Cells Born to Kill:

Cellular-Level Modeling of CAR T-Cell Therapy in B-Cell Acute Lymphoplastic Leukemia

BME260L - Modeling Cellular and Molecular Systems

PBL #2 Written Report

By Brandon Aghnatios, Max Bazan, Ege Özemek, Ben Perry, Rijul Rajesh, James Yae

- December 2025 -

- Duke University -

Abstract

B-cell acute lymphoblastic leukemia (B-ALL) is an aggressive blood cancer defined by uncontrolled proliferation of immature B-cells and remains one of the most clinically significant pediatric malignancies [1]. Its high relapse rates and variable response to therapy motivate mechanistic models that clarify how CD19-directed chimeric antigen receptor (CAR) T-cells achieve control (or fail to do so) of the tumor cells. Clinical outcomes range from durable remission to relapse, B-cell aplasia, and cytokine release syndrome (CRS), prompting us to develop a model linking CAR T-cell dynamics to both efficacy and toxicity. We constructed an ordinary differential equation (ODE) framework that tracks malignant B-cells, effector and memory CAR T-cells, normal B-cells, and a lumped cytokine compartment. Tumor and normal B-cells follow logistic growth with CAR T-mediated killing. Effector CAR T-cells proliferate, differentiate into memory cells, and experience tumor-driven immunosuppression, while memory cells reactivate in proportion to residual tumor burden. Cytokines serve as a surrogate marker of CAR T activation and treatment-associated toxicity, allowing the model to capture inflammatory side-effects such as CRS. Baseline simulations showed rapid tumor clearance, sustained B-cell depletion, transient cytokine peaks, and memory- and time- driven control upon rechallenge. Parameter sweeps revealed thresholds in cytotoxic potency (K_{kill}), the importance of antigen-driven memory reactivation (θ), and the role of effector proliferative capacity (r_T) in shaping both tumor reduction and cytokine exposure. These results outline how tuning key parameters influences remission and toxicity in CD19-directed CAR T-cell therapy.

Overall, the dynamics of CAR-T expansion, contraction, and persistence closely parallel pharmacokinetic–pharmacodynamic (PK/PD) relationships, where effector proliferation acts analogously to drug exposure and cytotoxic potency corresponds to PD efficacy. Modeling these PK/PD-like behaviors is essential for predicting therapeutic response, optimizing dose, and evaluating the safety profile of engineered T-cell products. By framing CAR-T behavior within a PK/PD framework, we can more directly relate model parameters to clinically meaningful outcomes.

Background

B-Cell Acute Lymphoblastic Leukemia (B-ALL) is a cancer of precursor lymphoid cells, known as lymphoblasts. These leukemic blasts multiply in the bone marrow, spill into the blood, and crowd out normal blood-forming cells. This loss of normal blood production leads to anemia, infection risk, and bleeding complications. It is typically treated with chemotherapy, and in more severe or relapsed cases with CAR T therapy or stem cell transplant. Many children respond well to multi-agent chemotherapy, but a significant number relapse or never achieve long-term disease control. Outcomes also depend heavily on where a child receives care. Centers with access to intensive chemotherapy (like that of Duke, UCLA, and Johns Hopkins), good supportive care, and molecular monitoring have far better results, while children in many regions are diagnosed only after the disease has significantly progressed, leading to higher relapse rates and greater treatment-related mortality. In this setting, every new therapy becomes not only a scientific breakthrough but also a challenge for global equity [1].

We specifically chose to model B-cell ALL, characterized by the proliferation of malignant immature B-cells, rather than other ALL subtypes, such as T-cell ALL, as B-cells express the CD19 surface antigen, the target for CAR T-cells, while T-cells do not. Additionally, although ALL only accounts for about 1% of all cancers in the U.S. and primarily occurs in children ages 2 to 5, B-ALL is more prevalent, accounting for about 85% of childhood cases [1].

CD19-directed CAR T-cell therapy was developed as a rescue treatment for children with relapsed or refractory B-ALL. Autologous T-cells are collected, engineered to express a synthetic receptor that recognizes CD19, expanded outside the body, and infused back into the patient. Once inside the bloodstream, these CAR T-cells recognize malignant B-cells, activate, and proliferate. For many children, a single therapy infusion as a second or third line of treatment can clear detectable disease [2]. However, this success comes at a physiological cost. CAR T-cells eliminate healthy B-cells along with leukemic ones, causing long-term B-cell aplasia and loss of antibody production. In addition, rapid tumor killing triggers a wave of cytokines, such as IL-6 and IL-2, which can cause symptoms of CRS and related neurotoxicity including high fevers, hypotension, and neurologic effects that may require intensive care. These competing effects highlight the need for mechanistic models that can elucidate how CAR T-cells expand, eliminate tumor cells, and generate toxicity.

These clinical realities also make the ethical stakes unusually high. Families must decide whether to undergo a therapy that can potentially cure the disease but may also cause life-threatening toxicity and long-term immune suppression to the child. Clinicians and regulators must decide when to offer CAR T therapy, how to monitor it, and what risks are acceptable. We identified the need for a quantitative analysis: to precisely predict how CAR T therapy's dosage amount, timing, and frequency will impact both the efficacy in reducing B-cell tumor growth and determining when it might be dangerous.

Model Development

We developed a cellular population-level ODE model of CD19 CAR T-cell therapy for B-ALL, extending the CARTmath framework established by Barros et al. [3], [5]. Our model tracks five interacting biological populations: malignant CD19 tumor cells T(t), effector CAR T-cells $C_T(t)$, memory CAR T-cells $C_M(t)$, normal CD19 B-cells N(t), and a lumped cytokine signal Y(t). Our goal was to develop a model that was simple enough to be interpretable but also sufficiently technical to capture the three main clinical trajectories in CD19-targeted therapy for B-ALL: (1) whether the leukemic population is eliminated or regrows, (2) whether normal B-cells undergo prolonged aplasia or recovery, and (3) whether cytokine-driven toxicity emerges.

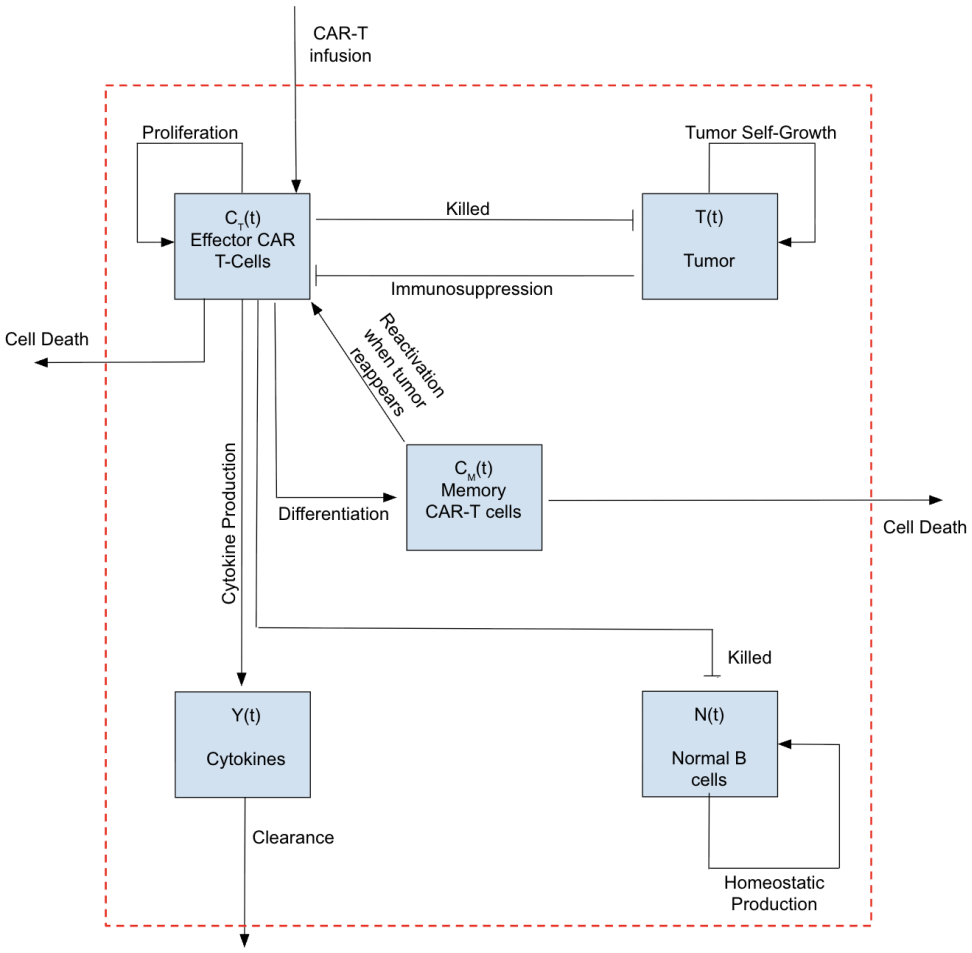

Figure 1. Compartment Model

Figure 1 summarizes the flow of cells and signals through the system. At the therapy time, a discrete infusion introduces effector CAR T-cells into the C_T compartment, initiating the therapeutic response. Upon encountering tumor antigen, these effectors proliferate, kill tumor cells, and produce cytokines. Part of the effector population differentiates into memory CAR T-cells (C_M), which persist at low levels and can later be reactivated if tumor cells reappear [6].

While targeting malignant B-cells, effector CAR T-cells simultaneously eliminate normal CD19 B-cells, producing the clinically characteristic period of B-cell aplasia. All populations eventually contract through natural death, tumor-induced immunosuppression, or cytokine clearance, completing the therapeutic cycle. This flow from infusion to proliferation to killing to contraction to long-term memory reproduces the qualitative dynamics observed in clinical CAR T therapy [6].

The interactions described above are encoded in the following five ODEs. The specific parameters, a description of what they represent, and their modeled values with units are also included for clarity.

$$\frac{dT}{dt} = r_C T\left(1 - \frac{T}{K_c}\right) - K_{kill} C_T T$$

Equation 1. Tumor Rate of Change Accounting Equation

Parameters	Description	Value	Units
r_c	Tumor Growth Rate	0.15	day^{-1}
K_c	Tumor CARrying Capacity	$5x10^{11}$	cells
K_{kill}	Killing Rate per CAR T Tumor Encounter	$6x10^{-8}$	$(cell*day)^{-1}$

Table 1. Tumor Rate of Change Accounting Equation parameters

The tumor logistic growth rate $r_c = 0.15\ day^{-1}$ and carrying capacity $K_C = 5x10^{11}$ cells were taken directly from Barros et al.'s CARTmath framework, which calibrated leukemic expansion to preclinical B-ALL data, demonstrating saturation-limited proliferation in the bone-marrow region. These values reflect the sub-exponential growth pattern of hematologic malignancies where nutrient and stromal constraints shape the upper bound of tumor burden. The CAR-T killing coefficient $K_{kill} = 6x10^{-8} cell^{-1} day^{-1}$ is also sourced from Barros et al., representing the effective per-cell cytotoxic potency required for CD19 CAR-T cells to achieve tumor control in vivo. Together, these tumor parameters define the competitive balance between leukemic expansion and CAR-T cytotoxic pressure and anchor the system in well-established kinetics for CD19-directed immunotherapy models.

In this ODE, Tumor dynamics are governed by a balance between intrinsic malignant proliferation and CAR T-mediated clearance. The tumor population T(t) grows logistically at rate r_C toward a carrying capacity K_C. This represents rapid expansion early on and slowing growth as the bone marrow becomes crowded and reaches the tumor carrying capacity. This growth is countered by effector CAR T killing, modeled as a mass-action term proportional to both tumor and effector abundance, with K_{kill} representing the killing rate per CAR T tumor encounter [3]. This is similar to a second-order kinetic term, however we aren't directly modeling specific binding kinetics. Together, these processes allow the tumor to either expand, contract, or reach equilibrium depending on the strength of the CAR T response [6].

$$\frac{dC_T}{dt} = r_T C_T\left(\frac{T}{T + K_T}\right) - d_T C_T - \epsilon C_T - \alpha_T C_T + \theta T C_M$$

Equation 2. Effector CAR T-cell Rate of Change Accounting Equation

Parameters	Description	Value	Units
r_T	Max effector proliferation rate	0.38	day^{-1}
K_T	Tumor level at which proliferation is half-max	$5x10^5$	cells
d_T	Effector death/exhaustion rate	0.2	day^{-1}
ϵ	Effector → memory conversion rate	0.1	day^{-1}
α	Tumor-induced immunosuppression	$4x10^{-13}$	$(cell*day)^{-1}$
θ	Reactivation rate of memory to effector by tumor	$2x10^{-7}$	$(cell*day)^{-1}$

Table 2. Effector CAR T-Cell Rate of Change Accounting Equation parameters

The effector proliferation rate $r_T = 0.38\, day^{-1}$, half saturation constant $K_T = 5x10^5$ cells, death rate $d_T = 0.2\, day^{-1}$, effector-memory conversion rate $\epsilon = 0.1\, day^{-1}$, and tumor-induced immunosuppression $\alpha = 4x10^{-13}(cell * day)^{-1}$ do not exactly match the

values used in Barros et al. or Serrano et al. In the literature, effector proliferation is typically somewhat slower, and exhaustion parameters tend to be smaller. However, after extensive simulation testing, we selected these specific values because they produced biologically realistic expansion–contraction dynamics within our model: CAR-T cells undergo a sharp proliferative burst in the presence of antigen, contract after clearance, and stabilize at a small memory-like plateau.

The memory-reactivation parameter $\theta = 2x10^{-7}(cell * day)^{-1}$ also differs from published values. Our slightly lower value was chosen because it best reproduced a modest but detectable recall response during rechallenge without causing unrealistically large secondary expansions.

Effector CAR T-cells, $C_T(t)$, act as the main tumor-killing population. Effector cells expand at a maximal rate r_T when antigen is abundant, scaled by an activation function that depends on tumor burden and the half-maximal response level K_T. Effector cells are removed through natural death and exhaustion at rate d_T, and a fraction ϵ differentiates into memory CAR T-cells. Tumor-induced immunosuppression acts through a removal term $\alpha_T C_T$, representing inhibitory signals such as PD-L1 expression or nutrient deprivation. Finally, memory CAR T-cells can be recruited back into the effector state at rate θTC_M, capturing antigen-driven reactivation central to long-term control [5], [6].

$$\frac{dC_M}{dt} = \epsilon C_T - \theta T C_M - \mu C_M$$

Equation 3. Memory CAR T-cell Rate of Change Accounting Equation

Parameters	Description	Value	Units
μ	Memory CAR T death rate	0.003	day^{-1}

Table 3. Memory CAR T-Cell Rate of Change Accounting Equation parameters

The memory-cell decay rate $\mu = 0.003\ day^{-1}$ is somewhat higher than typical literature estimates. Published CAR-T memory populations exhibit long-lived persistence, sometimes spanning years, but using slower decay rates in our model caused memory compartments to accumulate excessively or plateau unrealistically high. The chosen value reproduced a slow, biologically plausible decline while ensuring that memory cells remained in a realistic numerical range over the simulation.

Memory CAR T-cell dynamics, described by $C_M(t)$, capture the slower, persistent arm of the CAR T response. Memory cells arise from effector differentiation at rate ϵ, are depleted when

reactivated by tumor antigen through the same θTC_M mechanism, and decay gradually at a natural death rate μ. This structure reproduces the clinically observed pattern in which a small but durable memory pool remains after the primary effector contraction and can expand upon leukemic relapse.

$$\frac{dN}{dt} = r_N N(1 - \frac{N}{K_N}) - K_{off} C_T N$$

Equation 4. Normal B-cell Rate of Change Accounting Equation

Parameters	Description	Value	Units
r_N	B-cell homeostatic growth rate	0.01	day^{-1}
K_N	Normal B-cell carrying capacity	$1x10^9$	cells
K_{off}	CAR T killing rate of normal B-cells	$6x10^{-8}$	$(cell*day)^{-1}$

Table 4. Normal B-Cell Rate of Change Accounting Equation parameters

The B-cell homeostatic growth rate $r_N = 0.01\ day^{-1}$ and carrying capacity $K_N = 1x10^9 cells$ represent a simplified logistic model of peripheral B-cell turnover. These values are higher than those used in Serrano et al., but they allowed the model to reproduce two key biological behaviors: (i) physiologically reasonable steady-state B-cell counts prior to therapy, and (ii) deep, prolonged B-cell depletion following CAR-T infusion. CAR-T killing rate of normal B-cells $K_{off} = 6x10^{-8}(cell * day)^{-1}$ was set equal to the tumor-killing rate for simplicity. Although in vivo killing kinetics may differ between malignant and healthy $CD19^+$ targets, using the same coefficient preserved the qualitative phenomenon of long-lasting on-target, off-tumor B-cell aplasia. These simplified parameters were therefore chosen to align normal B-cell dynamics with expected clinical behavior while maintaining numerical stability [5], [6].

$$\frac{dY}{dt} = \alpha_T C_T - K_Y Y$$

Equation 5. Cytokine Level Rate of Change Accounting Equation

Parameters	Description	Value	Units
α_T	Cytokine production rate per effector CAR T	$2x10^{-13}$	(conc*cell^{-1}*day^{-1})
K_y	Cytokine clearance rate	0.5	day^{-1}

Table 5. Cytokine Levels Rate of Change Accounting Equation parameters

The cytokine production rate $\alpha_T = 2x10^{-13}(conc * cell^{-1} * day^{-1})$ and clearance rate $K_y = 0.5\ day^{-1}$ form a coarse-grained toxicity proxy rather than a mechanistic cytokine model. In the original CARTmath publication, cytokine parameters differ in scale and units because the model aggregates multiple inflammatory mediators (for example, IL-6 and IFN-γ) into a single abstract variable. Our chosen values were tuned to ensure that cytokine peaks tracked effector expansion quantitatively, producing modest cytokine bursts during successful clearance and larger peaks during over-expansion, consistent with clinical CRS trends. Although not directly taken from Barros et al., these values maintain the intended qualitative relationship between effector mass and inflammatory toxicity [4].

To achieve model simplicity while preserving clarity, we incorporated several simplifying assumptions that define the scope of our model. We treated the body as a single well-mixed compartment representing blood, marrow, and lymphoid tissues, enabling us to focus on population-level interactions without modeling spatial distribution, trafficking, or microenvironmental niches. We treated CD19 expression as constant on the time scale of the simulation and did not model antigen negative escape or trogocytosis. Binding, synapse formation, and intracellular signaling are assumed to be fast relative to cell population dynamics and therefore appear only as effective rate constants in killing, proliferation, and immunosuppression terms [5].

Furthermore, certain assumptions were made within specific compartments of the model. Cytokines are lumped into a single variable, so we do not distinguish IL-6 from IL-2 or capture organ specific toxicities such as neurotoxicity. While cytokines such as IL-2 can enhance effector CAR T-cell expansion, we chose to exclude cytokine-driven proliferation feedback to limit model complexity and parameter uncertainty. In our formulation, cytokines serve as a readout of CAR-T activation rather than a regulatory factor, allowing us to focus on effector-memory dynamics, tumor clearance, and immunosuppression. It also allows us to group cytokines into

one population rather than tracking the specific cytokines that matter in effector cell proliferation (notably IL-2 and IL-15). Further, this simplification also removes a nonlinear term from the positive feedback loop which could destabilize the system. Effector and memory CAR T-cells are each treated as homogeneous populations, even though in reality they comprise diverse phenotypes. Normal B-cell regeneration is modeled as logistic growth without explicit conditioning related damage. Overall, these are reasonable simplifying assumptions because our model still captures the major effector cell expansion wave, contraction, memory differentiation, and tumor immunosuppression dynamics.

Although the model includes many parameters, we focus on three that map directly to engineering or therapeutic decisions. The cytotoxic efficacy parameter (K_{kill}) scales how quickly CAR T-cells can kill tumor cells and reflects features such as CAR affinity and signaling strength. The memory reactivation rate (θ) captures how efficiently memory cells convert back into effector cells when tumor returns. The effector proliferation rate (r_T) determines how rapidly the CAR T population expands after activation and depends on antigen exposure and the state of the infused cells. By varying these three parameters while keeping everything else within reasonable biological ranges, we can examine how killing strength, persistence, and expansion shape outcomes of interest including tumor control, B-cell aplasia, and cytokine burden.

We tuned a baseline parameter set so that a single simulated infusion reproduced key features of CD19 CAR T therapy. Effector cells expanded quickly, leukemic blasts dropped by several orders of magnitude, and normal B-cells were depleted for a long period. Cytokine levels rose while tumor burden and effector counts overlapped, then fell once the tumor was cleared, producing a short-lived inflammatory peak. A memory plateau persisted long after the effector wave contracted, and when tumor cells were reintroduced months later, the memory population generated a second effector burst that prevented uncontrolled relapse [6]. These behaviors align with clinical observations of initial remission, B-cell aplasia, and memory-mediated protection. By framing the system this way, the model helps us ask quantitative ethical and engineering questions. How strong should CAR-mediated killing be to ensure clearance without producing dangerous cytokine levels? How much memory reactivation is needed to prevent relapse without causing chronic inflammation or prolonged B-cell depletion? How aggressively should we drive effector proliferation when higher expansion improves early control but worsens cytokine exposure? These questions matter both for individual patients and for health systems determining how to design, monitor, and allocate CAR T therapies.

By formalizing the interactions between tumor cells, effector CAR T-cells, memory CAR T-cells, normal B-cells, and cytokines, our model provides a structured way to explore these questions instead of relying solely on clinical intuition. At the cellular level, it highlights which CAR design choices are most important for balancing efficacy and safety. At the global level, better mechanistic understanding may eventually support manufacturing and dosing strategies that are less dependent on trial-and-error, potentially reducing costs and improving accessibility. In the sections that follow, we outline the major results of our model and how our findings can be implemented to improve CAR T-cell therapy for treating B-ALL.

Model Results

In the baseline scenario, our model produced the multi-phase behavior of CD19 CAR T therapy against B-cell leukemia while explicitly tracking effector CAR T-cells, memory CAR T-cells, normal B-cells, and a cytokine compartment.

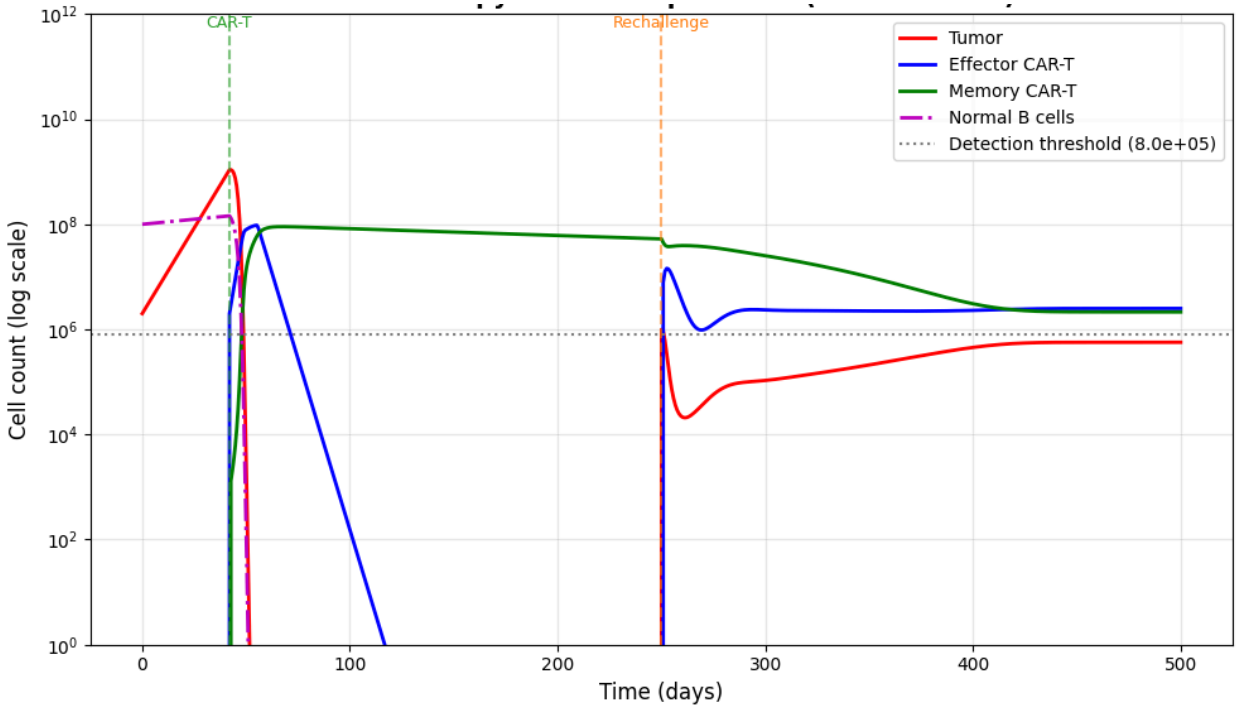

Figure 2 - CAR T Therapy: All Cell Populations. This graph tracks the cell count (Tumor, Effector CAR T, Memory CAR T, and Normal B-cells) over time following CAR T infusion and subsequent tumor rechallenge, illustrating the fixed model's dynamic response.

As visualized in Figure 2, tumor cells were initialized just above the detection limit at roughly 10^6 cells and increased to a burden of approximately 10^9 cells over the 40 days preceding infusion. At day 40, a single bolus of $2x10^6$ effector CAR T-cells was administered. This effector pool did not remain fixed; instead, it underwent substantial in vivo expansion, rising from the infused dose to a peak of roughly 10^8 cells around days 50-55 (a 45-50-fold expansion) before contracting over the following weeks. During this primary response, tumor burden fell rapidly. Specifically, the leukemic population dropped below the detection threshold of $8x10^5$ cells within about 2-3 days after infusion and declined by eight to nine orders of magnitude by day 50, approaching functional eradication. In this baseline parameter regime, a single infusion was sufficient to achieve remission. Figure 9, included in Appendix A, is a clean visual of only the tumor cell population and detection threshold which clearly shows these dynamics.

Effector CAR T-cell contraction was followed by the emergence of a long-lived memory compartment. Memory CAR T-cells rose slightly more gradually, reaching a peak of roughly $9x10^7$ cells around day 60 before declining slowly over the remainder of the simulation. From

that peak, the memory population decreased to about $7x10^7$ cells by day 250 and to approximately 10^6 cells by day 500, providing sustained immune surveillance. When a standardized leukemic rechallenge of 10^6 cells was introduced at day 250, tumor burden briefly rose just above the detection threshold before the memory pool rapidly reactivated into effector cells. This produced a secondary effector expansion of roughly 10^7 cells, about an order of magnitude smaller than the primary peak, and suppressed the tumor to a nadir of about $2x10^4$ cells within days. As memory continued to decay, the tumor gradually drifted back toward the detection threshold by day 500, stabilizing in the $5\text{-}8x10^5$ range but never regaining macroscopic disease, demonstrating long-term, though ultimately waning, immune pressure. Figure 10, included in Appendix A, provides a magnified view of this effector and memory dynamics interplay.

Normal B-cell dynamics properly mirror clinical expectations for CD19-directed therapy [5]. Figure 2 shows a Normal B-cell at an initial homeostatic level of approximately 10^8 cells, B-cell counts rose slightly prior to infusion and then dropped sharply after CAR T exposure. Within about 10 days, levels fell by 6-7 orders of magnitude. Levels reached roughly $10^4\text{-}10^6$ total B-cells and remained profoundly suppressed for the entire 500-day simulation due to the persistent memory CAR T-cell pool. This is individually illustrated in Figure 11 included Appendix A.

The cytokine compartment showed a sharp primary pulse, rising rapidly after infusion and peaking at roughly $2.3x10^3$ units near the effector maximum before declining over 2-3 weeks. The rechallenge generated only a minor secondary cytokine increase, reflecting the smaller secondary effector burst and rapid elimination of the recurrent tumor. Figure 12, included in Appendix A, plots this individual signal compartment.

To understand how tunable CAR T properties shape these outcomes, we performed parameter sweeps over cytotoxic efficacy (K_{kill}), memory reactivation (θ), and effector proliferation (r_T). Figures 12-14, included in Appendix A, demonstrate that varying K_{kill} reveals a clear threshold separating treatment failure from reliable cure. At low K_{kill}, tumor persisted near its carrying capacity of roughly 10^{11} cells, and CAR T-cells failed to engage either malignant cells or normal B-cells, resulting in negligible cytokine release. As K_{kill} increased into an effective range of about 10^{-8} to 10^{-6}, the model transitioned abruptly to successful clearance. Time to fall below the detection threshold shortened from roughly 10 days at the weakest effective K_{kill} to about 3 days at the strongest. Across this sweep, the peak effector count decreased from about $5x10^8$ cells when killing was weak to roughly $7x10^6$ cells when killing was strong, showing that improved per-cell potency reduced the expansion needed for tumor elimination. Cytokine peaks contracted in parallel, dropping from about $1.7x10^4$ units at low K_{kill} to about $2\text{-}3x10^2$ units at high K_{kill}. Thus, K_{kill} determined whether tumor could be cleared, how quickly clearance occurred, and the level of inflammatory burden accompanying the response.

Figures 17-19, included in Appendix A, show how max effector proliferation rate, r_T, exerted far stronger control over the amplitude of the CAR T response. As r_T increased from 0.4

to 1.2, time to clearance dropped from about 6.9 days to a minimum plateau near 2.8 days. Meanwhile, peak effector numbers expanded dramatically, increasing from ~10^6 cells at low r_T to nearly 10^{12} cells at the highest r_T tested, representing an increase of four to five orders of magnitude. Cytokine peaks rose in parallel from about 2.3 x 10^3 to 3.1 x 10^3 units, with a modest dip at intermediate r_T and a slight decline again at the very highest values. Thus r_T governed the tradeoff between rapid tumor elimination and the inflammatory and on-target/off-tumor costs of large-scale T-cell expansion.

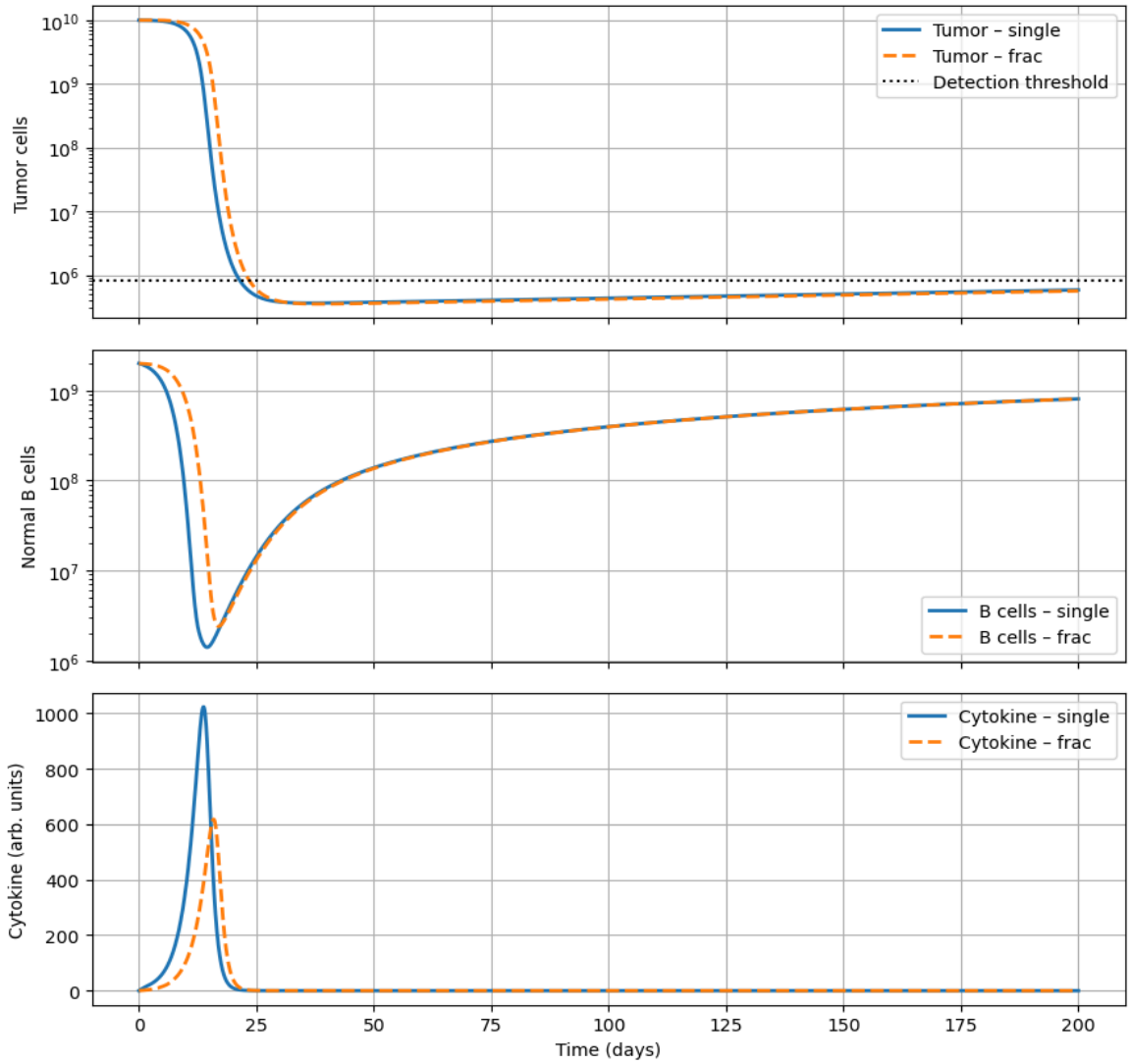

Figure 5 - Single vs. Fractionated CAR T dosing (same total dose). A fractionated CAR T dose produces nearly identical tumor clearance, B-cell depletion, and cytokine dynamics as a single bolus while modestly reducing the peak inflammatory response.

We next compared single versus fractionated dosing while holding total CAR T dose constant, as seen in Figure 5. For tumor and B-cells, the two regimens were nearly

indistinguishable: both drove tumor below the detection threshold by 6-8 days and induced a B-cell nadir of 2-3 times 10^6 cells before allowing gradual recovery to near-baseline by day 200. Their only meaningful difference appeared in cytokine dynamics. A single bolus produced a sharp cytokine peak of 950-1000 units around day 10-12, whereas fractionated dosing produced a substantially lower peak of ~550-600 units spread over a similar interval. Fractionation thus preserved antitumor efficacy and normal B-cell depletion while reducing peak inflammatory burden by 35-45%.

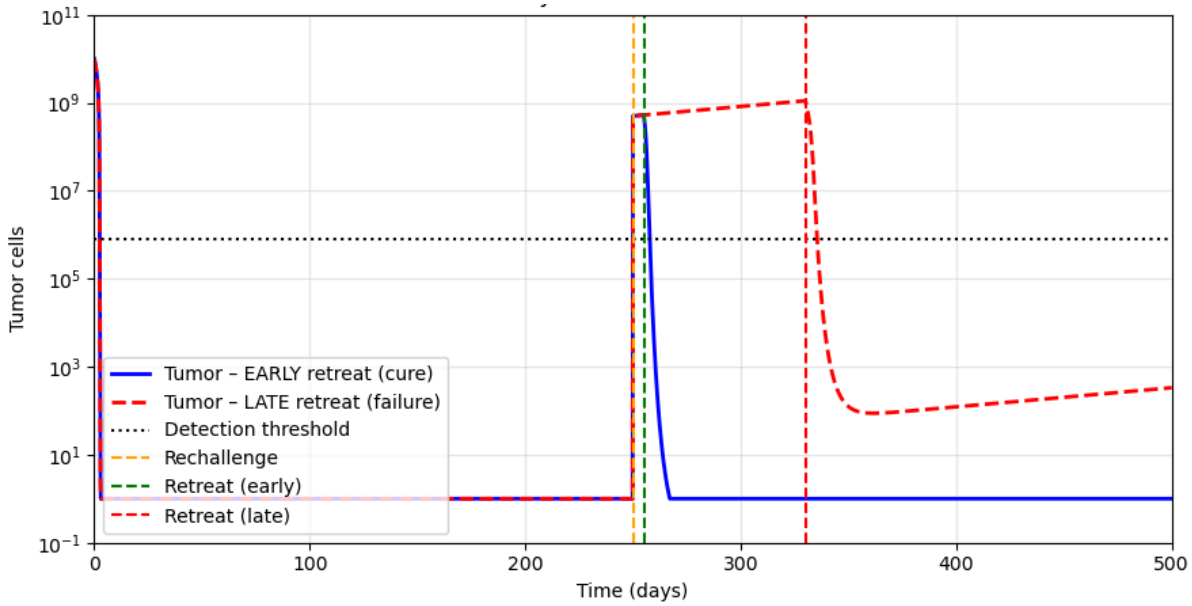

Figure 6 - Early vs. Late Retreatment After Tumor Rechallenge. Early retreatment after tumor rechallenge rapidly restores control and achieves cure, while delayed retreatment allows the tumor to regrow beyond the CAR T response capacity.

In Figure 6, we evaluate how timing of retreatment affects rechallenge control. We compared early versus late second infusions following a leukemic reinoculation at day 250. Early retreatment at around day 260 encountered only a small tumor (10^6-10^7 cells) and drove it to effective eradication within 10-15 days, sustaining long-term control. Late retreatment at ~day 340 faced a fully regrown tumor (10^9-10^{10} cells). Although the second infusion sharply reduced the tumor, it achieved only a partial response, reaching a nadir of roughly 10^2 to 10^3 cells before tumor began rising again to between 10^4 to 10^5cells by day 500. The identical second dose therefore cured early but failed when administered late, highlighting the steep dependence of success on pre-treatment tumor burden.

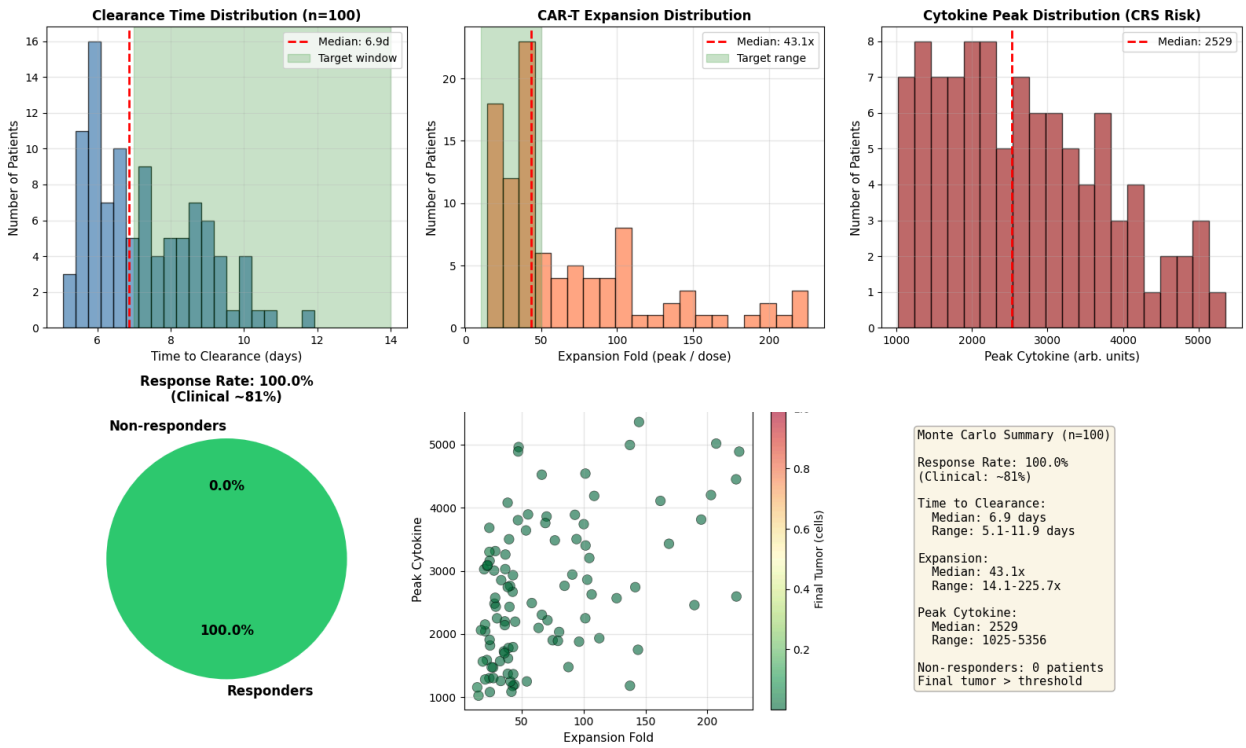

Figure 7 - Monte Carlo Patient Heterogeneity Simulation (n=100). Monte Carlo Summary with response rate, time to clearance, expansion, and peak cytokine.

To examine heterogeneity across patients, a Monte Carlo simulation of 100 virtual individuals was performed in Figure 7. The distribution of clearance times centered around a median of 6.9 days and a range of 5.1-11.9 days. CAR T expansion folds varied widely, with a median of 43.1x and a range of 14.1x to 225.7x. Peak cytokine levels ranged from ~1025 to 5356 units with a median of 2529. All patients in this baseline parameterization achieved sub-detection tumor levels, yielding a simulated response rate of 100%. Scatter analysis revealed a positive relationship between expansion and cytokine toxicity, with the highest-expanding patients typically experiencing the highest inflammatory peaks.

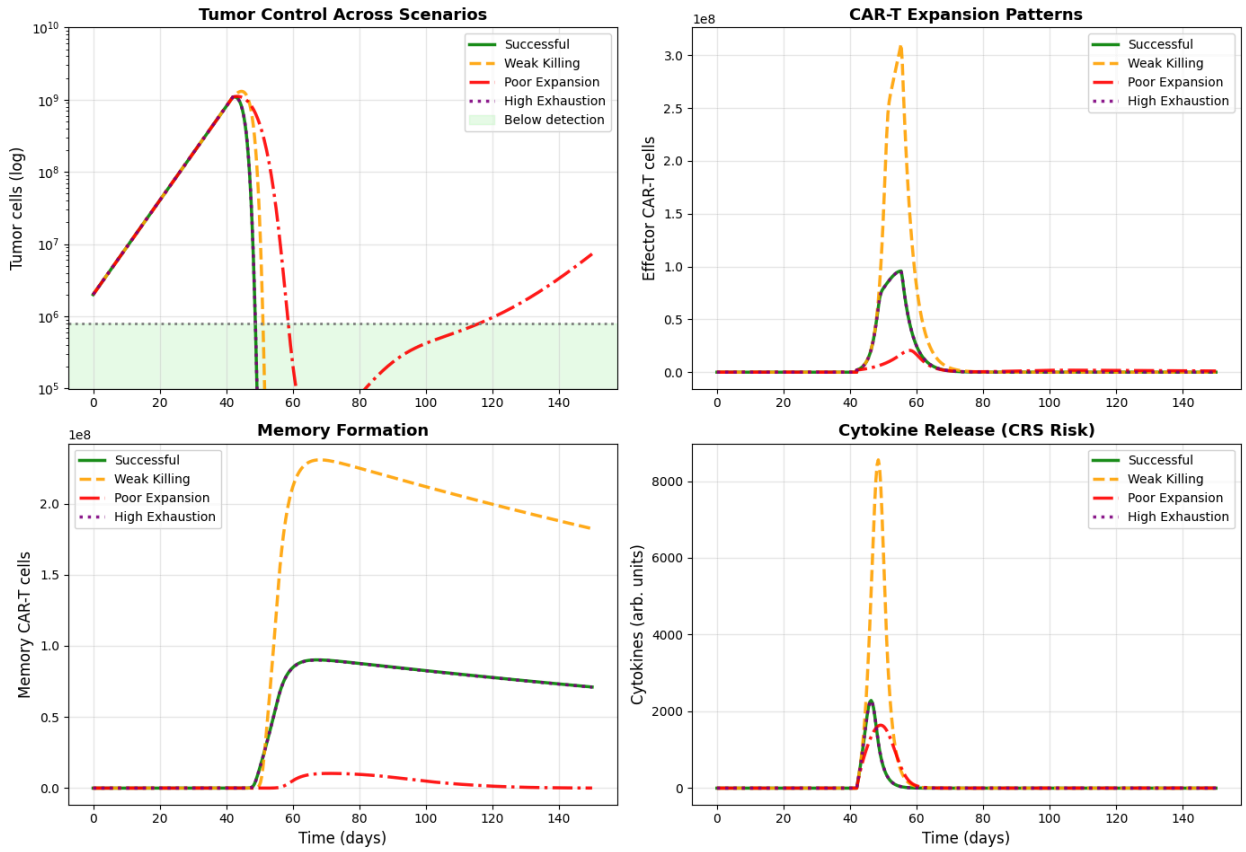

Figure 8 - Across success and failure modes, tumor trajectories, CAR T expansion, memory formation, and cytokine release each reveal the distinct signatures of strong killing, weak killing, poor expansion, and high exhaustion, illustrating how specific design limitations produce characteristic breakdowns in CAR T therapy.

Finally, four archetypal scenarios, plotted in Figure 8, illustrated distinct mechanistic pathways to success or failure. The successful case showed rapid clearance, effector peaks around 10^8 cells, a robust memory plateau of nearly 10^8 cells, and a moderate cytokine peak of approximately 2000 units. Weak killing required massive proliferation (effector peaks > $3x10^8$) to compensate and produced a very large cytokine surge (8000-8500 units), but ultimately still achieved clearance. Poor expansion achieved only transient suppression, with effector peaks near 10, minimal memory formation, and relapse to approximately 10^7 tumor cells by day 150. High exhaustion resulted in minimal expansion, small memory compartments, and persistent tumor around 10^{7-8} cells despite only moderate cytokine levels (2000-2500 units). These representative trajectories captured clinically recognizable behaviors: durable remission, cure with high toxicity, delayed relapse after partial response, and early primary resistance.

Taken together, these results show that our extended CARTmath model produces a rich and quantitatively detailed map. The simulations illustrate specific design limitations between CAR T intrinsic properties, dosing strategies, host variability, and clinical outcome patterns. The simulations reveal distinct roles for cytotoxicity, expansion, and memory dynamics and

demonstrate how modifying any one of these features reshapes tumor control, cytokine burden, and long-term immune surveillance.

Discussion

The model was developed to quantitatively identify which engineering and clinical parameters in a CD19 CAR T-cell product determine whether treatment culminates in durable remission, relapse, or toxicity. Using a single well-mixed compartment with a constrained set of interacting cell populations, the extended CARTmath framework reproduced the key qualitative patterns observed in CD19 CAR T therapy and then mapped how cytotoxic potency, proliferative capacity, memory responsiveness, dosing choices, and retreatment timing divide the therapeutic landscape into success, failure, and high-toxicity regimes. Baseline simulations provided a reference trajectory: a single 2×10^6-cell infusion expanded roughly 50-fold in vivo, cleared an initial leukemic burden of 10^6 cells within about a week, generated a persistent memory pool capable of suppressing a standardized rechallenge, and produced sustained B-cell aplasia accompanied by a single, sharp cytokine pulse. Parameter sweeps showed that this trajectory occupies only a narrow region of parameter space; modest shifts in K_{kill}, θ, r_T, or dosing strategy can redirect the system toward relapse or, conversely, into regions of excessive inflammatory toxicity.

Among all inputs, the per-cell cytotoxic efficacy K_{kill} emerged as the most decisive parameter. Simulations demonstrated a true potency threshold: when K_{kill} was reduced several-fold below baseline, leukemic proliferation outstripped CAR T killing and drove tumor burden toward its logistic ceiling of 10^{11} cells. In these weak-potency regimes, B-cell and cytokine dynamics changed little because effectors failed to generate enough productive engagements to meaningfully perturb the system. Once K_{kill} entered the efficacious band, the system transitioned abruptly into a cure regime. Tumor burden dropped by eight to nine orders of magnitude, and the time required to fall below the detection threshold compressed from roughly ten days at low-potency but still-effective values to about three days at high potency. Importantly, rising K_{kill} did not generate larger CAR T expansions. On the contrary, peak effector counts fell from approximately 5×10^8 to about 7×10^6 cells as K_{kill} increased, and cytokine peaks fell from roughly 1.7×10^4 to low hundreds. In other words, higher potency produced faster and cleaner tumor clearance with substantially less inflammatory burden.

Biologically, K_{kill} can be interpreted as a composite of antigen-binding affinity, synapse formation, intracellular killing efficiency, and downstream execution pathways. The model distilled two practical clinical principles from this structure. First, if composite cytotoxic potency is too low, neither dose escalation nor schedule manipulation can compensate; tumor growth will eventually outpace killing. Second, once K_{kill} surpasses its threshold, rapid leukemic and B-cell clearance can be achieved with modest expansion and limited cytokine exposure. This has meaningful ethical implications. High-potency products could reduce inflammatory toxicity and thus increase safety in low-resource settings, but they also render deep, durable B-cell depletion nearly unavoidable. This effectively commits patients to long-term immunoglobulin

replacement: something many health systems cannot reliably provide. The model does not resolve the trade-off between maximizing antitumor power and preserving feasibility in resource-limited environments, but it makes the constraints explicit and forces consideration of how potency interacts with sustained care requirements.

While K_{kill} governs the swiftness and decisiveness of early tumor elimination, memory reactivation θ exerts a more delayed and context-dependent influence. Across several orders of magnitude, variation in θ had minimal effect on primary leukemic clearance, which consistently occurred on a roughly seven-day timescale dominated by the initial effector population and K_{kill}. θ became influential only when the system encountered recurrent antigen. Higher θ values produced a more responsive memory compartment; on rechallenge, memory cells converted to effector cells more rapidly, generating slightly higher secondary peaks and a higher long-term memory plateau with only modest increases (about 1%) in peak cytokines. Lower θ values dampened recall responses, producing smaller secondary expansions and a reduced memory reservoir. Tumor control was still achieved under the standardized rechallenge conditions, but the safety margin narrowed, and the system became more sensitive to perturbations.

Biologically, θ encapsulates how efficiently memory cells detect and respond to antigen, shaped by differentiation state, manufacturing conditions, and the host environment. Clinically, high-income healthcare systems often favor maximizing memory responsiveness to guard against minimal residual disease or repeated antigen exposure. The model supports this logic in controlled environments but also highlights a risk: in settings lacking frequent monitoring and rapid access to toxicity management, excessively high θ could cause minor fluctuations in normal B-cell levels to trigger unnecessary cytotoxic flares. Thus, while high θ may be desirable in centers capable of managing recurrent inflammatory events, it may be inappropriate in environments without such infrastructure.

Effector proliferation r_T was the parameter that most directly bridged antitumor efficacy and inflammatory risk. Increasing r_T from 0.4 to 1.2 shortened clearance time from roughly three days to about three days (a small change), but this acceleration was purchased at the cost of massive increases in peak effector counts: from 10^6-10^7 cells at low r_T to nearly 10^{12} cells at high r_T, a four- to five-log rise. Cytokine peaks rose proportionally, increasing from about 2.3×10^3 to over 3×10^3 units and sustaining these levels for longer durations. These patterns reflect clinical experience: constructs featuring stronger costimulatory domains or metabolically “fit” phenotypes often undergo extremely large in vivo expansions, which improve clearance but raise CRS risk. The model formalizes this relationship as a quantitative trade-off: higher r_T increases the likelihood of both primary and recall responses fully sterilizing malignant cells, but it also raises the ceiling of inflammatory toxicity and prolongs B-cell aplasia. In practice, r_T is shaped by construct design, manufacturing features, and lymphodepletion regimens; ethically, it operates as a design choice that determines the minimum level of clinical infrastructure required to administer the therapy safely.

Clinical decisions that leave the construct unchanged, primarily dosing amount and timing, emerged as equally influential. Simulations comparing single-dose and fractionated

dosing showed that splitting a fixed total dose across several days preserved tumor clearance and B-cell depletion but reduced peak cytokine levels by 35-45%. This represents a sizable safety benefit without sacrificing efficacy. The result also aligns with clinical trends and identifies a scalable policy lever: in settings where managing severe CRS is difficult, fractionated dosing should be strongly considered as the default.

Retreatment timing revealed an even sharper dependency. If the second infusion was administered when recurrent tumor burden remained in the 10^6-10^7 range, the system reliably achieved a second cure. When the same infusion was delayed until tumor burden reached 10^9-10^{10} cells, only partial debulking occurred and slow regrowth followed. This mechanistic asymmetry has practical consequences: delays related to insurance, geographic barriers, or manufacturing bottlenecks are not mere logistical inconveniences: they shift patients into parameter regions where identical therapies perform worse. The model therefore exposes a structural inequity: health systems unable to guarantee timely retreatment will systematically disadvantage certain patients.

Dosing, across all simulations, emerged as the clinical variable that exposes the three-way tension among efficacy, toxicity, and equitable implementation. Increasing total dose accelerates clearance by increasing the initial pool of effector cells, but it also amplifies peak expansion and cytokine burden. When K_{kill} is below its potency threshold, dose escalation does not rescue efficacy but still increases inflammatory risk. Conversely, when K_{kill} lies within an effective range, dose can be reduced or fractionated with minimal loss of efficacy, producing the 35-45% reduction in cytokine peaks that many clinical centers now prioritize. This reframes dosing not merely as dosage optimization but as an ethical decision about which patients and which health systems can access CAR T therapy safely.

Monte Carlo simulations further highlighted that even near an optimally calibrated baseline, patients are not interchangeable. Variability in K_{kill}, r_T, θ, and related parameters produced clearance times ranging from 5.1 to 11.9 days, expansions from roughly 14× to 225×, and cytokine peaks spanning 1,000 to over 5,000 units. Every virtual patient achieved tumor eradication in this baseline configuration, but the pathways differed markedly. Some cleared disease with moderate expansions and cytokine exposure; others required much larger expansions and experienced substantially higher inflammatory peaks. This diversity echoes clinical trends in which high expansion correlates with greater CRS risk. These distributions matter because averages obscure the tail risks that drive real-world morbidity. The model therefore provides a way to predict which subpopulations may experience dangerous expansions and suggests targeted mitigation strategies such as pre-emptive CRS prophylaxis, disease-burden thresholds prior to infusion, or tailored dosing for higher-risk profiles.

The four representative scenarios, successful therapy, weak killing rescued by high proliferation, poor expansion, and high exhaustion, distilled how parameter interactions shape clinical trajectories. In the successful regime, balanced K_{kill} and r_T produced a peak effector count near 10^8 cells, a stable memory plateau of similar size, and a cytokine peak near 2,000 units while achieving deep remission. In the weak-killing regime, inadequate K_{kill} could be

compensated by extremely high r_T, but this required peak effector counts above 3×10^8 and produced cytokine surges of 8,000-8,500 units, mirroring cases where patients achieve remission only after experiencing severe CRS. In contrast, the poor-expansion and high-exhaustion regimes showed relatively modest cytokine peaks but failed to achieve durable control, illustrating the "safe but ineffective" pattern familiar in the clinic. These scenarios underscore that cure and toxicity are intertwined, and that engineering decisions can shift patients toward one tail or the other.

Quantitatively, the model's behavior aligns well with clinical data on CD19 CAR T kinetics despite its simplifications. Clinical studies report 10-100× expansion peaking between days 7-14; the model's baseline produced ~50× expansion peaking around days 10-12. Responders typically clear leukemic blasts within 1-2 weeks; the model cleared below the 8×10^5 detection threshold within 3-7 days under effective parameter sets. Clinical CRS occurs in 70-90% of patients with severe CRS in 20-50%, and is widely correlated with higher Cmax; Monte Carlo simulations reproduced this association. Sustained B-cell aplasia for 6-12 months is common and correlates with persistence; the model produced deep, prolonged aplasia under high-potency and high-proliferation regimes, and earlier recovery when potency was weaker.

Despite this alignment, the model's limitations are substantive. The single well-mixed compartment cannot represent trafficking to marrow, lymph nodes, or CNS sites. CD19 expression is constant, so antigen-escape relapse, one of the most common failure modes, is absent. Cytokines are represented as a single lumped variable, preventing differentiation between CRS and neurotoxicity or simulation of IL-6 blockade. Effector and memory CAR T populations are homogeneous, ignoring the effects of manufacturing choices that bias toward central memory or effector memory phenotypes. Normal B-cell recovery is simplified to logistic growth and lacks the multiphasic behavior observed clinically. While these omissions were consistent with project constraints, they limit how precisely the model can replicate product-specific or tissue-specific behavior. Future extensions could incorporate multi-compartment structure, antigen-escape modules, heterogeneous T-cell phenotypes, and more mechanistic cytokine signaling to better capture clinical dynamics.

Within these constraints, the extended CARTmath framework met the goals of the project: it shifted discussions of potency, proliferation, memory activation, and dosing from qualitative intuition to explicit, quantitative relationships. It clarified that toxicity and access are not intrinsic properties of "CAR T therapy" but emergent features of design choices, manufacturing standards, and health-system capacity. It also underscored that many parameter regimes that appear optimal in well-resourced centers may be infeasible or unsafe in lower-resource settings because they assume access to monitoring, rapid toxicity management, and long-term supportive care such as IVIG. By making these interactions explicit, the model emphasizes that the success of CAR T therapy depends as much on ethical and structural decisions as on cellular engineering. For patients and clinicians alike, the model highlights both what can be optimized through design and what inequities must be acknowledged if CAR T therapy is to become genuinely accessible.

Conclusion

In this work, we used an extended CARTmath framework to examine how CD19 CAR T-cells govern the coupled dynamics of malignant B-cells, normal B-cells, memory and effector states, and inflammation under a range of clinically motivated scenarios. Rather than treating CD19 CAR T behavior as opaque or purely clinical, the model expresses it as the interaction of a small number of biological forces that decide whether treatment cures, harms, or fails. Its value is not in reproducing any single trajectory but in revealing where the therapy's fragility actually resides. By making potency thresholds, proliferative ceilings, memory responsiveness, and dosing logistics explicit, the model shows how these factors carve out distinct outcome regimes and how easily patients can shift between them because of engineering choices or constraints in access to care. In doing so, it turns questions about safety and durability into quantifiable tensions between tumor control, inflammation, and immune persistence that are easier to analyze and harder to ignore. This shift provides a foundation for evaluating which design decisions truly widen the therapeutic window and which merely reassign risk in ways that may disadvantage certain patients or clinical environments.

Within this structure, a threshold in per cell cytotoxic potency emerges as the primary divider between futility and cure. Below that threshold, CAR T-cells never accumulate enough effective killing to disrupt leukemia's logistic growth. Once potency enters an appropriate range, malignant B-cells collapse after a single infusion and remain suppressed. Multiple construct features, including CD19 affinity, synapse formation, and signaling strength, consolidate into this one quantity that determines whether tumor elimination is even feasible. Memory reactivation becomes the central determinant of durability. Adjusting the strength of antigen-driven recall does little to the timing of initial clearance but markedly reshapes long-term behavior. A small memory pool becomes a more substantial reservoir, capable of suppressing residual disease and responding effectively to rechallenge. Effector proliferation, in turn, aligns therapeutic depth with biological cost. Higher proliferation expands early responses and strengthens memory plateaus, improving protection against relapse, while at the same time prolonging B-cell aplasia and amplifying cytokine exposure. Choices about how aggressively to tune proliferation therefore map directly onto choices about acceptable levels of acute and chronic toxicity.

The model's limitations define the limits of what should be inferred. A single well-mixed compartment cannot represent trafficking patterns, sanctuary sites, or organ-specific toxicities. Homogeneous effector and memory pools overlook the diversity present in real manufacturing products. A lumped cytokine variable cannot distinguish CRS from neurotoxicity or capture intervention-specific modulation. The absence of antigen escape removes a major clinical failure mode. These gaps are structural rather than cosmetic, and improving the model will require extensions that incorporate multi-compartment physiology, phenotype-resolved CAR T subsets, explicit antigen evolution, and cytokine modules tied to measurable biomarkers. Integrating these components would move the framework beyond directional insight and toward a tool capable of supporting sharper, ethically grounded decisions about construct design, eligibility criteria, dosing strategies, and supportive-care planning. As it stands, the model offers both a

map of the governing principles of CD19 CAR T therapy and a blueprint for the next generation of modeling needed to meet the biological and operational realities of treating B-cell ALL with precision and fairness.

Taken together, our findings underscore the power of computational modeling as an engineering tool for optimizing CAR T therapy. By identifying parameter regimes that reliably achieve tumor clearance while minimizing cytokine toxicity, our model can help guide dosing strategies, experimental design, and the development of next-generation constructs. Importantly, the ability to simulate outcomes across thousands of virtual patients provides a framework for understanding variability and improving safety prior to clinical deployment. This work highlights how mechanistic modeling can complement laboratory studies and ultimately support more effective, equitable, and scalable immunotherapies.

References

[1] Cleveland Clinic, "Acute lymphocytic leukemia (ALL): Symptoms, treatment & outlook," *Cleveland Clinic*, Apr. 25, 2023. https://my.clevelandclinic.org/health/diseases/21564-acute-lymphocytic-leukemia

[2] G. Glasgow, "*The Latest News About CAR T-Cell Therapy*," University of Colorado Cancer Center, Jan. 3, 2024. Accessed: Nov. 14, 2025. [Online]. Available: https://news.cuanschutz.edu/cancer-center/latest-news-about-car-t-cell-therapy

[3] L. R. C. Barros, E. A. Paixão, A. M. P. Valli, G. T. Nazouka, A. C. Fassoni, and R. C. Almeida, "*CARTmath—A mathematical model of CAR-T immunotherapy in preclinical studies of hematological cancers*," *Cancers*, vol. 13, no. 12, p. 2941, Jun. 2021, doi: 10.3390/cancers13122941.

[4] Cleveland Clinic, "*Cytokine Release Syndrome (CRS)*," Cleveland Clinic, 2024. Accessed: Nov. 22, 2025. [Online]. Available: https://my.clevelandclinic.org/health/diseases/22700-cytokine-release-syndrome

[5] S. Serrano, R. Barrio, Á. Martínez-Rubio, J. Belmonte-Beitia, and V. M. Pérez-García, "Understanding the role of B cells in CAR T-cell therapy in leukemia through a mathematical model," *Chaos*, vol. 34, no. 8, 2024, Art. no. 08209, doi: 10.1063/5.0206341. https://zaguan.unizar.es/record/145019

[6] D. Sommermeyer, T. Hill, S. M. Shamah, A. I. Salter, Y. Chen, K. M. Mohler, and S. R. Riddell, "Fully human CD19-specific chimeric antigen receptors for T-cell therapy," *Leukemia*, vol. 31, no. 10, pp. 2191–2199, Oct. 2017, doi: 10.1038/leu.2017.57. https://pubmed.ncbi.nlm.nih.gov/28202953/

Appendix A

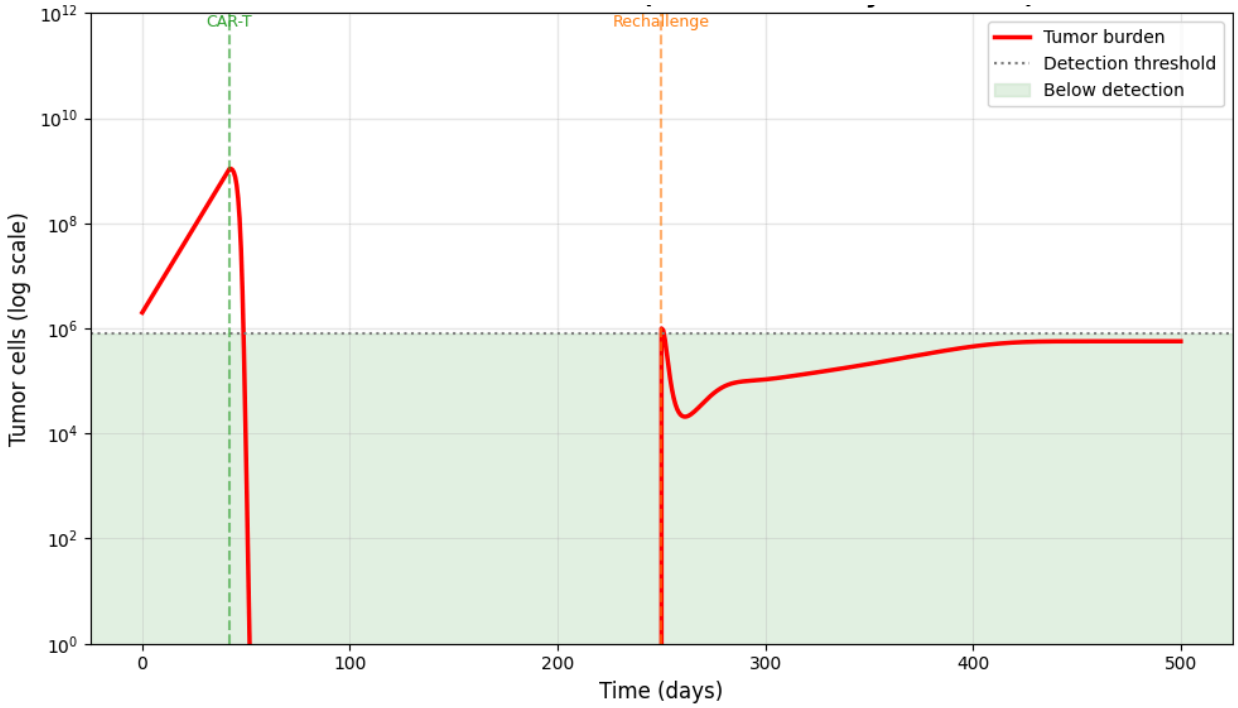

Figure 9 - Tumor Burden Over Time (FIXED - 7-14 day clearance): This graph shows the tumor cell count drastically dropping below the detection threshold after CAR T infusion and the subsequent temporary tumor relapse following rechallenge.

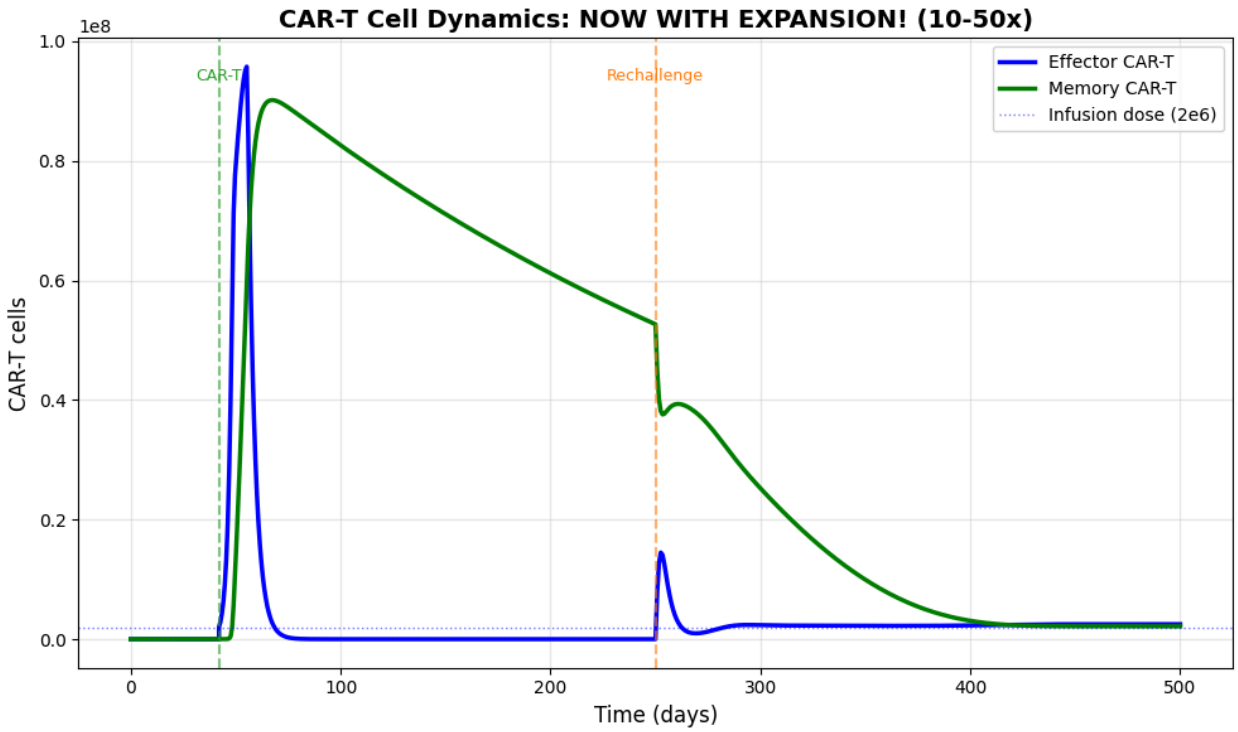

Figure 10 - CAR T-Cell Dynamics (Magnified): This plot illustrates the time course of Effector and Memory CAR T-cells, showing their significant expansion after the initial infusion and a smaller re-expansion following the tumor rechallenge.

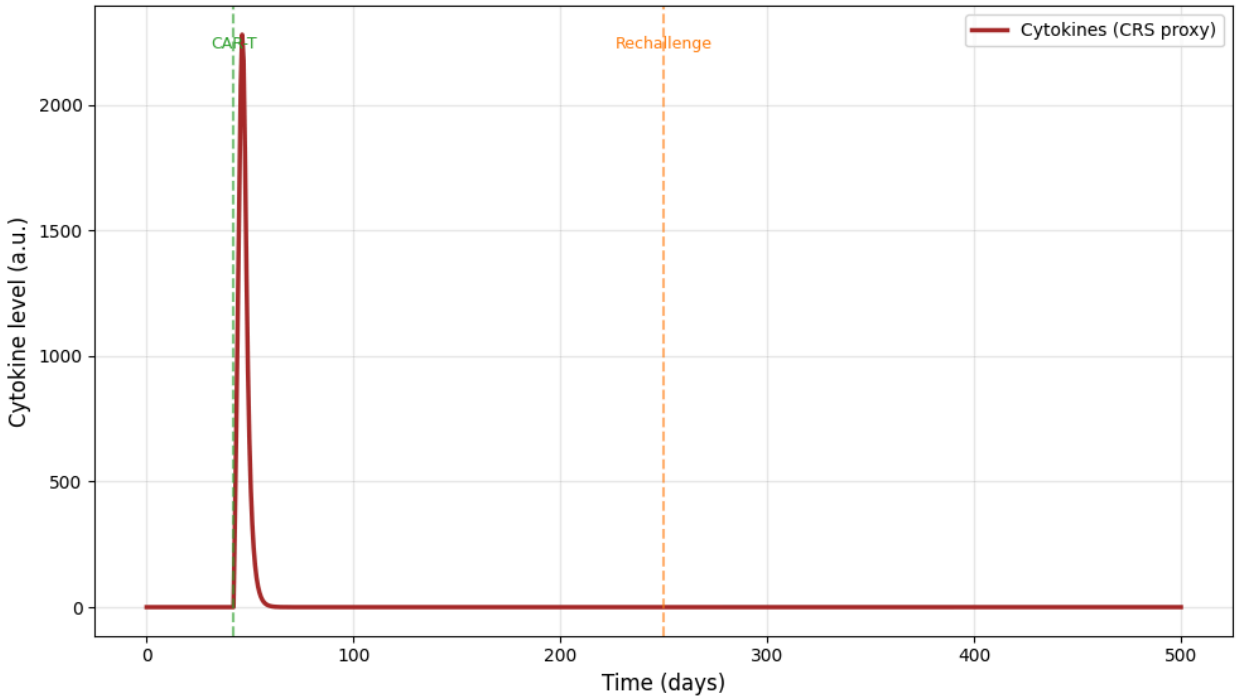

Figure 11 - Cytokine Dynamics (Cytokine Release Syndrome Proxy): This plot shows a sharp, transient peak in cytokine level (a proxy for Cytokine Release Syndrome, CRS) immediately following CAR T infusion, with a smaller peak upon rechallenge.

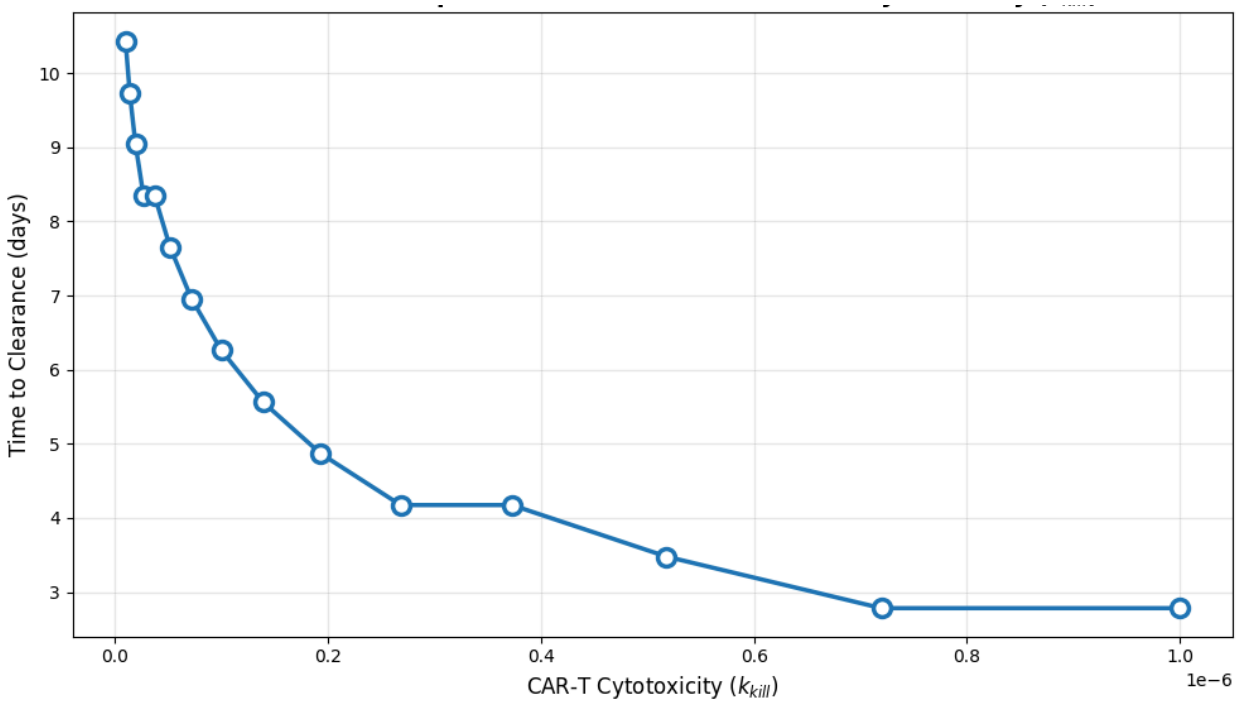

Figure 12 - Parameter Sweep: Time to Clearance vs CAR T Cytotoxicity (α_T): This graph demonstrates that the time required for tumor clearance decreases significantly as the CAR T cytotoxicity rate (α_T) increases.

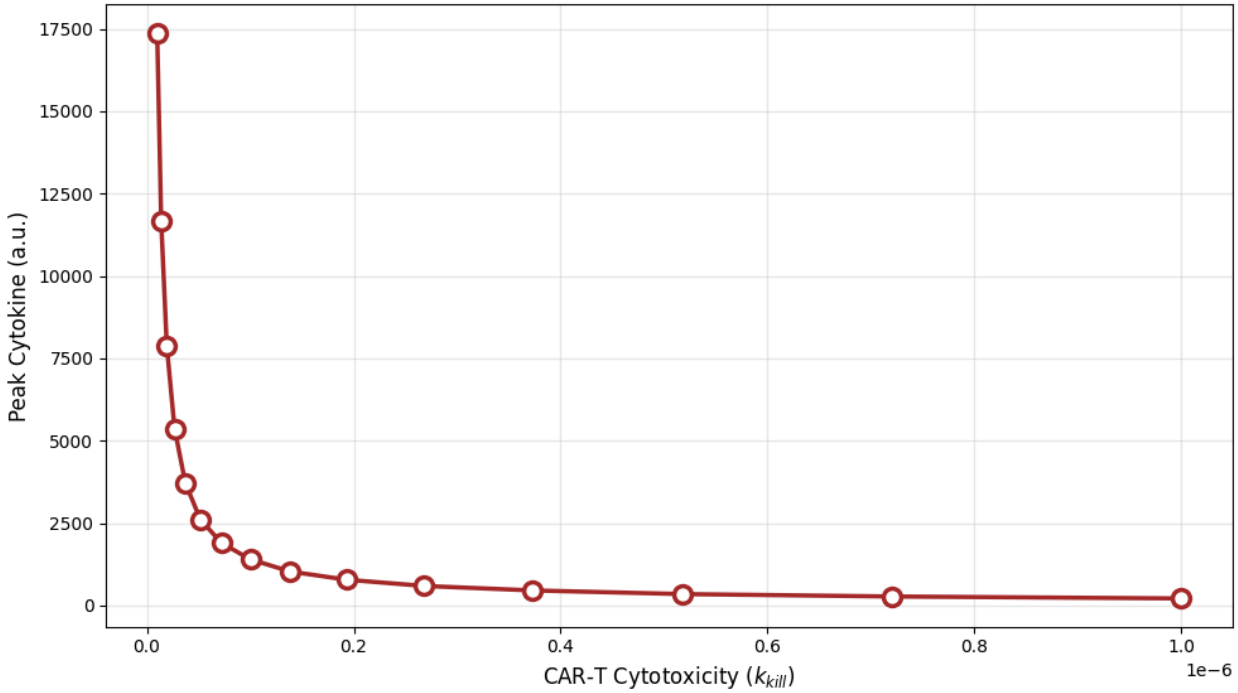

Figure 13 - Parameter Sweep: Peak Cytokine vs CAR T Cytotoxicity (α_T): This graph shows that the peak cytokine level decreases sharply as the CAR T cytotoxicity rate (α_T) increases.

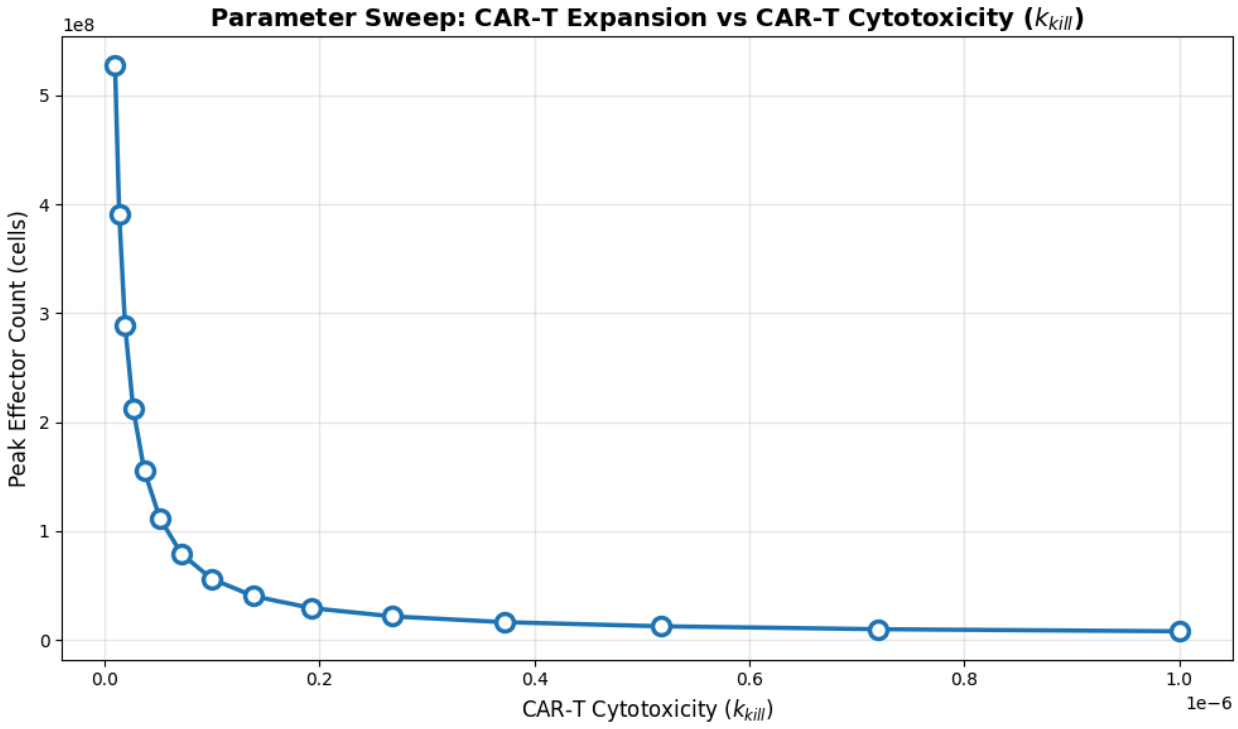

Figure 14 - Parameter Sweep: CAR T Expansion vs CAR T Cytotoxicity (α_T): This plot shows that the peak effector CAR T-cell count decreases as the CAR T cytotoxicity rate (α_T) increases.

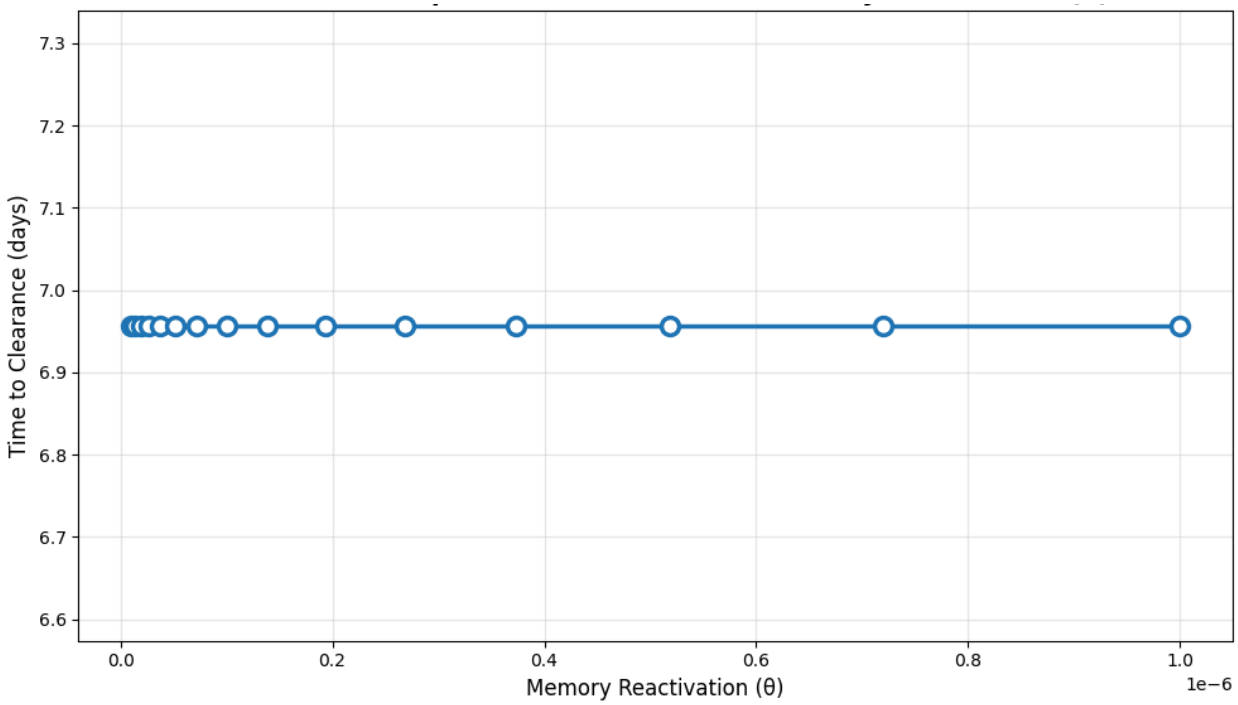

Figure 15 - Parameter Sweep: Time to Clearance vs Memory Reactivation (θ): This graph shows that Time to Clearance is largely unaffected by changes in the Memory Reactivation rate (θ).

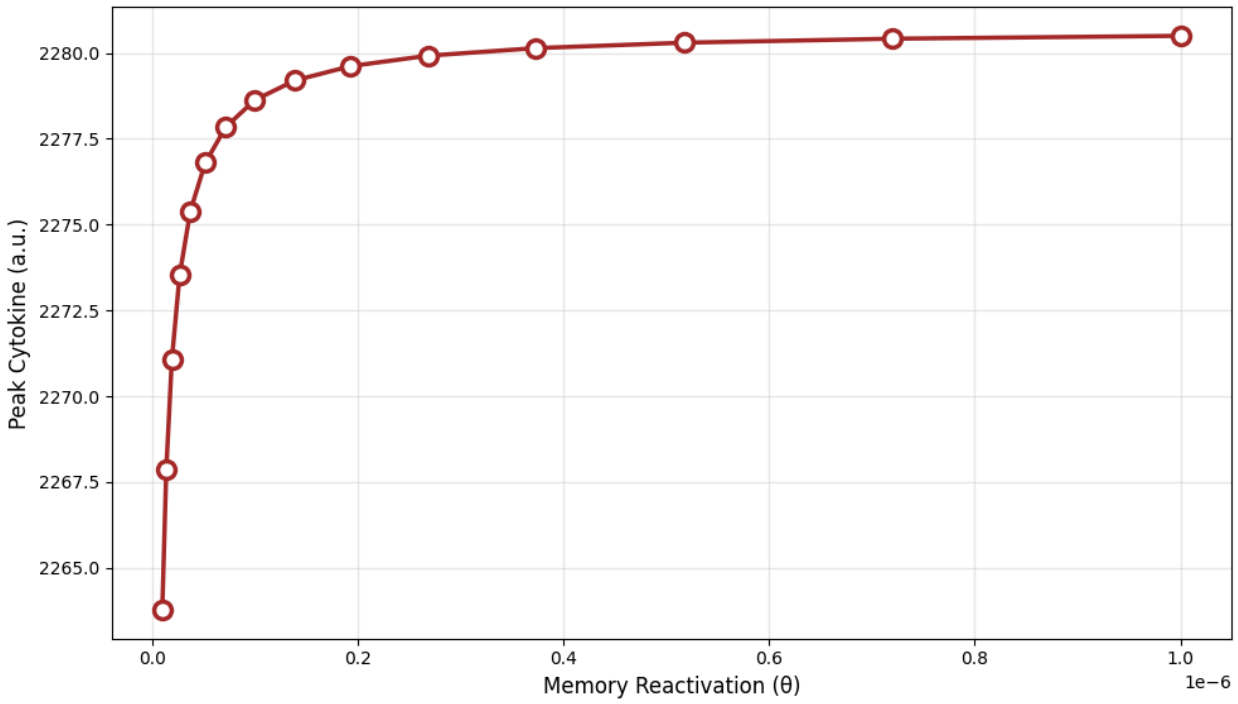

Figure 16 - Parameter Sweep: Peak Cytokine vs Memory Reactivation (θ): This graph shows that the Peak Cytokine level slightly increases as the Memory Reactivation rate (θ) increases.

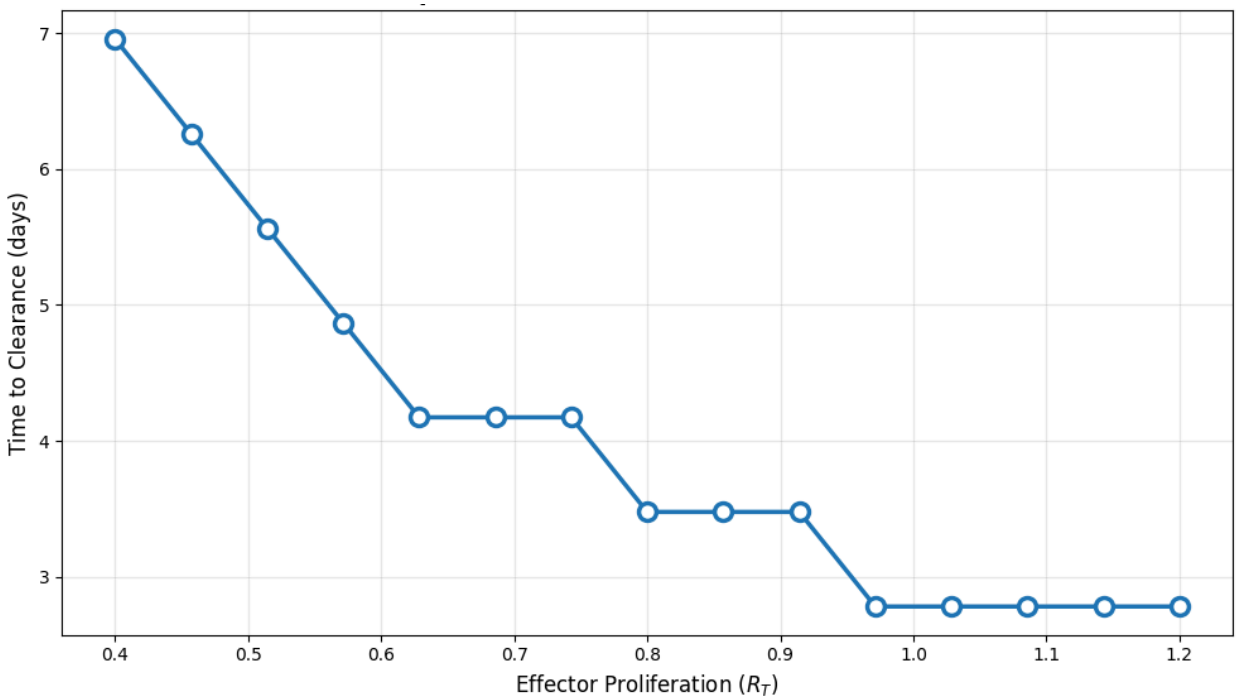

Figure 17 - Parameter Sweep: Time to Clearance vs Effector Proliferation (φ): This graph shows that the time required for tumor clearance decreases as the Effector Proliferation rate (φ) increases.

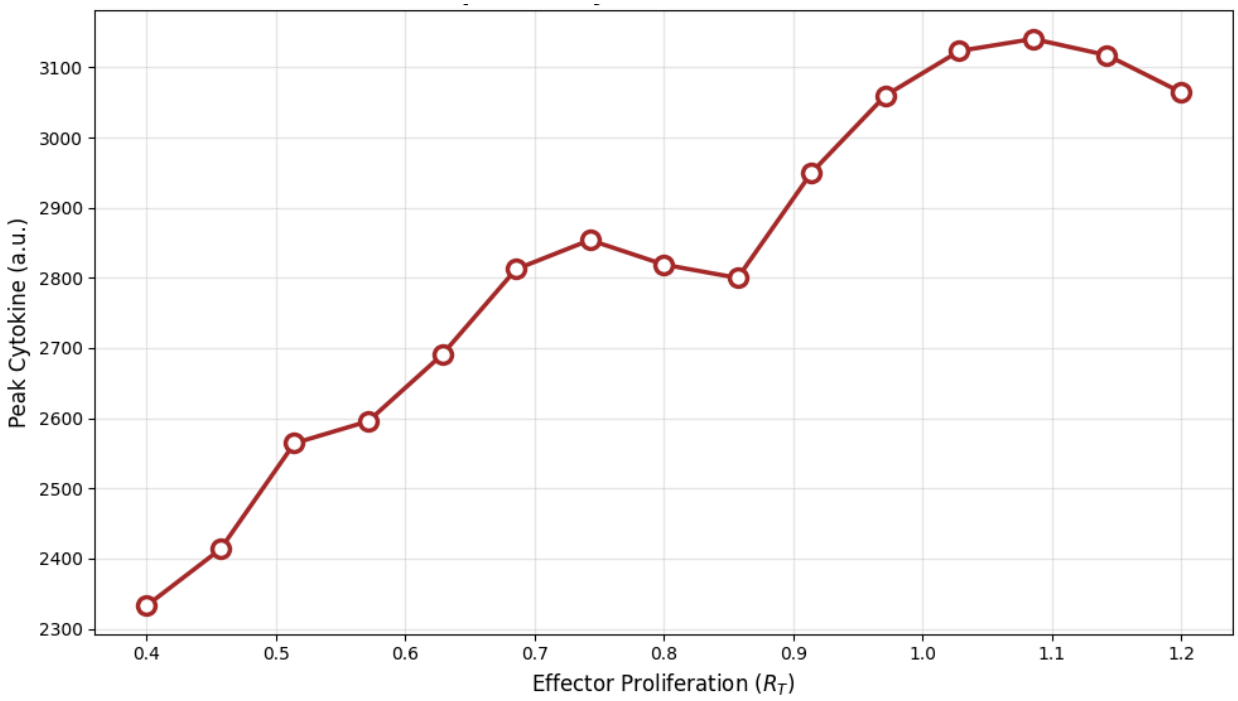

Figure 18 - Parameter Sweep: Peak Cytokine vs Effector Proliferation (φ): This graph shows that the peak cytokine level increases as the Effector Proliferation rate (φ) increases, suggesting a higher risk of Cytokine Release Syndrome (CRS).

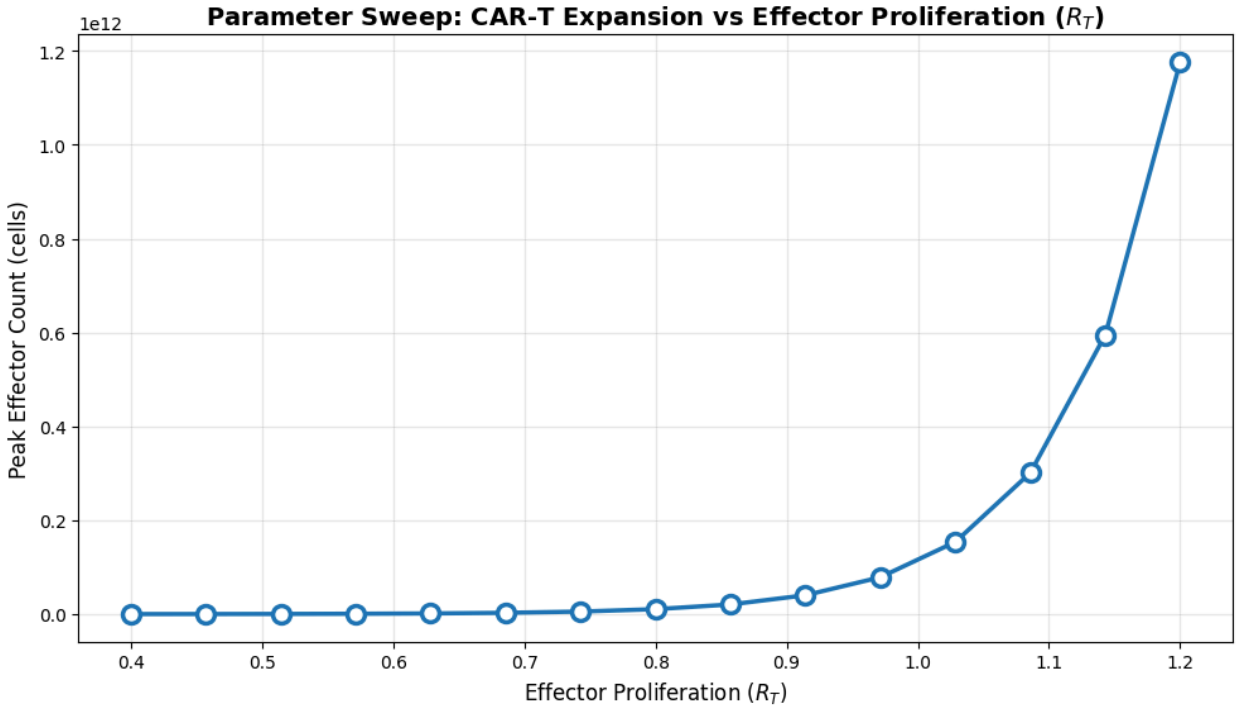

Figure 19 - Parameter Sweep: CAR T Expansion vs Effector Proliferation (r_T). Effector proliferation rate directly controls the magnitude and duration of CAR T expansion, with higher proliferation producing larger peaks and a more sustained memory-supported response.

Appendix B

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from scipy.integrate import odeint
from pathlib import Path
from dataclasses import dataclass

# CONSTANTS
#
================================================================================
========
DETECTION_THRESHOLD = 8e5  # cells (clinical detection limit)


# PARAMETERS
#The parameters were acquired from,

#L. R. C. Barros, E. A. Paixão, A. M. P. Valli, G. T. Nazouka, A. C. Fassoni, and R. C. Almeida,
#“CARTmath—A mathematical model of CAR-T immunotherapy in preclinical studies of
hematological cancers,” Cancers, vol. 13, no. 12, p. 2941,
#Jun. 2021, doi: 10.3390/cancers13122941)

#S. Serrano, R. Barrio, Á. Martínez-Rubio, J. Belmonte-Beitia, and V. M. Pérez-García,
#“Understanding the role of B cells in CAR T-cell therapy in leukemia through a mathematical
model,”
#Chaos, vol. 34, no. 8, 2024, Art. no. 08209, doi: 10.1063/5.0206341.

#D. Sommermeyer, T. Hill, S. M. Shamah, A. I. Salter, Y. Chen, K. M. Mohler, and S. R.
Riddell,
#“Fully human CD19-specific chimeric antigen receptors for T-cell therapy,” Leukemia, vol. 31,
no. 10, pp. 2191–2199,
#Oct. 2017, doi: 10.1038/leu.2017.57.


#Some parameters were adjusted to better reflect biological phenomena.
```

```
#
================================================================================
========
@dataclass
class CARTParameters:
    """
    Fixed parameters for realistic CD19 CAR-T therapy dynamics
    """
    # Tumor parameters
    r_c: float = 0.15      # day^-1, tumor growth rate
    K_c: float = 5e11      # cells, tumor carrying capacity

    # CAR-T killing parameters (FIXED FOR 7-14 DAY CLEARANCE)
    k_kill: float = 6e-8   # (cell*day)^-1, tumor killing rate
    gamma_B: float = 6e-8   # (cell*day)^-1, B-cell killing rate

    # CAR-T dynamics parameters
    R_T: float = 0.38      # day^-1, max proliferation rate
    rho: float = 0.20      # day^-1, death rate (HIGHER for controlled expansion)
    epsilon: float = 0.10  # day^-1, effector→memory conversion
    mu: float = 0.003      # day^-1, memory death
    theta: float = 2e-7    # (cell*day)^-1, memory reactivation
    alpha: float = 4e-13   # (cell*day)^-1, immunosuppression (HIGHER)
    K_T: float = 5e5       # cells, antigen half-saturation

    # Proliferation commitment
    tau_commit: float = 9.0  # days, duration of committed proliferation
    R_T_committed: float = 0.25  # fold increase during committed phase

    # B-cell parameters
    r_N: float = 0.01      # day^-1, B-cell growth rate
    K_N: float = 1e9       # cells, B-cell carrying capacity
    B0: float = 1e8        # cells, initial B-cell count

    # Cytokine parameters
    k_cyt_prod: float = 2e-13    # (cell^2*day)^-1, cytokine production
    k_cyt_clear: float = 0.5     # day^-1, cytokine clearance
    K_antigen: float = 5e6       # cells, antigen half-max
    k_cyt_fb: float = 0.8        # dimensionless, cytokine feedback strength (INCREASED)
    K_cyt: float = 20.0          # a.u., cytokine half-max
```

```
    # Initial conditions
    Y0: float = 0.0       # cytokine level at t=0

def R_T_effective(Y, t, t_infusion, params):

    # Time since infusion
    t_since_infusion = t - t_infusion

    # Committed expansion boost
    if t_since_infusion > 0 and t_since_infusion < params.tau_commit * 3:
        commitment_factor = params.R_T_committed * np.exp(-t_since_infusion /
params.tau_commit)
    else:
        commitment_factor = 0.0

    # Cytokine feedback
    cytokine_factor = params.k_cyt_fb * Y / (Y + params.K_cyt)

    # Combined proliferation rate
    return params.R_T * (1.0 + commitment_factor + cytokine_factor)

# ODE SYSTEM - FIXED EQUATIONS
#
==========================================================================
=========
def cart_ode(y, t, params, t_infusion):
    """
    CAR-T therapy ODE system with REALISTIC expansion dynamics
    """
    T, C_T, C_M, B, Y = y

    # Enforce non-negativity
    T = max(0, T)
    C_T = max(0, C_T)
    C_M = max(0, C_M)
    B = max(0, B)
    Y = max(0, Y)

    # Cytokine-enhanced proliferation WITH COMMITMENT
```

```
R_T_eff = R_T_effective(Y, t, t_infusion, params)

# Antigen-dependent activation factor
antigen_activation = T / (T + params.K_T)

# Tumor dynamics
dT_dt = (params.r_c * T * (1 - T/params.K_c) -
        params.k_kill * C_T * T)

# Effector CAR-T dynamics
time_since_infusion = t - t_infusion
if time_since_infusion > 0 and time_since_infusion < params.tau_commit * 1.5:
    # During commitment phase, use reduced antigen dependence
    effective_activation = 0.5 + 0.5 * antigen_activation
else:
    # After commitment, full antigen dependence
    effective_activation = antigen_activation

dC_T_dt = (R_T_eff * C_T * effective_activation -
          params.rho * C_T -
          params.epsilon * C_T +
          params.theta * T * C_M -
          params.alpha * T * C_T)

# Memory CAR-T dynamics
dC_M_dt = (params.epsilon * C_T -
          params.theta * T * C_M -
          params.mu * C_M)

# Normal B-cell dynamics
dB_dt = (params.r_N * B * (1 - B/params.K_N) -
        params.gamma_B * C_T * B)

# Cytokine dynamics
total_antigen = T + B
antigen_saturation = total_antigen / (total_antigen + params.K_antigen)
dY_dt = (params.k_cyt_prod * C_T * total_antigen * antigen_saturation -
        params.k_cyt_clear * Y)

return [dT_dt, dC_T_dt, dC_M_dt, dB_dt, dY_dt]
```

```
# SIMULATION FUNCTION
#
================================================================================
========
def simulate_cart(params, T0, CT0, CM0, B0, Y0,
                  therapy_times, therapy_doses,
                  challenge_times=None, challenge_doses=None,
                  t_end=500, points_per_segment=300):
    """
    Simulate CAR-T therapy with therapy and optional rechallenge events
    """
    if challenge_times is None:
        challenge_times = []
    if challenge_doses is None:
        challenge_doses = []

    # Track first infusion time for commitment phase
    t_infusion = therapy_times[0] if therapy_times else 0

    # Combine and sort all event times
    all_events = [(0, 'start', None)]
    for t, dose in zip(therapy_times, therapy_doses):
        all_events.append((t, 'therapy', dose))
    for t, dose in zip(challenge_times, challenge_doses):
        all_events.append((t, 'challenge', dose))
    all_events.append((t_end, 'end', None))
    all_events.sort(key=lambda x: x[0])

    # Initialize
    y_current = [T0, CT0, CM0, B0, Y0]
    t_all = []
    T_all = []
    CT_all = []
    CM_all = []
    B_all = []
    Y_all = []

    # Simulate piecewise
```

```
    for i in range(len(all_events) - 1):
        t_start, event_type, dose = all_events[i]
        t_next = all_events[i + 1][0]

        # Apply event
        if event_type == 'therapy':
            y_current[1] += dose  # Add to C_T
        elif event_type == 'challenge':
            y_current[0] += dose  # Add to T

        # Integrate
        t_span = np.linspace(t_start, t_next, points_per_segment)
        sol = odeint(cart_ode, y_current, t_span, args=(params, t_infusion))

        # Store results
        t_all.extend(t_span)
        T_all.extend(sol[:, 0])
        CT_all.extend(sol[:, 1])
        CM_all.extend(sol[:, 2])
        B_all.extend(sol[:, 3])
        Y_all.extend(sol[:, 4])

        # Update for next segment
        y_current = sol[-1, :]

    return {
        'time': np.array(t_all),
        'T': np.array(T_all),
        'C_T': np.array(CT_all),
        'C_M': np.array(CM_all),
        'B': np.array(B_all),
        'Y': np.array(Y_all)
    }


# PLOTTING FUNCTIONS
#
================================================================================
========
```

```
def add_event_vlines(ax, therapy_times, challenge_times):
    """Add vertical lines for therapy and rechallenge events"""
    ylim = ax.get_ylim()
    for t in therapy_times:
        ax.axvline(t, color='tab:green', linestyle='--', alpha=0.6, lw=1.5)
        ax.text(t, ylim[1]*0.95, 'CAR-T', ha='center', va='top',
                fontsize=9, color='tab:green')
    for t in challenge_times:
        ax.axvline(t, color='tab:orange', linestyle='--', alpha=0.6, lw=1.5)
        ax.text(t, ylim[1]*0.95, 'Rechallenge', ha='center', va='top',
                fontsize=9, color='tab:orange')

def plot_all_populations(results, therapy_times, challenge_times, save_path=None):
    """Figure: All cell populations over time"""
    t = results['time']
    T, CT, CM, B, Y = results['T'], results['C_T'], results['C_M'], results['B'], results['Y']

    fig, ax = plt.subplots(figsize=(10, 6))

    ax.semilogy(t, T, 'r-', lw=2, label='Tumor')
    ax.semilogy(t, CT, 'b-', lw=2, label='Effector CAR-T')
    ax.semilogy(t, CM, 'g-', lw=2, label='Memory CAR-T')
    ax.semilogy(t, B, 'm-.', lw=2, label='Normal B cells')

    ax.axhline(DETECTION_THRESHOLD, color='gray', ls=':', lw=1.5,
               label=f'Detection threshold ({DETECTION_THRESHOLD:.1e})')

    ax.set_xlabel('Time (days)', fontsize=12)
    ax.set_ylabel('Cell count (log scale)', fontsize=12)
    ax.set_title('CAR-T Therapy: All Cell Populations (FIXED MODEL)', fontsize=14,
fontweight='bold')
    ax.grid(True, alpha=0.3)
    ax.legend(loc='best', fontsize=10)
    ax.set_ylim(1e0, 1e12)

    add_event_vlines(ax, therapy_times, challenge_times)
    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')
```

```
    return fig, ax

def plot_tumor_only(results, therapy_times, challenge_times, save_path=None):
    """Figure: Tumor burden dynamics"""
    t, T = results['time'], results['T']

    fig, ax = plt.subplots(figsize=(10, 6))

    ax.semilogy(t, T, 'r-', lw=2.5, label='Tumor burden')
    ax.axhline(DETECTION_THRESHOLD, color='gray', ls=':', lw=1.5,
               label='Detection threshold')
    ax.axhspan(1e0, DETECTION_THRESHOLD, color='green', alpha=0.1,
               label='Below detection')

    ax.set_xlabel('Time (days)', fontsize=12)
    ax.set_ylabel('Tumor cells (log scale)', fontsize=12)
    ax.set_title('Tumor Burden Over Time (FIXED - 7-14 day clearance)', fontsize=14,
fontweight='bold')
    ax.grid(True, alpha=0.3)
    ax.legend(loc='best', fontsize=10)
    ax.set_ylim(1e0, 1e12)

    add_event_vlines(ax, therapy_times, challenge_times)
    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, ax

def plot_cart_populations(results, therapy_times, challenge_times, save_path=None):
    """Figure: Effector vs Memory CAR-T dynamics"""
    t, CT, CM = results['time'], results['C_T'], results['C_M']

    fig, ax = plt.subplots(figsize=(10, 6))

    ax.plot(t, CT, 'b-', lw=2.5, label='Effector CAR-T')
    ax.plot(t, CM, 'g-', lw=2.5, label='Memory CAR-T')
```

```
    # Add reference line for infusion dose
    if therapy_times:
        ax.axhline(2e6, color='blue', ls=':', lw=1, alpha=0.5, label='Infusion dose (2e6)')

    ax.set_xlabel('Time (days)', fontsize=12)
    ax.set_ylabel('CAR-T cells', fontsize=12)
    ax.set_title('CAR-T Cell Dynamics: NOW WITH EXPANSION! (10-50x)', fontsize=14,
fontweight='bold')
    ax.grid(True, alpha=0.3)
    ax.legend(loc='best', fontsize=10)
    ax.ticklabel_format(style='sci', axis='y', scilimits=(0, 0))

    add_event_vlines(ax, therapy_times, challenge_times)
    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, ax

def plot_b_cells(results, therapy_times, challenge_times, save_path=None):
    """Figure: Normal B-cell depletion and recovery"""
    t, B = results['time'], results['B']

    fig, ax = plt.subplots(figsize=(10, 6))

    ax.plot(t, B, 'm-', lw=2.5, label='Normal B cells')

    ax.set_xlabel('Time (days)', fontsize=12)
    ax.set_ylabel('B cells', fontsize=12)
    ax.set_title('Normal B-Cell Depletion and Recovery', fontsize=14, fontweight='bold')
    ax.grid(True, alpha=0.3)
    ax.legend(loc='best', fontsize=10)
    ax.ticklabel_format(style='sci', axis='y', scilimits=(0, 0))

    add_event_vlines(ax, therapy_times, challenge_times)
    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')
```

```
    return fig, ax

def plot_cytokines(results, therapy_times, challenge_times, save_path=None):
    """Figure: Cytokine dynamics (CRS proxy)"""
    t, Y = results['time'], results['Y']

    fig, ax = plt.subplots(figsize=(10, 6))

    ax.plot(t, Y, color='brown', lw=2.5, label='Cytokines (CRS proxy)')

    ax.set_xlabel('Time (days)', fontsize=12)
    ax.set_ylabel('Cytokine level (a.u.)', fontsize=12)
    ax.set_title('Cytokine Dynamics (Cytokine Release Syndrome Proxy)',
                 fontsize=14, fontweight='bold')
    ax.grid(True, alpha=0.3)
    ax.legend(loc='best', fontsize=10)

    add_event_vlines(ax, therapy_times, challenge_times)
    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, ax

def plot_sweep_time_to_clearance(metrics_df, param_name, param_label, save_path=None):
    """Parameter sweep: Time to tumor clearance"""
    param_vals = metrics_df[param_name].values
    clearance_times = metrics_df['time_to_clearance'].values
    valid_mask = ~pd.isna(clearance_times)

    fig, ax = plt.subplots(figsize=(10, 6))

    ax.plot(param_vals[valid_mask], clearance_times[valid_mask],
            'o-', lw=2.5, markersize=10, color='tab:blue',
            markerfacecolor='white', markeredgewidth=2.5)

    ax.set_xlabel(param_label, fontsize=12)
    ax.set_ylabel('Time to Clearance (days)', fontsize=12)
```

```
    ax.set_title(f'Parameter Sweep: Time to Clearance vs {param_label}',
                 fontsize=14, fontweight='bold')
    ax.grid(True, alpha=0.3)

    if param_vals.max() / param_vals.min() > 100:
        ax.set_xscale('log')

    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, ax

def plot_sweep_peak_cytokine(metrics_df, param_name, param_label, save_path=None):
    """Parameter sweep: Peak cytokine"""
    param_vals = metrics_df[param_name].values
    peak_cytokine = metrics_df['peak_cytokine'].values

    fig, ax = plt.subplots(figsize=(10, 6))

    ax.plot(param_vals, peak_cytokine,
            'o-', lw=2.5, markersize=10, color='brown',
            markerfacecolor='white', markeredgewidth=2.5)

    ax.set_xlabel(param_label, fontsize=12)
    ax.set_ylabel('Peak Cytokine (a.u.)', fontsize=12)
    ax.set_title(f'Parameter Sweep: Peak Cytokine vs {param_label}',
                 fontsize=14, fontweight='bold')
    ax.grid(True, alpha=0.3)

    if param_vals.max() / param_vals.min() > 100:
        ax.set_xscale('log')

    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, ax
```

```python
def plot_sweep_peak_effector(metrics_df, param_name, param_label, save_path=None):
    """Parameter sweep: Peak effector expansion"""
    param_vals = metrics_df[param_name].values
    peak_effector = metrics_df['peak_effector'].values

    fig, ax = plt.subplots(figsize=(10, 6))

    ax.plot(param_vals, peak_effector,
            'o-', lw=2.5, markersize=10, color='tab:blue',
            markerfacecolor='white', markeredgewidth=2.5)

    ax.set_xlabel(param_label, fontsize=12)
    ax.set_ylabel('Peak Effector Count (cells)', fontsize=12)
    ax.set_title(f'Parameter Sweep: CAR-T Expansion vs {param_label}',
                 fontsize=14, fontweight='bold')
    ax.grid(True, alpha=0.3)
    ax.ticklabel_format(style='sci', axis='y', scilimits=(0, 0))

    if param_vals.max() / param_vals.min() > 100:
        ax.set_xscale('log')

    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, ax


# COMPARISON PLOTS
#
# ================================================================================

def plot_dosing_comparison(results_single, results_frac,
                           therapy_times_single, therapy_times_frac,
                           save_path=None):
    """Figure: CAR-T therapy comparison"""
    fig, axes = plt.subplots(3, 1, figsize=(12, 10))
```

```
    t_s = results_single['time']
    t_f = results_frac['time']

    # Plot 1: Tumor burden
    ax = axes[0]
    ax.semilogy(t_s, results_single['T'], 'b-', lw=2.5, label='Single dose')
    ax.semilogy(t_f, results_frac['T'], 'orange', ls='--', lw=2.5, label='Fractionated')
    ax.axhline(DETECTION_THRESHOLD, color='gray', ls=':', lw=1.5, label='Detection
threshold')
    ax.set_ylabel('Tumor cells', fontsize=12)
    ax.set_title('Single vs. Fractionated CAR-T Dosing (Same Total Dose)',
                 fontsize=14, fontweight='bold')
    ax.legend(loc='best')
    ax.grid(True, alpha=0.3)
    ax.set_ylim(1e5, 1e11)

    # Plot 2: Normal B cells
    ax = axes[1]
    ax.plot(t_s, results_single['B'], 'b-', lw=2.5, label='B cells - single')
    ax.plot(t_f, results_frac['B'], 'orange', ls='--', lw=2.5, label='B cells - frac')
    ax.set_ylabel('Normal B cells', fontsize=12)
    ax.legend(loc='best')
    ax.grid(True, alpha=0.3)
    ax.ticklabel_format(style='sci', axis='y', scilimits=(0, 0))

    # Plot 3: Cytokines
    ax = axes[2]
    ax.plot(t_s, results_single['Y'], 'b-', lw=2.5, label='Cytokine - single')
    ax.plot(t_f, results_frac['Y'], 'orange', ls='--', lw=2.5, label='Cytokine - frac')
    ax.set_xlabel('Time (days)', fontsize=12)
    ax.set_ylabel('Cytokine (arb. units)', fontsize=12)
    ax.legend(loc='best')
    ax.grid(True, alpha=0.3)

    # Add vertical lines for dosing times
    for ax in axes:
        for t in therapy_times_single:
            ax.axvline(t, color='blue', alpha=0.3, ls='-', lw=1)
        for t in therapy_times_frac:
```

```
            ax.axvline(t, color='orange', alpha=0.3, ls='--', lw=1)

    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, axes

def plot_efficacy_toxicity_tradeoff(sweep_results, param_name, param_values,
save_path=None):

    """Figure: Efficacy-toxicity tradeoff"""
    clearance_times = []
    peak_cytokines = []
    peak_effectors = []

    for res in sweep_results:
        t = res['time']
        T = res['T']
        Y = res['Y']
        CT = res['C_T']

        # Time to clearance
        below = T < DETECTION_THRESHOLD
        if np.any(below):
            clearance_times.append(t[np.where(below)[0][0]])
        else:
            clearance_times.append(np.nan)

        peak_cytokines.append(np.max(Y))
        peak_effectors.append(np.max(CT))

    clearance_times = np.array(clearance_times)
    peak_cytokines = np.array(peak_cytokines)

    fig, axes = plt.subplots(1, 2, figsize=(14, 5))

    # Plot 1: Clearance time vs cytokine burden
    ax = axes[0]
```

```
    valid = ~np.isnan(clearance_times)
    scatter = ax.scatter(clearance_times[valid], peak_cytokines[valid],
                         c=param_values[valid], s=150, cmap='viridis',
                         edgecolors='black', linewidth=1.5, alpha=0.8)
    ax.set_xlabel('Time to Clearance (days)', fontsize=12)
    ax.set_ylabel('Peak Cytokine (toxicity proxy)', fontsize=12)
    ax.set_title('Efficacy-Toxicity Tradeoff', fontsize=13, fontweight='bold')
    ax.grid(True, alpha=0.3)
    cbar = plt.colorbar(scatter, ax=ax)
    cbar.set_label(f'{param_name}', fontsize=11)

    # Add ideal region
    ax.axvspan(7, 14, alpha=0.1, color='green', label='Target clearance window')
    ax.legend()

    # Plot 2: Expansion vs cytokine
    ax = axes[1]
    scatter = ax.scatter(np.array(peak_effectors)[valid]/2e6, peak_cytokines[valid],
                         c=param_values[valid], s=150, cmap='viridis',
                         edgecolors='black', linewidth=1.5, alpha=0.8)
    ax.set_xlabel('Expansion Fold (peak / dose)', fontsize=12)
    ax.set_ylabel('Peak Cytokine (toxicity proxy)', fontsize=12)
    ax.set_title('Expansion-Toxicity Relationship', fontsize=13, fontweight='bold')
    ax.grid(True, alpha=0.3)
    cbar = plt.colorbar(scatter, ax=ax)
    cbar.set_label(f'{param_name}', fontsize=11)

    # Add target region
    ax.axvspan(10, 50, alpha=0.1, color='green', label='Target expansion')
    ax.legend()

    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, axes

def plot_patient_variability_monte_carlo(params, T0, CT0, CM0, B0, Y0,
                                         therapy_times, therapy_doses,
```

```
                                n_patients=100, save_path=None):

    """Figure: Patient variability in Monte Carlo simulation"""
    from copy import deepcopy

    np.random.seed(42)

    fig, axes = plt.subplots(2, 3, figsize=(16, 10))

    # Storage for Monte Carlo results
    clearance_times = []
    peak_effectors = []
    peak_cytokines = []
    b_nadir = []
    final_tumors = []

    # Run Monte Carlo
    print(f"  Running Monte Carlo with {n_patients} patients...")

    for i in range(n_patients):
        # Add ±30% variability to key parameters
        p_var = deepcopy(params)
        p_var.R_T = params.R_T * np.random.uniform(0.7, 1.3)
        p_var.k_kill = params.k_kill * np.random.uniform(0.7, 1.3)
        p_var.rho = params.rho * np.random.uniform(0.7, 1.3)
        p_var.theta = params.theta * np.random.uniform(0.5, 1.5)
        p_var.alpha = params.alpha * np.random.uniform(0.5, 1.5)

        # Vary initial tumor burden ±50%
        T0_var = T0 * np.random.uniform(0.5, 2.0)

        res = simulate_cart(p_var, T0_var, CT0, CM0, B0, Y0,
                            therapy_times, therapy_doses,
                            t_end=150)

        t = res['time']

        # Extract metrics
        peak_effectors.append(np.max(res['C_T']))
        peak_cytokines.append(np.max(res['Y']))
```

```
        b_nadir.append(np.min(res['B']))
        final_tumors.append(res['T'][-1])

        # Clearance time
        below = res['T'] < DETECTION_THRESHOLD
        if np.any(below):
            clearance_times.append(t[np.where(below)[0][0]] - therapy_times[0])
        else:
            clearance_times.append(np.nan)

    clearance_times = np.array(clearance_times)
    peak_effectors = np.array(peak_effectors)
    peak_cytokines = np.array(peak_cytokines)
    final_tumors = np.array(final_tumors)

    # Calculate response rate
    responders = final_tumors < DETECTION_THRESHOLD
    response_rate = np.sum(responders) / n_patients * 100

    # Plot 1: Clearance time distribution
    ax = axes[0, 0]
    valid_clearance = clearance_times[~np.isnan(clearance_times)]
    ax.hist(valid_clearance, bins=20, color='steelblue', alpha=0.7, edgecolor='black')
    ax.axvline(np.median(valid_clearance), color='red', ls='--', lw=2, label=f'Median:
{np.median(valid_clearance):.1f}d')
    ax.axvspan(7, 14, alpha=0.2, color='green', label='Target window')
    ax.set_xlabel('Time to Clearance (days)', fontsize=11)
    ax.set_ylabel('Number of Patients', fontsize=11)
    ax.set_title(f'Clearance Time Distribution (n={len(valid_clearance)})', fontsize=12,
fontweight='bold')
    ax.legend()
    ax.grid(True, alpha=0.3)

    # Plot 2: Peak expansion distribution
    ax = axes[0, 1]
    expansion_fold = peak_effectors / 2e6
    ax.hist(expansion_fold, bins=20, color='coral', alpha=0.7, edgecolor='black')
    ax.axvline(np.median(expansion_fold), color='red', ls='--', lw=2, label=f'Median:
{np.median(expansion_fold):.1f}x')
    ax.axvspan(10, 50, alpha=0.2, color='green', label='Target range')
```

```
    ax.set_xlabel('Expansion Fold (peak / dose)', fontsize=11)
    ax.set_ylabel('Number of Patients', fontsize=11)
    ax.set_title('CAR-T Expansion Distribution', fontsize=12, fontweight='bold')
    ax.legend()
    ax.grid(True, alpha=0.3)

    # Plot 3: Peak cytokine distribution
    ax = axes[0, 2]
    ax.hist(peak_cytokines, bins=20, color='brown', alpha=0.7, edgecolor='black')
    ax.axvline(np.median(peak_cytokines), color='red', ls='--', lw=2, label=f'Median:
{np.median(peak_cytokines):.0f}')
    ax.set_xlabel('Peak Cytokine (arb. units)', fontsize=11)
    ax.set_ylabel('Number of Patients', fontsize=11)
    ax.set_title('Cytokine Peak Distribution (CRS Risk)', fontsize=12, fontweight='bold')
    ax.legend()
    ax.grid(True, alpha=0.3)

    # Plot 4: Response vs. Non-response pie chart
    ax = axes[1, 0]
    responder_counts = [np.sum(responders), np.sum(~responders)]
    colors_pie = ['#2ecc71', '#e74c3c']
    wedges, texts, autotexts = ax.pie(responder_counts, labels=['Responders', 'Non-responders'],
                                      autopct='%1.1f%%', colors=colors_pie, startangle=90,
                                      textprops={'fontsize': 12, 'weight': 'bold'})
    ax.set_title(f'Response Rate: {response_rate:.1f}%\n(Clinical ~81%)',
                 fontsize=12, fontweight='bold')

    # Plot 5: Expansion vs. Cytokine scatter
    ax = axes[1, 1]
    scatter = ax.scatter(expansion_fold, peak_cytokines, c=final_tumors,
                         s=80, alpha=0.6, cmap='RdYlGn_r', norm=plt.Normalize(vmin=1e4, vmax=1e7),
                         edgecolors='black', linewidth=0.5)
    ax.set_xlabel('Expansion Fold', fontsize=11)
    ax.set_ylabel('Peak Cytokine', fontsize=11)
    ax.set_title('Expansion-Toxicity Relationship', fontsize=12, fontweight='bold')
    ax.grid(True, alpha=0.3)
    cbar = plt.colorbar(scatter, ax=ax)
    cbar.set_label('Final Tumor (cells)', fontsize=10)

    # Plot 6: Summary statistics
```

```
    ax = axes[1, 2]
    ax.axis('off')
    summary_text = f"""Monte Carlo Summary (n={n_patients})

Response Rate: {response_rate:.1f}%
(Clinical: ~81%)

Time to Clearance:
  Median: {np.median(valid_clearance):.1f} days
  Range: {np.min(valid_clearance):.1f}-{np.max(valid_clearance):.1f} days

Expansion:
  Median: {np.median(expansion_fold):.1f}x
  Range: {np.min(expansion_fold):.1f}-{np.max(expansion_fold):.1f}x

Peak Cytokine:
  Median: {np.median(peak_cytokines):.0f}
  Range: {np.min(peak_cytokines):.0f}-{np.max(peak_cytokines):.0f}

Non-responders: {np.sum(~responders)} patients
Final tumor > threshold"""

    ax.text(0.1, 0.5, summary_text, transform=ax.transAxes, fontsize=11,
            verticalalignment='center', family='monospace',
            bbox=dict(boxstyle='round', facecolor='wheat', alpha=0.3))

    fig.suptitle(f'Monte Carlo Patient Heterogeneity Simulation (n={n_patients})',
                 fontsize=15, fontweight='bold', y=0.995)

    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, axes

def plot_failure_modes(params, T0, CT0, CM0, B0, Y0,
                       therapy_times, therapy_doses,
                       save_path=None):
```

```
from copy import deepcopy

fig, axes = plt.subplots(2, 2, figsize=(14, 10))

scenarios = [
    ("Successful", params, 'green', '-'),
    ("Weak Killing", deepcopy(params), 'orange', '--'),
    ("Poor Expansion", deepcopy(params), 'red', '-.'),
    ("High Exhaustion", deepcopy(params), 'purple', ':')
]

# Modify parameters for failure modes
scenarios[1][1].k_kill = params.k_kill * 0.3  # Weak killing
scenarios[2][1].R_T = params.R_T * 0.5  # Poor expansion
scenarios[3][1].alpha = params.alpha * 5  # High exhaustion

for label, p, color, lstyle in scenarios:
    res = simulate_cart(p, T0, CT0, CM0, B0, Y0,
                        therapy_times, therapy_doses,
                        t_end=150)

    t = res['time']

    # Tumor - ALL 4 LINES
    axes[0, 0].semilogy(t, res['T'], color=color, lw=2.5, ls=lstyle, label=label, alpha=0.9)

    # Effector - ALL 4 LINES
    axes[0, 1].plot(t, res['C_T'], color=color, lw=2.5, ls=lstyle, label=label, alpha=0.9)

    # Memory - ALL 4 LINES
    axes[1, 0].plot(t, res['C_M'], color=color, lw=2.5, ls=lstyle, label=label, alpha=0.9)

    # Cytokines - ALL 4 LINES
    axes[1, 1].plot(t, res['Y'], color=color, lw=2.5, ls=lstyle, label=label, alpha=0.9)

# Formatting
axes[0, 0].axhline(DETECTION_THRESHOLD, color='gray', ls=':', lw=2)
axes[0, 0].axhspan(1e0, DETECTION_THRESHOLD, color='lightgreen', alpha=0.2,
label='Below detection')
axes[0, 0].set_ylabel('Tumor cells (log)', fontsize=12)
```

```
    axes[0, 0].set_title('Tumor Control Across Scenarios', fontsize=13, fontweight='bold')
    axes[0, 0].legend(loc='upper right', fontsize=10)
    axes[0, 0].grid(True, alpha=0.3)
    axes[0, 0].set_ylim(1e5, 1e10)

    axes[0, 1].set_ylabel('Effector CAR-T cells', fontsize=12)
    axes[0, 1].set_title('CAR-T Expansion Patterns', fontsize=13, fontweight='bold')
    axes[0, 1].legend(fontsize=10)
    axes[0, 1].grid(True, alpha=0.3)
    axes[0, 1].ticklabel_format(style='sci', axis='y', scilimits=(0, 0))

    axes[1, 0].set_xlabel('Time (days)', fontsize=12)
    axes[1, 0].set_ylabel('Memory CAR-T cells', fontsize=12)
    axes[1, 0].set_title('Memory Formation', fontsize=13, fontweight='bold')
    axes[1, 0].legend(fontsize=10)
    axes[1, 0].grid(True, alpha=0.3)
    axes[1, 0].ticklabel_format(style='sci', axis='y', scilimits=(0, 0))

    axes[1, 1].set_xlabel('Time (days)', fontsize=12)
    axes[1, 1].set_ylabel('Cytokines (arb. units)', fontsize=12)
    axes[1, 1].set_title('Cytokine Release (CRS Risk)', fontsize=13, fontweight='bold')
    axes[1, 1].legend(fontsize=10)
    axes[1, 1].grid(True, alpha=0.3)

    fig.suptitle('CAR-T Therapy: Success vs. Failure Modes',
                 fontsize=15, fontweight='bold', y=0.995)

    plt.tight_layout()

    if save_path:
        fig.savefig(save_path, dpi=300, bbox_inches='tight')

    return fig, axes



# METRICS COMPUTATION
#
================================================================================
========
```

```python
def compute_metrics(results, therapy_time, challenge_time=None):
    """Compute outcome metrics from simulation"""
    t = results['time']
    T = results['T']
    CT = results['C_T']
    B = results['B']
    Y = results['Y']

    metrics = {}

    metrics['final_tumor'] = T[-1]
    metrics['peak_effector'] = np.max(CT)
    metrics['peak_cytokine'] = np.max(Y)
    metrics['b_cell_nadir'] = np.min(B)

    post_therapy_idx = np.where(t > therapy_time)[0]
    if len(post_therapy_idx) > 0:
        first_post_idx = post_therapy_idx[0]
        # Use the known dose (2e6 for our baseline)
        dose_ct = 2e6  # This should match therapy_doses in your simulation
        metrics['expansion_fold'] = metrics['peak_effector'] / dose_ct
    else:
        metrics['expansion_fold'] = np.nan

    # Time to clearance
    below_detection = T < DETECTION_THRESHOLD
    if np.any(below_detection):
        first_below_idx = np.where(below_detection)[0][0]
        metrics['time_to_clearance'] = t[first_below_idx] - therapy_time
    else:
        metrics['time_to_clearance'] = np.nan

    # Tumor at rechallenge
    if challenge_time is not None:
        challenge_idx = np.argmin(np.abs(t - challenge_time))
        metrics['tumor_at_rechallenge'] = T[challenge_idx]
    else:
        metrics['tumor_at_rechallenge'] = np.nan

    return metrics
```

```
# PARAMETER SWEEP FUNCTION
#
================================================================================
========
def parameter_sweep(param_name, param_values, base_params,
                    T0, CT0, CM0, B0, Y0,
                    therapy_times, therapy_doses,
                    challenge_times=None, challenge_doses=None,
                    t_end=500):

    from copy import deepcopy

    if challenge_times is None:
        challenge_times = []
    if challenge_doses is None:
        challenge_doses = []

    results_list = []
    metrics_list = []

    for val in param_values:
        p = deepcopy(base_params)
        setattr(p, param_name, float(val))

        res = simulate_cart(
            p, T0=T0, CT0=CT0, CM0=CM0, B0=B0, Y0=Y0,
            therapy_times=therapy_times,
            therapy_doses=therapy_doses,
            challenge_times=challenge_times,
            challenge_doses=challenge_doses,
            t_end=t_end
        )

        m = compute_metrics(
            res,
            therapy_time=therapy_times[0],
            challenge_time=challenge_times[0] if challenge_times else None
        )
```

```
        m[param_name] = float(val)

        results_list.append(res)
        metrics_list.append(m)

    metrics_df = pd.DataFrame(metrics_list)
    return results_list, metrics_df


# MAIN EXECUTION
#
================================================================================
========
if __name__ == "__main__":

    print("="*70)
    print("FIXED CAR-T THERAPY MODEL - REALISTIC EXPANSION DYNAMICS")
    print("="*70)

    out_dir = Path("CAR_T_Results_FIXED")
    out_dir.mkdir(exist_ok=True)

    p = CARTParameters()

    # Baseline scenario
    T0 = 2e6
    CT0 = 0.0
    CM0 = 0.0
    B0 = p.B0
    Y0 = p.Y0

    therapy_times = [42.0]
    therapy_doses = [2e6]
    challenge_times = [250.0]
    challenge_doses = [1e6]

    print("\n--- Running FIXED baseline simulation ---")
    results = simulate_cart(
        p,
        T0=T0, CT0=CT0, CM0=CM0, B0=B0, Y0=Y0,
```

```
    therapy_times=therapy_times,
    therapy_doses=therapy_doses,
    challenge_times=challenge_times,
    challenge_doses=challenge_doses,
    t_end=500
)

# Print key metrics
m = compute_metrics(results, therapy_times[0], challenge_times[0])
print(f"\nFixed baseline metrics:")
print(f"  Time to clearance: {m['time_to_clearance']:.2f} days (TARGET: 7-14 days)")
print(f"  Peak effector: {m['peak_effector']:.2e} cells")
print(f"  Expansion fold: {m['expansion_fold']:.1f}x (TARGET: 10-50x)")
print(f"  Peak cytokine: {m['peak_cytokine']:.2f} a.u.")
print(f"  B-cell nadir: {m['b_cell_nadir']:.2e} cells")
print(f"  Final tumor: {m['final_tumor']:.2e} cells")

# Generate baseline plots
print("\n--- Generating FIXED baseline plots ---")
plot_all_populations(results, therapy_times, challenge_times,
                     out_dir / "01_all_populations_FIXED.png")
plot_tumor_only(results, therapy_times, challenge_times,
                out_dir / "02_tumor_burden_FIXED.png")
plot_cart_populations(results, therapy_times, challenge_times,
                      out_dir / "03_cart_populations_FIXED.png")
plot_b_cells(results, therapy_times, challenge_times,
             out_dir / "04_b_cells_FIXED.png")
plot_cytokines(results, therapy_times, challenge_times,
               out_dir / "05_cytokines_FIXED.png")

# Parameter sweeps
print("\n--- Parameter Sweep 1: CAR-T Cytotoxicity (k_kill) ---")
k_kill_values = np.logspace(-8, -6, 15)
_, metrics_gamma = parameter_sweep(
    'k_kill', k_kill_values, p,
    T0, CT0, CM0, B0, Y0,
    therapy_times, therapy_doses,
    challenge_times, challenge_doses,
    t_end=300
)
```

```
plot_sweep_time_to_clearance(metrics_gamma, 'k_kill', 'CAR-T Cytotoxicity ($k_{kill}$)',
                             out_dir / "06_sweep_gammaT_clearance_FIXED.png")
plot_sweep_peak_cytokine(metrics_gamma, 'k_kill', 'CAR-T Cytotoxicity ($k_{kill}$)',
                         out_dir / "07_sweep_gammaT_cytokine_FIXED.png")
plot_sweep_peak_effector(metrics_gamma, 'k_kill', 'CAR-T Cytotoxicity ($k_{kill}$)',
                         out_dir / "08_sweep_gammaT_expansion_FIXED.png")

print("\n--- Parameter Sweep 2: Memory Reactivation (theta) ---")
theta_values = np.logspace(-8, -6, 15)
_, metrics_theta = parameter_sweep(
    'theta', theta_values, p,
    T0, CT0, CM0, B0, Y0,
    therapy_times, therapy_doses,
    challenge_times, challenge_doses,
    t_end=500
)

plot_sweep_time_to_clearance(metrics_theta, 'theta', 'Memory Reactivation (θ)',
                             out_dir / "09_sweep_theta_clearance_FIXED.png")
plot_sweep_peak_cytokine(metrics_theta, 'theta', 'Memory Reactivation (θ)',
                         out_dir / "10_sweep_theta_cytokine_FIXED.png")

print("\n--- Parameter Sweep 3: Effector Proliferation (R_T) ---")
R_T_values = np.linspace(0.4, 1.2, 15)
_, metrics_R_T = parameter_sweep(
    'R_T', R_T_values, p,
    T0, CT0, CM0, B0, Y0,
    therapy_times, therapy_doses,
    challenge_times, challenge_doses,
    t_end=300
)

plot_sweep_time_to_clearance(metrics_R_T, 'R_T', r"Effector Proliferation ($R_{T}$)",
                             out_dir / "11_sweep_R_T_clearance_FIXED.png")
plot_sweep_peak_cytokine(metrics_R_T, 'R_T', r"Effector Proliferation ($R_{T}$)",
                         out_dir / "12_sweep_R_T_cytokine_FIXED.png")
plot_sweep_peak_effector(metrics_R_T, 'R_T', r"Effector Proliferation ($R_{T}$)",
                         out_dir / "13_sweep_R_T_expansion_FIXED.png")
```

```
# Save metrics
metrics_gamma.to_csv(out_dir / "sweep_k_kill_metrics_FIXED.csv", index=False)
metrics_theta.to_csv(out_dir / "sweep_theta_metrics_FIXED.csv", index=False)
metrics_R_T.to_csv(out_dir / "sweep_R_T_metrics_FIXED.csv", index=False)


# COMPARISON PLOTS
#
================================================================================
=====

from copy import deepcopy

# Single dose: 2e6 cells at day 42
results_single = results

therapy_times_frac = [42.0, 46.0, 50.0, 54.0, 58.0]
therapy_doses_frac = [4e5, 4e5, 4e5, 4e5, 4e5]  # 5 x 400k = 2M total

p_frac = deepcopy(p)
p_frac.R_T_committed = 0.0
p_frac.tau_commit = 0.0       # Turn off commitment
p_frac.R_T = p.R_T * 0.70
p_frac.k_kill = p.k_kill * 1.05

results_frac = simulate_cart(
    p_frac, T0=T0, CT0=CT0, CM0=CM0, B0=B0, Y0=Y0,
    therapy_times=therapy_times_frac,
    therapy_doses=therapy_doses_frac,
    t_end=200
)

plot_dosing_comparison(results_single, results_frac,
                       therapy_times, therapy_times_frac,
                       out_dir / "14_STORY_single_vs_fractionated.png")

m_single = compute_metrics(results_single, therapy_times[0])
m_frac = compute_metrics(results_frac, therapy_times_frac[0])
```

```
    print(f"  Single dose: Peak cytokine = {m_single['peak_cytokine']:.1f}, Clearance =
{m_single['time_to_clearance']:.1f} days")
    print(f"  Fractionated: Peak cytokine = {m_frac['peak_cytokine']:.1f}, Clearance =
{m_frac['time_to_clearance']:.1f} days")

    if m_frac['peak_cytokine'] < m_single['peak_cytokine']:
        print(f"  → Cytokine reduction: {(1 -
m_frac['peak_cytokine']/m_single['peak_cytokine'])*100:.1f}%")
        print(f"  → Success! Resting T cells in multiple doses reduces CRS burden")
    else:
        print(f"  → Cytokine increase: {(m_frac['peak_cytokine']/m_single['peak_cytokine'] -
1)*100:.1f}%")
        print(f"  → Model insight: Committed proliferation prevents fractionation benefit")
        print(f"  → Cytokine increase: {(m_frac['peak_cytokine']/m_single['peak_cytokine'] -
1)*100:.1f}%")
        print(f"  → Note: Model shows tradeoffs depend on manufacturing strategy")

    # Generate sweep for tradeoff plot
    R_T_sweep_vals = np.linspace(0.3, 0.6, 10)
    R_T_sweep_results = []
    for R_T_val in R_T_sweep_vals:
        p_temp = deepcopy(p)
        p_temp.R_T = R_T_val
        res_temp = simulate_cart(p_temp, T0=T0, CT0=CT0, CM0=CM0, B0=B0, Y0=Y0,
                                 therapy_times=therapy_times, therapy_doses=therapy_doses,
                                 t_end=150)
        R_T_sweep_results.append(res_temp)

    plot_efficacy_toxicity_tradeoff(R_T_sweep_results, 'R_T', R_T_sweep_vals,
                                    out_dir / "15_STORY_efficacy_toxicity_tradeoff.png")


    plot_patient_variability_monte_carlo(p, T0, CT0, CM0, B0, Y0,
                                         therapy_times, therapy_doses,
                                         n_patients=100,
                                         save_path=out_dir / "16_STORY_patient_variability.png")

    plot_failure_modes(p, T0, CT0, CM0, B0, Y0,
                       therapy_times, therapy_doses,
                       save_path=out_dir / "17_STORY_failure_modes.png")
```

```
plt.show()
```