Paper AS03

Faster Patient Recruitment in Clinical Trials through Simulations

Keerthana Udupa, Novo Nordisk Service Centre India Private Ltd., Bengaluru, India

ABSTRACT

Large-scale clinical trials typically require a randomized, double-blind approach and must be conducted across multiple centres and countries to accurately verify the safety and effectiveness of treatments for a diverse range of patient groups. Efficient recruitment is essential for minimizing both the time and cost associated with trials.
While Monte Carlo simulations have been utilized in clinical trial planning for various purposes, a focus on recruitment represents a new approach. Important applications include evaluating treatment effects, assessing cost-effectiveness, and estimating long-term benefits. By analyzing these methods, the aim is to improve recruitment and reduce trial duration. Ultimately, these simulations may provide valuable insights for developing more efficient and effective clinical trials.

This paper introduces Monte Carlo simulation Markov models designed to enhance recruitment processes and effectively estimate and adjust parameters. Both continuous and discrete time simulation models for recruitment strategies will be explored, illustrated with straightforward hypothetical examples.

INTRODUCTION

Clinical trials are the cornerstone of obtaining marketing authorization for new medical products and indications. These large-scale, often multi-country experiments are designed to verify hypotheses regarding the safety and efficacy of treatments. Planning a clinical trial is a massive undertaking, and one of the most critical steps is figuring out how to get enough volunteer patients to participate. The recruitment process is frequently the most significant bottleneck in trial progression.

We often think of recruitment as a simple schedule, but it is "stochastic. Why is this so important? Because a trial's success hinges on recruiting the right number of people in a reasonable amount of time. If recruitment is too slow or falls short, the entire study can be delayed, become too expensive, or even fail to produce a meaningful result. Designing the recruitment process properly is essential because failing to meet the required sample size can reduce the probability of success or even make it impossible to confirm the trial's hypothesis. The "secret" to a successful trial is finding an equilibrium-a perfect balance between how much the trial costs and how long it takes to finish.

This paper introduces a powerful tool that helps researchers plan for this challenge with the objective of providing a detailed, instructive overview of several Monte Carlo simulation models for clinical trial patient recruitment. We will explore a range of techniques, from simple deterministic approaches to more complex stochastic models. These methodologies are designed to equip researchers and trial managers with the tools needed to plan, analyze, and optimize the critical recruitment phase

WHAT IS A SIMULATION MODEL?

In the context of clinical trial recruitment, a simulation model is like creating a "virtual trial" on a computer. It's a method for predicting how patient enrolment might unfold over time based on a set of key assumptions. By running the simulation many times, researchers can explore different possible outcomes and get a sense of the potential risks and timelines.

A core principle of effective simulation is a clear understanding of the model's underlying assumptions such as:

- Records of the exact calendar time when volunteers were contacted in previous trials.
- The average number of patients recruited per week or per month in similar studies.

These assumptions define the recruitment scenario and dictate which modeling approach is most appropriate. This section defines the key principles that differentiate recruitment dynamics and guide the selection of a suitable simulation framework.

THE BUILDING BLOCKS OF RECRUITMENT MODELS

TIME MEASUREMENT: CONTINUOUS AND DISCRETE TIME

The way a model measures time is another critical distinction, often dictated by the granularity of available data.

- **Continuous-Time Models:** These models track the precise moment each patient arrives. The key variable is the inter-arrival time; the duration between one successful recruitment and the next. This approach is most effective when individual-level data is available, providing the exact calendar time each participant was contacted or enrolled.
- **Discrete-Time Models:** These models aggregate patient arrivals into defined periods, such as weeks or months. The focus is not on the exact moment of arrival but on the number of patients recruited within each interval. This approach is practical when the only available data is an average recruitment rate or periodic counts of enrolled patients.

Table 1 Continuous vs. Discrete Recruitment

Feature	Continuous Time Model	Discrete Time Model
Core Idea	Assumes patients arrive one-by-one, sequentially over a continuous timeline.	Assumes a certain number of patients arrive within fixed time periods.
What it Measures	The exact time *between* each patient's arrival (inter-arrival time).	The *total number* of patients who arrive in each period.
When to Use It	When you have detailed, individual-level data on patient arrival times.	When you only have access to average recruitment rates.

The choice between these two approaches often depends on the type and quality of historical data available to the trial planners.

RECRUITMENT DYNAMICS: ACUTE AND CHRONIC DISEASE

- **Acute Diseases:** Characterized by a relatively short duration, recruitment for acute conditions is primarily constrained by incidence; the rate of new cases arising in the population. Potential participants are typically identified when they seek new symptoms. The recruitment rate is therefore a function of how frequently the disease occurs and the capacity of the research center to enroll new patients.
- **Chronic Diseases:** These conditions have a long duration, often lasting a lifetime. Consequently, recruitment is constrained by prevalence; the proportion of the population currently living with the disease. The process involves drawing from an existing, finite pool of eligible individuals. As recruitment progresses, this pool can become depleted, often causing the recruitment rate to decrease over time.

Table 2 Acute vs. Chronic Disease

Disease Type	Acute Diseases	Chronic Diseases
Patient Pool	Constrained by incidence.	Constrained by prevalence.
Recruitment Dynamic	The pool of potential patients is constantly being replenished by new cases.	The pool of potential patients is often finite and can be depleted over time, making recruitment harder as the trial progresses. This is a "non-replacement" scenario.
How Patients are Found	Typically approached when they visit a health center for symptoms.	Approached through advertising, invitation letters, or at scheduled visits from a pre-existing population.

This fundamental difference between a replenishing patient pool (acute) and a depleting one (chronic) is not just a theoretical point; it directly determines whether we should build a model "with replacements" or "without replacements," a critical choice.

CONTINUOUS-TIME SIMULATION MODELS

Continuous-time models are most appropriate when the analysis requires a granular understanding of the recruitment timeline and the exact timing between patient arrivals is a key variable. These models are particularly valuable when individual-level recruitment data is available to inform their parameters. This section details three continuous-time models of increasing complexity, each building upon the last to add layers of realism.

THE SIMPLE DETERMINISTIC MODEL

This is the most basic approach to recruitment modeling, serving as a foundational baseline for more complex simulations. Its primary value lies in its simplicity for establishing an initial, idealized timeline.

The model's core assumption is that there is a fixed, constant time (T) between each successive patient arrival. The total time required to recruit the target sample size is calculated deterministically, without any element of randomness.

The key variables and their relationships are as follows:

- t: The total accumulated time from the start of recruitment.
- i: A counter for the number of patients who have arrived.
- N: The total required sample size for the trial.
- T: The constant inter-arrival time. This can be derived from the expected number of patients per period (np), using the formula T = 1 / np.

The procedural logic is straightforward: the total time required to recruit N patients is calculated as t = N * T. This provides a single, fixed estimate of the recruitment duration.

STOCHASTIC MODEL WITH A CONSTANT RECRUITMENT PROBABILITY (S1)

This model introduces a crucial element of realism: uncertainty. It acknowledges that not every patient who is contacted will ultimately be recruited into the trial.

The core assumption is that each contacted patient has a constant probability (P) of being successfully recruited. This probabilistic element makes the recruitment duration a random variable rather than a fixed outcome.

The key variables introduced are:

- P: A constant probability of successful recruitment for any contacted patient.
- R: A random number drawn from a uniform distribution between 0 and 1, used to simulate the recruitment outcome.

The simulation advances time in fixed increments of T, representing each contact attempt. At each step, a random number R is generated. If R <= P, an arrival is recorded. If R > P, no arrival is recorded, but time still advances by T to the next attempt. This mechanism correctly models that failed attempts consume time. This process repeats until the target sample size N is reached.

The impact of this probabilistic element is significant. As the recruitment probability P decreases, the average time between successful arrivals increases. This leads to a longer and, critically, more variable total recruitment duration, which can be quantified by running the simulation multiple times.

STOCHASTIC MODEL WITH A RECRUITMENT PROBABILITY DISTRIBUTION (S2)

This model represents a more sophisticated approach by accounting for uncertainty about the recruitment probability itself. It is useful in scenarios where P is not known with certainty and may vary across patients or over time.

Its core assumption is that for each contacted patient, the recruitment probability P is not a fixed constant but is itself a random variable drawn from a probability distribution. For example, P could be drawn from a Uniform {0, 1} distribution for each contact attempt.

The simulation logic involves a two-step random process for each contact:

1. A value for the recruitment probability P is generated from its specified distribution (e.g., Uniform {0, 1}).
2. A second random number R is generated from a Uniform {0, 1} distribution.
3. The patient is recruited if R <= P.

This hierarchical structure models two layers of uncertainty: the inherent randomness of a patient's decision (modeled by R) and the operational uncertainty about the true recruitment probability for any given contact (modeled by P). This approach results in a marginal recruitment probability equal to the mean of the chosen distribution for P (e.g., an average probability of 0.5 for a Uniform {0, 1} distribution). However, because P itself varies with each attempt, this model introduces significantly more variability into the recruitment timeline compared to the fixed-probability model.

Having explored models focused on inter-arrival times, we now shift our attention to models that forecast recruitment based on arrivals aggregated into defined periods.

DISCRETE-TIME SIMULATION MODELS

Discrete-time models are employed when recruitment data is aggregated into periods, such as weekly or monthly counts. The primary goal of these models is to forecast the number of patients expected to be recruited in each successive interval of the trial, rather than the precise timing of each arrival.

DISCRETE MODEL WITH REPLACEMENT (S3)

This model operates under the core assumption that patients are drawn from a population where those willing to participate are fully replaced. This implies that the pool of potential candidates is not depleted over time, and the probability of recruitment does not degrade due to prior enrollment activity. This is often a reasonable assumption for trials targeting acute conditions with high incidence.

The simulation generates aggregate period counts by modeling one potential patient at a time. For each individual patient, the model attempts recruitment in Period 1. If the attempt fails (R > P), it re-attempts for the same patient in Period 2, and so on, until the patient is recruited or a maximum number of periods is reached. After the patient's final outcome is determined (i.e., they are assigned to an arrival period or are never recruited), the simulation proceeds to the next individual patient, starting again from Period 1. This process is repeated until the target sample size N is accumulated across all periods.

This model can be adapted to use either a constant probability P across all periods or a varying P that changes from one period to the next to reflect factors like seasonal effects or the impact of advertising campaigns.

DISCRETE MODEL WITHOUT REPLACEMENT (S4)

This model is specifically designed for scenarios where the pool of eligible and willing patients is finite and is not replenished, a common situation in trials for chronic or rare diseases.

The key implication of the "without replacement" assumption is that the probability of recruiting a new patient is dependent on the number of patients already recruited. As the finite pool of candidates is depleted, the recruitment rate is expected to decrease over time.

The primary objective of this model is to assess two critical planning metrics that arise from this dynamic:

- The expected delay in recruitment caused by the diminishing pool of candidates.
- The expected effective duration of the recruitment phase, accounting for these delays.

These metrics provide a more realistic forecast for trials where finding willing participants becomes progressively more challenging.

Having outlined the theoretical frameworks for both continuous and discrete-time models, the next section will demonstrate their application to a practical scenario.

ILLUSTRATIVE APPLICATION AND ANALYSIS

Applying theoretical models to a concrete example is invaluable for understanding their practical outputs and implications. This section uses a hypothetical clinical trial scenario to demonstrate the results generated by the different simulation models and provides an interpretation of their contrasting forecasts.

CASE STUDY PARAMETERS

The simulation models were applied to a hypothetical case study with the following parameters:

- **Trial Objective**: A clinical trial to confirm the safety and effectiveness of a treatment for the prevention of cytomegalovirus, a rare disease infection in women.
- **Required Sample Size (N):** 720 patients.
- **Proposed Recruitment Period**: 2 years (24 months).

INTERPRETING MODEL OUTPUTS

The outputs from the continuous and discrete models provide different but complementary insights into the recruitment timeline and patient flow.

Continuous Model Results

The continuous-time models were used to estimate the total time required to recruit all 720 patients under different assumptions about the recruitment probability.

The continuous-time models (Simple, S1, and S2) were used to estimate the total time required to recruit all 720 patients under different assumptions about the recruitment probability.

Table 3 Estimated Total Recruitment Time for Continuous Models

Model	Probability (P)	Mean Time (months)	Standard Deviation
Simple Model	1	24.00	0.00
S1	0.96	25.00	0.19
S1	0.8	30.00	0.51
S1	0.5	48.03	1.29
S2	Uniform {0,1}	47.97	1.24

As shown in Table 1, the simple deterministic model provides a baseline estimate of 24 months, assuming perfect recruitment efficiency (P=1). However, as realism is introduced, the forecast changes dramatically. In model S1, as the probability of successfully recruiting a contacted patient (P) decreases from 0.96 to 0.5, the mean recruitment time increases from 25 months to 48 months. This demonstrates the profound impact of recruitment efficiency on the trial timeline. Notably, the results for S2, where P is drawn from a Uniform {0,1} distribution, are very similar to S1 with a fixed P=0.5, as the average probability in the uniform distribution is 0.5.

Discrete Model Results

The discrete-time models (S3 and S4) forecast the number of patients expected to be recruited in each of the 24 one-month periods.

Table 4 Estimated Number of Patients Recruited per Period for Discrete Models

Period (T)	S3 (a) Mean	S3 (b) Mean	S3 (c) Mean
1	29.9	151.3	95.0
2	30.1	149.7	82.0
3	29.8	89.9	71.6
4	29.9	65.1	63.6
5	30.1	59.0	54.9
6	30.0	23.5	48.0
7	30.1	31.8	41.6
8	30.2	35.9	36.4
9	29.9	17.3	32.1
10	30.0	13.8	27.9
11	29.8	16.2	24.8
12	29.9	9.6	21.5
13	30.0	7.9	18.9
14	30.0	3.9	16.4
15	30.2	5.9	14.1
16	29.9	4.9	12.5
17	29.8	2.6	11.1
18	29.9	0.6	9.8
19	29.9	5.1	8.7
20	30.4	2.2	7.6
21	30.3	5.5	6.4
22	30.0	9.9	5.7
23	30.1	0.4	4.9
24	29.8	7.8	4.4
Table notes: (a) Assumes a constant recruitment probability across all periods. (b) Assumes a unique, pre-specified probability for each period. (c) Demonstrates a declining recruitment trend under a constant probability due to the simulation's logic of repeated attempts per patient.			

The patterns in Table 4 reveal different recruitment dynamics. Model S3(a), assuming a fixed recruitment probability, shows a stable average recruitment of approximately 30 patients per month. In contrast, model S3(c), which uses a constant probability within the "with replacement" framework, shows a sharp decline from 95 patients in the first month to just 4.4 in the last. This front-loaded pattern is a direct consequence of the simulation logic, which gives each patient multiple opportunities to enroll across successive periods; most successful recruitments will naturally occur during the first few attempts.

Table 5 Recruitment Expected Values for Discrete Model without Replacement (S4)

Metric	Arrivals at End of Period	Arrivals Within Period
Expected Delay (Months)	6.77	6.49
Effective Duration (Months)	17.23	17.51
Standard Deviation	6.48	6.44

Finally, Table 5 quantifies the impact of patient pool depletion using model S4. This forecast allows managers to distinguish between the planned 24-month recruitment window and the effective 17.5-month duration during which recruitment is active, followed by an expected 6.5-month "tail" where finding the remaining patients becomes highly inefficient. This insight is critical for resource planning, allowing for a de-escalation of recruitment efforts in the final months to control costs.

CONCLUSION

This report has detailed a suite of Monte Carlo simulation models designed to bring quantitative rigor to the planning of clinical trial patient recruitment. By moving from simple deterministic calculations to more nuanced stochastic simulations, these models provide a powerful framework for forecasting timelines, allocating resources, and managing risk.

The key insights reveal a clear distinction in the application of these models. Continuous-time models are best suited for detailed analyses of inter-arrival times and are most powerful when individual-level data is available. In contrast, discrete-time models provide a practical approach for period-based forecasting, making them ideal when data is aggregated into weekly or monthly counts. The choice between a "with replacement" and "without replacement" framework is critical and should be guided by whether the trial targets an acute condition with a constant influx of new patients or a chronic condition with a finite, depleting pool of candidates.

The proposed Monte Carlo simulation Markov models are versatile and powerful tools. They enable trial managers to optimize recruitment processes, estimate key operational parameters, and calibrate timelines with greater confidence. By embracing these simulation methodologies,

REFERENCES

[1] Abbas I, Rovira J, Casanovas J. Clinical trial optimization: Monte Carlo Simulation Markov model for planning clinical trials recruitment. Contemp Clin Trials. 2007 may;28(3):220-31. Doi: 10.1016/j.cct.2006.08.002. Epub 2006 aug 10. Pmid: 16979387.

[2] Halpern EF,Weinstein MC, Hunink MJM, Gazelle GS. Representation both first and second-order uncertainties by Monte Carlo Simulation for groups of patients. Med Decis Mak 2000;20:314–22.

[3] Moir R. A Monte Carlo analysis of the Fisher randomization technique: reviving randomization for experimental economists. Exp Econ 1998;1:87–100.

[4] OpenAI. (2025). ChatGPT (GPT-4o, 2 July version) [Large language model].https://chat.openai.com

DISCLAIMER: The views and opinions expressed in this paper belong solely to the author and do not necessarily reflect those of the author's employer, organization, committee, or any other group or individual. No data or confidential information belonging to Novo Nordisk has been utilized in this paper.

CONTACT INFORMATION (HEADER 1)

Your comments and questions are valued and encouraged. Contact the author at:
Author Name: Keerthana Udupa
Company: Novo Nordisk Service Centre India Private Ltd.

Address: 9th, 10th,11th floor, Cignus (K Raheja), Bengaluru, Karnataka, India, 560066
Work Phone: 8951899017
Email: kqzu@novonordisk.com