Paper AS04

AI-Augmented Bayesian Methods for Late-Onset Toxicity in Early-Phase Trials

Rachana Lele, Syneos Health, San Antonio, Texas, US
Arfan Afzal, Syneos Health, Calgary, Canada

ABSTRACT

Late-onset toxicities present a persistent challenge in early-phase clinical trials, especially those involving molecularly targeted therapies. Adaptive and traditional dose-finding designs like CRM, TITE-CRM, and Rolling 6 have been proposed to address delayed toxicity profiles, but their reliability is often compromised by poor parameter specification, censored data handling, and unvalidated model assumptions. This review complements these design approaches by highlighting Bayesian simulation techniques that use posterior predictive probabilities to evaluate and anticipate future toxicity risks. These approaches enable proactive safety monitoring through formal stopping rules and accrual suspension based on predictive metrics. Despite strong performance in simulations, model-based designs remain underutilized due to implementation complexity and regulatory hesitation. To advance this space, we explore AI-driven adaptive Bayesian models that dynamically update parameters as new data emerge. These methods improve prediction accuracy for late-onset events and offer a promising framework for more responsive, data-informed early-phase trial designs.

INTRODUCTION

Late-onset toxicities create a major challenge in modern early-phase cancer trials. This is especially true with immunotherapies and targeted agents, which often lead to delayed immune-related or cumulative side effects. These toxicities and treatment responses can appear at any time during follow-up across multiple cycles [1]. This complicates the use of traditional short-term toxicity models to establish safe dose levels. The delay presents a statistical problem: when assigning doses to new patients, many previously treated patients have not finished their dose-limiting toxicity (DLT) window. This results in censored or partially observed data. Traditional frequentist dose escalation methods like 3+3, Rolling 6, or fixed cohort rules cannot use partial information. They require complete DLT outcome observations before moving forward, which either delays patient enrollment or leads to unsafe dose increases by ignoring pending data. Rapid enrollment and late-onset toxicity cause major challenges for rule-based designs [2], as these methods do not consider time-to-event information. Model-based adaptive designs like CRM and TITE CRM were developed to address these issues. For example, TITE CRM uses time-to-event weighting, allowing patients with only partial follow-up to provide fractional information to the likelihood. This is particularly helpful when toxicity happens over several cycles [1]. However, these designs remain weak against errors due to model misspecification, prior selection, and assumptions about the toxicity curve. If these assumptions are incorrect, it can negatively affect their performance. Also, many existing model-based designs require precise calibration and may become unstable when outcomes are delayed [3]. To boost the reliability of these designs, recent studies have looked at Bayesian simulation methods, especially posterior predictive modeling. These methods assess and predict future toxicity risks by considering realistic accrual and censoring patterns. They work alongside CRM and TITE CRM rather than replacing them, permitting scenario-based evaluations, sensitivity analyses, and predictive safety monitoring. Simulation-based accrual suspension rules (for example, stopping when more than 50% of data are pending) are essential when applying TITE BOIN to late-onset toxicity cases. Predictive probabilities help guide decisions on halting enrollment and managing overdoses [5,8,9]. These advancements suggest a more flexible framework for early dose-finding phases. This includes time-to-event modeling, posterior predictive probabilities, and simulation-based decision support to handle uncertainty from delayed toxicities. This review points out the limitations of traditional methods, details how Bayesian simulation enhances model-based design evaluation, and examines new opportunities for machine learning-enhanced Bayesian models that adjust toxicity predictions as fresh data comes in.

ADAPTIVE AND TRADITIONAL DOSE-FINDING APPROACHES

CONTINUAL REASSESSMENT METHOD (CRM)

The Continual Reassessment Method (CRM) was developed to address the known issues with rule-based designs like 3+3, which increase doses without fully using the available toxicity data. CRM uses a model-driven approach that continuously updates the estimated dose-toxicity relationship as new patient outcomes are observed. This allows for a more effective and accurate identification of the maximum tolerated dose (MTD). CRM works best when all patients have finished their DLT assessment period. This way, the model can include complete toxicity data and select the dose with an estimated toxicity probability closest to the target [4]. However, CRM is less effective in cases with late-onset toxicities because it requires fully observed DLT outcomes before updating the model. When toxicities are delayed, many outcomes remain pending. This situation forces CRM to either delay enrollment or overlook

incomplete follow-up, both of which can endanger patient safety and trial efficiency. The process includes treating a patient at an initial dose, observing the toxicity outcome, updating the toxicity estimates across doses, selecting the next dose based on the updated model, and repeating this until the trial concludes.

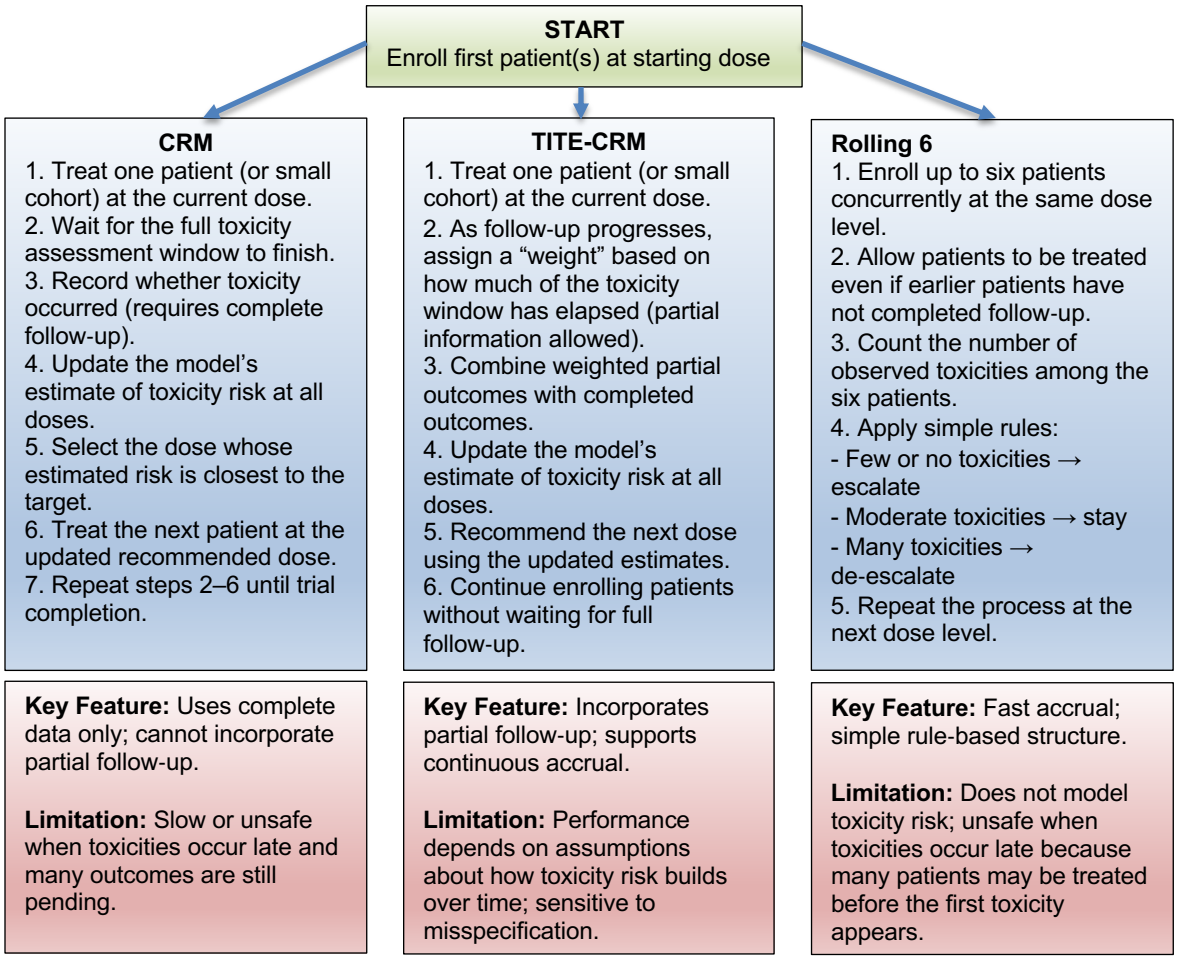
TIME-TO-EVENT CRM (TITE-CRM)

Time to Event CRM (TITE CRM) was developed to overcome the limits of traditional CRM in handling incomplete follow-up data when side effects show up late. In modern oncology trials, side effects can happen at any time across multiple cycles, making it impractical to wait for the full dose-limiting toxicity (DLT) window before adjusting doses. TITE CRM addresses this issue by using a weighting system that considers how much of the DLT window each patient has completed. This allows for ongoing patient enrollment and real-time dose assignment, even when many outcomes are still pending. TITE CRM is particularly suited for immunotherapies and targeted agents with longer toxicity windows. However, it faces challenges in cases where side effects appear late because its weighting depends on assumptions about how the risk of toxicity builds up over time. Incorrect assumptions can lead to biased dose recommendations. Model-based designs remain sensitive to such mistakes and require careful calibration and simulation to ensure safety. The method involves enrolling a patient, assigning a weight based on follow-up time, combining this weighted partial data with complete outcomes, updating toxicity estimates, and selecting the next dose without waiting for all data.

ROLLING 6

The Rolling 6 design was introduced in pediatric oncology to allow faster patient enrollment than the traditional 3+3 design while keeping the rules simple and easy to follow. This method permits up to six patients to be treated at the same dose level at once. This is helpful in situations where quick enrollment is crucial and side effects are likely to show up early. However, Rolling 6 is mainly a rule-based design and doesn't estimate toxicity risk or consider partial follow-up. Because of this, it is not suitable for treatments that have late side effects. If side effects happen later, many patients might receive an unsafe dose before the first one is reported. Late side effects make it harder to determine the right dose since results can emerge during several treatment cycles, a situation where Rolling 6 lacks a statistical way to handle pending results. The design recruits up to six patients at a dose, waits for dose-limiting toxicity (DLT) results according to its rules, makes straightforward decisions to increase or decrease the dose based on observed side effects, and continues this until it finds the maximum tolerated dose (MTD). While these designs represent important steps in finding the right dose, their effectiveness drops when toxicity results are delayed, only partially seen, or spread over multiple cycles. Because the Continuous Reassessment Method (CRM) and Rolling 6 cannot fully make use of censored or pending results, they require additional tools to evaluate safety under realistic patterns of enrollment and follow-up. Time-to-event CRM (TITE CRM) can use partial follow-up, but it still depends on assumptions about how toxicity risk increases over time—assumptions that may not be accurate when side effects emerge in complicated or delayed ways. Due to these limitations, these methods can benefit from Bayesian simulation techniques that assess safety under realistic patterns of enrollment and late-onset side effects.

WORKFLOWS FOR CRM, TITE-CRM, AND ROLLING 6 DESIGNS



BAYESIAN SIMULATION TECHNIQUES

Bayesian simulation techniques involve scenario-based, model-driven assessments. They use repeated simulated trials to understand how a design performs under real-world patterns of accrual, censoring, and late-onset toxicity. In this approach, a Bayesian model defines the dose–toxicity (and sometimes dose–efficacy) relationship. Posterior or posterior predictive distributions generate future outcomes based on different assumptions about the timing and occurrence of events. These simulations help researchers evaluate operating characteristics like the chance of choosing an unsafe dose, the likelihood of early stopping, and how patients are distributed across doses before applying the design in practice.

These techniques are especially useful when traditional methods like CRM, TITE CRM, or Rolling 6 are used in situations with late-onset toxicity. Although CRM and TITE CRM can, in principle, model time-to-event data, their accuracy relies heavily on assumptions about how toxicity develops over multiple cycles and how partial follow-up is weighted. Bayesian simulation makes these assumptions clear and examines them across various plausible scenarios. This process reveals whether a design might be too aggressive, too cautious, or too vulnerable to incorrect assumptions.

When developing the Joint TITE CRM, extensive simulation evaluated how the design performs when both toxicity and efficacy can occur at any time during multi-cycle follow-up. It also helped establish decision rules. Central to these methods are the use of posterior predictive probabilities. Given the data observed so far, the model generates the distribution of future toxicity outcomes at each dose. This information can then define stopping rules, accrual suspension criteria, or overdose control boundaries. In designs like TITE BOIN12, Bayesian predictive calculations manage late-onset toxicity and efficacy by assessing, at each interim, whether continuing at a certain dose is likely to breach pre-specified safety or efficacy thresholds.

The process includes (1) defining a Bayesian model for toxicity (and possibly efficacy), (2) fitting the model to existing data, including censored and time-to-event information, (3) simulating future patient outcomes using the fitted model, (4) calculating predictive probabilities of key events like excessive toxicity, and (5) applying decision rules based on these probabilities to escalate, de-escalate, stop, or suspend accrual. Embedding traditional designs within a Bayesian simulation framework allows for more transparent and safer management of late-onset toxicities.

For CRM and TITE CRM, simulation helps identify strong prior choices, appropriate weighting schemes for partial follow-up, and decision thresholds that ensure safety while maintaining efficiency. For simpler rule-based designs like Rolling 6, Bayesian simulation can show their limitations with delayed toxicity—such as the risk of treating many patients at unsafe doses before observing any dose-limiting toxicities—and encourage the adoption of more model-based or predictive methods. Bayesian simulation techniques do not replace these designs but enhance them into better-understood, better-calibrated tools for trials where toxicity and efficacy develop over long periods.

Table 1. Comparison of traditional designs and Bayesian-enhanced approaches [1,5,6]

Design/ Approach	Core idea	How it handles late-onset toxicity	Role of Bayesian simulation	Typical use case
CRM	Model-based dose–toxicity curve updated as outcomes accrue	Requires mostly complete DLT data; struggles with many pending outcomes	Used to calibrate priors, assess safety, and quantify operating characteristics under delayed events	Early-phase trials with relatively short DLT windows
TITE-CRM	Extends CRM with time-to-event weighting for partial follow-up	Uses weighted partial data; still sensitive to assumptions about risk over time	Simulation evaluates different weighting schemes, accrual patterns, and misspecification scenarios	Trials with multi-cycle or delayed toxicities where model-based design is acceptable
Rolling 6	Rule-based design allowing up to six patients per dose	Does not model time-to-event; vulnerable when toxicities are late	Simulation highlights risks of overdosing and motivates more advanced designs	Pediatric or simple settings with expected early toxicities
Joint TITE-CRM / TITE-BOIN-type designs	Bayesian time-to-event models with predictive decision rules	Explicitly model late-onset toxicity (and often efficacy) over multiple cycles	Simulation central to design: calibrates rules, evaluates safety/efficacy trade-offs, and tunes predictive thresholds	Modern immunotherapy/ targeted trials with late-onset safety and activity outcomes

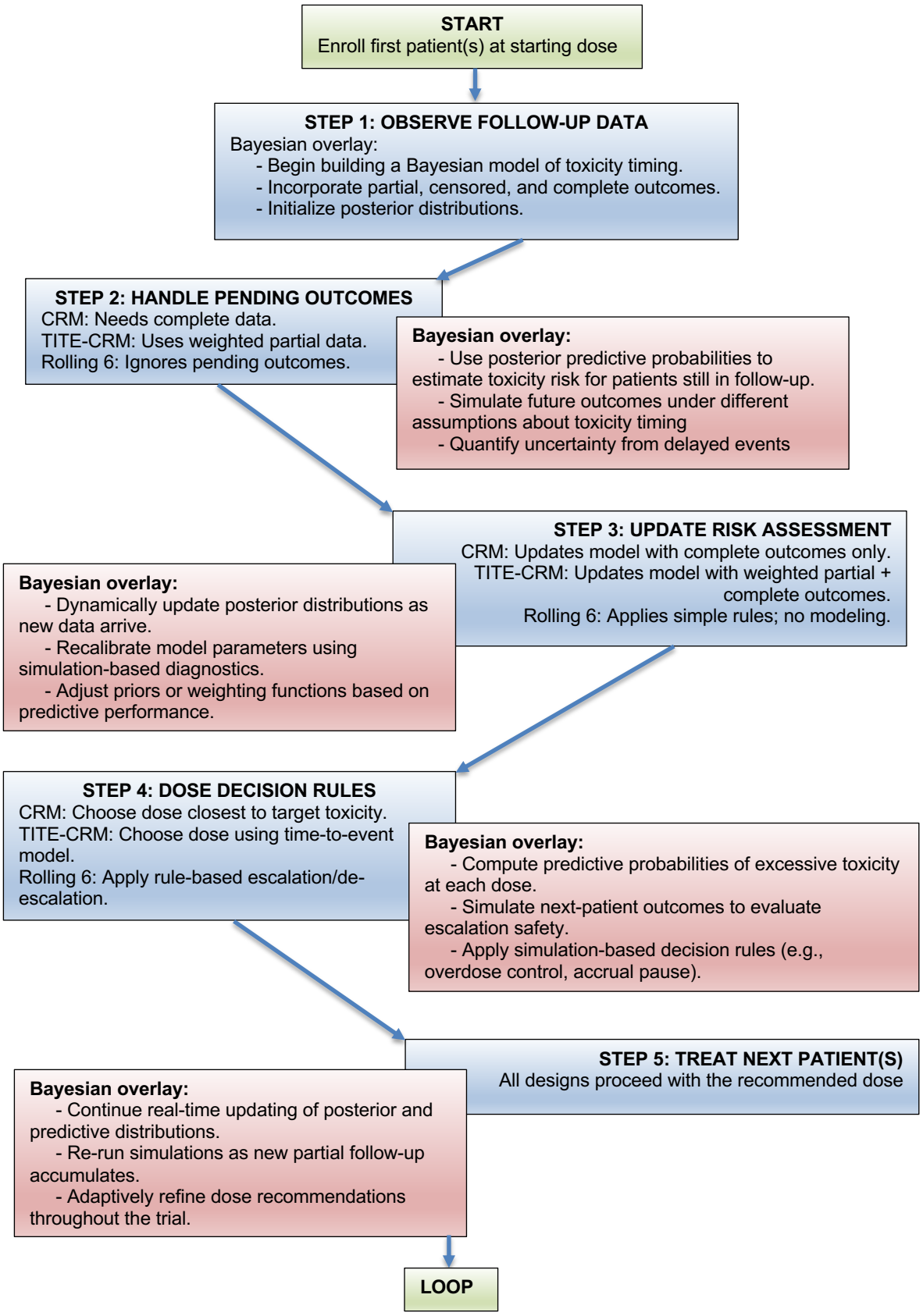
POSTERIOR PREDICTIVE PROBABILITY FRAMEWORK

Posterior predictive probabilities are an important part of Bayesian simulation methods. They estimate the chances of future toxicity outcomes based on the data we have, including censored or partially completed follow-up data. After we update a Bayesian dose-toxicity model with the latest information, we use the posterior predictive distribution to simulate possible outcomes if more patients receive treatment at each dose. This includes the probability of a dose-limiting toxicity (DLT), the risk that a dose surpasses a safety limit, or whether ongoing recruitment would violate toxicity control rules. This forward-looking aspect is particularly helpful in trials with late-onset toxicities, where many results are still pending, and traditional methods cannot safely guide dose escalation. Designs like TITE BOIN12 use posterior predictive probabilities to address delayed toxicity and effectiveness. They establish real-time decision rules that remain valid even when follow-up is incomplete. In contrast, broader Bayesian dose escalation frameworks rely on predictive simulations to ensure patient safety and maintain effective operating characteristics.

WHY THIS MATTERS FOR LATE-ONSET TOXICITY

Posterior predictive probabilities help researchers include uncertainty about outcomes that have not been observed. This makes them perfect for situations where dose-limiting toxicities (DLTs) might happen weeks or months after treatment starts. Instead of waiting for complete follow-up, the design uses predictive probabilities to estimate future risks and make informed decisions that prioritize safety. This ability is a key reason why Bayesian simulation methods perform better than traditional approaches when toxicities are delayed.

BAYESIAN SIMULATION OVERLAY ON DOSE-FINDING DESIGNS



ARTIFICIAL INTELLIGENCE (AI)-DRIVEN ADAPTIVE BAYESIAN MODELS FOR LATE-ONSET TOXICITIES

AI-driven adaptive Bayesian models improve traditional dose-finding methods by combining flexible machine learning components with Bayesian updating. This allows the dose-toxicity relationship to change as new data, including censored and late-onset events, is collected. Unlike TITE CRM, which relies on a fixed parametric form and a set weighting scheme for partial follow-up, AI-enabled Bayesian models can learn complex, nonlinear time-to-event patterns directly from the data. This makes them especially useful for modern oncology trials, where toxicities can happen across multiple cycles [1]. Three representative AI-driven Bayesian approaches show how dynamic parameter updating provides a stronger framework for managing late-onset toxicities.

BAYESIAN NEURAL NETWORK (BNN) TIME-TO-EVENT MODELS

Bayesian neural networks model the risk or overall chance of toxicity as a flexible function of dose, time, and patient characteristics. The approach begins by assigning prior distributions to network weights. These weights are updated through Bayesian inference as partial or full follow-up data from each patient becomes available. The model then generates predictions of toxicity probabilities for future patients. This framework can capture nonlinear and delayed risk patterns that traditional models cannot. Since the parameters are continuously updated, the model easily accommodates late-onset toxicity without fixed weighting assumptions. This flexibility reduces bias when the risk of toxicity increases late in the cycle or varies across different patient groups. These are scenarios where TITE CRM is known to be sensitive to errors.

GAUSSIAN PROCESS (GP) BAYESIAN HAZARD MODELS

Gaussian process models view the dose-time toxicity surface as a smooth, nonparametric function. This allows the hazard to develop in complex ways without needing a specific parametric form. The method uses a Gaussian process prior for the latent hazard function and updates the posterior as toxicity times and censored observations accumulate. It then uses the posterior to estimate predictive toxicity probabilities at each dose and time point [11,12]. These models are particularly helpful for late-onset toxicity. They can directly capture delayed, plateauing, or multi-phase hazard shapes from the data. This contrasts with the TITE CRM, which assumes a steady accumulation of information over the DLT window. It may misjudge risk when the timing of toxicity does not fit this assumption.

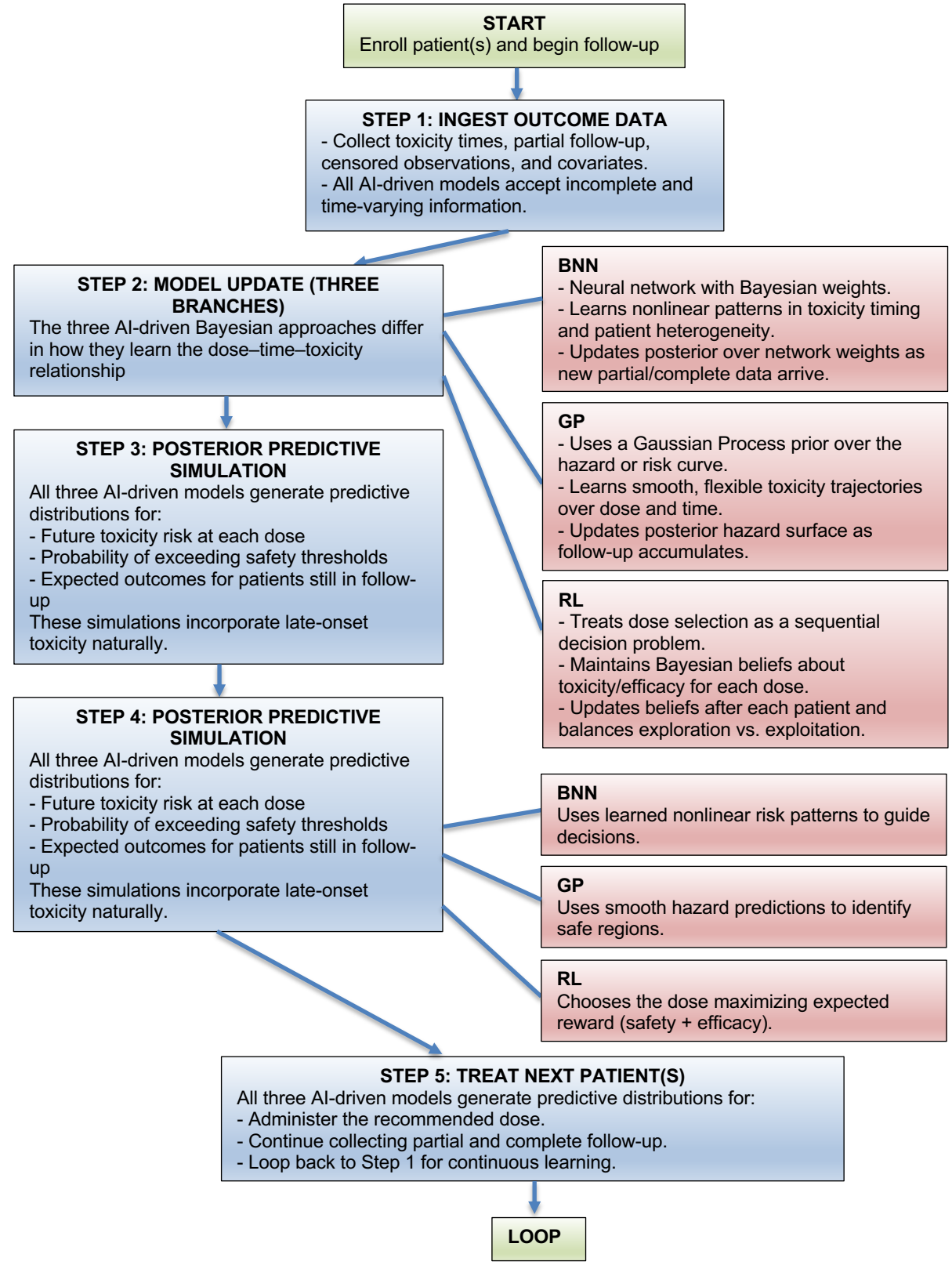
REINFORCEMENT LEARNING (RL) / CONTEXTUAL BANDIT DOSE-ESCALATION MODELS

Bayesian reinforcement learning methods treat dose selection as a step-by-step decision-making process. The algorithm maintains a belief about toxicity and possibly efficacy profiles, updating this belief as new outcomes are observed, whether they are complete or censored. At each interim, the model uses posterior predictive probabilities to evaluate the risk of future toxicity for each dose. It selects the next dose by balancing exploration and exploitation. The use of Bayesian multi-armed bandits in clinical dosing offers a way to optimize the balance between exploring different doses and exploiting the best estimated dose for treating patients. These models address late-onset toxicity by directly modeling time-to-event outcomes and conducting predictive simulations to estimate the chances of exceeding toxicity thresholds, even when many patients are still in follow-up. Unlike traditional dose-finding methods that aim to match a target toxicity rate using a parametric model, reinforcement learning frameworks can optimize broader goals by balancing safety and efficacy in a step-by-step decision-making process.

WHY THESE AI-DRIVEN BAYESIAN MODELS OUTPERFORM TITE-CRM FOR LATE-ONSET TOXICITY

Across all three approaches, the main advantage over TITE CRM is the ability to update flexible models based on data, rather than depending on a fixed parametric curve and predetermined weighting scheme. Model-based designs are prone to errors when toxicity timing varies across cycles [1], and AI-driven Bayesian models tackle this problem by learning the hazard function's shape from the data. Additionally, posterior predictive simulation, which is key to all three methods, has been shown to improve safety monitoring and overdose control in late-onset settings [2,5]. These models therefore provide a better framework for early phase trials, where toxicity develops gradually, accrual is rapid, and decisions must be made amid substantial uncertainty.

AI-DRIVEN BAYESIAN DOSE-FINDING WORKFLOW



CONCLUSION

The challenges of late-onset toxicities show the need for dose-finding strategies that work well when outcomes are delayed, censored, or only seen partially. Traditional methods like CRM, TITE-CRM, and Rolling 6 have made important progress, but their effectiveness relies on assumptions about when toxicities occur and how models are built. These assumptions may not apply to modern immunotherapies and multi-cycle targeted treatments. Bayesian simulation techniques have become essential, allowing researchers to assess operating characteristics under realistic accrual and follow-up patterns. They also help quantify uncertainty through posterior predictive probabilities and implement clear stopping and overdose-control rules.

AI-driven adaptive Bayesian models build on these techniques by providing flexible, data-based ways to learn the dose-time-toxicity relationship during the trial. Instead of replacing CRM or TITE-CRM, these models work alongside them to tackle their persistent weaknesses. Bayesian neural networks create a flexible framework for modeling complex nonlinear relationships. When used with survival data, these models can detect delayed hazard patterns that traditional parametric models might overlook. By doing this, AI-enabled Bayesian frameworks maintain the clarity and statistical strength of traditional designs while improving their adaptability to complex late-onset toxicity profiles.

Overall, combining AI, Bayesian simulation, and established dose-finding methods offers a practical and forward-thinking approach for early-phase oncology trials. These hybrid strategies improve the identification of safe and effective doses in situations where toxicity develops over long periods. They achieve this by dynamically updating parameters, using predictive modeling, and promoting transparent decision-making. As therapies continue to advance, the collaboration between traditional designs and AI-driven Bayesian methods will be crucial for ensuring patient safety and trial efficiency.

REFERENCES

1. Barnett H, Boix O, Kontos D, Jaki T. Joint TITE-CRM: A Design for Dose Finding Studies for Therapies with Late-Onset Safety and Activity Outcomes. Statistics in Biopharmaceutical Research. 2025;17(1):149–160.
2. Chen K, Chen TY, Zhang Y, Lin R, Yuan Y. Practical Considerations for Using the TITE-BOIN Design to Handle Late-Onset Toxicity or Fast Accrual in Phase I Trials. Clin Cancer Res. 2025;31(13):2573–2580.
3. Jin H, Yin G. Time-to-event calibration-free odds design: A robust efficient design for phase I trials with late-onset outcomes. Pharmaceutical Statistics. 2023;22(5):773–783.
4. Pan H, Yuan Y. Phase I Designs for Late-Onset Toxicity. In: Bayesian Adaptive Design for Immunotherapy and Targeted Therapy. Springer, Singapore; 2023. https://doi.org/10.1007/978-981-19-8176-0_2
5. Zhou Y, Lin R, Lee JJ, Li D, Wang L, Li R, Yuan Y. TITE-BOIN12: A Bayesian phase I/II trial design to find the optimal biological dose with late-onset toxicity and efficacy. Stat Med. 2022;41(11):1918–1931.
6. Barnett H, Boix O, Kontos D, Jaki T. Dose finding studies for therapies with late-onset toxicities: A comparison study of designs. Stat Med. 2022;41(30):5767–5788.
7. Fu C, Zhou S, Lee JJ. Posterior Predictive Design for Phase I Clinical Trials. Journal of the American Statistical Association. 2025;120(552):2646–2657.
8. Weber S, Widmer L, Bean A. OncoBayes2: Bayesian Logistic Regression for Oncology Dose-Escalation Trials. R package version 0.9-4; 2025.
9. Neal RM. Bayesian learning for neural networks. New York: Springer; 1996.
10. Chapfuwa P, Tao C, Li C, Page C, Goldstein B, Carin L, Henao R. Adversarial Time-to-Event Modeling. Proc Mach Learn Res. 2018;80:735–744.
11. Rasmussen CE, Williams CKI. Gaussian Processes for Machine Learning. MIT Press; 2005.
12. Fernández T, Rivera N, Teh YW. 30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.
13. Villar SS, Bowden J, Wason J. Multi-armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges. Stat Sci. 2015;30(2):199–215.
14. Zhou Y, Li R, Yan F, Lee JJ, Yuan Y. A Comparative Study of Bayesian Optimal Interval (BOIN) Design With Interval 3 + 3 (i3 + 3) Design for Phase I Oncology Dose-Finding Trials. Statistics in Biopharmaceutical Research. 2021;13(2):147–155.

CONTACT INFORMATION

(In case a reader wants to get in touch with you, please put your contact information at the end of the paper.)
Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Rachana Lele, Arfan Afzal
Company: Syneos Health
Email: rachana.lele@syneoshealth.com, arfan.afzal@syneoshealth.com
Website: Early Phase | Syneos Health