A Patient Recruitment Mixed Distribution Model with Center Initiation Cycle Times

James M. Powers*

September 28, 2012

Abstract

The paper presents two useful results for planning of enrollment in a clinical trial while accounting for site initiation times: one for when target recruitment occurs after all sites have been initiated and another for when recruitment target may possibly complete before all sites have been initiated. We summarize the development of the two sets of formulae and provide an application to the planning of global, multi-center clinical trials.
Key Words: Multi-center clinical trials, enrollment model, site initiation times.

1 Introduction

Models for prediction of patient recruitment in global, multi-center clinical trials have been considered in a number of papers including (Senn (1998), Carter et. al (2005), Anisimov and Fedorov (2007), Gajewski et al. (2008) and Anisimov (2011)). The potential impact to study timelines warrants sponsors of clinical trials to carefully consider enrollment rates as a variable factor in planning of the trial. Further, it is also essential to consider models that include site (clinical center) initiation times. Regulatory processes required to initiate a site, such that it can enroll patients in a clinical trial, are variant amongst different countries and regions. Therefore for multi-center, global trials an enrollment plan must include accounting for variability in site initiation.

The following defines the notation used throughout the remainder of the paper. First we define λ_i the enrollment rate and t_i as the initiation time for the i-th site, $n_\bullet(T) = \sum_{i=1}^{N} n_i(T)$ as the total number of enrolled subjects at time T across N centers. The cumulative rate at T is defined as

$$\Lambda_\bullet(T) = \sum_{i=1}^{N} \lambda_i(T - t_i) = \sum_{i=1}^{N} \Lambda_i(T). \tag{1.1}$$

The subscript $\bullet$ is used here and what follows to indicate summation with respect to the corresponding underlying indices. It is assumed that in each center i enrollment follows a Poisson process with rate λ_i.

*Innovation Advanced Analytics, Quintiles, RTP, NC, USA

The legitimacy of the normal approximations used later in (2.13) and (3.12) is based on the asymptotic normality of mixed Poisson distributions (Gurland (1958), Willmot (1986) and Johnson et al. (1993), Chapter 8). For (3.12) we also employ asymptotic behavior of random sums (Feller (1968), V.9 and VIII.5, Theorem 4).

The paper focuses on derivation of explicit formulae for two cases. In Section 2 the formulae for prediction of enrollment, accounting for staggered site initiation, are presented in the case where enrollment ends after all sites have been initiated. In Section 3, the results are extended to the case where the enrollment target of interest occurs before all sites being initiated.

2 Case 1: All sites initiated before target enrollment at moment T

We define the interval by which sites are expected to be initiated in by $[a, b]$, where a is a predetermined lower bound for site initiation and b is the upper bound. Practically these bounds are typically known by clinical operations and/or regulatory experts. For example, it may be quite known that a certain country will not approve site initiation processes to begin before x months. Therefore a lower bound a is created by the reality of regulatory submissions. Each regulatory jurisdiction will have lower and upper bounds that can be estimated from historical data and from study personnel who are updated on changing regulatory environments.

We define Case I as follows: $a \leq t_i \leq b$, where $T > b$. The expected value and variance of λ_i and t_i are

$$\mathbf{E}[\lambda_i] = \bar{\lambda},\ \mathbf{Var}[\lambda_i] = \sigma_\lambda^2 \tag{2.1}$$
$$\mathbf{E}[t_i] = \bar{t},\ \mathbf{Var}[t_i] = \sigma_t^2 \tag{2.2}$$

We assume that $\{\lambda_i\}$ and $\{t_i\}$ are mutually independent.

The expected value and variance of $n_\bullet(T)$, the number of total subjects enrolled at time T are determined using the law of iterated expectations and variances

$$\mathbf{E}[n_\bullet(T)] = \mathbf{E}[\Lambda_\bullet(T)] \tag{2.3}$$
$$\mathbf{Var}[n_\bullet(T)] = \mathbf{E}[\Lambda_\bullet(T)] + \mathbf{Var}[\Lambda_\bullet(T)] \tag{2.4}$$

Recall that $\Lambda_\bullet(T)$ is a sum of independent $\lambda_i(T-t_i)$. Therefore we need to find $\mathbf{E}[\lambda_i(T-t_i)]$ and $\mathbf{Var}[\lambda_i(T-t_i)]$, then sum these up.

$$\mathbf{E}[\lambda_i(T-t_i)] = \bar{\lambda}(T-\bar{t}) \tag{2.5}$$
$$\mathbf{Var}[\lambda_i(T-t_i)] = \sigma_\lambda^2\sigma_t^2 + \bar{\lambda}^2\sigma_t^2 + \sigma_\lambda^2(T-\bar{t})^2 \tag{2.6}$$

Combining 2.3-2.6, we obtain

$$\mathbf{E}[n_\bullet(T)] = N\bar{\lambda}(T-\bar{t}) \tag{2.7}$$
$$\mathbf{Var}[n_\bullet(T)] = N[\bar{\lambda}(T-\bar{t}) + \sigma_\lambda^2(T-\bar{t})^2 + \sigma_t^2(\sigma_\lambda^2+\lambda^2)] \tag{2.8}$$

These results are flexible to different assumptions about the distribution of λ_i and t_i. For example we may assume that λ_i are Gamma distributed, and also that t_i are Beta distributed. These two distributions are flexible to handle various historical data assumptions. For these assumptions we simply equate the previous results with the mean and variance results for these distributions.
If λ_i are Gamma (α, β) distributed then

$$\bar{\lambda} = \alpha\beta^{-1} \tag{2.9}$$
$$\sigma_\lambda^2 = \alpha\beta^{-2} \tag{2.10}$$

If t_i are Beta (p, q, a, b) distributed then

$$\bar{t} = (1-w)a + wb, \text{ where } w = p(p+q)^{-1} \tag{2.11}$$
$$\sigma_t^2 = (b-a)^2 pq[(p+q)^{-2}(p+q+1)]. \tag{2.12}$$

When $p = q = 1/2$ the Beta distribution is the Uniform and therefore (2.11) and (2.12) coincide with the results in Anisimov (2009).

In practical application, the parameters in (2.9) - (2.12) are replaced by their estimators based on historical data. Strictly speaking, one must incorporate their uncertainties but this is beyond the scope of the paper. If prior data are of substantial volume those uncertainties do not contribute much to the variance of $n_\bullet(T)$.

With the mean and variance now calculated it is possible to determine the probability of $n(T)$ given T is a certain value p_γ. Therefore the number of enrolled patients at a given time with a lower bound of probability is

$$n_\gamma(T) = N\bar{\lambda}(T - \bar{t}) - z_\gamma\sqrt{\mathbf{Var}[n_\bullet(T)]} \tag{2.13}$$

where z_γ is the value of the cumulative standard normal at a certain percentile p_γ.
We now may extend the above to the slightly more complex case of an enrollment target of interest before the end of the site initiation window.

3 Case 2: Not all sites initiated before enrollment target at moment T

The second case to consider is one where target enrollment is estimated before the end of site initiation. In this case we may view target enrollment as some interim target that occurs before the full enrollment target is met. In reality, it would be rare that study sites would be initiated right up to the end of enrollment. However it may be the case that add-on sites have the potential to be added to push enrollment to the final target.

Specifically the case under consideration is that where $a \leq T \leq b$, where the interval $[a, b]$ is large enough to have N centers open during the interval. In what follows we use the following properties of the random sum ($X_\bullet = \sum_{i=1}^{N} X_i$) of identically distributed, independent random variables (Feller, 1968):

$$\mathbf{E}[X_\bullet] = \mathbf{E}[N]\mathbf{E}[X_i] \tag{3.1}$$

$$\mathbf{Var}[X_\bullet] = \mathbf{E}[N]\mathbf{Var}[X_i] + [\mathbf{E}[X_i]]^2\mathbf{Var}[N] \tag{3.2}$$

The expected value of $n_\bullet(T)$ is defined as

$$\mathbf{E}[n_\bullet(T)] = \mathbf{E}[N(T)]\mathbf{E}[n_i(T)] \tag{3.3}$$

We then solve for each term separately. First, for $\mathbf{E}[N(T)]$, let $\varphi(t)$ be a density function of initiation times with the support set $[a, b]$. Then by definition of a density function

$$\mathbf{E}[N(T)] = N\int_a^T \varphi(t)dt = N\Phi(T), \tag{3.4}$$

where Φ is the cumulative distribution function. For a center that is opened before T,

$$\mathbf{E}[n_i(T)] = \mathbf{E}[\lambda_i(T - t_i)] = \mathbf{E}[\lambda_i]\mathbf{E}[(T - \mathbf{E}[t_i])] = \bar{\lambda}(T - \bar{t}(T)), \tag{3.5}$$

where the conditional expectation $\bar{t}(T)$ is

$$\bar{t}(T) = \mathbf{E}[t_i|t_i \le T] = \Phi^{-1}(T)\int_a^T t\varphi(t)dt \tag{3.6}$$

Combining (3.1)-(3.3)

$$\mathbf{E}[n_\bullet(T)] == N\Phi(T)\bar{\lambda}(T - \bar{t}(T)) \tag{3.7}$$

Observing that

$$\mathbf{Var}[N(T)] = N\Phi(T)(1 - \Phi(T)) \tag{3.8}$$

and ,

$$S^2(T) = \mathbf{Var}[n_i(T)]] = \bar{\lambda}(T - \bar{t}(T)) + \sigma_\lambda^2(T - \bar{t}(T)) + \sigma_t^2(T)(\sigma_\lambda^2 + \bar{\lambda}^2) \tag{3.9}$$

where

$$\mathbf{Var}[t_i|t_i \le T] = \sigma_t^2(T) = \Phi^{-1}(T)\int_a^T (T - \bar{t}(T))^2\varphi(t)dt, \tag{3.10}$$

one can verify that (see (3.2))

$$\mathbf{Var}[n_\bullet(T)] = N\Phi(T)S^2(T) + N\Phi(T)[1 - \Phi(T)]\bar{\lambda}(T - \bar{t}(T)). \tag{3.11}$$

Finally, using the results (see Introduction) on asymptotic normality, similar to (2.13) we have

$$n_\gamma(T) = N\bar{\lambda}(T - \bar{t}) - z_\gamma\sqrt{\mathbf{Var}[n_\bullet(T)]} \tag{3.12}$$

Note that $\bar{t}(T)$ is always less than T, see (3.6).

4 Conclusions

We have presented formulae for calculation of target number of enrolled subjects in a global, multi-center clinical trial, while accounting for the uncertainty inherent in site initiation times due to differing regulatory processes. The proposed formulae for the two cases are flexible to distribution assumptions about enrollment rate and site initiation times.

The analyst must carefully consider the particular trial and perform sensitivity analyses to examine alternative scenarios. These formulae provide for easy calculations using various assumptions.

Acknowledgements

The author would like to thank Dr. Valerii Fedorov for guidance in development of formulae presented; Dr. Vladimir Anisimov for review comments and discussion concerning enrollment predictive modeling; and, Cara Woodruff, Steve Southwick, Xiaoqiang Xue and Teresa Marhsall, as contributing members of the Advanced Analytics group at Quintiles.

References

Anisimov, V.V., and Fedorov, V.V. (2007), Modelling, prediction and adaptive adjustment of recruitment in multicentre trials, *Statistics in Medicine*, 26, 4958-4975.

Anisimov, V.V. (2009), Predictive modelling of recruitment and drug supply in multicenter clinical trials, *JSM Proceedings, Biopharmaceutical Section*, 1248-1259.

Anisimov, V.V. (2011),Statistical modelling of clinical trials (recruitment and randomization), *Communications in Statistics-Theory and Methods*, 40, 3684-3699.

Carter, R.E., Sonne, S.C., and Brady, K.T. (2005),Practical considerations for estimating clinical trial accrual periods: application to a multi-center effectiveness study, *BMC Medical Research Methodology*, 5, 11.

Feller, W. (1968),An introduction to probability theory and its applications. Volume II, 3rd edition, New York: John Wiley and Sons.

Gajewski, B.J., Simon, S.D., and Carlson, S.E. (2008), Predicting accrual in clinical trials with Bayesian posterior predictive distributions, *Statistics in Medicine*, 27, 2328-2340.

Gurland, J.A. (1958), A generalized class of contagious distributions. Biometrics, 14, 229-249.

Johnson, N.L., Kotz, N., and Kemp, A.W. (1993), Univariate discrete distributions. 2nd edition, New York: John Wiley and Sons.

Senn, S. (1998), Some controversies in planning and analyzing multi-center trials, *Statistics in Medicine*, 17, 1753-1765.

Willmot, G. (1986), Mixed compound Poisson distributions, *Astin Bulletin*, 16S, S59-S79.