

Enrollment Simulation in Clinical Trials

Venita DePuy, Bowden Analytics

ABSTRACT

Enrollment predictions for clinical trials are often as simple as estimating A subjects per site per month $\times$ B sites $\times$ C months = D target enrollment. This prediction does not model the uncertainty of the estimated enrollment duration. The commonly used method provides an exact number, but in reality, while we may average a certain number of subjects per month, we are likely to see some variability in enrollment. In addition, this method does not incorporate possible enrollment delays, such as delaying enrollment between each subject due to dose toxicity concerns.

We propose a simulation method to allow for the creation of estimated enrollment durations with associated confidence intervals, while also allowing for optional delays in recruitment. This method simulates exponentially distributed times to subject enrollment. In addition, this approach also allows us to take into consideration variables such as start-up times or enrollment holds due to data monitoring committee meetings.

INTRODUCTION

The largest time-limiting step in any clinical trial is the duration of the enrollment period. Current practice to estimate enrollment timelines is typically no more than estimating a given number of people per month enrolled per site; for example, clinicians approximate that 5 subjects per month per site are recruited in similar trials, so it is calculated that it will take 10 months for 4 sites to recruit 200 subjects. If this timeline is not met, it will affect not only schedules, budgets, and submission deadlines for the current trial, but can also delay an entire program of work, from affecting the start-up of other studies to delay in time until market.

The current prediction method provides a single estimate: it will take a certain number of months to recruit the target number of subjects. While enrollment may average a certain number of subjects per month, exact enrollment will fluctuate on a month-to-month basis.

We propose the use of Markov simulation models to develop an enrollment strategy, incorporating the optimal number of sites and type of sites, which we expect to enroll within a period of time. The historic approach provides an average estimate of enrollment duration. We will, on average, enroll the target number of subjects within that time period. The proposed simulation method allows us to provide an estimate for which we can say, we will enroll within this time period, or earlier, 95% of the time.

STUDY START-UP SIMULATION

Prior to enrolling subjects, each site will require a certain amount of time to review and approve the protocol, receive institutional review board (IRB) approval, receive study medication and other materials, and complete other start-up activities. This is typically calculated as the time between the pre-study screening visit (PSSV) and site initiation visit (SIV). After the SIV occurs, the site is ready to begin enrolling subjects. The amount of time this process requires typically varies between academic sites, such as medical practices associated with a university, and non-academic sites, such as private practices. Start-up times may also vary by therapeutic area and study phase.

We utilize the RANEXP in SAS® to generate random study start-up times from an exponential distribution. User-defined parameters for mean start-up duration for both academic and non-academic sites are used. Start-up times are generated for each site, to incorporate the variability in start-up times that often occurs between sites. We assume that all sites' PSSV occur at the same time; in other words, all start-up periods begin at the same time. Each site's random start-up time is then combined with random times to subject enrollment for that site.

ENROLLMENT SIMULATION

Times to subject enrollment are generated in the same manner as study start-up times. The average time to subject enrollment, often based on the estimates provided by sites at the PSSV or clinicians' estimates of enrollment in similar trials, is defined by the user and used to generate enrollment times from an exponential distribution. The first random number generated for a given site represents the time from the end of study start-up until the first subject enrolled; the second number represents the time from the first subject enrolled until the second subject enrolled; and so on. Subject enrollment times are assigned to each specific site, as they are generated, to allow appropriate incorporation of start-up times.

We recommend generating enrollment times for the total target enrollment at each site. In other words, if the target enrollment is 100, and there are 5 sites, 100 times to enrollment should be generated for each site (500 subjects total). While this does generate an excess number of data, it prevents the possible introduction of bias if insufficient random subjects are generated for a site. If sites will not enroll more than a certain number of subjects, however, that restriction may be incorporated by restricting the number of random subjects generated for each site.

Cumulative times are then generated for each subject, reflective of the actual time from that site's SIV until that subject is enrolled. In other words, suppose Site 1 had a randomly generated start-up time of x months, and the first subject at that site took y months to enroll; that subject would enroll at $x + y$ months after the site initiation visit. Similarly, suppose the second subject at that site took z months to enroll; that subject's time to enrollment would not begin until the previous subject at that site had enrolled, making the second subject's cumulative time $x + y + z$ months.

Subjects are then sorted by cumulative time, inclusive of all sites. One site might randomly enroll several subjects before another site had enrolled any subjects. Since the goal is the enrollment of the target number of subjects, this method allows us to determine the shortest period of time until the target number of subjects is enrolled, in each iteration of the simulation. We then determine the cumulative amount of time, from SIV until enrollment, for the final subject necessary to achieve target enrollment; this is the duration of enrollment for this iteration.

In addition, this type of simulation allows the inclusion of additional terms to describe scheduled delays in enrollment. For example, a Phase I trial might require that certain amount of time elapses between one subject's enrollment and the enrollment of the next subject, due to dose toxicity concerns. Another trial might enroll 3 subjects in each cohort, and require a delay after each third subject to allow for a data review meeting before enrolling the next cohort at a higher dose level.

RESULTS

This simulation incorporates the following user-defined values: target enrollment, number of total sites, number of academic sites, average enrollment per site per month, and start-up duration for both academic and non-academic sites. The amount of scheduled delay between enrollments and the frequency of those delays (such as every 3rd subject) are also defined by the user. For simplicity, user-entered values for initial random number seed and number of iterations of the simulation may also be utilized.

This simulation should be executed for a minimum of 100 iterations, to allow determination of a one-sided 95% confidence interval for the estimated enrollment duration. We recommend using multiples of 10 (100 iterations, 1,000 iterations, et cetera) to simplify the calculation of the confidence intervals. It is also important to use a RETAIN statement or other method to ensure that each iteration uses a different seed value for the random numbers.

The month in which target enrollment is reached is selected from each iteration of the simulation. We recommend rounding the duration up to the next highest integer, to simplify summarization and discussion of the results. The distribution of those months is examined to determine the mean enrollment time. In addition, the one-sided 95% confidence interval is created by determining the range of months which include the shortest 95% of observations. We recommend producing a one-sided confidence interval because, typically, only the upper bound is of interest. Alternately, a two-sided 95% confidence interval may be produced, to provide a lower limit as well; i.e., we can then say, "95% of the time, we

expect the enrollment duration to fall between x and y months.” If desired, the median, minimum, maximum, and other statistics may also be produced.

It is recommended that the mean and one-sided 95% confidence interval be shared with business development, project management, and other appropriate team members. Budgeting, timelines, and other planning should be based on the upper bound of the one-sided 95% confidence interval. Numbers and types of sites, average enrollment, and other values can be adjusted to explore possible outcomes if the original estimates from the simulation do not meet the goals of the study. Team members can then explore what options - increased recruitment effort and associated higher enrollment rate, decreased duration of enrollment delays between subjects due to more timely data review meetings, and so on - can be utilized to meet those goals.

EXAMPLE

A Phase I cancer study with a target enrollment of 40 subjects among 6 sites is being planned. Due to the particular indication, clinical and project management staff estimate that each site would enroll 1 subject every 4 months. Start-up duration estimates of 6 weeks for non-academic sites and 14 weeks for academic sites were provided. Initial estimates of 0.25 subjects per site per month, multiplied by 6 sites, provided a rough estimate of 27 months' enrollment duration (ignoring start-up period duration). However, due to potential toxicity issues, the most recently enrolled subject must complete at least 1 week on study before the next subject can enroll. In addition, team members wanted to consider scenarios in which each site enrolled 1 subject every 3 months and 1 subject every 2 months. It was assumed that 3 sites would be academic and 3 would be non-academic. For simplicity's sake, we will assume that 1 month equal 4 weeks exactly.

A naïve estimator for enrollment duration, including start-up duration, could be calculated as follows:

- 1) 1.5 months in which all sites are in the start-up period.
- 2) 2 months in which the non-academic sites are enrolling and academic sites are not enrolling, due to longer start-up period
- 3) Remainder of enrollment period in which all sites are actively enrolling.

Therefore, we anticipate no enrollment for the first 1.5 months and $0.25 \text{ subjects per site per month} \times 3 \text{ sites} \times 2 \text{ months} = 1.5 \text{ subjects}$ enrolled after the first 3.5 months. The remaining enrollment period duration is calculated as the remaining 38.5 subjects divided by 6 sites divided by 0.25 subjects per site per month = 25.67 months. So, a total of 29.17 months, which we round up to 30 months. This figure does not take into account the delays between subjects; the project team may or may not pad this estimate slightly to account for any delays due to the 1 week lag period.

This estimate reflects the average duration of the enrollment period. Assuming accurate parameter values were used, the target number of subjects can be enrolled during that period, *on average*. For a particular trial, the actual enrollment duration could be shorter or longer, with no measure of the uncertainty of the estimate. The first iteration of the simulation (described below) required 36 months to enroll the target number of subjects. That six month delay can easily equate to hundreds of thousands of dollars in immediate costs. The following results describe the calculation of the mean enrollment duration and associated confidence interval for the same scenario.

For the first iteration of the simulation, the following numbers were generated. Cumulative times from the beginning of study start-up until subject enrollment are presented below, with raw enrollment times displayed in parentheses. We assume that start-up periods begin at the same time for all sites.

Site	Start-up Duration	Cumulative Time to 1 st Subject (Raw Time)	Cumulative Time to 2 nd Subject (Raw Time)	Cumulative Time to 3 rd Subject (Raw Time)
1*	1.87	9.06 (7.19)	13.39 (4.33)	16.07 (2.68)
2*	2.81	3.54 (0.73)	5.27 (1.73)	8.35 (3.08)
3*	1.51	5.46 (3.95)	6.45 (1.00)	13.16 (6.71)
4	2.50	5.09 (2.58)	12.21 (7.12)	13.04 (0.83)
5	5.82	9.18 (3.36)	11.63 (2.44)	13.32 (1.70)
6	0.41	4.15 (3.74)	4.45 (0.29)	7.50 (3.06)

* indicates academic site

The first subject to be enrolled would be the 1st subject at Site 2, at 3.54 months after the beginning of the study start-up period. He would be followed by the 1st subject at Site 6, then the 2nd subject at Site 6, and so on. However, the 4th and 5th subjects, at 5.09 and 5.27 months from the beginning of the start-up periods respectively, fall within the mandatory 1 week (0.25 month) lag between subject enrollments. Therefore, the 5th subject will be delayed and additional 0.07 months, until 5.34 months, to maintain the necessary lag time. The 6th subject, scheduled to enroll at 5.46 months, will then need delayed 0.13 months until 5.59 months, to ensure that necessary lag time is present as well. The 7th subject, at 6.45 months, would not be delayed since a sufficient gap is already present. All enrollment times are programmatically examined and adjusted to ensure that the required enrollment lag is present.

Since enrollment ends when target enrollment is reached, we expect that the number of subjects enrolled will vary between sites. In this iteration, Sites 1 through 6 enrolled 6, 9, 4, 7, 9, and 5 subjects, respectively. The 40th subject was enrolled at 35.1 months, which is rounded up to 36 months. In this particular iteration of the scenario, the 1 week lag time between subjects did not affect the total enrollment period.

While this particular realization resulted in a 36 month duration (compared to the naïve estimate of 30 months), different iterations provide different results. 66.9% of realizations had enrollment durations of 30 months or less. While this supports the idea that, on average, we could enroll the necessary subjects in 30 months or less, it also indicates that, in 33.1% of simulated results, it took longer than 30 months to reach target enrollment. In two iterations, it took 43 months to achieve target enrollment, simply due to random variation.

For this particular scenario, estimates ranged from 6 months to 43 months, with both a mean and a median of 28 months. Since 1000 iterations were calculated, we determine the one-sided 95% confidence interval by finding the time point with 5% (50 realizations) having durations less than that; in this case, 16 months. Similarly, 949 iterations of the simulation have enrollment duration of 35 months or less, 20 have durations of 36 months, and 31 have durations longer than 36 months, so the 95% confidence bound is 36 months. Therefore, we recommend that budgets, timelines, and other resourcing be based on the upper bound of 36 months, with an estimated 0.25 subjects enrolled per site per month.

The table below provides comparisons between 0.25, 0.33, and 0.5 subjects per site per month estimates.

Statistic	0.25 subjects per site per month	0.33 subjects per site per month	0.5 subjects per site per month
Minimum	6	5	4
Mean (Median)	28 (28)	22 (22)	15 (16)
95% Upper Bound	36	28	20
Maximum	43	35	25

These results allow the project team to provide more robust estimates of enrollment duration than the naïve estimate. The team should plan budgets and timelines incorporating the upper one-sided 95% confidence bound of 36 months for subject enrollment. In our simulations of 0.25 subjects per month enrollment rate, 50% of iterations resulted in enrollment durations of 28 months or less, and 66.9% resulted in durations of 30 months or less. It is important to note that in approximately one-third of iterations, however, the naïve estimate was exceeded. Those extended enrollment durations would result in requests for additional funding, reallocation of staff, and even delays in time to market. However, if trial planning incorporated the recommended 95% bound of 36 months, we would expect that - if the 0.25 subjects per site per month estimate is accurate, and no other delays (lack of study materials, protocol amendments, etc.) occur, it is unlikely (<5%) that enrollment will require longer than scheduled.

CONCLUSION

Enrollment projections are a key component of clinical trial planning. It is much more cost-efficient to build timelines for the longest estimated enrollment timeline (the upper 95% confidence bound), with the possibility of completing enrollment earlier, than it is to build timelines around a shorter projected enrollment period and be forced to extend timelines. If projected timelines are unacceptable to the project team, it is straightforward to vary enrollment rates or number of sites to calculate revised enrollment durations.

While no simulation can account for the variety of possible delays, from lack of drug availability to IRB delays due to protocol amendments, we believe this simulation method and its resulting confidence intervals can improve the planning process by providing estimates that account for variations in enrollment.

We recognize that most sites would not enroll on weekends, and that infinitesimal delays - such as those of less than one day - are not of clinical relevance, but for the purposes of simulation we treat time as a continuous variable with no differentiation for weekends.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Venita DePuy
Owner/Principal Biostatistician, Bowden Analytics
Bowden.analytics@gmail.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.