Predicting Accrual in Clinical Trials: Bayesian Posterior Predictive Distribution

Stephen D. Simon and Byron Gajewski
Children's Mercy Hospital and University of Kansas Medical Center

Introduction:

The most common reason **why clinical trials fail** is that they fall well below their goals for patient accrual. Accrual rates are often developed in an ad hoc fashion. If they are monitored at all, accrual rates are examined using a purely subjective approach. There are not any good quantitative tools for planning and monitoring accrual rates.

With good planning tools, researchers would be able to construct **realistic targets** for their sample sizes rather than promising a sample size that could not be delivered in a reasonable time frame and within a limited research budget. With good monitoring tools, researchers would get an **early warning when accrual rates are suffering**. This would allow them to take appropriate corrective action before too much harm was done.

Notation:

T	Planned trial duration
n	Planned number of patients
P	Prior certainty about accrual
t_0	Start of the trial
t_i	Observed arrival times
m	Current number of patients
w_i	Observed waiting times
W_i	Unobserved future waiting times
Exp	Exponential distribution
θ	Parameter of Exp()
IG	Inverse Gamma distribution
k, V	Parameters of IG()

Suppose that the trial director wishes to assess the accrual process after m patients have been recruited. Let t_0 represent the time that the study started and let t_1, t_2, …, t_m represent the times that each new patient enters the trial. Without loss of generality, we can assume that the study starts at time t_0=0. Compute **waiting times**

$$w_i = t_i - t_{i-1}$$

We will assume that

$$w \mid \theta \sim Exp(\theta)$$

> "Globally, more than 80 per cent of clinical trials fail to enrol on time, and this recruitment problem is extremely costly for drug companies, contributing to 85-95 per cent of the lost days in a clinical trial." Barnes (2006).

and specify an **inverse gamma as the prior distribution for theta**. Eliciting a prior distribution is a difficult task. We start our elicitation by asking two simple questions (see below). Let T represent the answer to the first question. Divide the answer to the second question by 10 to get P. This produces the following prior distribution

$$\theta \sim IG(k = nP, V = TP)$$

You can then present various properties of this prior distribution to the researcher to assess how realistic this prior distribution is.

The posterior distribution is

$$\theta \mid w \sim IG(nP + m, TP + t_m)$$

which has posterior mean

$$E[\theta \mid \mathbf{w}] = \frac{TP + t_m}{nP + m - 1} \approx \left(\frac{nP}{nP+m}\right)\frac{T}{n} + \left(\frac{m}{nP+m}\right)\frac{t_m}{m}$$

> Questions to elicit a prior distribution:
>
> (1) How long will it take to accrue n subjects? (This is T).
>
> (2) On a scale of 1-10, how confident are you in your answer to (1)? (Divide by 10 to get P).

This is approximately a weighted average of the prior mean and the mean waiting time for the observed data. The relative weights depend on the prior sample size (nP) and the sample size of patients observed so far (m).

Example:

To illustrate the proposed method, consider an unnamed current phase III clinical trial (randomized, double-blind, and placebo-controlled) used to examine the efficacy and safety of a dietary supplement. This study was planned and accrual started prior to our development of these methods, but still serves to illustrate how this approach would work. **The current protocol requires *n*=350 subjects,** with balanced randomization to either treatment or placebo control. In the previous study, investigators were able to recruit, from a similar population, 350 subjects across 3 years.

At the design phase of the study it was felt that the previous clinical trial offered strong prior information. Setting ***P*=0.5** results in k=175 and V=1.5 years. This corresponds to a prior mean of $V/(k\text{-}1)$=0.0086 years or **3.1 days** for the average waiting time.

After the study was funded and the protocol approved, the investigative team began recruiting subjects. After **239 days** the project director compiled a report that displays enrolled dates of **41 subjects**. This represents an average waiting time of **5.8 days**, much longer than expected.

We are interested in **predicting the next *n-m* waiting times** $W_{m+1},\ldots,W_n$. A simple and easily generalizable approach for each W is to randomly select θ_1 from the posterior distribution and then randomly select waiting time n-m random variables from $W_{m+1,1}$,…,$W_{n,1}$ from an exponential distribution with parameter θ_1. Repeat this process for θ_2, $\theta_3,\ldots,\theta_b$,, where b is a large number (typically 1,000). The sum of observed and simulated waiting times, $S_b(n) = w_1 + w_2 + \ldots + w_m + W_{m+1,b} + \ldots + W_{n,b}$ will represent b estimates of the total duration of the clinical trial. The results from this simulation are shown in Table 1 and Figure 1.

> We calculated three predictive distributions:
>
> A) Informative prior at the start of the study (m=0, P=0.5);
>
> B) non-informative prior after 39 patients (m=39, P=0); and
>
> C) informative prior after 39 patients (m=39, P=0.5).

We can also use this process to obtain a **posterior predictive sample size**. Let T represent the time point at which the study must end. Compute partial sums $S_b(m+1)$, $S_b(m+2)$, . . until the partial sum exceeds T. The values n_b which represent the largest values where the partial sums do not exceed T, provide a predictive distribution of sample sizes. See Figure 2 for an example.

This simulation approach also allows you to examine **more complex accrual patterns**, such an accrual goal of recruiting until 50 patients have volunteered, or until 6 months have elapsed, whichever comes first.

Future work:

The model we propose is easily extended in a variety of ways:

1. Use of alternatives to the exponential distribution for modeling waiting times.
2. Examination of alternative prior distributions.
3. Use hierarchical models to predict accrual across multiple centers in a multi-center trial.

> Table 1. Median predicted trial duration and 95% Credible Interval (yrs)
>
> (A) 3.0 (2.5,3.7)
>
> (B) 5.6 (4.3,7.6)
>
> (C) 3.7 (3.3,4.4)

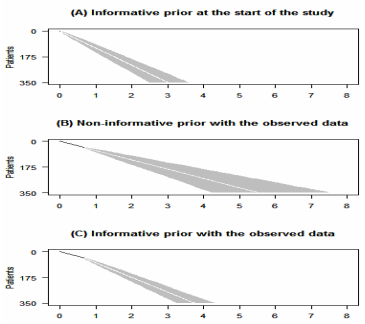

Figure 1. Estimated study completion times (in years) for a study that accrues 300 patients. The top figure represents the Bayesian predictive distribution using the prior distribution. The middle figure represents the predictive distribution using the observed data and a non-informative prior distribution. It represents a simple extrapolation of the existing accrual pattern in the data. The bottom figure represents the predictive distribution using the observed data and an informative prior distribution. Notice that this represents a compromise between prior distribution and the data.

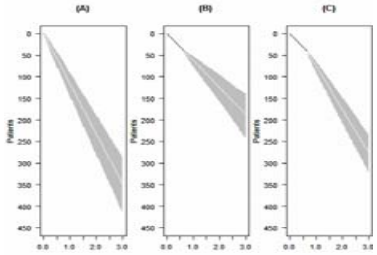

Figure 2. Estimated sample sizes for a study that ends at exactly 3 years.

Conclusion:

Predicting accrual across a fixed time period of a planned study is not an easy procedure. As demonstrated in this paper there are uncertainties that should be incorporated into this prediction. We hope that the method in this paper encourages investigators to account for these uncertainties when planning and monitoring accrual in a clinical trial.