Predicting patient recruitment in multicenter trials in JMP

Xiaotong Jiang[1] and Richard C. Zink, PhD[1,2]

[1] Department of Biostatistics

University of North Carolina at Chapel Hill

Chapel Hill, North Carolina, USA

[2] JMP Life Sciences

SAS Institute Inc.

Cary, North Carolina, USA

Introduction

Clinical trials are experiments that study whether a new drug, device, or therapy works and is safe for people. Different clinical trial phases have different sample size requirements, particularly if we want to ensure that the trial is adequately powered for the primary endpoint. The need ranges from tens of patients in the short run to thousands of patients and more for long term safety trials. Relying heavily on patients, clinical trials can be conducted only after enough patient information has been collected and treatment arms randomly assigned (if necessary). Thus, patient recruitment is a very crucial part of the initial design of any clinical trials.

Recruiting patients is a long and complex process with many roles involved, especially for multicenter trials. It requires active collaboration and timely communication among patients, physicians, researchers, lab workers, sponsors, and administrative staff. However, clinical trials often run under tight timelines and limited budget. According to the National Institutes of Health (NIH) and Food and Drug Administration (FDA), at least 80% of clinical trials in the US fail to meet their patient recruitment goals [1][2]. Such failures to recruit patients on time can cause additional costs, shortage in resources, and significant delay in trials. Moreover, it is more likely to see a high type II error when there are not enough observations, which would lead to false results and misleading conclusions.

There are many reasons why it is hard to enroll and retain eligible patients, and many strategies have been developed to conquer this recruitment challenge. Power/sample-size analysis is usually conducted in the proposal; a detailed recruitment strategy should be developed and tailored specifically to the research question (Thoma A, 2010); and a series of considerations need to be taken into account by physicians, researchers, and staff when recruitment starts. However, it is often unclear how far the current recruitment is in the progress as time goes. For example, suppose the goal is to recruit 1000 people from July 10 to September 12, but only 800 people have been enrolled by the end of August. Is it possible to recruit 200 more people in 12 days? Is the team going to miss the deadline? If yes, should more centers be added to increase

recruitment rates? It is difficult to know because recruitment rate is not constant but varies by center and by time. This kind of uncertainty makes it complicated to plan ahead and take measures to adjust for the recruitment rates. It seems that an interactive and flexible tool is in need to predict the remaining enrollment so investigators could follow the state of recruitment and react to unexpectedly low or high recruits without delay.

In this paper, we implement a patient recruitment model proposed by Anisimov and Fedorov (2007) in JMP Clinical 6.0 to assess and report the current condition of an ongoing patient recruitment, using the recruitment information collected since the first patient. After users enter the recruitment specifications and other additional parameters, the predictive model will generate a series plot summarizing the finished recruitment and predict future recruitment with a 95% confidence interval. Together with other graphs and tables, such visualization draws a clear picture of the entire recruitment process for investigators to consult with and make appropriate decisions.

Methodology

Lots of predictive models have been developed by statisticians to solve the recruitment prediction problem (Barnard et al, 2010). A commonly used method is the Poisson-Gamma distribution (Anisimov and Fedorov, 2007).

Let T_1 be the current interim time, K_1 be the number of patients enrolled up to T_1, T be the enrollment target time, n be the total number of patients needed, and N be the total number of centers. These values are known in advance. Suppose that up to time T_1 there have been k patients enrolled in τ_i days for center i, where $i = 1,2,3, \ldots N$. We assume that

- The number of patients recruited by center i up to time T_1 follows a Poisson distribution with recruitment rate λ_i;
- The rate λ_i is a sample from a Gamma distribution (α, β).

This means that patients are enrolled in a clinical trial according to independent Poisson processes, each of which is defined by an unknown, nonconstant parameterλ_i. λ_i is the recruitment rate (the number of patients recruited for each unit of time period) for center i, a measure that indicates how fast the recruiting process is, and belongs to a gamma distribution with unknown parameters, α and β.

The two gamma parameters, α and β, could be estimated using the enrollment data that researchers have in hand so far. Using the properties of the Poisson and Gamma distributions, the probability that each center has recruited k_i patients follows a negative binomial distribution with parameters (p, r) where $p = \frac{\alpha}{\alpha+m\tau_i}$ and $r = \alpha$:

$$p(k_i = k) = p(k, \tau_i, \alpha, m) = \frac{\Gamma(k+\alpha)}{k!\Gamma(\alpha)} \left(\frac{m\tau_i}{\alpha+m\tau_i}\right)^k \left(\frac{\alpha}{\alpha+m\tau_i}\right)^\alpha, k = 0, 1, 2, \ldots \text{where } m = \frac{\alpha}{\beta} \qquad (1)$$

Parameters α, β, and m are estimated by maximum likelihood method where the log-likelihood functions is the sum of the logarithm of equation (1) over all centers, given the current enrollment data points (k_i, τ_i) for $i = 1, 2, \cdots, N$.

Patient Recruitment

We now focus on the remaining time period and predict how much more time is needed to finish recruiting the rest of the patients. Note that $K_2 = n - K_1$ is the number of subjects left to enroll and $T_2 = T - T_1$ is the remaining recruitment time. Using the estimated α and β and following the gamma distribution, researchers could then simulate different values of recruitment rate (a.k.a. λ_i) and thus different values of remaining times that are derived from the λ_i's .

A Bayesian approach is applied when there are 20 or more centers, and maximum likelihood estimation is used when there are fewer than 20 centers. By default, the number of simulations is 10,000. The predicted remaining time is represented as

- $\tilde{T}_1 = \frac{\text{Gamma}(K_2,1)}{\sum_i \text{Gamma}(\alpha+k_i,\beta+\tau_i)}$ for $N \geq 20$
  For each simulation, a random value are generated from Gamma$(\alpha + k_i, \beta + \tau_i)$, the posterior distribution of λ_i, and summed over all centers to form the denominator. For the numerator Gamma$(K_2, 1)$, the value of K_2 is chosen to be $K_1 + 1, K_1 + 11, \cdots, n$ (the default interval is 10 but users could specify other values), in which n should always be included. Ratios are taken to form $\widetilde{T_1}$ for each value of K_2. There are 10,000 simulations generated for each value of K_2. The mean for each 10,000 values is presented as the predicted mean, and the 2.5 and 97.5 percentiles are presented to provide a 95% confidence interval for each value of K_2.

- $\tilde{T}_2 = \frac{\text{Gamma}(K_2,1)}{\sum_i \frac{k_i}{\tau_i}}$ for N < 20
  For each simulation, the MLE of the rate λ_i is $\frac{k_i}{\tau_i}$ which is then summed over all centers to form the denominator. The numerator Gamma$(K_2, 1)$ is computed where the value of K_2 is chosen to be $K_2 + 1, K_2 + 11, \cdots, n$ (the default interval is 10 but users could specify other values), in which n should always be included. Ratios are taken to form $\widetilde{T_2}$ for each value of K_2. 10,000 simulations are generated for each value of K_2. The mean for the 10,000 values is presented as the predicted mean, and the 2.5 and 97.5 percentiles are presented to provide a 95% confidence interval for each value of K_2 considered.

The day values (mean, 2.5 percentile, and 97.5 percentile) are rounded up and added to T_1 so they represent actual dates. These simulated values will be used to predict the additional number of sites in the following subsection, adaptive adjustment.

Adaptive Adjustment

Given a user-defined probability, p_1 (0.85 by default), conclusions can be drawn with respect to whether the enrollment target will be reached by the target time. Ideally, we want $100\times p_1\%$ or greater than $100\times p_1\%$ of the simulations ($\tilde{T}_1, \tilde{T}_2$ from previous part) to be less than or equal to the target time. This means $100\times(1 - p_1)\%$ or less of the simulated predicted remaining times are greater than the actual remaining time, which also means that there is a $100\times(1 - p_1)\%$ (or

less) chance that the recruitment deadline will not be met (see Figure 1 as an example). If p_1 is not met (in other words, if we cannot meet the deadline with a high probability), the program initiates adaptive adjustment (see Figure 2 as an example). Adaptive adjustment calculates the number of additional centers needed for the recruiters to complete the enrollment before the deadline (with a high probability p_2 introduced and defined in a few paragraphs).

Assume new centers cannot start recruiting right away and all new centers have the same delay. Define d (30 by default, user-specified) to be the delay (in days) for new sites to come online. Let M be the number of new centers to be added, and assume the parameters α and m are unknown. The remaining time is estimated as

$$\tilde{T}(M) = d + \frac{Gamma(K_3,1)}{\widetilde{\Lambda}_1 + Gamma(\alpha M,1)m/\alpha},$$

where $K_3 = K_2 - \Pi_{\widetilde{\Lambda}}(d)$ is the number of patients left to recruit after time point $T_1 + d$. Note that the value of $\widetilde{\Lambda}_1$ depends on the number of sites before the addition. As it is explained above, $\widetilde{\Lambda}_1 = \sum_i Gamma(\alpha + k_i, \beta + \tau_i)$ when there are 20 or more centers and $\widetilde{\Lambda}_1 = \sum_i k_i/\tau_i$ when there are fewer than 20 centers.

To find K_3, we first need to find $\Pi_{\widetilde{\Lambda}}(d)$, the number of patients recruited during the delay, between T_1 and $T_1 + d$. Because the recruiting time τ_i is d for all sites between T_1 and $T_1 + d$, $\Pi_{\widetilde{\Lambda}}(d)$ is simulated from a negative binomial distribution with parameters (p, r) where $p = \frac{\hat{\alpha}}{\hat{\alpha}+\hat{m}d}$ and $r = N\hat{\alpha}$:

$$p(k_i = k) = p(k, d, \hat{\alpha}, \hat{m}) = \frac{\Gamma(k+N\hat{\alpha})}{k!\Gamma(N\hat{\alpha})}\left(\frac{\hat{m}d}{\hat{\alpha}+\hat{m}d}\right)^k \left(\frac{\hat{\alpha}}{\hat{\alpha}+\hat{m}d}\right)^{N\hat{\alpha}}, k = 1, 2, \ldots \text{ where } m = \frac{\alpha}{\beta}.$$

There are 10,000 (default) random negative binomial values $\Pi_{\widetilde{\Lambda}}(d)$ generated, and thus 10,000 K_3 values are generated. For each value of K_3, we simulate from the Gamma distribution to form the numerator of $\tilde{T}(M)$. The denominator is composed of two parts, the first of which is simulated the same way as the denominator of $\tilde{T}_1$ (or $\tilde{T}_2$, depending on the number of centers) is and the second of which is 10,000 simulated random values Gamma$(\alpha M, \beta)$, the total recruitment rate with rates of the M new centers included. Ratios are then taken and added with the delayed time d. The ceilings of the predicted dates $\tilde{T}(M)$ are taken and then added to T_1 so they represent the actual dates.

We have added M new centers, but how do we know that we have had enough centers? The time limit T will be met if the probability of the 10,000 expected dates of completion that are earlier than the target date (see Figure 3 as an example) is greater than a user-defined value, p_2 (0.90 by default). If there is not a high probability that the deadline will be met, M will be incremented by 1 and the same calculation will be repeated until p_2 is reached. The initial value of M is 1 by default. The mean of the simulated dates of completion is computed and so are the 2.5 and 97.5 percentiles for each value of M. The date values are rounded up (see Figure 4 as an example). The entire calculation process is recorded in a table in JMPC 6.0 (see Table 1 as an example).

Applications

We use the Nicardipine data, which is usually registered during the installation of JMPC, as an example of the application of the patient recruitment feature. First, users click on the button **Patient Recruitment** when building a review to get to the following dialog (Figure 1). The **General** outline is where users specify the target date, target enrollment, and start date. JMPC automatically uses the last randomization rate as current date but another date could be selected to be the current date. Under the **Option** outline, users are able to specify other values such as the number of simulations, M the initial number of new centers, d the number of days delayed at new centers, p_1 the probability that determines whether there is a need to add additional centers, p_2 the probability that determines when to stop adding new centers, etc.

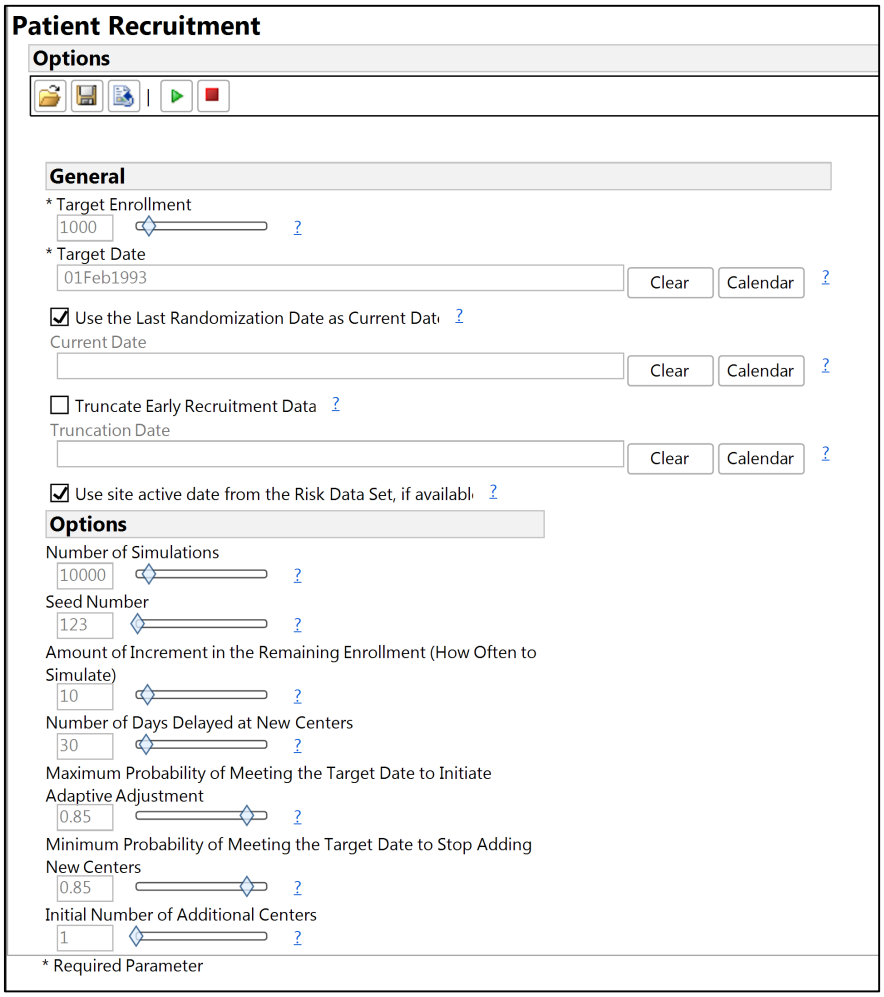

Figure 1. Patient recruitment dialog

As it is mentioned in the beginning of the subsection *Adaptive Adjustment*, below are examples of one case where the deadline is not missed (Figure 2) given p_1 and another case where the deadline is missed (Figure 3) given p_1. In both figures, we can see the fixed values (in blue)

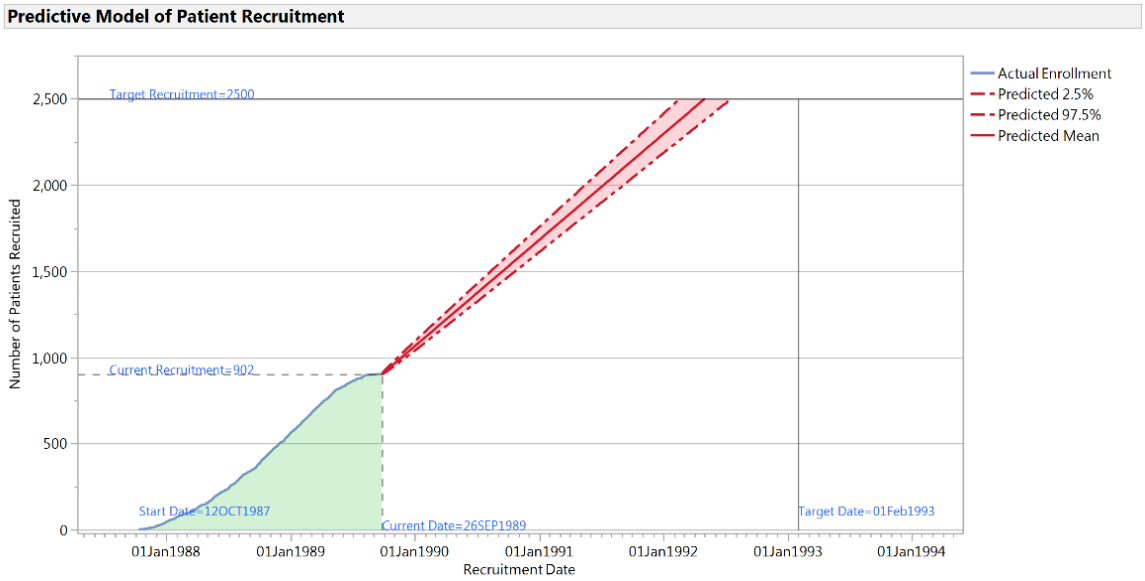

Figure 2. An example of predictive model of patient recruitment where the deadline is not missed

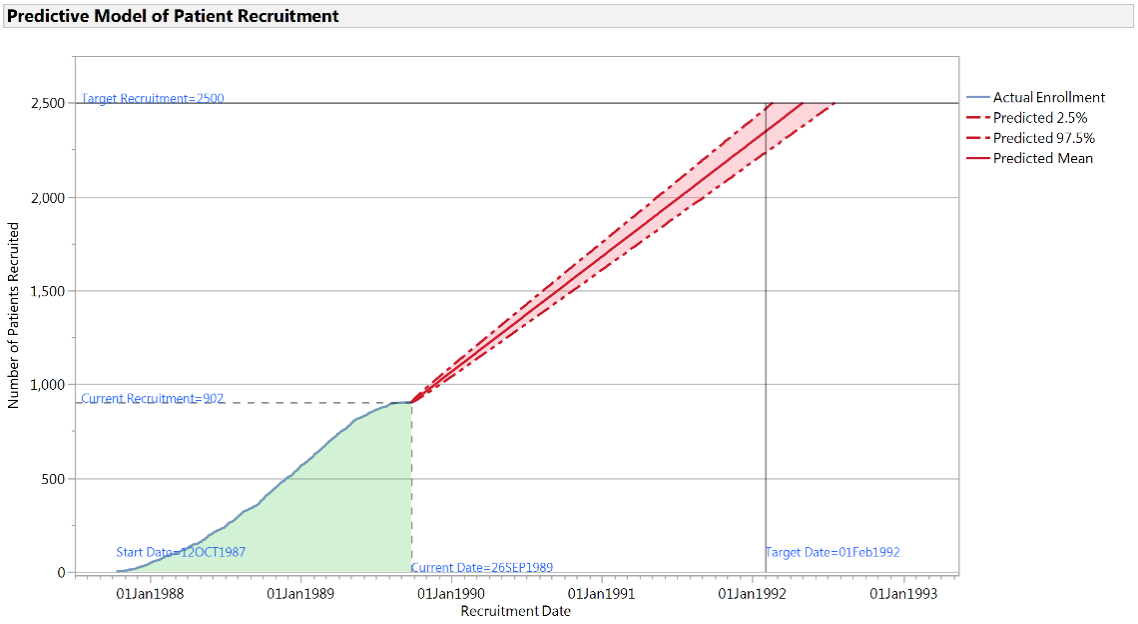

Figure 3. An example of predictive model of patient recruitment where the deadline is missed

specified in the dialog: target recruitment ($n = 2500$), current recruitment ($K_1 = 902$), start date, current date $T_1 = 26\text{Sep}1989$ and target date $T = 01\text{Feb}1993$. The blue curve is the actual recruitment, and the red curves are the mean, the 2.5th percentile, and the 97.5th percentile of the simulations, served as the prediction band of the remaining recruitment. When the target date is 01Feb1993, both mean and the 95% confidence intervals are on the left side of the deadline, indicating that recruitment will meet the deadline for at least 97.5 of the time. When the target date is selected to be a year earlier, both mean and 95% confidence interval are on the right side of the deadline, indicating that recruitment will miss the deadline for at least 97.5 of the time.

Hence the deadline can no longer be met if no additional centers are added. Here, it is easy to decide because the red band does not contain the target date. Whether the target time will be missed or not can sometimes be not so apparent if we rely on the graphs only, but it can be determined by computing the underlying probability and comparing it to p_1.

Since the case in Figure 3 is predicted to fail the deadline with a higher probability than p_1, adaptive adjustment is initiated. By default, JMPC first starts with adding one center. Table 1 presents the entire adaptive adjustment process and calculation. We see that adding one center is not enough because the probability of meeting the target date is still 0.00%. JMPC then continues adding more centers one at a time until p_2 is met. As it is explained in *Adaptive Adjustment*, p_2 is the cutoff where we stop adding more centers and are satisfied with the current probability of meeting the deadline with new centers added. In this case, the default value of p_2 is 85%, so adaptive adjustment stops at 9. It means that 9 new centers are recommended to be added for the clinical trial to reach target recruitment before the deadline with a probability of at least p_2 (currently, there are 40 centers in the trial). For this example, the team may need to change their expectations as to when the trial will complete enrollment.

Table 1. Adaptive adjustment with initiation probability p_1 and termination probability p_2

Number of new centers	Probability of meeting the target date with the additional centers added	Average predicted date of enrollment completion	2.5th percentile of predicted date of enrollment completion	97.5th percentile of predicted date of enrollment completion	Target Date
1	0.77%	06May1992	19Feb1992	25Jul1992	01Feb1992
2	3.61%	14Apr1992	26Jan1992	04Jul1992	01Feb1992
3	11.10%	24Mar1992	05Jan1992	14Jun1992	01Feb1992
4	22.68%	04Mar1992	15Dec1991	26May1992	01Feb1992
5	39.52%	14Feb1992	26Nov1991	05May1992	⋮
6	56.42%	27Jan1992	09Nov1991	17Apr1992	01Feb1992
7	72.17%	09Jan1992	22Oct1991	30Mar1992	01Feb1992
8	83.97%	24Dec1991	06Oct1991	13Mar1992	01Feb1992
9	92.22%	07Dec1991	20Sep1991	24Feb1992	01Feb1992

Figure 4 shows how fast the probability of meeting the deadline (the 1st column in Table 1) increases as more new centers are added. The more centers are added, the faster the increase is. It is also worth mentioning that the curve is flat for the first two or three centers. This tells us that adding a few new centers is not enough for a really close deadline. Similar patterns and similar conclusions can be drawn from Figure 5. Figure 5 illustrates how quickly the predicted date of completion converges to the target date as more new centers are added (from right to left).

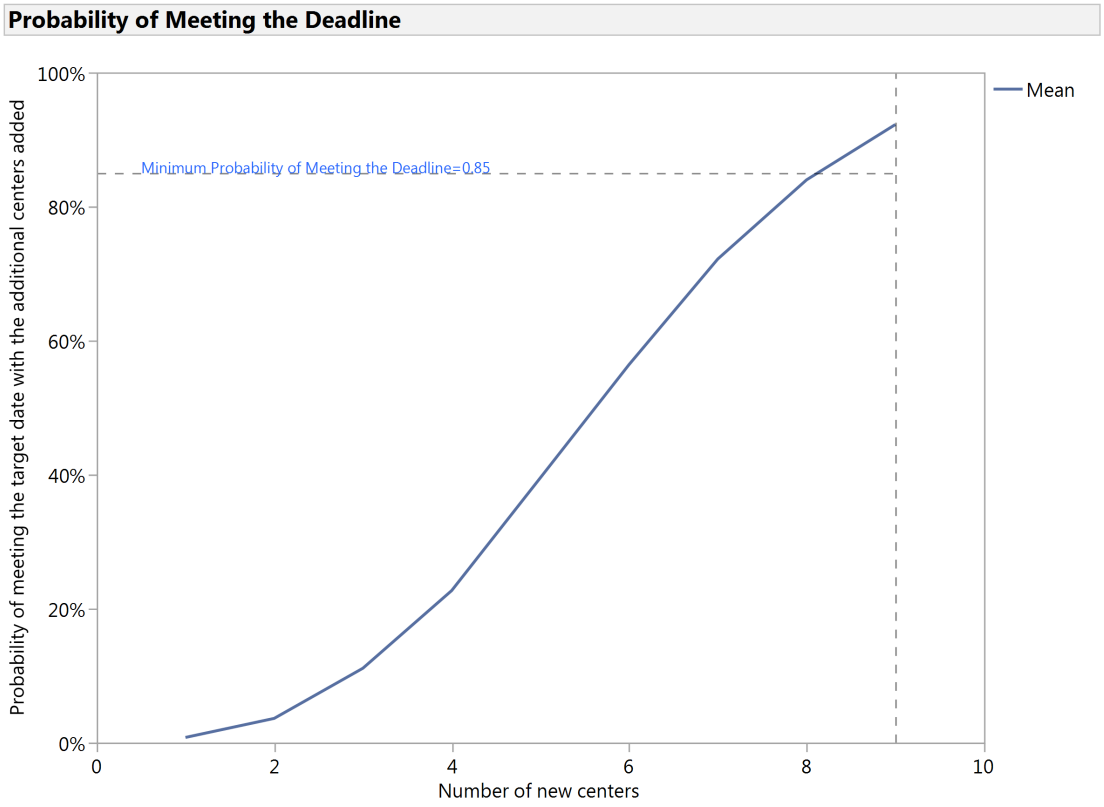

Figure 4. Probability of meeting the deadline with new centers added

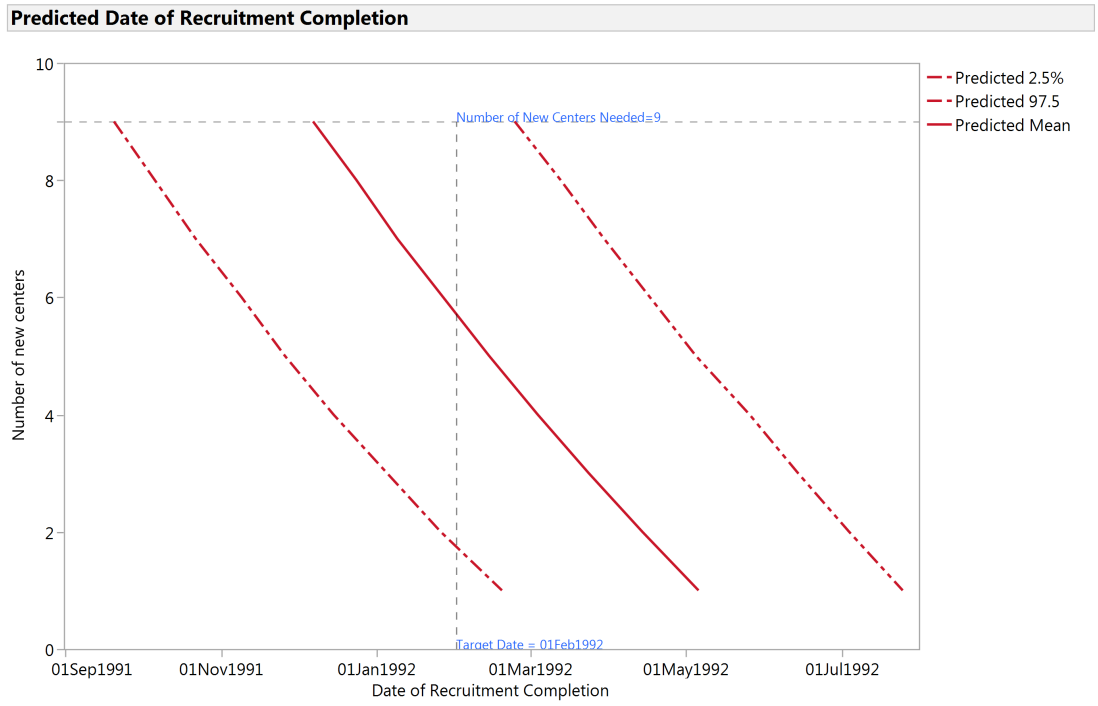

Figure 5. Predicted date of recruitment completion with new centers added

Conclusions

In JMP Clinical 6.0, we implement Anisimov and Fedorov's predictive modeling method to predict patient recruitment for ongoing, multi-center clinical trials. The model assumes that the recruitment of each patient is independent and follows a Poisson process with a Gamma prior. The parameters estimates are calculated by maximum likelihood estimation. An empirical Bayesian approach or MLE is applied to simulations to predict the remaining recruitment time for centers less than 20 or centers more than 19, respectively. If there is a high probability of missing the deadline, then adaptive adjustment is initiated by hypothetically adding new centers until high probability of meeting the deadline. The simulation and prediction are summarized in graphs and tables.

Using the current recruitment data and the new feature in JMP Clinical, investigators are able to have a more straightforward and visual idea of how far they are in the recruitment process. Investigators can use this new feature to manage expectations of patient recruitment and determine the number of new centers to add for them to stay on schedule. The predictive model can be extended to a Pareto process, more complicated than Poisson, or other distributions that fit the setting of the particular clinical trial (Mijoule et al, 2012). It can also be assumed that the recruitment rate is not constant for each center over time, which is closer to the real multi-center recruitment situation, by defining a number of time intervals where recruitment tends to be constant within center. More complicated assumptions involve continuously changing λ_i's and different choices of non-conjugate priors. In principle, the closer the model is to existing conditions, the more accurate the predictions are.

References

[1] Patient recruitment. Quintiles. Retrieved from http://www.quintiles.com/services/patient-recruitment on August 19, 2016.

[2] Insurance against clinical study enrollment delays. Clinical trial patient recruitment, Three Wire. Retrieved from http://www.threewire.com/clinical-trial-patient-recruitment/ on August 19, 2016.

[3] Anisimov, V. V. and Fedorov, V. V. (2007), Modelling, prediction and adaptive adjustment of recruitment in multicentre trials. Statist. Med., 26: 4958–4975.

[4] Barnard, Katharine D., Louise Dent, and Andrew Cook. "A systematic review of models to predict recruitment to multicentre clinical trials." BMC medical research methodology 10.1 (2010): 1.

[5] Mijoule, Guillaume, Stéphanie Savy, and Nicolas Savy. "Models for patients' recruitment in clinical trials and sensitivity analysis." Statistics in medicine31.16 (2012): 1655-1674.

[6] Thoma A, Farrokhyar F, McKnight L, Bhandari M. How to optimize patient recruitment. Canadian Journal of Surgery. 2010; 53(3):205-210.