Monte Carlo Simulation: Clinical Trial Forcasting

Team: Anaiysaa Deswal, Adit Doppa, Hanna Akita, Nivan Gujarl, Nitya Sunkarapalli, Seoyeon Jung

PURDUE UNIVERSITY | The Data Mine

Johnson&Johnson

BACKGROUND

Our project develops a simulation-driven web application designed to help Johnson & Johnson forecast clinical trial enrollment more accurately and support better data-driven decision making.

Key Objectives:

- Build a Monte Carlo simulation framework to model uncertainty in patient enrollment across clinical trial sites.
- Use Generative AI tools to accelerate code generation and improve development iteration speed.
- Generate probabilistic forecasts to better estimate potential clinical trial timelines.
- Create a structured, data-driven forecasting approach to improve overall clinical trial planning and progress tracking.

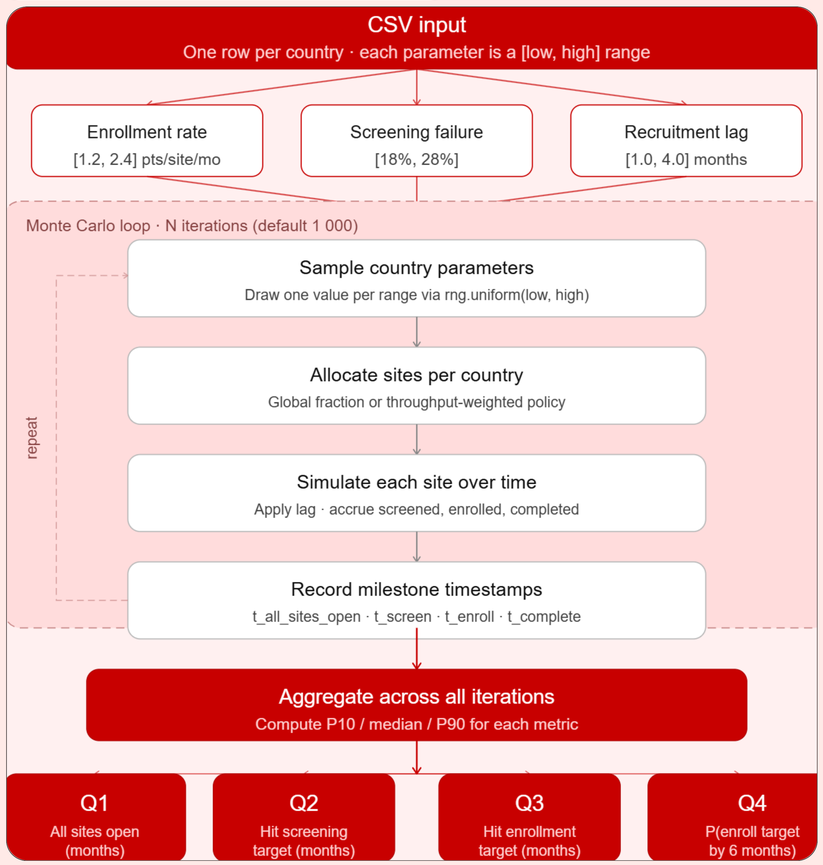
PROJECT GOALS

Clinical trial enrollment is difficult to forecast accurately, often leading to timeline delays, increased operational costs, and nearly 80% of trials fail to meet their initial enrollment targets, potentially costing pharmaceutical companies up to $8 million per day. These challenges are driven by factors such as site variability in recruitment performance, operational fatigue among trial teamsand limited forecasting tools.

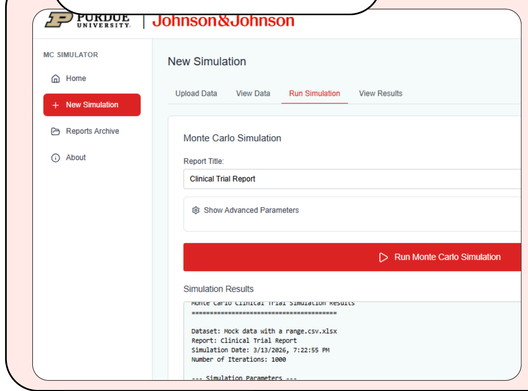
TECHNOLOGIES

Tools: Gemini CLI, OpenAI API
Front-End: React, Java Script
Development: AGILE flow, AIDLC
Backend: Python

SIMULATION METHODOLOGY



DASHBOARD VIEW



CONCLUSION

- Developed a probabilistic simulation model to forecast clinical trial enrollment using Monte Carlo methods
- Incorporated site capacity, enrollment variability, and patient-level attributes to improve timeline estimation under uncertainty
- Evaluated GenAI frameworks (Gemini CLI) for internal development workflows
  - Improved development efficiency and iteration speed
  - Human validation required to ensure accuracy and reliability
- Gained experience in simulation modeling, data-driven decision making, and full-stack development within a pharmaceutical context
- Model captures key system dynamics, but does not fully represent real-world constraints such as regulatory delays, patient dropout behavior, and operational variability
- Future work: integrate real clinical data, enhance model fidelity, and optimize for production deployment

DEVELOPMENT METHODOLOGY

Data Creation

- **Dataset Generation:** Simulated clinical trial data using country-level inputs (sites, timelines, capacity)
- **Patient Simulation:** Generated patient-level attributes and enroliment patterns

Analysis & Modeling

- **Sampling & Milestones:** Extracted key timepoints and calculated trial milestones
- **Monte Carlo Simulation:** Modeled uncertainty and outcome variability
- **Outcome Aggregation:** Averaged results to estimate timelines and performance

Enhancements & Optimization

- **Model Enhancements:** Added weighted distributions and validation data
- **Advanced Analytics:** Improved realism and accuracy of simulations

AI Integration

- **AI Models:** Assisted code generation and workflow automation
- **APIs:** Enabled seamless data access
- **Reproducibility:** Improved consistency of simulation results

SAMPLE INPUT DATA

Country	Num Sites	Screening Failure	Enrollment Rate	Recruitment Lag	Max Patients	Follow-up Duration	Dropout Rate (%)	Max Patients
US	86	24.5–25.5	1.64–1.70	1.86–1.94	73.5–76.5	20.58–21.42	21.17–22.03	5160

DEVELOPMENT OPTIMIZATION

In addition to technical project development, our team was requested by J&J to explore how GenAI tools like CLIs could be integrated into our development framework as a model for accelerating J&J's internal software development timelines.

We Explored:

- Used Gemini CLI to navigate, refactor, and extend the existing simulation codebase through natural language commands, simulating how a new J&J developer might ramp up on an unfamiliar internal system
- Assessed output consistency and reliability, identifying where AI assistance reduced friction and where human oversight remained necessary
- Compared against multi-model approach (ChatGPT, Claude, Deepseek, Ollama) to evaluate single deeply-integrated tool vs. broad multi-model workflow

Impact For J&J:

- Reduces engineer ramp-up time on large, complex internal codebases across clinical operations and regulatory systems. AI-assisted refactoring could accelerate technical debt reduction without dedicated engineering sprints
- Standardizing on a GenAI development framework gives J&J consistent, auditable AI assistance, critical in regulated pharmaceutical software environments where traceability matters

FUTURE GOALS

Real patient-level clinical data is highly restricted; Which is why in this project the team relied on synthetic, statistically-grounded mock data, which limits direct validation against true trial records. To more accurately forecast enrolment and increase usability as well as project scalability, we aim to:

1. Extend Patient-Journey Modeling :
   - Integrate site fatigue, site-level seasonality, and patient dropout distributions. Track operational needs forecasting: drug supply, staff allocation, resource budgeting
2. Expand to Terminal-Based Exported Package
   - Package simulation as a standalone CLI tool for use in existing J&J pipelines. Support batch runs across trial config and output structured JSON/CSV for reporting

ACKNOWLEDGEMENTS

We are grateful to our TA, Khushboo Tewari, and the J&J team - Fei Chen, for their ongoing support and guidance. Special thanks to Nicholas Lenfesti, Ashley Arroyo, and the entire Data Mine staff for their resources and assistance throughout the project.