A Study on Optimization Measurement Policies for Quality Control Improvements in Gene Therapy Manufacturing

by

Michaela Starkey

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved June 2020 by the
Graduate Supervisory Committee:

Giulia Pedrielli, Chair
Jing Li
Teresa Wu

ARIZONA STATE UNIVERSITY

August 2020

ABSTRACT

With the increased demand for genetically modified T-cells in treating hematological malignancies, the need for an optimized measurement policy within the current good manufacturing practices for better quality control has grown greatly. There are several steps involved in manufacturing gene therapy. These steps are for the autologous-type gene therapy, in chronological order, are harvesting T-cells from the patient, activation of the cells (thawing the cryogenically frozen cells after transport to manufacturing center), viral vector transduction, Chimeric Antigen Receptor (CAR) attachment during T-cell expansion, then infusion into patient. The need for improved measurement heuristics within the transduction and expansion portions of the manufacturing process has reached an all-time high because of the costly nature of manufacturing the product, the high cycle time (approximately 14-28 days from activation to infusion), and the risk for external contamination during manufacturing that negatively impacts patients post infusion (such as illness and death).

The main objective of this work is to investigate and improve measurement policies on the basis of quality control in the transduction/expansion bio-manufacturing processes. More specifically, this study addresses the issue of measuring yield within the transduction/expansion phases of gene therapy. To do so, it was decided to model the process as a Markov Decision Process where the decisions being made are optimally chosen to create an overall optimal measurement policy; for a set of predefined parameters.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. Giulia Pedrielli for her support and guidance on this study. Her help has been immeasurable, and her comments on this manuscript have been extremely useful. The other members of my committee, Dr. Jing Li and Dr. Teresa Wu, also deserves recognition for their encouragement and consultation.

I would also like to thank Mr. Gaurav Sharma for his sound advice and Mr. Colin Lynch for his assistance in formulating and coding the sensitivity analysis.

Finally, I want to thank my grandmother Fran and my three cats, Mia, Maya, and Mona for giving me the home support I needed to complete this study.

TABLE OF CONTENTS

Page

LIST OF TABLES iv

LIST OF FIGURES v

CHAPTER

1 INTRODUCTION 1

1.1 Background and Motivation 3

1.2 Challenges to Consider 5

1.3 Thesis Structure 6

2 METHODOLOGIES 7

2.1 CAR-T Manufacturing Process 7

2.2 Problem Formulation 9

2.2.1 Markov Decision Process Components 10

2.3 Learning Model for the Yield Function 11

2.3.1 State Transition 11

2.3.2 Example for Bayesian Update 12

2.4 Policy Improvement 14

3 NUMERICAL RESULTS 20

3.1 Design of Experiments 20

3.2 Results and Discussion 23

3.2.1 Sensitivity Analysis 28

4 CONCLUSIONS AND FUTURE WORK 31

4.1 Future Work 32

REFERENCES 33

LIST OF TABLES

Table Page

2.1 Bayesian Update Example Parameter Values 12

3.1 Parameters and Associated Levels for Sensitivity Analysis 21

3.2 Parameter Level Combinations and Associated Mode Optimal Policies . 27

3.3 Overall Mode Policies Generated by MDP 28

3.4 Comparison of Performance Across Optimality Categories 28

LIST OF FIGURES

Figure		Page
2.1	General Manufacturing Process Map with Tests	9
2.2	Initial Prior and Likelihood (left) for Resulting Initial Posterior (right)	13
2.3	Initial Sigmoid Yield Curve (top) and Yield Curve with Updates (bottom)	14
3.1	Free Parameter Significance Levels per Optimality Category	24
3.2	Mode Policy per Optimality Category	26
3.3	Normalized Total-order Sobol Index per Parameter	30

Chapter 1

INTRODUCTION

CAR T-cell therapy is a targeted product which does not require bone marrow matching. It is highly efficient and specifically kills malignant cells [13]. CAR T-cell therapies typically use autologous T-lymphocytes, although clinical trials are evaluating allogeneic products (i.e. products harvested from a healthy donor to then the patient), to treat patients with hematologic malignancies and selected solid tumors. These adoptive cell therapies consist of T-cells manipulated ex-vivo (outside of patient) to target tumor antigens before infusion back into the patient [3]. CARs (i.e. chimeric antigen receptors) have two important roles: first, to recognize and bind to a tumor cell via a ligand, and second, to transfer intracellular signals that result in T-cell activation. More than 1,000 patients have received CAR T-cells in the United States alone, with more than 250 CAR T-cell trials listed on ClinicalTrials.gov, primarily from the United States and China [3]. This cellular manipulation portion of the gene therapy development is in the realm of quality control, as it houses the activation, transduction, and expansion processes which can be a major source of failing FDA regulations/ general quality expectations.

Developing a quality control heuristic will be a difficult feat. Many different factors play a role in the manufacturing of the therapy; mostly due to the fact it is a "living-drug". These factors can be in the CAR design, gene introduction, cell culture, cell purification technologies from each manufacturer, and also patient differences (since it is an autologous therapy). In terms of the patient factors, these can be classification/stage of tumor, time interval between collection of initial sample

and current treatment regimen, routine physical examination items associated with the treatment, but also the total number of peripheral blood cells or T-cells with a specific phenotype, pathogen screening, etc. It is clear that the materials used for production are directly related to the quality of the products [15].

There is more than enough motivation to research and establish a good, standardized quality management system for production materials, including risk assessment of their use, auditing of suppliers, and quality testing [15]. The justification for creating such a heuristic is to have control of the overall process and product quality of the CAR T-cells in order to ensure the repeatability of the production process and batch-to-batch consistency of the final products [15]. By instituting testing methods for certain steps within the process and establishing acceptance criteria for the in-process testing, in combination with the release testing of final products, achieving overall control/quality is possible. Literature suggests that a simplified release test be done (this is for the end of the entire manufacturing process) and it should be clearly explained. The missing information will be supplemented by the testing done throughout the process itself (and also by the process validation). For the whole process of CAR T-cell product preparation, comprehensive process characterization and process validation of multiple continuous batches (at least three batches) of production are required [15]. Currently, a complete quality control test procedure for gene therapy is not possible due to the short shelf life of the product and the time-sensitive nature of delivering the treatment to the critically ill patient. For this reason, a simulation was developed to test several factors that can affect the way quality testing is done.

1.1 Background and Motivation

Within each of the activation, transduction, and ex vivo expansion processes, many factors have to be controlled in order to minimize contamination risks that can then cause infections in the infused patient. Furthermore, ensuring the therapy doesn't exceed the level of toxicity allowed nor contains excess cytokines is an additional difficulty to the manufacturing process. CAR T-cells carry a more extensive toxicity profile than dendritic cell vaccines, namely the potential for tumor lysis; cytokine release syndrome requiring intensive unit care; neurologic toxicity ranging from encephalopathies characterized by aphasia, seizures, and in rare instances, fatal cerebral edema; as well as unusual immunologic manifestations such as hemophagocytic lymphohistiocytosis (HLH) [9]. The primary side effect of CAR-T therapy, however, is cytokine release syndrome (CRS), also known as a cytokine storm, a syndrome caused by excessive cytokine release that is usually observed in CAR-T related therapy [13]. Cytokine release syndrome, a systemic inflammatory response observed with monoclonal antibody drugs and adoptive T-cell treatments, has become a major issue for CAR-T therapy [12]. Studies have also shown a positive correlation between cytokine surges with associated clinical toxicities in patients treated with CD19-targeted CAR T-cells to the degree of tumor burden at the time of CD19-targeted T-cell infusions [4].

Other side-effects of the CAR-T therapy administration include myalgia, respiratory distress syndrome, B-cell deficiency, capillary leak syndrome, coagulation disorders, nausea/ vomiting, diarrhea, infection due to contamination during manufacturing and/or infusion site, tumor lysis syndrome, neurotoxicity, and macrophage activation syndrome as a result of severe CRS. In particular, macrophage activation syndrome and/or severe levels of CRS have shown rapid response to the use

of steroids, a study shows. However, steroids can greatly damage CAR T-cells and even eliminate them [13]; patients always relapse when treated with steroids and the relapse may or may not be handled by the CAR T-cells existing in the body (if any are left post steroids).

Outside of direct patient health-risk, there is a financial burden that accompanies gene therapy. Even though CAR T-cell therapy is reported to be covered by Medicare/Medicaid Services, this does not mean costs accrued at a hospital providing the therapy will be covered as well. On October 1, 2018, the 2019 Inpatient Prospective Payment System rule went into effect, giving CAR T-cell therapy a diagnosis-related group with higher payment, MS-DRG 16, and the ability to receive additional reimbursement for the CAR T-cell product through a New Technology Add on Payment (NTAP). To capture the full amount of the NTAP ($186,500), the CAR T-cell product charge must be set high enough that when it is reduced by the facility's cost-to-charge ratio, the NTAP can be recovered [3]. Also, since Medicaid varies from state to state, a state's Medicaid plan may only cover CAR T-cell therapy if performed as an outpatient, whereas another state may require an inpatient admission. Some Medicaid plans provide reimbursement on a per diem basis only that is inadequate to cover hospitals' costs; others reimburse only a percentage of the CAR T-cell product cost [3].

From start to finish (i.e. from harvest to infusion), a patient can receive a bill upwards of $500,000 for a single infusion treatment. This, in part, is not only due to hospital charges or insurance coverage, but because of the initial cost of intellectual property generation, research and development, and creation of manufacturing facilities. If CAR T-cell therapy is to be widely used across hematologic malignancies and solid tumors, the price of the product will need to be lowered to compare more

closely to existing therapeutic options [3]. All in all, if the financial load that coincides with the gene therapy is ignored in quality management tactics, it will and presently does deter patients from this treatment option. Patients will choose current therapy regimes outside of gene therapy because they want to avoid carrying this financial burden themselves or having family/friends carry this burden if they pass during or after infusion; passing on the option of gene therapy could ultimately result in their death.

Given all of this information, it is clear to see the strong need for an efficient, safe, and affordable treatment option for a life-threatening condition. Developing a better measurement policy for the transduction and expansion processes within the bio-manufacturing of gene therapy is crucial to achieving this goal. We aim to accomplish this with the methodologies explained in the up-coming chapters.

1.2 Challenges to Consider

The main objectives for this study in terms of preset goals are explained below.

Research Goal: Develop a Markov decision process model that represents the measurement part of the quality control process for individualized cancer therapy. The result is then incorporated into BIO-Man, a simulation platform to be adopted to evaluate solutions to scale up production of gene therapies at a higher quality.

Long Term Goal: To achieve an overall quality control process that implements this study's optimal measurement policies with current FDA regulations.

Developing a simulation model of this nature is difficult and comes with a few constraints.

Individualized Nature: Yield is almost entirely dependent on patient cells' response to reagents during the transduction/expansion manufacturing steps. As such, new data for every simulation must be generated because of the autologous nature of the therapy-product.

Timing Constraints: If preparations and/or testing take too long, cells will die. In general, the longer the wait the lower the yield. Also, waiting too long can be critical for a terminal patient.

General Constraints: I have limited knowledge on dependent factors affecting yield/ production in general as I am not a subject matter expert.

1.3 Thesis Structure

The content of thesis is organized as follows :

1. **Chapter 2:** We describe the technical approach to model and optimize the yield measurement process.
2. **Chapter 3:** Walks-through the experimentation phase of the study and discusses results of the experiments.
3. **Chapter 4:** Concludes the thesis and covers next steps/ future work.

Chapter 2

METHODOLOGIES

In this section, we detail the technical approach for the modeling and optimization of the measurement process. In particular, we model measurement as a Markov Decision Process (MDP) integrated with a Bayesian update function for the update of the yield statistical learning model. This then feeds into a linear optimization program to maximize the Q-factor function at each state for the entire action space which selects the highest factor for the associated action as the decision at time t. Subsequently, we discuss the key functions, yield curve approximation, and policy improvement of the built simulation.

2.1 CAR-T Manufacturing Process

As previously mentioned, in this work we aim to discover the optimal combination of parameters that help control the quality outcome of the therapy-product during the transduction/expansion manufacturing steps. In order to do so, we first present the CAR-T manufacturing process with current practices, which we will optimally monitor and discuss later on how to include the suggested optimal ranges for the parameters included in this study.

A general process of gene therapy manufacturing good manufacturing process (GMP) was taken from literature and is discussed in this section. The simplified version of the process is as follows: Isolate T-cells from patient's blood collected, T-cells then genetically modified in GMP (good manufacturing process). CAR's are delivered

to the T-cells in 3 ways: Viral Vectors (retroviral or lentiviral), transposon systems (Sleeping Beauty), or mRNA transduction. Then the CAR T-cells are amplified and infused in the patient at the end of the process. At present, gene modification in the production of CAR T-cells is mainly achieved by viral vectors (retroviral vectors or lentiviral vectors), transposition subsystems, or direct mRNA electroporation, so that CARs are expressed on the surface of T cells in order to specifically recognize and bind antigens on the surface of tumor cells, and finally kill the tumor cells [15].

Given this information there are 3 possible processes for CAR production to consider when formulating a GMP for total quality control. Also, dependent on the needs of the original T-cell harvest location, a stability test should be conducted for samples that need to be temporarily preserved during the production process. If there is no need to cryogenically freeze the cells, a sterility test would be needed to ensure no contamination or flaw occurred during the transportation portions (before infusion to the patient also). The issue explained in the literature is that the current sterility testing approach can take up to 2 weeks, but there are alternative quick-detection methods that are required in process control and release testing (i.e. gram staining or acridine orange staining) [15]. As mentioned earlier, validation is a necessary part of model formation, and before the quick methods are fully validated the gram and orange staining methods should be performed in parallel. It is recommended to perform a sterility test on samples taken from culture supernatant at regular intervals after three or four days of culture, covering 48–72hr before release. If contamination is found in the early stages of cell preparation, further preparation should be terminated [15]. The following state/process map (Figure 1) was created based off of all possibilities mentioned in the literature and it is a low-level version of the overall manufacturing process. The possible state space for the total quality control process

would be stability testing, termination of production, sterility testing, release testing, and validation [7].

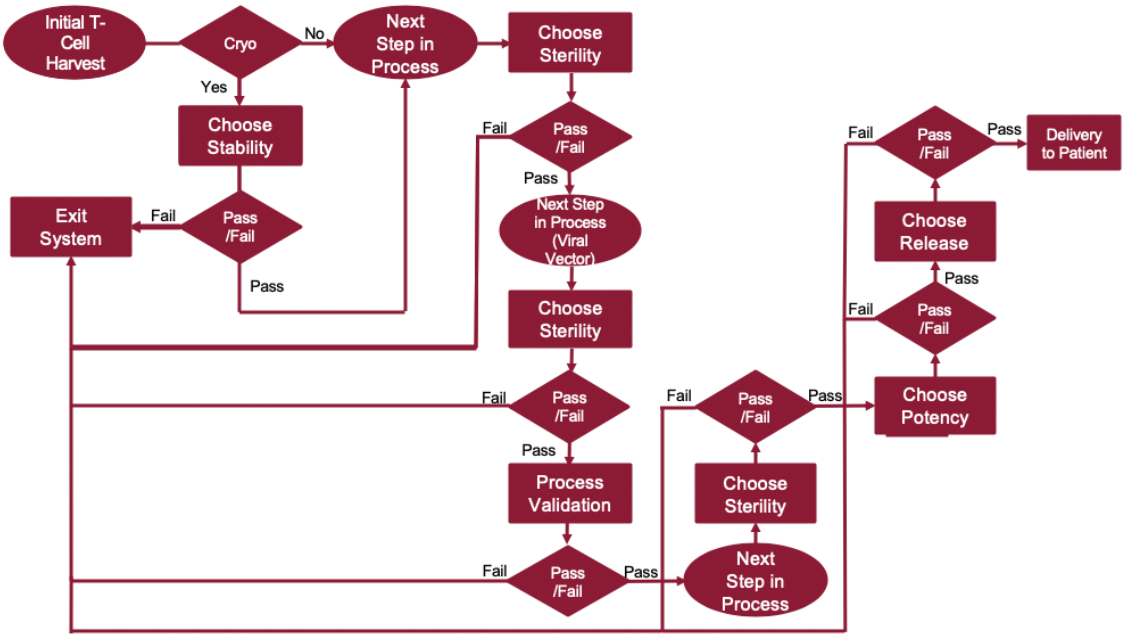

Figure 2.1: General Manufacturing Process Map with Tests

Even though the above information is about a total quality control process, this paper mainly pinpoints how to find an optimal measurement policy for the transduction/ expansion portion of the manufacturing process. The main testing done in this part of the process are the stability tests done by lab technicians. These tests are considered measurements in this study.

2.2 Problem Formulation

The objective of this section is to derive the Markov Decision Processes (MDP). MDPs are a mathematical framework for a decision-making stochastic control pro-

cess that has random outcomes and where the system is partly affected by an outside decision maker. In our case, it would be a lab technician in charge of the transduction/expansion manufacturing step only and will be the setting throughout this paper.

2.2.1 Markov Decision Process Components

States: States represent the percent yield achieved during the manufacturing process at the time of measurement. Such percent yield is a finite support continuous random variable. The state can be observed upon measurement, only at discrete times. In particular, we assume a measurement can be performed every 10 time-units. Timestamp, t, is based on the total length of the transduction process over the number of measurements to be taken. To be more specific, the transduction process takes 4 hours, converted to 240 minutes and divided into 24 equal parts; thus, resulting in 24 timestamps (Equation 2.1).

$$s_t = \{0, 100\}; t = \{0, 1, ...23\} \tag{2.1}$$

Actions: Actions are defined as the following: Measure then Continue (0), Measure then Stop the Process (1), Do Not Measure and Continue (2), and Do Not Measure and Stop the Process (3) (Equation 2.2). This is a simplified action set as this can get computationally complicated as more actions are added; i.e. the number of possible states and paths can grow exponentially.

$$A = \{0, 1, 2, 3\} \tag{2.2}$$

2.3 Learning Model for the Yield Function

In the Bayesian update the lower and upper bounds of the truncated normal distributions being used, for the prior and likelihood distributions, are set to 5 and 100 respectively. A vector of 5 sampled measurements, x, at time t, are generated from a single yield value taken from the assumed yield function, Y(t), at the same time t, but have the added variance as mentioned earlier. The mean of this vector is the mean for the likelihood distribution in the Bayesian update. The prior is another truncated normal distribution ranging from 5 to 100 with mean equal to the state yield value of the assumed yield function measured at time t. To obtain the posterior distribution of the Bayesian update, the prior probabilities are multiplied by the likelihood probabilities. It is then normalized by the sum of probabilities for each observation, given by:

$$P(\theta(t)|Y(t)) = \frac{P(Y(t)|\theta(t))P(\theta(t))}{\sum P(Y(t)|\theta_i(t))P(\theta_i(t))} \tag{2.3}$$

The mean of the resulting posterior distribution is $\hat{Y}(t)$.

2.3.1 State Transition

$\boldsymbol{P_{s,a}} = \boldsymbol{P}(\boldsymbol{s_{t+1}}, \boldsymbol{s_t})$: Probability transition matrix; The probability of going from any state $\boldsymbol{s_t}$ at time t, to the future state $\boldsymbol{s_{t+1}}$. The transition probabilities are not known in this problem, so it is given random values such that the rows and columns of the matrix sum to 1. The transition matrix is utilized in the value-function later on in this paper. The yield is a function growing in a sigmoidal manner over time [14].

$$\boldsymbol{Y(t) = \frac{\bar{Y}}{(1 + \frac{\bar{Y} - Y_0}{Y_0}) * e^{-rt}}} \tag{2.4}$$

The above equation, Equation 2.4, is the sigmoid function for the assumed yield where $\boldsymbol{\bar{Y}}$ is the yield threshold and $\boldsymbol{Y_0}$ is initial yield. Samples that are taken for the Bayesian update, are from the assumed yield generated by the function in Equation 2.4. We then take 5 measurements at each timestamp being evaluated and calculate the mean and variance of measurements that are inputted into Bayesian Update portion of the design.

We are trying to simulate the measurements taken by lab techs for each measurement action that has been decided. We keep a count of measurements for an entire policy and each time a measurement is taken the counter is increased by 1 and then we add randomness to emulate the general measurement variance that occurs in lab settings.

2.3.2 Example for Bayesian Update

The following example was simulated to showcase how the Bayesian update portion of the model functions. The parameter values used in this example are set to the following levels:

Table 2.1: Bayesian Update Example Parameter Values

$\boldsymbol{r}$ (Growth Rate)	$\boldsymbol{\gamma}$ (Discount Factor)	$\boldsymbol{C}$ (Instantaneous Cost)	$\boldsymbol{P(risk)_t}$
0.05	0.9	2	0.1

In this example, we modeled cell population growth with a sigmoidal growth curve in equation 2.4 for the transduction/expansion phases of manufacturing. The resulting sigmoid, with r = 0.05, is presented in the top graph of Figure 2.3.

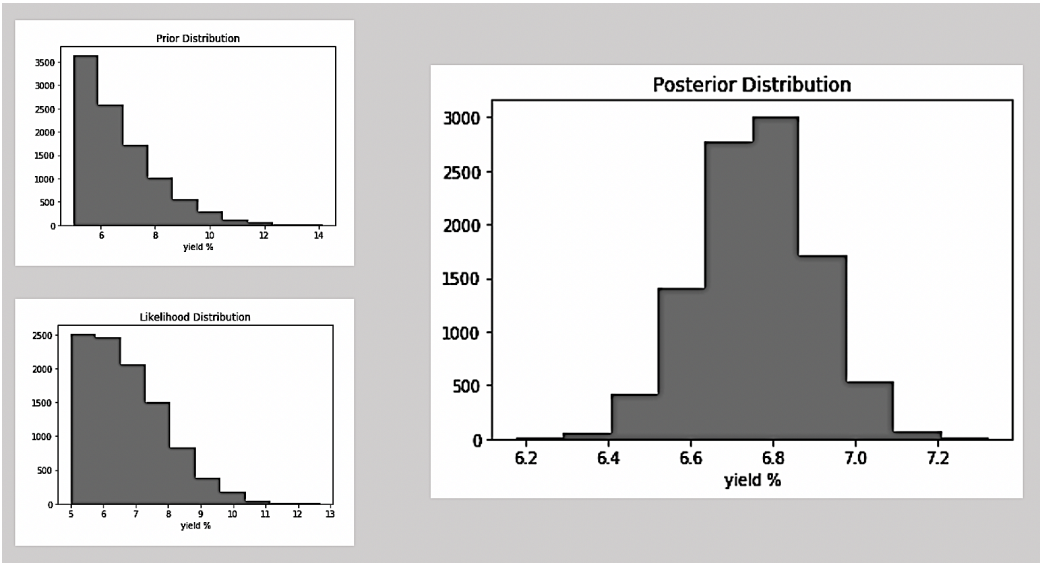

Figure 2.2: Initial Prior and Likelihood (left) for Resulting Initial Posterior (right)

The MDP made a measurement at t = 10, so we present the resulting likelihood and prior distributions from that measurement in Figure 2.2 on the left. These distributions are then multiplied together to produce the posterior distribution, which is presented on the right in Figure 2.2. The mean of this distribution gives $\hat{\boldsymbol{Y}}(\mathbf{10})$. Repeating this measurement for all measurements results in $\hat{\boldsymbol{Y}}(\boldsymbol{t})$, presented on the right in Figure 2.2.

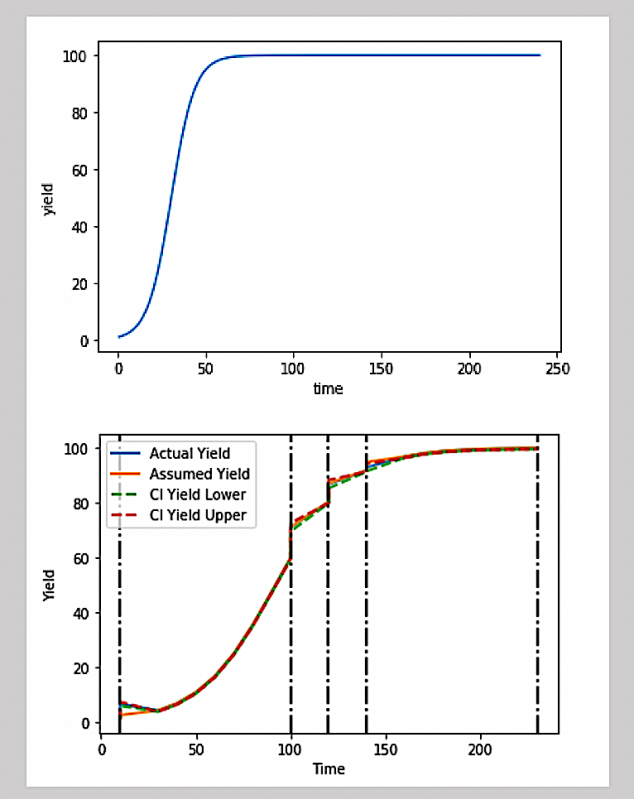

Figure 2.3: Initial Sigmoid Yield Curve (top) and Yield Curve with Updates (bottom)

In Figure 2.3 above, the bottom graph shows the resulting fitted yield of the MDP after all the Bayesian updates are completed. The dotted lines show the time stamps at which measurements were taken and updates occurred, the blue curve shows the actual yield, and the orange line gives the fitted yield.

2.4 Policy Improvement

$\boldsymbol{\pi}$: A policy; "optimal" when reward is maximized and costs/risks are minimized, making the policy greedy choosing, for each state s at time t, the action with maximum Q-factor [10].

$$\boldsymbol{\pi^*} = \text{optimal set of actions in a policy} \tag{2.5}$$

$\boldsymbol{V^{\pi}(s, a)}$: The state-value function; defines the value of a state as the expected cumulative future discounted reward starting from that state for the associated action at the specific time for the entire generated policy (Equation 2.6). It is based on the Bellman iterative equation from MDP literature.

$$\boldsymbol{V_i^{\pi}(s, a) = E\left[\sum_{t=1}^{T} \gamma^t R_{a_t}(s_t, s_{t+1})\right]} \tag{2.6}$$

Costs: Costs considered at this time, specifically for cost of measurement, can change depending on user and fluctuating prices of the test used for measurements and will be varied in the experiments section of this work.

Instantaneous Cost: An instantaneous cost occurs when the action 0 or 1 is chosen, representing the decision to measure which incurs the cost of testing (cost of materials and labor).

Discounted Future Costs: It is important to note that the other additional costs are formulated in the reward function to calculate the relative cost to a penalty cost for each measurement as well. This means that for every measurement there will be an additional penalty accounted for to help deter the model from making more measurements than necessary; since it runs the risk of contamination and lowering the yield that must be taken into account (another negative cost). The final cost being considered is the positive cost associated with the information gain when tak-

ing the action of "Measure". This is to help offset the negative costs of making the measurement in the first place and for the risk of contaminating the therapy-product; thus, the objective function contains this reward function that seeks to maximize the overall reward for a policy.

$\boldsymbol{Q^{\pi}(s,a)}$: The Q-factor; Takes the action a, in state s at time t, under the policy $\boldsymbol{\pi}$ and it returns the utility of that state-action pair. $\boldsymbol{Q^{\pi}(s,a)}$ is defined as the expected return starting from s, taking the action a and following policy $\boldsymbol{\pi}$ (Equation 2.7).

$$\boldsymbol{Q^{\pi}(s,a)} = \text{argmax}(\boldsymbol{Q[s_t,a_0], Q[s_t,a_1], Q[s_t,a_2], Q[s_t,a_3]}) \tag{2.7}$$

To explain $\boldsymbol{Q^{\pi}(s,a)}$ further, the framework uses Equation 2.8 to calculate the each of the 4 elements in the $\boldsymbol{Q^{\pi}(s,a)}$ equation in Equation 2.7.

$$\boldsymbol{Q_i(s_t,a)} = \sum_{\boldsymbol{a=0}}^{\boldsymbol{3}} \boldsymbol{P_{s,a}R(s_t) + \gamma P_{s,a}V^{\pi}(s_t,a)} \tag{2.8}$$

Where,

$\boldsymbol{Q(s_t,a)}$: Action-value function in state s at time t with action a

$\boldsymbol{P_{s,a}}$: Transition probability matrix; Contains random numbers between 0 and 1, and whose rows/columns sum to 1

$\boldsymbol{R(s_t)}$: Reward value in state s at time t

$\boldsymbol{\gamma}$: Discount factor

$\boldsymbol{V^{\pi}(s_t,a)}$: State-value function at time t

After completing the reward calculations for the state-value function in Equation 2.6, the Python 3.7 package 'PuLP', is being utilized and it takes in the reward, gamma, and transition matrix to solve the linear program. The evaluated linear program output is V, the state-value function result.

Before covering the reward function, one of the major components of the objective function in the linear program, there are a few parameters and constraints that should be considered. The parameters being probability of contamination, rate of increase for the sigmoid function in data generation, discount factor on reward, instantaneous cost, and information gain is the action being evaluated is measure related. As for the constraints, one that we are considering is that the rate of increase for the sigmoid function, r, cannot ascertain values higher than 0.2 as that can make the cells volatile/ unstable and is difficult biologically to obtain in manufacturing. Another constraint is that if the probability of contaminating the product is higher than 0.4, we should choose to not measure the product during the transduction or expansion phases. This comes with a few additional issues, one being you have no information on how the yield curve is behaving and thus not able to determine if you need to add more viral vector serum to increase the CAR attachment, the other issue being since you don't know how the cells are behaving the manufacturer is not able to determine if the yield threshold has been met. The yield threshold is a critical component of the therapy production because the patient's physician determined the amount of therapy required for their specific patient and failing to produce the amount needed could mean certain death for the patient and a large bill of unnecessary costs.

The reward function portion of the MDP design considers instantaneous cost of measuring as well as the cost associated with probability of contamination and the

reward for taking a measurement to gain information about how the real yield is behaving. It also considers a discount factor on how confident the MDP should be when making a decision based on the reward (Equation 2.9).

$$\boldsymbol{R_t = \gamma_t * R_{a_t}(s_t, s_{t+1}); a_t \in \{0, 1, 2, 3\}}; \text{ Action chosen at time t} \tag{2.9}$$

The function implemented in the model acts as the 'true' reward function, without the discount factor, is shown below:

$$\boldsymbol{R_{a_t}(s_t, s_{t+1}) = \begin{cases} -C + [C(1 + \Delta CC_t)] - [C(1 - P(risk)_t)] & a_t = 0 \\ -C & a_t = 1 \\ 0 & a_t = 2, a_t = 3 \end{cases}}$$

Where,

$\boldsymbol{R_{a_t}}$: Reward at time t for the action a

$\boldsymbol{C}$: Instantaneous cost

$\boldsymbol{\Delta CC_t = \frac{Y(t) - Y(t-1)}{100}}$; Change in percent yield

$\boldsymbol{a_t}$: Action taken at time t

$\boldsymbol{P(risk)_t = P_t(risk|a_t, TCM_t) = P_{t-1}(risk|a_t, TCM_t) + P_0(TCM_t)}$; Probability of contamination

$\boldsymbol{TCM_t}$: Total count of measurements taken until time t

$\boldsymbol{P_0}$: Initial probability of contamination

A final cumulative reward for a simulated policy is the expected utility based

on the actions taken for that path, expected costs/penalties for the path. will be calculated and that will determine which actions on which paths to take as a good policy.

Chapter 3

NUMERICAL RESULTS

In this chapter, the experimentation and sensitivity analysis conducted will be covered initially, then the results and discussions will be included before discussing conclusions and future work in Chapter 4.

The objective of this analysis is to determine the effect the free parameters have on policy selection and to separate optimal policies from sub-optimal policies.This will ultimately result in a recommendation for the range free parameters should be restricted to. If restricting these parameters is not possible, then we have further recommendations for policies to follow within different parameter ranges.

Chapter 3.1 will cover the design of the numerical simulations, discussing what the free parameters of the MDP are, what their respective ranges are, and how we define optimality. Chapter 3.2 will cover the numeric differences in policies across different levels of optimality and how optimality is influenced by the free parameters. Finally, Chapter 3.2.1 will cover the formal sensitivity analysis of the MDP.

3.1 Design of Experiments

There are several free parameters in this process that drive the decision to choose one of the generated policies over another; by this it is meant that the model will generate different policies for the set of values you input in addition to the randomized data being created for the yield function. While the created simulation model can

be constrained by experimentally derived values of these aforementioned parameters, these values are often uncertain in real-life scenarios. For instance, the probability that a medical instrument is contaminated by mycoplasma and could thus infect a CAR T-Cell harvest can vary widely between hospitals despite the prevalence of sterilization technology. Mycoplasma contamination rates usually vary between 15% and 30% but could be high as 70% [5]. We therefore simulate the decision process across a wide array of parameter values in order to estimate the degree to which these parameters determine the outcome of the MDP. We did not have access to real data, so we treat the rate of increase as a free parameter over the interval (0.04, 0.5) in the random data generation to achieve a wide range of scenarios. The parameters we vary, as well as their respective ranges, are given in Table 3.1.

Table 3.1: Parameters and Associated Levels for Sensitivity Analysis

$\boldsymbol{r}$ (Growth Rate)	$\boldsymbol{\gamma}$ (Discount Factor)	$\boldsymbol{C}$ (Instantaneous Cost)	$\boldsymbol{P(risk)_t}$
(0.04,0.5)	(0,1)	(0,1000)	(0,1)

Specifically, we focused our analysis on 4 parameters that we freely varied in this study. The first is the sigmoidal growth rate (r), which we used to model the population growth of CAR T-Cells in solution. While in theory this parameter can take on any positive real number (as we assume cell count always increases with time), in practice we found that extremely low values never result in a yield of 100% throughout the course of the simulation, while high values represented biologically unfeasible scenarios where cell counts almost instantly achieved maximum yield. We found that values between 0.04 and 0.5 were reasonable given the length of time our simulation represents (240 minutes).

The second parameter we varied was the probability of contamination for every measurement. Regardless of how clean the instruments are, there is always some non-zero probability that the instrument will be infected. Conversely, this probability will also never be 1, so we simulate the range between 0 and 1.

The third parameter is the instantaneous cost of the measurement, or the monetary cost of paying a technician to make the measurement, the resources extended to make that measurement, and so on. We chose a range between 0 and an arbitrarily high $1000 so the MDP could identify the cost at which the sum of this cost and the contamination risk outweigh the benefits of measuring at all.

Finally, we varied the discount factor ($\boldsymbol{\gamma}$), which quantifies how important future reward is to the current state. This influences the decision being made in the current state based on future rewards. In our case we only consider the next state. We varied this value between 0 and 1 as this is the natural range of this parameter.

There were 3 quantifications taken for every policy generated by the MDP to use when we compare all the generated policies: stop time ($\boldsymbol{t_{stop}}$), mean squared error ($\boldsymbol{MSE}$), and total number of measurements ($\boldsymbol{TCM}$). Stop time is taken when the process is stopped by either the action 1 or by action 3. Mean squared error is between the actual yield and the assumed yield (the yield curve being updated). The total number of measurements are the count of action 0 and 1 instances in a generated policy by the stop time.

It is important to also note how the optimal policy is being identified. In this study, the MDP generates an optimal policy based off of the defined set of param-

eter values. As mentioned earlier, the ranges for each parameter are being divided into segments and randomly combined with the other parameter values creating 1000 unique combinations in total. After running the analysis, we gather the combination of levels for each of the parameters and document their associated MDP policy. To simplify the MDP optimal policies for each of these, we then find the mode policy within each parameter level. This will be the optimal policy generated by the MDP for those level combinations. Using this information, we then recategorize these policies based off our own set of constraints on what we think is most optimal within that set of generated policies (see Table 3.2). The optimal policy, $\boldsymbol{\pi}^*$, is being chosen based on whether the probability of contamination is less than or equal to 0.4, total cost is less than or equal to \$854 (lower 50th percentile), and an $\boldsymbol{MSE}$ less than or equal to 2E-4 (lower 35th percentile). We define feasible solutions as anything that is less than optimal, but still meeting requirements (makes at least 1 measurement and reaches the yield threshold within time T). Last, we define a policy as infeasible when there are no measurements taken because there needs to be at least one measurement.

3.2 Results and Discussion

In this analysis, we want to determine which free parameters had a significant effect on optimality. To do this, we ran 6 different multinomial logistic regressions between the categories of optimality (i.e. if a policy was determined to be optimal, feasible, or infeasible) and the free parameters. These regression models represented all possible interactions, and then selected the most parsimonious model by comparing the AIC (Akaike Information Criterion) values that are generated with the 'multinom' (within the R 4.0 package 'nnet 7.3-14') and AIC functions in R 4.0 [11]. The model without interactions between free parameters had the lowest AIC value.

In this model, each parameter had a significant effect ($\boldsymbol{p < 0.001}$) on determining optimality. Gamma was the lone exception ($\boldsymbol{p > 0.05}$). The parameter space for each optimality category is given in Figure 3.1.

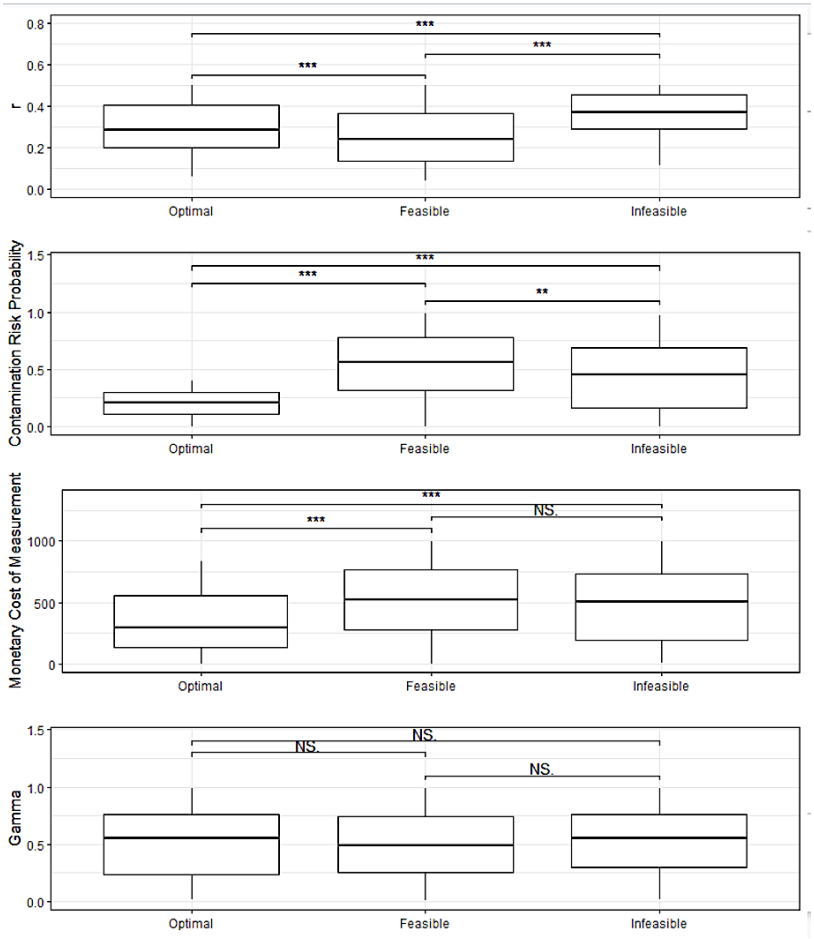

Figure 3.1: Free Parameter Significance Levels per Optimality Category

Each row represents the spread of a particular free parameter within each optimality category. ANOVAs were conducted between each optimality category across parameter space, and all but gamma were significant. A post-hoc Tukey test was performed for pairwise comparisons, where the significance levels (p-values) are presented on the graph as: NS = Not significant, ** $\boldsymbol{p < 0.01}$, and *** $\boldsymbol{p < 0.001}$.

Optimal policies are correlated with low values $\boldsymbol{P(risk)_t}$ and $\boldsymbol{C}$. However, it is also correlated with intermediate values of r. When r is small, therefore cell growth is slow, many measurements are necessary to consistently monitor yield until it reaches threshold. On the other hand, if r is too high, then yield increases to threshold before any measurements can be taken at t = 10.

We can see this effect in the mode policies for each kind of optimality (Figure 3.2). When r is low in the feasible case, measurements need to be taken throughout the duration of the experiment. Therefore the probability of choosing action 0 or 1 are higher at later time steps than at earlier time steps relative to the optimal case. When r is high in the infeasible case and the yield increases to threshold quickly, taking measurements only results in a small information gain. This small gain is outweighed by the cost of taking the measurement, so no measurements ever get taken. As we will see in the sensitivity analysis in Chapter 3.2.1, this effect is the primarily driven by r, as the model is the most sensitive to that parameter.

This effect results in different mean behaviors between the different optimality levels. The average values for various metrics are given in Table 3.4. From this table, we can see that optimal policies take fewer measurements than feasible policies, and thus they tend to stop sooner and are cheaper than feasible policies. They also tend to reduce error relative to feasible policies.

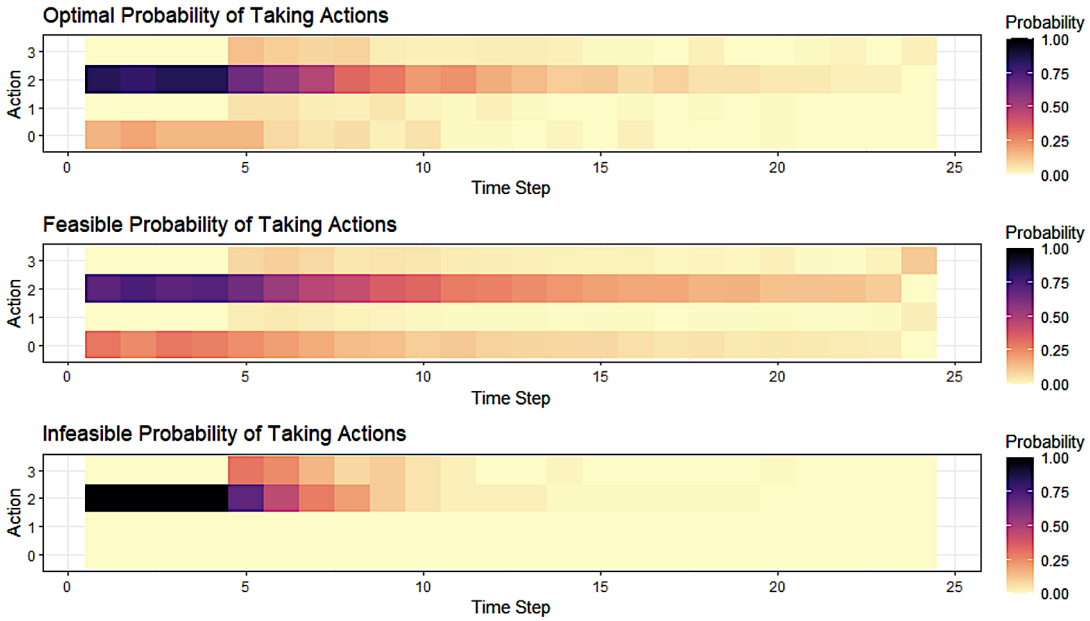

Figure 3.2: Mode Policy per Optimality Category

Visualization of the mode policy for each optimization level are shown in the figure above (Figure 3.2). The top panel gives the optimal mode policy, the middle gives the feasible policy, and the bottom gives the infeasible policy. The rows within each panel represent an action taken at the time step given in the y-axis. The color gives the probability of doing that action at that time step. At the beginning, the sum of the column = 1, however columns past t = 5 do not sum to 1 because not all policies last for more than 5 time steps.

Since the model is sensitive to the free parameters it will produce very different policies based on the parameter settings. Some of these parameters might be modifiable within the manufacturing process, however they may also be unalterable constraints (e.g. technicians might be able to control the growth rate by adding more viral vector solution or they cannot). We have therefore produced a set of mode policies for different parameter combinations in Table 3.2. This table can act as a

guideline for technicians looking to measure yield in the expansion/transduction processes.

Table 3.2: Parameter Level Combinations and Associated Mode Optimal Policies

r	γ	C	$P(risk)_t$	Mode Policy	Optimal	Feasible
0	0	0	0	22220222222202022222223	Y	Y
0	0	0	1	22202022000222022222223	N	Y
0	0	1	0	20022222022222220202023	N	Y
0	0	1	1	22222002220222222222221	N	Y
0	1	0	0	22222002220222222222221	Y	Y
0	1	0	1	22222222222222222222203	N	Y
0	1	1	0	22222002002022022022223	N	Y
0	1	1	1	22222220022222222022223	N	Y
1	0	0	0	02223	Y	Y
1	0	0	1	22221	N	Y
1	0	1	0	22221	Y	Y
1	0	1	1	20223	N	Y
1	1	0	0	22202223	Y	Y
1	1	0	1	20223	N	Y
1	1	1	0	20223	Y	Y
1	1	1	1	022223	N	Y

In Table 3.2 above, parameter level combinations with their associated mode optimal policies for different parameter combinations generated from the MDP are enumerated. The level 0 corresponds to a low value (lower 50th percentile) within the ranges for the parameters and level 1 corresponds to a high value (upper 50th percentile) within that same ranges for the parameters. The lower 50th percentile range of r is [0.04, 0.09), and the upper 50th percentile range is [0.09, 0.5]. $\boldsymbol{\gamma}$ lower range = [0, 0.5), upper range = [0.5, 1]. $\boldsymbol{C}$ lower range = [0, 500), and the upper range = [500, 1000]. $\boldsymbol{P(risk)_t}$ lower range = [0, 0.5), and the upper range = [0.5, 1]. The far right columns show whether or not a particular policy is optimal and/or feasible (if both are denoted with 'N' for no, then the policy is infeasible).

Table 3.3: Overall Mode Policies Generated by MDP

	POLICY	Optimal	Feasible
Overall Mode Policy	22223	N	N
Optimal Mode Policy	22221	Y	Y

Additionally, we included a table (Table 3.3) that shows the overall mode policy (pooled mode policy) and the overall optimal mode policy and a table that summarizes the main differences across optimality categories (Table 3.4).

Table 3.4: Comparison of Performance Across Optimality Categories

Optimality Category	$\boldsymbol{t_{stop}}$	$\boldsymbol{TCM}$	$\boldsymbol{C_{total}}$	$\boldsymbol{MSE}$
Optimal	85.93	1.35	\$426.51	5.22E-5
Feasible	118.1	3.2	\$1639.17	2.94E-4
Infeasible	69.22	0	0	0

Each cell in Table 3.4 gives the average value of the metric given in the column for the policy optimality type given in the row. Recall that $\boldsymbol{t_{stop}}$ is the stopping time stamp of the policy, $\boldsymbol{TCM}$ is the total count of measurements taken in a policy, $\boldsymbol{MSE}$ is for the mean square error, and $\boldsymbol{C_{total}}$ in this summary table is for total cost of entire policy.

3.2.1 Sensitivity Analysis

In this section, we will discuss how the analysis is conducted by enumerating the general steps taken on how we achieved a completed analysis on this study.

In order to measure sensitivity, we perform a variance-based sensitivity analysis (sometimes called the Sobol method). Under this paradigm, sensitivity is understood as the amount of variance of an output caused by variance of an input, as well as

interactions between inputs. The proportion of variance caused by a single parameter is called its first-order Sobol index, while the contribution made by two parameters is the second-order index. The sum of first-order and higher-order indices for a particular parameter is called the total-order index, which is what we report here. We calculate these indices with the 'SALib' package in Python 3.7.3 [6]. Random values for each parameter were selected with the 'Saltelli.Sample' function in this package and were run for 1000 replications, Figure 3.3 below depicts the normalized total-order Sobol index for each parameter associated with the 3 quantifying metrics mentioned earlier that we are considering ($\boldsymbol{MSE}$, $\boldsymbol{TCM}$, and $\boldsymbol{t_{stop}}$). Since we are normalizing the total-order Sobol index, these values cannot be used to directly measure the true sensitivity. However, they can be used to approximate the effect size each parameter may have on the overall system. Each free parameter has the same relative effect on measurements, in this case however, the rate of increase, r, for the sigmoid yield function contributes the most to stop time, only some to total measurements, and even less to $\boldsymbol{MSE}$. In general, the rate r is the main source of variation in this simulation because of how the data is being generated (i.e. it is entirely dependent on the yield function values generated and the rate is what creates these values of yield).

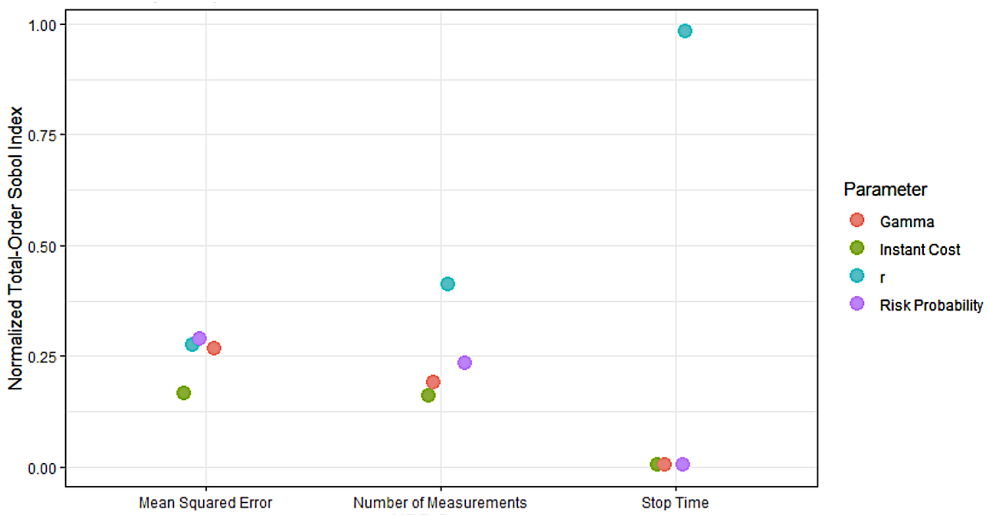

Figure 3.3: Normalized Total-order Sobol Index per Parameter

The optimality of a generated policy from the MDP can be measured by mean squared error ($\boldsymbol{MSE}$), total number of measurements ($\boldsymbol{TCM}$), and stop time ($\boldsymbol{t_{stop}}$) as mentioned earlier. During the sensitivity analysis, these quantifications are used and the normalized sensitivity of each free parameter included in both the model as well as the analysis (the color of each point) are given on the x-axis in Figure 3.3. Please note that jitter was added to avoid overlap between points in the figure to make it easier to analyze.

Chapter 4

CONCLUSIONS AND FUTURE WORK

In addition to the optimal policies discussed throughout this study, the Federal Drug Association (FDA) and the Foundation for the Accreditation of Cellular Therapy (FACT-JACIE) have published many standards regarding good manufacturing practices (GMPs) in the realm of gene therapy production that should be followed to maintain overall quality and adhere to all regulations. Moreover, they state that in the U.S., minimally manipulated cellular therapy products from first- or second-degree related donors are regulated under the 21 CFR 1271 Good Tissue Practices (GTP) regulations and section 361 of the Public Health ServiceAct [2], [1], [8]. They list general guidelines that must be followed when submitting accreditation applications for manufacturing facilities, both centralized and decentralized, that are wanting to manufacture gene therapy. These guidelines include documentation from all participating parties from harvest to infusion, training on a yearly basis (at minimum), descriptions of procedures that are submitted to the FDA and are available to all participating parties, etc. before a patient can receive the infusion.

We, as well as the industrial manufacturing labs that make these therapy-products, have pushed for innovation in many different areas and tools for bio-production. These tools are: Single Use Machines/ disposable, continuous/ batch processing (hybrid), and establish an overall quality control process with optimized measurement policies; which is the purpose of this study. We suggest that the policies generated in this model are taken in conjunction with already established GMPs to ensure that the FDA regulations are being adhered to and to achieve all the goals set in the begin-

ning of this work; Make the manufacturing of the gene therapy more efficient, safe, and affordable.

4.1 Future Work

More work can still be done to further improve the quality control process. We make the following recommendations:

Redefine 'error': Redefine the definition of 'error' as the difference in the areas underneath the actual and fitted yield curves. Optimality can in part be defined by this difference instead of the mean squared error ($\boldsymbol{MSE}$). To do this, we could implement a function to calculate the area between the two yield curves to get a more accurate idea of the measurement ability of the design framework and identify which policies are most optimal. Based on the minimization of that area in combination with satisfying the other optimality constraints.

Increase Parameter Space: Increase the number of free parameters, as some factors may have a stronger influence on policy than the ones considered here. One example of a parameter to include is the yield threshold.

Model Scaling: Scale models to fit the demand/size of the system in such a way as the models always adhere to all regulations.

REFERENCES

[1] "Immune effector cells: Accreditation manual", Foundation for the Accreditation of Cellular Therapy **1**.

[2] "Common standards for cellular therapies", Foundation for the Accreditation of Cellular Therapy **1**.

[3] B. Santomasso, J. W. K. R., C. Bachier and E. J. Shpall, "The other side of car t-cell therapy: Cytokine release syndrome, neurologic toxicity, and financial burden", American Society of Clinical Oncology Educational Book **39**, 433–444.

[4] Brentjens, R. J. e. a., "Cd19-targeted t cells rapidly induce molecular remissions in adults with chemotherapy-refractory acute lymphoblastic leukemia", Science Translational Medicine **5**, 177.

[5] Degeling MH, B. M. T. B., Maguire CA, "Sensitive assay for mycoplasma detection in mammalian cell culture", Analytical Chemistry **84**, 9, 4227–4232.

[6] Herman, J. and W. S. Usher, "An open-source python library for sensitivity analysis", Journal of Open Source Software **2**, 9.

[7] Janicijevic, I. e. a., "Using a markov chain for product quality improvement simulation", U.P.B. Sci. Bull. **76**, 1.

[8] Marks, P., "The fda's regulatory framework for chimeric antigen receptor-t cell therapies", Journal of ASCPT **1**, 428–430.

[9] Maus, M. V. and S. Nikiforow, "The why, what, and how of the new fact standards for immune effector cells", Journal for ImmunoTherapy of Cancer **5**, 1.

[10] Patacchiola, M., "Dissecting reinforcement learning-part.2", Github **2**, 9.

[11] Prabhu, A. e. a., "Bioproduction of succinic acid from xylose by engineered yarrowia lipolytica without ph control", Federation of European Biochemical Societies .

[12] Quan, Q., "An umbrella for the cytokine storm: Enabling precision detection of crs in car-t therapy", Journal of Clinical Oncology **37**, 15.

[13] S. Li, J. S. J. S., Z. Yang and C. Qian, "Adoptive therapy with car redirected t cells for hematological malignancies", Science China Life Sciences **59**, 4, 370–378.

[14] Webb, G., "Logistic models of structured population growth," hyperbolic partial differential equations", Hyperbolic Partial Differential Equations **2**, 9, 527—-539.

[15] Y. Li, L. Y., Y. Huo and J. Wang, "Quality control and nonclinical research on car-t cell products: General principles and key issues", Engineering **5**, 1, 122–131.