)(T - t_1)/(\alpha/m + \tau)$, $B = m(T - t_1 - d)/\alpha$, $Q = K_2 + (\alpha N + K_1)(T - t_1)^2/(\alpha/m + \tau)^2$, K_1 is the number of patients recruited up to time t_1, $K_2 = n - K_1$, and $z_{1-\delta}$ is the $(1 - \delta)$-quantile of a standard normal distribution.

To prove (20), note that after time t_1 the existing N centres continue to recruit with the overall rate Λ, where we use the predicted rate $\widetilde{\Lambda}_1$ instead. If we add M new centres initiated with the same delay d, then after time $t_1 + d$ these centres add the total rate $\Lambda_+ = \mathrm{Ga}(\alpha M, \beta)$, where $\beta = \alpha/m$. Thus, the number of patients recruited in the interval $[t_1, T]$ can be represented as $X = \Pi_{\widetilde{\Lambda}_1}(T - t_1) + \Pi_{\Lambda_+}(T - t_1 - d)$, where $\widetilde{\Lambda}_1 = \mathrm{Ga}(\alpha N + K_1, \beta + \tau)$ (we use Bayesian re-estimation). The recruitment will be completed before deadline if $X \geqslant K_2$, and our task is to find M such that $\mathbf{P}(X \geqslant K_2) \geqslant 1 - \delta$. Finally, using the normal approximation of Poisson and gamma distributions and relation (3), we prove (20).

When parameters (α, m) are not known, we can use the values $(\widehat{\alpha}, \widehat{m})$ instead and formula (20) provides the first approximation very close to the actual number of centres required.

For a particular number of additional centres, the quality of the adjustment of the recruitment can be checked by simulation. If M new centres should be added with the same delay d, then the predicted remaining time $\widetilde{T}(M)$ can be represented as

$$\widetilde{T}(M) = d + \frac{\mathrm{Ga}(K_3, 1)}{\widetilde{\Lambda}_1 + \mathrm{Ga}(\alpha M, 1)m/\alpha} \tag{21}$$

where $K_3 = K_2 - \Pi_{\widetilde{\Lambda}}(d)$ is the number of patients left to recruit after time point $t_1 + d$. Note that as $d < T - t_1$, at large N, $K_3 > 0$ with probability close to 1.

Different characteristics of $\widetilde{T}(M)$ can be calculated numerically using very fast simulation of gamma variables involved in (21). In this case, for a large number of simulation runs, we recommend using the value $K_3 = [K_2 - \Pi_{\widehat{\Lambda}}(d)]_+$ to avoid negative values.

6. CONCLUSIONS

We propose a recruitment model where the patients arrive at different centres according to Poisson processes, with the rates viewed as a sample from a gamma distribution. This model mirrors the natural variation in recruitment rates observed in practice. Statistical analysis of several tens of completed trials shows that this model is in good agreement with real data and can serve as a basic recruitment model.

A statistical technique for predictive modelling of recruitment time for ongoing trials using current recruitment information is developed. It allows estimating various characteristics of the predicted remaining time, in particular, confidence (or prediction) intervals. The technique for adaptive adjustment of recruitment is suggested as well, and the approximate formula for calculating the number of additional centres that should be added with the purpose of accomplishing the study by a certain deadline with a pre-specified probability is provided.

Theoretical results are illustrated for different scenarios using Monte Carlo simulation.

ACKNOWLEDGEMENTS

The authors thank Prof. Stephen Senn for valuable comments and discussions and Dr James Roger for providing data on enrolment for a number of completed GSK studies.

DOI: 10.1002/sim

REFERENCES

1. Senn S. Some controversies in planning and analysing multi-centre trials. *Statistics in Medicine* 1998; **17**: 1753–1756.
2. Rukhin AL. Gamma-distribution order statistics, maximal multinomial frequency and randomization designs. *Journal of Statistical Planning and Inference* 2005; **136**:2213–2226.
3. Anisimov V, Fedorov V, Jones B. Optimal design of multicentre clinical trials with random enrolment. *GSK BDS Technical Report 2003-03*, 2003 (Available from: http://www.biometrics.com/5DBE1E43769243B380257146004342D6.html).
4. Anisimov V, Fedorov V. Design of multicentre clinical trials with random enrolment. In *Advances in Statistical Methods for the Health Sciences*, Balakrishnan N, Auget J-L, Mesbah M, Molenberghs G (eds). Birkhauser: Basel, 2006; 387–400.
5. Bates GE, Neyman J. Contributions to the theory of accident proneness. *University of California Publications in Statistics* 1952; **1**(9):215–254.
6. Haight F. *Handbook of the Poisson Distributions*. Wiley: New York, 1967.
7. Hougaard P, Lee M-LT, Whitmore GA. Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics* 1997; **53**:1225–1238.
8. Johnson NL, Kotz S, Adrienne WK. *Univariate Discrete Distributions* (2nd edn). Wiley: New York, 1993.
9. Kemp AW, Kemp CD. A rapid and efficient estimation procedure for the negative binomial-distribution. *Biometrical Journal* 1987; **29**(7):865–873.
10. Cox D. Some statistical methods connected with series of events (with discussion). *Journal of Royal Statistical Society B* 1955; **17**:129–164.
11. Cox D, Isham V. *Point Processes*. Chapman & Hall: London, 1980.
12. Snyder D, Miller M. *Random Point Processes in Time and Space* (2nd edn). Springer: Berlin, 1991.
13. Anisimov V, Fedorov V. Modeling of enrolment and estimation of parameters in multicentre trials. *GSK BDS Technical Report 2005-01*, 2005 (Available from: http://www.biometrics.com/7E5C600B533062FD80257146004333DC.html).
14. Senn S. *Statistical Issues in Drug Development*. Wiley: Chichester, 1997.
15. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions* (2nd edn), vol. 1. Wiley: New York, 1994.
16. Johnson NL, Kotz S, Balakrishnan N. *Discrete Multivariate Distributions*. Wiley: New York, 1997.
17. Feller W. *An Introduction to Probability Theory and Its Applications* (3rd edn), vol. 1. Wiley: New York, 1968.
18. Johnson NL, Kotz S, Balakrishnan N. *Continuous Univariate Distributions* (2nd edn), vol. 2. Wiley: New York, 1995.
19. Bernardo JM, Smith AFM. *Bayesian Theory*. Wiley: Chichester, 2004.
20. Stuart A, Ord K, Arnold S. *Kendall's Advanced Theory of Statistics*, vol. 2A. Edward Arnold: London, 1999.
21. Abramowitz M, Stegun IA. *Handbook of Mathematical Functions*. Dover: New York, 1972.

DOI: 10.1002/sim