Available online at www.sciencedirect.com

SCIENCE DIRECT®

Controlled Clinical Trials 25 (2004) 429–436

Controlled Clinical Trials

www.elsevier.com/locate/conclintrial

Application of stochastic processes to participant recruitment in clinical trials

Rickey Edward Carter*

Department of Biometry and Epidemiology, Medical University of South Carolina, 135 Cannon Street, Ste 303, P.O. Box 250835, Charleston, SC 29425, USA

Received 12 December 2003; accepted 1 July 2004

Abstract

The allocation of sufficient time for participant recruitment is one of the fundamental aspects in planning a clinical trial. This paper illustrates how a Poisson process can be used to determine an optimal period of time for participant recruitment by simulating Poisson arrival into a clinical trial. The simulation study provides the means to generate of an empirical probability density function for the recruitment time based on time-dependent changes in the accrual rate. From this empirical distribution, a clinical trial recruitment period can be planned to provide a high level of confidence (e.g., 90% probability) of enrolling the sample size within the planned amount of time given the simulation assumptions.

Keywords: Poisson process; Clinical trial; Accrual time

1. Introduction

A clinical trial can be simply defined as an experiment in which human subjects are exposed to medical treatments to better understand the effects of the treatments on the well being of the human subjects. Broadly speaking, clinical trials are commonly divided into many phases representing differing specific aims of the research. For example, a Phase III clinical trial is one that is commonly used to

* Tel.: +1 843 876 1114; fax: +1 843 876 1126.
E-mail address: carterre@musc.edu.

0197-2456/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.cct.2004.07.002

further demonstrate safety and efficacy of a treatment in a large number of patients who will eventually be candidates for the treatment [1]. Often, a Phase III clinical trial is a randomized, double blind, multicenter experiment in which hundreds to thousands of human subjects are enrolled and followed longitudinally over time [1]. The payments to cover the medical expenses the human subjects incur and the associated manpower to adequately conduct and monitor a Phase III clinical trial yield a final clinical trial budget that can easily reach tens of millions of dollars. Careful and regulated planning is employed to ensure that the trial's scientific objectives are fulfilled; however, one vital component of the trial planning is often overlooked.

The successful and timely recruitment of human subjects into a clinical trial is vital to the success of the trial [2]. For example, if an inadequate number of subjects are enrolled, then the study may fail to reject the null hypothesis due to lack of power. Similarly, if the recruitment period is too long, then other competing treatments may be released to the general market and human subject protection may be compromised [2,3]. A common approach to improving the generalizability of the study results and minimizing the recruitment time is to enlist multiple clinical sites to increase the rate of participant accrual. However, a careful balance between the number of sites and the expected number of subjects per site must be struck as to not compromise the scientific integrity of the clinical trial. Therefore, a clinical trial must be planned with an adequate understanding of the capabilities of each participating center.

The late Louis Lasagna summarized one of the key pitfalls in estimating the participant accrual rate. He is quoted as saying, "Investigators overestimate, many fold, the pool of available patients who meet the inclusion criteria and would be willing to enroll in a particular trial". The problem is so widespread that it is known as "Lasagna's law" in the clinical trial field [4]. However, limited quantitative exploration of the phenomenon has been conducted to date [4]. A second factor that has not been previously discussed in the literature is how to adequately plan the recruitment period given a reasonable estimate of the accrual rate.

Experience has shown that even when tangible efforts are made to estimate the accrual rate for a clinical trial, the period of time required to accrue the necessary number of subjects is longer than planned. It is proposed that one reason for this delay is not the estimated accrual rate but rather how the rate is used. A common approach to estimating the expected time to accrue the sample size in a clinical trial is to divide the sample size by the expected rate. For example, if subjects are eligible for recruitment at a rate of 10 subjects/month, then one could estimate that it would take 10 months to enroll 100 subjects. There are two factors that are not often explicitly explored when this method of estimating the recruitment period is employed. First, the rate of accrual is often assumed fixed when in actuality it often represents an average rate. Second, when the estimated rate represents an average, little or no attention is directed towards the variance of the estimated accrual rate. When these factors are combined and applied to the previous example, it becomes evident that "on average" it may take 10 months to enroll 100 subjects, but the actual amount of time to accrue 100 subjects may vary. The statistical framework for the methodology will be developed using the Poisson distribution. This discrete distribution has widespread applications that often include modeling the number of counts [5,6]. In the context of clinical trials, the Poisson distribution can be used to estimate the accrual probabilities associated with the assumed accrual rate. The approach this paper follows is to provide an empirical technique to approximate the distribution of the accrual time so that a probabilistic approach to planning the recruitment period of a clinical trial can be used. Stochastic process theory will serve as the foundation for the technique.

The remainder of this paper is organized as follows. In the next section, the basic principles of the stochastic model are formalized and developed. Next, the methods are illustrated by example. Finally, a discussion and general extensions are discussed as a conclusion.

2. Methodology

2.1. Poisson process

A stochastic process is a collection of random variables indexed by time [7]. Generally, stochastic processes are either classified as a discrete time or a continuous time process. One particular continuous time stochastic process that has widespread application is a Poisson process. A Poisson process is a continuous time stochastic process that satisfies the following three assumptions: (1) the number of events occurring at a particular time is independent of the number of events that occurred previously; (2) the average rate at which events occur is constant over time; and (3) events occur one at a time [7]. Simulations can be used to ease the limitation imposed by assumption (2) with the trade off of loosing the ability to reduce the Poisson process probability to a Poisson distribution. However, as it will be shown later, allowing for a time-dependent variation in the rate is advantageous in planning a clinical trial.

2.2. Notation and model

Let λ equal the expected number of participants enrolled per day across all sites averaged over all days and let T denote the time in days to accrue N subjects, where N represents the total planned sample size. Denote N_t as the total number of participants enrolled before time t, $t<T$ and let X_t denote the number of new subjects enrolled on day t such that $N_{t+1}=N_t+X_t$. Denote $P(N_t=n)$ as the probability that there have been exactly n participants enrolled before time t. Then,

$$P(X_t = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

is a Poisson distribution with mean λt. To account for the randomness or systematic changes in λ, denote λ_{kt} as the expected number of participants enrolled on day t for site k, $k=1,\ldots,K$. Let the rate $\lambda_t^* = \sum_K \lambda_{kt}$ represent the mean number of subjects enrolled on day t across all sites.

Therefore, the total number of new subjects, x, enrolled on day t with rate λ_t^* under the Poisson distribution is expected to follow the following distribution:

$$P(X_t = x) = \frac{e^{-\lambda_t^*}\left(\lambda_t^*\right)^x}{x!} \tag{1.1}$$

Utilizing the Poisson distribution specified in Eq. (1.1), it is possible to simulate random values for x. In particular, one could empirically estimate the distribution of T by considering the following pseudo-code:

Step 1 Initialize simulation iteration j, $j=1$–1000.
Step 2 Generate a random Poisson value, x, for X_t using λ_t^* as the Poisson mean.

Step 3 Update $N_j=N_j+x$, $T_j=T_j+1$.
Step 4 Continue Steps 1 through 3 until N_j reaches the desired sample size for all j iterations.
Step 5 Calculate percentiles for T based 1000 T_j estimates for T.

To illustrate this plan literally, on the first simulation iteration, Step 2 would produce random, non-negative integers that represent the number of people enrolled on a given day. For example, the first 10 numbers generated could be 1, 0, 2, 0, 0, 1, 0, 0, 0 and 1 with the interpretation of a total of 5 subjects were enrolled in the first 10 days with 1 subject on day 1, 2 subjects on day 3, etc.

3. Example

A multicenter clinical trial evaluating the palliation benefits of radiation therapy and esophageal stents in subjects with terminal esophageal cancer requires a planned sample size of 324 subjects. The anticipated accrual rate for the trial is assumed to be a total of 18 participants/month across multiple centers yielding a daily estimate of λ=0.591 participants/day. The goal is to estimate the total number of days (or months) needed to recruit the 324 required subjects into the trial. One could easily estimate that the recruitment will take 18 months (324/18) by assuming that the rate of 18 subjects/month will remain constant during the course of the study; however, the application of the proposed method will explore how feasible this estimate is.

Before the simulation is considered, one could assume that the subjects will arrive according to a Poisson process. Then by assuming the rate remains constant over the 18-month period (or 548 days), a Poisson distribution with mean λ*=0.591*548=324 participants per 18-month period is defined. One practical use of this approach is that it enables a probability to be associated with the recruitment period, T. In particular by the normal approximation to the Poisson distribution, $T\sim N$ (μ=324, σ^2=324). Therefore, the probability of recruiting the desired 324 participants in the 18-month period is 0.5. This probability actually is an upper bound for the true probability since the maximum rate of 18 participants/month is not likely to occur over the entire study period. For example, during the first part of the trial, not all centers may be open for enrollment, and as such the actual enrollment rate should be lower during the initial months of the trial. The fact that the study has at best a 50% likelihood of obtaining the desired sample size in the 18 months suggests two questions need to be answered. First, given a rate of 0.591 participants/day, how long should the recruitment period be planned to give a level of confidence, for example 90% confidence, of recruiting the 324 subjects? Second, if 18 months are fixed, what does the expected rate need to be to give 90% certainty of enrolling 324 subjects within 18 months? Solving the following probability statement in terms of t yields the recruitment time needed to ensure, with a probability of 0.90, that the sample size will be obtained.

$$P(N\geq 324|\lambda = 0.591,\ T = t) = 0.90,$$

or equivalently

$$P(N<324|\lambda = 0.591,\ T = t) = 0.10.$$

By the normal approximation to the Poisson,

$$P(N<324|\lambda = 0.591,\ T = t) \approx P\left(Z<(323.5 - \lambda t)/(\lambda t)^{1/2}\right), \tag{1.2}$$

which is equivalent to solving

$$\lambda t + 1.282(\lambda t)^{1/2} - 323.5 = 0$$

$$\left(t^{1/2} + 22.56999\right)\left(t^{1/2} - 24.23706\right) = 0$$

in terms of t. Therefore, when λ=0.591, t would be approximately 587.4 days or about 20 months.

The solution of the second question would be similar to the above solution with the exception of solving Eq. (1.2) for λ instead of t. However, this application of a Poisson process requires that the rate is constant with respect to time. Simulations and the associated empirical distribution alleviate this assumption form the calculations.

The empirical density function (Fig. 1) generated from the simulation indicates the mean recruitment time is about 550 days (~18 months) and gives a visual indication of the variance of the potential accrual period. The cumulative distribution (Fig. 2) can be used to quickly determine the number of days needed to recruitment based on a probability. For example, to have about a 90% chance of obtaining the desired sample size would require recruitment to last about 580 days. Note that this approximation is very

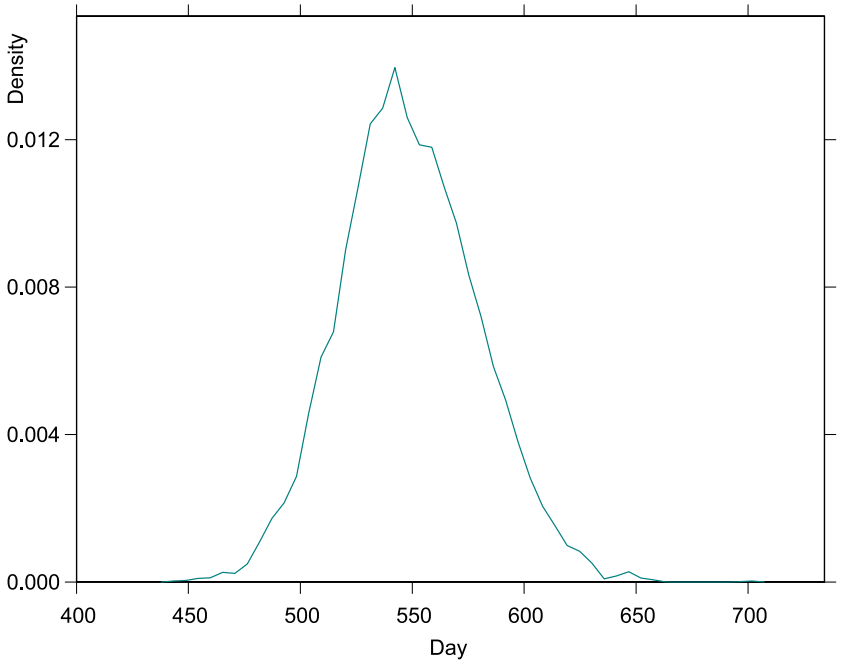

Fig. 1. Empirical density function for the time needed to recruit 324 participants based on a Poisson process with rate 0.591 participants/day (18/month).

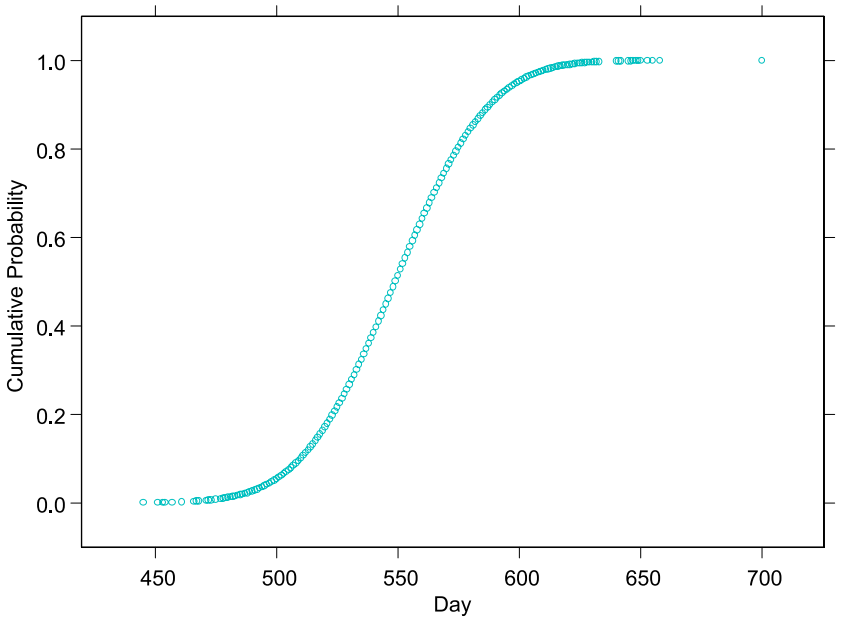

Fig. 2. Empirical cumulative distribution function for the time needed to recruit 324 participants based on a Poisson process with rate 0.591 participants/day.

similar to time of 587.4 days calculated earlier, but as illustrated in the following example, the simulation allows for greater flexibility in trial planning.

Suppose a sample size of 50 is needed in a pilot study and suppose that one participant is expected per month. Therefore, a crude estimate of recruitment time is 50 months. Common practice in clinical trials is to enlist multiple sites when the recruitment time is longer than feasible. Fifty months for a pilot study is unreasonable, but 10 months may be suitable. It can be assumed that five sites enrolling participants at the rate of one per month could enroll the desired sample size of 50 in 10 months. However, not all sites are likely to begin enrolling participants on the same day. Suppose it is known in advance that sites 1 and 2 will both begin on the same day. Site 3 will start 1 week later. Site 4 will begin 2 weeks after sites 1 and 2, and finally site 5 will begin 4 weeks after sites 1 and 2. Under these assumptions, a 10-month recruitment time is no longer realistic. Suppose further that a 90% chance of obtaining the sample size of 50 within T days is desired.

The rate in this model varies according to the day. Days 1 through 7, the rate is 2 participants/month (sites 1 and 2 active). Days 8 though 14, the rate is 3 participants/month. It is not until day 29 that the trial is operating at the maximum rate of 5 participants/month. Fig. 3 illustrates the distribution of the days needed to reach 50 participants based on simulation incorporating the variable rates. The simulation indicates that approximately 12 months (or 375 days) are needed to obtain the desired sample size with a probability of 0.90. Suppose 10 months is the maximum time available for the pilot study. If one more site is enrolled on day 1, then the 375 days is reduced to around 305 days, which is approximately 10 months.

This simulation suggests that if all six sites begin enrolling subjects at their respected rates within a reasonable amount of time (this simulation assumes that all sites begin enrolling subjects within the first month), then there is a high probability of meeting the 10-month deadline without using an excessive

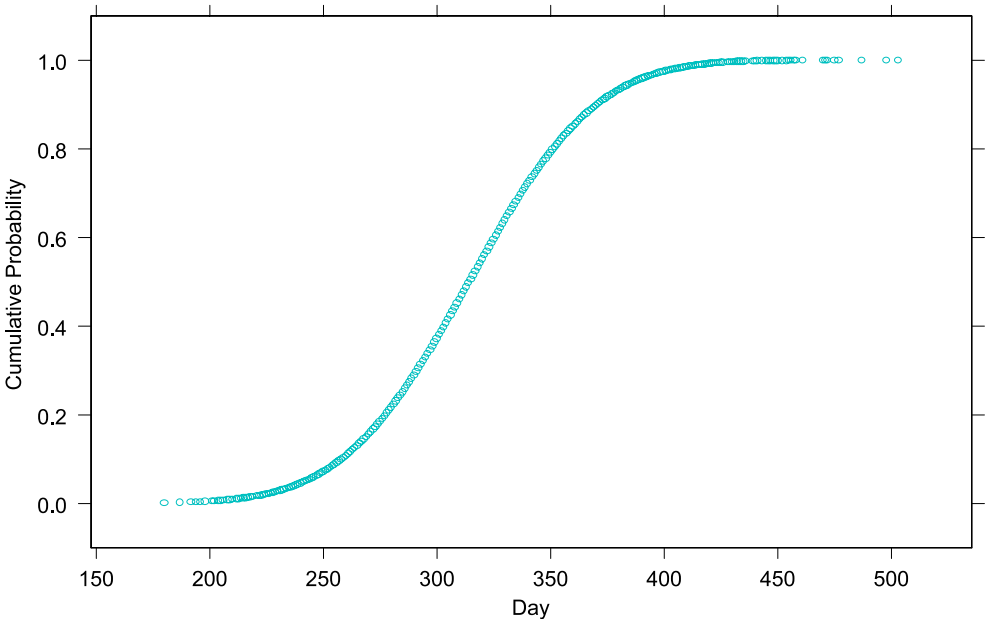

Fig. 3. Empirical cumulative distribution function for the time needed to recruit 50 participants based on a Poisson process with a rate conditioned on the number of sites recruiting participants.

number of sites. Additional simulations can be conducted to apply principles such as Lasagna's law to the estimated rates in order to learn more about how the assumptions affect the trial duration.

4. Conclusion

This paper applied Poisson processes and simulations to clinical trials by providing a method to empirically estimate the distribution of accrual time in clinical trials. In the context presented in the examples, the accrual rate was assumed to be known and constant for a single site. It is acknowledged that the determination of the appropriate accrual rate may be the most difficult aspect of the simulation since there are numerous milestones in a clinical trial (such as screening and the informed consent process) that may result in the reduction of the expected accrual rate. The approach of a constant rate was chosen for brevity and clarity; however, simple extensions of the model are possible to allow for the fluctuation in the estimated accrual rate. One such extension would be to periodically modify the accrual rate in response to planned clinical trial advertisements or seasonal trends in disease incidence. An approach to account for unforeseen reductions in accrual rate is to reduce the expected accrual rate by multiplying it by a random, uniformly distributed number on the interval (0,1). This effectively reduces the expected accrual rate, on average, by 50%, but allows instances where the rate is reduced more substantially. The simulation study presented in this manuscript has been performed using the SAS System [8], and selected SAS programs from this manuscript have been made available at http://people.musc.edu/~carterre.

The simulation approach to estimating the recruitment period allows for additional calculations that are valuable during a trial's planning stage. In particular, similar simulations can facilitate management

of clinical trials that have low accrual rates and long follow up times. For example, if one incorporates participant payments into to the simulated enrollment, the simulations could calculate estimated cash flow associated with the trial's operation. This would be an important consideration if the capital for conducting the trial were received over time. All of these techniques discussed are illustrated as being conducted in the planning phase, but conditional models can be developed as a trial progresses to assess whether the assumed rates were reasonable estimates of the actual number of eligible participants and whether or not the trial is likely to finish on schedule. In summary, utilizing a Poisson distribution to estimate the empirical distribution of the accrual period has several practical uses in clinical trials.

References

[1] Friedman LM, Furberg CD, DeMets DL. Fundamentals of clinical trials. 3rd ed. St. Louis, Missouri: Mosby-Year Book; 1996.
[2] U.S. Department of Health and Human Services, Draft guidance for clinical trial sponsors: on the establishment and operation of clinical trial data monitoring committees http://www.fda.gov/cber/guidelines.htm; 11-15-2001.
[3] DeMets DL, Fleming TR. The independent statistician for data monitoring committees. Stat Med 2004;23:1513–7.
[4] Spilker Bert. Patient recruitment. Guide to clinical trials. First edition. Philadelphia: Lippincott-Raven Publishers; 1996. p. 85–92.
[5] Ross S. A first course in probability. Fourth ed. New York: Macmillan College Publishing; 1994.
[6] Casella G, Berger RL. Statistical inference. 2nd ed. North Scituate, MA: Duxbury Press; 2002.
[7] Lawler Gregory F. Introduction to stochastic processes. First edition. London: Chapman & Hall; 1995.
[8] The SAS System for Windows. SAS Institute 8. Cary, NC; 1999.