Proceedings of the World Congress on Engineering 2008 Vol II
WCE 2008, July 2 - 4, 2008, London, U.K.

Using Mixed Poisson Models in Patient Recruitment in Multicentre Clinical Trials

Vladimir V. Anisimov *

Abstract— **To address different uncertainties associated with patient recruitment in multicentre clinical trials a mixed Poisson model is elaborated (patients arrive at different clinical centres according to delayed doubly stochastic Poisson processes with gamma distributed rates). Analytic approach to modeling and predicting the recruitment at the initial and ongoing stages is suggested. A Bayesian approach to re-estimating recruitment rates using data at intermediate time point and predicting the remaining recruitment time and the number of recruited patients in centres/regions is proposed. Different performance measures of centres are considered.**

Keywords: ***multicentre trial, patient recruitment, mixed Poisson model, Bayesian technique, estimation, prediction***

1 Introduction

The recruitment time (time required to complete patient recruitment) is one of the key decision variables at the design stage of multicentre clinical trials. Existing techniques of recruitment planning are mainly deterministic and do not account for various uncertainties associated with stochasticity in patient arrival, variation in recruitment rates between different centres and centre initiation delays.

It is quite natural to model the patient recruitment in a particular centre i as a delayed Poisson process $\Pi_{\lambda_i}(t-u_i)$. In real trials the recruitment rates $\{\lambda_i\}$ vary between different centres and it is suggested to model this variation using a gamma distribution. Centre delays u_i are not known in advance and can be also considered as some random variables. Therefore we come to a Poisson-gamma recruitment model considered in [1, 2, 3] and associated with mixed Poisson model with three levels of stochasticity. Statistical analysis of many studies shows a good fit of this model to real data [2].

The analytic approach for predicting recruitment time and the number of patients in different centres/regions is proposed. Section 2 describes the recruitment modeling technique. Section 3 is devoted to the prediction of the recruitment time at the initial and ongoing stages of the clinical trial. Section 4 deals with the prediction of the number of recruited patients in different centres/regions till the end of the trial or for a particular time interval at the initial and ongoing stages.

*Research Statistics Unit, GlaxoSmithKline, New Frontiers Science Park (South), Third Avenue, Harlow, Essex, CM19 5AW, United Kingdom Tel +44-1279-631579 Email: Vladimir.V.Anisimov@gsk.com

2 Modeling Recruitment

Consider a multicentre clinical study where n patients have to be recruited by N clinical centres. At study design stage it is important to predict the duration of the recruitment and how many patients will be recruited in particular regions as this may affect the power of statistical tests and the amount of drug supply required to satisfy patient demand in regions.

Denote by $n_i(t)$ the number of patients recruited by centre i up to time t and let $T(n,N)$ be the recruitment time. Assume that centre i is initiated at a random time u_i and patients arrive according to a Poisson process with rate λ_i. Thus, $n_i(t) = \Pi_{\lambda_i}(t-u_i)\chi(u_i \le t)$, where $\chi(A)$ is the indicator of the event A. Let

$$n(t) = \sum_{i=1}^{N} n_i(t) \tag{1}$$

be the total number of patients recruited up to time t by all N centres. The overall recruitment rate is

$$\Lambda(t) = \sum_{i=1}^{N} \lambda_i \chi(u_i \le t). \tag{2}$$

Therefore the process $n(t)$ is a nonhomogeneous mixed Poisson process with instantaneous rate $\Lambda(t)$. Assume that recruitment is described by a Poisson-gamma recruitment model [1, 2, 3], that is, the rates $\{\lambda_i\}$ are viewed as a sample from a gamma distributed population. This model accounts for the natural variation in recruitment over time and in recruitment rates between different centres and has been validated for many real trials with large enough (≥ 20) number of centres . This model can be also viewed in the empirical Bayesian setting as the rates are not known in advance and can be considered as random variables with a prior gamma distribution.

ISBN:978-988-17012-3-7

3 Recruitment Time Prediction

There are two basic stages of the recruitment prediction: the design stage (initial stage before study start) and the intermediate stage (ongoing study).

3.1 Initial prediction

Assume first that all centres are initiated simultaneously at time $t_0 = 0$. Denote by $\gamma(\alpha, \beta)$ a gamma distributed random variable with parameters (α, β) (shape and rate parameters, respectively). Assume that $\{\lambda_i\}$ are i.i.d.r.v. and in distribution $\lambda_i = \gamma(\alpha, \beta)$. Notice that parameters (α, β) have a simple interpretation: $\mathrm{E}\,\lambda_i = \alpha/\beta$, $\mathrm{Var}\,\lambda_i = \alpha/\beta^2$ and $\alpha = 1/c_v^2$, where c_v is the coefficient of variation of the recruitment rates, $c_v = sd(\lambda_i)/\mathrm{E}\,(\lambda_i)$.

Then $n(t) = \Pi_\Lambda(t)$, where $\Lambda = \gamma(\alpha N, \beta)$, and we can write a representation $T(n, N) = \beta\gamma(n, 1)/\gamma(\alpha N, 1)$. Thus, the variable $T(n, N)$ has a Pearson type VI distribution [5] with p.d.f.

$$p(x, n, N, \alpha, \beta) = \frac{1}{\mathcal{B}(n, \alpha N)} \frac{x^{n-1}\beta^{\alpha N}}{(x+\beta)^{n+\alpha N}},\ x \geq 0, \quad (3)$$

where $\mathcal{B}(a, b) = \int_0^1 x^{a-1}(1-x)^{b-1}\mathrm{d}x$ is a beta function.

In the general case, where centre i is initiated at some (possibly random) time u_i, the process $n(t)$ has a local rate $\Lambda(t)$ (see (1),(2)). Denote by

$$\Sigma(t) = \sum_{i=1}^{N} \lambda_i (t - u_i)\chi(u_i \leq t)$$

a cumulative rate in the interval $[0, t]$. Using properties of a mixed Poisson process and a gamma distribution, we can write the relation for the cumulative distribution function of the time $T(n, N)$: for any $T > 0$,

$$\mathrm{P}(T(n, N) \leq T) = \mathrm{P}(\gamma(n, 1) \leq \Sigma(T)). \quad (4)$$

For any particular type of distribution of u_i, the right-hand side in (4) can be evaluated very quickly with high precision using Monte Carlo simulation.

The expressions (3) and (4) at the design stage also allow to calculate the minimal number of centres needed to complete the recruitment in time with a pre-specified probability. e.g. to solve the optimization problem: find the least N_* such that $\mathrm{P}(T(n, N_*) \leq T) \geq p$, where p is some prescribed probability.

In real trials the parameters (α, β) of the recruitment model are not known in advance. At the initial stage the parameters can be evaluated using planned by study managers data on recruited number of patients in different regions, or using historical data from similar studies. At the ongoing stage the parameters can be estimated using recruitment data.

3.2 Prediction of the ongoing study

Consider study at some intermediate time point t_1 and develop the technique of the recruitment prediction for the remaining period using interim data. Suppose that up to time t_1 a centre i has recruited k_i patients. Let τ_i be the actual duration of recruitment in centre i, where $\tau_i = t_1 - u_i$, and $K_1 = \sum_{i=1}^{N} k_i$ be the total number of patients recruited up to time t_1. The aim is to construct the prediction of the remaining recruitment period using the interim data $\{k_i, \tau_i, i = 1, .., N\}$. Denote by $T(K_2, N)$ the remaining recruitment time where $K_2 = n - K_1$ is the remaining number of patients left to recruit.

Assume for simplicity that all centres belong to the same pool (the parameters (α, β) of a Poisson-gamma recruitment model are the same for all centres). If the parameters (α, β) were known, then k_i, as a Poisson mixed with gamma variable has a negative binomial distribution with parameters $(\alpha, \tau_i/\beta)$ ([4], p.199). Thus, given data $\{k_i, \tau_i,\ i = 1, .., N\}$, the log-likelihood function up to a constant has the form

$$\begin{aligned}\mathcal{L}(\alpha, \beta) \quad = \quad & \sum_{i=1}^{N} \ln\Gamma(k_i + \alpha) - N\ln\Gamma(\alpha) \\ & -K_1 \ln\beta - \sum_{i=1}^{N}(k_i + \alpha)\ln(1 + \tau_i/\beta),\end{aligned}$$

and ML estimators $(\widehat{\alpha}, \widehat{\beta})$ can be numerically calculated using two-dimensional optimization procedures.

Consider now the prediction of the remaining recruitment time. If the recruitment rates $\{\lambda_i\}$ were known, then the overall recruitment rate is $\Lambda = \sum_i \lambda_i$ and the remaining time can be represented as

$$T(K_2, N) = \gamma(K_2, 1)/\Lambda.$$

Using the estimators of Λ based on actual recruitment data we can construct the estimators of $T(K_2, N)$.

Let us use an empirical Bayesian technique. Suppose first that parameters (α, β) are known. As λ_i has a prior gamma distribution with parameters (α, β), then, given data, the posterior estimator of λ_i is $\widehat{\lambda}_i = \gamma(\alpha + k_i, \beta + \tau_i)$, and the posterior estimator of Λ is

$$\widehat{\Lambda} = \sum_{i=1}^{N} \gamma(\alpha + k_i, \beta + \tau_i). \quad (5)$$

Thus, the predicted remaining time can be represented as

$$\widehat{T}(K_2, N) = \gamma(K_2, 1)/\widehat{\Lambda}.$$

First, assume for simplicity that $\tau_i \equiv \tau$. Then $\widehat{\Lambda} = \gamma(\alpha N + K_1, \beta + \tau)$,

$$\widehat{T}(K_2, N) = \frac{\gamma(K_2, 1)}{\gamma(\alpha N + K_1, 1)}(\beta + \tau),$$

and therefore $\widehat{T}(K_2, N)$ has a Pearson type VI distribution (see (3)). In realistic situations, where τ_i can be different, $\widehat{\Lambda}$ at large enough N ($N \geq 10$) can be well approximated by a gamma distributed variable with the same mean and variance as of the variable $\widehat{\Lambda}$. Therefore, in general case $\widehat{T}(K_2, N)$ can be also approximated by a Pearson type VI distribution with corresponding parameters.

For the values of parameters (α, β) in these formulae we can take ML estimators $(\widehat{\alpha}, \widehat{\beta})$ using data in the interval $[0, t_1]$. Simulation results show that if $t_1/T > 0.5$, then the impact of additional errors in estimating parameters on the precision of prediction is practically negligible [2].

These results also allow to evaluate numerically at any stage of the trial the minimal number of centres N_* needed to complete recruitment with a given confidence, e.g. satisfy the relation

$$\mathrm{P}(\widehat{T}(K_2, N_*) \leq T - t_1) \geq p,$$

where p is some prescribed probability.

4 Patient Recruitment Prediction

4.1 Prediction at the initial stage

Consider the design stage of the study and assume for simplicity that all centres were initiated at the initial time. Let us first study the prediction of the total number of patients to be recruited in different centres and regions.

Consider some region I_s with N_s clinical centres. Denote by n_i the total number of patients recruited in centre i ($n(I_s)$ – number of patients recruited in region I_s), respectively, where $n(I_s) = \sum_{i \in I_s} n_i$. Then, given rates λ_i, $n(I_s)$ has a binomial distribution with parameters $(n, p(I_s))$, where

$$p(I_s) = \lambda(I_s)/\Lambda, \quad \lambda(I_s) = \sum_{i \in I_s} \lambda_i, \ \Lambda = \sum_{i=1}^{N} \lambda_i.$$

As according to Poisson-gamma model the recruitment rates λ_i are i.i.d.r.v. having a gamma distribution with parameters (α, β), then we can write representations: $\lambda(I_s) = \gamma(\alpha N_s, \beta)$ and $\Lambda = \gamma(\alpha N, \beta)$, respectively. Furthermore, $p(I_s)$ has a beta distribution with parameters $(\alpha N_s, \alpha(N - N_s))$ and, consequently, $n(I_s)$ as a mixed binomial with beta random variable has a beta-binomial distribution. Thus, in region I_s, for any $k = 0, .., n$,

$$\begin{aligned}&\mathbf{P}(n(I_s) = k)\\ &= \binom{n}{k} \frac{\mathcal{B}\Big(\alpha N_s + k, \alpha(N - N_s) + n - k\Big)}{\mathcal{B}\Big(\alpha N_s, \alpha(N - N_s)\Big)}.\end{aligned}$$

4.1.1 Prediction for a fixed time interval

Let us study now a prediction for a fixed time interval $[0, \Delta]$. Consider some region I_s with N_s clinical centres and denote by $n(I_s, t)$ the total number of patients recruited in this region in the interval $[0, t]$.

If all centres were initiated at time $t_0 = 0$, then in centre i, $n_i(t) = \Pi_{\lambda_i}(t)$ is a mixed Poisson process. Therefore, the process $n(I_s, t)$ is a doubly stochastic Poisson process with rate $\Lambda(I_s) = \gamma(\alpha N_s, \beta)$ and the variable $n(I_s, \Delta)$ has a negative binomial distribution with parameters $(\alpha N_s, \Delta/\beta)$. According to [4], p.199,

$$\mathrm{P}(n(I_s, \Delta) = k) = \frac{\Gamma(k + \alpha N_s)}{k!\Gamma(\alpha N_s)} \frac{\Delta^k \beta^{\alpha N_s}}{(\beta + \Delta)^{\alpha N_s + k}}, \ k = 0, 1, ...$$

This relation can be used for evaluating critical boundaries of the number of patients supposed to be recruited in different centres/regions.

If the centres are initiated at different times u_i, then $n(I_s, t)$ is developing as a nonhomogeneous mixed Poisson process $\Pi_{\Lambda(I_s,t)}(t)$ with instantaneous rate

$$\Lambda(I_s, t) = \sum_{i \in I_s} \lambda_i \chi(u_i \leq t).$$

In this case we have two levels of stochastic mixing, by random rates and by random delay times. If we know the distribution of delay times u_i, all characteristics of $n(I_s, t)$ can be computed very fast using Monte Carlo simulation.

In some special cases the mean and the variance of $n(I_s, t)$ can be calculated in the closed form. For example, if for centre i the time u_i is uniformly distributed in interval $[a, b]$, then as $t > b$,

$$\mathrm{E}\, n_i(t) = mt - m(a + b)/2,$$

$$\mathrm{Var}\, n_i(t) = (m^2 + s^2)(b - a)^2/12 + s^2(t - (a + b)/2)^2,$$

where $m = \alpha/\beta$, $\sigma^2 = \alpha/\beta^2$ [3]. Similar formulae can be derived for $t < b$. Correspondingly, the mean and the variance of $n(I_s, t)$ can be easy calculated as $n(I_s, t) = \sum_{i \in I_s} n_i(t)$ and $n_i(t), i \in I_s$, are independent.

4.2 Ongoing stage

Consider at some intermediate time t_1 the prediction of the number of recruited patients in the remaining period. Let (α, β) be the parameters of the model estimated using recruitment data up to time t_1. Consider some region I_s with N_s clinical centres. To predict future behaviour of recruitment we use empirical Bayesian approach and re-estimate rates in each centre.

Given data, the patients in centre i after time t_1 arrive according to a mixed Poisson process $\Pi_{\widehat{\lambda}_i}(t)$ with a random posterior rate $\widehat{\lambda}_i = \gamma(\alpha + k_i, \beta + \tau_i)$. Denote by $\widehat{n}_i$

and $\widehat{n}(I_s)$ the predicted number of patients supposed to be recruited in centre i and region I_s, respectively, in the remaining period. Then the variable $\widehat{n}(I_s)$ has a mixed binomial distribution with parameters $(K_2, \widehat{p}(I_s))$, where $\widehat{p}(I_s) = \sum_{i\in I_s} \widehat{\lambda}_i/\widehat{\Lambda}$ (see (5)) is a random variable. Assume for simplicity that $\tau_i \equiv \tau$. Let $k(I_s) = \sum_{i\in I_s} k_i$ be the number of patients recruited in the region I_s up to time t_1. Then $\widehat{p}(I_s)$ can be represented as

$$\frac{\gamma(\alpha N_s + k(I_s), 1)}{\gamma(\alpha N_s + k(I_s), 1) + \gamma(\alpha(N - N_s) + K_1 - k(I_s), 1)}$$

and has a posterior beta distribution with parameters $(\alpha N_s + k(I_s), \alpha(N - N_s) + K_1 - k(I_s))$. Thus, $\widehat{n}(I_s)$ has a posterior beta-binomial distribution and for any $j = 0, .., K_2$,

$$\mathbf{P}(\widehat{n}(I_s) = j \mid \text{data})$$
$$= \binom{K_2}{j} \frac{\mathcal{B}\Big(\alpha N_s + k(I_s) + j, \alpha(N - N_s) + n - k(I_s) - j\Big)}{\mathcal{B}\Big(\alpha N_s + k(I_s), \alpha(N - N_s) + K_1 - k(I_s)\Big)}.$$

In the realistic situations, where τ_i can be different, the approximate formulae of a similar form can be derived using at large enough N ($N \geq 10$) the approximation of the variable $\widehat{\Lambda}$ in (5) by a gamma distributed variable.

4.2.1 Prediction for a fixed time interval

Consider now prediction on some fixed interval $[t_1, t_1+\Delta]$. Denote by $\widehat{n}_i(t_1, \Delta)$ the predicted number of patients supposed to be recruited in centre i in this interval. Given recruitment data up to time t_1, the posterior distribution of $\widehat{n}_i(t_1, \Delta)$ is negative binomial ([4], p.199) with parameters $(\alpha+k_i,\ \Delta/(\beta+\tau_i))$, and for any $j = 0, 1, ...,$

$$\begin{aligned} &\mathrm{P}(\widehat{n}_i(t_1, \Delta) = j \mid \text{data}) \\ = \ &\frac{\Gamma(j + \alpha + k_i)}{j!\Gamma(\alpha + k_i)} \frac{\Delta^j(\beta + \tau_i)^{\alpha+k_i}}{(\beta + \tau_i + \Delta)^{\alpha+k_i+j}}, \end{aligned} \tag{6}$$

where (α, β) are estimated using recruitment data up to time t_1. Correspondingly,

$$\mathrm{E}\left[\widehat{n}_i(t_1, \Delta) \mid \text{data}\right] = \frac{(\alpha + k_i)\Delta}{\beta + \tau_i}. \tag{7}$$

This result can be used for evaluating different performance measures of recruitment, for example, the probability $P_i(0, \Delta, \text{data})$ that a centre i will not recruit any patients in the interval $[t_1, t_1 + \Delta]$ given that this centre did not recruit any patients during the recruitment duration τ_i, the mean predicted number of patients $M_i(\Delta, \text{data})$ to be recruited by this centre in the interval $[t_1, t_1 + \Delta]$, etc. Using relations (6), (7) we can easy calculate that

$$P_i(0, \Delta, \text{data}) = \frac{(\beta + \tau_i)^\alpha}{(\beta + \tau_i + \Delta)^\alpha},\ M_i(\Delta, \text{data}) = \frac{\alpha\Delta}{\beta + \tau_i}.$$

For example, $P_i(0, \Delta, \text{data})$ is increasing if the duration of the empty period τ_i during which a centre did not recruit any patients is increasing. This illustrates the fact that it is less likely to recruit in future patients by centres with longer empty periods.

These measures can be extended to groups of centres and can be used to alert low performing centres/regions.

5 Conclusions

The innovative analytic approach to evaluating various distributions associated with the prediction of the recruitment time and the number of patients to be recruited in different centres/regions at the design and ongoing stages of multicentre clinical trials is proposed. The approach is based on using mixed Poisson models and empirical Bayesian technique. These predictive distributions can be used for evaluating various trial characteristics, in particular, critical boundaries of the number of patients to be recruited in particular regions, critical supply levels needed to cover patient demand in these regions, the maximum number of places in hospitals, different associated costs, etc.

References

[1] Anisimov, V., Fedorov, V., "Design of multicentre clinical trials with random enrolment," in: N. Balakrishnan, J.L. Auget, M. Mesbah, and G. Molenberghs (Eds). *Advances in Statistical Methods for the Health Sciences*, Birkhuser, pp. 387-400, 2006.

[2] Anisimov, V., Fedorov, V., "Modeling, prediction and adaptive adjustment of recruitment in multicentre trials," *Statistics in Medicine*, **26**, 27, pp. 4958-4975, 2007.

[3] Anisimov, V., Downing, D., Fedorov, V., "Recruitment in multicentre trials: prediction and adjustment," in: J. Lopez-Fidalgo, J.M. Rodriguez-Diaz, B. Torsney (Eds.): *mODa 8 - Advances in Model-Oriented Design and Analysis*, Physica-Verlag, pp. 1-8, 2007.

[4] Johnson, N.L., Kotz, S., Adrienne W.K., *Univariate Discrete Distributions*, 2nd Edition, Wiley & Sons, New York, 1993.

[5] Johnson, N.L., Kotz, S., Balakrishnan, N., *Continuous Univariate Distributions*, V.1, 2nd Edition, Wiley & Sons, New York, 1994.