Contemporary Clinical Trials 46 (2016) 7–10

Contents lists available at ScienceDirect

Contemporary Clinical Trials

journal homepage: www.elsevier.com/locate/conclintrial

Letter to the Editor

Discussion on the paper "Real-Time Prediction of Clinical Trial Enrollment and Event Counts: A Review", by DF Heitjan, Z Ge, and GS Ying

Keywords:
Patient enrollment
Poisson process
Mixed models
Event count
Prediction

The paper by Heitjan et al. [11] provides a very interesting and useful review of the methods for predicting patient enrollment and event counts in clinical trials. The aim of this letter is to raise an additional discussion on some points and to provide readers with more comprehensive information and clarification of particular methods/techniques.

First, it would be useful to specify that there are two basic stages in predicting patient enrollment and various events:

1. Start-up (baseline) prediction before trials starts and therefore there is no real trial data available yet, and
2. Interim prediction where it is possible to use real trial data and update (re-project) trial behaviour for the remaining period.

At both stages, good predictive techniques potentially can use similar models, only input parameters will be evaluated differently.

Since the trial start-up stage is not reflected in detail in [11], it seems expedient to devote some time to this.

1. Trial start-up stage

This stage may also include an early stage of the trial where not many centres are initiated and not many patients have been recruited yet. Typically during this stage, the basic input information that is provided by clinical teams for enrollment predicting includes the following key elements:

(a) total number of randomized patients (sample size); (b) expected number of screened patients and screening duration; (c) list of regions and countries to be involved into the study: (d) planned number of centres to be initiated in each country and some information about the expected schedule of initiation; (e) expected enrollment rates in centres or countries (this may include screening/enrollment rates and dropout probabilities).

This information has many uncertainties. In particular, at the start-up stage we may not know the exact schedule of center's initiation and we especially cannot predict the exact screening/enrollment rates and dropout probabilities.

Therefore, one of the main problems at this stage is how to account for these uncertainties and evaluate trial enrollment feasibility. There is no universal approach since the solution may depend on data availability. If we have a similar historical study (similar therapeutic indication, inclusion/exclusion criteria, etc.) conducted in the same regions, then this information can be used to create the initial trial enrollment design. Specifically, the enrollment rates for a new trial can be treated as random variables with some prior distribution where parameters can be evaluated using historical data in these regions and some prior information. As the rates are positive, it is natural to use a gamma distribution.

For a new trial, we can also assume that the centres in the regions can be initiated in time according to some distributions where parameters are estimated using historical data. As usually teams for each country/region may provide some time intervals where a given number of centres is planned to be initiated, then at the first instance we can assume that the times of initiation are distributed uniformly in these intervals [3,4]. If some historical information about initiation dates is available, other types of distributions can be also used.

Note that during the start-up stage there can be a rather long transient period until most of the centres will be initiated. Thus, the total number of patients and centres may not be too large. Therefore, during this period it is important to account for the process of centres initiation and the methods based on modelling enrollment in the individual centres are more preferable compared to models based on global prediction.

1.1. Poisson-gamma enrollment model

On this way we are naturally coming to using a so-called Poisson-gamma enrollment model (P-G model) developed in [1–5]. This model assumes that the patients arrive at clinical centres according to delayed doubly stochastic Poisson processes where the variation in rates between different centres is modelled using a gamma distribution. The delays in center's initiation also can be random.

This model is very flexible as it provides the opportunity to model the enrollment on different levels (center, country, region, trial) and has many additional features, e.g. predicting with credibility bounds, predicting probability to complete in time, evaluate effects of changing the number of centres, etc. One of the additional advantages of P-G model is that most of these characteristics can be calculated using closed-form expressions, thus, there is no need to use Monte Carlo simulation.

Note that Carter et al. [10] also modelled variation in rates of corresponding Poisson processes but using a uniform distribution. However, this approach has some limitations as it assumes that the rates are bounded in some interval. Moreover, the analysis of many real trials shows that the empirical distributions of the rates are rather far from uniform distribution and heavy tailed.

In the framework of P-G model, at the start-up stage as input data it should be provided the expected means and standard deviations of the enrollment rates (on center or country level) to estimate the prior parameters of the rates used in prediction. These values can be evaluated using historical data from similar trials and information provided by clinical teams.

If there is no information from similar trials, then we can use the planned/expected rates provided by clinical teams weighted with some expert estimators. This data can be used as sample statistics for

http://dx.doi.org/10.1016/j.cct.2015.11.008
1551-7144/

evaluating the prior parameters of P-G model (on country or regional level). Some discussion on using baseline estimates of rates at the trial start-up was provided in [3,5]. Bakhshi et al. [9] investigated P-G model further and suggested the empirical way to set the prior parameters by using the results of the meta-analysis.

As a separate set of input data for P-G model, the information about the process of the centre's initiation should be provided. The case where the times of initiation have uniform distribution was considered in [3,4]. In this case, the closed-form expressions for predictive characteristics were derived.

Note that at start-up and early stages other approaches based on models for global enrollment, e.g. using Poisson models with global gamma distributed rate [13,17], and Brownian (Lai et al. [14]) or fractional Brownian (Zhang & Lai [18]) motions may not be appropriate as in general at these stages there is a small number of active centres and patients recruited.

Therefore, on my opinion, P-G model is rather flexible and can be applied to the vast majority of trials at start-up and early stages.

2. Interim stage

At this stage, it is natural to use real data and re-estimate parameters of the model with the purpose to adjust to real data and improve accuracy of prediction of the remaining enrollment. Thus, it is typically assumed that there is already some number of active centres that enrolled a reasonable number of patients (enough to use statistical estimations). Therefore, the methods and results may depend on trial goals and data availability.

There can be other tasks at the interim stage including evaluating enrollment performance and other operational characteristics, detecting outliers, etc. However, this interesting direction may lead us outside the current discussion.

Most papers by other authors are mainly dealing with prediction of global enrollment and there are two basic directions. One is using mixed Poisson processes where the global rate is modelled using different approaches [12,13,16,17]. Another one uses Brownian or fractional Brownian motions [14,18].

A brief review of the papers related to these directions is provided by the authors [11] in Sec. 3 “Predicting Accrual”. However, I would argue with the classification of the models (or two streams) proposed in Sec. 3.1. It seems rather artificial as actually the first stream should also involve modelling of enrolment. The second stream potentially can use modelling for predicting future trends and therefore time to reach targets, as well. In addition, the description of the papers related to using random-effect models in Sec. 3.3 is done rather schematically. As the use of P-G model is receiving further attention and development in papers of different authors, it seems expedient to provide more details here.

2.1. Use of a Poisson-gamma enrollment model

In the framework of P-G model [1–5], the enrollment processes at different levels are modelled as non-homogeneous Poisson processes with time-dependent and in general random rates, which are governed by the processes of centre's opening and closing as well as individual center's data. Together with modelling enrollment at the start-up stage, P-G model can be efficiently applied to an interim prediction. The input is the enrollment data (for each centre, the duration of active enrollment and the number of patients recruited). Using this data, the parameters of a gamma distribution of the enrollment rates are estimated using ML procedure (on global or regional level). Then in each centre the posterior rate is re-estimated using individual data and the Bayesian procedure. The posterior rates also have gamma distributions with different parameters depending on interim data due to the property of conjugate distributions (Poisson and gamma). These rates can be used to create the predictions of the remaining enrollment on different levels and evaluate other characteristics.

The technique based on using P-G model has several advantages compared to other approaches: it accounts for multiple centre's effects, different times for opening and closing centres, allows predicting in a closed form the mean number of recruited patients with credibility bounds (on different levels), predicting credibility bounds for time to complete enrollment and probability to complete in time. One of the essential features is the opportunity to evaluate the interim adaptive adjustment (if enrollment is going slower as expected, evaluate the number of new centres needed to be added with the purpose to complete enrollment in time with a given confidence). In addition, this technique has several other features that are available only in this framework, e.g., predicting centre/country performance, number of “empty” centres, creating optimal enrollment design [1–5].

It also seems expedient to raise some discussion on using formulae compared to Monte Carlo simulation. As for rather general scenarios the most of characteristics can be calculated using closed-form expressions (explicit formulae), then there is no need to use Monte Carlo simulation. The availability of formulae has advantages as it allows us to investigate the functional dependence on different parameters (number of sites, vector of rates, centre's delays, etc.) and, thus, analyse in real time the impact of various factors, perform sensitivity analysis and find the optimal solutions, which would be hard to archive using simulation. In addition, simulation may not work well for evaluating small tail and risk probabilities, P-values and also may lead to large errors in small regions.

Note also that in some cases of special restrictions on enrollment and more complicated assumptions, it may be difficult to derive formulae. In these cases Monte Carlo simulation can be the natural choice.

I would also like to correct the author's statement in Sec. 3.3 [11] that “Mijoule et al. [15] proposed replacing the gamma with a Pareto mixture.” Actually in [15] the authors investigated further properties of P-G model, compared them with the Pareto-Poisson model using real datasets from [2], and also investigated the feasibility of the model. In final conclusions the authors “recommend the use of the Poisson-gamma, which is easier to handle”, and also recommend using a uniform distribution for centres initiation when the opening dates of the centres are not known precisely, which is proposed in [3,4].

It would also be interesting to provide some parallel between Williford et al. [17], Gajewski et al. [13] and P-G model. In both papers the authors use a Poisson process with gamma distributed rate to model the global enrollment and also a Bayesian interim adjustment. Note that in the framework of P-G model the global enrollment in general is not described by a Poisson-gamma process as the sum of gamma distributed variables in general does not follow a gamma distribution. Nevertheless, this sum for a large number of summands can be well approximated by a gamma distributed variable [5]. Thus, P-G model can also serve as a justification of model [13,17] on a global level.

It is also worth noting that in the framework of P-G model, the interim prediction can account for the opportunity in the future to open or close some centres [5]. Thus, the predictive processes at different levels are in general non-homogeneous doubly stochastic Poisson processes where the global and individual rates depend on the processes of initiation, closing centres and individual rates. This feature of P-G model profitably differentiates it from the other models for global prediction based on using Poisson models [12,13,17] and Brownian and fractional Brownian motions [14,18], as in these papers it is assumed that the predictive process is time-homogeneous with constant parameters estimated at interim time.

2.2. Modelling trends

Here I would like to add some discussion to Sec. 3.4 [11] "Modelling trends in the Poisson rate”. Actually the author's statement “The models described thus far all assume a constant mean enrollment rate per