Applications of Data Science and Artificial Intelligence to Decision Making in Healthcare and Finance

by

Chi Heem Wong

B.Eng. (1st Class Hons), National University of Singapore (2014)
B.Soc.Sci. (1st Class Hons), National University of Singapore (2014)
S.M., Massachusetts Institute of Technology (2017)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Electrical Engineering and Computer Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2021

Author .....
Department of Electrical Engineering and Computer Science
March 1, 2021

Certified by .....
Andrew W. Lo
Charles E. and Susan T. Harris Professor of Sloan School of Management
Thesis Supervisor

Accepted by .....
Leslie A. Kolodziejski
Professor of Electrical Engineering and Computer Science
Chair, Department Committee on Graduate Students

Applications of Data Science and Artificial Intelligence to Decision Making in Healthcare and Finance

by

Chi Heem Wong

Submitted to the Department of Electrical Engineering and Computer Science
on March 1, 2021, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Electrical Engineering and Computer Science

Abstract

Decision-making requires timely and accurate information in order to understand the implications of the actions and to manage the potential risk. This thesis presents computational methods to quantify risk in drug development programs, address current challenges in health economics, and investigate and predict rare events in finance. The thesis is split into three major parts.

Part I addresses a core issue in accessing the risk and value of drug development programs: the probability of success (PoS). We introduce a Markov chain model of a drug development program that allows us to fill in missing data and infer phase transitions from clinical trial metadata. We investigate the PoSs across various therapeutic areas, and then conduct further analysis for areas that are of public interest (e.g., oncology, vaccine, and anti-infective therapeutic) in order to understand the bottlenecks in the drug development process.

Part II of the thesis focuses on the use of modeling and simulations to make informed predictions and drive policy-making in healthcare. One chapter in this Part is devoted to the use of data to estimate the financial impact of gene therapy in the U.S. between 2020 and 2035, while another chapter is dedicated to estimating the cost and benefit of various clinical trial designs for the development of a vaccine to prevent COVID-19.

Part III presents a novel 'big data' analysis and machine learning prediction model of panic selling behavior by retail investors. We document the frequency and timing of panic selling, analyze the demographics of investors who tend to freak out and panic sell, and determine if panic selling is a detrimental or optimal action financially. We also develop machine learning models to predict if an investor might panic sell in the near future given the demographic characteristics of the investor, their portfolio history, and the current and past market conditions.

Thesis Supervisor: Andrew W. Lo
Title: Charles E. and Susan T. Harris Professor of Sloan School of Management

Acknowledgments

There is an old Chinese proverb (**百年树人**), that says that it takes ten years to grow a tree, but a hundred years to bring up a person. The phrase implies that it takes a long time and tremendous effort to raise a child. This cannot be truer. First and foremost, I have to attribute most of my achievements in life so far to the sacrifices – such as giving up a career – that mom gave in order to raise me. Thank you, Tan Siew Hoon, my mom.

I would like to thank the many mentors who have guided and helped me over the years. I would like to thank my PhD advisor, Prof. Andrew W. Lo, for this opportunity and for the guidance that he has given me over the past few years. I would also like to thank Prof. Albert Tsui, my undergraduate thesis advisor in NUS, for his encouragements that gave me the confidence to explore and publish my ideas. Without them, I will not be where I am as a researcher. Shout-outs have to be given Prof. Leslie A Kolodziejski, Prof. Polina Gollands, Janet Fischer, and Jayna Cummings for navigating me through the different obstacles in the PhD program. Lastly, I want to thank my fellow collaborators – Prof. Jonathan Gruber, Prof. Rena Conti, Dr. Sean Khozin, Dr. Kathryn Kaminski, Dr. Donald Berry, Scott Berry, Peter Hale, Dr. Christine Blazynski and the cannot-be-named people from the brokerage firm – for their guidance in the projects.

I have to mention and thank Ru-Chen Ang, a mentor who encouraged me to think big and gave me the confidence to pursue my dreams when I was just a lost college kid.

Pursuing a PhD in a foreign land is tough, but it isn't too bad when you have good friends around when you need them. Thank you Tien-Ju Yang, my roommate for tolerating me and my quirky habits. Thank you Wei-Ning Hsu, Ge Liu (Saber), Ruizhi Liao (Ray), Jeehyun Yang, Kosuke Yoshinaga, Ryuji Takagi, Gabriel Loke, Lisa Chou, "PG" Shen, Anthony Kim, Jeffrey Ho, John Tedrow, Manish Singh, Zied Ben Chaouch & Sarah Kefí, and all my friends for the good times. Lastly, thank you Kien Wei Siah for being my trusted co-worker and best friend over the years.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

Decision-making requires timely and accurate information in order to understand the implications of the actions and to manage the potential risk. This thesis presents computational methods to quantify risk in drug development programs, simulate and forecast possible outcomes of healthcare policies, and investigate and predict rare events in finance.

The thesis is split into three major parts.

Part I addresses a core issue in the assessment of the risk and value of drug development programs: the probability of success (PoS). While the PoSs of drug development programs are critical inputs to economic models that allow biopharma investors to make smart investment decisions and policymakers to allocate resources prudently, previous studies of success rates suffer from the lack of accurate information on phase transitions. The collection of such data is expensive, time-consuming, and susceptible to error. To overcome this, we introduce a Markov chain model of a drug development program – defined as the set of clinical trials with the same drug and indication – that allows us to fill in missing data and infer phase transitions from clinical trial metadata. We apply the model on a database of clinical trial metadata to create a dataset that is of a magnitude larger than any prior study. This allows us to estimate PoSs more accurately and to update them quickly when the data is refreshed. We investigate the PoSs across various therapeutic areas, and then conduct further analysis for areas that are of public interest (e.g., oncology, vaccine, and anti-infective therapeutic)

in order to understand the bottlenecks in the drug development process.

Part II of the thesis focuses on the use of modeling and simulations to make informed predictions and drive policy-making in healthcare.

One chapter in this Part will be devoted to the use of data to estimate the financial impact of gene therapy in the U.S. between 2020 and 2035. Gene therapy is a new class of medical treatment that alters part of a patient's genome through the replacement, deletion, or insertion of genetic material to treat a disease. While still in its infancy, gene therapy has demonstrated immense potential to treat and even cure previously intractable diseases. However, the high prices of the few approved treatments have raised concerns regarding affordability among budget constrained payers and patients alike. By building and using a new model that estimates the future number of gene-therapy approvals, the size of their potential patient populations, and the prices of these future treatments, we estimate the patient impact and spending on gene therapy on private and public sector payers. It is our hope that this study will clarify some of these unknowns, and allow policymakers, healthcare providers, insurance companies and patients to make better informed decisions about the future of this important therapeutic class.

A subsequent chapter will estimate the cost and benefit of various clinical trial designs for the development of a vaccine for the prevention of COVID-19. The COVID-19 pandemic has caused the deaths of hundreds of thousands, upended the lives of billions, and caused trillions of dollars in economic loss. It has been hoped that the rapid approval of an effective vaccine can reduce human suffering, save lives and allow economies to resume normal activities. There are several clinical trial designs, such as a randomized clinical trial (RCT), a vaccine efficacy RCT with an optimized surveillance period that maximizes the benefits of the trial (O-RCT), an adaptive clinical trial (A-RCT) based on the group sequential design or a human challenge trial (HCT), that can be used to develop a vaccine. Each trial design requires a different length of time, number of patients and financial cost. In this study, we compare the costs and benefits – as measured by the number of deaths and infections avoided – of confirming the safety and efficacy of a COVID-19 vaccine using the four aforementioned clinical trial designs. We calibrate our simulations using a set of estimated epidemiological models for the SARS-CoV-2 virus (one for each of the 50 states and Washington, D.C.)

to determine attack rates and cumulative numbers of infections and deaths in the United States under various scenarios. We believe that the framework we proposed in this chapter will allow stakeholders to make better informed practical and ethical decisions in regards to accelerating the development of a COVID-19 vaccine amidst the ongoing pandemic.

Part III of the thesis presents a novel 'big data' analysis and machine learning prediction model of panic selling behavior by retail investors. Panic selling is often discussed in the financial press and media, but is rarely defined or quantified. To investigate this phenomenon, we developed a method to identify panic selling and apply it to one of the large dataset of brokerage account information spanning between 2003 to 2015 to examine panic selling and 'freakout' behavior . We document the frequency and timing of panic selling, analyze the demographics of investors who tend to panic sell and freak out, and determine if panic selling is a detrimental or optimal action financially. Finally, we develop machine learning models to predict if an investor might panic sell in the near future given the demographic characteristics of the investor, his portfolio history, and the current and past market conditions.

Part I

Quantifying Risk in Drug Development Programs

Chapter 2

Modeling the Probabilities of Success of Development Programs

2.1 Background

The typical clinical development program spans decades [1], costs billions of dollars [3] and faces great uncertainty on its returns. These attributes, together with the allure of higher returns in other sectors [8], make investors reluctant to invest in translational research that brings new drugs, medical devices and diagnostic tools from the bench to the bedside. This funding gap is sometimes referred to as the biotechnology "valley of death".

Many ideas, ranging from forging academic-industry partnerships [5] to creating mega-funds to diversify the investment risk, have been proposed to bridge the valley of death. All of them rely on being able to generate an acceptable risk-to-reward ratio – a metric that is highly dependent on the technical probabilities of success (PoSs) of the underlying projects.

Getting accurate and updated PoSs of drug development programs is not an easy task and previous studies of success rates have been constrained by the data in several respects. Abrantes-Metz et al. [1] surveyed 2,328 drugs using 3,136 phase transitions (e.g., from Phase 1 to Phase 2 in the approval process) while DiMasi et al. [2] studied 1,316 drugs from just 50 companies. A landmark study in this area, Hay et al. [4], analyzed 7,372 development paths of 4,451 drugs using 5,820 phase transitions. Smietana et al. [7] computed statistics using 17,358 phase transitions for 9,200 compounds while Thomas et al. [9] used 9,985 phase

transitions for 7,455 clinical drug development programs. In contrast, ClinicalTrials.gov, the clinical trial repository maintained by the National Institutes of Health (NIH), contains over 356,000 clinical trial entries submitted by various organizations as of 30 October 2020 (see www.clinicaltrials.gov). It is estimated that trained analysts would require tens of thousands of man-hours to manually assimilate its full information to produce PoS estimates.

We seek an alternative way to infer phase transitions quickly and accurately. This is done using a Markov chain model of a drug development program – defined as the set of clinical trials with the same drug and indication – that allows us to fill in missing data and infer phase transitions from clinical trial metadata. Applying the model on a database of clinical trial metadata allows us to create a dataset that is a magnitude larger than any prior study.

In this chapter, we will describe the drug development process, introduce our Markov chain model, define key terms, and explain the difference between how PoSs are estimated. The description of our data, results, and analyses of PoSs and other drug development statistics will be presented in the next chapter.

2.2 The Drug Development Process

A *drug development program*, also known as a drug development path, is the clinical investigation of the use of a drug for a disease. It typically consists of a sequence of clinical trials, separated into three phases. Phase 1 trials test mainly the safety and tolerance of a drug while phase 2 trials test the efficacy of the drug for a given indication. Phase 3 trials attempt to confirm the drug's efficacy for larger populations and against alternatives. Some trials involve the combination of two phases into a single protocol, denoted 'Phase 1/2' (a combination of Phases 1 and 2) and 'Phase 2/3' (a combination of Phases 2 and 3).

We say that a drug development program has reached phase i if it is observed, or can be inferred, that there is at least one trial in phase i. It is possible that a clinical trial can be repeated in multiple development paths. In Figure 2-1, we show an example in which a single phase 1 trial for a drug is involved in four different development paths, each targeting a different disease. It is not uncommon that the result of the phase 1 trial is used as supporting evidence for the safe use of a drug, allowing that drug to be used for different

indications without additional phase 1 testing. For example, hydroxychloroquine — already approved for the treatment of malaria — was tested for effectiveness against COVID-19 without another phase 1 clinical trial. There also exist clinical trials where different drug combinations are tested for the same indication in different arms.

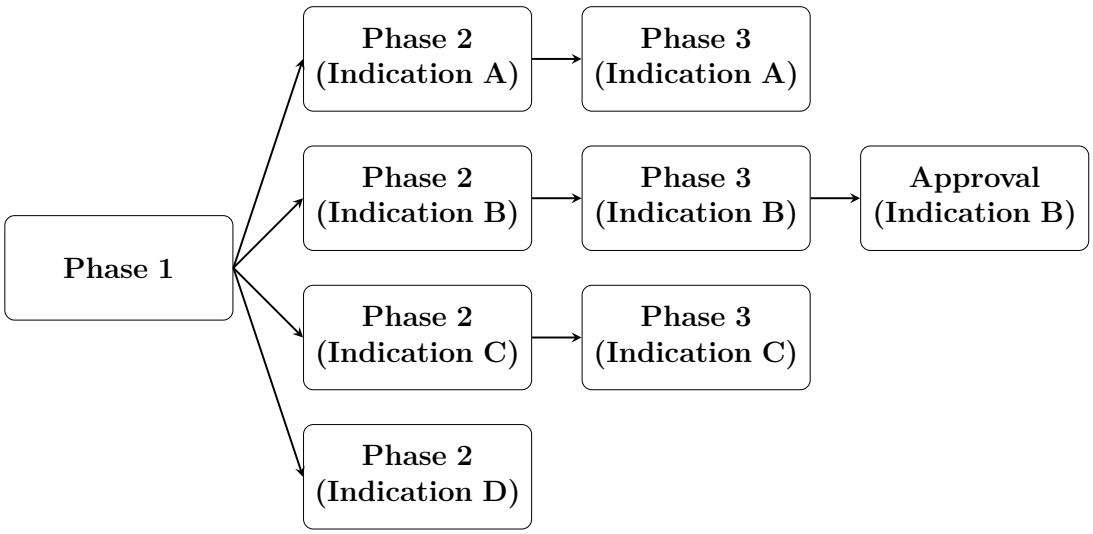


Figure 2-1: We define a drug development path as the development of a drug for a specific indication. A single clinical trial can belong to multiple drug development programs. We illustrate a hypothetical example where 4 drug development paths, all using the same drug, share the same phase 1 clinical trial.

2.3 A Model of Drug Development

A *phase transition* between phase i and j is the change between the states of a drug development program. We make the assumption that each program must transition from phase 1 to phase 2 to phase 3 to approval in this order, and model the possible states in a drug development program as a Markov chain shown in Figure 2-2. Every drug development path in our study must start from phase 1 (or 'missing' phase 1) and end up in one of the nodes labeled as 'in progress', 'terminated' or 'approval'.

We infer missing transitions in the development paths arising from incomplete records. This is plausible since each of these stages involves distinct predefined tests, all of which are required by regulators in any new drug application (NDA). If we observe data for phases 1

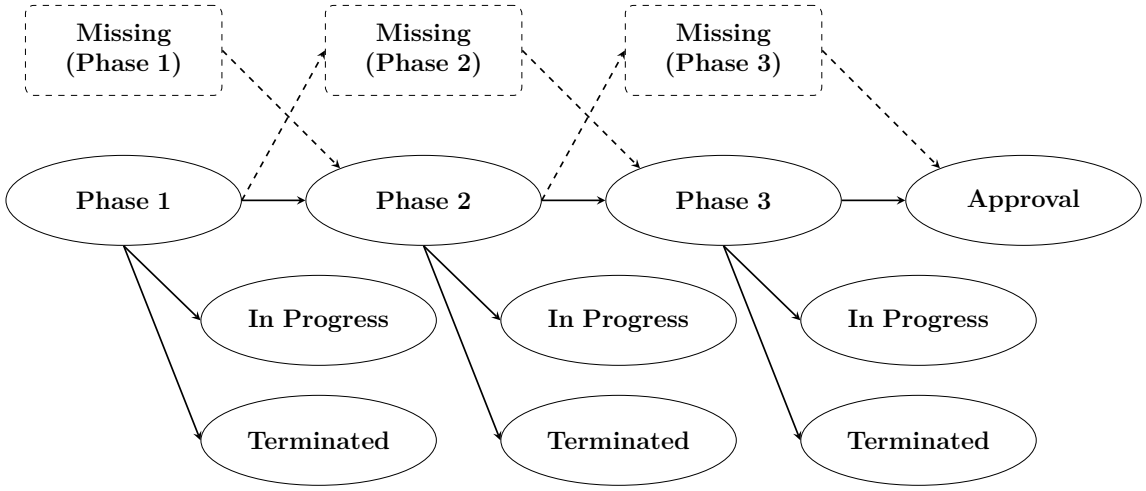


Figure 2-2: Observed and unobserved states in a drug development, from phase 1 to approval. A drug development is in phase i if it has at least one trial in phase i. The 'missing' states represent phases where we do not observe any clinical trial in that phase for the drug-indication pair, but where we know must have occurred. Every drug development path in our study must start from phase 1 (or 'missing' phase 1) and end up in one of the nodes labeled as 'in progress', 'terminated' or 'approval'.

and 3 but not phase 2 trials for a given drug-indication pair, our idealized process implies that there was at least one phase 2 trial that occurred, but is missing from our dataset. Accordingly, we impute the successful completion of phase 2 in these cases. There exist some rare cases where phase 2 trials are skipped, as with the example of aducanumab (BIIB037), Biogen's Alzheimer's candidate, as reported by Root [6]. Since skipping phase 2 trials is motivated by compelling phase 1 data and is approved by the regulatory authorities, imputing the successful completion of phase 2 trials in these cases to trace drug development paths is a reasonable approximation. We make the standard assumption that phase 1/2 and phase 2/3 trials are to be considered as phase 2 and phase 3, respectively.

2.4 Computing the Probability of Success

We call the estimated probability of a drug development program transitioning from phase i to phase $i+1$ the "phase i PoS", and the "estimated overall PoS" is defined as the estimated probability of a drug development program going from phase 1 to regulatory approval in at least one country.

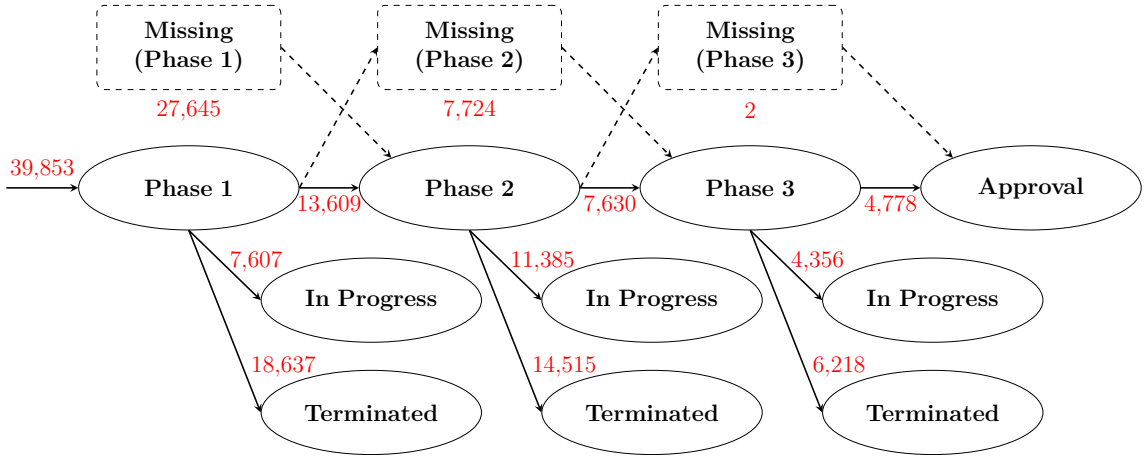


Figure 2-3: An example of the number of transitions computed.

The probability of a drug development program transitioning from phase i to phase j (PoS_{ij}) can be computed using the simple ratio N_j/N_i, where N_i is the number of drug development programs that have reached phase i (where $i = 1, 2, 3$) of the drug development process and are not in active development between phase i and phase j (where $j = 2, 3$, or "A" which denotes regulatory approval, $i < j$), and Nj is the number of drug development programs among the former that made it to phase j. PoS_{1A} is also known as the "overall PoS".

For clarity, we will walk our readers through some calculations using the example shown in Figure 2-3. In that figure, we see that 13,609 development programs have conducted phase 1 testing whereas 27,645 vaccine programs have skipped phase 1 to go to phase 2 or 3 testing directly. This is not uncommon in many drug development programs, where drug candidates move directly to the higher phases based on safety studies conducted for other indications. Among these 67,498 drug development programs, we know that 7,607 have yet to conclude phase 1 testing while 59,891 have completed phase 1. Of these 59,891 programs, 41,254 have gone on to phase 2 while 18,637 have failed. In the notation introduced earlier, N_1= 59,891 and N_2=41,254, yielding an estimate of $\frac{41,254}{59,891}$, or 68.9%, for PoS_{12}. Repeating the logic for the transitions between phase 2 and phase 3, and between phase 3 and approval, gives 51.4% and 43.5% as estimates of PoS_{23} and PoS_{3A} respectively.

In order to compute the probability of a drug development program making it all the way from phase 1 to approval, we consider only the drug development programs that have

definite outcomes. In other words, we do not consider development programs that are 'in progress' in the denominator. In our example, the number of such programs is 27,645 + 39,853 - 7,607 - 11,385 - 4,356 = 44,150. Since 4,778 programs made it to approval, the estimated PoS_{1A} is $\frac{4,778}{44,150} = 10.8\%$.

2.5 Path-by-path vs Phase-by-phase

The estimated probability of a drug development program transitioning from phase 1 to approval – estimated directly using the method described above – is called the 'path-by-path' estimate of the overall PoS, and is reported for all PoS calculations in this thesis.

It should be emphasized that because of the treatment of in-progress drug development programs, path-by-path PoS estimates are not multiplicative:

$$\text{PoS}_{1\text{A}} \text{ (path-by-path)} \neq \text{PoS}_{12} \times \text{PoS}_{23} \times \text{PoS}_{3\text{A}} \tag{2.1}$$

In contrast, the 'phase-by-phase' estimates used in prior studies (DiMasi et al., 2010, 2020; Hay et al., 2014; Thomas et al., 2016) do multiply:

$$\text{PoS}_{ij} = \prod_{x=i,\ldots,j-1} \text{PoS}_{x,x+1} \tag{2.2}$$

The phase-by-phase approach is valid under some circumstances, such as when one does not have any development programs that are active in any of the phases in his database. This is easily seen if one simply set the number of "in progress" development programs in all phases in Figure 6 to zero and recomputing the PoS.

The path-by-path approach can also obtain the same results as the phase-by-phase approach if one is willing to make an additional assumption: programs that are "in progress" in phase i will transit to phase $i+1$ or to "terminated" with the same probability as going from phase i to phase $i+1$, or from phase i to "terminated", without "in progress" programs.

We will illustrate this with Figure 2-4, which shows the different states of a drug development program with hypothetical transitions from "in progress" states to the next phase or to the "terminated" state. The proportions are derived from the numbers shown in Figure 2-3.

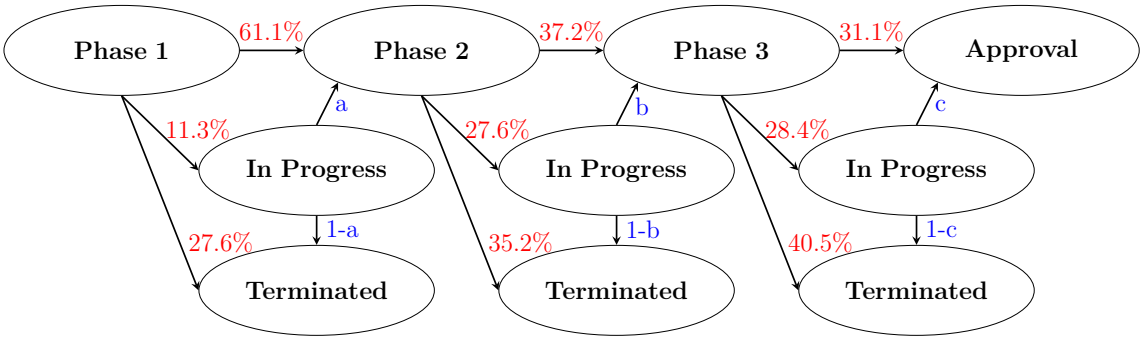


Figure 2-4: A Markov chain that includes hypothetical transitions from "in progress" states to the next phase or to the "terminated" state. The proportions are derived from the numbers shown in Figure 2-3.

Without considering "in progress" programs, the probability of transiting from phase 1 to phase 2 is 61.1/(61.1+27.6) = 68.9% and the probability of transiting from phase 1 into the "terminated" state is 10.8%. Similarly, without considering "in progress" programs between phase 2 and phase 3, or phase 3 to Approval, the probabilities of transiting from phase 2 to phase 3, or phase 3 to approval are 51.4% and 43.5% respectively. If we set a, b and c to be 68.9%, 51.4% and 43.5% respectively, we will obtain a PoS_{1A} of 15.4%, which is exactly $\text{PoS}_{12}\times$ $\text{PoS}_{23}\times$ PoS_{3A}. In contrast, the path-by-path approach obtains a PoS_{1A} of 10.8% as it does not make any assumptions and ignores programs that are "in progress" in either phase 1, phase 2 or phase 3.

We believe that our method of inferring unobserved clinical development stages and then applying the path-by-path approach is a better measure of the PoSs of clinical development programs as it does not overestimate the PoSs, and makes no assumption about the programs that are in active development.

That said, the path-by-path method of computation is only applicable when the variable being investigated is constant throughout the drug development process. This may not be true when we look at subsets of the data, such as when we investigate whether the use of biomarkers for patient selection impacts the success rate of clinical drug development. This is because biomarkers may not be used at all stages of a clinical drug development program. In such cases, we default back to the 'phase-by-phase' estimation to get an insight into the trend. This is done by considering only those drug development programs with phases that ended between t_1 and t_2 in the computation of the PoS.

TrialID	Drug Name	Phase	Start Date	End Date	Disease Type	Therapeutic Area	Sponsor
48391	Loratadine	1/2	NULL	7/6/2003	Allergic Rhinitis	Autoimmune/ Inflammation	(Other Hospital/ Academic/ Medical Center)
70538	Loratadine	3	NULL	18/9/2007	Allergic Rhinitis	Autoimmune/ Inflammation	(Other Hospital/ Academic/ Medical Center)
100378	Loratadine	3	NULL	29/10/2008	Asthma	Autoimmune/ Inflammation	Merck & Co.
122164	Loratadine	4	1/1/2010	1/3/2012	Allergic Rhinitis	Autoimmune/ Inflammation	(Other Hospital/ Academic/ Medical Center)
151465	Loratadine	3	1/5/2011	14/5/2014	Pain (nociceptive)	CNS	Cancer and Leukemia Group B (CALGB)
153368	Loratadine	1	NULL	1/7/2006	Asthma	Autoimmune/ Inflammation	(Other Hospital/ Academic/ Medical Center)

Table 2.1: A sample of Citeline data entries.

2.6 Estimating the PoSs from data

Given our Markov-chain model of drug development, we will be able to estimate the PoSs from clinical trial metadata.

2.6.1 Data

We use Citeline data provided by Informa Pharma Intelligence, which combines individual clinical trial information from TrialTrove and drug approval data from Pharmaprojects. Citeline is a superset of the most commonly data sources. In addition to incorporating multiple data streams – including nightly feeds from official sources such as ClinicalTrials.gov – Citeline contains data from primary sources such as institutional press releases, financial reports, study reports, and drug marketing label applications, and secondary sources such as analyst reports by consulting companies. Secondary sources are particularly important for reducing potential biases that can arise from the tendency of organizations to report only successful trials (especially prior to the FDA Amendments Act of 2007 requiring all clinical trials to be registered and tracked via ClinicalTrials.gov).

We reproduce a sample of the clinical trial metadata used in Table 2.1. Each entry contains the trial identification number, the drug being used, the phase of the clinical trial, the start and the end dates, the disease being targeted, the therapeutic area that the disease belongs to, and the sponsor(s).

It is possible that the same clinical trial, as classified by the trial's unique identifier, may appear multiple times in our analysis. For example, if a clinical trial targets two diseases, we consider it to be two data points in which each entry contains the name of only one disease. This reflects our assumption that the trial is part of different development programs.

Several snapshots of the data were used for the analyses in this thesis. We will describe

and summarize the data used in the respective sections before we present the results of the analyses.

2.6.2 An Algorithm for Computing PoSs

Given our Markov-chain model of drug development, we can compute PoSs using the algorithm presented in Algorithm 1. The algorithm recursively considers all possible drug-indication pairs and determines the maximum observed phase. Reaching phase i would imply that all lower phases were completed. To determine if a drug development program has been terminated in the last observed phase or is still ongoing, we use a simple heuristic: if the time elapsed between the end date of the most recent phase i and the end of our sample exceeds a certain threshold t_i, we conclude that the trial has terminated. Based on practical considerations, we set t_i, to be 360, 540 and 720 days for Phases 1, 2, and 3, respectively. For example, we assume that it can take up to 12 months to prepare documents for an NDA filing after a Phase 3 trial has been completed. Since the FDA has a 60 days period to decide if it wishes to follow up on a filing, and an additional 10 months to deliver a verdict, this places the overall time between Phase 3 to approval to a maximum of 24 months, hence we set $t_3 = 720$ days.

This heuristic allows us to impute missing trial data, and by counting the number of phase transitions, we can estimate the phase and overall PoSs.

Algorithm 1: An algorithm for identifying phase transitions in drug development and computing the path-by-path probability of success.

Input : data: A list of clinical trial metadata
T: date of database snapshot
Output: PoS_{12}, PoS_{23}, PoS_{3A}
Initialize $n_{12,s} = n_{12,f} = n_{23,s} = n_{23,f} = n_{3A,s} = n_{3A,f} = 0$
for *every (drug, indication) in data* **do**
  Filter and populate a list of trials conducted by organization on indication using drug;
  if *(drug is approved for indication)* **then**
    $n_{12,s}$++;
    $n_{23,s}$++;
    $n_{3A,s}$++;
  end
  else
    if ∃*(at least one phase 3 trial)* **then**
      $n_{12,s}$++;
      $n_{23,s}$++;
      if *(Most recent end date of phase 3 trials is* $< T - t_3$*)* **then**
        $n_{3A,f}$++;
      end
    end
    else
      if ∃*(at least one phase 2 trial)* **then**
        $n_{12,s}$++;
        if *(Most recent end date of phase 2 trials is* $< T - t_2$*)* **then**
          $n_{23,f}$++;
        end
      end
      else
        if ∃*(at least one phase 1 trial) &*
        (Most recent end date of phase 1 trials is $< T - t_1$*)* **then**
          $n_{12,f}$++;
        end
      end
    end
  end
end
$\text{PoS}_{12} = n_{12,s}/(n_{12,s} + n_{12,f})$
$\text{PoS}_{23} = n_{23,s}/(n_{23,s} + n_{23,f})$
$\text{PoS}_{3A} = n_{3A,s}/(n_{3A,s} + n_{3A,f})$
return PoS_{12}, PoS_{23}, PoS_{3A}

Chapter References

[1] Rosa M Abrantes-Metz, Christopher Adams, and Albert D Metz. Pharmaceutical development phases: a duration analysis. *Journal of Pharmaceutical Finance, Economics and Policy*, 14:19–42, 2005.

[2] Joseph A DiMasi, Lanna Feldman, Abraham Seckler, and Andrew Wilson. Trends in risks associated with new drug development: success rates for investigational drugs. *Clinical Pharmacology & Therapeutics*, 87(3):272–277, 2010.

[3] Joseph A DiMasi, Henry G Grabowski, and Ronald W Hansen. Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of Health Economics*, 47:20–33, 2016.

[4] Michael Hay, David W Thomas, John L Craighead, Celia Economides, and Jesse Rosenthal. Clinical development success rates for investigational drugs. *Nature Biotechnology*, 32(1):40–51, 2014.

[5] Deanna Pogorelc. The biotech valley of death has become the uncrossable canyon. here's one innovative approach to funding. *MedCity News*, Sept 2012. URL `https://medcitynews.com/2012/09/the-biotech-valley-of-death-has-become-the-uncrossable-canyon-heres-one-innovative-approach-to-funding/`. Accessed: 2020-10-13.

[6] Cyndi Root. Biogen Idec moves aggressively, advances alzheimer drug into phase 3, Dec 2014. URL `https://www.clinicalleader.com/doc/biogen-idec-moves-aggressively-advances-alzheimer-drug-into-phase-0001`.

[7] Katarzyna Smietana, Marcin Siatkowski, and Martin Møller. Trends in clinical success rates. *Nature Reviews Drug Discovery*, 15:379–380, 2016.

[8] Richard T Thakor, Nicholas Anaya, Yuwei Zhang, Christian Vilanilam, Kien Wei Siah, Chi Heem Wong, and Andrew W Lo. Just how good an investment is the biopharmaceutical sector? *Nature Biotechnology*, 35(12):1149–1157, 2017.

[9] David W Thomas, Justin Burns, John Audette, Adam Carroll, Corey Dow-Hygelund, and Michael Hay. Clinical development success rates 2006–2015. *BIO Industry Analysis*, 1:16, 2016.

Chapter 3

Data Analytics For The Understanding Of The Drug Development Process

3.1 Introduction

In the previous chapter, we explained our model of drug development, introduced a model which can be used to infer phase transitions from clinical trial metadata and described the computation of the probability of success (PoS) of drug development programs.

In this chapter, we will apply the methods, and others, onto data to derive meaningful insights into the challenges confronting drug development. Section 3.2 investigates drug development statistics across all the indications and compares them with prior studies. We will also present the PoSs conditioned on the use of biomarkers and the PoSs of orphan development programs. Section 3.3 presents the statistics for the development of oncological drugs, an area that has always held the attention of the public and policymakers due to the high cost of treatment drugs and massive spending by Medicare and Medicaid. Section 3.3 gives insights into the business of vaccine development, which has become a topic of great interest due to the COVID-19 pandemic.

3.2 General Drug Development Statistics

We will first summarize the general characteristics of drug development programs. This section takes reference from Wong et al. [5], which was performed between 2016 and 2017 using data between 2000 and 2015. After the publication of the paper, we continued to produce quarterly updates on the MIT Laboratory for Financial Engineering's (LFE) Project ALPHA (Analytics for Life-sciences Professionals and Healthcare Advocates) website, which can be accessed via www.projectalpha.mit.edu/pos. In this thesis, we update the results of the prior paper with data obtained in 2020 Q3.

3.2.1 Summary of Data

The clinical trials used in this analysis have start dates between January 1, 2000, and October 05, 2020, the latter being the date that we receive the snapshot of the data.

In Figure 3-1, we plot the cumulative number of clinical trials conducted over time, with further breakdown by sponsor type. Our dataset consists of 210,755 clinical trials and 142,646 unique drug-indication pairs. We count 662,822 unique data points, defined as unique {drug, indication, clinical trial identifier} triplets. Of these, 281,835 (42.5%) are industry-sponsored – defined as clinical trials involving at least one industry sponsor – while 380,987 (57.5%) do not involve any industry sponsorship.

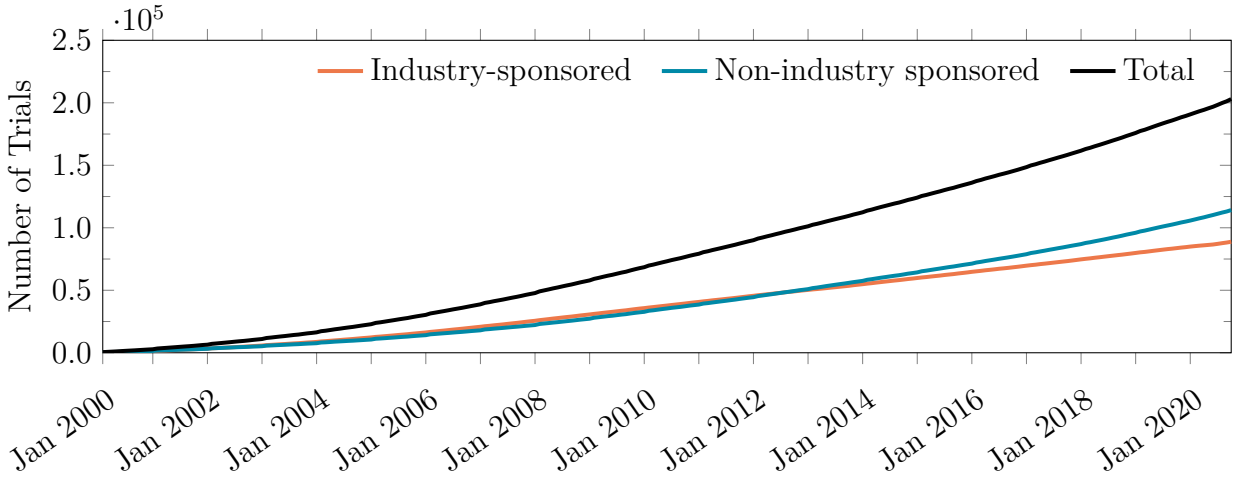


Figure 3-1: The cumulative number of clinical trials being conducted over time, by sponsor type.

3.2.2 PoSs by Therapeutic Area

Table 3.1 contains our estimates of the aggregate PoS for each clinical phase across all indications. Corresponding estimates from the prior literature are also included for comparison. We find that 10.8% of all drug development programs eventually lead to approval, which is slightly higher than the 10.4% reported by Hay et al. [2] and the 9.6% reported by Thomas et al. [4]. The overall PoS presented in this study, Hay et al. [2] and Thomas et al. [4] are much higher than the 1 to 3 percent that is colloquially seen as it is conditioned on the drug development entering Phase 1. Our phase-specific PoS estimates are higher in all phases. The largest increase is seen in $PoS_{2,3}$, where we obtained a value of 51.4% compared to 32.4% in Hay et al. [2] and 30.7% in Thomas et al. [4]. These differences may be due to our method of imputing missing clinical trials.

	This study - All indications (Industry)			**Thomas *and others* (2016) - All Indications**		**Hay *and others* (2014) - All Indications**		**Hay *and others* (2014) - Lead Indications**		**DiMasi *and others* (2010) - Lead Indications**	
Method	$PoS_{i,i+1}$	Path-by-Path $PoS_{i,A}$	Phase-by-Phase $PoS_{i,A}$	Phase-by-Phase $PoS_{i,i+1}$	$PoS_{i,A}$	Phase-by-Phase $PoS_{i,i+1}$	$PoS_{i,A}$	Phase-by-Phase $PoS_{i,i+1}$	$PoS_{i,A}$	Phase-by-Phase $PoS_{i,i+1}$	$PoS_{i,A}$
Phase 1 to 2	68.9%	10.8%	15.4%	63.2%	9.6%	64.5%	10.4%	66.5%	15.3%	71.0%	19.0%
Phase 2 to 3	51.4%	16.0%	22.4%	30.7%	15.2%	32.4%	16.2%	39.5%	23.1%	45.0%	26.8%
Phase 3 to Approval	43.5%	43.5%	43.5%	49.6%	49.6%	50.0%	50.0%	58.4%	58.4%	60.0%	59.5%
Phase 1 to Approval		10.8%	15.4%		9.6%		10.4%		15.3%		19.0%
Number of Drugs		22,760		?		5,820		4,736		1,316	
Years of source data (time-span)		2000 - 2020 (20 years)		2006-2015		2003-2011 (9 years)				1993-2009 (17 years)	
Number of Companies		6,271		1,103		835				50	

Table 3.1: Comparison of the results of our paper with previous publications using data from January 1, 2000, to October 8, 2020.

Table 3.2 contains phase and overall PoS estimates by therapeutic group. The overall PoS ($PoS_{1,A}$) ranges from a minimum of 4.0% for oncology to a maximum of 40.5% for vaccines (infectious disease). The overall PoS for oncology drug development programs is about two-thirds the previously reported estimates of 5.1% in Thomas et al. [4] and 6.7% in Hay et al. [2].

Table 3.2: The probability of success by therapeutic group, using data from January 1, 2000, to October 8, 2020. We computed this using the path-by-path method. SE denotes the standard error.

	Phase 1			Phase 2					Phase 3			Overall	
Therapeutic Area	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	PoS_{1A}	S.E.
Oncology	29746	65.8	0.3	11390	37.1	0.5	7.0	0.3	2776	28.9	0.9	4.0	0.1
Metabolic/Endocrinology	4697	72.7	0.6	2884	60.1	0.9	21.1	0.8	1350	45.0	1.4	16.1	0.6
Cardiovascular	3640	72.3	0.7	2364	70.4	0.9	25.8	1.0	1285	47.5	1.4	20.4	0.7
CNS	6593	70.6	0.6	4003	56.0	0.8	14.5	0.6	1607	36.0	1.2	10.9	0.4
Autoimmune/Inflammation	6755	70.3	0.6	3924	52.7	0.8	17.0	0.7	1433	46.6	1.3	12.6	0.5
Genitourinary	1160	69.7	1.3	760	61.6	1.8	25.9	1.7	380	51.8	2.6	19.3	1.2
Infectious Disease	4552	67.6	0.7	2515	66.7	0.9	21.3	0.9	1102	48.6	1.5	15.7	0.6
Ophthalmology	746	88.3	1.2	549	54.6	2.1	16.8	1.8	194	47.4	3.6	17.4	1.6
Vaccines (Infectious Disease)	2002	83.6	0.8	1480	66.3	1.2	46.5	1.3	871	79.0	1.4	40.5	1.2
Overall	59891	68.9	0.2	29869	51.4	0.3	16.0	0.2	10998	43.5	0.5	10.8	0.1
All except oncology	30145	71.9	0.3	18479	60.2	0.4	21.5	0.3	8222	48.4	0.6	16.5	0.2

3.2.3 PoSs Conditioned on the Use of Biomarkers

As the use of biomarkers to select patients, enhance safety, and serve as surrogate clinical endpoints has become more common, it has been hypothesized that trials using biomarkers are more likely to succeed. We test this hypothesis by comparing the PoS of drugs with and without the use of biomarkers for patient selection, an approach that is similar to Thomas et al. [4]

In our database, only a small proportion ($< 10\%$) of all drug development programs that use biomarkers use them in all stages of development. As such, we adopt the phase-by-phase approach instead of using the path-by-path approach.

Table 3.3 shows only trials that use biomarkers to stratify patients. As can be seen, there is substantial variation in the use of biomarkers across therapeutic areas. Biomarkers are seldom used outside of oncology. Trials using biomarkers exhibit almost twice the overall PoS ($PoS_{1,A}$) compared to trials without biomarkers (13.0% vs. 4.3%). While the use of biomarkers in the stratification of patients improves the POS in all phases, it is most significant in Phases 1 and 2. (We caution against over-interpreting the results for therapeutic areas outside oncology due to their small sample size.) These findings are similar in spirit to the analysis by Thomas et al. [4], which also found substantial improvement in the PoS when biomarkers were used.

Table 3.3: Probability of success of drug development programs with and without biomarkers, using data from January 1, 2005, to October 8, 2020, computed using the phase-by-phase method. These results consider only trials that use biomarkers in patient stratification.

		Phase 1			Phase 2			Phase 3			Overall	
Therapeutic Area		Paths	PoS	S.E.	Paths	PoS	S.E.	Paths	PoS	S.E.	PoS	S.E.
Oncology	With Biomarker	4542	52.2	0.7	2821	41.7	0.9	6785	18.6	0.5	12.6	0.9
	Without Biomarker	11638	31.4	0.4	666	57.8	1.9	2108	19.7	0.9	1.1	0.1
Metabolic/Endocrinology	With Biomarker	27	59.3	9.5	19	84.2	8.4	1778	35.4	1.1	14.3	8.4
	Without Biomarker	2151	41.0	1.1	42	28.6	7.0	1308	45.6	1.4	6.6	0.6
Cardiovascular	With Biomarker	15	73.3	11.4	30	63.3	8.8	1202	42.7	1.4	29.6	14.5
	Without Biomarker	1673	40.0	1.2	33	63.6	8.4	1252	47.0	1.4	8.0	0.8
CNS	With Biomarker	73	54.8	5.8	77	41.6	5.6	2510	31.6	0.9	10.2	4.7
	Without Biomarker	3025	37.0	0.9	47	44.7	7.3	1560	35.8	1.2	4.2	0.4
Autoimmune/Inflammation	With Biomarker	17	64.7	11.6	52	34.6	6.6	2615	30.3	0.9	15.7	10.9
	Without Biomarker	3203	37.5	0.9	10	70.0	14.5	1423	46.5	1.3	5.3	0.4
Genitourinary	With Biomarker	16	43.8	12.4	0	–	–	425	31.3	2.2	–	–
	Without Biomarker	488	29.9	2.1	7	42.9	18.7	373	52.0	2.6	4.9	1.0
Infectious Disease	With Biomarker	13	61.5	13.5	87	74.7	4.7	1355	39.9	1.3	36.4	13.5
	Without Biomarker	2239	34.3	1.0	77	79.2	4.6	1025	46.3	1.6	6.3	0.6
Ophthalmology	With Biomarker	1	100.0	0.0	6	33.3	19.2	372	34.1	2.5	33.3	19.2
	Without Biomarker	203	57.1	3.5	3	100.0	0.0	191	46.6	3.6	9.1	2.0
Vaccines (Infectious Disease)	With Biomarker	8	37.5	17.1	24	20.8	8.3	836	42.6	1.7	–	–
	Without Biomarker	638	49.2	2.0	0	–	–	871	79.0	1.4	16.6	1.7
Overall	With Biomarker	4712	52.3	0.7	3116	42.8	0.9	885	58	1.7	13	0.8
	Without Biomarker	25258	35.1	0.3	17878	28.8	0.3	10111	42.2	0.5	4.3	0.1

3.2.4 PoSs of Orphan Development Programs

Table 3.4 contains PoS estimates for drugs that treat rare diseases, also known as 'orphan drugs'. The classifications for rare diseases are obtained from both the EU and US rare disease resources: OrphaNet and NIH GARD. Rare diseases may belong to any therapeutic group, and the computation of the statistics for orphan drugs is identical to that used for the trials in Table 3.2.

Broadly speaking, orphan drug development has slightly higher success rates, with 12.0% of drug development projects reaching the market. Comparing these results against those for all drug development, we see that while the Phase 3 success rate falls from 43.5% to 42.5%, the Phase 1 and Phase 2 PoS increase from 68.9% to 83.0% and from 51.4% to 52.1%, respectively, leading to an increase in the overall POS.

Our data reveals that most orphan drug trials are in oncology. Our overall PoS of 12.0% is much lower than the 25.3% reported in Thomas et al. [4]. This discrepancy can be attributed to their identification of only non-oncology indications as 'rare diseases', and their use of the phase-by-phase method of computing the POS. Our estimated orphan drug PoS increases to

19.4% after excluding all oncology indications from the calculations, which is more in line with the findings of Thomas et al. [4].

Table 3.4: The probability of success of orphan drug development programs. We computed the results using the path-by-path method. We used the entire dataset from January 1, 2000, to October 08, 2020.

	Phase 1			Phase 2					Phase 3			Overall	
Therapeutic Area	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	PoS_{1A}	S.E.
Oncology	4258	79.7	0.6	1886	45.3	1.1	10.2	0.8	512	37.7	2.1	8.0	0.6
Metabolic/Endocrinology	401	88.8	1.6	228	67.5	3.1	17.5	2.9	93	43.0	5.1	18.9	2.7
Cardiovascular	279	90.3	1.8	215	69.3	3.1	31.6	3.5	112	60.7	4.6	33.2	3.3
CNS	415	89.6	1.5	305	59.3	2.8	12.1	2.2	87	42.5	5.3	14.6	2.2
Autoimmune/Inflammation	475	89.1	1.4	346	61.6	2.6	8.1	1.7	118	23.7	3.9	9.2	1.7
Genitourinary	27	88.9	6.0	24	62.5	9.9	33.3	10.1	13	61.5	13.5	32.0	9.3
Infectious Disease	292	93.8	1.4	225	55.6	3.3	21.3	3.0	84	57.1	5.4	23.8	3.0
Ophthalmology	40	80.0	6.3	16	68.8	11.6	6.2	7.3	6	16.7	15.2	5.3	5.1
Vaccines (Infectious Disease)	82	92.7	2.9	71	32.4	5.6	31.0	5.5	22	100.0	0.0	28.9	5.2
Overall	6269	83.0	0.5	3316	52.1	0.9	13.4	0.7	1047	42.5	1.5	12.0	0.5
All except oncology	2011	90.0	0.7	1430	60.9	1.3	17.6	1.2	535	47.1	2.2	19.4	1.1

3.2.5 Discussion

Compared to Hay et al. [2] and Thomas et al. [4], we obtain higher PoS estimates for all phases. Our numbers will result in a lower estimated drug development cost, especially in Phase 3, where the cost of conducting a trial significantly exceeds that of other phases.

In Table 3.4, we show that orphan development programs have a slightly higher overall PoS compared to the PoS of all drug development programs. This is a surprising finding, given that we found lower PoS for orphan development programs in our previous study [5]. The slightly higher success rates of orphan drug development may be due to several factors, including the maturation of new technologies such as gene replacement therapy.

We see heterogeneity in the PoSs across the different therapeutic areas, with oncology development programs having the lowest success rates (4.0%, SE=0.1%) and vaccines having the highest (40.5%, SE=1.2%). The variance in the PoSs will affect the optimal composition of investors' portfolio.

All else being equal, we would expect that drug development programs with lower PoSs will face greater difficulty in raising sufficient capital to propel them into market. However, reality cannot be more different; over the past decade, oncology drug development has at-

tracted the lion's share of all venture capitalist (VC) investments [3], while the number of companies producing vaccines has dwindled over the years [1]. We will explore these areas in detail over the next two sections in this chapter.

Section References

[1] Evaluatepharma world preview 2018, outlook to 2024, 2018.

[2] Michael Hay, David W Thomas, John L Craighead, Celia Economides, and Jesse Rosenthal. Clinical development success rates for investigational drugs. *Nature Biotechnology*, 32(1):40–51, 2014.

[3] David Thomas and Chad Wessel. 2019 emerging therapeutic company trend report. *Biotechnology Innovation Organization (BIO)*, 2019.

[4] David W Thomas, Justin Burns, John Audette, Adam Carroll, Corey Dow-Hygelund, and Michael Hay. Clinical development success rates 2006–2015. *BIO Industry Analysis*, 1:16, 2016.

[5] Chi Heem Wong, Kien Wei Siah, and Andrew W Lo. Estimation of clinical trial success rates and related parameters. *Biostatistics*, 20(2):273–286, 2019.

3.3 Oncology Development Programs Statistics

Despite the billions of dollars spent annually on oncology research, effective treatments for numerous types of cancer remain as elusive as ever. In the previous section, we find that even though 49.7% of all drug development programs are being conducted in the field of oncology, only 4.0% of all cancer drug development programs make it from phase 1 to regulatory approval. This average masks the huge variation in the probability of success (PoS) for treatment of the different diseases within oncology. Given the heterogeneous nature of cancer [1], one would expect that some cancers are better studied than others, and that more drugs and therapies would have been attempted for these better-known diseases. Nevertheless, to the best of our knowledge, a comprehensive and fine-grained study of the PoS for drug development programs in oncology has not been conducted.

In this section, we focus specifically on the performance of oncology drug development. We shed light on the landscape and productivity of oncology clinical drug development programs by calculating several different metrics, including the number of development programs, the probabilities of success, and the duration of trials for the various disease groups within oncology. In addition, we investigate the change in these statistics considering only the programs related to rare and orphan diseases, and similarly, considering only the programs using biomarkers in some phase of drug development.

3.3.1 Data Summary

In this analysis, we used an snapshot of the database where the trials have start dates between January 1, 2000, and September 30, 2018. We filter our database for only industry-sponsored trials relevant to the development of cancer drugs. Our dataset contains 108,248 unique {trial identification number, drug, disease} triplets after removing data points with an unknown phase or disease.

Our path-by-path computation identifies 24,448 drug development programs across 40 different disease groups. From Figure 3-2, we can see that the three diseases with the most drug development programs are non-small cell lung cancer (1,501), breast cancer (1,373) and colorectal cancer (1,351), while the three diseases with the fewest number of drug develop-

ment programs are unspecified hematological cancer (141), testicular cancer (123) and basal cell carcinoma (123). The median number of drug development programs across all diseases is 538.5.

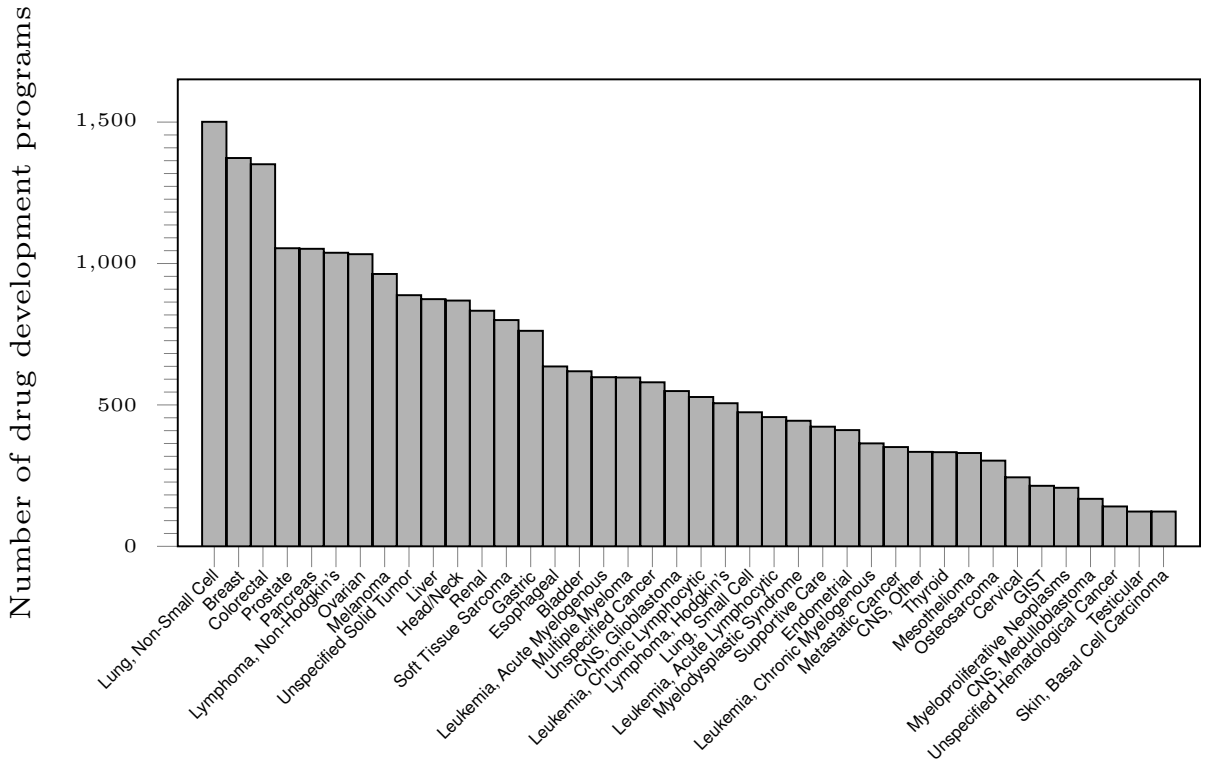


Figure 3-2: Number of drug development programs for disease groups in oncology.

3.3.2 Duration of Oncology-related Clinical Trials

We compute the duration of clinical trials and produce boxplots by disease group and trial phase (see Figures 3-3, 3-4, 3-5, 3-6, and 3-7). The median trial durations are 0.75–4.34 years for phase 1, 2.12–4.38 years for phase 2, 2.12-4.95 years for phase 3, 2.17–4.73 years for phase 1/2 and 1.82–7.81 years for phase 2/3. We do not detect any pattern for the duration of trials across disease groups. Within each disease group, we can see a large range of trial durations, with some trials taking as much as 16 times the median duration of other trials within the same disease group.

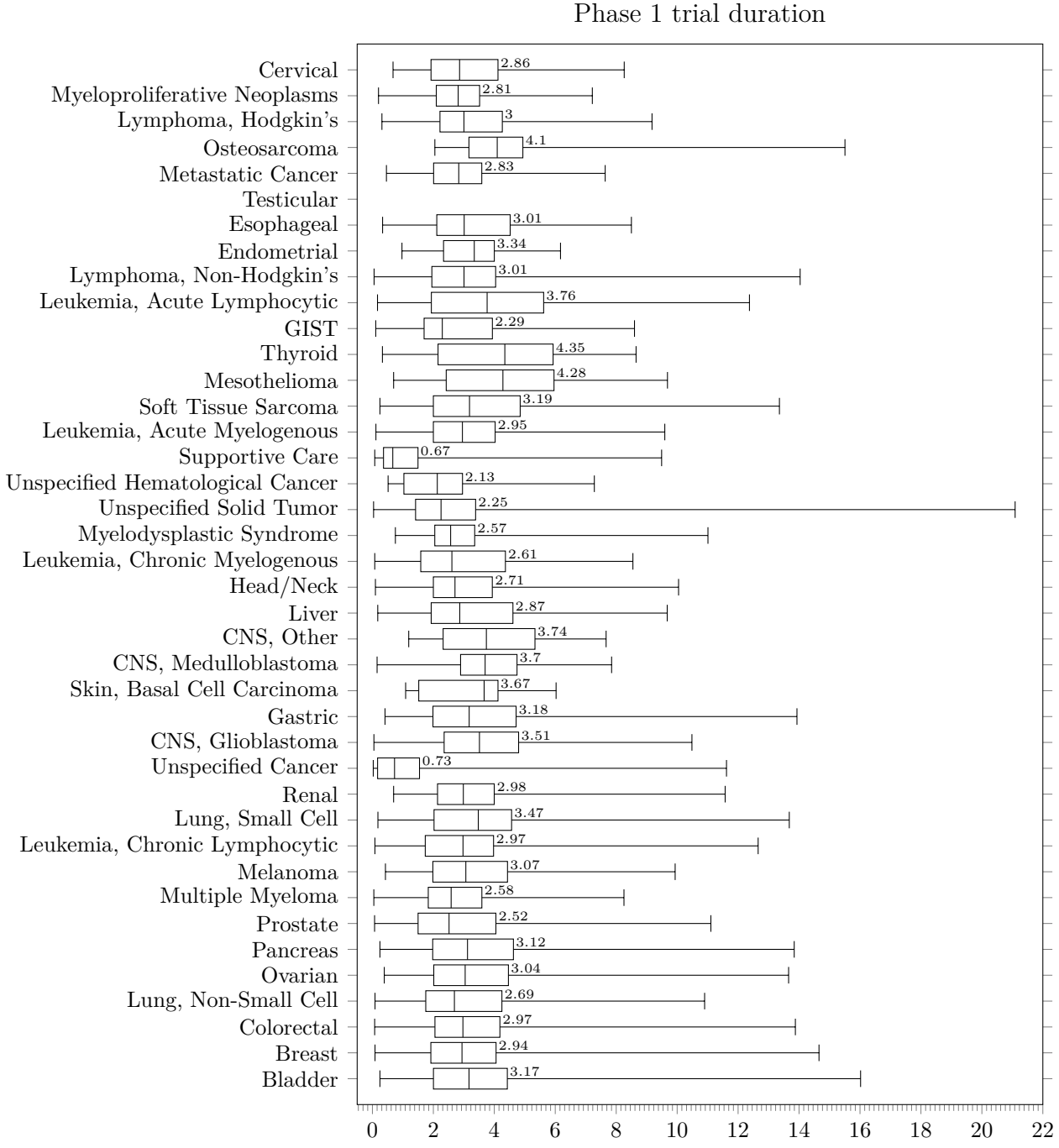


Figure 3-3: Boxplot diagram of the duration of phase 1 clinical trials, in years, by disease. The median value is displayed at the end of the box. We only compute the boxplot when there are more than 5 trials.

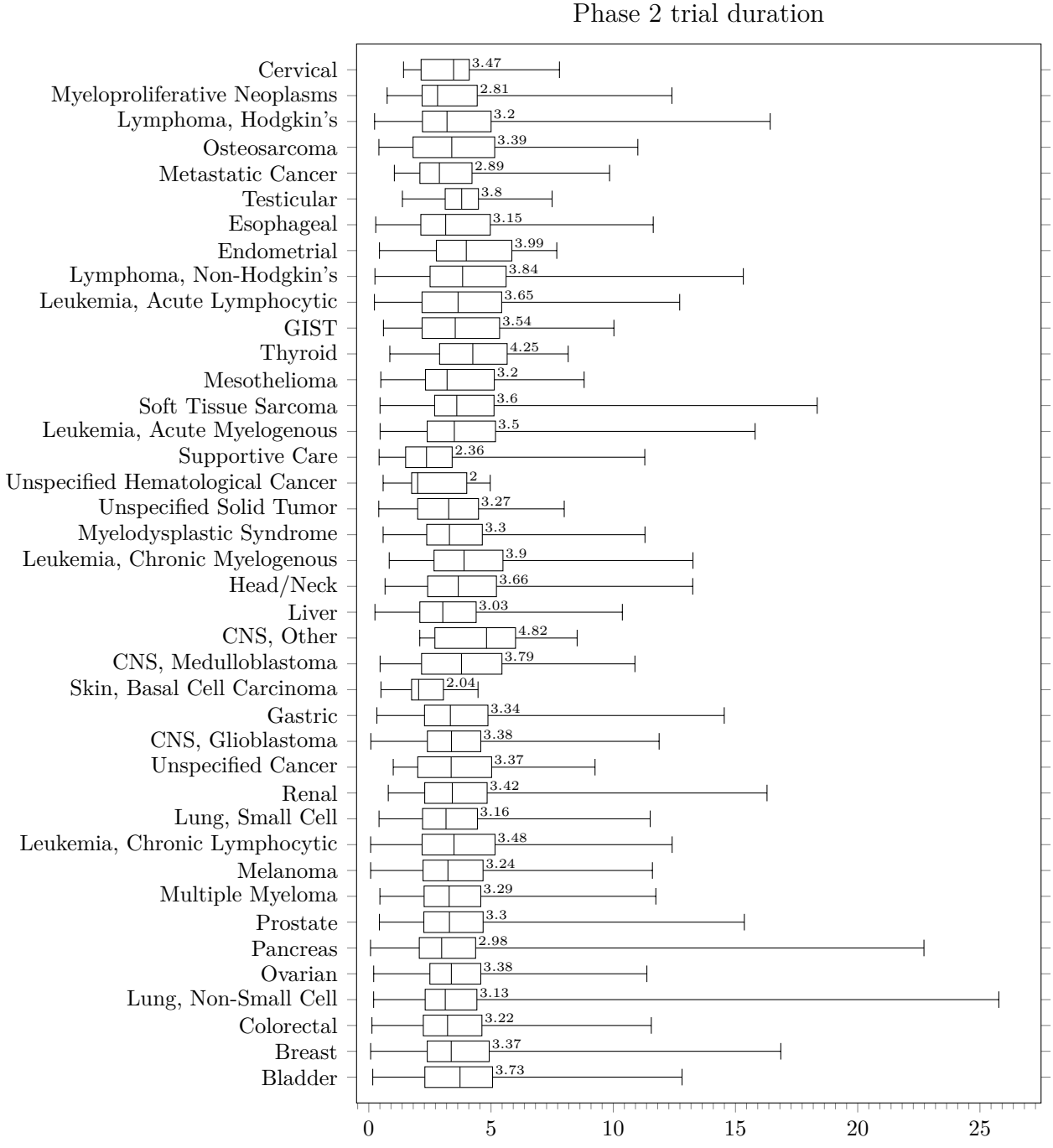


Figure 3-4: Boxplot diagram of the duration of phase 2 clinical trials, in years, by disease. The median value is displayed at the end of the box. We only compute the boxplot when there are more than 5 trials.

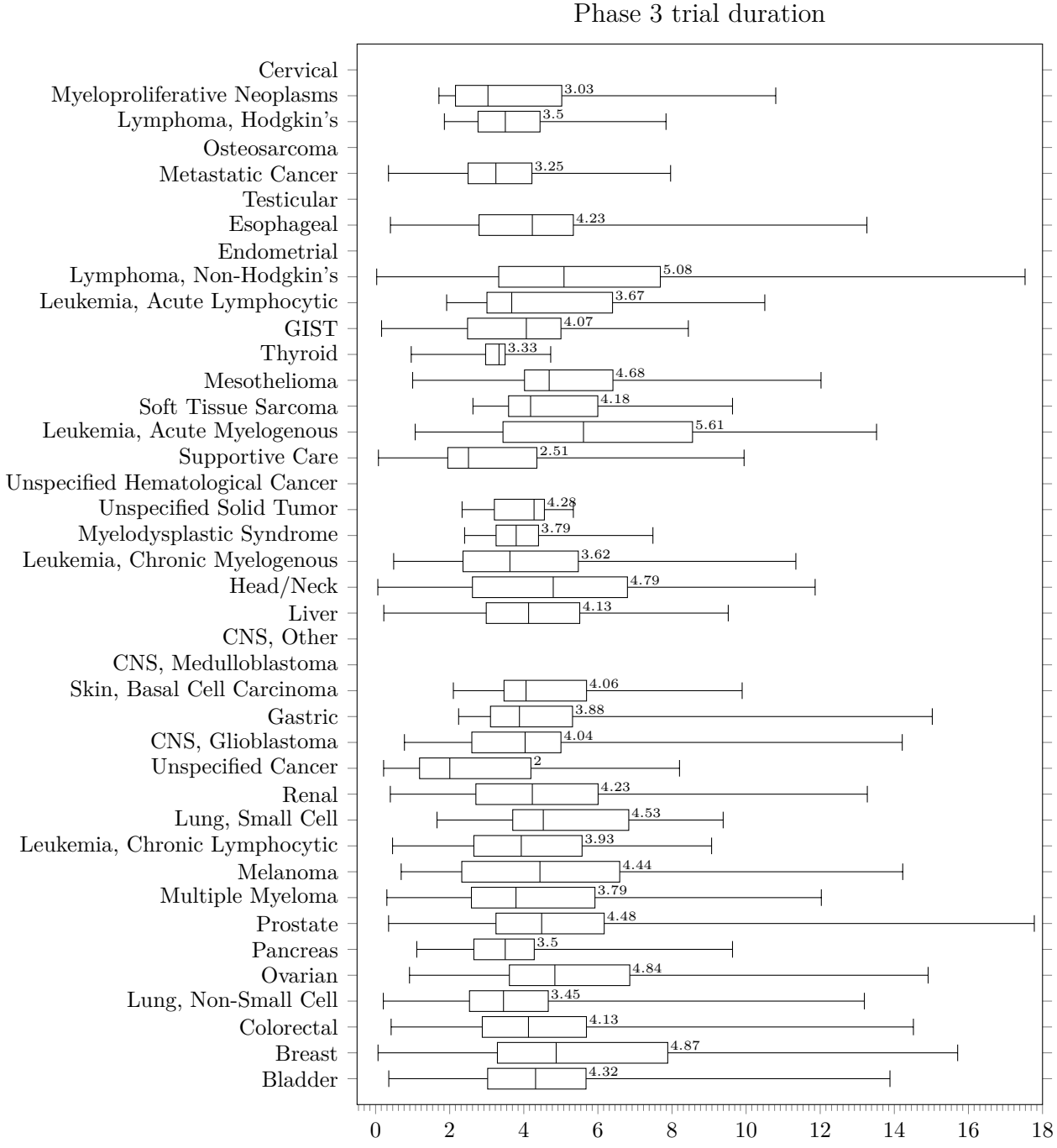


Figure 3-5: Boxplot diagram of the duration of phase 3 clinical trials, in years, by disease. The median value is displayed at the end of the box. We only compute the boxplot when there are more than 5 trials.

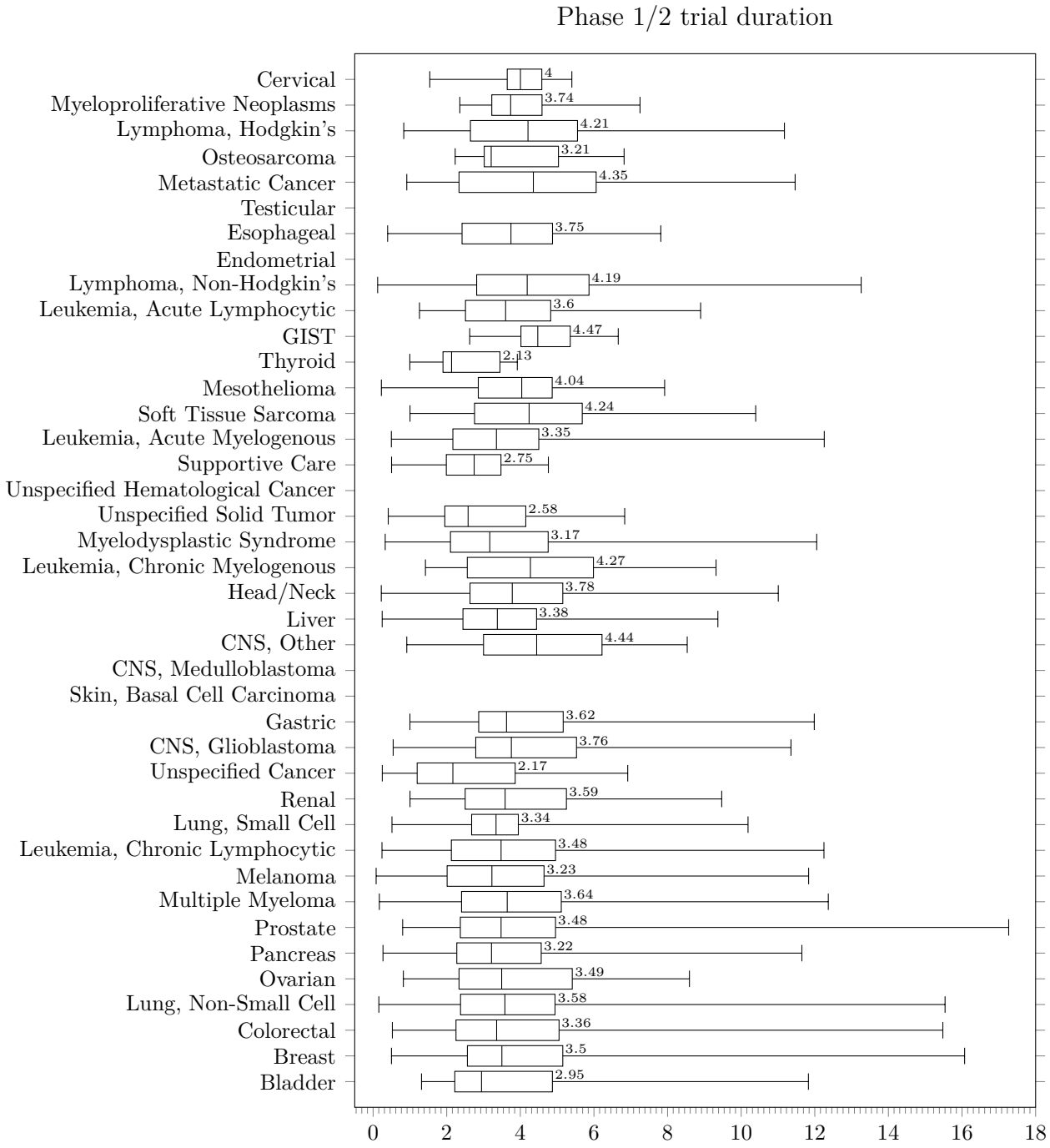


Figure 3-6: Boxplot diagram of the duration of phase 1/2 clinical trials, in years, by disease. The median value is displayed at the end of the box. We only compute the boxplot when there are more than 5 trials.

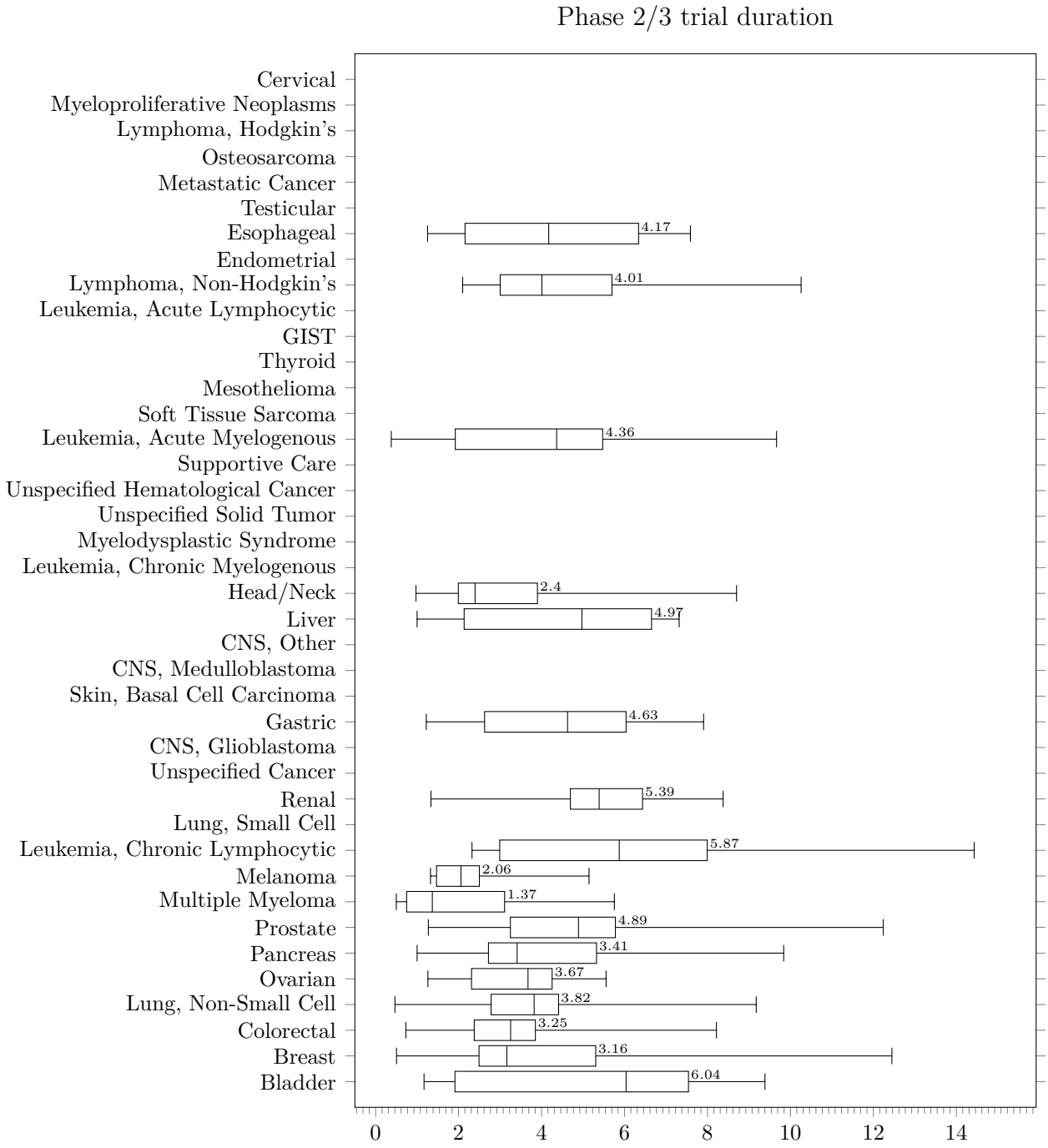


Figure 3-7: Boxplot diagram of the duration of phase 2/3 clinical trials, in years, by disease. The median value is displayed at the end of the box. We only compute the boxplot when there are more than 5 trials.

3.3.3 PoS across Oncology Indications

The PoS for all diseases is displayed in Table 3.5. We can see that the PoS vary quite differently between phases and disease types. After eliminating drugs for use in supportive care, we see that breast cancer has the highest overall PoS, with 9.5% of all drug development programs making it from phase 1 to marketing approval. On the other hand, there are disease types, such as osteosarcoma and uncategorized central nervous system-related cancer, for which no drug development path has made it to approval. We do not detect any linear relation between the number of development paths and the overall PoS (see Figure 3-8).

Disease	Paths	PoS12	PoS12err	Paths	PoS23	PoS23err	PoS2A	PoS2Aerr	Paths	PoS3A	PoS3Aerr	PoS1A	PoS1Aerr
Lung, Non-Small Cell	1501	71.0	1.2	690	42.0	1.9	2.2	0.6	172	8.7	2.2	1.5	0.4
Breast	1373	70.0	1.2	641	47.1	2.0	15.1	1.5	214	45.3	3.4	10.1	1.0
Colorectal	1351	56.9	1.3	482	33.6	2.2	5.8	1.1	114	24.6	4.0	2.8	0.5
Prostate	1054	63.7	1.5	465	36.1	2.2	8.6	1.4	120	33.3	4.3	5.0	0.8
Pancreas	1052	57.7	1.5	342	27.5	2.4	2.9	1.0	60	16.7	4.8	1.3	0.4
Lymphoma, Non-Hodgkin's	1038	75.3	1.3	474	44.5	2.3	3.4	0.9	101	15.8	3.6	2.6	0.6
Ovarian	1033	61.1	1.5	367	32.2	2.4	5.4	1.3	75	26.7	5.1	2.8	0.6
Melanoma	963	59.6	1.6	347	26.5	2.4	7.2	1.5	60	41.7	6.4	3.6	0.7
Unspecified Solid Tumor	888	54.2	1.7	163	38.0	3.8	1.2	1.0	31	6.5	4.4	0.4	0.3
Liver	874	63.6	1.6	328	36.6	2.7	2.7	1.0	58	15.5	4.8	1.5	0.5
Head/Neck	869	61.6	1.7	305	35.1	2.7	3.6	1.1	67	16.4	4.5	1.8	0.5
Renal	833	61.7	1.7	303	30.7	2.6	4.0	1.2	64	18.8	4.9	2.0	0.6
Soft Tissue Sarcoma	800	57.9	1.7	260	23.1	2.6	0.8	0.6	25	8.0	5.4	0.4	0.3
Gastric	762	61.9	1.8	296	35.5	2.8	13.5	2.1	73	54.8	5.8	7.2	1.1
Esophageal	636	58.5	2.0	245	36.3	3.1	3.3	1.3	45	17.8	5.7	1.7	0.6
Bladder	619	63.5	1.9	199	39.2	3.5	8.5	2.1	52	32.7	6.5	4.3	1.0
Leukemia, Acute Myelogenous	598	78.3	1.7	313	44.4	2.8	5.4	1.5	54	31.5	6.3	4.7	1.1
Multiple Myeloma	597	80.2	1.6	324	40.7	2.7	9.0	1.7	83	34.9	5.2	7.4	1.3
Unspecified Cancer	580	41.6	2.0	167	57.5	3.8	7.2	2.1	80	15.0	4.0	2.4	0.7
CNS, Glioblastoma	549	81.4	1.7	246	30.5	2.9	5.3	1.5	49	26.5	6.3	4.0	1.1
Leukemia, Chronic Lymphocytic	528	74.6	1.9	248	36.3	3.1	5.2	1.5	72	18.1	4.5	3.6	1.0
Lymphoma, Hodgkin's	506	64.8	2.1	201	36.3	3.4	5.5	1.7	47	23.4	6.2	3.1	0.9
Lung, Small Cell	474	65.6	2.2	213	42.7	3.4	2.8	1.2	58	10.3	4.0	1.7	0.7
Leukemia, Acute Lymphocytic	457	78.3	1.9	227	41.0	3.3	7.0	1.9	43	37.2	7.4	5.8	1.4
Myelodysplastic Syndrome	444	81.8	1.8	229	37.6	3.2	0.0	0.0	30	0.0	0.0	0.0	0.0
Supportive Care	423	89.1	1.5	333	67.6	2.6	11.7	1.9	176	22.2	3.1	11.8	1.8
Endometrial	411	57.7	2.4	141	25.5	3.7	0.7	0.8	20	5.0	4.9	0.3	0.3
Leukemia, Chronic Myelogenous	364	71.7	2.4	174	27.6	3.4	6.9	2.0	39	30.8	7.4	4.5	1.3
Metastatic Cancer	351	86.3	1.8	231	55.4	3.3	6.1	1.8	78	17.9	4.3	6.1	1.6
CNS, Other	334	73.1	2.4	151	22.5	3.4	6.6	2.1	28	35.7	9.1	4.3	1.3
Thyroid	333	45.0	2.7	82	26.8	4.9	3.7	2.2	15	20.0	10.3	1.2	0.7
Mesothelioma	330	49.7	2.8	107	36.4	4.7	1.9	1.4	22	9.1	6.1	0.8	0.6
Osteosarcoma	303	48.2	2.9	87	12.6	3.6	0.0	0.0	10	0.0	0.0	0.0	0.0
Cervical	244	67.2	3.0	68	42.6	6.0	0.0	0.0	4	0.0	0.0	0.0	0.0
GIST	214	52.3	3.4	59	35.6	6.2	3.4	2.5	15	13.3	8.8	1.3	0.9
Myeloproliferative Neoplasms	207	83.1	2.6	105	37.1	4.7	2.9	1.7	26	11.5	6.3	2.4	1.3
CNS, Medulloblastoma	168	56.5	3.8	55	40.0	6.6	5.5	3.5	9	33.3	15.7	2.6	1.5
Unspecified Hematological Cancer	141	73.0	3.7	59	30.5	6.0	0.0	0.0	18	0.0	0.0	0.0	0.0
Testicular	123	65.0	4.3	58	32.8	6.2	1.7	1.7	19	5.3	5.1	1.0	1.0
Skin, Basal Cell Carcinoma	123	41.5	4.4	28	32.1	8.8	10.7	6.1	7	42.9	18.7	3.1	1.7
Total	24448	65.0	0.3	9813	38.0	0.5	5.7	0.3	2333	24.1	0.9	3.3	0.1

Table 3.5: Probability of success (PoS) for different disease groups in oncology.

Variable	Estimate	Std Error	t-statistic	p-value
Intercept	4.15 $\times 10^{-2}$	8.57$\times 10^{-3}$	4.83	2.20 $\times 10^{-5}$
x	-1.74 $\times 10^{-5}$	1.21 $\times 10^{-5}$	-1.44	0.158

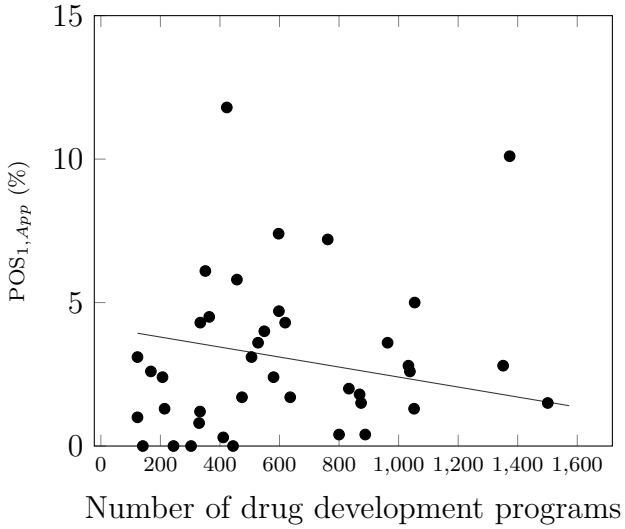


Figure 3-8: Scatterplot of PoS vs. number of paths.

3.3.4 PoS of Oncology Trials using Biomarkers (Patient Selection)

It has been shown that the use of biomarkers in patient selection greatly enhances the efficiency of the clinical trial process. We investigate if this holds true in oncology by computing the PoS of drug development programs that use biomarkers in selecting patients and comparing them to those that do not (see Table 3.6). As can be seen, for the vast majority of the disease types and phases, the use of biomarkers in patient selection improved the PoS. On average across disease types, the use of biomarkers improves the $\text{PoS}_{1,App}$ by 13.3%.

3.3.5 PoS of Orphan Oncology Programs

We make further subsets of our data in order to investigate the probability of success of oncology development programs involving rare diseases. We observe that for some diseases, such as acute lymphocytic leukaemia, myelodysplastic syndrome and acute myelogenous leukaemia, orphan drug development overlaps with more than 30% of all oncology drug development programs (see Figure 3-10).

The breakdown of the PoS for orphan drug development programs is contained in Table

3.7. We provide a comparison of the overall PoS of orphan diseases to the overall PoS of all oncology trials in Figure 3-9. Orphan drug development programs in oncology seldom make it to approval, with only 14 out of 40 disease groups having one approval or more.

In general, the overall PoS for orphan drug development programs across disease types is lower than the PoS of all oncology programs across the same diseases. A closer look at the individual phase probabilities reveals that, while orphan drugs have a higher phase 1 PoS (76.9% vs 65.0%) and a higher phase 2 PoS (44.6% vs 38.0%), a lower proportion of development programs make it from phase 3 to approval (10.1% vs 24.1%). This results in a lower overall PoS for orphan drugs. Admittedly, our analysis of orphan drugs suffers from a low number of data points and a high uncertainty, and may not reflect the true situation.

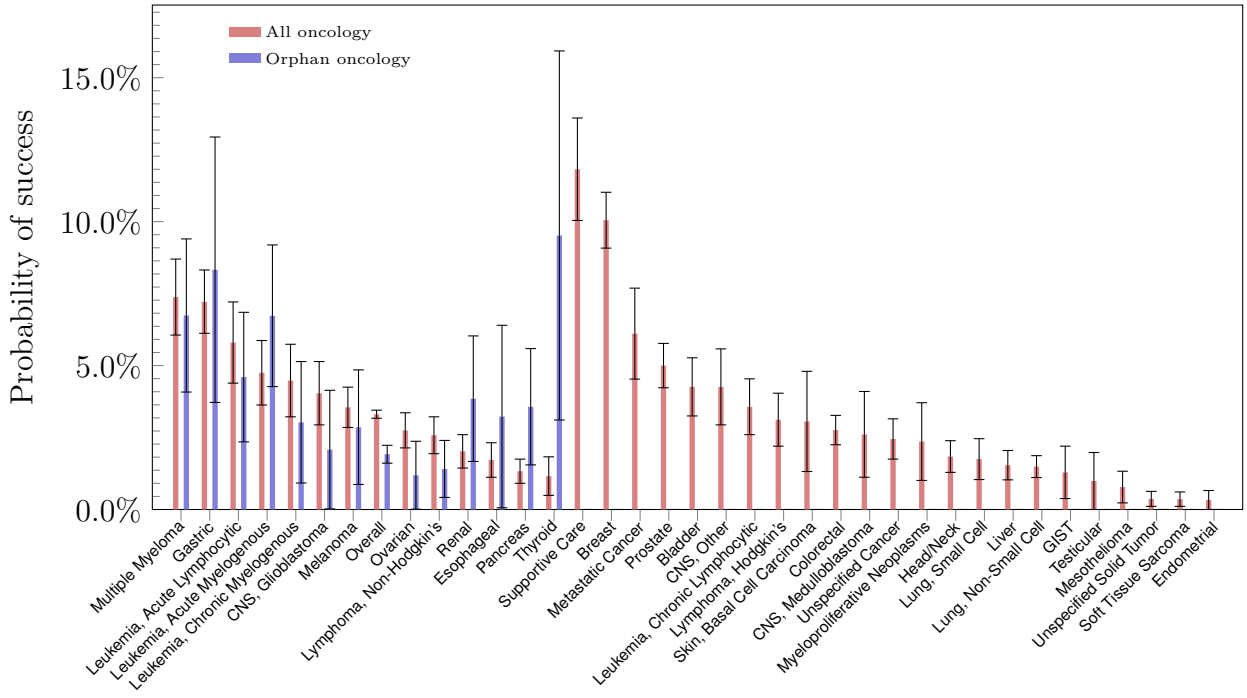


Figure 3-9: Comparison of the orphan drug PoS against the PoS for all oncology development programs. Only 14 diseases have at least one approval. Full results for orphan drug development programs are in Table 3.7

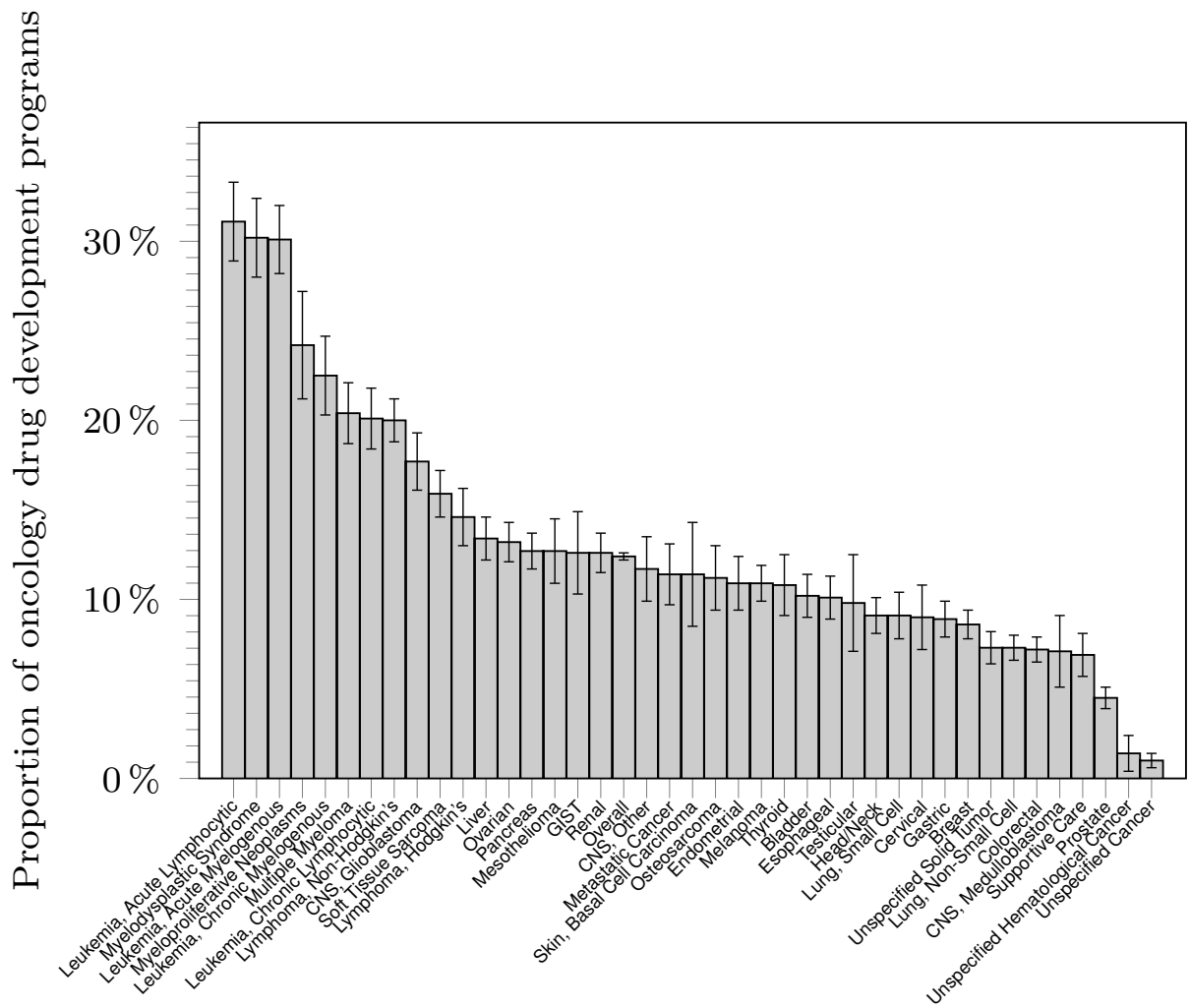


Figure 3-10: Orphan drug development programs as a proportion of oncology drug development programs.

Section References

[1] R Fisher, L Pusztai, and C Swanton. Cancer heterogeneity: implications for targeted therapeutics. *British journal of cancer*, 108(3):479, 2013.

Disease	Biomarkers?	Paths	PoS12	PoS12err	Paths	PoS23	PoS23err	Paths	PoS3A	PoS3Aerr	PoS1A	PoS1Aerr
Breast	No biomarkers	500	34.8	2.1	194	20.1	2.9	96	32.3	4.8	2.3	0.9
	With biomarkers	252	65.9	3.0	317	42.0	2.8	118	55.9	4.6	15.5	3.2
Lung, Non-Small Cell	No biomarkers	595	40.8	2.0	398	23.9	2.1	144	5.6	1.9	0.5	0.3
	With biomarkers	224	62.5	3.2	181	46.4	3.7	28	25.0	8.2	7.3	3.7
Colorectal	No biomarkers	658	27.5	1.7	276	16.3	2.2	84	14.3	3.8	0.6	0.3
	With biomarkers	201	47.8	3.5	124	28.2	4.0	30	53.3	9.1	7.2	3.1
Ovarian	No biomarkers	474	31.2	2.1	248	18.5	2.5	64	17.2	4.7	1.0	0.5
	With biomarkers	168	54.8	3.8	74	36.5	5.6	11	81.8	11.6	16.3	6.7
Melanoma	No biomarkers	429	27.5	2.2	231	13.4	2.2	39	23.1	6.7	0.9	0.5
	With biomarkers	150	48.0	4.1	89	38.2	5.2	21	76.2	9.3	14.0	5.3
Pancreas	No biomarkers	505	26.5	2.0	240	15.4	2.3	59	15.3	4.7	0.6	0.4
	With biomarkers	138	46.4	4.2	59	23.7	5.5	1	100.0	0.0	11.0	3.8
Lymphoma, Non-Hodgkin's	No biomarkers	385	45.5	2.5	285	22.5	2.5	79	7.6	3.0	0.8	0.5
	With biomarkers	138	66.7	4.0	127	66.9	4.2	22	45.5	10.6	20.3	7.9
Unspecified Solid Tumor	No biomarkers	457	21.0	1.9	105	16.2	3.6	31	6.5	4.4	0.2	0.3
	With biomarkers	131	64.9	4.2	22	40.9	10.5	0	–	–	–	–
Gastric	No biomarkers	328	28.7	2.5	172	10.5	2.3	43	51.2	7.6	1.5	0.8
	With biomarkers	120	53.3	4.6	69	46.4	6.0	30	60.0	8.9	14.8	6.1
Esophageal	No biomarkers	275	24.4	2.6	142	13.4	2.9	25	4.0	3.9	0.1	0.2
	With biomarkers	118	52.5	4.6	61	45.9	6.4	20	35.0	10.7	8.4	5.2
Head/Neck	No biomarkers	390	27.7	2.3	192	16.1	2.7	61	9.8	3.8	0.4	0.3
	With biomarkers	110	52.7	4.8	62	40.3	6.2	6	83.3	15.2	17.7	8.7
Liver	No biomarkers	395	31.9	2.3	239	21.3	2.7	57	14.0	4.6	1.0	0.6
	With biomarkers	108	54.6	4.8	34	41.2	8.4	1	100.0	0.0	22.5	7.0
Prostate	No biomarkers	500	33.8	2.1	349	24.4	2.3	114	29.8	4.3	2.5	0.8
	With biomarkers	102	49.0	4.9	53	37.7	6.7	6	100.0	0.0	18.5	5.5
Renal	No biomarkers	411	33.3	2.3	224	14.7	2.4	56	14.3	4.7	0.7	0.5
	With biomarkers	91	50.5	5.2	40	52.5	7.9	8	50.0	17.7	13.3	9.5
Bladder	No biomarkers	256	27.0	2.8	119	16.0	3.4	43	23.3	6.4	1.0	0.7
	With biomarkers	87	55.2	5.3	39	46.2	8.0	9	77.8	13.9	19.8	10.2
Soft Tissue Sarcoma	No biomarkers	458	34.5	2.2	199	14.6	2.5	17	5.9	5.7	0.3	0.4
	With biomarkers	76	51.3	5.7	45	33.3	7.0	8	12.5	11.7	2.1	3.4
Leukemia, Acute Lymphocytic	No biomarkers	128	39.1	4.3	124	16.1	3.3	19	21.1	9.4	1.3	1.2
	With biomarkers	72	70.8	5.4	78	61.5	5.5	24	50.0	10.2	21.8	9.0
Endometrial	No biomarkers	170	21.8	3.2	91	12.1	3.4	16	6.3	6.1	0.2	0.3
	With biomarkers	71	42.3	5.9	32	21.9	7.3	4	0.0	0.0	0.0	0.0
Leukemia, Acute Myelogenous	No biomarkers	217	49.8	3.4	200	26.0	3.1	41	14.6	5.5	1.9	1.2
	With biomarkers	69	69.6	5.5	78	66.7	5.3	13	84.6	10.0	39.2	11.9
CNS, Glioblastoma	No biomarkers	173	54.9	3.8	180	13.3	2.5	37	18.9	6.4	1.4	1.0
	With biomarkers	59	59.3	6.4	34	55.9	8.5	12	50.0	14.4	16.6	10.7
Lymphoma, Hodgkin's	No biomarkers	207	28.5	3.1	127	9.4	2.6	44	18.2	5.8	0.5	0.4
	With biomarkers	56	46.4	6.7	36	63.9	8.0	3	100.0	0.0	29.7	8.5
Leukemia, Chronic Lymphocytic	No biomarkers	186	34.9	3.5	166	23.5	3.3	50	8.0	3.8	0.7	0.6
	With biomarkers	54	75.9	5.8	54	42.6	6.7	22	40.9	10.5	13.2	7.5
Multiple Myeloma	No biomarkers	193	50.8	3.6	234	25.2	2.8	75	28.0	5.2	3.6	1.5
	With biomarkers	47	51.1	7.3	46	63.0	7.1	8	100.0	0.0	32.2	8.7
Thyroid	No biomarkers	200	21.5	2.9	55	14.5	4.8	14	14.3	9.4	0.4	0.7
	With biomarkers	44	40.9	7.4	17	23.5	10.3	1	100.0	0.0	9.6	6.7
Lung, Small Cell	No biomarkers	212	34.9	3.3	142	22.5	3.5	58	10.3	4.0	0.8	0.6
	With biomarkers	42	40.5	7.6	17	29.4	11.1	0	–	–	–	–
Cervical	No biomarkers	77	20.8	4.6	33	15.2	6.2	4	0.0	0.0	0.0	0.0
	With biomarkers	39	51.3	8.0	20	45.0	11.1	0	–	–	–	–
Myelodysplastic Syndrome	No biomarkers	141	51.8	4.2	159	20.1	3.2	29	0.0	0.0	0.0	0.0
	With biomarkers	37	64.9	7.8	38	57.9	8.0	1	0.0	0.0	0.0	0.0
Osteosarcoma	No biomarkers	165	20.0	3.1	69	8.7	3.4	10	0.0	0.0	0.0	0.0
	With biomarkers	34	26.5	7.6	13	0.0	0.0	0	–	–	–	–
Mesothelioma	No biomarkers	194	24.2	3.1	71	14.1	4.1	22	9.1	6.1	0.3	0.4
	With biomarkers	31	38.7	8.7	19	63.2	11.1	0	–	–	–	–
Leukemia, Chronic Myelogenous	No biomarkers	148	37.8	4.0	103	8.7	2.8	29	17.2	7.0	0.6	0.6
	With biomarkers	30	63.3	8.8	45	28.9	6.8	10	70.0	14.5	12.8	8.9
Metastatic Cancer	No biomarkers	67	34.3	5.8	110	32.7	4.5	61	18.0	4.9	2.0	1.4
	With biomarkers	20	80.0	8.9	57	49.1	6.6	17	17.6	9.2	6.9	6.4
CNS, Other	No biomarkers	136	42.6	4.2	125	10.4	2.7	28	35.7	9.1	1.6	1.2
	With biomarkers	19	36.8	11.1	8	37.5	17.1	0	–	–	–	–
GIST	No biomarkers	121	23.1	3.8	30	10.0	5.5	10	0.0	0.0	0.0	0.0
	With biomarkers	17	47.1	12.1	21	47.6	10.9	5	40.0	21.9	9.0	12.5
Myeloproliferative Neoplasms	No biomarkers	43	34.9	7.3	60	16.7	4.8	26	11.5	6.3	0.7	0.9
	With biomarkers	15	53.3	12.9	23	30.4	9.6	0	–	–	–	–
Unspecified Cancer	No biomarkers	377	12.7	1.7	86	29.1	4.9	80	15.0	4.0	0.6	0.4
	With biomarkers	13	23.1	11.7	11	9.1	8.7	0	–	–	–	–
Unspecified Hematological Cancer	No biomarkers	35	14.3	5.9	40	10.0	4.7	18	0.0	0.0	0.0	0.0
	With biomarkers	11	27.3	13.4	5	0.0	0.0	0	–	–	–	–
Testicular	No biomarkers	43	18.6	5.9	51	25.5	6.1	19	5.3	5.1	0.2	0.6
	With biomarkers	10	20.0	12.6	1	0.0	0.0	0	–	–	–	–
Skin, Basal Cell Carcinoma	No biomarkers	70	8.6	3.3	13	15.4	10.0	7	42.9	18.7	0.6	1.3
	With biomarkers	9	11.1	10.5	9	11.1	10.5	0	–	–	–	–
CNS, Medulloblastoma	No biomarkers	87	23.0	4.5	35	28.6	7.6	9	33.3	15.7	2.2	2.7
	With biomarkers	7	14.3	13.2	8	0.0	0.0	0	–	–	–	–
Supportive Care	No biomarkers	120	61.7	4.4	195	45.6	3.6	160	20.6	3.2	5.8	1.9
	With biomarkers	0	–	–	12	83.3	10.8	16	37.5	12.1	–	–

Table 3.6: Phase-by-phase probability of success (PoS) for drug development programs, with and without biomarkers for patient selection.

Disease	P1totalPaths	PoS12	PoS12err	P2totalPaths	PoS23	PoS23err	PoS2A	PoS2Aerr	Paths	PoS3A	PoS3Aerr	PoS1A	PoS1Aerr
Lymphoma, Non-Hodgkin's	208	84.1	2.5	129	39.5	4.3	1.6	1.2	31	6.5	4.4	1.4	1.0
Leukemia, Acute Myelogenous	180	85.0	2.7	116	50.9	4.6	6.0	2.7	20	35.0	10.7	6.7	2.5
Leukemia, Acute Lymphocytic	142	83.1	3.1	87	55.2	5.3	4.6	2.6	24	16.7	7.6	4.6	2.2
Ovarian	136	75.0	3.7	57	47.4	6.6	1.8	1.9	20	5.0	4.9	1.2	1.2
Pancreas	134	74.6	3.8	71	50.7	5.9	4.2	2.8	15	20.0	10.3	3.6	2.0
Myelodysplastic Syndrome	134	79.9	3.5	72	36.1	5.7	0.0	0.0	5	0.0	0.0	0.0	0.0
Soft Tissue Sarcoma	127	69.3	4.1	58	39.7	6.4	0.0	0.0	8	0.0	0.0	0.0	0.0
Multiple Myeloma	122	77.0	3.8	73	53.4	5.8	8.2	3.5	27	22.2	8.0	6.7	2.7
Breast	118	72.9	4.1	56	37.5	6.5	0.0	0.0	15	0.0	0.0	0.0	0.0
Liver	117	79.5	3.7	59	61.0	6.3	0.0	0.0	13	0.0	0.0	0.0	0.0
Lung, Non-Small Cell	109	73.4	4.2	57	56.1	6.6	0.0	0.0	19	0.0	0.0	0.0	0.0
Leukemia, Chronic Lymphocytic	106	71.7	4.4	56	42.9	6.6	0.0	0.0	12	0.0	0.0	0.0	0.0
Melanoma	105	81.0	3.8	62	50.0	6.4	3.2	2.5	19	10.5	7.0	2.9	2.0
Renal	105	77.1	4.1	55	38.2	6.6	5.5	3.1	20	15.0	8.0	3.8	2.2
Colorectal	97	63.9	4.9	30	13.3	6.2	0.0	0.0	4	0.0	0.0	0.0	0.0
CNS, Glioblastoma	97	93.8	2.4	61	60.7	6.3	1.6	2.0	18	5.6	5.4	2.1	2.1
Leukemia, Chronic Myelogenous	82	70.7	5.0	44	22.7	6.3	4.5	3.2	8	25.0	15.3	3.0	2.1
Head/Neck	79	75.9	4.8	36	41.7	8.2	0.0	0.0	5	0.0	0.0	0.0	0.0
Lymphoma, Hodgkin's	74	60.8	5.7	31	6.5	4.4	0.0	0.0	2	0.0	0.0	0.0	0.0
Gastric	68	79.4	4.9	32	40.6	8.7	9.4	6.2	3	100.0	0.0	8.3	4.6
Unspecified Solid Tumor	65	67.7	5.8	8	0.0	0.0	–	–	0	–	–	–	–
Esophageal	64	89.1	3.9	37	37.8	8.0	2.7	3.3	1	100.0	0.0	3.2	3.2
Bladder	63	66.7	5.9	26	38.5	9.5	0.0	0.0	2	0.0	0.0	0.0	0.0
Myeloproliferative Neoplasms	50	90.0	4.2	34	73.5	7.6	0.0	0.0	20	0.0	0.0	0.0	0.0
Prostate	47	53.2	7.3	16	0.0	0.0	–	–	0	–	–	–	–
Endometrial	45	75.6	6.4	19	26.3	10.1	0.0	0.0	5	0.0	0.0	0.0	0.0
Lung, Small Cell	43	90.7	4.4	33	33.3	8.2	0.0	0.0	3	0.0	0.0	0.0	0.0
Mesothelioma	42	69.0	7.1	19	57.9	11.3	0.0	0.0	5	0.0	0.0	0.0	0.0
Metastatic Cancer	40	87.5	5.2	27	63.0	9.3	0.0	0.0	17	0.0	0.0	0.0	0.0
CNS, Other	39	84.6	5.8	23	43.5	10.3	0.0	0.0	3	0.0	0.0	0.0	0.0
Thyroid	36	72.2	7.5	14	71.4	12.1	14.3	10.6	7	28.6	17.1	9.5	6.4
Osteosarcoma	34	61.8	8.3	15	6.7	6.4	0.0	0.0	1	0.0	0.0	0.0	0.0
Supportive Care	29	62.1	9.0	12	91.7	8.0	0.0	0.0	7	0.0	0.0	0.0	0.0
GIST	27	81.5	7.5	14	35.7	12.8	0.0	0.0	3	0.0	0.0	0.0	0.0
Cervical	22	81.8	8.2	8	100.0	0.0	–	–	0	–	–	–	–
Skin, Basal Cell Carcinoma	14	92.9	6.9	4	0.0	0.0	0.0	0.0	0	–	–	–	–
CNS, Medulloblastoma	12	75.0	12.5	4	75.0	21.7	0.0	0.0	3	0.0	0.0	0.0	0.0
Testicular	12	100.0	0.0	7	0.0	0.0	–	–	0	–	–	–	–
Unspecified Cancer	6	16.7	15.2	0	–	–	–	–	0	–	–	–	–
Unspecified Hematological Cancer	2	0.0	0.0	0	–	–	–	–	0	–	–	–	–
Overall	3032	76.9	0.8	1562	44.6	1.3	2.4	0.4	365	10.1	1.6	1.9	0.3

Table 3.7: Probability of success (PoS) for orphan development programs in oncology.

3.4 Statistics of Vaccine and Other Anti-Infective Therapeutic Development Programs

In this section, we provide estimates of the historical probabilities of success (PoSs) of clinical trials for vaccines and other therapeutic drugs for infectious diseases to inform discussions on the planning and financing of the fight against one of humanity's oldest foes. This is of particular importance in light of the recent havoc wreaked by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease (COVID-19).

While the PoSs of therapeutic drugs for various disease groups are well-documented [2, 5, 11, 23, 24, 28, 29], relatively little has been published on treatments for infectious diseases and vaccine development despite their importance [4, 20]. Prior studies have focused on narrower subsets relevant to their specific interests and have relied on much more limited data sets. For example, Young et al. [31] employed 10 to 25 data points per estimated value from the Bill and Melinda Gates Foundation to estimate the PoSs of vaccines for neglected diseases, and DiMasi et al. [6] reported PoS estimates on a per-drug basis using 2,575 trials for diseases of interest to the Gates Foundation. In contrast, we employ a much larger and broader data set of 16,328 unique clinical trials to estimate the PoSs of vaccines and nonvaccine therapeutics targeting 29 different infectious diseases using all available drug-indication pairs – a methodology that has been argued to be more relevant for evaluating drug development R&D risk and productivity [29].

Vaccination is commonly recognized as one of the most cost-effective public health measures for combating infectious diseases [3, 7, 15, 20, 21, 22]. In developed countries, routine prophylactic vaccination and effective treatment options have led to the control or complete elimination of several deadly infectious diseases through individual and herd immunity, preventing millions of deaths and untold suffering each year. This prophylaxis dramatically reduces the burden on the health care system and society as a whole. In addition, the deaths, hospitalizations, and treatment costs avoided by these measures have led to significant economic savings [7, 18, 22].

As technology continues to advance, one expects that the human species will be better

able to cope with these diseases. The fact remains, however, that we still do not have effective treatments or vaccines for many infectious diseases. While the discovery of antibiotics has reduced the mortality rate of bacterial infection, and the development of the smallpox vaccine has led to the eradication of the devastating disease [19], other infectious diseases, such as acquired immunodeficiency syndrome (AIDS) and malaria, still take the lives of tens of millions every year. According to the World Health Organization, there are currently only 26 infectious diseases that are preventable by available vaccines [30].

By developing better risk measures for therapeutic development, we hope to facilitate greater investment, identification of underserved areas that require public sector support, and more efficient business and financing models in this critical field.

3.4.1 Data Summary

For this study, we used clinical trials metadata from the January 7, 2020, snapshots of Citeline's PharmaProjects and TrialTrove databases. We filter our data to include only trials that have been tagged by Citeline as being in the 'Infectious Disease' or 'Vaccines (Infectious Diseases)' therapeutic areas. Since the two therapeutic areas may overlap in data points, we define clinical trials that are involved in any vaccine development as part of a 'vaccine' development program. In addition, we process the data such that more specific diseases (e.g., rabies) can be identified instead of broad vaccine classes (e.g., vector-borne disease vaccines). Clinical trials that are not involved in any vaccine development program will be deemed to be part of a 'nonvaccine' drug development program. We derive 43,414 data points in total. We define an 'industry-sponsored' development program as one where there is at least one commercial company involved in any stage of clinical development. The complement – in which there is no commercial company involved in any stage of the vaccine or drug development program – shall be referred to as 'non-industry-sponsored'. Given these definitions, a drug or vaccine development program (and the clinical trials in the program) can belong to only one of these mutually-exclusive sets: industry-sponsored vaccines, industry-sponsored nonvaccine therapeutics, non-industry-sponsored vaccines, and, non-industry-sponsored nonvaccine therapeutics.

The vaccines in TrialTrove are identified by broad categories such as "respiratory vac-

cines", "other viral vaccines", or "hepatitis vaccines". We attempt to infer the diseases targeted by the vaccines by cross-referencing the disease tags for each clinical trial. For example, a clinical trial may be tagged with both "hepatitis vaccines" and "HBV", allowing us to conclude that the vaccine is indicated for HBV (hepatitis B virus). Those clinical trials that have only vaccine tags will have their disease labeled as "others." Manual inspection of some of the clinical trial titles shows that this category includes diseases such as measles and tuberculosis.

We plot the number of development programs known to start in each month from January 2000 through December 2019 in Figure 3-11. There are 1,838 and 706 industry-sponsored and non-industry-sponsored vaccine development programs, respectively, and, 3,851 and 2,978 industry-sponsored and non-industry-sponsored nonvaccine drug development programs targeting infectious diseases, respectively. As can be seen from Figure 3-11, the number of industry-sponsored clinical programs attempting to treat infectious diseases is often greater than the number of vaccine development programs. We see a precipitous fall in the number of infectious disease treatment development programs initiated between late 2018 and mid-2019, which is likely to be related to declining investment in the research and development (R&D) of novel antibiotics, precipitated by the closure of antibiotics biotechnology firms and the withdrawal of pharmaceutical companies from the antibiotics business [13, 16].

Between January 2000 and June 2011, the number of non-industry-sponsored vaccine development programs initiated is on par with the number of non-industry-sponsored, non-vaccine anti-infective drug development programs initiated (see Figure 3-12). However, the number of nonvaccine drug development programs initiated begin to outpace the number of vaccine development programs after January 2012, and experience a rapid boom between mid-2015 and mid-2018 before declining rapidly between October 2018 and January 2019.

3.4.2 Result

Vaccines

Overall, 2,544 vaccine development programs are observed in our data set, of which 1,838 are sponsored by industry and 706 do not involve any industry sponsor in any stage of

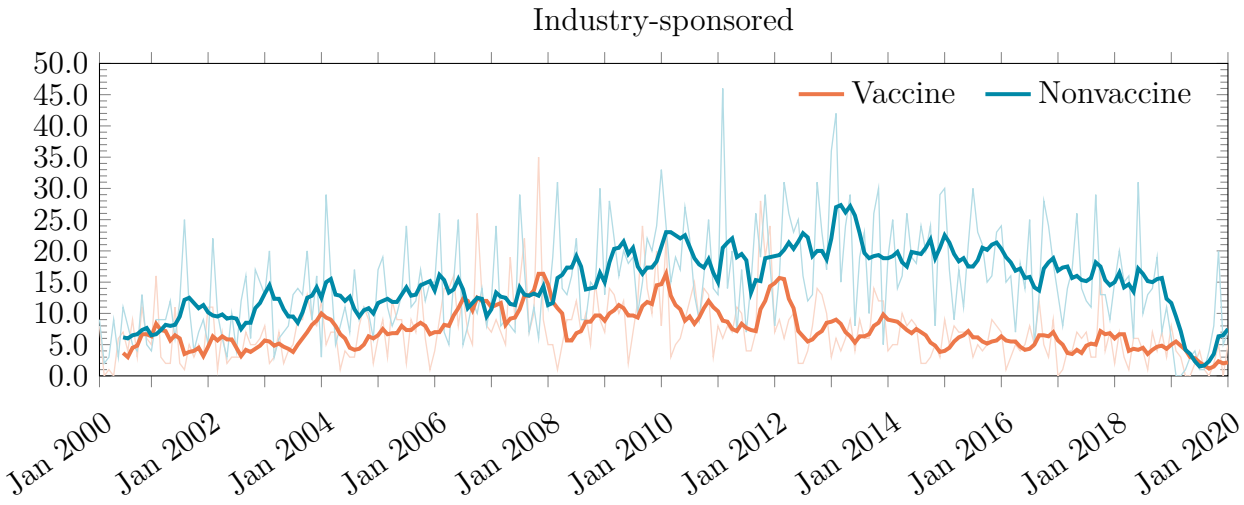


Figure 3-11: The number of industry-sponsored development programs initiated per month from January 2000 through December 2019 in the areas of vaccine and nonvaccine treatment for infectious diseases (thin, light colored lines). The darker, thicker lines represent the 6-month moving average of the individual series.

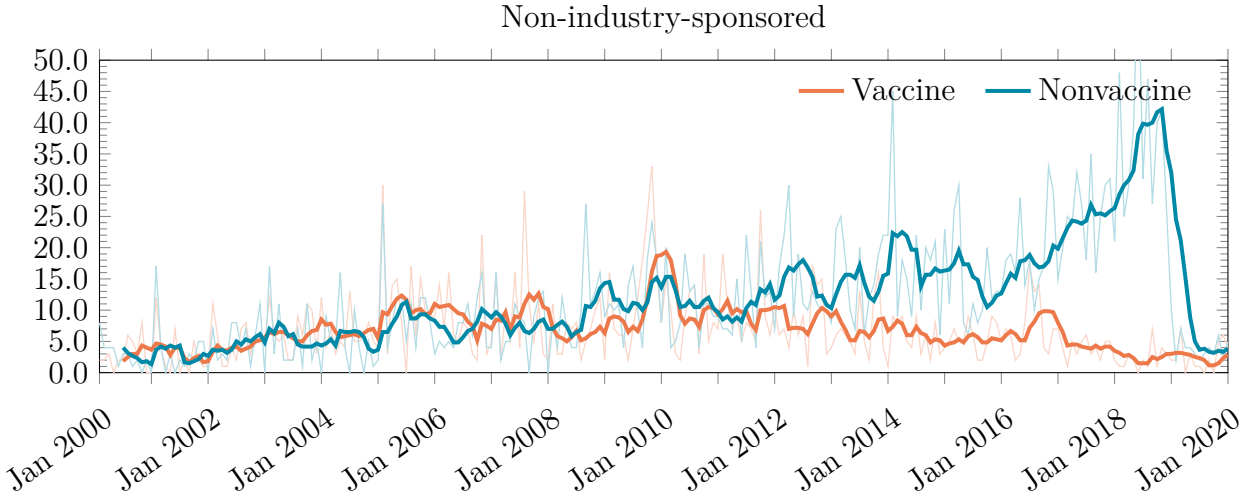
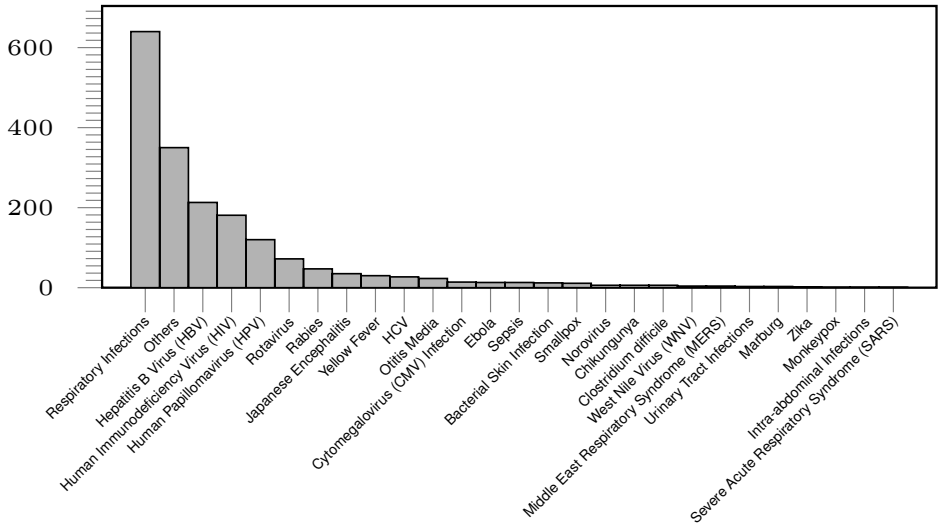


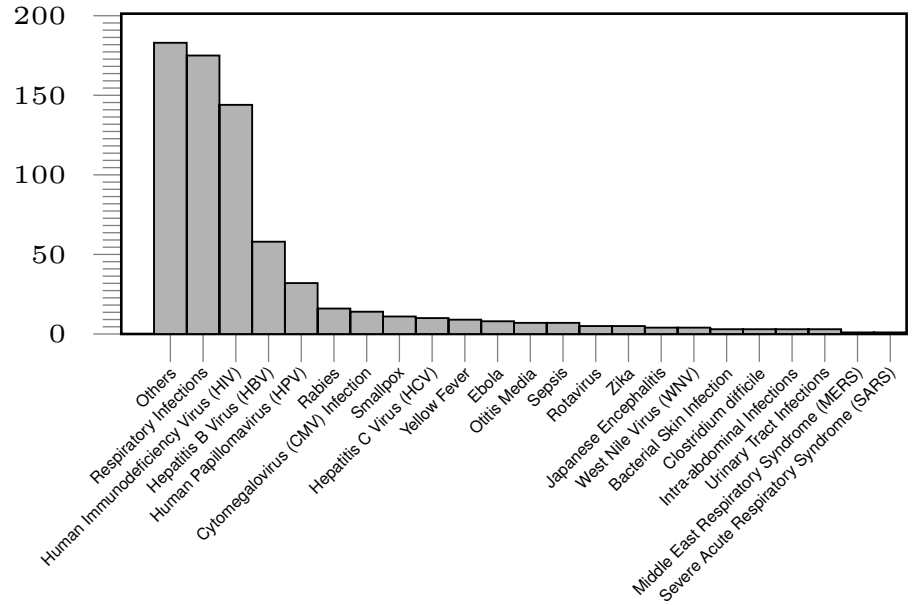
Figure 3-12: The number of non-industry-sponsored development programs initiated per month from January 2000 through December 2019 in the areas of vaccine and nonvaccine treatment for infectious diseases (thin, light colored lines). The darker, thicker lines represent the 6-month moving average of the individual series.

development. For industry-sponsored drug development programs, respiratory infections is the most actively researched vaccine category, comprising 34.8% (n = 640) of all vaccine development programs (see Figure 3-13). HBV and human immunodeficiency virus (HIV) vaccines represent 11.6% (n=213) and 9.8% (n=181) of all vaccine development programs,

respectively, whereas intra-abdominal infections, monkeypox, and severe acute respiratory syndrome (SARS) vaccines are the least researched fields, with only one development path observed per disease.



(a) Number of industry-sponsored, vaccine drug development programs



(b) Number of non-industry-sponsored, vaccine drug development programs

Figure 3-13: Number of vaccine development programs observed for each vaccine type. HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; HPV, human papillomavirus.

From Table 3.8, we can see that the overall PoS for industry-sponsored vaccine development programs is 39.6% (standard error, or SE=1.2%), which is substantially higher than the average overall PoS of 10.8% (SE=0.1%) across all industry-sponsored drug development programs (see Table 3.2). These findings are largely in line with the results of Wong et al. [29], who first observed this trend, and of DiMasi et al. [6], despite the fact that the latter computed their estimates using a different method (a "phase-by-phase" approach) and considered only lead indications. We estimate PoS_{12}, PoS_{23}, and PoS_{3A} to be 82.5% (SE=0.9%), 65.4% (SE=1.3%), and 80.1% (SE=1.4%), respectively.

Across all industry-sponsored vaccine development programs, we can see that monkeypox vaccines have had the most developmental success, followed by rotavirus and Japanese encephalitis vaccines (see Table 3.8). Their overall success rates are 100% (SE=0.0%), 78.7% (SE=5.2%), and 67.6% (SE=8.0%), respectively. The overall PoS for monkeypox is based on only one sample. Only 12 diseases out of the 27 disease categories with at least one development path observed have seen at least one approved vaccine.

Table 3.8: The probabilities of success (PoSs) of industry-sponsored vaccine development programs.

	Phase 1			Phase 2					Phase 3			Overall		
Disease	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	Paths	PoS_{1A}	S.E.
Bacterial Skin Infection	12	83.3	10.8	7	14.3	13.2	0.0	0.0	1	0.0	0.0	9	0.0	0.0
Chikungunya	6	83.3	15.2	0	–	–	–	–	0	–	–	1	0.0	0.0
Clostridium difficile	6	100.0	0.0	6	33.3	19.2	0.0	0.0	0	–	–	4	0.0	0.0
Cytomegalovirus Infection (CMV)	14	57.1	13.2	3	33.3	27.2	0.0	0.0	0	–	–	8	0.0	0.0
Ebola	13	53.8	13.8	7	57.1	18.7	28.6	20.2	2	100.0	0.0	11	18.2	11.6
Hepatitis B Virus (HBV)	213	94.8	1.5	187	74.9	3.2	54.5	3.7	132	77.3	3.6	190	53.7	3.6
Hepatitis C Virus (HCV)	27	70.4	8.8	15	0.0	0.0	0.0	0.0	0	–	–	23	0.0	0.0
Human Immunodeficiency Virus (HIV)	181	65.2	3.5	95	36.8	4.9	0.0	0.0	21	0.0	0.0	144	0.0	0.0
Human Papillomavirus (HPV)	120	88.3	2.9	69	52.2	6.0	36.2	6.1	30	83.3	6.8	77	32.5	5.3
Intra−abdominal Infections	1	100.0	0.0	1	100.0	0.0	0.0	0.0	1	0.0	0.0	1	0.0	0.0
Japanese Encephalitis	35	100.0	0.0	35	71.4	7.6	65.7	8.1	24	95.8	4.1	34	67.6	8.0
Marburg	3	0.0	0.0	0	–	–	–	–	0	–	–	3	0.0	0.0
Middle East Respiratory Syndrome (MERS)	4	50.0	25.0	0	–	–	–	–	0	–	–	2	0.0	0.0
Monkeypox	1	100.0	0.0	1	100.0	0.0	100.0	0.0	1	100.0	0.0	1	100.0	0.0
Norovirus	6	100.0	0.0	5	0.0	0.0	0.0	0.0	0	–	–	5	0.0	0.0
Otitis Media	23	95.7	4.3	22	81.8	8.2	45.5	10.6	18	55.6	11.7	23	43.5	10.3
Rabies	47	91.5	4.1	40	87.5	5.2	65.0	8.1	30	86.7	6.2	39	66.7	7.5
Respiratory Infections	640	79.1	1.6	465	66.9	2.2	50.1	2.4	287	81.2	2.3	575	40.5	2.0
Rotavirus	72	97.2	1.9	70	91.4	3.3	68.6	6.0	53	90.6	4.0	61	78.7	5.2
Sepsis	13	38.5	13.5	5	80.0	17.9	0.0	0.0	4	0.0	0.0	13	0.0	0.0
Severe Acute Respiratory Syndrome (SARS)	1	0.0	0.0	0	–	–	–	–	0	–	–	1	0.0	0.0
Smallpox	11	81.8	11.6	8	62.5	17.1	50.0	17.7	5	80.0	17.9	10	40.0	15.5
Urinary Tract Infections	3	100.0	0.0	3	100.0	0.0	0.0	0.0	1	0.0	0.0	1	0.0	0.0
West Nile Virus (WNV)	4	25.0	21.7	1	100.0	0.0	0.0	0.0	1	0.0	0.0	4	0.0	0.0
Yellow Fever	30	90.0	5.5	26	73.1	8.7	57.7	10.5	15	100.0	0.0	25	60.0	9.8
Zika	2	0.0	0.0	0	–	–	–	–	0	–	–	2	0.0	0.0
Others	350	87.1	1.8	268	63.4	2.9	47.0	3.2	142	88.7	2.7	285	44.2	2.9
Total	1,838	82.5	0.9	1,339	65.4	1.3	45.9	1.4	768	80.1	1.4	1552	39.6	1.2

In contrast, non-industry-sponsored vaccine development programs have an overall PoS of

only 6.8% (SE=1.0%), with PoS_{12}, PoS_{23}, and PoS3A estimates of 63.3% (SE=1.8%), 37.3% (SE=2.6%), and 39.8% (SE=4.9%), respectively (Table 3.9). The top three indications with the highest overall success rates for non-industry-sponsored drug development programs are otitis media (28.6%, SE=17.1%), rabies (25.0%, SE=10.8%), and Japanese encephalitis (25.0%, SE=21.7%). The latter estimates are derived from only a handful of samples and must be interpreted with caution as their large standard errors suggest.

Table 3.9: The probabilities of success (PoSs) of non-industry-sponsored vaccine development programs.

	Phase 1			Phase 2					Phase 3			Overall		
Disease	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	Paths	PoS_{1A}	S.E.
Bacterial Skin Infection	3	100.0	0.0	0	–	–	–	–	0	–	–	0	0.0	0.0
Clostridium difficile	3	66.7	27.2	0	–	–	–	–	0	–	–	1	0.0	0.0
Cytomegalovirus Infection (CMV)	14	50.0	13.4	5	40.0	21.9	40.0	21.9	2	100.0	0.0	12	16.7	10.8
Ebola	8	12.5	11.7	0	–	–	–	–	0	–	–	7	0.0	0.0
Hepatitis B Virus (HBV)	58	91.4	3.7	48	47.9	7.2	8.3	4.3	16	25.0	10.8	46	8.7	4.2
Hepatitis C Virus (HCV)	10	70.0	14.5	5	0.0	0.0	0.0	0.0	0	–	–	8	0.0	0.0
HIV	144	48.6	4.2	62	3.2	2.2	0.0	0.0	2	0.0	0.0	136	0.0	0.0
Human Papillomavirus (HPV)	32	87.5	5.8	16	56.3	12.4	6.3	6.5	7	14.3	13.2	18	5.6	5.4
Intra−abdominal Infections	3	100.0	0.0	0	–	–	–	–	0	–	–	0	0.0	0.0
Japanese Encephalitis	4	100.0	0.0	4	100.0	0.0	25.0	21.7	4	25.0	21.7	4	25.0	21.7
Middle East Respiratory Syndrome (MERS)	1	100.0	0.0	0	–	–	–	–	0	–	–	0	0.0	0.0
Otitis Media	7	100.0	0.0	7	28.6	17.1	28.6	17.1	2	100.0	0.0	7	28.6	17.1
Rabies	16	81.3	9.8	13	53.8	13.8	30.8	12.8	7	57.1	18.7	16	25.0	10.8
Respiratory Infections	175	66.9	3.6	101	51.5	5.0	16.8	3.9	41	41.5	7.7	148	11.5	2.6
Rotavirus	5	80.0	17.9	4	50.0	25.0	0.0	0.0	1	0.0	0.0	4	0.0	0.0
Sepsis	7	42.9	18.7	2	0.0	0.0	0.0	0.0	0	–	–	6	0.0	0.0
Severe Acute Respiratory Syndrome (SARS)	1	0.0	0.0	0	–	–	–	–	0	–	–	1	0.0	0.0
Smallpox	11	63.6	14.5	6	16.7	15.2	16.7	15.2	1	100.0	0.0	10	10.0	9.5
Urinary Tract Infections	3	100.0	0.0	1	0.0	0.0	0.0	0.0	0	–	–	1	0.0	0.0
West Nile Virus (WNV)	4	0.0	0.0	0	–	–	–	–	0	–	–	4	0.0	0.0
Yellow Fever	9	66.7	15.7	6	33.3	19.2	0.0	0.0	1	0.0	0.0	8	0.0	0.0
Zika	5	40.0	21.9	0	–	–	–	–	0	–	–	3	0.0	0.0
Others	183	57.9	3.6	71	35.2	5.7	9.9	3.8	14	50.0	13.4	137	5.1	1.9
Total	706	63.3	1.8	351	37.3	2.6	11.1	1.8	98	39.8	4.9	577	6.8	1.0

Nonvaccine Anti-Infective Therapeutics

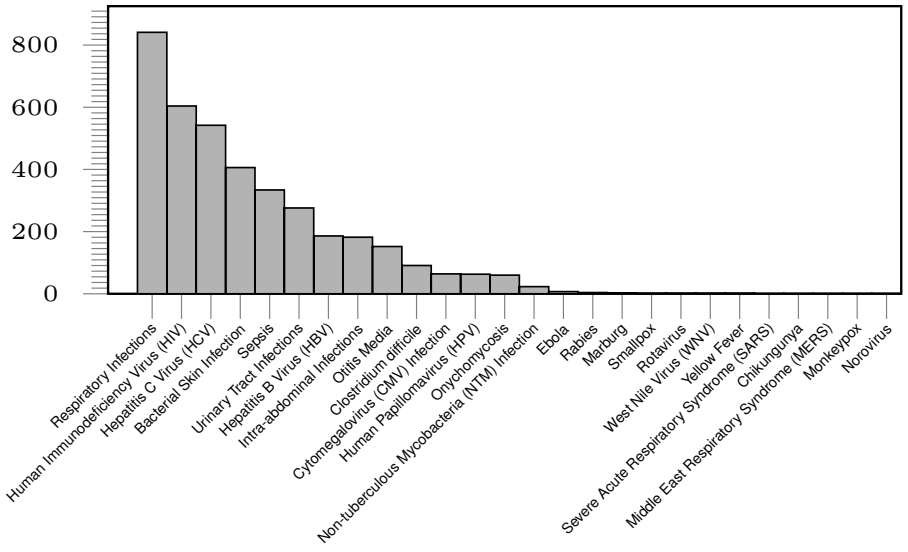
In contrast to vaccines, which are intended to prevent disease, a number of alternatives have been developed to treat–and, in some cases, cure–patients afflicted with an infectious disease. According to our data set, 3,851 and 2,978 industry-sponsored and non-industry-sponsored nonvaccine drug development programs, respectively, have been initiated in the area of infectious disease (see Figure 3-14). The top three diseases with the greatest number of industry-sponsored drug development programs are respiratory infections (21.8%), HIV (15.7%) and hepatitis C virus, or HCV (14.1%). Together, they comprise 51.6% of all industry-sponsored nonvaccine development programs. Non-industry anti-infectious-disease

drug development programs focus on treating respiratory infections (20.5%), HIV (13.9%), and bacterial skin infection (12.1%).

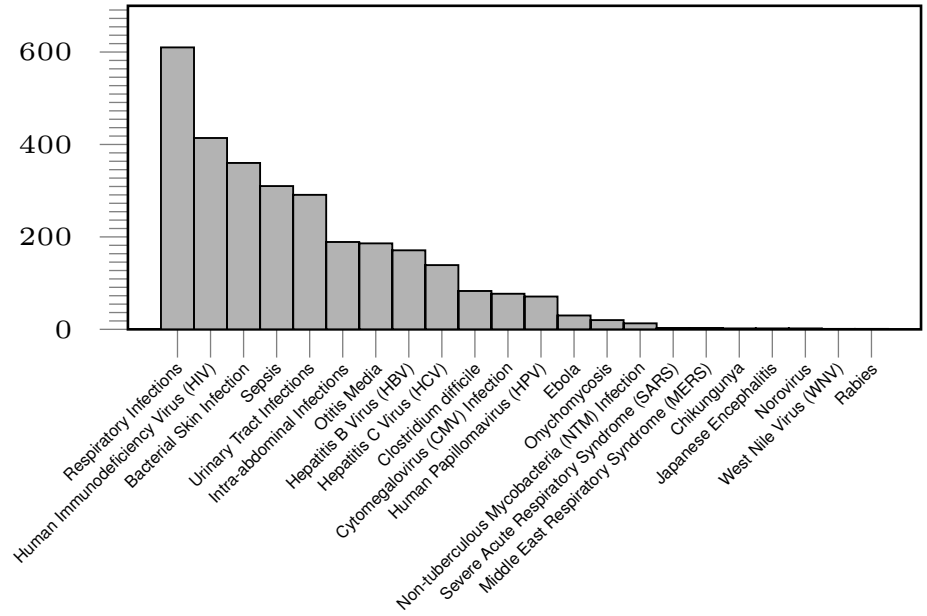
With respect to addressing the most recent virus outbreaks–MERS, SARS, Ebola, and Zika–a total of nine industry-sponsored and 36 non-industry-sponsored nonvaccine drug development programs were initiated over the past 20 years, and there have been no approved therapies to date.

From Table 3.10, we can see that the overall PoS across all industry-sponsored drug development programs treating infectious diseases is 16.3% (SE=0.7%). The PoS_{12}, PoS_{23}, and PoS3A are 65.0% (SE=0.8%), 64.3% (SE=1.0%), and 51.1% (SE=1.6%), respectively. Based on our data, the highest success rates for industry-sponsored nonvaccine development programs have been for smallpox (100.0%, SE=0.0%), cytomegalovirus (CMV) infection (31.8%, SE=7.0%), and onychomycosis (29.8%, SE=6.7%). There are currently no nonvaccine therapies approved for rotavirus, SARS, rabies, Ebola, West Nile virus, Marburg, yellow fever, chikungunya, MERS, monkeypox, or norovirus. With the exception of norovirus and MERS, these diseases without any vaccine are predominantly prevalent in nonindustrialized nations, and thus represent neglected diseases. It is also interesting that for the latter eight diseases, even the PoS_{12} is low. Since phase 1 trials in the development of anti-infective therapies focus primarily on safety, understanding the pharmacokinetics of the compound, and maximum tolerable dose levels, it can be inferred that the drugs tested are either of high toxicity or lack the necessary characteristics required for optimal absorption, distribution, metabolism, and excretion (ODME), or perhaps failed to advance due to financial constraints.

For non-industry-sponsored nonvaccine development programs, the overall PoS is 8.2% (SE=0.6%) while PoS12, PoS23, and PoS3A are 61.0% (SE=0.9%), 65.2% (SE=1.2%), and 30.0% (SE=1.8%), respectively (see Table 3.11). The top three indications with the highest overall success rates for non-industry-sponsored nonvaccine development programs are CMV infection (23.5%, SE=5.9%), clostridium difficile (20.5%, SE=6.5%), and sepsis (17.4%, SE=2.6%).



(a) Number of industry-sponsored, nonvaccine drug development programs



(b) Number of non-industry-sponsored, nonvaccine drug development programs

Figure 3-14: Number of nonvaccine development programs observed for each vaccine type. HBV, hepatitis B virus; HCV, hepatitis C virus; HIV, human immunodeficiency virus; HPV, human papillomavirus.

PoS by Biological Family and Transmission Routes (Industry-Sponsored)

In an attempt to shed more light on the industry-sponsored vaccine and nonvaccine drug development programs, we classify the diseases by their biological family and transmission

Table 3.10: The probabilities of success (PoSs) of industry-sponsored nonvaccine drug development programs.

	Phase 1			Phase 2					Phase 3			Overall		
Disease	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	Paths	PoS_{1A}	S.E.
Bacterial Skin Infection	406	54.9	2.5	207	72.9	3.1	19.8	3.2	104	39.4	4.8	343	12.0	1.8
Chikungunya	1	0.0	0.0	0	–	–	–	–	0	–	–	1	0.0	0.0
Clostridium difficile	91	83.5	3.9	66	53.0	6.1	4.5	2.8	25	12.0	6.5	71	4.2	2.4
Cytomegalovirus Infection (CMV)	64	87.5	4.1	43	60.5	7.5	32.6	7.8	19	73.7	10.1	44	31.8	7.0
Ebola	7	28.6	17.1	1	0.0	0.0	0.0	0.0	0	–	–	6	0.0	0.0
Hepatitis B Virus (HBV)	186	77.4	3.1	105	68.6	4.5	36.2	5.2	54	70.4	6.2	129	29.5	4.0
Hepatitis C Virus (HCV)	542	68.8	2.0	348	52.3	2.7	23.6	2.4	155	52.9	4.0	490	16.7	1.7
Human Immunodeficiency Virus (HIV)	604	63.2	2.0	326	59.8	2.7	39.3	2.8	167	76.6	3.3	520	24.6	1.9
Human Papillomavirus (HPV)	63	85.7	4.4	34	23.5	7.3	11.8	5.7	6	66.7	19.2	41	9.8	4.6
Intra−abdominal Infections	182	68.7	3.4	113	72.6	4.2	2.7	2.0	35	8.6	4.7	123	2.4	1.4
Marburg	3	0.0	0.0	0	–	–	–	–	0	–	–	3	0.0	0.0
Middle East Respiratory Syndrome (MERS)	1	0.0	0.0	0	–	–	–	–	0	–	–	1	0.0	0.0
Monkeypox	1	0.0	0.0	0	–	–	–	–	0	–	–	1	0.0	0.0
Non−tuberculous Mycobacteria (NTM) Infection	23	87.0	7.0	16	62.5	12.1	6.3	7.7	4	25.0	21.7	13	7.7	7.4
Norovirus	1	0.0	0.0	0	–	–	–	–	0	–	–	1	0.0	0.0
Onychomycosis	60	85.0	4.6	44	63.6	7.3	31.8	7.6	22	63.6	10.3	47	29.8	6.7
Otitis Media	152	48.0	4.1	68	80.9	4.8	51.5	6.2	51	68.6	6.5	143	24.5	3.6
Rabies	4	75.0	21.7	0	–	–	–	–	0	–	–	1	0.0	0.0
Respiratory Infections	841	64.2	1.7	476	70.0	2.1	22.9	2.2	222	49.1	3.4	666	16.4	1.4
Rotavirus	2	100.0	0.0	2	0.0	0.0	0.0	0.0	0	–	–	2	0.0	0.0
Sepsis	334	66.8	2.6	206	64.6	3.3	10.2	2.4	81	25.9	4.9	265	7.9	1.7
Severe Acute Respiratory Syndrome (SARS)	1	100.0	0.0	1	0.0	0.0	0.0	0.0	0	–	–	1	0.0	0.0
Smallpox	2	100.0	0.0	2	100.0	0.0	100.0	0.0	2	100.0	0.0	2	100.0	0.0
Urinary Tract Infections	276	55.1	3.0	143	72.0	3.8	10.5	3.2	51	29.4	6.4	215	7.0	1.7
West Nile Virus (WNV)	2	50.0	35.4	1	0.0	0.0	0.0	0.0	0	–	–	2	0.0	0.0
Yellow Fever	2	0.0	0.0	0	–	–	–	–	0	–	–	2	0.0	0.0
Total	3,851	65.0	0.8	2,202	64.3	1.0	23.2	1.0	998	51.1	1.6	3133	16.3	0.7

Table 3.11: The probabilities of success (PoSs) of non-industry-sponsored nonvaccine drug development programs.

	Phase 1			Phase 2					Phase 3			Overall		
Disease	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	Paths	PoS_{1A}	S.E.
Bacterial Skin Infection	360	46.4	2.6	151	81.5	3.2	19.2	3.7	85	34.1	5.1	306	9.5	1.7
Chikungunya	2	100.0	0.0	2	50.0	35.4	0.0	0.0	1	0.0	0.0	2	0.0	0.0
Clostridium difficile	83	94.0	2.6	51	76.5	5.9	15.7	6.2	22	36.4	10.3	39	20.5	6.5
Cytomegalovirus Infection (CMV)	77	83.1	4.3	51	51.0	7.0	23.5	6.9	13	92.3	7.4	51	23.5	5.9
Ebola	30	96.7	3.3	28	14.3	6.6	0.0	0.0	2	0.0	0.0	27	0.0	0.0
Hepatitis B Virus (HBV)	171	49.1	3.8	73	47.9	5.8	1.4	1.4	31	3.2	3.2	156	0.6	0.6
Hepatits C Virus (HCV)	139	84.2	3.1	112	43.8	4.7	8.9	2.9	33	30.3	8.0	118	8.5	2.6
Human Immunodeficiency Virus (HIV)	414	61.1	2.4	195	49.2	3.6	13.3	2.6	75	34.7	5.5	335	7.8	1.5
Human Papillomavirus (HPV)	71	88.7	3.8	42	42.9	7.6	2.4	2.7	8	12.5	11.7	40	2.5	2.5
Intra-abdominal Infections	189	66.1	3.4	112	76.8	4.0	12.5	3.8	51	27.5	6.2	141	9.9	2.5
Japanese Encephalitis	2	100.0	0.0	2	0.0	0.0	0.0	0.0	0	-	-	2	0.0	0.0
Middle East Respiratory Syndrome (MERS)	3	100.0	0.0	3	66.7	27.2	0.0	0.0	0	-	-	1	0.0	0.0
Non-tuberculous Mycobacteria (NTM) Infection	13	84.6	10.0	9	44.4	16.6	11.1	11.9	2	50.0	35.4	9	11.1	10.5
Norovirus	2	100.0	0.0	1	0.0	0.0	0.0	0.0	0	-	-	1	0.0	0.0
Onychomycosis	20	75.0	9.7	15	66.7	12.2	0.0	0.0	6	0.0	0.0	16	0.0	0.0
Otitis Media	186	30.1	3.4	53	56.6	6.8	7.5	3.9	24	16.7	7.6	177	2.3	1.1
Rabies	1	0.0	0.0	0	-	-	-	-	0	-	-	1	0.0	0.0
Respiratory Infections	610	58.5	2.0	323	72.8	2.5	11.5	2.1	141	26.2	3.7	482	7.7	1.2
Sepsis	310	80.0	2.3	227	77.5	2.8	15.9	3.0	94	38.3	5.0	207	17.4	2.6
Severe Acute Respiratory Syndrome (SARS)	3	100.0	0.0	3	100.0	0.0	0.0	0.0	2	0.0	0.0	2	0.0	0.0
Urinary Tract Infections	291	46.7	2.9	126	73.8	3.9	10.3	3.4	49	26.5	6.3	237	5.5	1.5
West Nile Virus (WNV)	1	100.0	0.0	1	0.0	0.0	0.0	0.0	0	-	-	1	0.0	0.0
Total	2,978	61.0	0.9	1,580	65.2	1.2	12.2	0.9	639	30.0	1.8	2351	8.2	0.6

type. The classifications are presented in Table 3.12. We then compute the PoSs using these classifications.

Looking at the vaccine PoSs by transmission route (see Table 3.13), we see that vaccines

Table 3.12: List of transmission routes and biological family for the infectious diseases.

Disease	Transmission Route	Biological Family
Hepatitis B virus (HBV)	Human-human (Others)	Hepadnavirus
Other	Multiple or others	Multiple or others
Otitis Media	Multiple or others	Multiple or others
Bacterial Skin Infection	Multiple or others	Multiple or others
Sepsis	Multiple or others	Multiple or others
Human papillomavirus (HPV)	Human-human (Others)	Flaviviridae
Human immunodeficiency virus (HIV)	Human-human (Others)	Retroviridae
Intra-abdominal Infections	Multiple or others	Multiple or others
Onychomycosis	Multiple or others	Multiple or others
Clostridium difficile	Multiple or others	Clostridiaceae
Cytomegalovirus (CMV) Infection	Human-human (Others)	Herpesviridae
Hepatitis C virus (HCV)	Human-human (Others)	Flaviviridae
Respiratory Infections	Multiple or others	Multiple or others
Urinary Tract Infections	Multiple or others	Multiple or others
Rotavirus	Human-human (Others)	Reoviridae
Ebola	Human-human (Others)	Filoviridae
Marburg	Human-human (Others)	Filoviridae
Smallpox	Human-human (Airborne)	Poxvirus
Zika	Animal bites	Flaviviridae
Rabies	Animal bites	Rhabdoviridae
Yellow Fever	Animal bites	Flaviviridae
Chikungunya	Animal bites	Togaviridae
Norovirus	Contaminated food or water	Caliciviridae
Japanese Encephalitis	Animal bites	Flaviviridae
Non-tuberculous Mycobacteria (NTM) Infection	Multiple or others	Multiple or others
West Nile Virus (WNV)	Animal bites	Flaviviridae
Middle East Respiratory Syndrome (MERS)	Human-human (Airborne)	Coronaviridae
Severe Acute Respiratory Syndrome (SARS)	Human-human (Airborne)	Coronaviridae
Monkeypox	Animal bites	Poxviridae

for diseases transmitted through animal bites have the highest overall PoS (61.3%, SE=4.7%), whereas no vaccine has been developed for diseases transmitted through contaminated food or water.

Table 3.13: The probabilities of success (PoSs) of industry-sponsored vaccine development programs, grouped by the transmission route.

	Phase 1			Phase 2					Phase 3			Overall		
Transmission route	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	Paths	PoS_{1A}	S.E.
Animal bites	125	89.6	2.7	103	78.6	4.0	63.1	5.0	71	91.5	3.3	106	61.3	4.7
Contaminated food or water	6	100.0	0.0	5	0.0	0.0	0.0	0.0	0	-	-	5	0.0	0.0
Human-human (Others)	643	82.4	1.5	446	62.8	2.3	39.7	2.4	238	74.4	2.8	517	34.2	2.1
Human-human (Airborne)	16	68.8	11.6	8	62.5	17.1	50.0	17.7	5	80.0	17.9	13	30.8	12.8
Multiple or others	1,048	81.9	1.2	777	65.6	1.7	47.5	1.9	454	81.3	1.8	911	40.5	1.6
Total	1,838	82.5	0.9	1,339	65.4	1.3	45.9	1.4	768	80.1	1.4	1552	39.6	1.2

We find that companies have been most successful in developing nonvaccine treatments for diseases transmitted between humans through the air, with 50.0% (SE=25.0%) of all drug development programs making it from phase 1 to regulatory approval (see Table 3.14). Unfortunately, this is based on only four drug development programs and may not be indicative of the general trend. Treatments for diseases that transmit through 'human to human (others)' have an overall PoS of 21.5% (SE=1.2%) while no approval is observed for diseases transmitted through 'animal bites' or 'contaminated food or water'.

Table 3.14: The probabilities of success (PoSs) of industry-sponsored, nonvaccine anti-infective drug development programs, grouped by the transmission route.

	Phase 1			Phase 2					Phase 3			Overall		
Transmission Route	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	Paths	PoS_{1A}	S.E.
Animal bites	10	40.0	15.5	1	0.0	0.0	0.0	0.0	0	-	-	7	0.0	0.0
Contaminated food or water	1	0.0	0.0	0	-	-	-	-	0	-	-	1	0.0	0.0
Human-human (Others)	1,471	68.9	1.2	859	56.2	1.7	31.0	1.7	401	66.3	2.4	1235	21.5	1.2
Human-human (Airborne)	4	75.0	21.7	3	66.7	27.2	66.7	27.2	2	100.0	0.0	4	50.0	25.0
Multiple or others	2,365	62.7	1.0	1,339	69.5	1.3	18.1	1.2	595	40.7	2.0	1886	12.8	0.8
Total	3,851	65.0	0.8	2,202	64.3	1.0	23.2	1.0	998	51.1	1.6	3133	16.3	0.7

When we classify the vaccines by the biological family of the infectious agent (Table 3.15), we see that reoviridae (e.g., rotavirus), rhabdoviridae (e.g., rabies), and hepadnaviridae (e.g., HBV) are the three biological families with the highest overall PoSs for vaccines at 78.7%, (SE=5.2%), 66.7% (SE=7.5%), and 53.7% (SE=3.6%), respectively. We have yet to see a vaccine for diseases caused by agents in the biological families of retroviridae (e.g., HIV), caliciviridae (e.g., norovirus), clostridiaceae (e.g., clostridium difficile), coronaviridae (e.g., SARS, MERS), herpesviridae (e.g., CMV infection), or togaviridae (e.g., chikungunya).

When we consider nonvaccine PoSs by biological family of the infectious agent (see Table 3.16), we see that nonvaccine therapies for poxviridae (e.g., smallpox), herpesviridae (e.g., CMV infection), and hepadnaviridae (e.g., HBV) have the highest overall PoS at 66.7% (SE=27.2%), 31.8% (SE=7.0%), and 29.5% (SE=4.0%), respectively. For viruses in the reoviridae (e.g., rotavirus), coronaviridae (e.g., SARS, MERS), caliciviridae (e.g., norovirus), rhabdoviridae (e.g., rabies), and togaviridae (e.g., chikungunya) families, there have been less than five development programs each, and no approved treatment.

Table 3.15: The probabilities of success (PoSs) of industry-sponsored vaccine development programs, grouped by the biological family.

	Phase 1			Phase 2					Phase 3			Overall		
Biological Family	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	Paths	PoS_{1A}	S.E.
Caliciviridae	6	100.0	0.0	5	0.0	0.0	0.0	0.0	0	-	-	5	0.0	0.0
Clostridiaceae	6	100.0	0.0	6	33.3	19.2	0.0	0.0	0	-	-	4	0.0	0.0
Coronaviridae	5	40.0	21.9	0	-	-	-	-	0	-	-	3	0.0	0.0
Filoviridae	16	43.8	12.4	7	57.1	18.7	28.6	20.2	2	100.0	0.0	14	14.3	9.4
Flaviviridae	218	86.2	2.3	146	55.5	4.1	43.2	4.3	70	90.0	3.6	165	38.2	3.8
Hepadnaviridae	213	94.8	1.5	187	74.9	3.2	54.5	3.7	132	77.3	3.6	190	53.7	3.6
Herpesviridae	14	57.1	13.2	3	33.3	27.2	0.0	0.0	0	-	-	8	0.0	0.0
Multiple or others	1,042	81.8	1.2	771	65.9	1.7	47.9	1.9	454	81.3	1.8	907	40.7	1.6
Poxviridae	12	83.3	10.8	9	66.7	15.7	55.6	16.6	6	83.3	15.2	11	45.5	15.0
Reoviridae	72	97.2	1.9	70	91.4	3.3	68.6	6.0	53	90.6	4.0	61	78.7	5.2
Retroviridae	181	65.2	3.5	95	36.8	4.9	0.0	0.0	21	0.0	0.0	144	0.0	0.0
Rhabdoviridae	47	91.5	4.1	40	87.5	5.2	65.0	8.1	30	86.7	6.2	39	66.7	7.5
Togaviridae	6	83.3	15.2	0	-	-	-	-	0	-	-	1	0.0	0.0
Total	1,838	82.5	0.9	1,339	65.4	1.3	45.9	1.4	768	80.1	1.4	1552	39.6	1.2

Table 3.16: The probabilities of success (PoSs) of industry-sponsored, nonvaccine anti-infective drug development programs, grouped by the biological family.

	Phase 1			Phase 2					Phase 3			Overall		
Biological Family	Paths	PoS_{12}	S.E.	Paths	PoS_{23}	S.E.	PoS_{2A}	S.E.	Paths	PoS_{3A}	S.E.	Paths	PoS_{1A}	S.E.
Caliciviridae	1	0.0	0.0	0	-	-	-	-	0	-	-	1	0.0	0.0
Clostridiaceae	91	83.5	3.9	66	53.0	6.1	4.5	2.8	25	12.0	6.5	71	4.2	2.4
Coronaviridae	2	50.0	35.4	1	0.0	0.0	0.0	0.0	0	-	-	2	0.0	0.0
Filoviridae	10	20.0	12.6	1	0.0	0.0	0.0	0.0	0	-	-	9	0.0	0.0
Flaviviridae	609	70.3	1.9	383	49.6	2.6	22.5	2.2	161	53.4	3.9	535	16.1	1.6
Hepadnaviridae	186	77.4	3.1	105	68.6	4.5	36.2	5.2	54	70.4	6.2	129	29.5	4.0
Herpesviridae	64	87.5	4.1	43	60.5	7.5	32.6	7.8	19	73.7	10.1	44	31.8	7.0
Poxviridae	3	66.7	27.2	2	100.0	0.0	100.0	0.0	2	100.0	0.0	1815	66.7	27.2
Reoviridae	2	100.0	0.0	2	0.0	0.0	0.0	0.0	0	-	-	3	0.0	0.0
Retroviridae	604	63.2	2.0	326	59.8	2.7	39.3	2.8	167	76.6	3.3	2	24.6	1.9
Rhabdoviridae	4	75.0	21.7	0	-	-	-	-	0	-	-	520	0.0	0.0
Togaviridae	1	0.0	0.0	0	-	-	-	-	0	-	-	1	0.0	0.0
Multiple or others	2,274	61.9	1.0	1,273	70.3	1.3	18.8	1.3	570	41.9	2.1	1	13.2	0.8
Total	3,851	65.0	0.8	2,202	64.3	1.0	23.2	1.0	998	51.1	1.6	3133	16.3	0.7

3.4.3 Discussion

Companies producing vaccines and other therapeutics for infectious diseases have gradually been retreating from these spaces in recent years. The number of companies producing vaccines has dwindled over the past few decades, and the top four vaccine companies now make up more than 90% of the global market [1]. Similarly, the top four companies producing antiviral drugs occupy about 80% of the global market [1]. Antibiotic developers such Achaogen and Melinta Therapeutics have filed for bankruptcy in the past year, while large pharmaceutical companies such as Novartis and Sanofi have withdrawn from the space [14],

leading the Infectious Diseases Society of America to sound the alarm about the availability of effective antibiotics [17].

It should be no surprise that investors are unwilling to invest in companies producing vaccines and treatments for infectious diseases given the economics of this market [26]. These have been generally regarded as low-margin products, and they have low expected growth potential compared to chronic treatments in other therapeutic areas, such as oncology or cardiovascular diseases. For example, Merck's oncology assets are estimated to have contributed $11.8 billion in incremental revenues from 2017 to 2020; for the same period, the incremental contribution of its vaccines portfolio is estimated to be $2.7 billion [25]. And Merck is the second largest vaccine maker in the world. This lack of investment has resulted in a relatively low number of development programs for vaccines and treatments of infectious diseases; only about 9.5% (5,869 programs out of about 59,891; see Table 3.2) of all industry-sponsored drug development programs launched in the past two decades are in these areas.

Our study indicates that the technical success rate is unlikely to be a barrier to investments in new vaccines and treatments for infectious diseases, unlike cancer drugs, where the financial risk of new RD projects comes from the reduced chance of bringing a drug-indication pair from phase 1 to market. The overall PoS of industry-sponsored vaccines and treatments for infectious diseases are above the average for all therapeutic groups (see Table 3.2).

It is often suggested that the fundamental issue behind this lack of investment is that the market for vaccines and treatments for infectious diseases is simply not lucrative enough. Despite the expense of research and development and the need for large-scale production [27], anti-infective disease treatments are used only occasionally, while vaccine companies face an avalanche of liability lawsuits [12]. Furthermore, the companies are at the mercy of government pricing decisions [13].

Apart from financial considerations, the dearth of vaccines and other treatments for infectious diseases may be due to the lack of available subjects for testing these therapeutics, especially during non-epidemic periods. This may be alleviated by having faster preclinical and clinical pathways in cases of severe infectious diseases with no existing treatments. One such pathway is the Animal Rule ([9, 10] whereby the "FDA may grant marketing approval

based on adequate and well-controlled animal efficacy studies when the results of those studies establish that the drug is reasonably likely to produce clinical benefit in humans" [8]. This has been used to approve smallpox and monkeypox vaccines, and can be expanded for the investigation of therapeutics for other potentially deadly infectious diseases with low incidence rates, such as SARS.

Even though this pathway can expedite the development of vaccines and anti-infective treatments, it still requires considerable development time as one needs to establish the equivalence of the drug mechanism between animal models and humans. While it is desirable to hasten the development of vaccines and medical product during an epidemic, biological breakthroughs and science will ultimate drive the efficiency of our ability to fight pandemics of novel pathogens.

It remains to be seen if more non-industry-sponsored research can alleviate the issue. Our study shows that only 6.8% (SE=1.0%) and 8.2% (SE=0.6%) of non-industry-sponsored vaccines and nonvaccine infectious disease development programs transition from phase 1 to approval, respectively. However, this may be a result of selection bias: promising vaccine and therapeutics initiated in non-industry settings are often pursued in conjunction with industry-sponsored sponsors, whereas commercially less promising projects are more likely to be pursued by non-profit organizations.

Section References

[1] Evaluatepharma world preview 2018, outlook to 2024, 2018.

[2] Rosa M Abrantes-Metz, Christopher Adams, and Albert D Metz. Pharmaceutical development phases: a duration analysis. *Journal of Pharmaceutical Finance, Economics and Policy*, 14:19–42, 2005.

[3] FE Andre. How the research-based industry approaches vaccine development and establishes priorities. *Developments in biologicals*, 110:25, 2002.

[4] Matthew M Davis, Amy T Butchart, John RC Wheeler, Margaret S Coleman, Dianne C Singer, and Gary L Freed. Failure-to-success ratios, transition probabilities and phase lengths for prophylactic vaccines versus other pharmaceuticals in the development pipeline. *Vaccine*, 29(51):9414–9416, 2011.

[5] Joseph A DiMasi, Lanna Feldman, Abraham Seckler, and Andrew Wilson. Trends in risks associated with new drug development: success rates for investigational drugs. *Clinical Pharmacology & Therapeutics*, 87(3):272–277, 2010.

[6] Joseph A DiMasi, Maria I Florez, Stella Stergiopoulos, Yaritza Peña, Zachary Smith, Michael Wilkinson, and Kenneth A Getz. Development times and approval success rates for drugs to treat infectious diseases. *Clinical Pharmacology & Therapeutics*, 107 (2):324–332, 2020.

[7] Jenifer Ehreth. The global value of vaccination. *Vaccine*, 21(7-8):596–600, 2003.

[8] Food and Drug Administration. Guidance for industry product development animal rule, October 2015. URL `https://www.fda.gov/files/drugs/published/Product-Development-Under-the-Animal-Rule.pdf`. Accessed: Apr 31 2020.

[9] Food and Drug Administration. Approval of new drugs when human efficacy studies are not ethical or feasible, 21 c.f.r. §§600–650, 2019.

[10] Food and Drug Administration. Approval of biological products when human efficacy studies are not ethical or feasible, 21 c.f.r. §§90–95, 2019.

[11] Michael Hay, David W Thomas, John L Craighead, Celia Economides, and Jesse Rosenthal. Clinical development success rates for investigational drugs. *Nature Biotechnology*, 32(1):40–51, 2014.

[12] Scott Hensley and Bernard Wysocki Jr. As industry profits elsewhere, u.s. lacks vaccines, antibiotics, November 2005. URL `https://www.wsj.com/articles/SB113141787830190837`.

[13] Charlotte Hu. Pharmaceutical companies are backing away from a growing threat that could kill 10 million people a year by 2050, July 2018. URL `https://www.businessinsider.com/major-pharmaceutical-companies-dropping-antibiotic-projects-superbugs-2018-7`.

[14] Andrew Jacobs. Crisis looms in antibiotics as drug makers go bankrupt. *The New York Times*, 2019.

[15] Marie Paule Kieny and Marc P Girard. Human vaccine research and development: an overview. *Vaccine*, 23(50):5705–5707, 2005.

[16] R Langreth. Antibiotics aren't profitable enough for big pharma to make more. *Bloomberg Businessweek*, 2019.

[17] Infectious Diseases Society of America et al. Correction: Antibiotic company starts bankruptcy proceedings; highlights urgent need for investment in infection fighting drugs, 2019.

[18] US Department of Health, Human Services, et al. Encouraging vaccine innovation: Promoting the development of vaccines that minimize the burden of infectious diseases in the 21st century: Report to congress, 2017.

[19] World Health Organization. *The global eradication of smallpox: final report of the Global Commission for the Certification of Smallpox Eradication, Geneva, December 1979*. World Health Organization, 1980.

[20] Esther S Pronker, Tamar C Weenen, Harry Commandeur, Eric HJHM Claassen, and Albertus DME Osterhaus. Risk in vaccine research and development quantified. *PloS one*, 8(3):e57755, 2013.

[21] OECD. Publishing, Organization for Economic Cooperation, and Development (OECD) Staff. *Health at a glance 2013: OECD Indicators*. OECD Publishing, 2013.

[22] Vanessa Rémy, York Zöllner, and Ulrike Heckmann. Vaccination: the cornerstone of an efficient healthcare system. *Journal of market access & health policy*, 3(1):27041, 2015.

[23] Katarzyna Smietana, Marcin Siatkowski, and Martin Møller. Trends in clinical success rates. *Nature Reviews Drug Discovery*, 15:379–380, 2016.

[24] David W Thomas, Justin Burns, John Audette, Adam Carroll, Corey Dow-Hygelund, and Michael Hay. Clinical development success rates 2006–2015. *BIO Industry Analysis*, 1:16, 2016.

[25] Trefis. Mrk revenues: How does merck make money? URL `https://dashboards.trefis.com/no-login-required/qVh09Fjc/MRK-Revenues-How-Does-Merck-Make-Money-`.

[26] Jonathan T Vu, Benjamin K Kaplan, Shomesh Chaudhuri, Monique K Mansoura, and Andrew W Lo. Financing vaccines for global health security. Technical report, National Bureau of Economic Research, 2020.

[27] Jerry P Weir and Marion F Gruber. An overview of the regulation of influenza vaccines in the united states. *Influenza and other respiratory viruses*, 10(5):354–360, 2016.

[28] Chi Heem Wong, Kien Wei Siah, and Andrew W. Lo. Estimating clinical trial success rates and related parameters in oncology. *Available at SSRN 3355022*, 2019.

[29] Chi Heem Wong, Kien Wei Siah, and Andrew W. Lo. Estimation of clinical trial success rates and related parameters. *Biostatistics*, 20(2):273–286, 2019.

[30] World Health Organization. Vaccines and diseases, 2020. URL `https://www.who.int/immunization/diseases/en/`.

[31] Ruth Young, Tewodros Bekele, Alexander Gunn, Nick Chapman, Vipul Chowdhary, Kelsey Corrigan, Lindsay Dahora, Sebastián Martinez, Sallie Permar, Johan Persson, et al. Developing new health technologies for neglected diseases: a pipeline portfolio review and cost model. *Gates open research*, 2, 2018.

Part II

Analytics for Healthcare Economics

Chapter 4

Estimating the Financial Impact of Gene Therapy

4.1 Introduction

Gene therapy is a new class of medical treatment that alters part of a patient's genome through the replacement, deletion, or insertion of genetic material to treat a disease. While still in its infancy, gene therapy has demonstrated immense potential to treat and even cure previously intractable diseases. The introduction of voretigene neparvovec (marketed as Luxturna®) for inherited retinal disease and onasemnogene abeparvovec-xioi (marketed as Zolgensma®) for spinal muscular atrophy (SMA) in the U.S. have improved the lives of patients [114, 124]. Yet the price per treatment of $425,000 per eye for Luxturna, and $2.1 million per patient for Zolgensma, have raised concerns regarding affordability among budget-constrained payers and patients alike.

Stakeholders have expressed concerns that gene therapy will be too expensive for individual patients to afford on their own, especially if they continue to be priced at more than 30 times the median household income of $61,937 [103], (equivalently, at several multiples of the average U.S. home mortgage of $354,400 [126]). Insurance coverage for gene therapy also varies by policy type. Many insurance policies do not cover access to gene therapy, or they impose very restrictive policies to limit the number of patients who might be treated [78, 156]. Widespread underinsurance in the U.S. [116]—requiring substantial

out-of-pocket costs in the form of deductibles and coinsurance payments—may place gene therapy out of reach for American patients who might benefit from treatment. Many health plans, especially those facing fixed annual budgets, including state Medicaid policies, some employer-offered plans, and insurance offered on the federal and state-sponsored exchanges, have warned they may not be able or willing to absorb the additional spending should a greater number of people become eligible for expensive gene therapy once it reaches the market [159]. While some spending on novel gene therapies would likely be paid for by Medicare, the taxpayer-supported health insurance for Americans over the age of 65, other spending on these treatments might have an impact on the wages that private corporations pay to their employees [70].

In this paper, we estimate the potential fiscal impact of gene therapy on the U.S. market. To do so, we create a new model to estimate the future number of gene therapy approvals, the size of their potential patient populations, and the prices of these future treatments. We begin by surveying the clinical trial databases for late-stage gene therapy trials, defined here as phase 2/3 or phase 3, and compile the prevalence and incidence of the diseases targeted in these trials from a meta-analysis of published sources. We develop a novel method to estimate the price of each gene therapy under consideration by calculating the expected quality-adjusted life years gained for each therapy in the relevant patient population. Combining these results and previously published probabilities of technical success by therapeutic area, we simulate whether a disease will have an approved gene therapy over the next fifteen years, the expected number of treated patients and the expected spending from January 2020 to December 2034.

4.2 Simulation Design

A critical element in our simulation analysis is the number of gene therapies that will receive regulatory approval over the next few years. Therefore, we begin with a brief review of the approval process in the U.S.

Since the passing of the Food, Drug, and Cosmetic Act in 1938, pharmaceuticals developed by companies have to be reviewed by the Food and Drug Administration (FDA) for

safety and efficacy before they can be marketed in the U.S. The application for marketing approval differs slightly by the type of therapy: New Drug Applications (NDAs) are for small molecules, and Biologics License Applications (BLAs) are for biologics. Gene therapy is considered a biologic product, hence the BLA designation applies.

Clinical investigations in human subjects typically take place in three phases—phases 1, 2 and 3—before marketing approvals are sought. Phase 1 trials are designed to investigate the dosage and safety of the treatment, while phase 2 trials attempt to detect early signs of efficacy and possible side effects in a relatively small sample of patients. Phase 3 trials are intended to demonstrate a statistically significant treatment effect when compared to the best standard of care in a broader population of patients. Some clinical trials combine multiple phases into a single design, with the phase numbers separated by a slash. For example, a phase 2/3 trial combines elements of phase 2 and phase 3 investigations into a single trial design in order to reduce the overall development time and cost, and maximize the participation of subjects with orphan disease willing to participate in trials. The clinical development of therapeutics is a tedious and costly process that may span decades and cost billions of dollars, with the bulk of the cost and time spent conducting phase 3 clinical trials [63, 84]. The process is also very risky, with only 13.8% of therapeutic development programs entering phase 1 reaching approval [166].

To estimate the financial impact of gene therapies on the U.S. healthcare system, we first identify all existing late-stage clinical trials of gene therapies, simulate their successes or failures from phase 2/3 or 3 to approval, then estimate the spending on the successful ones by summing the product of their expected prices and number of patients, as outlined in Figure 4-1. By using simulation analysis rather than purely deterministic methods, we are able to capture the inherent uncertainty in costs, revenues, and other parameters of this new therapeutic class.

We organize the simulation into the following five distinct modules, and describe each of these in some detail: (1) identifying the number of gene therapies currently in clinical trials; (2) estimating the probabilities of success of these trials; (3) estimating the time to approval; (4) simulating the expected number of patients treated by these therapies if approved; and (5) estimating the expected market prices of the approved therapies. We

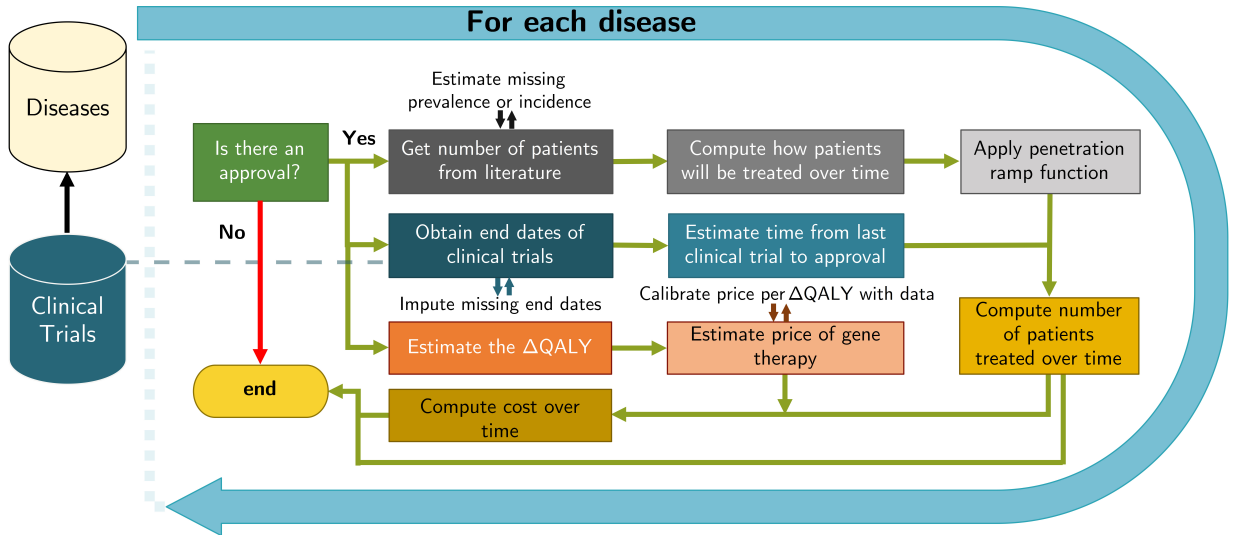


Figure 4-1: A flowchart showing the performance of the simulation. After extracting the information on each disease from the clinical trial databases, we simulate whether the disease will obtain an approval. If it fails to do so, the simulation will end for this disease in this iteration. Otherwise, we will estimate the expected number of patients to be treated, compute the corresponding cost of treatment, and store the results. At each step of the computation, we source data from literature and impute missing information.

describe the first four modules in Sections 4.2.1–4.2.4. Given the importance and potential controversies surrounding the pricing of gene therapies, we devote a standalone section to this issue in Section 4.3.

4.2.1 Clinical Trial Data

We use clinical trial metadata from the Citeline TrialTrove database and the U.S. National Library of Medicine's ClinicalTrials.gov database to determine the number of gene therapies currently under development and their potential number of patients.

We download data from the Citeline database, isolating trials tagged with 'gene therapy' under the 'therapeutic class' field. We supplement this information by searching for trials on the *clinicaltrials.gov* main page using the key words 'gene therapy', then reading the trial description to determine if the trial is in fact related to a gene therapy. All database queries were made before May 31, 2019. Clinical trials from both sources are merged before filtering for clinical trials that are in either phase 2/3 or phase 3 of the development process and

are not known to be compassionate uses of the treatment.[1] The outcomes of compassionate use are rarely used as data points in the clinical development process. Even though adverse events from compassionate use are reported to the FDA and may, in rare cases, be used to characterize the risk and benefits of a therapy, the FDA is cognizant that these uses often occur outside of clinical trial settings, and has almost never given an unfavorable decision to a product labeling because of an adverse outcome of compassionate use [88, 115]. We include clinical trials without U.S. trial site in our dataset because it is currently possible for the FDA, as empowered by Federal administrative law 21 CFR Part 312.120, to grant marketing approval using evidence from foreign clinical trials [68].

Our filtering criteria are intended to remove trial entries unrelated to the clinical development process, and to isolate gene therapies that are most likely to seek regulatory approval in the U.S. in the near future.

We remove repeated entries, and identify the diseases and therapeutic areas targeted by each gene therapy. Each clinical trial entry in our dataset contains a brief title of the trial, its clinical phase, the disease being targeted, the start and end dates of the clinical trial, the therapy name, and the companies involved in the clinical trial.

This process yields 109 trials investigating 57 distinct diseases, listed in Table 4.11 in Supplementary Materials. We classify the diseases into three categories: cancer (oncology),[2] rare disease, and general disease. The distribution of disease and the clinical trials by category and therapeutic area are shown in Table 4.1 and Table 4.2. The majority of trials and diseases are in the area of oncology, followed by rare diseases. These therapeutic areas are notoriously risky. Only 3.1% and 6.2% of the drug development programs in oncology and rare diseases go from Phase 1 to approval, respectively, compared to the baseline of 13.8% across all drugs and indications [166].

[1]Compassionate use, also known as 'expanded access', refers to the administration of investigational treatments outside of the clinical trial to treat patients with serious or immediately life-threatening diseases, or conditions when there are no comparable or satisfactory alternative treatment options.

[2]We classify Ewing's Sarcoma—a rare form of cancer—as a rare disease instead of cancer.

Table 4.1: Count of number of clinical trials by category and therapeutic area.

Therapeutic Area	Cancer	General	Rare Disease	Subtotal
Autoimmune/Inflammation		3	2	5
Cardiovascular	-	15	1	16
CNS	-	3	7	10
Metabolic/Endocrinology	-	3	15	18
Oncology	52	-	1	53
Ophthalmology	-	-	7	7
Subtotal	52	24	33	109

Table 4.2: Count of number of diseases by category and therapeutic area.

Therapeutic Area	Cancer	General	Rare Disease	Subtotal
Autoimmune/Inflammation	-	2	1	3
Cardiovascular	-	6	1	7
CNS	-	1	4	5
Metabolic/Endocrinology	-	3	6	9
Oncology	28	-	1	29
Ophthalmology	-	-	4	4
Subtotal	28	12	17	57

4.2.2 Probability of Success Estimates

We define a gene therapy development program as the set of clinical trials by a sponsor testing a therapeutic for a disease. We consider whether a gene therapy will be developed for a disease by simulating correlated coin flips' for each gene therapy program, and observing if there is at least one approval.

Our computational method assumes that clinical trials are always perfectly correlated within the same development program. This is logical, since the FDA requires findings from at least two pivotal trials in the BLA review process[171][3]. Our assumption can also be justified for clinical trials run by the same sponsor that target the same disease, but for different patient segments. We reason that the sponsors are risking multiple expensive late-stage trials for the same disease, thus have confidence that the treatment will work on all patient sub-populations, and therefore any marketing licensing approval (or denial) will be

[3]It must be noted that several pathways, such as the priority review program, are exceptions to this rule. The assumption of perfect correlation still holds if only one trial is required for regulatory review.

Table 4.3: The probability of success of drug development programs from phase 3 to approval (PoS_{3A}), categorized by therapeutic area. We assume that the probability of success for gene therapy follows a similar pattern.

Therapeutic Area	PoS_{3A} (%)
Autoimmune/Inflammation	48.5
Cardiovascular	50.1
Central Nervous System (CNS)	37.0
Metabolic/Endocrinology	45.7
Oncology	28.5
Ophthalmology	45.9

similar for all the patient segments.

It can be argued that different gene therapy treatments for a disease are highly correlated, since they operate on similar platforms (e.g. CAR-T or in-vivo gene delivery using adeno-associated virus vectors), even though different gene sequences may be targeted. To reflect this, we consider a correlation of 90% between development programs in our simulations. Our sensitivity analysis, however, demonstrates that these computations are insensitive to this parameter (see Section 4.4.3).

The phase 3 to approval probability of success (PoS_{3A}) for each disease is informed by prior studies on the probabilities of success of drug development programs by therapeutic area from the MIT Laboratory of Financial Engineering's Project ALPHA website [96]. These estimates for the probabilities of success are derived from over 55,000 drug development programs between January 2000 and January 2020, and computed using the path-by-path methods as introduced in Wong et al. [166]. The PoS_{3A} values used in our simulations are given in Table 4.3 and the mapping of diseases to therapeutic areas is shown in Table 4.12.

4.2.3 Time to Approval

We also require an estimate of the time to approval for gene therapy treatments in order to assess the patient impact and cost over time. Typically, companies submit their Biologics License Application (BLA) to the U.S. Food and Drug Administration (FDA) some time after the end of the clinical trial period. We assume that the time between the end of the

last clinical trial for the disease and the submission of the BLA is a variable drawn from a triangular distribution between 0 and 365 days, with a median of 182.5 days. This is informed by the practical knowledge that it takes an average of 6 months to prepare the documents for the BLA submission [166].

In addition, there will be a lag time between the submission of the BLA and the decision of the FDA. The FDA has 60 days to decide if it will follow up on a BLA filing [91], and it can take another 10 months to deliver its verdict [90]. This implies the maximum possible time between BLA submission and FDA approval will be 12 months. We thus assume that the time between the BLA submission and the FDA decision is drawn from a triangular distribution between 0 and 365 days, with a median of 182.5 days. Our assumptions are also valid for therapies that use the priority review pathways.

We also assume that the BLA will be filed only after the last clinical trial for a disease has ended. Trials with missing declared end dates will have their end dates imputed by adding random durations to the trial start date, drawn from a gamma distribution fitted to clinical trials with complete date information in our data (see Figure 4-2).

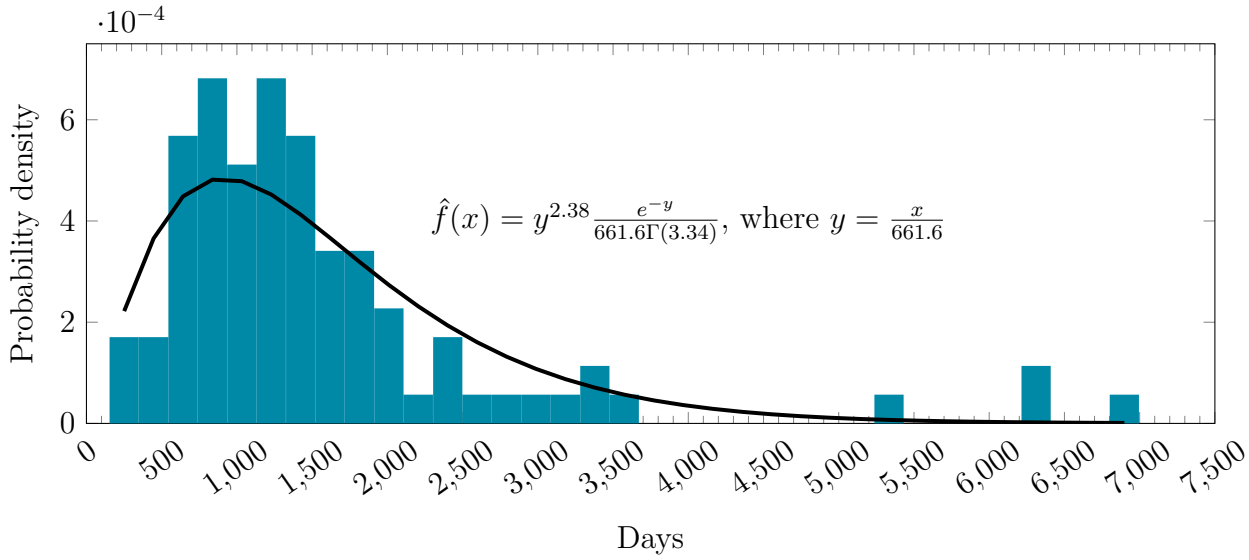


Figure 4-2: The empirical distribution of duration against our fitted gamma distribution.

Diseases with a prior approved therapy will automatically be considered to be approved as of January 1st 2020. For some diseases, the last clinical trial will have ended more than three years ago (i.e., before January 2017). For diseases that match this criterion, we treat

them as though they have failed.

4.2.4 Number of Patients

The second module simulates the number of new and existing patients that will be treated over time, conditioned on the disease receiving an approved gene therapy.

We consider only the superset of the patient segments listed in the clinical trials for each disease. For example, if there are two clinical trials, one targeting 'patients above the age of 40' and the other targeting 'patients above the age of 18', we only consider the latter when estimating the patient population for the disease. If insufficient information about the sub-population is given, we assume that all the patients with that disease are eligible. The proportion of patients who are eligible for treatment and are willing to do so will be taken into account later, as we will explain in Section 4.2.4.

Incidence and Prevalence

For the number of currently affected patients and the number of new patients per year for each indication, we source medical journals and online data repositories, such as the Surveillance, Epidemiology, and End Results (SEER) website and *cancer.net*. If we are able to find an estimated patient population, we cite it directly. Otherwise, we multiply the prevalence and incidence rates by the population of the U.S., which we take to be 327.7 million [74]. When necessary, we also make the assumption that the female to male ratio is 1:1.

In cases in which we are able to find estimates for the disease incidence but not the prevalence, we combine the incidence of the disease (i.e., i new patients a year) and the disease survival rate (i.e., $p\%$ of the people with a disease will be alive after k years) to obtain the steady-state estimate of the prevalence (j) using Equation Equation 4.1. The incidence can also be estimate from the prevalence by rearranging Equation Equation 4.1 to yield Equation 4.2.

$$\text{Prevalence } (j) = \frac{ki}{1-p} \tag{4.1}$$

$$\text{Incidence } (i) = \frac{j(1-p)}{k} \tag{4.2}$$

The equations can be derived by assuming that the number of patients will be constant through the years at a level j. Since ki new patients are added over k years and $j(1-p)$ patients that will die over the same period, $ki = j(1-p)$ for the number of patients to be constant over time. Rearranging this equation will yield Equation 4.1 and Equation 4.2. The number of patients for each disease are presented in Table 4.13 in the Supplementary Materials. We adjust these estimates to avoid double-counting in cases of overlapping patient populations, e.g., the number of patients for 'Spinal Muscular Atrophy' is the difference between 'Spinal Muscular Atrophy' and 'Spinal Muscular Atrophy I' (a sub-category of the former).

Treatment of Patients over Time

In our simulation, we assume that newly diagnosed patients are treated immediately upon diagnosis. We further assume that the proportion of existing patients who seek treatment do so in such a way that the existing stock of patient declines exponentially, with a half-life of λ. Mathematically, the proportion of existing patients that seek treatment between time t and $t+\delta$ after approval is given by $E(t, \delta, \lambda)$, where:

$$E(t, \delta, \lambda) = e^{-\frac{t \ln 2}{\lambda}} - e^{-\frac{(t+\delta) \ln 2}{\lambda}}, t \geq 0 \tag{4.3}$$

In the face of limited information, we assume that 25% of the existing stock of patients will seek treatment in the first year of our simulation. This requires that the half-life be set to 28.91 months, which in turn implies that 95% of all patients who are diagnosed prior to the approval of the gene therapy want treatments within 10.5 years. Admittedly, this is just an assumption, and we perform a sensitivity analysis to determine its impact on our results in Section 4.4.3. Not everyone who seeks treatment will be given one; the effective number of patients treated is determined by the patient penetration rate, as we shall describe next.

Patient Penetration

It is unlikely that all the patients under consideration will receive gene therapy treatments. This may be due to ineligibility, lack of awareness of the treatment, or simply lack of interest in gene therapy, among many other reasons. We term the percentage of the patients that receive gene therapy treatments the 'patient penetration rate', and model it using a ramp function, $\rho(t, \Theta_{max}, T_{max})$. The ramp function is frequently used by the industry to model the rate of adoption of a product or technology [142], and is given by:

$$\rho(t, \Theta_{max}, T_{max}) = \begin{cases} \dfrac{t \cdot \Theta_{max}}{T_{max}}, 0 \leq t \leq T_{max} \\ \Theta_{max}, \text{ otherwise} \end{cases} \tag{4.4}$$

An illustration of the ramp function is given in Figure 4-3.

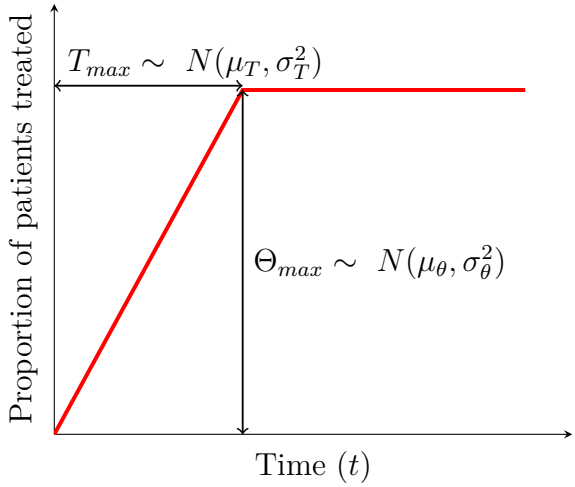


Figure 4-3: An illustration of the ramp function used to model the patient penetration rate over time.

Θ_{max} and T_{max} are assumed to follow Gaussian distributions $N(\mu_\theta, \sigma_\theta^2)$ and $N(\mu_T, \sigma_T^2)$, respectively. The parameter settings are listed in Tables 4.4 and 4.5. When setting μ_θ and μ_T, we need to take into the account the nature of the diseases. At one extreme, we have rare diseases, which are often life-threatening, and affect a relatively small number of people. Faced with these prospects of survival, more patients are willing to enroll in new treatments quickly after they are approved. In addition, since the number of patients is relatively small, insurers are more willing to cover these therapies and manufacturers are more able to cope

with a larger proportion of patients. Given this, we assign a high value of 40% for μ_θ, and a low value of 6 months for μ_T.

On the other hand, general diseases are seldom deadly, and affect a large number of patients, possibly even in the millions. Since an acceptable standard of care is available for these conditions, patients may be less inclined to use new treatments due to fear of the unfamiliar. In addition, it is often financially difficult for insurance companies to cover so many patients. We thus assume that the maximum penetration rate will be 1%, and the ramp-up period, 5 years.

As an intermediate case, cancers have characteristics that fall between these two extremes, but in general, they are more similar to rare diseases. We therefore assign values of 10% and 12 months to the maximum penetration rate and ramp-up period, respectively. All variances are set to 10% of the means to model moderate uncertainty in our numbers. They do not affect our mean estimates of the number of impacted patients or spending on gene therapy.

Classification	μ_θ	σ_θ^2
General	0.01	0.002
Rare Diseases	0.4	0.08
Cancer	0.1	0.02

Table 4.4: Parameter settings for $\Theta_{max} \sim \ N(\mu_\theta, \sigma_\theta^2)$.

Classification	μ_T	σ_T^2
General	60	6
Rare Diseases	6	0.6
Cancer	12	0.12

Table 4.5: Parameter settings for $T_{max} \sim \ N(\mu_T, \sigma_T^2)$. We consider the severity of the disease and the number of patients when making the assumptions.

The net number of patients to be treated for the disease at time t after the approval of a gene therapy is given by:

$$\texttt{Patients}_t = \rho(t, \theta_{max}, T_{max}) \cdot [\texttt{New patients}_t + E(t, \delta, \lambda) \cdot \texttt{Existing patients}_t] \quad (4.5)$$

We do not consider the effect of market competition among different therapies for the

same disease and patient groups on the number of treated patients. In our model, there is only one approval per disease, and a fraction of the eligible patients will receive that treatment.

4.3 Pricing

The cost to the healthcare system of providing the gene therapy for a disease for all patients being treated at time t after approval is given by $C(t)$, where

$$C(t) = \texttt{Patients}_t \times \texttt{Price of gene therapy} \tag{4.6}$$

The price of each treatment is crucial to computing the expected total spending, and a source of considerable controversy because of the high price of gene therapies relative to many conventional therapeutics. The Institute for Clinical and Economic Review (ICER)—an independent nonprofit organization that aims to evaluate the clinical and economic value of healthcare innovation—has advocated pricing drugs and gene therapies by the relative risk and benefit to the patient. This is typically done by comparing the quality-adjusted life years (QALY) with and without the treatment, then multiplying the change in QALY (ΔQALY) by a constant, typically set between \$50,000 and \$150,000 per ΔQALY [135].

$$\texttt{Price of gene therapy} = \frac{\texttt{Price}}{\Delta\text{QALY}} \times \Delta\text{QALY} \tag{4.7}$$

ICER has published reports containing its estimates of QALY gained by patients with vision loss associated with biallelic RPE65-mediated retinal disease following treatment with Luxturna® [124], and with SMA Type I following treatment with Zolgensma® [82]. These reports compute ΔQALY using the results of clinical trials to make informed estimates about the potential improvements in the quality of life and life expectancy of the patients.

While ICER's methods are considered by some stakeholders to be the gold standard for this type of calculation, replicating its methods for all the clinical trials under consideration is not feasible in this paper, given the fact that all the clinical trials in our analysis are still pending. As an alternative, we develop a mathematical model to estimate the expected

increase in QALY for each disease in our sample.

4.3.1 Estimating ΔQALY

We consider a representative patient who is expected to live to the age of x with a probability of $l(x)$. The function $l(x)$ is also known as the survival curve of the population. The patient enjoys a quality of life, $f(s, x)$, that is dependent on his age, x, and his state of health, s. The expected QALY of a typical person in the baseline state of s_0 (the 'healthy' state) can be computed by integrating $l(x)f(s_0, x)$ over x.

$$\text{Expected QALY (healthy)} = \int_0^\infty l(x)f(s_0, x)dx \tag{4.8}$$

Suppose that the patient is afflicted with a disease at time a, which changes his survival curve after time a from $l(x)$ to $\widetilde{l}(x)$. Likewise, his quality of life after diagnosis changes from $f(s_0, x)$ to $f(s_d, x)$. This patient will then have an expected QALY of:

$$\text{Expected QALY (unhealthy)} = \int_0^a l(x)f(s_0, x)dx + \int_a^\infty \widetilde{l}(x)f(s_d, x)dx \tag{4.9}$$

The change in the expected QALY due to the disease can then be expressed as:

$$\begin{aligned}
\Delta\text{QALY} &= \text{Expected QALY (unhealthy)} - \text{Expected QALY (healthy)} && (4.10)\\
&= \int_0^a l(x)f(s_0, x)dx + \int_a^\infty \widetilde{l}(x)f(s_d, x)dx - \int_0^\infty l(x)f(s_0, x)dx && (4.11)\\
&= \int_a^\infty \widetilde{l}(x)f(s_d, x) - l(x)f(s_0, x)dx && (4.12)\\
&\leq 0 && (4.13)
\end{aligned}$$

It is customary in the literature to incorporate time preferences into the model. This is done by multiplying the integrand by the discount factor, $r(x-a)$. There is a normalization term $l(a)$ to reflect conditional survival to age x.

$$\Delta\text{QALY} = \int_a^\infty \frac{r(x-a)}{l(a)} \left[\widetilde{l}(x)f(s_d, x) - l(x)f(s_0, x)\right] dx \tag{4.14}$$

If the distribution of age when the patient population contracts the disease is given by

$A(a)$, then the expected decrease in QALY over the patient population is given by:

$$\mathrm{E}(\Delta\mathrm{QALY}) = \int_0^\infty A(a) \int_a^\infty \frac{r(x-a)}{l(a)} \left[\widetilde{l}(x) f(s_d, x) - l(x) f(s_0, x)\right] dx da \tag{4.15}$$

Equation 4.15 is a general formula that accounts for the expected value of the changes in QALY between two states of health using only three variables: the time of disease onset, and the utility of the two health states. By making the relevant substitutions, we can also apply this formula to compute the expected changes in QALY given a gene therapy (gt) and an alternative treatment (alt).

$$\mathrm{E}(\Delta\mathrm{QALY}) = \int_0^\infty A(a) \int_a^\infty \frac{r(x-a)}{l(a)} \left[\widetilde{l}_{gt}(x) f(s_{gt}, x) - \widetilde{l}_{alt}(x) f(s_{alt}, x)\right] dx da \tag{4.16}$$

While death and patient statistics can be collected to determine $l(x)$ and $A(a)$ empirically, determining $f(s, x)$ and $\widetilde{l}(x)$ is challenging. Therefore, we use simple functions to modify these variables. In particular, we assume that being afflicted with a severe disease will modify the survival curve by a multiplicative factor, $D(t)$. That is, the survival curve of a patient after he is diagnosed at age a is given by:

$$\widetilde{l}(x) = l(x) \cdot D(x-a) \tag{4.17}$$

This functional form assumes that the disease is age-agnostic, and affects the survival curve only through the time elapsed since the patient has been diagnosed. For example, if the disease does not affect mortality (e.g. blindness), then $D(x-a) = 1$ for all $x - a > 0$. On the other hand, if the condition causes death immediately, then $D(x-a) = 0$ for all $x - a > 0$.

For the utility function, $f(s, x)$, we assume that it can be decomposed into two multiplicative factors, one dependent only on age, $f_a(x)$, and the other dependent only on the state of health, $f_h(s)$:

$$f(s, x) = f_a(x) \cdot f_h(s) \tag{4.18}$$

Assuming that Equations 4.17 and 4.18 hold, Equation 4.16 can be simplified to:

$$\mathrm{E}(\Delta\mathrm{QALY}) = \int_0^\infty A(a) \int_a^\infty \frac{l(x)}{l(a)} f_a(x) r(x-a) K(x,a) dx da \tag{4.19}$$

where $K(x,a)$ is the change in the quality-adjusted life years:

$$K(x,a) = D_{gt}(x-a) f_h(s_{gt}) - D_{s_{alt}}(x-a) f_h(s_{alt}) \tag{4.20}$$

4.3.2 Calibration of ΔQALY

For each of these variables, we attempt to obtain empirical values from the literature as much as possible. When necessary, we interpolate values, briefly explaining our assumptions and the data collection methods for the inputs to the model.

For the age-dependent QoL, $f_a(x)$, we extract the general population utility values from Institute for Clinical and Economic Review [114] and fit a linear model across the data. The QoL values and the fitted model are shown in Figure 4-4.

Age Group	Mean
18–29	0.922
30–39	0.901
40–49	0.871
50–59	0.842
60–69	0.823
70–79	0.790
≥80	0.736

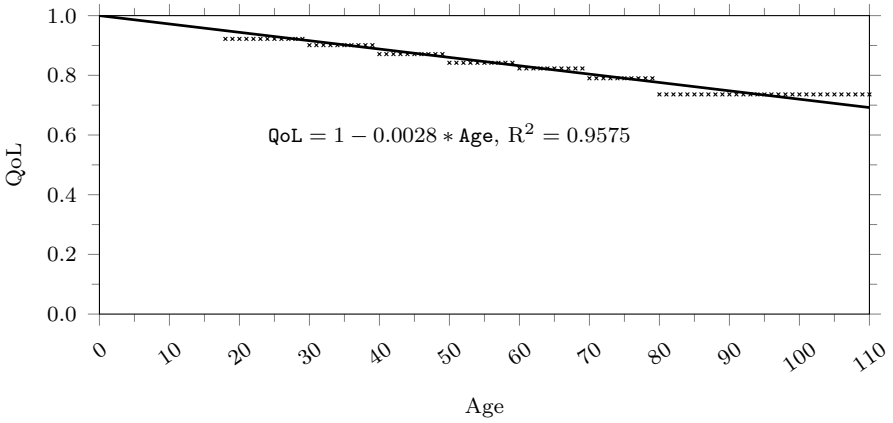


Figure 4-4: Age-dependent QoL, $f_a(x)$. The values extracted from ICER's SMA final report [114] are replicated in the table on the left and are presented as crosses in the figure on the right. A linear line, with its intercept set to 1, is fitted with the data points.

Since we are unable to know the patient outcomes for these potential gene therapies ahead of their approval, we assume that the gene therapy treatments will restore a person's survivability to that of a normal individual. This implies that $D_{gt}(x-a) = 1$. To estimate the impact of a disease on patient survivability, we model its survival curve, $D_{alt}(x-a)$, using

the exponential survival curve shown in Equation 4.21. In the equation, λ is the force of mortality, and μ is the normalization factor. We estimate λ and μ by matching the function to T-year survival rates, which are the proportions of the patients (k) who will be alive after T years, from data. The parameter values and their sources are listed in Table 4.14 in the Supplementary Material.

$$D_{alt}(x-a) = D_{alt}(t) = \lambda e^{-\lambda(t-\mu)}, \text{ where } \mu = \frac{1}{\lambda}\ln\frac{1}{\lambda} \;\&\; \lambda = -\frac{\ln k}{T} \;\&\; t = x - a \qquad (4.21)$$

The health-related quality of life variables, $f_h(s_{gt})$ & $f_h(s_{alt})$, are treated separately, depending on the disease classification. For cancerous indications, we assume that the quality of life of the patients is not affected by the disease. For non-cancerous indications, we source the medical literature for the available quality of life (QoL) estimates. We use the QoL for the typical disease condition to approximate the 'before treatment' QoL, $f_h(s_{alt})$, and use the best possible outcome for each condition as the 'post-treatment' QoL, $f_h(s_{gt})$. We interpolate the missing values using linear regressions of the sourced QoLs against disease severity. To do this, we first give scores, ζ, ranging from one to five for each disease, based on our perception of disease severity. We then fit a line of $f_h(s_{alt})$ against ζ in order to estimate the missing 'before treatment' QoL values, $f_h(s_{alt})$ (see Figure 4-5). We define $\Delta\text{QoL} = f_h(s_{gt}) - f_h(s_{alt})$. Separately, we regress ΔQoL against ζ to interpolate the change in QoL (see Figure 4-6). Given ΔQoL and $f_h(s_{alt})$, we can then estimate the missing values of $f_h(s_{gt})$. Our estimated values are reported in Table 4.15 in the Supplementary Material.

We are able to extract the distribution of the age of cancer onset from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program website. However, empirical age distributions are practically nonexistent for non-cancerous diseases. To overcome this lack, we search the literature for the average age of diagnosis of each disease, and fit a triangle distribution for each disease using the optimization program shown in Figure 4-7.

This program maximizes the domain's interval (Equation 4.22) while imposing the requirement that the distribution's mode, c, has to be in the domain (Equation 4.23). In addition, the area under the curve has to be equal to 1 (Equation 4.24), and the mean of

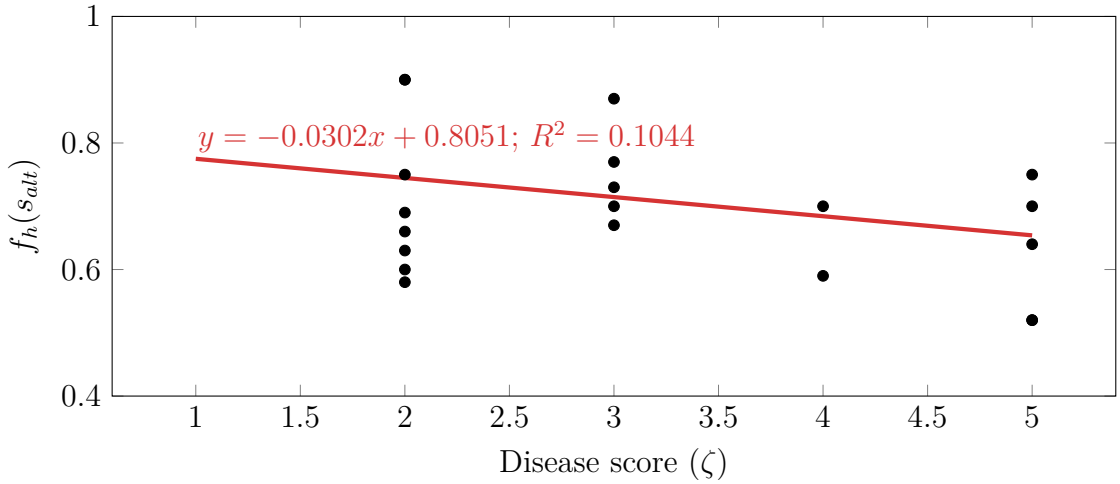


Figure 4-5: Scatter plot of $f_h(s_{alt})$ against disease score (ζ)

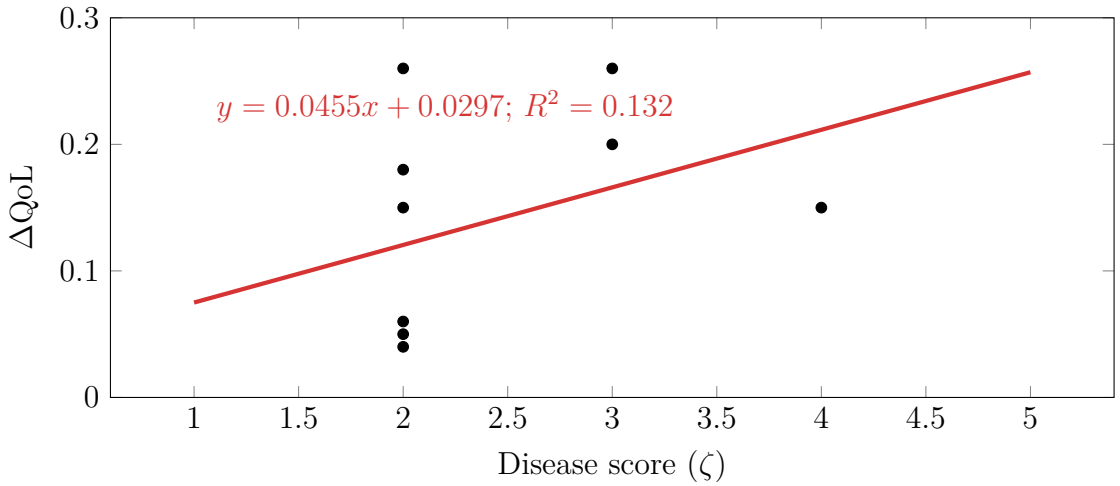


Figure 4-6: Scatter plot of ΔQoL against disease score (ζ)

the distribution has to be equal to the average age (Equation 4.25).

We experimented with an impulse function and an uniform distribution to model the age distribution, but these functions created unrealistic scenarios. Modeling the age distribution with the impulse function, while simple, will force Equation 4.15 to collapse into a single point, and lose any nuance in the QALY gained by patients of different ages. On the other hand, estimating a uniform distribution from the average age creates distributions with narrow support. The distributions from our optimization program have a wider base of support and avoid sharp changes in density. We illustrate this with some examples in Figure 4-14 in the Supplementary Materials.

$$\underset{x_{min},\, x_{max},\, c,\, z}{\text{maximize}} \quad x_{max} - x_{min} \tag{4.22}$$

$$\text{subject to} \quad x_{min} \leq c \leq x_{max}, \tag{4.23}$$

$$\frac{1}{2} z(x_{max} - x_{min}) = 1, \tag{4.24}$$

$$\int_{x_{min}}^{x_{max}} x f(x) dx = \mu_{age} \tag{4.25}$$

$$\text{where} \quad f(x) = \begin{cases} z\dfrac{x - x_{min}}{c - x_{min}} \text{ if } x \leq c \\ z(1 - \dfrac{x - c}{x_{max} - c}) \text{ otherwise} \end{cases} \tag{4.26}$$

Figure 4-7: An optimization program to obtain the triangle distribution given the average age of diagnosis, μ_{age}. x_{min} and x_{max} are coordinates of the base of the triangle. c and z are the mode and height of triangle.

We assume a 3% per annum discount rate, as suggested by ICER for high-impact single or short-term therapy (SST) [92].

4.3.3 Price per ΔQALY

To estimate as realistic a market price of gene therapy as possible, we calibrate our assumed price per ΔQALY with the 4 data points currently available: Zolgensma, priced at $2.1 million per patient [132], Luxturna, priced at $0.425 million per eye treated [157], Kymriah, priced at $0.475 million for a one-time dose [67], and Yescarta, priced at $0.373 million for a one-time dose [67]. Separately, Zynteglo, sold at a cost of 1.6 million Euros (approximately $1.8 million), has been approved in the European Union. The data points are listed in Table 4.6.

Table 4.6: Diseases under consideration, approved gene therapy treatments used as proxy, prices of approved treatments, countries/areas in which treatments have been approved, and computed expected change in QALY.

Disease	Approved treatment	Country approved	List price	E(ΔQALY)
Beta-Thalassemia	Zynteglo	E.U.	1.8M	4.58
Diffuse Large B Cell Lymphoma (DLBCL)	Yescarta	U.S.	0.373M	6.19
Leber Congenital Amaurosis due to RPE65 Mutations	Luxturna	U.S.	0.425M	4.63
Leukemia (Acute Lymphoblastic)	Kymriah	U.S.	0.475M	13.02
Spinal Muscular Atrophy Type 1	Zolgensma	U.S.	2.125M	20.56

We calibrate the price per ΔQALY by minimizing the mean-squared error (MSE) between

the estimated price given the expected change in QALY and the actual price. We report the mean absolute percentage error (MAPE) between the estimated price and the actual price in addition to the MSE. We note that Zolgensma, Zynteglo, and Luxturna are gene replacement therapies for rare diseases, while Kymriah and Yescarta are chimeric antigen receptor T-cell (CAR-T) therapies indicated for cancers. As such, we perform two separate calibrations, one for rare diseases and the other for cancerous indications. We assume that the price per ΔQALY for general diseases is identical to that for cancerous indications.

Considering only the therapies approved in the United States, we estimate a price per E(ΔQALY) of \$101,663 (MSE: 2.18×10^9, MAPE: 11.2%) for rare diseases and \$40,797 (MSE: 1.77×10^10, MAPE: 44.2%) for other diseases. Using all the data points, the price per E(ΔQALY) for rare diseases increases to \$114,781 (MSE: 1.70×10^{12}, MAPE: 108%). In this paper, we use the former for our calculations since it has a smaller mean-squared error and better reflects prices in the U.S., our focus. This value will give us estimates of \$2.09M per patient for Zolgensma and \$0.470M per eye for Luxturna.

Our calibrated price per E(ΔQALY) for cancerous indications is just slightly below ICER's \$50,000 to \$100,000 range for 'intermediate care value'. The higher price per E(ΔQALY) for rare diseases also reaffirms the general belief that developers of treatments for rare diseases should be compensated more for their elevated R&D risk and the low financial prospects of serving a small population of patients. It is assumed that the clinical cost of delivering the gene therapy is a negligible fraction of the overall cost of development (though they are considerably higher than the delivery cost of conventional therapeutics). It is also highly likely that the outside option cost will be similar.

The expected increases in QALY computed by our model are close to those provided by the ICER reports for the treatments [114, 124]. For example, we estimate that treatments for Spinal Muscular Atrophy Type 1 and Leber Congenital Amaurosis due to RPE65 Mutations provide 20.56 and 4.63 incremental QALYs, whereas ICER estimates Zolgensma and Luxturna to provide 12.23 to 26.58 and 1.3 to 2.7 incremental QALYs[4], respectively. We have deliberately applied the same methods and assumptions used for the all other diseases

[4]ICER provides a range of ΔQALY estimates corresponding to different age groups. We have considered the distribution of ages to produce a weighted average estimate.

to estimate the expected changes in QALY for Spinal Muscular Atrophy Type 1 and Leber Congenital Amaurosis due to RPE65 Mutations even though we could have obtained these numbers directly from ICER reports. By doing so, our price per ΔQALY calibration will correct for potential biases in our data, and our price estimates will be more realistic.

Our estimated change in QALY, the price per unit change in QALY, and the estimated price of therapy for each disease are shown in Table 4.7.

Table 4.7: Estimated ΔQALY, assumed price per ΔQALY and estimated price of gene therapies per disease. Prices are given to 3 significant figures for display in this table.

Disease	$\mathbf{\Delta QALY}$	$\frac{\mathbf{Cost}}{\mathbf{\Delta QALY}}$ **($)**	**Price ($)**
General Diseases:			
Arteriosclerosis Obliterans	2.96	41K	121K
Critical Limb Ischemia	7.32	41K	299K
Degenerative Arthritis	3.53	41K	144K
Diabetic Foot Ulcers	7.92	41K	323K
Diabetic Peripheral Neuropathy	3.95	41K	161K
Heart Failure	6.92	41K	282K
Knee Osteoarthritis with Kellgren & Lawrence Grade 3	10.62	41K	433K
Parkinson's Disease	8.26	41K	337K
Peripheral Artery Disease	4.52	41K	184K
Refractory Angina due to Myocardial Ischemia (AFFIRM)	3.80	41K	155K
Stable Angina	3.85	41K	157K
Rare Diseases:			
Beta-Thalassemia	4.58	102K	466K
Cerebral Adrenoleukodystrophy (CALD)	20.33	102K	2.07M
Choroideremia	4.24	102K	431K
Cystic Fibrosis	13.20	102K	1.34M
Ewing's Sarcoma	14.04	102K	1.43M
Hemophilia A	11.18	102K	1.14M
Hemophilia B	10.63	102K	1.08M
Leber Congenital Amaurosis due to RPE65 Mutations	4.63	102K	470K
Leber Hereditary Optic Neuropathy	3.97	102K	404K
Lipoprotein Lipase Deficiency (LPLD)	5.74	102K	584K
Metachromatic Leukodystrophy	21.06	102K	2.14M
Mucopolysaccharidosis Type IIIa	16.27	102K	1.65M
Recessive Dystrophic Epidermolysis Bullosa	5.89	102K	599K

Continued on next page

Table 4.7 – continued from previous page

Disease	ΔQALY	$\frac{\text{Cost}}{\Delta\text{QALY}}$ ($)	Price ($)
Retinitis Pigmentosa	3.28	102K	333K
Sickle Cell Anemia	7.36	102K	748K
Spinal Muscular Atrophy	19.23	102K	1.96M
Spinal Muscular Atrophy Type 1	20.56	102K	2.09M
Cancer:			
B-Cell Non-Hodgkin's Lymphoma	4.90	41K	200K
BCG Unresponsive NMIBC	2.86	41K	117K
Bladder Cancer, in situ concurrent with Papillary Tumors	0.66	41K	26.9K
Diffuse Large B Cell Lymphoma (DLBCL)	6.19	41K	253K
Head and Neck Cancer	6.13	41K	250K
Hepatocellular Carcinoma	9.30	41K	380K
High-Grade Glioma	12.56	41K	512K
Leukemia (Acute Lymphoblastic)	13.04	41K	532K
Leukemia (Acute Myelogenous)	8.55	41K	349K
Lymphoma	4.90	41K	200K
Melanoma (Locally Advanced Cutaneous)	6.23	41K	254K
Melanoma (Metastatic)	9.22	41K	376K
Multiple Myeloma (Newly Diagnosed)	5.90	41K	241K
Nasopharyngeal Carcinoma	5.20	41K	212K
NSC Lung Cancer	7.04	41K	287K
NSC Lung Cancer Stage 3	6.52	41K	266K
Oral Cancer (Advanced)	8.21	41K	335K
Ovarian Cancer (Platinum-Resistant)	10.83	41K	442K
Ovarian Cancer, Primary Peritoneal Cavity Cancer	7.93	41K	324K
Pancreatic Cancer (Locally Advanced)	7.64	41K	312K
Prostate Cancer	0.42	41K	17.1K
Prostate Cancer (Localized)	0.42	41K	17.1K
Prostate Cancer (Metastatic Hormone-Refractory)	7.75	41K	316K
Prostate Cancer (Newly Diagnosed)	0.99	41K	40.4K
Recurrent Glioblastoma	12.55	41K	512K
Relapsed and Refractory Multiple Myeloma (RRMM)	8.33	41K	340K
Squamous Cell Cancer of Head and Neck or Esophagus	5.51	41K	225K
Synovial Sarcoma	8.65	41K	353K

4.4 Results

All the equations are discretized from their continuous form in our computations. When solving integrals using the trapezoidal rule to obtain the ΔQALY, we use strip widths of 1 year for an age range from 0 to 110 years old, the resolution offered by the life tables. When simulating the number of patients and the cost over time, we use time intervals of 1 month. Our code is implemented on Python 3.6 backed by Numpy, and executed on a single 2.2GHz CPU core. Pseudo-code and further details of computation can be found in Section 4.6.8 in the Supplementary Materials.

We perform 1,000,000 iterations of the simulation to compute the mean number of patients and the total spending. With this number of iterations, one can expect the computed mean to be within 1.89% of the true mean 95% of the time (see Section 4.6.7 in the Supplementary Materials). We also report the 5th and 95th percentiles of the computed values as our upper and lower bounds respectively.

In the following section, we define a 'minor' to be a patient below the age of 18 and an 'elderly' patient to be one who is older than 62 years old. The remainder of the patients are labeled as 'adults'.

4.4.1 Expected Number of Approvals and Patients

Our simulations indicate that the expected number of gene therapies approved between January 2020 and January 2034 is 18.3, with a 90% confidence interval of [14.0, 23.0] (see Figure 4-8).

Table 4.8 shows the annual number of patients over time by age groups. Our simulations expect the number of patients treated to grow from 16,244 in 2020 to 94,696 in 2025 before declining to 65,612 in 2034. The decline can be attributed to the declining stock of existing patients as they are treated, and the fact that we do not consider new development programs launched in the future. The proportions of patients who are minors, adults and elderly are 17.9%, 35.4%, and 46.7% respectively.

We show the number of patients treated by month in Figure 4-9a. We can see that our simulations expect the number of patients treated to peak at around 7911 (CI: [3978, 12477])

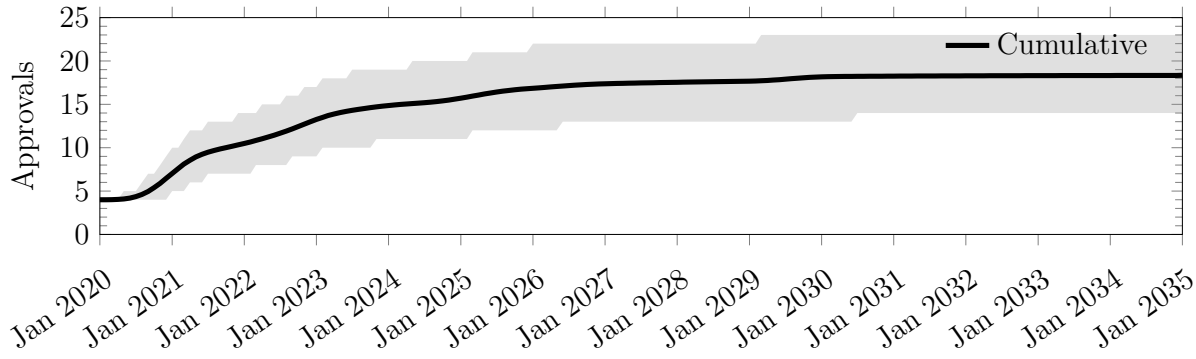


Figure 4-8: Cumulative number of approvals between January 2020 and December 2034, obtained from 1,000,000 simulation runs.

per month in Jul 2025 before declining to 5424 (CI: [2778, 8350]) by December 2034. The monthly number of existing patients treated exceeds the monthly number of newly-diagnosed patients treated until Sep 2024, when this trend is expected to reverse (see Figure 4-9b). Only 7% of all patients treated in December 2034 are preexisting patients. Cancer patients are expected to form the biggest group of patients receiving gene therapy treatments, simply due to the number of cancer indications being targeted. We expect the relative proportions of cancer, general disease, and rare disease patients to be 48.0%, 30.0%, and 22.0%, respectively, in December 2034. The cumulative number of patients to be treated is expected to be 1.09 million (CI: [0.595M, 1.66M]) by the end of December 2034 (see Figure 4-9c).

4.4.2 Expected Spending

We expect an increase in spending up to $2.11 billion per month (CI: [1.01B, 3.88B]) in Apr 2026, before decreasing slowly to a steady-state rate of $1.62 billion (CI: [0.624B, 2.9B]) per month (see Figure 4-10a). We emphasize that the total spending eventually declines because our simulations analyze a fixed stock of innovations, and do not account for new development programs that may be launched in the future. Treating existing cancer patients initially consumes over 45.6% of the total monthly expenditure, but declines to only 0.99% by December 2034. In contrast, the proportion of spending on new patients in the 'general disease' and 'rare disease' groups will increase from 0.0% and 4.26%, respectively, in Feb 2020 to 21.2% and 46.2% by December 2034. The monthly spending on treating existing patients

shall exceed the monthly spending on treating newly diagnosed patients in Nov 2023. The cumulative discounted spending on treating patients with approved gene therapy products is expected to reach $241 billion (CI: [123B, 402B]) by December 2034, 15 years after the start of our simulation.

In terms of annual spending on approved gene therapies, we expect that $5.15 billion will be spent in 2020, increasing to $25.3B in 2026 before declining to $21.0B in 2034 (see Table 4.9). Minors, adults and the elderly will consume 43.2%, 26.0%, and 30.9%, respectively, of the total spending. In the U.S., all elderly people are covered by Medicare. It is also estimated that two in five children and one in seven adults in the U.S. are covered by Medicaid [98]. The remainder of the spending is expected to come from private sources such as employer-provided or private insurance. Using these proportions, we estimate that the expected annual spending by Medicare, Medicaid [5] and private sources respectively may reach $8.1, $5.44, and $12.2 billion (see Table 4.10). We discuss the implications of these estimates in Section 4.4.4.

The total expected increase in QALY over these 15 years is 5.59 million (see Figure 4-11), which translates to an average increase of 5.12 in QALY per patient. This comes at an average 2020 present value cost of $43,110 per unit change in QALY.

[5] The spending estimates for Medicaid do not take into account the 23.1% drug rebate that it is expected to receive [97].

Table 4.8: Expected annual number of patients treated by gene therapy between 2020 and 2035, conditioned on the age group and patient type. 'Minor', 'adult' and 'elderly' are defined to be 'below the age of 18', 'between the ages of 18 and 62', and 'greater than 62 years old', respectively.

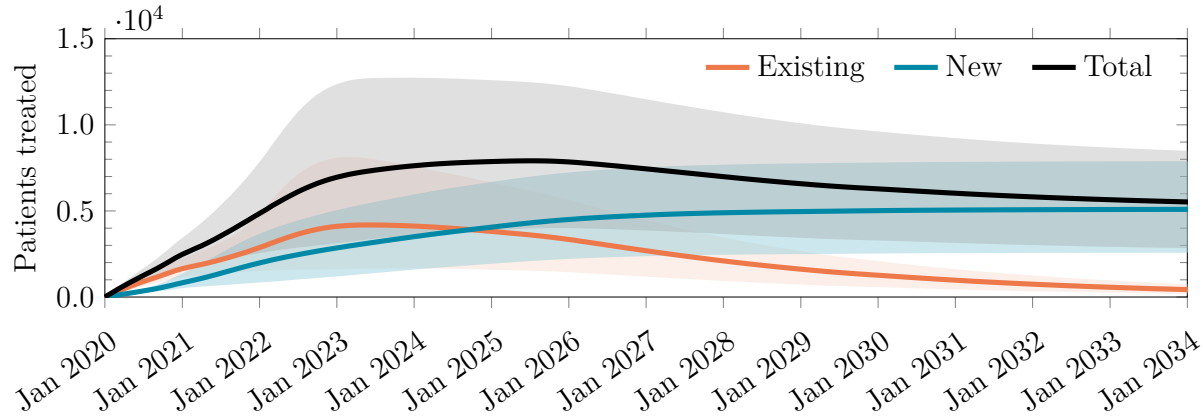
Year	**Minor**			**Adult**			**Elderly**			**Total**
	Existing	New	Subtotal	Existing	New	Subtotal	Existing	New	Subtotal	
2020	1,630	417	2,047	3,682	1,657	5,339	6,006	2,853	8,859	16,244
	(1,283, 1,995)	(322, 522)	(1,614, 2,500)	(2,557, 4,943)	(1,080, 2,454)	(3,670, 7,237)	(3,928, 8,282)	(1,813, 4,142)	(5,785, 12,250)	(11,349, 21,685)
2021	2,833	1,294	4,127	9,921	6,399	16,320	14,278	9,540	23,818	44,265
	(2,001, 4,357)	(681, 1,950)	(2,770, 6,326)	(5,857, 14,948)	(2,970, 11,725)	(9,380, 24,834)	(8,578, 21,422)	(4,783, 17,438)	(14,050, 36,234)	(26,876, 65,911)
2022	3,832	3,235	7,067	16,809	11,041	27,849	23,112	15,513	38,625	73,543
	(2,000, 8,352)	(947, 14,178)	(3,224, 22,601)	(6,965, 31,751)	(4,364, 19,680)	(12,843, 47,995)	(9,343, 46,360)	(6,561, 28,346)	(17,258, 69,494)	(35,001, 126,974)
2023	4,722	5,612	10,334	19,364	13,890	33,254	25,862	19,031	44,893	88,482
	(2,001, 11,596)	(1,243, 25,879)	(3,614, 37,353)	(7,340, 36,474)	(5,999, 23,487)	(15,161, 56,955)	(9,222, 52,202)	(8,407, 32,988)	(19,347, 80,954)	(41,055, 151,872)
2024	4,922	7,734	12,656	18,580	16,230	34,810	24,122	21,781	45,902	93,371
	(1,832, 11,745)	(1,490, 29,056)	(3,681, 40,322)	(7,284, 32,906)	(7,683, 26,300)	(16,794, 56,353)	(8,817, 45,862)	(10,261, 36,233)	(21,011, 77,784)	(45,504, 151,799)
2025	5,235	9,621	14,856	16,570	18,159	34,728	21,115	23,994	45,110	94,696
	(1,865, 11,736)	(1,741, 30,664)	(3,996, 41,541)	(6,585, 28,364)	(9,026, 28,683)	(17,320, 54,531)	(7,795, 38,698)	(11,748, 38,878)	(21,451, 73,552)	(47,833, 148,985)
2026	4,998	11,086	16,085	13,653	19,350	33,003	17,220	25,370	42,592	91,682
	(1,667, 11,079)	(1,918, 31,601)	(3,996, 41,494)	(5,511, 23,012)	(9,839, 30,163)	(16,868, 50,992)	(6,444, 30,948)	(12,692, 40,529)	(20,887, 67,946)	(47,432, 141,917)
2027	4,246	12,120	16,366	10,687	19,915	30,604	13,402	26,032	39,433	86,401
	(1,323, 9,859)	(2,004, 32,129)	(3,676, 40,559)	(4,403, 17,871)	(10,254, 30,847)	(15,948, 46,843)	(5,129, 23,798)	(13,206, 41,300)	(19,847, 62,104)	(45,218, 132,708)
2028	3,522	12,842	16,364	8,203	20,129	28,332	10,215	26,242	36,457	81,153
	(1,024, 8,638)	(2,052, 32,455)	(3,369, 39,508)	(3,402, 13,700)	(10,409, 31,105)	(14,899, 43,197)	(3,945, 18,068)	(13,367, 41,550)	(18,583, 57,103)	(42,510, 124,357)
2029	2,978	13,373	16,351	6,259	20,219	26,478	7,744	26,315	34,059	76,888
	(851, 7,552)	(2,114, 32,684)	(3,217, 38,620)	(2,594, 10,485)	(10,480, 31,214)	(13,969, 40,339)	(3,002, 13,682)	(13,432, 41,632)	(17,475, 53,231)	(40,221, 117,723)
2030	2,656	13,807	16,463	4,764	20,276	25,040	5,863	26,361	32,224	73,726
	(712, 6,628)	(2,222, 32,909)	(3,228, 38,011)	(1,969, 8,027)	(10,526, 31,277)	(13,223, 38,155)	(2,275, 10,366)	(13,473, 41,682)	(16,597, 50,343)	(38,538, 112,779)
2031	2,116	14,039	16,154	3,620	20,313	23,933	4,434	26,391	30,826	70,914
	(540, 5,433)	(2,241, 32,998)	(3,032, 37,097)	(1,490, 6,146)	(10,558, 31,324)	(12,635, 36,496)	(1,720, 7,852)	(13,501, 41,715)	(15,896, 48,190)	(36,887, 108,603)
2032	1,654	14,185	15,840	2,746	20,338	23,084	3,349	26,411	29,759	68,684
	(408, 4,332)	(2,251, 33,049)	(2,861, 36,273)	(1,125, 4,698)	(10,577, 31,355)	(12,174, 35,255)	(1,298, 5,938)	(13,518, 41,736)	(15,350, 46,585)	(35,539, 105,370)
2033	1,286	14,282	15,567	2,079	20,354	22,434	2,528	26,424	28,953	66,952
	(308, 3,400)	(2,257, 33,084)	(2,726, 35,611)	(847, 3,575)	(10,591, 31,374)	(11,813, 34,313)	(978, 4,482)	(13,529, 41,751)	(14,926, 45,369)	(34,479, 102,883)
2034	993	14,344	15,337	1,572	20,364	21,937	1,906	26,432	28,338	65,612
	(232, 2,644)	(2,262, 33,106)	(2,619, 35,063)	(638, 2,713)	(10,601, 31,387)	(11,533, 33,602)	(735, 3,379)	(13,536, 41,761)	(14,597, 44,479)	(33,666, 100,944)

Table 4.9: Expected annual spending on gene therapy between 2020 and 2035, conditioned on the age group and patient type. 'Minor', 'adult' and 'elderly' are defined to be 'below the age of 18', 'between the ages of 18 and 62', and 'greater than 62 years old' respectively. Numbers in billions.

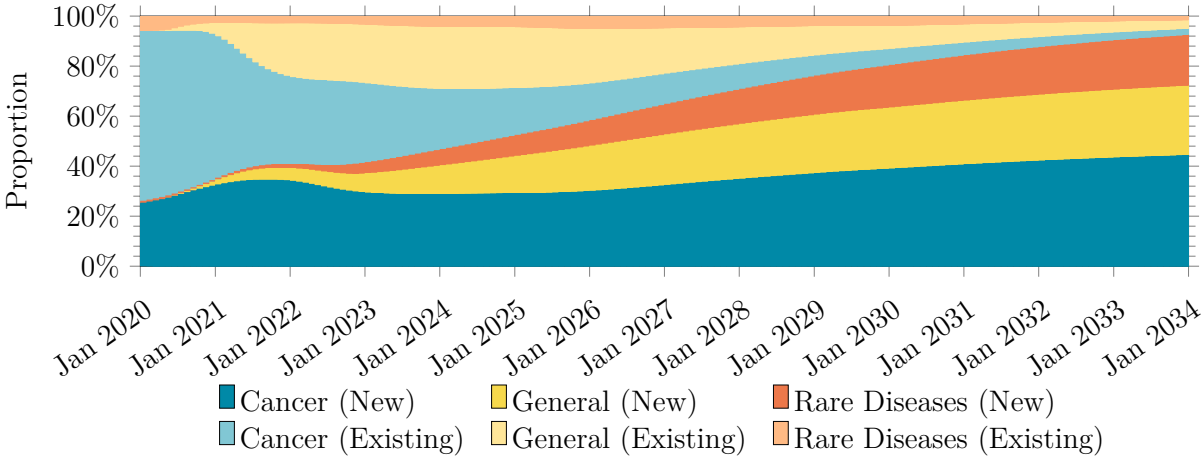
Year	**Minor**			**Adult**			**Elderly**			**Total**
	Existing	New	Subtotal	Existing	New	Subtotal	Existing	New	Subtotal	
2020	1.77	0.32	2.10	0.91	0.37	1.29	1.20	0.57	1.77	5.15
	(1.31, 2.25)	(0.25, 0.40)	(1.55, 2.65)	(0.67, 1.18)	(0.26, 0.53)	(0.93, 1.69)	(0.79, 1.65)	(0.36, 0.80)	(1.15, 2.45)	(4.00, 6.39)
2021	2.23	0.72	2.94	2.19	1.31	3.51	2.76	1.80	4.56	11.01
	(1.60, 3.19)	(0.42, 0.86)	(2.03, 4.06)	(1.43, 3.16)	(0.68, 2.33)	(2.17, 5.31)	(1.73, 3.98)	(0.96, 3.42)	(2.77, 7.20)	(7.47, 15.85)
2022	2.37	1.79	4.16	2.91	2.09	5.01	3.52	2.70	6.22	15.38
	(1.39, 5.76)	(0.50, 9.98)	(1.97, 15.78)	(1.63, 4.41)	(0.94, 3.70)	(2.72, 7.76)	(1.92, 5.30)	(1.31, 5.03)	(3.39, 9.99)	(8.76, 27.15)
2023	2.91	3.33	6.25	3.37	2.67	6.04	3.79	3.34	7.13	19.41
	(1.26, 8.02)	(0.60, 18.48)	(1.97, 26.43)	(1.80, 5.12)	(1.30, 4.42)	(3.34, 9.10)	(1.97, 5.77)	(1.71, 5.81)	(3.88, 11.05)	(10.30, 39.86)
2024	3.06	4.79	7.85	3.46	3.26	6.72	3.84	4.02	7.86	22.43
	(1.06, 8.12)	(0.67, 20.70)	(1.84, 28.46)	(1.80, 5.27)	(1.66, 5.17)	(3.75, 9.99)	(1.94, 5.89)	(2.11, 6.65)	(4.31, 11.92)	(11.27, 44.10)
2025	3.50	6.11	9.61	3.18	3.76	6.94	3.51	4.59	8.10	24.65
	(0.98, 8.37)	(0.77, 21.79)	(1.95, 29.48)	(1.62, 4.88)	(1.96, 5.81)	(3.85, 10.27)	(1.74, 5.43)	(2.42, 7.36)	(4.44, 12.19)	(12.04, 46.08)
2026	3.49	7.15	10.64	2.67	4.08	6.75	2.94	4.96	7.89	25.28
	(0.88, 8.08)	(0.85, 22.43)	(1.94, 29.57)	(1.35, 4.12)	(2.14, 6.25)	(3.74, 10.00)	(1.46, 4.53)	(2.62, 7.84)	(4.33, 11.85)	(11.98, 45.94)
2027	2.97	7.89	10.86	2.14	4.27	6.40	2.34	5.17	7.51	24.77
	(0.70, 7.18)	(0.88, 22.79)	(1.77, 28.88)	(1.08, 3.32)	(2.24, 6.51)	(3.53, 9.49)	(1.17, 3.59)	(2.74, 8.11)	(4.13, 11.27)	(11.28, 44.52)
2028	2.47	8.42	10.89	1.66	4.34	6.00	1.80	5.24	7.04	23.92
	(0.54, 6.30)	(0.89, 23.01)	(1.62, 28.11)	(0.83, 2.63)	(2.28, 6.61)	(3.29, 8.93)	(0.90, 2.75)	(2.78, 8.20)	(3.87, 10.61)	(10.48, 42.84)
2029	2.25	8.86	11.10	1.28	4.36	5.64	1.37	5.26	6.63	23.37
	(0.48, 5.71)	(0.95, 23.22)	(1.62, 27.70)	(0.64, 2.07)	(2.30, 6.64)	(3.08, 8.42)	(0.69, 2.10)	(2.79, 8.22)	(3.64, 10.04)	(10.03, 41.59)
2030	2.37	9.31	11.68	0.98	4.38	5.36	1.04	5.27	6.31	23.35
	(0.39, 5.52)	(1.02, 23.53)	(1.58, 27.89)	(0.48, 1.62)	(2.31, 6.66)	(2.91, 8.02)	(0.52, 1.60)	(2.80, 8.24)	(3.45, 9.60)	(9.98, 41.08)
2031	1.90	9.49	11.39	0.75	4.39	5.14	0.79	5.28	6.07	22.59
	(0.30, 4.51)	(1.03, 23.61)	(1.47, 27.11)	(0.36, 1.26)	(2.32, 6.67)	(2.78, 7.71)	(0.39, 1.22)	(2.81, 8.25)	(3.31, 9.28)	(9.47, 39.80)
2032	1.47	9.60	11.07	0.57	4.40	4.97	0.60	5.28	5.88	21.92
	(0.22, 3.57)	(1.03, 23.64)	(1.38, 26.37)	(0.28, 0.97)	(2.32, 6.68)	(2.68, 7.47)	(0.30, 0.93)	(2.82, 8.25)	(3.20, 9.03)	(9.02, 38.69)
2033	1.14	9.67	10.81	0.43	4.40	4.84	0.45	5.29	5.74	21.38
	(0.17, 2.78)	(1.04, 23.67)	(1.30, 25.78)	(0.21, 0.74)	(2.33, 6.69)	(2.60, 7.29)	(0.22, 0.70)	(2.82, 8.25)	(3.11, 8.84)	(8.67, 37.83)
2034	0.87	9.72	10.59	0.33	4.41	4.73	0.34	5.29	5.63	20.95
	(0.13, 2.15)	(1.04, 23.68)	(1.24, 25.31)	(0.16, 0.57)	(2.33, 6.69)	(2.53, 7.15)	(0.17, 0.53)	(2.82, 8.26)	(3.04, 8.70)	(8.40, 37.13)

Table 4.10: Expected annual spending on gene therapy between 2020 and 2035 by funding source. Numbers in billions.

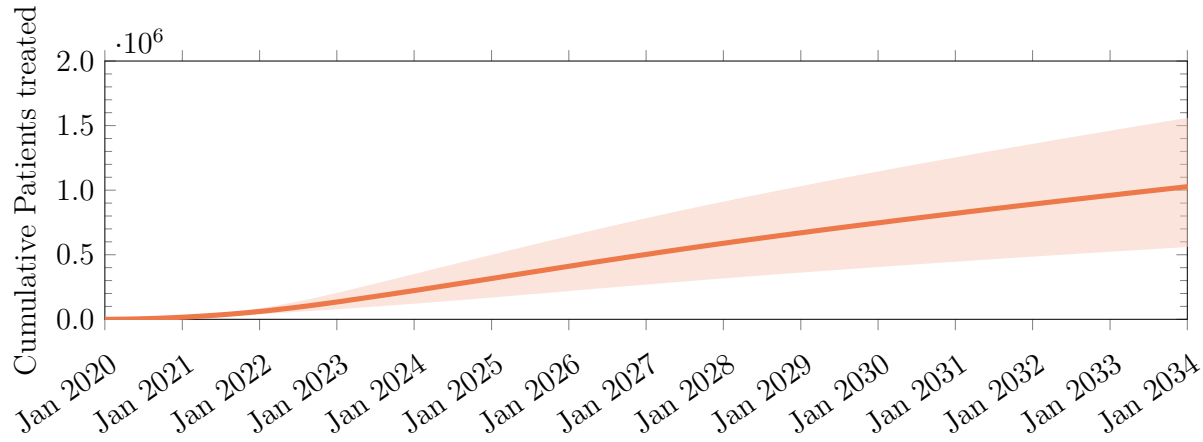
	Medicare	Medicaid	Private
2020	1.77 (1.15, 2.45)	1.02 (0.79, 1.26)	2.36 (1.87, 2.87)
2021	4.56 (2.77, 7.20)	1.68 (1.20, 2.28)	4.77 (3.29, 6.78)
2022	6.22 (3.39, 9.99)	2.38 (1.27, 7.02)	6.79 (3.79, 13.75)
2023	7.13 (3.88, 11.05)	3.36 (1.40, 11.44)	8.92 (4.46, 21.11)
2024	7.86 (4.31, 11.92)	4.10 (1.43, 12.37)	10.47 (4.84, 23.19)
2025	8.10 (4.44, 12.19)	4.83 (1.55, 12.82)	11.71 (5.22, 24.14)
2026	7.89 (4.33, 11.85)	5.22 (1.55, 12.83)	12.17 (5.20, 24.08)
2027	7.51 (4.13, 11.27)	5.26 (1.43, 12.50)	12.01 (4.86, 23.37)
2028	7.04 (3.87, 10.61)	5.21 (1.32, 12.14)	11.68 (4.50, 22.55)
2029	6.63 (3.64, 10.04)	5.25 (1.28, 11.92)	11.50 (4.32, 21.97)
2030	6.31 (3.45, 9.60)	5.44 (1.26, 11.96)	11.60 (4.33, 21.81)
2031	6.07 (3.31, 9.28)	5.29 (1.19, 11.61)	11.24 (4.09, 21.13)
2032	5.88 (3.20, 9.03)	5.14 (1.13, 11.29)	10.90 (3.89, 20.54)
2033	5.74 (3.11, 8.84)	5.01 (1.08, 11.03)	10.63 (3.73, 20.06)
2034	5.63 (3.04, 8.70)	4.91 (1.04, 10.83)	10.41 (3.60, 19.69)



(a) Monthly number of patients treated with gene therapy across all diseases. The line represents the mean and the shaded region represents the 5th and 95th percentiles of our simulation.

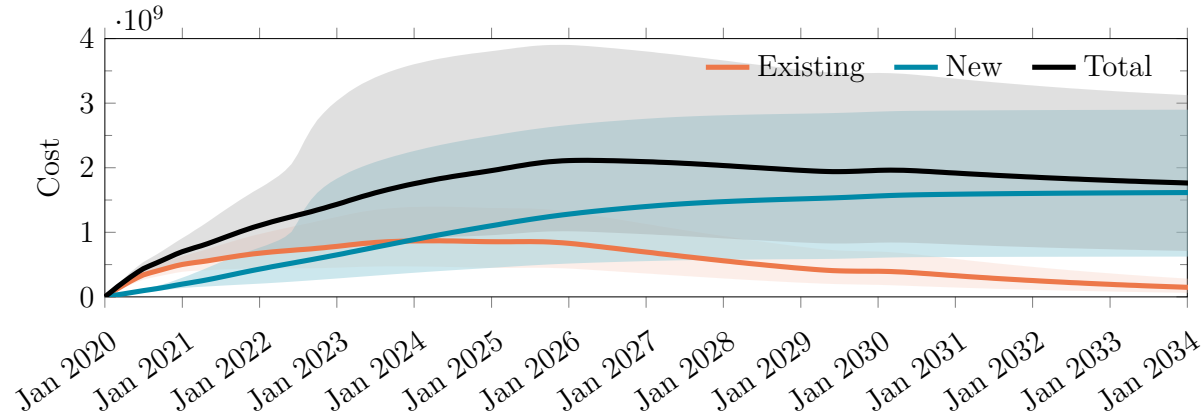


(b) Stacked chart depicting the proportion of existing and new patients treated in that month, by disease category.

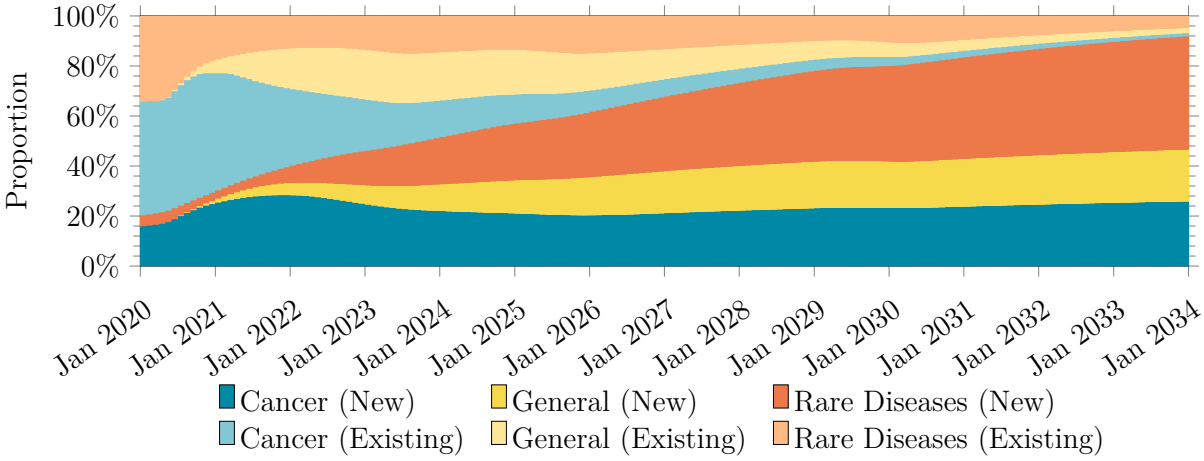


(c) Cumulative number of patients treated. The line represents the mean and the shaded region represents the 5th and 95th percentiles of our simulation.

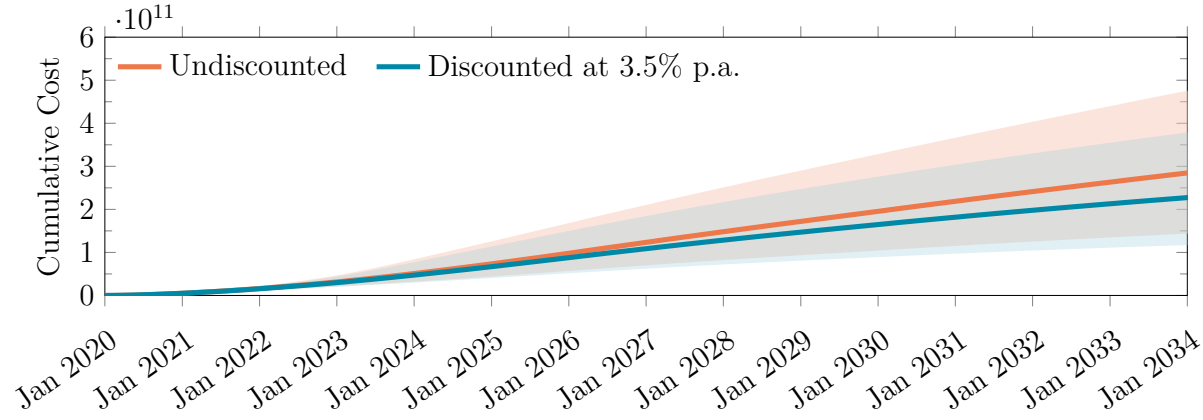
Figure 4-9: Number of patients treated between January 2020 and December 2034, obtained from 1,000,000 simulation runs.



(a) Monthly spending on treating existing and new patients with gene therapy. The line represents the mean and the shaded region represents the 5th and 95th percentiles from our simulation.

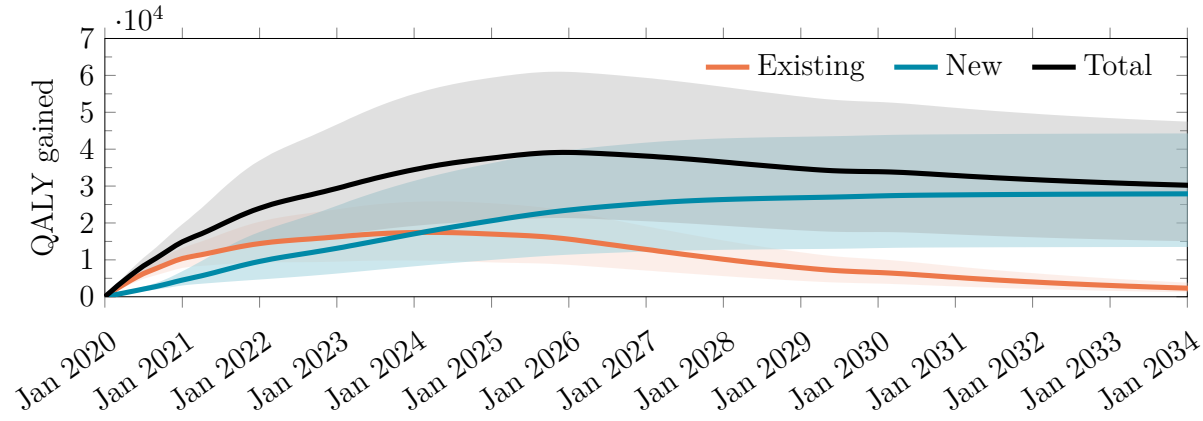


(b) Stacked chart depicting the proportion of spending on treating existing and new patients in that month, by disease category.

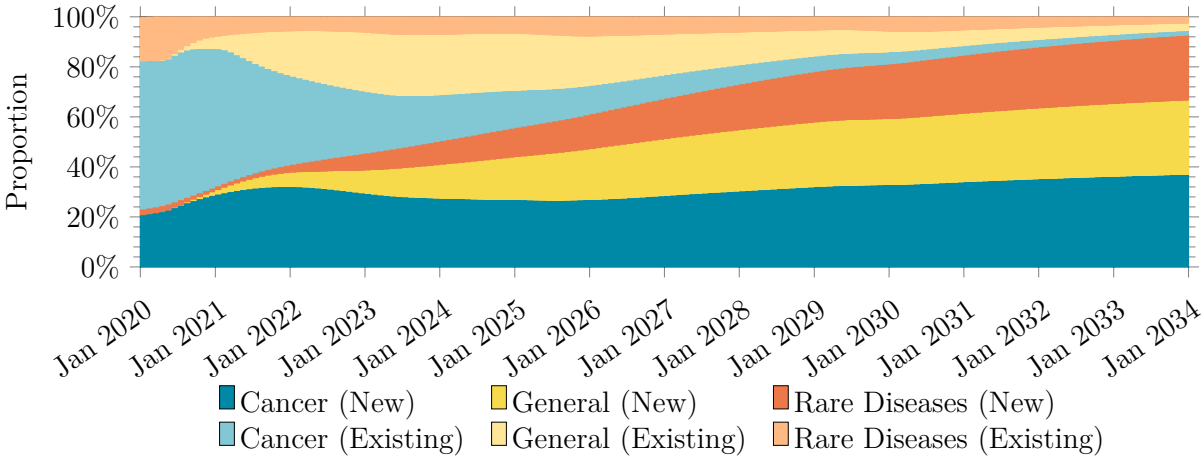


(c) Cumulative spending on treating patients with gene therapy. The line represents the mean and the shaded region represents the 5th and 95th percentiles of our simulation.

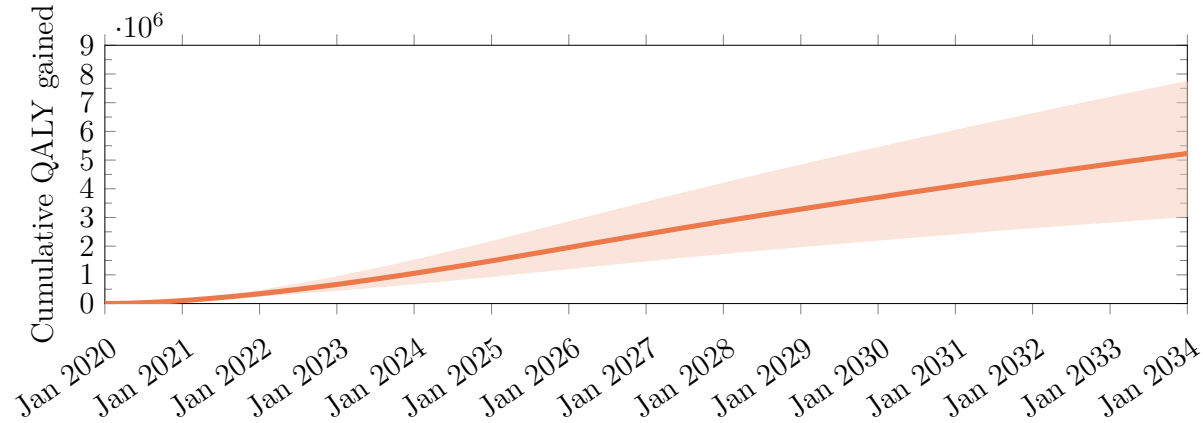
Figure 4-10: Spending on gene therapy between January 2020 and December 2034, obtained from 1,000,000 simulation runs.



(a) QALY gained by treating existing and new patients with gene therapy. The line represents the mean and the shaded region represents the 5th and 95th percentiles from our simulation.



(b) Stacked chart depicting the QALY gained by treating existing and new patients in that month, by disease category.



(c) Cumulative QALY gained by treating patients with gene therapy. The line represents the mean and the shaded region represents the 5th and 95th percentiles of our simulation.

Figure 4-11: Expected ΔQALY made possible by gene therapy treatments between January 2020 and December 2034, obtained from 1,000,000 simulation runs.

4.4.3 Sensitivity Analysis

To test the sensitivity of our results to initial conditions, we simulate ±20% changes in the following variables, analyzing their impact on our results.

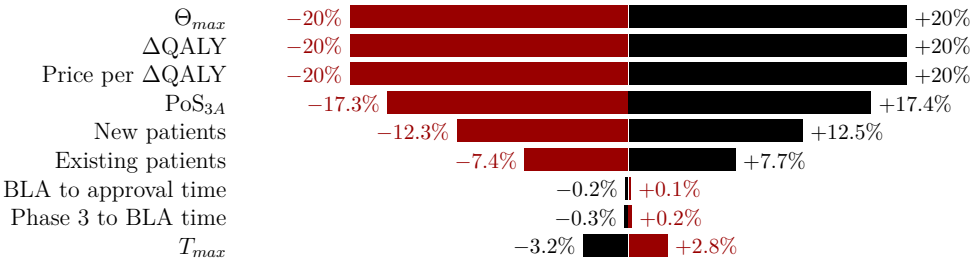
1. The maximum penetration rate in the ramp function, Θ_{max}
2. The time to maximum penetration rate in the ramp function, T_{max}
3. The amount of QALY gained in each disease
4. The price per ΔQALY
5. The phase-3-to-approval probability of success (PoS_{3A})
6. The number of new patients of each disease
7. The number of existing patients of each disease
8. The time from phase 3 to BLA
9. The time from BLA to approval

For each of these factors, we consider its impact on the the peak monthly spending and the cumulative spending from January 2020 to December 2034 of patient treatment. We also look at how the variables change the timing of the peak monthly spending.

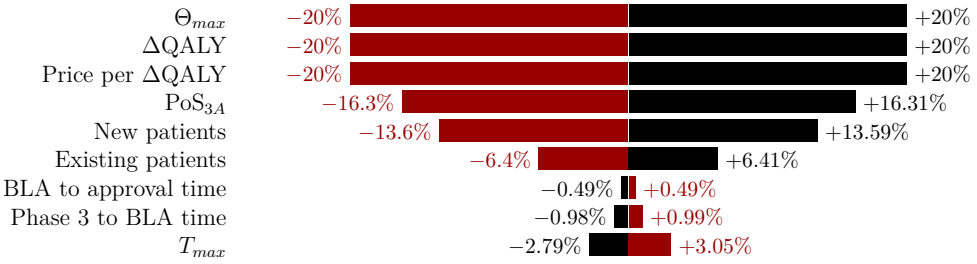
As can be seen from Figure 4-12, the percentage change in the discounted cumulative spending and the maximum monthly spending on treating all patients with gene therapy scale linearly with the percentage change in several variables: the maximum penetration rate (Θ_{max}) and the QALY gained (ΔQALY), the price per ΔQALY. Increasing or decreasing the transition probability from phase 3 to approval, or the number of new or existing patients only leads to sublinear increases or decreases in the discounted cumulative spending and the maximum monthly spending. However, changing the time variables, such as the number of days from phase 3 to BLA, from BLA to approval, or the ramp-up period (T_{max}), induce a small change in the opposite direction.

Introducing perturbations of 20% in the probability of success, the number of new patients, the number of days from Phase 3 to BLA or from BLA to approval, or the time to maximum penetration rate in the ramp function (T_{max}) will change the date of the peak monthly spending in the same direction as the perturbation, by up to 10 months. Increasing or decreasing the number of existing patients, on the other hand, will cause a shift of up to

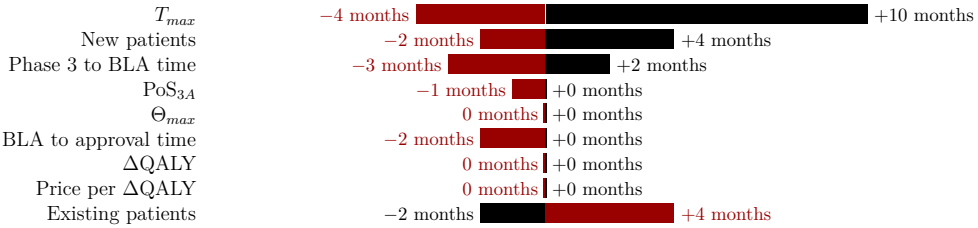
4 months in the date of peak spending in the opposite direction. Perturbing the maximum penetration rate (Θ_{max}), the QALY gained (ΔQALY), and the price per ΔQALY will not change the date of peak spending.



(a) Tornado chart of the impact of the variables on the peak value.



(b) Tornado chart of the impact of the variables on the cumulative spending (both nominal and discounted).



(c) Tornado chart of the impact of the variables on the date of peak value. Since we compute by calendar month, a small machine precision error may change the results by 1 month.

Figure 4-12: Tornado charts showing the sensitivity of the variables on the different metrics. The black bars represent the effect of increasing the variable by 20% and the red bars represent the effect of decreasing the variable by 20%.

We also study the effect of changing the correlation between development programs. Changing the correlation from our assumed value of 0.9 to 0 (i.e., perfectly uncorrelated) increases the mean discounted cumulative spending by 3.4%, from \$241 billion to \$245 billion. Increasing the correlation to 1.0 instead will decrease the mean discounted cumulative spending by 0.4% to \$236 billion.

We vary the proportion of existing patients seeking treatment in the first year – which

determines the λ parameter in Equation 4.3 – and observe that mean discounted cumulative spending changes by between -32% and +0.08% (see Figure 4-13). We can expect the results to differ by less than 5% from the baseline if the proportion of existing patients seeking treatments in the first year is between 8% and 45%.

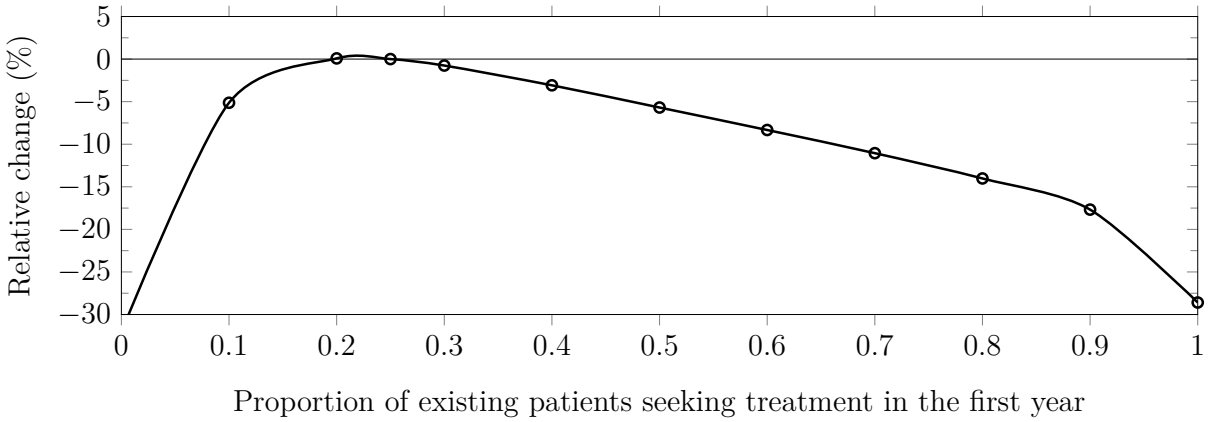


Figure 4-13: Percentage change in the discounted cumulative spending compared to the baseline when the proportion of existing patients seeking treatment in the first year changes.

Our study is independent of the results by Quinn et al. [142], who estimated that 341,775 patients will have been treated with gene therapy by December 2030, and increases by approximately 50,000 per year in the steady state. The authors of this report did not attempt to quantify the cost of providing gene therapy to these patients. In contrast, our simulation expects that about 820,425 patients will be treated by the end of December 2030, with a steady-state increase of around 61,170 per year in the long run.

Some of the differences between our estimates and this other report are due to differences in sample inclusion criteria and the use of different data for patient prevalence and incidence of disease. For example, Quinn et al. [142] considers “durable" gene therapies under all phases of clinical investigation whereas we consider any therapy with late-stage clinical trial(s). Furthermore, they assume that the ‘potentially treatable pool in oncology is entirely incident—there is no prevalence’. Another difference arises from our decision to start with the broadest range of patients and then deflate these numbers through the penetration rate, rather than attempting to estimate the prevalence and incidence for each patient segment. If we removed existing oncology patients from our simulation, the cumulative number of patients treated by December 2030 becomes 666,895, approximately 1.95 times the estimate

in Quinn et al. [142]. We can obtain similar patient estimates to Quinn et al. [142] simply by reducing our penetration rates by 48.8%, which will lower our estimated cumulative spending on gene therapy between January 2020 and December 2034 to $149 billion.

4.4.4 Discussion

We estimate that 1.09 million patients are to be treated with gene therapy by the end of December 2034, spending up to $25.3B annually. These estimates are likely to be lower bounds since our simulation employs conservative assumptions about the speed and volume of gene therapy development. Specifically, we consider only late-stage gene therapy development programs, defined as those already in phase 2/3 or phase 3, and do not account for the possibility that a program in phase 1 or phase 2 may be fast-tracked or granted accelerated approval. We also do not account for any new gene therapies programs being added and approved between 2020 and 2034.

A potential criticism of our approach is that estimating the cost of gene therapies solely based on the change in QALY will overestimate the aggregate spending in the U.S. For example, we do not take into account the potential cost savings to gene therapies due to avoiding multiple costly therapeutic sessions over time based on the current standard of care, or to the recovery of the opportunity cost of caregivers. We have omitted the clinical costs of delivering gene therapy in our analysis, which are often higher than conventional therapeutics due to the need for inpatient hospital care. While there are cases where gene therapy is predicted to provide net cost-savings in treatment after accounting for the direct medical cost (e.g., valoctocogene roxaparvovec for the management of hemophilia A [80]), there is not yet evidence showing that gene therapy will result in net long-term cost savings. In addition, research has shown that new medical technologies generally raise health costs, and that cost-increasing changes in treatments outweigh cost-saving changes the majority of the time [101]. We also do not consider any markup that happens under the prevalent 'buy-and-bill' process in the U.S. These considerations suggest that our approach is indeed conservative, and that our estimates are likely to be lower bounds for realized costs over the next 15 years. Nonetheless, we have taken care to calibrate our price per ΔQALY using actual prices for approved therapeutics and estimating QALYs for those diseases, thereby

allowing us to produce price estimates that closely track past data.

Another potential criticism is that we fail to consider the possibility that having multiple gene therapies for the same disease may lower the prices of the therapies. However, there is no analogous evidence that the presence of multiple brand-name drugs in the same class lowers the list prices of the drugs [149].

Based on our assumptions, the annual spending on gene therapy will average $20.4 billion and may reach $25.3 billion in 2026. The cumulative spending on all future gene therapies from from January 2020 to January 2034 will be approximately $306 billion, or $241 billion when discounted at a cost of capital of 3% per annum over the next 15 years. We estimate the cost of gene therapy to average $43,110 per unit QALY, several times the average annual expenditure of $16,346 for American cancer patients between 2010 and 2014 [138].

However, when viewed from the broader perspective of aggregate U.S. spending, these figures seem less daunting. In 2018, the U.S. tax revenue was $3.33 trillion, of which individual income tax and payroll tax revenues were $1.68 and $1.17 trillion, respectively [136]. If the average spending of $20.4 billion were to be fully funded through income and payroll taxes, an increase of 0.612% would cover the expense. From a purely budgetary perspective, universal access to gene therapy should be feasible if taxpayers are willing to pay for it. This is probably the simplest and most direct way to provide access to gene therapy to all Americans, though it may not be politically feasible.

Since all elderly patients are covered by Medicare, we estimate that the program would need to increase its annual budget by up to $8.1 billion, or 1.1% of its 2018 spending of $750.2 billion [19]. Funding this increase would require either an increase in payroll taxes or a reduction in other expenditures.

We estimate that annual gene therapy spending by Medicaid may reach $5.44 billion. This is approximately 0.9% of its 2018 spending of $597.4 billion [19]. Since Medicaid must be provided to all eligible Americans without any preset cap, managing this increase will require either raising funds from state and federal governments to pay for these additional costs, or cutting benefits.

Annual spending by minors and adults who are ineligible for Medicare or Medicaid—and therefore must rely on private insurers—is predicted to reach $12.2 billion. This spending

poses a significant challenge for insurers and companies, who face annual budgets and competing priorities. In order to manage spending, many insurance policies might choose not to cover spending on gene therapy, or impose very restrictive policies to limit the number of potential patients who might be treated [156]. Many private insurers are already warning they may not be able or willing to absorb the additional spending should a greater number of people become eligible for expensive gene therapy treatments once new ones reach the market [159].

Many novel methods to finance gene therapy treatments through the existing healthcare infrastructure have been proposed, such as outcome-based payments, whereby the manufacturers would be paid only if the patients achieve predefined outcomes after treatment [72]. We note that both Zolgensma and Luxturna have offered outcome-based payment methods to payers. There have also been proposals to allow mortgage-like payments, and even performance-based annuity payments, as ways to finance gene therapy treatments [133]. In September 2019, Cigna, one of the largest U.S. health insurance companies, announced a program called Embarc Benefit Protection in which employers, health plans, and unions pay a monthly per-member premium that provides members with access to the two FDA-approved gene therapies, Luxturna and Zolgensma, at no out-of-pocket costs to them if their physicians authorize treatment. At the time of writing, Luxturna is not provided under Embarc Benefit Protection, and it is unclear if the program is in effect. Cigna hopes to keep the monthly cost of the program to below $1 per member [164], but if our simulations are accurate, this will be financially infeasible.

A more ambitious proposal involves creating a national and possibly international gene therapy reinsurance company that performs a similar function to Embarc, but which serves a large number of primary health insurance providers. By allowing multiple primary insurers to cede the specific risk of gene therapy patients to the reinsurer, these risks can be diversified over a much larger pool members, lowering the cost of capital. The capital required for such a reinsurer can be raised through securitization techniques as described in [133], who simulated such a structure, and concluded that the returns to investors would be quite attractive under a broad range of assumptions. However, their simulations were not specifically calibrated for gene therapy, hence our framework may provide a useful complement to their analysis.

Also, it may be more cost-effective for the reinsurer to assume the responsibility of delivering the gene therapies it reinsures through nationally distributed Centers of Excellence (CoEs). This may seem too far afield for a reinsurance company, but the ability to have direct control over the quality of delivery, and to be able to collect data on the performance of these therapies over time, are two compelling reasons for the reinsurer to take this on. The data collected from these centers will be critical, not only for assessing the actuarial risk of reinsurance, but also for implementing performance-related contractual agreements, e.g., if a gene therapy ceases to be effective, then any remaining payments for the therapy will be cancelled.

An additional benefit of a single reinsurer to manage the risk and responsibility of delivering gene therapy is the ability of that reinsurer to avoid the adverse selection problem that often plagues individual insurers [71]. This problem arises when some insurers are willing to pay for gene therapy treatments while others are not, leading patients who require gene therapies to enroll en masse with those insurers providing coverage. Since these policies will likely have higher premiums to cover the high cost of gene therapy, patients have an incentive to leave the policy after receiving the treatment, leaving the insurers to pay the remaining cost without being able to recover the expenses. If a single reinsurer can aggregate this risk across a large pool of gene therapy patients and coordinate payouts across all the insurers, this adverse selection problem can be greatly mitigated, or altogether avoided. The viability of such a reinsurance vehicle would depend critically on the various parameters of the modules in our simulation, as well as the ability to engage with the largest health insurer of all, the U.S. government.

4.5 Conclusion

In this paper, we estimate the number of patients who will be treated by gene therapy between January 2020 and December 2034 using various data sources. We also develop a mathematical model to estimate the cost of these gene therapies, and calibrated the model to yield realistic cost estimates. It is our hope that this study, and our estimates of the potential financial impact of gene therapy in the U.S., will clarify some of the unknowns

surrounding the impact of this new class of treatment, and allow policymakers, healthcare providers, insurance companies and patients alike to make more informed financial decisions about the future of this important therapeutic class.

Chapter References

[1] Diffuse large b-cell lymphoma - cancer stat facts. URL https://seer.cancer.gov/statfacts/html/dlbcl.html. Accessed: 2020-04-10.

[2] Key statistics for multiple myeloma. URL https://www.cancer.org/cancer/multiple-myeloma/about/key-statistics.html. Accessed: 2020-05-29.

[3] High grade gliomas. URL https://www.aboutkidshealth.ca/Article?contentid=1312&language=English. Accessed: 2020-04-10.

[4] Cancer of the Oral Cavity and Pharynx - Cancer Stat Facts. URL https://seer.cancer.gov/statfacts/html/oralcav.html. Accessed: 2020-04-10.

[5] Heart Failure. URL https://www.cdc.gov/dhdsp/data_statistics/fact_sheets/fs_heart_failure.htm. Accessed: 2020-04-10.

[6] Acute Lymphocytic Leukemia - Cancer Stat Facts. URL https://seer.cancer.gov/statfacts/html/alyl.html. Accessed: 2020-04-10.

[7] Acute Myeloid Leukemia - Cancer Stat Facts. URL https://seer.cancer.gov/statfacts/html/amyl.html. Accessed: 2020-04-10.

[8] Anaplastic Astrocytoma brain cancer. URL http://www.orbustherapeutics.com/anaplastic. Accessed: 2020-04-10.

[9] Diffuse Large B-Cell Lymphoma., . URL https://www.lymphoma.org/aboutlymphoma/nhl/dlbcl/. Accessed: 2020-04-10.

[10] Non-Hodgkin Lymphoma - Cancer Stat Facts., . URL https://seer.cancer.gov/statfacts/html/nhl.html. Accessed: 2020-04-10.

[11] Thalassemia Awareness. URL https://www.cdc.gov/features/international-thalassemia/index.html. Accessed: 2020-04-10.

[12] Cancer of the Urinary Bladder - Cancer Stat Facts. URL https://seer.cancer.gov/statfacts/html/urinb.html. Accessed: 2020-04-10.

[13] Key Statistics for Ewing Tumors., . URL https://www.cancer.org/cancer/ewing-tumor/about/key-statistics.html. Accessed: 2020-04-10.

[14] Ewing Sarcoma., . URL https://rarediseases.org/rare-diseases/ewing-sarcoma/. Accessed: 2020-04-10.

[15] Hepatocellular Carcinoma. URL https://rarediseases.org/rare-diseases/hepatocellular-carcinoma/. Accessed: 2020-04-10.

[16] Hodgkin lymphoma - cancer stat facts. URL https://seer.cancer.gov/statfacts/html/hodg.html. Accessed: 2020-04-10.

[17] Melanoma of the Skin - Cancer Stat Facts. URL https://seer.cancer.gov/statfacts/html/melan.html. Accessed: 2020-04-10.

[18] Cancer of the Prostate - Cancer Stat Facts. URL https://seer.cancer.gov/statfacts/html/prost.html. Accessed: 2020-04-10.

[19] NHE Fact Sheet. URL https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet. Accessed: 2020-04-10.

[20] NIH Research Portfolio Online Reporting Tools (RePORT). URL https://report.nih.gov/NIHfactsheets/ViewFactSheet.aspx?csid=109. Accessed: 2019-06-10.

[21] Cancer of the Urinary Bladder - Cancer Stat Facts, . URL https://seer.cancer.gov/statfacts/html/urinb.html. Accessed: 2020-03-20.

[22] Cancer of the Oral Cavity and Pharynx - Cancer Stat Facts, . URL https://seer.cancer.gov/statfacts/html/oralcav.html. Accessed: 2020-03-20.

[23] Cancer of the Ovary - Cancer Stat Facts, . URL https://seer.cancer.gov/statfacts/html/ovary.html. Accessed: 2020-03-20.

[24] Cancer of the Pancreas - Cancer Stat Facts, . URL https://seer.cancer.gov/statfacts/html/pancreas.html. Accessed: 2020-03-20.

[25] Cancer of the Prostate - Cancer Stat Facts, . URL https://seer.cancer.gov/statfacts/html/prost.html. Accessed: 2020-03-20.

[26] Diffuse Large B-Cell Lymphoma - Cancer Stat Facts, . URL https://seer.cancer.gov/statfacts/html/dlbcl.html. Accessed: 2020-03-20.

[27] Leber Hereditary Optic Neuropathy, . URL https://rarediseases.org/rare-diseases/leber-hereditary-optic-neuropathy/. Accessed: 2020-04-10.

[28] Melanoma of the Skin - Cancer Stat Facts, . URL https://seer.cancer.gov/statfacts/html/melan.html. Accessed: 2020-03-20.

[29] Non-Hodgkin Lymphoma - Cancer Stat Facts, . URL https://seer.cancer.gov/statfacts/html/nhl.html. Accessed: 2020-03-20.

[30] Orphanet: Mucopolysaccharidosis type 3, . URL https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Expert=581. Accessed: 2020-02-27.

[31] Survival Rates and Factors That Affect Prognosis (Outlook) for Non-Hodgkin Lymphoma, . URL https://www.cancer.org/cancer/non-hodgkin-lymphoma/detection-diagnosis-staging/factors-prognosis.html. Accessed: 2020-03-20.

[32] Survival Rates for Nasopharyngeal Cancer, . URL https://www.cancer.org/cancer/nasopharyngeal-cancer/detection-diagnosis-staging/survival-rates.html. Accessed: 2020-03-20.

[33] Synovial Sarcoma, . URL https://www.stjude.org/disease/synovial-sarcoma.html. Accessed: 2020-03-20.

[34] Non-hodgkin lymphoma - cancer stat facts. URL https://seer.cancer.gov/statfacts/html/nhl.html. Accessed: 2020-04-10.

[35] Cancer of the Lung and Bronchus - Cancer Stat Facts. URL https://seer.cancer.gov/statfacts/html/lungb.html. Accessed: 2020-04-10.

[36] Ovarian Cancer Statistics – Ovarian Cancer Research Alliance., . URL https://ocrahope.org/patients/about-ovarian-cancer/statistics/. Accessed: 2020-04-10.

[37] Recurrence and Treatment – Ovarian Cancer Research Alliance., . URL https://ocrahope.org/patients/about-ovarian-cancer/recurrence/. Accessed: 2020-04-10.

[38] Cancer of the Ovary – Cancer Stat Facts., . URL https://seer.cancer.gov/statfacts/html/ovary.html. Accessed: 2020-04-10.

[39] Cancer of the Pancreas - Cancer Stat Facts. URL https://seer.cancer.gov/statfacts/html/pancreas.html. Accessed: 2020-04-10.

[40] Primary peritoneal cancer: Know the basics. URL https://www.curetoday.com/publications/cure/2019/womens-cancers/primary-peritoneal-cancer-know-the-basics. Accessed: 2020-04-10.

[41] Multiple Myeloma Prognosis - Multiple Myeloma Survival Rate. URL https://themmrf.org/multiple-myeloma/prognosis/. Accessed: 2020-04-10.

[42] About SMA. URL https://smafoundation.org/about-sma/. Accessed: 2020-04-10.

[43] Synovial sarcoma. URL https://rarediseases.info.nih.gov/diseases/7721/synovial-sarcoma. Accessed: 2020-04-10.

[44] Angina by the Numbers - Mortality, Incidence, Prevalence, and other Angina Statistics., Dec 2009. URL https://www.mdmag.com/medical-news/angina_statistics. Accessed: 2020-04-10.

[45] Lung Cancer - Non-Small Cell - Statistics, June 2012. URL https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/statistics. Accessed: 2020-03-20.

[46] Treatment of Relapsed or Refractory Multiple Myeloma in the Era of Novel Agents, May 2012. URL https://touchoncology.com/treatment-of-relapsed-

or-refractory-multiple-myeloma-in-the-era-of-novel-agents/. Accessed: 2020-03-20.

[47] Fast Facts, Jul 2015. URL https://www.hemophilia.org/About-Us/Fast-Facts. Accessed: 2020-04-10.

[48] The Epidemiology of Lysosomal Storage Disorders | DRG Blog, October 2016. URL https://decisionresourcesgroup.com/blog/epidemiology-lysosomal-storage-disorders/. Accessed: 2020-04-10.

[49] Glioblastoma and Malignant Astrocytoma, 2017.

[50] Cancer Facts & Figures 2017, 2017. URL https://www.cancer.net/cancer-types/oral-and-oropharyngeal-cancer/statistics. Accessed: 2020-04-10.

[51] Brain tumours., Oct 2019. URL https://www.cancerresearchuk.org/about-cancer/brain-tumours/types/glioma-adults. Accessed: 2020-04-10.

[52] What is the mortality rate for heart failure?, Nov 2019. URL https://www.medscape.com/answers/163062-86190/what-is-the-mortality-rate-for-heart-failure. Accessed: 2020-04-10.

[53] Laryngeal and Hypopharyngeal Cancer - Statistics, Aug 2019. URL https://www.cancer.net/cancer-types/laryngeal-and-hypopharyngeal-cancer/statistics. Accessed: 2020-04-10.

[54] Nasal Cavity and Paranasal Sinus Cancer - Statistics, Feb 2019. URL https://www.cancer.net/cancer-types/nasal-cavity-and-paranasal-sinus-cancer/statistics. Accessed: 2020-04-10.

[55] Oral and Oropharyngeal Cancer - Statistics, Dec 2019. URL https://www.cancer.net/cancer-types/oral-and-oropharyngeal-cancer/statistics. Accessed: 2020-04-10.

[56] Nasopharyngeal cancer: Statistics, January 2019. URL https://www.cancer.net/cancer-types/nasopharyngeal-cancer/statistics. Accessed: 2020-04-10.

[57] Lung Cancer - Non-Small Cell - Statistics., Mar 2019. URL https://www.cancer.net/cancer-types/lung-cancer-non-small-cell/statistics. Accessed: 2020-04-10.

[58] Data & Statistics on Sickle Cell Disease, Oct 2019. URL https://www.cdc.gov/ncbddd/sicklecell/data.html. Accessed: 2020-04-10.

[59] Incidence of Sickle Cell Trait in the US, Oct 2019. URL https://www.cdc.gov/ncbddd/sicklecell/features/keyfinding-trait.html. Accessed: 2020-04-10.

[60] Multiple myeloma - statistics, Mar 2020. URL https://www.cancer.net/cancer-types/multiple-myeloma/statistics. Accessed: 2020-05-29.

[61] Ewing Sarcoma - Childhood and Adolescence - Statistics., Feb 2020. URL https://www.cancer.net/cancer-types/ewing-sarcoma-childhood-and-adolescence/statistics. Accessed: 2020-04-10.

[62] What is the mortality rate for diabetic neuropathy?, Jan 2020. URL https://www.medscape.com/answers/1170337-4921/what-is-the-mortality-rate-for-diabetic-neuropathy. Accessed: 2020-04-10.

[63] Rosa M Abrantes-Metz, Christopher Adams, and Albert D Metz. Pharmaceutical development phases: a duration analysis. *Journal of Pharmaceutical Finance, Economics and Policy*, 14:19–42, 2005.

[64] Sarah Acaster, Binny Pinder, Clara Mukuria, and Amanda Copans. Mapping the EQ-5D index from the cystic fibrosis questionnaire-revised using multiple modelling approaches. *Health and Quality of Life Outcomes*, 13(1):33, 2015.

[65] Frank J. Accurso. 89 - Cystic Fibrosis. In Lee Goldman and Andrew I. Schafer, editors, *Goldman's Cecil Medicine (Twenty Fourth Edition)*, pages 544 – 548. W.B. Saunders, Philadelphia, twenty fourth edition edition, 2012. ISBN 978-1-4377-1604-7. doi: 10.1016/B978-1-4377-1604-7.00089-0. URL http://www.sciencedirect.com/science/article/pii/B9781437716047000890. Accessed: 2020-04-10.

[66] MP Adam, HH Ardinger, RA Pagon, SE Wallace, LJH Bean, K Stephens, and A Amemiya. X-Linked Adrenoleukodystrophy–GeneReviews®.

[67] Michelle Andrews. Staggering prices slow insurers' coverage of car-t cancer therapy, Jul 2018. URL https://khn.org/news/staggering-prices-slow-insurers-coverage-of-car-t-cancer-therapy/. Accessed: 2020-06-28.

[68] O'connell Ann Meeker, Abruzzini Anthony F., Hamill Caitilin, and Zakar Jessica. Global approaches to drug development: When ex-us clinical data can support US drug approvals, 2019. URL https://www.iqvia.com/library/white-papers/global-approaches-to-drug-development. Accessed: 2020-05-20.

[69] American Diabetes Association. Statistics About Diabetes. URL http://www.diabetes.org/diabetes-basics/statistics/. Accessed: 2020-04-10.

[70] David I Auerbach and Arthur L Kellermann. A decade of health care cost growth has wiped out real income gains for an average US family. *Health Affairs*, 30(9):1630–1636, 2011.

[71] Jane F Barlow, Mo Yang, and J Russell Teagarden. Are payers ready, willing, and able to provide access to new durable gene therapies? *Value in Health*, 22(6):642–647, 2019.

[72] Troyen A Brennan and James M Wilson. The special case of gene therapy pricing. *Nature biotechnology*, 32(9):874–876, 2014.

[73] Anh L Bui, Tamara B Horwich, and Gregg C Fonarow. Epidemiology and risk profile of heart failure, Jan 2011. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3033496/. Accessed: 2020-04-10.

[74] U.S. Census Bureau. U.S. Census Bureau QuickFacts: United States, Jan 2019. URL https://www.census.gov/quickfacts/fact/table/US/PST045219. Accessed: 2020-05-22.

[75] John R Burnett. Familial Lipoprotein Lipase Deficiency, Jun 2017. URL `https://www.ncbi.nlm.nih.gov/books/NBK1308/`. Accessed: 2020-04-10.

[76] Jaime Caro, Kristen Migliaccio-Walle, Khajak J Ishak, and Irina Proskorovsky. The morbidity and mortality following a diagnosis of peripheral arterial disease: long-term follow-up of a large database. *BMC Cardiovascular Disorders*, 5(1):14, 2005.

[77] Liz Carroll, Gary Benson, Jeremy Lambert, Khadra Benmedjahed, Marek Zak, and Xin Ying Lee. Real-world utilities and health-related quality-of-life data in hemophilia patients in France and the United Kingdom. *Patient Preference and Adherence*, 13: 941, 2019.

[78] James D Chambers, Ari D Panzer, David D Kim, Nikoletta M Margaretos, and Peter J Neumann. Variation in us private health plans' coverage of orphan drugs. *Am J Manag Care*, 25(10):508–512, 2019.

[79] Artur V Cideciyan. Leber congenital amaurosis due to RPE65 mutations and its treatment with gene therapy. *Progress in Retinal and Eye Research*, 29(5):398–427, 2010.

[80] Keziah Cook, Shaun P Forbes, Kelly Adamski, Janice J Ma, Anita Chawla, and Louis P Garrison Jr. Assessing the potential cost-effectiveness of a gene therapy for the treatment of hemophilia a. *Journal of Medical Economics*, 23(5):501–512, 2020.

[81] RF Cornell and AA Kassim. Evolving paradigms in the treatment of relapsed/refractory multiple myeloma: increased options and increased complexity. *Bone marrow transplantation*, 51(4):479–491, 2016.

[82] Comparative Effectiveness Public Advisory Council. ICER reports on spinal muscular. *PharmacoEconomics & Outcomes News*, 823:1–9, 2019.

[83] Adele D'Amico, Eugenio Mercuri, Francesco D Tiziano, and Enrico Bertini. Spinal muscular atrophy. *Orphanet journal of rare diseases*, 6(1):1–10, 2011.

[84] Joseph A DiMasi, Henry G Grabowski, and Ronald W Hansen. Innovation in the pharmaceutical industry: new estimates of r&d costs. *Journal of health economics*, 47: 20–33, 2016.

[85] Anahita Dua and Cheong J. Lee. Epidemiology of Peripheral Arterial Disease and Critical Limb Ischemia. *Techniques in Vascular and Interventional Radiology*, 19(2): 91–95, 2016. doi: 10.1053/j.tvir.2016.04.001.

[86] Steve Duff, Michael S Mafilios, Prajakta Bhounsule, and James T Hasegawa. The burden of critical limb ischemia: a review of recent literature. *Vascular health and risk management*, 15:187, 2019.

[87] Xiaojing Fan, Duolao Wang, Bruce Hellman, Mathieu F Janssen, Gerben Bakker, Rupert Coghlan, Amelia Hursey, Helen Matthews, and Ian Whetstone. Assessment of Health-Related Quality of Life between People with Parkinson's Disease and Non-Parkinson's: Using Data Drawn from the '100 for Parkinson's' Smartphone-Based Prospective Study. *International Journal of Environmental Research and Public Health*, 15(11):2538, 2018.

[88] FDA. Expanded access | information for industry, May 2019. URL `https://www.fda.gov/news-events/expanded-access/expanded-access-information-industry`. Accessed: 2020-04-10.

[89] Andrea Ferrari and Paola Collini. What is synovial sarcoma? URL `http://sarcomahelp.org/synovial-sarcoma.html`. Accessed: 2020-04-10.

[90] U.S. Food and Drug Administration. Step 4: Fda drug review, Apr 2018. URL `https://www.fda.gov/patients/drug-development-process/step-4-fda-drug-review`. Accessed: 2020-04-10.

[91] U.S. Food and Drug Administration. CFR - Code of Federal Regulations Title 21, Apr 2019. URL `https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?fr=314.101`. Accessed: 2020-05-04.

[92] Institute for Clinical and Economic Review. Adapted Value Assessment Methods for High-Impact "Single and Short-Term Therapies" (SSTs), Nov 2020. URL https://icer-review.org/wp-content/uploads/2019/01/ICER_SST_FinalAdaptations_111219.pdf.

[93] Centre for Clinical Practice at NICE (UK et al. Sickle Cell Acute Painful Episode: Management of an Acute Painful Sickle Cell Episode in Hospital. 2012.

[94] Centers for Disease Control and Prevention. Peripheral arterial disease (PAD). URL https://www.cdc.gov/heartdisease/PAD.htm. Accessed: 2020-04-10.

[95] Centers for Disease Control and Prevention. Diabetes Quick Facts, Aug 2019. URL https://www.cdc.gov/diabetes/basics/quick-facts.html. Accessed: 2020-04-10.

[96] MIT Laboratory for Financial Engineering. Estimates of Clinical Trial Probabilities of Success (PoS), Feb 2019. URL https://projectalpha.mit.edu/pos/. Accessed: 2020-04-10.

[97] Centers for Medicare & Medicaid Services. Medicaid drug rebate program, Nov 2018. URL https://www.medicaid.gov/medicaid/prescription-drugs/medicaid-drug-rebate-program/index.html. Accessed: 2020-09-29.

[98] Kaiser Family Foundation. Medicaid state fact sheets, May 2020. URL https://www.kff.org/interactive/medicaid-state-fact-sheets/. Accessed: 2020-08-09.

[99] Michael L Ganz, Sean Stern, Alex Ward, Luba Nalysnyk, Martin Selzer, Alaa Hamed, and Neal Weinreb. A new framework for evaluating the health impacts of treatment for Gaucher disease type 1. *Orphanet Journal of Rare Diseases*, 12(1):38, 2017.

[100] Kate Gardner, Abdel Douiri, Emma Drasar, Marlene Allman, Anne Mwirigi, Moji Awogbade, and Swee Lay Thein. Survival in adults with sickle cell disease in a high-income setting. *Blood, The Journal of the American Society of Hematology*, 128(10): 1436–1438, 2016.

[101] Annetine C Gelijns, Ethan A Halm, et al. The diffusion of new technology: Costs and benefits to health care. In *The Changing Economics of Medical Technology*. National Academies Press (US), 1991.

[102] Yezaz Ahmed Ghouri, Idrees Mian, and Julie H Rowe. Review of hepatocellular carcinoma: Epidemiology, etiology, and carcinogenesis. *Journal of carcinogenesis*, 16, 2017.

[103] Guzman Gloria. New data show income increased in 14 states and 10 of the largest metros, Sept 2019. URL `https://www.census.gov/library/stories/2019/09/us-median-household-income-up-in-2018-from-2017.html`. Accessed: 2020-06-27.

[104] Pegah Golabi, Sofie Fazel, Munkhzul Otgonsuren, Mehmet Sayiner, Cameron T Locklear, and Zobair M Younossi. Mortality assessment of patients with hepatocellular carcinoma according to underlying disease and treatment modalities. *Medicine*, 96(9), 2017.

[105] Kimberley A Goldsmith, Matthew T Dyer, Peter M Schofield, Martin J Buxton, and Linda D Sharples. Relationship between the EQ-5D index and measures of clinical outcomes in selected studies of cardiovascular interventions. *Health and quality of life outcomes*, 7(1):96, 2009.

[106] Luis Henrique Wolff Gowdak. Prevalence of refractory angina in clinical practice. *Heart Metabolism*, (72):9–12, 2017.

[107] Mohit Gupta. Developments in the Management of BCG-Unresponsive NMIBC, Jan 2019. URL `https://www.renalandurologynews.com/home/news/urology/bladder-cancer/developments-in-the-management-of-bcg-unresponsive-nmibc/`. Accessed: 2020-04-10.

[108] Christian J Hendriksz, Kenneth I Berger, Christina Lampe, Susanne G Kircher, Paul J Orchard, Rebecca Southall, Sarah Long, Stephen Sande, and Jeffrey I Gold. Health-related quality of life in mucopolysaccharidosis: looking beyond biomedical issues. *Orphanet Journal of Rare Diseases*, 11(1):119, 2016.

[109] Timothy D Henry, Daniel Satran, James S Hodges, Randall K Johnson, Anil K Poulose, Alex R Campbell, Ross F Garberich, Bradley A Bart, Rachel E Olson, Charlene R Boisjolie, et al. Long-term survival in patients with refractory angina. *European heart journal*, 34(34):2683–2688, 2013.

[110] Paul Hjemdahl, Sven V Eriksson, Claes Held, Lennart Forslund, Per Näsman, and Nina Rehnqvist. Favourable long term prognosis in stable angina pectoris: an extended follow up of the angina prognosis study in stockholm (apsis). *Heart*, 92(2):177–182, 2006.

[111] Sarah Houben-Wilke, Rudolf A Joerres, Robert Bals, Frits ME Franssen, Sven Gläser, Rolf Holle, Annika Karch, Armin Koch, Helgo Magnussen, Anne Obst, et al. Peripheral artery disease and its clinical relevance in patients with chronic obstructive pulmonary disease in the COPD and Systemic Consequences–Comorbidities Network Study. *American Journal of Respiratory and Critical Care Medicine*, 195(2):189–197, 2017.

[112] Mohammad Houshyari, Farzaneh Hajalikhani, Afshin Rakhsha, and Parastoo Hajian. A comparative study of survival rate in high grade glioma tumors being treated by radiotherapy alone versus chemoradiation with nitrosourea. *Global journal of health science*, 7(6):33, 2015.

[113] Inserm. Orphanet: Recessive dystrophic epidermolysis bullosa inversa. URL `https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Lng=GB&Expert=79409`. Accessed: 2020-04-10.

[114] Institute for Clinical and Economic Review. Spinraza® and zolgensma® for spinal muscular atrophy: effectiveness and value: final evidence report. 2019.

[115] Jonathan P Jarow, Steven Lemery, Kevin Bugin, Sean Khozin, and Richard Moscicki. Expanded access of investigational drugs: the experience of the center of drug evaluation and research over a 10-year period. *Therapeutic innovation & regulatory science*, 50(6):705–709, 2016.

[116] Tolbert Jennifer, Orgera Kendal, Singer Natalie, and Anthony Damico. Key facts about the uninsured population, Dec 2019. URL https://www.kff.org/uninsured/issue-brief/key-facts-about-the-uninsured-population/. Accessed: 2020-06-27.

[117] In Kyung Jeon, Hye Rang On, and Soo-Chan Kim. Quality of life and economic burden in recessive dystrophic epidermolysis bullosa. *Annals of Dermatology*, 28(1):6–14, 2016.

[118] de Jong, Joep J., Hendricksen, Kees, Hugh, and Joost L. Hyperthermic Intravesical Chemotherapy for BCG Unresponsive Non-Muscle Invasive Bladder Cancer Patients, Jan 2018. URL https://content.iospress.com/articles/bladder-cancer/blc180191. Accessed: 2020-04-10.

[119] Ashish M Kamat, Marc Colombel, Debasish Sundi, Donald Lamm, Andreas Boehle, Maurizio Brausi, Roger Buckley, Raj Persad, Joan Palou, Mark Soloway, et al. BCG-unresponsive non-muscle-invasive bladder cancer: recommendations from the IBCG. *Nature Reviews Urology*, 14(4):244–255, 2017.

[120] Eitan Kerem, Joseph Reisman, Mary Corey, Gerard J Canny, and Henry Levison. Prediction of mortality in patients with cystic fibrosis. *New England Journal of Medicine*, 326(18):1187–1191, 1992.

[121] Christine G Kohn, Matthew W Parker, Brendan L Limone, and Craig I Coleman. Impact of angina frequency on health utility values of patients with chronic stable angina. *Health and quality of life outcomes*, 12(1):39, 2014.

[122] Michel Lacroix, Dima Abi-Said, Daryl R Fourney, Ziya L Gokaslan, Weiming Shi, Franco DeMonte, Frederick F Lang, Ian E McCutcheon, Samuel J Hassenbusch, Eric Holland, et al. A multivariate analysis of 416 patients with glioblastoma multiforme: prognosis, extent of resection, and survival. *Journal of neurosurgery*, 95(2):190–198, 2001.

[123] Cathy Lally, Cynthia Jones, Wildon Farwell, Sandra P Reyna, Suzanne F Cook, and W Dana Flanders. Indirect estimation of the prevalence of spinal muscular atrophy

type i, ii, and iii in the United States. *Orphanet journal of rare diseases*, 12(1):175, 2017.

[124] Byron L Lam and Stephen R Russell. Voretigene Neparvovec for Biallelic RPE65-Mediated Retinal Disease: Effectiveness and Value.

[125] Christine Lavery, Chris J Hendriksz, and Simon A Jones. Mortality in patients with sanfilippo syndrome. *Orphanet journal of rare diseases*, 12(1):168, 2017.

[126] Richard Leong. Average U.S. mortgage size hits record-high $354,500 - mba, Mar 2019. URL `https://www.reuters.com/article/us-usa-mortgages/average-u-s-mortgage-size-hits-record-high-354500-mba-idUSKBN1QU1VA`. Accessed: 2020-04-10.

[127] Ian M MacDonald, Natalia Binczyk, Alina Radziwon, and Ioannis Dimopoulos. Choroideremia. In *Hereditary Chorioretinal Disorders*, pages 99–106. Springer, 2020.

[128] Asif Mahmood, Jay Berry, David A Wenger, Maria Escolar, Magdi Sobeih, Gerald Raymond, and Florian S Eichler. Metachromatic leukodystrophy: a case of triplets with the late infantile variant and a systematic review of the literature. *Journal of child neurology*, 25(5):572–580, 2010.

[129] David J Margolis, D Scot Malay, Ole J Hoffstad, Charles E Leonard, Thomas MaCurdy, Karla López de Nava, Yang Tan, Teresa Molina, and Karen L Siegel. Incidence of diabetic foot ulcer and lower extremity amputation among Medicare beneficiaries, 2006 to 2008. In *Data Points Publication Series [Internet]*. Agency for Healthcare Research and Quality (US), 2011.

[130] David J Margolis, D Scot Malay, Ole J Hoffstad, Charles E Leonard, Thomas MaCurdy, Karla López de Nava, Yang Tan, Teresa Molina, and Karen L Siegel. Prevalence of diabetes, diabetic foot ulcer, and lower extremity amputation among Medicare beneficiaries, 2006 to 2008. In *Data Points Publication Series [Internet]*. Agency for Healthcare Research and Quality (US), 2011.

[131] Benjamin J Miller, Charles F Lynch, and Joseph A Buckwalter. Conditional survival is greater than overall survival at diagnosis in patients with osteosarcoma and ewing's sarcoma. *Clinical Orthopaedics and Related Research®*, 471(11):3398–3404, 2013.

[132] John Miller and Caroline Humer. Novartis $2 million gene therapy for rare disorder is world's most expensive drug, May 2019. URL `https://www.reuters.com/article/us-novartis-genetherapy/novartis-2-million-gene-therapy-for-rare-disorder-is-worlds-most-expensive-drug-idUSKCN1SU1ZP`. Accessed: 2020-06-28.

[133] Vahid Montazerhodjat, David Weinstock, and Andrew W. Lo. Buying Cures vs. Renting Health: Financing Healthcare via Consumer Loans. *Science Translational Medicine*, 8:327ps6, 2016.

[134] S Muraki, T Akune, H Oka, Y En-Yo, M Yoshida, A Saika, T Suzuki, H Yoshida, H Ishibashi, F Tokimura, et al. Association of radiographic and symptomatic knee osteoarthritis with health-related quality of life in a population-based cohort study in Japan: the ROAD study. *Osteoarthritis and Cartilage*, 18(9):1227–1234, 2010.

[135] P. J. Neumann, J. T. Cohen, and M. C. Weinstein. Updating cost-effectiveness—the curious resilience of the $50,000-per-qaly threshold. *New England Journal of Medicine*, 371(9):796–797, 2014. URL `https://doi.org/10.1056/NEJMp1405158`.

[136] Congressional Budget Office. The budget and economic outlook: 2019 to 2029, 2019.

[137] Orphanet. Beta thalassemia major. URL `https://www.orpha.net/consor/cgi-bin/OC_Exp.php?Lng=EN&Expert=231214`. Accessed: 2020-05-04.

[138] Joohyun Park and Kevin A Look. Health care expenditure burden of cancer care in the United States. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 56:0046958019880696, 2019.

[139] Rui Pinto, Carla Caseiro, Manuela Lemos, Lurdes Lopes, Augusta Fontes, Helena Ribeiro, Eugénia Pinto, Elisabete Silva, Sonia Rocha, Ana Marcao, et al. Prevalence

of lysosomal storage diseases in Portugal. *European Journal of Human Genetics*, 12 (2):87–92, 2004.

[140] Iris Plug, JG Van Der Bom, Marjolein Peters, EP Mauser-Bunschoten, Arja DE GOEDE-BOLDER, Lily Heijnen, Cees Smit, José Willemse, and FR Rosendaal. Mortality and causes of death in patients with hemophilia, 1992–2001: a prospective cohort study 1. *Journal of Thrombosis and Haemostasis*, 4(3):510–516, 2006.

[141] Mariano Provencio, Dolores Isla, Antonio Sánchez, and Blanca Cantos. Inoperable stage iii non-small cell lung cancer: Current treatment and role of vinorelbine. *Journal of thoracic disease*, 3(3):197, 2011.

[142] Casey Quinn, Colin Young, Jonathan Thomas, Mark Trusheim, et al. Estimating the clinical pipeline of cell and gene therapies and their potential economic impact on the US healthcare system. *Value in Health*, 22(6):621–626, 2019.

[143] Gerald V Raymond, Patrick Aubourg, Asif Paker, Maria Escolar, Alain Fischer, Stephane Blanche, André Baruchel, Jean-Hugues Dalle, Gérard Michel, Vinod Prasad, et al. Survival and functional outcomes in boys with cerebral adrenoleukodystrophy with and without hematopoietic stem cell transplantation. *Biology of Blood and Marrow Transplantation*, 25(3):538–548, 2019.

[144] Genetics Home Reference. Retinitis pigmentosa. URL `https://ghr.nlm.nih.gov/condition/retinitis-pigmentosa`. Accessed: 2020-04-10.

[145] Peter Reichardt, Michael Leahy, Xavier Garcia del Muro, Stefano Ferrari, Javier Martin, Hans Gelderblom, Jingshu Wang, Arun Krishna, Jennifer Eriksson, Arthur Staddon, et al. Quality of life and utility in patients with metastatic soft tissue and bone sarcoma: the sarcoma treatment and burden of illness in North America and Europe (SABINE) study. *Sarcoma*, 2012, 2012.

[146] Jeffrey M Robbins, Gerald Strauss, David Aron, Jodi Long, Jennifer Kuba, and Yelena Kaplan. Mortality rates and diabetic foot ulcers: is it time to communicate mortality

risk to patients with diabetic foot ulceration? *Journal of the American Podiatric Medical Association*, 98(6):489–493, 2008.

[147] Sanjoy Roy, Debarshi Lahiri, Tapas Maji, and Jaydip Biswas. Recurrent glioblastoma: where we stand. *South Asian journal of cancer*, 4(4):163, 2015.

[148] Alette Ruarus, Laurien Vroomen, Robbert Puijk, Hester Scheffer, and Martijn Meijerink. Locally Advanced Pancreatic Cancer: A Review of Local Ablative Therapies, Jan 2018. URL `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5789366/`. Accessed: 2020-04-10.

[149] Ameet Sarpatwari, Jonathan DiBello, Marie Zakarian, Mehdi Najafzadeh, and Aaron S Kesselheim. Competition and price among brand-name drugs in the same class: A systematic review of the evidence. *PLoS medicine*, 16(7):e1002872, 2019.

[150] Bjoern Schwander. Early health economic evaluation of the future potential of next generation artificial vision systems for treating blindness in Germany. *Health Economics Review*, 4(1):27, 2014.

[151] Meysam Seyedifar, Farid Abedin Dorkoosh, Amir Ali Hamidieh, Majid Naderi, Hossein Karami, Mehran Karimi, Masoomeh Fadaiyrayeny, Masoumeh Musavi, Sanaz Safaei, Mohammad Mahdi Ahmadian-Attari, et al. Health-related quality of life and health utility values in beta thalassemia major patients receiving different types of iron chelators in Iran. *International Journal of Hematology-Oncology and Stem Cell Research*, 10(4):224, 2016.

[152] James Shearer, Colin Green, Carl E Counsell, and John P Zajicek. The impact of motor and non motor symptoms on health state values in newly diagnosed idiopathic Parkinson's disease. *Journal of Neurology*, 259(3):462–468, 2012.

[153] Jayne Smith-Palmer, Jay P Bae, Kristina S Boye, Kirsi Norrbacka, Barnaby Hunt, and William J Valentine. Evaluating health-related quality of life in type 1 diabetes: a systematic literature review of utilities for adults with type 1 diabetes. *ClinicoEconomics and outcomes research: CEOR*, 8:559, 2016.

[154] Jin Sothornwit, Gulapar Srisawasdi, Atchara Suwannakin, and Apiradee Sriwijitkamol. Decreased health-related quality of life in patients with diabetic foot problems. *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, 11:35, 2018.

[155] J Richard Steadman, Karen K Briggs, Lauren M Matheny, and Henry B Ellis. Ten-year survivorship after knee arthroscopy in patients with kellgren-lawrence grade 3 and grade 4 osteoarthritis of the knee. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, 29(2):220–225, 2013.

[156] Sullivan Thomas. How are insurers treating the $2m drug, zolgensma?, Oct 2019. URL `https://www.policymed.com/2019/10/how-are-insurers-treating-the-2m-drug-zolgensma.html`. Accessed: 2020-06-27.

[157] Meg Tirrell. A US drugmaker offers to cure rare blindness for $850,000, Jan 2019. URL `https://www.cnbc.com/2018/01/03/spark-therapeutics-luxturna-gene-therapy-will-cost-about-850000.html`. Accessed: 2020-06-28.

[158] Soili Törmälehto, Mika E Mononen, Emma Aarnio, Jari PA Arokoski, Rami K Korhonen, and Janne Martikainen. Health-related quality of life in relation to symptomatic and radiographic definitions of knee osteoarthritis: data from Osteoarthritis Initiative (OAI) 4-year follow-up study. *Health and Quality of Life Outcomes*, 16(1):154, 2018.

[159] John Tozzi. Employers fear squeeze from genetic cures that cost millions, September 2019. URL `https://www.bloomberg.com/news/articles/2019-09-11/employers-fear-squeeze-from-genetic-cures-that-cost-millions`. Accessed: 2020-04-10.

[160] Vascular Surgery Unit. Atherosclerosis obliterans of the lower extremities in thai patients. *J Med Assoc Thai*, 89(10):1612–20, 2006.

[161] U.S. Department of Health and Human Services. NIH Research Portfolio Online Reporting Tools (RePORT). URL `https://report.nih.gov/nihfactsheets/viewfactsheet.aspx?csid=55`. Accessed: 2019-06-10.

[162] Belinda van Zyl, Denise Tang, and Nikola A Bowden. Biomarkers of platinum resistance in ovarian cancer: what can we use to improve treatment. *Endocrine-related cancer*, 25(5):R303–R318, 2018.

[163] Peter Vorlat, Guy Putzeys, Dominique Cottenie, Tom Van Isacker, Nicole Pouliart, Frank Handelberg, Pierre-Paul Casteleyn, Filip Gheysen, and René Verdonk. The oxford unicompartmental knee prosthesis: an independent 10-year survival analysis. *Knee Surgery, Sports Traumatology, Arthroscopy*, 14(1):40–45, 2006.

[164] Joseph Walker and Anna Wilde Mathews. Insurers Pitch New Ways to Pay for Million-Dollar Therapies, Sep 2019. URL `https://www.wsj.com/articles/insurers-pitch-new-ways-to-pay-for-million-dollar-therapies-11567677600`. Accessed: 2020-05-04.

[165] Allison W Willis, Mario Schootman, Nathan Kung, Bradley A Evanoff, Joel S Perlmutter, and Brad A Racette. Predictors of survival in patients with parkinson disease. *Archives of neurology*, 69(5):601–607, 2012.

[166] Chi Heem Wong, Kien Wei Siah, and Andrew W Lo. Estimation of clinical trial success rates and related parameters. *Biostatistics*, 20(2):273–286, 2019.

[167] Jing Wu, Yuerong Han, Judy Xu, Yang Lu, Hongliang Cong, Junyi Zheng, and He Sun. Chronic stable angina is associated with lower health-related quality of life: evidence from Chinese patients. *PLoS One*, 9(5), 2014.

[168] M. J. Young, A. J. M. Boulton, A. F. Macleod, D. R. R. Williams, and P. H. Sonksen. A multicentre study of the prevalence of diabetic peripheral neuropathy in the United Kingdom hospital clinic population. *Diabetologia*, 36(2):150–154, 1993. doi: 10.1007/bf00400697.

[169] Reza Zamani, Salman Khazaei, and Shahab Rezaeian. Survival analysis and its associated factors of beta thalassemia major in hamadan province. *Iranian journal of medical sciences*, 40(3):233, 2015.

[170] Klaus Zerres and Sabine Rudnik-Schöneborn. Natural history in proximal spinal muscular atrophy: clinical analysis of 445 patients and suggestions for a modification of existing classifications. *Archives of neurology*, 52(5):518–523, 1995.

[171] Audrey D Zhang, Jeremy Puthumana, Nicholas S Downing, Nilay D Shah, Harlan M Krumholz, and Joseph S Ross. Assessment of clinical trials supporting US food and drug administration approval of novel therapeutic agents, 1995-2017. *JAMA Network Open*, 3(4):e203284–e203284, 2020.

[172] Santiago Zuluaga-Sanchez, Megan Teynor, Christopher Knight, Robin Thompson, Thomas Lundqvist, Mats Ekelund, Annabelle Forsmark, Adrian D Vickers, and Andrew Lloyd. Cost effectiveness of nusinersen in the treatment of patients with infantile-onset and later-onset spinal muscular atrophy in Sweden. *PharmacoEconomics*, 37(6): 845–865, 2019.

4.6 Supplementary Materials

4.6.1 Current Gene Therapy Clinical Trials

Here, we list the clinical trials that are used in this study in the following table.

Table 4.11: List of clinical trials used in this study. 'TT' and 'CT' indicates 'TrialTrove' and '*clinicaltrials.gov*' respectively.

Trial Title	**Disease**	**Sponsors**	**Source**
Randomized, double-blind, placebo-controlled study of AMG0(HGF plasmid) for patients with arteriosclerosis obliterans	Arteriosclerosis Obliterans	AnGes	TT
Tisagenlecleucel Versus Standard of Care in Adult Patients With Relapsed or Refractory Aggressive B-cell Non-Hodgkin Lymphoma: A Randomized, Open Label, Phase III Trial (BELINDA)	B-Cell Non-Hodgkin's Lymphoma	Novartis	TT
A Global Randomized Multicenter Phase III Trial of JCAR017 Compared to Standard of Care in Adult Subjects With High-risk, Second-line, Transplant-eligible Relapsed or Refractory Aggressive B-cell Non-Hodgkin Lymphomas (TRANSFORM).	B-Cell Non-Hodgkin's Lymphoma	Celgene	TT
A Phase III, Open Label Study to Evaluate the Safety and Efficacy of INSTILADRIN (rAd-IFN)/Syn3) Administered Intravesically to Patients With High Grade, BCG Unresponsive Non-Muscle Invasive Bladder Cancer (NMIBC)	BCG Unresponsive NMIBC	FKD Therapeutics	TT
A Phase III Study of BC-819 in Patients with Bladder Cancer who Failed Initial Treatment of BCG	BCG Unresponsive NMIBC	Anchiano Therapeutics	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Phase 3 Single Arm Study Evaluating the Efficacy and Safety of Gene Therapy in Subjects With Transfusion-dependent beta-Thalassemia, Who do Not Have a beta0/beta0 Genotype, by Transplantation of Autologous CD34+ Stem Cells Transduced Ex Vivo With a Lentiviral betaA-T87Q-Globin Vector in Subjects < or = 50 Years of Age	Beta-Thalassemia	bluebird bio	TT
A Phase 3 Single Arm Study Evaluating the Efficacy and Safety of Gene Therapy in Subjects With Transfusion-dependent beta-Thalassemia, Who Have a beta0/beta0 Genotype, by Transplantation of Autologous CD34+ Stem Cells Transduced Ex Vivo With a Lentiviral betaA-T87Q-Globin Vector in Subjects < or = 50 Years of Age	Beta-Thalassemia	bluebird bio	TT
An Integrated Phase II/III, Open Label, Randomized and Controlled Study of the Safety and Efficacy of CG0070 Adenovirus Vector Expressing GM-CSF in Patients With Non-Muscle Invasive Bladder Cancer With Carcinoma In Situ Disease Who Have Failed BCG Bladder Oncolytic virus for Non-muscle invasive bladder cancer Disease (BOND)	Bladder Cancer, in situ concurrent with Papillary Tumors	Cold Genesys	TT
A Phase 2/3 Study of the Efficacy and Safety of Hematopoietic Stem Cells Transduced With Lenti-D Lentiviral Vector for the Treatment of Cerebral Adrenoleukodystrophy (CALD)	Cerebral Adrenoleukodystrophy (CALD)	bluebird bio	CT
Efficacy and Safety of AAV2-REP1 for the Treatment of Choroideremia	Choroideremia	Nightstar Therapeutics	CT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Phase 3 Double-Blind, Randomized, Placebo-Controlled Study to Evaluate the Safety and Efficacy of AMG0 in Subjects With Critical Limb Ischemia Efficacy and Safety of AMG0 in Subjects With Critical Limb Ischemia (AGILITY)	Critical Limb Ischemia	AnGes	TT
Safety and Efficacy of Recombinant Adeno-Associated Virus Containing the CFTR Gene in the Treatment of Cystic Fibrosis	Cystic Fibrosis	Targeted Genetics Corporation/ Cystic Fibrosis Foundation Therapeutics	CT
A Placebo Controlled, Double-blind, Randomized, Parallel-group, Multi-center Phase III study to determine the Efficacy and Safety of TisssueGene-C in Patients with Degenerative Arthritis	Degenerative Arthritis	Kolon Life Science	TT
Safety and Efficacy Study of Pl-VEGF165 to Treat Diabetic Foot Syndrome	Diabetic Foot Syndrome	Human Stem Cells Institute	TT
A Phase III, Double-blind, Randomized, Placebo-controlled, Multicenter Study to Asses the Safety and Efficacy of VM202 to Treat Chronic Nonhealing Foot Ulcers in Diabetic Patients With Concomitant Peripheral Arterial Disease (PAD)	Diabetic Foot Ulcers	Helixmith	TT
A Phase III, Double-Blind, Randomized, Placebo-Controlled, Multicenter Study to Assess the Safety and Efficacy of VM202 in Subjects With Painful Diabetic Peripheral Neuropathy	Diabetic Peripheral Neuropathy	Helixmith	TT
A Phase III, Randomized, Open-Label Study Evaluating Efficacy of Axicabtagene Ciloleucel Versus Standard of Care Therapy in Subjects With Relapsed/Refractory Diffuse Large B Cell Lymphoma	Diffuse Large B Cell Lymphoma (DLBCL)	Gilead Sciences/Kite Pharma	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Multi-center Phase III, Randomized, Open-Label Trial of Vigil (Bi-shRNAfurin and GMCSF Augmented Autologous Tumor Cell Immunotherapy) in Combination With Irinotecan and Temozolomide as a Second-Line Regimen for Ewing's Sarcoma	Ewing's Sarcoma	Gradalis	TT
A Phase III Study of INGN 241 in Combination with Radiation Therapy in Patients with Advanced Solid Tumors and Head and neck cancer.	Head and Neck Cancer	Introgen Therapeutics	TT
An Open-Label, Randomized, Multi-Center Phase III Clinical Trial Comparing E10A Plus Chemotherapy And Chemotherapy Alone For Treatment Of Head And Neck Cancer	Head and Neck Cancer	Marsala Biotech	TT
A Randomized, Open-label, Multi-center Phase III Study Designed to Evaluate the Safety and Efficacy of E10A in Patients With Recurrent/Unresectable Squamous Cell Carcinoma of the Head and Neck Region	Head and Neck Cancer	Guangzhou Double Bioproducts Co.	TT
A Phase III, Pivotal, Randomized, Placebo-controlled, Double-Blind, Multicenter Study to Evaluate RT-100 AC6 Gene Transfer in Patients with Heart Failure and Reduced Left Ventricular Ejection Fraction;Heart Failure with Reduced Left Ventricular Ejection Fraction: One-time Gene Transfer Using RT-100 Intracoronary Administration of Adenovirus 5 encoding Human AC6 (FLOURISH)	Heart Failure	Renova Therapeutics	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Phase 3 Open-Label, Single-Arm Study To Evaluate The Efficacy and Safety of BMN 270, an Adeno-Associated Virus Vector-Mediated Gene Transfer of Human Factor VIII in Hemophilia A Patients With Residual FVIII Levels = 1 IU/dL Receiving Prophylactic FVIII Infusions	Hemophilia A	BioMarin	TT
Phase 3 Study To Evaluate Efficacy/Safety of Valoctocogene Roxaparvovec an AAV Vector-Mediated Gene Transfer of hFVIII at a Dose of 4E13vg/kg in Hemophilia A Patients With Residual FVIII Levels < or = 1IU/dL Receiving Prophylactic FVIII Infusions	Hemophilia A	BioMarin	TT
A Phase III Run In trial to Evaluate SPK-8011 in Patients with Hemophilia A	Hemophilia A	Roche/Spark Therapeutics	TT
An open-label, single-dose, multi-center, multi-national, Phase III pivotal trial to investigate efficacy and safety of AMT-061 in severe or moderately severe hemophilia B; HOPE-B: Trial of AMT-061 in Severe or Moderately Severe Hemophilia B Patients; Phase III, Open-label, Single-dose, Multi-center, Multinational Trial Investigating a Serotype 5 Adeno-associated Viral Vector Containing the Padua Variant of a Codon-optimized Human Factor IX Gene (AAV5-hFIXco-Padua, AMT-061) Administered to Adult Subjects With Severe or Moderately Severe Hemophilia B	Hemophilia B	uniQure	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Pivotal Phase III Study of PF-06838435 in Patients with Hemophilia B; Phase 3, Open Label, Single Arm Study To Evaluate Efficacy And Safety Of Fix Gene Transfer With Pf-06838435 (Raav-Spark100-Hfix-Padua) In Adult Male Participants With Moderately Severe To Severe Hemophilia B (Fix:C < or =2%)	Hemophilia B	Pfizer	TT
A Pivotal Phase III Study to Evalaute AMT-060 in Patients with Hemophilia B	Hemophilia B	uniQure	TT
Multicenter Randomized Controlled Trial of Adenovirus-mediated Adjuvant Gene Therapy Improving Outcome of Liver Transplantation in Patients With Advanced Hepatocellular Carcinoma	Hepatocellular Carcinoma	Wuhan Tiandakang Bio-Tech Engineering Co./ Shenzhen Tiandakang Gene Engineering Co.	TT
A Phase III Randomized, Open-Label Study Comparing Pexa Vec (Vaccinia GM CSF / Thymidine Kinase-Deactivated Virus) Followed by Sorafenib Versus Sorafenib in Patients With Advanced Hepatocellular Carcinoma (HCC) Without Prior Systemic Therapy	Hepatocellular Carcinoma	Transgene/ Sillajen Biotherapeutics /Jennerex/ Lees Pharmaceutical	TT
Phase III Prospective, Open-Label, Parallel-Group, Randomized, Multicenter Trial Comparing the Efficacy of Surgery, Radiation, and Injection of Murine Cells Producing Herpes Simplex Thymidine Kinase Vector Followed by Intravenous Ganciclovir Against the Efficacy of Surgery and Radiation in the Treatment of Newly Diagnosed, Previously Untreated Glioblastoma Multiforme	High-Grade Glioma	Novartis/Sandoz	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Controlled, Randomised, Parallel Group Study Of The Efficacy And Safety Of Herpes Simplex Virus Thymidine Kinase Gene Therapy (Cerepro) with Subsequent Ganciclovir For The Treatment Of Patients With Operable High-Grade Glioma.	High-Grade Glioma	Trizell	TT
A Randomized, Double-Blind, Placebo-Controlled, Multi-Center, Phase 3 Study to Determine the Efficacy of TG-C in Subjects With Kellgren and Lawrence Grade (KLG) 2 or 3 Osteoarthritis of the Knee	Knee Osteoarthritis with Kellgren & Lawrence Grade 2 or 3	Kolon TissueGene	TT
A Multicenter, Randomized, Placebo Controlled, Double-blind, Parallel, Phase III Clinical Trial to Evaluate the Efficacy and Safety of Invossa K Injection in Patients Diagnosed as Knee Osteoarthritis With Kellgren & Lawrence Grade 2	Knee Osteoarthritis with Kellgren & Lawrence Grade 2 or 3	Kolon Life Science	TT
Safety and Efficacy Study in Subjects With Leber Congenital Amaurosis	Leber Congenital Amaurosis due to RPE65 Mutations	Spark Therapeutics	CT
Efficacy Study of GS010 for Treatment of Vision Loss From 7 Months to 1 Year From Onset in LHON Due to the ND4 Mutation	Leber Hereditary Optic Neuropathy	GenSight Biologics	CT
Efficacy Study of GS010 for the Treatment of Vision Loss up to 6 Months From Onset in LHON Due to the ND4 Mutation	Leber Hereditary Optic Neuropathy	GenSight Biologics	CT
Efficacy and Safety Study of Bilateral Intravitreal Injection of GS010 for the Treatment of Vision Loss up to 1 Year From Onset in LHON Due to the ND4 Mutation	Leber Hereditary Optic Neuropathy	GenSight Biologics	CT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
Tisagenlecleucel Versus Blinatumomab or Inotuzumab for Adult Patients With Relapsed/Refractory B-cell Precursor Acute Lymphoblastic Leukemia: A Randomized Open Label, Multicenter, Phase III Trial	Leukemia (Acute Lymphoblastic)	Novartis	TT
Phase IIIb Study for Relapsed/Refractory Pediatric/Young Adult Acute Lymphoblastic Leukemia Patients to be Treated With CTL019	Leukemia (Acute Lymphoblastic)	Novartis	TT
A Phase II/III Prospective, Open Label Study to Evaluate Safety and Efficacy of Intravenous Autologous CD19 CAR-T Cells for Relapsed/ Refractory B-Acute Lymphoblastic Leukemia	Leukemia (Acute Lymphoblastic)	Gaia Science	TT
A Randomized Phase II/III Study of $\alpha\beta$ T Cell-Depleted, Related, Haploidentical Hematopoietic Stem Cell Transplant (Haplo-HSCT) Plus Rivogenlecleucel vs. Haplo-HSCT Plus Post-Transplant Cyclophosphamide (PTCy) in Patients With AML or MDS	Leukemia (Acute Myelogenous)	Bellicum Pharmaceuticals	TT
Randomized, Registrational, Controlled Study of BPX-501 with Allogeneic Hematopoietic Stem Cells (Allo-HSCT) in Patients with Acute Myelogenous Leukemia	Leukemia (Acute Myelogenous)	Bellicum Pharmaceuticals	TT
TK008: Randomized Phase III Trial of Haploidentical HCT With or Without an Add Back Strategy of HSV-Tk Donor Lymphocytes in Patients With High Risk Acute Leukemia	Leukemia (Acute Myelogenous)	Molmed	TT
A Phase IIb/III Study of AST-VAC1 in Patients with Acute Myelogenous Leukemia (AML)	Leukemia (Acute Myelogenous)	Asterias/Lineage Cell Therapeutics	TT
A Study of Glybera for the Treatment of Lipoprotein Lipase (LPL) Deficiency	Lipoprotein Lipase Deficiency (LPLD)	uniQure	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Study to Determine the Safety and Efficacy in Lipoprotein Lipase-Deficient Subjects After Intramuscular Administration of AMT-011, an Adeno-Associated Viral Vector Expressing Human Lipoprotein LipaseS447X.	Lipoprotein Lipase Deficiency (LPLD)	uniQure	TT
An Open-label Study to Assess the Efficacy and Safety of Alipogene Tiparvovec (AMT-011), Human LPL [S447X], Expressed by an Adeno-Associated Viral Vector After Intramuscular Administration in LPL-deficient Adult Subjects	Lipoprotein Lipase Deficiency (LPLD)	uniQure	TT
A Study of AMT-011 in Patients With LPL Deficiency	Lipoprotein Lipase Deficiency (LPLD)	uniQure	TT
A Phase III Trial of Glybera for Dyslipidemia	Lipoprotein Lipase Deficiency (LPLD)	uniQure	TT
Phase II/III study of Ad-IFNg in Cutaneous T-cell lymphoma	Lymphoma	Transgene	TT
A Safety and Efficacy Study of Cryopreserved GSK2696274 for Treatment of Metachromatic Leukodystrophy (MLD)	Metachromatic Leukodystrophy	GlaxoSmithKline	CT
PV-10 Intralesional Injection vs Systemic Chemotherapy or Oncolytic Viral Therapy for Treatment of Locally Advanced Cutaneous Melanoma	Melanoma (Locally Advanced Cutaneous)	Provectus Biopharmaceuticals	TT
A Phase Ib/III, Multicenter, Trial of Talimogene Laherparepvec in Combination With Pembrolizumab (MK-3475) for Treatment of Unresectable Stage IIIB to IVM1c Melanoma (MASTERKEY-265/KEYNOTE-034)	Melanoma (Metastatic)	Amgen/ Merck & Co./Merck Sharp & Dohme (MSD)	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Phase III Clinical Trial to Evaluate the Safety and Efficacy of Treatment With 2 mg Intralesional Allovectin-7 Compared to Dacarbazine (DTIC) or Temozolomide (TMZ) in Subjects With Recurrent Metastatic Melanoma;Allovectin-7 Immunotherapeutic for Metastatic Melanoma (AIMM).	Melanoma (Metastatic)	Brickell Biotech, AnGes	TT
A Randomized Phase III Clinical Trial to Evaluate the Efficacy and Safety of Treatment With OncoVEXGM-CSF Compared to Subcutaneously Administered GM-CSF in Melanoma Patients With Unresectable Stage IIIb, IIIc and IV Disease	Melanoma (Metastatic)	Amgen	TT
An Extension Protocol to Evaluate the Efficacy and Safety of Extended Use Treatment With OncoVEXGM-CSF for Eligible Melanoma Patients Participating in Study 005/05	Melanoma (Metastatic)	Amgen	TT
A Controlled, Randomized Phase III Trial Comparing the Response to Dacarbazine With and Without Allovectin-7 in Patients With Metastatic Melanoma.	Melanoma (Metastatic)	Brickell Biotech	TT
Open-label, Single-arm, Multi-center Study of Intracerebral Administration of Adeno-associated Viral (AAV) Serotype rh.10 Carrying Human N-sulfoglucosamine Sulfohydrolase (SGSH) cDNA for Treatment of Mucopolysaccharidosis Type IIIA	Mucopolysaccharidosis Type IIIa	LYSOGENE	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Phase III, Single Arm, Multi-Center Study to Assess the Efficacy and Safety of Clarithromycin(Biaxin)-Lenalidomide-Low-Dose-Dexamethasone (BiRd) Combined With B-cell Muturation Antigen (BCMA)-Directed Chimeric Antigen Receptor (CAR) T-cell Therapy in Patients With Newly Diagnosed Multiple Myeloma	Multiple Myeloma (Newly Diagnosed)	Shanghai Unicar-Therapy Bio-medicine	TT
Clinical Trial of Recombinant Adenovirus-p53 (Gendicine) Combined with Radiotherapy in Nasopharyngeal Carcinoma Patients.;	Nasopharyngeal Carcinoma	Shenzhen SiBiono GeneTech Co.	TT
A Phase II/III, Multi-Center, Open-Label, Randomized Study to Compare the Effectiveness and Safety of Intralesional Administration of RPR/INGN 201 in Combination with Taxotere and Carboplatin and Radiotherapy Versus Taxotere and Carboplatin and Radiotherapy Alone in Patients with Locally Advanced Unresectable Non-Small Cell Lung Cancer (NSCLC)	NSCLC	Introgen Therapeutics	TT
A Phase IIB/III Randomized, Double-blind, Placebo Controlled Study Comparing First Line Therapy With or Without TG4010 Immunotherapy Product in Patients With Stage IV Non-Small Cell Lung Cancer (NSCLC)	NSCLC	Transgene	TT
Phase III multi-center, open, randomized clinical trial of percutaneous intratumoral injection of genetically engineered adenovirus (injection of H101), IL-2, TB hyperthermia and systemic chemotherapy in the treatment of advanced non-small cell lung cancer	NSCLC	Shanghai Sunway Biotech	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
Phase III Study of Lucanix (Belagenpumatucel-L) in Advanced Non-small Cell Lung Cancer: An International Multicenter, Randomized, Double-blinded, Placebo-controlled Study of Lucanix Maintenance Therapy for Stages III/IV NSCLC Subjects Who Have Responded to or Have Stable Disease Following One Regimen of Front-line, Platinum-based Combination Chemotherapy; Survival, Tumor-free, Overall and Progression-free (STOP)	NSCLC Stage 3	Activate Immunotherapy	TT
rAd-p53 Combined Chemotherapy Via Selective Arterial Cannula in The Treatment of Advanced Oral Cancer, A Randomized Controlled Trial	Oral Cancer (Advanced)	Shenzhen SiBiono GeneTech Co.	TT
A Randomized, Controlled, Double-Arm, Open-Label, Multi-Center Study of Ofranergene Obadenovec (VB-111) Combined With Paclitaxel vs. Paclitaxel Monotherapy for the Treatment of Recurrent Platinum-Resistant Ovarian Cancer	Ovarian Cancer (Platinum-Resistant)	Gynecologic Oncology Group (GOG)/ VBL Therapeutics	TT
A Phase II/III Trial of Chemotherapy Alone Versus Chemotherapy Plus SCH 58500 in Newly Diagnosed Stage III Ovarian and Primary Peritoneal Cancer Patients With Greater Than or Equal to 0.5 cm and Less Than or Equal to 2 cm Residual Disease Following Surgery	Ovarian Cancer, Primary Peritoneal Cavity Cancer	Merck & Co./Merck Sharp & Dohme (MSD)	TT
A Randomized, Phase II/III, Study of TNFerade Biologic With 5-FU and Radiation Therapy for First-Line Treatment of Unresectable Locally Advanced Pancreatic Cancer	Pancreatic Cancer (Locally Advanced)	Precigen	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
Phase II/III Study of ProSavin for the Treatment of Parkinson's Disease	Parkinson's Disease	Oxford BioMedica	TT
Phase III Trial of CERE-120 for Parkinson's Disease	Parkinson's Disease	Sanofi/Sanofi Genzyme, Sangamo Therapeutics	TT
A Randomized, Placebo-controlled Phase IIIa Pivotal Confirmatory Study to Evaluate Safety and Efficacy of VY-AADC in Patients with Parkinson's Disease	Parkinson's Disease	Neurocrine Biosciences	TT
A Randomized Double-Blind Placebo-Controlled Parallel Group Study of the Efficacy and Safety of XRP0038/NV1FGF on Amputation or Any Death in Critical Limb Ischemia Patients With Skin Lesions	Peripheral Artery Disease	Sanofi	TT
Efficiency, Safety and Portability of Neovasculgen	Peripheral Artery Disease	Human Stem Cell Institute, Russia	CT
Gene Therapy using Intramuscular Administration of AMG0001 in Patients with Peripheral Arterial Disease;	Peripheral Artery Disease	AnGes	TT
Hepatocyte Growth Factor to Improve Functioning in Peripheral Artery Disease: The HI-PAD Study;	Peripheral Artery Disease	Helixmith	TT
A phase III study of HGF Plasmid in Peripheral Arterial Disease (PAD) in the US	Peripheral Artery Disease	AnGes	TT
Phase 3 Study of Efficiency, Safety and Portability of Gene Therapy Drug Neovasculgen (DNA Encoding the 165-amino-acid Isoform of Human Vascular Endothelial Growth Factor (pCMV - VEGF165) for Peripheral Arterial Disease Complex Treatment	Peripheral Artery Disease	Human Stem Cells Institute	TT
Provenge (Sipuleucel-T) Active Cellular Immunotherapy Treatment of Metastatic Prostate Cancer After Failing Hormone Therapy	Prostate Cancer	Dendreon	CT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Randomized, Controlled Trial of Replication-Competent Adenovirus-Mediated Suicide Gene Therapy in Combination With IMRT Versus IMRT Alone for the Treatment of Newly-Diagnosed Prostate Cancer With an Intermediate Risk Profile	Prostate Cancer (Localized)	Henry Ford Health System	TT
A Randomized Controlled Trial of ProstAtak as Adjuvant to Up-front Radiation Therapy For Localized Prostate Cancer	Prostate Cancer (Localized)	Candel Therapeutics	TT
A Phase III Randomized, Open-Label Study of CG1940 and CG8711 Versus Docetaxel and Estramustine in Patients with Metastatic Hormone-Refractory Prostate Cancer Who are Chemotherapy-Naive.	Prostate Cancer (Metastatic Hormone-Refractory)	ANI Pharmaceuticals, Takeda	TT
A Phase III Randomized, Open-Label Study of CG1940 and CG8711 Versus Docetaxel and Prednisone in Patients With Metastatic Hormone-Refractory Prostate Cancer Who Are Chemotherapy-Naive.	Prostate Cancer (Metastatic Hormone-Refractory)	ANI Pharmaceuticals, Takeda	TT
A Phase III Randomized, Open-Label Study of Docetaxel in Combination With CG1940 and CG8711 Versus Docetaxel and Prednisone in Taxane-Nave Patients With Metastatic Hormone-Refractory Prostate Cancer With Pain.	Prostate Cancer (Metastatic Hormone-Refractory)	ANI Pharmaceuticals, Takeda	TT
A Randomized Controlled Trial Of AdV-tk + Valacyclovir Administered During Active Surveillance For Newly Diagnosed Prostate Cancer	Prostate Cancer (Newly Diagnosed)	Candel Therapeutics	TT
An Open label,Randomized, Multi-Centered, Intra-Patient Controlled Phase III Study of FCX-007 in Patients with Recessive Dystrophic Epidermolysis Bullosa (RDEB)	Recessive Dystrophic Epidermolysis Bullosa	Fibrocell Science	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
VITAL: A Pivotal Phase 3 Study of EB-101 for the Treatment of Recessive Dystrophic Epidermolysis Bullosa (RDEB) (GENE TRANSFER)	Recessive Dystrophic Epidermolysis Bullosa	Stanford University Medical Center/ Abeona Therapeutics	TT
A Phase III, Randomized, Controlled, Double-Arm, Open-Label, Multi-center Study of VB-111 Combined With Bevacizumab vs. Bevacizumab Monotherapy in Patients With Recurrent Glioblastoma	Recurrent Glioblastoma	VBL Therapeutics	TT
A Phase II/III Randomized, Open-Label Study of Toca 511, a Retroviral Replicating Vector, Combined With Toca FC Versus Standard of Care in Subjects Undergoing Planned Resection for Recurrent Glioblastoma or Anaplastic Astrocytoma	Recurrent Glioblastoma	Tocagen	TT
A Randomized, Double-Blind, Placebo-Controlled, Parallel Group, Multicenter, Phase 3 Study to Evaluate the Safety and Efficacy of Ad5FGF-4 in Patients With Refractory Angina Due to Myocardial Ischemia;Ad5FGF-4 In Patients With Refractory Angina Due to Myocardial Ischemia (AFFIRM)	Refractory Angina due to Myocardial Ischemia (AFFIRM)	Gene Biotherapeutics/Angionetics	TT
A Phase III, Multicenter, Randomized, Open-label Study to Compare the Efficacy and Safety of bb2121 Versus Standard Triplet Regimens in Subjects With Relapsed and Refractory Multiple Myeloma (RRMM) (KarMMa-3)	Relapsed and Refractory Multiple Myeloma (RRMM)	Celgene	TT
A Single Global Phase 3 trial of RST-001 in Patients With Retinitis Pigmentosa (RP)	Retinitis Pigmentosa	Abbvie/Allergan	TT
A Phase II/III Expansion Study to Evaluate Safety and Efficacy of NSR-RPGR in Patients with a Diagnosis of X - Linked Retinitis Pigmentosa due to RPGR mutations	Retinitis Pigmentosa	NightstaRx	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
Phase 3 HGB-210 study of LentiGlobin in patients with SCD	Sickle Cell Anemia	bluebird bio	TT
Open-label, historical controlled study of AVXS-101 for treatment of spinal muscular atrophy	Spinal Muscular Atrophy	Novartis/AveXis	TT
A Multi-National Study of a One-Time Intrathecal Dose of AVXS-101 in Patients with Spinal Muscular Atrophy Types 1, 2, 3	Spinal Muscular Atrophy	Novartis/AveXis	TT
A Global Study of a Single, One-Time Dose of AVXS-101 Delivered to Infants With Genetically Diagnosed and Pre-symptomatic Spinal Muscular Atrophy With Multiple Copies of SMN2	Spinal Muscular Atrophy Type 1	Novartis/AveXis	TT
European, Phase 3, Open-Label, Single-Arm, Single-Dose Gene Replacement Therapy Clinical Trial for Patients With Spinal Muscular Atrophy Type 1 With One or Two SMN2 Copies Delivering AVXS-101 by Intravenous Infusion	Spinal Muscular Atrophy Type 1	Novartis/AveXis	TT
Phase 3, Open-Label, Single-Arm, Single-Dose Gene Replacement Therapy Clinical Trial for Patients With Spinal Muscular Atrophy Type 1 With One or Two SMN2 Copies Delivering AVXS-101 by Intravenous Infusion	Spinal Muscular Atrophy Type 1	Novartis/AveXis	TT
A Phase Ib/III Multicenter, Randomized, Trial of Talimogene Laherparepvec in Combination With Pembrolizumab for the Treatment of Subjects With Recurrent or Metastatic Squamous Cell Carcinoma of the Head and Neck	Squamous Cell Cancer of Head and Neck or Esophagus	Amgen/ Merck & Co./Merck Sharp & Dohme (MSD)	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
Phase III randomized clinical trial of intratumoral injection of E1B gene-deleted adenovirus (H101) combined with cisplatin-based chemotherapy in treating squamous cell cancer of head and neck or esophagus.	Squamous Cell Cancer of Head and Neck or Esophagus	Shanghai Sunway Biotech	TT
Phase III Randomized Study of Ad5CMV-p53 Gene Therapy (INGN 201) Versus Methotrexate in Patients With Refractory Squamous Cell Carcinoma of the Head and Neck (T301).	Squamous Cell Cancer of Head and Neck or Esophagus	Sanofi, Introgen Therapeutics	TT
A Phase III, Multi-Center, Open-Label, Randomized Study to Compare the Effectiveness and Safety of Intratumoral Administration of INGN 201 in Combination with Chemotherapy Versus Chemotherapy Alone in Patients with Squamous Cell Carcinoma of the Head and Neck (SCCHN)	Squamous Cell Cancer of Head and Neck or Esophagus	Introgen Therapeutics	TT
A Randomized, Controlled, Parallel Group, Multicenter Phase 3 Study to Evaluate the Efficacy and Safety of Ad5FGF-4 Using SPECT Myocardial Perfusion Imaging in Patients With Stable Angina Pectoris	Stable Angina	Gene Biotherapeutics/ Angionetics/ Gene Biotherapeutics	TT
A Randomized, Double Blind, Placebo Controlled, Parallel Group, Multicenter Study to Evaluate the Efficacy and Safety of Ad5FGF-4 in Female Patients With Stable Angina Pectoris Who Are Not Candidates for Revascularization;Angiogenesis in Women with Angina pectoris who are not candidates for Revascularization [AWARE]	Stable Angina	Gene Biotherapeutics/ Angionetics/ Gene Biotherapeutics	TT

Continued on next page

Table 4.11 – continued from previous page

Trial Title	Disease	Sponsors	Source
A Multinational Multicenter, Randomized, Double Blind, Placebo Controlled Study to Evaluate the Efficacy and Safety of Ad5FGF-4 in Patients With Stable Angina;(The Angiogenic Gene Therapy Trial - 4 [AGENT 4]).	Stable Angina	Bayer AG/Bayer HealthCare, Gene Biotherapeutics	TT
A Multicenter, Randomized, Double-Blind, Placebo Controlled Study to Evaluate the Efficacy and Safety of Ad5FGF-4 in Patients With Stable Angina (The Angiogeneic Gene Therapy Trial - 3 [AGENT 3])	Stable Angina	Bayer AG/Bayer HealthCare, Gene Biotherapeutics	TT
Multicentre, Randomized,Double Blind, Placebo Controlled Trial of Myocardial Angiogenesis Using VEGF165, Intramyocardial Gene Delivery in Patients With Severe Angina Pectoris	Stable Angina	Johnson & Johnson	TT
A Pivotal Study of NY-ESO-1 in Patients with Synovial Sarcoma including Myxoid Round Cell Liposarcoma	Synovial Sarcoma	GlaxoSmithKline/ AdaptImmune	TT

4.6.2 Disease-to-Therapeutic Area Mapping

As mentioned in the main paper, we show how the diseases are related to the therapeutic areas in the table below.

Table 4.12: Diseases with ongoing gene therapy trials and their associated therapeutic areas.

Disease	**Therapeutic Area**
– **General Conditions** –	
Arteriosclerosis Obliterans	Cardiovascular
Critical Limb Ischemia	Cardiovascular
Degenerative Arthritis	Autoimmune/Inflammation
Diabetic Foot Symptoms	Metabolic/Endocrinology
Diabetic Foot Ulcers	Metabolic/Endocrinology
Diabetic Peripheral Neuropathy	Metabolic/Endocrinology
Heart Failure	Cardiovascular
Knee Osteoarthritis with Kellgren & Lawrence Grade 3	Autoimmune/Inflammation
Parkinson's Disease	CNS
Peripheral Artery Disease	Cardiovascular
Refractory Angina due to Myocardial Ischemia (AFFIRM)	Cardiovascular
Stable Angina	Cardiovascular
– **Rare Diseases** –	
Beta-Thalassemia	Metabolic/Endocrinology
Cerebral Adrenoleukodystrophy (CALD)	CNS
Choroideremia	Ophthalmology
Cystic Fibrosis	Cardiovascular
Ewing's Sarcoma	Oncology
Hemophilia A	Metabolic/Endocrinology
Hemophilia B	Metabolic/Endocrinology
Leber Congenital Amaurosis due to RPE65 Mutations	Ophthalmology
Leber Hereditary Optic Neuropathy	Ophthalmology
Lipoprotein Lipase Deficiency (LPLD)	Metabolic/Endocrinology
Metachromatic Leukodystrophy	Metabolic/Endocrinology
Mucopolysaccharidosis Type IIIa	CNS
Recessive Dystrophic Epidermolysis Bullosa	Autoimmune/Inflammation
Retinitis Pigmentosa	Ophthalmology
Sickle Cell Anemia	Metabolic/Endocrinology
Spinal Muscular Atrophy	CNS
Spinal Muscular Atrophy Type 1	CNS
– **Cancer** –	
B-Cell Non-Hodgkin's Lymphoma	Oncology
BCG Unresponsive NMIBC	Oncology

Continued on next page

Table 4.12 – continued from previous page

Disease	**Therapeutic Area**
Bladder Cancer, in situ concurrent with Papillary Tumors	Oncology
Diffuse Large B Cell Lymphoma (DLBCL)	Oncology
Head and Neck Cancer	Oncology
Hepatocellular Carcinoma	Oncology
High-Grade Glioma	Oncology
Leukemia (Acute Lymphoblastic)	Oncology
Leukemia (Acute Myelogenous)	Oncology
Lymphoma	Oncology
Melanoma (Locally Advanced Cutaneous)	Oncology
Melanoma (Metastatic)	Oncology
Multiple Myeloma (Newly Diagnosed)	Oncology
Nasopharyngeal Carcinoma	Oncology
NSCLC	Oncology
NSCLC Stage 3	Oncology
Oral Cancer (Advanced)	Oncology
Ovarian Cancer (Platinum-Resistant)	Oncology
Ovarian Cancer, Primary Peritoneal Cavity Cancer	Oncology
Pancreatic Cancer (Locally Advanced)	Oncology
Prostate Cancer	Oncology
Prostate Cancer (Localized)	Oncology
Prostate Cancer (Metastatic Hormone-Refractory)	Oncology
Prostate Cancer (Newly Diagnosed)	Oncology
Recurrent Glioblastoma	Oncology
Relapsed and Refractory Multiple Myeloma (RRMM)	Oncology
Squamous Cell Cancer of Head and Neck or Esophagus	Oncology
Synovial Sarcoma	Oncology

4.6.3 Patient Population Estimation

We source the patient prevalence and incidence of the diseases from different sources. When necessary, we compute the prevalence from the incidence using Equation 4.1, or vice versa, using Equation 4.2. Our results are shown in Table 4.13. These numbers do not reflect the adjustments we make to NSC lung cancer, prostrate cancer and spinal muscular atrophy in order to minimize overlapping patient groups.

Table 4.13: Number of current patients and annual new patients for each disease. An asterisk (∗) indicates that either the prevalence is computed from the incidence using Equation 4.1, or vice versa, using Equation 4.2.

Disease	**Current patients**	**New patients per year**
– General Conditions –		
Arteriosclerosis Obliterans	[94] 8500000	*192100
Critical Limb Ischemia	[85] 975000	[85] 300000
Degenerative Arthritis	[161] 27000000	*486000
Diabetic Foot Ulcers	[130] 2250000	[129] 112500
Diabetic Peripheral Neuropathy	[69,95,168] 9441480	[69,95,168] 467400
Heart Failure	[106] 5800000	[106] 812000
Knee Osteoarthritis with Kellgren & Lawrence Grade 2 or/and 3	*2929730	[161] 542000
Parkinson's Disease	[20] 500000	[20] 50000
Peripheral Artery Disease	[94] 8500000	*564400
Refractory Angina due to Myocardial Ischemia (AFFIRM)	[5] 8200000	[5,73] 565000
Stable Angina	[44] 10000000	[44] 500000
– Rare Diseases –		
Beta-Thalassemia	[11] 1000	[137] 3277
Cerebral Adrenoleukodystrophy (CALD)	[66] 411	[66] 37
Choroideremia	[127] 6554	[127] 77
Cystic Fibrosis	[65] 30000	[65] 1000
Ewing's Sarcoma	[13,14,61] 15003	[13,14,61] 200
Hemophilia A	[47] 16000	[47] 360
Hemophilia B	[47] 4000	[47] 90
Leber Congenital Amaurosis due to RPE65 Mutations	[79] 187	[79] 19
Leber Hereditary Optic Neuropathy	[27] 6540	[27] 654
Lipoprotein Lipase Deficiency (LPLD)	[75] 328	*33

Continued on next page

Table 4.13 – continued from previous page

Disease	Current patients	New patients per year
Metachromatic Leukodystrophy	[48,139]9333	[48,139]771
Mucopolysaccharidosis Type IIIa	[30]1638	[30]39
Recessive Dystrophic Epidermolysis Bullosa	[113]100	*10
Retinitis Pigmentosa	[144]87387	[144]8739
Sickle Cell Anemia	[58]100000	[59]58745
Spinal Muscular Atrophy	[123]8526	[123]290
Spinal Muscular Atrophy Type 1	[42]17500	[82]500
– Cancer –		
B-Cell Non-Hodgkin's Lymphoma	[9,10]694704	[9,10]74200
BCG Unresponsive NMIBC	[12,107,118,119]371933	[12,107,118,119]42625
Bladder Cancer, in situ concurrent with Papillary Tumors	[12]356720	[12]41040
Diffuse Large B Cell Lymphoma (DLBCL)	[1]257	[1]18351
Head and Neck Cancer	[50,53,54,55]134337	[50,53,54,55]75275
Hepatocellular Carcinoma	[15,102,104]11287	[15,102,104]2032
High-Grade Glioma	[3,8,49,51,147]87540	[3,8,49,51,147]16334
Leukemia (Acute Lymphoblastic)	[6]95764	[6]5930
Leukemia (Acute Myelogenous)	[7]61048	[7]21450
Lymphoma	[16,34]905678	[16,34]82310
Melanoma (Locally Advanced Cutaneous)	[17]107605	[17]8683
Melanoma (Metastatic)	[17]47824	[17]3859
Multiple Myeloma (Newly Diagnosed)	0	[2]32270
Nasopharyngeal Carcinoma	[56]5390	[56]327
NSC Lung Cancer	[35,57]454469	[57] 191646
NSC Lung Cancer Stage 3	[35,57,141]151490	[57,141]63882
Oral Cancer (Advanced)	[4]250000	[4]53000
Ovarian Cancer (Platinum-Resistant)	[36,37,38]141150	[36,37,38]13956
Ovarian Cancer, Primary Peritoneal Cavity Cancer	[40]2290	[40]240
Pancreatic Cancer (Locally Advanced)	[39,148]22066	[39,148]17031
Prostate Cancer	[18]3110403	[18]174650
Prostate Cancer (Localized)	[18]2395010	[18]134481
Prostate Cancer (Metastatic Hormone-Refractory)	[18]186624	[18]10479
Prostate Cancer (Newly Diagnosed)	0	[18]174650
Recurrent Glioblastoma	[8,49,51,147]64127	[8,49,51,147]12120
Relapsed and Refractory Multiple Myeloma (RRMM)	[41,81]48840	[41,81]16280
Squamous Cell Cancer of Head and Neck or Esophagus	[50,53,54,55]120903	[50,53,54,55]67747

Continued on next page

Table 4.13 – continued from previous page

Disease	**Current patients**	**New patients per year**
Synovial Sarcoma	[43,89]7282	[89]655

4.6.4 Calibration of Survival Functions $D_{alt}(x-a)$

We source either the survival or mortality rate from literature and use them to compute λ, the time parameter in the exponential survival function. We show our result in the table below.

Table 4.14: List of survival rate or mortality rate and λ, for each disease. An asterisk ($*$) under λ denotes that the disease does not affect mortality directly.

Disease	k years survival rate		k years mortality rate			λ
	$k=5$	$k=10$	$k=1$	$k=5$	$k=10$	
Arteriosclerosis Obliterans				[160]11.3		0.024
Critical Limb Ischemia				[86]50		0.139
Degenerative Arthritis		[163]82				0.020
Diabetic Foot Ulcers				[146]49		0.135
Diabetic Peripheral Neuropathy					[62]5	0.005
Heart Failure				[52]42.3		0.110
Knee Osteoarthritis with Kellgren & Lawrence Grade 2	[155]7.5					0.518
Knee Osteoarthritis with Kellgren & Lawrence Grade 3	[155]7.5					0.518
Parkinson's Disease	[165]40					0.174
Peripheral Artery Disease				[76]33.2		0.081
Refractory Angina due to Myocardial Ischemia (AFFIRM)			[109]3.9			0.040
Stable Angina	[110]90					0.021
Beta-Thalassemia		[169]98.3				0.002
Cerebral Adrenoleukodystrophy (CALD)	[143]55					0.120
Choroideremia						$*$
Cystic Fibrosis					[120]28	0.033
Ewing's Sarcoma	[131]70					0.071
Hemophilia A					[140]9.7	0.010

Continued on next page

Table 4.14 – continued from previous page

Disease	k years survival rate		k years mortality rate			λ
	$k = 5$	$k = 10$	$k = 1$	$k = 5$	$k = 10$	
Hemophilia B					[140]9.7	0.010
Leber Congenital Amaurosis due to RPE65 Mutations						*
Leber Hereditary Optic Neuropathy						*
Lipoprotein Lipase Deficiency (LPLD)						*
Metachromatic Leukodystrophy	[128]52					0.131
Mucopolysaccharidosis Type III		[125]60				0.051
Recessive Dystrophic Epidermolysis Bullosa						*
Retinitis Pigmentosa						*
Sickle Cell Anemia		[100]96				0.004
Spinal Muscular Atrophy	[83]40					0.183
Spinal Muscular Atrophy Type 1	[170]10.13					0.458
B-Cell Non-Hodgkin's Lymphoma	[31]72					0.066
BCG Unresponsive Non-Muscle Invasive Bladder Cancer	[119]78					0.050
Bladder Cancer, Transitional Cell Carcinoma	[21]95.8					0.009
Diffuse Large B-Cell Lymphoma (DLBCL)	[26]63.2					0.092
Head and Neck Cancer	[22]64					0.089
Hepatocellular Carcinoma	[104]10					0.461
High-Grade Glioma	[112]9.87					0.463
Leukemia (Acute Lymphoblastic)	[6]68.8					0.075
Leukemia (Acute Myelogenous)	[7]28.7					0.250
Lymphoma	[29]72					0.066
Melanoma (Locally Advanced Cutaneous)	[28]64					0.089
Melanoma (Metastatic)	[28]23					0.294
Multiple Myeloma (Newly Diagnosed)	[60]52					0.131
Nasopharyngeal Carcinoma	[32]72					0.066
NSCLC	[45]23					0.294

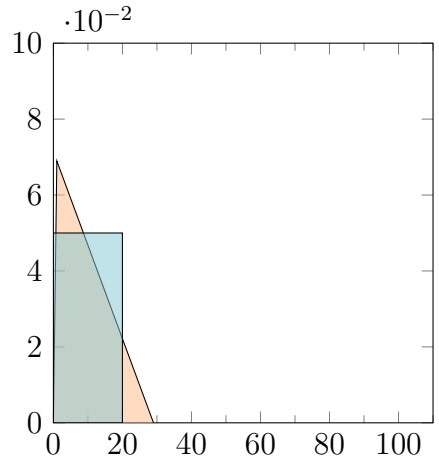
Continued on next page

Table 4.14 – continued from previous page

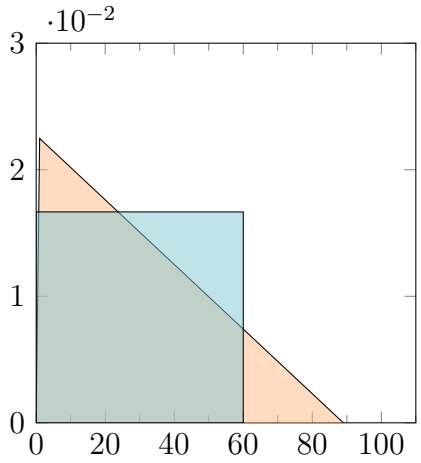
Disease	k years survival rate		k years mortality rate			λ
	$k = 5$	$k = 10$	$k = 1$	$k = 5$	$k = 10$	
NSCLC (Stage 3)	[45]33					0.222
Oral Cancer (Advanced)	[22]39.1					0.188
Ovarian Cancer (Platinum-Resistant)	[162]1.9					0.793
Ovarian Cancer, Primary Peritoneal Cavity Cancer	[23]47.6					0.148
Pancreatic Cancer (Locally Advanced)	[24]12.4					0.417
Prostate Cancer	[25]98					0.004
Prostate Cancer (Localized)	[25]98					0.004
Prostate Cancer (Metastatic Hormone Refractory)	[25]30.5					0.237
Prostate Cancer (Newly Diagnosed)	[25]95.1					0.010
Recurrent Glioblastoma	[122]10					0.461
Relapsed and Refractory Multiple Myeloma (RRMM)	[46]9.92					0.462
Squamous Cell Cancer of Head and Neck or Esophagus	[22]64					0.089
Synovial Sarcoma	[33]55					0.120

4.6.5 Calibration of Age Distribution $A(x)$

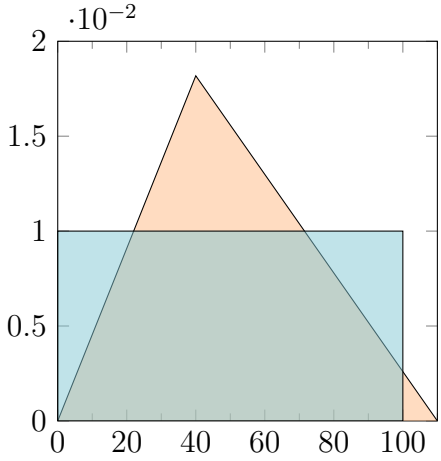
As mentioned in the main paper, our optimization program produces triangular age distributions that conforms to data, have wider support compared to fitting uniform distributions and, avoids sharp changes in the probability density. We illustrate some examples that compare triangle distributions with the uniform distributions with the same average age.



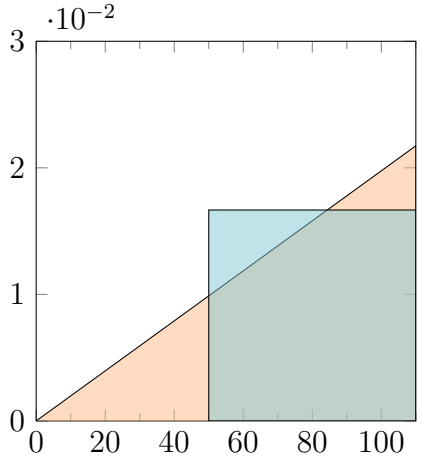
(a) Comparative probability distributions when $\mu_{age} = 10$



(b) Comparative probability distributions when $\mu_{age} = 30$



(c) Comparative probability distributions when $\mu_{age} = 50$



(d) Comparative probability distributions when $\mu_{age} = 80$

Figure 4-14: Age distributions given various mean ages, μ_{age}. The red triangles represent the solutions obtained by our optimization program, while the blue rectangles represent the solutions given by an uniform distribution. The distributions from the optimization program have a wider base of support and avoid sharp changes in density.

4.6.6 Quality of Life Estimation

The results of our literature search and estimation for the change in QoL for each disease is shown in the table below.

Table 4.15: Table of disease scores (ζ), estimated quality of life values before treatment $\hat{f}_h(s_{alt})$, after treatment $\hat{f}_h(s_{gt})$, and the change in quality of life (ΔQoL). Asterisks ($*$) indicate that the values are interpolated. Cancers are not included, as we assume that the gains in survival dominate the gains in QoL.

Non-Cancer Disease	ζ	$\hat{\mathbf{f}}_\mathbf{h}(\mathbf{s_{alt}})$	**ΔQoL**	$\hat{\mathbf{f}}_\mathbf{h}(\mathbf{s_{gt}})$
Arteriosclerosis Obliterans	1	*0.775	*0.075	*0.850
Beta-Thalassemia	3	[151]0.870	*0.166	*1.000
Cerebral Adrenoleukodystrophy (CALD)	5	*0.654	*0.257	*0.911
Choroideremia	3	*0.715	*0.166	*0.881
Critical Limb Ischemia	4	*0.684	*0.212	*0.896
Cystic Fibrosis	3	[64]0.671	*0.166	*0.837
Degenerative Arthritis	3	*0.715	*0.166	*0.881
Diabetic Foot Ulcers	3	[154]0.703	[154]0.258	[154]0.961
Diabetic Peripheral Neuropathy	2	[153]0.630	[153]0.180	[153]0.810
Ewing's Sarcoma	2	[145]0.690	*0.121	*0.811
Heart Failure	4	*0.684	*0.212	*0.896
Hemophilia A	5	[77]0.750	*0.257	*1.000
Hemophilia B	5	[77]0.700	*0.257	*0.957
Knee Osteoarthritis, Kellgren & Lawrence Grade 2	2	[134]0.900	[134]0.042	*0.942
Knee Osteoarthritis, Kellgren & Lawrence Grade 3	2	[134]0.900	[134]0.048	*0.948
Leber Congenital Amaurosis (RPE65 Mutations)	3	*0.715	*0.166	*0.881
Leber Hereditary Optic Neuropathy	3	*0.715	*0.166	*0.881
Lipoprotein Lipase Deficiency	4	*0.684	*0.212	*0.896
Lysosomal Storage Disease	5	[99]0.640	*0.257	*0.897
Mucopolysaccharidosis Type IIIa	2	[108]0.582	[108]0.264	[108]0.846
Osteoarthritis	2	[158]0.900	[158]0.040	*0.940
Parkinson's Disease	4	[87,152]0.700	[87]0.150	[87]0.850
Peripheral Artery Disease	2	[111]0.660	[111]0.060	[111]0.720
Recessive Dystrophic Epidermolysis Bullosa	4	[117]0.590	*0.212	*0.802
Refractory Angina due to Myocardial Ischemia	2	[105]0.600	*0.121	*0.721
Retinitis Pigmentosa	3	[150]0.770	*0.166	*0.936
Sickle Cell Anemia	3	[93]0.732	[93]0.198	[93]0.930
Spinal Muscular Atrophy[6]	5	0.520	*0.257	*0.777
Spinal Muscular Atrophy Type 1	5	[172]0.520	*0.257	*0.777

Continued on next page

[6]We are unable to find QoL values for SMA only and assume that they are the same as SMA Type 1.

Table 4.15 – continued from previous page

Non-cancer Disease	ζ	$\hat{f}_h(s_{alt})$	ΔQoL	$\hat{f}_h(s_{gt})$
Stable Angina	2	[121,167]0.750	[167]0.150	[167]0.900

4.6.7 Simulation Convergence Criteria

Let X_k be the results of the k-th simulation. X_k has a true mean of μ and variance σ^2. Let the mean of the Monte Carlo simulations over n runs be $\hat{\mu}_n = \frac{1}{n}\sum_k^n X_k$. Then, by Lindeberg–Lévy's Central Limit Theorem, $\hat{\mu}_n$ converges in distribution to a normal distribution with mean μ and variance of $n\sigma^2$. The 95 percent confidence interval for μ is given by:

$$\hat{\mu}_n \pm \frac{1.96 s_n}{\sqrt{n}} \tag{4.27}$$

where s_n is the sample variance of $\{X_1, \cdots, X_n\}$.

Since we are using 1-by-T vectors, we investigated the error in our simulation by dividing the half-range of the confidence interval in each time-step by $\hat{\mu}_n$ before taking the maximum across the time series. As can be seen from Figure 4-15, we should expect the simulated mean to be within 1.89% of the true mean 95% of the time with 1,000,000 iterations.

Figure 4-15: Plot of $\frac{1.96 s_n}{\mu\sqrt{n}}$ against the number of iterations of simulations of the cost.

4.6.8 Pseudo-Code and Implementation Details

Pseudo-code

We perform a Monte Carlo simulation to determine the total number of patients undergoing gene therapy and the cost of these gene therapies at specific points in time. The sequence of computations for each iteration of the simulation is detailed in Algorithm 2.

Algorithm 2: Pseudocode for one iteration of the simulation.

Input : $\mathbb{D}$: A list of diseases
Output: Arrays of $[1 \times T]$, where T is the number of time steps.
$\mathbb{P}$: Number of patients over time
$\mathbb{C}$: Total cost over time

```
ℙ ← 1 × T array of zeros
ℂ ← 1 × T array of zeros
for d in 𝔻 do
    p ← getPoS(d) // Get probability of success
    if random.uniform(0,1) ≤ p then
        // If the disease gets an approval...
        existing ← getExistingPatients(d) ; // Get existing patients (1 ×
         T)
        new ← getNewPatients(d) // Get new patients (1 × T)
        ρ ← getRampFunction(d) // Get penetration ramp (1 × T)
        price ← getPrice(d) // Get price of GT (scalar)
        ℙ+ = (existing + new)⊗ρ // Store number of patients
        ℂ+ = ℙ× price // Store cost of treatment
    end
end
return ℙ, ℂ
```

Implementation

All the equations are discretized for computation from their continuous forms. When solving the integrals using the trapezoidal rule to obtain ΔQALY, we use strip widths of 1 year across a range from 0 to 110 years old, the resolution offered by the life tables. When simulating the number of patients and the cost over time, we use steps of 1 month.

Our codes are implemented on Python 3.6 backed by Numpy. Our vectorized implementation averages 6.120ms per iteration over 1,000,000 runs on a single thread of an Intel

Xeon Gold 5120, clocked at 2.20GHz with 20GB of RAM. We attempted to use PyTorch to speed up the computations using a GPU, but it ran more slowly than a single-threaded CPU. We determined this took place for two reasons. First, generating random numbers must be sequential, since PyTorch delegates it solely to the CPU, which limits the amount of parallelization that can be achieved, as dictated by Amdahl's law. Second, because our computations require a large amount of data from different sources, they must be batched due to the GPU's limited RAM. The constant movement of data through the PCIe bridge, however, turns out to be a massive bottleneck to the overall speed.

4.6.9 Visualization of the Cost over Time

In this section, we visualize how the monthly cost of treating patients with gene therapy will be affected by changes to the variables. The results are summarized in the tornado chart presented in the main paper.

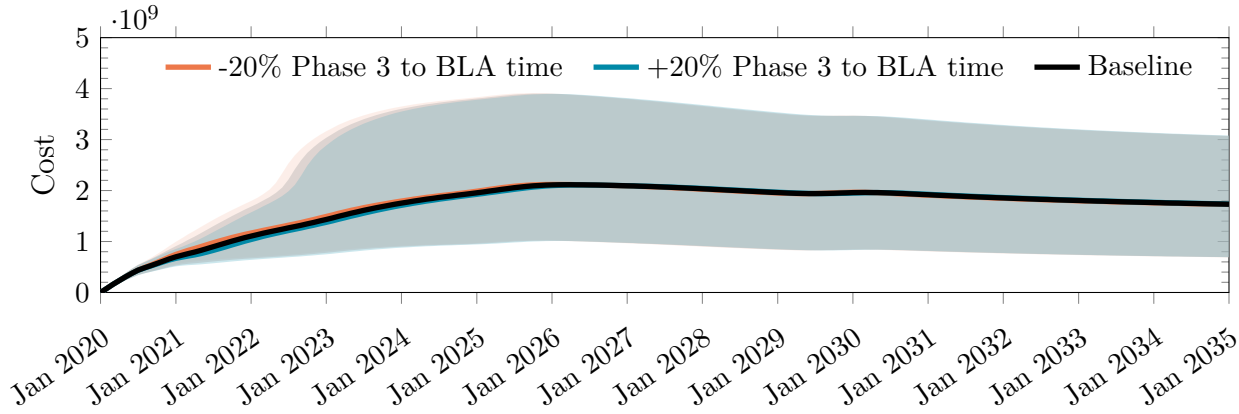


Figure 4-16: Impact on monthly cost of treating patients given a ±20% change in the time from phase 3 to BLA.

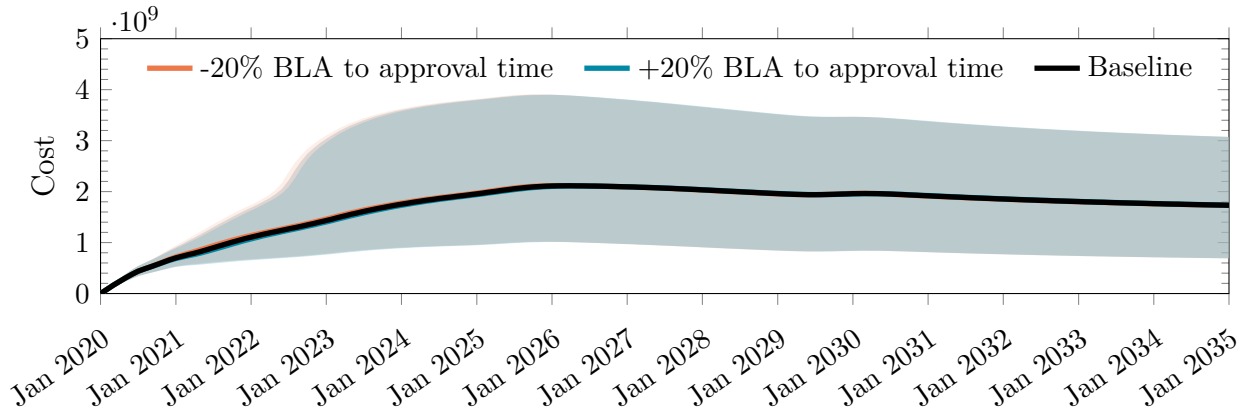


Figure 4-17: Impact on monthly cost of treating patients given a ±20% change in the time from BLA to approval.

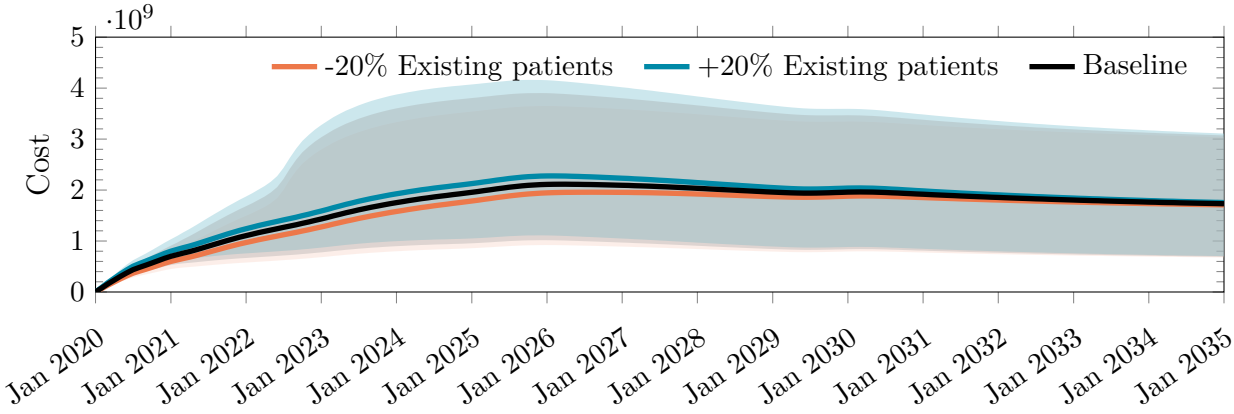


Figure 4-18: Impact on monthly cost of treating patients given a ±20% change in the number of existing patients.

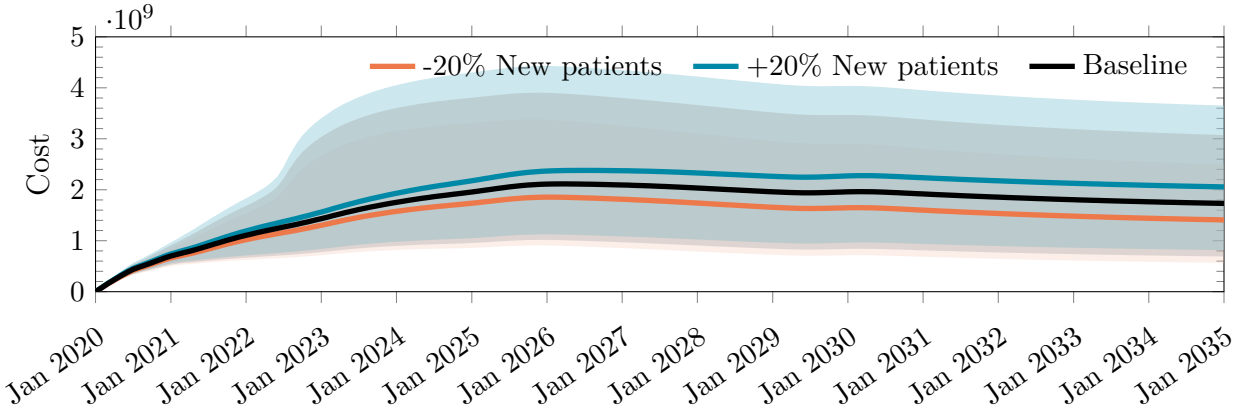


Figure 4-19: Impact on monthly cost of treating patients given a ±20% change in the number of new patients.

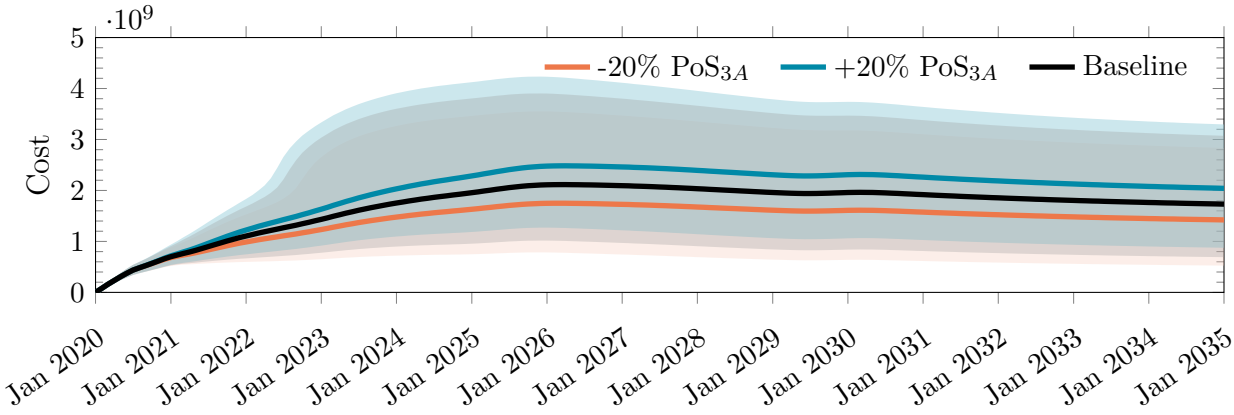


Figure 4-20: Impact on monthly cost of treating patients given a ±20% change in the PoS_{3A}.

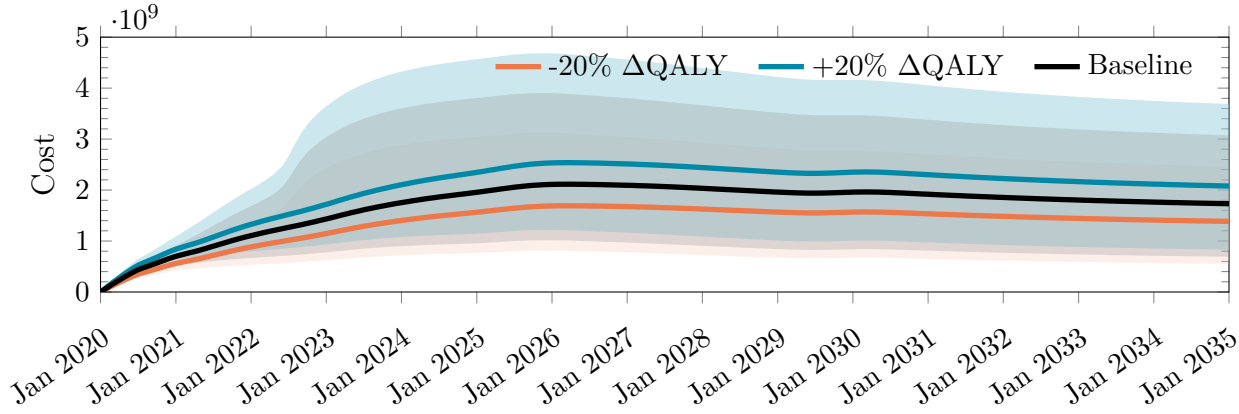


Figure 4-21: Impact on monthly cost of treating patients given a ±20% change in ΔQALY gained.

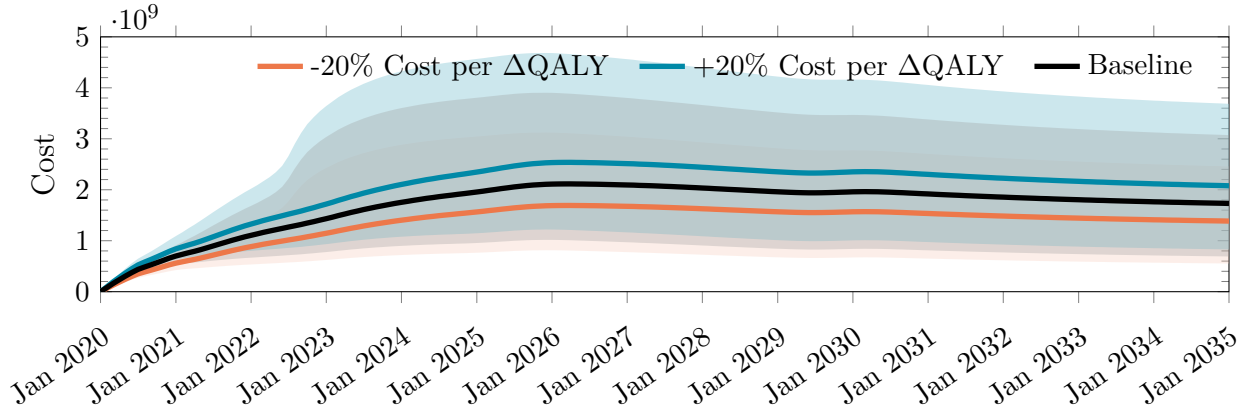


Figure 4-22: Impact on monthly cost of treating patients given a ±20% change in the cost per ΔQALY.

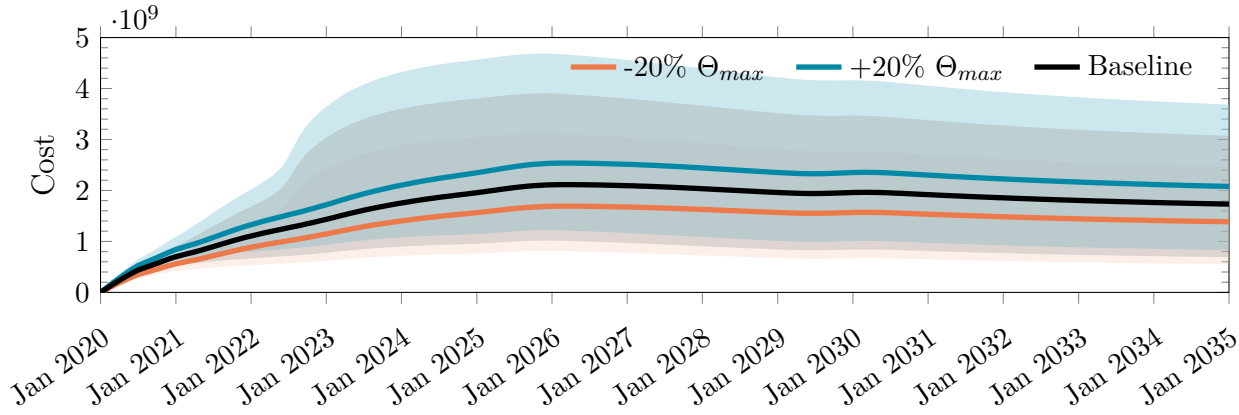


Figure 4-23: Impact on monthly cost of treating patients given a ±20% change in Θ_{max}.

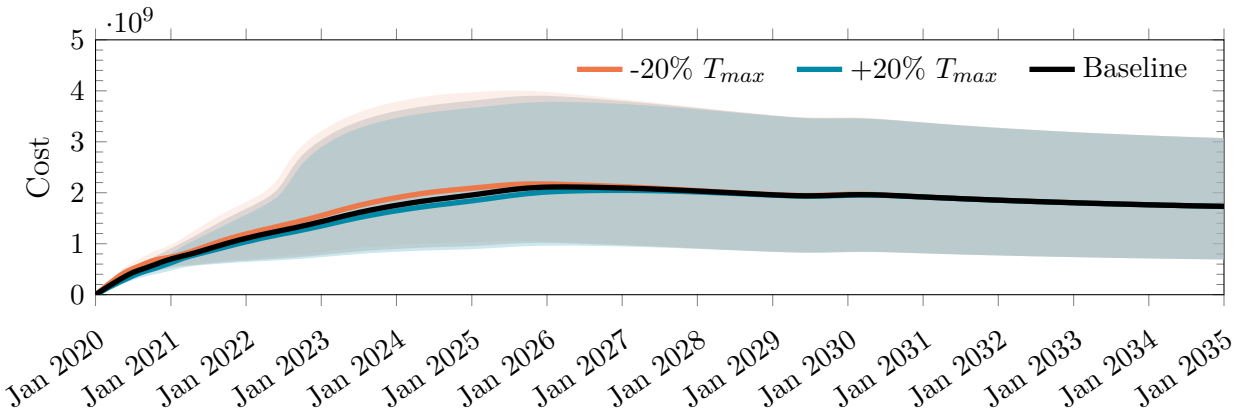


Figure 4-24: Impact on monthly cost of treating patients given a ±20% change in T_{max}.

Chapter 5

Cost/Benefit Analysis of Clinical Trial Designs for COVID-19 Vaccine Candidates

5.1 Introduction

The COVID-19 pandemic has caused the deaths of hundreds of thousands, and led to an economic fallout which has upended the lives of billions, and caused trillions of dollars in financial losses. And life may not return to normalcy until a vaccine is found [12]. Despite the many candidates undergoing testing, an approved vaccine is not expected until 2021, even with substantially compressed development timelines [1], smooth proceeding of clinical trials, and not accounting for possible failures [33]. It is possible—though considered highly unlikely at the present time—that, like many non-influenza epidemics, the crisis may be over before a successful vaccine is developed [34].

Unlike typical therapeutics that are administered to sick patients, vaccines are intended for the healthy. Therefore, confirming the safety and effectiveness of a vaccine is of critical importance [5]. The two primary methods for demonstrating vaccine safety and efficacy are through either a vaccine efficacy randomized clinical trial (RCT) or a vaccine immunogenicity RCT. In the former, large numbers of recruited healthy volunteers are randomly selected to

receive either the vaccine or a placebo/active control and then monitored for a period of time. At the end of the surveillance period, the difference in the proportion of infections between the treatment and control arms is computed to demonstrate the ability of the vaccine to prevent infection or disease. A phase 3 vaccine efficacy RCT typically takes five to ten years to complete [23].

In a vaccine immunogenicity RCT, the primary endpoint is an immunity measurement or surrogate marker which is known to correlate with protection against infection or a disease. This type of trial involves a smaller number of volunteers and requires a shorter follow-up period, and as a result, is quicker and less costly [41]. Given that SARS-CoV-2 is a novel pathogen for which we do not yet know how to determine whether a subject is protected, vaccine efficacy must be confirmed using the longer and more costly vaccine efficacy RCT. While there exists the possibility of an expedited (conditional) licensure based on immunogenicity results with post-approval commitments, we find it unlikely to occur given the latest information. The U.S. Food and Drug Administration (FDA) has also issued a guidance stating that "the goal of development programs should be to pursue traditional approval via direct evidence of vaccine efficacy" [19].

A human challenge trial (HCT), in which volunteers are randomized into either the vaccine or placebo arm and then infected deliberately with live virus in a controlled setting, has been proposed as an alternative to accelerate the vaccine development process. Upon challenge, HCTs can quickly demonstrate safety and efficacy of candidate vaccines in preventing infection or disease. Depending on sample size, HCTs could also help to establish functional immune correlates of protection to inform the design of future vaccines. Since an HCT allows comparison of immune responses in vaccinated and unvaccinated individuals, precise measurements of post-vaccination viral loads, characterization of immune responses (innate, adaptive, cell-mediated) and antibody titers, and close monitoring and care of patients, it can help establish the correlates of protection and prove vaccine efficacy concurrently. Moreover, a properly designed HCT can determine transmission risk of the infected in a controlled setting with minimal exposure to investigators and the public. While concerns have been raised regarding the ethics and morality of HCTs, it is generally accepted that HCTs are ethically permissible when the benefits to society outweigh acknowledged risks [50], and they

have been deemed acceptable for developing vaccines for multiple infectious diseases such as influenza [51], malaria [53], typhoid [47], cholera [52], and dengue fever [59]. To the best of our knowledge, there have been no published studies of a quantitative analysis of the potential societal value of a COVID-19 HCT.

In this chapter, we compare the costs and benefits—as measured by the number of deaths and infections avoided—of confirming the safety and efficacy of a COVID-19 vaccine using four clinical trial designs: a traditional vaccine efficacy RCT, a vaccine efficacy RCT with an optimized surveillance period that maximizes the benefits of the trial (ORCT), an adaptive vaccine efficacy RCT (ARCT), and an HCT. Although our framework applies more broadly to any vaccine candidate for any infectious disease, we calibrate our simulations using a set of estimated epidemiological models for the SARS-CoV-2 virus (one for each of the 50 states and Washington, D.C.) to determine attack rates and cumulative numbers of infections and deaths in the U.S under various scenarios.

A summary of our simulation framework is displayed in Figure 5-1. We first estimate baseline models from data and make assumptions for the evolution of the epidemic in order to predict the attack rates over the course of the clinical trials. We then combine the attack rates with the assumptions for the vaccine trial design to simulate the outcomes for the clinical trials. Conditioned on the vaccine being approved, we make assumptions on the vaccination schedule and simulate the path of the epidemic in order to compute the net infections and deaths prevented.

We will review the designs and assumptions for the four vaccine trials considered in Section 5.2 and explain our cost/benefit calculations in Section 5.4.1. We will then present the epidemiological model used in our forecasts in Section 5.3 and report our simulation results in Section 5.4. Finally, we discuss our findings and some broader issues of COVID-19 clinical trials in Section 5.5 and conclude in Section 5.6.

5.2 Vaccine Trial Design

We begin by describing the assumptions and calibrations used in each of the four vaccine trial designs we considered in our simulations.

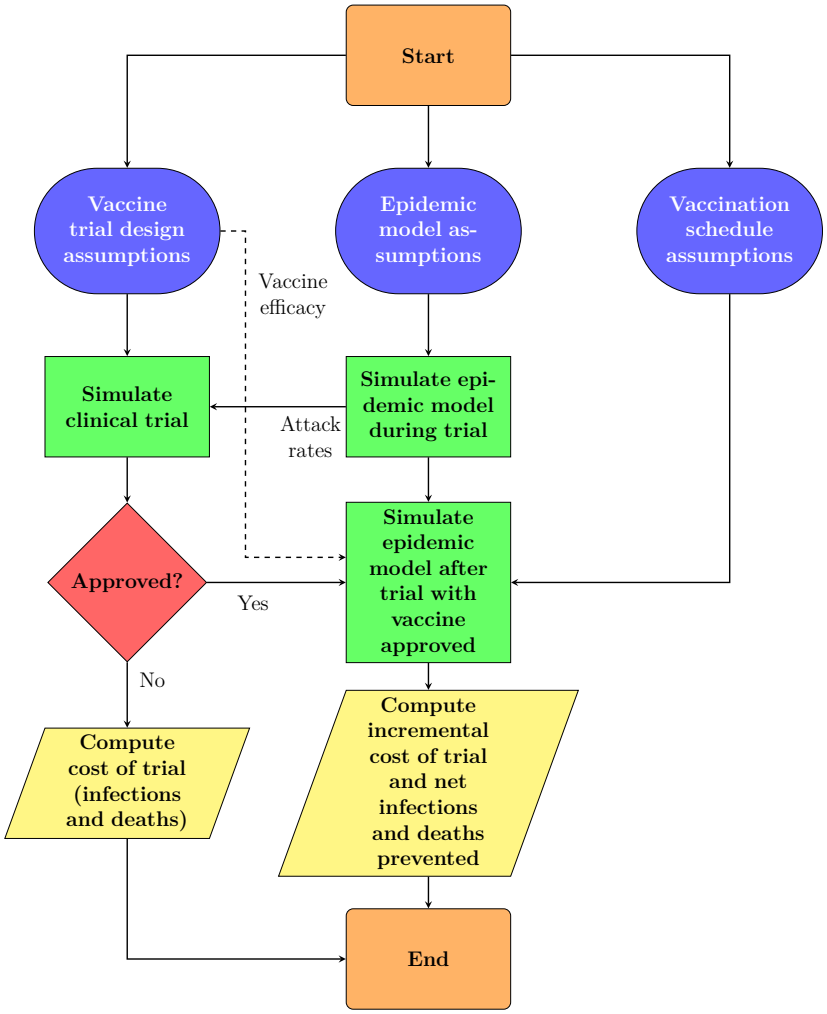


Figure 5-1: Simulation framework. For each Monte Carlo simulation path, we simulate patient-level infections data based on input trial design assumptions and attack rates from the population epidemiological model (for an RCT, ORCT, and ARCT). At the end of the trial (or, at each interim analysis for an ARCT), we determine if the vaccine candidate is approved under superiority or superiority-by-margin testing. Finally, we compute the expected net value of the trial design over 100,000 simulation paths using Equation 5.24.

5.2.1 Traditional Vaccine Efficacy RCT

First, we consider a traditional double-blind vaccine efficacy trial design. We assume that a closed cohort of 30,000 infection-free but at-risk healthy U.S. adults aged between 18 and 50 years will be enrolled for the study, similar to the phase 3 studies planned or underway for the COVID-19 vaccines developed by Moderna [37], AstraZeneca [2], Pfizer/BioNTech [6], and others. The participants will be randomized equally between the treatment and control arms, receiving either the vaccine candidate or an active control vaccine[1] (e.g., vaccine against meningococcal bacteria), respectively. Unlike clinical trials for cancer therapeutics where patient accrual can be a challenge due to the small pool of afflicted patients and strict inclusion/exclusion criteria, subject enrollment for vaccine efficacy studies are often accelerated because there is a large pool of healthy adult volunteers to recruit from. Therefore, we assume an accrual rate of 250 patients per day in our simulations.

Similar to the design of study protocols adopted for phase 3 clinical trials of current leading SARS-CoV-2 vaccine candidates, we assume a hypothetical COVID-19 vaccine candidate that will be administered to subjects in two doses, 28 days apart, i.e., the prime-boost regimen [36, 54]. Furthermore, we assume that it takes approximately 28 days after the booster dose for antibodies to develop (i.e., seroconversion) before surveillance can begin.

We consider efficacy in the prevention of infection by SARS-CoV-2 as the primary endpoint in our study.[2] To draw meaningful conclusions from the trial results, volunteers must be monitored long enough for a sufficient number of infections to occur. Here, we assume a fixed post-vaccination surveillance period of 180 days for all subjects in the cohort, after which a single safety and primary efficacy analysis will be performed to determine licensure (see Section 5.7.1).

Finally, we assume an interval of 120 days after surveillance for the preparation of a biologics license application (BLA) submission to the FDA, including an analysis and publication of safety, immunogenicity, and efficacy results; collection of chemistry, manufacturing,

[1]The use of an active vaccine (e.g., vaccine against meningococcal bacteria) as control provides some benefit to the participants, making it more ethical. It also serves to ensure that the participants are unable to tell whether they received the COVID-19 vaccine based on side effects such as soreness at the injection site, reducing the possibility of behavioral changes that can bias the results of the study.

[2]We note that secondary endpoints include the prevention of COVID-19.

and controls (CMC) data; the writing of a clinical study report; and subsequent review by the FDA. Under these assumptions, we estimate the time to licensure of our hypothetical candidate under a traditional RCT to be approximately 476 days. This is the baseline value against which we will compare the other three trial designs.

5.2.2 Optimized Vaccine Efficacy RCT

Depending on the transmission rate of COVID-19 during the trial and the assumed efficacy of the hypothetical candidate, a shorter surveillance period might be sufficient to observe significant results.[3] Therefore, we consider an optimized version of the traditional vaccine efficacy RCT design (ORCT) in which the surveillance period is optimized between 30 to 180 days based on different epidemiological scenarios and vaccine efficacies to maximize the expected number of incremental infections and deaths prevented.[4] Apart from the surveillance period, we assume that the ORCT is identical to the RCT in all other aspects.

5.2.3 Adaptive Vaccine Efficacy RCT

An adaptive version of the traditional vaccine efficacy RCT design (ARCT) is based on group sequential methods [27]. Instead of a fixed study duration with a single final analysis at the end, we allow for early stopping for efficacy via periodic interim analyses of accumulating trial data (see Section 5.7.2). While this reduces the expected duration of the trial, we note that adaptive trials typically require more complex study protocols which can be operationally challenging to implement for test sites unfamiliar with this framework. In our simulations, we assume a maximum of six interim analyses spaced 30 days apart, with the first analysis performed when the first 10,000 subjects have been monitored for at least 30 days. While we have assumed interim analyses at periodic calendar time points here, we note that most

[3] In general, the higher the transmission rate, the shorter the surveillance period required to observe a statistically significant difference in infection risk between the treatment arm and the control arm (or the lack of thereof) at the same level of significance and power, assuming a constant sample size and vaccine efficacy.

[4] There is a trade-off between time and power: A shorter surveillance period will, *ceteris paribus*, reduce the power of the RCT. However, it will also reduce the time to licensure of the vaccine (if approved), which can potentially prevent more infections and save more lives. Conversely, a longer surveillance period will increase the power of the RCT and prolong the time it takes for the vaccine to be approved. See Figure 5-6 for an illustration.

vaccine efficacy trials are event based, e.g., performing interim analyses when pre-specified numbers of events occur. In addition, we have adopted the Pocock's test for sequential testing in our analysis (see Algorithm 3), but we note that some companies are using variants of the O'Brien-Fleming test [35], which have stricter requirements for early stopping, and therefore may lead to longer studies [27].

5.2.4 HCT

Unlike traditional vaccine efficacy field trials which require large sample sizes to observe significant results, we assume that the HCT requires only 250 volunteers, randomized 4:1 between the treatment and control arms. Furthermore, to minimize the risk to participants, we assume that this study will recruit only young and healthy adults aged between 18 and 25 years without any underlying chronic conditions because this group of individuals has the lowest risk of mortality and complications after recovering from the infection [8, 43, 56].

It is clear that extensive preparations are required to set up an HCT: selecting, developing, testing an appropriate challenge virus strain;[5] manufacturing a batch of the selected challenge strain under good manufacturing practices (GMP); and identifying the dose level required to achieve satisfactory attack risk of non-severe clinical illness [56]. From discussions with challenge trial experts, there seems to be a lack of consensus on the appropriate set-up time for HCTs. We reflect this uncertainty in our simulations by incorporating a lag time for HCTs ("set-up time") that ranges between 30 to 120 days.

In the challenge study, volunteers are deliberately exposed to the SARS-CoV-2 virus, reducing post-vaccination monitoring times because investigators do not need to wait for infections to occur naturally as with non-challenge RCTs. Therefore, we assume a surveillance period of only 14 days (the incubation period for COVID-19 [21, 30, 31]) for the challenge study. Moreover, the attack rate in the control arm will be independent of the population epidemiological model since the study will be conducted in isolated facilities. In our simulations, we assume that 90% of the subjects in the control arm will be infected after the challenge.[6]

[5]There are multiple lineages of SARS-CoV-2 to choose from. In addition, a decision must be made between using a fully virulent or an attenuated strain of the SARS-CoV-2 virus.

[6]We do not assume a 100% attack rate since the challenge strain used is likely weakened to reduce risk

We note that the FDA is unlikely to approve an experimental vaccine tested in only 200 subjects (versus thousands in non-challenge RCTs), hence we assume that a large-scale safety study will be performed immediately after the conclusion of the challenge study—conditional on positive efficacy results—to evaluate the safety of the hypothetical vaccine candidate in a broader population. Assuming a single-arm study with 5,000 subjects followed for 30 days, we expect the process to be completed in 106 days. To accelerate licensure, we assume that the collection of safety data will be performed in parallel with BLA submission and FDA review. Since the latter is assumed to take 120 days, the additional safety study does not actually add to the time to licensure of the vaccine candidate. It does, however, add to the financial costs of the HCT (see Section 5.7.4).

Apart from the sample size, randomization ratio, set-up time, surveillance period, and safety data requirement, we assume that the HCT is identical to the RCT in all other respects. See Section 5.7.3 for a summary of our assumptions.

We anticipate similar post-marketing commitments for both the HCT and the RCTs, in terms of the collection of additional safety and effectiveness data, and supplementary studies to support the effectiveness of the vaccine in populations not included in the initial efficacy study, e.g., infants. However, we do not model them here because they do not affect our cost/benefit computations.

5.3 Epidemiological Model

To estimate the attack rate encountered by subjects in a given clinical trial—a key component for our cost/benefit calculations—we require information about the spread of the COVID-19 epidemic in the U.S. We use the Susceptible-Infected-Resolving-Dead-ReCovered with social distancing (SIRDC-SD) model proposed by Fernandez-Villaverde and Jones [16], chosen because it is able to fit both the cumulative and daily number of deaths in all the states well despite being a simple model, to establish a baseline for the epidemic.

to volunteers, and some individuals might have innately stronger immune systems that can counteract the virus.

SIRDC with Social Distancing (SIRDC-SD) Model

The model assumes that there is a constant population of N people. The number of people who are susceptible to infection, infected, resolving their infected status, dead, and recovered are denoted as S_t, I_t, R_t, D_t, and C_t respectively. By definition, the sum of all the people must be equals to N.

$$N = S_t + I_t + R_t + D_t + C_t \tag{5.1}$$

The dynamics of the epidemic are governed by the following differential equations:

$$\frac{dS_t}{dt} = -\frac{\beta(t) S_t I_t}{N} \tag{5.2}$$

$$\frac{dI_t}{dt} = \frac{\beta(t) S_t I_t}{N} - \gamma I_t \tag{5.3}$$

$$\frac{dR_t}{dt} = \gamma I_t - \theta R_t \tag{5.4}$$

$$\frac{dD_t}{dt} = \delta \theta R_t \tag{5.5}$$

$$\frac{dC_t}{dt} = (1-\delta)\theta R_t \tag{5.6}$$

Unlike most epidemiological models, the SIRDC-SD model assumes a contact rate parameter, $\beta(t)$, that decreases exponentially over time at a rate of λ from an initial value of β_0 to β^* instead of a static one.

$$\beta(t) = \beta_0 e^{-\lambda t} + \beta^*(1 - e^{\lambda t}) \tag{5.7}$$

This dynamic $\beta(t)$ incorporates the belief that social distancing over time will lead to a lower contact rate. This is particularly true in the U.S., where many cities have issued stay-at-home orders. Many people are also voluntarily wearing masks and are avoiding crowded places, which serve to reduce the contact rate.

The model also assumes that infections resolve at a Poisson rate γ, which implies that a person is infectious for a period of $1/\gamma$ on average. Thereafter, he will stop being infectious and transition into the 'resolving' state. Resolving cases will clear up at a Poisson rate of θ. There is an implicit assumption that people who recovered from the virus gain immunity to

the virus and cannot be reinfected.

Parameter Estimation/Calibration for SIRDC-SD Model

We estimate the model for each of the 50 states in the U.S. and Washington, D.C., using the time series of deaths in the U.S. obtained from the John Hopkins Center for Systems Science and Engineering (CSSE) COVID-19 repository [7]. Our data was downloaded on June 16, 2020. We do not scale the number of deaths but continue to perform a centered moving average smoothing on the daily number of deaths, as described in Fernandez-Villaverde and Jones [16].

Let D_t and d_t be the cumulative and daily number of deaths from data at time t, respectively. Let variables with hats denote the model's estimated values. We use the following optimization program to estimate the parameters of the model.

$$\underset{\beta_0,\beta^*,\lambda,I_0,\eta}{\text{minimize}} \quad \ln\left(\sum_t (D_t - \hat{D}_t)^2\right) + \ln\left(\sum_t (d_t - \hat{d}_t)^2\right) \tag{5.8}$$

subject to:

$$I_0 < N \ , \tag{5.9}$$

$$R_0 = \eta I_0 \ , \tag{5.10}$$

$$S_0 = N - R_0 - I_0 \ , \tag{5.11}$$

$$\beta_0 > \beta^* \ . \tag{5.12}$$

Our loss function is given by Equation 5.8, which says that we minimize the sum of 1) the natural logarithm of the sum of squared errors for the cumulative deaths, and 2) the natural logarithm of the sum of squared errors for the daily deaths. The minimization program is subjected to the four constraints. Equation 5.9 says that the initial number of infected must be less than the entire population. Equation 5.10 imposes that the number of initial resolving cases must be less than the number of initial infected cases. Equation 5.11 states that the conservation of population must hold at time = 0 and Equation 5.12 constrains the initial contact rate to be greater than the final contact rate.

We set γ, δ, and θ to 0.2, 0.008, and 0.1, respectively, as suggested by [16].

The optimization program is solved using the constrained Trust-Region algorithm as implemented in the SciPy Optimize package for each of the 50 U.S. states and Washington, D.C. Our estimated parameters for each state are reported in Table 5.1.

Table 5.1: Estimated parameters of the SIRDC model.

State	N	β_0	β^*	η	λ
Alabama	4,903,185	0.211	0.211	0.000	21.159
Alaska	731,545	0.799	0.000	0.947	0.430
Arizona	7,278,717	2.841	0.218	0.999	0.410
Arkansas	3,017,804	0.255	0.001	1.000	0.008
California	39,512,223	1.546	0.188	0.002	0.100
Colorado	5,758,736	1.961	0.188	0.511	0.149
Connecticut	3,565,287	3.006	0.177	0.006	0.169
Delaware	973,764	0.228	0.222	0.000	53.755
District of Columbia	705,749	0.699	0.171	0.999	0.078
Florida	21,477,737	1.712	0.185	0.975	0.122
Georgia	10,617,423	3.491	0.191	0.824	0.223
Hawaii	1,415,872	3.621	0.110	0.006	0.404
Idaho	1,787,065	2.871	0.134	0.994	0.462
Illinois	12,671,821	3.895	0.208	0.275	0.238
Indiana	6,732,219	1.270	0.188	0.993	0.128
Iowa	3,155,070	3.813	0.223	0.507	0.332
Kansas	2,913,314	1.594	0.157	0.379	0.132
Kentucky	4,467,673	4.129	0.185	0.140	0.269
Louisiana	4,648,794	4.324	0.181	0.370	0.257
Maine	1,344,212	7.164	0.169	0.991	0.962
Maryland	6,045,680	1.976	0.183	0.369	0.138
Massachusetts	6,892,503	2.258	0.182	0.412	0.148
Michigan	9,986,857	4.154	0.163	0.547	0.246
Minnesota	5,639,632	0.829	0.184	0.999	0.089
Mississippi	2,976,149	3.150	0.217	0.988	0.343
Missouri	6,137,428	0.882	0.189	1.000	0.125
Montana	1,068,778	0.149	0.149	1.000	3.169
Nebraska	1,934,408	4.622	0.201	0.541	0.396
Nevada	3,080,156	3.501	0.189	0.810	0.292
New Hampshire	1,359,711	1.506	0.221	0.866	0.236
New Jersey	8,882,190	2.766	0.179	0.048	0.130
New Mexico	2,096,829	0.421	0.148	1.000	0.043
New York	26,161,672	6.095	0.148	0.461	0.229
North Carolina	10,488,084	3.224	0.194	0.997	0.324

Continued on next page

Table 5.1 – continued from previous page

State	N	β_0	β^*	η	λ
North Dakota	762,062	1.789	0.213	0.984	0.391
Ohio	11,689,100	2.524	0.204	0.994	0.244
Oklahoma	3,956,971	3.219	0.168	0.867	0.316
Oregon	4,217,737	3.309	0.176	0.021	0.296
Pennsylvania	12,801,989	1.721	0.180	0.734	0.124
Rhode Island	1,059,361	3.872	0.214	1.000	0.499
South Carolina	5,148,714	2.219	0.192	0.488	0.180
South Dakota	884,659	0.587	0.000	0.999	0.021
Tennessee	6,829,174	0.198	0.196	0.000	84.504
Texas	28,995,881	5.141	0.200	0.279	0.311
Utah	3,205,958	1.390	0.212	0.999	0.447
Vermont	623,989	0.160	0.160	0.085	54.439
Virginia	8,535,519	6.097	0.216	0.000	0.315
Washington	7,614,893	1.490	0.175	0.968	0.138
West Virginia	1,792,147	0.194	0.193	0.000	26.549
Wisconsin	5,822,434	9.799	0.188	0.618	0.556
Wyoming	578,759	0.160	0.160	1.000	6.478

SIRDCV Model

The estimated models show how the epidemic has played out thus far but we will need to predict how it will evolve in the future after the lockdowns are relaxed and/or vaccines are developed. To do so, we extend the SIRDC-SD model to take into account semi-effective vaccination. The new model, which we shall name Susceptible-Infected-Resolving-Dead-ReCovered-Vaccinated with social distancing (SIRDCV).

We let $\bar{V}$ and ϵ be the number of persons vaccinated at every time step and the *effectiveness* of the vaccine, respectively. Effectiveness is defined as the performance of the vaccine under real-world conditions in a general population whereas efficacy is defined as the ability to protect against a virus under ideal conditions in a homogeneous population. The former is usually is less than the latter due to several reasons, e.g., improper storage of vaccines leading to loss of potency and non-compliance with the vaccine dosing schedule. For simplicity, we assume that the effectiveness of the vaccine in the epidemiological model is identical to the efficacy of the vaccine in the clinical trials. V_t^r and V_t^{nr} represent the stock of people

who are inoculated, and respond (r) and do not respond (nr) to the vaccine, respectively.

$$\frac{dS_t}{dt} = -\frac{\beta(t)S_tI_t}{N} - \bar{V} \tag{5.13}$$

$$\frac{dI_t}{dt} = \frac{\beta(t)(S_t + V_t^{nr})I_t}{N} - \gamma I_t \tag{5.14}$$

$$\frac{dV_t^{nr}}{dt} = (1-\epsilon)\bar{V} - \frac{\beta(t)V_t^{nr}I_t}{N} \tag{5.15}$$

$$\frac{dV_t^{r}}{dt} = \epsilon\bar{V} \tag{5.16}$$

$$\frac{dR_t}{dt} = \gamma I_t - \theta R_t \tag{5.17}$$

$$\frac{dD_t}{dt} = \delta\theta R_t \tag{5.18}$$

$$\frac{dC_t}{dt} = (1-\delta)\theta R_t \tag{5.19}$$

Equation 5.2 has been modified to remove vaccinated persons at every time step in Equation 5.13. We also modify Equation 5.3 to allow people who are vaccinated but do not respond to the inoculation to be infected in Equation 5.14. Equation 5.15 and Equation 5.16 keep track of the stock of people who are vaccinated. With this specification, the virus is allowed to spread even when the entire population is vaccinated because not everyone will respond to the mass inoculation.

Evolution of Epidemic with Reopening

We consider three different scenarios for the evolution of the epidemic over time. In the first, we assume that the current situation will continue indefinitely until the end of the epidemic ("status quo"). That is, stay-home orders and bans on social gatherings will be extended until there are no new infections. We simply forecast ahead of time using the estimated dynamic $\beta(t)$ in this scenario.

In the second, we consider that there will be a partial reopening with strict monitoring across all states starting from June 15, 2020 ("ramp"). To model this, we assume a ramp function for $\beta(t)$ that will increase to 0.22 over 90 days and maintain at that level until the end of the epidemic. The parameters are chosen to imply a final R_0 of 1.1, which reflects close monitoring and contact tracing, and if needed, temporary quarantines to arrest clusters of infections that may pop up. The contact rate parameter, β, in this scenario is described

by Equation 5.20.

$$\beta'(t) = \begin{cases} \beta(t) & \forall t < T_v \\ \beta(T_v) + \dfrac{\beta_{ss} - \beta(T_v)}{90} t & \forall T_v \leq t \leq T_v + 90 \\ \beta_{ss} & \text{otherwise} \end{cases} \tag{5.20}$$

In the third, we consider the behavioral-based response proposed by John Cochrane ("behavioral"), whereby people voluntarily reduce social contact when they perceive danger (e.g., when they observe that there is an uptick in the daily number of deaths) and increase social contact when they observe that there is a decrease in risk (e.g., when they observe a reduction in the daily number of deaths) [10]. This scenario is modeled by making the percentage change in contact rate parameter negatively proportionate to the change in the observed death rate over an interval of t_o. That is,

$$\frac{1}{\beta} \frac{d\beta}{d(\frac{\Delta D}{N})} = -k \tag{5.21}$$

Integrating Equation 5.21 will yield Equation 5.22.

$$\ln \beta = c - k \frac{\Delta D}{N} = c - k \frac{D_t - D_{t-t_o}}{N} \tag{5.22}$$

The exponent of c is the long term steady-state value of β. k can be interpreted as the percentage increase/decrease in β if there is a decrease/increase in the death rate. In our simulations, t_0, c, and k are set to 7, $\ln \beta_{ss}$, and 50,000, respectively. The default scenario of $c = \ln 0.2$ will correspond to a R_0 of 1 when approximately 16,000 deaths per week are observed in the U.S. This behavior will start immediately on June 15, 2020, to be consistent with the second scenario.

The new contact rate parameter in this case is defined by Equation 5.23.

$$\beta'(t) = \begin{cases} \beta(t) & \forall t < T_v \\ e^{c - k \frac{D_t - D_{t-t_o}}{N}} & \text{otherwise} \end{cases} \tag{5.23}$$

We give an example of how the basic reproduction number, or R_0, may look for each of the scenarios in Figure 5-2.

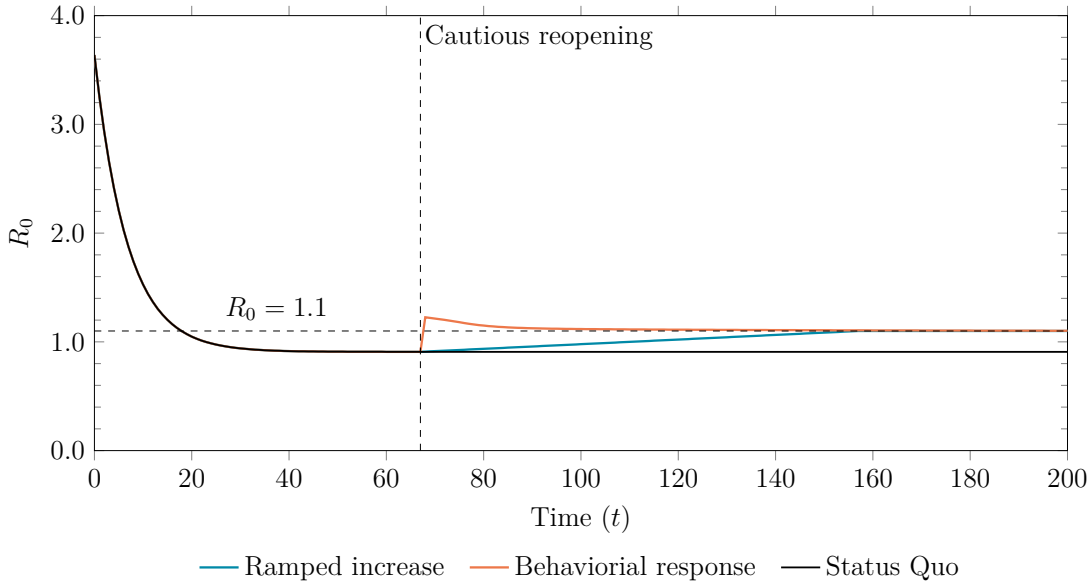


Figure 5-2: An illustration of how the $R_0 = \beta/\gamma$ changes over time for each of the three scenarios: status quo, a ramp increase, and behavioral-based response.

5.3.1 Population Vaccination Schedule

We assume that vaccines will be immediately available for distribution and inoculation upon licensure. This reflects how the leading vaccine companies have been scaling up their manufacturing capabilities and started producing millions of doses at industrial scale in parallel to the clinical trials [13, 38] and well before the demonstration of vaccine efficacy and safety. We model three ways that the susceptible population will be vaccinated upon vaccine licensure: 1M, 10M, and infinite doses administered per day. In the last case, the entire U.S. population is assumed to be vaccinated the day after licensure. While unrealistic, this gives an upper bound on the potential benefit of vaccine development. We assume that the vaccines are distributed proportionally to states according to their relative population at the start of the epidemic.

5.3.2 Forecasting Infections and Deaths

We forecast the cumulative number of infections and deaths in each state between February 29, 2020, and December 31, 2022, using the SIRDCV described by Equation 5.13 to Equation 5.19 before summing over all states in order to produce estimates for the entire

U.S. The attack rate at time t is the ratio of the number of new infections at time t to the number of susceptible persons at time $t-1$. Illustrations of how the cumulative number of infections and deaths change over time given the different evolution paths of the epidemic and vaccination schedules are shown in Figure 5-3

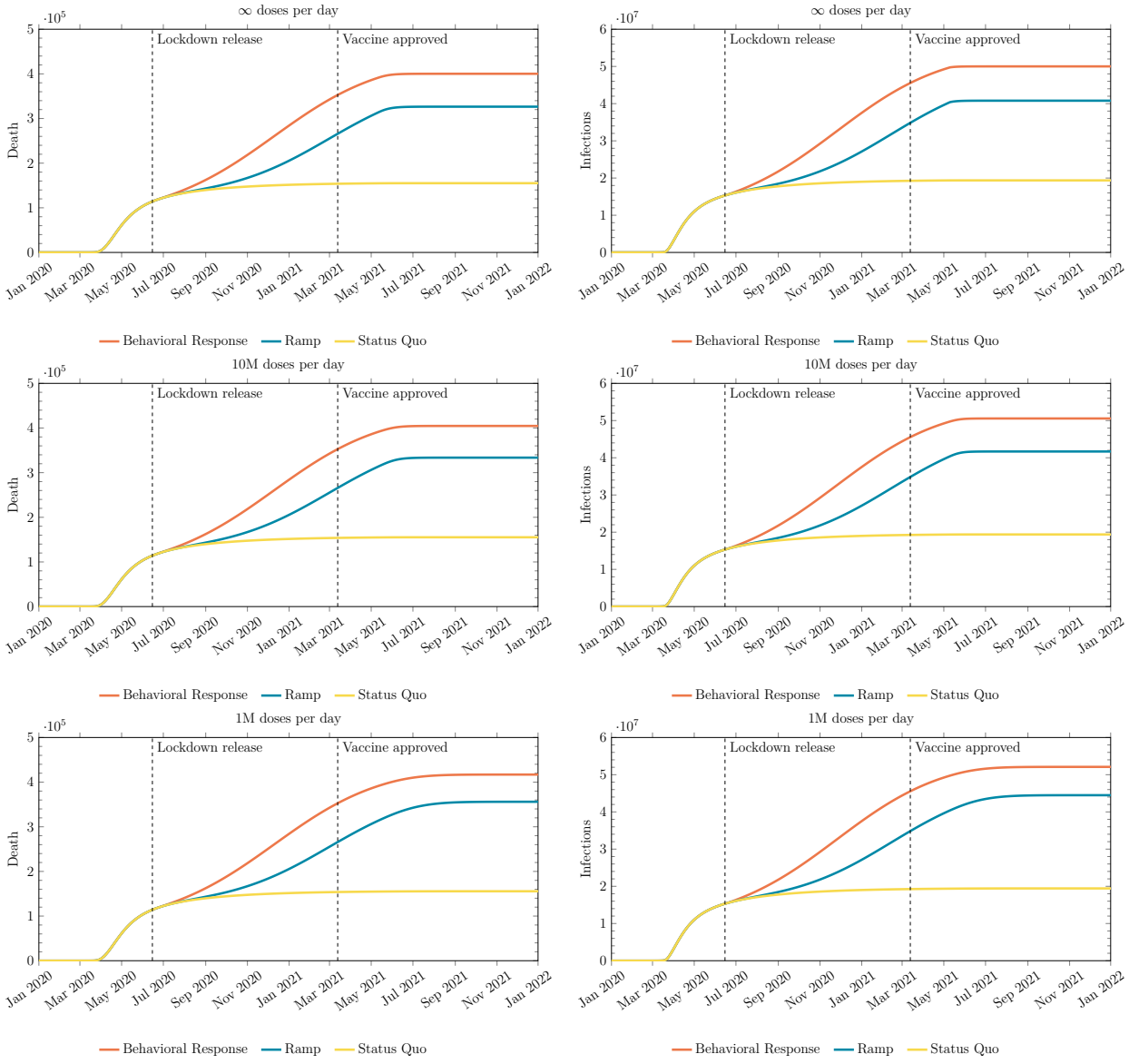


Figure 5-3: Illustration of how the cumulative number of infections and deaths change over time given the different evolution paths of the epidemic and vaccination schedules. We assume that the epidemic evolves based on our scenarios after June 15, 2020, and that the vaccine is approved on March 13, 2021. The vaccine efficacy assumed is 50%.

5.4 Results

Given the parameters for each trial design and an epidemiological model, we simulate the outcome of hypothetical clinical trials for all four designs and measure their incremental differences. Our cost/benefit methodology is described in Section 5.4.1, we report the numerical results in Section 5.4.2, and discuss them in Section 5.5.

5.4.1 Cost/Benefit Analysis

We apply cost benefit analysis to quantify and compare the net value of each trial design. We focus on public health outcomes—that is, the risks of mortality and morbidity—and provide a qualitative discussion of the societal and financial impact in Section 5.5.

As shown by Montazerhodjat et al. [39], Isakov et al. [24], and Chaudhuri et al. [9], the value associated with a pathway is computed as the difference between the post-trial benefit and the in-trial cost (Equation 5.24). The former estimates the net benefit of the trial to society at large while the latter measures the cost of conducting the study to volunteers in the trial.

$$\text{Net Value} = \text{Post-trial Benefit} - \text{In-trial Cost} \tag{5.24}$$

We quantify the cost of a trial design in terms of the number of COVID-19 infections and deaths observed in the clinical study. For post-trial benefit, we first consider a baseline scenario in which a vaccine is never developed and the epidemic is allowed to run its course. Next, we simulate the case where a vaccine is approved at some point in time depending on the duration of the trial design. The post-trial benefit is then the difference in the cumulative number of infections and deaths in the population between the two scenarios, i.e., the incremental number of infections and deaths prevented with a vaccine licensure. In cases where the vaccine candidate is rejected,[7] net value will be negative since post-trial benefit is zero but cost has been incurred for conducting the clinical trial. Lastly, we assume

[7]In our simulations, we consider a vaccine candidate with some efficacy ϵ and assume that infections in the clinical study follow a stochastic process (e.g., binomial distribution). Due to this randomness, false rejections of the efficacious vaccine might occur. This is also known as type II error. The false negative rate depends on the trial design (e.g., sample size, surveillance period, maximum type I error, superiority testing) and the epidemiological model (e.g., attack rate in the clinical study).

that the hypothetical vaccine candidate is generally well tolerated and any vaccine-related adverse reactions are mild and negligible with respect to in-trial costs and post-trial benefits [18, 25, 42].

5.4.2 Simulation Results

We compute the expected net value of different trial designs using Monte Carlo simulations and asymptotic distributions of the efficacy test statistics (see Section 5.7.1). Figure 5-1 illustrates the inputs, computations, and outputs of our simulation framework. We assume that all trials start on August 1, 2020, and simulate the epidemiological models until December 31, 2022. We perform sensitivity analysis over a wide range of trial design, epidemiological model, and population vaccination schedule assumptions (see Table 5.2), covering 756 different scenarios. We summarize our results in Table 5.3 and Section 5.7.6. In addition to our results, we release an open-source version of our simulation software, and encourage readers to rerun our simulations with their own preferred set of assumptions and inputs.

Assuming superiority testing and a vaccine efficacy of 50%, we estimate the date of licensure of the hypothetical vaccine candidate to be some time in November 2021 under an RCT (476 days), between June and August 2021 under an ORCT (326 to 380 days), between April and June 2021 under an ARCT (246 to 306 days), and between March and June 2021 under an HCT (221 to 311 days). Apart from an RCT which has a fixed trial duration, the dates of licensure from the ORCT and ARCT depend largely on the status of the epidemic during the clinical trial. If the transmission rate of the disease is low (e.g., due to social distancing or other non-pharmaceutical interventions), an extended surveillance period is required to accrue enough natural infections in order to observe a statistically significant difference in infection risk between the treatment arm and the control arm. Conversely, when the transmission rate is high, a short surveillance period is sufficient to observe significant results. We note that an HCT, on the other hand, does not depend on the epidemic situation but is instead limited by the time required to set up the challenge model. In general, we find that the time to licensure under ORCT and ARCT decreases with increasing vaccine efficacy: the greater the efficacy, the easier it is to observe a significant treatment effect.

We find that the ARCT provides the greatest expected net benefit among the three RCT

Table 5.2: Sensitivity analysis with respect to trial design, epidemiological model, and population vaccination schedule assumptions.

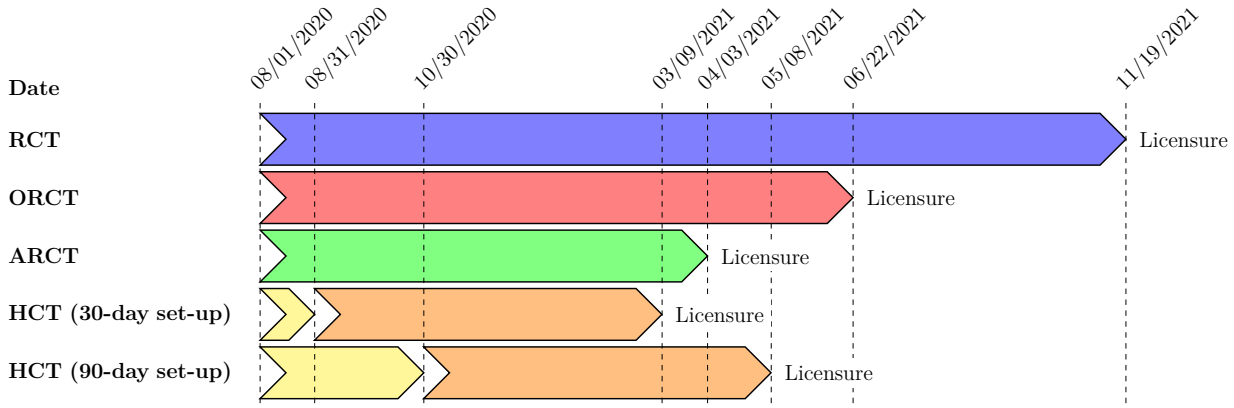
Parameter	**Values**
Trial design	RCT, ORCT, ARCT, HCT
Vaccine efficacy of hypothetical candidate (%)	30, 50, 70, 90
Set-up time for HCT (days)	30, 60, 90, 120
Efficacy requirement	Superiority, superiority by margin of 50% [5]
Epidemiological scenario	Status quo, ramp, behavioral
Population vaccination schedule (doses/day)	1M, 10M, infinite

designs in almost all scenarios. The utility of an HCT versus the RCTs, however, depends critically on the set-up time and the dynamics of the epidemic. For example, assuming superiority testing, a vaccine efficacy of 50%, the behavioral epidemiological model, and a population vaccination schedule of 10M doses per day, we estimate that the ARCT can help accelerate licensure by almost 8 months versus the RCT, thus preventing approximately 2.9M incremental infections and 23,000 incremental deaths from COVID-19 in the U.S. versus the latter.

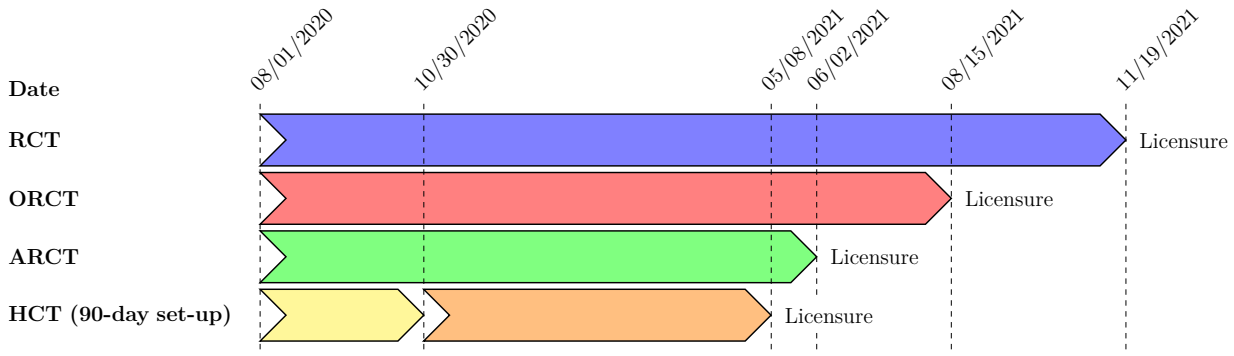
Under the same set of assumptions, an HCT that requires 30 days to set up can *further* reduce the time to licensure by a month, thus preventing approximately 1.1M more infections and 8,000 more deaths versus the ARCT. However, the advantage of the HCT vanishes when its set-up time is long: an HCT that requires 90 days to set up takes about one month longer to reach licensure as compared to the ARCT, leading to around 1.0M more infections and 8,000 more deaths versus the latter (see Figure 5-4a). Under such circumstances, the use of an HCT is worthwhile only when the prevalent transmission rate is low. If we consider the status quo scenario instead of the behavioral epidemiological model, the time to licensure is about one month shorter under the HCT than under the ARCT even with a 90 day set-up period (see Figure 5-4b). In this case, the HCT prevents approximately 60,000 incremental infections and 500 incremental deaths versus the ARCT. We observe similar trends under superiority-by-margin testing at a threshold of 50%.

Table 5.3: Expected number of incremental infections and deaths avoided in the U.S. under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority testing, and 10M doses of a vaccine per day are available after licensure, compared to the baseline case in which no vaccine is ever approved.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$
Status Quo								
RCT	3,914	31	11,539	92	19,130	151	21,557	170
ORCT	5,589	45	16,802	134	33,757	269	50,288	401
ARCT	9,596	76	31,473	250	66,641	531	83,522	665
HCT (30-day set-up)	140,731	1,124	152,263	1,216	156,885	1,254	159,876	1,277
HCT (60-day set-up)	110,046	879	118,937	950	122,482	979	124,777	997
HCT (90-day set-up)	86,466	690	93,370	745	96,111	768	97,886	782
HCT (120-day set-up)	68,213	544	73,611	587	75,747	605	77,132	615
Behavioral								
RCT	363,382	2,845	386,081	3,026	397,396	3,117	404,562	3,174
ORCT	1,139,585	9,061	1,377,157	10,955	1,426,014	11,345	1,457,500	11,598
ARCT	2,588,881	20,647	3,248,449	25,924	3,389,541	27,052	3,473,035	27,720
HCT (30-day set-up)	3,903,566	31,167	4,309,316	34,411	4,481,448	35,789	4,591,750	36,671
HCT (60-day set-up)	2,795,316	22,301	3,082,676	24,598	3,205,159	25,579	3,283,975	26,209
HCT (90-day set-up)	2,011,244	16,028	2,211,985	17,633	2,297,350	18,316	2,352,436	18,757
HCT (120-day set-up)	1,466,239	11,668	1,605,833	12,784	1,664,613	13,255	1,702,601	13,558
Ramp								
RCT	1,075,634	8,316	1,131,531	8,764	1,160,564	8,996	1,179,234	9,145
ORCT	2,853,202	22,569	3,839,945	30,432	3,973,769	31,501	4,050,013	32,111
ARCT	5,711,310	45,401	7,442,922	59,253	7,924,650	63,107	8,071,866	64,285
HCT (30-day set-up)	8,744,377	69,672	9,452,413	75,330	9,725,022	77,511	9,897,591	78,892
HCT (60-day set-up)	6,814,762	54,235	7,381,425	58,762	7,602,878	60,534	7,743,514	61,659
HCT (90-day set-up)	5,266,925	41,851	5,711,663	45,404	5,887,421	46,811	5,999,381	47,706
HCT (120-day set-up)	4,053,134	32,141	4,396,033	34,879	4,532,400	35,970	4,619,521	36,667



(a) Under the behavioral epidemiological model.



(b) Under the status quo epidemiological model.

Figure 5-4: Dates of licensure under RCT, ORCT, ARCT, HCT (30-day set-up time), and HCT (90-day set-up time), assuming superiority testing, a vaccine efficacy of 50%, and a population vaccination schedule of 10M doses per day.

5.5 Discussion

There has been a plethora of papers highlighting various ethical considerations for conducting HCTs [3, 20], some specifically for COVID-19 [11, 14, 26, 44, 50, 58]. Some of the main ethical concerns are: (1) what is the explicit scientific rationale for, and societal value of, an HCT; (2) whether the risks of harm to the subjects and the public at large are understood by the scientists and have been minimized; (3) whether informed consents have been obtained from subjects after they are given full disclosures of the risks involved; and (4) whether the subjects have been selected fairly and given appropriate compensation for both the risk and actual harm brought on by HCTs. Most bioethicists generally accept that these concerns can be addressed within the existing ethical framework for human medical research.

Our paper addresses the first and second of these ethical concerns. We provide scientific justifications for COVID-19 HCTs by considering how conducting them can allow companies to learn about the protection curves and accelerate the development of vaccines against SARS-CoV-2.

However, our analysis does not address the latter two ethical considerations as they concern the execution of HCTs, which is beyond the scope of this paper. Nonetheless, companies and scientists seeking to perform HCTs, and especially regulators, will have to address those concerns to preserve public trust and avoid a public backlash that could jeopardize other important medical research critical to addressing the current epidemic.

Some scientists argue that "a single death or severe illness in an otherwise healthy volunteer would be unconscionable" [11]. However, it can be argued that allowing tens of thousands of individuals to die by denying the consent of an informed individual to take a calculated risk is equally unconscionable. In this paper, we adopt the Benthamite approach [4], where every individual's utility is weighted equally in the aggregate utility function, as is the common convention in public economics analyses. Within this ethical perspective, our calculations show that an HCT can potentially provide substantial public health benefits in terms of accelerating vaccine development and reducing the burden of coronavirus-related mortality and morbidity in the U.S. —in some cases, by more than 1.1M infections and 8,000 deaths compared to the best performing RCT—when conducted early in the pan-

demic's life cycle and in cases where the spread of COVID-19 in the population is muted due to non-pharmaceutical interventions.

We also expect the financial costs of an HCT—which includes the cost of liability protection—to be lower than those of a traditional vaccine efficacy RCT, adding further support for a challenge design (see Section 5.7.4 for further discussion). While we have focused on public health outcomes here, it is clear that accelerated vaccine development provides tremendous societal and economic benefits as well—e.g., savings in insured medical costs, direct medical expenditures, and hospitalization costs, and accelerated economic recovery from an earlier reopening.

We emphasize that the expected costs and benefits of a clinical trial depend critically on many assumptions about existing conditions. For example, recruiting subjects in sufficient numbers and diversity can sometimes present a challenge for clinical trials involving experimental vaccines (although, in the case of HCTs for COVID-19, the organization 1Day Sooner reports over 32,000 registered volunteers as of July 27, 2020). Also, we do not include set-up time for non-challenge RCTs because phase 3 vaccine efficacy trials are already imminent as of now. Moreover, we assume a relatively short set-up time for HCTs because a challenge study can be set up relatively quickly using a wild-type strain [56], and the National Institute of Allergy and Infectious Diseases (NIAID) appears to have already made some headway in manufacturing challenge doses [48]. If, instead, we assume comparable set-up times (e.g., two months) and start dates for both an HCT and non-challenge RCTs, we expect that an HCT can accelerate licensure by two months when compared to an adaptive RCT.[8] Some have argued that at least one to two years is required to develop a robust model from scratch [11]. In this case, our results indicate that an ARCT will almost always be faster than an HCT. However, even if an HCT with a long set-up time does not lead to faster vaccine licensures over an ARCT given current conditions, the creation of a standing HCT agent and setting up an HCT now can provide a hedge against potential failures in the current crop of vaccine candidates. By having an approved, ready-to-go challenge virus and ready-to-go HCT sites that vaccine developers can access immediately, the approval process for as-yet-untested SARS-CoV-2 vaccine candidates can be accelerated when required. For

[8] Assuming superiority testing, a vaccine efficacy of 50%, and the behavioral epidemiological model.

a pandemic like COVID-19, such a hedge will almost always show substantial net benefits relative to its costs.

HCTs have several other benefits that will be more obvious as the pandemic progresses. They require many fewer eligible volunteers, whose numbers will dwindle as the pandemic progresses. They do not depend on attack rates at clinical trial sites which are notoriously difficult to estimate and highly dependent on non-pharmaceutical interventions such as lockdowns and other social-distancing policies. They also avoid logistical problems such as identifying subjects, obtaining subjects' consent, obtaining institutional review board's approval or tracking subjects, particularly when attempting large-scale clinical trials in places where contract research organizations (CROs) have little experience.

It is conceivable that multiple vaccines—instead of the single vaccine in our simulation study—are tested concurrently in a single trial design [57]. For example, five vaccines, such as those selected by Operation Warp Speed [28], could be tested concurrently in a six-arm trial (five vaccine arms and a control arm), requiring 40% fewer test subjects, thereby reducing in-trial expected morbidity and mortality costs by the same amount. The benefits can be increased if an adaptive platform clinical trial—designed to eliminate ineffective vaccines at the first signs of futility—is adopted. A clinical trial testing multiple vaccines can also reduce competition for volunteers, a problem that continues to plague vaccine developers [22].

We choose to quantify the cost and benefits of the clinical trials by measuring the number of infections and deaths avoided, and refrain from performing a traditional health technology assessment, such as comparing the economic value of an HCT versus an RCT using quality-adjusted life years measures or willingness to pay estimates such as the value of a statistical life. Performing such computations is straightforward given the output of our simulations, but we have refrained from doing so in deference to non-economist stakeholders who find it offensive to use any pecuniary measures when discussing the loss of human life.

Finally, our analysis focuses mainly on the U.S. for practical reasons involving access to data with which to calibrate our simulations and the broader goal of informing U.S. public health officials and policymakers as the country enters the final stages of vaccine development. However, a vaccine licensure may apply internationally. Given that the U.S. currently comprises 25% of all confirmed COVID-19 cases (as of July 7, 2020) [7], if the

assumptions made in our study also hold internationally, the net benefits for all the clinical trials will scale by a factor of 4, in which case HCTs can save an additional 4.4M infections and 32,000 deaths compared to the best performing RCT in certain situations.

We highlight that these figures depend heavily on the development of the epidemic in the U.S. moving forward. We have considered three simple scenarios, status quo, ramp, and behavioral, corresponding to low transmission, moderate transmission, and behavioral-based response, respectively. There are clearly many other sources of uncertainty that are not reflected here. For example, non-adherence to social distancing advisories and/or resistance to precaution recommendations such as wearing a mask in public will lead to an uncontrolled outbreak, which will help to accelerate non-challenge RCTs, making them attractive even when compared to an HCT with a short set-up time. We have found it difficult and impractical to incorporate these uncertainties in our assumptions due to the speed at which things are evolving and the unpredictability of public reaction. In addition, studies that have attempted to incorporate such uncertainties in their epidemic model report huge error bounds in their projections [46]. The wide confidence intervals prevent us from drawing any useful conclusions, which severely limit the usefulness of such models. Therefore, we recommend readers not to take our results as final or definitive, but to re-run our simulations with their own preferred set of assumptions, calibrated using the most current epidemiological data.

5.6 Conclusion

Our paper presents a systematic framework for quantitatively accessing the in-trial and societal cost/benefit trade-offs of various clinical trial designs in terms of infections and deaths averted. We hope that this framework will allow stakeholders to make more informed practical and ethical decisions regarding accelerating COVID-19 vaccine development in the ongoing pandemic.

Chapter References

[1] Yasmeen Abutaleb, Josh Dawsey, Laurie McGinley, and Carolyn Y Johnson. Trump pushing officials to speed up already-ambitious coronavirus vaccine timeline. *Washington Post*, Jun 2020. URL `https://www.washingtonpost.com/health/2020/06/17/trump-coronavirus-vaccine`. Accessed: 2020-06-15.

[2] AstraZeneca. Astrazeneca advances response to global covid-19 challenge as it receives first commitments for oxford's potential new vaccine, May 2020. URL `https://www.astrazeneca.com/media-centre/press-releases/2020/astrazeneca-advances-response-to-global-covid-19-challenge-as-it-receives-first-commitments-for-oxfords-potential-new-vaccine.html`. Accessed: 2020-07-03.

[3] Ben Bambery, Michael Selgelid, Charles Weijer, Julian Savulescu, and Andrew J Pollard. Ethical criteria for human challenge studies in infectious diseases. *Public Health Ethics*, 9(1):92–103, 2016.

[4] Jeremy Bentham. *A fragment on government*. The Lawbook Exchange, Ltd., 2001.

[5] Thomas M. Burton. Fda to require proof virus vaccine is effective before approving its use. *The Wall Street Journal*, Jun 2020. URL `https://www.wsj.com/articles/fda-to-issue-guidance-on-covid-19-vaccine-approval-11593516090`.

[6] Business Wire. Pfizer and biontech granted fda fast track designation for two investigational mrna-based vaccine candidates against sars-cov-2, July 2020. URL `https://www.businesswire.com/news/home/20200713005168/en/Pfizer-BioNTech-Granted-FDA-Fast-Track-Designation`. Accessed: 2020-07-14.

[7] Center for Systems Science and Engineering (CSSE). COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. `https://github.com/CSSEGISandData/COVID-19`, 2020.

[8] Centers for Disease Control and Prevention. Covid-19 provisional counts - weekly updates by select demographic and geographic characteristics, May 2020. URL `https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm`. Accessed: 2020-06-22.

[9] Shomesh E Chaudhuri, Martin P Ho, Telba Irony, Murray Sheldon, and Andrew W Lo. Patient-centered clinical trials. *Drug discovery today*, 23(2):395–401, 2018.

[10] John Cochrane. An SIR model with behavior, May 2020. URL `https://johnhcochrane.blogspot.com/2020/05/an-sir-model-with-behavior.html`. Accessed: 2020-07-17.

[11] Meagan E Deming, Nelson L Michael, Merlin Robb, Myron S Cohen, and Kathleen M Neuzil. Accelerating development of sars-cov-2 vaccines—the role for controlled human infection models. *New England Journal of Medicine*, 2020.

[12] Jillian Deutsch. Von der leyen: Life won't return to normal until vaccine. *Politico*, Apr 2020. URL `https://www.politico.eu/article/ursula-von-der-leyen-vaccine/`. Accessed: 2020-06-25.

[13] Emergent BioSolutions. Emergent biosolutions signs agreement to be u.s. manufacturing partner for astrazeneca's covid-19 vaccine candidate, Jun 2020. URL `https://investors.emergentbiosolutions.com/news-releases/news-release-details/emergent-biosolutions-signs-agreement-be-us-manufacturing-0`. Accessed: 2020-07-14.

[14] Nir Eyal, Marc Lipsitch, and Peter G Smith. Human challenge studies to accelerate coronavirus vaccine licensure. *The Journal of infectious diseases*, 221(11):1752–1756, 2020.

[15] Conor P Farrington and Godfrey Manning. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in medicine*, 9(12):1447–1454, 1990.

[16] Jesus Fernandez-Villaverde and Charles I Jones. Estimating and simulating a sird model of covid-19 for many countries, states, and cities. Technical report, National Bureau of Economic Research, 2020.

[17] Joseph L Fleiss, Bruce Levin, and Myunghee Cho Paik. *Statistical methods for rates and proportions*. john wiley & sons, 2013.

[18] Pedro M Folegatti, Katie J Ewer, Parvinder K Aley, Brian Angus, Stephan Becker, Sandra Belij-Rammerstorfer, Duncan Bellamy, Sagida Bibi, Mustapha Bittaye, Elizabeth A Clutterbuck, et al. Safety and immunogenicity of the chadox1 ncov-19 vaccine against sars-cov-2: a preliminary report of a phase 1/2, single-blind, randomised controlled trial. *The Lancet*, 396(10249):467–478, 2020.

[19] Food and Drug Administration. Development and licensure of vaccines to prevent covid-19, Jun 2020. URL `https://www.fda.gov/media/139638/download`. Accessed: 2020-07-10.

[20] G Miller Franklin and Christine Grady. The ethical challenge of infection-inducing challenge experiments. *Clinical Infectious Diseases*, 33(7):1028–1033, 2001.

[21] Wei Jie Guan, Zheng Yi Ni, Yu Hu, Wen Hua Liang, Chun Quan Ou, Jian Xing He, Lei Liu, Hong Shan, Chun Liang Lei, David SC Hui, et al. Clinical characteristics of coronavirus disease 2019 in china. *New England journal of medicine*, 382(18):1708–1720, 2020.

[22] Jared S Hopkins and Peter Loftus. Coronavirus researchers compete to enroll subjects for vaccine tests, Jul 2020. URL `https://www.wsj.com/articles/coronavirus-researchers-compete-to-enroll-subjects-for-vaccine-tests-11593968711`. Accessed: 2020-07-07.

[23] International Federation of Pharmaceutical Manufacturers & Associations. The complex journey of a vaccine. Technical report, International Federation of Pharmaceutical Manufacturers & Associations, Jul 2019. URL `https://www.ifpma.org/wp-content/uploads/2019/07/IFPMA-ComplexJourney-2019_FINAL.pdf`.

[24] Leah Isakov, Andrew W Lo, and Vahid Montazerhodjat. Is the fda too conservative or too aggressive?: A bayesian decision analysis of clinical trial design. *Journal of econometrics*, 211(1):117–136, 2019.

[25] Lisa A Jackson, Evan J Anderson, Nadine G Rouphael, Paul C Roberts, Mamodikoe Makhene, Rhea N Coler, Michele P McCullough, James D Chappell, Mark R Denison,

Laura J Stevens, et al. An mrna vaccine against sars-cov-2—preliminary report. *New England Journal of Medicine*, 383(20):1920–1931, 2020.

[26] Euzebiusz Jamrozik and Michael J Selgelid. Covid-19 human challenge studies: ethical issues. *The Lancet Infectious Diseases*, 2020.

[27] Christopher Jennison and Bruce W Turnbull. Group-sequential analysis incorporating covariate information. *Journal of the American Statistical Association*, 92(440):1330–1341, 1997.

[28] Alex Keown. 'operation warp speed' narrows list of potential covid-19 vaccine candidates down to five, Jun 2020. URL `https://www.biospace.com/article/white-house-narrows-covid-19-vaccine-candidates-down-to-5-for-operation-warp-speed/`. Accessed: 2020-07-02.

[29] Robert D Kirkcaldy, Brian A King, and John T Brooks. Covid-19 and postinfection immunity: Limited evidence, many remaining questions. *JAMA*, 2020.

[30] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9):577–582, 2020.

[31] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*, 2020.

[32] Donald W Light, Jon Kim Andrus, and Rebecca N Warburton. Estimated research and development costs of rotavirus vaccines. *Vaccine*, 27(47):6627–6633, 2009.

[33] Andrew W. Lo, Kien Wei Siah, and Chi Heem Wong. Estimating probabilities of success of vaccine and other anti-infective therapeutic development programs. *Harvard Data Science Review*, May 2020. URL `https://doi.org/10.1162/99608f92.e0c150e8`.

[34] Nicole Lurie, Melanie Saville, Richard Hatchett, and Jane Halton. Developing covid-19 vaccines at pandemic speed. *New England Journal of Medicine*, 382(21):1969–1973, 2020.

[35] Moderna. Cove study: Participate to make a world of difference, 2020. URL `https://www.modernatx.com/cove-study`. Accessed: 2020-11-23.

[36] Moderna. Moderna announces ind submitted to u.s. fda for phase 2 study of mrna vaccine (mrna-1273) against novel coronavirus, Apr 2020. URL `https://investors.modernatx.com/news-releases/news-release-details/moderna-announces-ind-submitted-us-fda-phase-2-study-mrna`. Accessed: 2020-06-29.

[37] Moderna. Moderna advances late-stage development of its vaccine (mrna-1273) against covid-19, Jun 2020. URL `https://investors.modernatx.com/news-releases/news-release-details/moderna-advances-late-stage-development-its-vaccine-mrna-1273`. Accessed: 2020-07-01.

[38] Moderna. Moderna and catalent announce collaboration for fill-finish manufacturing of moderna's covid-19 vaccine candidate, Jun 2020. URL `https://investors.modernatx.com/news-releases/news-release-details/moderna-and-catalent-announce-collaboration-fill-finish`. Accessed: 2020-07-01.

[39] Vahid Montazerhodjat, Shomesh E Chaudhuri, Daniel J Sargent, and Andrew W Lo. Use of bayesian decision analysis to minimize harm in patient-centered randomized clinical trials in oncology. *JAMA oncology*, 3(9):e170123–e170123, 2017.

[40] Thomas J Moore, Hanzhe Zhang, Gerard Anderson, and G Caleb Alexander. Estimated costs of pivotal trials for novel therapeutic agents approved by the us food and drug administration, 2015-2016. *JAMA internal medicine*, 178(11):1451–1457, 2018.

[41] Jozef Nauta. Statistics in clinical and observational vaccine studies, 2020.

[42] Anthony T Newall, Nathorn Chaiyakunapruk, Philipp Lambach, and Raymond CW

Hutubessy. Who guide on the economic evaluation of influenza vaccination. *Influenza and other respiratory viruses*, 12(2):211–219, 2018.

[43] Graziano Onder, Giovanni Rezza, and Silvio Brusaferro. Case-fatality rate and characteristics of patients dying in relation to covid-19 in italy. *Jama*, 323(18):1775–1776, 2020.

[44] Stanley A Plotkin and Arthur Caplan. Extraordinary diseases require extraordinary solutions. *Vaccine*, 38(24):3987, 2020.

[45] Stuart J Pocock. Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2):191–199, 1977.

[46] Debashree Ray, Maxwell Salvatore, Rupam Bhattacharyya, Lili Wang, Jiacong Du, Shariq Mohammed, Soumik Purkayastha, Aritra Halder, Alexander Rix, Daniel Barker, Michael Kleinsasser, Yiwang Zhou, Debraj Bose, Peter Song, Mousumi Banerjee, Veerabhadran Baladandayuthapani, Parikshit Ghosh, and Bhramar Mukherjee. Predictions, role of interventions and effects of a historic national lockdown in india's response to the the covid-19 pandemic: Data science call to arms. *Harvard Data Science Review*, 6 2020. doi: 10.1162/99608f92.60e08ed5. URL `https://hdsr.mitpress.mit.edu/pub/r1qq01kw`. https://hdsr.mitpress.mit.edu/pub/r1qq01kw.

[47] Meriel Raymond, Malick M Gibani, Nicholas PJ Day, and Phaik Yeong Cheah. Typhoidal salmonella human challenge studies: ethical and practical challenges and considerations for low-resource settings. *Trials*, 20(2):1–7, 2019.

[48] Owermohle Sarah. White house pressure for a vaccine raises risk the u.s. will approve one that doesn't work, Jun 2020. URL `https://www.politico.com/news/2020/06/15/pressure-coronavirus-vaccine-risk-approval-316094`. Accessed: 2020-07-15.

[49] Jeffrey Seow, Carl Graham, Blair Merrick, Sam Acors, Kathryn J.A. Steel, Oliver Hemmings, Aoife O'Bryne, Neophytos Kouphou, Suzanne Pickering, Rui Galao, Gilberto Betancor, Harry D Wilson, Adrian W Signell, Helena Winstone, Claire Kerridge, Nigel

Temperton, Luke Snell, Karen Bisnauthsing, Amelia Moore, Adrian Green, Lauren Martinez, Brielle Stokes, Johanna Honey, Alba Izquierdo-Barras, Gill Arbane, Amita Patel, Lorcan OConnell, Geraldine O Hara, Eithne MacMahon, Sam Douthwaite, Gaia Nebbia, Rahul Batra, Rocio Martinez-Nunez, Jonathan D. Edgeworth, Stuart J.D. Neil, Michael H. Malim, and Katie Doores. Longitudinal evaluation and decline of antibody responses in sars-cov-2 infection, July 2020. URL `https://www.medrxiv.org/content/10.1101/2020.07.09.20148429v1`. Accessed: 2020-07-14.

[50] Seema K Shah, Franklin G Miller, Thomas C Darton, Devan Duenas, Claudia Emerson, Holly Fernandez Lynch, Euzebiusz Jamrozik, Nancy S Jecker, Dorcas Kamuya, Melissa Kapulu, et al. Ethics of controlled human infection to address covid-19. *Science*, 368 (6493):832–834, 2020.

[51] Amy Caryn Sherman, Aneesh Mehta, Neal W Dickert, Evan J Anderson, and Nadine Rouphael. The future of flu: a review of the human challenge model and systems biology for advancement of influenza vaccinology. *Frontiers in cellular and infection microbiology*, 9:107, 2019.

[52] Debbie-Ann T Shirley and Monica A McArthur. The utility of human challenge studies in vaccine development: lessons learned from cholera. *Vaccine: development and therapy*, 2011(1):3, 2011.

[53] Danielle I Stanisic, James S McCarthy, and Michael F Good. Controlled human malaria infection: applications, advances, and challenges. *Infection and immunity*, 86(1), 2018.

[54] Nick Paul Taylor. Astrazeneca's covid-19 vaccine enters phase 2/3 clinical trial, May 2020. URL `https://www.fiercebiotech.com/biotech/astrazeneca-s-covid-19-vaccine-enters-phase-2-3-clinical-trial`. Accessed: 2020-06-28.

[55] Arianna Waye, Philip Jacobs, and Anthony B Schryvers. Vaccine development costs: a review. *Expert review of vaccines*, 12(12):1495–1501, 2013.

[56] World Health Organization. Feasibility, potential value and limitations of establishing a closely monitored challenge model of experimental covid-19 infection and illness in

healthy young adult volunteers. Technical report, World Health Organization, Jun 2020. URL https://www.who.int/publications/m/item/feasibility-potential-value-and-limitations-of-establishing-a-closely-monitored-challenge-model-of-experimental-covid-19-infection-and-illness-in-healthy-young-adult-volunteers.

[57] World Health Organization. An international randomised trial of candidate vaccines against covid-19. Technical report, World Health Organization, April 2020. URL https://www.who.int/blueprint/priority-diseases/key-action/Outline_CoreProtocol_vaccine_trial_09042020.pdf?ua=1.

[58] World Health Organization. Key criteria for the ethical acceptability of covid-19 human challenge studies. Technical report, World Health Organization, 2020.

[59] Wudan Yan. Challenge accepted: Human challenge trials for dengue, 2015.

5.7 Supplementary Materials

In this subsection, we include detailed results about clinical trial design (Sections 5.7.1 – 5.7.4), and additional simulation results (Section 5.7.6).

5.7.1 Efficacy Analysis

The protective effect of a vaccine—that is, vaccine efficacy—is defined as [41]:

$$\varepsilon \quad = \quad 1 - \frac{p_1}{p_0} \quad = \quad 1 - \frac{c_1/n_1}{c_0/n_0} \tag{5.25}$$

where ε refers to the vaccine efficacy, p_1 and p_0 are the attack rates observed in the treatment arm and the control arm, respectively, n_1 and n_0 refer to the sample sizes of the treatment arm and the control arm, respectively, and c_1 and c_0 refer to the number of infections observed in the treatment arm and the control arm, respectively. The attack rate is defined as the fraction of a cohort at risk that becomes infected during the surveillance period. There are conflicting views on the possibility of human reinfections [29, 49]; for simplicity, we rule out recurrent infections in our simulations.

Superiority Testing

First, we consider superiority testing to determine the licensure of a vaccine candidate at the end of a clinical study, e.g., RCT, ORCT, or HCT. The aim is to demonstrate that the efficacy of the candidate in the prevention of infections is greater than zero. Such a criteria might be appropriate for emergency use authorization during a pandemic where no alternative treatments are available. For this, we consider the following null and alternative hypotheses:

$$H_0 : p_1 - p_0 = 0 \qquad , \qquad H_1 : p_1 - p_0 \neq 0 \tag{5.26}$$

The test statistic under the null hypothesis is given by:

$$z = \frac{|p_1 - p_0| - a}{\sqrt{2\bar{p}\bar{q}a}} \quad , \quad a = \frac{r+1}{2rn_0} \quad , \quad r = \frac{n_1}{n_0} \tag{5.27}$$

$$\bar{p} = \frac{c_1 + c_0}{n_0(r+1)} = \frac{rp_1 + p_0}{r+1} \quad , \quad \bar{q} = 1 - \bar{p} \tag{5.28}$$

where z is the test statistic. For large samples, z is approximately the standard Normal distribution.

The power of a vaccine efficacy study under superiority testing is given by [15, 17]:

$$z_\beta = \frac{|P_1 - P_0|\sqrt{rn_0 - (r+1)/|P_1 - P_0|} - z_{\alpha/2}\sqrt{(r+1)\bar{P}\bar{Q}}}{\sqrt{P_1Q_1 + rP_0Q_0}} \tag{5.29}$$

$$\bar{P} = \frac{rP_1 + P_0}{r+1} \quad , \quad \bar{Q} = 1 - \bar{P} \tag{5.30}$$

$$P_1 = (1-\epsilon)P_0 \quad , \quad Q_i = 1 - P_i, \quad i \in \{0, 1\} \tag{5.31}$$

where α is the level of significance, β refers to the type II error under the alternative hypothesis, z_a is the $100(1-a)$ percentage points of the standard Normal distribution, P_1 and P_0 refer to the underlying (true) attack rate in the treatment arm and the control arm, respectively, and ϵ refers to the true vaccine efficacy.

Superiority-by-Margin Testing

Next, we consider the case where superiority by margin (also known as super-superiority)—that is, a vaccine efficacy that is greater than some minimum threshold—must be demonstrated for full licensure:

$$H_0 : \vartheta - \theta = 0 \quad , \quad H_1 : \vartheta - \theta \neq 0 \tag{5.32}$$

where $\vartheta = p_1/p_0$, and θ is a specified minimum threshold larger than 0 and smaller than 1.

The test statistic under the null hypothesis is given by [15]:

$$\chi^2 = \frac{(p_1 - \theta p_0)^2}{(\tilde{p}_1\tilde{q}_1 + r\theta^2\tilde{p}_0\tilde{q}_0)/rn_0} \qquad , \qquad \tilde{q}_i = 1 - \tilde{p}_i, \qquad i \in \{0, 1\} \tag{5.33}$$

where χ^2 is the test statistic, and $\tilde{p}_1$ and $\tilde{p}_0$ are the large sample approximations of the constrained maximum likelihood estimate of P_1 and P_0, respectively, under the null hypothesis (see below for closed-form solutions). For large samples, χ^2 is approximately the chi-square distribution on one degree of freedom.

The power of a vaccine efficacy study under superiority-by-margin testing is given by:

$$z_\beta = \frac{(\theta P_0 - P_1)\sqrt{rn_0} - z_{\alpha/2}\sqrt{\tilde{p}_1\tilde{q}_1 + r\theta^2\tilde{p}_0\tilde{q}_0}}{\sqrt{P_1Q_1 + r\theta^2P_0Q_0}} \tag{5.34}$$

Asymptotics for Superiority-by-Margin Testing

The constraint is:

$$\hat{p}_1 = \theta \hat{p}_0 \tag{5.35}$$

where $\hat{p}_1$ and $\hat{p}_0$ are the constrained maximum likelihood estimates of P_1 and P_0, respectively, under the null hypothesis.

The closed-form solution is given by:

$$\hat{p}_0 = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \tag{5.36}$$

$$A = (r+1)\theta n_0 \quad , \quad B = -(\theta r n_0 + c_1 + n_0 + \theta c_0) \quad , \quad C = c_1 + c_0 \tag{5.37}$$

The asymptotic approximation is:

$$\tilde{p}_0 = \frac{-B - \sqrt{B^2 - 4AC}}{2A} \quad , \quad \tilde{p}_1 = \theta \tilde{p}_0 \tag{5.38}$$

$$A = (r+1)\theta \quad , \quad B = -(\theta r + r P_1 + 1 + \theta P_0) \quad , \quad C = r P_1 + P_0 \tag{5.39}$$

5.7.2 Adaptive Vaccine Efficacy RCT

We propose an adaptive vaccine efficacy RCT design (ARCT) based on group sequential methods. First, we consider an alternative definition of vaccine efficacy based on relative force of infection, as opposed to relative risk of infection in Equation 5.25:

$$\varepsilon \approx 1 - \frac{\Lambda_1}{\Lambda_0} \quad , \quad \Lambda_i = \int_0^{t_s} \lambda_i(u)\, \mathrm{d}u, \quad i \in \{0, 1\} \tag{5.40}$$

where λ_1 and λ_0 refer to the force of infection in the treatment arm and the control arm, respectively, and t_s refers to the duration of the surveillance period. The force of infection of an infectious disease is defined as the expected number of new cases of the disease per unit person-time at risk. When the risk of infection is small, e.g., smaller than 0.10, the risk of infection is approximately equal to the cumulative force of infection [41].

Next, we note that the force of infection and the hazard function in survival analysis actually take the same functional form [41]. This suggests that infections can also be treated

as time-to-event data, in addition to binary variables as in Equation 5.25. By performing Cox regression on the time-to-infections data of a clinical trial, we can estimate the efficacy of the vaccine candidate from the hazard ratio of the treatment arm versus the control arm:

$$\varepsilon \approx 1 - \exp(\beta) \qquad , \qquad \lambda(t|z) = \lambda_{\text{baseline}}(t) \exp(\beta z) \tag{5.41}$$

where z refers to the treatment variable, i.e., whether the patient is vaccinated or not, $\lambda_{\text{baseline}}$ is the baseline hazard function, and β is the log hazard ratio. We note that the proportional hazards assumption is not unreasonable if we assume that the proportion of cases prevented by the vaccine is independent of the possibly non-homogeneous force of infection [41].

We consider the following null and alternative hypotheses based on the coefficient of the treatment variable in the Cox model:

$$H_0 : \beta - \beta_0 = 0 \qquad , \qquad H_1 : \beta - \beta_0 \neq 0 \tag{5.42}$$

where β_0 is 1 for superiority testing and smaller than 1 for superiority-by-margin testing.

The test statistic under the null hypothesis is given by:

$$z = \frac{\hat{\beta} - \beta_0}{\text{se}(\hat{\beta})} \tag{5.43}$$

where $\hat{\beta}$ is the maximum partial likelihood estimate of β and $\text{se}(\hat{\beta})$ is its standard error, and z is asymptotically Normal. This is also known the Wald test. It turns out this statistic satisfies the criteria for group sequential testing [27], allowing us to perform periodic interim analyses of accumulating trial data, rather than just a single final analysis at the end of a traditional vaccine efficacy RCT (see Figure 5-5). Under the group sequential testing framework, we estimate a new Cox model at each interim calendar time point based on the infections data that has accrued up to that point, over the course of the study surveillance period. At the interim analyses, we decide whether to stop the study early by rejecting the null hypothesis, i.e., approving the vaccine candidate, or to continue on to the next analysis by monitoring the subjects for a longer period of time [27].

We adopt Pocock's test for sequential testing [45]. It involves repeated testing at successive interim analyses at some constant nominal significance level over the course of the

study (see Algorithm 3). The critical value is chosen to satisfy the maximum type I error requirement, e.g., 5%.

Algorithm 3: Pocock's test. k refers to the k^{th} interim analysis, K refers to the maximum number of interim analyses planned, z_k refers to the test statistic at the k^{th} interim analysis, and $c(K, \alpha)$ refers to the nominal significance level which is a function of K and α, the maximum type I error allowed.

```
for k = 1, ..., K do
    if |z_k| ≥ c(K, α) then
        stop, reject H_0
    end
    else
        if k == K then
            stop, accept H_0
        end
        else
            continue
        end
    end
end
```

In our simulations, we consider a maximum of six interim analyses spaced 30 days apart, with the first analysis performed when the first 10,000 subjects enrolled have been monitored for at least 30 days. To keep the type I error at 5%, we consider a nominal significance level of 2.453 at each interim analyses [45].

For each of the epidemiological-model and population-vaccination schedule assumptions, we compute the expected net value of ARCT over 100,000 Monte Carlo simulation paths. For each path, we track the infections data of 30,000 patients for up to 180 days of surveillance. In addition, we estimate up to six Cox proportional hazards models, one at each interim analysis. The simulation process is computationally intensive despite parallelization, requiring approximately 8 hours to complete on the MIT Sloan "Engaging" high-performance computing cluster using over 400 processors.

While we have considered a simple adaptive design in this paper, we note that our framework can be easily extended to other sequential boundaries such as the O'Brien & Fleming's Test, to two-sided tests that allow for early stopping under the null hypothesis, i.e., early stopping for both futility and efficacy, and to flexible monitoring using the error

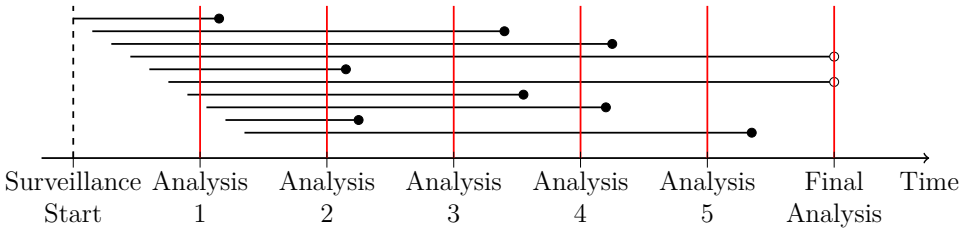


Figure 5-5: Infections as time-to-event data, measured from the start of surveillance. The horizontal lines represent the time to infection of ten subjects enrolled at different times. We monitor the subjects until an infection occurs or the end of study, whichever comes earlier. A solid circle at the right end denotes an infection, whereas a hollow circle indicates censoring. In the figure, we consider up to six analyses. At an interim analysis, subjects are considered censored if they are known to be uninfected and at risk at that point in time. Information on these subjects will continue to accrue through the surveillance period.

spending approach, instead of using a constant nominal significance level for all interim analyses [27].

5.7.3 Trial Design Assumptions

Table 5.4: Trial design assumptions common across RCT, ORCT, ARCT, and HCT.

Parameter	**Value**
Cohort	Closed and fixed
Accrual rate (patients/day)	250
Control arm	Vaccine for meningococcal bacteria
Treatment arm	Vaccine candidate for COVID-19
Vaccination schedule	Two doses administered 28 days apart
Vaccine efficacy (%)	30–90
Time for immune response (days)	28
Endpoint	Infection by SARS-CoV-2
Time for safety data collection, data analysis, and FDA review (days)	120
Type I error (%)	5

Table 5.5: Trial design assumptions specific to RCT, ORCT, ARCT, and HCT.

Parameter	**RCT**	**ORCT**	**ARCT**	**HCT**
Set-up time (days)	–	–	–	30–120
Sample size	30,000	30,000	30,000	250
Inclusion criteria	Healthy adults aged 18–50 years	Healthy adults aged 18–50 years	Healthy adults aged 18–50 years	Healthy adults aged 18–25 years
Randomization ratio (treatment:control)	1:1	1:1	1:1	4:1
Time for enrollment (days)	120	120	40–120	1
Surveillance period (days)	Fixed and constant for all subjects; 180	Fixed and constant for all subjects; 30–180	Calendar time interval	Fixed and constant for all subjects; 14
Attack rate (%)	Depends on epidemiological model	Depends on epidemiological model and surveillance period	Depends on epidemiological model and surveillance period	90
Efficacy analysis	Single analysis at end of study	Single analysis at end of study	Up to 6 interim analyses spaced 30 days apart	Single analysis at end of study
Additional safety study	–	–	–	Single-arm with 5,000 subjects
Estimated time to licensure (days)	476	326–476	246–396	221–311

5.7.4 Financial Cost of Vaccine Efficacy Studies

There are many sources of costs involved in a clinical trial, e.g., patient recruitment and retention, medical and administrative staff, clinical procedures and central laboratory, site management, and data collection and analysis. For a back-of-the-envelope calculation, we assume that the cost per subject in a phase 3 vaccine efficacy trial is around US$5,000. This suggests a cost of US$150M for a study with 30,000 subjects, close to that estimated for rotavirus vaccines [32] in one of the very few studies that estimate the cost of vaccine development [55]. The figure is very high as compared to the median expense of a phase 3 trial for novel therapeutic agents, estimated to be US$19M [40]. However, this is not surprising because vaccine efficacy studies are notorious for being costly due to the large sample sizes and lengthy follow-up durations. If we assume that challenge studies have a cost per subject that is ten times higher, i.e., US$50,000 per volunteer, the estimated cost of an HCT is approximately US$37.5M, where we have assumed a cost of US$5,000 per subject for the follow-up single-arm safety study comprising of 5,000 subjects. This makes up just 25% of the cost of an RCT with 30,000 subjects.

5.7.5 Trade-off Between Time and Power

As mentioned in the main text, there is a trade-off between time and power. A shorter surveillance period will, *ceteris paribus*, reduce the power of the RCT. However, it will also reduce the time to licensure of the vaccine (if approved), which would prevent more infections and save more lives. Conversely, a longer surveillance period would increase the power of the RCT but also prolong the time it takes for the vaccine to be approved. We illustrate the interaction between power and infections avoided over time in Figure 5-6.

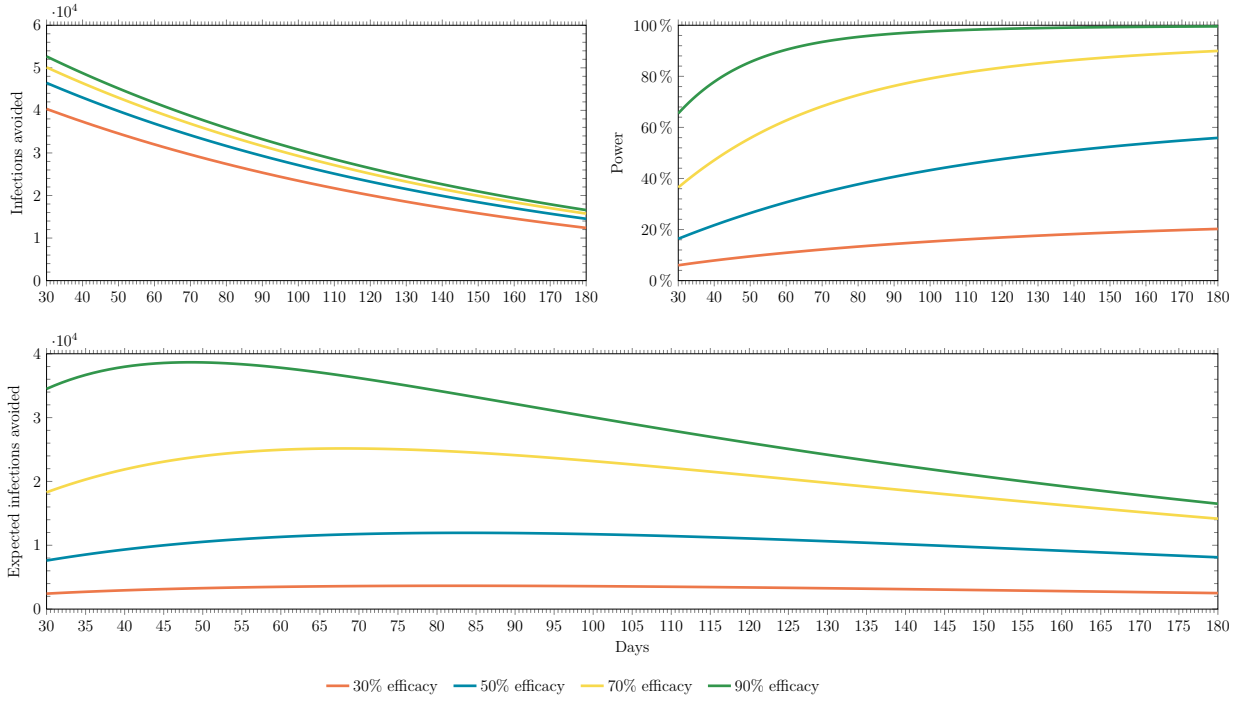


Figure 5-6: An illustration of the interaction between power and infections avoided over time. (Top left panel) The number of infections avoided decreases over time. (Top right panel) The power under the superiority test expected from the clinical trial increases with the surveillance time. (Bottom panel) The expected number of infections avoided—computed as the product of the power and infections avoided—as a function of the surveillance period.

5.7.6 Additional Simulation Results

Table 5.6: Expected number of incremental infections and deaths avoided in the U.S. under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority testing, and 1M doses of a vaccine per day are available after licensure, compared to the baseline case where no vaccine is ever approved.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$
Status Quo								
RCT	2,506	20	8,116	64	14,162	112	16,506	130
ORCT	3,654	29	11,947	95	25,167	200	38,663	308
ARCT	6,248	49	22,261	177	49,396	393	63,896	508
HCT (30-day set-up)	90,472	722	106,202	848	114,847	918	120,945	966
HCT (60-day set-up)	71,223	568	83,467	666	90,167	720	94,885	758
HCT (90-day set-up)	56,263	448	65,857	525	71,088	567	74,766	597
HCT (120-day set-up)	44,556	355	52,122	415	56,235	449	59,123	471
Behavioral								
RCT	224,835	1,736	264,810	2,056	289,168	2,251	306,050	2,386
ORCT	705,881	5,591	925,920	7,344	1,007,301	7,995	1,065,183	8,459
ARCT	1,502,846	11,959	2,051,223	16,346	2,269,753	18,094	2,423,075	19,321
HCT (30-day set-up)	2,209,905	17,618	2,695,582	21,502	2,982,094	23,794	3,189,157	25,451
HCT (60-day set-up)	1,611,969	12,834	1,951,336	15,548	2,150,531	17,142	2,294,765	18,295
HCT (90-day set-up)	1,190,836	9,465	1,429,078	11,370	1,566,872	12,473	1,666,446	13,269
HCT (120-day set-up)	894,225	7,092	1,065,008	8,457	1,161,296	9,228	1,230,321	9,780
Ramp								
RCT	756,692	5,764	845,731	6,477	899,765	6,909	937,666	7,212
ORCT	1,825,095	14,344	2,656,479	20,964	2,890,096	22,832	3,047,293	24,089
ARCT	3,594,521	28,466	5,131,954	40,766	5,768,903	45,861	6,091,608	48,443
HCT (30-day set-up)	5,526,735	43,930	6,565,535	52,235	7,130,975	56,759	7,523,068	59,896
HCT (60-day set-up)	4,282,314	33,975	5,086,688	40,404	5,528,656	43,941	5,837,268	46,409
HCT (90-day set-up)	3,311,292	26,206	3,926,171	31,120	4,265,392	33,834	4,503,392	35,738
HCT (120-day set-up)	2,564,645	20,233	3,031,075	23,959	3,288,349	26,018	3,469,234	27,465

Table 5.7: Expected number of incremental infections and deaths avoided in the U.S. under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority testing, and infinite doses of a vaccine per day are available after licensure, compared to the baseline case where no vaccine is ever approved.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$
Status Quo								
RCT	4,343	35	12,691	101	20,900	165	23,426	185
ORCT	6,190	50	18,462	147	36,872	294	54,672	436
ARCT	10,655	84	34,672	276	72,976	581	90,989	725
HCT (30-day set-up)	157,044	1,255	168,612	1,347	172,598	1,380	174,917	1,398
HCT (60-day set-up)	122,531	978	131,429	1,049	134,478	1,075	136,254	1,088
HCT (90-day set-up)	96,093	767	102,986	822	105,338	841	106,709	852
HCT (120-day set-up)	75,691	604	81,068	647	82,896	662	83,965	670
Behavioral								
RCT	401,196	3,147	422,644	3,318	432,235	3,396	437,725	3,439
ORCT	1,284,033	10,217	1,542,261	12,276	1,587,101	12,634	1,613,158	12,843
ARCT	2,957,024	23,592	3,683,384	29,403	3,813,885	30,447	3,881,898	30,991
HCT (30-day set-up)	4,466,352	35,669	4,884,898	39,016	5,039,465	40,253	5,128,348	40,964
HCT (60-day set-up)	3,196,408	25,510	3,494,817	27,895	3,605,985	28,786	3,670,305	29,300
HCT (90-day set-up)	2,291,219	18,268	2,500,498	19,941	2,578,527	20,566	2,623,871	20,928
HCT (120-day set-up)	1,659,356	13,214	1,805,003	14,377	1,858,914	14,809	1,890,330	15,060
Ramp								
RCT	1,174,517	9,107	1,229,484	9,547	1,255,157	9,752	1,270,085	9,871
ORCT	3,172,803	25,126	4,242,057	33,649	4,362,661	34,612	4,422,914	35,094
ARCT	6,347,189	50,488	8,191,884	65,245	8,662,725	69,012	8,776,472	69,922
HCT (30-day set-up)	9,669,217	77,070	10,366,266	82,641	10,597,019	84,487	10,728,517	85,539
HCT (60-day set-up)	7,564,062	60,228	8,126,045	64,719	8,315,537	66,236	8,423,946	67,103
HCT (90-day set-up)	5,860,161	46,598	6,304,440	50,146	6,456,348	51,362	6,543,545	52,059
HCT (120-day set-up)	4,512,448	35,815	4,857,257	38,569	4,976,272	39,521	5,044,819	40,070

Table 5.8: Expected number of incremental infections and deaths avoided in the U.S. under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority-by-margin testing at 50%, and 1M doses of a vaccine per day are available after licensure, compared to the baseline case where no vaccine is ever approved. We observe negative expected net values when vaccine efficacy is 30% because the candidate is almost never approved under superiority-by-margin testing. While a cost from conducting the trial is always incurred, the expected post-trial benefit is close to zero.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$
Status Quo								
RCT	−34	0	319	3	4,091	32	14,935	118
ORCT	239	2	1,149	9	6,123	49	26,189	208
ARCT	−39	0	199	1	3,840	30	27,107	215
HCT (30-day set-up)	−171	−1	2,523	20	113,800	910	120,945	966
HCT (60-day set-up)	−171	−1	1,955	16	89,345	713	94,885	758
HCT (90-day set-up)	−171	−1	1,515	12	70,439	562	74,766	597
HCT (120-day set-up)	−171	−1	1,171	9	55,722	445	59,123	471
Behavioral								
RCT	−1,461	−11	2,242	17	289,168	2,251	306,050	2,386
ORCT	−331	−2	21,526	171	955,088	7,581	1,065,183	8,459
ARCT	−1,384	−11	29,583	235	2,043,288	16,282	2,423,068	19,321
HCT (30-day set-up)	−171	−1	67,258	537	2,954,925	23,577	3,189,157	25,451
HCT (60-day set-up)	−171	−1	48,652	388	2,130,938	16,986	2,294,765	18,295
HCT (90-day set-up)	−171	−1	35,595	283	1,552,596	12,359	1,666,446	13,269
HCT (120-day set-up)	−171	−1	26,494	210	1,150,715	9,144	1,230,321	9,780
Ramp								
RCT	−1,406	−11	10,693	82	899,765	6,909	937,666	7,212
ORCT	−198	−1	64,285	508	2,467,656	19,477	3,047,293	24,089
ARCT	−1,196	−9	82,127	649	4,714,327	37,425	6,088,218	48,416
HCT (30-day set-up)	−171	−1	164,007	1,305	7,066,008	56,242	7,523,068	59,896
HCT (60-day set-up)	−171	−1	127,036	1,009	5,478,287	43,541	5,837,268	46,409
HCT (90-day set-up)	−171	−1	98,023	777	4,226,532	33,526	4,503,392	35,738
HCT (120-day set-up)	−171	−1	75,645	598	3,258,390	25,781	3,469,234	27,465

Table 5.9: Expected number of incremental infections and deaths avoided in the U.S. under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority-by-margin testing at 50%, and 10M doses of a vaccine per day are available after licensure, compared to the baseline case where no vaccine is ever approved. We observe negative expected net values when vaccine efficacy is 30% because the candidate is almost never approved under superiority-by-margin testing. While a cost from conducting the trial is always incurred, the expected post-trial benefit is close to zero.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$
Status Quo								
RCT	−25	0	471	4	5,536	44	19,507	154
ORCT	374	3	1,625	13	8,217	66	34,029	271
ARCT	−33	0	298	2	5,170	41	35,268	280
HCT (30-day set-up)	−171	−1	3,675	29	155,455	1,243	159,876	1,277
HCT (60-day set-up)	−171	−1	2,842	23	121,365	970	124,777	997
HCT (90-day set-up)	−171	−1	2,203	18	95,234	761	97,886	782
HCT (120-day set-up)	−171	−1	1,709	14	75,056	599	77,132	615
Behavioral								
RCT	−1,461	−11	3,852	30	397,396	3,117	404,562	3,174
ORCT	−331	−2	32,156	256	1,352,103	10,757	1,457,500	11,598
ARCT	−1,384	−11	46,267	368	3,037,771	24,238	3,473,025	27,720
HCT (30-day set-up)	−171	−1	107,601	859	4,440,619	35,463	4,591,750	36,671
HCT (60-day set-up)	−171	−1	76,935	614	3,175,958	25,346	3,283,975	26,209
HCT (90-day set-up)	−171	−1	55,168	440	2,276,419	18,149	2,352,436	18,757
HCT (120-day set-up)	−171	−1	40,014	319	1,649,447	13,134	1,702,601	13,558
Ramp								
RCT	−1,406	−11	14,720	115	1,160,564	8,996	1,179,234	9,145
ORCT	−183	−1	93,009	738	3,387,704	26,840	4,050,013	32,111
ARCT	−1,142	−9	119,304	947	6,492,110	51,647	8,067,450	64,250
HCT (30-day set-up)	−171	−1	236,179	1,882	9,636,422	76,805	9,897,591	78,892
HCT (60-day set-up)	−171	−1	184,404	1,468	7,533,612	59,983	7,743,514	61,659
HCT (90-day set-up)	−171	−1	142,660	1,134	5,833,783	46,385	5,999,381	47,706
HCT (120-day set-up)	−171	−1	109,769	871	4,491,107	35,642	4,619,521	36,667

Table 5.10: Expected number of incremental infections and deaths avoided in the U.S. under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority-by-margin testing at 50%, and infinite doses of a vaccine per day are available after licensure, compared to the baseline case where no vaccine is ever approved. We observe negative expected net values when vaccine efficacy is 30% because the candidate is almost never approved under superiority-by-margin testing. While a cost from conducting the trial is always incurred, the expected post-trial benefit is close to zero.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$	$\mathbb{E}[\Delta\text{Infections}]$	$\mathbb{E}[\Delta\text{Deaths}]$
Status Quo								
RCT	−22	0	523	4	6,050	48	21,198	168
ORCT	416	3	1,789	14	8,976	72	36,974	295
ARCT	−31	0	332	2	5,655	45	38,342	305
HCT (30-day set-up)	−171	−1	4,084	33	171,025	1,367	174,917	1,398
HCT (60-day set-up)	−171	−1	3,154	25	133,252	1,065	136,254	1,088
HCT (90-day set-up)	−171	−1	2,443	20	104,377	833	106,709	852
HCT (120-day set-up)	−171	−1	1,895	15	82,140	656	83,965	670
Behavioral								
RCT	−1,461	−11	4,337	34	432,235	3,396	437,725	3,439
ORCT	−331	−2	36,046	287	1,504,842	11,979	1,613,158	12,843
ARCT	−1,384	−11	52,340	417	3,416,029	27,264	3,881,886	30,991
HCT (30-day set-up)	−171	−1	121,991	974	4,993,552	39,886	5,128,348	40,964
HCT (60-day set-up)	−171	−1	87,239	696	3,573,132	28,524	3,670,305	29,300
HCT (90-day set-up)	−171	−1	62,381	498	2,555,035	20,379	2,623,871	20,928
HCT (120-day set-up)	−171	−1	44,993	358	1,841,978	14,674	1,890,330	15,060
Ramp								
RCT	−1,406	−11	16,101	126	1,255,157	9,752	1,270,085	9,871
ORCT	−178	−1	102,769	816	3,718,588	29,487	4,422,914	35,094
ARCT	−1,126	−9	131,636	1,045	7,109,717	56,588	8,771,717	69,884
HCT (30-day set-up)	−171	−1	259,025	2,065	10,500,475	83,717	10,728,517	85,539
HCT (60-day set-up)	−171	−1	203,020	1,617	8,239,778	65,633	8,423,946	67,103
HCT (90-day set-up)	−171	−1	157,479	1,253	6,397,527	50,894	6,543,545	52,059
HCT (120-day set-up)	−171	−1	121,300	963	4,930,935	39,161	5,044,819	40,070

Table 5.11: Estimated date of licensure and probability of approval under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority testing, and 1M doses of a vaccine per day are available after licensure. For ARCT, we report the median date of licensure over all Monte Carlo simulations. DoL: date of licensure (month/day/year); PoA: probability of approval.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)
Status Quo								
RCT	11/19/2021	20.2	11/19/2021	55.9	11/19/2021	89.9	11/19/2021	99.6
ORCT	08/14/2021	13.6	08/15/2021	38.9	07/30/2021	67.2	07/10/2021	84.3
ARCT	07/02/2021	14.5	06/02/2021	44.2	06/02/2021	83.8	06/02/2021	99.6
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0
Behavioral								
RCT	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0
ORCT	06/24/2021	90.5	06/22/2021	100.0	06/22/2021	100.0	06/22/2021	100.0
ARCT	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0
Ramp								
RCT	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0
ORCT	07/06/2021	88.9	06/22/2021	99.6	06/22/2021	100.0	06/22/2021	100.0
ARCT	05/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0

Table 5.12: Estimated date of licensure and probability of approval under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority testing, and 10M doses of a vaccine per day are available after licensure. For ARCT, we report the median date of licensure over all Monte Carlo simulations. DoL: date of licensure (month/day/year); PoA: probability of approval.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)
Status Quo								
RCT	11/19/2021	20.2	11/19/2021	55.9	11/19/2021	89.9	11/19/2021	99.6
ORCT	08/15/2021	13.8	08/15/2021	38.9	07/30/2021	67.2	07/10/2021	84.3
ARCT	07/02/2021	14.5	06/02/2021	44.2	06/02/2021	83.8	06/02/2021	99.6
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0
Behavioral								
RCT	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0
ORCT	06/23/2021	89.6	06/22/2021	100.0	06/22/2021	100.0	06/22/2021	100.0
ARCT	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0
Ramp								
RCT	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0
ORCT	07/06/2021	88.9	06/22/2021	99.6	06/22/2021	100.0	06/22/2021	100.0
ARCT	05/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0

Table 5.13: Estimated date of licensure and probability of approval under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority testing, and infinite doses of a vaccine per day are available after licensure. For ARCT, we report the median date of licensure over all Monte Carlo simulations. DoL: date of licensure (month/day/year); PoA: probability of approval.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)
Status Quo								
RCT	11/19/2021	20.2	11/19/2021	55.9	11/19/2021	89.9	11/19/2021	99.6
ORCT	08/14/2021	13.6	08/14/2021	38.6	07/30/2021	67.2	07/10/2021	84.3
ARCT	07/02/2021	14.5	06/02/2021	44.2	06/02/2021	83.8	06/02/2021	99.6
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0
Behavioral								
RCT	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0
ORCT	06/23/2021	89.6	06/22/2021	100.0	06/22/2021	100.0	06/22/2021	100.0
ARCT	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0
Ramp								
RCT	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0	11/19/2021	100.0
ORCT	07/06/2021	88.9	06/22/2021	99.6	06/22/2021	100.0	06/22/2021	100.0
ARCT	05/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)	03/09/2021	98.1	03/09/2021	100.0	03/09/2021	100.0	03/09/2021	100.0
HCT (60-day set-up)	04/08/2021	98.1	04/08/2021	100.0	04/08/2021	100.0	04/08/2021	100.0
HCT (90-day set-up)	05/08/2021	98.1	05/08/2021	100.0	05/08/2021	100.0	05/08/2021	100.0
HCT (120-day set-up)	06/07/2021	98.1	06/07/2021	100.0	06/07/2021	100.0	06/07/2021	100.0

Table 5.14: Estimated date of licensure and probability of approval under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority-by-margin testing at 50%, and 1M doses of a vaccine per day are available after licensure. For ARCT, we report the median date of licensure over all Monte Carlo simulations. A blank entry indicates that the vaccine candidate is never approved. DoL: date of licensure (month/day/year); PoA: probability of approval.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)
Status Quo								
RCT	11/19/2021	0.1	11/19/2021	2.5	11/19/2021	26.2	11/19/2021	90.1
ORCT	06/22/2021	0.3	06/22/2021	2.5	08/06/2021	16.3	07/31/2021	53.5
ARCT		0.0	07/02/2021	0.6	08/01/2021	9.3	08/01/2021	64.3
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0
Behavioral								
RCT		0.0	11/19/2021	1.3	11/19/2021	100.0	11/19/2021	100.0
ORCT		0.0	06/22/2021	2.4	06/22/2021	94.8	06/22/2021	100.0
ARCT		0.0	06/02/2021	2.4	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0
Ramp								
RCT		0.0	11/19/2021	1.4	11/19/2021	100.0	11/19/2021	100.0
ORCT		0.0	06/22/2021	2.4	06/30/2021	83.2	06/22/2021	100.0
ARCT		0.0	06/02/2021	2.5	05/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0

Table 5.15: Estimated date of licensure and probability of approval under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority-by-margin testing at 50%, and 10M doses of a vaccine per day are available after licensure. For ARCT, we report the median date of licensure over all Monte Carlo simulations. A blank entry indicates that the vaccine candidate is never approved. DoL: date of licensure (month/day/year); PoA: probability of approval.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)
Status Quo								
RCT	11/19/2021	0.1	11/19/2021	2.5	11/19/2021	26.2	11/19/2021	90.1
ORCT	06/22/2021	0.3	06/22/2021	2.5	08/06/2021	16.3	07/31/2021	53.5
ARCT		0.0	07/02/2021	0.6	08/01/2021	9.3	08/01/2021	64.3
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0
Behavioral								
RCT		0.0	11/19/2021	1.3	11/19/2021	100.0	11/19/2021	100.0
ORCT		0.0	06/22/2021	2.4	06/22/2021	94.8	06/22/2021	100.0
ARCT		0.0	06/02/2021	2.4	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0
Ramp								
RCT		0.0	11/19/2021	1.4	11/19/2021	100.0	11/19/2021	100.0
ORCT		0.0	06/22/2021	2.4	06/29/2021	83.2	06/22/2021	100.0
ARCT		0.0	06/02/2021	2.5	05/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0

Table 5.16: Estimated date of licensure and probability of approval under different trial designs, vaccine efficacies, and epidemiological scenarios, assuming trials start on August 1, 2020, superiority-by-margin testing at 50%, and infinite doses of a vaccine per day are available after licensure. For ARCT, we report the median date of licensure over all Monte Carlo simulations. A blank entry indicates that the vaccine candidate is never approved. DoL: date of licensure (month/day/year); PoA: probability of approval.

	Vaccine Efficacy (%)							
	30		**50**		**70**		**90**	
	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)	DoL	PoA (%)
Status Quo								
RCT	11/19/2021	0.1	11/19/2021	2.5	11/19/2021	26.2	11/19/2021	90.1
ORCT	06/22/2021	0.3	06/22/2021	2.5	08/06/2021	16.3	07/31/2021	53.5
ARCT		0.0	07/02/2021	0.6	08/01/2021	9.3	08/01/2021	64.3
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0
Behavioral								
RCT		0.0	11/19/2021	1.3	11/19/2021	100.0	11/19/2021	100.0
ORCT		0.0	06/22/2021	2.4	06/22/2021	94.8	06/22/2021	100.0
ARCT		0.0	06/02/2021	2.4	04/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0
Ramp								
RCT		0.0	11/19/2021	1.4	11/19/2021	100.0	11/19/2021	100.0
ORCT		0.0	06/22/2021	2.4	06/29/2021	83.2	06/22/2021	100.0
ARCT		0.0	06/02/2021	2.5	05/03/2021	100.0	04/03/2021	100.0
HCT (30-day set-up)		0.0	03/09/2021	2.5	03/09/2021	99.1	03/09/2021	100.0
HCT (60-day set-up)		0.0	04/08/2021	2.5	04/08/2021	99.1	04/08/2021	100.0
HCT (90-day set-up)		0.0	05/08/2021	2.5	05/08/2021	99.1	05/08/2021	100.0
HCT (120-day set-up)		0.0	06/07/2021	2.5	06/07/2021	99.1	06/07/2021	100.0

Part III

Big Data and Machine Learning in Finance

Chapter 6

When Do Investors Freak Out?: Machine Learning Predictions of Panic Selling

6.1 Introduction

Financial advisors have long advised their clients to stay calm and weather any passing financial storm in their portfolios. Despite this, a percentage of investors tend to 'freak out' and sell off a large portion of their risky assets in certain adverse market environments. This situation is often discussed in the financial press and media[1], but is rarely defined or quantified. In this chapter, we develop a method to identify panic selling and apply it to a novel large dataset of brokerage account information from 2003 to 2015 to examine panic selling and 'freakout' behavior.

We begin by characterizing the aggregate behavior of investors who make panic sales. First, we document the frequency and timing of panic selling. We see that, while panic sales are infrequent, with only 0.1% of the investors panic selling at any point in time, they occur at up to 3 times the baseline frequency when there are large market movements.

[1]Consider the typical CNBC headline, "The market may be swinging, but the last thing you should do is freak out: Wall Street trading coach". Source: https://www.cnbc.com/2018/02/09/the-market-may-be-swinging-but-dont-freak-out-says-trading-coach.html

Second, we find that 30.9% of the investors who panic sell never return to reinvest in risky assets. However, of those that do, more than 58.5% reenter the market within half a year.

Third, we analyze the investors by demographic group who tend to 'freak out' under our definition (that is, they make panic sales during periods of sharp market downturns), and find that investors who are males, or above the age of 45, or married, or with a greater number of dependents, or who have declared themselves having excellent investment experience or knowledge tend to freak out in higher proportions.

Fourth, we find that the median investor earns a zero to negative return after he panic sells. Calculating the opportunity cost of panic selling over time finds that panic selling is suboptimal if executed in an improving market, but it is beneficial as a stop-loss mechanism in rapidly deteriorating markets.

Finally, we develop machine learning models to predict when investors might panic sell in the near future. Our set of predictive features includes the demographic characteristics of the investor, their portfolio histories, and current and past market conditions. This task is made difficult by the extreme rarity of panic sales. Nonetheless, our best-performing deep neural network achieves a 69.5% true positive accuracy rate and a 81.2% true negative accuracy rate, demonstrating that artificial intelligence techniques can assist in identifying individuals at risk of panic selling in the near future.

6.2 Literature Review

Behavioral finance has documented a wide range of stylized actions of investors, including loss aversion, regret aversion, the snake-bite effect, overtrading, and the disposition effect [25, 35]. There has been renewed interest in these behavioral patterns since the financial crisis of 2007-2008, both within the academic community and among the general public. We summarize several documented investor behaviors, some which are related to panic selling, and others which are inconsistent with the phenomenon.

6.2.1 Panic Selling

Although widely discussed in the financial industry (for example, see Rotblot [29]), little of the available literature discusses the concept of panic selling during a period of lowered market performance. This is most likely due to the limited availability of datasets that cover a wide range of selling events and market environments. Using price and volume information as well as data from Chinese stock markets, Shi et al. [32] provide a theoretical model based on conditioning to explain investor behavior. Their model shows that investors can be either overconfident or panicked based on price momentum. The strongest positive correlation in behavior occurs during price reversals, when many investors are more likely to sell their risky assets in a panic.

In contrast to panic selling, however, Statman et al. [33] found that share turnover is positively correlated to lagged returns, which suggests overconfidence is a dominant factor. Barber et al. [6] demonstrated that investors tend to buy stocks with strong recent performance, and they tend to buy stocks with higher trading volume.

Our study presents evidence that investors occasionally panic and sell off a large portion of their portfolio. It attempts to address the above issues by using a larger and more fine-grained dataset over a longer time horizon than earlier studies. In this way, we hope to capture a broader range of circumstances under which investors may make panic sales.

6.2.2 Overtrading and the Disposition Effect

Overtrading is in some ways the opposite of panic selling, which causes the investor to leave the market, either temporarily or permanently. Several authors have documented that some investors tend to overtrade. For example, Benos [10] and Odean [25] suggest that overconfidence causes investors to trade too frequently. Using trading account data, Barber and Odean [4] document overtrading, and demonstrate that it is detrimental to the wealth of those investors who trade too frequently.

In addition, much of the behavioral finance literature has focused on the disposition effect [31], the tendency for investors to buy stocks with strong recent performance and hold onto their losing investments. This can also be considered as another near-opposite to panic

selling.

6.2.3 Stop-loss

A parallel to panic selling can be found in the use of stop-loss strategies by investors. Stop-loss strategies are rules used by investors to reduce their holdings in risky assets should the value of their holdings reach a certain predetermined threshold. Kaminski and Lo [19] and Lo and Remorov [24] examine the value of these rules under different market conditions. In some situations (for example, if market prices exhibit momentum), stop-loss strategies may outperform buy-and-hold strategies over certain time horizons. This, however, depends on the condition that investors return to the market at some point (see [23]). Since the performance of a stop-loss rule is dependent on investor reentry, in our empirical analysis, we also examine investor reentry after a 'freakout'.

6.2.4 Stock Market Crashes and Investor Overreaction

Many authors have studied stock market crashes. From an asset pricing perspective, several authors demonstrate that rare disaster risk can explain the equity risk premium and other puzzles in macro-finance [7][8, 11, 18][27]. Many other authors examine the impact of tail risk on total market returns [3][21].

The more relevant question for panic selling, however, is how investors behave during a stock market crash. Bondt and Thaler [13] and Bondt et al. [12] argue that investors tend to overreact to large economic events. Chopra et al. [16], Rozeff and Zaman [30], Bauman et al. [9], Wang et al. [34] and others provide empirical evidence for investor overreaction. There are many different explanations for overreaction during a market crash (for a summary, see Amini et al. [1]). These include changes in investor sentiment [2], herding behavior [28], market constraints [26][20], and changing risk preferences [14]).

Similarly, many studies of investor overreaction focus on price changes, but only in some cases do they use survey information to document these possible behavioral factors, while few studies have access to the entire portfolio of investors to consider their actual portfolio decisions.

6.3 Data Summary

We analyze the financial activity of 653455 anonymous accounts corresponding to 298556 households from one of the largest brokerage firms in the United States. These accounts are drawn at random from the population of U.S. brokerage accounts active as of December 31st, 2015, and have had their account identification numbers fully anonymized.

Our dataset consists of (i) monthly snapshots of positions and balances held in every sampled portfolio, (ii) every trade made through these accounts, and (iii) the demographic information of the account holder as reported on the initial application form, including age, income, and self-declared levels of experience and knowledge. The kinds of assets contained in the accounts include stocks, mutual funds, options, fixed income, and cash securities. Details of the composition of our dataset are included in the Supplementary Materials for the sake of brevity in the main text of this article. We have also been given a map from accounts to households that allows us to aggregate activities of related accounts. In this study, we analyze panic selling at the household level.

A household will consist of one or more individual accounts opening and closing at different points in time. We consider the time when the first individual account is opened as the account opening date of the household. Since the data given to us only records activities starting from January 2003, all households that were active prior to this date will be reported as though they were started on January 2003. While all the household accounts in our sample have at least one account that is open at the time of the study, there are some households who sold their assets and decided not to return to the market. We call these 'inactive' households. The number of active household accounts at time t, N_t, can be computed recursively with the following formula:

$$N_t = N_{t-1} + n_t^o - n_t^p \tag{6.1}$$

where n_t^p and n_t^o denote the number of households that panic sold and opened at time t, respectively. Figure 6-1 shows the cumulative number of household accounts that opened, exited the market and were active over time.

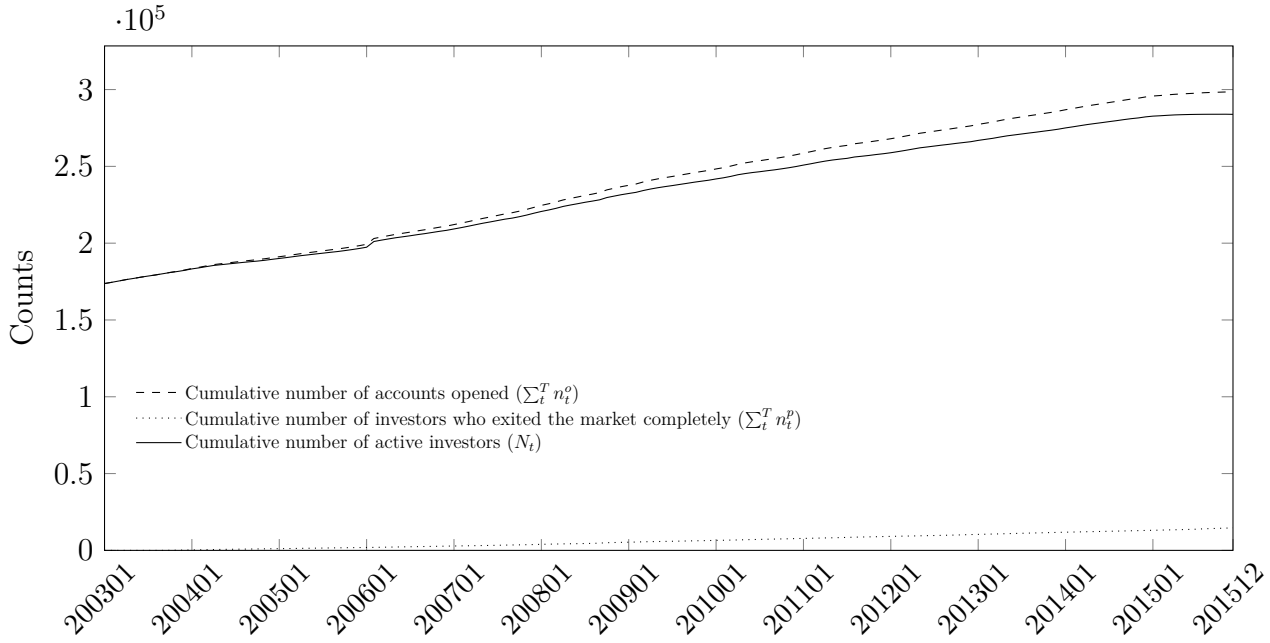


Figure 6-1: Number of household accounts versus time (YYYYMM format).

6.4 Methodology

6.4.1 Identifying panic sells

It is typically understood that an investor 'panic sells' when he *intentionally* sells off a substantial proportion of his risky assets abruptly. We develop rules in order to systematically capture such a behavior.

Consider a situation in which we are given monthly snapshots of portfolios and a view of every trade. Let V_t be the value of an investor portfolio at month t and $x_t = \frac{V_t - V_{t-1}}{V_{t-1}}$ be the percentage change in value of the portfolio between month t and $t+1$. Let T_t be the sum of value of all the trades in month t. Then $t_t = \frac{T_t}{V_{t-1}}$ is the proportion of the portfolio traded in this month. A positive t_t denotes a net buy and a negative t_t denotes a net sell. An investor is said to have made a panic sell in month t when

- Condition (1) The value of his portfolio declines by at least p_1 over one month (i.e. $x_t \leq -p_1$ for some $p_1 > 0$) *and*
- Condition (2) The investor makes a net sell of p_2 of his portfolio within the same period (i.e. $t_t \leq -p_2$ for some $p_2 > 0$).

Condition (1) states that the value of the portfolio falls substantially between two monthly snapshots of the portfolio. This is a necessary condition for any liquidation. However, it is not a sufficient condition, as large changes in a portfolio may be induced by natural market movements without any action by the investor. In order to identify that the investor *intentionally* reduced his holding of risky assets, we impose Condition (2). At first glance, it may seem that Condition (2) alone would be effective in detecting panic selling. This is untrue, however, as a portfolio may have depreciated substantially before an investor sells. For example, suppose that we let p_2 take a value of 0.8 in order to capture a large change in the portfolio. However, if the value of the investor's portfolio falls 25% from \$100,000 to \$75,000 due to market conditions before he liquidates the rest of it, it is impossible to fulfil the condition of p_2 of 0.80. Hence, just using Condition (2) by itself will cause us to miss this liquidation event. On the other hand, a lower value of p_2 can be used if we impose Condition (1).

In addition to identifying panic selling, we identify cases where investors who exited from their risky positions decide *intentionally* to take on risk again. We call such an event a 'return to market'. The following rules identify such events:

- Condition (3) The value of the portfolio must reach at least p_3 of the pre-liquidation value *and*
- Condition (4) The investor must have a cumulative net buy of p_4 of the amount that he sold during liquidation.

We set p_1 (the monthly portfolio decline), p_2 (the monthly portfolio net sell), p_3 (the portfolio rebound), and p_4 (the cumulative net buy) as 0.9, 0.5, 0.5, and 0.5, respectively, in this study. While setting both p_1 and p_2 to lower values will increase the number of panic sales identified, this does not change our analysis or the general pattern exhibited by household investors (see Section 6.8.6 of the Supplementary Materials).

6.4.2 Identifying risk factors for liquidations

In order to understand which groups of investors (G) are more susceptible to panic selling or freakout events (E), we compute the relative prevalence of the group given an event:

$P(G|E)/P(G)$. A number greater than 1 indicates that the group is more likely to have the event compared to other groups, while a value less than 1 signals the opposite.

We can test our null hypothesis of $P(G|E) = P(G)$ against the alternative hypothesis of $P(G|E) \neq P(G)$ using the two-proportion Z-test. Let $p_1 = P(G|E)$, $P(G) = p_2$, and the number of investors in each group be n_1 and n_2, respectively. The test statistic is given by $z = \frac{p_1 - p_2}{SE}$, where $SE^2 = p(1-p)(\frac{1}{n_1} + \frac{1}{n_2})$ and $p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n2}$.

6.5 Results

We counted 36374 panic sells by 26852 household investors (9.0% of all households) across a period of 13 years between January 2003 and December 2015, endpoints inclusive. A heat map for panic sales and returns to the market is given in Figure 6-2, while Figure 6-3 shows the distribution of panic sales per household. Of households with at least one panic-selling event, 21706 of them (80.8%) did so once within our sample period, while 3081 (11.4%) did so twice. The mean and standard deviation of the number of panic sells per investor are 1.35 and 0.98 respectively. These numbers suggest that we are seeing a behavioral pattern that is different from overtrading.

6.5.1 When do the investors panic sell?

As can be seen from Figure 6-4, panic sales occur regularly, with a base level around 0.10%. By overlaying the change in S&P 500 value against the plot, however, we noticed that the spikes in the proportion of households panic selling coincide with sharp falls in the stock market. Looking at the top ten months with the highest proportion of active investors panic selling, we see they include significant stock market events (Table 6.1), confirming the common idea that investors freak out in times of market uncertainty. In the rest of the chapter, we collectively refer to these months as 'crisis' periods, and panic selling specifically in these months as 'freakouts'.

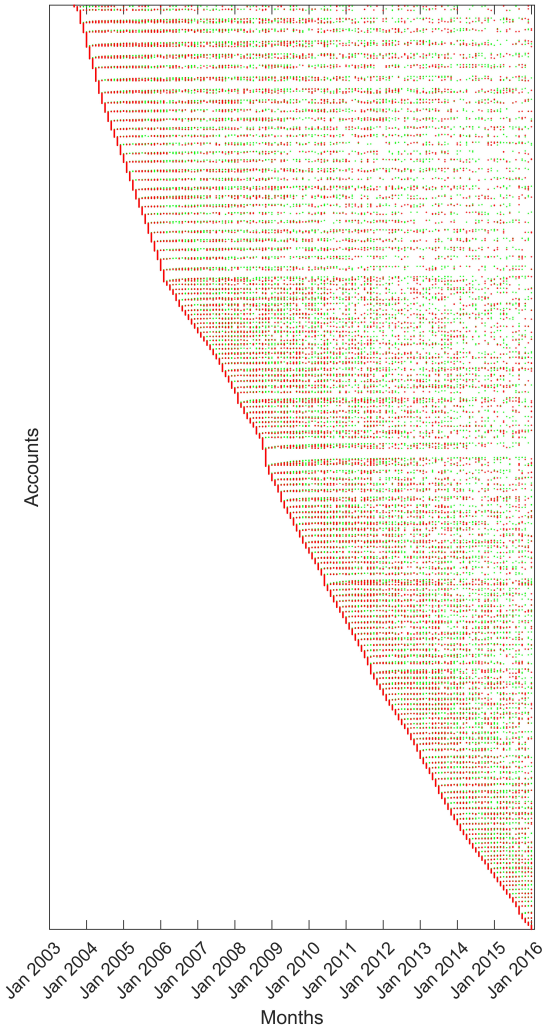


Figure 6-2: Heat map of panic-selling events and returns to market. Row entries are unique to households, while the horizontal axis denotes time in the YYYYMM format. Red denotes a panic sale, while green denotes a return to market for the household.

6.5.2 Returning to the market

We ask the question,"What happens to an investor after he panic sells?" As shown in Figure 6-8, as of December 31, 2015, 30.9% of these investors have not taken on risky assets since they freaked out. Of the freakouts that concluded with the investor reentering the market, 58.5% and 13.1% lasted 1 to 5 months and 6 to 10 months, respectively.

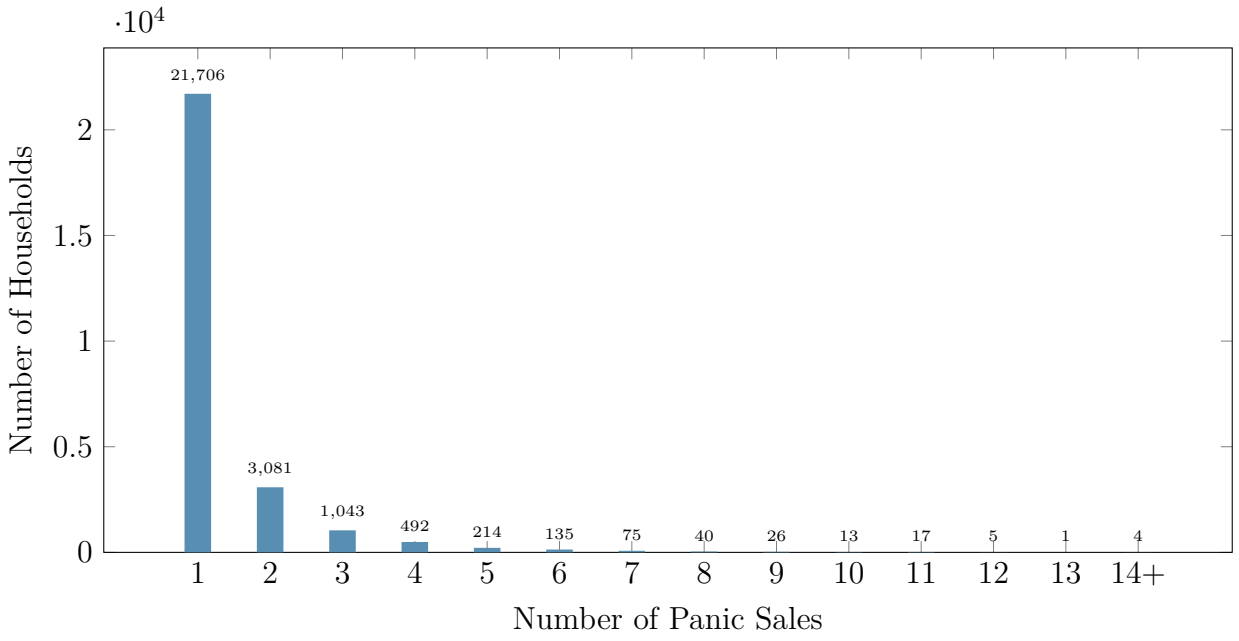


Figure 6-3: Frequency of panic sales. 80.8% and 11.4% of all investors made panic sales once and twice, respectively.

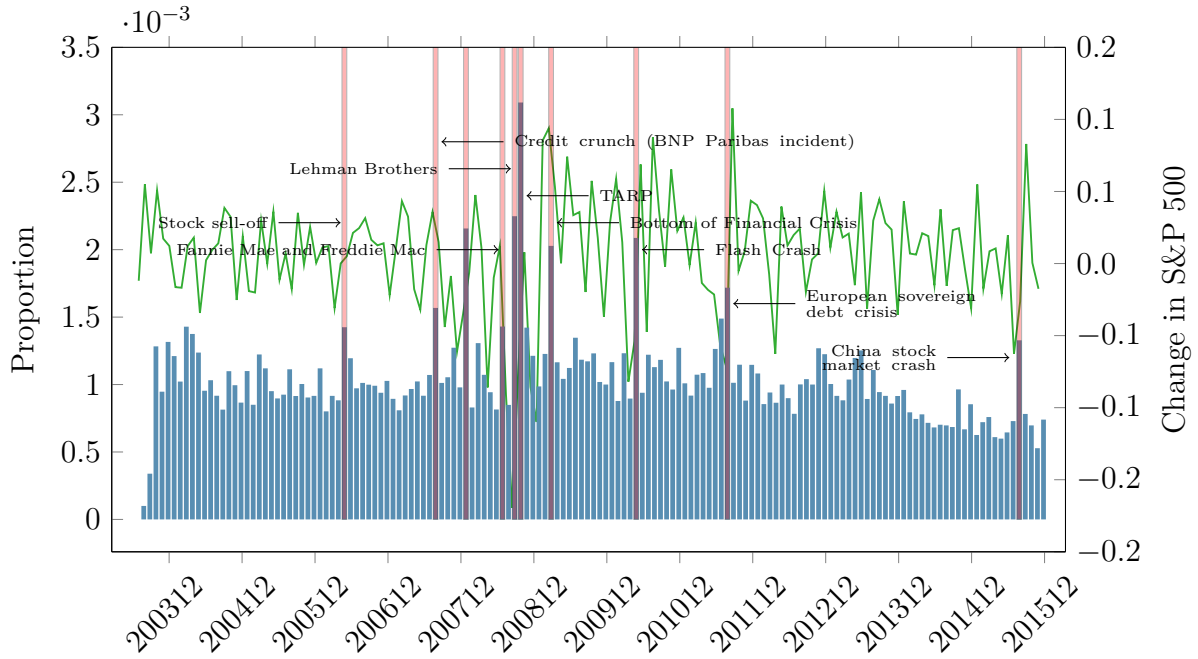


Figure 6-4: The proportion of active households who panic sold in each month (YYYYMM). The green line depicts the month-to-month change in S&P 500 over time.

YYYYMM	Counts	Active Accts	Pct Of Active Accts (%)	Known Event
200605	290	203534	0.142	Stock and commodity sell-off
200708	338	215628	0.157	Credit crunch (BNP Paribas incident)
200801	476	220646	0.216	
200807	324	226614	0.143	Fannie Mae and Freddie Mac
200809	513	228220	0.225	Lehman Brothers
200810	710	229703	0.309	TARP
200903	475	234268	0.203	Bottom of Financial Crisis
201005	512	245319	0.209	Flash Crash
201108	440	256098	0.172	European sovereign debt crisis
201508	377	283900	0.133	China stock market crash

Table 6.1: Months with the highest relative percentage of liquidations and the corresponding events.

6.5.3 Portfolio characteristics of investors who panic sold

Table 6.2 tabulates the distribution of the value of portfolios just prior to panic sales. 43.2% of the portfolios are less than $20000 in value. The 25th, 50th, 75th and 90th percentile portfolio values are $7688.78, $27605.35, $96387.94, and $277986.65, respectively.

Portfolio Value	Count	Percentage
0–20000	15714	43.20
20000–40000	5284	14.53
40000–60000	3049	8.38
60000–80000	1945	5.35
80000–100000	1549	4.26
100000–200000	3796	10.44
200000–400000	2625	7.22
400000–600000	976	2.68
600000–800000	479	1.32
800000–1000000	261	0.72
1000000–∞	696	1.91
Total	36374	100.01%

Table 6.2: Distribution of portfolio value immediately prior to panic sales. Percentages do not sum to exactly 100.00 due to rounding errors.

6.5.4 Is panic selling optimal?

Are investors wise to liquidate most of their risky assets over a short period of time? On the one hand, one may subscribe to the view that investors are rational actors who are optimally changing the composition of their portfolio. This behavior can be observed in 'stop-loss' or

'trailing-stop' trades, in which trades are executed to limit further losses when the market is plunging, or to lock in profits when the market is on the rise. On the other hand, it is possible that investors are panicked by changing market conditions and therefore sell, despite knowing that it is not in their best interest to do so.

Opportunity cost of panic selling

We examine if panic selling is an optimizing behavior by first calculating hypothetical returns over various time horizons in which the investor did not panic sell. That is, we assume that the panicked investor did not sell off his risky assets, and track the hypothetical returns of this portfolio until the investor actually returned to the market. We then average the returns to get an aggregate estimate of the potential returns that were forgone. If these hypothetical returns are negative, we conclude that panic selling is an optimizing behavior, as it prevented further losses. On the other hand, if the hypothetical returns are positive, this implies that investors could have profited if they simply left the money in their accounts.

Figures 6-5 shows the average hypothetical returns over 20-, 100-, 200-, 600- and 1000-day periods of investors who liquidated in the tabulated month. Negative values indicate that the average investor would have lost money, while positive values indicate that the average investor would have been better off had he not liquidated.

We found that the average hypothetical returns are highly correlated with the returns of the S&P 500 over the same time horizons. Our results also suggest that the experience of the individual investor depend on market conditions when he exited, and the duration of his exit.

This point is more obvious in Figure 6-6, where we compute the median of the hypothetical returns conditioned on the time of their liquidation and the duration of their exit. We plot the kernel regressions to smooth out variations over the time horizons. As can be seen, during the financial crisis, it was typically wise to liquidate one's entire portfolio of risky assets over the short to medium term (less than 35 months). A person who liquidated at the start of the crisis and then left the market for 15 months at that time saved himself from losing another 17%. Holding out for more than 34 months after liquidation, however, would have caused the investor to miss the post-2009 market rally and forgo potential profits.

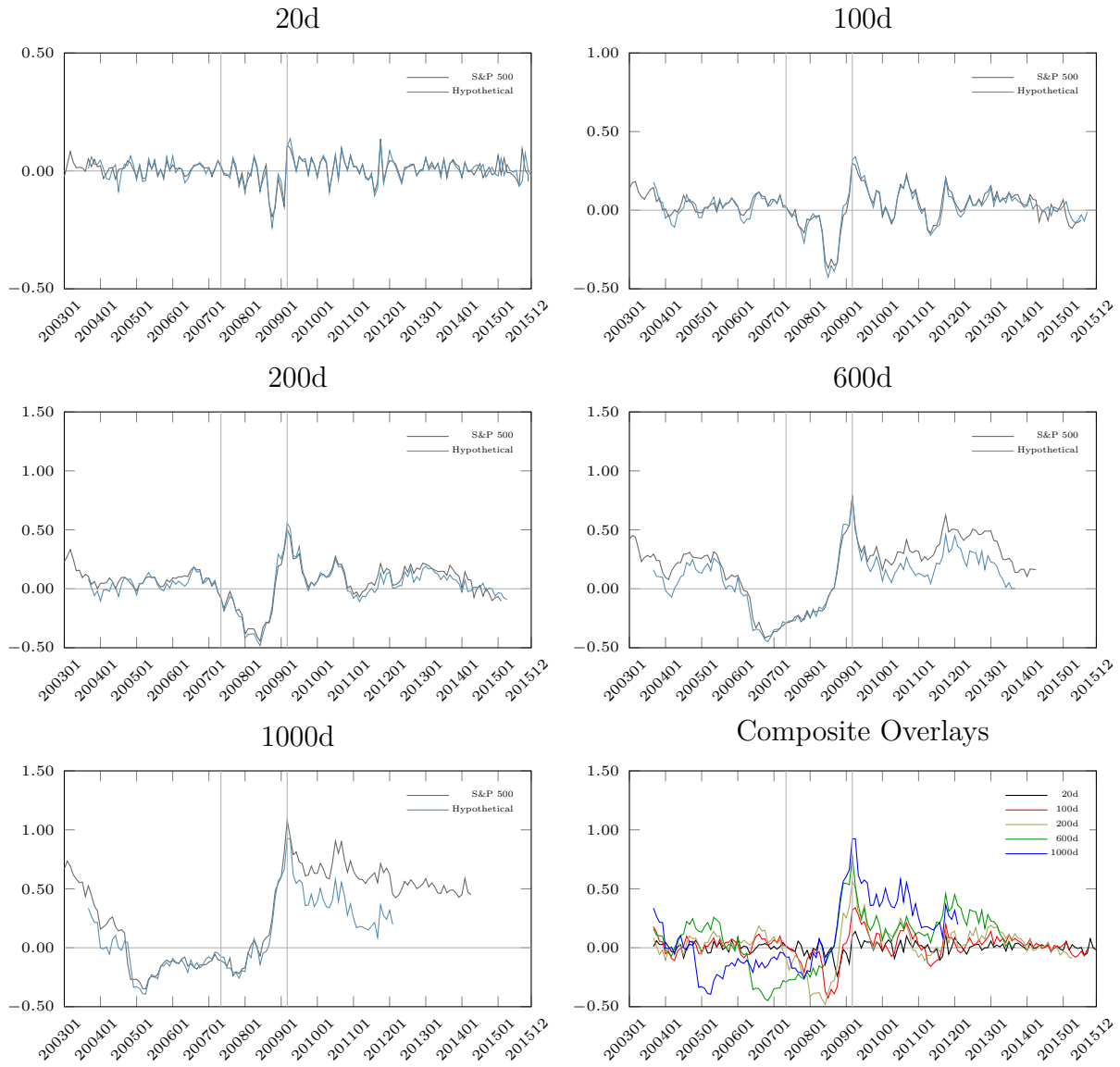


Figure 6-5: Median hypothetical returns of investors who liquidated in a particular month (YYYYMM) over d days. This is constructed by assuming that the investor did not panic sell and held his portfolio for d days.

The reverse is true after the financial crisis. Notably, the persistent rally of the financial markets after the financial crisis ensured that investors who liquidated then would pay a high price in terms of opportunity cost.

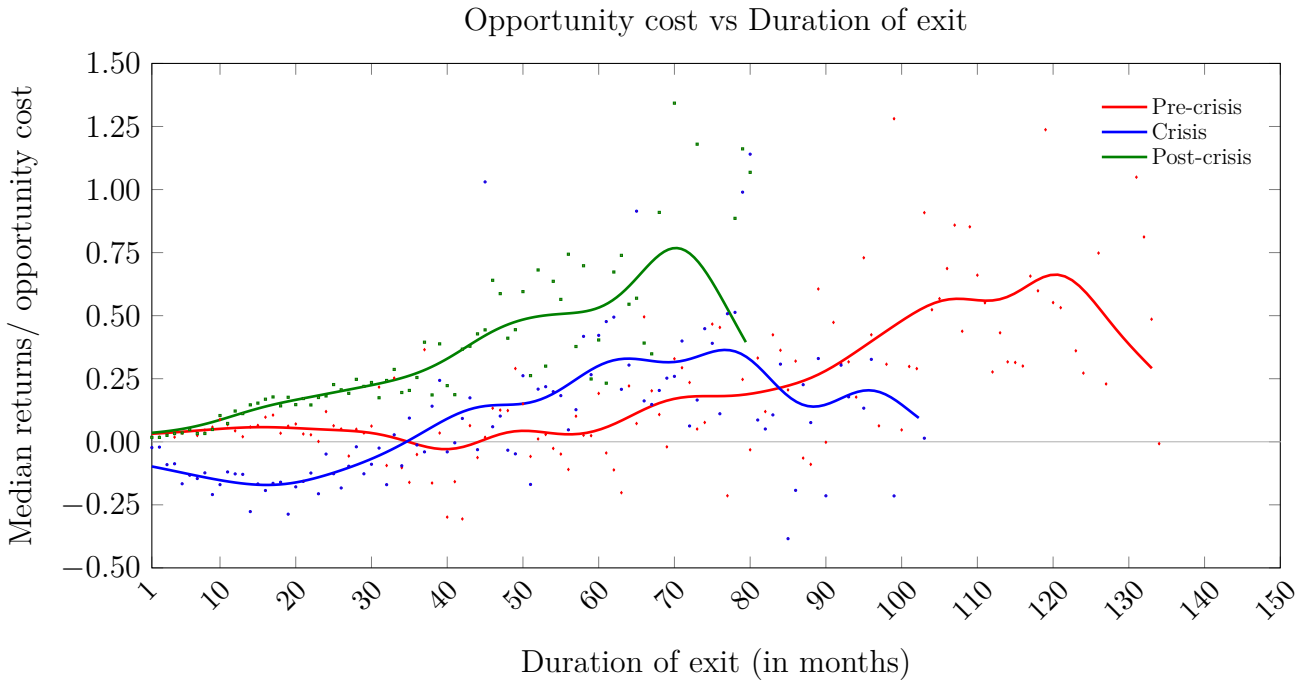


Figure 6-6: Median return of investors under the assumption that they held their portfolio over the duration of their exit. We define the pre-crisis, crisis and post-crisis periods to be Jan 2003–Apr 2007, May 2007–Feb 2009, and Mar 2009–Dec 2015 respectively. The smoothed lines are kernel regressions of the individual series. The number of data points drops exponentially with the duration of staying out (see Figure 6-8). Thus, values for a duration > 60 months are based only on a few data points.

Performance during liquidation

It might be argued that investors who made panic sales did so strategically, which in turn gave them better returns than the market. For example, they could have kept their outperforming stocks while selling the bulk of the underperforming ones, or invested the proceeds of the sales in assets with higher returns. Figure 6-7 shows that this is typically not the case. The median investor trades infrequently, and makes zero to negative returns when out of the market for periods between 1 month and 5 years.

6.5.5 Demographic profile of investors

In this section, we profile the demographic characteristics of investors who liquidated significant parts of their portfolio.

All demographic information in our dataset, with the exception of age, reflects the cus-

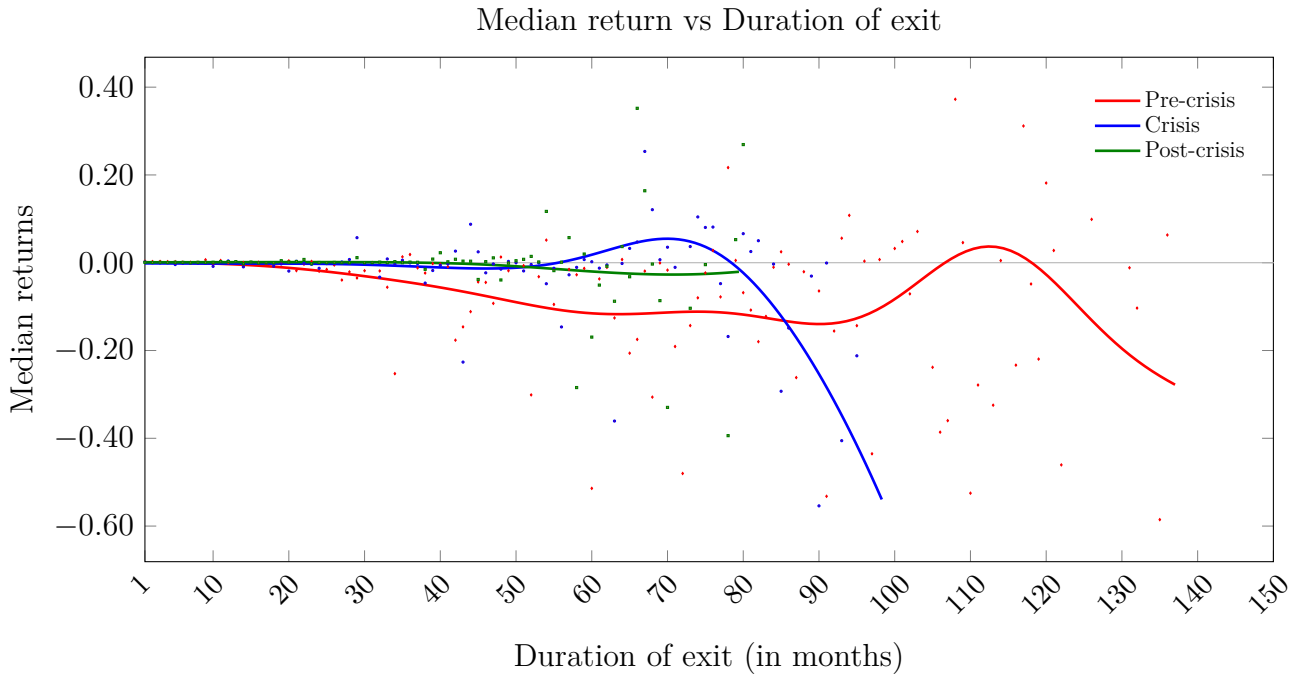


Figure 6-7: Median returns of investors based on their investment actions over the duration of their exit. We define the pre-crisis, crisis and post-crisis periods to be Jan 2003–Apr 2007, May 2007–Feb 2009, and Mar 2009–Dec 2015, respectively. The smoothed lines are kernel regressions of the individual series. The number of data points drops exponentially with the duration of exit (see Figure 6-8). Values for a duration > 60 months are thus based only on a few data points.

tomer profile at the time the brokerage accounts were opened, and is not updated over time. While this may produce inaccuracies in our analysis, we believe that it can still generate insights as to which kind of investors are more likely to panic sell. Certain fields are missing for some investors, as demographic information is collected on a voluntary basis.

Due to the structure of the data, in which a household can contain multiple investing accounts and multiple investing accounts can share a set of customers (please refer to Section 6.8.3 in the Supplementary Materials for more detail), care has to be taken to analyze customer demographics. For each customer in a household, we compute fractional weights based on the size of the portfolios to which they are linked. The computational method is elaborated in Section 6.8.5 in the Supplementary Materials.

Some floating point values may be imprecise in the tables, as we only give the results to two decimal places.

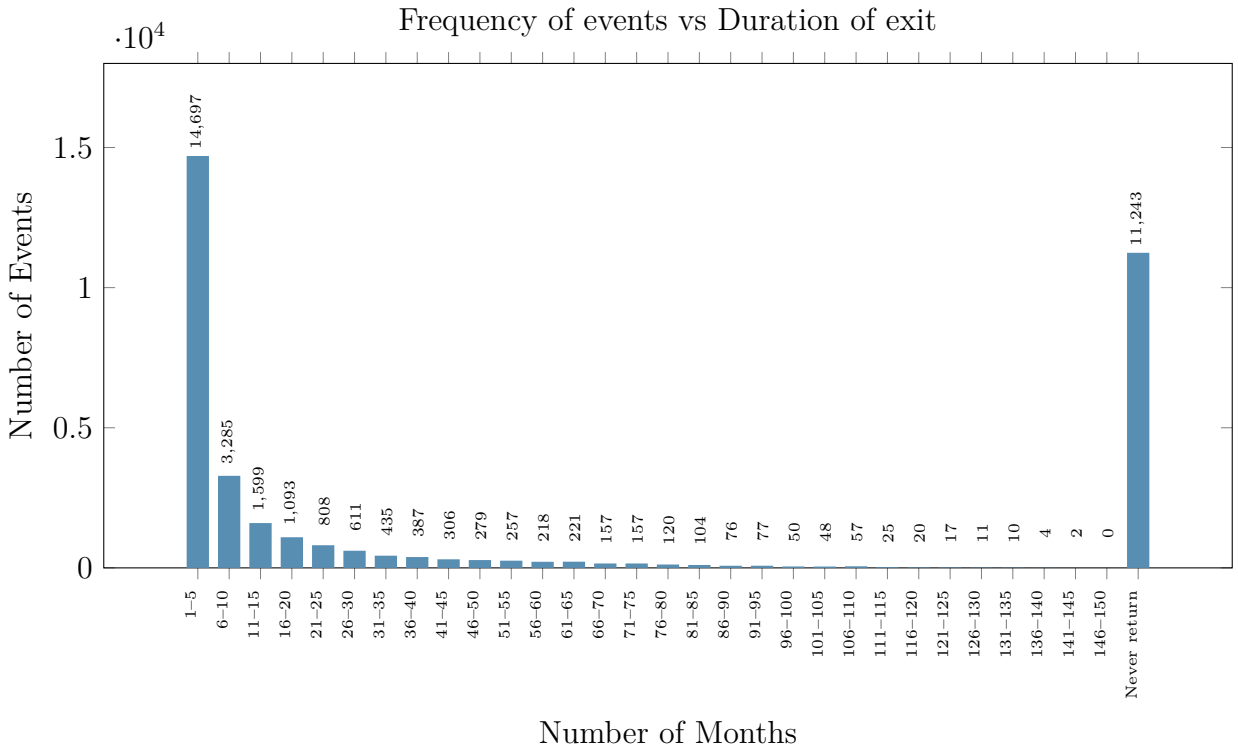


Figure 6-8: Frequency of duration of exit between panic selling and returning to the market.

Age As can be seen in Table 6.3, people between the ages of 45 and 100 have a heightened tendency to make panic sales, both across the entire sample and during crisis periods. Younger investors are less likely to make panic sales by a wide margin.

Marital status Table 6.4 shows that investors who are married or divorced are more likely to freak out across the entire sample period than other groups.

Gender We note that many investors in our dataset do not provide gender information. Among those who volunteered this information, males compose 56.2% of the sample. Previous behavioral finance studies have typically recorded a disproportionate proportion of males [5]. Our analysis shows that males are slightly more likely than females to freak out (i.e. panic sell during periods of high financial stress) but are less likely to panic sell in general (Table 6.5).

Number of dependents Among those with known information about having dependents, investors with no dependents are least likely to panic sell (Table 6.6). There seems to be a

Age group	(A) Liquidation (Full sample)	(B) Liquidation (Crisis periods)	All investors	Rel. prop for (A)	Rel. prop for (B)
Missing	3557.12	759.15	71174.87	0.56^{+}	0.55^{+}
age <21	136.99	18.14	2110.69	0.72^{+}	0.44^{+}
21≤ age <25	170.99	25.70	2819.12	0.68^{+}	0.47^{+}
25≤ age <30	466.03	56.38	6732.05	0.77^{+}	0.43^{+}
30≤ age <35	796.74	103.96	10414.13	0.85^{+}	0.52^{+}
35≤ age <40	869.40	134.75	11447.19	0.85^{+}	0.61^{+}
40≤ age <45	1269.09	241.84	14257.73	0.99	0.88
45≤ age <50	2226.95	454.52	20873.71	1.19^{+}	1.12^{+}
50≤ age <55	3051.63	663.69	27346.63	1.25^{+}	1.25^{+}
55≤ age <60	3475.76	756.39	30437.99	1.27^{+}	1.28^{+}
60≤ age <65	3381.66	736.02	29437.30	1.28^{+}	1.29^{+}
65≤ age <70	3013.68	730.55	26510.99	1.27^{+}	1.42^{+}
70≤ age <75	2042.53	483.52	18938.85	1.20^{+}	1.32^{+}
75≤ age <80	1164.50	298.71	11811.12	1.10^{+}	1.31^{+}
80≤ age <85	638.12	177.44	7442.69	0.96	1.23^{+}
85≤ age <90	359.30	95.05	4728.17	0.85^{+}	1.04
90≤ age <95	164.56	46.08	2238.36	0.82	1.06
95≤ age <100	45.86	15.08	653.64	0.78	1.19
100≤ age <infty	21.08	4.03	234.76	1.00	0.89
Total	26852	5801	299610		

Table 6.3: Distribution of investors by age groups. (A) shows the weights of investors that made panic sales across the entire sample period. (B) shows the weights of investors that freaked out. A proportion less than/greater than 1 indicates that members of the group are less likely/more likely to liquidate compared to members of other groups. $^{+}$ indicates significant at the 1% rejection level.

Category	(A) Liquidation (Full sample)	(B) Liquidation (Crisis periods)	All investors	Rel. prop for (A)	Rel. prop for (B)
Separated	14.88	1.00	191.57	0.87	0.27
Minor	96.80	15.02	1475.72	0.73^{+}	0.53
Widowed	333.90	72.97	4821.70	0.77^{+}	0.78
Missing	7868.55	1734.12	113814.44	0.77^{+}	0.79^{+}
Single	4496.66	901.05	47592.89	1.05^{+}	0.98
Divorced	1187.70	249.79	11464.65	1.16^{+}	1.13
Married	12853.50	2827.05	120249.03	1.19^{+}	1.21^{+}
Total	26852	5801	299610		

Table 6.4: Distribution of investors by marital status. (A) shows the weights of investors that made panic sales across the entire sample period. (B) shows the weights of investors that freaked out. A proportion less than/greater than 1 indicates that members of the group are less likely/more likely to liquidate compared to members of other groups. $^{+}$ indicates significant at the 1% rejection level.

Gender	(A) Liquidation (Full sample)	(B) Liquidation (Crisis periods)	All investors	Rel. prop for (A)	Rel. prop for (B)
Female	378.92	104.19	5943.39	0.71^{+}	0.91
Missing	25822.43	5525.99	286025.14	1.01^{+}	1.00
Male	650.66	170.82	7641.47	0.95	1.15
Total	26852	5801	299610		

Table 6.5: Distribution of investors by gender. (A) shows the weights of investors that made panic sales across the entire sample period. (B) shows the weights of investors that freaked out. A proportion less than/greater than 1 indicates that members of the group are less likely /more likely to liquidate compared to members of the other groups. $^{+}$ indicates significant at the 1% rejection level.

positive correlation between the likelihood of panic selling and the number of dependents.

Number of Dep.	(A) Liquidation (Full sample)	(B) Liquidation (Crisis periods)	All investors	Rel. prop for (A)	Rel. prop for (B)
Missing	3532.45	754.99	70801.04	0.56^{+}	0.55^{+}
0	14808.74	3186.52	156436.02	1.06^{+}	1.05^{+}
1	3090.31	670.57	26994.69	1.28^{+}	1.28^{+}
2	3055.74	710.32	27584.11	1.24^{+}	1.33^{+}
3	1514.31	319.54	11957.38	1.41^{+}	1.38^{+}
4	587.07	110.88	4117.34	1.59^{+}	1.39^{+}
≥ 5	263.39	48.19	1719.41	1.71^{+}	1.45
Total	26852	5801	299610		

Table 6.6: Distribution of investors by number of dependents. (A) shows the weights of investors that made panic sales across the entire sample period. (B) shows the weights of investors that freaked out. A proportion less than/greater than 1 indicates that members of the group are less likely/more likely to liquidate compared to members of the other groups. $^{+}$ indicates significant at the 1% rejection level.

Self-declared investing experience Table 6.7 shows that the likelihood of panic sales and freakouts is most pronounced when the investor has self-declared good or excellent investing experience. Interestingly, those for whom we lack this information, and those who declared themselves to have no investment experience, are less likely to panic sell or freakout.

Category	(A) Liquidation (Full sample)	(B) Liquidation (Crisis periods)	All investors	Rel. prop for (A)	Rel. prop for (B)
Missing	5281.71	1124.01	89774.51	0.66^{+}	0.65^{+}
None	2044.11	395.43	24317.28	0.94^{+}	0.84^{+}
Decline to report	853.97	163.33	9531.72	1.00	0.89
Limited	8972.14	1869.94	98277.98	1.02	0.98
Good	7216.77	1631.21	61775.62	1.30^{+}	1.36^{+}
Excellent	2483.30	617.08	15932.89	1.74^{+}	2.00^{+}
Total	26852	5801	299610		

Table 6.7: Distribution of investors by investment experience. (A) shows the weights of investors that made panic sales across the entire sample period. (B) shows the weights of investors that freaked out. A proportion less than/greater than 1 indicates that members of the group are less likely/more likely to liquidate compared to members of the other groups. $^{+}$ indicates significant at the 1% rejection level.

Self-declared investing knowledge Similar to investing experience, we find that investors who describe their investment knowledge as good or excellent panic sell or freak out in higher proportions compared to their baselines (Table 6.8).

Occupational Group The occupational groups with the three highest risks of panic selling are 'self-employed', 'owners' and 'real estate', while the three occupational groups with the least risk of panic selling are 'paralegal', 'minor' and 'social worker'.

Category	(A) Liquidation (Full sample)	(B) Liquidation (Crisis periods)	All investors	Rel. prop for (A)	Rel. prop for (B)
Missing	8757.51	1581.76	131408.17	0.74^{+}	0.62^{+}
Decline to report	1083.23	220.11	11902.07	1.02	0.96
Limited	7282.50	1650.57	77048.98	1.05^{+}	1.11^{+}
None	1750.25	401.47	16543.69	1.18^{+}	1.25^{+}
Good	6144.44	1480.24	51353.35	1.34^{+}	1.49^{+}
Excellent	1834.07	466.85	11353.75	1.80^{+}	2.12^{+}
Total	26852	5801	299610		

Table 6.8: Distribution of investors by investment knowledge. (A) shows the weights of investors that made panic sales across the entire sample period. (B) shows the weights of investors that freaked out. A proportion less than/greater than 1 indicates that members of the group are less likely/more likely to liquidate compared to members of the other groups. $^{+}$ indicates significant at the 1% rejection level.

Category	(A) Liquidation (Full sample)	(B) Liquidation (Crisis periods)	All investors	Rel. prop for (A)	Rel. prop for (B)
Paralegal	36.35	5.45	576.67	0.70	0.49
Minor	96.81	15.02	1476.89	0.73^{+}	0.53
Social worker	16.95	2.99	285.85	0.66	0.54
Missing	5934.96	1259.68	96257.00	0.69^{+}	0.68^{+}
Government	30.57	4.06	275.68	1.24	0.76
Police-military	133.37	18.39	1244.83	1.20	0.76
Artist	173.55	33.68	2222.40	0.87	0.78
Student	71.22	11.28	741.06	1.07	0.79
Medical	394.31	80.73	5279.71	0.83^{+}	0.79
Education	554.67	117.43	6544.05	0.95	0.93
Skilled labor	1020.02	171.87	9574.47	1.19^{+}	0.93
Scientist	114.69	29.41	1595.88	0.80	0.95
Secretary	232.82	54.37	2926.68	0.89	0.96
Unemployed	981.34	197.95	10505.23	1.04	0.97
Homemaker	765.36	158.54	8095.60	1.05	1.01
S-skilled office	350.59	77.65	3898.63	1.00	1.03
Computer	645.39	130.43	6481.64	1.11^{+}	1.04
Attorney	426.05	89.00	4417.51	1.08	1.04
Engineer	288.28	59.57	2951.42	1.09	1.04
Clergy	32.76	6.06	294.84	1.24	1.06
Cpa	301.43	63.07	3065.35	1.10	1.06
White-collar	1162.82	245.28	11625.54	1.12^{+}	1.09
Physician	496.35	109.52	5109.17	1.08	1.11
Retired	3817.83	908.75	42210.01	1.01	1.11^{+}
Pilot	73.48	13.73	624.20	1.31	1.14
Manager	1449.41	307.67	13569.83	1.19^{+}	1.17^{+}
Marketing	1142.77	219.33	9365.87	1.36^{+}	1.21^{+}
Executive	1412.26	293.41	10756.85	1.46^{+}	1.41^{+}
Financial	733.94	163.30	5903.92	1.39^{+}	1.43^{+}
Professional	1679.33	405.09	14271.18	1.31^{+}	1.47^{+}
Consultant	471.01	126.73	4427.47	1.19^{+}	1.48^{+}
Owner	534.88	118.20	3963.37	1.51^{+}	1.54^{+}
Self employed	967.02	225.22	6914.92	1.56^{+}	1.68^{+}
Real estate	309.39	78.13	2149.44	1.61^{+}	1.88^{+}
Disabled			6.85		
Total	26852	5801	299610		

Table 6.9: Distribution of investors by occupation groups, as classified by the broker. (A) shows the weights of investors that made panic sales across the entire sample period. (B) shows the weights of investors that freaked out. A proportion less than/greater than 1 indicates that members of the group are less likely/more likely to liquidate compared to members of the other groups. $^{+}$ indicates significant at the 1% rejection level.

6.6 Prediction of individual panic sells

Using logistic regression and deep neural network techniques, we attempt to predict panic sales for every individual in the *next* month in advance, given one's demographic attributes, past trading patterns, portfolio history and recent market conditions. In our logistic regression model, we seek a generalized linear model where the separating hyperplane is linear with respect to the input feature space. This allows an easy interpretation of the coefficients in terms of odd ratios, but the class of functions that it can model accurately is restricted. In our machine learning model, we push the limits of prediction by training neural network models of 5 hidden layers and 15 hidden layers of 60 neurons to find similarities between panic-selling events. However, doing so necessarily sacrifices easy interpretation[2]. Despite the drawbacks of each method, we hope to show that there exists significant information in the dataset that would allow us to predict panic selling.

In the rest of this section, we will refer to the occurrence of panic selling in the *next* month as a positive data point, and its absence as a negative data point.

6.6.1 Construction of training and testing datasets

We created a dataset for machine learning using demographic attributes, portfolio states, and market states. Among the demographic attributes used are age, marital status, number of dependents, self-declared investment experience, self-declared investment knowledge and occupational group. We assume equal weights for all the customers in a household when assigning scores to the one-hot categories. For example, if a household has three customers with ages 50, 50 and 70, the category 'Age:50' will have a score of 2/3, while the category 'Age:70' will have a score of 1/3.

For portfolio states, we consider the changes in portfolio balance, the distribution of the portfolio (in cash, equities, options and penny stocks), the nominal and net values of trades, and, the number of trades as functions of time. We incorporated lags of 6 months in order to allow the models to easily pick up time-series signals. We use the month-to-month change, the volatility of prices, and the trading volume of the S&P 500 as indicators of market

[2]Of course, the notion of 'interpretability' is itself up for debate [17]

conditions. Market information was downloaded from Yahoo! Finance. We considered lags of 12 months for these market variables.

A summary of the variables is shown in Table 6.10. For variables that are unbounded from either side (e.g. $\in \mathbb{Z}_+$), we shifted the midpoint value to zero, then scaled them to be within $[-1, 1]$. In total, the inputs into our models are vectors of length 507.

For the purpose of benchmarking the predictive power of the models, we perform a random 60-40 training testing split. This ratio is maintained for each of the two classes, so that the test set is representative of the entire sample. In order to prevent cross-contamination between the training and testing sets, which would falsely inflate the performance of the models, we ensure that all of a household's data points are either in the testing set or the training set. We use the training set of investors for both rounds of training, but evaluate the performance of the models only on the test set in order to detect over-fitting of data points. We do not require a validation set, as we do not perform any parameter optimization or model selection.

6.6.2 Evaluation

Panic sales are rare events. In all, we obtain 25,418,786 data points, of which only 33,226, or 0.131%, are panic sales (The number of panic sales is less than the number reported in the previous section because we wish to create a lagged series, which forces us to drop some data points). This extremely unbalanced dataset poses a significant problem for any binary classification algorithm. For example, if an algorithm made the prediction of 'not a panic sell' for any input, it would achieve an accuracy of 99.869%, an eye-popping but practically irrelevant number. To get a better sense of the performance of the models, we compute accuracy rates separately for both the negative and the positive examples. In addition, we display the receiver operating characteristic (ROC) and precision-recall (PR) curves and report the areas under them. An explanation of these metrics is given in Section 6.8.7 in the Supplementary Materials. Since the ROC and PR curves serve to answer different questions, we have included the results for both in order to allow our readers to decide if the models are useful for their applications.

Description	Variable type
Demographics	
Age	Discrete, 83 groups
Marital status	Discrete, 7 groups
Number of dependents	8 groups
Investment experience	Discrete, 6 groups
Investment knowledge	Discrete, 6 groups
Occupation group	Discrete, 35 groups
Portfolio Factors	
Risky assets balance	$\mathbb{R}_+$
Penny stocks balance	$\mathbb{R}_+$
Options balance	$\mathbb{R}_+$
Cash balance	$\mathbb{R}_+$
Portfolio balance	$\mathbb{R}_+$
Number of risky assets	$\mathbb{Z}_+$
Number of penny stocks	$\mathbb{Z}_+$
Pct of cash in portfolio	[0,1]
Pct of risky asset (value) in portfolio	[0,1]
Pct of penny stocks in risky asset (value)	[0,1]
Pct of options in risky asset (value)	[0,1]
Pct of penny stocks in risky asset (count)	[0,1]
Net value of trades	$\mathbb{R}$
Nominal value of trades	$\mathbb{R}_+$
Nominal value of intraday trades	$\mathbb{R}_+$
Pct of trades involved in intraday trades (value)	[0,1]
Number of trades	$\mathbb{Z}_+$
Number of intraday trades	$\mathbb{Z}_+$
Pct of trades involved in intraday trades (abs num)	[0,1]
Net value of trades as percentage of portfolio balance	[0,1]
Nominal value of trades as percentage of portfolio balance	[0,1]
Is the investor in or out of the market	$\{0, 1\}$
We consider 6 months running lags for all the portfolio factors except the last	
Market State	
Month-to-month change in the S&P 500	$\mathbb{R}$
Month-to-month change in the volume traded of GSPC	$\mathbb{R}$
Volatility of the volume traded of GSPC within the month	$\mathbb{R}_+$
Volatility of the price over the past 20 days	$\mathbb{R}_+$
Volatility of the price over the past 60 days	$\mathbb{R}_+$
Volatility of the price over the past 180 days	$\mathbb{R}_+$
We consider 12 months running lags for market factors	

Table 6.10: List of raw variables used to construct the machine learning data set

6.6.3 Computation

For all the models, we use the cross-entropy loss and train the models to optimality using batch gradient descent (GD) with Adaptive Momentum [22]. It can be shown that minimizing the cross-entropy loss yields the maximum likelihood estimate of the parameters. For each batch of 150000, we randomly draw half of the samples from each of the two classes (with replacement for the positive class and without replacement for the negative class) in order to prevent the classifiers from over-emphasizing either class, which would be the natural tendency of the classifier had we selected the training batch at random. We terminate the training when we determine that the accuracy and/or loss has been saturated. We note in passing that, given the appropriate training schedule, the solution converged on by GD for the logistic regression will be the global minimum solution with respect to the loss of the training set.

All the models were implemented on Tensorflow 1.6 with CUDA 9.0/ CuDNN 7.0, and training was executed on a single Microsoft Azure NC12 instance, which contains 2 Nvidia Tesla K80 GPUs.

6.6.4 Results

The accuracy curves, receiver operating characteristic curves and precision-recall curves on the testing set are shown in Figure 6-9, 6-10 and 6-11 respectively. As can be seen from Figure 6-9, all the models have been trained to convergence. The neural networks converge after approximately 2000 steps, while the logistic classifier converged after approximately 8000 steps. There is no evidence that there is any form of overfitting on the train set, despite the 15-layer neural network containing over 56000 parameters.

The final accuracy rates, areas under the ROC curves (AUROCs) and areas under the PR curves (AUPRCs) on the test set for all the models are reported in Table 6.11. We can see that the neural networks outperform the logistic classifier on all metrics. Between the neural networks, the 15-layer network showed an improvement of 1.3 percentage points over the 5-layer network on the positive data, but a deterioration of 1.5 percentage points on the negative data. We can see that the 5-layer neural network marginally outperforms the

15-layer neural network on the AUROC and AUPRC metrics, but the differences may be simply due to randomness in training. The comparable performance of the neural networks shows that a 5-layer neural network has enough capacity to approximate the function and a larger network is unnecessary.

Model	Accuracy		AUROC	AUPRC ($\times 10^{-3}$)
	Positive samples	Negative samples		
Random Predictor	–	–	0.500	1.307
Logistic Regression	57.9%	78.8%	0.739	5.521
Neural Net (5 hidden layers)	69.5%	81.5%	0.821	15.184
Neural Net (15 hidden layers)	70.8%	79.0%	0.813	13.819

Table 6.11: Performance of the models on the test set

Interpreting the logistic classifier

We attempt to interpret the coefficients of the logistic classifier. We group the variables according to their classification type (demographic factor vs. market factor vs. portfolio factor) and report the top 10 most important variables according to the absolute value of the weights of the coefficients. This works in our analysis, as we have monotonically transformed values to between -1 and 1. Our results are shown in Table 6.12.

Age dominates the list of the most important demographic variables. In general, being young or elderly decreases the risk of panic selling. Being disabled or a minor also lowers the likelihood of panic selling. While not shown, declaring oneself a member of the 'clergy', an 'owner' or an 'executive' increases the likelihood of panic selling. In addition, having self-declared 'excellent' investment experience increases the odds of panic selling. These results substantially agree with the analysis by demographic slices in Section 6.5.5.

Among all the market factors, lagged series of the 20-day S&P 500 volatility, the 60-day S&P 500 volatility and the volatility of the S&P 500 trading volume are the most important factors in predicting panic sales. The signs of the coefficients are mixed.

Our analysis of the coefficients for the portfolio factor shows that the likelihood of a panic sale increases with the percentage of daily trades made by the investor. Furthermore, an investor will be more likely to panic sell if options compose a larger proportion of the entire portfolio. The liquidation of the portfolio has been added as a variable to help the

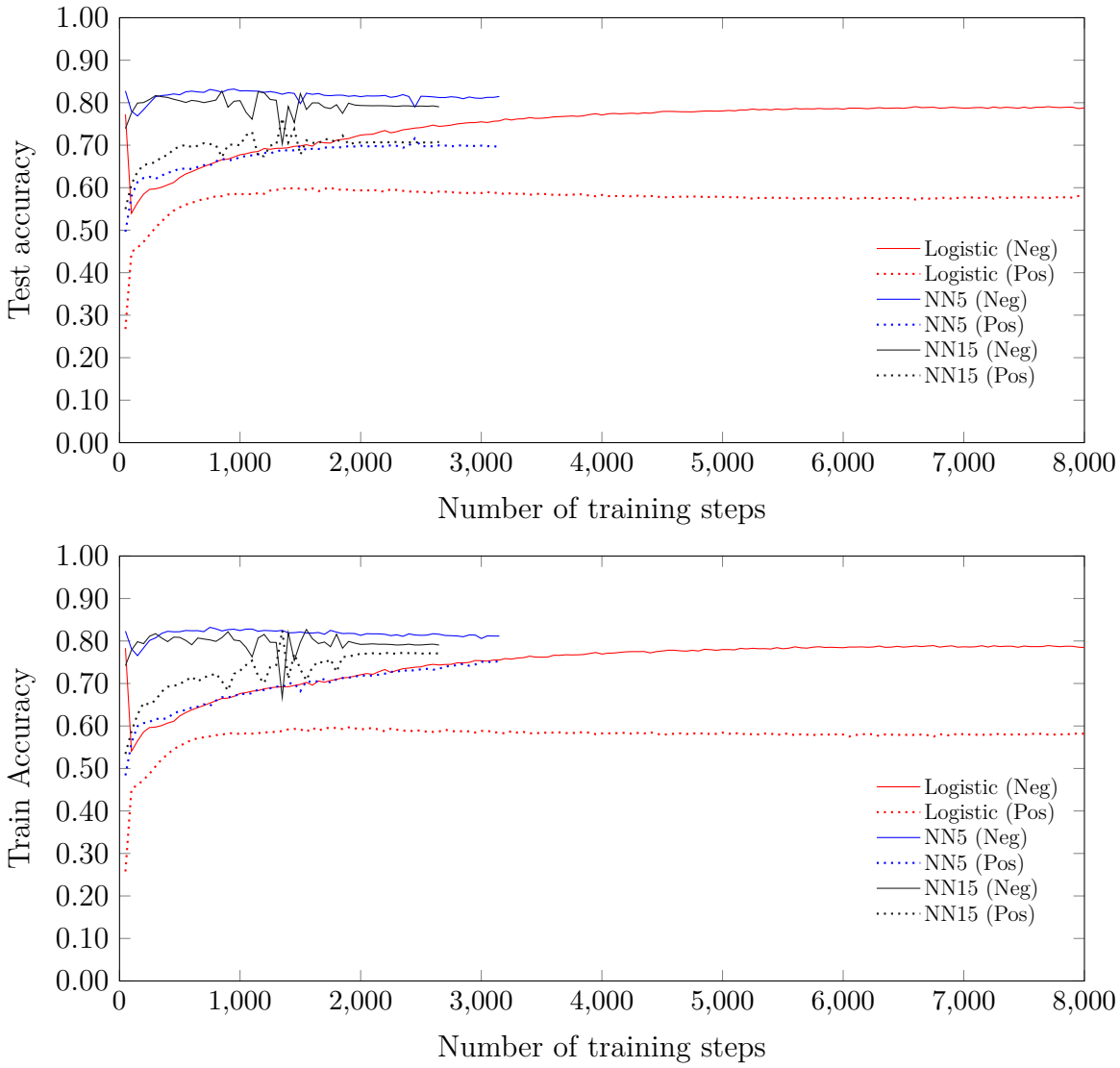


Figure 6-9: Accuracy curves over training steps. The training of the 15-layer and 5-layer neural networks were terminated at around the 2650th and 3150th step respectively as we deemed that they have converged. The logistic classifier was terminated at around the 8000th step.

convergence of the model, and the model accurately deciphered that the chance of a panic sale is high when the portfolio has not been liquidated. This serves as a sanity check that our model is picking up the correct signals.

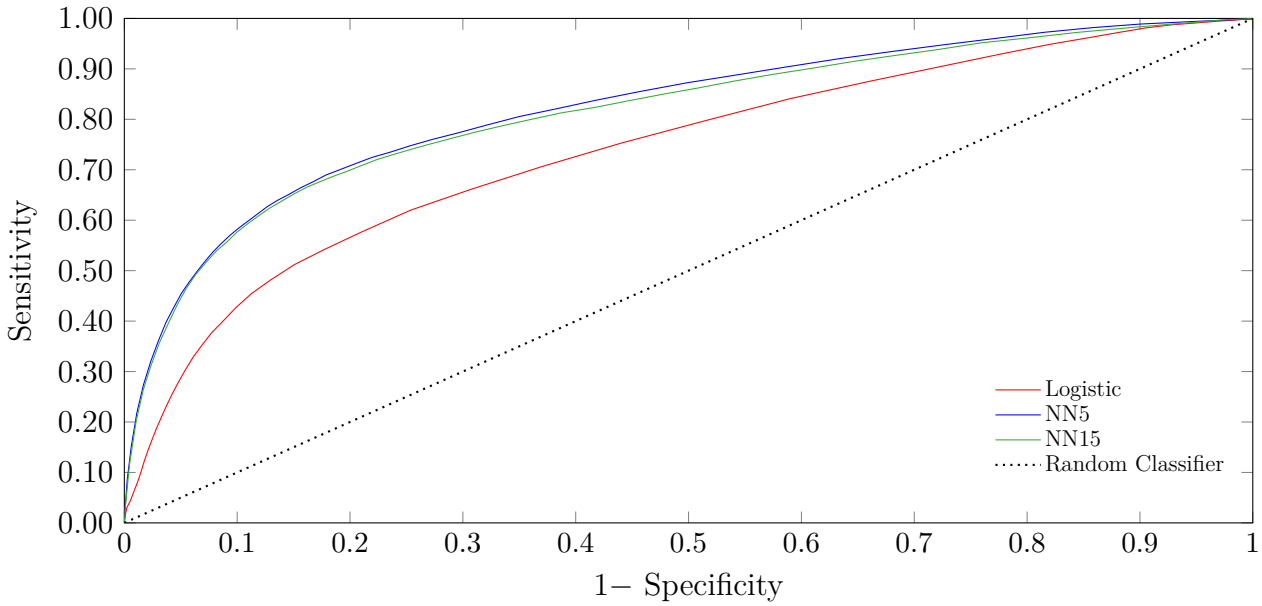


Figure 6-10: Receiver operating characteristic (ROC) curves of the trained models.

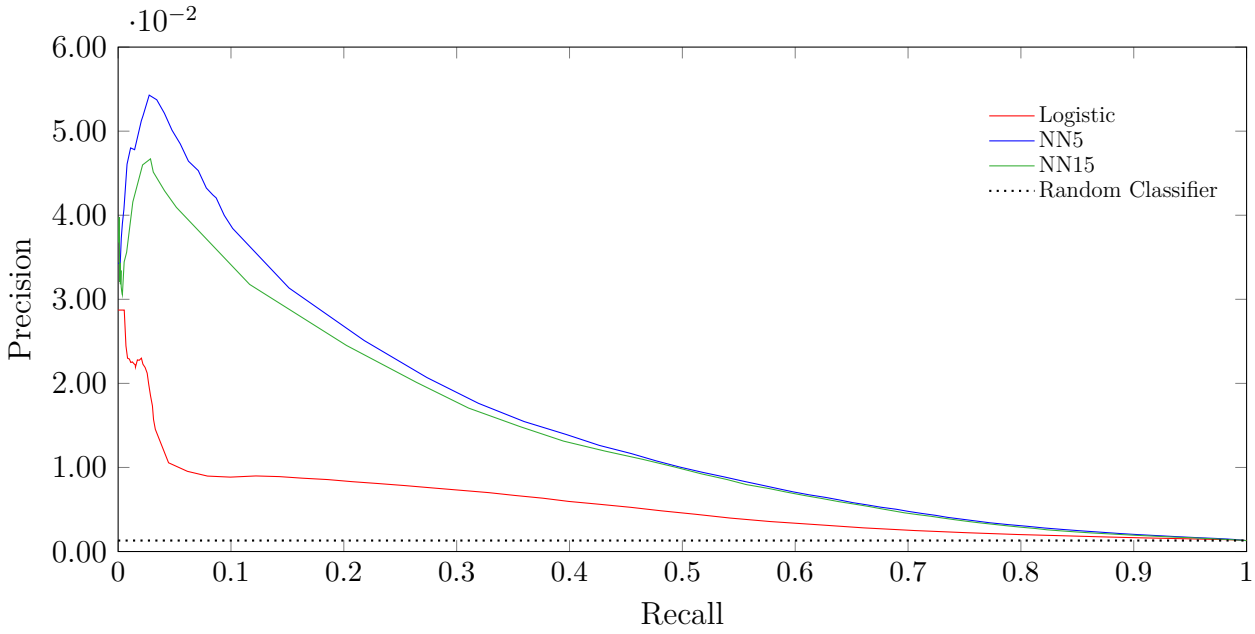


Figure 6-11: Precision-recall (PR) curves of the trained models.

6.7 Conclusion

The analyses in this chapter hinge on the heuristic we developed to identify panic sales. To test the robustness of our results, we performed additional runs with different parameters. We find that, although decreasing the thresholds will increase the number of panic sales identified

Variable Name	Description	Coefficient
Demographic factors		
Age:97	Age of 97	-0.885
dmsa_curr_occup_tx SOCIALWORKER	Occupation: Social Worker	-0.759
Age:<21	Age less than 21	-0.642
Age:23	Age of 23	-0.620
Age:26	Age of 26	0.588
Age:94	Age of 94	-0.587
Age:99	Age of 99	-0.586
dmsa_curr_occup_tx DISABLED	Occupation: Disabled	-0.546
dmsa_curr_occup_tx MINOR	Occupation: Minor	-0.531
invst_exprc_cd E	Investment experience: excellent	0.501
Market factors		
60d_price_vol_lag_8	60 days volatility of S&P500 (8 months ago)	0.758
20d_price_vol_lag_4	20 days volatility of S&P500 (4 months ago)	-0.748
20d_price_vol	20 days volatility of S&P500	0.730
volume_vol	Volatility of volume traded in S&P500 across one month	0.723
60d_price_vol_lag_5	60 days volatility of S&P500 (5 months ago)	-0.701
20d_price_vol_lag_6	20 days volatility of S&P500 (6 months ago)	-0.690
20d_price_vol_lag_9	20 days volatility of S&P500 (9 months ago)	0.687
20d_price_vol_lag_11	20 days volatility of S&P500 (11 months ago)	0.680
volume_vol_lag_7	Volatility of volume traded in S&P500 across one month (7 months ago)	0.671
volume_vol_lag_5	Volatility of volume traded in S&P500 across one month (5 months ago)	0.620
Portfolio factors		
pct_intra_day_trades_lag_5	Percentage of intra-day trades in a month (5 months ago), by counts	0.786
pct_val_options_lag_1	Percentage of portfolio that is options (1 month ago), by value	0.765
pct_val_options_lag_6	Percentage of portfolio that is options (6 months ago), by value	0.736
pct_intraday_val_lag_2	Percentage of intra-day trades in a month (2 months ago), by value	0.716
inMarket	'1' if the portfolio has not been liquidated	0.708
pct_intraday_val_lag_1	Percentage of intra-day trades in a month (1 month ago), by counts	0.699
pct_val_options_lag_3	Percentage of portfolio that is options (3 months ago), by value	0.687
pct_intraday_val	Percentage of intra-day trades in this month, by value	0.664
pct_intra_day_trades	Percentage of intra-day trades in this month, by counts	0.661
pct_val_options_lag_4	Percentage of portfolio that is options (4 months ago), by value	0.648

Table 6.12: Most important variables in the logistic classifier.

across all time periods, there is still a disproportionate number of accounts which panic sell in periods of high financial stress (see Section 6.8.6 of the Supplementary Materials).

Panic selling and freaking out are distinct behavioral patterns in finance that differ from other previously studied patterns. While the disposition effect claims that investors tend to hold on to their losers and keep their winners, we see that investors who made panic sales achieve only a slightly negative return after they liquidate. Also, in contrast to overtrading, investors who made panic sales did so infrequently. We see that panic selling spikes in periods of crisis, suggesting a relationship between panic selling and market conditions. Our logistic model suggests that recent market volatility influences panic selling behavior.

Panic selling and freakouts often have negative connotations. We show that this negativity may not always be warranted. While panic selling in normal market conditions is indeed harmful to the median retail investor, freaking out in environments of sustained market

decline prevents further losses and protects one's capital.

Panic sales are not random events. Specific types of investor, such as those with less than $20000 in portfolio value, tend to liquidate more frequently than others. Subtle patterns in portfolio history, past market movements, and demographic profile can be exploited by deep neural networks to accurately predict if an investor will panic sell in the near future.

Unfortunately, the problem of causation cannot be addressed with the data we have. Therefore, our study does not address *why* investors panic sell. This topic, however, would doubtless be an interesting direction for future research.

Chapter References

[1] Shima Amini, Bartosz Gebka, Robert Hudson, and Kevin Keasey. A review of the international literature on the short term predictability of stock prices conditional on large prior price changes: Microstructure, behavioral and risk related explanations. *International Review of Financial Analysis*, 26:1–17, 2013.

[2] Malcolm Baker and Jeffrey Wurgler. Investor sentiment and the cross-section of stock returns. *The Journal of Finance*, 61(4):1645–1680, 2006.

[3] Turan G Bali, K Ozgur Demirtas, and Haim Levy. Is there an intertemporal relation between downside risk and expected returns? *Journal of Financial and Quantitative Analysis*, 44(4):883–909, 2009.

[4] Brad M Barber and Terrance Odean. Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The Journal of Finance*, 55(2): 773–806, 2000.

[5] Brad M Barber and Terrance Odean. Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics*, 116(1):261–292, 2001.

[6] Brad M Barber, Terrance Odean, and Ning Zhu. Systematic noise. *Journal of Financial Markets*, 12(4):547–569, 2009.

[7] Robert J Barro. Rare disasters and asset markets in the twentieth century. *The Quarterly Journal of Economics*, 121(3):823–866, 2006.

[8] Robert J Barro. Rare disasters, asset prices, and welfare costs. *American Economic Review*, 99(1):243–64, 2009.

[9] W Scott Bauman, C Mitchell Conover, and Robert E Miller. Investor overreaction in international stock markets. *Journal of Portfolio Management*, 25:102–111, 1999.

[10] Alexandros V Benos. Aggressiveness and survival of overconfident traders. *Journal of Financial Markets*, 1(3-4):353–383, 1998.

[11] Tim Bollerslev and Viktor Todorov. Tails, fears, and risk premia. *The Journal of Finance*, 66(6):2165–2211, 2011.

[12] De Bondt, FM Werner, and Richard H Thaler. Further evidence on investor overreaction and stock market seasonality. *The Journal of Finance*, 42(3):557–581, 1987.

[13] Werner FM Bondt and Richard Thaler. Does the stock market overreact? *The Journal of Finance*, 40(3):793–805, 1985.

[14] Keith C Brown, WV Harlow, and Seha M Tinic. The risk and required return of common stock following major price innovations. *Journal of Financial and Quantitative Analysis*, 28(1):101–116, 1993.

[15] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[16] Navin Chopra, Josef Lakonishok, and Jay R Ritter. Measuring abnormal performance: do stocks overreact? *Journal of Financial Economics*, 31(2):235–268, 1992.

[17] F. Doshi-Velez and B. Kim. Towards a rigorous science of interpretable machine learning. *ArXiv e-prints*, Feb 2017.

[18] Xavier Gabaix. Variable rare disasters: An exactly solved framework for ten puzzles in macro-finance. *The Quarterly Journal of Economics*, 127(2):645–700, 2012.

[19] Kathryn M Kaminski and Andrew W Lo. When do stop-loss rules stop losses? *Journal of Financial Markets*, 18:234–254, 2014.

[20] Gautam Kaul and Mahendrarajah Nimalendran. Price reversals: Bid-ask errors or market overreaction? *Journal of Financial Economics*, 28(1-2):67–93, 1990.

[21] Bryan Kelly and Hao Jiang. Tail risk and asset prices. *The Review of Financial Studies*, 27(10):2841–2871, 2014.

[22] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[23] Joachim Klement. Assessing stop-loss and re-entry strategies. 2013. Available at SSRN: https://ssrn.com/abstract=2277722 or http://dx.doi.org/10.2139/ssrn.2277722.

[24] Andrew W Lo and Alexander Remorov. Stop-loss strategies with serial correlation, regime switching, and transaction costs. *Journal of Financial Markets*, 34:1–15, 2017.

[25] Terrance Odean. Are investors reluctant to realize their losses? *The Journal of Finance*, 53(5):1775–1798, 1998.

[26] Jinwoo Park. A market microstructure explanation for predictable variations in stock returns following large price changes. *Journal of Financial and Quantitative Analysis*, 30(2):241–256, 1995.

[27] Robert S Pindyck and Neng Wang. The economic and policy consequences of catastrophes. *American Economic Journal: Economic Policy*, 5(4):306–39, 2013.

[28] Andy Puckett and Xuemin Sterling Yan. Short-term institutional herding and its impact on stock prices. 2008. Unpublished manuscript, University of Missouri.

[29] Charles Rotblot. *The Danger of Getting Out of Stocks During Bear Markets*, 2004. URL `http://www.aaii.com/journal/article/the-danger-of-getting-out-of-stocks-during-bear-markets.touch`.

[30] Michael S Rozeff and Mir A Zaman. Overreaction and insider trading: Evidence from growth and value portfolios. *The Journal of Finance*, 53(2):701–716, 1998.

[31] Hersh Shefrin and Meir Statman. The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance*, 40(3):777–790, 1985.

[32] Leilei Shi, Liyan Han, Yiwen Wang, Ding Chen, Yan Piao, and Chengling Gou. Market crowd trading conditioning and its measurement. 2011. Available at SSRN: https://ssrn.com/abstract=1661515 or http://dx.doi.org/10.2139/ssrn.1661515.

[33] Meir Statman, Steven Thorley, and Keith Vorkink. Investor overconfidence and trading volume. *The Review of Financial Studies*, 19(4):1531–1565, 2006.

[34] Jia Wang, Gulser Meric, Zugang Liu, and Ilhan Meric. Investor overreaction to technical insolvencyand bankruptcy risks in the 2008 stock market crash. *The Journal of Investing*, 22(2):8–14, 2013.

[35] Martin Weber and Colin F Camerer. The disposition effect in securities trading: An experimental analysis. *Journal of Economic Behavior & Organization*, 33(2):167–184, 1998.

6.8 Supplementary Materials

6.8.1 Account security holding and portfolio allocations data

The raw position files consist of monthly snapshots that record the quantities and month-end prices of each security held in the portfolio of all accounts within the sample that were open on the last day of the month. Securities are uniquely identified either by CUSIP ID or ticker symbol, and accounts are uniquely identified by an anonymized numeric key ('acid key' or 'acid'). An internal asset class assignment for each security is also provided within the brokerage account files, which classifies each CUSIP/ticker as one of: 'equities', 'mutual funds', 'fixed income securities', 'cash or cash equivalents', or 'options'. Additionally, a separate identifier is provided distinguishing 'cash equities' from 'ETFs' within the equities category.

Key	Description	Format
month	Month of snapshot (all positions are those held at month-end)	YYYYMM (e.g. 201512)
settle_qty	Quantity of shares held in security	double
ticker_symbol	Ticker symbol	string (e.g. AAPL)
cusip_num	Security identification number registered with the US SEC	9-digit alpha-numeric (e.g. 17275R102)
issue_price	Exchange-listed close price on the last market day of the assigned month	double
product_grplvl1	Top level security type identifier	string (e.g. 'EQUITY')
product_grplvl2	Mid level security type identifier	string (e.g. 'EQUITY')
product_grplvl3	Bottom level security type identifier	string (e.g. 'EQUITY')
acid_key	Unique account identification number	integer (e.g. 9374629673)

Table 6.13: Summary of the data fields in the positions datafile.

6.8.2 Trading data

The raw trade files consist of annual records of all trades executed by the sampled accounts during the year. Each trade is timestamped by date, uniquely identified by acid and CUSIP/ticker, and includes the dollar principal (either positive or negative) expended on the trade (a buy or sell, respectively). The commission in dollar paid by the account for the trade is also recorded. The daily timestamped nature of the trading data is critical to our analysis because it enables computation of metrics based on intra-month trading decisions and returns, and therefore exposes granular patterns of behavior that would not be visible at fixed-interval monthly or quarterly frequencies. Furthermore, while portfolio holdings data reflect both individual allocation decisions as well as changes in asset values, making it

difficult to disentangle the effects of investor decision from the effects of changing prices, the trade data reflects the decision to allocate in a much more direct manner. For these reason, the availability of trade data differentiates this study from similar studies focusing on retail brokerage account or government stock holdings data.

Key	Description	Format
trade_date	Date of trade	YYYYMMDD (e.g. 20080317)
buy_sell	Indicator of buy or sell	string (e.g. 'B', 'S')
principal	Principal amount traded	double
quantity	Units of asset traded	integer
tcommission	Trade commission	double
cusip_nr	Security identification number registered with the US SEC	9-digit alpha-numeric (e.g. 17275R102)
ticker_symbol	Ticker symbol	string (e.g. AAPL)
product_grplvl1	Top level security type identifier	string (e.g. 'EQUITY')
product_grplvl2	Mid level security type identifier	string (e.g. 'EQUITY')
product_grplvl3	Bottom level security type identifier	string (e.g. 'EQUITY')
acid_key	Unique account identification number	integer (e.g. 9374629673)

Table 6.14: Summary of the data fields in the trades datafile.

6.8.3 Relationship between household, accounts and customers

An investing account can be co-owned by multiple customers. The brokerage firm has associated a group of accounts into a household based on the relationships between the customers. An investing account can only belong to one household whereas the map between investing accounts and customers can be many-to-many. Figure 6-12 illustrates the relationship between the accounts and customers for one of the households.

6.8.4 Demographic data

The demographic files record the personal information on the account application forms of the accounts selected by the random sampling procedure, and can be merged with the historical account data contained in the position and trade files using the anonymized key. Demographic fields include age, income, profession, investment knowledge ('knowledge'), investment experience ('experience'), and marital status. Knowledge and experience are survey questions included with the other components of the application questionnaire, and can receive values of 'Excellent', 'Good', 'Limited', 'None' or 'Decline to report'. These fields reflect the account holder's self-reported view of his or her familiarity with personal finance

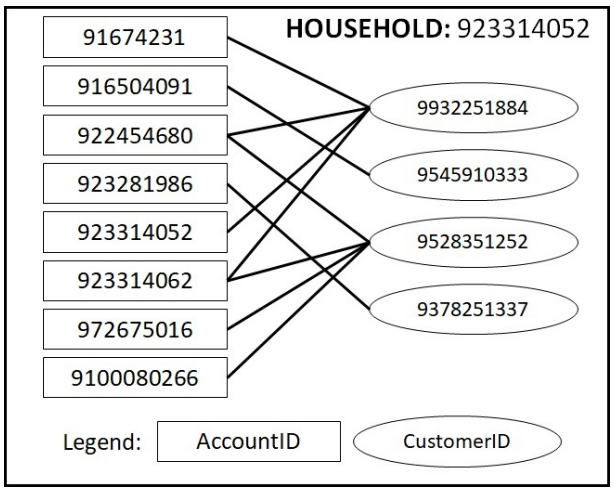


Figure 6-12: A graphical representation of how the households, accounts and customers are related.

and financial decision-making, and therefore they offer a novel way to measure the behavior and performance of investors as a function of their financial sophistication.

Key	Description	Format
month_last_record	Month of last record	integer (YYYYMM)
cust_age	Age	integer
dmsa_martl_stat_cd	Marital status	string
cust_depndt_qy	Number of dependents	integer
ps_gndr_cd	Gender	char
dmsa_curr_occup_tx	Occupation group	string
invst_knldg_cd	Investment knowledge	string
invst_exprc_cd	Investment experience	string
acctid_key	Account ID	integer

Table 6.15: Summary of the data fields in the demographic attributes datafile.

6.8.5 Computing the demographic distribution

The computation of the distribution of demographic features in our dataset is complicated by the fact that a household can consist of multiple customers. Furthermore, some customers in a household can be associated with more accounts than others. In Figure 10, customer 9932251884 is associated with four accounts, while customer 9378251337 is only associated with one account. One can conceptualize that the former customer is more 'influential,' and should be assigned a higher weight.

There are many ways to aggregate this information. The typical method used by the brokerage firm is to consider either the minimum or the maximum of all the customers for a single variable. For example, it will use the maximum age of all the customers in a household when analyzing the age of a household, or consider the highest level of investing experience declared by all the customers in a household to be the household's investing experience. While this is useful for marketing purposes, where one is only interested in finding a target audience (e.g. if the household has someone who needs retirement products), it does not suffice for our study. Furthermore, this method will fail when one attempts to apply it to unordered information, such as occupational groups.

We choose to take into account the portfolio weights of each customer to analyze the demographic distribution of our dataset. To do so, we will first compute the weight of every account based on their average portfolio value over its lifetime. Let the portfolio weight of account i be p_i. For every account, we assume an equal weight between all its registered customers. Denote the set of customers in account i by c_i. Thus, the effective weight of customer j will then be $\sum_{\forall i, j \in c_i} \frac{p_i}{|c_i|}$.

We demonstrate an example of the computational process using Figure 6-13. There, we have 3 investing accounts (in rectangles) and 3 customers (in ovals). First, we compute the average portfolio values across the entire time horizon to find that the portfolio weights of the accounts are $\frac{1}{3}$, $\frac{1}{2}$, $\frac{1}{6}$, respectively. For each account, we then assign a weight from the account to the customers on an equal basis. The results of this step are in green. Finally, for each customer, we can compute the total weight. For customer 9932251884, the overall weight is then $\frac{1}{2} \times \frac{1}{2} + \frac{1}{6} \times 1 = 0.417$.

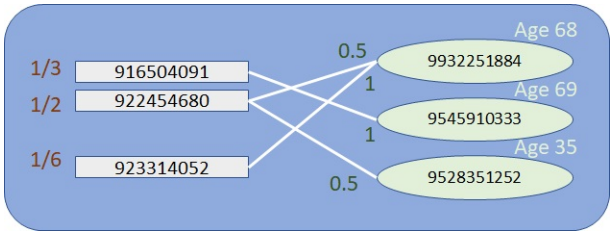


Figure 6-13: An example of how the demographic weights are computed. The rectangles represent investing accounts while the ovals represent customers. The numbers in red are the portfolio weights for each investing account, while the green numbers are the weights to a customer from an investing account.

We also attempted a method in which all the customers in a household were assigned equal weights. While the resulting numbers differed slightly, the conclusions drawn were similar.

6.8.6 Changing the parameters for the identification of panic sales

Our method of determining a panic sale requires us to define two parameters, p_1 and p_2, the monthly portfolio decline and the monthly portfolio net sell, which we set to 0.9 and 0.5, respectively. We conduct additional runs with different parameter pairs to determine how they affect the identification of panic selling. As the amount of computation required is immense, however, costing more than 5,000 CPU-hours per run, we performed only 2 additional runs with the parameter settings shown in Table 6.16. We did not vary p_3 and p_4, the portfolio rebound and the cumulative net buy, as they do not affect the identification of panic sales.

Run	p_1	p_2	p_3	p_4
1	0.9	0.5	0.5	0.5
2	0.5	0.25	0.5	0.5
3	0.25	0.1	0.5	0.5

Table 6.16: Summary of the parameters used in the various run

Figure 6-14 shows the results of our additional runs versus our baseline. As expected, decreasing the thresholds will increase the number of panic sales being captured. While we still observe the major spikes in reaction to major events remain across all runs, lowering the thresholds also amplifies 'noise' in our data.

6.8.7 Explanation of machine learning models

Issue of imbalanced data

One of the biggest issues encountered in training our machine learning models is the extremely imbalanced dataset. Given that the negative class comprises of 99.87% of all data points, a naive classifier that always predicts '0' will easily achieve an accuracy of 99.87%.

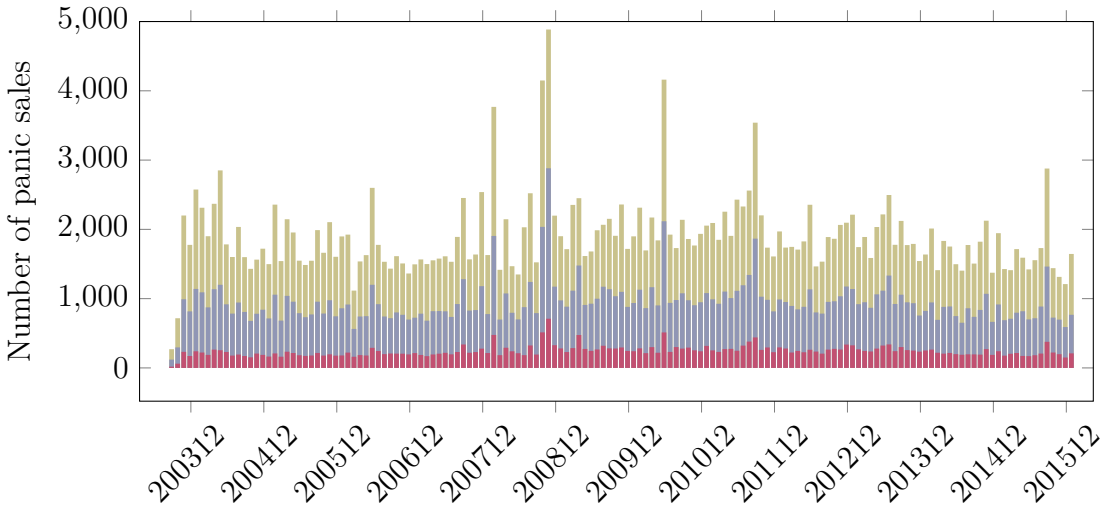


Figure 6-14: The number of panic sales over time for different parameter sweeps. The red, blue and gold bars represent the results for the parameter sets $\{0.9, 0.5, 0.5, 0.5\}$, $\{0.5, 0.25, 0.5, 0.5\}$ and $\{0.25, 0.1, 0.5, 0.5\}$ respectively.

Naively training the models based on the usual cross-entropy minimization will lead to this outcome.

To mitigate this problem, we oversampled the underrepresented class, which we achieved by creating training batches with equal weights. We also considered using SMOTE [15], but we found that interpolating variables generated nonsensical data points; our data was constructed in such a way that there are too many constraints that have to be fulfilled for this method to be applicable.

Metrics for evaluating models

As discussed, accuracy over the entire test set is not a valid measure for imbalanced data. Instead, we evaluated our models on the accuracy of both positive-labelled and negative-labelled data points to get a more useful idea of their real world performances.

In addition, we reported two other metrics that characterize the performance of machine learning models: the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPR).

We define the following measures:

$$\text{Sensitivity} = \text{Recall} = \frac{\#\text{True Positives}}{\#\text{True Positives} + \#\text{False Negatives}} \tag{6.2}$$

$$\text{Specificity} = \frac{\#\text{True Negatives}}{\#\text{True Negatives} + \#\text{False Positives}} \tag{6.3}$$

$$\text{Precision} = \frac{\#\text{True Positives}}{\#\text{True Positives} + \#\text{False Positives}} \tag{6.4}$$

The receiver operating characteristic (ROC) curve is the created by plotting the true positive rate of a classifier, also known as its 'sensitivity' or 'recall', against its false positive rate, or 1−specificity, at different thresholds. A naive classifier will have a ROC profile that is a diagonal from (0,0) to (1,1). In this case, the AUROC of the naive classifier will be 0.5. On the other hand, a perfect classifier will have an AUROC of 1. Mathematically, the AUROC is the probability that the score of a randomly selected positive example is higher than the score of a randomly selected negative example.

The ROC is not useful if one is interested in the rate that the models produce false alarms. In such cases, the precision-recall (PR) curve is more useful. A naive classifier will have a precision that is equal to the proportion of positive data points in the entire sample for all thresholds.

Part IV

Conclusion

6.9 Final Words

As the old adage goes: you can't manage what you don't measure. This thesis focuses on the using computational methods to collect and process data, investigate highly non-linear relationships and simulate complex scenarios in order to forecast possible outcomes. Through the chapters, we present computational methods to quantify risk in drug development programs, address current challenges in health economics and investigate and predict rare events in finance. We hope that the methods introduced will allow the respective stakeholders to make timely and informed decisions.