CLINICAL TRIALS ARTICLE Clinical Trials 2012; 9: 681–688

Modeling and prediction of subject accrual and event times in clinical trials: a systematic review

Xiaoxi Zhang[a] *and Qi Long*[b]

Background Modeling and prediction of subject accrual and event times in clinical trials has been a topic of considerable interest for important practical reasons. It has implications not only at the initial planning stage of a trial but also on its ongoing monitoring.

Purpose To provide a systematic view of the recent research in the field of modeling and prediction of subject accrual and event times in clinical trials.

Methods Two classes of methods for modeling and prediction of subject accrual are reviewed, namely, one that uses the Brownian motion and the other uses the Poisson process. Extensions of the accrual models in multicenter clinical trials are also discussed. Trials with survival endpoints require proper joint modeling of subject accrual and event/lost-to-follow-up (LTFU) times, the latter of which can be modeled either parametrically (e.g., exponential and Weibull) or nonparametrically.

Results Flexible stochastic models are better suited when modeling real trials that does not follow constant underlying enrollment rate. The accrual model generally improves as center-specific information is accounted for in multicenter trials. The choice between parametric and nonparametric event models can depend on confidence on the underlying event rates.

Limitations All methods reviewed in event modeling assume noninformative censoring, which cannot be tested.

Conclusions We recommend using proper stochastic accrual models, in combination with flexible event time models when applicable, for modeling and prediction of subject enrollment and event times in clinical trials. *Clinical Trials* 2012; **9**: 681–688. http://ctj.sagepub.com

Introduction

Modeling and prediction of subject enrollment as well as event times (for trials with survival endpoints) has long been a topic of considerable interest to the operational teams of clinical trials. It has important implications on budgeting and resource allocation not only at the design stage of a trial, but also to the execution of an ongoing trial.

In the area of accrual modeling, constant or piecewise constant assumptions are commonly made. Close examination of real trial accrual almost always reveals an enrollment pattern that is neither constant nor the shape of a step-function (i.e., piecewise constant). More often than not, the enrollment increases in the beginning of a trial, as the sites gradually get upto speed in the enrollment activities, or as additional sites start recruitment. In addition, the conventionally used enrollment monitoring and prediction models assume the accrual process to be "static" (i.e., free of random variations), and hence usually rely on linear extrapolation to project future accrual. However, in reality, the number of subjects enrolled fluctuates from one day to another. This is generally true even after all sites are actively recruiting and the overall accrual rate is roughly stable, which is due to the stochastic nature of the enrollment process. Since the subjects are independent individuals, their arrival and participation in a trial are random events by nature.

[a]Pfizer Inc., New York, USA [b]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, GA, USA
Author for correspondence: Xiaoxi Zhang, Pfizer Inc., 235 E 42nd Street, New York, NY 10017, USA.
Email: xiaoxi.zhang@pfizer.com

10.1177/1740774512447996

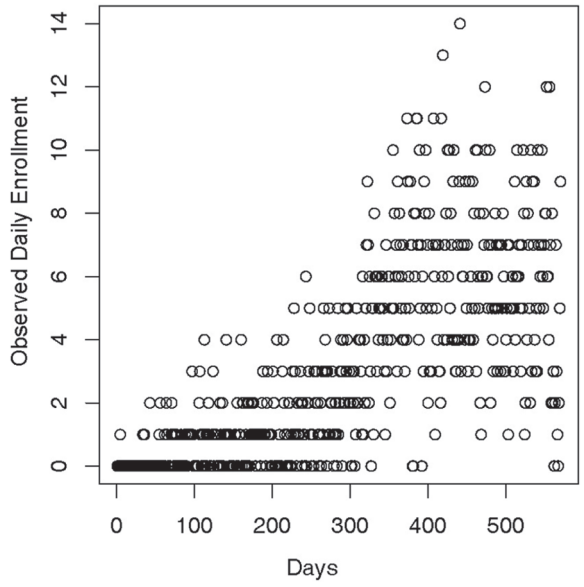

Figure 1. Observed daily accrual in a real oncology trial [1].

For this reason, stochastic models that capture the randomness in the enrollment process are more appropriate than the static approach that is widely used in practice. Consequently, an interval estimate that quantifies the uncertainty is more desirable than a point estimate (e.g., from simple linear extrapolation). Two main features of real life accrual process, time dependency and inherent random fluctuation, are evident in Figure 1, which depicts the observed daily accrual of a real oncology trial [1]. This multicenter randomized Phase III study was conducted to evaluate novel treatments of adjuvant colon cancer and will be discussed in more details later in this article.

If the enrollment runs faster than anticipated, the sponsor generally has the option of limiting subject recruitment, should faster enrollment be undesirable, for instance, in certain adaptive trials. It is generally of concern when subject enrollment in a trial trails behind and leads to an extended study duration. This becomes a more prominent issue when drug development becomes an oftentimes competitive field with multiple sponsors competing for the first-in-class status of a compound, in which case a delay in access to market may translate into a palpable impact on financial interests. In extreme cases, a trial may even have to conclude with an inadequate sample size as a result of slow accrual, for example, when it becomes impossible to enroll the prespecified number of subjects within a reasonable duration. The reduced sample size undermines the ability to draw scientifically sound inferences based on the study results. This is particularly problematic when large clinical trials fail to yield conclusive results to the scientific community.

In practice, slow accrual is not uncommon in clinical trials. It is reported that as many as 80% of clinical trials fail to meet their original accrual goals [2]. There are various reasons that contribute to slow accrual. Common reasons include, but are not limited to, overly optimistic anticipation at the design stage, similar trials competing for resources, stringent inclusion/exclusion criteria, limited subject pool, and slow center start-up and so on. The wide use of multinational–multicenter trials, as a consequence of large sample size and/or regulatory requirements, brings additional challenges in accrual modeling and prediction. Despite best efforts to prevent slow accrual at the design stage, it is difficult, if at all possible, to address all causes of slow accrual. Consequently, it is important to be able to monitor and predict accrual throughout the duration of a trial, which allows any detected issue to be addressed with effective measures and helps to ensure timely conclusion of subject recruitment.

In a trial with survival endpoint, the power of the study is driven by the number of events observed. That is to say, in such trials, it is equally, if not more, important to predict when a given number of events will occur. In this case, modeling and prediction of subject accrual alone is of limited use. Nevertheless, it is an indispensable component of proper modeling and prediction of event times when accrual activities are still ongoing at the time of event monitoring; in other words, proper accrual modeling is necessary to make accurate prediction for event times. In addition, different considerations are called upon when modeling subjects who have already enrolled and those who are yet to be enrolled. Similarly, subjects who are still at risk of the event of interest (i.e., still being followed) and those who have experienced the event of interest or have lost to follow-up also need to be modeled differently. On a practical note, if the enrollment activities are still ongoing, there are generally multiple ways to reach the prespecified number of events, as determined in the power calculation, that is, extending the minimum follow-up duration, increasing the target number of subjects to be enrolled, or a combination of the two. If the original accrual goal has been met and the enrollment activities have stopped, there generally exist extra hurdles to restart the enrollment activities. As a result, the available options to address detected issues may become limited, which further underlines the importance of timely monitoring and prediction of event times.

In a broader sense, modeling and prediction of subject accrual, or event times when relevant, is useful not only in determining the timing for final analysis but also in determining the timing of

interim analyses as well as in the planning of data safety monitoring committee (DSMC) meetings.

This article is intended to provide a review of recent research on modeling and prediction of subject accrual and event times in clinical trials. The goal is to present proper tools that can assist sponsors/investigators to detect issues early and subsequently take necessary measures to address them. We examine a collection of statistical models in the literature for modeling and prediction of subject accrual and event times in sections "Modeling and prediction of accrual" and "Modeling and prediction of event times," respectively, and summarize some relevant numerical results that have been reported in cited papers in section "Numerical Results." In section "Discussions," some discussion remarks are provided.

Modeling and prediction of accrual

As noted in "Introduction," the randomness in the subject accrual process calls for stochastic approaches. In what follows, we review two major classes of statistical models reported in the literature. One uses the Brownian motion and the other one is based on the Poisson process. Of note, there are other approaches taken in accrual modeling, for example, Ying [3] discusses a parametric modeling of subject accrual using a truncated exponential distribution.

Brownian motion–based models

In a Brownian motion, the increments between two time points are normally and independently distributed, and the variance of any increment is proportional to the length of the duration [4]. Lai *et al*. [5] propose to model the difference between the cumulative number of subjects enrolled and its expected value as a Brownian motion. In their article, they propose to model the expected value of the cumulative recruitment as a linear function of time, which essentially assumes the underlying enrollment rate to be constant over time. The slope of the linear function and the variance parameter in the Brownian motion model are estimated using the least squares (LS) method or method of moments.

In a follow-up work, Zhang and Lai [6] extend the original work into a fractional Brownian motion model, in which the increments are no longer assumed to be independent and the covariance structure is governed by a single Hurst parameter [7]. When the parameter is greater (or lesser) than 0.5, the fractional Brownian motion model implies positively (or negatively) correlated increments. When the Hurst parameter equals 0.5, it reduces to a standard Brownian motion. In the six trials they examine, they estimate the Hurst parameter to be ranging from 0.35 to 0.70. Therefore, the authors argue that the fractional Brownian motion model is more flexible and fits the data better. Since the (fractional) Brownian motion is continuous, they acknowledge that their model is more appropriate if one is interested in the cumulative accrual (rather than the incremental accrual), in which case the (fractional) Brownian motion serves as a better continuous *approximation* of the underlying enrollment process, which is discrete.

Poisson process–based models

Since the number of subjects enrolled in a trial is an integer that increases over time with random daily increments, its characteristics can be naturally captured by a Poisson process [4]. Throughout this article, we make the distinction of the *observed* accrual rate, which could vary considerably from one day to the next as a result of the innate randomness in the enrollment process, from the *underlying* accrual rate, which is the rate of the Poisson process and is assumed to be constant in a homogeneous Poisson process (HPP) or time varying in a nonhomogeneous Poisson process (NHPP).

Models using HPP

The Poisson process is a stochastic-counting process of events, in which the events are independent [4]. When one studies subject enrollment in a trial, it is recognized that the subjects are generally independent individuals and hence their enrollments are plausibly independent events. Even though the number of subjects enrolled in a trial usually varies from one day to the next, this may be largely due to the stochastic nature of this counting process. When the underlying accrual rate is constant, the enrollment process follows a HPP. It is the property of a HPP that the waiting times between two consecutive events are independently and identically distributed (i.i.d.) exponential random variables with the rate of the Poisson process.

Gajewski *et al*. [8] propose a statistical model for subject accrual, which assumes that the waiting time between the recruitment of two consecutive subjects are i.i.d. exponential random variables. Even though not explicitly stated as such, this approach essentially assumes a HPP. By definition, the longer the average waiting time between any two consecutive subjects, the lower the rate of the Poisson process is. While the rate of the Poisson process is assumed to be constant, the *observed* daily enrollment is still random and fluctuates from one day to another. Gajewski *et al*. [8] adopt a Bayesian

framework for inference and prediction. They use a conjugate prior distribution for the underlying enrollment rate and suggest asking the investigator two questions to elicit prior information, that is, how long one anticipates it would take to enroll all N subjects (T) and what level of confidence one has in the answer to the previous question ($0<P<1$). Under conjugate prior distribution, the above information is translated as a prior knowledge of enrolling NP subjects in a time duration of TP. The posterior predictive distribution is then used for accrual prediction. Compared to the naive methods in which the accrual is assumed to be constant or piecewise constant, the underlying randomness in enrollment is properly respected in this model. In addition, their Bayesian approach allows for incorporation of prior knowledge, which, as the authors argue, unlike in many other problems, is generally available in the problem at hand.

Under the same constant accrual rate assumption, Anisimov and Fedorov [9] propose a multicenter accrual model, which assumes the underlying accrual rate to be constant at each center and allows for staggered center start-up times. The authors assign a common Gamma prior distribution to the center-specific accrual rates. In the special case of all centers starting at the same time, the underlying overall accrual rate, which is the sum of multiple Gamma random variables, still follows a Gamma distribution. It follows that the overall enrollment across centers assumes a HPP where the overall Poisson rate is the sum of rates across centers. In a more general setting, when not all centers start at the same time, the underlying overall enrollment rate assumes the shape of step-function and the marginal distribution of the overall enrollment follows a doubly stochastic process, which is also known as a Cox process [10]. The authors suggest deriving the posterior distribution for each center in order to predict future recruitment. Notably, this approach addresses the issue with accrual that is present in many multicenter clinical trials, that is, possible staggered center start-up time and differential enrollment capacity across centers.

Models using NHPP

Despite the obvious improvement in the approaches that employ a HPP to model overall or center-specific accrual, these approaches have their limitations. In most real clinical trials, the assumption of constant underlying accrual rate (either overall or within each center) may not hold. To address this, Zhang and Long [11] propose a flexible model for subject accrual using a NHPP, which employs a smooth nonparametric fit of the underlying accrual rate, and propose to perform inference and prediction under the Bayesian framework. When the underlying enrollment rate is constant over time, their proposed model reduces to a homogenous Poisson process. Considering that the enrollment profile in different trials can differ considerably, Zhang and Long propose to use B-splines to model the underlying accrual rate over time. Assuming the independence of the prior parameters, they suggest two ways to specify the prior distributions. One is data driven and the other one is derived through inquiring the anticipated maximum underlying enrollment rate (i.e., the average accrual rate after the enrollment stabilizes) and the coefficient of variation (c.v.) of this estimate. In their work, it is assumed that the underling future enrollment rate is the same as the underlying rate at the time of the accrual modeling. In many trials, the underlying enrollment rate increases in the beginning, and once all sites reach their enrollment capacity, the underlying rate tends to stabilize. Hence, their proposal is more likely to err on the conservative side, that is, underestimating the underlying rate and overestimating the time needed to achieve full enrollment. This practice is similar to last observation carried forward (LOCF) but it accounts for the natural fluctuation in the accrual process through stochastic modeling. The estimation of the posterior predictive distribution is achieved through Markov chain Monte Carlo sampling. As a side note, it is possible to utilize parametric models in place of B-splines in the event that the underlying accrual rate can be approximated by suitable parametric functions. In this case, the inference for future accrual can be derived from the parametric function rather than using the estimated underlying accrual rate at the time of the enrollment modeling. However, as the authors note, the major difficulty lies in the appropriate specification of such parametric functions, the choice of which may not be obvious in many situations. We also comment that, unlike the work by Anisimov and Fedorov [9], one drawback of the NHPP model is that it does not explicitly utilize the center-specific accrual information, even though the flexible fit of the underlying accrual rate implicitly accommodates, to certain degrees, the staggered center start-up time and possible nonconstant accrual pattern within each center.

Following up on the original NHPP model, Zhang and Long extend their work into accrual modeling of multiregional trials [12], where the term "region" is used loosely to denote a center or a collection of centers that share certain common features. In the enrollment of multiregional trials, the overall and the region-specific underlying accrual rate usually vary over time and the regions may start enrollment at different times. To address these complicating issues, the authors propose to model accrual within each region using a NHPP process with a common prior distribution on the region-specific B-spline

Table 1. Comparison of Poisson process–based accrual models

	Multisites?	Type of Poisson process
Gajewski *et al*. [8]	No	HPP
Anisimov and Fedorov [9]	Yes	HPP within each site
Zhang and Long [11]	No	NHPP
Zhang and Long [12]	Yes	NHPP within each site

HPP: homogeneous Poisson process; NHPP: nonhomogeneous Poisson process.

coefficients. They show that the region-specific NHPP model further improves the precision of prediction compared to the original NHPP model that ignores the region-specific accrual information. Of note, there is usually a sudden increase in the underlying overall accrual rate when a new center starts recruitment, resulting in discontinuity in the underlying overall accrual rate. This potentially poses a problem to the original NHPP model, which assumes that the underlying overall accrual rate is, or can be approximated by, a continuous and smooth function over time. The region-specific NHPP model, however, properly addresses this issue. Table 1 provides a comparison of the Poisson process–based models that are presented in this section.

Modeling and prediction of event times

Clinical trials with survival endpoints usually require long follow-up duration, and lost-to-follow-up (LTFU) time is often present and complicates modeling and prediction of event times. When modeling event times, subjects can be divided into three groups. Specifically, subjects who have already experienced the event of interest, or have lost to follow-up, by the time of the monitoring have full information including event/LTFU time observed. No projection is needed for this group of subjects. Subjects who are still being followed at the time of the interim monitoring are at risk of the event of interest and only provide partial information (i.e., the subjects are free from the event and LTFU by the time of the monitoring). Consequently, the (future) events/LTFU time are unknown and need to be predicted for this second group of subjects. The projected survival time of such subjects is then the time elapsed from enrollment (or randomization, in randomized studies) to either the projected (future) event time or the projected (future) LTFU time, whichever is earlier, based on suitable event and LTFU models. Finally, if the enrollment is still ongoing at the time of the monitoring, the enrollment time of the future subjects are also unknown. Hence, accrual projection is needed in addition to the event/LTFU projection for this third group of subjects.

In what follows, we review two groups of approaches of event time modeling based on the accrual model used, that is, one uses HPP and the other one uses NHPP.

Models using HPP for accrual

Bagiella and Heitjan [13] propose a statistical model for prediction of landmark event times (e.g., when a certain number of events occur) assuming constant underlying enrollment rate and constant event rate over time (i.e., the event times follow an exponential distribution). The point estimate of the landmark time is obtained by plugging-in the maximum likelihood (ML) estimates of underlying accrual rate, event rate, and LTFU rate for each treatment arm. To quantify the variability in the aforementioned point estimate, conjugate prior distributions are introduced to each rate parameter so that Bayesian posterior predictive intervals can be obtained accordingly.

The aforementioned exponential event model proposed by Bagiella and Heitjan is attractive in its simplicity (one single parameter each for the event and LTFU model). However, it is well known that the assumption of constant hazard rate likely does not hold in real clinical trials. Subsequently, Ying and Heitjan [14] propose a more flexible parametric approach using Weibull distribution for the event/LTFU modeling. When the shape parameter of the Weibull distribution is 1, it becomes an exponential distribution. Hence, this model encompasses the exponential one mentioned earlier [13]. The choice of the Weibull distribution is also justified by the fact that it serves as a good approximation to a couple of other parametric distributions that are also used in survival analysis, for example, the Gamma distribution [15] and the lognormal distribution [16], the latter of which has longer tails. However, the two-parameter Weibull model does not have conjugate prior distributions in this setting, and hence numerical approximation or Monte Carlo simulation methods are required in the implementation. In this model, the future enrollment time is generated based on sampling of past arrival time intervals. One potential limitation is that the future intervals may follow different distributions than the past intervals. This is the case

when enrollment pattern changes with time, that is, nonstationary. As a remedy, it is possible to assign differential weights to the past intervals when drawing the future intervals, for example, assigning heavier weights to the more recent past intervals [3]. Of note, S-plus codes are available upon request to the authors of the original work [14].

Alternatively, Ying *et al*. [17] propose a flexible nonparametric model for the prediction of event times, which uses the Kaplan–Meier method to estimate the distribution of the event/LTFU processes. The point predictions directly come from the Kaplan–Meier estimator, and the interval predictions, as the authors suggest, come from Bayesian bootstrap samples [18]. This approach is generally robust since it is free of parametric assumptions, but it loses efficiency compared to the parametric approaches if the parametric assumptions are correctly specified.

In all the above-mentioned approaches, the treatment assignment information can be explicitly incorporated in the model. In the cases when treatment assignment must remain blinded, Donovan *et al*. [19] propose a parametric mixture model with a known mixture proportion (i.e., the randomization ratio). This approach can make a substantial difference compared to a simple model that considers all subjects as one group, especially when the monitoring occurs early in a trial and the number of events is few.

Models using NHPP for accrual

If the underlying enrollment rate deviates from a constant, as it usually occurs in real trials, the event time prediction based on a HPP accrual model is subject to bias. To address this, Zhang and Long [20] propose to combine the NHPP model for accrual with either parametric (exponential distribution) or nonparametric (Kaplan–Meier estimate) event/LTFU models. It is also possible to combine the model for multiregional accrual model [12] with suitable event/LTFU models.

In summary, prediction of event times requires proper joint modeling of subject accrual and event/LTFU processes. Either or both components can be tailored to allow for more robust or more efficient predictions.

Numerical results

In this section, we review the numerical results on comparisons between different models that have been reported in some of the cited articles. We first summarize the results for prediction of subject accrual and then for prediction of event times.

Prediction of accrual

In simulation studies, Zhang and Long [11] show that the original NHPP model outperforms the HPP model proposed by Gajewski *et al*. [8] with smaller root mean squared errors (rMSE) and biases, when the underlying accrual rate is not constant over time. Similar results are observed when hypothetical accrual prediction is conducted using accrual data from a real trial of adjuvant colorectal cancer treatments [1]. Figure 1 displays the observed accrual in this trial. Each dot represents the observed daily enrollment across 32 countries. Since accrual occurred during weekdays, weekends and holidays were excluded from this plot. The 95% confidence intervals (CI) of the HPP method do not cover the observed accrual duration in the scenarios investigated. Since the maximum underlying enrollment rate is considerably under-specified *a priori*, the 95% CI of the NHPP method at the early projection (i.e., with 30% enrollment) fails to include the observed accrual duration, but it clearly excludes the planned duration, and hence can still serve as a "red flag" of slow accrual to the sponsor. The projection, however, is closer to the observed accrual duration, when using a more diffused prior distribution on the B-spline coefficients (i.e., allowing for a larger prior variability of the underlying maximum accrual rate). That is to say, a more diffused prior distribution can mitigate the impact of a misspecified maximum underlying enrollment rate. Nevertheless, all estimates improve as more information accumulates.

In a recent working paper [12], the multiregional NHPP accrual model is compared with the original NHPP model that ignores the region-specific accrual information. The simulation studies show that the multiregional approach provides smaller rMSE and almost always tighter posterior CI (at least 88% of the times in the simulated scenarios investigated) compared to the original NHPP model. Accrual data from the same real trial [1] is reexamined. The results show that the multiregional NHPP approach improves the precision in predicting the accrual duration with tighter posterior CI, for example, the average width of the posterior CI of the multiregional approach ranges from 46% to 64% of the width of the original NHPP method.

In summary, the current numerical results suggest that the NHPP accrual model is more flexible than the HPP model as a result of better accommodation of the time-varying underlying accrual rate. Furthermore, the precision of accrual prediction can be improved by utilizing accrual information within each region in multiregional accrual models.

Prediction of event times

Ying *et al*. [17] compare their nonparametric event model with the exponential event model proposed

in the study by Bagiella and Heitjan [13] (both assume a HPP accrual model) in simulation studies and find that when the event data are generated from an exponential model, the exponential approach performs well with small bias and tight interquartile range at various time points when predicting the interim/final analysis times. However, when the event times are generated from a Gamma distribution (i.e., the parametric form of the event model is misspecified), the exponential approach can be severely biased, whereas the nonparametric approach still provides desired coverage rates in most cases, as long as the prediction does not occur very early in the trial with too few events. Ying *et al*. also compare the methods in a real trial for chronic granulomatous disease (CGD) [21] and find that the nonparametric approach almost always covers the landmark dates, despite wider CIs.

Ying and Heitjan [14] further compare the Weibull event model with the nonparametric approach [17]. When the parametric model is correctly specified in simulation studies, they find that the Weibull model provides CIs that are half the length of the CIs obtained using the nonparametric approach, despite satisfactory and comparable coverage probabilities. When the event data are generated from a Gamma model, the Weibull approach outperforms the nonparametric one with considerably shorter intervals on average, despite the misspecification of the parametric form. However, when the Weibull model is used on event data that are generated from a lognormal distribution, it provides lower coverage rates than the nonparametric approach, even though the results are more comparable in the prediction of the final analysis time. Similar to the simulation results, when the Weibull model is applied to the same CGD dataset [21], it leads to narrower CIs.

More recently, Zhang and Long [20] compare the exponential event model with the nonparametric approach when combined with a NHPP accrual model. The findings are largely consistent with those reported in Ref. [17] that the exponential event model is more prone to model misspecification but may gain efficiency/precision if the parametric model is correctly specified. Investigation of a real trial of first-line metastatic colon cancer treatment [22] also supported the conclusion. Whereas the underlying true hazard rate deviates substantially from the parametric assumption, the nonparametric method outperforms the parametric one.

In summary, the current numerical results suggest that, under the same accrual model, the parametric event models (e.g., exponential or Weibull) can be highly efficient when correctly specified but may be subject to severe bias when misspecified. This is especially true for the exponential model of event times, which is not as flexible as the Weibull model. In addition, as shown in Ref. [11], the performance of the HPP model for accrual can be poor when the assumption of constant underlying accrual rate is not met, and hence it follows that the models for event prediction that assume constant underlying accrual rates may perform poorly as well when such assumption is not met.

Discussion

The random fluctuation that is inherent in the accrual process calls for the use of stochastic approach in modeling subject accrual in clinical trials. In this article, we review a collection of research papers in this field. Should there be a strong reason to suggest a constant, or a close to constant, underlying enrollment rate over the entire accrual duration, the HPP method benefits from its model simplicity. Since the underlying accrual rate is not always constant over time in real trials, a more flexible NHPP model is generally more desirable. In addition, more precise prediction of multicenter trial accrual can be obtained using models with region/center-specific accrual rates, with or without modeling the underlying accrual rate at each region/center flexibly. This is because they better utilize all available accrual information and directly address the multiregion/center nature of such trials.

In the area of event time modeling, the choice between a flexible parametric function (e.g., a Weibull model) and a nonparametric model of the event time (e.g., the Kaplan–Meier method) depends on how confident investigators/sponsors are with the parametric assumptions. Nevertheless, the difference between the two generally diminishes as data accumulate. The benefits of joint modeling of accrual and event/LTFU time may be exploited to the full extent when it employs both a flexible accrual component and a flexible event/LTFU component. Of note, all the event models reviewed in this article, parametric or nonparametric, assume noninformative censoring, that is, the latent event and LTFU times are assumed independent. Since this assumption cannot be verified, it is possible to use the competing risk model as an alternative as in Refs. [23] and [24].

In general, the projection of accrual and event times is more precise with tighter CI when the prediction occurs late in a trial, but there is little room left to make any correction, even if there exist effective measures to address detected issues. On the other hand, an early projection is more likely to yield wider CI, especially when there lacks good prior information, which hence provides at most limited utility in practice. This is true regardless of the model used, and to address this issue, one can

argue in favor of more frequent monitoring of subject enrollment/event times.

It is also of interest on many occasions to model enrollment at the design stage, in which case the application of the models assuming constant underlying enrollment rate (either overall or within each region for multiregional models) is generally straight forward. It is, however, not always clear how the prediction should be applied using the NHPP model. We note that the prediction at the design stage based on the NHPP model is essentially the same as the HPP models when prior information on the shape of the underlying enrollment rate is not available. However, when proper prior information on the shape of the underlying enrollment rate (e.g., based on the enrollment pattern of similar trials in the past) exists, the NHPP model may yield a better prediction than simply ignoring such information in the prior predictive distribution.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgments

The authors would like to thank Dr Jonas Ellenberg and Dr Susan Ellenberg for their invitation to present at the University of Pennsylvania Annual Conference on Statistical Issues in Clinical Trials and contribute to this special issue.

References

1. **Van Cutsem E, Labianca R, Bodoky G,** *et al*. Randomized phase III trial comparing biweekly infusional 5-fluorouracil/leucovorin alone or with irinotecan in the adjuvant treatment of stage III colon cancer: PETACC-3. *J Clin Oncol* 2009; **27**: 2117–25.
2. **Barnes K.** Pharma giants risk reputation through clinical trial cost-cutting, Available at: http://www.outsourcing-pharma.com/Clinical-Development/Pharma-giants-risk-reputation-through-clinical-trial-cost-cutting (accessed 25 December 2010).
3. **Ying GS.** *Prediction of event times in randomized clinical trials*. Dissertations, ProQuest, Paper AAI3125925, 2004. Available at: http://repository.upenn.edu/dissertations/AAI3125925
4. **Ross SM.** *Stochastic Processes* (2nd edn). Wiley, New York, 1995.
5. **Lai D, Moye LA, Davis BR, Brown LE, Sacks FM.** Brownian motion and long-term clinical trial recruitment. *J Stat Plan Infer* 2001; **93**: 239–46. doi: 10.1016/S0378-3758(00)00203-2.
6. **Zhang Q, Lai D.** Fractional Brownian motion and long term clinical trial recruitment. *J Stat Plan Infer* 2010; **141**: 1783–88. doi: 10.1016/j.jspi.2010.11.028.
7. **Beran J.** *Statistics for Long-Memory Processes*. Chapman & Hall/CRC, 1994.
8. **Gajewski BJ, Simon SD, Carlson SE.** Predicting accrual in clinical trials with Bayesian posterior predictive distributions. *Stat Med* 2008; **27**: 2328–40. doi: 10.1002/sim.3128
9. **Anisimov VV, Fedorov VV.** Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. *Stat Med* 2007; **26**: 4958–75. doi: 10.1002/sim.2956.
10. **Cox DR.** Some statistical methods connected with series of events. *J Roy Stat Soc B* 1955; **17**: 129–64.
11. **Zhang X, Long Q.** Stochastic modeling and prediction for accrual in clinical trials. *Stat Med* 2010; **29**(6): 649–58. doi: 10.1002/sim.3847.
12. **Zhang X, Long Q.** *Bayesian modeling and prediction of patient accrual in multi-regional clinical trials*. 2012 Technical Report 12-01, Department of Biostatistics and Bioinformatics, Emory University.
13. **Bagiella E, Heitjan DF.** Predicting analysis times in randomized clinical trials. *Stat Med* 2001; **20**: 2055–63.
14. **Ying GS, Heitjan DF.** Weibull prediction of event times in clinical trials. *Pharm Stat* 2008; **7**: 107–20.
15. **Balasooriya U, Abeysinghe T.** Selecting between gamma and Weibull distribution—an approach based on predictions of order statistics. *J Appl Stat* 1994; **21**: 17–27.
16. **Carroll KJ.** On the use and utility of the Weibull model in analysis of survival data. *Control Clin Trials* 2003; **24**: 682–701.
17. **Ying GS, Heitjan DF, Chen TT.** Nonparametric prediction of event times in randomized clinical trials. *Clin Trials* 2004; **1**: 352–61.
18. **Rubin DB.** The Bayesian bootstrap. *Annals Stat* 1981; **9**: 130–34.
19. **Donovan JM, Elliott MR, Heitjan DF.** Predicting event times in clinical trials when treatment arm is masked. *J Biopharm Stat* 2006; **16**: 343–56.
20. **Zhang X, Long Q.** Joint monitoring and prediction of accrual and event times in clinical trials. *Biometrical Journal*, 2012. (in press).
21. **International Chronic Granulomatous Disease Cooperative Study Group.** A controlled trial of interferon gamma to prevent infection in chronic granulomatous disease. *N Engl J Med* 1991; **324**: 509–16.
22. **Saltz LB, Cox JV, Blanke C,** *et al*. Irinotecan plus fluorouracil and leucovorin for metastatic colorectal cancer. *N Engl J Med* 2000; **343**: 905–14.
23. **Beyersmann J, Latouche A, Buchholz A, Schumacher M.** Simulating competing risks data in survival analysis. *Stat Med* 2009; **28**: 956–71.
24. **Allignol A, Schumacher M, Wanner C, Drechsler C, Beyersmann J.** Understanding competing risks: a simulation point of view. *BMC Med Res Methodol* 2011; **11**: 86.