Risk and Uncertainty in Healthcare Finance, Investment Management, and Asset Pricing

by

Zied Ben Chaouch

B.S. (Hons.), McGill University (2014)
S.M., Massachusetts Institute of Technology (2016)

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

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2022

Author .....
Department of Electrical Engineering and Computer Science
August 26, 2022

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

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

Risk and Uncertainty in Healthcare Finance, Investment Management, and Asset Pricing

by

Zied Ben Chaouch

Submitted to the Department of Electrical Engineering and Computer Science
on August 26, 2022, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

Measuring and managing risk and uncertainty has been a ongoing challenge for academics and practitioners in finance and economics. At its core, the challenge requires to understand both the randomness in the underlying process as well as the way humans respond to it when making decisions. This thesis focuses on these challenges from three different perspectives.

First, from the point of view of healthcare finance, this thesis addresses key questions that occur during the drug development process and the drug deployment process. We begin by developing a systematic, quantitative, transparent, and reproducible framework that incorporates the patient's risk and uncertainty preferences into the regulatory and decision-making process, both theoretically and empirically, to improve the design and regulation of clinical trials. Then, we consider both the development and deployment of vaccines for emerging infectious diseases. Using the COVID-19 pandemic as a case study, we develop a quantitative method to simulate and evaluate various vaccine allocation strategies when the supply of vaccines is subject to stochastic shocks. We conclude this part of the thesis by proposing and analyzing the viability of a portfolio approach aimed to improve the risk/return trade off of investment when developing mRNA vaccine candidates for 11 emerging infectious diseases. Vaccine development is not only challenging due to the high scientific risk when developing a compound, but also due to the uncertainty in the occurrence of epidemics, leading to a lack of financial incentives for pharmaceutical firms to invest in vaccine research and development.

The second part of the thesis dives into the field of empirical asset pricing. If multi-factor models are routinely used in by finance academics and practitioners to understand and quantify the risk exposures of an asset, more than 150 factors have been proposed in the asset pricing literature, constituting a "factor zoo". This thesis develops linear and nonlinear techniques to construct latent factors from a set of 150 well-known risk factors using different types of autoencoders. We then compare the performance of these latent models to classical multi-factor models on various test assets.

The final part of the thesis explores an investor's risk profile and behavioral biases in the investment management landscape and aims to understand how different market participants and different types of individuals compare along the dimensions of risk aversion and investment style. To this end, we survey a large pool of individual investors, financial advisors, and institutional investors over three years about their investment decisions under various historical and hypothetical scenarios.

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

Acknowledgments

First, I would like to thank my advisor, Professor Andrew W. Lo, for his guidance and support at MIT and at the Laboratory for Financial Engineering (LFE). It has been a true privilege to meet regularly with Professor Lo and an honor to be a Teaching Assistant for his classes. His unique style in teaching, both in person and online, and in research have been a constant source of inspiration over the years. I am very grateful to have had the opportunity to explore a variety of topics spanning multiple fields, assist Professor Lo in his classes at Sloan and on edX, and be part of the LFE's friendly and vibrant community.

I also want to thank Professor Martha Gray and Dr. Shomesh Chaudhuri for being part of my thesis committee and providing very insightful feedback. It has been a privilege to collaborate with Sho on multiple projects throughout my PhD.

The list would of course not be complete without thanking Jayna Cummings, Crystal Myler and Kate Lyons for their constant support. Jayna is an outstanding assistant director, she coordinates every project at the LFE, ensures that the students have access to the appropriate datasets, and carefully reviews every research paper we draft.

The PhD is a long journey, and I have been very fortunate to be surrounded by amazing friends and collaborators. Thank you Manish, Chi Heem, Kien Wei, Jianmeng, and Amy for fostering a fun, lively, and collaborative environment at the 1 Broadway office (and outside the office!). Thank you Eren, Govind, Hsin Yu, Igor, and Tadayuki for keeping me sane and for joining me on many adventures on Boston's restaurant scene. And thank you to my friends from the MIT tango community, Adina, Alican, Berk & Amina, Dennis & Natasha, Gozde, Wendy, and most importantly Yu Wen, as well as Adam, Guillermo, Pamela & Steve, Simonida, and Vicky—D'Arienzo and Biagi also deserve a special mention!

I also want to send my deepest gratitude to Madame Lecocq and Monsieur Savino for believing in me when I was still very young, Monsieur Courbot for giving me a passion for mathematics, Madame Jalabert and Monsieur Kobbi for giving me a passion for physics, and Madame Ben Ammar for giving me a passion for writing and French literature.

Finally, I want to thank my family for their constant love and support. First and foremost, my parents Sonia and Sami: thank you for always being here for me, for being a source of inspiration and encouragement, and for bringing so much joy into my life. I owe you everything. Thank you Sarah for your love and support throughout my PhD. And thank you Ferida and Maherzia, my favorite grandmothers, as well as my late grandfathers Boubaker and Slah whom I miss dearly. And finally thank you to our dear friend Hichem.

Contents

List of Figures

List of Tables

Glossary

AE	autoencoder.
APT	arbitrage pricing theory.
BDA	Bayesian Decision Analysis.
CAE	clustered autoencoder.
CAPM	capital asset pricing model.
CDC	United States Centers for Disease Control and Prevention.
CDRH	Center for Devices and Radiological Health.
CEPI	Coalition for Epidemic Preparedness Innovations.
CISDM	Center for International Securities and Derivatives Markets.
CRSP	Center for Research in Security Prices.
DAE	deep autoencoder.
DBS	deep brain stimulation.
DCE	Discrete Choice Experiment.
EID	Emergent Infectious Disease.
ESKD	end-stage kidney disease.
EUA	emergency use authorization.
EVI	Expected Value of Information.
EVSI	Expected Value of Sample Information.
FDA	United States Food and Drug Administration.
FF_3	Fama-French Three-Factor Model.
FF_5	Fama-French Five-Factor Model.

FIFO	First-In-First-Out.
GAN	generative adversarial network.
GMM	generalized method of moments.
GRS	Gibbons-Ross-Shanken.
HCT	Human Challenge Trial.
I.I.D	independent and identically distributed.
ICG	International Coordinating Group.
KHI	Kidney Health Initiative.
KRT	kidney replacement therapy.
LASSO	least absolute shrinkage and selection operator.
MAR	maximum acceptable risk.
MDIC	Medical Device Innovation Consortium.
mRNA	messenger RNA.
NBTS	National Brain Tumor Society.
NDA	New Drug Application.
OLS	ordinary least squares.
PCA	principal components analysis.
PD	Parkinson's disease.
PoS	probability of success.
PPI	Patient Preferences Information.
q5	Hou-Xue-Zhang q-Factor Model.
R&D	Research & Development.
RCAE	recursive clustered autoencoder.
RCT	Randomized Controlled Trial.
ReLU	rectified linear unit.
RRT	renal replacement therapy.
SAE	split autoencoder.

U.S. United States of America.

WHO World Health Organization.

WRDS Wharton Research Data Services.

Chapter 1

Introduction

In 1921, the economist Frank Knight, from the University of Chicago, published a book *Risk, Uncertainty, and Profit* [143] based on his PhD dissertation in which he makes the distinction between two types of randomness: *risk* and *uncertainty*. According to Knight's taxonomy, risk represents the type of randomness which can be quantified, that is the type of randomness that can be fully represented by a known probability distribution. On the other hand, uncertainty represents the type of randomness whose probability distribution is unknown, an "unquantifyable" uncertainty, often referred to as the "unknown unknowns".

This distinction motivated Lo and Mueller [159] to create a taxonomy of uncertainty, distinguishing 5 levels of randomness. The first level represents *complete certainty* i.e., a deterministic situation in which everything is known and predictable. The second level represents *risk without uncertainty*, which is what Knight calls *risk*. The third level of randomness is *fully reducible uncertainty*, which describes situations in which some probabilities are unknown but gathering enough data can help us reduce the randomness to Level 2—for example through the use of statistical analysis, with the law of large numbers or the central limit theorem. The fourth level is called *partially reducible uncertainty*: this level describes situations in which we cannot fully determine the data generating process even with classical frequentist statistics as there are uncertainties in the model itself. This level is what Knight calls *uncertainty*. Although frequentist statistics is of little help here, Bayesian statistics can provide a useful framework to navigate the realm of Level 4 uncertainty. Starting at Level 5, we reach *irreducible uncertainty*, in which all attempts at understanding are futile.

This distinction becomes very important when we try to understand human behavior. While we dislike risk, we abhor uncertainty. More practically, if many investors are willing to take on risks, most investors would shed away from uncertainty. Ellsberg's paradox [70], described in [156], can help illustrate this distinction in a very practical context. Suppose that I present you with a well-mixed urn that contains 100 balls, 50 of which are red and 50 are black. You are asked to select a color (either red or black), and, blindfolded, I will randomly draw a ball from the urn. If the color of the ball selected matches the color you have chosen, you win $\$10,000$, otherwise you earn nothing. How much are you willing to pay (at most) to play this game once?

The expected value of this game is $0.5 \cdot \$10,000 + 0.5 \cdot \$0 = \$5,000$, suggesting that a fair price to pay would be $\$5,000$ (for a risk-neutral participant). However, many people would not go as high as $\$5,000$ to play this game and will prefer to bid slightly lower: this is because the game is risky (and can be played only once), so participants with a positive risk-aversion would require a *risk premium* (i.e., a discounted fee) to play the game .

Now, let me present you with a second urn. This urn also contains 100 balls, however I do not tell you how many balls out of the 100 are red and how many are black. How much are you willing to pay (at most) to play this game once? Still $\$5,000$? If this scenario is making you feel uncomfortable, do not be alarmed: you are entering the realm of Level 4 uncertainty. While you knew that the odds were 50-50 in the first scenario, you are now facing unknown odds in the second scenario, and playing this game only once will not allow you to estimate the unknown odds, leading to *partially reducible uncertainty*. Mathematically, it can be shown that both scenarios are equivalent and would give you an expected payoff of $\$5,000$. Yet, most participants bid a much lower amount to play the second game: participants need to face their own risk-aversion and also uncertainty-aversion in the second game, lowering the bid further away from the risk-neutral $\$5,000$ bid.

Measuring and managing risk and uncertainty has been a ongoing challenge for academics and practitioners in finance and economics as it merges human behavior and decision-making with probabilistic and statistical reasoning to quantify and understand the randomness in the underlying process. This thesis focuses on these challenges from three different perspectives: first in healthcare finance—during the drug development process and the drug deployment process, then in empirical asset pricing, and finally in investment management. While these fields may seem very different at first sight, they all rely at their core on the concepts of benefit, risk, and uncertainty: BDA explicitly trades off benefit, risk, and uncertainty in its formulation, empirical asset pricing aims to understand and quantify the nature of the different risk factors stocks are exposed to, and investment management is inevitably tied to an investor's risk and uncertainty profile and behavioral biases.

1.1 Organization of the Thesis

The thesis is divided into three major sections which we describe below:

Part I: Managing risks & uncertainty in healthcare finance

The first part of this thesis is devoted to the field of healthcare finance and the drug development process. We approach this section from two different angles.

Regulatory and decision-making process

The first direction focuses on the regulatory and decision-making process. In the United States of America (U.S.), a treatment (e.g., a drug or a medical device) often

needs to undergo three stages of clinical trials (commonly referred to as Phase I, Phase II, and Phase III) and, if successful, is eventually submitted to the United States Food and Drug Administration (FDA) for approval. A treatment becomes available to patients if it is granted FDA approval.

While there are multiple factors taken into account by the FDA when deciding whether or not to approve a treatment, the efficacy and safety of a treatment is often at the forefront of the debate. Efficacy and safety are commonly tested using traditional two-arms RCTs. The null hypothesis would consist in having an ineffective (or unsafe) treatment, and in this case the treatment group would be equivalent or worse-off than the control group. On the other hand, the alternative hypothesis would consist in having an effective (or safe) treatment, and in this case the treatment group would perform better than the control group. Traditionally, the FDA would reject the null hypothesis at a 2.5% or 5% significance level. While this threshold is convenient and has often been used for historical reasons, it is difficult to argue that the same threshold should be used in all applications. In fact, the appropriate threshold to use is, almost by definition, very closely linked to the amount of risk the stakeholders are willing to undertake. In the context of clinical trials, this means that patients may be willing, under some circumstances, to bear additional risks to gain access to a certain treatment rather than miss-out on the potential benefits this treatment can provide. Conversely, when alternative treatments are available, patients may be much less risk-tolerant and require a lower significance level threshold to assess the efficacy of the treatment under consideration.

In particular, BDA aims to incorporate patient preferences into the regulatory and decision-making process by trading off the impact of a type-I error (the event of incorrectly approving an ineffective and potentially harmful treatment) with a type-II error (the event of incorrectly rejecting an effective treatment) in a systematic, transparent, and reproducible way. This approach also helps improve the design of clinical trials by optimally selecting the significance level and sample size of a trial in order to reduce the duration of clinical trials and provide patients with a faster access to the treatment if it is found to be effective and/or safe.

In collaboration with regulators, we apply the BDA methodology to medical devices for patients with kidney failure [40] and Parkinson's disease [45]. An application to heart failure can be found in [44]. While the latter two estimate the optimal significance level for efficacy endpoints, the former focuses exclusively on estimating the optimal significance level for safety endpoints.

From a methodological perspective, we present a generalized version of the BDA procedure in [38]. This work extends the BDA approach in three distinct ways.

1. First, we generalize the discrete setting of the traditional BDA approach—in which the treatment effect of the treatment can only take two values (corresponding to scenarios in which the treatment is effective/safe or ineffective/unsafe)—by assuming a continuous prior distribution on the treatment effect.

2. Second, while the classical BDA carefully incorporates the risk preferences of the patients into the design, it fails to include their view on potential uncertainties in the treatment's efficacy. The objective of the extended BDA is twofold: we

use a mean-variance optimization framework to jointly minimize the expected loss to patients (as in the classical BDA)—while accounting for the uncertainty in the treatment's efficacy—and the variance of this loss. The variance term is important as it measures the sensitivity of the loss function to the uncertainty in the treatment's efficacy.

3. Third, we show how to obtain standard error estimates and confidence intervals on the optimal significance level and sample size calculated by the BDA using bootstrap estimators.

Through the extended BDA, we propose a unified framework which is consistent with the usual assumptions of a RCT, and incorporates both the risk and uncertainty preferences of patients obtained from a patient preference survey. As with the traditional BDA, this method is systematic, transparent, and repeatable.

Vaccine development and deployment

The second direction focuses on the development and deployment of vaccines for infectious diseases.

With more than 44.7 million infections in the U.S. and 219 million worldwide, and a death toll over $721,000$ in the United States and 4.55 million worldwide, the COVID-19 pandemic has profoundly altered the research agenda of the scientific community as a whole, launching an unprecedented race against the clock to develop a cure or a vaccine for the disease. Mass vaccination has become the critical pathway to alleviate the impact of the disease, as is apparent with the success of Israel's mass vaccination campaign. However, producing and distributing the vaccines has become a new challenge for manufacturers. Despite promising results regarding the ability to store the Pfizer-BioNTech vaccine in standard freezers over periods of two weeks rather than the initial storage constraint at $-80°$C, vaccine shortages and appointment cancellations have followed factory shutdowns, production mix-ups, delays in shipment, and power outages. Optimizing the allocation of vaccines has become crucial not only due to the limited supply of vaccines, but also due to the fact that Pfizer-BioNTech and Moderna vaccines need to be administered twice for each individual, over a recommended time interval of 3 or 4 weeks, respectively. Although supply constraints are important in the United States, they are even more binding in other regions such as Canada, Europe, Africa, Latin America, and India.

An important debate has arisen regarding the advantages of delaying the second dose to provide more first doses to susceptible individuals. While doses were held back under the Trump administration in order to guarantee a second dose to individuals who have received their first dose, the Biden administration has pledged to reverse this policy and release all available doses. Other countries, such as the United Kingdom and Canada, have adopted the policy of delaying the second dose up to three months, and Singapore is currently considering delaying the second dose up to 12 weeks. However, as Texas, Washington State, and Michigan experienced in mid-February 2021, releasing too many doses for first-time users could lead to delays for individuals eligible to receive their second dose (a "second-shot crunch").

In [17], we forecast the effect of various vaccine allocation strategies on the cumulative number of infections and deaths in the United States to quantify the impact of prioritizing first doses versus second doses. In particular, we extend the DELPHI model to account for vaccines, and use a simple model of shocks to the number of vaccines supplied to account for distributional constraints. While our analysis focuses on the United States, our recommendations can be generalized to other countries and especially those where the supply of vaccines is heavily limited. The framework provided here can be reused in the event of a future pandemic to improve the allocation of vaccines and reduce the number of infections and deaths.

From a vaccine development perspective, the human, social, and financial costs of the COVID-19 pandemic helped raise awareness and urge researchers and politicians to prepare for the next pandemic by proactively engaging in the Research & Development (R&D) of novel vaccines against Emergent Infectious Diseases (EIDs). A notable example of such an effort is the Coalition for Epidemic Preparedness Innovations (CEPI), which has added 14 vaccine candidates targeting COVID-19 as of January 11, 2022 to the portfolio of six other priority EIDs created in 2017.

Vaccine R&D has also undergone a revolution during the pandemic—exemplified by mRNA vaccine technology—which has demonstrated robust safety, high efficacy, and unprecedented speed of clinical development. This technology has the potential to significantly reduce the cost and duration of vaccine R&D, enabling much more rapid responses to future EIDs. It is also particularly suited to the portfolio-based approach of CEPI, since different mRNA vaccine candidates may share the same resources and facilities of preclinical animal studies, clinical testing, and post-approval manufacturing and delivery. An important challenge to the portfolio-based model of mRNA vaccine development is the lack of sufficient and sustainable funding to support the vaccine development pipeline over the extended period (typically over multiple years) from preclinical research to FDA approval, an issue known as the "valley of death" in translational biomedical research. Governments, international agencies, and non-governmental organizations have contributed significantly to create a sizeable portfolio of vaccine candidates, but their efforts have nevertheless fallen short. However, the private sector may provide the investment needed to finance the vaccine R&D pipeline, provided that the vaccine portfolio can generate financial returns for its investors.

In [13], we simulate a hypothetical vaccine megafund with a large portfolio of 120 mRNA vaccine candidates targeting a total of 11 diseases and ask whether the risk/return profile of the megafund is attractive to investors. To address the complexity of simulating the vaccine development process, we calibrate the simulation parameters with inputs from domain experts in mRNA technology and an extensive literature review. We illustrate the key factors affecting the financial performance of the vaccine megafund and discuss potential solutions to improve its financial returns to investors.

Part II: Managing risks & uncertainty in empirical asset pricing

The second part of this thesis is devoted to the field of empirical asset pricing [39]. Factor models are routinely used in the financial industry to identify and quantify sources of systematic risk in order to manage the risk of a portfolio of securities or hedge investment positions, or in valuation contexts to estimate the cost of capital of an asset. The capital asset pricing model (CAPM) [200, 153] simply identifies the systematic risk of a stock with its exposure to the market. However, the CAPM fails to explain the cross-section of asset prices and multiple anomalies (sources of systematic risk that are not captured by the CAPM) were documented [121]. More complex factor models were developed such as the Fama-French 3, 4, and 5-factors models (*FF3*, *FF5*, *FF6*) [73, 74, 75], the Carhart 4-factors model [31], or the 4 and 5 q-factors (*q4*, *q5*) by Hou, Xue, and Zhang [120, 119].

A large set of factors have been introduced in the literature to attempt to explain the cross-section of asset prices, however no single model has convincingly been able to capture most of the anomalies. Tools have been proposed to "tame the factor zoo" and compress the 150+ factors proposed in the literature into a parsimonious model. These techniques include the double selection least absolute shrinkage and selection operator (LASSO) approach [85] as well as principal components analysis (PCA) [151]. While the former consists of a model-selection approach, the latter aims to extract a set of "fundamental" latent factors. Our approach is similar to [151] as we aim to construct a set of latent factors that can explain away most of the anomalies in the cross section of asset prices.

In this chapter, we use different types of AEs to compress a set of 150 known factors (from 1973 to 2017) to obtain a parsimonious factor model that can explain the cross-section of asset prices. We then test our model on the 150 known factors, as well as the Fama-French Five-Factor Model (FF5) and the Hou-Xue-Zhang q-Factor Model (q5) factor models. In addition, we test our models on a set of portfolios, including the Fama-French anomaly portfolios and the Hou-Xue-Zhang anomaly portfolios.

Although these fundamental factors are usually difficult to interpret by design, we incorporate economic intuition into the design of some of our AE architectures. In particular, the 150 factors analyzed are not uncorrelated, and some groups of factors can even be highly correlated, potentially leading to an over-representation effect when compressing the 150 factors. To solve this issue, we introduce the CAE: this model first clusters the 150 factors into 40 to 80 groups, and then each node in the first hidden layer of the CAE is only connected to nodes belonging to a single factor cluster. The nodes in this hidden layer are then compressed into a parsimonious model.

Another contribution of our work is that we aim to capture nonlinear relationships between returns and factors. While linear factor models are often a good benchmark, latent factors obtained from machine-learning models tend to capture nonlinearities that linear models would fail to express. To that end, we use nonlinear activation functions and run polynomial regressions (of order 2 which only includes interaction terms) of returns on the latent factors learnt to capture interactions among pairs of

factors. Similarly, we capture asymmetric exposures to positive and negative factor returns by splitting factors into a positive component and a negative component, and then running a linear regression on this augmented set of factors.

Part III: Managing risks & uncertainty in investment management

The final part of this thesis is devoted to the field of investment management [160]. There are three major groups of participants in the investment management industry: individual investors, financial advisors, and institutions. Each of these groups has its own risk preferences and behavioral characteristics it uses in its investment decisions. We study the behavior of these groups by using the results of three comprehensive global surveys between 2015 and 2017, each of them covering over 7,000 individuals, over 2,300 advisors, and over 660 institutional investors.

The breadth of our dataset sets it apart from earlier survey data in the literature. To the best of our knowledge, these surveys are the first to present the same set of questions to three distinct groups of market participants over three consecutive years. This dataset covers over 17 countries in the Individual Investor surveys and over 14 countries in the Financial Advisor surveys every year. This global breadth provides us with insight into investment behavior by country, and allows us to compare survey results across countries. Finally, all our survey subjects have a significant stake in the market: all the surveyed individual investors have a net worth above $200,000, while the financial advisors and the institutional investors are employed in the financial industry. As a result, their answers will generally be more realistic and have greater relevance for modeling investor behavior than the results of surveying students in a laboratory setting, as many other studies have done. Our main goal is to understand how different market participants and different types of individuals compare along the dimensions of risk aversion and asset allocation. To this end, we poll members of these groups about their investment decisions under various historical and hypothetical scenarios.

We obtain two sets of results.

1. The first set of results shows that investors tend to be significantly more risk-averse and mostly extrapolative in their asset allocation, while institutions tend to be significantly less risk-averse and mostly contrarian in their investment decisions, with advisors falling in the middle of the risk aversion scale while also following a contrarian asset allocation strategy.

2. The second set of results focuses on just individual investors—using a clustering algorithm applied to survey responses, we are able to identify five distinct types of investors: *passive investors*, *extrapolators*, *risk avoiders*, *contrarians* and *optimistic investors*. *Extrapolators* tend to decrease allocation in equities following bad market performance, and tend to increase allocation following good returns, extrapolating past trends. *Passive investors* leave their allocation unchanged in either scenario. *Risk avoiders* significantly cut their allocation to equities when

they see large moves in the S&P 500 in either direction. *Contrarians* tend to increase allocation in equities following bad market performance, and tend to decrease allocation following good returns. *Optimistic investors* tend to increase their allocation in either scenario. While the largest cluster of individuals in our dataset corresponds to passive investors, it also contains a significant number of risk avoiders and extrapolators. Evidence for each of these types is found in the literature, although most papers focus only on one type at a time.

In comparison, we find that most financial advisors and institutional investors are contrarian in allocation strategy—that is, they would change equity allocation in the direction opposite to recent returns on the S&P 500. This contrasts with the overall behavior of individual investors, who on average are extrapolators. The differences in the reactions across these three groups of market participants are significant and very large. We note that a few earlier studies have viewed individuals as momentum traders, and institutions as contrarians. However, these studies consider shorter-term horizons than ours, and focus on trading behavior. Our survey asks about asset allocation, a strategic and longer-term investment decision, rather than short-term trading, which potentially could be affected by excessive speculation on the part of individual traders, or by liquidity considerations of institutions. Recent papers focusing on asset allocation of Dutch institutional investors actually concludes that they tend to be contrarians.

Our results have another important implication, one that arises from the differences in response between financial advisors and individual investors. We find that advisors generally advise their clients to change their allocation in the opposite direction of the typical preference of the individual investor. It may be that advisors recognize the excessive tendency of investors toward extrapolation, and try to mitigate this effect by giving "contrarian advice". Also, the proportion of advisors who suggest a significant decrease in equity allocation when seeing large S&P 500 moves is much smaller than the proportion of individual investors who would implement such a change. As a result, advisors may also provide the significant benefit of ensuring their investors stay invested in the markets despite periods of high volatility, and hence earn higher returns in the long run. Overall, our findings suggest that financial advisors are of direct benefit to most individual investors.

Finally, we compare risk aversion across the three groups, as well as within investor demographic categories. Individual investors are significantly more risk-averse than financial advisors, who are in turn more risk-averse than institutional investors. We find that individual risk aversion increases with age, which is consistent with previous literature linking risk aversion to age, wealth, and education.

1.2 List of Publications

Publications cited in the thesis

1. Chapter 2 includes: Ben Chaouch et al., *Use of Bayesian decision analysis in the design of clinical trials: Application to kidney failure devices*. Manuscript under preparation, 2022. [40]

2. Chapter 3 includes: Chaudhuri et al., *Use of Bayesian decision analysis to maximize value in patient-centered randomized clinical trials in Parkinson's disease*. Manuscript under Review (Journal of Biopharmaceutical Statistics), 2021. [45]

3. Chapter 4 includes: Ben Chaouch, Chaudhuri, and Lo. *Bayesian decision analysis under risk and uncertainty: A tale of two exposures*. Manuscript under Review (Statistics in Medicine), 2022. [38]

4. Chapter 5 includes: Ben Chaouch, Lo, and Wong. *Should we allocate more covid-19 vaccine doses to non-vaccinated individuals?*, **PLOS Global Public Health**, 2(7):1–17, 07 2022. [17]

5. Chapter 6 includes: Barberio et al., *Accelerating Vaccine Innovation for Emerging Infectious Diseases via Parallel Discovery*. In **Entrepreneurship and Innovation Policy and the Economy**, volume 2 of NBER Chapters. National Bureau of Economic Research, Inc, July 2022. [13]

6. Chapter 7 includes: Ben Chaouch, Lo, Singh, and Xiong. *Chimeras in the factor zoo*. Manuscript in Progress, 2022. [39]

7. Chapter 8 includes: Lo, Remorov, and Ben Chaouch. *Measuring risk preferences and asset-allocation decisions: A global survey analysis*. **Journal of Investment Management**, 18(3):5–50, 2020. [160]

Additional Publications

1. Abouarab et al., *Financing repurposed drugs for rare diseases: A case study of Unravel Biosciences*. Manuscript under Review (Orphanet), 2022. [2]

2. Chaudhuri et al., *Patient-centered clinical trial design for heart failure devices via bayesian decision analysis*. Manuscript submitted for publication (JACC: Heart Failure), 2022. [44]

3. Xu, Cho, Ben Chaouch, and Lo. *Incorporating patient preferences and burden-of-disease in evaluating als drug candidate AMX0035: A bayesian decision analysis perspective*. Manuscript under Review (ALSFD), 2022. [224]

Part I

Managing Risks & Uncertainty in Healthcare Finance

Chapter 2

An Introduction to Bayesian Decision Analysis

The first part of this thesis focuses on applications of risk management to healthcare finance. Chapters 2, 3, and 4 are devoted to the topic of Bayesian Decision Analysis (BDA). We introduce the concept of BDA in Chapter 2 with an application to kidney failure devices. Then, in Chapter 3 we briefly discuss an application to Parkinson's disease devices. Finally, we generalize the BDA approach in Chapter 4 by assuming a continuous distribution for the treatment effect and accounting for the patient's uncertainty aversion through a mean-variance optimization formulation. The generalized BDA developed in Chapter 4 uses the Parkinson's disease devices analyzed in Chapter 3 as a case study to emphasize the differences with the classical BDA approach used in Chapters 2 and 3.

2.1 Motivation

2.1.1 Randomized controlled trials

RCTs are widely used to assess the effectiveness and/or safety of a new drug or medical device. Typically, a treatment is deemed to provide adequate statistical evidence of effectiveness and/or safety if the clinical trial produces a statistical significance which is lower than a certain pre-defined threshold. Following the seminal work of Sir Ronald Fisher [87], it has been customary to use a fixed significance level threshold of 2.5% for one-sided hypothesis tests and 5% for two-sided hypothesis tests.

Using the same threshold irrespective of the application, while convenient, is a severe limitation. In fact, institutions such as the FDA do not rely solely on p-values to assess the effectiveness of a new treatment and instead prefer to carefully consider the potential benefits and risks of a treatment in a broader context. For example, we would expect BDA to recommend a more conservative significance level for less severe diseases for which there are existing approved treatments (i.e., lower than the traditional 2.5% threshold), and a less conservative threshold (i.e., higher than 2.5%) for more severe diseases with few or no treatment options. Making this tradeoff

between benefit and risk is equivalent to balancing the expected loss to patients due to incorrectly approving an unsafe device with the loss due to incorrectly rejecting a safe device, commonly referred to as type I (α) and type II (β) errors, respectively.

The FDA has been very vocal about its goal to better incorporate patient preferences into the regulatory and decision-making process. For example, the FDA has published a series of guidance documents to highlight the importance of Patient Preferences Information (PPI), suggest factors to consider when incorporating PPI into benefit/risk tradeoff decisions, and outline settings for which it is appropriate to use PPI in the regulatory and decision-making process [78, 80, 82]. However, the FDA does not yet endorse any quantitative method to inform benefit/risk tradeoff decisions in a systematic way.

2.1.2 Incorporating patient preferences into the regulatory and decision-making process

While other techniques have been developed, such as the Expected Value of Information (EVI) framework [182, 4], BDA has been proposed [126, 155, 47, 168, 43] as a way to bring more transparency into the choice of the appropriate significance level threshold. BDA was then successfully applied to weight-loss devices [46], heart-failure devices [44], and Parkinson's disease devices [45], in close collaboration with regulators. In fact, BDA is a systematic, transparent, and repeatable process aimed at incorporating patient preferences into the design of clinical trials to optimize the tradeoff between type I and type II error rates and minimize the expected loss in utility to patients for a specific treatment. Unlike the EVI methodology which relies on a fully Bayesian formulation and focuses on posterior credible intervals and Bayes' risk, the BDA framework is closer to a traditional frequentist hypothesis testing approach, preferred by researchers and regulators, and is mathematically related to the EVI as shown in [45]. Under the BDA framework, optimal clinical trial designs are selected by maximizing the expected gain in utility to patients, or equivalently by minimizing the expected loss in utility to patients.

2.1.3 Application to kidney replacement therapy (KRT) devices

In partnership with the Center for Devices and Radiological Health (CDRH) at the FDA and the American Society of Nephrology, the Kidney Health Initiative (KHI) has recently developed a survey [88] to elicit patient preference information regarding potential novel renal replacement therapy (RRT) from a representative population of patients suffering from end-stage kidney disease (ESKD). With over 700,000 individuals affected, high death rates, and a reduced quality of life [202] and with the support of patients advocacy leaders from the American Association of Kidney Patients [55], KHI has committed with the CDRH to collect patient-focused outcomes which would help drive RRT innovations into directions that improve the patient's quality and length of life [124, 209]. The [88] survey measures how patients evalu-

ate the potential benefits and risks of wearable dialysis devices and the differences in patient risk preferences based on their current modality of dialysis treatment and other characteristics. By quantifying the maximum level of risk patients are willing to accept and the maximum amount of time they are willing to wait in exchange for the benefits provided by a hypothetical wearable device, this survey can help inform future clinical trial designs of dialysis devices.

2.1.4 Summary of contributions

In this study, we apply the BDA methodology to the PPI collected through the [88] survey in order to determine the optimal significance level and sample size for wearable dialysis device clinical trials using two safety endpoints for a hypothetical dialysis device: the risk of serious bleeding and the risk of serious infection.

Overall, we find BDA-optimal type I error rates significantly lower than the classical 2.5% threshold used in practice with rates of 1.2% for the risk of serious bleeding and less than 0.1% for the risk of serious infection. This suggests that patients are not willing to bear either type of additional risk presented by the hypothetical device in exchange for the probable benefits described in the survey instrument. Considering patient subgroup populations individually, we find that the patients on peritoneal dialysis tend to be more risk-tolerant and would be comfortable with a significance level threshold of 3.8% and 1.3% for the risks of serious bleeding and serious infection respectively, which is respectively 1.5 times higher and 50% lower than the 2.5% one-sided threshold conventionally used. Relative to other subgroups, these patients perceive the potential benefits of the hypothetical device more strongly than the risks it presents and require less statistical evidence to adopt the device as an alternative to in-center dialysis. On the other side of the spectrum, home hemodialysis patients tend to be more risk-averse and would require a significance level threshold of 0.3% for the risk of serious bleeding and a threshold lower than 0.1% for the risk of serious infection, which is almost 10 times lower than a typical 2.5% significance level. For these patients, more statistical evidence is needed to justify transitioning from an in-center dialysis to the hypothetical device.

The remainder of the chapter is structured as follows. We develop in Section 2.2 the patient-centered utility-based model for the hypothetical RRT device considered and provide a BDA formulation to optimize the type-I and type-II error rate thresholds. We apply the method to the RRT device in Section 2.3, present the optimal clinical trial design for different patient subgroups under consideration, and discuss the significance of these results as well as their limitations. We conclude in Section 2.4 and provide some additional sensitivity analyses in Appendix A to highlight the robustness of the results to our assumptions.

2.2 The Bayesian Decision Analysis Methodology

In this section, we present the classical BDA approach and use it to determine the optimal significance level and the optimal sample size for the RCT.

2.2.1 Estimating type I and type I losses in utility

We define the net risk of the device ($NetRisk_{Device}$) as the hypothetical additional device-associated risk incurred by the patient and the maximum acceptable risk (MAR_{Device}) as the additional device-associated risk the patient is willing to bear in exchange for the probable benefits of the device described in the survey instrument. maximum acceptable risks (MARs) are estimated using the patient preference survey through interval regression models [197].

The patient's type I loss in utility i.e., the decrease in utility incurred by the patient if an unsafe device is approved, is defined as

$$L_0 = NetRisk_{Device}. \tag{2.1}$$

Similarly, the patient's type II loss in utility i.e., the decrease in utility incurred by the patient if a safe device is not approved, is defined as

$$L_1 = MAR_{Device} - NetRisk_{Device}. \tag{2.2}$$

The assumed risks of the device are outlined in Table 2.1. Both the device and baseline risks (and their respective standard errors) were calibrated by experts at KHI using the best available current estimates [197]. The robustness of our results to changes in these assumptions is discussed in Appendix A.

Type of Risk	Device Risk	Baseline Risk	Device Net Risk	Safety Margin (δ_0)
Risk of Serious Bleeding	8% (2%)	0.1% (1%)	7.9% (2.2%)	3%
Risk of Serious Infection	25% (5%)	6% (2%)	19% (5.4%)	6%

Table 2.1: Assumptions made on the net risk of the device and the safety margin used in the analysis. Standard errors are reported in the parentheses.

The estimated mean MARs of the patient preference survey were obtained from [197] and are summarized in Table 2.2. It is interesting to note that all patient subgroups (with the exception of Hispanic patients) have mean MARs for the risk of serious bleeding that are above the net risk of the device, suggesting that, on average, patients are willing to bear a higher level of risk of serious bleeding in exchange for the probable benefits of the device described in the survey instrument. Conversely, only peritoneal dialysis patients have a mean MAR for the risk of serious infection that is above the net risk of the device, suggesting that only the peritoneal dialysis patients are willing to bear the level of risk of serious infection, while the patients in other dialysis treatments are not willing to accept the high level of risk of serious infection.

Table 2.2: Mean maximum acceptable risk, mean temporal discount rate, and mean concern ratio estimated from the patient preference survey across patient subgroups. Standard errors are reported in the parentheses.

Population Subgroups	Risk of Serious Bleeding		Risk of Serious Infection		Concern Ratio (γ)
	MAR	r	MAR	r	
Overall	11.08% (0.54%)	90.9% (9.9%)	15.04% (1.02%)	57.5% (5.3%)	0.52 (0.01)
Home Hemodialysis	8.99% (0.74%)	57.5% (11.8%)	14.31% (1.60%)	48.1% (9.9%)	0.53 (0.01)
In-Center Hemodialysis	9.81% (0.71%)	129.9% (23.6%)	9.23% (1.55%)	57.8% (7.7%)	0.50 (0.01)
Peritoneal Dialysis	16.63% (1.23%)	101.0% (26.6%)	26.80% (2.12%)	69.4% (17.8%)	0.55 (0.01)
Age: Less than 40 years	9.38% (1.04%)	39.5% (8.6%)	16.50% (2.56%)	43.9% (7.3%)	0.52 (0.01)
Age: between 40 and 64 years old	10.90% (0.70%)	116.3% (107.1%)	15.05% (1.39%)	59.9% (19.8%)	0.53 (0.01)
Age: 65 years and above	12.24% (0.85%)	103.1% (84.2%)	14.30% (1.91%)	65.4% (25.2%)	0.51 (0.01)
Male	9.80% (0.66%)	97.1% (26.2%)	15.24% (1.44%)	59.2% (11.8%)	0.52 (0.01)
Female	12.49% (0.78%)	85.5% (10.2%)	14.87% (1.46%)	56.2% (6.3%)	0.53 (0.01)
Time on Dialysis: less than 1 year	10.92% (1.31%)	128.2% (52.6%)	15.14% (2.21%)	68.5% (15.0%)	0.57 (0.01)
Time on Dialysis: 1 to 4 years	10.75% (0.64%)	84.7% (34.1%)	15.16% (1.37%)	54.9% (0.15%)	0.52 (0.01)
Time on Dialysis: more than 5 years	11.93% (1.05%)	89.3% (39.6%)	14.72% (2.10%)	57.8% (17.9%)	0.51 (0.01)
Hispanic	7.68% (0.87%)	83.3% (27.7%)	10.55% (2.26%)	32.3% (5.5%)	0.53 (0.01)
Non-Hispanic (Overall)	12.18% (0.62%)	92.6% (10.3%)	16.48% (1.14%)	65.4% (6.0%)	0.52 (0.01)
White Non-Hispanic	12.83% (0.77%)	107.5% (15.0%)	16.44% (1.44%)	80.6% (9.8%)	0.52 (0.01)
Black Non-Hispanic	10.26% (0.91%)	65.4% (14.4%)	15.61% (1.89%)	47.2% (7.5%)	0.52 (0.01)
Other Non-Hispanic	14.12% (3.34%)	62.1% (17.0%)	22.02% (8.11%)	60.6% (18.5%)	0.48 (0.04)
Two or More Races Non Hispanic	19.94% (6.03%)	625.0% (746.3%)	23.32% (10.63%)	17.3% (1.4%)	0.55 (0.04)

We assess the willingness to wait of the patients using a temporal discount rate r estimated in [197]. This discount rate is obtained for the risk of serious bleeding and the risk of serious infection separately and are reported in Table 2.2. We can note a high impatience (i.e., low willingness to wait) for the hypothetical device across all patient subgroups.

Given that the device presents two sources of risk, we define a concern ratio parameter $\gamma \in [0, 1]$ to aggregate both risk measures into a single risk measure:

$$NetRisk_{Device} = NetRisk_{Bleeding} \cdot (1 - \gamma) + NetRisk_{Infection} \cdot \gamma, \tag{2.3}$$

$$MAR_{Device} = MAR_{Bleeding} \cdot (1 - \gamma) + MAR_{Infection} \cdot \gamma. \tag{2.4}$$

By definition, $\gamma = 0$ if the patient is only concerned about the risk of serious bleeding and $\gamma = 1$ if the patient is only concerned about the risk of serious infection. A concern ratio $\gamma = 0.5$ reflects an equal concern between the risk of serious bleeding and serious infection. The concern ratio is estimated from the survey data (using the bleeding concern score and the infection concern score for each patient) as

$$\gamma = \frac{\text{Infection Concern Score}}{\text{Bleeding Concern Score} + \text{Infection Concern Score}}, \tag{2.5}$$

where each concern score was provided by the patient in the survey. Estimates of the concern ratio are reported in Table 2.2 and show that, across patient subgroups, the mean concern ratio is consistent with a value of $\gamma = 0.5$.

2.2.2 Patient-centered utility-based model

For the purpose of this analysis, we choose the null hypothesis to represent an unsafe device. More precisely, we state that a device is unsafe if the risk of the device is higher the patient's MAR by more than a safety margin δ_0 (chosen to be 3% for the risk of serious bleeding and 6% for the risk of serious infections, Table 2.1). In fact, we have chosen the safety margin to be slightly larger than the standard errors on the net risk of the device. Sensitivity analysis on the choice of safety margin is presented in Appendix A. Mathematically, we define the null and alternative hypotheses as

$$H_0 \text{ (“Unsafe Device”)}: \quad NetRisk_{Device} - MAR_{Device} > \delta_0, \tag{2.6}$$

$$H_1 \text{ (“Safe Device”)}: \quad NetRisk_{Device} - MAR_{Device} \leq \delta_0. \tag{2.7}$$

Not approving an unsafe device is assumed to lead to a zero loss in utility while not approving a safe device decreases the patient's utility by an amount L_1. On the other hand, approving an unsafe device decreases the patient's utility by an amount $DF_t \cdot L_0$, where $DF_t = e^{-rt}$ is the temporal discount factor measuring the patient's willingness to wait for a device with lower risks of bleeding or infection (based on the patient's temporal discount rate r and the time t until the device is approved), and approving a safe device decreases the patient's utility by an amount $(1 - DF_t) \cdot L_1$. The discount factor decreases from 1 to 0 as we increase the time t. In particular,

$DF_t = 1$ when the wait time is equal to zero, so approving an unsafe device would decrease the patient's utility by an amount L_0 while approving a safe device would not affect the patient's utility. The potential losses in utility per patient of a fixed-sample RCT are summarized in Table 2.3.

Hypotheses	Not Approved	Approved
H_0: Unsafe Device	0	$DF_t \cdot L_0$
H_1: Safe Device	L_1	$(1 - DF_t) \cdot L_1$

Table 2.3: Loss in value per patient associated with a balanced fixed-sample RCT.

Aggregating all four scenarios in Table 2.3, we obtain an expected loss in utility to patients given by

$$\mathbb{E}\left[\text{Loss}\right] = p_0 \cdot \alpha \cdot DF_t \cdot L_0 + p_1 \cdot \left[\beta + (1 - \beta) \cdot (1 - DF_t)\right] \cdot L_1, \tag{2.8}$$

where p_0 (p_1) is the prior probability that the device is unsafe (respectively, safe). We recall that L_0 and L_1 correspond to the patient's loss in utility due to the approval of an unsafe device (Equation 2.1) and the non-approval of a safe device (Equation 2.2), respectively, α is the type I error rate (i.e., the probability of approving an unsafe device), β is the type II error rate (i.e., the probability of not approving a safe device), and $DF_t = e^{-rt}$ is the temporal discount factor reflecting the patient's willingness to wait for a device with lower risks of bleeding or infection (based on the patient's temporal discount rate r).

While L_0, L_1, and DF_t are estimated from the patient preference surveys, the type I and type II error rates α and β are associated with the clinical trial endpoint, and can be calculated as the conditional probability

$$\alpha = \mathbb{P}\left(\text{Accept Device}|\text{Unsafe}\right) = \mathbb{P}\left(T_n \leq \lambda|\text{Unsafe}\right), \tag{2.9}$$

$$\beta = \mathbb{P}\left(\text{Reject Device}|\text{Safe}\right) = \mathbb{P}\left(T_n > \lambda|\text{Safe}\right), \tag{2.10}$$

where T_n is the sufficient statistic used in the RCT to assess the safety of the device (see Section 2.2.3), and λ is the critical value which determines the rejection region of the hypothesis test.

2.2.3 Patients-centered clinical trial design via Bayesian decision analysis

For concreteness, we define the safety endpoints of the RCT for the hypothetical wearable device as the risks of serious bleeding and serious infection due to the device. Therefore, during the RCT, we would measure the average risk of serious bleeding and serious infection in the treatment arm $\overline{Risk_{Device}}$ and in the control arm $\overline{Risk_{Baseline}}$. We assume that the patient outcomes in the control and treatment arms are independent and identically distributed independent and identically distributed (I.I.D) and

follow a normal distribution:

$$Risk_{Baseline} \sim N\left(\mu_c, \sigma_c^2\right) \quad \text{and} \quad Risk_{Device} \sim N\left(\mu_t, \sigma_t^2\right) \tag{2.11}$$

where μ_c (μ_t) denotes the mean of the risk endpoints observed in the control (treatment) arm, and σ_c (σ_t) denotes the standard deviation of the risk endpoints observed in the control (treatment) arm, respectively. We define the sufficient statistic of interest as

$$T_n = \sqrt{I_n} \cdot \left(\overline{Risk_{Device}} - \overline{Risk_{Baseline}} - MAR_{Device}\right) \tag{2.12}$$

where $I_n = \frac{n}{\sigma_c^2+\sigma_t^2}$ is a scaling factor and n is the number of patients per trial arm.

Under the null hypothesis, the statistic T_n follows a standard t-distribution with mean zero and $2(n-1)$ degrees of freedom. This means that a device with a net risk no less than the MAR will be considered unsafe. Conversely, under the alternate hypothesis, the statistic T_n is follows a non-central t-distribution with non-centrality parameter $-\delta_0 \cdot \sqrt{I_n}$, for a safety margin $\delta_0 > 0$, and $2(n-1)$ degrees of freedom.

In this model, the type I and type II error rates are calculated as follows

$$\alpha = \mathbb{P}\left(T_n \leq \lambda | \mu = 0\right) = F_t\left(\lambda | \mu = 0\right), \tag{2.13}$$

$$\beta = \mathbb{P}\left(T_n > \lambda | \mu = 0\right) = 1 - F_t\left(\lambda | \mu = -\delta_0 \cdot \sqrt{I_n}\right), \tag{2.14}$$

where $F_t(\cdot)$ is the cumulative distribution function of a non-central t-distribution with mean μ and $2(n-1)$ degrees of freedom.

Following the BDA procedure, the optimal critical value λ^* and the optimal sample size n^* are obtained by jointly minimizing the expected loss in utility to patients (Equation 2.8) over λ and n (Equation 2.13 and Equation 2.14). We further constrain the statistical power $(1-\beta)$ to be lower than 90% to reflect practical considerations of the medical device industry (which include budget, time, personnel, and other resource limitations).

2.3 Kidney replacement therapy devices from a BDA perspective

2.3.1 BDA-optimal type I error rates and clinical trial sizes

The BDA-optimal type I error rates and clinical trial sample sizes are reported in Table 2.4 for two safety endpoints: the risk of serious bleeding and the risk of serious infection. The results are provided for the overall patient population (row 1 of Table 2.4) and for each patient subgroups (rows 2 to 18 of Table 2.4). A graphical representation of the optimal type I error rates obtained is provided in Figure 2-1 for the risk of serious bleeding endpoint, Figure 2-2 for the risk of serious infection endpoint, and in Figure 2-3 for the aggregated safety endpoint.

Overall, we find a BDA-optimal type I error rate of 1.2% for the risk of bleeding and one below 0.1% for the risk of serious infection. Both values are lower than the

traditional 2.5% threshold, suggesting that patients are not willing to bear either type of additional risk presented by the hypothetical device in exchange for the probable benefits described in the survey instrument. Furthermore, we find a BDA-optimal total sample size of 17 patients for the risk of serious bleeding and 102 patients for the risk of serious infection. However, with a statistical power below 1% for the risk of serious infection and a type I error rate below 0.1%, the sample size becomes unimportant since such a device would need to demonstrate exceedingly high safety effects in order to receive FDA approval.

The BDA-optimal thresholds and sample sizes are also calculated across patient subgroups to reflect heterogeneity in patient preferences. We include the patient's mode of dialysis (home hemodialysis, in-center hemodialysis, and peritoneal dialysis), age (below 40, between 40 and 64, and above 65), gender (male and female), time on dialysis (less than 1 year, between 1 and 4 years, and more than 5 years), and ethnicity (Hispanic, overall non-Hispanic, white non-Hispanic, black non-Hispanic, other non-Hispanic, and 2 or more races and non-Hispanic).

Mode of Dialysis. The BDA-optimal type I error for the risk of serious bleeding is 0.3% for home hemodialysis patients, 1.0% for in-center hemodialysis patients, and 3.8% for peritoneal dialysis patients, suggesting that only peritoneal dialysis patients have a more risk-tolerant threshold than the traditional 2.5% significance level threshold. The optimal sample sizes range from 12 patients for peritoneal dialysis patients to 23 patients for home hemodialysis patients and we always achieve a statistical power of 90%. The BDA-optimal type I error for the risk of serious infection is below 0.1% for home and in-center hemodialysis patients and is 1.3% for peritoneal dialysis patients. Peritoneal dialysis patients are the most risk-tolerant group while home hemodialysis patients are the most risk-averse. Optimal sample sizes range from 101 to 102 patients for home and in-center hemodialysis with a statistical power of less than 1%, while we obtain an optimal sample size of 23 patients for peritoneal dialysis patients and achieve a statistical power of 90%.

Age. We find that the BDA-optimal type I error for the risk of serious bleeding increases with age: 0.2%, 1.5%, and 1.9% for patients with age below 40, between 40 and 64, and above 65 years old, respectively. However, these thresholds are still below the traditional 2.5% significance level threshold. Conversely, the optimal sample sizes decrease with age: 23, 16, and 15 patients for patients below 40, between 40 and 64, and above 65 years old, respectively, and we always achieve a statistical power of 90%. The BDA-optimal type I error for the risk of serious infection is below 0.1% across all age groups and the optimal sample size varies between 101 to 102 patients across all age groups with a statistical power of less than 1%.

Gender. As is often the case, we find that the BDA-optimal type I error rate for the risk of serious bleeding is higher for male (1.9%) than female (0.7%) patients, suggesting that males tend to be less risk-averse than women. However, these thresholds are still below the traditional 2.5% significance level threshold. The optimal sample size is 15 and 19 patients for male and female patients, respectively, and we always achieve a statistical power of 90%. The BDA-optimal type I error for the risk of serious infection is less than 0.1% for both male and female patients and the optimal sample size is 102 patients across genders with a statistical power below 1%.

Time on Dialysis. The BDA-optimal type I error for the risk of serious bleeding varies with the amount of time the patient has been on dialysis, taking values of 1.6%, 1.0%, and 1.5% for less than 1 year, 1 to 4 years, and more than 5 years on dialysis, respectively. Once again, these thresholds are still below the traditional 2.5% significance level threshold. Conversely, the optimal sample size does not change significantly across different times on dialysis, ranging between 16 and 17 patients required to achieve a 90% statistical power. The BDA-optimal type I error for the risk of serious infection is below 0.1% across all groups and we obtain an optimal sample size of 102 patients across all groups with statistical power below 1%.

Race and ethnicity. The BDA-optimal type I errors for the risk of serious bleeding are 2.3%, 0.7%, 1.7%, and 28% for white non-Hispanic, black non-Hispanic, other non-Hispanic, and multiracial non-Hispanic patients, respectively. Overall, this BDA-optimal type-I error is 1.7% for non-Hispanic patients and is below 0.1% for Hispanic patients. The optimal sample size ranges from 4 to 19 patients for all non-Hispanic subgroups to achieve a statistical power of 90%. A 90% statistical power cannot be achieved for Hispanic patients. For the risk of serious infection, we find a BDA-optimal type I error below 0.1% for all subgroups except other non-Hispanic patients (0.4%) and multiracial non-Hispanic patients (0.2%). We only achieve a 90% statistical power for non-Hispanic patients and multiracial non-Hispanic patients, with an optimal sample size of 28 and 33 patients respectively.

Table 2.4: BDA-optimal type I error rate, total sample size, and statistical power of the trial across patient population subgroups.

Population Subgroups	Risk of Serious Bleeding			Risk of Serious Infection		
	α	Sample Size	Power	α	Sample Size	Power
Overall	1.2%	17	90%	$< 0.1\%$	102	$< 1\%$
Home Hemodialysis	0.3%	23	90%	<0.1%	101	<1%
In-Center Hemodialysis	1.0%	17	90%	<0.1%	102	<1%
Peritoneal Dialysis	3.8%	12	90%	1.3%	23	90%
Age: Less than 40 years	0.2%	23	90%	<0.1%	101	<1%
Age: between 40 and 64 years old	1.5%	16	90%	<0.1%	102	<1%
Age: 65 years and above	1.9%	15	90%	<0.1%	102	<1%
Male	1.9%	15	90%	<0.1%	102	<1%
Female	0.7%	19	90%	<0.1%	102	<1%
Time on Dialysis: less than 1 year	1.6%	16	90%	<0.1%	102	<1%
Time on Dialysis: 1 to 4 years	1.0%	17	90%	<0.1%	102	<1%

Continued on next page

Table 2.4 – *Continued from previous page*

Population Subgroups	Risk of Serious Bleeding			Risk of Serious Infection		
	α	Sample Size	Power	α	Sample Size	Power
Time on Dialysis: more than 5 years	1.5%	16	90%	<0.1%	102	<1%
Hispanic	<0.1%	101	<1%	<0.1%	102	<1%
Non-Hispanic (Overall)	1.7%	15	90%	<0.1%	102	<1%
White Non-Hispanic	2.3%	14	90%	<0.1%	102	<1%
Black Non-Hispanic	0.7%	19	90%	<0.1%	101	<1%
Other Non-Hispanic	1.7%	15	90%	0.4%	28	90%
Two or More Races Non Hispanic	28%	4	90%	0.2%	33	90%

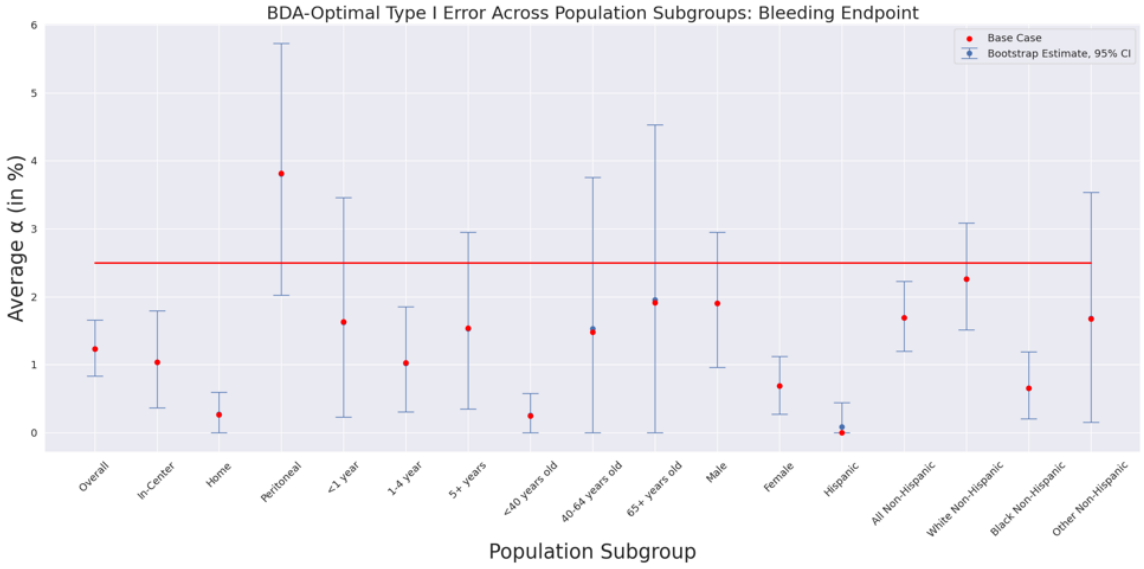


Figure 2-1: **BDA-Optimal Type I Error Rate for the Risk of Serious Bleeding.** The horizontal red line corresponds to the commonly used 2.5% significance level. Bootstrap estimates of 95% confidence intervals are plotted in blue.

2.3.2 Sensitivity Analysis

The BDA estimation procedure performed depends both on the estimated survey parameters and on our modeling assumptions. We perform a sensitivity analysis on three key parameters of the model to test the robustness of the BDA outputs. Additional considerations are provided in Appendix A.

First, the concern ratio has been defined as a simplifying tool to aggregate two risk measures when estimating the patient's loss in utility (Equation 2.3 and Equation 2.4). While this parameter is based on survey data, the linear aggregation performed

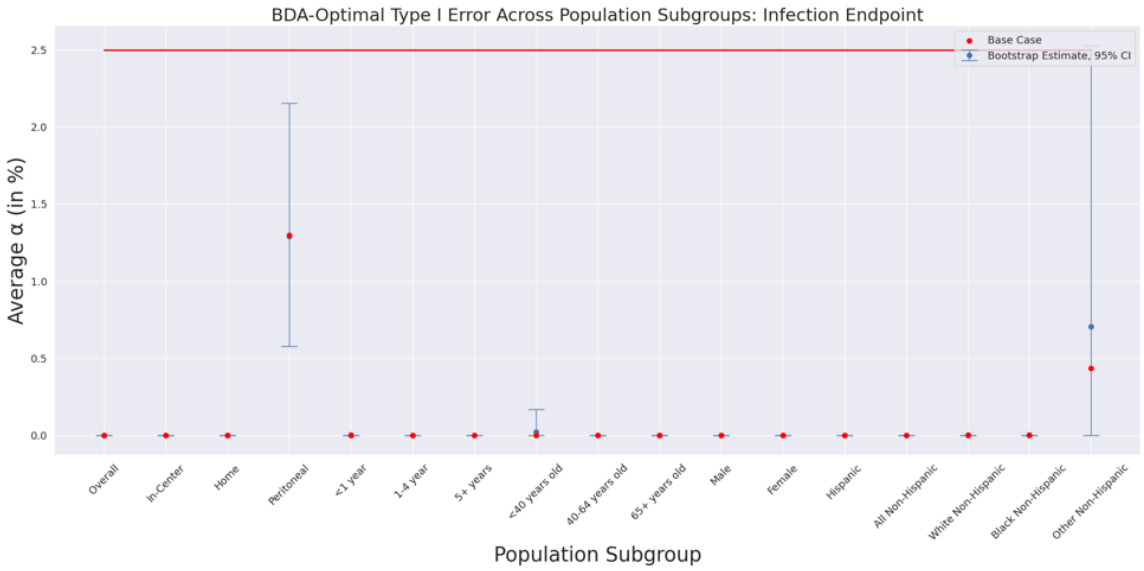


Figure 2-2: **BDA-Optimal Type I Error Rate for the Risk of Serious Infection.** The horizontal red line corresponds to the commonly used 2.5% significance level. Bootstrap estimates of 95% confidence intervals are plotted in blue.

Figure 2-3: **BDA-Optimal Type I Error Rate for the Overall Safety Endpoint.** The horizontal red line corresponds to the commonly used 2.5% significance level. Bootstrap estimates of 95% confidence intervals are plotted in blue.

in Equation 2.13 and Equation 2.14, albeit very simplistic and insightful, may not be entirely realistic. In Figure 2-4 and Figure 2-5, we plot the BDA outputs obtained for the entire patient population studied for the two safety endpoints considered when varying the concern ratio from 0 to 1. We observe that increasing the concern ratio from 0 to 1 reduces the optimal significance level threshold and increases the

optimal trial sample size calculated by the BDA. The results described in Table 2.4 are recovered by considering the risk of serious bleeding or the risk of serious infection as safety endpoint along with a concern ratio $\gamma = 0$ and $\gamma = 1$ respectively. We find that the results are still consistently lower than the 2.5% significance level threshold used in practice.

The concern ratio is also used to aggregate the two safety endpoints into an aggregate safety endpoint (Equation 2.12). Sensitivity to the concern ratio is highlighted in Figure 2-4 and Figure 2-5 when considering the aggregated safety endpoint. We also find that the BDA optimal significance level decreases with the concern ratio while the BDA optimal sample size increases with the concern ratio. The results described in Figure 2-3 are consistent with an aggregated safety endpoint when setting the concern ratio γ to the population's average ($\gamma = 0.52$).

Furthermore, we test the sensitivity of the BDA outputs to two key assumptions of the model: the safety margin δ_0 and the variability in the baseline risk and in the device risk. As shown in Figure A-1 and Figure A-2, reducing the safety margin leads to a less conservative significance level threshold while increasing the safety margin leads to a more conservative significance level threshold. The results are still consistently lower than the usual 2.5% significance level threshold used in practice. Finally, Figure A-3 and Figure A-4 show that reducing the variability in the risk of serious bleeding and in the risk of serious infection in the control and treatment groups leads to a more conservative significance level threshold while an increase in the variability parameters leads to a less conservative significance level threshold. Increasing the variability actually increases the significance level threshold above the 2.5% reference level when considering the risk of serious bleeding. The rest of the results remain consistently lower than the 2.5% reference point.

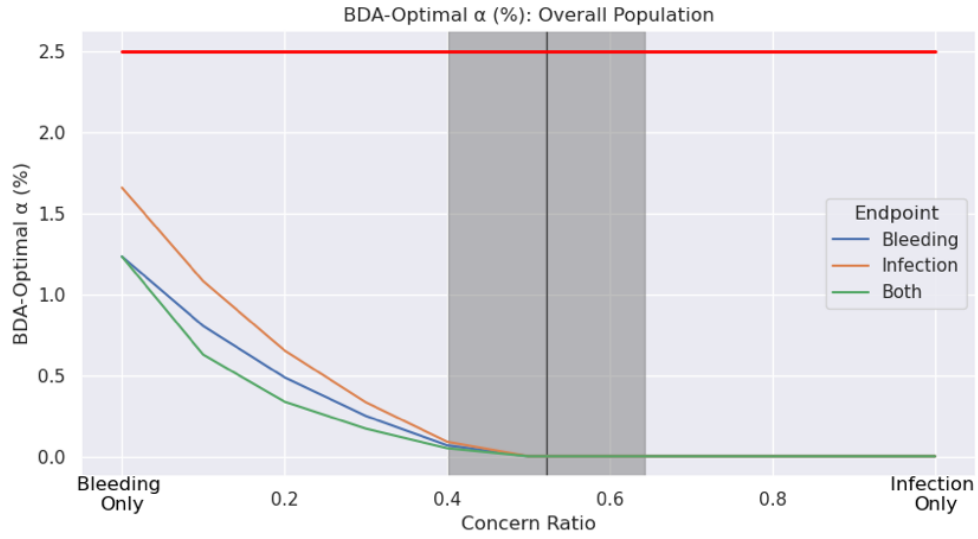


Figure 2-4: **BDA-Optimal Type I Error Rate for Three Safety Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents the range of plus/minus one standard deviation around the mean.

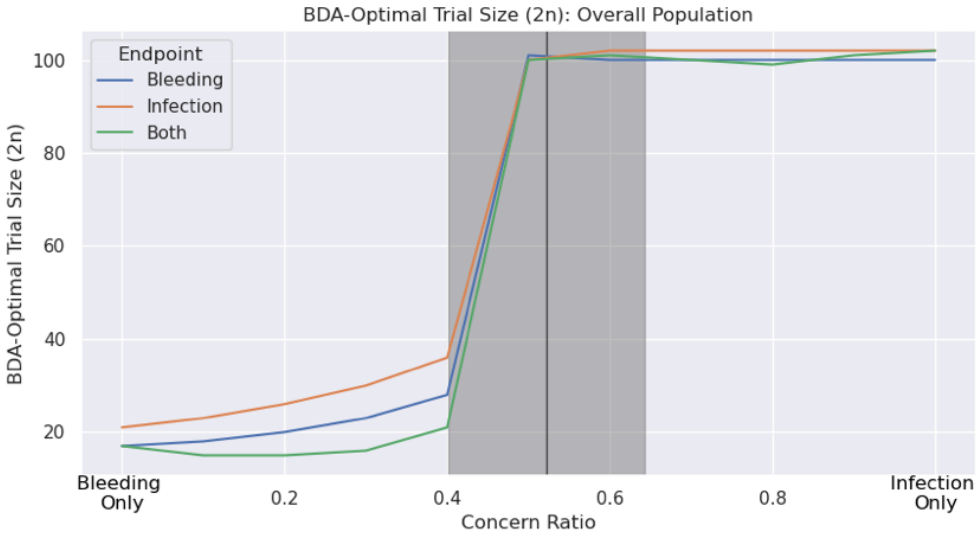


Figure 2-5: **BDA-Optimal Clinical Trial Total Sample Size for Three Safety Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents the range of plus/minus one standard deviation around the mean.

2.3.3 Discussion of the results

In this study, we found that a 2.5% fixed significance level threshold would not be an appropriate choice for evaluating the safety of a hypothetical kidney replacement therapy (KRT) device as it will not maximize the patient's utility. In fact, patients often require a much lower significance level threshold, reflecting a lower willingness to bear the risks of the device in exchange for its probable benefits and a stronger concern regarding the approval of a potentially harmful device over the potential fear of missing out on the device's benefits. Although we observe some heterogeneity across patient subgroups, only patients on peritoneal dialysis would be comfortable with a significance level threshold above 2.5% when evaluating the risk of serious bleeding endpoint. For the risk of serious infection, we found that all patient subgroups would require a much lower significance level threshold than 2.5%, with patient on peritoneal dialysis being the most permissive subgroup at a 1.3% threshold. These observations remain valid as we perturb key assumptions of the model as shown in Section 2.3.2.

Another interesting observation can be made from this study regarding the temporal preferences of the patients surveyed. As shown in Table 2.2, patients tend to present very high temporal discount rates (r), or equivalently very low willingness to wait ($1/r$). Most patients surveyed would not be willing to wait more than a year or two to obtain a similar device with lower levels of risks. This is counter-intuitive as patients seem to have a very low risk-tolerance and a high level of impatience. The former would lead to less devices being approved, and therefore longer wait times until a new device is approved, in contradiction with the latter. Such impatience is not uncommon in behavioral economics. In fact, researchers have found that individuals may, in certain situations, prefer small rewards now over larger rewards in the future, violating the time-consistency of exponential discounting models. In these cases, hyperbolic discounting has been proposed as an alternative [150]. Under hyperbolic discounting, the discount factor would no longer take the form e^{-rt} but rather $\frac{1}{rt}$, making the individual more biased towards immediate rewards (even if these rewards are small). For example, an individual with hyperbolic discounting temporal preferences would therefore prefer a riskier device now to a device that presents lower levels of risks but would be available more than two years later.

Finally, this application of BDA to KRT devices highlights the objectivity of the method: by maximizing the patient's utility, BDA is able to correctly reflect the patient's preferences into a quantifiable benefit-risk tradeoff and select an appropriate significance level threshold to evaluate the safety and/or the efficacy of a potential treatment. However, it is important to stress that the objectivity and usefulness of recommendation from the BDA depends critically on the quality of the patient preferences survey used. Ideally, the design of the survey would be prepared in partnership with both regulators and patient advocacy organizations and would contain survey questions that are consistent with the clinical endpoints that would eventually be used to assess the safety and/or efficacy of the hypothetical treatment in a clinical trial.

2.3.4 Limitations and sensitivity analysis

Our findings must be qualified in several respects. First, the survey instrument used to assess patient preferences cannot fully describe and quantify the probable benefits of the hypothetical KRT as there are no actual wearable KRT device on the market. To ensure that patients had a basic understanding of some key probable benefits of the device, a description of these benefits was included in the KHI survey along with comprehension questions that verified whether the patient had understood the outlined benefits. These questions were usually answered correctly, mitigating the risk of uninformed responses from the patients. In fact, a lack of appreciation of the device's benefits can lead patients to make more conservative choices and reduce the significance level threshold calculated by the BDA.

Second, unlike MARs which tend to be precisely estimated (Table 2.2), we observe large standard errors when estimating temporal discount rate (Table 2.2). This can be partially explained by the fact that patients were randomly asked to answer the time tradeoff questions for either the risk of serious bleeding or the risk of serious infection, lowering the number of responses from 577 (for MAR threshold regressions) to 212 and 183 (respectively for the risks of serious bleeding and serious infection). Another explanation relies on a large inherent variability in each patient's willingness to wait. We mitigate this effect by estimating standard errors on the BDA outputs as shown in Figure 2-1, Figure 2-2, and Figure 2-3 which reflect the uncertainty in the temporal discount rate and in the MARs estimated through threshold regressions.

2.4 Conclusion

Critics of Bayesian Decision Analysis have often incorrectly viewed this approach as a way to inflate significance level thresholds and lead researchers and regulators to be more permissive in their decision process. While an increase in the type I error rate threshold is often caused by the patient's desire to get access to a novel therapy, this application to kidney replacement therapy devices provides a concrete example of patients requiring significance level thresholds far below the typical 2.5% used in practice. Indeed, by design, BDA will lead to more permissive thresholds when patients are willing to accept a higher level of therapeutic risks in exchange for greater benefits but would lead to more conservative thresholds when the perceived risks outweigh the potential benefits. When correctly applied, BDA incorporates the patient's preferences in a truthful, objective, transparent, and repeatable way, and can provide useful recommendations to all stakeholders.

Chapter 3

An Application of Bayesian Decision Analysis to Patient-Centered Randomized Clinical Trials in Parkinson's Disease

Before generalizing the BDA approach in Chapter 4, we illustrate the classical BDA using a second case study: while Chapter 2 applied BDA to the safety endpoint of a hypothetical kidney replacement therapy device, we now turn our attention to efficacy endpoints for a hypothetical deep brain stimulation (DBS) device for patients suffering from Parkinson's disease (Parkinson's disease (PD)).

More concretely, we apply BDA to PD patient preference scores elicited from survey data. BDA allows us to choose a sample size n and significance level (α) that maximizes the overall expected value to patients of a balanced two-arm fixed-sample RCT, where the expected value is computed under both null and alternative hypotheses.

By explicitly incorporating patient preferences into clinical trial designs and the regulatory decision-making process, BDA provides a quantitative and transparent approach to combine clinical and statistical significance. For PD patients who have never received DBS treatment, we find that a 5% significance threshold may not be conservative enough to reflect their risk-aversion level. However, this study shows that patients who previously received DBS treatment present a higher tolerance to accept therapeutic risks in exchange for improved efficacy which is reflected in a higher statistical threshold.

This study is part of a larger collaborative initiative to identify and measure the factors that matter most to PD patients [166].

3.1 Introduction

In this chapter, we use quantitative patient preference data elicited from a preference study of PD patients to design pivotal clinical trials with patient-centered p-value thresholds and sample sizes [113]. Patient-centered design may be substantially different from the conventional approach, which may consider using a fixed one-sided significance level of 5%. Based on the device and indication considerations, the FDA typically uses a 2.5% one-sided significance level for studies presenting "conventional uncertainty" and a 5% one-sided significance level under "moderate uncertainty" [79]. To err on the conservative side, we use the 5% one-sided level as our benchmark. For patients who had previously received DBS treatment, the significance levels identified by the BDA framework are similar to or larger than 5%, while for patients with who had never received DBS, the significance levels are smaller than 5%. The optimal significance levels obtained for patients with prior experience with DBS are in fact always strictly larger than a 2.5% significance level benchmark. Moreover, in both these populations, the patients with more severe cognitive and motor function symptoms were willing to accept more uncertainty. This patient-centered design necessarily makes several assumptions, and a sensitivity analysis of these results is provided in the Appendix B.1. The usefulness of these recommendations relies on the appropriateness of these key assumptions and accurately calibrated model parameters.

3.2 Patient-centered clinical trial designs

We consider a quantitative framework that explicitly takes patient preferences into account across multiple device attributes when determining the optimal sample size and critical value of a balanced two-arm fixed-sample RCT. We first define a patient-centered value model associated with given medical device attributes. Our patient value model is based on preference data for a specific, hypothetical device [46]. This framework is derived from previous oncology analyses [168, 126] and can be used in areas where patient-preference data are available.

We assign prior probabilities to each possible combination of attributes and formulate the expected value of the trial. The optimal trial size ($2n$, where there are n patients in each arm of the study) and the one-sided significance level (α or critical value, λ_α) are then jointly determined to maximize the expected value of the trial. For expositional simplicity, we have assumed a balanced clinical trial, but the methodology also applies more generally to cases where the clinical trial is unbalanced. Note that maximizing the value of the trial means either providing access to a safe and effective treatment to patients, or reaching the conclusion that the treatment has not demonstrated a reasonable assurance of safety and effectiveness as soon as possible. This is equivalent to minimizing the potential losses, which include the consequences of incorrect decisions for all patients, and the time inefficiency of delaying access of a potentially safe and effective treatment.

3.2.1 Patient Value Model

A multidisciplinary team, including researchers from the Medical Device Innovation Consortium (MDIC), the Michael J. Fox Foundation for Parkinson's Research (MJFF), and the US Food and Drug Administration (FDA), developed and administered a survey to patients with PD to quantify the maximum level of risk that patients would accept to achieve different potential benefits of a neurostimulation device [113]. The ultimate aim of this study was to explore the feasibility of using patient preference information to optimize clinical trial design for medical device treatments for PD. This study began with qualitative research that incorporated input from patient partners and regulatory end-users in attribute development [20]. Based on this input from patients and regulators, a threshold technique survey was developed and administered to quantify the benefit-risk tradeoffs that are acceptable to people with PD and to assess their willingness to wait for a novel device that offers a greater therapeutic benefit. The threshold technique was used to quantify patients' risk tolerance for new or worsening depression or anxiety, brain bleed, or death in exchange for improvements in "on-time", motor symptoms, pain, cognition, and pill burden. The survey also elicited patients' maximum willingness to wait to receive treatment benefit. Equation 3.1 shows the patient preference model estimated from the survey data [20, 113].

$$\begin{aligned}\text{Risk threshold} = (\beta_1 + \beta_2 \cdot \text{Agecat2} + \beta_3 \cdot \text{Agecat3} + \beta_4 \cdot \text{Agecat4}) \cdot \text{Benefit} \\ + \gamma_1 \cdot \text{Non-ambulatory} + \gamma_2 \cdot \text{Cognitive symptom} + \gamma_3 \cdot \text{DBS} \\ + \gamma_4 \cdot \text{Dyskinesia} + \gamma_5 \cdot \text{Motor symptom} + \text{Error}\end{aligned} \tag{3.1}$$

The *Risk threshold* estimates the additional risk, in the form of mortality, brain bleed, and depression risk, the patient would be willing to accept above their baseline risk before becoming indifferent to receiving a given benefit. The absolute level of benefit is denoted Benefit, and can include improvements in "on time" (a common clinical trial endpoint in Parkinson's disease clinical trials), motor symptoms, cognition, and pain level. Descriptions of indicator variables that specify the model for different subgroups of patients are provided in Table 3.1. Finally, *Error* is a random error term with mean 0.

Indicator Variable	Description
Agecat2	Equals 1 if patients are between 61 and 66 years old
Agecat3	Equals 1 if patients are between 67 and 71 years old
Agecat4	Equals 1 if patients are 72 years old or older
Non-ambulatory	Equals 1 if patients report problems with balance or walking
Cognitive symptom	Equals 1 if patients report difficulty thinking clearly
DBS	Equals 1 if patients report received DBS
Dyskinesia	Equals 1 if patients experience dyskinesia as a side effect of their medication

Table 3.1: Definition of indicator variables in Equation 3.1.

The patient preferences reflected in Equation 3.1 can be aggregated across benefits and risks to compute an overall preference score for a specific treatment. For example, the patient value lost from an increase in mortality risk resulting from a more invasive surgery could be offset by any additional benefits. The benefits, such as the value gained from increase in on-time can be mapped to an equivalent decrease in mortality risk using Equation 3.1. Given both these changes, and holding other attributes constant, the net change in value could then be calculated to determine if the additional on-time would more than compensate for the increased mortality risk according to the patient preference information, suggesting the new benefit-risk profile is preferred. The relative loss of value per patient, L, of using a lower-scored intervention over another, is then defined in terms of the net difference in value.

Throughout this analysis, the null hypothesis H=0 will represent an ineffective but possibly harmful device and the alternative hypothesis $H = 1$ will represent an effective device. The value of a clinical trial for a superiority claim depends on the outcome of the trial, yet it affects patients beyond the scope of the trial. We assume there is no post-trial loss in value for a correct decision, i.e., rejecting (approving) a device that is less (more) preferred relative to the control, except for the wait time of the regulatory review process. We make a further distinction in the relative loss of value per patient, L, defining L_0 as the relative loss of value per patient from using an ineffective device under the null hypothesis ($H = 0$), and L_1 as the relative loss in value per patient from missing out on the opportunity to use an effective device under the alternative hypothesis ($H = 1$). The severity ratio, L_1/L_0 , thus provides a measure of their relative importance to a patient.

In addition, we can calculate the average loss in value per patient of a false approval or rejection as $DF_t \cdot L_0$ and L_1, respectively, where DF_t is a discount factor that decreases from 1 to 0 to account for the wait time, t, caused by the regulatory approval process. In other words, patients will place a lower value on an effective treatment if it is not accessible immediately. Therefore, the average per patient loss in value caused by the length of the regulatory approval process under the alternative hypothesis is $(1 - DF_t) \cdot L_1$.

Hypotheses	$\hat{H} = 0$ (Not Approved)	$\hat{H} = 1$ (Approved)
$H = 0$	0	$DF_t \cdot L_0$
$H = 1$	L_1	$(1 - DF_t) \cdot L_1$

Table 3.2: Loss in value per patient associated with a balanced fixed-sample RCT.

The potential losses in value per patient of a fixed-sample trial are shown in Table 3.2. Note that there is no loss in value (i.e., there is maximum value) in the hypothetically optimal scenario where the correct approval decision is made immediately and without running a trial. The number of patients affected in each subgroup can be used to scale these values to estimate a collective loss of value. In this case study, we assume that the patient populations affected by a Type 1 (false positive) or Type 2 (false negative) error are approximately equal, hence our focus on the per patient loss.

Time horizon preferences, which measure the capacity of patients to tolerate waiting, are obtained directly from patient survey responses. We assume time-consistent (i.e., exponential) discounting. If the annual discount rate is R, then the discount factor is given by $DF_t = e^{-R \cdot t}$, where t is length of the regulatory approval process. This assumption ensures that patient preferences do not change over time in such a way that they become inconsistent with one another. Equations 3.2 and 3.3 specify the fitted discount rate model that estimates the indifference thresholds of the wait time for present versus future benefits [113].

$$\text{Wait time threshold} = \text{Discount rate} \cdot \log(\text{Benefit}) + \text{Error}, \tag{3.2}$$

$$\begin{aligned} \text{Discount rate} = & [\delta_1 + \delta_2 \cdot \text{Agecat2} + \delta_3 \cdot \text{Agecat3} + \delta_4 \cdot \text{Agecat4} \\ & + \delta_5 \cdot \text{Non-ambulatory} + \delta_6 \cdot \text{Cognitive symptom} + \delta_7 \cdot \text{DBS} \\ & + \delta_8 \cdot \text{Dyskinesia} + \delta_9 \cdot \text{Motor symptom}]^{-1}. \end{aligned} \tag{3.3}$$

The duration of the regulatory approval process (t) is determined by the size of the study ($2n$), the patient accrual rate for the study (η), the time required to set up the study (s), the follow-up time for the final patient to complete the study (f) and the FDA review time (τ), such that $t = s + 2n/\eta + f + \tau$. The estimates of the discount rate calculated from this model range from 14.5% to 32.7% per year, increasing as a function of both age and disease severity. This discount rate governs the trade-off between the study accrual (and hence the trial duration) and the overall error rate (both Type 1 and Type 2 error). The more data that are collected, the greater the certainty in the regulatory decision, but also the longer patients may need to wait to access the new treatment.

3.2.2 Bayesian Decision Analysis

A quantitative primary endpoint that tests for efficacy is assumed for the trial; however, the same BDA framework can be applied to evaluate safety concerns. In our analysis, we assume the primary endpoint is based on motor function. We further assume that subjects in the treatment arm receive the investigational device, and each subject's response is independent of all other responses. In the same way, patients in the control arm are assumed to receive a standard of care treatment, including adjuvant medication and non-pharmacological therapy (e.g., physical, occupational, and speech therapies).

The response variables in the treatment arm, denoted by $\{T_1, \ldots, T_n\}$, are assumed to be independent and identically distributed, where $T_i \sim N(\mu_t, \sigma_t^2)$. Similarly, the control arm responses, represented by $\{P_1, \ldots, P_n\}$, are assumed to be independent and identically distributed as $P_i \sim N(\mu_p, \sigma_p^2)$. As part of the model, we further confine ourselves to superiority trials, where the device is likely to have either a positive effect ($\mu_t > \mu_p$), or no effect ($\mu_t \leq \mu_p$). In such cases, the treatment effect of the device, δ, is defined as the difference between the response means in the two arms (i.e., $\delta := \mu_t - \mu_p$). In fixed-sample trial with n subjects in each arm, we collect

observations from the treatment and control arms, and form the following t-statistic:

$$T = \frac{\hat{\mu}_t - \hat{\mu}_p}{\sqrt{\frac{\hat{\sigma}_t^2}{n} + \frac{\hat{\sigma}_p^2}{n}}}, \tag{3.4}$$

where $\hat{\mu}$ and $\hat{\sigma}$ represent the sample mean and standard deviation, respectively, and T has a noncentral t-distribution with noncentrality parameter $\delta \cdot \sqrt{\frac{n}{\sigma_t^2+\sigma_p^2}}$. Under the assumption that the two variances are equal, this distribution has $2(n-1)$ degrees of freedom. The t-statistic is then compared to the critical value, λ_α. Finding that $T > \lambda_\alpha$ supports the rejection of the null hypothesis (i.e., that the device has no effect). The probability of failing to reject the null hypothesis, for a device that provides a treatment effect δ with response variances σ_t^2 and σ_p^2, is therefore $\mathbb{P}(T \leq \lambda_\alpha)$.

Assuming previously observed probabilities p_0 and p_1 (where $p_0 + p_1 = 1$) for the cases where the investigational device is equally effective ($H = 0$) and more effective ($H = 1$) to the control treatment, respectively, and letting V_0 and V_1 be the value created in the hypothetically optimal scenarios where the correct approval decision is made immediately and without running a trial, the expected value associated with an RCT design with parameters (n, λ_α) is

$$\mathbb{E}[\text{Value}; n, \lambda_\alpha] = p_0 \cdot (V_0 - \mathbb{E}[\text{Loss}|H = 0]) + p_1 \cdot (V_1 - \mathbb{E}[\text{Loss}|H = 1]), \tag{3.5}$$

where

$$\mathbb{E}[\text{Loss}|H = 0] = \alpha \cdot DF_t \cdot L_0, \tag{3.6}$$

and

$$\mathbb{E}[\text{Loss}|H = 1] = \beta \cdot L_1 + (1 - \beta) \cdot (1 - DF_t) \cdot L_1, \tag{3.7}$$

α is the significance level, and $(1 - \beta)$ is the power of the trial. The optimal sample size (n^*) and critical value (λ_α^*) are jointly determined such that the expected value of the trial is maximized subject to an upper bound chosen by the user for the power level. This power constraint is intended to represent the typical practices in the medical device industry, which reflect budget, time, personnel, and other resource limitations. While a 90% power constraint is assumed throughout the analysis, we consider relaxing this constraint in the sensitivity analysis provided in Appendix B.1 by imposing no power constraint (Table B.1) and an 80% power constraint (Table B.2). Finally, in solving the constrained optimization problem, we observe that the expected value of the trial is maximized when the expected loss,

$$\mathbb{E}[\text{Loss}; n, \lambda_\alpha] = p_0 \cdot \mathbb{E}[\text{Loss}|H = 0] + p_1 \cdot \mathbb{E}[\text{Loss}|H = 1], \tag{3.8}$$

is minimized.

3.2.3 Bayesian Decision Analysis and Expected Value of Information

In this section, we demonstrate how it is possible to embed our BDA approach as a special case of the EVI. EVI involves maximizing the expected utility of collecting incremental information and is a general framework for applying Bayesian analysis to decision making. In particular, the EVI decision rule to approve a treatment if the expected net benefit given the sample data is greater than zero maximizes the expected net benefit (which is the objective function in our BDA framework). As such, we can show that the BDA-optimal statistical significance threshold is equivalent to the EVI decision rule to approve a treatment if the expected net benefit given the sample data is greater than zero.

Our BDA framework aims to maximize the expected value associated with an RCT design with a sample size n and a critical value λ_n,

$$\max_{n,\lambda_n} \mathbb{E}\left[\text{Value}; n, \lambda_n\right]. \tag{3.9}$$

This framework maximizes the value of the RCT by first finding the optimal critical value λ_n as a function of the sample size n, and then finding the optimal sample size n. If we assume a probability distribution for the clinical profile θ of the treatment, then we can express Equation 3.9 as follows,

$$\max_{n,\lambda_n} \mathbb{E}\left[\text{Value}; n, \lambda_n\right] = \max_{n,\lambda_n} \int_\theta p_\theta \cdot \mathbb{E}\left[\text{Value}|\theta\right] d\theta. \tag{3.10}$$

For example, if we assume that the treatment's profile can only take two values i.e., $\theta = 0$ with probability p_0 and $\theta = 1$ with probability p_1, then Equation 3.10 reduces to the framework used in this chapter,

$$\max_{n,\lambda_n} \mathbb{E}\left[\text{Value}; n, \lambda_n\right] = \max_{n,\lambda_n} \left\{p_0 \cdot \mathbb{E}\left[\text{Value}|\theta = 0\right] + p_1 \cdot \mathbb{E}\left[\text{Value}|\theta = 1\right]\right\}. \tag{3.11}$$

The maximization described in 3.10 is more commonly performed by minimizing the expected loss of the trial rather than by maximizing the expected value of the trial. Both methods are equivalent as,

$$\mathbb{E}\left[\text{Value}; n, \lambda_n\right] = \mathbb{E}\left[\text{Value}^* - \text{Loss}; n, \lambda_n\right] = \mathbb{E}\left[\text{Value}^*\right] - \mathbb{E}\left[\text{Loss}; n, \lambda_n\right], \tag{3.12}$$

where Value^* is the value created in the hypothetically optimal scenarios where the correct approval decision is made immediately and without running a trial. Since $\mathbb{E}\left[\text{Value}^*\right]$ is independent of n and λ_n, Equation 3.10 is equivalent to:

$$\min_{n,\lambda_n} \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] = \min_{n,\lambda_n} \int_\theta p_\theta \cdot \mathbb{E}\left[\text{Loss}|\theta\right] \, d\theta. \tag{3.13}$$

Given an observation $T_n = t_n$ of the trial's summary statistic, Equation 3.13 can be

reformulated as follows,

$$\begin{aligned}
\min_{n,\lambda_n} \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] &= \min_{n,\lambda_n} \int_\theta p_\theta \cdot \mathbb{E}\left[\text{Loss}|\theta\right]\, d\theta && (3.14)\\
&= \min_{n,\lambda_n} \int_\theta p_\theta \cdot \Big[\mathbb{P}\left(\text{Approve}|\theta\right) \cdot C_{\theta,app} + \mathbb{P}\left(\text{Reject}|\theta\right) \cdot C_{\theta,rej}\Big]\, d\theta && (3.15)\\
&= \min_{n,\lambda_n} \int_\theta p_\theta \cdot \Big[\mathbb{P}\left(\text{Approve}|\theta\right) \cdot C_{\theta,app} + \left(1 - \mathbb{P}\left(\text{Approve}|\theta\right)\right) \cdot C_{\theta,rej}\Big] d\theta && (3.16)\\
&= \min_{n,\lambda_n} \int_\theta \Big[p_\theta \cdot \mathbb{P}\left(\text{Approve}|\theta\right) \cdot \left(C_{\theta,app} - C_{\theta,rej}\right) + p_\theta \cdot C_{\theta,rej}\Big]\, d\theta && (3.17)\\
&= \min_{n,\lambda_n} \int_\theta \Big[p_\theta \cdot \mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot \left(C_{\theta,app} - C_{\theta,rej}\right) + p_\theta \cdot C_{\theta,rej}\Big]\, d\theta && (3.18)\\
&= \min_{n,\lambda_n} \int_\theta \int_{\lambda_n}^{\infty} p_\theta \cdot \mathbb{P}\left(T_n = t_n|\theta\right) \cdot \left(C_{\theta,app} - C_{\theta,rej}\right) dt_n\, d\theta + \int_\theta p_\theta \cdot C_{\theta,rej}\, d\theta && (3.19)\\
&= \min_{n,\lambda_n} \int_{\lambda_n}^{\infty} \int_\theta p_\theta \cdot \mathbb{P}\left(T_n = t_n|\theta\right) \cdot \left(C_{\theta,app} - C_{\theta,rej}\right) d\theta\, dt_n + \int_\theta p_\theta \cdot C_{\theta,rej}\, d\theta && (3.20)
\end{aligned}$$

Equation 3.20 is minimized if the first derivative with respect to λ_n is equal to zero. Applying Leibniz's integral rule we find,

$$\begin{aligned}
\frac{\partial}{\partial \lambda_n} \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] &= \frac{\partial}{\partial \lambda_n} \int_{\lambda_n}^{\infty} \int_\theta p_\theta \cdot \mathbb{P}\left(T_n = t_n|\theta\right) \cdot \left(C_{\theta,app} - C_{\theta,rej}\right) d\theta\, dt_n && (3.21)\\
&= -\int_\theta p_\theta \cdot \mathbb{P}\left(T_n = \lambda_n|\theta\right) \cdot \left(C_{\theta,app} - C_{\theta,rej}\right) d\theta. && (3.22)
\end{aligned}$$

Hence, the decision rule becomes:

$$-\int_\theta p_\theta \cdot \mathbb{P}\left(T_n = t_n|\theta\right) \cdot \left(C_{\theta,app} - C_{\theta,rej}\right) d\theta \quad \underset{\text{Reject}}{\overset{\text{Approve}}{\gtrless}} \quad 0, \qquad (3.23)$$

which is equivalent to,

$$\int_\theta p_\theta \cdot \mathbb{P}\left(T_n = t_n|\theta\right) \cdot \left(C_{\theta,app} - C_{\theta,rej}\right) d\theta \quad \underset{\text{Approve}}{\overset{\text{Reject}}{\gtrless}} \quad 0. \qquad (3.24)$$

Using Baye's rule again, we get:

$$\int_\theta \left(C_{\theta,app} - C_{\theta,rej}\right) \cdot \mathbb{P}\left(\theta|T_n = t_n\right)\, d\theta \quad \underset{\text{Approve}}{\overset{\text{Reject}}{\gtrless}} \quad 0. \qquad (3.25)$$

The term $\left(C_{\theta,app} - C_{\theta,rej}\right)$ can be interpreted as a net cost (i.e., loss of value) of approving a therapy with clinical profile θ. Therefore,

$$C_{\theta,app} - C_{\theta,rej} = -\text{Net Benefit}(\theta). \qquad (3.26)$$

Hence Equation 3.26 leads to the EVI decision rule,

$$\int_\theta \text{Net Benefit}(\theta) \cdot \mathbb{P}\left(\theta | T_n = t_n\right)\, d\theta \quad \underset{\text{Reject}}{\overset{\text{Approve}}{\gtrless}} \quad 0, \tag{3.27}$$

which says to approve a treatment if the expected net benefit given the sample data is greater than zero.

In this chapter, we consider a probability model with only two possible outcomes to connect the BDA framework (and by extension, the EVI framework) to the frequentist null and alternative hypothesis framework used in most clinical trials today. Given the additional information needed to apply EVI—the utility function, the probability distributions of the state variables, etc.—we believe that our approach is the closest alternative to the status quo and therefore a practical next step for allowing regulators to incorporate patient preferences in their current decision framework. The BDA framework described in this chapter bridges these two approaches enabling the classical clinical trial endpoint analysis to reflect multidimensional benefit-risk tradeoffs.

3.3 Deep brain stimulation case study

Using BDA and the estimated patient preference model, we formulated a patient-centered fixed-sample RCTs for Parkinson's devices. Table 3.3 summarizes the parameter values used in our analysis. These parameters have been calibrated based on regulatory reviews of DBS devices and literature reviews of efficacy and safety [11, 216, 140, 185]. In Appendix B.1, we use sensitivity analysis to investigate the robustness of our analysis to perturbations in key parameter values assumed by our model. To provide readers with greater transparency and intuition regarding our Bayesian decision model, we provide the source code of our analysis in the Supplementary Materials, allowing users to input their own parameter values of interest and see the results. Table 3.4 reports the optimal RCTs for population subsets, including patients who have previously received DBS, patients with problems walking, patients with difficulty thinking clearly, patients with severe movement symptoms, and patients who experience dyskinesia as a side effect of their Parkinson's medication. These results are illustrated in Figure 3-1. In this analysis, preferences across age cohorts have been weighted equally. (A sensitivity analysis is provided in Appendix B.1 in which age cohort preferences are considered individually.)

Table 3.3: Assumptions for the RCT design.

Parameter	Description	Value
Probability that the treatment is effective (p_1)	The estimated a priori probability that the treatment is effective ($H = 1$), which can be estimated from historical success rates or set to 50% when there is no prior information	50%
Motor symptom score mean difference (δ)	Average reduction in motor symptoms on a rating scale from 0 to 10 under $H = 1$	0.98 out of 10
Motor symptom score standard deviation (σ_t and σ_p)	Treatment response variability in both the treatment and control arms	2 out of 10
On-time mean difference	Average increase in on-time under $H = 1$	4.5 hours
Reduction in pain mean difference	Average reduction in pain symptoms on a rating scale from 0 to 10 under $H = 1$	0.83 out of 10
Cognitive symptom mean difference	Average improvement in cognitive symptoms on a rating scale from 0 to 10 under $H = 1$	0.53 out of 10
Depression risk	Increased depression risk from receiving treatment under $H = 1$	3%
Brain-bleed risk	Increased brain-bleed risk from receiving treatment under $H = 1$	1.6%
Mortality risk	Increased mortality risk from receiving treatment under $H = 1$	0.8%
Patient accrual rate (η)	The rate at which patients enroll in the trial. We assume a uniform accrual rate over time.	200 patients per year
Start-up time before patient enrollment (s)	The time before the trial starts needed for paperwork, etc.	6 months
Follow-up period after enrolling the last patient (f)	After the last patient is enrolled, patients are followed up for this amount of time before any data analysis is conducted.	1 year
FDA review time (τ)	The time required for the FDA to review an application and make an approval decision.	9 months

As can be seen, the optimal RCT designs vary substantially across patient subgroups. On the one hand, for patients who had previously received DBS, the sample sizes range from 110 to 154 subjects, and the significance level falls between 4.0% and 10.0%. Within this range, patients with more severe motor and cognitive symptoms prefer less conservative trial designs. This result reflects the willingness of these relatively more risk-tolerant patients to bear additional uncertainty regarding the effectiveness of a device in order to reduce the chance of missing out on its potential benefits. Moreover, the BDA framework also allows the design to grant faster access to this risk-tolerant subgroup, as a result of smaller clinical trials.

On the other hand, for the average patient who has never received DBS, the benefits of an implantable device do not significantly outweigh the risks, resulting in

clinical trials with much larger sample sizes (150 to 294 subjects) and more conservative significance levels (0.2% to 4.4%). For these patients, the additional risks of depression, brain bleed, and mortality associated with an implantable device weigh relatively heavier on their decision-making process. Here, more evidence of clinical effectiveness is required to outweigh the potential consequences of a false approval. It is important to note that [216] reported an improvement in on-time and motor function at a statistical significance of $p < 0.001$ for bilateral deep brain stimulation when compared against the standard of care, so even our most stringent threshold would not negate prior approvals.

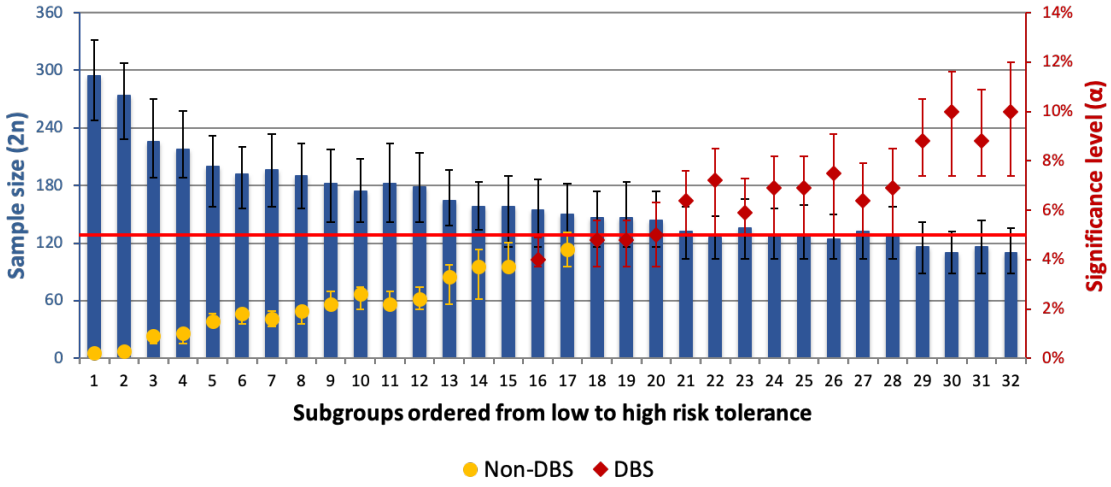


Figure 3-1: Patient-centered sample sizes (2n) and significance levels for patient subpopulations with increasing risk tolerance.

Table 3.4: BDA-optimal RCTs for Parkinson's devices for patient subpopulations with increasing risk tolerance. Indicator variables are set to 1 when a characteristic is present.

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
1	0	0	0	0	0	0.06	17.2%	294	0.2%	90%
2	0	0	0	0	1	0.08	18.8%	274	0.3%	90%
3	0	0	0	1	0	0.28	15.9%	226	0.9%	90%
4	0	0	0	1	1	0.30	17.2%	218	1.0%	90%
5	0	1	0	0	0	0.40	19.6%	200	1.5%	90%
6	0	1	0	0	1	0.42	21.6%	192	1.8%	90%
7	0	0	1	0	0	0.46	17.7%	196	1.6%	90%
8	0	0	1	0	1	0.48	19.3%	190	1.9%	90%
9	0	1	0	1	0	0.62	18.0%	182	2.2%	90%
10	0	1	0	1	1	0.64	19.6%	174	2.6%	90%
11	0	0	1	1	0	0.68	16.3%	182	2.2%	90%
12	0	0	1	1	1	0.70	17.7%	178	2.4%	90%
13	0	1	1	0	0	0.80	20.2%	164	3.3%	90%
14	0	1	1	0	1	0.82	22.3%	158	3.7%	90%
15	0	1	1	1	0	1.02	18.4%	158	3.7%	90%
16	1	0	0	0	0	1.03	19.6%	154	4.0%	90%
17	0	1	1	1	1	1.04	20.2%	150	4.4%	90%
18	1	0	0	0	1	1.05	21.5%	146	4.8%	90%
19	1	0	0	1	0	1.25	17.9%	146	4.8%	90%
20	1	0	0	1	1	1.27	19.6%	144	5.0%	90%
21	1	1	0	0	0	1.37	22.7%	132	6.4%	90%
22	1	1	0	0	1	1.39	25.4%	126	7.2%	90%

Continued on next page

Table 3.4 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
23	1	0	1	0	0	1.43	20.1%	136	5.9%	90%
24	1	0	1	0	1	1.46	22.2%	128	6.9%	90%
25	1	1	0	1	0	1.59	20.5%	128	6.9%	90%
26	1	1	0	1	1	1.61	22.7%	124	7.5%	90%
27	1	0	1	1	0	1.65	18.4%	132	6.4%	90%
28	1	0	1	1	1	1.67	20.1%	128	6.9%	90%
29	1	1	1	0	0	1.77	23.5%	116	8.8%	90%
30	1	1	1	0	1	1.80	26.4%	110	10.0%	90%

3.4 Discussion

In this chapter, we modify the traditional 5% significance level used in standard RCT designs to better reflect patient preferences and their medical needs. We quantify the loss of value to public health associated with false approvals and rejections, then use BDA to determine the optimal significance level and sample size that minimize this loss, and thereby, maximize this value. We apply this framework to a case study on Parkinson's disease using patient preferences obtained from survey data, and demonstrate that a fixed statistical significance level of 5% does not necessarily maximize overall value to patients.

For patients with severe symptoms and prior experience with DBS, value is lost due to larger, lengthier clinical trials that are too conservative about the false approval rate. Here, a missed opportunity to approve a potentially beneficial yet more risky treatment has a substantial negative impact on risk-tolerant patients. Conversely, for risk-averse patients, such as those with relatively mild symptoms, the traditional significance level of 5% is more permissive than the calculated patient-centered thresholds. In these cases, patients require clear demonstration of clinical effectiveness to reduce the probability of a false approval and subsequent harm to their health.

While we have made strong assumptions in this case study for illustrative purposes, these assumptions can be readily relaxed or modified in future applications. For example, when considering potential regulatory decisions for the broader population, it is possible to aggregate the preferences of patient subgroups by prevalence, incidence rates, and other epidemiological measures within this framework. In addition, patient value models that incorporate diminishing marginal returns or present-biased time preferences can also be incorporated into the BDA model. A more detailed discussion of these and other factors can be found in [46]. We believe that a nuanced consideration of these issues will be instructive in the design of future clinical trials.

Finally, although this framework provides a systematic and quantitative method of incorporating multifaceted tradeoffs, the usefulness of its recommendations relies on the appropriateness of these assumptions on accurately calibrated model parameters. For example, patient preference information can be representative of the broader indicated population, and participant demographic and clinical characteristics across cohorts can be compared to assess their similarity. To improve the validity of these assumptions, advisory committees consisting of key stakeholder groups such as patients, caregivers, physicians, medtech and biopharma executives, regulators, and policymakers may find it useful to formulate explicit models for the consequences of certain clinical trial actions. These estimates could then be incorporated into the FDA review of clinical trial designs as additional inputs to their qualitative and quantitative deliberations. The use of an objective, explicit and transparent method that directly relates information about patient preferences to an acceptable level of uncertainty associated with clinical evidence would provide the ecosystem with a tool to appropriately size clinical investigations.

In addition to model validation, the ethics of such designs require consideration. There is a growing body of research demonstrating that patients facing severe ill-

ness tend to exhibit "unrealistic optimism" about the potential outcome of a trial, irrespective of the information provided in any consent form [130]. In these circumstances, it is all the more critical for regulators to incorporate patient preferences using a rational and systematic method into their decision-making process. Regulatory agencies and trial sponsors already take many of these behavioral considerations into account. For example, shifting away from balanced clinical trials in which there is an equal number of patients in the treatment arm and in the control arm–when there is strong scientific conviction that a therapeutic candidate is likely to succeed to allow more patients enrolled in the trial to benefit from the investigational treatment. Our proposed framework is designed to provide regulators with another tool with which to address behavioral differences associated with burden of disease—one that provides greater transparency and reproducibility than discretionary deliberations behind closed doors.

There is a broader question of whether or not consumers have the necessary expertise to express their preferences over certain therapeutic choices and counterfactuals that are part of the patient preference elicitation process. The FDA encourages device manufacturers to collect data on patient experience and perspectives, and will incorporate this data, when available, into its decision-making process. Even with the limitations in both consumer expertise and elicitation technology, some forms of patient-preference information can be measured and reflected in regulatory decisions. As with other types of decisions facing today's consumers—financial, nutritional, technological, and so on—medical decisions have grown in complexity, requiring consumers to become better informed. Moreover, from an ethical perspective, it can be argued that, irrespective of a consumer's level of education, in matters concerning their physical well-being—such as volunteering for a clinical trial—the consumer's preferences have to be taken into consideration. This underlies the very foundations of informed consent as an ethical pre-requisite for human experimentation.

3.5 Conclusion

As shown in its benefit-risk guidance, the FDA currently considers a variety of factors beyond p-values when making its decisions [80, 78]. However, determining the maximum acceptable level of uncertainty for clinical evidence remains an unresolved question in regulatory science for all stakeholders. As such, the ability of the BDA framework to weigh tradeoffs that reflect a multiplicity of perspectives systematically and transparently, and from that, to calculate the optimal threshold for the significance level, makes it a potentially valuable tool for facilitating the clinical trial design process. The FDA encourages sponsors to discuss their novel clinical trial plans early through the Q-Submission process [81].

In the case of Parkinson's disease, our survey data suggests that a fixed statistical significance level of 5% does not necessarily maximize overall value to patients. Patients who have never received DBS treatment tend to present a higher level of risk-aversion which would be consistent with a lower statistical threshold, while patients who previously received DBS treatment present a higher risk-tolerance which

would be consistent with a higher statistical threshold.

Chapter 4

Extending the Bayesian Decision Analysis Framework under Risk and Uncertainty

BDA is a systematic, transparent, and repeatable process which can guide the regulatory decision-making process by incorporating the patient's benefit/risk preferences explicitly into the design of clinical trials.

In this chapter, we extend the methodology of BDA to reflect the patient's risk and uncertainty preferences when the hypothetical treatment's outcome (i.e., effectiveness or safety) is uncertain, as is usually the case when clinical trial are designed. This is done by assigning a continuous prior distribution to the treatment's outcome and expressing the BDA objective as a mean-variance optimization problem which accounts for the patient's uncertainty-aversion. The objective of the extended BDA setting is to produce an optimal significance level and sample size that minimizes the expected loss of value to patients as well as the variance of this loss.

As a proof-of-concept, we apply the extended BDA framework to previous studies and compare the outputs to those of the traditional BDA framework for a hypothetical treatment's efficacy endpoint. As expected, we find that the extended BDA setting produces optimal significance levels that are more conservative when the severity of the disease is lower and less conservative when the severity of the disease is higher, suggesting that patients may be willing to try a treatment with higher level of uncertainty around its efficacy when the disease is more severe and when there are no alternatives available. Furthermore, we investigate the sensitivity of the models to the uncertainty-aversion parameter and find the extended BDA setting (under a Gaussian prior) to be robust to uncertainties in the treatment's efficacy. Conversely, under a Bernoulli prior the BDA outputs tend to be more conservative when the severity is low and less conservative when the severity is high as we increase the uncertainty-aversion parameter.

4.1 Introduction

4.1.1 The challenge of incorporating patient preferences into the design of RCTs

RCTs are regarded as the gold standard when it comes to estimating the efficacy of a treatment. Following Sir Ronald Fisher's highly influential work [87], it has been customary to use a 5% significance level threshold when performing two-tailed hypothesis tests, and a 2.5% significance level threshold for single-tailed hypothesis tests. Although it is a convenient reference point, it is difficult to argue that the same threshold should be used in all applications. In fact, the appropriate threshold to use is, almost by definition, very closely linked to the amount of risk the stakeholders are willing to undertake. In the context of clinical trials, this means that patients may be willing, under some circumstances, to bear additional risks to gain access to a certain treatment rather than miss-out on the potential benefits this treatment can provide. Conversely, when alternative treatments are available, patients may be much less risk-tolerant and require a lower significance level threshold to assess the efficacy of the treatment under consideration.

In the U.S., there has been an increasing interest from the FDA to incorporate patient preferences into the regulatory and decision-making process. For example, the Alzheimer drug *aducanumab* (Aduhelm [Biogen]) has received an accelerated approval on June 2021 despite disagreements on the statistical efficacy of the drug and the meaningfulness of some clinical endpoints[84, 180]. Given the limited alternatives available to patients suffering from Alzheimer's disease, patient advocacy groups have been actively pushing for an approval. However, this decision resulted in a controversy that lead to the resignation of three FDA advisers [179].

4.1.2 Evidence from official FDA guidance documents

The example of *aducanumab* is interesting as it highlights both the growing importance of the patients' perspective as well as a lack of a systematic, transparent, and repeatable framework to incorporate PPI into the regulatory decision-making process. Furthermore, the FDA has published in September 2021 a draft guidance [83] to address how patient experience data can help the FDA with the benefit/risk assessment of new drugs. Similar points are addressed for medical devices in the August 2016 and 2019 final documents [78] and [80].

To directly quote the FDA, "Patients are experts in the experience of their disease or condition, and they are the ultimate stakeholders in the outcomes of medical treatment" [83] and "PPI may be best suited to inform regulatory decision-making when: 1) significant risks of treatment or uncertainty about risks exist relative to the expected benefits; 2) patients' views about the most important benefits and risks vary considerably within a population; and/or 3) when patients' views as to the most important benefits are expected to differ from those of healthcare professionals. If available, PPI would be considered within the context of FDA's assessment of the drug's efficacy and safety to the patient population, although it would not, for exam-

ple, overcome significant safety issues or lack of therapeutic benefit" [83]. However, no methodology has been officially endorsed yet by the FDA to incorporate patient preferences into the design and regulatory assessment of clinical trials.

Understanding this benefit/risk is even more important when the efficacy evidence is debatable and [129] highlights the lack of a systematic and transparent approach to resolve such issues.

4.1.3 A brief review of BDA and EVI

A quantitative decision theory method has been proposed in [105, 199] under the name of EVI to determine the optimal sample size required to satisfy a specified significance level threshold and power when designing a clinical trial. More recent surveys of the technique can be found in [3, 182, 4]. The technique has been refined through the development of the Expected Value of Sample Information (EVSI) [218] to determine the optimal significance level threshold and sample size by maximizing the expected net benefit of the trial [215, 115]. However, these approaches rely on the Bayes risk [108] and differ considerable from the classical frequentist hypothesis testing framework, which is often preferred by researchers and regulators.

As a middle-ground between a frequentist and a Bayesian approach, the BDA methodology has been proposed [155, 47, 126, 168, 43] and successfully applied to practical settings in collaboration with regulators for weight-loss devices [46], heart failure devices [44], and Parkison's disease devices [45]. The objective of the BDA is to determine the optimal trial design parameters of a hypothetical treatment by using PPI to trade off the impact of a type I error (the event of incorrectly approving an ineffective and potentially harmful treatment) with a type II error (the event of incorrectly rejecting an effective treatment) in a systematic, transparent, and reproducible way. The optimal sample size of a trial is also important to estimate to reduce the duration of clinical trials and provide patients with a faster access to the treatment if it is found to be effective. As shown in Chapter 3, BDA and EVI are related although they differ significantly in their formulation.

4.1.4 Advantages of the extended BDA framework

To design optimal trials, the classical BDA aims to maximize the expected benefit to patients, or equivalently minimize the expected loss, of a clinical trial. The PPI is carefully analyzed using Discrete Choice Experiments (DCEs) and incorporated into the loss function. As an input, the classical BDA uses the expected efficacy of the treatment under consideration, a quantity which is not known precisely at the time the clinical trial is designed. The device is therefore assumed to either be effective (with some probability p_1) and the efficacy matches the expected efficacy value, or ineffective (with probability $p_0 = 1 - p_1$) and the efficacy is zero.

We argue here that this dichotomy, although helpful as a simplifying assumption, can be refined by assuming a continuous probability distribution for the efficacy of the treatment rather than the aforementioned Bernoulli distribution. We show here that

assuming a Normal distribution for the treatment's efficacy is in fact completely consistent with the fundamental assumptions of RCTs. Furthermore, by fully embracing the uncertainty around the efficacy of the treatment, we can address simultaneously the risk preferences as well as the uncertainty-aversion of the patients. Understanding the patient's preferences regarding an uncertain efficacy (or safety) endpoint is in fact almost as important to understanding the patient's benefit/risk preferences.

The methodology we develop in this chapter aims to address these limitations by extending the BDA. While the classical BDA carefully incorporates the risk preferences of the patients into the design, it fails to include their view on potential uncertainties in the treatment's efficacy, unlike the extended BDA framework we propose here. The objective of the extended BDA is twofold: we use a mean-variance optimization framework to jointly minimize the expected loss to patients (as in the classical BDA)—while accounting for the uncertainty in the treatment's efficacy—and the variance of this loss. The variance term is important as it measures the sensitivity of the loss function to the uncertainty in the treatment's efficacy. Through the extended BDA, we propose a unified framework which is consistent with the usual assumptions of a RCT, and incorporates both the risk and uncertainty preferences of patients obtained from a DCE. As with the traditional BDA, this method is systematic, transparent, and repeatable.

4.1.5 Revisiting the case study for Parkinson's disease

As a proof-of-concept, we use a patient preference model estimated from a DCE for Parkinson's disease [113] to study a specific DBS medical device. A classical BDA framework has been used in [45] and Chapter 3 to determine optimal trial designs. Here, we apply instead the extended BDA methodology developed in this chapter and compare the results obtained under each framework. We find that while the traditional BDA tends to produce more conservative trial designs than the extended BDA framework, the latter is more robust to the uncertainty around the treatment's effectiveness as the patient's uncertainty-aversion increases. Furthermore, the patient's uncertainty-aversion tends to produce more conservative designs when the severity ratio of the disease is low, and less conservative designs when the severity ratio is high, suggesting that patients may be willing to try a treatment with higher level of uncertainty around its efficacy when the disease is more severe and when there are no alternatives available.

The remainder of the chapter is structured as follows. We present the extended BDA framework in Section 4.2, apply it to the Parkinson's patient preference study in Section 4.3, and discuss the advantages and limitations of this method in Section 4.4. Appendix C.1 covers omitted proofs, a theoretical toy model, as well as a discussion of in-trial costs and average statistical power. Finally, we include in Appendix C.2 further results on the Parkinson's patient preference study, as well as other applications of the extended BDA framework.

4.2 Extending the BDA framework under risk and uncertainty: a tale of two exposures

4.2.1 A review of the traditional BDA framework

BDA aims to maximize the expected value associated with a RCT design with a sample size n and a critical value λ_n [126]: mathematically, the BDA objective is formulated as

$$\max_{n,\lambda_n} \mathbb{E}\left[\text{Value}; n, \lambda_n\right]. \tag{4.1}$$

This means that BDA aims to maximize the value of the RCT by first finding the optimal critical value λ_n as a function of the sample size n, and then finding the optimal sample size n. In other words, BDA aims to solve

$$(n^*, \lambda_n^*) = \max_{n} \max_{\lambda_n} \mathbb{E}\left[\text{Value}; n, \lambda_n\right]. \tag{4.2}$$

Throughout this chapter, we assume that the test statistics and critical values we define are based on an efficacy endpoint. A similar analysis can be carried out for a safety endpoint (see [40] for an example). If we assume that the efficacy θ of the treatment can be either 0 (with probability p_0) or θ_0 (with probability $p_1 = 1 - p_0$), then we can express the BDA problem [126] as:

$$\max_{n,\lambda_n} \mathbb{E}\left[\text{Value}; n, \lambda_n\right] = \max_{n,\lambda_n} \left\{p_0 \cdot \mathbb{E}\left[\text{Value}|\theta = 0\right] + p_1 \cdot \mathbb{E}\left[\text{Value}|\theta = \theta_0\right]\right\}. \tag{4.3}$$

We note that Equation 4.3 implicitly assumes that we are conditioning on the safety parameter, hence holding it constant. The maximization in Equation 4.3 is more commonly performed by minimizing the expected loss of the trial rather than by maximizing the expected value of the trial [126]. Both methods are equivalent as

$$\mathbb{E}\left[\text{Value}; n, \lambda_n\right] = \mathbb{E}\left[\text{Value}^* - \text{Loss}; n, \lambda_n\right] = \mathbb{E}\left[\text{Value}^*\right] - \mathbb{E}\left[\text{Loss}; n, \lambda_n\right], \tag{4.4}$$

where Value^* is the value created in the hypothetically optimal scenario where the correct approval decision is made immediately and without running a trial. Since $\mathbb{E}\left[\text{Value}^*\right]$ is independent of n and λ_n, Equation 4.3 is equivalent to:

$$\min_{n,\lambda_n} \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] = \min_{n,\lambda_n} \left\{p_0 \cdot \mathbb{E}\left[\text{Loss}|\theta = 0\right] + p_1 \cdot \mathbb{E}\left[\text{Loss}|\theta = \theta_0\right]\right\}. \tag{4.5}$$

4.2.2 Choosing a prior distribution for the treatment effect

More generally, we can assume that the efficacy θ of the treatment can take a continuum of values to better reflect the uncertainty in the efficacy of the hypothetical treatment. We then express the BDA problem using an integral of the expected value of the trial *conditional on* the efficacy θ of the treatment:

$$\max_{n,\lambda_n} \mathbb{E}\left[\text{Value}; n, \lambda_n\right] = \max_{n,\lambda_n} \int_{\theta} p_\theta \cdot \mathbb{E}\left[\text{Value}|\theta\right] \, d\theta, \tag{4.6}$$

where θ is used as a shorthand to denote the hypothesis that the treatment has an efficacy θ. Equivalently, we can rewrite Equation 4.6 by minimizing the expected loss of the trial rather than by maximizing the expected value of the trial:

$$\min_{n,\lambda_n} \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] = \min_{n,\lambda_n} \int_{\theta} p_\theta \cdot \mathbb{E}\left[\text{Loss}|\theta\right] \, d\theta. \tag{4.7}$$

For example, if we assume that the treatment's efficacy can only take two values i.e., $\theta = \theta_0$ with probability p_1 and $\theta = 0$ with probability p_0, then Equation 4.6 reduces to Equation 4.3.

As formally discussed in Section 4.2.4, it is natural to assume that the treatment effect θ follows a normal distribution with mean μ_θ and variance σ_θ^2

$$\theta \sim N(\mu_\theta, \sigma_\theta^2). \tag{4.8}$$

The expected efficacy of the treatment μ_θ and its variance σ_θ^2 are specified in Equation 4.16 and are usually agreed upon by the researchers and regulators. For example, we refer the reader to previous BDA applications [44, 40] in which these parameters have been calibrated either by experts with domain-specific expertise based on multiple years of first-hand knowledge and participation in regulatory review [46, 44, 40], or based on regulatory reviews of targeted devices and literature reviews of efficacy and safety [45]. A thorough sensitivity analysis is usually performed to verify the robustness of the results to perturbations in these assumptions.

We use the prior in Equation 4.8 throughout the chapter, however we also consider a Bernoulli prior distribution

$$\theta \sim \mu_\theta \cdot \text{Bernoulli}(p_1), \tag{4.9}$$

with p_1 being the probability that the device is effective and $p_0 = 1 - p_1$ the probability that the device is ineffective. This case reduces to the traditional BDA framework and is a useful baseline to consider when discussing the results obtained under a Gaussian prior (Equation 4.8). When no prior information is available, it is common to use the uninformative prior $p_0 = p_1 = 0.5$, in line with the clinical equipoise assumption [89]. In fact, the Bernoulli prior can easily be generalized by assuming that θ is a mixture of normal distributions, where $\theta \sim N(0, \sigma_\theta^2)$ with probability p_0 and $\theta \sim N(\mu_\theta, \sigma_\theta^2)$ with probability $p_1 = 1 - p_0$. The probabilities p_0 and p_1 can be assumed to be 0.5 when no prior information is available, however the expected treatment effect conditional on the treatment being effective (i.e., μ_θ) and the variance σ_θ^2 can be agreed upon by the researchers and regulators as discussed above.

4.2.3 Incorporating uncertainty aversion via mean-variance optimization

Furthermore, the objective function of the BDA in Equation 4.1 (or equivalently Equation 4.7) can be generalized. In fact, maximizing the expected value of the trial (or minimizing the expected loss of the trial) highlights the benefit-risk tradeoff of

the treatment as perceived by patients. Additionally, we can consider the uncertainty around the benefit-risk tradeoff by accounting for the variance of the value (or loss) of the trial. More concretely, we can solve for:

$$(n^*, \lambda_n^*) = \arg\max_{n,\lambda_n} \left\{ \mathbb{E}\left[\text{Value}; n, \lambda_n\right] - \frac{1}{2} \cdot q \cdot \text{Var}\left(\text{Value}; n, \lambda_n\right) \right\}, \tag{4.10}$$

or equivalently,

$$(n^*, \lambda_n^*) = \arg\min_{n,\lambda_n} \left\{ \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] + \frac{1}{2} \cdot q \cdot \text{Var}\left(\text{Loss}; n, \lambda_n\right) \right\}, \tag{4.11}$$

where the parameter q is interpreted as the patient's *uncertainty aversion*. We impose an upper bound constraint $P_{\max} = 90\%$ on the statistical power to represent practical considerations of the industry, reflecting budget, time, and personnel constraints, as well as other resource limitations, ensuring that

$$\text{Power} \leq P_{\max}. \tag{4.12}$$

The statistical power turns out to be more complex to define when we assume a continuous prior distribution on θ instead of a Bernoulli prior. We say that a treatment is effective treatment if $\theta \geq \mu_\theta$ (as opposed to $\theta = \mu_\theta$ under a Bernoulli prior) and define the *average power* of the clinical trial by

$$\overline{\text{Power}} := \mathbb{P}\left(\text{Approve Treatment} | \theta \geq \mu_\theta\right). \tag{4.13}$$

We therefore impose the following constraint in Equation 4.11:

$$\overline{\text{Power}} \leq P_{\max}. \tag{4.14}$$

A detailed discussion of the average power is provided in Appendix C.1.4.

Intuitively, if the loss function is more sensitive to the uncertainty in the treatment effect and to the patient's benefit-risk preferences, we should expect a more conservative type I error rate α and a larger trial sample size n to gain enough confidence in the statistical procedure. The patient's uncertainty-aversion parameter q can be estimated using DCEs in the same way as the patient's discount rate r is estimated through time-tradeoff questions [45] and Chapter 3.

In fact, we will observe that, when the severity of the disease (as defined in [126] by the ratio of losses L_1/L_0) is "small" i.e., when the relative loss per patient is higher for a type I error than for a type II error, the variance term tends to decrease as we increase the critical value λ, and hence the BDA would favor solutions with high values of λ (i.e., small values of α), making the decision process more conservative. This decrease is more pronounced under a Bernoulli prior than under a Gaussian prior (which leads to more robust BDA outputs). Conversely, when the severity is "large", the variance term tends to increase with λ, which induces the BDA to choose smaller values of λ (i.e., larger values of α), making the decision process less conservative.

In other words, the variance term tends to amplify the effect of the expected loss. For example, a pancreatic cancer patient that has exhausted the standard of care may be willing to accept a higher type I error rate (estimated by [126] to be 27.8% under the traditional BDA framework) for a treatment despite the uncertainty around its expected efficacy while a diabetes patient may require a lower type I error rate (estimated by [126] to be 1.7% under the traditional BDA framework) as treatments that can help enhance the quality of life of patients with diabetes already exist. In the latter case, the cost of a type I error is much higher than that of a type II error and makes the patient less willing to bear additional levels of uncertainty, while the terminal nature of pancreatic cancer makes type II error much more costly than type I errors. A formal discussion is available in Appendix C.1.2.

In other words, higher values of the uncertainty-aversion parameter q in the mean-variance framework allow the decision-maker to be more conservative when the severity of the disease is high, and less conservative (under a Bernoulli prior) or as conservative (under a Gaussian prior) when the severity of the disease is low.

4.2.4 A toy model for the approval decision

The response variables observed in the RCT's control arm are denoted by $\{P_i^C\}_{i=1}^n$ and are assumed to be I.I.D and follow a normal distribution with mean μ_c and variance σ_c^2. Similarly, the response variables observed in the RCT's treatment arm are denoted by $\{P_i^T\}_{i=1}^n$ and are assumed to be I.I.D and follow a normal distribution with mean μ_t and variance σ_t^2. The treatment effect is defined as the difference between the treatment group's response variable and the control group's response variable:

$$\theta = P^T - P^C \sim N(\mu_\theta, \sigma_\theta^2), \tag{4.15}$$

where

$$\mu_\theta = \mu_t - \mu_c, \quad \text{and} \quad \sigma_\theta^2 = \sigma_c^2 + \sigma_t^2. \tag{4.16}$$

Equation 4.15 justifies the assumption we made earlier regarding the choice of prior distribution for the treatment effect (Equation 4.8). The parameters μ_c, μ_t, σ_c, and σ_t are usually agreed upon by researchers and regulators as discussed in Section 4.2.2, and this is enough to specify the prior distribution for θ.

To estimate the expected treatment effect, we define the following sufficient statistic:

$$T_n = \frac{\sqrt{I_n}}{n} \sum_{i=1}^{n} (P_i^T - P_i^C), \tag{4.17}$$

where I_n is defined as the *information in the trial*

$$I_n = \frac{n}{\sigma_c^2 + \sigma_t^2} = \frac{n}{\sigma_\theta^2}. \tag{4.18}$$

In [126], the sufficient statistic T_n is assumed to follow a normal distribution (condi-

tional on the treatment effect θ):

$$(T_n|\theta) \sim N\left(\theta\sqrt{I_n}, 1\right). \tag{4.19}$$

However, this assumption can be relaxed to a noncentral t-distribution with noncentrality parameter $\theta \cdot \sqrt{I_n}$ and $2 \cdot (n-1)$ degrees of freedom [46, 45].

Furthermore, the patient incurs a cost $c_{\theta,app}$ when a treatment with efficacy θ is approved and a cost $c_{\theta,rej}$ when a treatment with efficacy θ is not approved. We assume that a treatment is approved only when the observed treatment effect T_n is above the threshold λ_n calculated by the BDA. We recognize that this is a simplifying assumption and that, in practice, regulatory agencies take into account various contextual factors other than statistical significance when deciding to approve a drug (e.g., clinical significance, secondary endpoints, etc.).

4.2.5 Modeling the expected loss per patient

Given a treatment effect θ and an observation T_n of the trial's treatment effect, the trial's loss can be formulated as

$$\begin{aligned} L(\theta, T_n) &= \mathbb{1}\{\text{Treatment Approved}\} \cdot c_{\theta,app} + \mathbb{1}\{\text{Treatment Not Approved}\} \cdot c_{\theta,app} && (4.20)\\ &= \mathbb{1}\{T_n > \lambda_n|\theta\} \cdot c_{\theta,app} + (1 - \mathbb{1}\{T_n > \lambda_n|\theta\}) \cdot c_{\theta,rej} && (4.21)\\ &= \mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}. && (4.22) \end{aligned}$$

It is also possible to consider the "in-trial" loss $L^{in}(\theta)$ which represents the loss incurred by patients enrolled in the clinical trial and only depends on the treatment effect θ (the in-trial loss is independent of T_n because the approval or non-approval of the treatment has no impact on the patient's loss during the clinical trial period). In this case, the loss function considered becomes:

$$\text{Loss} = L(\theta, T_n) + \omega \cdot L^{in}(\theta), \tag{4.23}$$

for some scaling factor ω. In-trial costs are discussed in appendix (Section C.1.3) as we do not include them in the main analysis of the chapter. In fact, in-trial costs become very important to consider when the prevalence of the disease is small and that the number of patients enrolled in the trial constitute a significant fraction of the affected population. For example, this would be the case for a rare pediatric disease[56].

The expected value and variance of the trial's loss can be calculated as follows:

$$\begin{aligned} \mathbb{E}\left[L(\theta, T_n)\right] &= \int_\theta p_\theta \cdot \left(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\right) d\theta, && (4.24)\\ \text{Var}\left(L(\theta, T_n)\right) &= \int_\theta p_\theta \cdot \left(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c_{\theta,app}^2 - c_{\theta,rej}^2) + c_{\theta,rej}^2\right) d\theta && (4.25)\\ &\quad - \left(\int_\theta p_\theta \cdot \left(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\right) d\theta\right)^2 && (4.26) \end{aligned}$$

The proof is given in Appendix C.1.1.

4.2.6 Estimating the patient's loss

When a treatment with efficacy θ is approved, the patient incurs a cost $c_{\theta,app}$. Conversely, the patient incurs a cost $c_{\theta,rej}$ when a treatment with efficacy θ is not approved.

The goal of this section is to construct the cost functions $c_{\theta,app}$ and $c_{\theta,rej}$. One key feature is that the cost functions must depend on the treatment effect θ. Indeed, researchers and regulators do not know in advance the exact treatment effect of the hypothetical treatment. It is therefore crucial to vary the hypothetical effectiveness of the treatment in the PPI surveys to query patient preferences over a wide range of potential treatment effects θ, as shown in [113]. Imposing a prior distribution on θ (Equation 4.8) allows us to appropriately reflect the likelihood of these different scenarios.

We start by defining L_0 as the relative loss of value per person from using an ineffective treatment with treatment effect $\theta = 0$. Similarly, we define $L_1(\theta)$ as the relative loss in value per person from missing out on the opportunity to use a treatment with treatment effect θ. Given a treatment effect θ, the severity ratio $L_1(\theta)/L_0$ provides a measure of their relative importance.

When the treatment is ineffective (i.e., when $\theta \leq 0$), the average loss of value per patient of a rejection is equal to

$$c_{0,rej} = 0, \tag{4.27}$$

while the average loss of value per patient of an approval is equal to

$$c_{0,app} = DF_t(\theta) \cdot L_0, \tag{4.28}$$

where $DF_t(\theta)$ is the patient's discount factor which accounts for the wait time t caused by the regulatory review process of a treatment with a treatment effect θ. The discount factor $DF_t(\theta)$ decreases from 1 to 0 as the wait time increases. We provide an explicit formula for $DF_t(\theta)$ in Equation 4.39.

When the treatment is effective and has a treatment effect $\theta \geq \mu_\theta$, the average loss of value per patient of a rejection is equal to

$$c_{\theta,rej} = L_1(\theta), \tag{4.29}$$

while the average loss of value per patient of an approval is equal to

$$c_{\theta,app} = DF_t(\theta) \cdot L_1(\theta). \tag{4.30}$$

When the treatment effect θ lies in the interval $(0, \mu_\theta)$, we define the loss of value per patient using the following linear interpolations: the average loss of value per patient of a rejection is equal to

$$c_{\theta,rej} = L_1(\theta) \cdot \frac{\theta}{\mu_\theta}, \tag{4.31}$$

while the average loss of value per patient of an approval is equal to

$$c_{\theta,app} = L_0 \cdot e^{-rt} \cdot \left(1 - \frac{\theta}{\mu_\theta}\right) + L_1(\theta) \cdot \left(1 - e^{-rt}\right) \cdot \frac{\theta}{\mu_\theta}. \tag{4.32}$$

The relative loss of value per person under each scenario is summarized in Table 4.1.

Efficacy θ	Not Approved	Approved
$\theta \leq 0$	0	$DF_t(\theta) \cdot L_0$
$0 < \theta < \mu_\theta$	$L_1(\theta) \cdot \frac{\theta}{\mu_\theta}$	$L_0 \cdot e^{-rt} \cdot \left(1 - \frac{\theta}{\mu_\theta}\right) + L_1(\theta) \cdot \left(1 - e^{-rt}\right) \cdot \frac{\theta}{\mu_\theta}$
$\theta \geq \mu_\theta$	$L_1(\theta)$	$DF_t(\theta) \cdot L_1(\theta)$

Table 4.1: Loss in value per patient associated with a balanced fixed-sample RCT as a function of the treatment effect θ.

Estimating the loss in value. The loss L_0 can be viewed as the net risk of using the treatment and can be agreed upon by the researchers and regulators (e.g., [45, 46, 44, 40]). The loss $L_1(\theta)$ can be estimated using the patient survey results from the DCE. In particular, $L_1(\theta)$ can be viewed as the difference between the *risk threshold* $R(\theta)$ of using a treatment with treatment effect θ and L_0

$$L_1(\theta) = R(\theta) - L_0. \tag{4.33}$$

For example, as detailed in [45, 113] and Chapter 3, we can use a threshold technique to estimate the following patient preference model:

$$R(\theta) = \beta \cdot \theta + [\text{Covariates}] + \varepsilon, \tag{4.34}$$

where $R(\theta)$ can be viewed as the additional risk the patient would be willing to accept before becoming indifferent to receiving a treatment with treatment effect θ. In general, the coefficient β will be positive as individuals will tend to be willing to take more risks as the treatment effect increases. It is often preferable to avoid extrapolating $L_1(\theta)$ beyond the range of θ available in the data and set $L_1(\theta) = L_1(\theta_{\max})$ for all $\theta \geq \theta_{\max}$, where $\theta_{\max}$ is the highest value of θ available in the survey data.

If no regression data is available to estimate $L_1(\theta)$, as in [126, 168, 46], then we can use a linear interpolation between $L_1(0) = 0$ and $L_1(\mu_\theta) = L_{\mu_\theta}$ which is known. In this case,

$$L_1(\theta) = L_{\mu_\theta} \cdot \frac{\theta}{\mu_\theta}, \tag{4.35}$$

and we define

$$c_{\theta,rej} = \begin{cases} 0, & \theta \leq 0, \\ L_1(\theta) \cdot \frac{\theta}{\mu_\theta}, & 0 < \theta < \mu_\theta, \\ L_{\mu_\theta}, & \theta \geq \mu_\theta, \end{cases} \tag{4.36}$$

$$c_{\theta,app} = \begin{cases} DF_t(\theta) \cdot L_0, & \theta \leq 0, \\ L_0 \cdot e^{-rt} \cdot \left(1 - \frac{\theta}{\mu_\theta}\right) + L_1(\theta) \cdot \left(1 - e^{-rt}\right) \cdot \frac{\theta}{\mu_\theta}, & 0 < \theta < \mu_\theta, \\ DF_t(\theta) \cdot L_{\mu_\theta}, & \theta \geq \mu_\theta. \end{cases} \tag{4.37}$$

The only difference here is that we do not extrapolate beyond μ_θ and cap the costs $c_{\theta,rej}$ and $c_{\theta,app}$ at L_{μ_θ} and $DF_t(\theta) \cdot L_{\mu_\theta}$ respectively.

Estimating the patient's time horizon preferences. To account for the patient's time horizon preferences, we define a discount factor $DF_t(\theta)$ which decreases from 1 to 0 as we increase the wait time t caused by the regulatory process of a treatment with treatment effect θ. This reflects the fact that patients have a lower preference for effective treatments that are not accessible immediately. As in [46, 45], we define the wait time t as

$$t = s + \frac{2n}{\eta} + f + \tau, \tag{4.38}$$

where s is the time required to setup the study, η is the patient accrual rate for the study, f is the follow-up time of the final patient to complete the study, and τ is the FDA review time. Typical values [46, 45] are $s = 6$ months (or 1 year in [168]), $\eta = 100$ patients per year, $f = 12$ months, and $\tau = 9$ months.

In particular, following [46, 45], we want a discount factor of the form $DF_t(0) = e^{-rt}$ when the treatment is not effective and $DF_t(\mu_\theta) = 1 - e^{-rt}$ when the treatment has a treatment effect μ_θ (where r is the patient's temporal discount rate). A simple linear interpolation bounded between 0 and 1 produces the following definition:

$$DF_t(\theta) = \min\left\{\max\left[e^{-rt} \cdot \left(1 - \frac{\theta}{\mu_\theta}\right) + \left(1 - e^{-rt}\right) \cdot \frac{\theta}{\mu_\theta}, 0\right], 1\right\}. \tag{4.39}$$

Note that the term $\frac{\theta}{\mu_\theta}$ should also be restricted to the $[0, 1]$ interval in Equation 4.39 through the transformation

$$\frac{\theta}{\mu_\theta} \mapsto \min\left\{\max\left[\frac{\theta}{\mu_\theta}, 0\right], 1\right\}. \tag{4.40}$$

The temporal discount rate r can be estimated using the patient survey results from the DCE. For example, as detailed in [45, 113] and Chapter 3, we can use a threshold technique to estimate the following patient preference model:

$$\text{Wait time threshold} = \frac{1}{r} \cdot \log(\mu_\theta) + \varepsilon, \tag{4.41}$$

where the *Wait time threshold* represents the additional time the patient would be willing to wait before becoming indifferent to receiving a treatment with treatment effect $\theta = \mu_\theta$ in the future versus the control now.

4.2.7 Estimating standard errors for the BDA

When $L_1(\theta)$ is estimated from a threshold regression of the form of Equation 4.34, we obtain an estimate $\hat{\beta}$ of the coefficient β as well as standard errors σ_β^2. Based on the linear regression model, $\hat{\beta}$ follows a normal distribution with mean β_0 and variance σ_β^2:

$$\hat{\beta} \sim N(\beta_0, \sigma_\beta^2). \tag{4.42}$$

We should highlight here that the variance term σ_β^2 is a standard error that reflects the precision with which the patient preference model (Equation 4.34) is being estimated. Similarly, the discount rate r is estimated from a threshold regression of of the form of Equation 4.41: the estimate $\hat{r}$ has standard errors σ_r^2 and follows a normal distribution with mean r_0 and variance σ_r^2:

$$\hat{r} \sim N(r_0, \sigma_r^2). \tag{4.43}$$

Equations 4.42 and 4.43 allow us to use a bootstrap procedure to estimate a probabilistic distribution for the optimal BDA parameters (n^*, λ_n^*). When $L_1(\theta)$ is not estimated using a threshold regression (as in [126], see Appendix C.2.2 for a detailed treatment), we directly use the known standard deviation estimate σ_{L_1} of L_1 to generate bootstrap samples distributes as

$$\hat{L}_1 \sim N(L_1, \sigma_{L_1}^2). \tag{4.44}$$

Hence, we either generate 100 bootstrap samples for $\hat{\beta}$ and $\hat{r}$ or directly generate 100 bootstrap samples for $\hat{L}_1$ and run the BDA optimization procedure (Equation 4.11) for each bootstrapped value $\hat{\beta}$ and $\hat{r}$. We then produce a distribution for the optimal BDA parameters (n^*, λ_n^*) and calculate the corresponding standard errors on our estimates of the optimal sample size n^* and the optimal critical value λ_n^* (or equivalently α^*).

4.3 Results

In this section, we apply the extended BDA approach under the mean-variance framework to the DBS study conducted in [113]. A patient preference model has already been analyzed in Chapter 3 using a classical BDA framework i.e., using a Bernoulli prior and an uncertainty aversion coefficient $q = 0$. Other applications of the mean-variance BDA framework are described in Appendix C.2.

We use Table 3.3 to calibrate the parameters of the BDA. We assume that the average motor symptoms score decreases by 0.98 units (on a scale from 0 to 10) if the treatment is effective. The treatment response variability in the control and

treatment arm is assumed to be of 2 units (on a scale from 0 to 10). We summarize these assumptions in Table 4.2.

Parameter	Control Group	Treatment Group
Response Variable (out of 10)	$\mu_c = 0.16$	$\mu_t = 1.14$
Variability in the Response Variable (out of 10)	$\sigma_c = 2$	$\sigma_t = 2$
Average treatment effect (out of 10)	$\mu_\theta = 0.98$	
Variability in the treatment effect (out of 10)	$\sigma_\theta = 2.83$	

Table 4.2: Assumptions used for the BDA application to the DBS study. These values are based on Table 3.3.

Although our main analysis assumes a Gaussian prior distribution, we first perform the mean-variance analysis under a Bernoulli prior to compare it to the classical BDA framework used in Chapter 3. The Bernoulli prior assumes that there is a 50% probability that the treatment is effective while the Gaussian prior assumes the treatment effect to follow a Normal distribution with mean $\mu_\theta = 0.98$ and variance $\sigma_\theta = 8$.

Here, we estimate the optimal BDA type I error threshold α and trial size $2n$ for various population subsets across different age cohorts (which we weight equally). These patient subpopulations are sorted by their corresponding severity ratio to better represent patients who have previously received DBS, patients with severe movement symptoms, patients who present difficulty walking, thinking clearly, and/or experience dyskinesia as a side-effect of their Parkinson's medication.

To better understand the role played by the uncertainty-aversion parameter q, we run simulations for different values of q ranging from 0 to 2 in increments of 0.10. We then compare the results with the intuition derived in Appendix C.1.2 for a simple toy model.

We first assume a Bernoulli prior and present in Figure 4-1 and Figure 4-2 the estimated optimal BDA type I error threshold α and trial size $2n$ across the different population subsets studied for different values of uncertainty-aversion parameter q. We then repeat the procedure under a Gaussian prior and present the results in Figure 4-3 and Figure 4-4.

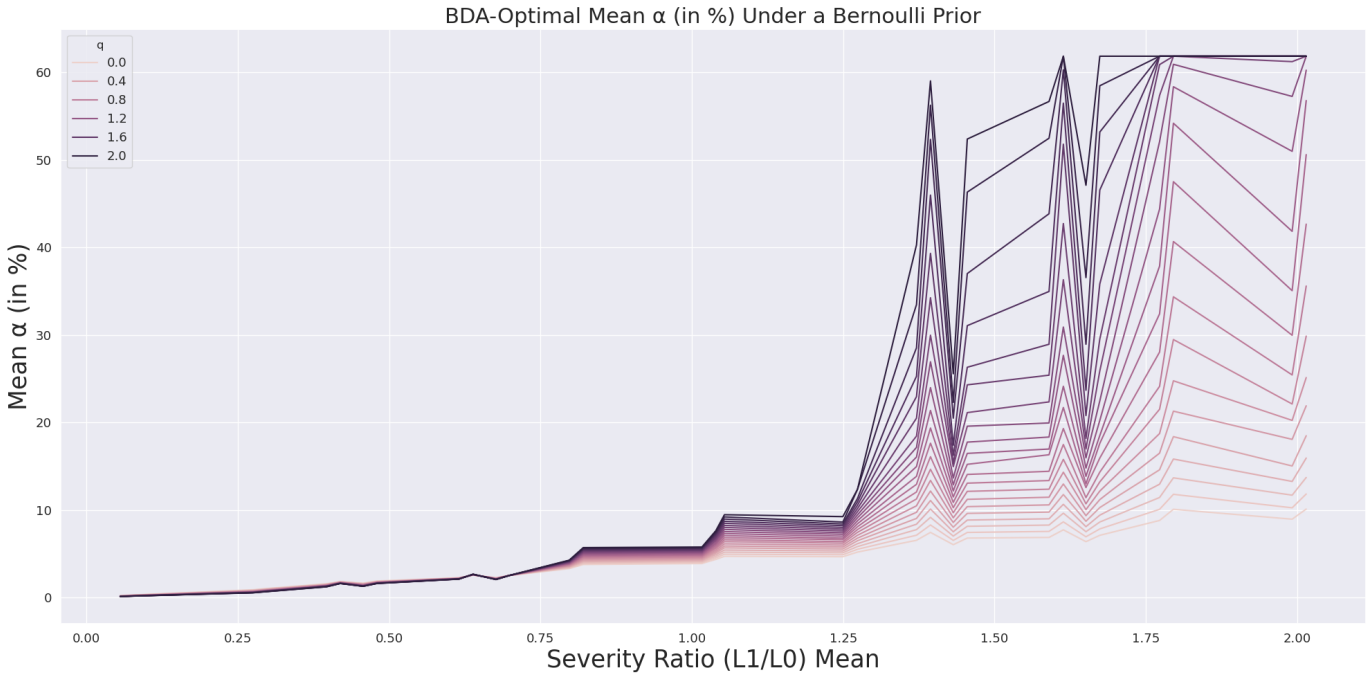


Figure 4-1: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Bernoulli prior.

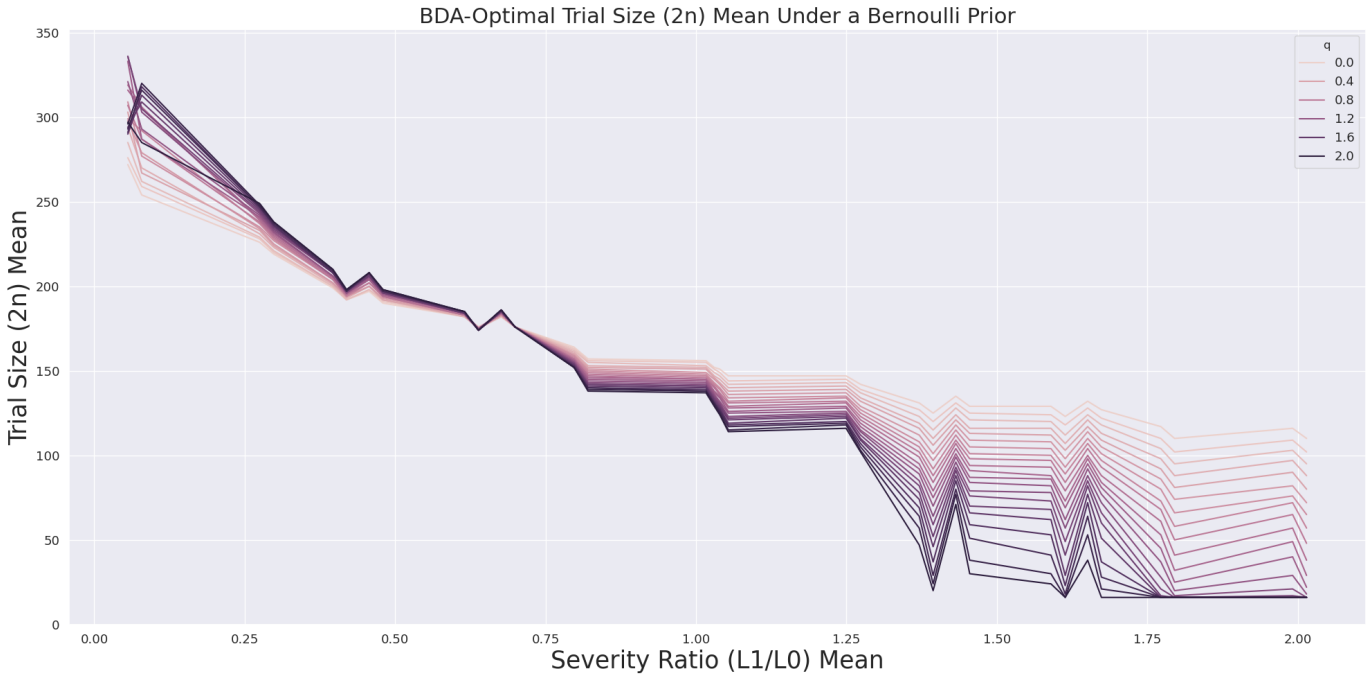


Figure 4-2: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Bernoulli prior.

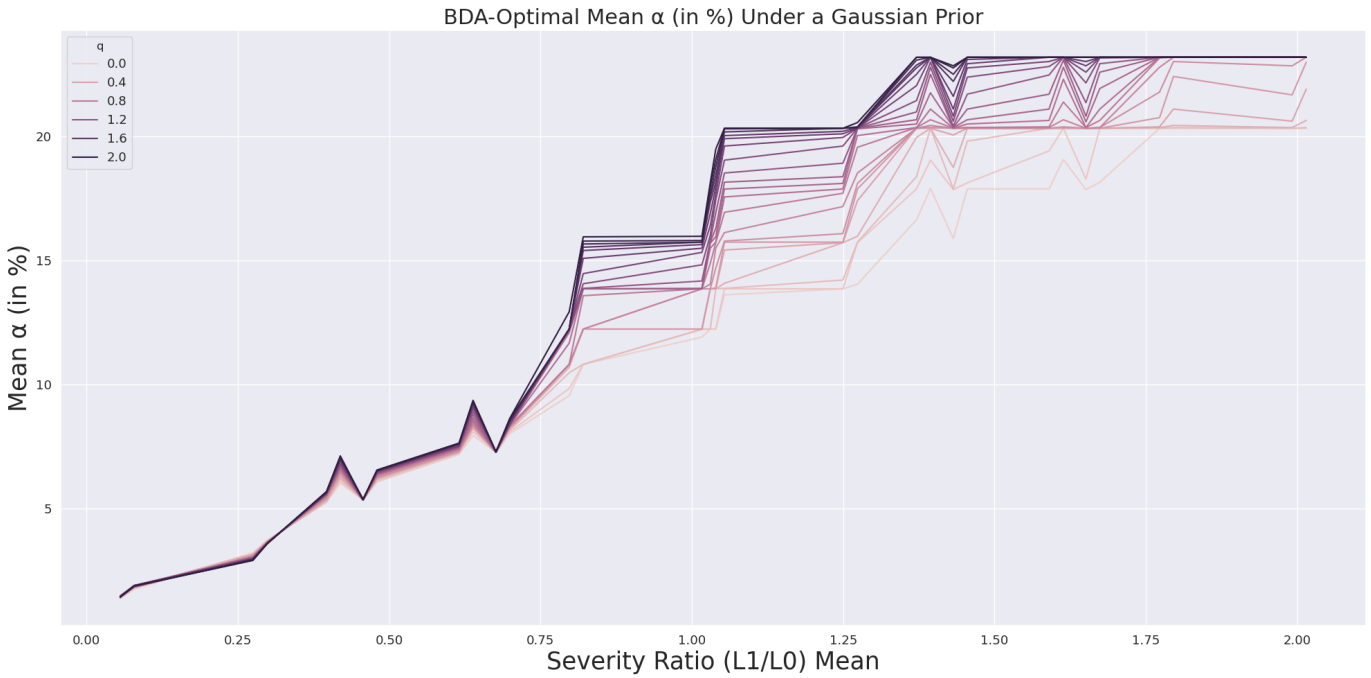


Figure 4-3: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Gaussian prior.

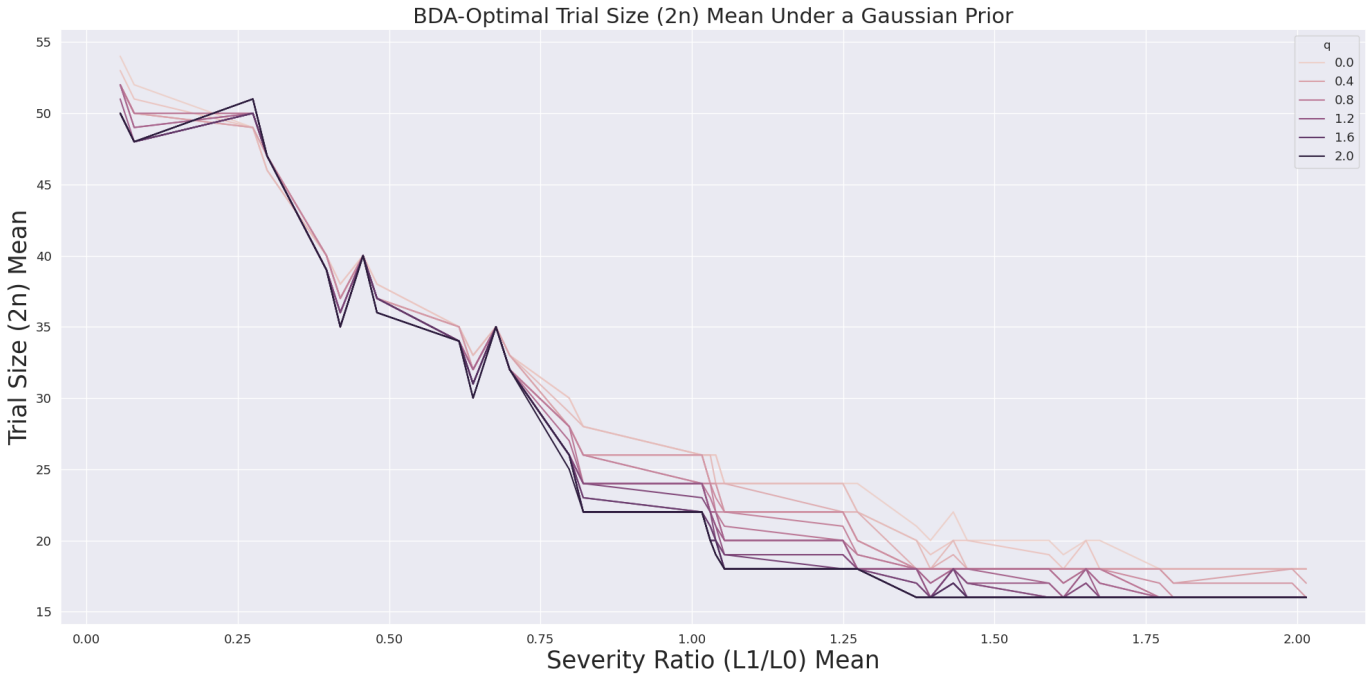


Figure 4-4: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Gaussian prior.

The case of $q = 0$ under a Bernoulli prior (Figure 4-1, Figure 4-2, and Table C.2) corresponds to the traditional BDA framework used in Chapter 3 and constitutes our baseline model. As suggested by the toy model described in Appendix C.1.2, the optimal significance level α decreases as we increase the uncertainty-aversion parameter q when the severity ratio is roughly below 0.6, leading to a more conservative decision. Similarly, the optimal trial size $2n$ increases with the uncertainty-aversion parameter q when the severity ratio is roughly below 0.6, which is also consistent with a more conservative decision. In other words, when the severity ratio is smaller, the uncertainty aversion parameter induces the BDA to design trials with a larger sample size and a smaller significance threshold because more effectiveness evidence is required to compensate for the uncertainty associated with an incorrect approval of the treatment.

Conversely, when the severity ratio is above 0.8, the optimal significance level α increases while the optimal trial size $2n$ decreases as we increase the uncertainty-aversion parameter q. Hence, when the severity ratio is higher, the uncertainty-aversion parameter induces the BDA to design trials with smaller sample sizes and a tolerance for a larger significance threshold, leading to less conservative decisions. This reflects a higher risk-tolerance from patients within these subgroups as the potential risk of missing-out on the potential benefits of the treatment outweighs the potential risks of an incorrect approval.

The same observations apply when we assume a Gaussian prior for higher values of severity ratio (Figure 4-3, Figure 4-4, and Table C.2): higher uncertainty-aversion coefficients lead to less conservative decisions when the severity ratio is larger. However, when the severity ratio is small (below 0.6), the optimal BDA outputs are much less sensitive to the uncertainty aversion parameter under a Gaussian prior than under a Bernoulli prior.

4.4 Discussion

Comparison of the Extended BDA Framework to the Classical BDA Framework. The classical BDA framework assumes a Bernoulli prior with an uncertainty-aversion parameter $q = 0$ while the extended BDA framework allows the treatment effect of the treatment to follow a Normal distribution. It is helpful to compare the results under a Bernoulli prior and a Gaussian prior to better understand the impact of modeling the inherent uncertainty in the treatment effect. Indeed, under a Bernoulli prior, we assume the treatment effect to be either 0 or μ_θ while we assume a continuum of possibilities under the Gaussian prior as the treatment effect follows a Normal distribution with mean μ_θ and variance σ_θ^2. The normality assumption is not only more realistic (as the treatment effect of the treatment is never known for certain in advance at the time the clinical trial is designed), but it is also entirely consistent with the assumption that the response variables in the control and treatment arms are I.I.D and normally distributed. Furthermore, the PPI conducted usually queries patients about their risk tolerance relative to a hypothetical treatment by varying the effectiveness and riskiness of the treatment. Hence, applying a continuous prior dis-

tribution rather than a discrete prior distribution can better reflect the uncertainty in the efficacy of the hypothetical treatment and assess the appropriate risk-preferences of patients given this uncertainty.

Comparing Figure 4-1 to Figure 4-3, we immediately see that the BDA optimal significance level α tends to be higher under a Gaussian prior than under a Bernoulli prior across patient subpopulations. However, when the severity ratio is above 1.25, the optimal α becomes much higher under a Bernoulli prior than under a Gaussian prior when the uncertainty-aversion parameter q increases. In fact, the outputs we obtain under a Gaussian prior are much more robust to q than the ones we obtain under a Bernoulli prior when the severity is higher. On the other hand, Figure 4-2 and Figure 4-4 show that the optimal trial size tends to be higher under a Bernoulli prior across patient subpopulations and across choices of q. Hence, while a Bernoulli prior tends to make more conservative decisions than a Gaussian prior when the severity is below 1.25 or when the uncertainty-aversion parameter is below 0.5, a Gaussian prior leads to more robust outputs. For example, when $q = 0$ and $s = 0.06$, an optimal RCT design uses a significance level $\alpha = 0.19\%$ and a trial size $2n = 294$ under a Bernoulli prior, but a significance level $\alpha = 1.33\%$ and trial size $2n = 54$ under a Gaussian prior. Hence, the optimal RCT parameters are about 7 times more conservative in the Bernoulli case. When $q = 0$ and $s = 2.02$, the optimal RCT parameters are $\alpha = 10.0\%$ and $2n = 110$ under a Bernoulli prior, but $\alpha = 20.3\%$ and trial size $2n = 18$ under a Gaussian prior, which makes the Bernoulli case about 2 times more conservative. In fact, the optimal RCT parameters obtained under the Bernoulli and Gaussian priors differ more as we increase the severity ratio.

Furthermore, we observe that the BDA outputs tend to be less sensitive to the uncertainty-aversion parameter under a Gaussian prior, especially when the severity of the disease is high. The extended BDA framework is therefore much more robust when we use the Gaussian prior. For example, when $s = 2.02$, the BDA suggests a significance level of $\alpha = 10.0\%$ when $q = 0$ and $\alpha = 59.5\%$ when $q = 1$ under a Bernoulli prior, corresponding to a 5-fold increase, while the corresponding outputs are $\alpha = 20.3\%$ when $q = 0$ and $\alpha = 23.2\%$ when $q = 1$ under a Gaussian prior, corresponding to a 14% increase. Similarly, the optimal sample size decreases from $2n = 110$ to $2n = 18$ when $q = 0$ and $q = 1$ respectively under a Bernoulli prior, corresponding to a decrease by an order of magnitude, while the optimal sample size remains relatively stable from $2n = 18$ to $2n = 16$ when $q = 0$ and $q = 1$ respectively under a Bernoulli prior. This low sensitivity to q is a very desirable property as the uncertainty-aversion parameter would be estimated from patient preferences surveys (which may not be very precise to estimate q) and rather than assumed and be constrained by limitations of the model.

Therefore, while the extended BDA frameworks is robust under a Gaussian prior, we would only recommend using the mean-variance optimization with $q < 0.5$ under a Bernoulli prior given the high sensitivity of the BDA optimal trial designs to the uncertainty-aversion parameter q under a Bernoulli prior.

Interpretation of the Uncertainty-Aversion Parameter. The other important difference between the traditional BDA framework and the extended BDA framework is the inclusion of the uncertainty-aversion parameter q. Indeed, while the traditional BDA aims to minimize the expected loss of value to the patient of the RCT design, the extended BDA aims to jointly minimize the expected loss of value and the variance of the loss. If the loss is more sensitive to the treatment effect θ, the inclusion of q induces the BDA to select a stricter significance level α and a larger sample size $2n$ to gain more confidence in the treatment's effectiveness. The mean-variance optimization allows us to better reflect how the uncertainty around the hypothesized benefit impacts the benefit/risk tradeoff and q is interpreted as the patient's uncertainty-aversion parameter.

In other words, the extended BDA framework allows us to distinguish two types of risks. The first is the inherent risk of the treatment, which is present in the loss function $L(\theta, T_n)$ and was accounted for by the classical BDA. The second is the risk due to the choice of sensitivity of the loss function to the uncertainty in the hypothesized treatment effect. This second type of risk, estimated through the variance term in Equation 4.11, is controlled by the parameter q. It is also possible to interpret q from the perspective of the regulator to provide more transparency and objectiveness in the decision process. In this case, the BDA would only take into consideration the risk-preferences of the patients in the loss function $L(\theta, T_n)$, but would incorporate the regulator's uncertainty-aversion regarding the statistical procedure.

While the mean-variance objective is powerful and often used in the economics literature, a helpful extension of the work would be to account for potential behavioral biases of the patient. For example, behavioral economics have highlighted asymmetries in the way individuals tend to prefer avoiding losses to acquiring equivalent gains. This bias is called *loss aversion* and will be explored in Chapter 8 in the context of asset-allocation decisions in investment management. Using an objective function that better captures this asymmetry may lead to more conservative outcomes as type I errors (considered as a direct loss) would dominate over type II errors (which can be interpreted as missing out on a potential gain).

Limitations. The framework we have developed here must be qualified in several respects. First, we have confined our attention to traditional two-arm, fixed-sample RCTs for expositional simplicity. The extended BDA methodology can also be generalized to optimize key parameters in more complex trial designs with multiple arms or adaptive features. Second, the outputs of of the BDA heavily rely on the quality of the patient preference model developed, and hence on the DCE conducted. It is therefore important to ensure that the surveyed patients are well informed and that the survey avoids potential biases. Ideally, the DCE would be designed and conducted in collaboration with regulators. Finally, we should note that regulators do not rely solely on p-values to approve a treatment candidate. While p-values are an important part of the process, regulators consider a variety of other factors when making their decisions.

4.5 Conclusion

The first objective of this analysis is to address a common limitation of the traditional BDA framework by considering a continuous prior distribution for the hypothetical treatment's effectiveness. Rather than assuming the treatment can either be ineffective or effective (with a specific hypothesized efficacy), we instead assume that the treatment's effectiveness can potentially take a wide range of values. We propose a way to properly quantify the patient's risk tolerance as a function of the effectiveness of the treatment and then weight the patient's benefit/risk tradeoff by the probability distribution of the treatment's effectiveness.

The second objective is to account for the patient's aversion to the uncertainty in the treatment's effectiveness. We have modelled this by extending the traditional BDA objective to account for the variance of the patient's loss in value. This mean-variance formulation allow us to introduce the patient's uncertainty-aversion parameter, which can also be estimated through patient preferences surveys.

As a proof-of-concept, we replicated results from a patient preference study on Parkinson's Disease [113, 45] and compared the outputs obtained under the traditional BDA setting to those obtained under the extended BDA setting (other BDA studies [126, 168, 46] are discussed in Appendix C.2). Although the former tends to produce results that are more conservative than the latter, the optimal significance levels obtained are more robust to the uncertainty around the treatment's efficacy. As we increase the uncertainty-aversion parameter, the traditional and extended BDA outputs tend to be more conservative for a low severity ratio but less conservative for a higher severity ratio. Furthermore, we observed that the extended BDA outputs tend to be less sensitive to the uncertainty-aversion parameter under a Gaussian prior as compared to using a Bernoulli prior (which corresponds to the classical BDA setting under a mean-variance objective function).

As patient preferences are gaining more importance within the regulatory decision process, this new methodology can provide a robust, transparent, and systematic way to incorporate the patient's tolerance to risk and uncertainty into the design of clinical trial.

Chapter 5

Improving the Deployment of COVID-19 Vaccines under Stochastic Supply Shocks

We developed benefit-risk methods in Chapters 2, 3, and 4 that can help researchers and regulators incorporate patient preferences into the regulatory and decision-making process for specific medical devices for kidney failure and Parkinson's disease as well as for more generic treatments. In Chapters 5 and 6, we turn our attention to informing policy-making under uncertainty in healthcare. In particular, we focus on vaccines targeting EIDs, a very specific class of drugs that presents diverse sources of risks, not only on the development side, but also on the manufacturing side, and on the demand side. If the stochastic nature of the outbreak of EIDs introduce high levels of uncertainty in the demand of vaccines, a feature that often scares away big biopharmaceutical companies from developing them, it is imperative to react quickly and efficiently once an outbreak occurs.

In Chapter 6, we focus on the development of a large portfolio of 120 mRNA vaccine candidates in the preclinical stage targeting 11 EIDs. In this chapter, we specifically consider two COVID-19 vaccines, which are administered in the form of two doses three to four weeks apart, and simulate the effects of various vaccine distribution policies on the cumulative number of infections and deaths in the United States in the presence of shocks to the supply of vaccines. The cost-benefit framework we develop here allows us to compare vaccine allocation strategies in order to reduce the number of deaths and infections caused by the COVID-19 pandemic.

Our forecasts suggest that allocating more than 50% of available doses to individuals who have not received their first dose can significantly increase the number of lives saved and significantly reduce the number of COVID-19 infections. We find that a 50% allocation saves on average 33% more lives, and prevents on average 32% more infections relative to a policy that guarantees a second dose within the recommended time frame to all individuals who have already received their first dose. In fact, in the presence of supply shocks, we find that the former policy would save on average $8,793$ lives and prevents on average $607,100$ infections while the latter policy would save on average $6,609$ lives and prevents on average $460,743$ infections.

5.1 Introduction

With more than 44.7 million infections in the U.S. and 219 million worldwide, and a death toll over 721,000 in the United States and 4.55 million worldwide, the COVID-19 pandemic has profoundly altered the research agenda of the scientific community as a whole, launching an unprecedented race against the clock to develop a cure or a vaccine for the disease.

5.1.1 Existing epidemiological models for the COVID-19 pandemic

To better contain the disease, and to design more efficient policies in combating it, the United States Centers for Disease Control and Prevention (CDC) has collected and combined an ensemble of models to forecast the spread of the epidemic [191]. These models range from traditional SIR and SEIR-type models, to agent-based models, mixture models, and machine-learning models. Some models, such as the DELPHI [152] model, explicitly account for the effects of government intervention, such as the implementation of social distancing policies. These models have quickly been applied practically: for example, at the clinical level, the DELPHI model helped to reallocate ventilators and alleviate shortages [24]; similarly, at the policy level, the DELPHI was used to propose a more efficient allocation of vaccines [25, 23]. Epidemiological models have also been used to optimize the design of vaccine clinical trials, and to quantify the potential advantages of using adaptive RCTs and Human Challenge Trials (HCTs) over traditional RCTs [21, 43]. Epidemiological models and simulations have helped researchers and policymakers answer pressing questions, such as how to prioritize the delivery of vaccine across demographics and medical conditions [165], and where should vaccination clinics be located to maximize the effectiveness of the vaccination campaign [25, 23].

5.1.2 Impact of delays and shortages in the supply chain of COVID-19 Vaccines

With a rising number of infections and deaths, and the emergence of COVID-19 variants despite extended periods of lockdown, mass vaccination has become the critical pathway to alleviate the impact of the disease, as is apparent with the success of Israel's mass vaccination campaign [57]. However, producing and distributing the vaccines has become a new challenge for manufacturers. Despite promising results regarding the ability to store the Pfizer-BioNTech vaccine in standard freezers over periods of two weeks [118] rather than the initial storage constraint at $-80°$C [37], vaccine shortages and appointment cancellations [116] have followed factory shutdowns [147], production mix-ups [149], delays in shipment [32], and power outages [27, 195]. Optimizing the allocation of vaccines has become crucial not only due to the limited supply of vaccines, but also due to the fact that Pfizer-BioNTech and Moderna vaccines need to be administered twice for each individual, over a recommended time interval of

3 or 4 weeks, respectively [35]. Although supply constraints are important in the United States, they are even more binding in other regions such as Canada [32, 116], Europe [147, 32, 194], Africa [58], Latin America [134], and India [16].

An important debate has also arisen regarding the advantages of delaying the second dose to provide more first doses to susceptible individuals [193, 145, 122, 162, 207]. While doses were held back under the Trump administration in order to guarantee a second dose to individuals who have received their first dose, the Biden administration has pledged to reverse this policy and release all available doses [8]. Other countries, such as the United Kingdom and Canada, have already adopted the policy of delaying the second dose up to three months [206, 172], and Singapore is considering delaying the second dose up to 12 weeks as of May 2021 [114]. However, as Texas, Washington State, and Michigan experienced in mid-February 2021, releasing too many doses for first-time users could lead to delays for individuals eligible to receive their second dose (a "second-shot crunch") [211].

5.1.3 Should we allocate more COVID-19 vaccine doses to non-vaccinated individuals?

Researchers, medical doctors, and clinicians have provided arguments for and against delaying the second dose [135]. On the one hand, while allocating more first doses may initially slow down the spread of the infections, and ultimately reduce the number of deaths by allowing a bigger proportion of the population to have *some* immunity, it is possible that protection will degrade over time, and delaying the second dose may leave at-risk individuals inadequately protected. From a disease evolutionary perspective, partial immunization could also contribute to the selection of vaccine-resistant variants of SARS-CoV-2 [154]. This point is now even more relevant with the spread of the Delta variant, the predominant variant in the U.S. as of September 2021, which is twice as contagious as the original strain of the virus, yet only modestly decreases the effectiveness of the two mRNA vaccines considered [161]. On the other hand, clinical trial results and data from the Israeli mass vaccination campaign on the efficacy of the first dose tend to support the policy of delaying the second dose up to three months, especially when the supply of vaccines is constrained [10, 123, 213, 184].

In this work, we forecast the effect of various vaccine allocation strategies on the cumulative number of infections and deaths in the United States to quantify the impact of prioritizing first doses versus second doses. In particular, we extend the DELPHI model to account for vaccines, and use a simple model of shocks to the number of vaccines supplied to account for distributional constraints. Similar questions have recently been studied by other researchers. For example, [100, 212] recommend a second dose deferral strategy in order to vaccinate more people faster even if the single-dose efficacy decays over time. Likewise, [167], using the agent-based epidemics model developed in [203], suggest a 9-week delay for the second dose, although the results are mixed for the Pfizer-BioNTech vaccine when the efficacy of the first dose decays over time. While our analysis focuses on the United States, our recommendations can be generalized to other countries and especially those where

the supply of vaccines is heavily limited. Furthermore, the framework provided here can be reused in the event of a future pandemic to improve the allocation of vaccines and reduce the number of infections and deaths.

The remainder of the chapter is structured as follows: we present the epidemiological model used to forecast the COVID-19 outbreak from October 1st, 2020 to August 1st, 2021 in Section 5.2, as well as the model used to account for supply shocks; our forecasts are presented in Section 5.3, and the policies under investigation are compared and discussed in Section 5.4; we conclude in Section 5.5. Finally, a more detailed description of our analysis is available in Supplementary Material.

5.2 Methodology

We begin by presenting the epidemiological model used to simulate the COVID-19 pandemic, the assumptions made in our forecasts, as well as the model used to simulate the supply of vaccine under random shocks.

5.2.1 Epidemiological Model

Many epidemiological models have been proposed to forecast the spread of COVID-19 [191]. In particular, [152] proposes a novel SEIR-based model, called the DELPHI model, that explicitly accounts for the effects of government intervention. As shown in Fig 5-1, the DELPHI model categorizes individuals into eight classes: Susceptible individuals who have not been infected (S); Exposed individuals who have been infected, and are currently within the incubation period (E); Infected and contagious individuals (I), who are then categorized into the Detected Hospitalized (DH), the Detected and home-Quarantined (DQ), and the Undetected and self-quarantined (U) classes; Recovered individuals (R); and individuals Deceased from the COVID-19 (D).

As we consider two hypothetical vaccines in this study (loosely modelled after the Moderna vaccine and the Pfizer-BioNTech vaccine), we augment the DELPHI model by including five vaccination categories for each vaccine brand X used: individuals receiving their first dose who respond to the first dose ($V^r_{X,1}$ for immediate "response"), individuals receiving their first dose who do not respond to the first dose but will respond to the second dose ($V^{dr}_{X,1}$ for "delayed response"), and individuals receiving their first dose who will neither respond to the first dose nor to the second dose ($V^{nr}_{X,1}$ for "no response"); individuals who receive their second dose and respond to the vaccine ($V^r_{X,2}$); and individuals who receive their second dose and do not respond to the vaccine ($V^{nr}_{X,2}$).

We assume that the exposed individuals (E) are not yet contagious, and that recovered individuals (R) and vaccinated individuals from the $V^r_{X,2}$ group have permanent immunity to COVID-19. We further assume that the infection rate of individuals depends on a government response function (see Appendix D.1.1) which models the effects of government intervention. The dynamics of the augmented DELPHI model are available in Appendix D.1.1.

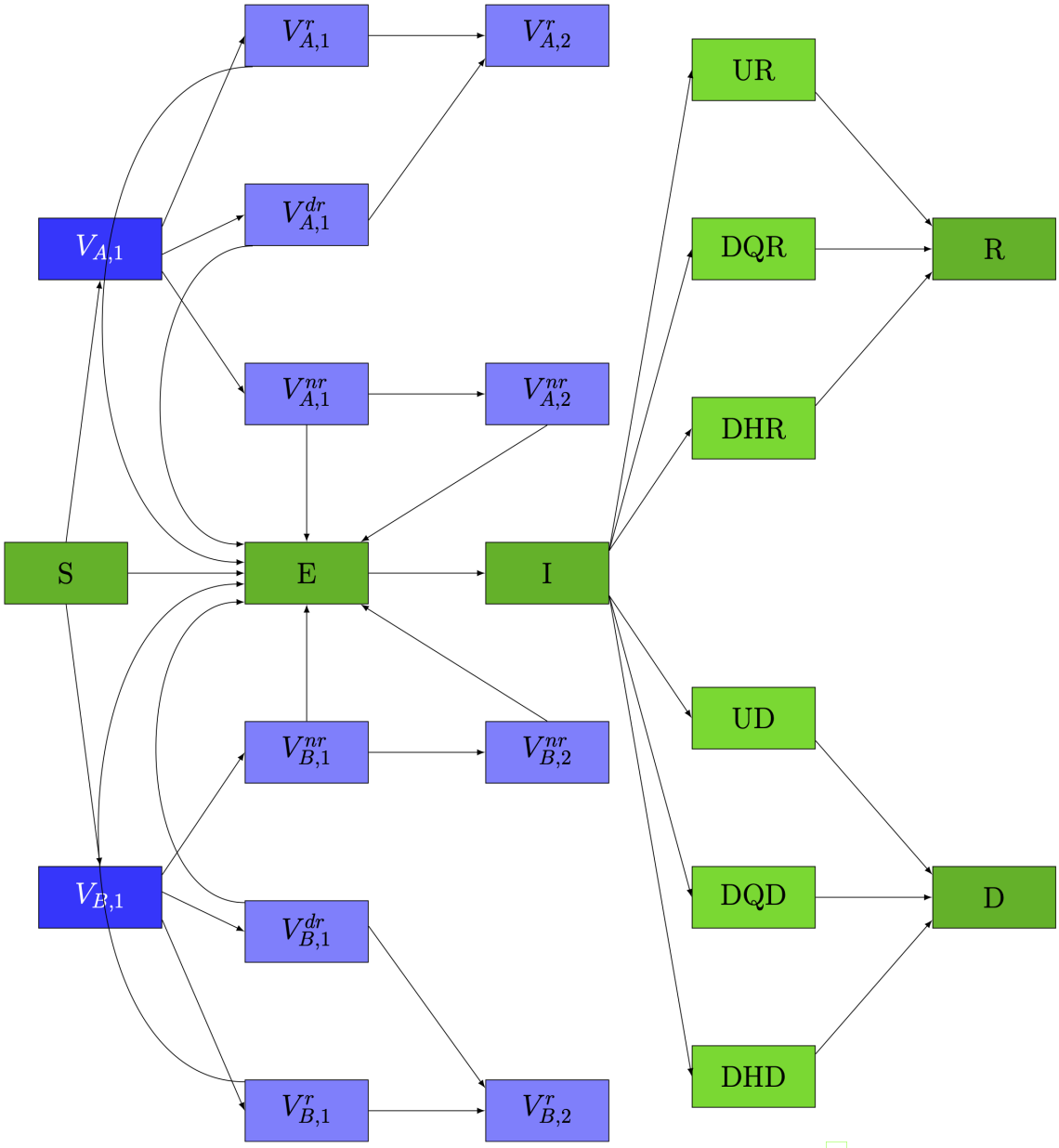


Figure 5-1: Flowchart of the original DELPHI model (in green) [152] and the additional vaccination states (in blue) for two hypothetical vaccines. For illustrative purposes, Vaccine A is loosely modelled after the Moderna vaccine, and Vaccine B after the Pfizer-BioNTech vaccine.

5.2.2 Data and Assumptions

The first step of the analysis consists in fitting the original DELPHI model to historical data using the dataset developed by [152]. After estimating the parameters of the original DELPHI model for each state of the U.S., we recalibrate these parameters to allow us to simulate a discretized version of the DELPHI model using a time step of 1 day. We then ensure that the discretized model yields the same output as the

original continuous-time model (see Appendix D.1.2 for more details). This step is crucial, as it considerably improves the speed of the simulation and allows us to run Monte Carlo analyses. For consistency of the results, all simulations are based on the set of discretized parameters.

	Brand A	Brand B	Perturbations
1st dose Efficacy[b]	80.20%	52%	±20%
2nd dose Efficacy[b]	95.60%	92%	+4%, −20%
Time Until 2nd dose[c]	21 days		28, 35, 49, & 63 days
Time to Develop Permanent Immunity[a]	14 days		21 days
Vaccination Start Date[d]	December 15th, 2020		

[a] [192]
[b] [96]
[c] The FDA recommends 21 days for the Pfizer-BioNTech vaccine and 28 days for the Moderna vaccine. As shown in Appendix D.2.1, this difference will have no impact on the analysis.
[d] [36]

Table 5.1: Vaccination parameters used in the augmented DELPHI model for two hypothetical vaccines. For illustrative purposes, Vaccine A is loosely modelled after the Moderna vaccine, and Vaccine B after the Pfizer-BioNTech vaccine.

The parameters used in the augmented DELPHI model are presented in Table 5.1. We assume a uniform daily infection rate among individuals in each vaccination state. Individuals who respond to the first dose (the "immediate response" group $V^r_{X,1}$) remain completely susceptible to an infection in the first 14 days of their first vaccination, but become permanently immune to the disease 14 days following their first dose. Similarly, individuals in the "delayed response" group i.e., the $V^{dr}_{X,1}$ group, remain completely susceptible to an infection 35 days after receiving their first dose (i.e., 21 days to receive their second dose after their first dose, followed by 14 days to develop permanent immunity), but develop permanent immunity immediately afterwards. Individuals who neither respond to the first nor second dose i.e., the $V^{nr}_{X,1}$ group, remain permanently susceptible to an infection. Although the FDA recommends a time interval between vaccine doses of 21 days for the Pfizer-BioNTech vaccine and 28 days for the Moderna vaccine, this difference has no impact on the analysis, as shown in Appendix D.2.1. Finally, we assume that the immune response to a vaccine does not decay over time.

5.2.3 Modeling the Supply of Vaccines

To explore the effect of shocks to the supply of vaccines on the vaccination policy adopted, we decompose the vaccine rollout into two phases: during the *ramp-up phase*, the number of new vaccine doses supplied increases at a linear rate, until it reaches a terminal value of 1.5 million new doses per day (President Biden's target [205]); This terminal value is reached on the 90^{th} day, when we enter the *steady-state phase* in which the supply rate of new doses becomes constant. The assumed terminal

value is on the conservative side, as the 7-day moving average of the number of doses administered daily (as reported to the CDC) increased from 1.5 million doses per day in February 2021 to 3 million in April 2021 [36, 217, 198] (a terminal rate of 3 million doses per day is explored in Appendix D.2.3). The black curve in Fig 5-2 represents the daily number of new vaccine doses supplied by one vaccine company. As shown in the plot, the number of doses supplied by this company increases linearly, until it reaches a value of 0.75 million doses (one half of 1.5 million , as we consider two vaccines in this study).

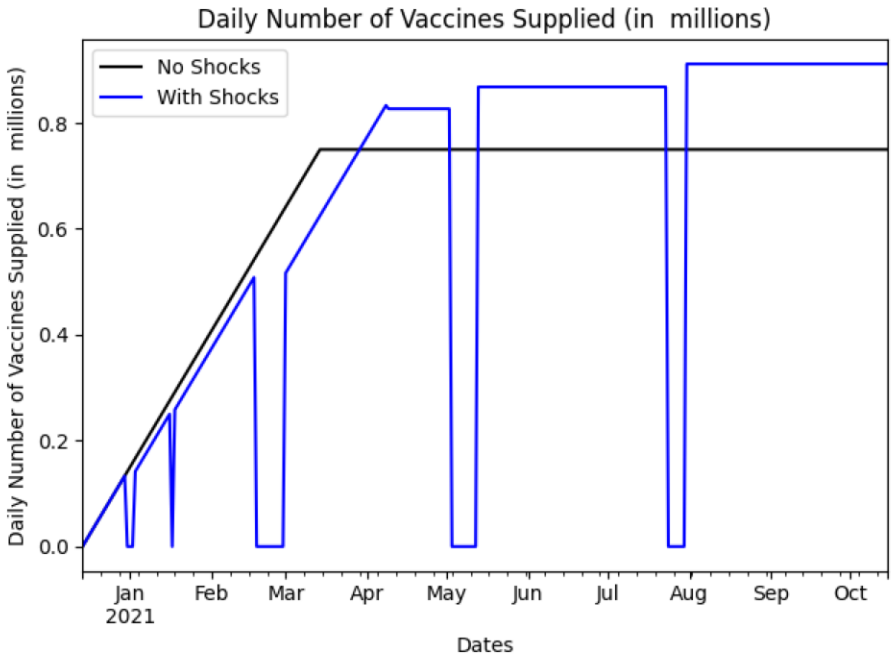


Figure 5-2: Example of the daily number of vaccines supplied by one company with and without supply shocks between December 15th, 2020 and August 1st, 2021.

To model supply shocks, we assume that shock occurrences follow a Poisson process with a rate of 1 shock per 30 days. Using a Poisson process is appropriate here as we assume that shocks that occur over disjoint time intervals are independent, and that the process is memoryless. Once a shock occurs, the supply of this particular vaccine drops to zero over a length of time drawn from a uniform distribution between 0 and 14 days. The supply then picks up at the previous positive level and continues to increase linearly. Furthermore, we assume that shocks lasting 7 days or more have a 50% probability of boosting the terminal supply rate by 5%. The blue curve in Fig 5-2 provides an example of a supply curve with such shocks. Supply shocks can simply represent delays in production or delivery of vaccines (which would tend to last a few days), but they can also model a factory shutdown aimed at improving the production of vaccines (which would tend to last longer and may increase the terminal supply rate).

Finally, each state receives a fraction of the number of available doses in proportion to its population size. We then run Monte Carlo simulations to investigate the

robustness of the vaccination policy to supply shocks. The number of cumulative deaths and cumulative infections are aggregated at the country level, and are used to compare vaccination policies.

5.3 Results

In this section, we explore the performance of various vaccination policies, and evaluate them based on the number of cumulative deaths and cumulative infections aggregated at the country level. This helps us understand whether we should store vaccine doses in order to guarantee a second dose to individuals who received a first dose, or if it is more efficient to allocate as many first doses as possible.

Storing doses ensures that individuals who received a first dose will be able to obtain their second dose according to the recommended vaccination schedule (here, 21 days) even if supply shocks occur. However, this strategy reduces the number of individuals that can be vaccinated each day, and may lead to a higher cumulative number of deaths and infections. We further assume that 1% of unused doses are lost each day in order to model spoilage or wastage due to unforeseen circumstances.

5.3.1 Vaccination Policies

The policies we investigate are described below.

Baseline Policy As a baseline policy, we consider the case of not vaccinating the population. This case is expected to present the highest number of cumulative infections and deaths.

Policy of Interest The vaccination policy of interest consists in allocating a fixed fraction of available doses to first-time users, and allocating the remaining doses to individuals who have already received their first dose and are eligible to receive their second dose. Furthermore, unused doses are reallocated to individuals eligible to get a vaccine. For example, under a policy of interest allocating 75% of doses to first-time users, 75% of the doses available today would be administered to individuals who have not received their first dose and 25% of doses will be administered to individuals who have received their first dose at least 21 days ago; if doses are unused because we have more second doses available today than eligible individuals for a second dose, we reallocate these unused doses to first-time users; if doses are unused because we have more first doses available today than individuals eligible for their first dose, these unused doses are reallocated to individuals eligible for a second dose today. In comparison, we also consider a scenario under which we do not allow for doses reallocation.

The policy of interest is then compared to the following alternatives.

Alternative Policy I: Strong Priority Scenario Doses are allocated by prioritizing all individuals who have received a first dose and will eventually need to receive the second dose in the future. This means that all individuals who receive their first dose are guaranteed to receive their second dose within the recommended time frame. Under the strong priority scenario, a second dose will immediately be placed in storage each time an individual receives their first dose, and this dose will be administered to this individual 21 days later.

Alternative Policy II: Weak Priority Scenario In contrast to the strong priority scenario, this policy consists of allocating doses in priority to individuals scheduled to receive their second dose on that specific day. In the weak priority scenario, doses available today are first administered to individuals eligible to receive their second dose; after clearing the second dose queue, the remaining doses are allocated to first-time users.

In all the vaccination policies described above, second doses are always allocated in a First-In-First-Out (FIFO) fashion: this gives higher priority to individuals eligible for a second dose who have not been able to receive their second dose yet over individuals who only became eligible for a second dose today.

A final point: it can be useful to view the weak priority scenario as a special case of the policy of interest in which we reallocate unused doses. In fact, under a policy of interest that allocates 0% of doses towards first-time users, individuals eligible to receive their second dose today will be given priority; then unused doses would be reallocated towards first-time users.

5.3.2 Policy Evaluation

To compare the four policies described in Section 5.3.1, we simulate the evolution of the epidemics in the absence and in the presence of random supply shocks. In particular, we run 1,000 simulations to obtain a distribution for the cumulative number of infections and the cumulative number of deaths between October 1st, 2020 and August 1st, 2021 under random supply shocks. After comparing the outputs obtained with various number of Monte Carlo simulations, we selected the number of Monte Carlo simulations to be large enough to reflect the uncertainty in the output while being parsimonious enough to retain a practical simulation runtime. The DELPHI parameters used in all our forecasts were estimated on February 7th, 2021.

We plot in Fig 5-3 the cumulative number of infections and the cumulative number of deaths under the policy of interest in the absence of supply shocks (green dots) and in the presence of supply shocks (blue dots) as we increase the allocation of doses towards first-time users. The numbers on the y-axis are negative as we display the number of infections and the number of deaths under the policy of interest relative to the baseline policy. In particular, we observe that allocating 50% of available doses to first-time users saves $11,632$ lives on average if there are no supply shocks, and saves an average of $8,793$ lives (95%CI: $[6,477; 10,803]$) under random supply shocks relative to a no-vaccination policy. Allocating 50% of doses to first-time uses also

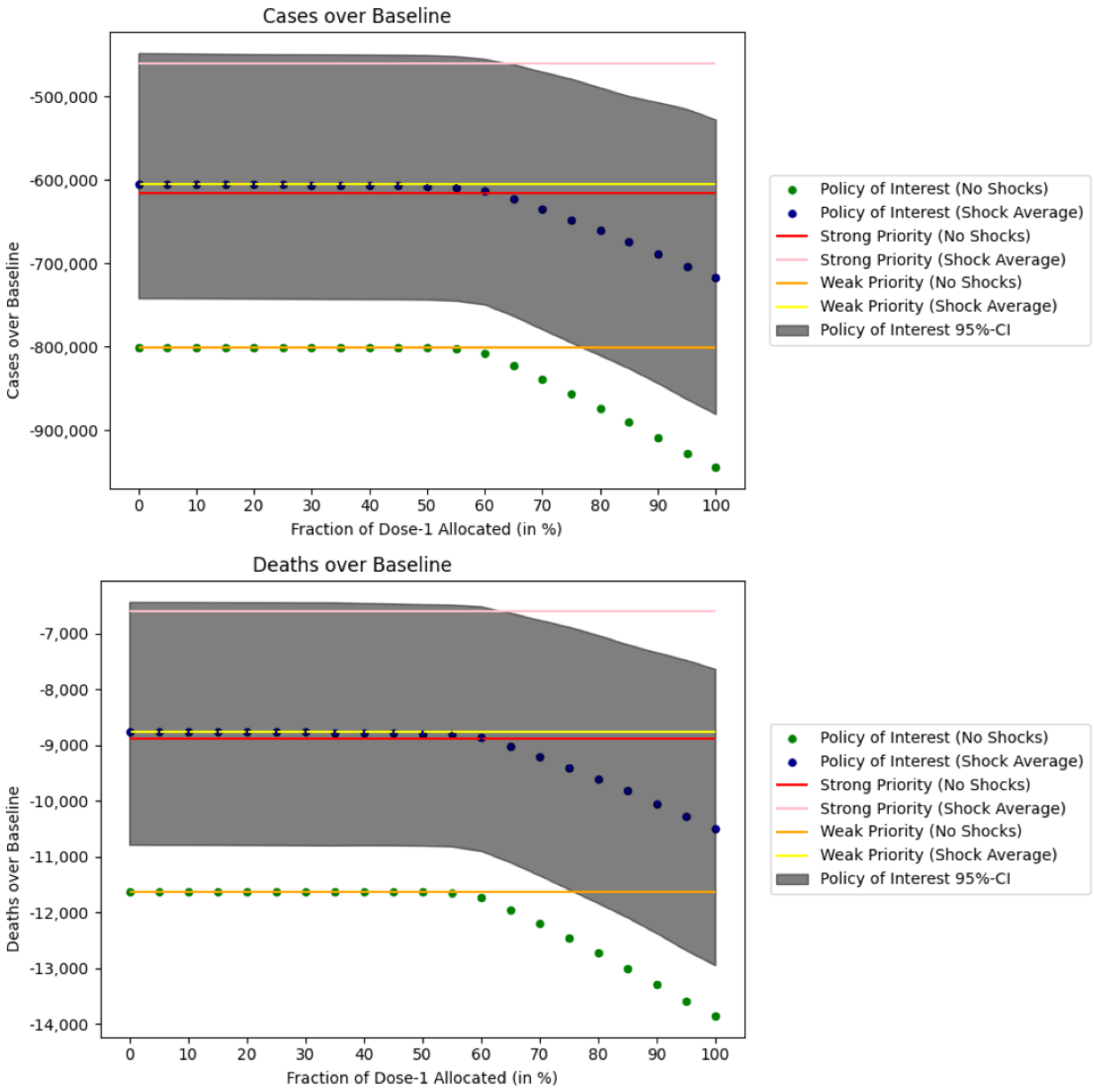


Figure 5-3: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of deaths and infections** between October 1st, 2020 and August 1st, 2021 relative to a no vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.

prevents $801,451$ infections under no supply shocks and $607,100$ infections (95%CI: $[450,024; 743,198]$) under supply shocks.

As we increase the fraction of doses allocated to first-time users, our forecasts predict a decrease in the cumulative number of infections and deaths. In fact, if we

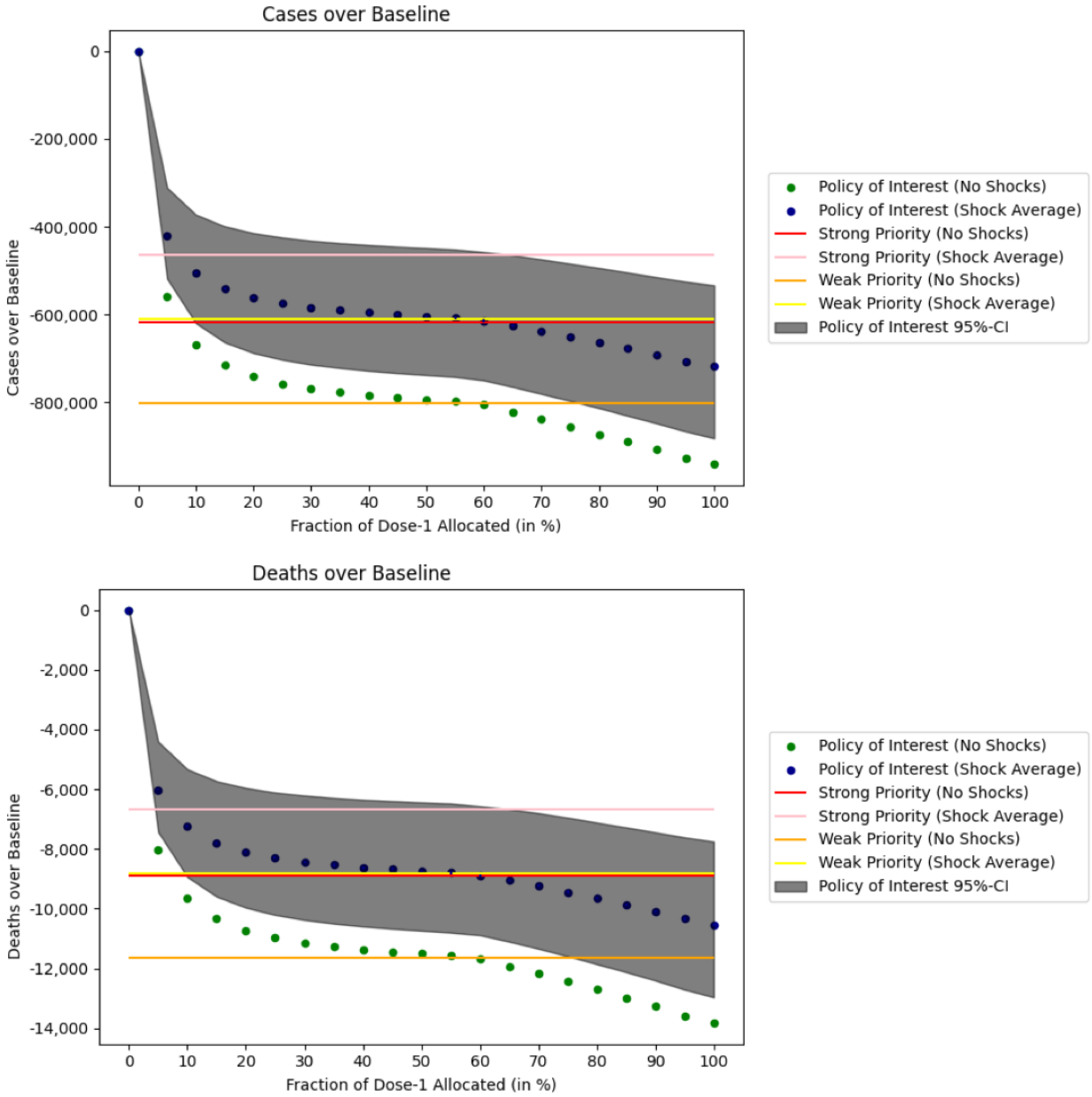


Figure 5-4: Simulation of the DELPHI model under supply shocks when we **do not reallocate excess doses** to individuals eligible to get a vaccine. We calculate the cumulative number of deaths and infections between October 1st, 2020 and August 1st, 2021 relative to a no vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.

allocate all available doses to first-time users, the policy of interest would save $13,859$ lives in the absence of supply shocks, and save on average $10,497$ lives (95%CI: $[7,634; 12,943]$) under supply shocks. This policy would also prevent $944,717$ in-

fections in the absence of supply shocks, and prevent on average 716,579 infections (95%CI: [527,296; 880,397]) under supply shocks. The estimates are reported in Table 5.2 and Table 5.3.

In contrast, the alternative policies considered do not allocate a fixed fraction of available doses to first-time users. Instead, under the strong priority policy (Table 5.4), a second dose is kept in storage as soon as an individual receives his or her first dose. This strategy is able to save only 8,876 lives under no supply shocks, and on average 6,609 lives (95%CI: [4,790; 8,213]) under supply shocks, while it prevents 616,315 infections in the absence of supply shocks and on average 460,743 infections (95%CI: [337,131; 569,924]) under supply shocks.

The weak priority policy (Table 5.5) relaxes this restriction and distributes the available doses each day in priority to individuals who already received their first dose and are eligible to receive their second dose on that day. Under this strategy, our forecasts predict that 11,631 lives would be saved under no supply shocks, and on average 8,759 lives (95%CI: [6,434; 10,786]) would be saved under supply shocks. This strategy would also prevent 801,387 infections in the absence of supply shocks and on average 604,926 infections (95%CI: [447,624; 741,807]) under supply shocks.

It is important to highlight here that under the policy of interest, the cumulative number of infections and deaths remains constant as we vary the fraction of doses allocated to first-time users from 0% to about 50%. This effect is due to the reallocation of unused doses modelled by our simulation. More concretely, if we allocate no doses to first-time users (i.e., we only give doses to individuals who have already received their first dose) and *do not* reallocate unused doses, then nobody would ever receive their first dose and hence nobody will ever be eligible to receive a second dose. Reallocating unused doses overcomes this issue. Furthermore, reallocating unused doses under a 0% first dose allocation policy exactly matches the outcome of the weak priority scenario (the orange and yellow lines in Fig 5-3), in which we always give priority to individuals who already received their first dose and are now eligible to receive their second dose.

When unused doses are not reallocated, we obtain the forecasts displayed in Fig 5-4. Allocating no doses to first-time users is identical to the no-vaccination policy, while increasing the dose allocation to first-time users beyond 30% produces similar results to Fig 5-3.

5.4 Discussion

5.4.1 Policy Comparison

As expected, vaccinating the population significantly reduces the number of infections and deaths under all the policies considered. However, the forecasts presented in Section 5.3.2 allow us to immediately rule out the efficiency of the first alternative policy (i.e., the strong priority scenario) relative to the other vaccination policies presented. In fact, both the policy of interest and the weak priority scenario are significantly better than the strong priority scenario in the presence and absence of

Dose 1	No Shocks	Shock Average	Shock SD	Shock s.e.	5th-perc.	95th-perc.
0	-801,387	-604,926	90,931	2,875	-741,807	-447,624
5	-801,387	-605,019	90,908	2,875	-741,879	-447,886
10	-801,387	-605,122	90,883	2,874	-741,998	-448,173
15	-801,387	-605,238	90,854	2,873	-742,126	-448,502
20	-801,387	-605,370	90,822	2,872	-742,261	-448,861
25	-801,387	-605,522	90,788	2,871	-742,416	-448,995
30	-801,387	-605,703	90,750	2,870	-742,592	-449,121
35	-801,387	-605,924	90,706	2,868	-742,722	-449,259
40	-801,387	-606,205	90,648	2,867	-742,795	-449,441
45	-801,387	-606,576	90,575	2,864	-742,879	-449,673
50	-801,451	-607,100	90,494	2,862	-743,198	-450,024
55	-802,656	-608,261	90,495	2,862	-744,474	-451,435
60	-808,134	-612,442	90,999	2,878	-749,287	-454,678
65	-822,703	-622,631	93,180	2,947	-762,924	-461,109
70	-839,272	-634,693	95,480	3,019	-779,156	-470,017
75	-856,478	-647,566	97,751	3,091	-795,018	-478,389
80	-873,507	-660,804	99,898	3,159	-810,243	-488,860
85	-890,732	-674,248	102,115	3,229	-825,871	-499,160
90	-908,812	-688,298	104,511	3,305	-843,509	-506,485
95	-927,952	-703,192	107,090	3,386	-862,812	-514,555
100	-944,717	-716,579	109,442	3,461	-880,397	-527,296

Table 5.2: Simulation of the DELPHI model under supply shocks as we vary the fraction of doses allocated to first-time users. We calculate the **number of infections** relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the 2021/02/07 DELPHI model parameters.

random supply shocks. This is also expected, as more individuals are able to receive their first dose under the policy of interest and the weak priority scenario and start to develop an immune response early. The magnitude of the improvement is even more striking: under supply shocks, the policy of interest allocating 50% of available doses to first-time users is expected to save on average 33% more lives and prevents on average 32% more infections than the strong priority scenario. Nevertheless, the strong priority scenario is still important to analyze, as individuals getting a vaccine in the U.S. usually obtain an appointment for their second dose as soon as they receive their first dose, and patients requiring a second dose are given priority [34].

In the absence of supply shocks, the weak priority scenario is dominated by the policy of interest when more than 60% of available doses are allocated to first-time users. In particular, if we compare the weak priority scenario to a policy of interest allocating 85% of available doses to first-time users, our forecasts predict an increase in the number of lives saved and of the number of infected averted of 11.7% and 11.1% respectively. This result also holds in the presence of supply shocks: we forecast

Dose 1	No Shocks	Shock Average	Shock SD	Shock s.e.	5pct	95pct
0	-11,631	-8,759	1,349	43	-10,786	-6,434
5	-11,631	-8,760	1,349	43	-10,789	-6,435
10	-11,631	-8,762	1,349	43	-10,789	-6,435
15	-11,631	-8,764	1,348	43	-10,790	-6,437
20	-11,631	-8,766	1,348	43	-10,793	-6,438
25	-11,631	-8,768	1,347	43	-10,795	-6,438
30	-11,631	-8,771	1,347	43	-10,798	-6,441
35	-11,631	-8,774	1,346	43	-10,799	-6,441
40	-11,631	-8,779	1,345	43	-10,797	-6,453
45	-11,631	-8,785	1,344	42	-10,798	-6,463
50	-11,632	-8,793	1,342	42	-10,803	-6,477
55	-11,648	-8,810	1,342	42	-10,817	-6,484
60	-11,730	-8,870	1,350	43	-10,895	-6,517
65	-11,946	-9,024	1,383	44	-11,098	-6,628
70	-12,198	-9,205	1,418	45	-11,334	-6,756
75	-12,454	-9,400	1,453	46	-11,585	-6,878
80	-12,723	-9,605	1,488	47	-11,831	-7,027
85	-12,996	-9,819	1,524	48	-12,087	-7,192
90	-13,286	-10,045	1,562	49	-12,370	-7,337
95	-13,593	-10,285	1,603	51	-12,675	-7,473
100	-13,859	-10,497	1,640	52	-12,943	-7,634

Table 5.3: Simulation of the DELPHI model under supply shocks as we vary the fraction of doses allocated to first-time users. We calculate the **number of deaths** relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the 2021/02/07 DELPHI model parameters.

an increase of 12.1% in the number of lives saved and an increase of 11.5% in the number of infections averted. These differences are statistically significant as a Welch t-test yields t-statistics of 16.4 and 16, respectively, for the number of lives saved and infections averted. This is also expected, as the weak priority scenario can be viewed as a special case of the policy of interest, where less than 50% of available doses are allocated to first-time users, with unused doses being reallocated. As a consequence, the number of lives saved and infections averted will always be higher under the policy of interest.

In summary, our forecasts suggest that allocating more than 50% of available doses towards first-time users, even at the cost of delaying the distribution of second doses, would be a better policy than guaranteeing a second dose within the recommended time frame to every individual receiving their first dose. The simulations also show that these results are robust to supply shocks. Although our analysis focuses on the United States, the forecasts and their interpretation can be generalized to any country. Prioritizing first doses would be even more relevant to countries where the

Setting	Cumulative Deaths	Cumulative Infections	Deaths over Baseline	Cases over Baseline
No Vaccination	610,626	30,877,115	0	0
No Shocks	601,750	30,260,800	-8,876	-616,315
Shock Average	604,017	30,416,372	-6,609	-460,743
Shock SD	1,054	71,403	1,054	71,403
Shock s.e.	33	2,258	33	2,258
t-Statistic	–	–	-198	-204
5pct	602,413	30,307,191	-8,213	-569,924
95pct	605,836	30,539,984	-4,790	-337,131

Table 5.4: Simulation of the DELPHI model under supply shocks as we vary the fraction of doses allocated to first-time users. We calculate the number of infections and deaths relative to a no-vaccination baseline under a **strong priority scenario**. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the 2021/02/07 DELPHI model parameters.

Setting	Cumulative Deaths	Cumulative Infections	Deaths over Baseline	Cases over Baseline
No Vaccination	610,626	30,877,115	0	0
No Shocks	598,995	30,075,728	-11,631	-801,387
Shock Average	601,867	30,272,189	-8,759	-604,926
Shock SD	1,349	90,931	1,349	90,931
Shock s.e.	43	2,875	43	2,875
t-Statistic	–	–	-205	-210
5pct	599,840	30,135,308	-10,786	-741,807
95pct	604,192	30,429,491	-6,434	-447,624

Table 5.5: Simulation of the DELPHI model under supply shocks as we vary the fraction of doses allocated to first-time users. We calculate the number of infections and deaths relative to a no-vaccination baseline under a **weak priority scenario**. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the 2021/02/07 DELPHI model parameters.

vaccine supply is severely limited (as shown in Appendix D.2.3).

5.4.2 Limitations and Sensitivity Analysis

Our forecasts are all based on an augmented version of the DELPHI epidemic model [152] that accounts for vaccinations. We should note that the model fails to account for demographics to assign different contact rates, hospitalization rates, and mortality rates across different age groups. Furthermore, some simplifying assumptions are used: for example, recovered individuals are assumed to have permanent immunity. However, among the top 10 models used by the CDC, DELPHI of-

ten displays the best performance with a low mean absolute percentage error (see https://www.covidanalytics.io/projections).

A critical limitation of our model is that we assume no decay in the efficacy of the vaccine over time if an individual has received their first dose, but are still waiting for their second dose. At this point, this decay in efficacy remains an open question [210]. Although our simulations begin on October 1st, 2020 and end on August 1st, 2021, vaccinations start on December 15th, 2020. If we consider a policy of interest that allocates 100% of available doses to first-time users, it would mean that individuals receiving their first dose at the end of December 2020 would not receive their second dose by August 1st, 2021. If the efficacy of the first dose decays over time, our forecast would be overly optimistic. However, knowing this decay rate would help determine the optimal fraction of doses that need to be allocated to first-time users under the policy of interest to balance the advantages of delaying the second dose against the efficiency loss due to the delay.

In hindsight, although the decay in efficacy is still difficult to quantify as of July 2022, most countries (including the U.S.) recommend a booster dose four to six months following the second dose. A useful extension of the current framework could be to consider all vaccinated individuals as susceptible six months after receiving their second dose. This modification would lead to a conservative result as it assumes a complete and instantaneous decay in efficacy at the six month mark, but can help select a policy that is robust to strong decays in efficacy.

Finally, we find that our results remain significant as we perturb some key assumed parameters. We show in Appendix D.2 the forecasts obtained as we increase the time interval between the first and second dose (from 4 weeks to 9 weeks), as we increase or decrease the efficacy of each vaccine dose, as we increase the time needed to develop permanent immunity, and as we increase the supply of vaccines. In particular, we observe that the curves obtained in Appendix D.2 tend to shift upwards as we increase the time interval between the doses, increase the time needed to develop permanent immunity, decrease the supply of vaccines, or decrease the efficacy of each vaccine dose, implying an overall reduction in the number of lives saved and infections averted. In addition, the curves become flatter, implying a lower sensitivity to the chosen fraction of available doses allocated to first-time users, especially as we increase the time needed to develop permanent immunity, decrease the supply of vaccines, or decrease the efficacy of the first dose.

5.5 Conclusion

We have developed a systematic framework to compare the efficiency of various vaccination policies. In particular, we extend the DELPHI model [152] to account for vaccination states, and explore the impact of prioritizing vaccines to first-time users instead of guaranteeing a second dose within the recommended time frame to individuals who have already received their first dose.

Our forecasts suggest that allocating more than 50% of available doses to first-time users significantly increases the number of lives saved and significantly reduces the

number of COVID-19 infections. It is important to highlight here that our forecasts are *not* recommending individuals to skip the second dose, a trend that has already raised some concerns as the efficacy of a single dose of mRNA vaccine over a long period of time remains unclear [148, 26, 201]. Instead, we suggest delaying the second dose to allow more individuals to receive the first dose in order to reduce the spread of the disease faster.

Chapter 6

Accelerating Vaccine Innovation for Emerging Infectious Diseases

The COVID-19 pandemic has raised awareness about the global imperative to develop and stockpile vaccines against future outbreaks of emerging infectious diseases. Prior to the pandemic, vaccine development for emerging infectious diseases was stagnant, largely due to the lack of financial incentives for pharmaceutical firms to invest in vaccine R&D. This R&D requires significant capital investment, most notably in conducting clinical trials, but vaccines generate much less profit for pharmaceutical firms compared to other therapeutics in disease areas such as oncology.

The portfolio approach of financing drug development has been proposed as a financial innovation to improve the risk/return tradeoff of investment in drug development projects through the use of diversification and securitization. By investing in a sizable and well-diversified portfolio of novel drug candidates, and issuing equity and securitized debt based on this portfolio, the financial performance of such a biomedical "megafund" can attract a wider group of private-sector investors.

To analyze the viability of the portfolio approach in expediting vaccine development against emerging infectious diseases, we simulate the financial performance of a hypothetical vaccine megafund consisting of 120 mRNA vaccine candidates in the preclinical stage, which target 11 emerging infectious diseases, including a hypothetical "disease X" that may be responsible for the next pandemic. We calibrate the simulation parameters with input from domain experts in mRNA technology and an extensive literature review, and find that this vaccine portfolio will generate an average annualized return on investment of -6.0% per annum and a negative net present value of $-\$9.5$ billion, despite the scientific advantages of mRNA technology and the financial benefits of diversification. We also show that clinical trial costs account for 94% of the total investment, while vaccine manufacturing costs account for only 6%. The most important factor of the megafund's financial performance is the price per vaccine dose, while other factors, such as the increased probability of success due to mRNA technology, the size of the megafund portfolio, and the possibility of conducting human challenge trials do not significantly improve its financial performance.

Our analysis indicates that continued collaboration between government agencies

and the private sector will be necessary if the goal is to create a sustainable business model and robust vaccine ecosystem for addressing future pandemics.

6.1 Introduction

The extraordinary human, social, and economic losses caused by the COVID-19 pandemic has heightened the global imperative to prepare for the next pandemic by proactively developing novel vaccines against EIDs. EIDs are a broad class of infectious agents that have either recently appeared for the first time, or whose incidence has rapidly increased in terms of size of the affected population or geographic area [223, 176]. A closely related threat is the reemergence of new variants of a previously identified EID, which may have become more transmissive or pathogenic through genetic mutation or shifting environmental conditions [170].

Given the dynamic and stochastic nature of EID outbreaks, the most effective strategy to prevent a future pandemic is to develop and stockpile vaccines before an outbreak occurs [131]. A notable example of proactive vaccine development is the CEPI, which has a portfolio of 32 vaccine candidates, as of April 14, 2022 targeting COVID-19 and six other priority EIDs [50]. Currently, the CEPI portfolio is diversified across 13 different therapeutic mechanisms (e.g., nucleic acid, recombinant protein, etc.) and five different stages of clinical development, from preclinical research to Emergency Use Listing by the World Health Organization (WHO). A similar example of proactive response was the International Coordinating Group (ICG) on Vaccine Provision's stockpiling of 2 million doses of yellow fever vaccines during a global shortage in 2000 [171]. In 2019, members of ICG renewed its pledge to maintain a stockpile of 6 million yellow fever vaccine doses [222]. Stockpiling vaccines well before an epidemic outbreak enables local governments and public health agencies to quickly address the sharp increase in vaccine demand following the outbreak, and facilitates more efficient vaccine allocation [132].

These considerations—and the remarkable effectiveness of mRNA vaccine technology against COVID-19—naturally lead to the question of the financial feasibility of a portfolio of mRNA vaccine candidates diversified across target EIDs, including both local EIDs and pathogens that may cause the next global pandemic. We address this question in this chapter by evaluating the financial performance of a hypothetical portfolio of 120 mRNA vaccine candidates targeting 11 EIDs, and determining whether the risk/return profile of such a portfolio might be attractive to private-sector investors. We do this by performing Monte Carlo simulations of the outcomes of hypothetical vaccine development programs that conform to a pre-specified set of parameters, and then examining the statistical distribution of these outcomes. We calibrate the parameters of these simulations using input from domain experts in mRNA technology and an extensive literature review.

We find that this vaccine portfolio yields an average annualized return on investment of -6.0% per annum, and a negative net present value of $-\$9.5$ billion, despite the scientific advantages of mRNA technology and the financial benefits of diversification. We also show that the clinical trial costs of this vaccine portfolio account for

94% of the total investment, while vaccine manufacturing costs account for only 6%. The most important factor of the portfolio's financial performance is the price per vaccine dose, while other factors, such as the increased probability of success due to mRNA technology, the size of the portfolio, and the possibility of conducting human challenge trials—in which healthy subjects are vaccinated and then deliberately infected with the virus to test vaccine efficacy—do not significantly improve its financial performance.

If the goal is to create a sustainable business model for addressing EIDs effectively, our results suggest that a likely pre-requisite will be continued collaboration between the public and private sector.

6.2 Brief Overview of Vaccine Development

6.2.1 The Past: A Decline in Vaccine R&D Prior to the COVID-19 Pandemic

Before the COVID-19 pandemic, pharmaceutical firms had pivoted away from vaccine R&D for EIDs, especially for small-scale but highly lethal agents such as the Ebola and Marburg viruses [137]. Several important factors were involved in this exodus, including high R&D costs [97], a low probability of success (PoS) in developing a vaccine candidate from preclinical studies to regulatory approval (estimated to be between 6% and 25% by [61, 187, 186, 214], the low list prices of vaccines [33], the uncertainty in vaccine demand and revenues [95, 183], and the lack of sustainable funding from public and private sectors in the absence of an imminent epidemic outbreak. Pharmaceutical firms have a greater financial incentive to develop and manufacture vaccines for common seasonal epidemics such as influenza compared to EIDs, since there is much less uncertainty in the estimated demand of these vaccines [69].

To illustrate the financial disincentives of vaccine R&D for EIDs more concretely, consider the following simplified model. Assume that the cost of developing a single vaccine candidate, from preclinical studies to regulatory approval or emergency use authorization (EUA), is $200 million, the probability of receiving regulatory approval is 25%, and the target EID occurs with probability 10% in any given year. If an outbreak does occur, we assume 10 million doses are manufactured, with a list price $20 per dose. Under these assumptions, the total expected revenues over the next 20 years (which is the duration of a vaccine patent):

$$25\% \cdot \$20 \cdot 10 \text{ million} \cdot 10\% \cdot 20 = \$100 \text{ million} \tag{6.1}$$

is only half of the R&D costs, despite rather optimistic assumptions about these costs and the EUA compared to more realistic estimates found in the literature [187, 186, 214]. This simple example also shows that the financial returns of vaccine R&D can be increased if the EUA can be improved due to scientific innovation (e.g., mRNA technology) or financial innovation (e.g., a portfolio approach to parallel

vaccine development), or a combination of both.

6.2.2 The Present: A Revolution in mRNA Vaccines

Vaccine R&D has gone through a scientific revolution during the pandemic, exemplified by mRNA technology, which has demonstrated robust levels of safety, high efficacy, and unprecedented speed in clinical vaccine development [41]. Once the genetic sequence of a pathogen is known, mRNA vaccine candidates can be designed more quickly than traditional vaccines. In addition, since mRNA vaccines do not require the production of inactivated or attenuated pathogens, they can be manufactured at large scale at higher efficiency, lower cost, and with more robust safety guarantees [181]. This technology has the potential to significantly reduce both the cost and the duration of vaccine R&D, enabling a much more rapid response to future EIDs. It is also particularly suited for the development of multiple mRNA vaccines in parallel, as in the portfolio approach taken by CEPI , since different mRNA vaccines may be able to share the same resources and facilities for preclinical studies, clinical testing, and post-approval manufacturing and delivery [208].

As an illustration of the success of mRNA vaccine development, consider the mRNA-1273 vaccine developed by Moderna for COVID-19, which was designed in 2 days, tested on the first human volunteer in 63 days, and received an EUA from the FDA in a little over 11 months after the genetic sequence of the original viral strain was first released [174, 109]. The R&D period of mRNA vaccines is significantly shorter than the usual 5 to 10 years for traditional vaccine development that were required before the COVID-19 pandemic.

We should note that the stunning successes of mRNA vaccine R&D against the COVID-19 virus was a result not only of technological advances, but also due to the close partnership between the public and private sectors in developing a mature mRNA technology well over a decade before the pandemic [67], as well as a product of the unprecedented collaboration between the government, regulatory agencies, scientists and clinicians around the world, and the pharmaceutical industry to expedite vaccine development in the midst of the COVID-19 outbreak. As we illustrate in subsequent sections, the continued collaboration and funding support from the public sector is critical to ensuring that vaccine R&D for EIDs can be financially sustainable.

6.2.3 The Future: Parallel R&D for mRNA Vaccines

mRNA technology brings a novel perspective to vaccine R&D in the portfolio approach used by CEPI by lowering the R&D and manufacturing costs through sharing resources on a common R&D platform, which improves the EUA of vaccine development by the "multiple-shots-on-goal" parallel strategy of discovery. However, a serious challenge to vaccine R&D remains in the lack of sufficient and sustainable funding to support the vaccine R&D pipeline over an extended period, typically multiple years from preclinical research to the regulatory approval of a vaccine, an issue known as the "valley of death" in translational medicine [29].

Governments, international agencies, and non-governmental organizations such as the Gates Foundation, Wellcome Trust, and CEPI have made significant contributions to the development of a portfolio of vaccine candidates, but these efforts are not sufficient due to the scale of the challenge (see Section 2 of [214] for a detailed discussion). The private sector does have sufficient resources to bridge this funding gap but will do so only if the portfolio can generate sufficiently attractive financial returns for its investors.

To illustrate the benefits and challenges of applying the portfolio approach to vaccine R&D, we return to our earlier back-of-the-envelope calculation. Suppose we invest in a portfolio of 10 mRNA vaccines candidates targeting local epidemics. The total cost increases to $10 \cdot \$200$ million = \$2 bilion, while the probability that at least one vaccine candidate receives regulatory approval (assuming statistically independent outcomes) increases substantially to $1 - (1 - 25\%)^{10} = 94.4\%$. The expected revenues over the next two decades becomes:

$$94.4\% \cdot \$20 \cdot 10 \text{ million} \cdot 10\% \cdot 20 = \$378 \text{ million}, \tag{6.2}$$

a financial loss of \$1.6 billion. However, if the vaccine targets an EID which causes a global pandemic with an annual probability 1%, and 1 billion vaccine doses are produced if a pandemic occurs, the expected revenues of the vaccine portfolio increases to:

$$94.4\% \cdot \$20 \cdot 1 \text{ billion} \cdot 1\% \cdot 20 = \$3.8 \text{ billion} \tag{6.3}$$

a profit of \$1.8 billion, while the expected revenues of investing in one vaccine is only:

$$25\% \cdot \$20 \cdot 1 \text{ billion} \cdot 1\% \cdot 20 = \$1.0 \text{ billion} \tag{6.4}$$

which implies a deficit of \$1 billion.

These numbers highlight both the advantages and the bottlenecks to applying a portfolio approach to funding vaccine R&D. First, the parallel discovery strategy improves the EUA of vaccine R&D. Even if vaccine development outcomes are correlated to each other, the probability of having an approved vaccine in a portfolio is still higher than the EUA of investing in a single vaccine program (assuming that the pairwise correlations are not equal to 1). An increased EUA can make vaccine R&D profitable for those EIDs capable of causing global pandemics. However, it is insufficient to generate financial value for vaccines against local EIDs, since the revenues of local vaccine sales is limited. In addition, since the mRNA vaccines share the same therapeutic mechanism, it is reasonable to assume that there will be no significant difference in efficacy between different approved mRNA vaccines for the same EID (as in the case of COVID-19). As a result, there will be considerable cannibalization of demand for vaccines targeting the same EID, since the demand for vaccines will not increase with the number of approved vaccines. Finally, the stochastic nature of EID outbreaks induces large variance in the revenues of vaccine sales. For vaccine R&D aimed at preventing a global pandemic, even though the expected financial return is positive, there is still a significant probability in our illustrative model of $(1-1\%)^{20} = 81.8\%$ that a global pandemic will not occur in the next 20 years, leading

to a financial loss of $2 billion.

6.3 Portfolio Approach to Financing Drug Development

6.3.1 Challenges of the Drug Development Process

To develop a novel therapeutic candidate from laboratory discovery to regulatory approval, a drug developer needs to conduct multiple clinical trials to test the safety and efficacy of the therapeutic candidate on the target patient population. These clinical trials are conducted in sequence through four stages (preclinical, phase 1, phase 2, and phase 3)[1]. Trials in a more advanced phase typically require a larger patient enrollment and a longer time to complete, and are correspondingly more expensive. If the phase 3 clinical trial shows clear safety and efficacy, the drug developer files a New Drug Application (NDA) to the FDA for regulatory approval. If the FDA approves the NDA, the drug developer may manufacture the drug and collect revenues from drug sales. Sometimes, the FDA may require an additional phase 4 clinical trial after regulatory approval, in order to test the long-term benefits and side effects of the drug on a large patient population.

Despite the tremendous breakthroughs in biomedicine over the past decades, new drug development has become slower, more expensive, and less likely to succeed, causing a significant funding gap for early-stage drug development programs. The lack of sufficient funding for translational biomedical R&D is due to several institutional features of drug development, including a low EUA, a long investment horizon, high clinical trial costs, and a high cost of capital, especially for small biotechnology companies which do not have marketed drugs that generate revenues and must rely on external financing to sustain its R&D pipeline[2]. The declining efficiency of translating scientific discoveries in research laboratories into novel products has also been observed in other industries in the US [12].

[1]Phase 1 trials typical involve 10 to 50 patients, with the only goal of establishing the safety and the maximum tolerable dose of a given drug candidate. If no significant side effects are encountered in Phase 1, a Phase 2 trial is initiated in which 50 to 500 patients who suffer from the targeted disease are carefully selected to test the drug candidate's efficacy. If significant benefits are detected in that trial, a much larger Phase 3 trial involving thousands of patients is launched to test the drug candidate's efficacy in a broader and less carefully curated sample of patients, and if significant benefits are detected in Phase 3 with no serious side effects, the drug is approved for general use. Because vaccines are administered much more widely than other drugs, and given to healthy subjects rather than only those with a given disease, the regulatory hurdle for determining safety and efficacy is considerably higher—a typical Phase 3 trial for a vaccine involves 30,000 subjects (as in the case of the COVID-19 vaccines), hence the outsized costs of Phase 3 trials. See Chapter 8 from [156] for further details.

[2]See [158] for a systematic review of financing issues in the biopharma industry.

6.3.2 Advantages of Financing Vaccine R&D via the "Vaccine Megafund"

To address the challenge of funding translational medicine, [86] proposed a novel financing vehicle, the biomedical "megafund", which invests in a sizable portfolio of drug candidates diversified across different clinical stages and therapeutic areas. Using financial engineering techniques such as securitization, the authors show that the risk/return profile of the megafund is attractive to a wide group of investors. Originally proposed to finance oncology drug development, the megafund model was subsequently applied to other disease areas, including orphan diseases [72], Alzheimer's disease [157], pediatric cancer [69], ovarian cancer [42], glioblastoma [204], and vaccines against EIDs [214]. It is currently being applied by the National Brain Tumor Society (NBTS) to finance novel drug candidates to treat glioblastoma [173].

The key idea behind the megafund is to reduce the financial risks of its assets and improve its expected returns by raising capital to acquire a portfolio of vaccine candidates, issuing equity and securitized debt with different risk/return profiles that appeal to a wide range of private-sector investors. The vaccine candidates are used as collateral, and the revenues generated by future vaccine sales are used to service its debt and interest payments. The residual equity is then distributed among its equity holders. If the future cash flows are insufficient to service the debt, the megafund declares bankruptcy and the collateral is transferred to its bondholders. The main advantage of portfolio diversification is that by increasing the EUA of having at least one approved drug candidate, the megafund is able to lower the financial risks and attract large amounts of capital from the bond market, whose size is much larger than the venture capital, public equity, or private equity market [125]. In 2020, a total of $12.2 trillion worth of fixed income securities were issued in the US, compared to $390 billion of equity. In the same year, the total private placement was $330.1 billion in the US, of which $314.4 billion was in the form of debt and $15.8 billion in the form of equity [125].

6.3.3 Evaluating the Financial Performance of the Vaccine Megafund

In the vaccine megafund simulation analysis of [214], the financial performance of a vaccine-focused portfolio is extremely unattractive to for-profit investors, with an expected annualized return of -61% and a standard deviation (SD) of 4%. Multiple factors lead to this negative financial return, including a low EUA of vaccine trials, high clinical trial costs, and limited revenues from vaccine sales. Based on these findings, the authors propose several strategies to finance the vaccine megafund, including higher vaccine prices, public sector funding, and a novel subscription model in which subscribers would pay annual fees for priority access to the vaccines in case of future outbreaks.

In this chapter, we extend the work of [214] in several important ways. First, [214] simulated vaccine trial outcomes stochastically, but used a single fixed expected value to estimate the annual profit for approved vaccines. We implement a more realistic

simulation framework in which the entire value chain of vaccine development, manufacturing, and sales is simulated under the stochastic occurrence of EID outbreaks. The uncertainty in future EID outbreaks increases the variance of megafund cash flows, which directly impacts its risk/return profile. In addition, we use improved EUA estimates of mRNA vaccines to adjust the cash flows of the megafund, and calibrate the cost structure of mRNA vaccine manufacturing with input from domain experts and an extensive literature review. Finally, while [214] mainly focused on the annualized return of the vaccine megafund, we systematically investigate a wide spectrum of metrics to gauge its financial and social impact, such as the net present value and the number of EID outbreaks prevented. We also provide a detailed breakdown of the cost structure for the vaccine megafund to identify the main drivers of its financial performance.

The risk/reward profile of the vaccine megafund hinges on the scientific and business expertise of fund managers to select promising drug candidates and diversify the portfolio [204]. For a real-world vaccine portfolio such as CEPI 's, active portfolio management is critical, given budget constraints, to select a limited number of vaccine candidates. [98] apply multi-criteria decision analysis to select promising vaccine candidates for the CEPI portfolio in the context of multiple trade-offs and heterogeneous stakeholder preferences. In a subsequent study [99], the authors apply portfolio decision analysis to optimize the investment of CEPI in 16 vaccine technology platforms. [7] analyzed the optimal investment strategy of vaccine manufacturing capacity for countries with different socioeconomic characteristics.

While we fully recognize the importance of active portfolio management in improving the financial performance of a vaccine megafund, we do not impose exogenous budget constraints or perform any portfolio optimization in our simulation analysis since our goal is to understand the relationships between the investment and revenues of the vaccine megafund and its endogenous factors such as the improvement in the EUA of mRNA vaccine development, the cost structure of mRNA vaccine manufacturing, the size of the megafund portfolio, and the possibility of conducting human challenge trials to expedite vaccine clinical trials.

6.4 Simulation Methods

6.4.1 Vaccine Megafund Portfolio

We simulate the financial performance of a large portfolio of mRNA vaccine candidates using an adaptation of [214]'s portfolio structure and probability of outbreak P_a of each EID, as shown in Table 6.1. We also include 10 vaccine candidates which target "disease X", the unknown pathogen which may cause the next pandemic, in accordance with the updated CEPI portfolio [50]. We assume that disease X has a low annual probability of outbreak $P_a = 1\%$, and the number of infected cases will be 400 million, close to that of COVID-19.

Targeted EID	Number of Vaccine Candidates (N_{vac})	Annual Probability of Outbreak (P_a, in %)	Average Number of Infections (n_I)
Disease X	10	1.0	400,000,000
Chikungunya	16	10.8	523,600
Zika Virus	18	4.3	500,062
Lassa Fever	7	100.0	300,000
Rift Valley Fever	3	10.5	79,414
SARS-CoV-1	2	7.1	8,098
West Nile Virus	23	10.0	500
MERS-CoV	8	40.0	436
Crimean-Congo Haemorrhagic Fever	7	12.5	320
Nipah Virus	20	15.8	136
Marburg Virus	6	12.0	75

Table 6.1: Portfolio for simulated mRNA vaccine megafund [50, 214].

6.4.2 Vaccine Clinical Trials

We use the simulation framework in [204] to model the correlated outcomes of vaccine clinical trials. The assumed values of the simulation parameters of a vaccine clinical trial are summarized in Table 6.2. The simulated trial outcomes depend on two critical sets of parameters. First, the EUA to reach each stage in the clinical development process is estimated using historical industry average values [186, 214]. In addition, since the mRNA vaccine for COVID-19 is known to induce humoral immune protection by producing neutralizing antibodies [128], we assume that mRNA vaccines will have a higher EUA for the six EIDs in the portfolio whose correlates of protection are also neutralizing antibodies (Chikungunya virus, SARS-CoV-1, Marburg virus, Rift Valley Fever, Nipah virus, and Zika virus). To reflect the increased EUA due to mRNA technology for these diseases, we multiply the historical EUA by a technology factor α_{tech}. We set α_{tech} to 1.2 in the baseline model, which reflects a 20% increase in the EUA over the industry average. We do not increase the EUA for the other five diseases with cellular or unknown immune responses, including disease X. We vary α_{tech} in the sensitivity analysis to gauge the effect of increased EUA on financial performance.

In addition, the correlations between vaccine trial outcomes play a major role in the simulation outcomes. If two vaccine trial outcomes are highly correlated, e.g., due to the same target pathogen or therapeutic mechanism, they are more likely to simultaneously succeed or fail, which leads to lower diversification benefits from the portfolio, greater variance in the cash flows of the megafund, and thus greater overall financial risk. Using the input of domain experts in mRNA technology, we construct a biologically motivated metric to estimate these correlations.

Specifically, we propose a novel distance metric d_{ij} between pathogens i and j, defined as the average of similarity scores based on four biological factors: taxonomy,

Parameter	PRE to P1	P1 to P2	P2 to P3	P3 to EUA	Source
Probability of Success (EUA, in %)	60.0	83.6	65.8	80.9	[214, 186, 220]
Duration (months) of Standard clinical trial	18.0	24.0	18.0	14.0	[214, 22]
Development cost ($M) of Standard clinical trial	26.0	14.0	28.0	150.0	[97]
Duration (months) of Human challenge trial	—	—	—	8.0	[22]
Development cost ($M) of Human challenge trial	—	—	—	12.5	[22]

Abbreviations — PRE: preclinical phase, P1: Phase 1, P2: Phase 2, P3: Phase 3, EUA : Emergency Use Authorization.
Note — We assume that a vaccine receives EUA once it successfully completes phase 3 clinical trial. Furthermore, we assume human challenge trials are only applicable to phase 3.

Table 6.2: Simulation parameters for vaccine clinical trials.

qualitative features (e.g., type of disease vector, strand direction, nucleic acid topology), quantitative features (e.g., number of strands, total genome size), and the edit distance of protein sequences. Simply put, the more similar two pathogens are to each other, the more correlated we assume their trial outcomes will be. This value of d_{ij} is normalized between 0 and 1, with d_{ij} closer to 0 if pathogens i and j are more biologically similar, and $d_{ij} = 0$ if they are identical. Given the values of d_{ij}, a natural way to define the correlation ρ_{ij} between the outcomes of vaccine trials targeting pathogens i and j is $\rho_{ij} = 1 - d_{ij}$, i.e., the vaccine trial outcomes have a higher correlation if their target EIDs are more biologically similar, and vice versa.

Figure 6-1 shows the heatmap of ρ_{ij} between each pair of pathogens, excluding disease X (which we assume to be independent of the other pathogens, to reflect its a priori unknown biological properties). The correlation matrix ρ_{ij} defined this way is positive definite (PD) in our calibration, although it is not guaranteed to be PD in general and may need to be transformed into a PD matrix by an appropriate method [189][3]. Since this metric does not specify the correlation between two vaccine trials targeting the same pathogen, we assume this correlation to be 0.8, which is higher than the maximum correlation of 0.64 across different pathogens (Figure 6-1). To gauge the impact of correlation on the financial performance, we vary the assumed values of correlation in the sensitivity analysis.

[3]Positive definiteness is a mathematical property that guarantees the positivity of the variance of a weighted average of random variables. Given that the risk (as measured by variance) of a portfolio is never negative, it is important to impose this property on any correlation matrix otherwise, nonsensical numerical results like negative risk may occur.

	Chikun.	SARS	MERS	Marburg	RVF	Lassa	Nipah	CCHF	WNV	Zika
Chikun.	1.00	0.30	0.30	0.37	0.27	0.39	0.38	0.29	0.38	0.33
SARS	0.30	1.00	0.58	0.32	0.21	0.25	0.28	0.26	0.29	0.28
MERS	0.30	0.58	1.00	0.33	0.20	0.25	0.28	0.26	0.29	0.28
Marburg	0.37	0.32	0.33	1.00	0.27	0.37	0.46	0.37	0.36	0.35
RVF	0.27	0.21	0.20	0.27	1.00	0.48	0.29	0.52	0.27	0.26
Lassa	0.39	0.25	0.25	0.37	0.48	1.00	0.36	0.35	0.40	0.40
Nipah	0.38	0.28	0.28	0.46	0.29	0.36	1.00	0.32	0.39	0.39
CCHF	0.29	0.26	0.26	0.37	0.52	0.35	0.32	1.00	0.29	0.28
WNV	0.38	0.29	0.29	0.36	0.27	0.40	0.39	0.29	1.00	0.64
Zika	0.33	0.28	0.28	0.35	0.26	0.40	0.39	0.28	0.64	1.00

Figure 6-1: Estimated correlations between vaccine candidates. We assume that vaccine candidates for disease X are uncorrelated with vaccines for the other diseases and that vaccine candidates targeting the same disease have a 0.8 correlation.

6.4.3 Human Challenge Trials

Given the demonstrated safety and efficacy of mRNA vaccines for COVID-19, it is conceivable that HCTs may be ethically justified for mRNA vaccine candidates in our portfolio. The HCT is an efficient yet highly controversial clinical trial design, in which healthy participants with no previous exposure to a disease are deliberately infected with the live pathogen in a controlled clinical environment (e.g., an isolated ward in a hospital). The controlled setting of a HCT allows much more precise and rapid testing of the safety and efficacy of vaccines with a smaller number of trial participants than standard vaccine trials. As a result, an HCT may significantly reduce the cost and duration of clinical trials and lead to expedited regulatory approval of effective vaccines. In a simulation analysis, [22] showed that conducting an HCT for COVID-19 vaccines may significantly reduce the number of infected and deceased patients in the US compared to other clinical trial designs, provided that the vaccine is effective and the HCT is initiated in a timely manner.

Although conducting an HCT is in principle more efficient in time and cost than traditional vaccine trials, in practice it still faces multiple challenges. First and foremost, the ethical justification of deliberatively injecting healthy participants with a live EID agent is highly controversial, due to the absence of well-established ethical guidelines to specify the conditions under which an HCT may be deemed ethical. In addition, HCTs require more time and resources during their initial preparation stage (e.g., identifying and manufacturing low-risk virus strains, identifying low-risk populations, and establishing an acceptable HCT protocol with regulators). As a result, the first HCTs for COVID-19 were initiated after the mRNA vaccine candidates had already received EUA from the FDA in US and Europe [30, 190].

Although we recognize the ethical and practical challenges of HCTs, we model an

idealized scenario when an HCT is authorized for mRNA vaccine R&D and may be conducted in an ethical and timely manner. We use the binary variable HCT_i to denote whether an HCT is authorized by the FDA during an outbreak of disease i (i.e., $HCT_i = 1$ with probability p_{HCT} if the HCT is authorized by the FDA, and $HCT_i = 0$ with probability $1 - p_{HCT}$ if otherwise). If $HCT_i = 1$, we use the reduced cost and duration of HCT (rows 4 and 5 of Table 6.2) instead of the corresponding values of standard trials. We assume $p_{HCT} = 0$ in the baseline model (i.e., no HCT is conducted) and gauge the effect of p_{HCT} in the sensitivity analysis.

6.4.4 Vaccine Manufacturing and Supply Chain

The cost structures of mRNA vaccine manufacturing and its supply chain are key to simulating the cash flows of the megafund. Since mRNA vaccine manufacturers do not disclose this information, we use publicly available estimates in the literature [141, 142] to calibrate these cost structures. The line-item budget of mRNA vaccine manufacturing is summarized in Table 6.3. The main factor driving the manufacturing costs is the amount of mRNA raw material needed to produce the target number of vaccines. We assume that each production line consists of a bioreactor with a 30-liter working volume and mRNA titer 5g/L [142]. We also assume that each vaccine dose contains 65μg of mRNA, the average of the Pfizer/BioNTech and Moderna vaccines for COVID-19.

Category	Item	Unit Cost (USD)	Quantity
Fixed costs	Production line	58 million	1 bioreactor of 30L working volume
	Raw materials	456.6 million per (year · production line)	29,162 grams of mRNA per production line per year
	Consumables	150 million per (year · production line)	
Variable costs	Labor	20 per hour	113,186 labor hours per production line per year
	Quality control	10 per hour	10-dose vials
	Fill-and-finish	0.27 per dose	Not modeled here
	Lab, utility, waste management, etc.	$< 1\%$ total cost	

Table 6.3: Cost structure of mRNA vaccine production. [141, 142]

Using the estimates in Table 6.3, the variable cost of producing each mRNA vaccine dose is \$1.60. We assume that each local EID outbreak requires 10 million vaccine doses. It takes 8.1 days to produce the mRNA needed with one production line, and an additional 4 to 5 weeks to perform quality control for each batch produced. The total manufacturing cost is \$16 million if one uses the existing production line, and \$75 million if one builds a new production line. Similarly, we assume that a disease X pandemic requires 1 billion vaccine doses. It takes 81.4 days to produce the

mRNA needed with 10 production lines. The total cost is $1.6 billion with existing production lines, and $2.2 billion with new ones. Furthermore, we assume that the variable cost of delivering each vaccine dose in the supply chain is $1.00 (of the same order of magnitude as the manufacturing cost). We make a conservative assumption about the supply chain cost due to the lack of publicly available estimates in the literature. Our simulation results show that the supply chain costs constitute only 2% of total costs (Figure 6-4), so the financial performance is not sensitive to the detailed structure of supply chain costs, as long as it does not exceed $1.00 per dose by an order of magnitude.

To estimate the revenues generated by vaccine sales, we use the list prices of mRNA vaccines for COVID-19. As of October 26, 2021, the Pfizer/BioNTech vaccine is priced at $24.00 per dose in the US, and the Moderna vaccine at $15.00 per dose [133]. We assume that the price per vaccine dose is $20.00. This is likely to be an underestimate, since it is below the prices of all adult vaccines listed in the vaccine price list of Centers for Disease Control and Prevention except for influenza vaccines [33]. To gauge the impact of the list price of vaccines, we vary the price in the sensitivity analysis.

6.4.5 Simulating Correlated Clinical Trial Outcomes

The key to simulating the financial performance of the vaccine megafund is to simulate the correlated binary outcomes of vaccine clinical trials. As in the previous biomedical megafund simulations (e.g., [204]), we use the technique proposed by [71] to simulate correlated Bernoulli variables. Vaccine clinical trials have five development phases (preclinical, phase 1, phase 2, phase 3, and emergency use authorization, or EUA), and need to go through four phase transitions before receiving the EUA . Let the Bernoulli variable $B_{ij} \in \{0, 1\}$ denote whether vaccine candidate i has entered the development phase j, with $j \in \{0, 1, 2, 3, 4\}$. Initially all vaccines are in preclinical stage, i.e., we set $B_{i0} = 1$. If the vaccine trial advances from phase $j - 1$ to j where $j \in \{1, 2, 3\}$, we set $B_{ij} = 1$. If the vaccine receives EUA from the FDA, we set $B_{i4} = 1$.

To simulate the correlated phase transitions of clinical trials from phase j to $j+1$, we first draw a vector of multivariate standard normal variables $\varepsilon_j = [\varepsilon_{1j}, ..., \varepsilon_{nj}]$ with independent components ε_{ij}, where the length n is the number of vaccines in the portfolio. Next, we compute $z_j = \Sigma^{1/2}\varepsilon_j$ where $\Sigma^{1/2}$ is the Cholesky decomposition of the correlation matrix σ (Figure 6-1). The resulting vector z_j follows a multivariate normal distribution with zero mean and covariance matrix equal to Σ. Given the probability of success p_j for phase transition from j to $j + 1$ (Table 6.2), we simulate the binary clinical trial outcome as:

$$B_{i,j+1} = \begin{cases} 1, & \text{if } z_{ij} > \alpha_j, \\ 0, & \text{if } z_{ij} \leq \alpha_j, \end{cases} \tag{6.5}$$

where z_{ij} is the i^{th} component of z_j, $\alpha_j = \Phi^{-1}(1 - p_j)$, and Φ^{-1} is the inverse cumulative distribution function of the standard normal variable. The clinical trial

outcomes B_{ij} generated this way are positively correlated in each phase transition and used in the financial calculations. In each Monte Carlo simulation, if we observe $B_{ij} = 0$, the clinical trial for vaccine i terminates in phase j and all subsequent B_{ik} (with $k > j$) are set to 0. If we observe $B_{ij} = 1$, the megafund incurs the clinical trial cost for phase j. If an epidemic outbreak occurs and there is at least one vaccine i with $B_{i4} = 1$ (i.e., it has received EUA), we manufacture the vaccine and collect the revenue from vaccine sales.

6.4.6 Overview of the Simulation Framework

At the initial time $t = 0$, all vaccine candidates enter the preclinical stage. For simplicity, we assume that the development costs of each phase are incurred at the start of the phase. In each subsequent year from $t = 1$ to $t = T$, we simulate whether any EID outbreaks (including the disease X pandemic) occur in year t. In the absence of any outbreaks, we develop each vaccine candidate (except the ones for "disease X") from the preclinical stage to the completion of phase 2, assuming the cost and timeline of a standard clinical trial (rows 2 and 3 of Table 6.2). We do not initiate a large-scale phase 3 clinical trial unless an outbreak has occurred, since there will not be enough infected subjects with which to test vaccine efficacy until then. From a financial perspective, this also reduces the significant late-stage clinical trial costs compared to the simulation analysis of [214].

If an EID outbreak occurs in year t, we assume that one of the four scenarios below will occur (Figure 6-2):

1. At least one vaccine candidate targeting the disease has successfully completed a phase 3 trial during a previous outbreak of the same disease and received approval or an EUA from the FDA. We manufacture the vaccines, supply them to the point of distribution, and collect the revenues from the vaccine sales.

2. At least one vaccine candidate targeting the disease has successfully completed a phase 2 trial. We initiate the phase 3 clinical trial. If the phase 3 trial is successful, the vaccine receives an EUA from the FDA. We manufacture and supply the vaccines, and collect the revenues from the vaccine sales.

3. At least one vaccine candidate for the epidemic is in the preclinical or phase 1 stage. We initiate an accelerated phase 1/2 trial, which costs \$28 million (the same as a standard phase 2 trial) and completes in 3 months, followed by a standard phase 3 trial, which completes in 14 months. If the phase 3 trial is successful, the vaccine receives an EUA. We manufacture and supply the vaccines, and collect the revenues.

4. No vaccine candidates for the disease have previously completed a phase 3 trial or remain in the R&D pipeline. In this case, no cash flows are generated, since all vaccine candidates have failed in the clinical trial process.

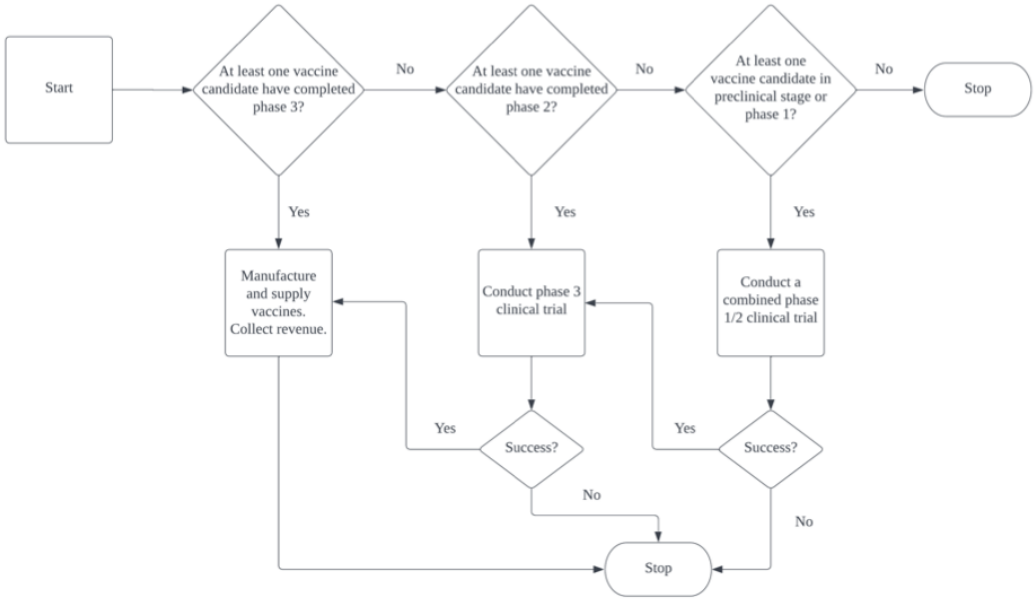


Figure 6-2: Overview of the simulation framework in the event of an epidemic outbreak.

6.5 Results

There are four key observations and insights from the results of the simulation analysis:

- Despite the improved EUA of mRNA vaccines, the vaccine megafund does not generate financial value for the investors, and is not a financially self-sustainable business model for the pharmaceutical industry.
- From the perspective of public policy, the vaccine megafund will require $9.5 billion funding from the public sector at its initiation to generate positive financial value for investors (given an average revenue of $7.5 billion from vaccine sales and an average cost of $17.7 billion from the clinical trial development and vaccine manufacturing).
- The main bottlenecks of the financial performance are the limited and uncertain revenues generated by the vaccine sales and the significant costs of clinical trials, which account for 94% of the total investments in the megafund.
- The vaccine megafund generates significant social benefits by preventing, on average, 31 epidemic outbreaks out of 45 over the next two decades. In addition, there is a 66% probability that the next "disease X pandemic" will be prevented by vaccines developed from the megafund portfolio.

The performance of the baseline portfolio is summarized in Table 6.4. We find that this portfolio has a negative expected annualized return $E[R_a] = -6.0\%$ (standard deviation $SD[R_a] = 6.7\%$) and a negative expected net present value (NPV) of −$9.5

billion (standard error SE $13 million). The vaccine megafund does not generate positive financial value for its investors, since the revenues generated by the vaccine sales ($7.5 billion on average) is insufficient to recover the investment in clinical trial development and vaccine manufacturing ($17.7 billion on average). However, the financial value to private-sector investors does not capture the benefits generated by the megafund to society. On average, 45 infectious disease outbreaks will occur in the simulation period, 31 of which will be prevented or contained by vaccines developed from the portfolio. In addition, there is a 66% probability that vaccines in the portfolio will prevent the next "disease X pandemic", should one occur. Using even the most conservative "quality adjusted life year" estimate (e.g., [175]), the lives saved and socioeconomic losses avoided by the vaccines far exceed the negative financial value of the megafund.

Metric	Mean	Standard Error	Standard Deviation	Median	25% Qt.	75% Qt.
Annualized Return (R_a)	−6.0%	0.021%	6.7%	−5.7%	−7.4%	−4.4%
Net Present Value (NPV, USD, billion)	−9.5	0.013	4.1	−9.9	−12.1	−7.4
Investment (USD, billion)	17.7	0.017	5.3	17.8	14.0	21.4
Revenues (USD, billion)	7.5	0.024	7.7	5.8	3.4	7.0
Profit (USD, billion)	−10.0	0.023	7.4	−11.5	−14.9	−7.5
Number of Prevented Epidemics (N_{ep})	31	0.04	13	34	19	42

Note — NPV is computed with an annual discount rate $r = 10\%$. The standard deviation of preclinical trial cost is zero since the megafund invests in the preclinical trials of all 120 vaccine candidates at the initial time 0.

Table 6.4: Performance of the baseline portfolio computed with 100K Monte Carlo simulations.

The distribution of key performance metrics of the megafund is displayed in the histograms of Figure 6-3. We find that, although R_a and NPV are negative in most simulations, there is a 9.8% probability that $R_a > 0$, and a 3.1% probability that $NPV > 0$. In addition, the distribution of megafund investments is smooth with a single peak (i.e., this is a unimodal distribution), while the distribution of revenues has two peaks (i.e., a bimodal distribution): although revenues are mostly likely to fall below $10 billion, there is a sizeable probability that revenues exceed $20 billion. The latter corresponds to the rare scenarios when a disease X pandemic occurs, generating revenues of $20 billion from vaccine sales. This bimodality of revenues leads to significant variance in the annualized return and NPV of the megafund.

To gain additional insight into the major costs that reduce the financial performance of the megafund, we present a breakdown of megafund investment in Figure 6-4, and find that the costs of clinical trials constitute 94% of the total cost, with phase 3 trials alone accounting for 59%. The net cost of vaccine manufacturing and its supply chain constitute only 6% of the total cost, and the higher efficiency of

mRNA vaccine manufacturing is not sufficient to generate financial profits for the investors. Our finding is consistent with the "valley of death" in financing translational medicine [29], in which the main bottleneck is the risk associated with the uncertainty of revenues at the early stages of drug discovery versus the enormous cost of clinical trials. Even with more efficient vaccine manufacturing technologies and supply chain designs, the significant cost of clinical trials still prevents the vaccine megafund from generating positive financial value to its investors.

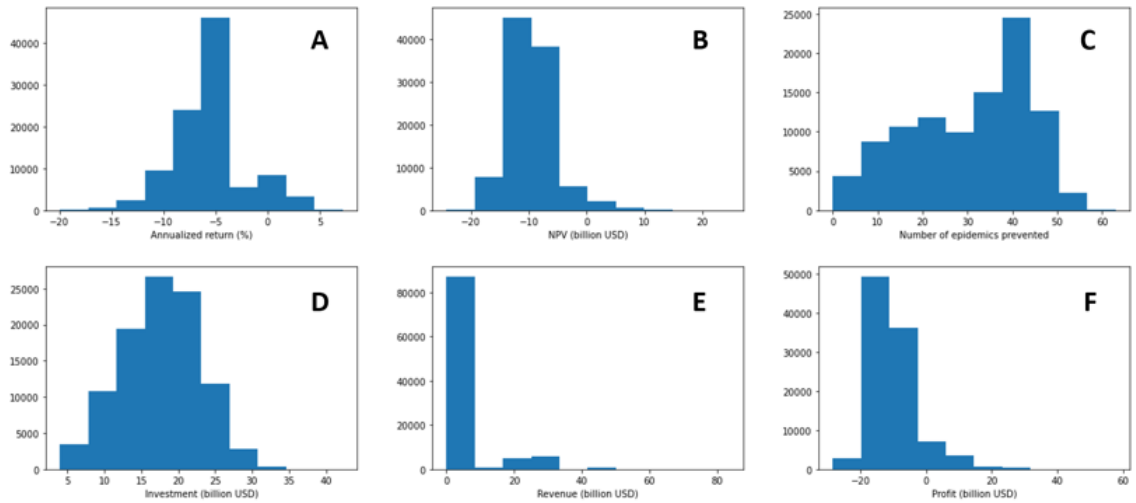


Figure 6-3: Histograms of key performance metrics of vaccine megafund. (A) Annualized return. (B) Net present value (NPV). (C) Number of epidemics prevented. (D) Total investment. (E) Total revenue. (F) Net profit.

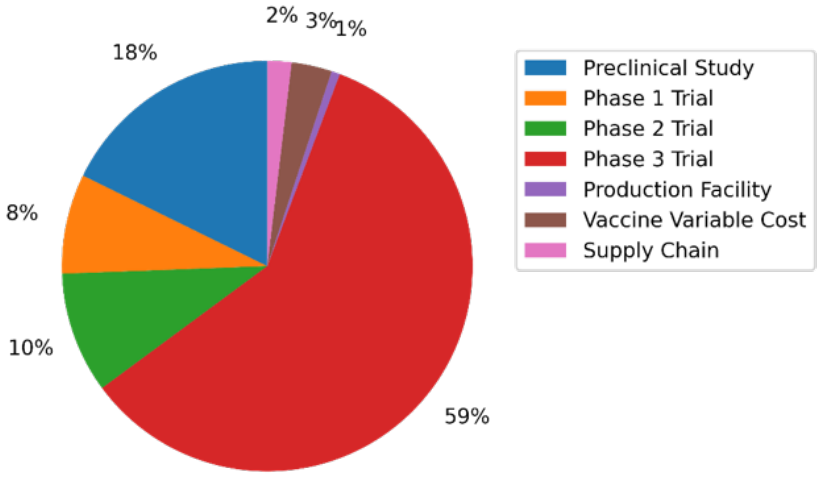


Figure 6-4: Breakdown of cost structure of the vaccine megafund. Clinical trial costs constitute 94% of all costs, while manufacturing costs constitute only 6%.

6.6 Sensitivity Analysis

The simulated financial performance of the vaccine megafund hinges on the assumed values of key simulation parameters calibrated using inputs from mRNA domain experts and estimates from the literature. We perform a sensitivity analysis to test the robustness of the simulation results against the assumed parameter values. The results discussed below are summarized in Table 6.5.

6.6.1 Vaccine Price

The price per vaccine dose π is the key driver of the financial performance. In the baseline model, we assume $\pi = \$20.00$, where both the annualized return and NPV are negative. Increasing π to \$69.00 (row 2 of Table 6.5) achieves the breakeven point for the annualized return. Increasing π further to \$78.00 (row 3 of Table 6.5) achieves the breakeven point for NPV. Assuming $\pi = \$100.00$ (row 4 of Table 6.5), the megafund generates a small but positive expected annualized return of 1.9%, with a volatility of 7.2% and an expected NPV of \$3.6 billion (SE \$55 million). Such a high list price of \$100.00 per vaccine dose is not unusual in the US. As of April 14, 2022, thirteen common adult vaccines have list prices above \$100.00 in the US [33]. However, these may be impossible to afford in low-to-middle income countries, and may even increase vaccine hesitancy among the affected population.

6.6.2 Improved Probability of Success of mRNA Vaccines

To test whether the increased EUA of mRNA vaccines leads to improved financial performance, we multiply the EUA of vaccine trials for six diseases by the technology factor α_{tech} to reflect the higher efficacy of mRNA vaccines for diseases with humoral immune protection. In the baseline model, we set $\alpha_{tech} = 1.2$ (i.e., a 20% increase in EUA). Surprisingly, increasing α_{tech} from 1.0 to 1.3 (rows 5 to 7 of Table 6.5) achieves a mixed effect: the expected annualized return increased from -6.7% to -5.8%, while the expected NPV decreased from $-\$8.1$ to $-\$9.9$ billion. As we increase α_{tech} from 1.0 to 1.3, the average number of approved vaccine candidates increases from 28 to 49, and the expected investment also increases from \$15.2 to \$18.4 billion. However, the reason for the mixed effect is that the expected revenue undergoes a much smaller increase, from \$7.1 to \$7.6 billion, since on average only 3 additional EID outbreaks are prevented by the approved vaccines (due to the stochastic occurrence of EID outbreaks). The smaller ratio of revenue to investment causes the annualized return to be less negative and increase, while the larger increase in investment causes the NPV to be more negative and decrease. We conclude that the higher EUA of mRNA technology alone does not generate positive financial value for the megafund unless we also reduce the clinical trial costs or raise the price of the vaccine.

6.6.3 Correlations between Clinical Trial Outcomes

The correlation between vaccine trial outcomes measures the tendency for multiple vaccine trials to simultaneously succeed or fail due to a common target disease or mechanism of action. In the baseline model, we estimate the correlation via the novel virus distance metric d_{ij}. However, we cannot simply rescale d_{ij} in the sensitivity analysis, since the resulting correlation matrix is not guaranteed to remain positive definite. Instead, we gauge the impact of correlation by assuming an equi-correlated correlation matrix in which $\rho_{ij} = \rho$ is the same for all diseases, and vary the value of ρ from 0 (independent) to 80% (highly correlated), as shown in rows 8 to 12 in Table 6.5. As expected, we observe that higher values of ρ lead to worse financial performance, as the expected annualized return decreases from -3.5% to -11.7% and the expected NPV decreases from $-\$8.3$ to $-\$9.5$ billion. In addition, the volatility of the annualized return dramatically increases from 2.5% to 23.6%. This shows the importance of diversity in the megafund portfolio to generate positive financial value.

6.6.4 Human Challenge Trials

If deemed ethical, an HCT may be able to significantly reduce the cost and duration of the clinical development of vaccine candidates by testing a smaller group of participants than traditional vaccine trials. We investigate the effect of HCTs on the megafund performance by assigning the probability p_{HCT} that HCT is allowed for each EID. The baseline portfolio does not utilize HCT, i.e., $p_{HCT} = 0$. Increasing p_{HCT} from 0 to 30% (rows 13 to 14 of Table 6.5) reduces the expected investment and increases both the annualized return and NPV, although both remain negative. We find that utilizing HCT alone is also insufficient to generate positive financial value for the investors.

6.6.5 Megafund Portfolio Size

The parallel vaccine development strategy increases the probability that at least one vaccine candidate will be approved, but it also increases the investment in clinical trials. To investigate the effect of portfolio size, we multiply the number of vaccine candidates for each infectious disease by a factor γ. The baseline portfolio corresponds to $\gamma = 1$. Increasing the portfolio size by 50% ($\gamma = 1.5$, row 16 of Table 6.5) leads to worse financial performance, since the expected investment increases from \$17.7 to \$25.7 billion, while the expected revenue only increases by a much smaller amount, from \$7.5 to \$7.9 billion, as the natural occurrence of EID outbreaks remains the same. Decreasing the portfolio size by 50% ($\gamma = 0.5$, row 15 of Table 6.5) increases both expected return and NPV, though both remain negative. In addition, the average number of epidemics prevented decreases from 31 to 27, which reflects a higher loss to society not captured by our financial analysis.

Portfolio	$E[R_a]$	$SD[R_a]$	$E[NPV]$	$SD[NPV]$	$E[Inv]$	$SD[Inv]$	$E[Rev]$	$SD[Rev]$	$E[N_{ep}]$	$SD[N_{ep}]$
Baseline	−6.0%	6.7%	−9.5	4.1	17.7	5.3	7.5	7.7	31	13
π = \$69/dose	0.0%	7.1%	−1.4	11.9	17.7	5.3	25.8	26.7	31	13
π = \$78/dose	0.7%	7.1%	0.0	13.5	17.7	5.3	29.2	30.2	31	13
π = \$100/dose	1.9%	7.2%	3.6	17.4	17.7	5.3	37.4	38.7	31	13
$\alpha_{tech} = 1.0$	−6.7%	11.9%	−8.1	4.1	15.2	5.3	7.1	7.8	28	14
$\alpha_{tech} = 1.1$	−6.2%	9.1%	−8.8	4.1	16.4	5.4	7.3	7.8	29	14
$\alpha_{tech} = 1.3$	−5.8%	4.8%	−9.9	4.1	18.4	5.1	7.6	7.7	31	13
$\rho = 0\%$	−3.5%	2.5%	−8.3	3.7	18.1	2.5	10.7	8.9	43	7
$\rho = 20\%$	−3.8%	2.7%	−8.5	4.0	18.0	3.9	10.2	8.7	41	9
$\rho = 40\%$	−4.2%	4.2%	−8.7	4.3	17.9	5.0	9.6	8.6	38	11
$\rho = 60\%$	−5.9%	11.1%	−9.0	4.6	17.8	6.0	8.7	8.3	35	14
$\rho = 80\%$	−11.7%	23.6%	−9.5	4.8	17.7	7.1	7.5	7.9	31	17
$p_{HCT} = 10\%$	−5.7%	6.7%	−8.8	4.1	16.7	5.1	7.5	7.7	31	13
$p_{HCT} = 30\%$	−5.1%	6.7%	−7.6	3.9	14.7	4.6	7.5	7.7	31	13
$\gamma = 0.5$	−4.1%	8.9%	−3.7	3.0	9.3	2.9	6.5	7.3	27	14
$\gamma = 1.5$	−7.3%	5.7%	−15.3	5.4	25.7	7.6	7.9	7.9	32	13

Note — R_a denotes annualized return (p.a.); NPV denotes net present value, Inv denotes net investment, Rev denotes net revenue, in billion USD; N_{ep} denotes the number of EID outbreaks contained by vaccines from the portfolio; π denotes the price per vaccine dose in USD; α_{tech} denotes the technology factor; p_{HCT} denotes the probability of HCT; ρ denotes the pairwise correlation between vaccine trial outcomes; γ denotes portfolio size factor. NPV is computed with an annual discount rate $r = 10\%$.

Table 6.5: Sensitivity analysis of key simulation parameters computed with 100K Monte Carlo simulations.

6.7 Discussion

Our analysis illustrates three major challenges to the portfolio approach of financing mRNA vaccines for EIDs. First, the portfolio approach reduces the supply side risk of vaccine R&D by increasing the probability of having at least one effective vaccine against an EID. However, it does not mitigate the demand side risk in the revenues generated by vaccine sales since vaccine demand is mainly determined by the natural occurrence of EID outbreaks. The stochastic nature of outbreaks limits the revenues generated by the approved vaccines, unless we increase the list price to \$78.00 per dose. But with such a high list price, local governments and populations may not be able to afford the vaccines, which further reduces their demand and revenues. In addition, since mRNA vaccines share the same therapeutic mechanism, it is reasonable to expect that there will be no differentiated efficacy of different vaccines against the same disease. As a result, there will be significant market cannibalization between approved vaccines since the total revenues of vaccine sales will not increase if there is more than one approved vaccine. Finally, the significant costs of clinical trials constitute 94% of megafund investment and severely limit its financial performance. One potential solution is to use more cost-effective clinical trial designs such as adaptive trials (Berry 2011) and platform trials [221], which simultaneously test multiple vaccine candidates using a shared control arm. These innovative trial designs have been shown to significantly reduce clinical trial costs and expedite the R&D process for glioblastoma therapeutic candidates [204]. In addition, they do not elicit the ethical controversies of human challenge trials.

We also note that the primary goal of the vaccine megafund is to prevent future EID outbreaks and minimize the overall burden of disease. In light of this goal, our

simulation assumes that we invest in clinical trials for all vaccine candidates simultaneously without optimizing for financial performance using sophisticated investment strategies [99] or financial engineering techniques such as dynamic leverage [169]. For example, if three vaccine candidates for the same infectious disease successfully complete their phase 2 trials, we may instead first conduct phase 3 trials for two vaccine candidates, initiating the phase 3 trial for the third vaccine only if the first two have failed. This will reduce the costs of late-stage clinical trial development and improve its financial value. However, the increased financial value must be weighed against potential delays in FDA approvals of life-saving vaccines. A robust and multi-criteria optimization framework is needed to ensure that their value to society is not compromised by optimizing financial returns for the investors.

6.8 Conclusion

Despite an increased probability of success due to mRNA vaccine technology, diversification across a large number of vaccine candidates, and the potential benefits of conducting human challenge trials, the vaccine megafund model does not generate positive financial value for private-sector investors. The three bottlenecks of its financial performance are the limited revenues of vaccine sales, the cannibalization of approved vaccines for the same infectious disease, and the significant costs of late-stage clinical trials. Nonetheless, the vaccine megafund does generate tremendous social value by preventing future epidemic outbreaks; if endowed with public sector funding of $10 billion, it may also generate positive financial value for investors.

Our analysis indicates that continued collaboration between government agencies and the private sector will be necessary if the goal is to create a sustainable business model and robust vaccine ecosystem for addressing future pandemics. Strategies such as stockpiling vaccines for the most dangerous EIDs, putting in place advance market commitments or subscription fees to purchase/reserve mass quantities of vaccines in case of outbreaks, creating government-sponsored manufacturing and distribution facilities that can supplement private-sector resources, and providing limited government guarantees to investors funding vaccine programs for a pre-specified list of priority diseases may all play a role in helping us reduce the impact of, or even prevent, future pandemics.

Part II

Managing Risks & Uncertainty in Empirical Asset Pricing

Chapter 7

Chimeras in the Factor Zoo: Constructing Latent Asset Pricing Factors via Autoencoders

The first part of this thesis applied risk management tools to the field of healthcare finance to inform policy and decision-making under uncertainty. The remainder of the thesis is devoted to key issues in empirical asset pricing (Part 2, Chapter 7) and investment management (Part 3, Chapter 8) involving risk and uncertainty management.

Empirical asset pricing aims to understand and quantify the nature of the different risk factors stocks are exposed to. For example, we estimate the risk exposure of a stock to a well-chosen set of risk factors using statistical inference techniques on the parameters of a linear (or nonlinear) factor model for the stock's returns. In this case, uncertainty would be associated with a time-varying nature of the factor model's parameters, including regime changes, non-stationarity, and potential black swans (e.g., a pandemic, economy-wide shocks, the introduction of disruptive technologies, a sudden flight to liquidity, etc.).

7.1 Introduction

Factor models are routinely used in the financial industry to identify and quantify sources of systematic risk in order to manage the risk of a portfolio of securities or hedge investment positions, or in valuation contexts to estimate the cost of capital of an asset. From an academic perspective, the empirical asset pricing literature reflects an unending quest for factors that could explain the cross section and time-series of expected returns [52, 112]. The importance of factors, whether traded or non-traded, is justified theoretically in the arbitrage pricing theory (APT) framework developed by Stephen Ross in 1976 [196, 54] by linking return factor structures with risk premia. In other words, according to APT, the cross section of expected returns should be explained by systematic factors who carry a nonzero risk premium.

The CAPM[200, 153] is the simplest and most popular single-factor model used in

both academia and industry. Developed in 1964, the CAPM identifies the systematic risk of a stock with its exposure to the market. However, the CAPM fails to explain a large portion of the cross-section of stock returns and multiple anomalies were documented over the past 50 years [121]. More complex factor models were developed based on stock characteristics such as the Fama-French 3, 4, and 5-factors models (*FF3*, *FF5*, *FF6*) [73, 74, 75], the Carhart 4-factors model [31], or the 4 and 5 q-factors (*q4*, *q5*) by Hou, Xue, and Zhang [120, 119]. For example, the Fama-French Three-Factor Model (FF3) model, developed in 1993, includes two additional sources of systematic risk: the value premium and the size premium.

A large set of factors has been introduced in the literature to attempt to explain the cross-section of asset prices, however no single model has convincingly been able to capture most of the anomalies [52, 112, 121]. Tools have been proposed to "tame the factor zoo" by testing the validity of the 300+ factors proposed in the literature and develop powerful parsimonious models (see [9, 139] for a more comprehensive review). We can group these techniques into two broad categories: a model-selection approach and a latent factor approach.

The first approach constructs a factor model by carefully selecting a subset out of the existing factors in the literature. For example, [85] use a double-selection LASSO method to evaluate the incremental contribution of a new factor to an existing factor model. A bootstrap-based method has also been used by [111] as a step-wise model selection method to select new factors. Similarly, [91] seek for factors with incremental explanatory power using the adaptive group LASSO. From a Bayesian perspective, [15] evaluates the contribution of a traded factor to an existing model by calculating Bayes factors and posterior model probabilities. [28] extend this idea to traded and non-traded factors and apply it to quadrillions candidate factor models.

The second approach aims to extract a set of latent factors by compressing the information contained in a large cross section of stock returns, portfolio returns, or factor returns. In contrast with the previous scenario in which factors are selected from the set of anomaly factors, the latent factors are now obtained by combining the anomaly factors in a linear or nonlinear way. Latent factors are no longer members of the factor zoo, but "chimeras" of anomaly factors. For example, PCA methods have been attempted by relating principal components of the covariance matrix of returns with risk premia [146, 188, 94, 93]. More recently, [151] extends the PCA framework to a risk premia PCA framework by choosing factors that explain most of the correlation and also minimize the pricing error (since, according to APT, average returns should be explained by the risk premia and factor risks estimated and no residual risk should be left). Time-varying techniques have been proposed by [138] and [77] to estimate conditional factor models using instrumented PCA and local PCA respectively. Similarly, [106] and [48] use deep neural networks to forecast asset returns. In particular, [48] learn time-varying factor models directly from stock returns using generative adversarial networks (GANs). Finally, [107] extend the instrumented PCA approach from [138] to nonlinear conditional factor models using AEs.

Our approach falls within the second category as we aim to construct a set of latent factors. In contrast with [151] and [107], we obtain parsimonious factor models by directly compressing the information contained in 150 factors described in [85] (from

1976 to 2017) using different types of AEs. We then test our model on the 150 known factors and compare its performance to benchmark factor models such as FF5 and q5. In addition, we test our models on a set of portfolios, including Fama-French anomaly portfolios [90] and Hou-Xue-Zhang anomaly portfolios [225]. Although latent factors are usually difficult to interpret by design, we incorporate economic intuition into the design of some of our AE architectures. In particular, among the 150 factors from [85] we use as inputs, we observe groups of factors that are correlated and can potentially lead to an over-representation bias during the compression phase relative to other factors that capture unique features of the cross-section and only appear. To solve this issue, we introduce a CAE that first clusters the 150 factors into smaller groups and then compresses the clusters into a small number of latent factors. This approach allows us to train deeper AEs more efficiently than the usual deep AEs with fully connected layers. Similar to [151], we also impose an APT condition in the loss function to reduce the pricing error. While linear AEs and PCA are similar in various respects, the latent factors obtained through PCA are uncorrelated while AEs usually do not generate uncorrelated latent factors. We address this issue by training the AE in a recursive way, learning one latent factor at the time, each time removing the factor exposure from the input. An advantage of using AEs over PCA is that we can construct latent factors that depend on the input factors in a nonlinear way. While linear factor models are often a good benchmark, latent factors obtained from machine-learning models tend to capture nonlinearities that linear models would fail to reflect. We show here that using tanh activation functions can help improve the performance of the model. In fact, we find that latent factors obtained using nonlinear AEs perform better than linear latent factors in polynomial regressions of order 2.

The remainder of the chapter is structured as follows: we describe the approach proposed in this work in Section 7.2; we carry out an empirical analysis for different AE factor models, present the results in Section 7.3, and discuss their relevance in Section 7.4; we conclude in Section 7.5. Additional sensitivity analysis and robustness checks are available in Appendix E.

7.2 Methodology

In this section, we discuss latent factor models, develop the different type of autoencoders used to learn a latent structure from a large set of anomaly factors, and describe metrics to evaluate the performance of the AE models.

7.2.1 Latent factor models

Given a set of N anomaly factors and a set of K latent factors observed over T time periods, we denote by $r_{t,i}$ the excess return of anomaly factor i at time t and by f_t^k the excess return of latent factor k at time t. In line with the APT framework, we assume that the anomaly factor returns approximately follow a linear factor model

of the form

$$r_{i,t} \approx \sum_{k=1}^{K} \beta_i^k f_t^k + \varepsilon_{i,t}, \quad \text{with} \quad i \in \{1, \ldots, N\} \text{ and } t \in \{1, \ldots, T\}, \tag{7.1}$$

where β_i^k represent the loading of anomaly factor i on latent factor k and $\varepsilon_{i,t}$ is an idiosyncratic error term. The errors are assumed to be homoskedastic with variance σ_ε^2 and uncorrelated from the latent factor returns: $\mathbb{E}\left[f_{k,t}\varepsilon_{i,t}\right] = 0$ for all k, i, t. In vector form, this becomes

$$r_t \approx \beta f_t + \varepsilon_t, \tag{7.2}$$

where $r_t \in \mathbb{R}^N$, $\beta \in \mathbb{R}^{N \times K}$, $f_t \in \mathbb{R}^K$, and $\varepsilon_t \in \mathbb{R}^N$.

In this chapter, we construct a parsimonious set of K latent factors with returns f_t that can explain the returns of the anomaly factors, and estimate the associated factor loadings β. More formally, since the anomaly factors are traded factors (i.e., f_t are interpreted as excess returns), we know from APT that $\mathbb{E}[\varepsilon_t]$ should be equal to zero and

$$\mathbb{E}[r_t] = \beta \cdot \mathbb{E}[f_t]. \tag{7.3}$$

The factor loadings β are interpreted as risk premia.

Following [151], the latent factors should minimize both the unexplained time-series variation in the anomaly factors as well as the pricing error relative to the APT model (Equation 7.3). More formally, if we denote by $\hat{r}_t$ the estimated anomaly factor returns (Equation 7.2) and by r_t the observed anomaly factor returns, the unexplained variation is given by

$$\begin{aligned}
\text{Unexplained Variation} &= \frac{1}{T} \cdot \sum_{t=1}^{T} \left[\frac{1}{N} \sum_{i=1}^{N} (\varepsilon_{i,t})^2 \right], && (7.4)\\
&= \frac{1}{T} \cdot \sum_{t=1}^{T} \left[\frac{1}{N} \sum_{i=1}^{N} (r_{i,t} - \hat{r}_{i,t})^2 \right], && (7.5)\\
&= \frac{1}{T \cdot N} \cdot \sum_{t=1}^{T} \sum_{i=1}^{N} (r_{i,t} - \beta_i f_t)^2, && (7.6)
\end{aligned}$$

where $\beta_i \in \mathbb{R}^{1 \times K}$, and the pricing error is given by

$$\begin{aligned}
\text{Pricing Error} &= \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \cdot \sum_{t=1}^{T} \varepsilon_{i,t} \right]^2, && (7.7)\\
&= \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \cdot \sum_{t=1}^{T} r_{i,t} - \frac{1}{T} \cdot \sum_{t=1}^{T} \hat{r}_{i,t} \right]^2, && (7.8)\\
&= \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \cdot \sum_{t=1}^{T} r_{i,t} - \frac{1}{T} \cdot \sum_{t=1}^{T} \beta_i f_t \right]^2. && (7.9)
\end{aligned}$$

Using the notation from [151], this joint minimization can be expressed as

$$
\begin{aligned}
(\beta^*, f^*) &= \underset{\beta,f}{\operatorname{argmin}} \frac{1}{T} \cdot \sum_{t=1}^{T} \left[\frac{1}{N} \sum_{i=1}^{N} (r_{i,t} - \hat{r}_{i,t})^2 \right] \\
&\quad + (1+\gamma) \cdot \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \cdot \sum_{t=1}^{T} r_{i,t} - \frac{1}{T} \cdot \sum_{t=1}^{T} \hat{r}_{i,t} \right]^2 \qquad (7.10) \\
&= \underset{\beta,f}{\operatorname{argmin}} \frac{1}{T \cdot N} \cdot \sum_{t=1}^{T} \sum_{i=1}^{N} (r_{i,t} - \beta_i f_t)^2 \\
&\quad + (1+\gamma) \cdot \frac{1}{N} \sum_{i=1}^{N} \left[\frac{1}{T} \cdot \sum_{t=1}^{T} r_{i,t} - \frac{1}{T} \cdot \sum_{t=1}^{T} \beta_i f_t \right]^2, \qquad (7.11)
\end{aligned}
$$

where the parameter γ controls the importance we want to attribute to the pricing error term relative to the unexplained variation. As a special case, $\gamma = -1$ corresponds to the usual PCA objective [151]. Intuitively, the unexplained variation term aims to improve the explainability power of the factor model in the time-series dimensions (i.e., improving the R^2) and obtain a better model of *returns*. Conversely, the pricing error term tries to improve the explainability power of the factor model in the cross-section (i.e., reducing the number of significant alphas) and obtain a better model of *average returns*. The parameter γ is then viewed as a trade-off between these two objectives. The APT loss can be written more compactly as

$$
\frac{1}{T \cdot N} \Sigma_{t,i} (r_{t,i} - \hat{r}_{t,i})^2 + (1+\gamma) \frac{1}{N} \Sigma_i (\bar{r}_i - \hat{\bar{r}}_i)^2, \qquad (7.12)
$$

where

$$
\bar{r}_i = \frac{1}{T} \cdot \sum_{t=1}^{T} r_{i,t} \quad \text{and} \quad \hat{\bar{r}}_i = \frac{1}{T} \cdot \sum_{t=1}^{T} \hat{r}_{i,t}. \qquad (7.13)
$$

The factor structure described in Equation 7.11 corresponds to the risk premium PCA objective developed in [151]. The factor structure we use here differs from [151] as we relax the linearity assumption when learning latent factors. In fact, we compress the information present in the anomaly factors using an autoencoder neural network. An autoencoder is composed of an input layer, at least one hidden layers, and a reconstruction layer (see Figure 7-1). Throughout the chapter, we set the input layer and the reconstruction layer to have N neurons and be interpreted as anomaly factor returns. Figure 7-1 describes a shallow autoencoder, which is the simplest type of autoencoder: this network contains only one hidden layer (ignoring the input layer and the reconstruction layer) with K neurons which corresponds to the latent factors learnt by the model. Since the number of neurons K in the hidden layer is small compared to the number of neurons N in the input/reconstruction layers, the network compresses the information contained in the anomaly factor returns into a set of K latent factors. More formally, the latent factors are obtained from the input

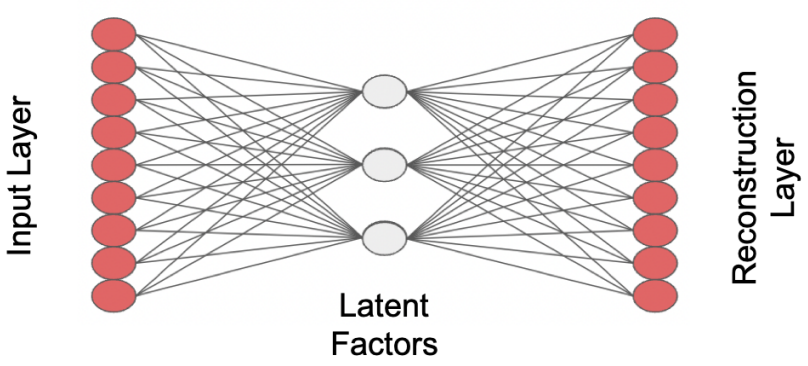


Figure 7-1: Architecture of a shallow autoencoder (AE). We represent here a shallow autoencoder with one hidden layer between the input and reconstruction layers. The hidden layer corresponds to the learnt latent factors while the input/reconstruction layers correspond to the anomaly factors.

data by learning a weight matrix $W^0 \in \mathbb{R}^{K\times N}$ and a bias vector $b^0 \in \mathbb{R}^K$ such that

$$f_t = g\left(b^0 + W^0 \cdot r_t\right). \tag{7.14}$$

The activation function $g(\cdot)$ can be chosen to be linear or nonlinear. The latter would allow the latent factors to depend on interactions of the input data in a nonlinear fashion. Common examples of nonlinear activation functions include the sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}$, the hyperbolic tangent function $\tanh(x)$, and the rectified linear unit (ReLU) function $\text{ReLU}(x) = \max(0, x)$. Although we mainly consider linear and $\tanh(\cdot)$ activation functions throughout the chapter, we explore the sensitivity of the autoencoder to different different choices of activation functions in Appendix E.5. The anomaly factor returns are then recovered from the latent factors as latent factors get decoded into the reconstruction layer through the learnt matrix $W^1 \in \mathbb{R}^{N\times K}$ and the bias vector $b^1 \in \mathbb{R}^N$ using

$$\hat{r}_t = g\left(b^1 + W^1 \cdot f_t\right). \tag{7.15}$$

The parameters b^0, b^1, W^0, and W^1 are obtained by minimizing the APT loss function

$$\min_{b^0,b^1,W^0,W^1} \frac{1}{T}\cdot\sum_{t=1}^{T}\left[\frac{1}{N}\sum_{i=1}^{N}(r_{i,t}-\hat{r}_{i,t})^2\right] + (1+\gamma)\cdot\frac{1}{N}\sum_{i=1}^{N}\left[\frac{1}{T}\cdot\sum_{t=1}^{T}r_{i,t} - \frac{1}{T}\cdot\sum_{t=1}^{T}\hat{r}_{i,t}\right]^2, \tag{7.16}$$

where

$$\hat{r}_t \;=\; g\left(b^1 + W^1 \cdot f_t\right), \tag{7.17}$$

$$= \quad g\left(b^1 + W^1 \cdot g\left(b^0 + W^0 \cdot r_t\right)\right). \tag{7.18}$$

The shallow AE can be easily generalized to a DAE by increasing the number of layers between the input/reconstruction layers and the layer of latent factors, as shown in Figure 7-2. Hidden layers are obtained from the input layer by recursively applying the transformation given in Equation 7.14 (with weight matrices and bias vectors of appropriate dimensions).

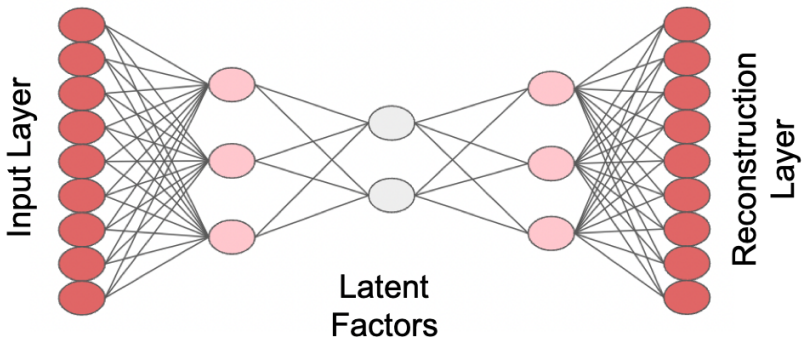


Figure 7-2: Architecture of the deep autoencoder (DAE). We represent here a deep autoencoder as there are now additional hidden layers (the light pink nodes) between the input/reconstruction layers and the layer of latent factors (in grey).

Autoencoders can be viewed as a generalization of PCA. In fact, [107] show that a shallow linear AE is equivalent to PCA up to a rotation matrix when the AE objective is to minimize the unexplained variation (see Proposition 1 of [107]). In our case, the shallow linear AE is related to the risk premia PCA from [151] given the APT loss function used. One key difference between AEs and PCA is that the latent factors learnt by an AE are not uncorrelated. We address this issue in the next section by using a recursive AE.

7.2.2 Autoencoder models

We now describe the different autoencoder architecture used to construct latent factor models. All autoencoder networks in this chapter are designed and trained using *Tensorflow* [1]. As described earlier, we instill economic intuition to the models by always using the APT loss function (Equation 7.12 and Equation 7.16) when training an autoencoder. This allows us to minimize the unexplained variation while penalizing the pricing error relative to APT. We use cross-validation to select the parameter γ using the metrics described in Section 7.2.3 (see Appendix E.3) and set $\gamma = 10$ in all autoencoder models. In fact, as shown in Appendix E.3, the autoencoder models are robust to variations in the choice of γ, with the better choices lying between

$\gamma = 5$ and $\gamma = 15$. A value $\gamma = 10$ is not an unreasonable choice and this value has been used successfully in [151]. The idea here is that we want to focus mainly on the pricing error while ensuring that the unexplained variation would be small. It is not advisable to ignore the unexplained variation penalty term as we want to capture some time-series variability. In fact, ignoring completely the unexplained variation term could lead to erroneous inferences: we could construct a factor model that prices all assets correctly simply by including a constant factor in the model. This constant factor would play the role of the constant term in the regression and capture all the alpha rather than being interpreted as a beta.

Throughout the chapter, we will consider both linear activation functions as our baseline and hyperbolic tangent functions to learn nonlinear dependencies. ReLU and sigmoid activation functions are discussed in Appendix E.5.

Shallow autoencoders The baseline model we consider consists of a shallow autoencoder (Figure 7-1) with linear activation functions and an APT loss function. This consists of an AE with a single hidden layer of K neurons: the layer of latent factors we are learning. The input and reconstruction layers are composed of $N = 150$ neurons each, with each neuron representing the returns of an anomaly factor. This linear AE is related to the risk premia PCA approach as shown in Proposition 1 (proof in Appendix E.1). We then extend this AE by consider a shallow AE with nonlinear activation functions.

Proposition 1. *Denote by R and $\hat{R}$ the matrices in $\mathbb{R}^{N\times T}$ with entries equal to $r_{i,t}$ and $\hat{r}_{i,t}$ respectively. Then Equation 7.16 can be written more compactly as:*

$$\min_{b^0,b^1,W^0,W^1} \frac{1}{N\cdot T}\cdot\|R-\hat{R}\|_F^2+(1+\gamma)\cdot\frac{1}{N\cdot T^2}\|(R-\hat{R})\cdot\iota\|_F^2, \tag{7.19}$$

where $\iota\in\mathbb{R}^{T\times 1}$ is a vector of ones, $\|\cdot\|_F$ is the Frobenius norm of a matrix, and

$$\hat{R}=g\left(b^1\cdot\iota'+W^1\cdot g\left(b^0\cdot\iota'+W^0\cdot R\right)\right). \tag{7.20}$$

Equivalently, Equation 7.19 can be written in terms of the matrix trace operator as:

$$\min_{b^0,b^1,W^0,W^1} \frac{1}{N\cdot T}\cdot\text{trace}\left((R-\hat{R})'(R-\hat{R})\right)+(1+\gamma)\cdot\frac{1}{N\cdot T^2}\text{trace}\left(\iota'\cdot(R-\hat{R})'(R-\hat{R})\cdot\iota\right). \tag{7.21}$$

Under a linear activation function $g(x)=x$, the factors obtained through Equation 7.19 as $F=g\left(b^0\cdot\iota'+W^0\cdot R\right)$ are equivalent, up to a rotation, to those obtained by applying PCA to

$$\frac{1}{N\cdot T}\cdot R\cdot\left(I_T+\frac{1+\gamma}{T}\iota\iota'\right)\cdot R', \tag{7.22}$$

where $I_T\in\mathbb{R}^{T\times T}$ is the $T\times T$ identity matrix.

In other words, a shallow linear autoencoder with APT loss is equivalent to the risk premia PCA proposed by [151].

We incorporate two additional modifications to all the autoencoders used in the chapter to learn latent factors more efficiently. First, the input returns from anomaly factors are standardized but not demeaned. We observe that dividing input returns by their standard deviation makes the network much more robust, especially when we consider nonlinear interaction functions or deeper autoencoders, while demeaning input returns leads to less efficient training. Second, we set all bias terms in the network (e.g., b^0 and b^1) to zero. This allows latent factors to better capture the anomalous part of the input returns instead of viewing it as an unpriced intercept term. Zero-bias autoencoders also have regularization advantages when using an appropriate shrinkage function [144].

Deep autoencoders (DAEs) As described in Figure 7-2, a DAE is a shallow autoencoder but with additional hidden layers between the input/reconstruction layers and the layer of latent factors. In general, DAEs are harder to train than shallow AEs as there are more network weights to learn, especially when the amount of data is limited. For this reason, we only consider a DAE with 3 hidden layers (Figure 7-2): the middle layer (the grey neurons in Figure 7-2) corresponds to the latent factors learnt by the network and the two other hidden layers (the light pink neurons in Figure 7-2) comprise of 20 neurons each. Although other DAE architectures can be explored such as asymmetric DAEs (with hidden layers of different sizes), we did not find it to improve the performance of the network. To reduce the number of weights in the network and improve its trainability, we refine the DAE into a CAE.

Clustered autoencoders (CAEs) Figure 7-3 CAEs can be viewed as a special case of DAE. Instead of using dense connections between the first two and the last two layers of the network, we instead cluster the input nodes into groups and only connect nodes within a group to the neurons of the first hidden layer (Figure 7-3). The same architecture holds for the reconstruction layer. More precisely, we cluster the 150 anomaly factors from the input layer into 60 groups, connect inputs to their corresponding 60 nodes in the first hidden layer, and compress them into K latent factors using a dense layer. When a group contains 5 or more anomaly factors, we connect densely it to 2 nodes in the first hidden layer.

CAEs are an improvement over DAEs for two main reasons. First, we reduce the number of parameters the model needs to learn, serving as a powerful regularization tool. Second, we instill economic intuition into the model by grouping similar anomaly factors together and extracting their common component. In fact, having groups of anomaly factors that are correlated can lead to ineffective feature extraction as the latent factors will tend to capture only the components of anomaly factors that are over-represented in the data, at the expense of other anomaly factors that capture unique but important components of the cross section of asset returns. Hence, by design, the CAE will tend to mitigate this over-representation bias.

The other degree of freedom in the CAE is the number of clusters we want to use to group anomaly factors together. To this end, we use k-means clustering and analyze the inertia and silhouette score of the clusters (see Figure E-6 in Appendix E.4).

We find that the optimal number of cluster to use falls between 60 to 80 clusters. Throughout the chapter, we fix the number of clusters to 60 to reduce the size of the first and third hidden layers. Out of 60 clusters, 34 represent single anomaly factors, 10 represent two anomaly factors and 5 represent three anomaly factors (see Table E.2 in Appendix E.4). On the other side of the spectrum, one cluster represents 13 anomaly factors, one cluster represents 12 anomaly factors, and two clusters represents 11 anomaly factors. A description of the clusters obtained are reported in Table E.3 (Appendix E.4) and sensitivity of the model to the number of clusters is discussed in Appendix E.4.

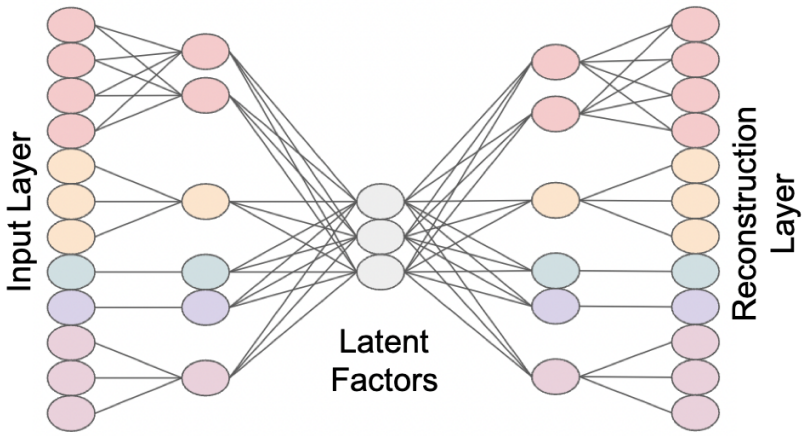


Figure 7-3: Architecture of the clustered autoencoder (CAE).

Recursive autoencoders As discussed above, one main difference between latent factors obtained through PCA and AEs is that PCA latent factors will be uncorrelated. We address this issue by proposing recursive AEs which aim to reduce correlations among latent factors by learning one latent factor at a time and immediately removing the exposure of anomaly factors to this latent factor. We find that recursive autoencoders tend to perform better than regular AEs. This can be attributed to different reasons. For example, using uncorrelated latent factors can improve the efficiency when estimating factor loadings, not only by avoiding cases of multicollinearity, but also by leading to more efficient standard error estimations. Furthermore, learning one factor at a time reduces the complexity of the AE, providing a regularization advantage. It is also desirable to use a recursive AE if we want to obtain latent factors that are sorted by their importance. With a recursive AE, we know that the first factor will be the dominant one, the second factor will be important but slightly less than the first one, etc. This ordering related the recursive AE latent factors to PCA latent factors which are always sorted in decreasing level of importance. This property can become helpful if we perturb the inputs to the network, for example during a bootstrap procedure.
In this chapter, we consider recursive versions of the shallow AE, the DAE, and the CAE. We call the latter a recursive clustered autoencoder (RCAE). A potential

drawback of using recursive AEs occurs when we use nonlinear activation functions. While the latent factors learnt in the linear case would be equivalent (up to a rotation) for a recursive and non-recursive autoencoder, they become very different for networks with nonlinear activation functions.

Split autoencoders (SAEs) Finally, the last variation on the autoencoder we explore is a split autoencoder (SAE). The SAE can be based on shallow AEs, DAEs, and CAEs under general activation functions without affecting the network architecture. The only difference resides in the input and reconstruction layers: instead of having N nodes per layer, we use $2N$ nodes in each layer. The duplication happens as follows: for each anomaly factor return, if the return is positive at time t, then the first neuron of the pair will take the value of the return while the second will be set to zero; conversely, we set the first neuron to zero and the second return to the value of the return if the anomaly factor return is negative at time t. This type of network is designed to learn asymmetric factor loadings: the latent factors can now depend on positive and negative anomaly factor returns in different ways.

7.2.3 Performance evaluation

The factor models obtained are then tested through time series regressions. There are two ways of testing asset pricing models. The first involves comparing the adjusted R^2 of different models. The R^2 metric quantifies how well the proposed factor model is able to explain *returns* (instead of *average returns*). In other words, R^2 is about capturing time series effects. Although we provide R^2 metrics in the results, this consideration is secondary. Our objective is to explain the cross section of the returns, which means that we are interested in a model of *average returns* [51]. In this case, we test a factor model using a time series regression with the test returns as dependent variables and latent factors as explanatory variables. If latent factors are obtained in a nonlinear way, we can no longer assume that we have traded factors and need to use a cross-sectional regression. The asset pricing model would be valid if the intercepts obtained are jointly equal to zero. Although a Gibbons-Ross-Shanken (GRS) test [92] or a Fama-Macbeth approach [76] can be used, we estimate robust standard errors and relevant test statistics using generalized method of moments (GMM) estimators [51, 139]. It is also helpful to estimate the intercept's t-statistic obtained for each test return to understand how many assets can and cannot be explained by the proposed factor model. Assets with an intercept t-statistic above 1.96 are considered to be *unexplained* by the proposed factor model.

In addition to linear time series regressions, we perform polynomial time series regressions of degree 2 and split time series regressions. In a polynomial regression, we include all interaction terms between covariates but do not include quadratic terms of covariates. Split regressions consist in splitting covariates into their positive component and their negative component. In other words, instead of using x_t as a regressor, we would use both $x_t^+ = \max(x_t, 0)$ and $x_t^- = \min(x_t, 0)$. These nonlinear regressions are used to detect whether a nonlinear structure is needed to explain the cross section of returns.

7.3 Empirical analysis

In this section, we apply the methodology described in Section 7.2 to anomaly factors between 1976 and 2017 and test the latent factor models on classical anomaly portfolios.

7.3.1 Data description

Latent factor models are constructed using the 150 publicly available anomaly factors replicated by [85]. The data consists of monthly returns for each factor from July 1976 to December 2017, as described in Table E.1 in Appendix E.2. Each of the 150 factors contains 498 time-series return values over time with no missing values. As a preprocessing step, we normalize anomaly factor returns by their variance.

The latent factor models are then tested on various anomaly portfolios between July 1976 and December 2017. We first use 75 Fama-French portfolios, obtained from Kenneth French's online data library [90]. This dataset includes 25 double-sorted portfolios formed on size and book-to-market, 25 double-sorted portfolios formed on size and operating profitability, and 25 double-sorted portfolios formed on size and investment. Furthermore, we use 374 Hou-Xue-Zhang anomaly portfolios obtained from the *q-data* library [225]. This dataset covers 187 anomalies across 6 categories: momentum (41), value-versus-growth (32), investment (29), profitability (45), intangibles (30), and frictions (10). We obtain $187 \times 2 = 374$ portfolios from the lowest and highest decile for each anomaly.

Finally, we test our latent factor models on mutual fund and hedge fund returns. We consider $8,866$ mutual fund portfolios from the Center for Research in Security Prices (CRSP) Mutual Funds database through Wharton Research Data Services (WRDS) and from the Lipper databse through the Refinitiv workspace. These funds are filtered in the following way: funds must have at least \$100 million in net asset value and must have at least 12 active months between 1976 and 2017. These mutual funds cover the following 6 asset classes: equity $(4,480)$, bond $(2,242)$, mixed assets $(1,132)$, money market (604), alternatives (356), and commodity (47). Hedge fund returns are obtained from the Morningstar Center for International Securities and Derivatives Markets (CISDM) Database and cover $6,619$ active and defunct funds. Similarly to mutual funds, hedge funds must have at least \$10 million in net asset value and must have at least 12 active months between 1976 and 2017.

7.3.2 Empirical results

We now test the latent factors constructed in Section 7.2 on three test assets: the set of 150 anomaly factors (which we trained our models on), the 75 Fama-French anomaly portfolios (FF), the 374 Hou-Xue-Zhang anomaly portfolios (HXZ), and the $8,866$ mutual fund returns (MF). The $6,619$ hedge fund returns (HF) will be analyzed in the forthcoming paper. We find that the latent factors tend to perform better than the classical CAPM, FF5, and q5 factor models. In fact, we report in Table 7.1 the fraction of test assets that are explained by each factor model: we consider an asset

to be explained by a factor model if the regression's intercept is associated with a t-statistic below 1.96 (corresponding to a type I error rate of 2.5%). Standard errors are estimated using heteroskedasticity robust continuously updated GMM estimators with Bartlett kernel[1]. The regressions are performed initially only with the latent factors, however we find that the market factor is not always captured by the latent factor models (see Table 7.3). For this reason, we also include the regression results when the market factor is added to the latent factors. We find that, in a few cases, including a market factor improves the performance of the latent factor models. While these results correspond to linear regressions, we also explore the performance of latent factor models when we allow interaction terms in the regression (this is a polynomial regression of order 2 with interaction terms but without second order terms). The results are displayed in Table E.4 in Appendix E.6.

Table 7.1: Fraction of explained excess returns (in %). Brackets indicate the fraction obtained if we include a market factor.

Model	Factors	FF	HXZ	MF
CAPM	50	65	48	70
FF5	63	80	57	67
q5	83	87	80	69
Linear 6	83 (81)	4 (48)	6 (82)	68 (68)
RLinear 5	69 (67)	72 (96)	61 (77)	68 (68)
RLinear 6	87 (85)	5 (72)	8 (83)	71 (71)
R-Linear 9	93 (90)	32 (87)	53 (80)	72 (72)
CAE 7	83 (82)	9 (64)	16 (73)	67 (67)
CAE 8	81 (82)	4 (53)	11 (77)	67 (67)
RCAE 5	69 (65)	69 (95)	65 (77)	68 (68)
RCAE 7	84 (85)	5 (63)	11 (84)	70 (70)
Deep 5	71 (69)	7 (65)	13 (72)	66 (66)
Deep 8	81 (79)	5 (72)	7 (80)	68 (68)
Tanh-Tanh 6	82 (81)	80 (99)	58 (83)	72 (72)
R-Tanh-Tanh 5	95 (95)	17 (56)	20 (82)	72 (72)
R-Tanh-Tanh 8	95 (95)	5 (76)	10 (84)	71 (71)
CAE 6 Tanh-Tanh	75 (77)	97 (81)	94 (76)	66 (66)
RCAE 4 Tanh-Tanh	75 (75)	73 (99)	57 (76)	68 (68)
Deep 5 Tanh-Tanh	79 (77)	20 (99)	28 (73)	68 (68)
Tanh 6	81 (81)	61 (99)	52 (83)	71 (71)
R-Tanh 4	83 (82)	97 (99)	85 (78)	70 (70)
CAE 6 Tanh	78 (78)	27 (39)	52 (72)	66 (66)
RCAE 5 Tanh	72 (72)	45 (97)	48 (78)	69 (69)

Continued on next page

[1]GMM estimators are obtained from the *linearmodels* Python library `https://bashtage.github.io/linearmodels/`.

Table 7.1 – continued from previous page

Model	Factors	FF	HXZ	
Model	Factors	FF	HXZ	MF
Deep 6 Tanh	71 (70)	7 (93)	19 (71)	68 (68)
ReLU 4	87 (77)	17 (97)	7 (87)	88 (88)
ReLU 9	79 (75)	60 (93)	37 (85)	83 (83)
R-ReLU 6	67 (64)	19 (95)	20 (79)	82 (82)
CAE 6 ReLU	76 (70)	92 (41)	90 (77)	84 (84)
RCAE 4 ReLU	61 (55)	16 (83)	36 (43)	69 (69)
Deep 5 ReLU	54 (60)	7 (48)	45 (67)	80 (80)
Sigmoid 4	44 (43)	95 (76)	82 (62)	65 (65)
R-Sigmoid 5	65 (65)	27 (76)	46 (77)	73 (73)
CAE 5 Sigmoid	51 (49)	41 (60)	12 (70)	83 (83)
RCAE 5 Sigmoid	21 (25)	0 (51)	1 (27)	57 (57)
Deep 5 Sigmoid	43 (45)	59 (56)	67 (46)	75 (75)

Furthermore, The average adjusted R^2 are evaluated for each set of test assets, as reported in Table 7.2. In terms of adjusted R^2, we find that FF5 and q5 factor models tend to perform better than the latent factors we constructed. However, including a market factor in the latent factor model improves significantly the performance of the models. Analogous results for polynomial regressions are presented in Table E.5 of Appendix E.6.

Table 7.2: Regression adjusted R-squared (in %). Brackets indicate the fraction obtained if we include a market factor.

Model	Factors	FF	HXZ	MF
CAPM	18	75	76	52
FF5	54	93	83	57
q5	46	90	83	57
Linear 6	65 (68)	62 (86)	59 (86)	17 (58)
RLinear 5	63 (66)	59 (85)	57 (85)	16 (58)
RLinear 6	65 (68)	61 (86)	59 (86)	17 (58)
R-Linear 9	70 (72)	64 (89)	63 (87)	22 (60)
CAE 7	67 (70)	62 (88)	60 (86)	19 (59)
CAE 8	68 (71)	62 (88)	61 (86)	20 (59)
RCAE 5	63 (66)	58 (85)	57 (85)	16 (58)
RCAE 7	67 (70)	62 (88)	61 (86)	20 (59)
Deep 5	64 (66)	61 (86)	59 (86)	16 (58)

Continued on next page

Table 7.2 – continued from previous page

Model	Factors	FF	HXZ	MF
Deep 8	69 (71)	63 (89)	62 (87)	21 (59)
Tanh-Tanh 6	45 (48)	48 (81)	49 (82)	13 (56)
R-Tanh-Tanh 5	50 (52)	57 (83)	54 (83)	11 (56)
R-Tanh-Tanh 8	56 (59)	59 (85)	57 (84)	17 (58)
CAE 6 Tanh-Tanh	58 (61)	56 (85)	55 (84)	15 (57)
RCAE 4 Tanh-Tanh	52 (55)	53 (83)	53 (83)	11 (56)
Deep 5 Tanh-Tanh	57 (60)	55 (85)	53 (84)	14 (57)
Tanh 6	46 (49)	49 (81)	49 (82)	13 (57)
R-Tanh 4	50 (52)	52 (82)	52 (83)	10 (55)
CAE 6 Tanh	62 (65)	58 (87)	57 (85)	17 (58)
RCAE 5 Tanh	60 (63)	57 (85)	56 (85)	15 (58)
Deep 6 Tanh	62 (65)	58 (86)	56 (86)	17 (58)
ReLU 4	52 (56)	39 (82)	43 (82)	9 (55)
ReLU 9	64 (66)	61 (86)	59 (86)	17 (58)
R-ReLU 6	55 (59)	45 (84)	48 (84)	12 (56)
CAE 6 ReLU	54 (58)	37 (82)	41 (83)	12 (56)
RCAE 4 ReLU	46 (53)	25 (81)	28 (82)	10 (55)
Deep 5 ReLU	55 (59)	46 (83)	47 (84)	14 (57)
Sigmoid 4	38 (42)	42 (80)	44 (81)	10 (56)
R-Sigmoid 5	39 (43)	39 (80)	41 (81)	8 (55)
CAE 5 Sigmoid	45 (49)	40 (80)	45 (82)	10 (55)
RCAE 5 Sigmoid	37 (42)	32 (77)	38 (80)	10 (56)
Deep 5 Sigmoid	48 (52)	42 (80)	47 (82)	11 (56)

We now compare the latent factor models directly to the three baseline factor models. In Table 7.3, we show how many factors from the CAPM, the FF5, and the q5 factor models can be explained by each latent factor model. This is done by regressing the each baseline factor on a latent factor models and testing whether the intercept is significantly different from zero. We also add a $(*)$ to the number displayed when one of the factors that cannot be explained by a latent factor model is the market factor. Table 7.3 clearly shows that many latent factor models do not capture the market factor, which explains why adding the market factor to these models can significantly improve their performance. Analogous results for polynomial regressions are presented in Table E.6 of Appendix E.6.

Table 7.3: Number of unexplained CAPM, FF5, and q5 factors. (*) indicates that one of the unexplained factors is the market factor.

Model	CAPM	FF5	q5
Linear 6	1*	2*	3*
RLinear 5	0	1	2
RLinear 6	1*	1*	3*
R-Linear 9	0	0	1
CAE 7	1*	2*	3*
CAE 8	1*	2*	3*
RCAE 5	0	0	2
RCAE 7	1*	1*	3*
Deep 5	1*	2*	3*
Deep 8	1*	2*	3*
Tanh-Tanh 6	0	0	1
R-Tanh-Tanh 5	1*	1*	2*
R-Tanh-Tanh 8	1*	1*	3*
CAE 6 Tanh-Tanh	0	0	1
RCAE 4 Tanh-Tanh	0	0	1
Deep 5 Tanh-Tanh	1*	2*	2*
Tanh 6	0	0	1
R-Tanh 4	0	0	1
CAE 6 Tanh	0	1	3
RCAE 5 Tanh	1*	1*	2*
Deep 6 Tanh	1*	2*	3*
ReLU 4	1*	1*	3*
ReLU 9	1*	1*	2*
R-ReLU 6	1*	2*	3*
CAE 6 ReLU	0	4	2
RCAE 4 ReLU	1*	2*	4*
Deep 5 ReLU	1*	3*	4*
Sigmoid 4	0	2	3
R-Sigmoid 5	1*	3*	2*
CAE 5 Sigmoid	1*	4*	4*
RCAE 5 Sigmoid	1*	5*	5*
Deep 5 Sigmoid	0	2	4

Conversely, we explore in Table 7.4 whether the CAPM, the FF5, and the q5 factor models can explain each latent factor constructed. We find that, while the CAPM and the FF5 models fail to explain most of the latent factor models, the q5 model can explain many latent factors obtained under linear activation functions but fails to explain most latent factors obtained under nonlinear activation functions.

Analogous results for polynomial regressions are presented in Table E.7 of Appendix E.6.

Table 7.4: Number of factors unexplained by CAPM, FF5, and q5.

Model	CAPM	FF5	q5
CAPM	—	1	0
FF5	3	—	0
q5	4	2	—
Linear 6	3	4	1
RLinear 5	2	3	2
RLinear 6	3	3	2
R-Linear 9	3	5	3
CAE 7	3	3	1
CAE 8	3	4	1
RCAE 5	3	3	2
RCAE 7	5	4	4
Deep 5	3	2	0
Deep 8	3	3	1
Tanh-Tanh 6	5	5	4
R-Tanh-Tanh 5	3	4	2
R-Tanh-Tanh 8	4	4	2
CAE 6 Tanh-Tanh	4	5	2
RCAE 4 Tanh-Tanh	3	3	2
Deep 5 Tanh-Tanh	3	2	2
Tanh 6	5	5	4
R-Tanh 4	3	4	2
CAE 6 Tanh	2	3	2
RCAE 5 Tanh	2	2	2
Deep 6 Tanh	5	4	2
ReLU 4	4	4	4
ReLU 9	9	9	9
R-ReLU 6	6	6	6
CAE 6 ReLU	5	5	5
RCAE 4 ReLU	2	3	3
Deep 5 ReLU	4	5	5
Sigmoid 4	4	4	4
R-Sigmoid 5	5	5	5
CAE 5 Sigmoid	5	5	5
RCAE 5 Sigmoid	5	5	5
Deep 5 Sigmoid	5	5	5

7.4 Discussion

As discussed in Section 7.2, latent factor models can be constructed in several ways. The empirical analysis conducted in Section 7.3 allows us to make a few interesting observations when comparing different categories of latent factor models.

First, among linear latent factor models, we find that the recursive autoencoders tend to outperform non-recursive autoencoders, suggesting that the recursive procedure has beneficial regularization effects for linear networks. This is no longer the case when we consider nonlinear activation functions. Second, clustered autoencoders tend to outperform deep autoencoders and generalize better, as expected since CAEs are less complex than DAEs and incorporate economic intuition in the clustering step.

Third, linear, hyperbolic tangent, and ReLU activation functions tend to perform better than sigmoid activation functions. This can be explained by the fact that sigmoid activation functions constrain the output to the $[0, 1]$ interval while the similar *tanh* function is symmetric around zero and has a range of $[-1, 1]$. Hence, returns that are close to zero will remain close to zero under *tanh* (as well as linear and ReLU) activation functions, but will be closer to 0.5 under the sigmoid.

Overall, we find that the latent factor models tend to perform better than the benchmark CAPM, FF5, and q5 factor models. Based on Table 7.3, we observe that latent factors can explain most of the FF5 factors except for the market factor. However, latent factor models with nonlinear activation functions are often needed to explain q5 factors. Conversely, based on Table 7.4, FF5 factors fail to explain latent factors while q5 factors can mainly explain latent factors with linear activation functions.

The additional analyses presented in Appendix E.5 show that the above observations are robust to the choice of the number of latent factors and that, although the performance tends to improve as we increase the number of factors in the model, linear and hyperbolic tangent activation functions outperform the other models and tend to be robust to the size of the factor model. Latent factor models with linear activation functions tend to present the highest values of R^2, suggesting that the factors are able to better capture the time series variation in the returns. The models are also robust to the choice of γ in the APT loss function as shown in Appendix E.3.

We also consider nonlinear regression models through the polynomial regressions presented in Appendix E.6. Nonlinear stochastic discount factor models are equivalent to nonlinear factor models, as shown by [66] (Theorem 1). In fact, [66] argue that linear factor models assume a gaussian distribution in the returns while nonlinear factor models are important to consider when the distribution of returns is non-gaussian. Comparing Table 7.1 to Table E.4 (Appendix E.6), interaction terms do not significantly improve the fraction of explained excess returns for latent factors constructed from linear networks. However, we find interaction terms helpful when considering factors derived using nonlinear activation functions. This can also be seen by comparing the middle row to the top row of the plots displayed in Appendix E.5. We also observe that interaction terms allow nonlinear latent factors to explain better the benchmark factor models (Table E.6).

7.5 Conclusion

Unsupervised learning is a powerful tool for dimensionality reduction. While PCA is helpful to project high-dimensional data on a low-dimensional subspace, AEs are also able to account for nonlinear interactions in the data. We have proposed a methodology to apply autoencoders directly to a large set of anomaly factor returns to construct latent factor models that capture most of the information in the factor zoo. In addition, we incorporate economic intuition by using a modified loss function inspired by [151] and by clustering the input factors when using deep autoencoders. We have also found recursive autoencoders to be helpful for obtaining uncorrelated factors, but also for regularization purposes. As a proof-of-concept, we have applied this methodology to the set of anomaly factors used in [85].

This framework can also be extended with a focus on interpretability [64, 65]. For example, similarly to PCA outputs, latent factors derived from recursive autoencoders are ranked by their relative importance (by design). This stability in the output of the recursive autoencoder can allow for confidence interval estimation through a block bootstrap approach [111].

Part III

Managing Risks & Uncertainty in Investment Management

Chapter 8

Measuring Risk Preferences and Asset-Allocation Decisions

In the final portion of this thesis, we turn our attention to the space of investment management from a risk management perspective. More specifically, we return to a topic explored in Chapters 2, 3, and 4: estimating the risk and uncertainty-preferences of individuals. Instead of looking at patients, the individuals we consider here are regular investors, financial advisors, and institutional investors. The goal of the procedure is to understand the risk preferences and asset-allocation decision of each class of investor under various hypothetical scenarios.

In particular, we use a global survey of individual investors, financial advisors, and institutional investors between 2015 and 2017 to elicit their asset allocation behavior and risk preferences. We find drastically different behavior among these three groups of market participants. Most institutional investors exhibit highly contrarian reactions to past returns in their equity allocations. Financial advisors are also mostly contrarian; a few of them demonstrate passive behavior. Individual investors are, on average, extrapolative. To investigate further, we use a clustering algorithm to partition individuals into five distinct types: passive investors, risk avoiders, extrapolators, contrarians, and optimistic investors. Across demographic categories, older investors tend to be more passive and more risk-averse.

8.1 Background

There are three major groups of participants in the investment management industry: individual investors, financial advisors, and institutions. Each of these groups has its own risk preferences and behavioral characteristics it uses in its investment decisions. We study the behavior of these groups by using the results of three comprehensive global surveys between 2015 and 2017[1], each of them covering over 7,000 individuals, over 2,300 advisors, and over 660 institutional investors.

[1]The survey participants were randomly sampled each year. Some participants may be surveyed multiple times across the three years.

The breadth of our dataset sets it apart from earlier survey data in the literature. To the best of our knowledge, these surveys are the first to present the same set of questions to three distinct groups of market participants over three consecutive years. This dataset covers over 17 countries in the Individual Investor surveys and over 14 countries in the Financial Advisor surveys every year. This global breadth provides us with insight into investment behavior by country, and allows us to compare survey results across countries. Finally, all our survey subjects have a significant stake in the market: all the surveyed individual investors have a net worth above $200,000, while the financial advisors and the institutional investors are employed in the financial industry. As a result, their answers will generally be more realistic and have greater relevance for modeling investor behavior than the results of surveying students in a laboratory setting, as many other studies have done. Our main goal is to understand how different market participants and different types of individuals compare along the dimensions of risk aversion and asset allocation. To this end, we poll members of these groups about their investment decisions under various historical and hypothetical scenarios. We obtain two sets of results. The first set of results shows that investors tend to be significantly more risk-averse and mostly extrapolative in their asset allocation, while institutions tend to be significantly less risk-averse and mostly contrarian in their investment decisions, with advisors falling in the middle of the risk aversion scale while also following a contrarian asset allocation strategy. The second set of results focuses on just individual investors—using a clustering algorithm applied to survey responses, we are able to identify five distinct types of investors: *passive investors*, *extrapolators*, *risk avoiders*, *contrarians* and *optimistic investors*. *Extrapolators* tend to decrease allocation in equities following bad market performance, and tend to increase allocation following good returns, extrapolating past trends. *Passive investors* leave their allocation unchanged in either scenario. *Risk avoiders* significantly cut their allocation to equities when they see large moves in the S&P 500 in either direction. *Contrarians* tend to increase allocation in equities following bad market performance, and tend to decrease allocation following good returns. *Optimistic investors* tend to increase their allocation in either scenario.

While the largest cluster of individuals in our dataset corresponds to passive investors, it also contains a significant number of risk avoiders and extrapolators. Evidence for each of these types is found in the literature, although most papers focus only on one type at a time. [5] and [59] document that a large proportion of investors make no changes to allocations within their retirement portfolios over spans of several years. They note that this phenomenon may be linked to *inertia*, a widely recognized behavioral bias. Some major papers documenting extrapolation include [62], [101] using survey data, as well as [19] and [49] which look at historical 401(k) account allocations. Finally, [18] investigate individual trading records and derive a V-shaped probability distribution of selling a stock as a function of profit. This in part may be driven by the cluster of risk avoiders identified from our survey.

In comparison, we find that most financial advisors and institutional investors are contrarian in allocation strategy—that is, they would change equity allocation in the direction opposite to recent returns on the S&P 500. This contrasts with the overall

behavior of individual investors, who on average are extrapolators.[2] The differences in the reactions across these three groups of market participants are significant and very large. We note that a few earlier studies have viewed individuals as momentum traders, and institutions as contrarians.[3] However, these studies consider shorter-term horizons than ours, and focus on trading behavior. Our survey asks about asset allocation, a strategic and longer-term investment decision, rather than short-term trading, which potentially could be affected by excessive speculation on the part of individual traders, or by liquidity considerations of institutions. A recent paper by [63] does focus on asset allocation of Dutch institutional investors, and concludes that they tend to be contrarians.

Our results have another important implication, one that arises from the differences in response between financial advisors and individual investors. We find that advisors generally advise their clients to change their allocation in the opposite direction of the typical preference of the individual investor. It may be that advisors recognize the excessive tendency of investors toward extrapolation, and try to mitigate this effect by giving "contrarian advice". Also, the proportion of advisors who suggest a significant decrease in equity allocation when seeing large S&P 500 moves is much smaller than the proportion of individual investors who would implement such a change. As a result, advisors may also provide the significant benefit of ensuring their investors stay invested in the markets despite periods of high volatility, and hence earn higher returns in the long run.[4] Overall, our findings suggest that financial advisors are of direct benefit to most individual investors.

Finally, we compare risk aversion across the three groups, as well as within investor demographic categories. Individual investors are significantly more risk-averse than financial advisors, who are in turn more risk-averse than institutional investors. We find that individual risk aversion increases with age, which is consistent with previous literature linking risk aversion to age, wealth, and education; see [164], [178], and [110].

In Section 8.2 we outline the survey methodology and the estimation of risk aversion. We compare survey responses across individuals, advisors, and institutions in Section 8.3. We focus on individual investors in more detail in Section 8.4 and we conclude in Section 8.5.

8.2 Methodology

We use data from three separate but closely related surveys: the Natixis Global Survey of Individual Investors (from 2015 to 2017), the Natixis Global Survey of Financial Advisors (from 2015 to 2016), and the Natixis Global Survey of Institutional Investors (from 2015 to 2017). Each survey involved two sets of questions. The first

[2]While a large number of individual investors are passive or risk avoiders, both of these groups have symmetric reactions to large moves in the S&P 500: they either do nothing, or they significantly decrease equity allocation.

[3]See, for example, [102], Jackson (2003) [127], and Kaniel, Saar, and Titman (2008) [136].

[4][219] find that investors using a financial advisor are 1.5 times more likely to stick to long-term investment decisions.

set originated with Natixis Asset Management for their own research purposes. The second set was created by us, in coordination with Natixis, for studying the behavioral aspects of investor decision-making.

8.2.1 Survey Questions

We asked three similar questions in each survey for Individual Investors, Financial Advisors and Institutional Investors between 2015 and 2017. Furthermore, we have asked one additional question specifically to Individual Investors which differed each year. The first involved preferences among potential gambles, and was used to elicit the risk aversion coefficient among respondents. The second and third asked how respondents would change their investment allocation as a result of large negative or positive moves in the S&P 500. Possible responses were "significantly decrease equity allocation", "slightly decrease equity allocation", "do nothing", "slightly increase equity allocation", and "significantly increase equity allocation". In the case of financial advisors, we asked them how they would advise their clients to act in such situations. The additional question specifically targeted Individual Investors varied every year: in 2015, investors were asked if and when they decreased their allocation during the 2007–2009 Financial Crisis; in 2016, investors were asked how they modified their allocation based on the market volatility in January 2016; in 2017, investors were asked how they adjusted their allocation immediately after the U.S. presidential election on November 8, 2016. The exact formulation of the questions is included in the Appendix.

Natixis Asset Management commissioned CoreData Research to conduct each survey via an online questionnaire. The Individual Investor Survey was first carried out in March 2015 and involved 7,000 individuals from 17 countries. Each investor needed to have a minimum net worth of $200,000 (or Purchasing Power Parity equivalent) to participate. The survey was carried out again in February–March 2016 with 7,100 individuals across 19 countries, and in February–March 2017 with 8,300 individuals across 22 countries. The Financial Advisor Survey was first conducted in June–July 2015 and involved 2,400 advisors from 16 countries. Since some advisors opted to not complete the behavioral part of the survey, we had a total of 2,342 advisor observations. The survey was conducted again in July 2016 with 2,550 advisors across 15 countries. The Institutional Investor Survey was first carried out in October 2015 and involved 660 respondents from 29 countries. All respondents had to be decision makers working in the institutional investment industry, such as Chief Investment Officers, pension fund managers, and investment portfolio managers. The survey was carried out again in October–November 2016 with 700 respondents across 29 countries, and in September–October 2017 with 700 respondents across 31 countries.

Table 8.1 provides summary statistics across the three groups on age, gender, and net worth/assets. In the Appendix we also include a breakdown of respondents by country in the individual and advisor surveys, as well as the list of institution types covered in the institutional survey.

It is important to keep in mind that the surveys give us information about what investors *think they would do* under various market scenarios. Thus the results pertain

Year	Survey Group	# Subjects	Summary		
2015	Individual Investors	7000	**Age**	**Gender**	**Net Worth**
			Generation X: 32% Generation Y: 37% Baby Boomers: 28% Pre-Baby Boomers: 3%	Male: 59% Female: 41%	Mass Market: 21% Mass Affluent: 24% Emergning HNW: 25% High Net Worth: 30%
	Financial Advisors	2342	**Advisor Characteristics**		
			Average age: 46 years Male: 79%; Female: 21% Average personal book of business: $28.3 million		
	Institutional Investors	660	**Organization Assets**		
			Less than $2 billion: 18% $2 billion – $5 billion: 25% More than $5 billion: 57%		
2016	Individual Investors	7100	**Age**	**Gender**	**Net Worth**
			Generation X: 35% Generation Y: 33% Baby Boomers: 28% Pre-Baby Boomers: 4%	Male: 61% Female: 39%	Mass Market: 22% Mass Affluent: 29% Emergning HNW: 24% High Net Worth: 26%
	Financial Advisors	2550	**Advisor Characteristics**		
			Average age: 47 years Male: 84%; Female: 16% Average personal book of business: $29.5 million		
	Institutional Investors	700	**Organization Assets**		
			Less than $5 billion: 45% More than $5 billion: 55%		
2017	Individual Investors	8300	**Age**	**Gender**	**Net Worth**
			Generation X: 29% Generation Y: 34% Baby Boomers: 31% Pre-Baby Boomers: 5%	Male: 56% Female: 44%	Mass Market: 27% Mass Affluent: 26% Emergning HNW: 24% High Net Worth: 23%
	Institutional Investors	700	**Organization Assets**		
			Less than $5 billion: 44% More than $5 billion: 56%		

Table 8.1: Summary statistics for the three survey groups on age, gender, and net worth/assets. Different groups were presented with different demographic questions, and the summary statistics are not directly comparable across groups. Gender and age were not asked for institutional investors.
The definitions of generations of investors and their net worth classifications are in the Appendix. *HNW* stands for High Net Worth. *Organization Assets* means the size of assets for which the respondent's organization is responsible.

to investors' conditional expected changes to their portfolios; to obtain actual changes, we would need to use historical data on portfolio allocations or on investor trades. However, we obtain highly significant differences across various investor groups, so

that even in the presence of potential noise in self-reporting how investors anticipate their future decisions, we may still draw reliable conclusions from these comparisons.

8.2.2 Estimating Risk Aversion

We estimate the risk aversion of survey respondents using the popular technique introduced by [117], and later modified by [60]. The idea by Holt and Laury is to present subjects with pairs of gambles and ask which gamble they would choose within each pair. Dave et al. extend this idea by presenting a list of six different gambles and asking subjects to pick one. Table 8.2 lists the six gambles used in the Natixis surveys, with gamble 1 being the safest and gamble 6 being the riskiest. Note that these are the same gambles as in the Dave et al. study, except they are of greater magnitude, given that our subjects are professional investors or industry professionals, who tend to care about relatively large bets when investing.

The choice of one gamble among six gambles is converted to a choice of one gamble in each of five different pairs of gambles: gamble 1 vs. gamble 2, gamble 2 vs. gamble 3, ..., gamble 5 vs. gamble 6. The conversion is done as follows. If a person chose a particular gamble k among the six gambles, then he would choose gamble $i+1$ over gamble i as long as $i+1 \leq k$, and would choose gamble i over gamble $i+1$ as long as $i \geq k$. This is valid because gambles $1, 2, \ldots, 6$ are increasing in risk, in that order.

Gamble	**Outcome 1**		**Outcome 2**		**Mean**	**Standard Deviation**
	Probability	Payoff	Probability	Payoff		
Gamble 1	50%	$28,000	50%	$28,000	$28,000	$0
Gamble 2	50%	$36,000	50%	$24,000	$30,000	$6,000
Gamble 3	50%	$44,000	50%	$20,000	$32,000	$12,000
Gamble 4	50%	$52,000	50%	$16,000	$34,000	$18,000
Gamble 5	50%	$60,000	50%	$12,000	$36,000	$24,000
Gamble 6	50%	$70,000	50%	$2,000	$36,000	$34,000

Table 8.2: List of six gambles presented to survey participants. The subjects were asked to choose which one of the gambles they would prefer. Each gamble involves two outcomes, each of which has a 50% probability of occurring. The first gamble can be viewed as a "sure" outcome.

We assume that respondents have constant relative risk aversion (CRRA) utility. The risk aversion coefficient r will be used as a proxy for risk aversion among different groups throughout the chapter. When considering a particular gamble i, a subject first evaluates their expected utility:

$$E(U_i) = p_{i,1}U(x_{i,1}) + p_{i,2}U(x_{i,2}) \; ; \qquad U(x) = \frac{x^{1-r} - 1}{1 - r} \tag{8.1}$$

where $x_{i,1}$, $x_{i,2}$ are the two possible payoffs for the gamble, and $p_{i,1}$, $p_{i,2}$ are their associated probabilities of occurring. Then, for each pair of gambles i, j, after calculating

expected utility, the subject picks gamble i with probability:

$$\mathbb{P}\left(\text{Choose gamble } i\right) = \frac{[E(U_i)]^{1/\mu}}{[E(U_i)]^{1/\mu} + [E(U_j)]^{1/\mu}} \tag{8.2}$$

where μ is a noise parameter, since a subject may actually pick gamble j even if gamble i has higher expected utility. Values close to 0 signify little deviation from expected utility theory.

Using Equations 8.1 and 8.2, and survey responses, we calculate the likelihood function for parameters μ, r. We then estimate the parameters using the maximum likelihood method, and use the Hessian to estimate standard errors. A more detailed description of the estimation of the risk aversion coefficient is given in the appendix. The risk aversion estimates are compared across the three groups of subjects as well as across individual investor demographic categories.

8.3 Investors, Advisors, and Institutions

In this section, we present our comparison of allocation decisions and risk aversion across investors, advisors, and institutions. The results are striking. Investors appear to be mostly extrapolative in their changes in equity allocation, while advisors and institutions are predominantly contrarian. The proportion of "passive" respondents is much higher for individuals and financial advisors than for institutional investors. Finally, investors are more risk-averse than advisors, who are in turn more risk-averse than institutions.

8.3.1 Asset Allocation Decisions

All individual and institutional investors were asked two questions pertaining to investment decisions. The first asked how they would change their allocation to equities if the S&P 500 declined by 10–20% during the next six months, and other assets performed as expected. The second asked the same question in a scenario where the S&P increased by 10–20%. Financial advisors were asked two similar questions about how they would *advise their clients* to change equity allocation. There were five possible responses: "large decrease", "slight decrease", "do nothing", "slight increase", and "large increase".

Figures 8-1 and 8-2 plot the distributions of responses for the two questions. To generate these plots, we have aggregated the responses between 2015 and 2017. We present year by year results in Figures F-1 and F-2 of the appendix, along with plots displaying only the responses collected from investors, advisors and institutions in the United States. In the scenario of an S&P 500 fall, 44% of individuals would decrease equity allocation[5], in comparison to just 17% for advisors and 17% for institutions. At the same time, 67% of institutions and 52% of advisors would increase equity

[5]The proportion of respondents decreasing their equity allocation is calculated by counting all respondents who answered with "large decrease" or "slight decrease" to the survey question.

allocation, a much higher proportion than the corresponding 17% for individuals. For a rise in the S&P 500, the results are basically reversed. For individuals, 30% would decrease allocation and 32% would increase it, whereas the corresponding numbers are 50% and 13% for advisors, and 68% and 9% for institutions. Thus, on average, individual investors would change their allocation in the same direction as a recent S&P 500 move, while financial advisors and institutional investors would change allocation in the opposite direction.

These results do not seem to depend heavily on the country in which the respondent is based, as we get very similar numbers for respondents from the the United States and globally, as displayed in Figures F-1 and F-2 of the appendix. We get the following results in the case of the United States. For an S&P 500 fall, 31% of the individuals, 11% of the advisors and 9% of the institutions would decrease their equity allocation, while 26% of the individuals, 66% of the advisors and 79% of the institutions would increase their equity allocation. For a rise in the S&P 500, 23% of the individuals, 59% of the advisors and 82% of the institutions would decrease their equity allocation, while 34% of the individuals, 9% of the advisors and 6% of the institutions would increase their equity allocation. We observe that, on average, the individuals surveyed in the United States tend to be extrapolators while the financial advisors and institutional investors tend to be contrarians.

We obtain more insight into these differences by looking at the response distributions in more detail. One aspect that stands out is the "extreme contrarian" responses of institutional investors. 27% would significantly increase allocation following a fall in the S&P 500, whereas 22% would significantly decrease allocation following a rise in the S&P 500. This is much higher than the corresponding 3% and 17% for individuals, or the 13% and 11% for advisors. While advisors also give contrarian responses, their predominant response is to change allocation only *slightly*: 39% of advisors would favor a slight increase after a fall in the S&P 500, and 39% of advisors would favor a slight decrease after a rise in the S&P 500—much higher than the comparable rates among individual investors. Finally, we notice that while investors on average extrapolate, there is a significant fraction in both scenarios that prefers to significantly decrease equity allocation: 20% for an S&P 500 fall and 17% for an S&P 500 rise. We will later show, using clustering techniques on the individual investor dataset, that these numbers are mainly driven by the "risk avoider" class of investors (Figure F-4 of the appendix).

As a robustness check, we also compare the responses across individuals and advisors in 2015 and in 2016 for each country included in both the Individual Investor and the Financial Advisor Surveys of each year. We also include a comparison between the responses across individuals and institutional investors in 2016 and in 2017 for each country included in both the Individual Investor and the Financial Advisor Surveys of each year.[6] The results are shown in Figures 8-3 and 8-4. We again see that there is a much larger proportion of individual investors who would decrease their equity allocation after seeing an S&P 500 fall than there is of advisors; at the same

[6]We do not include institutional investors in the 2015 comparison because the sample sizes at the country level are small in the Institutional Investor Survey.

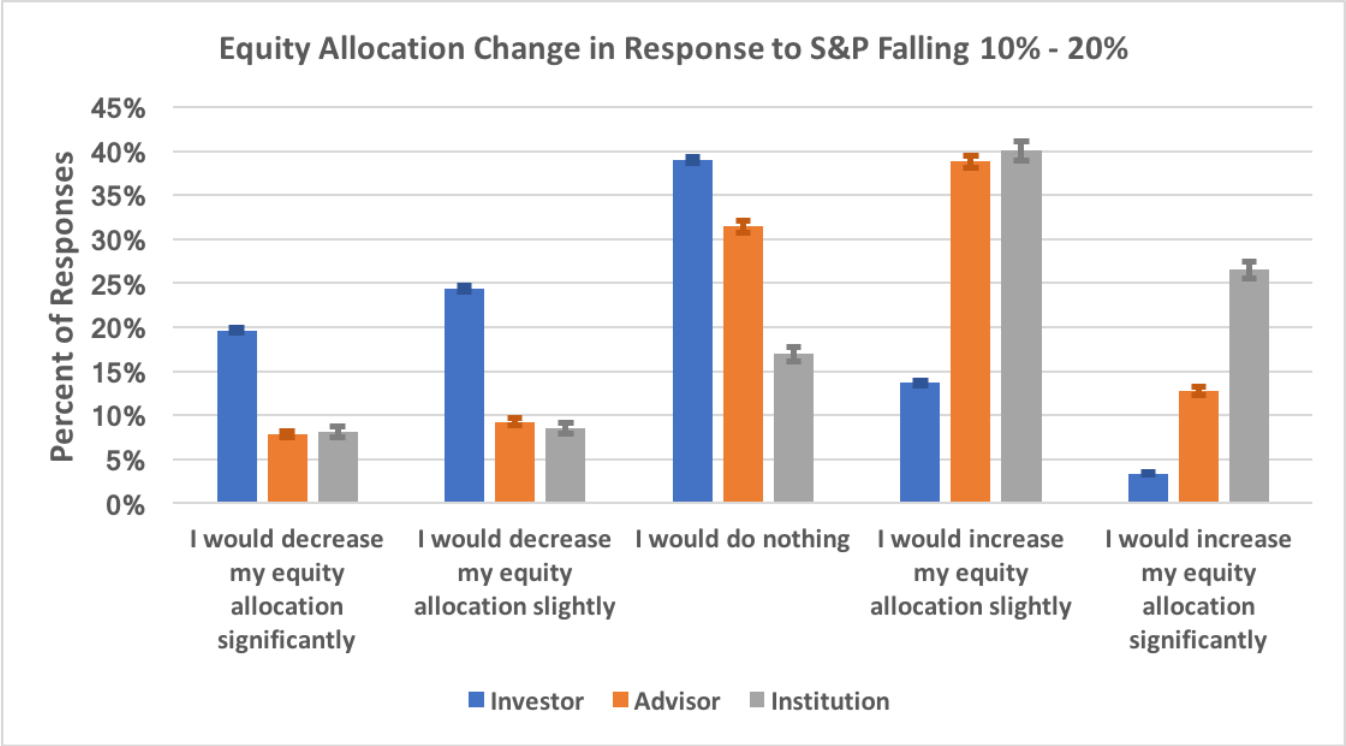


Figure 8-1: Reactions to a decrease in the S&P 500 across three groups between 2015 and 2017 (2015 and 2016 for advisors). For each group and each possible answer, we show error bars corresponding to one standard error calculated assuming each respondent chooses either that particular answer, or any other answer.

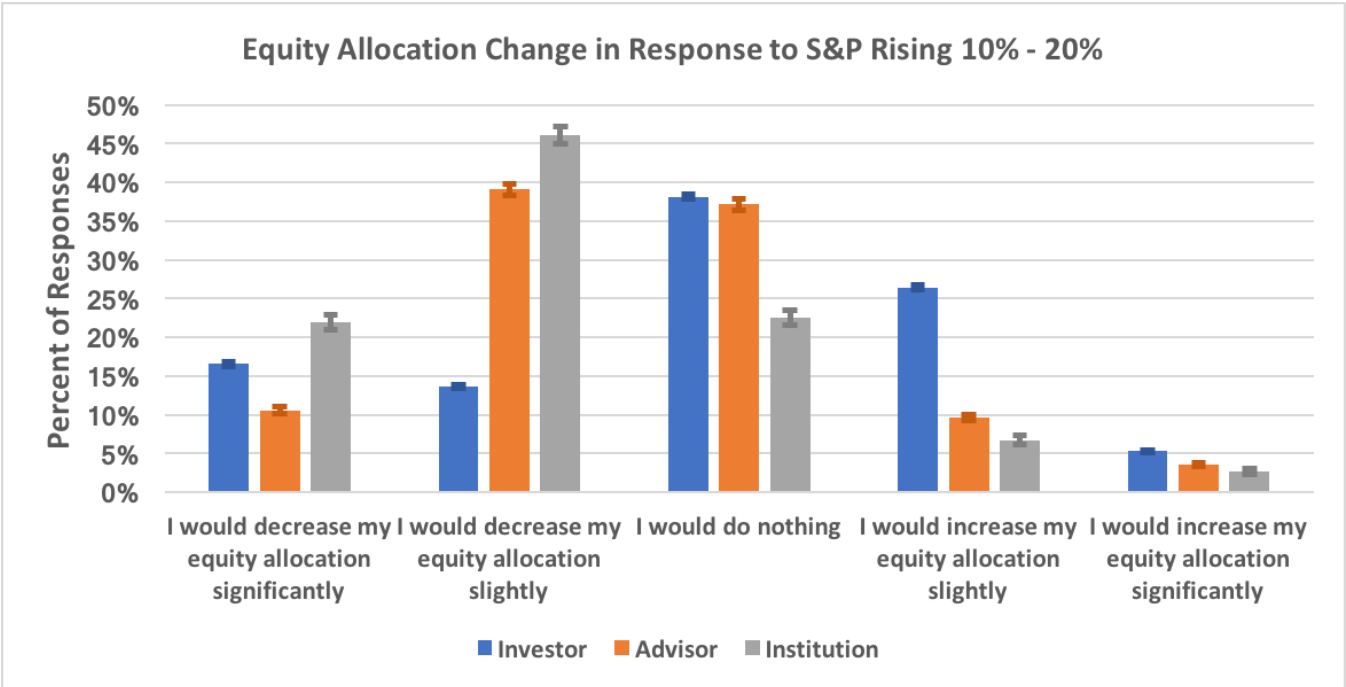


Figure 8-2: Reactions to an increase in the S&P 500 across three groups between 2015 and 2017 (2015 and 2016 for advisors).

Figure 8-3: Reactions to a decrease in the S&P 500 across individual investors and financial advisors, and across individual investors and institutional investors, split up by country. We only plot the results for countries that appear for both investor types each year. Error bars correspond to one standard error. Standard errors are large for some countries due to small sample size.

Figure 8-4: Reactions to an increase in the S&P 500 across individual investors and financial advisors, and across individual investors and institutional investors, split up by country. We only plot the results for countries that appear for both investor types each year. Error bars correspond to one standard error. Standard errors are large for some countries due to small sample size.

time, many more advisors would increase their allocation in this scenario compared to individuals. The differences are significant for each country at the 1% level, except for Canada in 2016 for a decrease in equity allocation.In the scenario of an S&P 500 rise, the results are reversed; this time, advisors are the ones who are more likely to decrease equity allocation. It is important to note that the differences between advisors and individual investors who decrease equity allocation are now smaller (although still significant at the 1% level for most countries). This again is caused by "risk avoider" investors who significantly decrease equity allocation following a rise in the S&P 500.

By comparing the responses from the Individual Investors and the Institutional Investors, we observe a similar behavioral difference. In 2016, for a decrease in the S&P 500, a much larger proportion of individual investors would decrease their equity allocation (with the exception of Australia) while a much larger proportion of institutional investors would increase their equity allocation (with the exception of Taiwan). These differences are significant for most countries at the 1% level. For an increase in the S&P 500, we observe the opposite reaction. While a larger portion of institutional investors decrease their equity allocation, a larger portion of individual investors actually increase their equity allocation (except for the Netherlands and Hong Kong). These differences are significant for most countries at the 1% level. Similarly, in 2017, for a decrease in the S&P 500, a much larger proportion of individual investors would decrease their equity allocation (with the exception of Taiwan) while a much larger proportion of institutional investors would increase their equity allocation (with the exception of Korea and Taiwan). These differences are significant for most countries at the 1% level. In the case of an increase in the S&P 500, we observe the opposite reaction. While a larger portion of institutional investors decrease their equity allocation (with the exception of Korea), a larger portion of individual investors actually increase their equity allocation (with the exception of Taiwan and Korea). These differences are significant for most countries at the 1% level.

The fifth type of response in these two scenarios is the "do nothing" response. Its levels for individual investors and financial advisors are similar in both scenarios, and relatively large: 39% and 31% in case of an S&P 500 fall, and 38% and 37% in case of an S&P 500 rise. In contrast, only 17% of institutional investors would do nothing following a decrease in the S&P 500; this number is 23% following an increase in the S&P 500. The more passive responses for individuals and advisors may be explained by the fact that many have long-term investment objectives that do not require significant changes to their asset mix. As a result, they would generally not be affected by large fluctuations in asset prices. In contrast, institutional investors often have shorter investment horizons, and their performance may be evaluated at shorter frequencies. Because of this, they may need to react to changes in the relative prices of different asset classes more often.

The observed differences in active response between groups have several potential explanations. The apparent tendency of a large number of individual investors toward extrapolative allocation may be an inherent aspect of their behavior. Our survey is insufficient to understand why this behavior comes about. However, we note that pervasive evidence for similar behavior has been documented in other studies; see

[101] for a comprehensive survey. Financial advisors may recognize that some of their individual investor clients excessively extrapolate, and instead advise them to apply a contrarian allocation strategy, as observed in the survey. Also, should the S&P 500 decline significantly, an advisor may view that circumstance as a good entry point for a client with a long investment horizon.

The contrarian behavior of financial advisors may also be explained by long-term investment objectives, which are typically planned to maintain a target asset mix over several years. For example, if equities move significantly relative to bonds over the short term, then client allocation will experience a large deviation from its target mix, and advisors may propose a contrarian reallocation to return it to the target.

The target-mix story may also explain the extreme contrarian response of institutions, especially if their performance is evaluated relative to a benchmark. However, because a large proportion of institutional respondents would *significantly* decrease or *significantly* increase allocation, there are probably more factors at play. It is possible that some institutional investors are employing value strategies or are engaged in distressed investing, which results in contrarian trading when asset class prices deviate from their earlier relationship. Another possibility may be that some investors (e.g. pension plans) have a target return in mind. If recent performance is very good, they may cut portfolio risk by moving out of equities in order to have a safer portfolio for the rest of the year, while likely still hitting their target.

Overall, because individual investors tend to have extrapolative reactions, while institutional investors usually have contrarian ones, we conclude that institutions generally take the other side of individual investor trades in broad asset allocations. Note that [101] propose firms may be also involved in accommodating individual investor demand through equity issuance.

8.3.2 Risk Aversion

We next compare risk aversion across the three survey groups. Figure 8-5 shows the distribution of preferences for the six gambles across respondents between 2015 and 2017. Recall that gamble 1 is the safest, and gamble 6 is the riskiest. We see that 40% of the individuals choose gamble 1, which is much higher than 29% for advisors, which in turn is higher than 12% for institutions; the differences are significant at the 1% level. At the other end of the spectrum, a significantly higher proportion of institutions choose gambles 5 and 6 in comparison to individuals and advisors. These observations strongly suggest that individual investors are the most risk-averse of the three groups, while institutional investors are the least risk-averse. Note that we are able to make this conclusion from the distribution of responses alone, without making assumptions about the utility functions of respondents.

To investigate further, we add the assumption that all subjects have CRRA preferences, and carry out the estimation procedure for the risk aversion coefficient as discussed in the methodology. The results are shown in Figure 8-6. In 2015, individual investors have the highest risk aversion coefficient at 1.14, followed by financial advisors at 0.85 and institutional investors at 0.39. In 2016, the numbers are 1.07, 0.78, and 0.34, respectively, for individual investors, financial advisors and institu-

tional investors. In 2017, the numbers are 1.07, and 0.33, respectively, for individual investors and institutional investors. If we consider the three years dataset as a whole, we get a risk aversion coefficient at 1.07 for individual investors, followed by financial advisors at 0.78 and institutional investors at 0.36. The respective noise parameters μ are 0.08, 0.07 and 0.04. If we consider only the respondents from the United States, we get a risk aversion coefficient at 1.08 for individual investors, followed by financial advisors at 0.73 and institutional investors at 0.29, and noise parameters μ at 0.07, 0.06 and 0.03 respectively. The pairwise differences in coefficient estimates are very large, and are again significant at the 1% level each year, and when considering the three years as a whole. Our results are consistent with the general intuition that individuals are generally the most risk-averse group of market participants, while institutional investors are the least risk-averse one. Likewise, it is plausible for financial advisors to fall in the middle of the spectrum, given the fact that, while they do work in the investment management industry, they generally do not directly manage money for their clients and so do not take outright bets in the markets, whereas institutions do.

We have presented evidence that individual investors are, on average, extrapolators, while financial advisors and institutional investors are contrarians. Contrarian behavior is particularly strong for institutions. Furthermore, institutions would usually reallocate portfolios more actively in response to large S&P 500 moves, while a significant proportion of individuals and advisors would do nothing in those situations. Individuals are by far more risk-averse than financial advisors, and advisors are much more risk-averse than institutional investors.

8.4 Individual Investor Decisions

We have seen that a large proportion of individual investors tend to extrapolate market performance when making their equity allocation decisions. However, we have also observed that a few of them tend to consistently decrease equity allocation. This suggests that we cannot label all investors as extrapolators, and it makes sense to study the individual dataset in more detail. We start by looking at the dependence of investor risk aversion and preference for active investing on the demographic factors of age, gender, and net worth. The strongest results are observed for age, older investors tending to be more risk-averse and also more passive. We then run a clustering algorithm to partition investors into five groups: passive investors, extrapolators, risk avoiders, optimistic investors and contrarians. We also compare the demographic breakdowns of the different investor types.

8.4.1 Asset Allocation

We used three questions from the survey to elicit the degree of investor passivity. The first two questions of the 2015-2017 surveys asked about their allocation under different S&P 500 moves, as discussed in the last section. In 2015, the third question asked when during the Financial Crisis of 2007-2009 investors decreased their equity

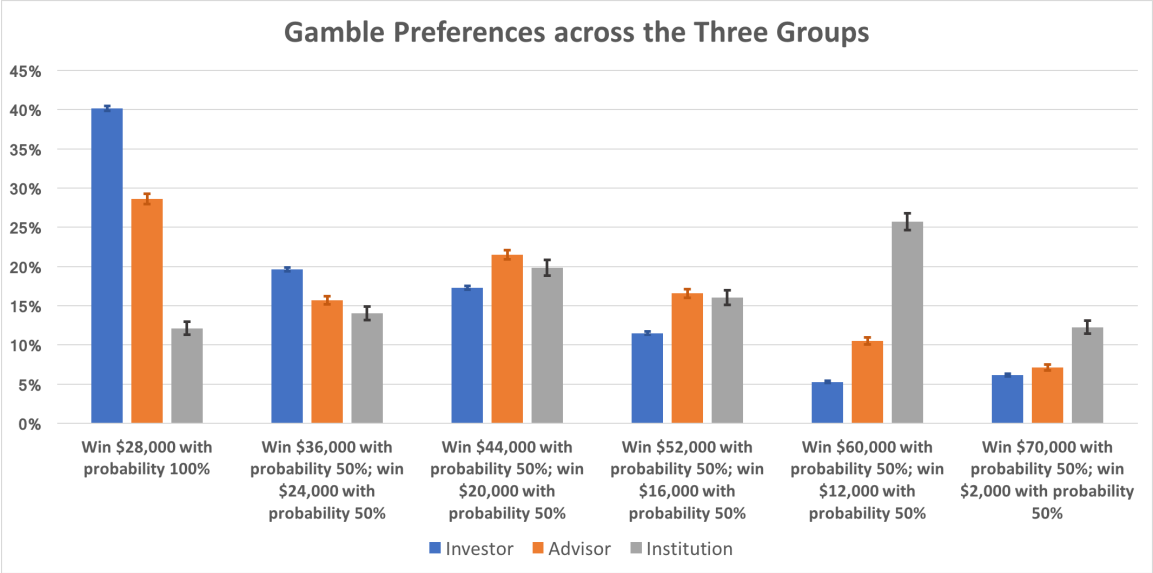


Figure 8-5: Distributions of gamble preferences across three groups between 2015 and 2017 (2016 for advisors). For each group and each possible answer, we show error bars corresponding to one standard error calculated assuming each respondent chooses either that particular gamble, or any other gamble.

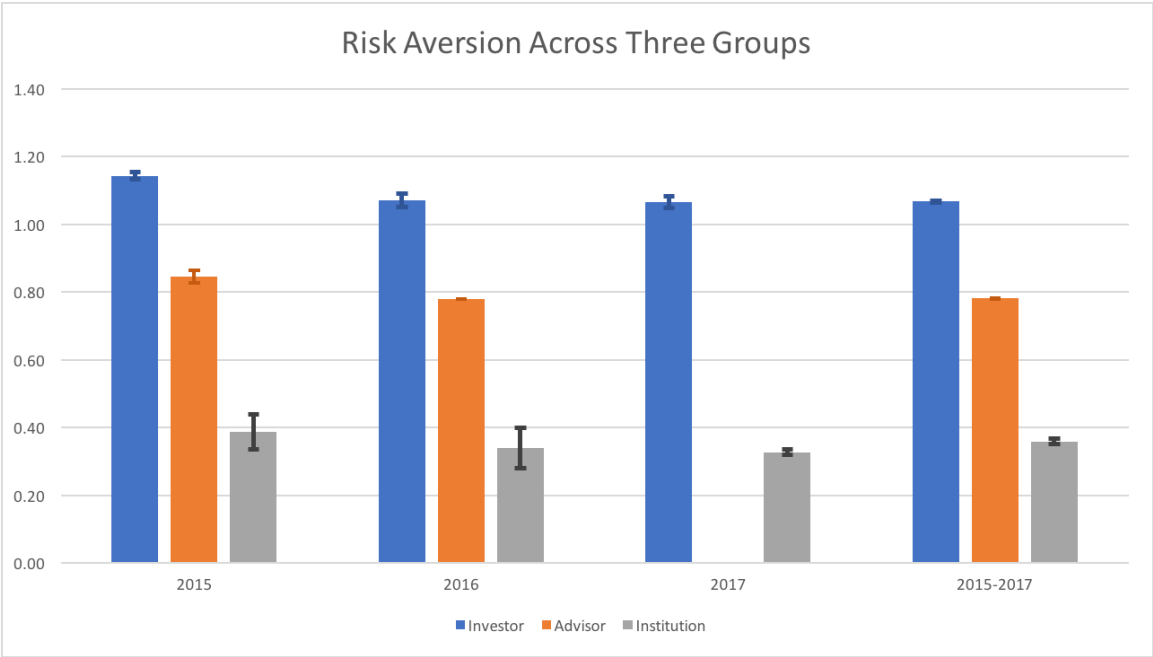


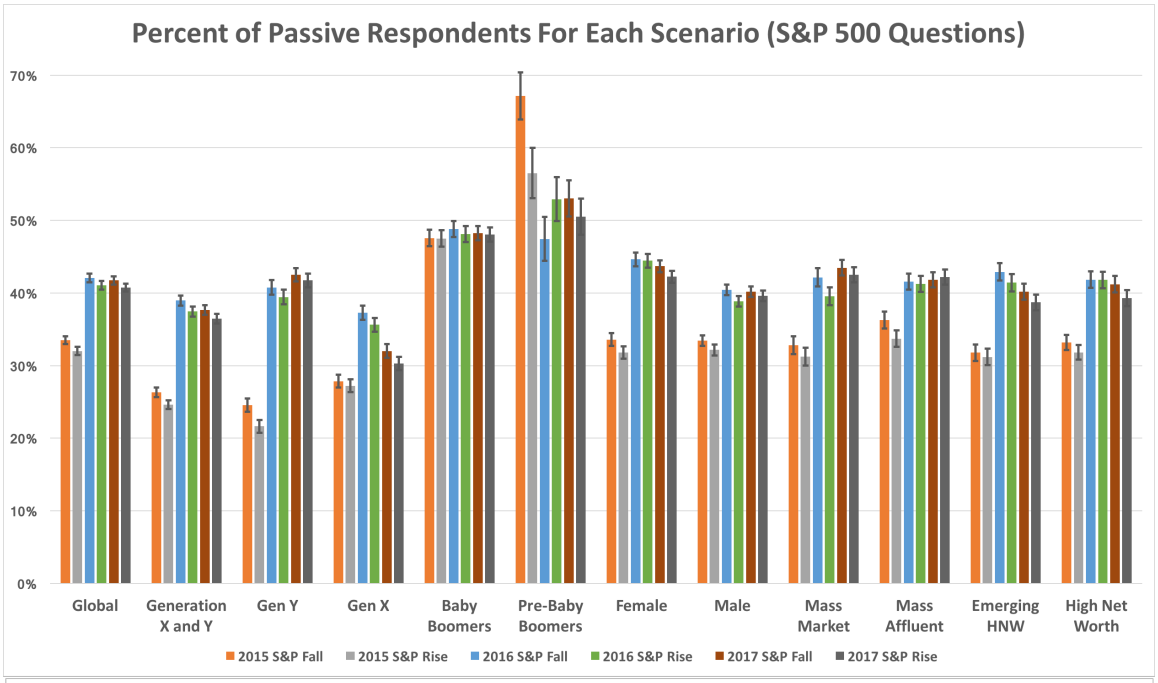
Figure 8-6: Estimated Risk Aversion coefficients across the three groups between 2015 and 2017, and averaged over 2015-2017. Financial advisors have not been surveyed in 2017. Error bars correspond to one standard error.

allocation. Possible answers ranged from “the second half of 2007" to “the second half of 2009". We also included an “I did nothing" response for investors who did not significantly decrease their allocation. In 2016, individual investors were asked how they have modified their equity allocation in response to the market volatility in January 2016. In 2017, they were asked how they modified their equity allocation in response to the U.S. presidential on November 8th, 2016. Possible answers for both questions ranged from “I decreased significantly my stock or shares allocation” to “I increased significantly my stock or shares allocation”, and included a “I did nothing” response. The first chart in Figure 8-7 compares the proportions of investors choosing the passive response for the first two questions, while the second chart in Figure 8-7 compares the proportions of investors choosing the passive response for the third question. We consider these responses across the different demographic categories. The classification we use for age and net worth is detailed in Table F.7 of the Appendix.

Within each demographic category, the percent of respondents who choose to do nothing for the S&P scenarios is similar in 2016 and 2017, but sometimes differs slightly in 2015. We can also note that, in most cases, the percentage responding passively to the Financial Crisis, the January Volatility and the Presidential Election are slightly higher than the same responses to the hypothetical moves in the S&P 500. By looking at pairwise differences across age categories, for each year, we observe that older investors are more likely to be passive. However, we observe in 2016 and 2017 that Generation X investors (34 to 49 years old) are slightly less passive than Generation Y investors (18 to 33 years old). In most of the other cases, younger investors are more passive than older investors at a 1% confidence level. Comparing the different ends of the spectrum in the lower chart of Figure 8-7, between 2015 and 2017, we observe that 79%, 62% and 79% of pre-Baby Boomers responded that they have behaved in a passive way, while this number is just 29% for Generation Y investors in 2015, and 46% and 56% for Generation X investors between 2016 and 2017. The greater observed percentage of older investors being more passive may be a result of their inherent behavior; this is consistent with other empirical evidence, for example, by [68] and [104]. Another potential explanation is that younger investors generally tend to have a higher allocation of equities in their portfolios, and therefore would react more to a large change in equity prices. The differences across gender and net worth categories are not large, and almost in all cases not significant at the 5% level (except for the 2016 results, for which the pairwise differences across gender are actually significant at the 1% level).

8.4.2 Risk Aversion

Figure 8-8 compares the estimated risk aversion coefficient across demographic categories. First, we observe that risk aversion increases with age, as documented earlier in [178]. In 2015, the pairwise differences across generations are significant at the 1% level (except for pre-Baby Boomers in 2015; standard errors for this cohort are large, in part, due to the small sample of 207 respondents). In 2016 and 2017, except for pre-Baby Boomers, these differences are significant at the 5% level (except



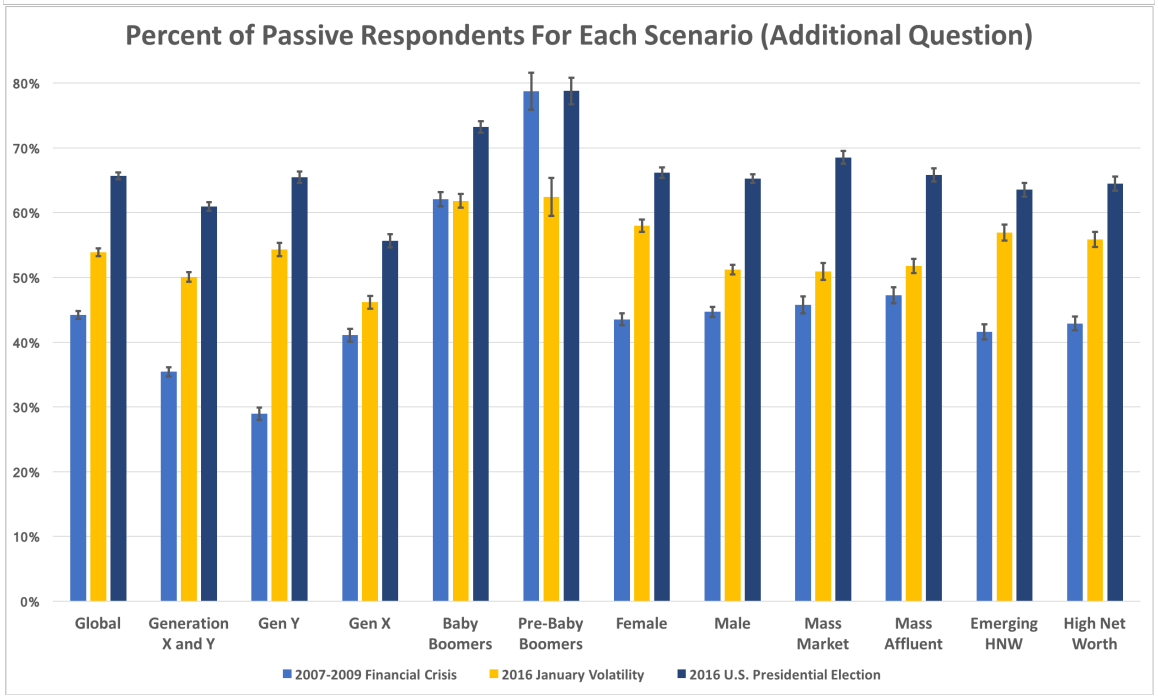


Figure 8-7: Percent of passive respondents for each scenario across different individual investor demographic categories between 2015 and 2017. The definitions of investor demographic categories are in Table F.7 of the Appendix. *HNW* stands for High Net Worth. Error bars correspond to one standard error, assuming the respondent chose either the "passive" response or any other response.

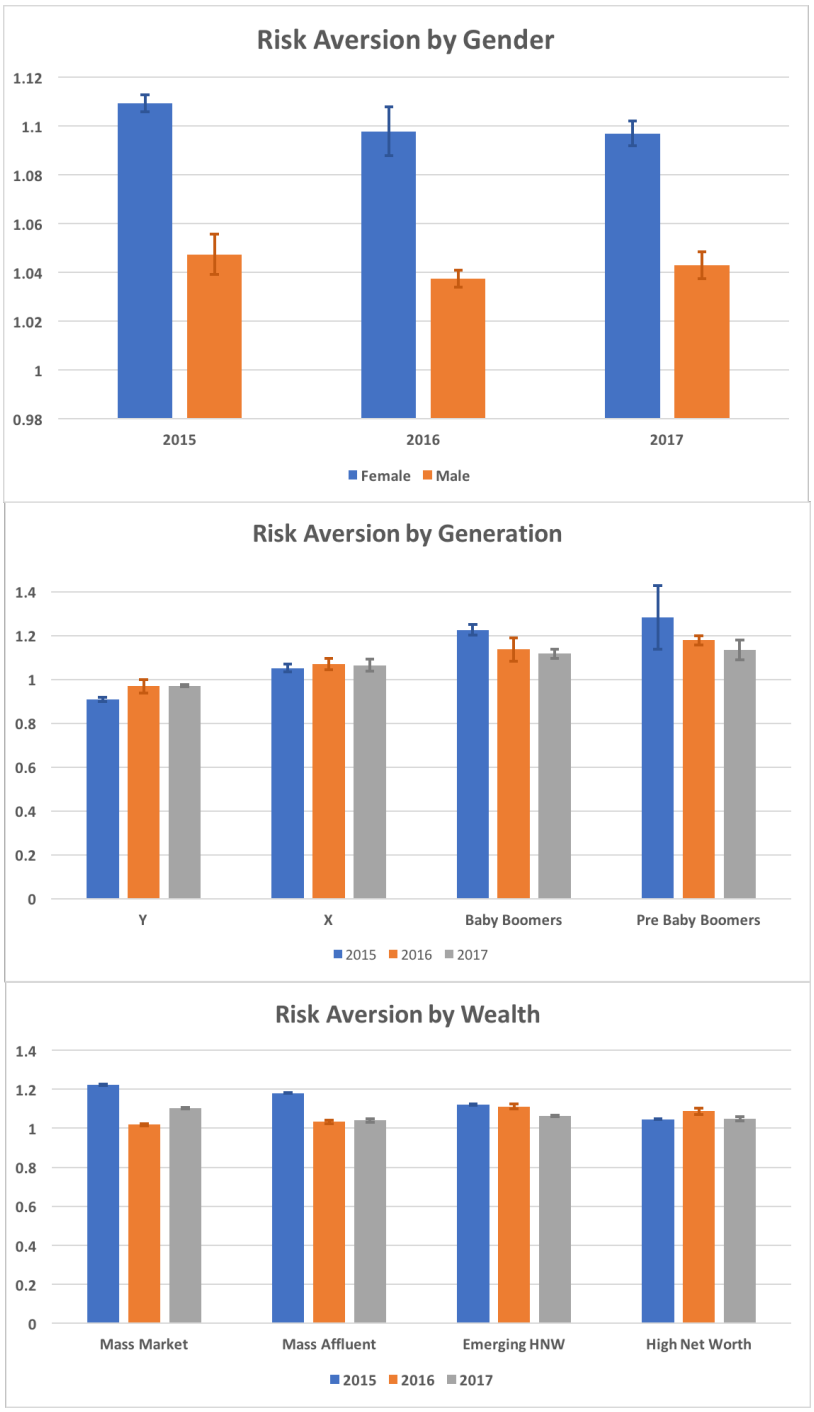


Figure 8-8: Estimated Risk Aversion coefficients across individual investor demographic categories between 2015 and 2017. Error bars correspond to one standard error.

when comparing Generation X with Baby Boomers). Furthermore, women appear to be slightly more risk-averse than men; these differences are significant at the 1% level. Studies such as [53] suggest that risk aversion decreases with net worth. In our surveys, we can only confirm this claim at the 5% level in 2015 (except for the High

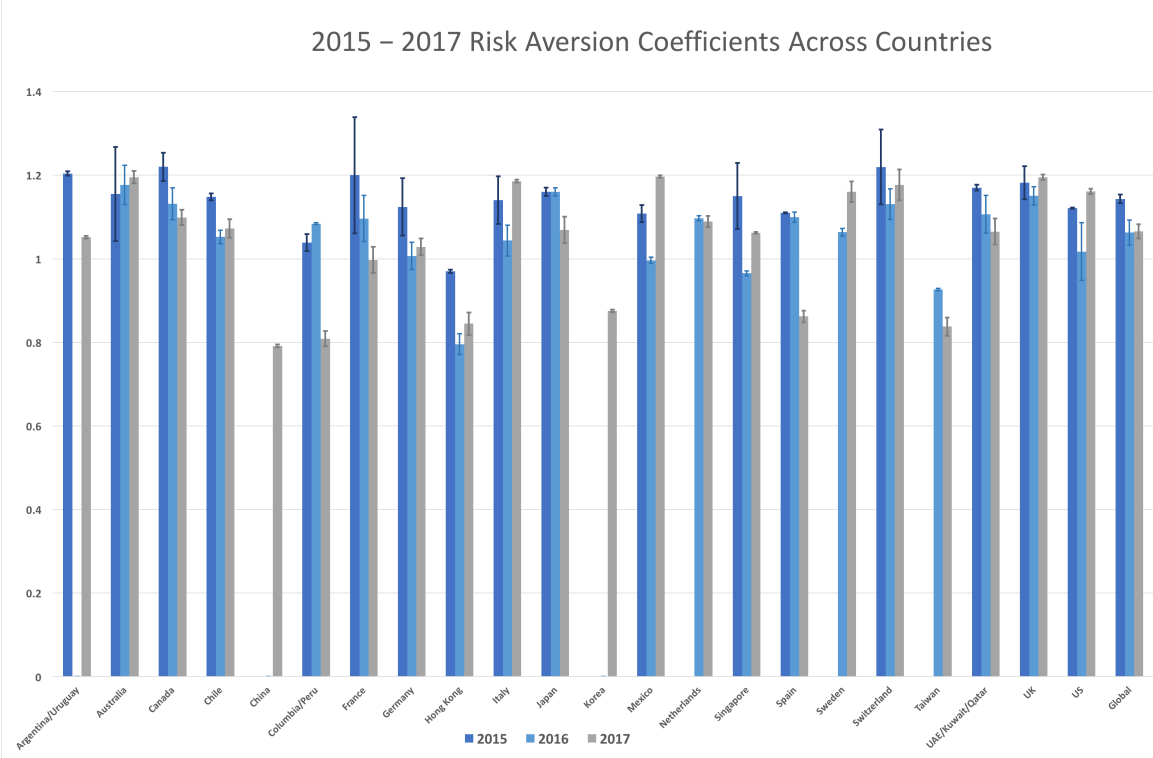


Figure 8-9: Estimated risk aversion coefficients for individual investors across countries between 2015 and 2017. As some countries have only been surveyed one or twice, the missing data has been left blank on the chart. Error bars correspond to one standard error.

Net Worth category). In 2016 and 2017, the pairwise differences across net worth categories are not significant at the 5% level.

We also include a comparison of individual risk aversion across countries in Figure 8-9; Figure F-3 of the Appendix presents a year by year comparison of individual risk aversion across countries. Standard errors are large for a few countries due to the small sample size, and large differences between countries are generally not seen. However, in 2017, and to some extent in 2016, pairwise differences are significant at the 5% level and even at the 1% level. The first outlier is Hong Kong in 2015 and 2016, where investors are significantly less risk-averse than in almost every other country considered (the differences are significant at the 5% level in 2015, and at the 1% level in 2016). In 2017, China is significantly less risk-averse than all other countries (except for Columbia and Peru), at the 5% level.

8.4.3 Cluster Analysis

We now perform clustering on the individual dataset of 22,400 investors pooled across all three years of the study. We use the responses to the two S&P 500 asset allocation questions described earlier. The verbal responses are transformed into numbers as shown in Table 8.3; the numbers range from -2 to 2 for the two questions on reactions to S&P 500 moves. We then run a k-means clustering algorithm assuming six clusters.

We choose six clusters since this partition gives the most distinct and cleanest clusters in comparison to using two to five clusters. For robustness we ran the algorithm for different values of the random seed, and performed hierarchical clustering on the dataset; the results were generally consistent in all cases. [7]

Question	Response Coding
Reaction to S&P 500 Fall	Change in Equity Allocation: −2 - Significant Decrease −1 - Slight Decrease 0 - I Did Nothing 1 - Slight Increase 2 - Significant Increase
Reaction to S&P 500 Rise	Change in Equity Allocation: Same as for Reaction to S&P 500 Fall
Reaction to Financial Crisis	Timing of Decrease in Equity Allocation: 1 - Second Half of 2007 2 - First Half of 2008 3 - Second Half of 2008 4 - First Half of 2009 5 - Second Half of 2009 6 - No Significant Decrease
Reaction to the January 2016 Volatility	Change in Equity Allocation: Same as for Reaction to S&P 500 Fall
Reaction to the 2017 Presidential Election	Change in Equity Allocation: Same as for Reaction to S&P 500 Fall

Table 8.3: Coding of investor responses for the clustering algorithm. The exact formulation of the questions and possible responses is in Sections F.1 and F.2 of the appendix.

Table 8.4 shows the results of the clustering procedure. Clusters 1 and 6 are composed of *extrapolators* who tend to change equity allocation in the same direction as a recent S&P 500 move. In cluster 1, extrapolators react more to a rise in the S&P 500, with average responses of −0.7 for an S&P 500 fall and 1.1 for an S&P 500 rise. In cluster 6, extrapolators react more to a fall in the S&P 500, with average responses of −2.0 for an S&P 500 fall and 0.8 for an S&P 500 rise. Cluster 2 consists predominantly of *passive investors*, since the average responses are "close" to the response corresponding to not changing equity allocation in each of the two S&P 500 scenarios: −0.1 for an S&P 500 fall and −0.1 for an S&P 500 rise. Cluster 3 contains mostly *optimistic investors*. They increase their allocation in both S&P 500 scenarios: average responses are 1.2 for an S&P 500 fall and 1.0 for an S&P

[7] We have also performed separate clustering analysis on each year from 2015 to 2017 and have included in the analysis the additional asset allocation question from each year. The verbal responses were also transformed into numbers as shown in Table 8.3; the numbers range from 1 to 6 for the reaction to the Financial Crisis (2015 survey), from −2 to 2 for the reaction to the January Volatility (2016 survey), and from −2 to 2 for the reaction to the U.S. Presidential Election (2017 survey). These results are presented in Section F.6 of the appendix.

500 rise. Cluster 4 is composed of *contrarians* who tend to change equity allocation in the opposite direction as a recent S&P 500 move: average responses are 0.7 for an S&P 500 fall and −1.7 for an S&P 500 rise. Finally, cluster 5 contains mostly *risk avoiders* who significantly cut allocation following large moves in the S&P 500: average responses are −1.5 for an S&P 500 fall and −1.6 for an S&P 500 rise.

To further validate our clustering approach, we look at the distributions of responses within each cluster, shown in Figure F-4 of the appendix. It is evident that the distributions for the risk avoider and both extrapolator clusters are tightly clustered around the corresponding means for the reactions to S&P 500 moves. The same is true for the passive, contrarian and optimistic clusters. We conclude that the clustering into six groups is indeed successful, and the descriptions of the clusters are appropriate.

It is important to note the relative sizes of the clusters from Table 8.4. Passive investors are the largest cluster and make up 35% of the whole sample. Extrapolators consist of 27% of all investors, and risk avoiders are at 19%. The rest is composed of optimistic investors (19%) and contrarians (8%). After running the clustering algorithm several times, we have obtained consistent results. Because passive investors and risk avoiders have symmetric responses to the two different questions about S&P 500 moves, the group of individual investors, taken as a whole, appears to exhibit the extrapolative behavior that we discussed when comparing this group to financial advisors and institutions.

To further motivate the significance of our clustering analysis, we have simulated completely random reactions (i.e., we took reactions to S&P 500 movements to be uniformly distributed over the set $\{-2, -1, 0, 1, 2\}$), and have assigned labels to the clusters obtained. If a cluster's mean reaction to an S&P 500 decline or increase was smaller than 0.5, we assigned the "Passive Investor" label to the cluster. The four other quadrants of the 2-dimensional response space were equally divided between Extrapolators, Risk-Avoiders, Optimistic and Contrarian Investors. The clusters obtained artificially (by running 10,000 simulations out of 50,000 responses) are drastically different from the clusters obtained from the survey data. In the former, we obtain, on average, a small fraction of Passive Investors, and the rest of the population is equally divided between Extrapolators, Contrarian Investors, Risk-Avoiders and Optimistic Investors. In the latter, we observe a dominance of Passive Investors, as well as a large number of Extrapolators and Risk-Avoiders. The proportion of Optimistic and Contrarian Investors is smaller.

In Table 8.4 we look at demographic patterns across different clusters. We assume that the clusters are independent collections of investors and compare their demographic averages; tests for statistical significance are given in Table F.11 of the appendix. There are no significant differences in gender. Passive investors are the oldest and risk avoiders as well as Optimistic investors are the youngest among the six clusters (statistically significant at the 0.1% level, except for the pairwise comparison between risk-avoiders and optimistic investors). Unsurprisingly, passive investors are also the ones that are most likely to be retired (significant at the 0.1% level). This is consistent with the results from Figure 8-7, which compared the degree of activity between generations. Here we have a more detailed breakdown, showing that the more

Clustering Analysis of Individual Investors (2015-2017)						
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
% Respondents	14%	35%	11%	8%	19%	13%
	Allocation Decisions					
Reaction to S&P 500 Fall	−0.7	−0.1	1.2	0.7	−1.5	−2.0
Reaction to S&P 500 Rise	1.1	−0.1	1.0	−1.7	−1.6	0.8
Behavior	Extrapolators+	Passive	Optimistic	Contrarian	Risk Avoiders	Extrapolators−
	Demographics					
Gender (% of Female)	42%	44%	40%	34%***	43%	39%
Age	1.16	1.33***	1.00	1.09	0.97	1.10
Net Worth	1.53	1.51	1.64*	1.54	1.51	1.56
Advised	65%	54%***	71%*	63%	62%	67%
Retired	8%	12%***	5%	6%	5%	6%
From U.S.	8%	12%	18%***	8%	8%	7%
Gamble Preference	2.41	2.12***	2.80***	2.58	2.54	2.53
Risk Aversion Coefficient	1.07	1.20***	0.82***	0.96	0.95	1.00

Table 8.4: Clustering of allocation decision responses from the 2015-2017 Individual Investor Surveys. For each cluster, we present the percent of respondents and the mean response based on the response coding in Table 8.3.
We also list the mean values of demographic categories across clusters. For *Age*, *Generation Y* = 0, *Generation X* = 1, *Baby Boomers* = 2, *Pre-Baby Boomers* = 3. For *Net Worth*, *Mass Market* = 0, *Mass Affluent* = 1, *Emerging HNW* = 2, *High Net Worth* = 3. The definitions of demographic categories are in Table F.7 in the Appendix. *Advised* is an indicator for if an investor uses a financial advisor. *Retired* is an indicator for if an investor is retired. *From U.S.* is an indicator for if an investor is from the United States. *Gamble Preference* corresponds to the one of six gambles from Table 8.2 chosen by the investor; Responses range from 1 to 6. *Risk Aversion Coefficient* is the estimated risk aversion coefficient based on the responses in each cluster.
For each category, we color in green the cell corresponding to the cluster with the highest mean value. We test for how significant the difference is between the highest mean and second-highest mean across the clusters; the result of the test is reported in terms of number of stars in the cell. * means significance at the 5% level, *** means significance at the 0.1% level; no stars means no significance at the 5% level. We color in red the cell corresponding to the cluster with the lowest mean value, and perform the same test comparing the lowest mean and second-lowest mean across the clusters.

active group, made up of younger investors, tends to exhibit risk avoidance behavior following large S&P 500 changes.

There are no significant differences in net worth across the clusters—except that the “optimistic" group has the highest average net worth (significant at the 5% level). An interesting distinction arises when we look at the proportion of investors in different clusters who use a financial advisor. We also observe that passive investors are much less likely to use an advisor (significant at the 1% level), while optimistic investors are much more likely to use an advisor (significant at the 5% level). Extrapolators are also more likely to use an advisor compared to passive investors,

contrarians and risk avoiders (significant at the 1% level). This makes intuitive sense, given that a financial advisor would likely try to discourage a risk avoider from significantly decreasing equity allocation in response to all large changes, as well as likely encouraging some fraction of passive investors into more active allocation. Interestingly enough, the optimistic and extrapolator clusters are the most likely to use a financial advisor—even though, as mentioned earlier, advisors usually provide contrarian advice to these individuals. We do not have sufficient evidence to explain this finding. However, it may be that advisors often deal with clients who extrapolate, and consequently advise an "opposite" allocation strategy to mitigate the bias of these investors.

Another dimension on which we compare investors is the degree of satisfaction with their 2014 and 2015 investment returns, as shown in Tables F.8 and F.9 of the appendix. Not surprisingly, in the 2015 Survey, risk avoiders were much less satisfied in comparison to optimistic investors (significant at the 0.1% level). This is likely explained by their risk avoidance: they saw large positive returns on their investments, and thus decreased their equity allocation—and missed out on very good market returns over the full year, the S&P 500 returning 13.5% in 2014. We note that financial advisors would be particularly beneficial to risk avoiders, since they would encourage investors to stay invested in the market despite large swings, and in this way earn higher returns over longer time horizons. In the 2016 Survey, Risk Avoiders have been more satisfied with their 2015 returns. The S&P 500 returned 1% in 2015, which could explain the satisfaction of investors with conservative strategies.

Finally, we compare risk aversion across the clusters. Passive investors are the most risk-averse cluster and optimistic investors are the most risk-seeking cluster (both are significant at the 0.1% level when comparing directly the gamble preferences as well as risk aversion coefficients). This is most likely due to its composition of predominantly older investors, who are more risk-averse than younger cohorts, as seen earlier. Ironically, risk avoiders do not appear to be one of the most risk-averse cluster: the differences are not significant compared to the contrarians group, but are significant at the 0.1% level compared to the optimistic group. We observe that the risk-avoiders group has the lowest average age, and predominantly consists of millennials. Although these younger investors may not recall the golden returns of the 1990s, they may still be psychologically affected by the 2008 financial crisis. This can cause large movements in the S&P 500 to trigger fear, and incentivize these investors to decrease their equity allocation. However, the risk aversion coefficient is calculated based on a win-win scenario by presenting six risky winning gambles to the participants. This suggests that our group of "risk avoiders" are actually more sensitive to loss-aversion than the other clusters, while being ready to take riskier win-win gambles than passive investors or extrapolators. On the other hand, passive investors, consisting of the eldest investors, do not react to any movement in the S&P 500 and are not interested in the riskier winning gambles. The latter is responsible for their higher risk-aversion coefficient. The idea that an individual's experience can affect his investment preferences has been investigated in [163]. It is also interesting to note that the least risk-averse cluster corresponds to the group of "optimistic" investors. These investors are willing and are able to take riskier asset allocation

decisions and riskier gambles. It is not a surprise to observe that this cluster has the highest net worth average. Optimistic investors also correspond to the group with the highest proportion of respondents from the United States. This could suggest that investors from the Unites States have a different and more optimistic perception of the S&P 500 than investors from the rest of the world.

We briefly comment on clustering results for the financial advisor and institutional investor responses. Because the sample sizes for these groups are much smaller than for the individual investor sample, we cannot perform the in-depth analysis that we do for individuals. However, we can still look at the largest cluster in each group. In particular, we use the responses from the two questions about hypothetical S&P 500 moves to form three clusters (data exploration found three–or four in the case of 2016 Financial Advisors–to be the most useful number). Tables F.15 to F.19 in the Appendix outline the relative sizes of the clusters and the average responses within each cluster. We see that the largest cluster in each sample clearly corresponds to a contrarian reaction; it makes up 47% of the sample for advisors and 68% for institutions in 2015, as well as 34% and 61% respectively for advisors and institutions in 2016. In 2017, 71% of the institutional investors appear to be contrarian. This further confirms our findings from the previous section that advisors and institutions are predominantly contrarian.

8.5 Conclusion

Using a comprehensive global survey over three consecutive years, we have identified differences in the investment behavior and risk tolerance preferences of individual investors, financial advisors, and institutional investors. Advisors and institutions exhibit contrarian strategies in their behavior whereby they tend to change equity allocation in the opposite direction of recent returns on the S&P 500. This reaction is particularly strong for institutional respondents, 67% of whom would increase equity allocation following a fall in the S&P 500, and 68% would decrease allocation following a rise. This behavior is not as pronounced among financial advisors, because a large proportion of advisors tend to act passively and not change allocation at all.

By asking for preferences among six hypothetical gambles we are able to estimate the risk aversion coefficient for each of the three groups. Consistent with general intuition, individual investors are by far the most risk-averse, whereas institutional investors are the least risk-averse. We also compare risk aversion across different individual demographic categories, and find that risk aversion increases with age.

We observe significant heterogeneity among individual investors in terms of their allocation decisions. Using a clustering algorithm we classify investors into five distinct types. The first corresponds to *passive investors* and makes up about 35% of the sample. The next three types are *extrapolators* (27% of the sample), *risk avoiders* (19% of the sample), and *optimistic investors* (11% of the sample). Finally we also observe a few *contrarians*, which constitute about 8% of all respondents.

Risk avoiders tend to significantly decrease equity allocation following large changes in the S&P 500 (both positive and negative), while extrapolators shift allocation in

the same direction as those changes. Passive investors are, on average, older and more risk-averse than the other types, while risk avoiders tend to consist of younger investors.

Our results using this novel survey dataset have important implications for future research. First, this data gives us new insight into the allocation decisions of market participants over medium-term time horizons. While there have been been extensive studies of short-term trading by individual and institutional investors (e.g. by [103] and [102]), few have looked at the broader decisions of asset allocation. The fact that institutions and advisors are largely contrarian in their allocation strategies, while investors are on average slightly extrapolative, may be important in understanding the process of strategic asset allocation and the trading between these different market participants on the asset class level.

Our other important insight comes from our breakdown of individual investors into clusters of different behavioral types. Our "risk avoider" type has seldom been documented in the literature, and it may prove to be a useful component of future economic models. Recent papers already incorporate extrapolators (e.g. [14]) into their models, as well as fundamental investors, which correspond to our institutional investors. It would be interesting to see the equilibrium dynamics of all five different individual investor types play out, as well as the dynamics of potentially contrarian institutions taking the other side of trades made by extrapolators.

Finally, it is of significant benefit to study the behavior of financial advisors, and in particular, why so many of them advocate contrarian strategies to their clients. It also appears that individuals who use a financial advisor are more likely to exhibit extrapolative behavior. Do they ignore financial advice, or are advisors trying to "balance out" their extrapolation with contrarian suggestions? Further surveys and studies of historical financial advice and associated client action are needed to answer these questions.

Part IV

Concluding Remarks

Chapter 9

Concluding Remarks

In this thesis, we approached the topic of risk management from three different perspectives: first in healthcare finance—during the drug development process and the drug deployment process, then in empirical asset pricing, and finally in investment management.

Although healthcare finance, empirical asset pricing, and investment management may appear as very different fields at first sight, they all rely at their core on the concepts of benefit, risk, and uncertainty. If BDA explicitly trades off benefit, risk, and uncertainty in its formulation, empirical asset pricing aims to understand and quantify the nature of the different risk factors stocks are exposed to, and investment management is inevitably tied to an investor's risk and uncertainty profile and behavioral biases.

9.1 Summary of contributions

In Chapters 2, 3, and 4, we focused on the regulatory and decision-making process by trading off type I and type II error rates to design optimal clinical trials that incorporate the patient's risk and uncertainty preferences. Chapter 2 presented the classical Bayesian framework with an application to safety endpoints of kidney replacement therapy devices and Chapter 3 presented an application to effectiveness endpoints of Parkinson's disease devices. In the former, we found that, except for patients on peritoneal dialysis, most patient groups presented a low risk tolerance with respect to the kidney replacement therapy device for the risk of serious bleeding and the risk of serious infection, as well as a low willingness to wait. This is reflected in the conservative 1.2% type I error rate threshold found for the risk of serious bleeding and a threshold below 0.1% for the risk of serious infection, which are both below the 2.5% threshold conventionally used. In the case of devices for Parkinson's disease, we found that a 5% threshold may not be conservative enough for patients with no prior experience with deep brain stimulation treatments but too conservative for patients who have previously received deep brain stimulation treatments.

Chapter 4 extended the Bayesian formulation by allowing the treatment effect to have a continuous prior distribution and by incorporating the patient's uncertainty

aversion into a mean-variance objective. We used the Parkinson's device analysis from Chapter 3 as a case study to understand the differences between the traditional BDA and extended BDA formulations. As expected, we found that the extended BDA setting produces optimal significance level thresholds that are more conservative when the severity of the disease is lower and less conservative when the severity of the disease is higher, suggesting that patients may be willing to try a treatment with higher level of uncertainty around its efficacy when the disease is more severe and when there are no alternatives available. Furthermore, we found the extended BDA framework to be robust, under a Gaussian prior, to the patient's uncertainty-aversion parameter and to uncertainties in the treatment's efficacy.

We then turned our attention to the development and deployment of vaccines for emerging infectious diseases. Chapter 5 developed a systematic method to evaluate and simulate vaccine allocation strategies under supply shocks, with an application to the COVID-19 pandemic. We found that, under some mild assumptions, allocating more than 50% of available doses to individuals who have not received their first dose can significantly increase the number of lives saved and significantly reduce the number of COVID-19 infections. More precisely, a 50% allocation saves on average 33% more lives, and prevents on average 32% more infections relative to a policy that guarantees a second dose to all first-dose recipients.

Chapter 6 focuses on the development of a large portfolio of mRNA vaccine candidates targeting the 11 EIDs selected by CEPI. We found that this vaccine portfolio yields an average annualized return on investment of -6.0% per annum, and a negative net present value of $-\$9.5$ billion, despite the scientific advantages of mRNA technology and the financial benefits of diversification. However, 94% of the total investment is used to fund the clinical trial for the vaccine candidates, while vaccine manufacturing costs account for only 6%. We found that the price per vaccine dose had a first order impact on the portfolio's financial performance, while other factors such as the increased probability of success due to mRNA technology, the size of the portfolio, and the possibility of conducting human challenge trials do not significantly improve its financial performance.

Moving away from healthcare finance, we considered an important problem in the empirical asset pricing literature: shrinking the set of risk factors proposed in the literature. In Chapter 7, we proposed linear and nonlinear techniques to construct latent factors from a set of 150 well-known risk factors from the asset pricing literature and compared the performance of these latent models to classical multi-factor models used routinely by finance academics and practitioners. We found that, by incorporating economic intuition into the loss function and architecture of the autoencoder, our methodology was able to capture most of the information contained in the 150 anomaly risk factors and performed better than the Fama-French 5-factors and Hou-Xue-Zhang 5-factors as a model of returns and average returns. We have also found recursive autoencoders to be powerful, not only for obtaining uncorrelated factors, but also as regularization tool that can produce factor models that are more robust to changes in the autoencoder's architecture and hyperparameters.

Finally, we explored investor risk profiles and behavioral biases in the investment management landscape by surveying a large pool of over $7,000$ individual investors,

2,300 financial advisors, and 660 institutional investors, in 2015, 2016, and 2017, about their investment decisions under various historical and hypothetical scenarios. The goal was to understand how different market participants and different types of individuals compare along the dimensions of risk aversion and investment style. We found that investors tend to be significantly more risk-averse and mostly extrapolative in their asset allocation, while institutions tend to be significantly less risk-averse and mostly contrarian in their investment decisions, with advisors falling in the middle of the risk aversion scale while also following a contrarian asset allocation strategy. Focusing solely on individual investors, we also identified five distinct types of investors: *passive investors*, *extrapolators*, *risk avoiders*, *contrarians* and *optimistic investors*. Although we found most of the surveyed individuals to be passive investors, a significant portion of investors are risk avoiders or extrapolators.

Part V

Appendices

Appendix A

Chapter 2 Supplementary Material

A.1 BDA-optimal type I error rates and clinical trial sample sizes across patient subgroups

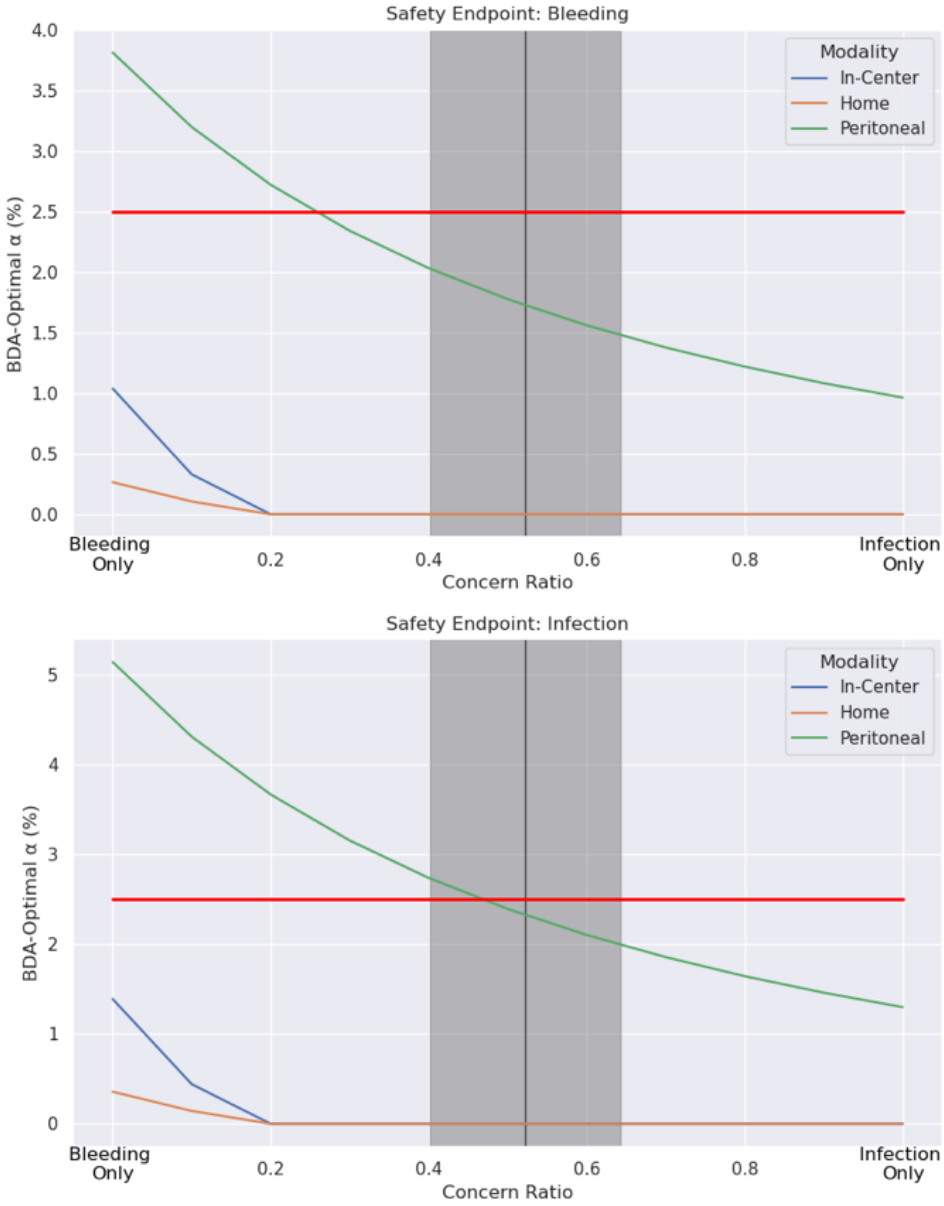


Figure A-1: **BDA-Optimal Type I Error Rate Across Dialysis Modality Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

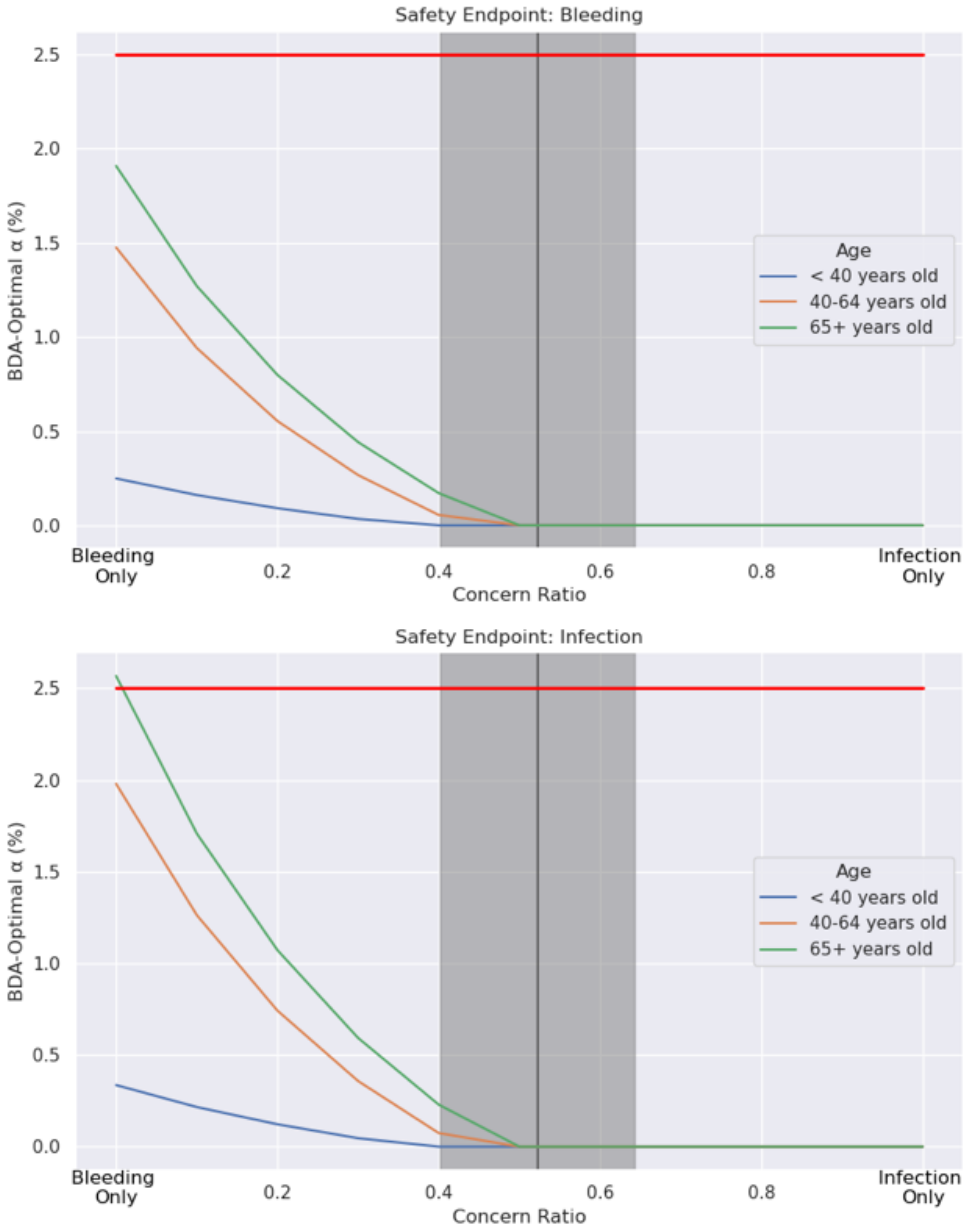


Figure A-2: **BDA-Optimal Type I Error Rate Across Age Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

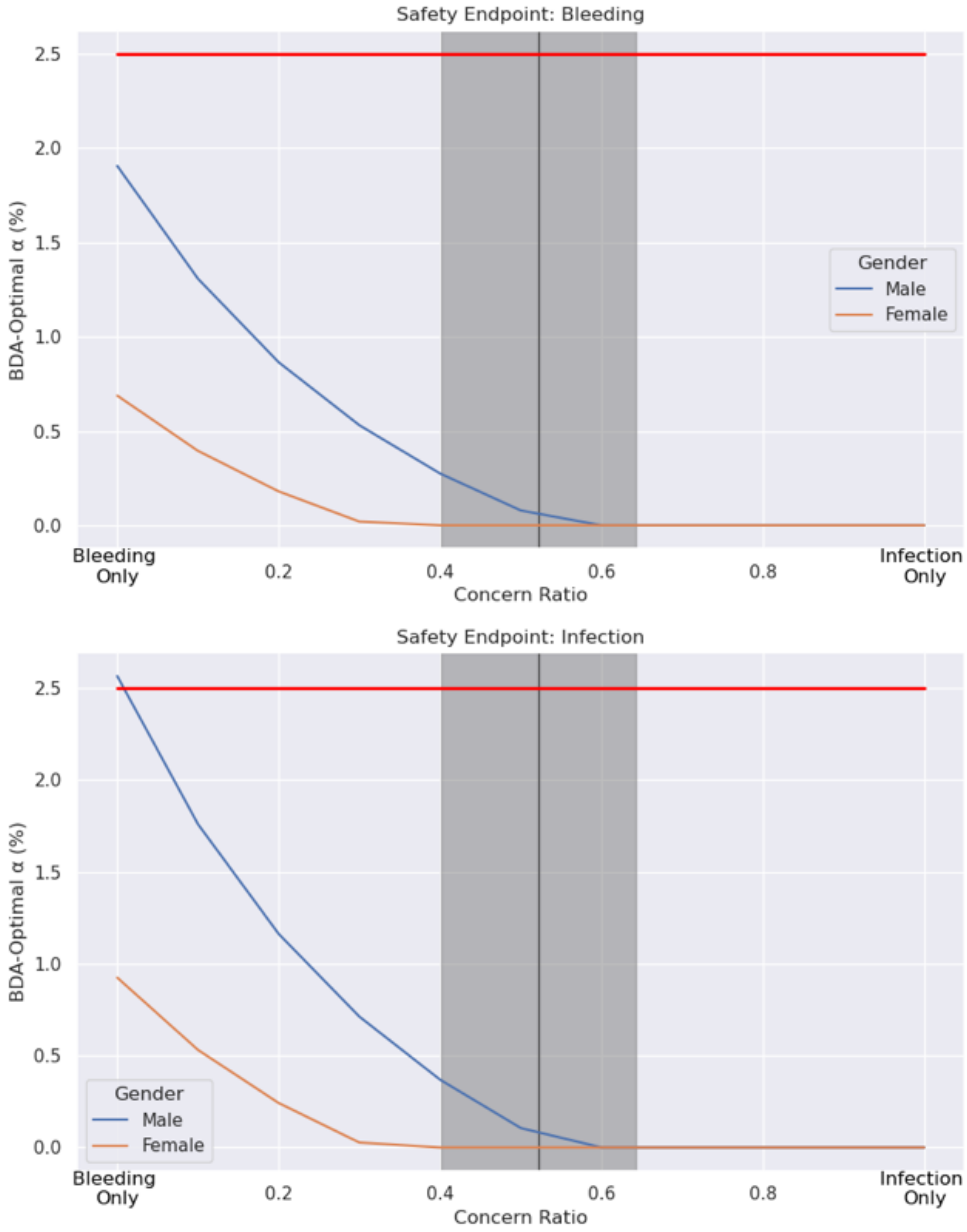


Figure A-3: **BDA-Optimal Type I Error Rate Across Gender Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

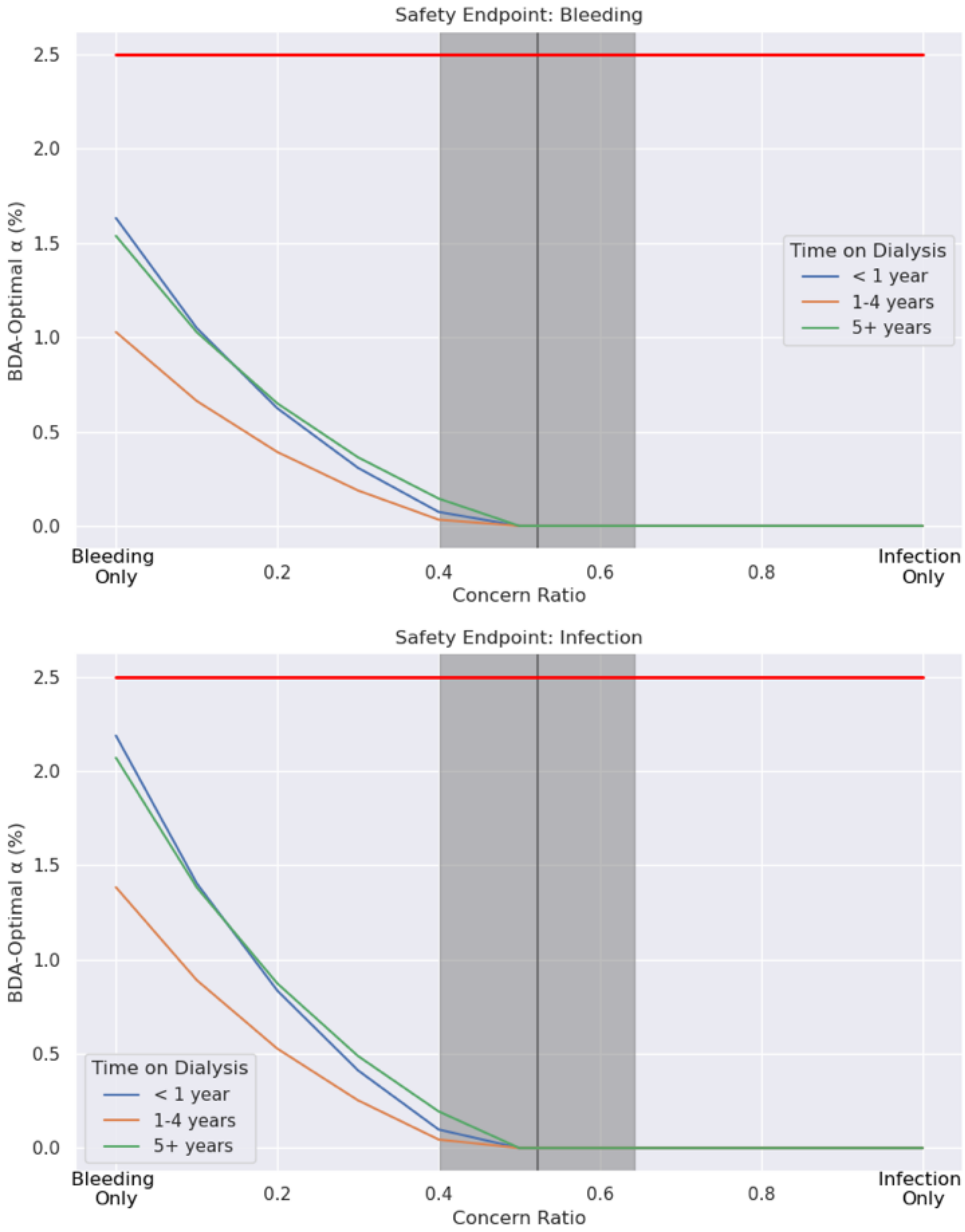


Figure A-4: **BDA-Optimal Type I Error Rate Across Time on Dialysis Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

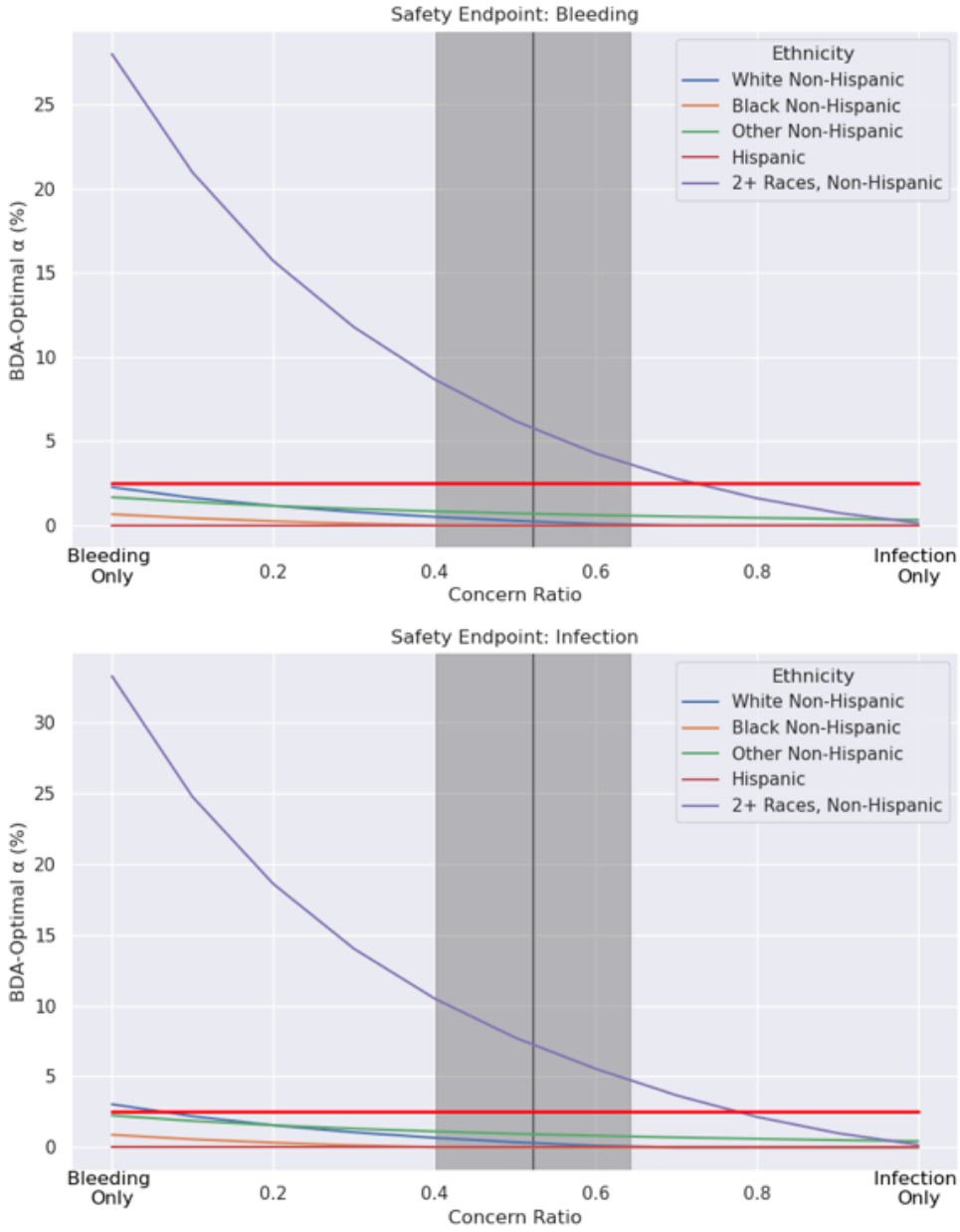


Figure A-5: **BDA-Optimal Type I Error Rate Across Ethnicity Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

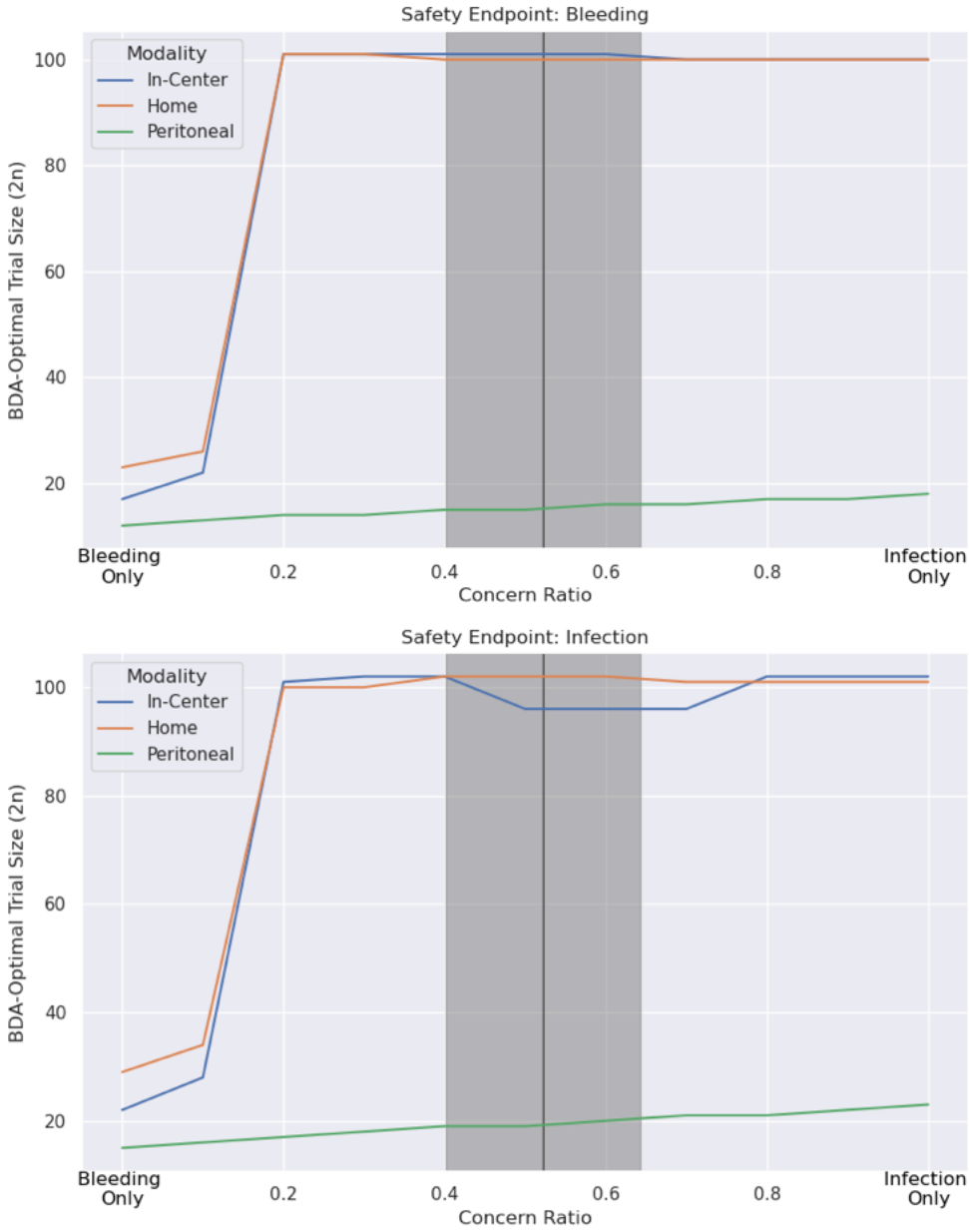


Figure A-6: **BDA-Optimal Total Sample Size Across Dialysis Modality Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

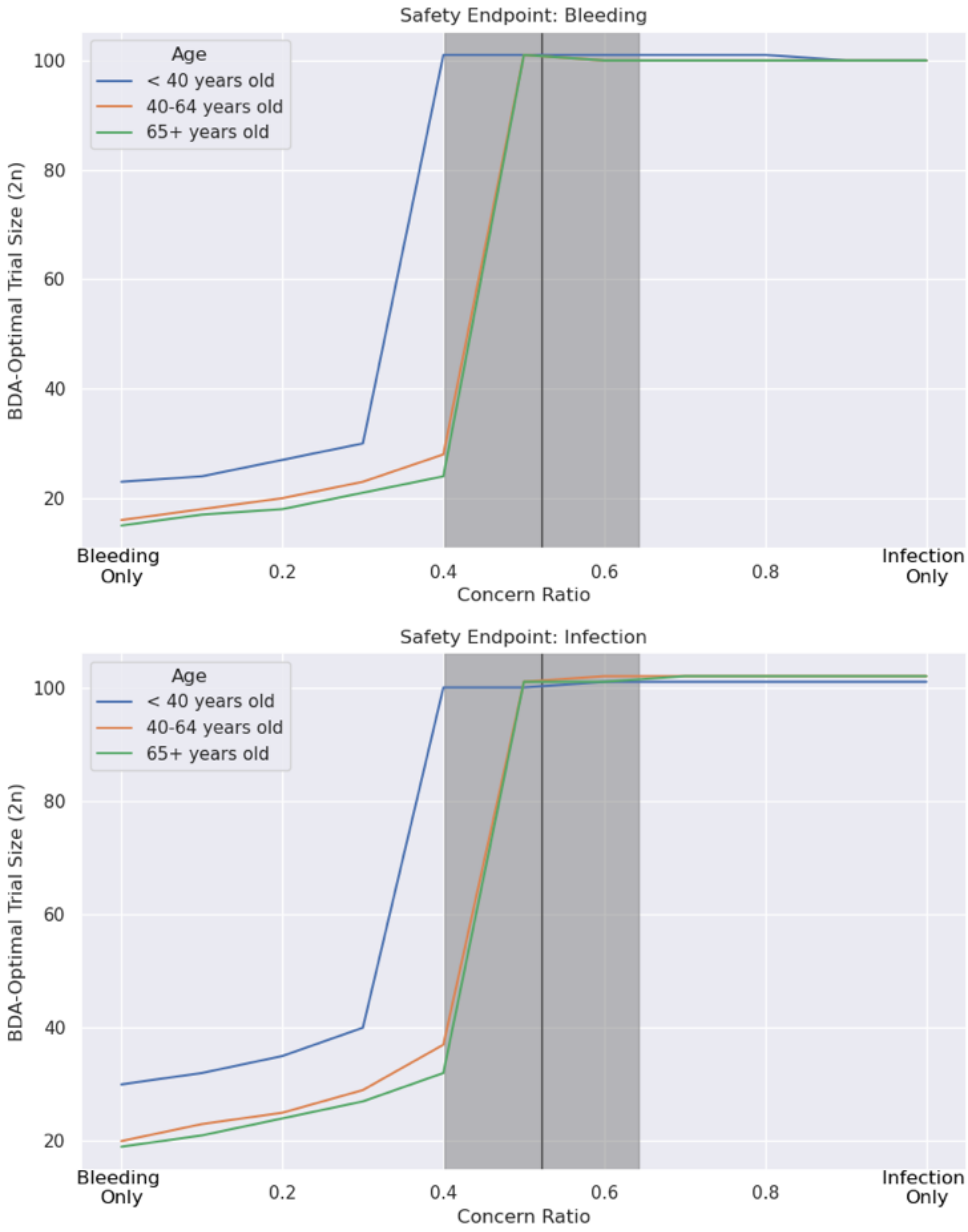


Figure A-7: **BDA-Optimal Total Sample Size Across Age Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

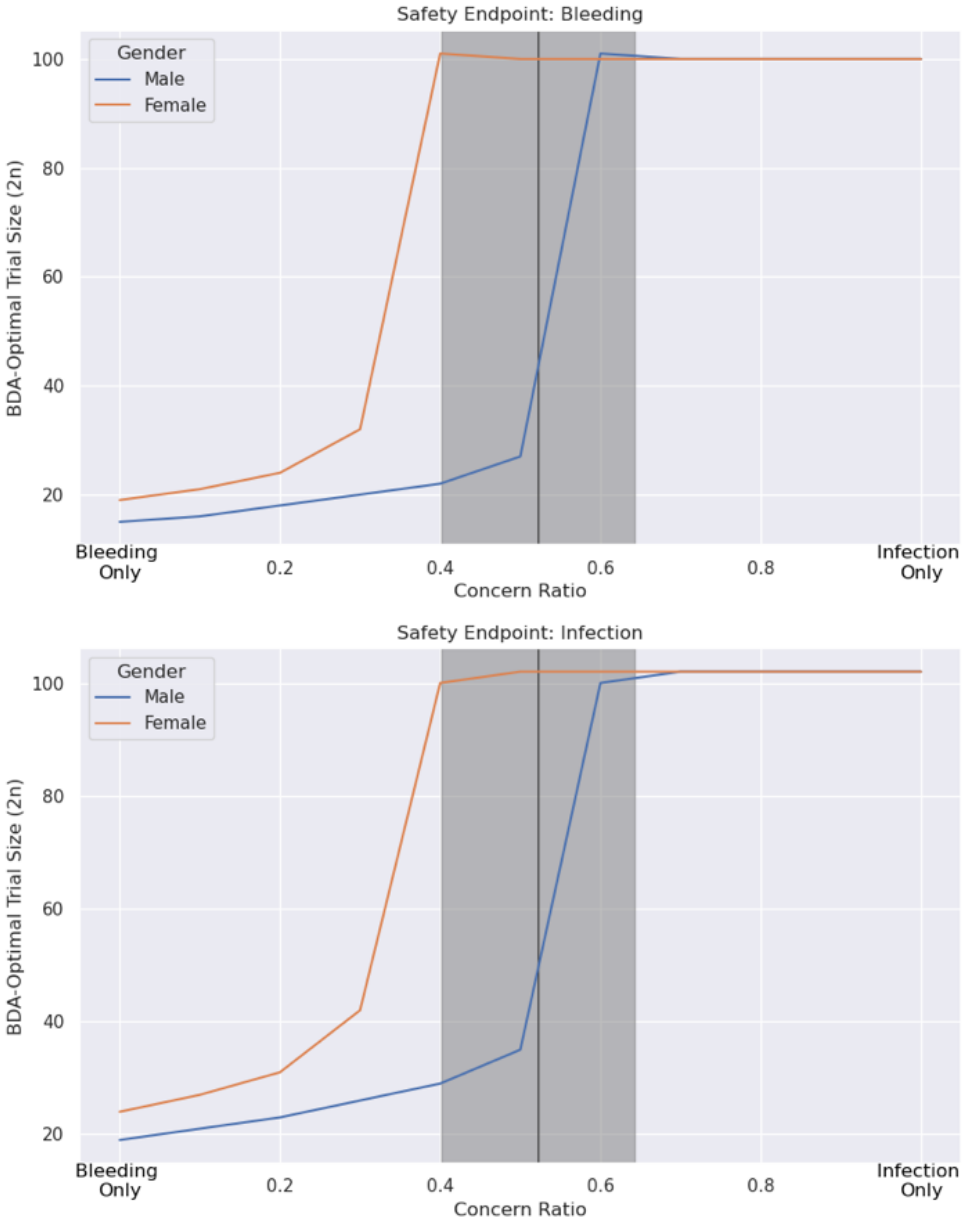


Figure A-8: **BDA-Optimal Total Sample Size Across Gender Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

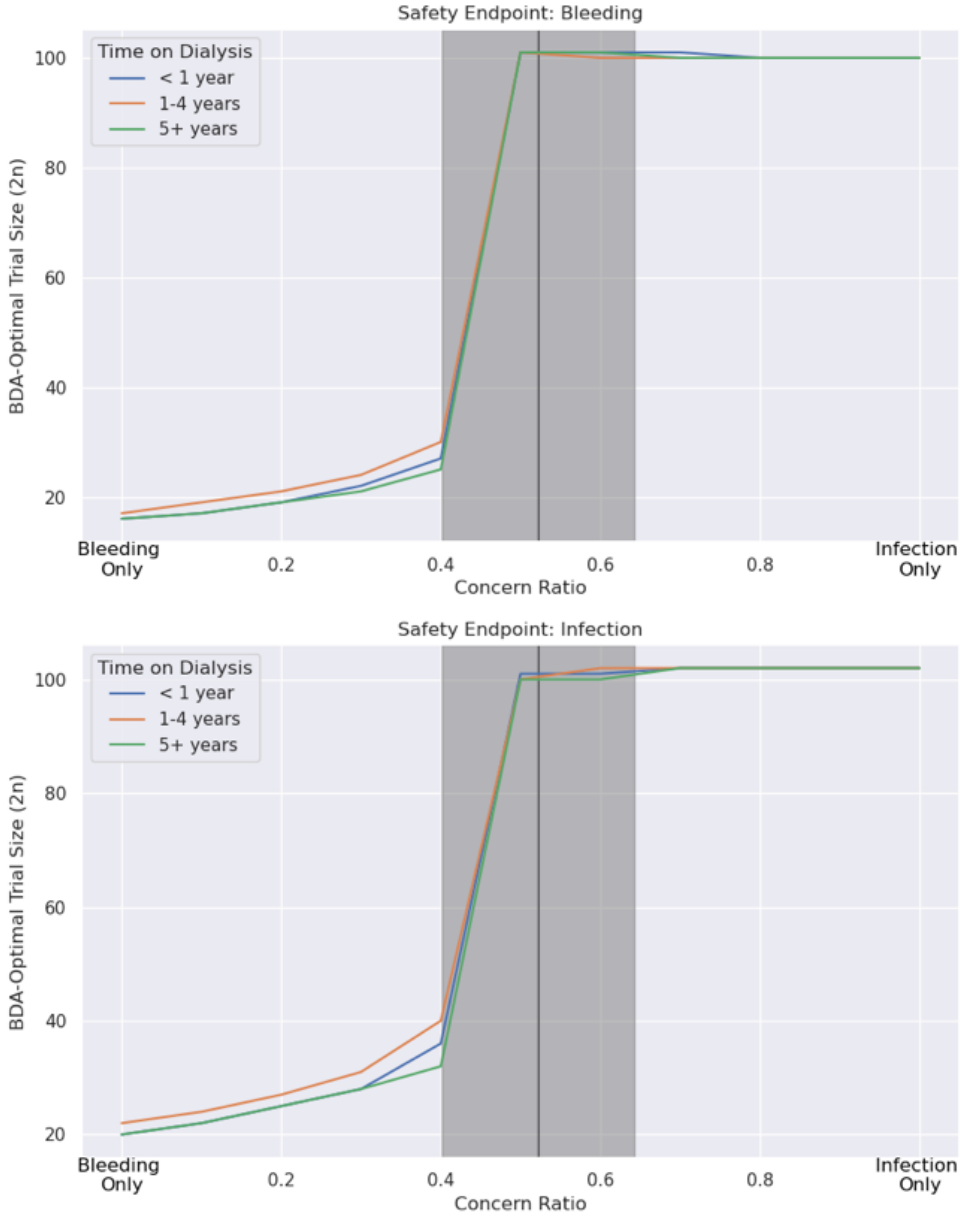


Figure A-9: **BDA-Optimal Total Sample Size Across Time on Dialysis Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

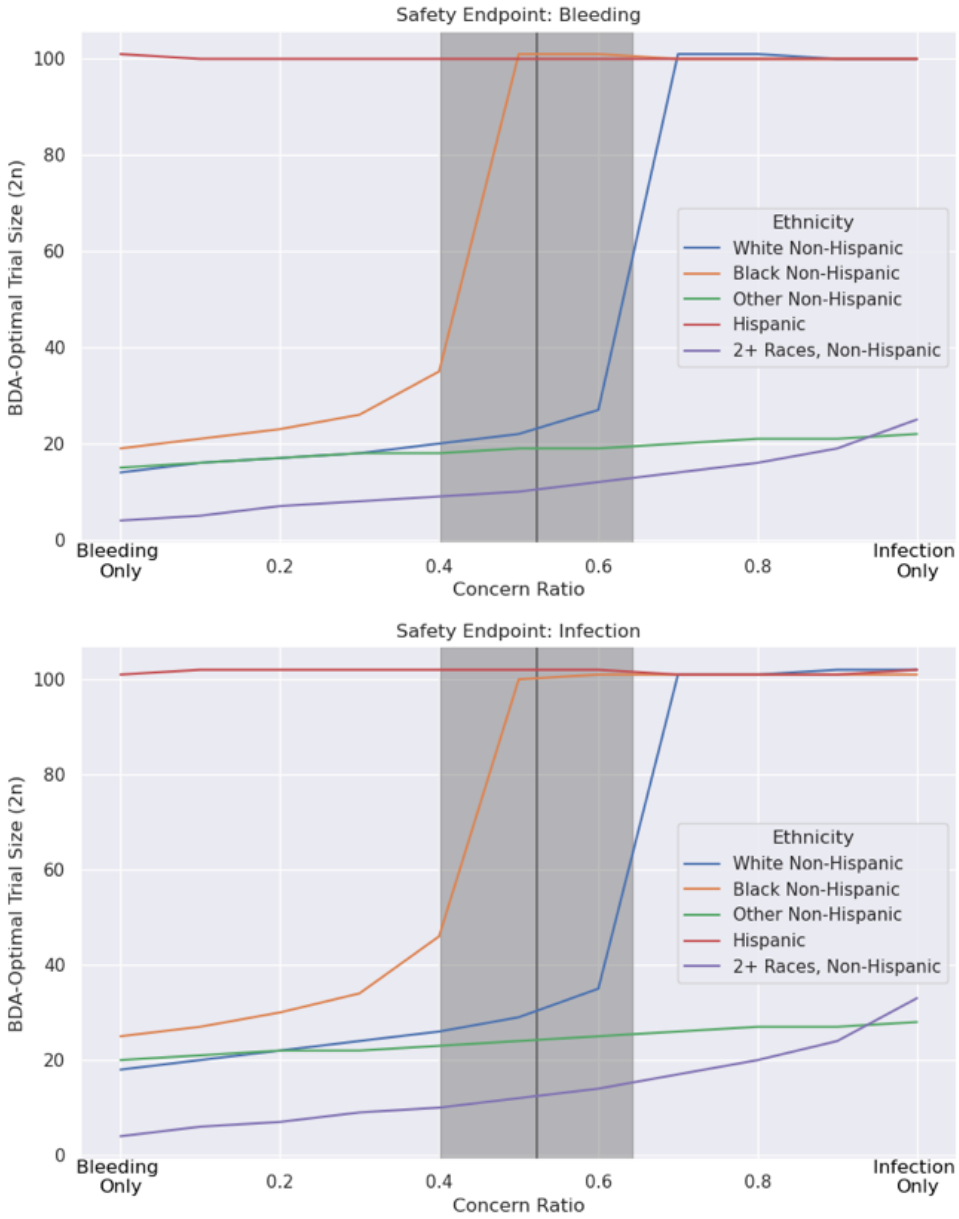


Figure A-10: **BDA-Optimal Total Sample Size Across Ethnicity Subgroups for the Risk of Serious Bleeding (top) and the Risk of Serious Infection (bottom) Endpoints.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents plus/minus one standard deviation around the mean.

A.2 Sensitivity Analysis: Safety margin

The baseline safety margin δ_0 is assumed to be 3% and 6% for the risk of serious bleeding and the risk of serious infection respectively. We consider here the impact of using a low safety margin by reducing δ_0 by 50% and a high safety margin by increasing δ_0 by 50%.

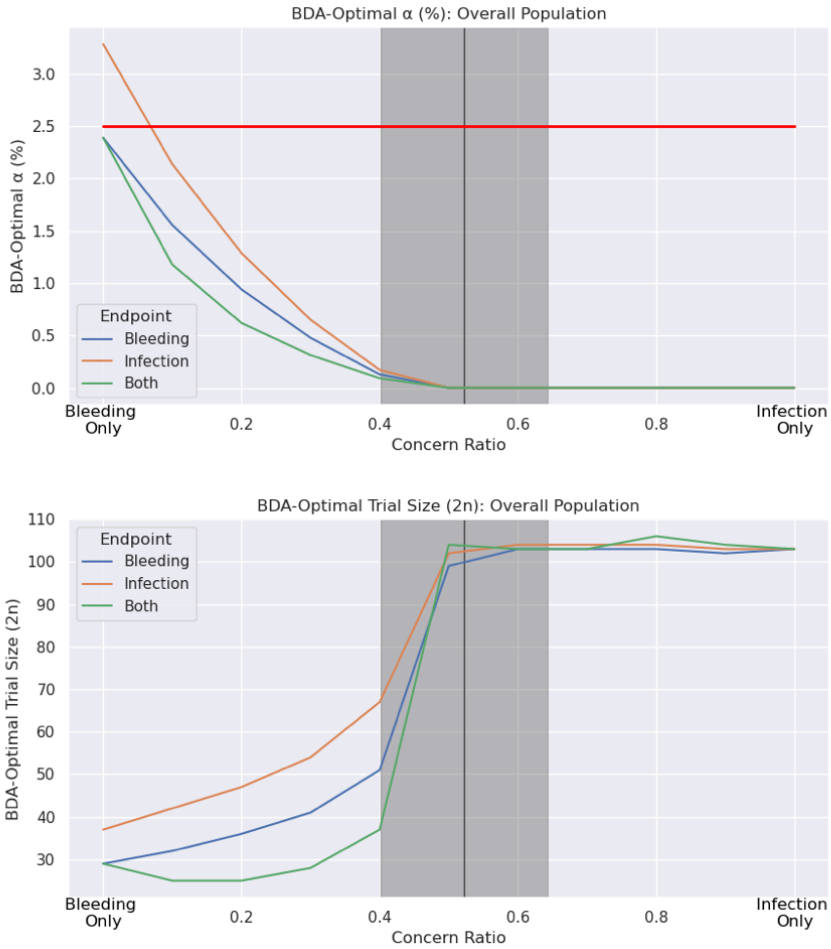


Figure A-11: **BDA-Optimal Type I Error Rate (Top) and Clinical Trial Total Sample Size (Bottom) for Three Safety Endpoints at a Low Safety Margin.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents the range of plus/minus one standard deviation around the mean.

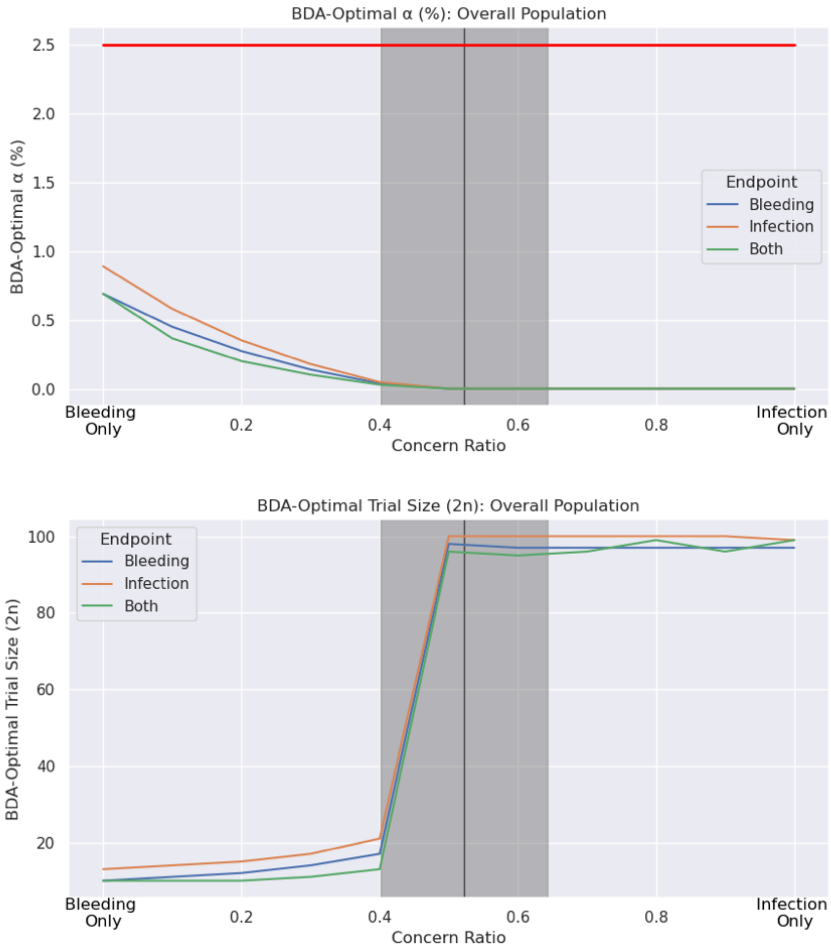


Figure A-12: **BDA-Optimal Type I Error Rate (Top) and Clinical Trial Total Sample Size (Bottom) for Three Safety Endpoints at a High Safety Margin.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents the range of plus/minus one standard deviation around the mean.

A.3 Sensitivity Analysis: Variability in the baseline and device risks

The baseline variability in the risk of serious bleeding and the risk of serious infection is assumed to be 1% and 2% respectively in the control group and 2% and 5% respectively in the treatment group. We consider here the impact of using a low variability by reducing δ_0 by 50% and a high variability by increasing δ_0 by 50%.

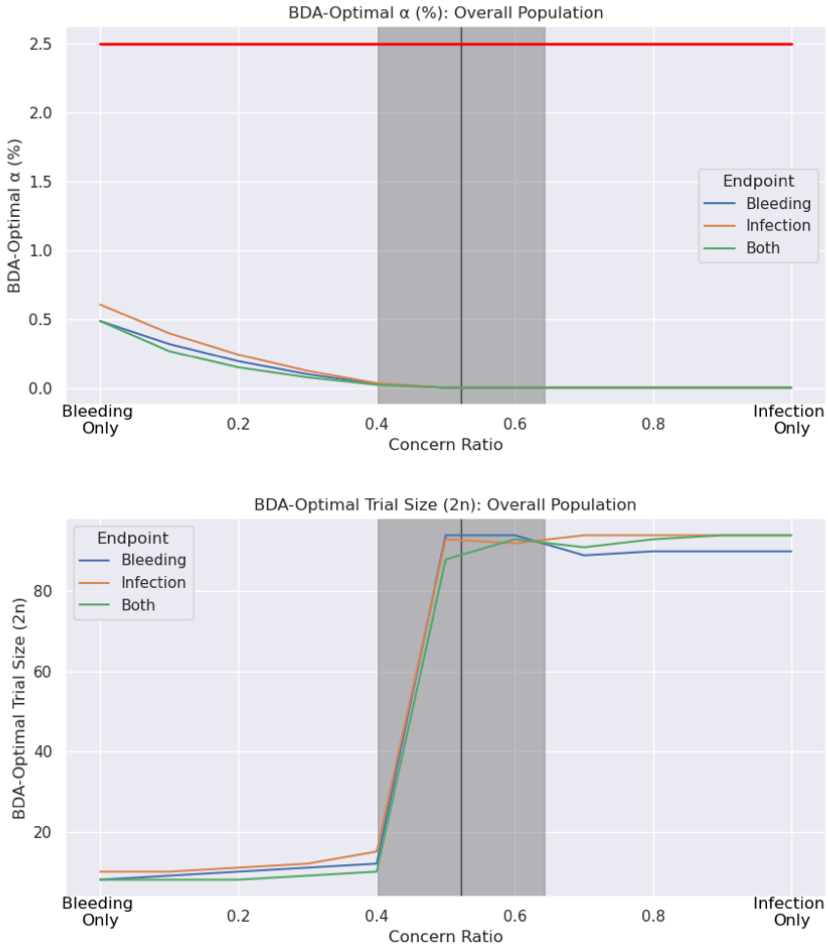


Figure A-13: **BDA-Optimal Type I Error Rate (Top) and Clinical Trial Total Sample Size (Bottom) for Three Safety Endpoints for a Low Variability.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents the range of plus/minus one standard deviation around the mean.

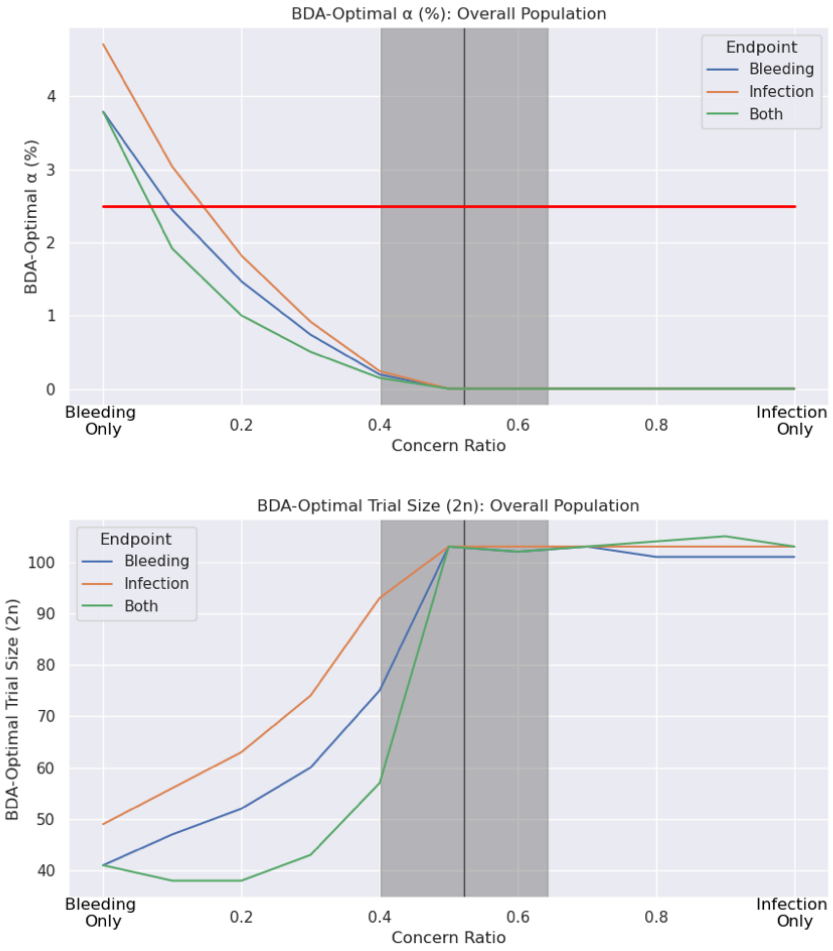


Figure A-14: **BDA-Optimal Type I Error Rate (Top) and Clinical Trial Total Sample Size (Bottom) for Three Safety Endpoints for a High Variability.** The average concern ratio across patients is represented by a vertical black line, and the shaded region represents the range of plus/minus one standard deviation around the mean.

Appendix B

Chapter 3 Supplementary Material

B.1 Sensitivity Analysis

In this section, we investigate the robustness of our results to the parameter assumptions in our model. We update the optimal balanced two-arm fixed-sample RCT for each of the subgroups in our study as we vary the power constraint ($\text{Power}_{\max}$) and the age of the patient subpopulation (Agecat). The optimal significance level and sample size associated with the perturbed parameters are given in Tables B.1 through B.6.

In Tables B.1 and B.2, we find that both the sample size and significance level decrease (increase) as the power constraint decreases (increases). By increasing the probability of a false rejection, the optimization is able to decrease both the trial length and the probability of a false approval. Furthermore, as the age increases in Tables B.3 through B.6, the delay caused by the trial length becomes more harmful as the discount rate increases. Here, the optimization attempts to further reduce the sample size of the trial at the expense of larger false approval rates.

Finally, in addition to providing recommendations for specific patient-centered RCT designs, this sensitivity analysis highlights the need for carefully considered assumptions and accurately calibrated preference models when implementing these methods.

Table B.1: BDA-optimal RCTs for Parkinson's devices with no power constraint. Indicator variables are set to 1 when a characteristic is present.

	Subpopulation							RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
1	0	0	0	0	0	0.06	17.2%	332	0.2%	94.3%
2	0	0	0	0	1	0.08	18.8%	308	0.3%	94.0%
3	0	0	0	1	0	0.28	15.9%	270	1.0%	95.4%
4	0	0	0	1	1	0.30	17.2%	258	1.2%	95.1%
5	0	1	0	0	0	0.40	19.6%	232	1.8%	94.6%
6	0	1	0	0	1	0.42	21.6%	220	2.0%	94.2%
7	0	0	1	0	0	0.46	17.7%	234	1.9%	95.2%
8	0	0	1	0	1	0.48	19.3%	224	2.1%	94.8%
9	0	1	0	1	0	0.62	18.0%	218	2.7%	95.3%
10	0	1	0	1	1	0.64	19.6%	208	2.9%	94.9%
11	0	0	1	1	0	0.68	16.3%	224	2.7%	95.8%
12	0	0	1	1	1	0.70	17.7%	214	2.9%	95.4%
13	0	1	1	0	0	0.80	20.2%	196	3.8%	95.1%
14	0	1	1	0	1	0.82	22.3%	184	4.4%	94.6%
15	0	1	1	1	0	1.02	18.4%	190	4.7%	95.5%
16	1	0	0	0	0	1.03	19.6%	186	4.9%	95.4%
17	0	1	1	1	1	1.04	20.2%	182	5.1%	95.2%
18	1	0	0	0	1	1.05	21.5%	174	5.6%	94.9%
19	1	0	0	1	0	1.25	17.9%	184	5.6%	95.8%
20	1	0	0	1	1	1.27	19.6%	174	6.3%	95.5%
21	1	1	0	0	0	1.37	22.7%	158	7.6%	95.0%
22	1	1	0	0	1	1.39	25.4%	148	8.5%	94.5%

Continued on next page

Table B.1 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
23	1	0	1	0	0	1.43	20.1%	166	7.3%	95.5%
24	1	0	1	0	1	1.46	22.2%	156	8.2%	95.2%
25	1	1	0	1	0	1.59	20.5%	160	8.2%	95.5%
26	1	1	0	1	1	1.61	22.7%	150	9.1%	95.2%
27	1	0	1	1	0	1.65	18.4%	166	7.9%	95.9%
28	1	0	1	1	1	1.67	20.1%	158	8.5%	95.6%
29	1	1	1	0	0	1.77	23.5%	142	10.5%	95.2%
30	1	1	1	0	1	1.80	26.4%	132	11.6%	94.7%
31	1	1	1	1	0	1.99	21.1%	144	10.9%	95.6%
32	1	1	1	1	1	2.02	23.5%	136	12.0%	95.3%

Table B.2: BDA-optimal RCTs for Parkinson's devices with 80% power constraint. Indicator variables are set to 1 when a characteristic is present.

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
1	0	0	0	0	0	0.06	17.2%	248	0.1%	80%
2	0	0	0	0	1	0.08	18.8%	228	0.2%	80%
3	0	0	0	1	0	0.28	15.9%	188	0.6%	80%
4	0	0	0	1	1	0.30	17.2%	188	0.6%	80%
5	0	1	0	0	0	0.40	19.6%	158	1.3%	80%
6	0	1	0	0	1	0.42	21.6%	156	1.4%	80%

Continued on next page

Table B.2 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
7	0	0	1	0	0	0.46	17.7%	158	1.3%	80%
8	0	0	1	0	1	0.48	19.3%	156	1.4%	80%
9	0	1	0	1	0	0.62	18.0%	142	2.0%	80%
10	0	1	0	1	1	0.64	19.6%	142	2.0%	80%
11	0	0	1	1	0	0.68	16.3%	142	2.0%	80%
12	0	0	1	1	1	0.70	17.7%	142	2.0%	80%
13	0	1	1	0	0	0.80	20.2%	138	2.2%	80%
14	0	1	1	0	1	0.82	22.3%	134	2.4%	80%
15	0	1	1	1	0	1.02	18.4%	116	3.7%	80%
16	1	0	0	0	0	1.03	19.6%	116	3.7%	80%
17	0	1	1	1	1	1.04	20.2%	116	3.7%	80%
18	1	0	0	0	1	1.05	21.5%	116	3.7%	80%
19	1	0	0	1	0	1.25	17.9%	116	3.7%	80%
20	1	0	0	1	1	1.27	19.6%	116	3.7%	80%
21	1	1	0	0	0	1.37	22.7%	104	5.0%	80%
22	1	1	0	0	1	1.39	25.4%	104	5.0%	80%
23	1	0	1	0	0	1.43	20.1%	104	5.0%	80%
24	1	0	1	0	1	1.46	22.2%	104	5.0%	80%
25	1	1	0	1	0	1.59	20.5%	104	5.0%	80%
26	1	1	0	1	1	1.61	22.7%	104	5.0%	80%
27	1	0	1	1	0	1.65	18.4%	104	5.0%	80%
28	1	0	1	1	1	1.67	20.1%	104	5.0%	80%
29	1	1	1	0	0	1.77	23.5%	88	7.4%	80%
30	1	1	1	0	1	1.80	26.4%	88	7.4%	80%

Continued on next page

Table B.2 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
31	1	1	1	1	0	1.99	21.1%	88	7.4%	80%
32	1	1	1	1	1	2.02	23.5%	88	7.4%	80%

Table B.3: BDA-optimal RCTs for Parkinson's devices for patients less than 61 years old. Indicator variables are set to 1 when a characteristic is present.

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
1	0	0	0	0	0	-0.06	15.6%	—	—	—
2	0	0	0	0	1	-0.03	16.8%	—	—	—
3	0	0	0	1	0	0.20	14.5%	242	0.6%	90%
4	0	0	0	1	1	0.23	15.6%	238	0.7%	90%
5	0	1	0	0	0	0.35	17.5%	210	1.2%	90%
6	0	1	0	0	1	0.38	19.1%	204	1.4%	90%
7	0	0	1	0	0	0.43	15.9%	204	1.4%	90%
8	0	0	1	0	1	0.45	17.2%	196	1.6%	90%
9	0	1	0	1	0	0.62	16.2%	188	1.9%	90%
10	0	1	0	1	1	0.65	17.5%	176	2.5%	90%
11	0	0	1	1	0	0.69	14.8%	188	1.9%	90%
12	0	0	1	1	1	0.72	15.9%	176	2.5%	90%
13	0	1	1	0	0	0.84	18.0%	164	3.2%	90%
14	0	1	1	0	1	0.87	19.6%	160	3.5%	90%

Continued on next page

Table B.3 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
15	0	1	1	1	0	1.10	16.6%	156	3.9%	90%
16	1	0	0	0	0	1.12	17.5%	154	4.0%	90%
17	0	1	1	1	1	1.13	18.0%	154	4.0%	90%
18	1	0	0	0	1	1.15	19.0%	146	4.8%	90%
19	1	0	0	1	0	1.39	16.2%	146	4.8%	90%
20	1	0	0	1	1	1.41	17.5%	144	5.0%	90%
21	1	1	0	0	0	1.53	20.0%	134	6.1%	90%
22	1	1	0	0	1	1.56	22.0%	124	7.5%	90%
23	1	0	1	0	0	1.61	17.9%	138	5.6%	90%
24	1	0	1	0	1	1.64	19.6%	124	7.5%	90%
25	1	1	0	1	0	1.80	18.2%	124	7.5%	90%
26	1	1	0	1	1	1.83	20.0%	124	7.5%	90%
27	1	0	1	1	0	1.87	16.5%	136	5.8%	90%
28	1	0	1	1	1	1.90	17.9%	124	7.5%	90%
29	1	1	1	0	0	2.02	20.5%	116	8.8%	90%
30	1	1	1	0	1	2.05	22.7%	110	10.0%	90%
31	1	1	1	1	0	2.29	18.7%	116	8.8%	90%
32	1	1	1	1	1	2.31	20.5%	110	10.0%	90%

Table B.4: BDA-optimal RCTs for Parkinson's devices for patients between 61 and 66 years old. Indicator variables are set to 1 when a characteristic is present.

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
1	0	0	0	0	0	0.04	16.1%	314	0.1%	90%
2	0	0	0	0	1	0.06	17.4%	294	0.2%	90%
3	0	0	0	1	0	0.24	14.9%	238	0.7%	90%
4	0	0	0	1	1	0.26	16.1%	228	0.8%	90%
5	0	1	0	0	0	0.35	18.1%	210	1.2%	90%
6	0	1	0	0	1	0.37	19.8%	204	1.4%	90%
7	0	0	1	0	0	0.40	16.4%	204	1.4%	90%
8	0	0	1	0	1	0.43	17.8%	196	1.6%	90%
9	0	1	0	1	0	0.55	16.7%	188	1.9%	90%
10	0	1	0	1	1	0.57	18.1%	188	1.9%	90%
11	0	0	1	1	0	0.60	15.3%	188	1.9%	90%
12	0	0	1	1	1	0.62	16.4%	188	1.9%	90%
13	0	1	1	0	0	0.71	18.6%	176	2.5%	90%
14	0	1	1	0	1	0.74	20.4%	164	3.2%	90%
15	0	1	1	1	0	0.91	17.1%	162	3.4%	90%
16	1	0	0	0	0	0.93	18.1%	160	3.5%	90%
17	0	1	1	1	1	0.93	18.6%	158	3.7%	90%
18	1	0	0	0	1	0.95	19.8%	156	3.9%	90%
19	1	0	0	1	0	1.13	16.7%	154	4.0%	90%
20	1	0	0	1	1	1.15	18.1%	154	4.0%	90%
21	1	1	0	0	0	1.24	20.7%	142	5.2%	90%
22	1	1	0	0	1	1.26	23.0%	138	5.6%	90%

Continued on next page

Table B.4 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
23	1	0	1	0	0	1.29	18.6%	146	4.8%	90%
24	1	0	1	0	1	1.31	20.3%	140	5.4%	90%
25	1	1	0	1	0	1.43	18.9%	140	5.4%	90%
26	1	1	0	1	1	1.46	20.7%	136	5.8%	90%
27	1	0	1	1	0	1.49	17.1%	142	5.2%	90%
28	1	0	1	1	1	1.51	18.6%	138	5.6%	90%
29	1	1	1	0	0	1.60	21.4%	124	7.5%	90%
30	1	1	1	0	1	1.62	23.8%	124	7.5%	90%
31	1	1	1	1	0	1.80	19.4%	124	7.5%	90%
32	1	1	1	1	1	1.82	21.4%	116	8.8%	90%

Table B.5: BDA-optimal RCTs for Parkinson's devices for patients between 67 and 71 years old. Indicator variables are set to 1 when a characteristic is present.

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
1	0	0	0	0	0	0.16	18.1%	242	0.6%	90%
2	0	0	0	0	1	0.18	19.8%	238	0.7%	90%
3	0	0	0	1	0	0.37	16.7%	210	1.2%	90%
4	0	0	0	1	1	0.39	18.1%	204	1.4%	90%
5	0	1	0	0	0	0.49	20.8%	188	1.9%	90%
6	0	1	0	0	1	0.51	23.1%	176	2.5%	90%

Continued on next page

Table B.5 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
7	0	0	1	0	0	0.55	18.6%	188	1.9%	90%
8	0	0	1	0	1	0.57	20.4%	176	2.5%	90%
9	0	1	0	1	0	0.70	19.0%	176	2.5%	90%
10	0	1	0	1	1	0.72	20.8%	164	3.2%	90%
11	0	0	1	1	0	0.76	17.1%	176	2.5%	90%
12	0	0	1	1	1	0.78	18.6%	166	3.1%	90%
13	0	1	1	0	0	0.88	21.5%	156	3.9%	90%
14	0	1	1	0	1	0.90	23.9%	154	4.0%	90%
15	0	1	1	1	0	1.09	19.5%	154	4.0%	90%
16	1	0	0	0	0	1.10	20.7%	146	4.8%	90%
17	0	1	1	1	1	1.11	21.5%	146	4.8%	90%
18	1	0	0	0	1	1.13	23.0%	142	5.2%	90%
19	1	0	0	1	0	1.32	18.9%	144	5.0%	90%
20	1	0	0	1	1	1.34	20.7%	138	5.6%	90%
21	1	1	0	0	0	1.43	24.3%	124	7.5%	90%
22	1	1	0	0	1	1.46	27.5%	116	8.8%	90%
23	1	0	1	0	0	1.49	21.4%	134	6.1%	90%
24	1	0	1	0	1	1.51	23.8%	124	7.5%	90%
25	1	1	0	1	0	1.65	21.8%	124	7.5%	90%
26	1	1	0	1	1	1.67	24.3%	116	8.8%	90%
27	1	0	1	1	0	1.70	19.4%	124	7.5%	90%
28	1	0	1	1	1	1.73	21.4%	124	7.5%	90%
29	1	1	1	0	0	1.82	25.2%	110	10.0%	90%
30	1	1	1	0	1	1.84	28.6%	104	11.3%	90%

Continued on next page

Table B.5 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
31	1	1	1	1	0	2.03	22.5%	110	10.0%	90%
32	1	1	1	1	1	2.06	25.2%	104	11.3%	90%

Table B.6: BDA-optimal RCTs for Parkinson's devices for patients 72 years old or older. Indicator variables are set to 1 when a characteristic is present.

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
1	0	0	0	0	0	0.01	19.7%	368	0.0%	90%
2	0	0	0	0	1	0.03	21.7%	314	0.1%	90%
3	0	0	0	1	0	0.21	18.1%	228	0.8%	90%
4	0	0	0	1	1	0.23	19.7%	222	0.9%	90%
5	0	1	0	0	0	0.32	22.9%	204	1.4%	90%
6	0	1	0	0	1	0.34	25.7%	196	1.6%	90%
7	0	0	1	0	0	0.37	20.3%	196	1.6%	90%
8	0	0	1	0	1	0.40	22.4%	196	1.6%	90%
9	0	1	0	1	0	0.52	20.7%	188	1.9%	90%
10	0	1	0	1	1	0.54	22.9%	176	2.5%	90%
11	0	0	1	1	0	0.57	18.5%	188	1.9%	90%
12	0	0	1	1	1	0.59	20.3%	176	2.5%	90%
13	0	1	1	0	0	0.68	23.7%	162	3.4%	90%
14	0	1	1	0	1	0.71	26.7%	156	3.9%	90%

Continued on next page

Table B.6 – *Continued from previous page*

Subpopulation								RCT Design		
No.	DBS	Non-ambulatory	Cognitive symptom	Motor symptom	Dyskinesia	Severity ratio L_1/L_0	Discount rate R	Trial size $2n$	Significance level α	Power $1-\beta$
15	0	1	1	1	0	0.88	21.3%	156	3.9%	90%
16	1	0	0	0	0	0.90	22.8%	154	4.0%	90%
17	0	1	1	1	1	0.90	23.7%	154	4.0%	90%
18	1	0	0	0	1	0.92	25.6%	146	4.8%	90%
19	1	0	0	1	0	1.10	20.6%	146	4.8%	90%
20	1	0	0	1	1	1.12	22.8%	142	5.2%	90%
21	1	1	0	0	0	1.21	27.2%	124	7.5%	90%
22	1	1	0	0	1	1.23	31.3%	124	7.5%	90%
23	1	0	1	0	0	1.26	23.6%	136	5.8%	90%
24	1	0	1	0	1	1.28	26.5%	124	7.5%	90%
25	1	1	0	1	0	1.40	24.2%	124	7.5%	90%
26	1	1	0	1	1	1.43	27.2%	124	7.5%	90%
27	1	0	1	1	0	1.46	21.2%	134	6.1%	90%
28	1	0	1	1	1	1.48	23.6%	124	7.5%	90%
29	1	1	1	0	0	1.57	28.3%	116	8.8%	90%
30	1	1	1	0	1	1.59	32.7%	104	11.3%	90%
31	1	1	1	1	0	1.77	25.0%	116	8.8%	90%
32	1	1	1	1	1	1.79	28.3%	110	10.0%	90%

Appendix C

Chapter 4 Supplementary Material

C.1 Methodology Details and Proofs

C.1.1 Mean Variance Framework: Proof

In Section 4.2.5, we claimed that the expected value and variance of the trial's loss can be calculated as follows:

$$\begin{aligned}
\mathbb{E}\left[L(\theta, T_n)\right] &= \int_\theta p_\theta \cdot \Big(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\Big)\, d\theta, && \text{(C.1)}\\
\mathrm{Var}\left(L(\theta, T_n)\right) &= \int_\theta p_\theta \cdot \Big(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c^2_{\theta,app} - c^2_{\theta,rej}) + c^2_{\theta,rej}\Big)\, d\theta && \text{(C.2)}\\
&\quad - \left(\int_\theta p_\theta \cdot \Big(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\Big)\, d\theta\right)^2 && \text{(C.3)}
\end{aligned}$$

The proof is given below.

Proof. First, we calculate the expected value of the loss.

$$\begin{aligned}
\mathbb{E}\left[L(\theta, T_n)\right] &= \mathbb{E}\Big[\mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\Big] && \text{(C.4)}\\
&= \mathbb{E}_\theta\Big[\mathbb{E}\Big[\mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}|\theta\Big]\Big] && \text{(C.5)}\\
&= \mathbb{E}_\theta\Big[\mathbb{E}\left[\mathbb{1}\{T_n > \lambda_n|\theta\}|\theta\right] \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\Big] && \text{(C.6)}\\
&= \mathbb{E}_\theta\Big[\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\Big] && \text{(C.7)}\\
&= \int_\theta p_\theta \cdot \Big(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\Big)\, d\theta. && \text{(C.8)}
\end{aligned}$$

To calculate the variance, we use the fact that

$$\mathrm{Var}\left(L(\theta, T_n)\right) = \mathbb{E}\left[L(\theta, T_n)^2\right] - \mathbb{E}\left[L(\theta, T_n)\right]^2. \qquad \text{(C.9)}$$

Indeed,

$$\mathbb{E}\left[L(\theta, T_n)^2\right] = \mathbb{E}\left[\Big(\mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\Big)^2\right] \qquad \text{(C.10)}$$

$$
\begin{aligned}
&= \mathbb{E}\left[\mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c_{\theta,app} - c_{\theta,rej})^2 + c^2_{\theta,rej} + 2 \cdot \mathbb{1}\{T_n > \lambda_n|\theta\} \cdot c_{\theta,rej} \cdot (c_{\theta,app} - c_{\theta,rej})\right] && \text{(C.11)}\\
&= \mathbb{E}\left[\mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) \cdot (c_{\theta,app} - c_{\theta,rej}) + 2 \cdot c_{\theta,rej}) + c^2_{\theta,rej}\right] && \text{(C.12)}\\
&= \mathbb{E}\left[\mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c^2_{\theta,app} - c^2_{\theta,rej}) + c^2_{\theta,rej}\right] && \text{(C.13)}\\
&= \mathbb{E}_\theta\left[\mathbb{E}\left[\mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c^2_{\theta,app} - c^2_{\theta,rej}) + c^2_{\theta,rej}|\theta\right]\right] && \text{(C.14)}\\
&= \mathbb{E}_\theta\left[\mathbb{E}\left[\mathbb{1}\{T_n > \lambda_n|\theta\}|\theta\right] \cdot (c^2_{\theta,app} - c^2_{\theta,rej}) + c^2_{\theta,rej}\right] && \text{(C.15)}\\
&= \mathbb{E}_\theta\left[\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c^2_{\theta,app} - c^2_{\theta,rej}) + c^2_{\theta,rej}\right] && \text{(C.16)}\\
&= \int_\theta p_\theta \cdot \left(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c^2_{\theta,app} - c^2_{\theta,rej}) + c^2_{\theta,rej}\right) d\theta. && \text{(C.17)}
\end{aligned}
$$

□

C.1.2 Mean Variance Framework: Toy Example

In this section, we consider a toy model to gain more intuition about the mean-variance framework. We assume throughout this section a Bernoulli prior and a discount rate $r = 0$. This means that the loss function can be written as

$$L(\theta, T_n) = \mathbb{1}\{T_n > \lambda_n|\theta\} \cdot c_{\theta,app} + (1 - \mathbb{1}\{T_n > \lambda_n|\theta\}) \cdot c_{\theta,rej}, \tag{C.18}$$

where

$$\theta = \begin{cases} 0, & \text{with probability 0.5,} \\ \mu_\theta, & \text{with probability 0.5.} \end{cases} \tag{C.19}$$

The per-patient costs simply become L_0 if we incorrectly approve an ineffective treatment (and 0 if we correctly reject an ineffective treatment), and L_1 if we incorrectly reject an effective treatment (and 0 if we correctly approve an effective treatment), as summarized in Table C.1.

Efficacy θ	Not Approved	Approved
$\theta = 0$	0	L_0
$\theta = \mu_\theta$	L_1	0

Table C.1: Loss in value per patient associated with a balanced fixed-sample RCT as a function of the treatment effect θ, assuming a Bernoulli prior and a discount rate $r = 0$.

If we denote by α and β the type I and type II errors respectively, then we know that

$$
\begin{aligned}
\alpha &= \mathbb{P}\left(T_n > \lambda_n|\theta = 0\right), && \text{(C.20)}\\
\beta &= \mathbb{P}\left(T_n \leq \lambda_n|\theta = 1\right). && \text{(C.21)}
\end{aligned}
$$

Hence, the expected loss can be written as

$$\mathbb{E}\left[\text{Loss}; n, \lambda_n\right] = \frac{1}{2} \cdot \mathbb{P}\left(T_n > \lambda_n | \theta = 0\right) \cdot L_0 + \frac{1}{2} \cdot \mathbb{P}\left(T_n \leq \lambda_n | \theta = 1\right) \cdot L_1, \tag{C.22}$$

$$= \frac{1}{2} \cdot \alpha \cdot L_0 + \frac{1}{2} \cdot \beta \cdot L_1. \tag{C.23}$$

Similarly, the expected squared loss is

$$\mathbb{E}\left[\text{Loss}^2; n, \lambda_n\right] = \frac{1}{2} \cdot \alpha \cdot L_0^2 + \frac{1}{2} \cdot \beta \cdot L_1^2, \tag{C.24}$$

and so the variance of the loss is

$$\text{Var}\left(\text{Loss}; n, \lambda_n\right) = \mathbb{E}\left[\text{Loss}^2; n, \lambda_n\right] - \mathbb{E}\left[\text{Loss}; n, \lambda_n\right], \tag{C.25}$$

$$= \frac{1}{2} \cdot \alpha \cdot L_0^2 + \frac{1}{2} \cdot \beta \cdot L_1^2 - \left(\frac{1}{4} \cdot \alpha^2 \cdot L_0^2 + \frac{1}{4} \cdot \beta^2 \cdot L_1^2 + \frac{1}{2} \cdot \alpha \cdot \beta \cdot L_0 \cdot L_1\right) \tag{C.26}$$

If we denote the severity ratio $s := L_1/L_0$, then

$$\mathbb{E}\left[\text{Loss}; n, \lambda_n\right] = \frac{L_0}{2} \cdot \left[\alpha + \beta \cdot s\right], \tag{C.27}$$

$$\text{Var}\left(\text{Loss}; n, \lambda_n\right) = \frac{L_0^2}{2} \cdot \left[\alpha + \beta \cdot s^2 - \frac{1}{2} \cdot \alpha^2 - \frac{1}{2} \cdot \beta^2 \cdot s^2 - \alpha \cdot \beta \cdot s\right] \tag{C.28}$$

$$= \frac{L_0^2}{2} \cdot \left[\alpha \cdot \left(1 - \frac{\alpha}{2}\right) + \beta \cdot \left(1 - \frac{\beta}{2}\right) \cdot s^2 - \alpha \cdot \beta \cdot s\right]. \tag{C.29}$$

Intuitively, if we view the expectation term as a function of s and the variance term as a function of s, then the former corresponds to an upwards sloping line (assuming L_0 and s are both positive) and the latter corresponds to a parabola which opens upwards. If the two curves intersect at some values s_- and s_+, then the mean-variance framework will mostly minimize the variance term for values of s below s_- and above s_+, and mostly minimize the expectation term when s is between s_- and s_+. However, if the curves do not intersect, the mean-variance framework will mostly try to minimize the variance term. Indeed, the two curves would intersect if and only if

$$\alpha + \beta \cdot s = L_0 \cdot \left[\alpha \cdot \left(1 - \frac{\alpha}{2}\right) + \beta \cdot \left(1 - \frac{\beta}{2}\right) \cdot s^2 - \alpha \cdot \beta \cdot s\right] \tag{C.30}$$

$$\iff 0 = \left[L_0 \cdot \left(1 - \frac{\alpha}{2}\right) - 1\right] \cdot \alpha - \left[1 + L_0 \cdot \alpha\right] \cdot \beta \cdot s + L_0 \cdot \beta \cdot \left(1 - \frac{\beta}{2}\right) \cdot s^2 \tag{C.31}$$

This equation has two real roots if and only if the discriminant is positive:

$$\Delta^2 := \left[1 + L_0 \cdot \alpha\right]^2 \cdot \beta^2 - 4 \cdot L_0 \cdot \alpha \cdot \beta \cdot \left(1 - \frac{\beta}{2}\right) \cdot \left[L_0 \cdot \left(1 - \frac{\alpha}{2}\right) - 1\right], \tag{C.32}$$

$$= [1 + L_0 \cdot \alpha]^2 \cdot \beta^2 - L_0 \cdot \alpha \cdot \beta \cdot (2 - \beta) \cdot [L_0 \cdot (2 - \alpha) - 2], \quad \text{(C.33)}$$

$$= \beta^2 \cdot \left[(1 + L_0 \cdot \alpha)^2 - L_0 \cdot \frac{\alpha}{\beta} \cdot (2 - \beta) \cdot [L_0 \cdot (2 - \alpha) - 2] \right]. \quad \text{(C.34)}$$

Variance. We now turn our attention to the variance term only in order to gain more intuition about it as a function of λ for different values of s. For simplicity of notation, we denote $V(\lambda) := \text{Var}(\text{Loss}; n, \lambda_n)$, $\alpha_\lambda := \frac{d\alpha}{d\lambda}$ and $\beta_\lambda := \frac{d\beta}{d\lambda}$. Then,

$$\frac{d}{d\lambda} V(\lambda) = \frac{\partial V}{\partial \alpha} \alpha_\lambda + \frac{\partial V}{\partial \beta} \beta_\lambda, \quad \text{(C.35)}$$

$$= \alpha_\lambda \cdot \frac{L_0^2}{2} \cdot [1 - \alpha - \beta \cdot s] + \beta_\lambda \cdot \frac{L_0^2}{2} \cdot \left[(1 - \beta) \cdot s^2 - \alpha \cdot s \right], \quad \text{(C.36)}$$

$$= \frac{L_0^2}{2} \cdot \left[\alpha_\lambda \cdot [1 - \alpha - \beta \cdot s] + \beta_\lambda \cdot \left[(1 - \beta) \cdot s^2 - \alpha \cdot s \right] \right], \quad \text{(C.37)}$$

$$= \frac{L_0^2}{2} \cdot \left[\alpha_\lambda \cdot (1 - \alpha) - (\alpha_\lambda \cdot \beta + \beta_\lambda \cdot \alpha) \cdot s + \beta_\lambda \cdot (1 - \beta) \cdot s^2 \right]. \quad \text{(C.38)}$$

By definition of α and β (Equations C.20 and C.21) Therefore, both $\alpha(\lambda)$ and $\beta(\lambda)$ are positive, however $\alpha(\lambda)$ is decreasing while $\beta(\lambda)$ is increasing. Hence, $\alpha_\lambda < 0$ while $\beta_\lambda > 0$, and so

$$\frac{d}{d\lambda} V(\lambda) = \frac{L_0^2}{2} \cdot \left[\underbrace{\alpha_\lambda \cdot (1 - \alpha)}_{<0} - \left(\underbrace{\alpha_\lambda \cdot \beta}_{<0} + \underbrace{\beta_\lambda \cdot \alpha}_{>0} \right) \cdot s + \underbrace{\beta_\lambda \cdot (1 - \beta)}_{>0} \cdot s^2 \right]. \quad \text{(C.39)}$$

Then when $s << 1$,

$$\frac{d}{d\lambda} V(\lambda) \sim \frac{L_0^2}{2} \cdot \alpha_\lambda \cdot (1 - \alpha) \leq 0, \quad \text{(C.40)}$$

and when $s >> 1$,

$$\frac{d}{d\lambda} V(\lambda) \sim \frac{L_0^2}{2} \cdot \beta_\lambda \cdot (1 - \beta) \cdot s^2 \geq 0. \quad \text{(C.41)}$$

Hence, this shows that for a small severity ratio, the variance term will tend to decrease with λ, and so the BDA will tend to select large values of λ (i.e., small values of α which corresponds to a more conservative decision). Conversely, for a large severity ratio, the variance term will tend to increase with λ, and so the BDA will tend to select small values of λ (i.e., large values of α which corresponds to a less conservative decision).

Regarding the intermediate case, it can be shown that the term $\alpha_\lambda \cdot \beta + \beta_\lambda \cdot \alpha$ is positive when $\lambda < \bar{\lambda}_n$ and negative when $\lambda > \bar{\lambda}_n$, where $\bar{\lambda}_n$ is defined as the root of the equation

$$\alpha_\lambda \cdot \beta + \beta_\lambda \cdot \alpha = 0. \quad \text{(C.42)}$$

Therefore,

$$\frac{d}{d\lambda}V(\lambda) = \frac{L_0^2}{2} \cdot \left[\underbrace{\alpha_\lambda \cdot (1-\alpha)}_{<0} - \underbrace{(\alpha_\lambda \cdot \beta + \beta_\lambda \cdot \alpha)}_{>0 \text{ then } <0} \cdot s + \underbrace{\beta_\lambda \cdot (1-\beta)}_{>0} \cdot s^2 \right]. \tag{C.43}$$

When $s \sim 1$, we conclude that $V(\lambda)$ is decreasing and then increasing. In this case, the BDA will tend to select an intermediate value of λ.

Expectation. We now turn our attention to the expectation term only in order to gain more intuition about it as a function of λ for different values of s. For simplicity of notation, we denote $E(\lambda) := \mathbb{E}[\text{Loss}; n, \lambda_n]$, $\alpha_\lambda := \frac{d\alpha}{d\lambda}$ and $\beta_\lambda := \frac{d\beta}{d\lambda}$. Then,

$$\begin{aligned} \frac{d}{d\lambda}E(\lambda) &= \frac{\partial E}{\partial \alpha}\alpha_\lambda + \frac{\partial E}{\partial \beta}\beta_\lambda, & \text{(C.44)} \\ &= \frac{L_0}{2} \cdot \left[\underbrace{\alpha_\lambda}_{<0} + \underbrace{\beta_\lambda \cdot s}_{>0} \right]. & \text{(C.45)} \end{aligned}$$

Hence, for a small severity ratio, the expectation term will tend to decrease with λ, and so the BDA will tend to select large values of λ (i.e., small values of α which corresponds to a more conservative decision). Conversely, for a large severity ratio, the expectation term will tend to increase with λ, and so the BDA will tend to select small values of λ (i.e., large values of α which corresponds to a less conservative decision).

Mean-Variance. Putting everything together through the mean-variance framework, the BDA aims to minimize

$$E(\lambda) + \frac{1}{2} \cdot q \cdot V(\lambda). \tag{C.46}$$

As we have seen previously, for a small severity ratio, both $E(\lambda)$ and $V(\lambda)$ tend to decrease with λ. However, for a large severity ratio, both $E(\lambda)$ and $V(\lambda)$ tend to increase with λ. Therefore, as we increase q, the BDA will tend to put much more weight on the variance term and select larger values of λ (i.e., smaller values of α, leading to a more conservative decision) when s is small and smaller values of λ (i.e., larger values of α, leading to a less conservative decision) when s is large.

This phase transition is verified in Section 4.3 under a Bernoulli prior: as we increase q, the BDA tends to lead to more conservative decisions when the severity ratio is small, and to less conservative decisions when the severity ratio is high. In this case, the phase transition occurs roughly when s is between 0.6 and 0.8.

C.1.3 In-Trial Costs

We expressed the trial's loss (given a treatment effect θ and an observation T_n of the trial's treatment effect) in Section 4.2.5 as:

$$\begin{aligned} L(\theta, T_n) &= \mathbb{1}\{\text{Treatment Approved}\} \cdot c_{\theta,app} + \mathbb{1}\{\text{Treatment Not Approved}\} \cdot c_{\theta,rej} && \text{(C.47)} \\ &= \mathbb{1}\{T_n > \lambda_n|\theta\} \cdot c_{\theta,app} + (1 - \mathbb{1}\{T_n > \lambda_n|\theta\}) \cdot c_{\theta,rej} && \text{(C.48)} \\ &= \mathbb{1}\{T_n > \lambda_n|\theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}. && \text{(C.49)} \end{aligned}$$

The loss function can be generalized to account for the loss incurred by patients enrolled in the clinical trial [126]. Indeed, the loss function can be expressed as the sum of the "post-trial" loss $L(\theta, T_n)$ and the "in-trial" loss $L^{in}(\theta)$. It is important to note here that the in-trial loss is independent of T_n because the approval or non-approval of the treatment has no impact on the patient's loss during the clinical trial period.

The post-trial loss $L(\theta, T_n)$ models the loss per person incurred by all patients affected by the disease while the in-trial loss models the loss per person incurred only by the patients enrolled in the clinical trial. Defining N as the prevalence of the disease and n as the number of patients enrolled in each arm of the trial, we can express the total loss function as:

$$\text{Loss} = L(\theta, T_n) + \frac{n}{N} \cdot L^{in}(\theta). \tag{C.50}$$

The BDA's optimization problem now becomes

$$(n^*, \lambda_n^*) = \arg\min_{n,\lambda_n} \left\{ \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] + \frac{1}{2} \cdot q \cdot \text{Var}\left(\text{Loss}; n, \lambda_n\right) \right\}, \tag{C.51}$$

where

$$\begin{aligned} \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] &= \mathbb{E}\left[L(\theta, T_n)\right] + \frac{n}{N} \cdot \mathbb{E}\left[L^{in}(\theta)\right], && \text{(C.52)} \\ \text{Var}\left(\text{Loss}; n, \lambda_n\right) &= \text{Var}\left(L(\theta, T_n)\right) + \left(\frac{n}{N}\right)^2 \cdot \text{Var}\left(L^{in}(\theta)\right) && \text{(C.53)} \\ &\quad + 2 \cdot \frac{n}{N} \cdot \text{Cov}\left(L(\theta, T_n),\ L^{in}(\theta)\right). && \text{(C.54)} \end{aligned}$$

The expressions for $\mathbb{E}\left[L(\theta, T_n)\right]$ and $\text{Var}\left(L(\theta, T_n)\right)$ are given in Section 4.2.5. The expressions for $\mathbb{E}\left[L^{in}(\theta)\right]$, $\text{Var}\left(L^{in}(\theta)\right)$, and $\text{Cov}\left(L(\theta, T_n),\ L^{in}(\theta)\right)$ are given by:

$$\begin{aligned} \mathbb{E}\left[L^{in}(\theta)\right] &= \int_\theta p_\theta \cdot L^{in}(\theta)\, d\theta, && \text{(C.55)} \\ \text{Var}\left(L^{in}(\theta)\right) &= \int_\theta p_\theta \cdot L^{in}(\theta)\, d\theta - \left(\int_\theta p_\theta \cdot L^{in}(\theta)\, d\theta\right)^2, && \text{(C.56)} \\ \text{Cov}\left(L(\theta, T_n),\ L^{in}(\theta)\right) &= \int_\theta p_\theta \cdot \left(\mathbb{P}\left(T_n > \lambda_n|\theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}\right) \cdot L^{in}(\theta)\, d\theta && \text{(C.57)} \end{aligned}$$

$$- \int_\theta \int_{\theta'} p_\theta \cdot p_{\theta'} \cdot \Big(\mathbb{P}\left(T_n > \lambda_n | \theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej} \Big) \cdot L^{in}(\theta') d\theta' d\theta \quad \text{(C.58)}$$

The proof is given below.

Proof. The expected value and variance of the in-trial loss are straightforward to calculate as $L^{in}(\theta)$ is only a function of θ. To calculate the covariance term, we use the fact that

$$\text{Cov}\left(L(\theta, T_n),\ L^{in}(\theta)\right) = \mathbb{E}\left[L(\theta, T_n) \cdot L^{in}(\theta)\right] - \mathbb{E}\left[L(\theta, T_n)\right] \cdot \mathbb{E}\left[L^{in}(\theta)\right]. \quad \text{(C.59)}$$

We know from Section 4.2.5 that

$$\mathbb{E}\left[L(\theta, T_n)\right] = \int_\theta p_\theta \cdot \Big(\mathbb{P}\left(T_n > \lambda_n | \theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej} \Big)\ d\theta, \quad \text{(C.60)}$$

therefore the only term we need to calculate is:

$$\begin{aligned}
\mathbb{E}\left[L(\theta, T_n) \cdot L^{in}(\theta)\right] &= \mathbb{E}\left[\Big(\mathbb{1}\{T_n > \lambda_n | \theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej} \Big) \cdot L^{in}(\theta)\right] && \text{(C.61)} \\
&= \mathbb{E}_\theta\left[\mathbb{E}\left[\mathbb{1}\{T_n > \lambda_n | \theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) \cdot L^{in}(\theta) + c_{\theta,rej} \cdot L^{in}(\theta) | \theta\right]\right] && \text{(C.62)} \\
&= \mathbb{E}_\theta\left[\mathbb{E}\left[\mathbb{1}\{T_n > \lambda_n | \theta\}\right] \cdot (c_{\theta,app} - c_{\theta,rej}) \cdot L^{in}(\theta) + c_{\theta,rej} \cdot L^{in}(\theta)\right] && \text{(C.63)} \\
&= \mathbb{E}_\theta\left[\mathbb{P}\left(T_n > \lambda_n | \theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) \cdot L^{in}(\theta) + c_{\theta,rej} \cdot L^{in}(\theta)\right] && \text{(C.64)} \\
&= \int_\theta p_\theta \cdot \Big(\mathbb{P}\left(T_n > \lambda_n | \theta\right) \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej} \Big) \cdot L^{in}(\theta)\, d\theta. && \text{(C.65)}
\end{aligned}$$

□

For example, in [126], the in-trial cost is defined at the $\theta = 0$ and at $\theta = \mu_\theta$ as:

$$\begin{aligned}
L^{in}(0) &= L_0, && \text{(C.66)} \\
L^{in}(\mu_\theta) &= \gamma \cdot N \cdot L_1, && \text{(C.67)}
\end{aligned}$$

for some appropriately defined costs L_0 and L_1 (denoted respectively $N \cdot c_1$ and $N \cdot c_2$ in [126]), and the incremental cost γ incurred per extra patient added to each arm. Using a linear interpolation yields

$$L^{in}(\theta) = L_0 \cdot \left(1 - \frac{\theta}{\mu_\theta}\right) + (\gamma \cdot N \cdot L_1) \cdot \frac{\theta}{\mu_\theta}. \quad \text{(C.68)}$$

C.1.4 Average Type I and Type II Error Rates

The type I error rate α is defined as the probability of approving an ineffective (and possibly harmful) treatment. Conversely, the type II error rate β is defined as the probability of not approving an effective treatment. Mathematically, these rates are

defined as:

$$\alpha = \mathbb{P}\left(T_n > \lambda_n | \theta = 0\right) \quad \text{and} \quad \beta = \mathbb{P}\left(T_n \leq \lambda_n | \theta = \mu_\theta\right). \tag{C.69}$$

The statistical power is then given by.

$$\text{Power} = 1 - \beta. \tag{C.70}$$

If we instead view α and β as functions of θ,

$$\alpha(\theta) = \mathbb{P}\left(T_n > \lambda_n | \theta\right) \quad \text{and} \quad \beta(\theta) = \mathbb{P}\left(T_n \leq \lambda_n | \theta\right), \tag{C.71}$$

then $\alpha(\theta) \leq \alpha(0) := \alpha$ for all $\theta \leq 0$, and $\beta(\theta) \leq \beta(\mu_\theta) := \beta$ for all $\theta \geq \mu_\theta$.

We can also define average type I and type II error rates as the expected value of α and β conditional on the treatment being ineffective ($\theta \leq 0$) or effective ($\theta \geq \mu_\theta$) respectively:

$$\bar{\alpha} = \frac{1}{\mathbb{P}\left(\theta \leq 0\right)} \cdot \int_{-\infty}^{0} p_\theta \cdot \alpha(\theta) \cdot d\theta = \frac{1}{\int_{-\infty}^{0} p_\theta \cdot d\theta} \cdot \int_{-\infty}^{0} p_\theta \cdot \mathbb{P}\left(T_n > \lambda_n | \theta\right) \cdot d\theta, \tag{C.72}$$

$$\bar{\beta} = \frac{1}{\mathbb{P}\left(\theta \geq \mu_\theta\right)} \cdot \int_{\mu_\theta}^{\infty} p_\theta \cdot \beta(\theta) \cdot d\theta = 2 \cdot \int_{\mu_\theta}^{\infty} p_\theta \cdot \mathbb{P}\left(T_n \leq \lambda_n | \theta\right) \cdot d\theta, \tag{C.73}$$

where we used the fact that θ is normally distributed with mean μ_θ, so $\mathbb{P}\left(\theta \geq \mu_\theta\right) = 0.5$. The average power is therefore defined as

$$\overline{\text{Power}} = 1 - \bar{\beta} = 1 - 2 \cdot \int_{\mu_\theta}^{\infty} p_\theta \cdot \mathbb{P}\left(T_n \leq \lambda_n | \theta\right) \cdot d\theta. \tag{C.74}$$

We can use the definition of the average power to constrain the BDA minimization

$$(n^*, \lambda_n^*) = \arg\min_{n, \lambda_n} \left\{ \mathbb{E}\left[\text{Loss}; n, \lambda_n\right] + \frac{1}{2} \cdot q \cdot \text{Var}\left(\text{Loss}; n, \lambda_n\right) \right\}, \tag{C.75}$$

by imposing an upper bound $P_{\max}$ on the average power: $\overline{\text{Power}} \leq P_{\max}$. Typically, $P_{\max}$ would be set to 80% or 90% to represent practical considerations of the industry, reflecting budget, time, and personnel constraints, as well as other resource limitations. Mathematically, the average power constraint can be rewritten as:

$$\int_{\mu_\theta}^{\infty} p_\theta \cdot \mathbb{P}\left(T_n \leq \lambda_n | \theta\right) \cdot d\theta \geq \frac{1}{2} \cdot (1 - P_{\max}). \tag{C.76}$$

Since $\bar{\alpha}$ and $\bar{\beta}$ will be lower than α and β respectively, the average power $\overline{\text{Power}}$ will be higher than the power estimated at $\theta = \mu_\theta$.

C.2 Additional Results

C.2.1 Results for the Parkinson's Disease Study

Outputs as a function of uncertainty-aversion

Standard errors. The following plots show the sensitivity of the BDA output's standard errors.

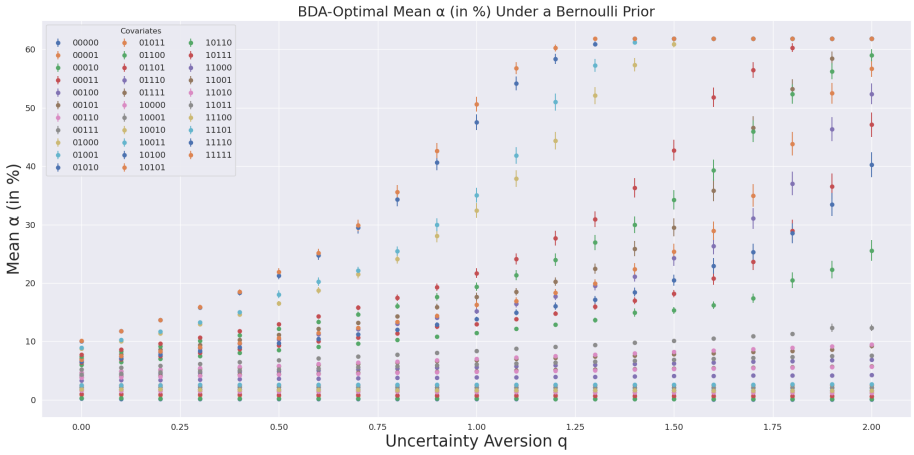


Figure C-1: BDA-Optimal significance level α (in %) as a function of the uncertainty-aversion parameter q, under a Bernoulli prior. Standard errors are calculated based on 100 bootstrap samples.

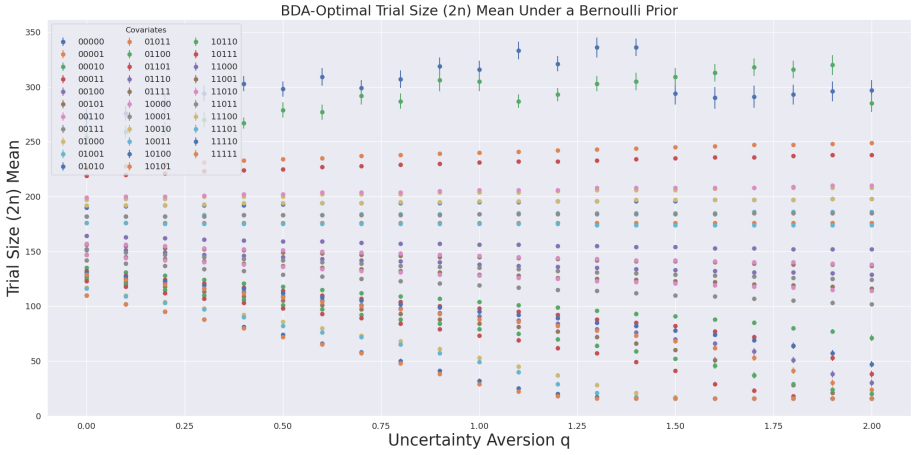


Figure C-2: BDA-Optimal trial size ($2n$) as a function of the uncertainty-aversion parameter q, under a Bernoulli prior. Standard errors are calculated based on 100 bootstrap samples.

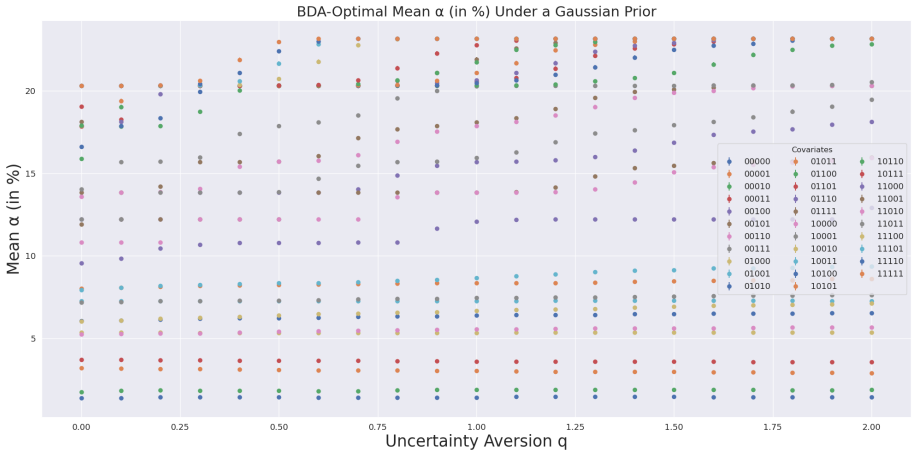


Figure C-3: BDA-Optimal significance level α (in %) as a function of the uncertainty-aversion parameter q, under a Gaussian prior. Standard errors are calculated based on 100 bootstrap samples.

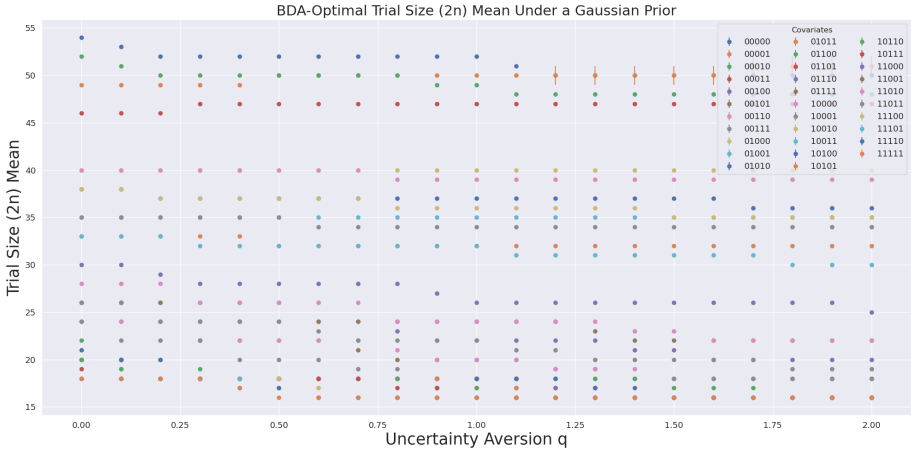


Figure C-4: BDA-Optimal trial size ($2n$) as a function of the uncertainty-aversion parameter q, under a Gaussian prior. Standard errors are calculated based on 100 bootstrap samples.

Confidence intervals. The following plots show the sensitivity of the BDA output's 95%-confidence intervals.

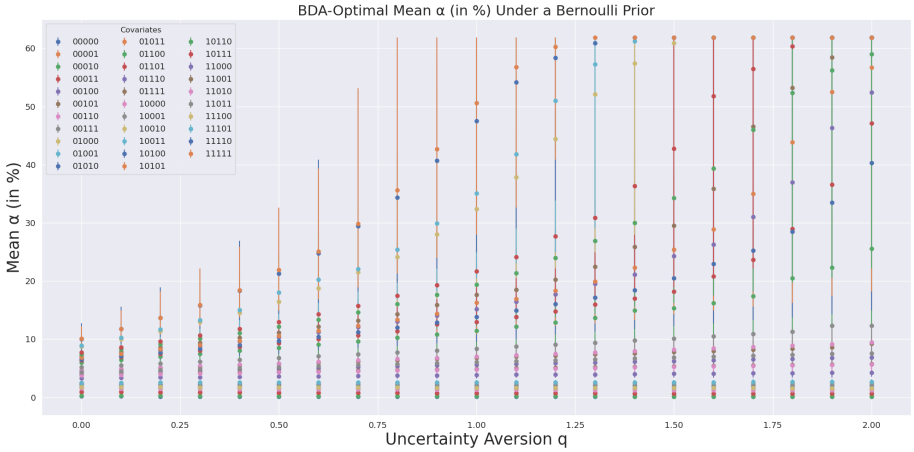


Figure C-5: BDA-Optimal significance level α (in %) as a function of the uncertainty-aversion parameter q, under a Bernoulli prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

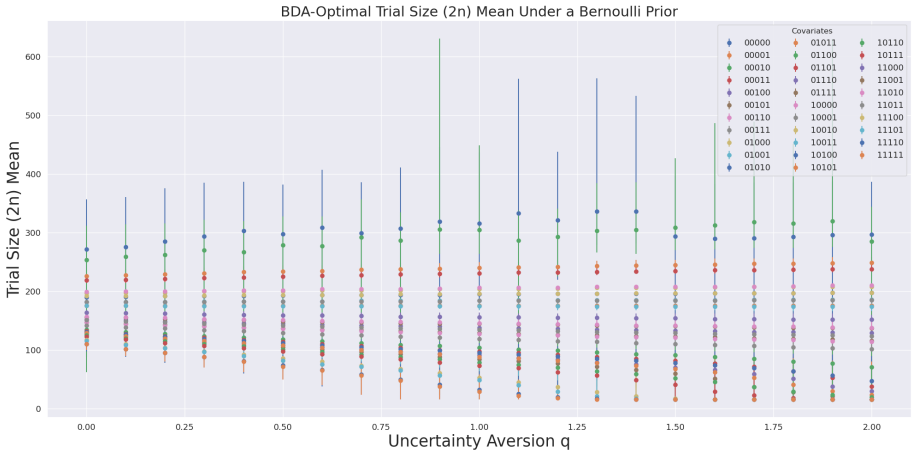


Figure C-6: BDA-Optimal trial size ($2n$) as a function of the uncertainty-aversion parameter q, under a Bernoulli prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

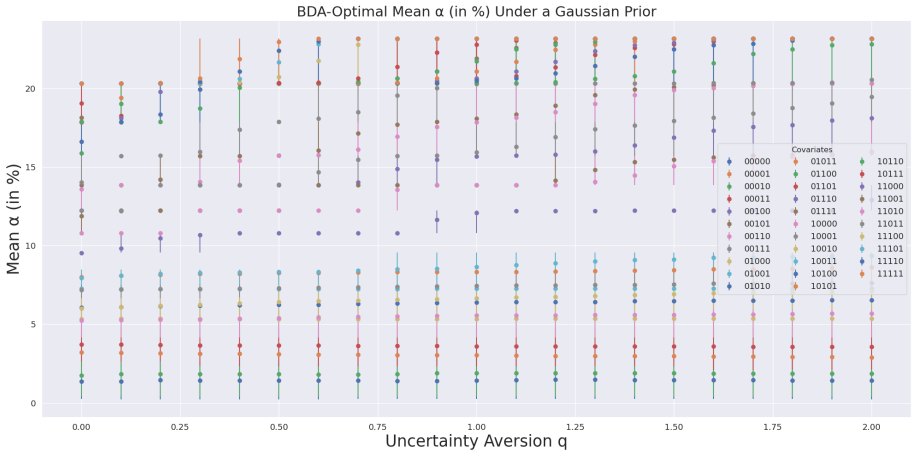


Figure C-7: BDA-Optimal significance level α (in %) as a function of the uncertainty-aversion parameter q, under a Gaussian prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

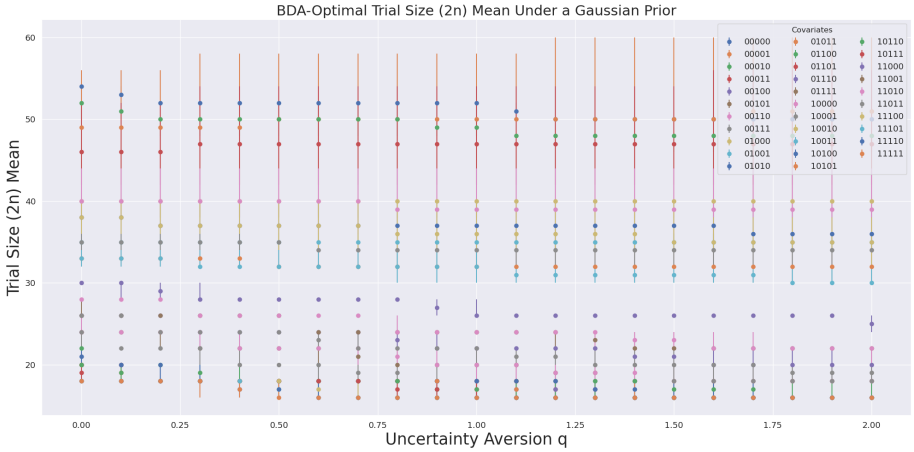


Figure C-8: BDA-Optimal trial size ($2n$) as a function of the uncertainty-aversion parameter q, under a Gaussian prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

Outputs as a function of severity

Standard errors. The following plots show the sensitivity of the BDA output's standard errors.

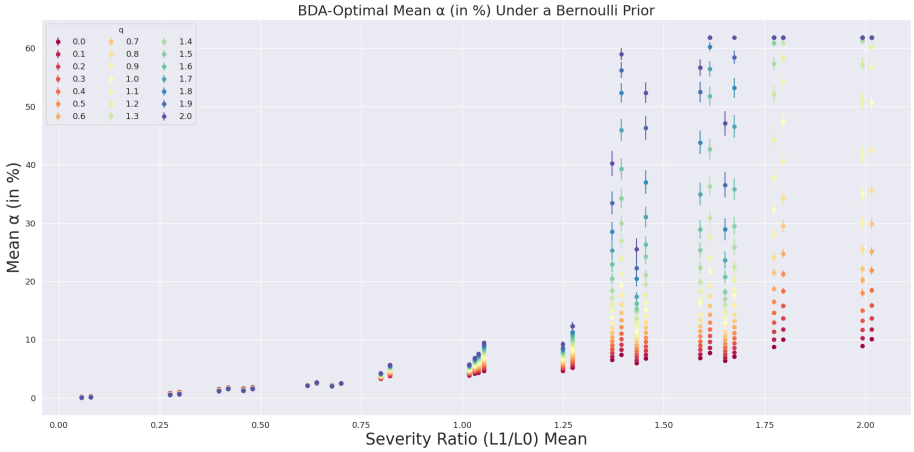


Figure C-9: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Bernoulli prior. Standard errors are calculated based on 100 bootstrap samples.

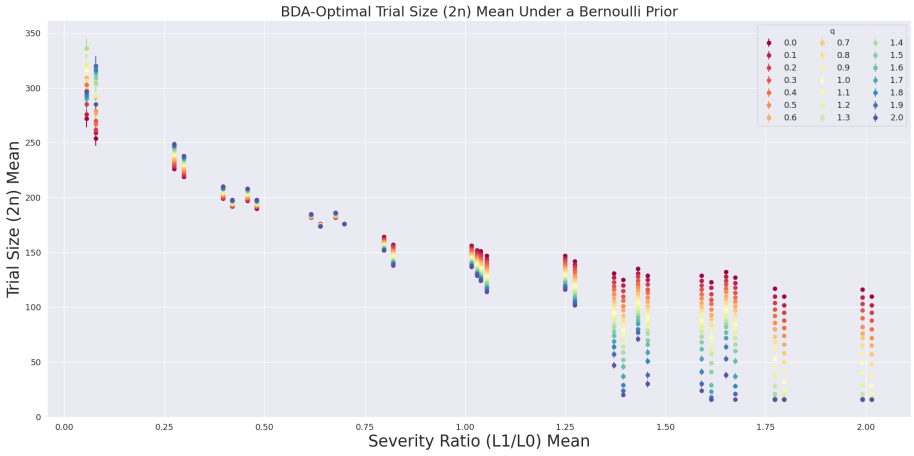


Figure C-10: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Bernoulli prior. Standard errors are calculated based on 100 bootstrap samples.

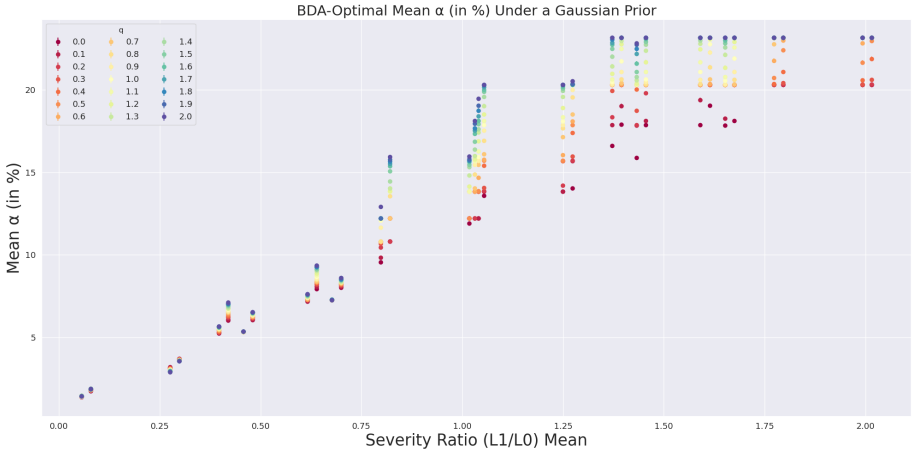


Figure C-11: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Gaussian prior. Standard errors are calculated based on 100 bootstrap samples.

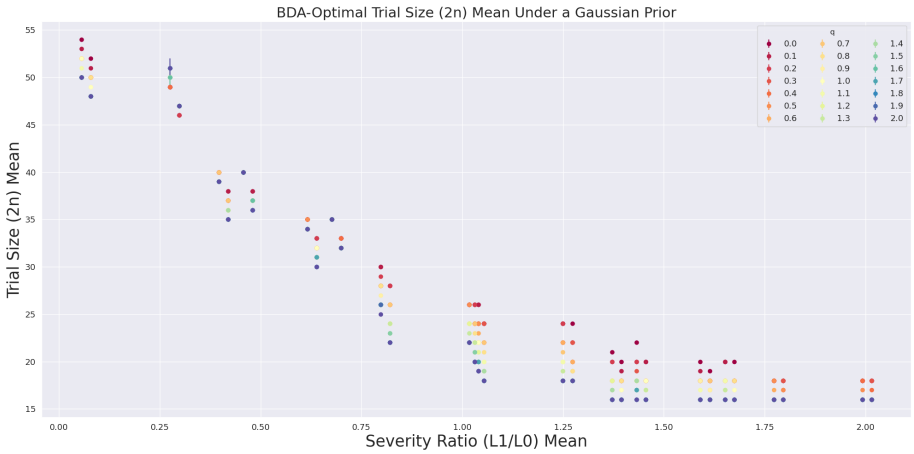


Figure C-12: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Gaussian prior. Standard errors are calculated based on 100 bootstrap samples.

Confidence intervals. The following plots show the sensitivity of the BDA output's 95%-confidence intervals.

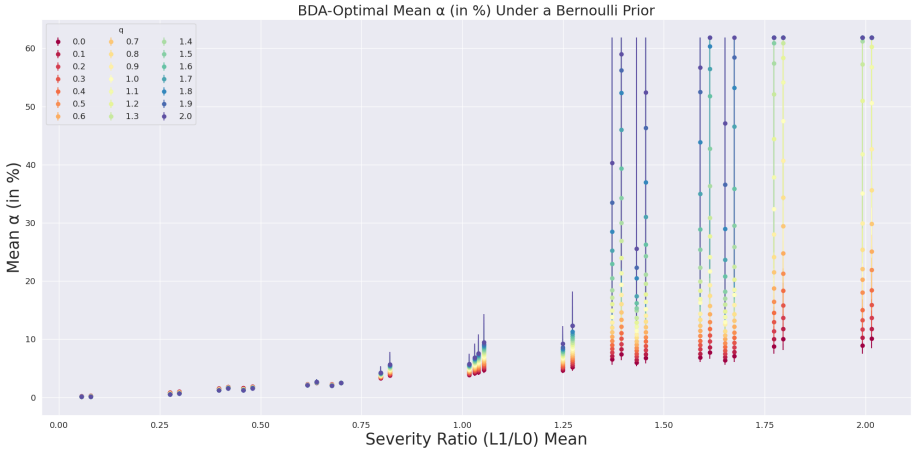


Figure C-13: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Bernoulli prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

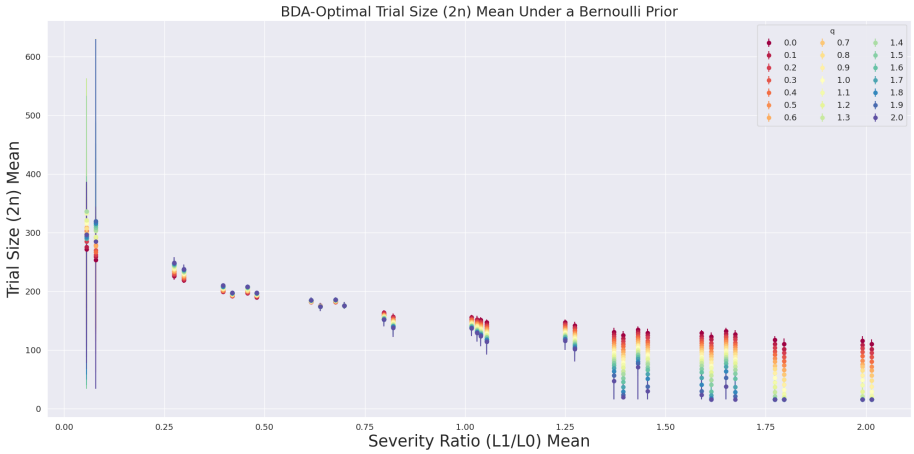


Figure C-14: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Bernoulli prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

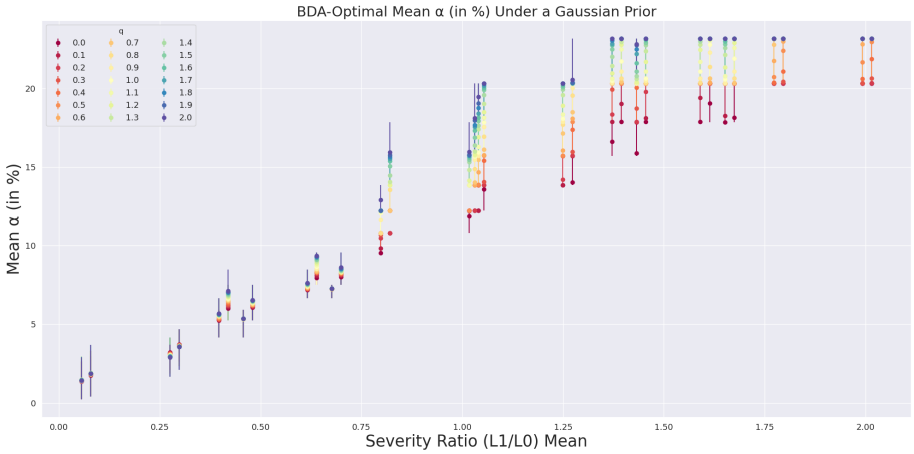


Figure C-15: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Gaussian prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

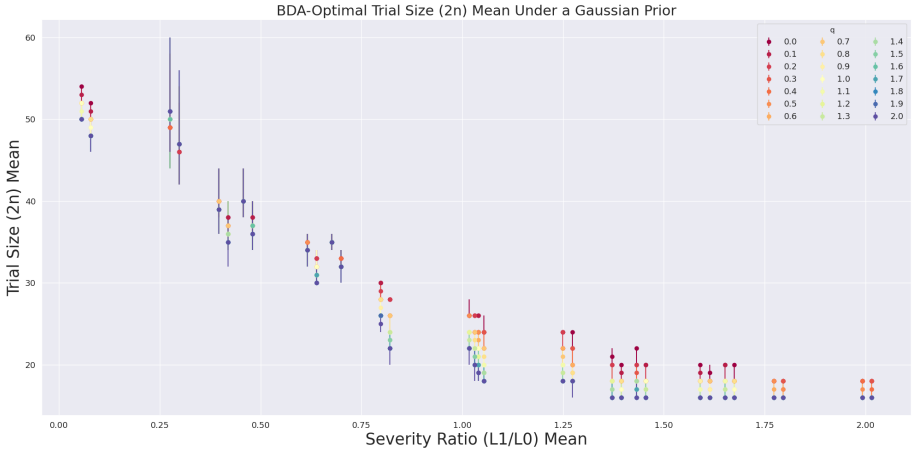


Figure C-16: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Gaussian prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

Numerical Results

Table C.2: BDA-Optimal significance levels and trial sizes across different subpopulations for different values of uncertainty-aversion q and under different prior distributions. For conciseness, we only show the results for q=0.0, 0.5, 1.0, 1.5, and 2.0. We use 100 bootstrap samples to obtain the distribution of the BDA outputs. For comparison, we present in the "NoBS" columns the non-bootstrapped BDA outputs when using the original input values from [113, 45].

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
0.06	00000	0.0	Bernoulli	0.19	0.19	0.01	0.00	0.40	294	272	8	88	357	79.20
0.06	00000	0.5	Bernoulli	0.15	0.15	0.01	0.00	0.32	304	298	7	144	382	79.20
0.06	00000	1.0	Bernoulli	0.13	0.13	0.01	0.00	0.27	312	316	8	177	436	79.20
0.06	00000	1.5	Bernoulli	0.11	0.11	0.01	0.00	0.24	320	294	10	34	396	79.20
0.06	00000	2.0	Bernoulli	0.10	0.10	0.01	0.00	0.21	324	297	9	78	387	79.20
0.06	00000	0.0	Normal	1.33	1.38	0.07	0.24	2.62	54	54	0	52	54	86.10
0.06	00000	0.5	Normal	1.37	1.43	0.07	0.23	2.62	52	52	0	50	52	85.75
0.06	00000	1.0	Normal	1.32	1.42	0.08	0.25	2.94	52	52	0	50	52	85.55
0.06	00000	1.5	Normal	1.38	1.46	0.08	0.24	2.93	50	50	0	50	50	85.24
0.06	00000	2.0	Normal	1.35	1.43	0.08	0.23	2.89	50	50	0	50	50	85.14
0.08	00010	0.0	Bernoulli	0.28	0.28	0.01	0.00	0.51	276	254	7	62	312	81.00
0.08	00010	0.5	Bernoulli	0.24	0.23	0.01	0.00	0.40	284	279	7	136	328	81.00
0.08	00010	1.0	Bernoulli	0.20	0.19	0.01	0.00	0.35	292	305	9	264	449	81.00
0.08	00010	1.5	Bernoulli	0.17	0.17	0.01	0.00	0.31	298	309	8	270	427	81.00
0.08	00010	2.0	Bernoulli	0.15	0.15	0.01	0.00	0.27	304	285	8	34	344	81.00
0.08	00010	0.0	Normal	1.73	1.75	0.08	0.42	2.94	52	52	0	50	52	86.85
0.08	00010	0.5	Normal	1.79	1.83	0.08	0.40	3.30	50	50	0	48	50	86.52
0.08	00010	1.0	Normal	1.87	1.90	0.09	0.39	3.30	48	49	0	48	50	86.20
0.08	00010	1.5	Normal	1.81	1.89	0.09	0.41	3.70	48	48	0	46	48	86.05

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
0.08	00010	2.0	Normal	1.77	1.88	0.09	0.40	3.70	48	48	0	46	48	85.92
0.28	00001	0.0	Bernoulli	0.86	0.87	0.01	0.69	0.98	226	226	1	220	236	90.00
0.28	00001	0.5	Bernoulli	0.72	0.73	0.01	0.58	0.82	234	234	1	228	244	90.00
0.28	00001	1.0	Bernoulli	0.63	0.64	0.01	0.51	0.72	240	240	1	234	250	90.00
0.28	00001	1.5	Bernoulli	0.58	0.57	0.01	0.46	0.66	244	245	1	238	254	90.00
0.28	00001	2.0	Bernoulli	0.53	0.52	0.01	0.42	0.60	248	249	1	242	258	90.00
0.28	00001	0.0	Normal	3.30	3.21	0.06	2.09	4.16	48	49	0	44	56	89.99
0.28	00001	0.5	Normal	3.30	3.10	0.06	1.86	4.16	48	50	0	44	58	89.99
0.28	00001	1.0	Normal	2.94	3.03	0.07	1.86	4.16	50	50	0	44	58	89.99
0.28	00001	1.5	Normal	2.94	2.97	0.07	1.66	4.16	50	50	1	44	60	89.99
0.28	00001	2.0	Normal	2.94	2.90	0.07	1.66	3.70	50	51	1	46	60	89.99
0.30	00011	0.0	Bernoulli	1.02	1.01	0.01	0.86	1.11	218	219	0	214	226	90.00
0.30	00011	0.5	Bernoulli	0.90	0.87	0.01	0.75	0.98	224	225	0	220	232	90.00
0.30	00011	1.0	Bernoulli	0.79	0.78	0.01	0.66	0.86	230	231	0	226	238	90.00
0.30	00011	1.5	Bernoulli	0.72	0.71	0.01	0.60	0.79	234	235	0	230	242	90.00
0.30	00011	2.0	Bernoulli	0.66	0.66	0.01	0.55	0.72	238	238	0	234	246	90.00
0.30	00011	0.0	Normal	3.70	3.71	0.06	2.62	4.68	46	46	0	42	52	90.00
0.30	00011	0.5	Normal	3.70	3.66	0.07	2.34	4.68	46	47	0	42	54	90.00
0.30	00011	1.0	Normal	3.70	3.61	0.07	2.34	4.68	46	47	0	42	54	90.00
0.30	00011	1.5	Normal	3.70	3.59	0.07	2.34	4.68	46	47	0	42	54	89.99
0.30	00011	2.0	Normal	3.70	3.58	0.08	2.09	4.68	46	47	0	42	56	89.99
0.40	10000	0.0	Bernoulli	1.58	1.56	0.00	1.51	1.58	198	199	0	198	200	90.00
0.40	10000	0.5	Bernoulli	1.44	1.43	0.00	1.38	1.44	202	202	0	202	204	90.00
0.40	10000	1.0	Bernoulli	1.32	1.32	0.00	1.32	1.32	206	206	0	206	206	90.00
0.40	10000	1.5	Bernoulli	1.27	1.27	0.00	1.27	1.27	208	208	0	208	208	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
0.40	10000	2.0	Bernoulli	1.21	1.21	0.00	1.21	1.21	210	210	0	210	210	90.00
0.40	10000	0.0	Normal	5.26	5.24	0.06	4.16	5.92	40	40	0	38	44	90.00
0.40	10000	0.5	Normal	5.26	5.41	0.07	4.16	5.92	40	40	0	38	44	90.00
0.40	10000	1.0	Normal	5.92	5.56	0.08	4.16	6.66	38	39	0	36	44	90.00
0.40	10000	1.5	Normal	5.92	5.61	0.08	4.16	6.66	38	39	0	36	44	90.00
0.40	10000	2.0	Normal	5.92	5.68	0.08	4.16	6.66	38	39	0	36	44	90.00
0.42	10010	0.0	Bernoulli	1.79	1.79	0.00	1.79	1.79	192	192	0	192	192	90.00
0.42	10010	0.5	Bernoulli	1.72	1.73	0.00	1.72	1.79	194	194	0	192	194	90.00
0.42	10010	1.0	Bernoulli	1.64	1.66	0.00	1.58	1.72	196	196	0	194	198	90.00
0.42	10010	1.5	Bernoulli	1.64	1.62	0.01	1.58	1.72	196	197	0	194	198	90.00
0.42	10010	2.0	Bernoulli	1.58	1.59	0.01	1.51	1.72	198	198	0	194	200	90.00
0.42	10010	0.0	Normal	5.92	6.02	0.05	5.26	6.66	38	38	0	36	40	90.00
0.42	10010	0.5	Normal	6.66	6.41	0.07	5.26	7.51	36	37	0	34	40	90.00
0.42	10010	1.0	Normal	6.66	6.67	0.07	5.26	7.51	36	36	0	34	40	90.00
0.42	10010	1.5	Normal	6.66	6.93	0.07	5.26	7.51	36	35	0	34	40	90.00
0.42	10010	2.0	Normal	7.51	7.12	0.08	5.92	8.47	34	35	0	32	38	90.00
0.46	01000	0.0	Bernoulli	1.64	1.62	0.00	1.58	1.64	196	197	0	196	198	90.00
0.46	01000	0.5	Bernoulli	1.51	1.50	0.00	1.44	1.51	200	200	0	200	202	90.00
0.46	01000	1.0	Bernoulli	1.38	1.38	0.00	1.38	1.38	204	204	0	204	204	90.00
0.46	01000	1.5	Bernoulli	1.32	1.32	0.00	1.32	1.32	206	206	0	206	206	90.00
0.46	01000	2.0	Bernoulli	1.27	1.27	0.00	1.21	1.27	208	208	0	208	210	90.00
0.46	01000	0.0	Normal	5.26	5.35	0.05	4.16	5.92	40	40	0	38	44	90.00
0.46	01000	0.5	Normal	5.26	5.33	0.06	4.16	5.92	40	40	0	38	44	90.00
0.46	01000	1.0	Normal	5.26	5.34	0.06	4.16	5.92	40	40	0	38	44	90.00
0.46	01000	1.5	Normal	5.26	5.36	0.07	4.16	5.92	40	40	0	38	44	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
0.46	01000	2.0	Normal	5.26	5.35	0.07	4.16	5.92	40	40	0	38	44	90.00
0.48	01010	0.0	Bernoulli	1.87	1.87	0.00	1.87	1.87	190	190	0	190	190	90.00
0.48	01010	0.5	Bernoulli	1.72	1.75	0.00	1.72	1.79	194	193	0	192	194	90.00
0.48	01010	1.0	Bernoulli	1.64	1.68	0.01	1.64	1.79	196	195	0	192	196	90.00
0.48	01010	1.5	Bernoulli	1.64	1.63	0.00	1.58	1.72	196	196	0	194	198	90.00
0.48	01010	2.0	Bernoulli	1.58	1.59	0.01	1.51	1.72	198	198	0	194	200	90.00
0.48	01010	0.0	Normal	5.92	6.07	0.05	5.26	6.66	38	38	0	36	40	90.00
0.48	01010	0.5	Normal	6.66	6.24	0.06	5.26	6.66	36	37	0	36	40	90.00
0.48	01010	1.0	Normal	6.66	6.39	0.07	5.26	7.51	36	37	0	34	40	90.00
0.48	01010	1.5	Normal	6.66	6.49	0.07	5.26	7.51	36	37	0	34	40	90.00
0.48	01010	2.0	Normal	6.66	6.55	0.07	5.26	7.51	36	36	0	34	40	90.00
0.62	10001	0.0	Bernoulli	2.22	2.24	0.00	2.22	2.32	182	182	0	180	182	90.00
0.62	10001	0.5	Bernoulli	2.22	2.18	0.01	2.04	2.32	182	183	0	180	186	90.00
0.62	10001	1.0	Bernoulli	2.13	2.15	0.01	2.04	2.32	184	184	0	180	186	90.00
0.62	10001	1.5	Bernoulli	2.13	2.11	0.01	1.95	2.32	184	184	0	180	188	90.00
0.62	10001	2.0	Bernoulli	2.04	2.09	0.01	1.87	2.32	186	185	0	180	190	90.00
0.62	10001	0.0	Normal	7.51	7.17	0.05	6.66	7.51	34	35	0	34	36	90.00
0.62	10001	0.5	Normal	7.51	7.29	0.04	6.66	7.51	34	35	0	34	36	90.00
0.62	10001	1.0	Normal	7.51	7.45	0.05	6.66	8.47	34	34	0	32	36	90.00
0.62	10001	1.5	Normal	7.51	7.55	0.06	6.66	8.47	34	34	0	32	36	90.00
0.62	10001	2.0	Normal	7.51	7.63	0.06	6.66	8.47	34	34	0	32	36	90.00
0.64	10011	0.0	Bernoulli	2.52	2.54	0.01	2.42	2.75	176	176	0	172	178	90.00
0.64	10011	0.5	Bernoulli	2.52	2.58	0.01	2.42	2.87	176	175	0	170	178	90.00
0.64	10011	1.0	Bernoulli	2.63	2.60	0.02	2.32	2.99	174	175	0	168	180	90.00
0.64	10011	1.5	Bernoulli	2.63	2.62	0.02	2.32	2.99	174	174	0	168	180	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
0.64	10011	2.0	Bernoulli	2.63	2.64	0.03	2.32	3.12	174	174	0	166	180	90.00
0.64	10011	0.0	Normal	7.51	7.94	0.05	7.51	8.47	34	33	0	32	34	90.00
0.64	10011	0.5	Normal	8.47	8.35	0.03	7.51	8.47	32	32	0	32	34	90.00
0.64	10011	1.0	Normal	8.47	8.65	0.06	7.51	9.56	32	32	0	30	34	90.00
0.64	10011	1.5	Normal	9.56	9.14	0.06	8.47	9.56	30	31	0	30	32	90.00
0.64	10011	2.0	Normal	9.56	9.36	0.05	8.47	9.56	30	30	0	30	32	90.00
0.68	01001	0.0	Bernoulli	2.22	2.24	0.00	2.22	2.32	182	182	0	180	182	90.00
0.68	01001	0.5	Bernoulli	2.13	2.17	0.01	2.04	2.32	184	183	0	180	186	90.00
0.68	01001	1.0	Bernoulli	2.13	2.12	0.01	1.95	2.32	184	184	0	180	188	90.00
0.68	01001	1.5	Bernoulli	2.04	2.07	0.01	1.95	2.22	186	185	0	182	188	90.00
0.68	01001	2.0	Bernoulli	2.04	2.04	0.01	1.87	2.22	186	186	0	182	190	90.00
0.68	01001	0.0	Normal	7.51	7.27	0.04	6.66	7.51	34	35	0	34	36	90.00
0.68	01001	0.5	Normal	7.51	7.26	0.04	6.66	7.51	34	35	0	34	36	90.00
0.68	01001	1.0	Normal	7.51	7.27	0.04	6.66	7.51	34	35	0	34	36	90.00
0.68	01001	1.5	Normal	7.51	7.28	0.04	6.66	7.51	34	35	0	34	36	90.00
0.68	01001	2.0	Normal	7.51	7.28	0.04	6.66	7.51	34	35	0	34	36	90.00
0.70	01011	0.0	Bernoulli	2.52	2.51	0.01	2.42	2.63	176	176	0	174	178	90.00
0.70	01011	0.5	Bernoulli	2.52	2.52	0.01	2.32	2.75	176	176	0	172	180	90.00
0.70	01011	1.0	Bernoulli	2.52	2.52	0.02	2.32	2.75	176	176	0	172	180	90.00
0.70	01011	1.5	Bernoulli	2.52	2.51	0.02	2.32	2.87	176	176	0	170	180	90.00
0.70	01011	2.0	Bernoulli	2.52	2.51	0.02	2.22	2.87	176	176	0	170	182	90.00
0.70	01011	0.0	Normal	7.51	8.01	0.05	7.51	8.47	34	33	0	32	34	90.00
0.70	01011	0.5	Normal	8.47	8.26	0.04	7.51	8.47	32	32	0	32	34	90.00
0.70	01011	1.0	Normal	8.47	8.35	0.03	7.51	8.47	32	32	0	32	34	90.00
0.70	01011	1.5	Normal	8.47	8.46	0.05	7.51	9.56	32	32	0	30	34	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
0.70	01011	2.0	Normal	8.47	8.62	0.06	7.51	9.56	32	32	0	30	34	90.00
0.80	11000	0.0	Bernoulli	3.26	3.29	0.02	2.99	3.70	164	164	0	158	168	90.00
0.80	11000	0.5	Bernoulli	3.55	3.61	0.03	3.25	4.20	160	159	0	152	164	90.00
0.80	11000	1.0	Bernoulli	3.86	3.85	0.04	3.26	4.56	156	156	1	148	164	90.00
0.80	11000	1.5	Bernoulli	4.02	4.08	0.05	3.40	4.96	154	154	1	144	162	90.00
0.80	11000	2.0	Bernoulli	4.20	4.25	0.07	3.55	5.39	152	152	1	140	160	90.00
0.80	11000	0.0	Normal	9.56	9.55	0.01	9.56	9.56	30	30	0	30	30	90.00
0.80	11000	0.5	Normal	10.81	10.80	0.01	10.81	10.81	28	28	0	28	28	90.00
0.80	11000	1.0	Normal	12.23	12.09	0.04	10.81	12.23	26	26	0	26	28	90.00
0.80	11000	1.5	Normal	12.23	12.23	0.00	12.23	12.23	26	26	0	26	26	90.00
0.80	11000	2.0	Normal	13.85	12.93	0.08	12.23	13.85	24	25	0	24	26	90.00
0.82	11010	0.0	Bernoulli	3.70	3.76	0.03	3.40	4.20	158	157	0	152	162	90.00
0.82	11010	0.5	Bernoulli	4.20	4.31	0.05	3.70	5.17	152	151	0	142	158	90.00
0.82	11010	1.0	Bernoulli	4.76	4.81	0.07	4.02	6.11	146	146	1	134	154	90.00
0.82	11010	1.5	Bernoulli	5.17	5.27	0.10	4.20	6.91	142	142	1	128	152	90.00
0.82	11010	2.0	Bernoulli	5.39	5.69	0.12	4.37	7.82	140	138	1	122	150	90.00
0.82	11010	0.0	Normal	10.81	10.81	0.00	10.81	10.81	28	28	0	28	28	90.00
0.82	11010	0.5	Normal	12.23	12.23	0.00	12.23	12.23	26	26	0	26	26	90.00
0.82	11010	1.0	Normal	13.85	13.85	0.00	13.85	13.85	24	24	0	24	24	90.00
0.82	11010	1.5	Normal	15.71	15.08	0.09	13.85	15.71	22	23	0	22	24	90.00
0.82	11010	2.0	Normal	15.71	15.95	0.07	15.71	17.85	22	22	0	20	22	90.00
1.02	11001	0.0	Bernoulli	3.86	3.87	0.02	3.55	4.37	156	156	0	150	160	90.00
1.02	11001	0.5	Bernoulli	4.37	4.44	0.04	3.86	5.17	150	149	0	142	156	90.00
1.02	11001	1.0	Bernoulli	4.76	4.92	0.06	4.20	6.11	146	145	1	134	152	90.00
1.02	11001	1.5	Bernoulli	5.17	5.38	0.08	4.37	6.91	142	141	1	128	150	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
1.02	11001	2.0	Bernoulli	5.62	5.76	0.10	4.56	7.51	138	137	1	124	148	90.00
1.02	11001	0.0	Normal	12.23	11.90	0.06	10.81	12.23	26	26	0	26	28	90.00
1.02	11001	0.5	Normal	12.23	12.23	0.00	12.23	12.23	26	26	0	26	26	90.00
1.02	11001	1.0	Normal	13.85	13.85	0.00	13.85	13.85	24	24	0	24	24	90.00
1.02	11001	1.5	Normal	15.71	15.49	0.06	13.85	15.71	22	22	0	22	24	90.00
1.02	11001	2.0	Normal	15.71	15.97	0.07	15.71	17.85	22	22	0	20	22	90.00
1.03	00100	0.0	Bernoulli	4.20	4.16	0.03	3.70	4.76	152	152	0	146	158	90.00
1.03	00100	0.5	Bernoulli	4.96	4.93	0.05	4.20	5.86	144	145	0	136	152	90.00
1.03	00100	1.0	Bernoulli	5.62	5.63	0.08	4.56	7.20	138	138	1	126	148	90.00
1.03	00100	1.5	Bernoulli	6.11	6.27	0.11	4.96	8.15	134	133	1	120	144	90.00
1.03	00100	2.0	Bernoulli	6.63	6.88	0.14	5.17	9.21	130	129	1	114	142	90.00
1.03	00100	0.0	Normal	12.23	12.23	0.00	12.23	12.23	26	26	0	26	26	90.00
1.03	00100	0.5	Normal	13.85	13.85	0.00	13.85	13.85	24	24	0	24	24	90.00
1.03	00100	1.0	Normal	15.71	15.69	0.02	15.71	15.71	22	22	0	22	22	90.00
1.03	00100	1.5	Normal	17.85	16.89	0.11	15.71	17.85	20	21	0	20	22	90.00
1.03	00100	2.0	Normal	17.85	18.12	0.08	17.85	20.32	20	20	0	18	20	90.00
1.04	11011	0.0	Bernoulli	4.37	4.34	0.03	3.86	4.96	150	151	0	144	156	90.00
1.04	11011	0.5	Bernoulli	5.17	5.23	0.06	4.37	6.36	142	142	1	132	150	90.00
1.04	11011	1.0	Bernoulli	5.86	6.09	0.09	4.96	7.82	136	135	1	122	144	90.00
1.04	11011	1.5	Bernoulli	6.63	6.86	0.13	5.39	9.21	130	129	1	114	140	90.00
1.04	11011	2.0	Bernoulli	7.20	7.62	0.17	5.62	10.83	126	124	1	106	138	90.00
1.04	11011	0.0	Normal	12.23	12.23	0.00	12.23	12.23	26	26	0	26	26	90.00
1.04	11011	0.5	Normal	13.85	13.89	0.03	13.85	13.85	24	24	0	24	24	90.00
1.04	11011	1.0	Normal	15.71	15.95	0.07	15.71	17.85	22	22	0	20	22	90.00
1.04	11011	1.5	Normal	17.85	17.94	0.09	15.71	20.32	20	20	0	18	22	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
1.04	11011	2.0	Normal	20.32	19.48	0.12	17.85	20.32	18	19	0	18	20	90.00
1.05	00110	0.0	Bernoulli	4.56	4.69	0.04	4.20	5.39	148	147	0	140	152	90.00
1.05	00110	0.5	Bernoulli	5.86	5.88	0.08	4.96	7.20	136	136	1	126	144	90.00
1.05	00110	1.0	Bernoulli	6.91	7.06	0.13	5.62	9.21	128	128	1	114	138	90.00
1.05	00110	1.5	Bernoulli	7.82	8.23	0.19	6.11	11.75	122	121	1	102	134	90.00
1.05	00110	2.0	Bernoulli	8.84	9.45	0.29	6.63	14.35	116	114	1	92	130	90.00
1.05	00110	0.0	Normal	13.85	13.61	0.06	12.23	13.85	24	24	0	24	26	90.00
1.05	00110	0.5	Normal	15.71	15.73	0.02	15.71	15.71	22	22	0	22	22	90.00
1.05	00110	1.0	Normal	17.85	17.87	0.08	15.71	20.32	20	20	0	18	22	90.00
1.05	00110	1.5	Normal	20.32	19.90	0.09	17.85	20.32	18	18	0	18	20	90.00
1.05	00110	2.0	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.25	00101	0.0	Bernoulli	4.56	4.65	0.03	4.20	5.17	148	147	0	142	152	90.00
1.25	00101	0.5	Bernoulli	5.62	5.76	0.06	4.96	6.91	138	137	0	128	144	90.00
1.25	00101	1.0	Bernoulli	6.63	6.77	0.10	5.62	8.49	130	129	1	118	138	90.00
1.25	00101	1.5	Bernoulli	7.51	7.80	0.14	6.11	10.40	124	123	1	108	134	90.00
1.25	00101	2.0	Bernoulli	8.49	9.23	0.45	6.63	12.26	118	116	1	100	130	89.94
1.25	00101	0.0	Normal	13.85	13.85	0.00	13.85	13.85	24	24	0	24	24	90.00
1.25	00101	0.5	Normal	15.71	15.73	0.02	15.71	15.71	22	22	0	22	22	90.00
1.25	00101	1.0	Normal	17.85	18.10	0.07	17.85	20.32	20	20	0	18	20	90.00
1.25	00101	1.5	Normal	20.32	20.10	0.07	17.85	20.32	18	18	0	18	20	90.00
1.25	00101	2.0	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.27	00111	0.0	Bernoulli	5.17	5.18	0.04	4.56	5.86	142	142	0	136	148	90.00
1.27	00111	0.5	Bernoulli	6.63	6.77	0.09	5.62	8.49	130	129	1	118	138	90.00
1.27	00111	1.0	Bernoulli	8.15	8.41	0.15	6.63	11.28	120	119	1	104	130	90.00
1.27	00111	1.5	Bernoulli	9.59	10.14	0.24	7.51	14.35	112	110	1	92	124	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
1.27	00111	2.0	Bernoulli	11.28	12.35	0.57	8.49	18.21	104	102	1	80	118	90.00
1.27	00111	0.0	Normal	13.85	14.04	0.06	13.85	15.71	24	24	0	22	24	90.00
1.27	00111	0.5	Normal	17.85	17.87	0.02	17.85	17.85	20	20	0	20	20	90.00
1.27	00111	1.0	Normal	20.32	20.30	0.02	20.32	20.32	18	18	0	18	18	90.00
1.27	00111	1.5	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.27	00111	2.0	Normal	20.32	20.55	0.08	20.32	23.18	18	18	0	16	18	90.00
1.37	10100	0.0	Bernoulli	6.36	6.52	0.07	5.62	7.82	132	131	0	122	138	90.00
1.37	10100	0.5	Bernoulli	9.59	9.73	0.18	7.51	12.73	112	112	1	98	124	90.00
1.37	10100	1.0	Bernoulli	13.25	13.84	0.41	9.59	20.49	96	95	1	74	112	90.00
1.37	10100	1.5	Bernoulli	17.51	20.48	0.95	12.23	38.15	82	78	2	42	100	89.94
1.37	10100	2.0	Bernoulli	61.86	40.29	2.14	14.94	61.86	16	47	3	16	90	90.00
1.37	10100	0.0	Normal	15.71	16.63	0.11	15.71	17.85	22	21	0	20	22	90.00
1.37	10100	0.5	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.37	10100	1.0	Normal	20.32	20.49	0.07	20.32	23.18	18	18	0	16	18	90.00
1.37	10100	1.5	Normal	23.18	22.49	0.12	20.32	23.18	16	16	0	16	18	90.00
1.37	10100	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.39	10110	0.0	Bernoulli	7.20	7.43	0.09	6.36	9.21	126	125	1	114	132	90.00
1.39	10110	0.5	Bernoulli	11.75	12.15	0.28	9.21	16.83	102	101	1	84	114	90.00
1.39	10110	1.0	Bernoulli	17.51	19.36	0.72	12.23	32.58	82	79	2	50	100	90.00
1.39	10110	1.5	Bernoulli	27.95	34.25	1.64	16.18	61.86	58	52	2	16	86	89.94
1.39	10110	2.0	Bernoulli	61.86	59.03	1.04	22.16	61.86	16	20	1	16	70	90.00
1.39	10110	0.0	Normal	17.85	17.90	0.03	17.85	17.85	20	20	0	20	20	90.00
1.39	10110	0.5	Normal	20.32	20.35	0.03	20.32	20.32	18	18	0	18	18	90.00
1.39	10110	1.0	Normal	23.18	21.75	0.14	20.32	23.18	16	17	0	16	18	90.00
1.39	10110	1.5	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
1.39	10110	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.43	01100	0.0	Bernoulli	5.86	6.02	0.05	5.39	6.91	136	135	0	128	140	90.00
1.43	01100	0.5	Bernoulli	8.49	8.56	0.13	6.91	10.83	118	118	1	106	128	90.00
1.43	01100	1.0	Bernoulli	10.83	11.48	0.26	8.49	16.18	106	104	1	86	118	90.00
1.43	01100	1.5	Bernoulli	14.35	15.32	0.60	10.40	23.95	92	91	1	66	108	90.00
1.43	01100	2.0	Bernoulli	18.21	25.56	1.79	12.23	61.86	80	71	3	16	100	90.00
1.43	01100	0.0	Normal	15.71	15.88	0.06	15.71	17.85	22	22	0	20	22	90.00
1.43	01100	0.5	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.43	01100	1.0	Normal	20.32	20.35	0.03	20.32	20.32	18	18	0	18	18	90.00
1.43	01100	1.5	Normal	20.32	21.09	0.13	20.32	23.18	18	17	0	16	18	90.00
1.43	01100	2.0	Normal	23.18	22.84	0.09	20.32	23.18	16	16	0	16	18	90.00
1.46	01110	0.0	Bernoulli	6.63	6.78	0.07	5.86	8.15	130	129	0	120	136	90.00
1.46	01110	0.5	Bernoulli	9.99	10.37	0.19	8.15	13.79	110	109	1	94	120	90.00
1.46	01110	1.0	Bernoulli	14.35	15.21	0.49	10.40	23.04	92	91	1	68	108	90.00
1.46	01110	1.5	Bernoulli	20.49	24.29	1.35	13.25	61.86	74	70	2	16	96	90.00
1.46	01110	2.0	Bernoulli	61.86	52.39	1.79	16.83	61.86	16	30	3	16	84	90.00
1.46	01110	0.0	Normal	17.85	17.87	0.02	17.85	17.85	20	20	0	20	20	90.00
1.46	01110	0.5	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.46	01110	1.0	Normal	20.32	20.66	0.09	20.32	23.18	18	18	0	16	18	90.00
1.46	01110	1.5	Normal	23.18	22.92	0.08	20.32	23.18	16	16	0	16	18	90.00
1.46	01110	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.59	10101	0.0	Bernoulli	6.91	6.85	0.07	6.10	8.15	128	129	0	120	134	90.00
1.59	10101	0.5	Bernoulli	10.40	10.57	0.18	8.49	13.79	108	108	1	94	118	90.00
1.59	10101	1.0	Bernoulli	14.94	16.30	0.66	10.83	23.95	90	88	2	66	106	89.87
1.59	10101	1.5	Bernoulli	21.31	25.40	1.37	14.35	61.86	72	68	2	16	92	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
1.59	10101	2.0	Bernoulli	61.86	56.68	1.41	18.21	61.86	16	24	2	16	80	90.00
1.59	10101	0.0	Normal	17.85	17.87	0.02	17.85	17.85	20	20	0	20	20	90.00
1.59	10101	0.5	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.59	10101	1.0	Normal	20.32	21.09	0.13	20.32	23.18	18	17	0	16	18	90.00
1.59	10101	1.5	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.59	10101	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.61	10111	0.0	Bernoulli	7.51	7.72	0.09	6.63	9.21	124	123	1	114	130	90.00
1.61	10111	0.5	Bernoulli	12.73	12.96	0.27	9.99	17.54	98	98	1	82	110	90.00
1.61	10111	1.0	Bernoulli	19.70	21.70	0.80	13.79	36.58	76	73	2	44	94	89.94
1.61	10111	1.5	Bernoulli	36.51	42.73	1.78	18.94	61.86	44	41	2	16	78	90.00
1.61	10111	2.0	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00
1.61	10111	0.0	Normal	20.32	19.06	0.12	17.85	20.32	18	19	0	18	20	90.00
1.61	10111	0.5	Normal	20.32	20.35	0.03	20.32	20.32	18	18	0	18	18	90.00
1.61	10111	1.0	Normal	23.18	22.78	0.10	20.32	23.18	16	16	0	16	18	90.00
1.61	10111	1.5	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.61	10111	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.65	01101	0.0	Bernoulli	6.36	6.37	0.05	5.62	7.20	132	132	0	126	138	90.00
1.65	01101	0.5	Bernoulli	9.21	9.34	0.14	7.51	11.75	114	114	1	102	124	90.00
1.65	01101	1.0	Bernoulli	12.73	12.98	0.28	9.59	18.21	98	98	1	80	112	90.00
1.65	01101	1.5	Bernoulli	16.83	18.17	0.65	12.23	29.05	84	82	1	56	100	90.00
1.65	01101	2.0	Bernoulli	61.86	47.11	2.11	14.94	61.86	16	38	3	16	90	90.00
1.65	01101	0.0	Normal	17.85	17.85	0.00	17.85	17.85	20	20	0	20	20	90.00
1.65	01101	0.5	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.65	01101	1.0	Normal	20.32	20.58	0.08	20.32	23.18	18	18	0	16	18	90.00
1.65	01101	1.5	Normal	23.18	22.84	0.09	20.32	23.18	16	16	0	16	18	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
1.65	01101	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.67	01111	0.0	Bernoulli	6.91	7.09	0.07	6.11	8.49	128	127	0	118	134	90.00
1.67	01111	0.5	Bernoulli	10.83	11.18	0.19	8.84	14.94	106	105	1	90	116	90.00
1.67	01111	1.0	Bernoulli	16.18	17.61	0.71	11.75	25.93	86	84	2	62	102	89.87
1.67	01111	1.5	Bernoulli	23.95	29.53	1.54	15.55	61.86	66	60	2	16	88	90.00
1.67	01111	2.0	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00
1.67	01111	0.0	Normal	17.85	18.15	0.08	17.85	20.32	20	20	0	18	20	90.00
1.67	01111	0.5	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.67	01111	1.0	Normal	23.18	21.92	0.14	20.32	23.18	16	17	0	16	18	90.00
1.67	01111	1.5	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.67	01111	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.77	11100	0.0	Bernoulli	8.49	8.79	0.11	7.51	10.83	118	117	1	106	124	90.00
1.77	11100	0.5	Bernoulli	15.55	16.48	0.41	12.23	23.95	88	86	1	66	100	90.00
1.77	11100	1.0	Bernoulli	29.05	32.41	1.26	18.90	61.86	56	53	2	16	78	90.00
1.77	11100	1.5	Bernoulli	61.86	60.90	0.55	61.86	61.86	16	17	1	16	16	90.00
1.77	11100	2.0	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00
1.77	11100	0.0	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.77	11100	0.5	Normal	20.32	20.75	0.10	20.32	23.18	18	18	0	16	18	90.00
1.77	11100	1.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.77	11100	1.5	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.77	11100	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.80	11110	0.0	Bernoulli	9.99	10.07	0.14	8.15	12.73	110	110	1	98	120	90.00
1.80	11110	0.5	Bernoulli	19.70	21.28	0.64	14.35	32.58	76	74	1	50	92	90.00
1.80	11110	1.0	Bernoulli	49.33	47.51	1.34	24.90	61.86	28	32	2	16	64	90.00
1.80	11110	1.5	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
1.80	11110	2.0	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00
1.80	11110	0.0	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.80	11110	0.5	Normal	23.18	22.41	0.13	20.32	23.18	16	17	0	16	18	90.00
1.80	11110	1.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.80	11110	1.5	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.80	11110	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.99	11101	0.0	Bernoulli	8.84	8.94	0.10	7.51	10.83	116	116	1	106	124	90.00
1.99	11101	0.5	Bernoulli	16.18	18.06	0.67	12.73	24.95	86	82	2	64	98	89.81
1.99	11101	1.0	Bernoulli	31.36	35.06	1.26	20.49	61.86	52	49	2	16	74	90.00
1.99	11101	1.5	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00
1.99	11101	2.0	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00
1.99	11101	0.0	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
1.99	11101	0.5	Normal	23.18	21.66	0.14	20.32	23.18	16	17	0	16	18	90.00
1.99	11101	1.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.99	11101	1.5	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
1.99	11101	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
2.02	11111	0.0	Bernoulli	9.99	10.09	0.13	8.49	12.23	110	110	1	100	118	90.00
2.02	11111	0.5	Bernoulli	20.49	21.89	0.67	14.94	32.58	74	72	1	50	90	89.94
2.02	11111	1.0	Bernoulli	59.54	50.61	1.27	27.95	61.86	18	29	2	16	58	90.00
2.02	11111	1.5	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00
2.02	11111	2.0	Bernoulli	61.86	61.86	0.00	61.86	61.86	16	16	0	16	16	90.00
2.02	11111	0.0	Normal	20.32	20.32	0.00	20.32	20.32	18	18	0	18	18	90.00
2.02	11111	0.5	Normal	23.18	22.98	0.07	20.32	23.18	16	16	0	16	18	90.00
2.02	11111	1.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00
2.02	11111	1.5	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00

Continued on next page

Table C.2 – continued from previous page

Population		q	Prior	α (in %)					Trial Size (2n)					Average Power
L_1/L_0	Group			NoBS	Mean	se	5%	95%	NoBS	Mean	se	5%	95%	(in %)
2.02	11111	2.0	Normal	23.18	23.18	0.00	23.18	23.18	16	16	0	16	16	90.00

C.2.2 Application to Isakov et al. (2019)

In this section, we apply the extended BDA framework to [126] and compare the results under a Bernoulli prior and a Gaussian prior In [126], the formulation also accounts for in-trial costs, which we discuss in detail in Appendix C.1.3.

Background

The post-trial loss (given a treatment effect θ and an observation T_n of the trial's treatment effect) is defined as

$$L(\theta, T_n) \quad = \quad \mathbb{1}\{T_n > \lambda_n | \theta\} \cdot (c_{\theta,app} - c_{\theta,rej}) + c_{\theta,rej}, \tag{C.77}$$

and models the loss per person incurred by all patients affected by the disease.

The in-trial loss $L^{in}(\theta)$ models the loss per person incurred only by the patients enrolled in the clinical trial and is independent of T_n because the approval or non-approval of the treatment has no impact on the patient's loss during the clinical trial period.

Defining N as the prevalence of the disease and n as the number of patients enrolled in each arm of the trial, the total loss function becomes

$$\text{Loss} = L(\theta, T_n) + \frac{n}{N} \cdot L^{in}(\theta). \tag{C.78}$$

In the setting described in [126], the cost of rejecting and approving a treatment with efficacy θ are given respectively by

$$\begin{cases} c_{\theta,rej} = L_1 \cdot \max\left(0, \min\left(1, \frac{\theta}{\mu_\theta}\right)\right), \\ c_{\theta,app} = L_0 \cdot \max\left(0, \min\left(1, 1 - \frac{\theta}{\mu_\theta}\right)\right), \end{cases} \tag{C.79}$$

where the costs L_0 and L_1 are denoted respectively $N \cdot c_1$ and $N \cdot c_2$ in [126], and the treatment effect μ_θ is denoted $\delta_0 = 2^{-3} \cdot \sigma_\delta$. Equation C.79 immediately follows from Equations 4.36, 4.37 and 4.39 by setting $r = 0$.

Notice that, in [126], the costs are assumed to be

$$\begin{cases} L_0 = N \cdot s_1, \\ L_1 = N \cdot \min\left(1, \frac{\mu_\theta}{\sigma_\theta}\right) \cdot s_2, \end{cases} \tag{C.80}$$

where $s_1 = 0.07$ is the current cost of adverse effects of medical treatment per patient (obtained from [177]), and s_2 is the severity of the disease, defined as a function of the years lived with disability YLD attributed to the disease considered, the number of deaths D, and the age-standardized prevalence rate $\tilde{N}$ of the disease:

$$s_2 = \frac{D + YLD}{D + \tilde{N}}. \tag{C.81}$$

To be able to compare different diseases together, we follow the methodology used in [126, 6] and use age-standardized rates per 100,000 individuals for D, YLD, and $\tilde{N}$. Estimates of D, YLD, and $\tilde{N}$ are obtained from the web appendix of [177], in particular Table 1 for $\tilde{N}$, Table 3 for D, and Table 5 for YLD.

The in-trial costs at $\theta = 0$ and $\theta = \mu_\theta$ are given by

$$L^{in}(0) = L_0, \tag{C.82}$$
$$L^{in}(\mu_\theta) = \gamma \cdot N \cdot L_1, \tag{C.83}$$

for some constant $\gamma = 4 \cdot 10^{-3} \cdot \frac{\mu_\theta}{\sigma_\theta}$ which represents the incremental cost incurred per extra patient added to each arm. Using a linear interpolation yields

$$L^{in}(\theta) = L_0 \cdot \max\left(0, \min\left(1, 1 - \frac{\theta}{\mu_\theta}\right)\right) + (\gamma \cdot N \cdot L_1) \cdot \max\left(0, \min\left(1, \frac{\theta}{\mu_\theta}\right)\right). \tag{C.84}$$

Standard Error Estimation

As the trial losses are not estimated from a DCE, we do not have the usual regression formulation with estimates of the standard deviation of the L_1. Notice that L_1 depends on s_2, and we can estimate the standard deviation of D, YLD, and $\tilde{N}$ by using the 95%-confidence intervals provided in the appendix in [177]. For simplicity, we approximate the standard deviation of D using its mean $\hat{D}$, its 5^{th}-percentile value D_5 and its 95^{th}-percentile value D_{95} as:

$$\sigma_D \approx \frac{1}{1.96} \cdot \frac{D_{95} - D_5}{2}. \tag{C.85}$$

Standard deviations of YLD and $\tilde{N}$ are calculated similarly.

We then use the delta method to calculate the distribution of s_2. We define

$$\hat{B} = \begin{bmatrix} \hat{D} \\ \hat{YLD} \\ \hat{\tilde{N}} \end{bmatrix}, \quad \text{and} \quad B_0 = \begin{bmatrix} D_0 \\ YLD_0 \\ \tilde{N}_0 \end{bmatrix}, \tag{C.86}$$

and the function

$$s(B) = \frac{D + YLD}{D + \tilde{N}}. \tag{C.87}$$

Then assuming that

$$\sqrt{k} \cdot \left[\hat{B} - B_0\right] \xrightarrow{D} N(0, \Sigma), \tag{C.88}$$

and assuming a diagonal covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_D^2 & 0 & 0 \\ 0 & \sigma_{YLD}^2 & 0 \\ 0 & 0 & \sigma_{\tilde{N}}^2 \end{bmatrix}, \tag{C.89}$$

we obtain

$$\sqrt{k} \cdot \left[s(\hat{B}) - s(B_0) \right] \xrightarrow{D} N\left(0, \nabla s(B_0)^T \cdot \Sigma \cdot \nabla s(B_0)\right). \tag{C.90}$$

The gradient term is equal to

$$\nabla s(B_0) = \frac{1}{(D_0 + \tilde{N}_0)^2} \cdot \begin{bmatrix} \tilde{N}_0 - YLD_0 \\ D_0 + \tilde{N}_0 \\ -(D_0 + YLD_0) \end{bmatrix}, \tag{C.91}$$

therefore the asymptotic variance simplifies to

$$\sigma^2_{s_2} = \frac{1}{(D_0 + \tilde{N}_0)^4} \cdot \left[\sigma^2_D \cdot (\tilde{N}_0 - YLD_0)^2 + \sigma^2_{YLD} \cdot (D_0 + \tilde{N}_0)^2 + \sigma^2_{\tilde{N}} \cdot (D_0 + YLD_0)^2 \right]. \tag{C.92}$$

Hence, we assume here that

$$\hat{s}_2 \sim N\left(\hat{s}_2, \sigma^2_{s_2}\right), \tag{C.93}$$

and generate 100 bootstrap samples of $\hat{s}_2$. This allows us to run the BDA optimization procedure (Equation 4.11) for each bootstrapped value $\hat{s}_2$ and obtain a distribution for the optimal BDA parameters (n^*, λ_n^*). We then calculate the corresponding standard errors on our estimates of the optimal sample size n^* and the optimal critical value λ_n^* (or equivalently α^*).

Results

Standard errors. The following plots show the sensitivity of the BDA output's standard errors.

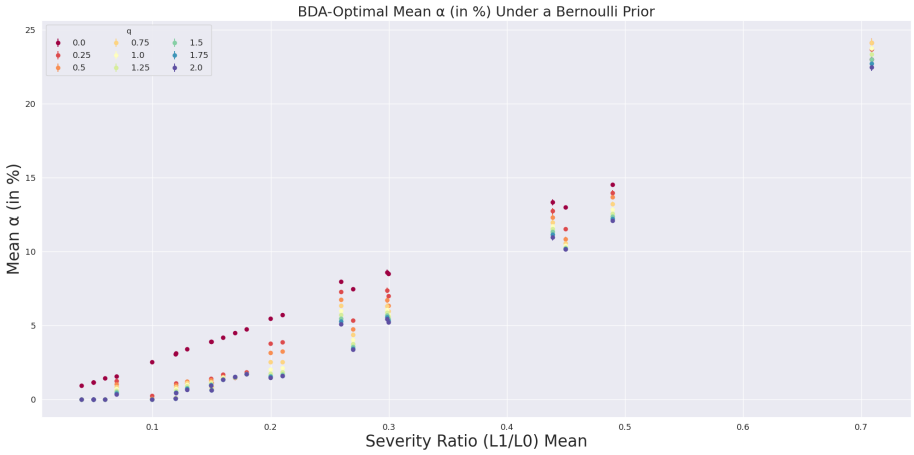


Figure C-17: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Bernoulli prior. Standard errors are calculated based on 100 bootstrap samples.

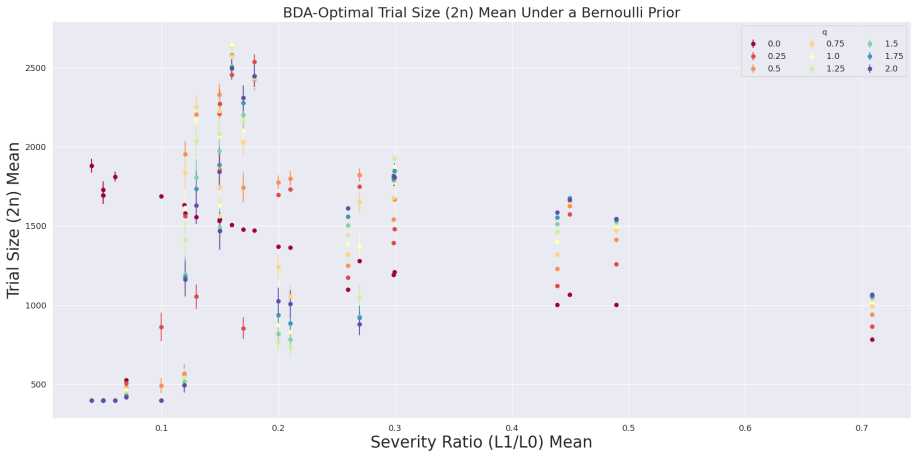


Figure C-18: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Bernoulli prior. Standard errors are calculated based on 100 bootstrap samples.

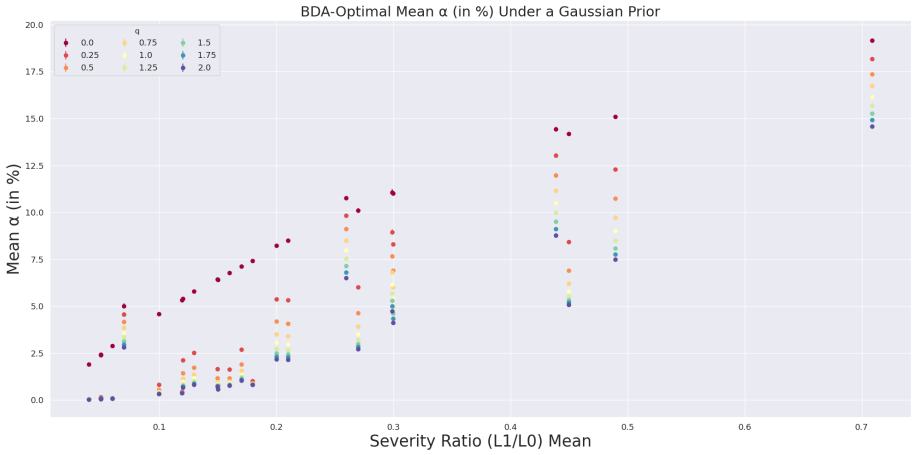


Figure C-19: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Gaussian prior. Standard errors are calculated based on 100 bootstrap samples.

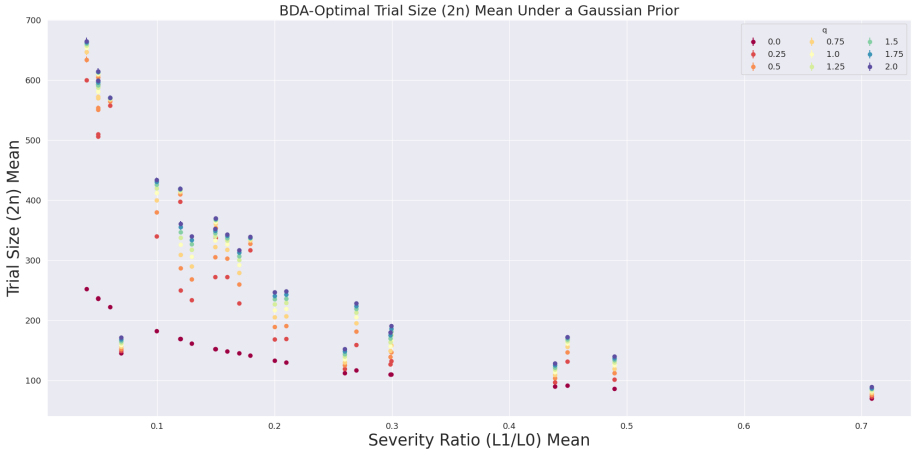


Figure C-20: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Gaussian prior. Standard errors are calculated based on 100 bootstrap samples.

Confidence intervals. The following plots show the sensitivity of the BDA output's 95%-confidence intervals.

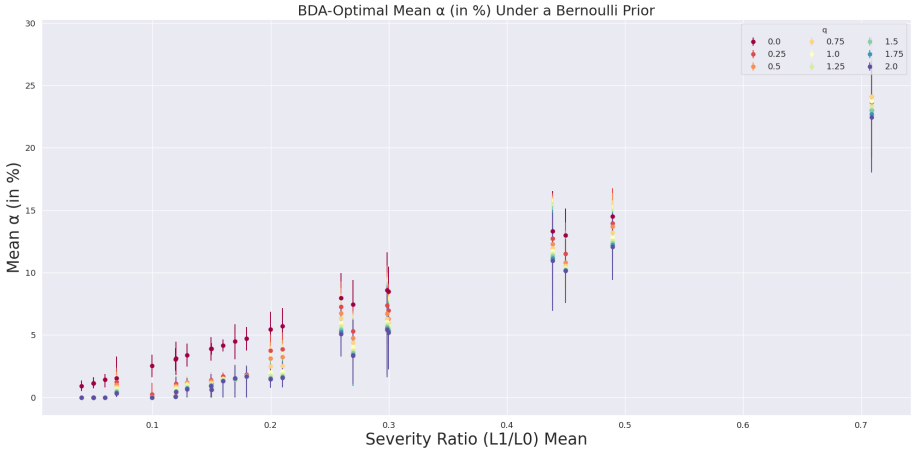


Figure C-21: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Bernoulli prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

Figure C-22: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Bernoulli prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

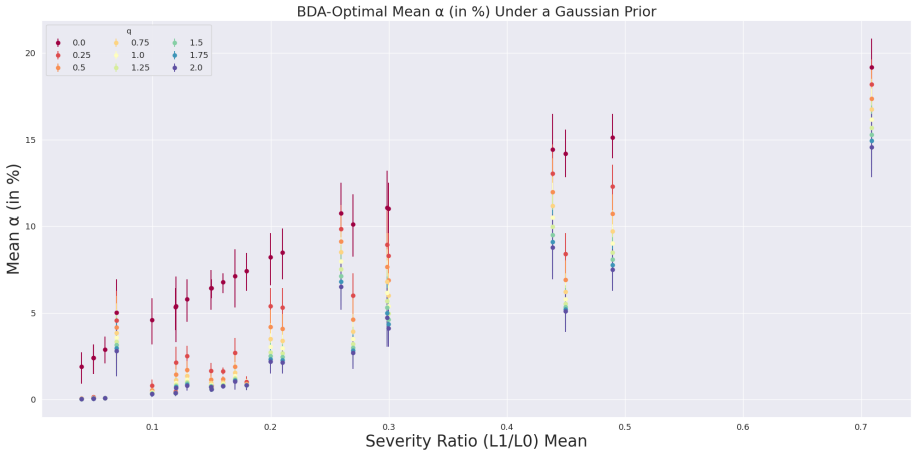


Figure C-23: BDA-Optimal significance level α (in %) as a function of the severity ratio, under a Gaussian prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

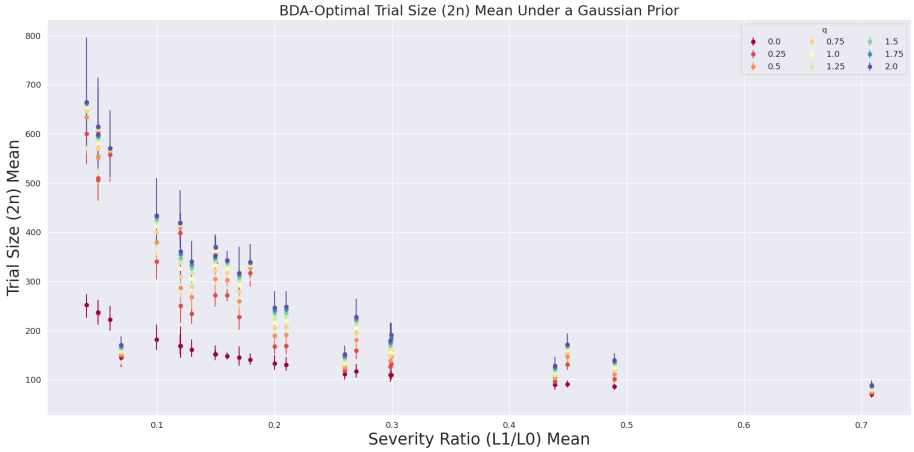


Figure C-24: BDA-Optimal trial size ($2n$) as a function of the severity ratio, under a Gaussian prior. The 95%-confidence intervals are calculated based on 100 bootstrap samples.

Appendix D

Chapter 5 Supplementary Material

D.1 Dynamics of the Augmented DELPHI Model

We review in D.1.1 the additions made to the DELPHI model introduced in [152] to include vaccination states. We then describe in D.1.2 our discretization technique to maintain the performance of the model while greatly reducing the running time of the simulation.

D.1.1 Dynamics of the DELPHI Model

The Original DELPHI Model

The DELPHI model (represented by the green nodes in Fig D-1) is composed of the following components [152]: the susceptible (not yet infected) population (S); exposed individuals that have been infected, are not contagious, and are within the incubation period (E); infected individuals that are currently contagious (I); infected individuals that self-quarantine at home but were not tested (UD and UR), where UR corresponds to the individuals that recover, while UD corresponds to individuals that perish; infected individuals that were detected and hospitalized (DHR and DHD), where DHR corresponds to the individuals that recover, while DHD corresponds to individuals that perish; infected individuals that were detected and quarantined at home (DQR and DQD), where DQR corresponds to the individuals that recover, while DQD corresponds to individuals that perish; individuals that recover from the disease and have permanent immunity (R); and individuals that perish from the disease (D).

The following helpful definitions can also be found in [152]: total number of hospitalized cases (TH); total number of detected deaths (DD); total number of detected cases (DT).

The dynamics of the original DELPHI model are outlined in Eq (D.1) to Eq (D.14) [152]:

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t) \;=\; -\alpha \cdot \gamma(t) \cdot S(t) \cdot I(t) \tag{D.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}E(t) = \alpha \cdot \gamma(t) \cdot S(t) \cdot I(t) - r_i \cdot E(t) \tag{D.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I(t) = r_i \cdot E(t) - r_d \cdot I(t) \tag{D.3}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}UR(t) = r_d \cdot (1 - p_{dth}(t)) \cdot (1 - p_d) \cdot I(t) - r_{ri} \cdot UR(t) \tag{D.4}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DHR(t) = r_d \cdot (1 - p_{dth}(t)) \cdot p_d \cdot p_h \cdot I(t) - r_{rh} \cdot DHR(t) \tag{D.5}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DQR(t) = r_d \cdot (1 - p_{dth}(t)) \cdot p_d \cdot (1 - p_h) \cdot I(t) - r_{ri} \cdot DQR(t) \tag{D.6}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}UD(t) = r_d \cdot p_{dth}(t) \cdot (1 - p_d) \cdot I(t) - r_{dth} \cdot UD(t) \tag{D.7}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DHD(t) = r_d \cdot p_{dth}(t) \cdot p_d \cdot p_h \cdot I(t) - r_{dth} \cdot DHD(t) \tag{D.8}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DQD(t) = r_d \cdot p_{dth}(t) \cdot p_d \cdot (1 - p_h) \cdot I(t) - r_{dth} \cdot DQD(t) \tag{D.9}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}TH(t) = r_d \cdot p_d \cdot p_h \cdot I(t) \tag{D.10}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DD(t) = r_{dth} \cdot (DHD(t) + DQD(t)) \tag{D.11}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DT(t) = r_d \cdot p_d \cdot I(t) \tag{D.12}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}R(t) = r_{ri} \cdot (UR(t) + DQR(t)) + r_{rh} \cdot DHR(t) \tag{D.13}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}D(t) = r_{dth} \cdot (UD(t) + DQD(t) + DHD(t)) \tag{D.14}$$

Model Calibration To calibrate the DELPHI model, we need to estimate the following quantities [152]:

- α: Infection rate, assumed to be constant across all countries.
- $\gamma(t)$: Government response, modelled as

$$\gamma(t) = \frac{2}{\pi} \cdot \arctan\left(-\frac{b \cdot (t-a)}{20}\right) + 1 + j_0 \cdot \exp\left(-\frac{(t - t_{jump})^2}{2 \cdot \sigma^2}\right), \tag{D.15}$$

  where a controls the time at which the measure starts; b controls the strength of the measure; j_0 controls the magnitude of the jump; t_{jump} is the median day when the jump occurs; and σ is the rate at which the resurgence in the cases occurred.
- r_i: Rate of infection leaving the incubation phase. Assumption: $r_i = \frac{\log 2}{T_i}$, where $T_i = 5$ days (the median time to leave incubation).
- r_d: Rate of detection. Assumption: $r_d = \frac{\log 2}{T_d}$, where $T_d = 2$ days (the median time to detection).

- r_{ri}: Rate of recovery not under hospitalization. Assumption: $r_{ri} = \frac{\log 2}{T_{ri}}$, where $T_{ri} = 10$ days (the median time to recovery not under hospitalization).

- r_{rh}: Rate of recovery under hospitalization. Assumption: $r_{rh} = \frac{\log 2}{T_{rh}}$, where $T_{rh} = 15$ days (the median time to recovery under hospitalization).

- r_{dth}: Rate of death. Assumption: $r_{dth} = \frac{\log 2}{T_{dth}}$, where T_{dth} is the time till death for dying patients.

- $p_{dth}(t)$: Mortality percentage over time, modelled as the following declining function (reflecting an improved ability to detect milder cases and increased standards of care for COVID-19 patients)

$$p_{dth}(t) = (p_{dth0} - \underline{p_{dth}}) \cdot \left[\frac{2}{\pi} \cdot \arctan\left(-\frac{t}{20} \cdot r_{ddec}\right) + 1\right] + \underline{p_{dth}}, \tag{D.16}$$

  where p_{dth0} is the initial mortality percentage; $\underline{p_{dth}}$ is a lower bound on mortality percentage assuming perfect detection and perfect treatment; r_{ddec} is the rate of decay of mortality percentage.

- p_d: Percentage of detected infection cases. Assumption: $p_d = 0.2$ is constant.

- p_h: Percentage of hospitalized detected infection cases. Assumption: $p_h = 0.03$ is constant.

Parameters to Fit The only parameters we need to fit to historical data are:

1. α
2. $r_{dth} = \frac{\log 2}{T_{dth}}$
3. p_{dth0}
4. r_{ddec}
5. a
6. b
7. j_0
8. t_{jump}
9. σ
10. k_1
11. k_2

where k_1 and k_2 are "internal" parameters used for initial conditions.

Inclusion of Vaccination States

In addition to the DELPHI model (represented by green nodes in Fig D-1), we add the following components to account for vaccination states:

- $X \in \{A, B\}$: Brand of the vaccine distributed.

- $V_{X,1}$: Individuals receiving the first dose for vaccine X.

- $V^r_{X,1}$ (immediate response): Individuals who respond to the first dose of vaccine X. We assume they have permanent immunity to the disease after a period of $T_{X,I}$ days after receiving the first dose.

- $V^{dr}_{X,1}$ (delayed response): Individuals who do not respond to the first dose of vaccine X but who respond to the second dose of vaccine X. We assume they have permanent immunity to the disease after a period of $T_{X,I}$ days after receiving the second dose.

- $V^{nr}_{X,1}$ (no response): Individuals who do not respond to the first and second doses of vaccine X.

We also need the following helpful definitions:

- $V_{X,a}(t)$: Number of vaccines X available at time t.

- $\varepsilon_{X,i}$: Effectiveness of i^{th} dose of vaccine X [96].

  - Pfizer-BioNTech: 52%efficacy after a single dose, 92%efficacy after two doses.
  - Moderna: 80.20%efficacy after a single dose, 95.60%efficacy after two doses.

- $V^S(t)$: Susceptible vaccinated individuals. We assume:

$$V^S(t) = \sum_{X \in \{A,B\}} V^r_{X,1}(t) + V^{dr}_{X,1}(t) + V^{nr}_{X,1}(t) + V^{nr}_{X,2}(t). \tag{D.17}$$

- $V^r_{X,2}$: Individuals who respond to the first and/or second doses of vaccine X, and who were not infected in the first $T_{X,I}$ days following their successful vaccination date.

- $V^{nr}_{X,2}$: Individuals who have still not responded to the second dose of vaccine X after $T_{X,I}$ days following their second vaccination date.

- $V_{X,1r\to 2r}$: individuals that were in the $V^r_{X,1}$ state and become $V^r_{X,2}$ when they receive their second dose.

- $V_{X,1dr\to 2r}$: individuals that were in the $V^{dr}_{X,1}$ state and become $V^r_{X,2}$ when they receive their second dose.

- $V_{X,1nr\to 2nr}$: individuals that were in the $V^{nr}_{X,1}$ state and become $V^{nr}_{X,2}$ when they receive their second dose.

Assumptions In addition, we make the following assumptions:

1. The immune response to a vaccine does not decay over time.
2. All vaccinated individuals receive two doses.
3. Individuals in the $V^r_{X,1}$ group are still susceptible to an infection in the first $T_{X,I}$ days of their first vaccination. Starting on day $T_{X,I}+1$, they have permanent immunity to the disease and join the $V^r_{X,2}$ group.
4. Individuals in the $V^{dr}_{X,1}$ group are still susceptible to an infection until day $T_{X,I}$ following their second vaccination. Starting on day $T_{X,I}+1$ since they receive the second dose, they have permanent immunity to the disease.
5. Individuals in the $V^{nr}_{X,2}$ group will never respond positively to the vaccine, and remain susceptible to an infection.
6. We assume $T_{X,I} = 14$ days [192].
7. We assume a uniform daily infection rate among individuals in each vaccination state.

Dynamics of the DELPHI Model with Vaccination States

We augment the dynamics of the DELPHI model with vaccination states, the blue terms appearing in Eq (D.18) to Eq (D.36):

$$\frac{\mathrm{d}}{\mathrm{d}t}S(t) = -\alpha\cdot\gamma(t)\cdot S(t)\cdot I(t) - \sum_{X\in\{A,B\}} V_{X,1}(t) \tag{D.18}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}E(t) = \alpha\cdot\gamma(t)\cdot(S(t)+V^S(t))\cdot I(t) - r_i\cdot E(t) \tag{D.19}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V^r_{X,1}(t) = \varepsilon_{X,1}\cdot V_{X,1}(t) - \alpha\cdot\gamma(t)\cdot V^r_{X,1}\cdot I(t) - V_{X,1r\to 2r} \tag{D.20}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V^{dr}_{X,1}(t) = (\varepsilon_{X,2}-\varepsilon_{X,1})\cdot V_{X,1}(t) - \alpha\cdot\gamma(t)\cdot V^{dr}_{X,1}\cdot I(t) - V_{X,1dr\to 2r} \tag{D.21}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V^{nr}_{X,1}(t) = (1-\varepsilon_{X,2})\cdot V_{X,1}(t) - \alpha\cdot\gamma(t)\cdot V^{nr}_{X,1}\cdot I(t) - V_{X,1nr\to 2nr} \tag{D.22}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V^r_{X,2}(t) = V_{X,1r\to 2r} + V_{X,1nr\to 2r} \tag{D.23}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}V^{nr}_{X,2}(t) = V_{X,1nr\to 2nr} - \alpha\cdot\gamma(t)\cdot V^{nr}_{X,2}\cdot I(t) \tag{D.24}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}I(t) = r_i\cdot E(t) - r_d\cdot I(t) \tag{D.25}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}UR(t) = r_d\cdot(1-p_{dth}(t))\cdot(1-p_d)\cdot I(t) - r_{ri}\cdot UR(t) \tag{D.26}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DHR(t) = r_d\cdot(1-p_{dth}(t))\cdot p_d\cdot p_h\cdot I(t) - r_{rh}\cdot DHR(t) \tag{D.27}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DQR(t) = r_d \cdot (1 - p_{dth}(t)) \cdot p_d \cdot (1 - p_h) \cdot I(t) - r_{ri} \cdot DQR(t) \quad \text{(D.28)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}UD(t) = r_d \cdot p_{dth}(t) \cdot (1 - p_d) \cdot I(t) - r_{dth} \cdot UD(t) \quad \text{(D.29)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DHD(t) = r_d \cdot p_{dth}(t) \cdot p_d \cdot p_h \cdot I(t) - r_{dth} \cdot DHD(t) \quad \text{(D.30)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DQD(t) = r_d \cdot p_{dth}(t) \cdot p_d \cdot (1 - p_h) \cdot I(t) - r_{dth} \cdot DQD(t) \quad \text{(D.31)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}TH(t) = r_d \cdot p_d \cdot p_h \cdot I(t) \quad \text{(D.32)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DD(t) = r_{dth} \cdot (DHD(t) + DQD(t)) \quad \text{(D.33)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}DT(t) = r_d \cdot p_d \cdot I(t) \quad \text{(D.34)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}R(t) = r_{ri} \cdot (UR(t) + DQR(t)) + r_{rh} \cdot DHR(t) \quad \text{(D.35)}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}D(t) = r_{dth} \cdot (UD(t) + DQD(t) + DHD(t)) \quad \text{(D.36)}$$

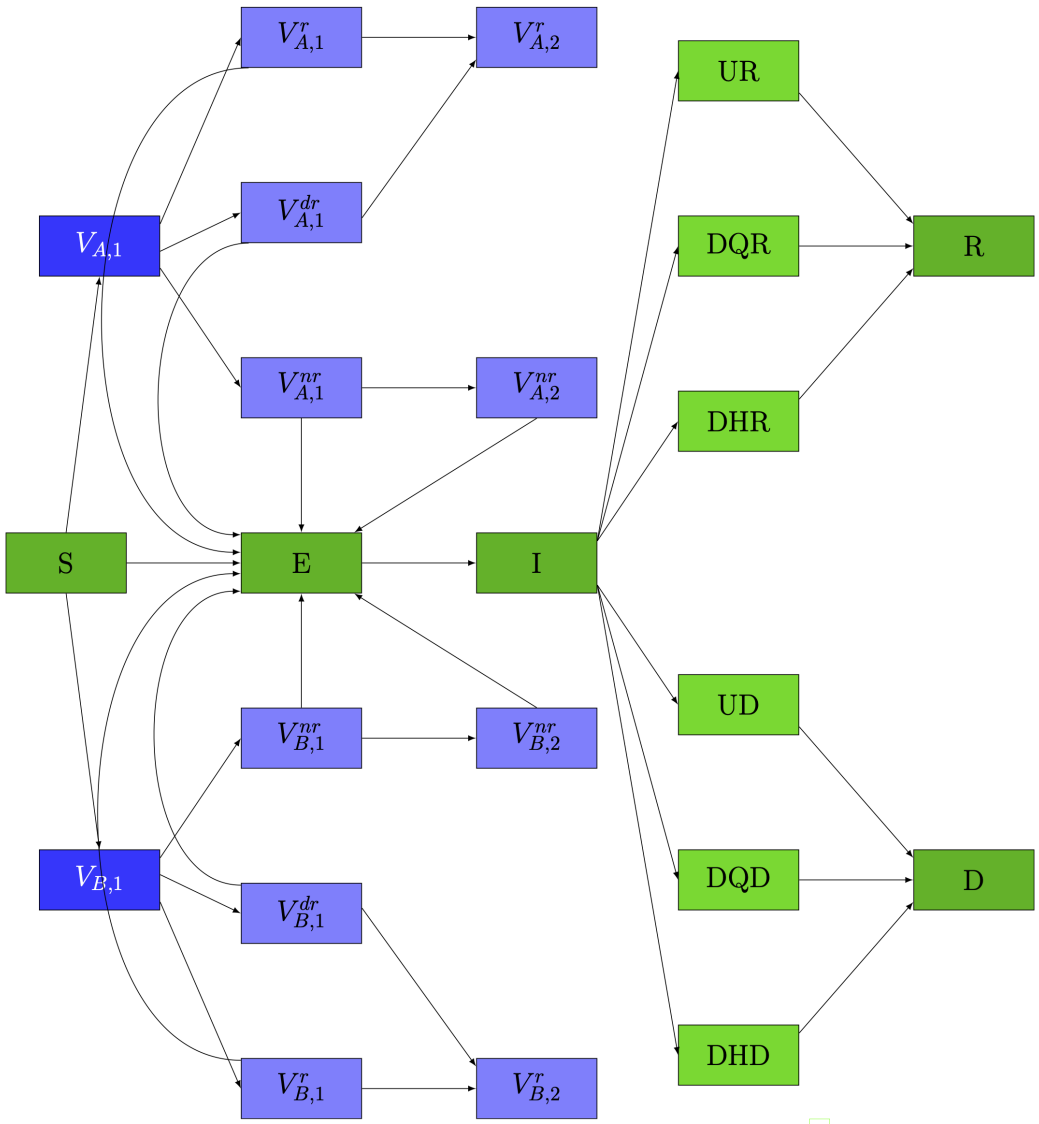


Figure D-1: Flowchart of the original DELPHI model (in green) [152] and the additional vaccination states (in blue). For illustrative purposes, brand A denotes the Moderna vaccine and brand B denotes the Pfizer-BioNTech vaccine.

D.1.2 Discretization of the DELPHI Model

We simulate the DELPHI model using a time step of 0.01 days. Although these simulations provide a very accurate solution to the DELPHI's system of ODEs, it is not a very practical approach. In fact, running a single simulation for the United States takes a few minutes, which is not an ideal setting for Monte Carlo simulation. To resolve this issue, we recalibrate our parameters and ensure that a discretized version of the DELPHI that uses these parameters and a time step of 1 day will yield the same output as the original non-discretized model. We plot in Fig D-2 a comparison of DELPHI outputs using continuous time steps and discrete time steps of 1 day. These outputs make us confident that using a discretized DELPHI will not affect our results.

We have observed that the parameters $\gamma(t)$, p_h, p_d, and p_{dth} can be left unchanged; however, it was crucial to re-estimate the parameters α, r_i, r_d, r_{ri}, r_{rh}, and r_{dth}. This can be done by noticing that the DELPHI equations yield the following properties:

$$\tilde{\alpha} = -\frac{1}{\gamma(t)} \cdot \frac{S(t+1) - S(t)}{S(t) \cdot I(t)} \tag{D.37}$$

$$\tilde{r_i} = -\frac{[S(t+1) + E(t+1)] - [S(t) + E(t)]}{E(t)} \tag{D.38}$$

$$\tilde{r_d} = -\frac{[S(t+1) + E(t+1) + I(t+1)] - [S(t) + E(t) + I(t)]}{I(t)} \tag{D.39}$$

$$\tilde{r_{dth}} = \frac{D(t+1) - D(t)}{UD(t) + DQD(t) + DHD(t)} = \frac{DD(t+1) - DD(t)}{DQD(t) + DHD(t)} \tag{D.40}$$

Furthermore, we notice that the discretized versions of r_i and r_d are related to their non-discretized versions through the same proportionality coefficient. We use this proportionality coefficient to estimate the discretized version of r_{rh} and r_{ri} from their non-discretized values. In other words,

$$\tilde{r_{ri}} = \frac{\tilde{r_d}}{r_d} \cdot r_{ri} = \frac{\tilde{r_i}}{r_i} \cdot r_{ri} \tag{D.41}$$

$$\tilde{r_{rh}} = \frac{\tilde{r_d}}{r_d} \cdot r_{rh} = \frac{\tilde{r_i}}{r_i} \cdot r_{rh} \tag{D.42}$$

To verify that p_h, p_d, and p_{dth} are not affected by discretization, we estimate them using:

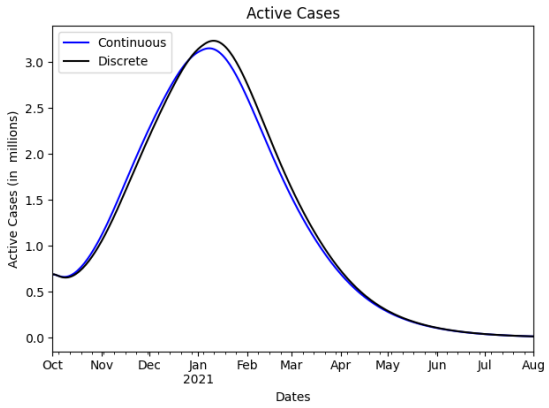
$$\tilde{p_h} = \frac{TH(t+1) - TH(t)}{DT(t+1) - DT(t)} \tag{D.43}$$

$$\tag{D.44}$$

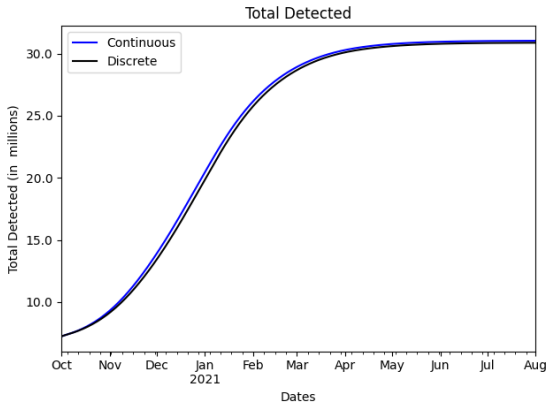
$$\tilde{p_d} = -\frac{DT(t+1) - DT(t)}{[S(t+1) + E(t+1) + I(t+1)] - [S(t) + E(t) + I(t)]} \tag{D.45}$$

$$\tag{D.46}$$

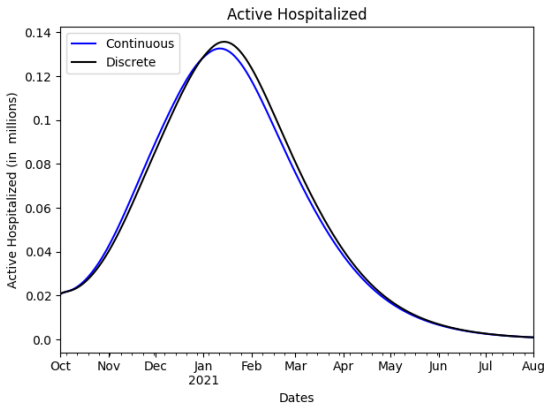
$$p_{\tilde{dth}} = -\frac{\begin{array}{r}[UD(t+1)+DHD(t+1)+DQD(t+1)+D(t+1)] \\ -[UD(t)+DHD(t)+DQD(t)+D(t)]\end{array}}{[S(t+1)+E(t+1)+I(t+1)]-[S(t)+E(t)+I(t)]} \quad \text{(D.47)}$$



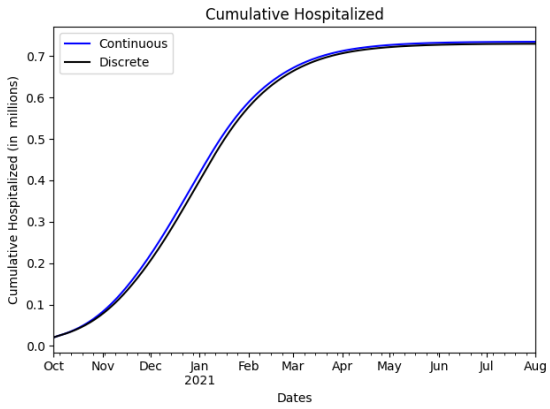
(a) Number of detected cases.



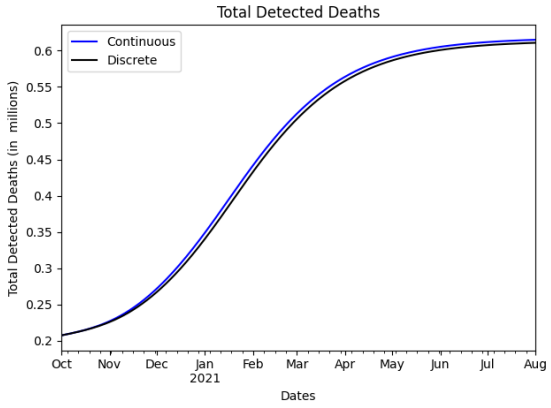
(b) Cumulative number of detected cases.



(c) Number of hospitalized cases.



(d) Cumulative number of hospitalized cases.



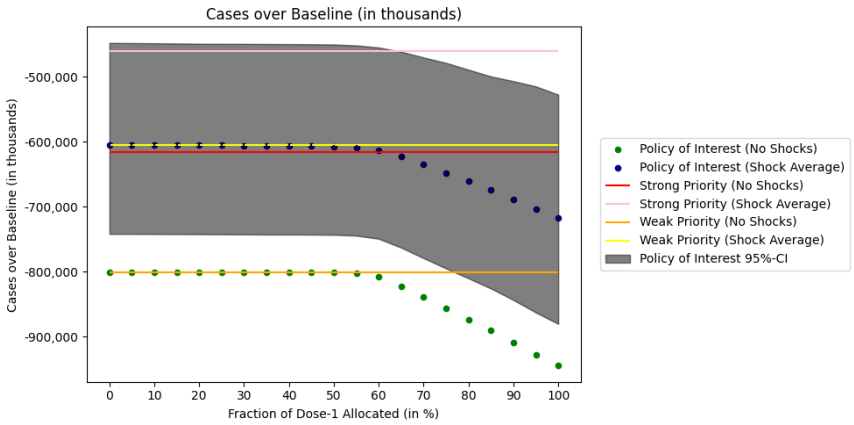
(e) Cumulative number of deaths.

Figure D-2: Comparison of the output for a non-discretized and a discretized (with a time step of 1 day) simulation of the DELPHI. We use the 2021/02/07 DELPHI model parameters.

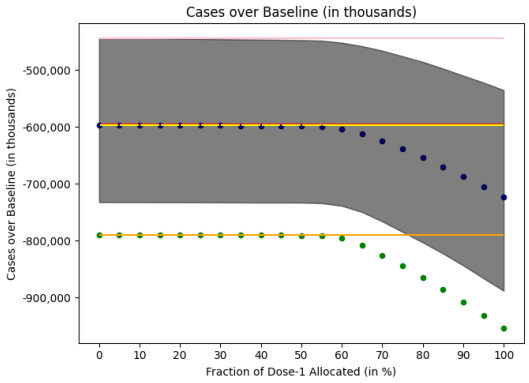
D.2 Sensitivity Analysis

In this section, we explore the sensitivity of our results to key parameters of the model. In D.2.1, we delay the second dose by changing the recommended time frame between two doses from 21 days (3 weeks), to 4 weeks, 5 weeks, 7 weeks, and 9 weeks. In D.2.2, we increase the time to reach permanent immunity after responding positively to a vaccine from 14 days to 21 days. In D.2.3, we modify the terminal supply rate of vaccines from 1.5 million doses per day to 3.0 million doses per day and 0.75 million doses per day. In D.2.4, we increase the efficacy of the first dose of the vaccine by 20% and also decrease the efficacy of the first dose of the vaccine by 20%. In D.2.5, we increase the efficacy of the second dose of the vaccine by 4% and also decrease the efficacy of the first dose of the vaccine by 20%. In D.2.6, we increase the frequency of supply shocks to 1 per 15 days and also decrease the frequency of supply shocks to 1 per 45 days.

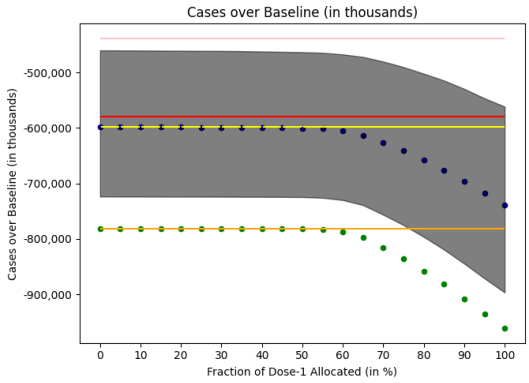
D.2.1 Impact of Delaying the Second Dose



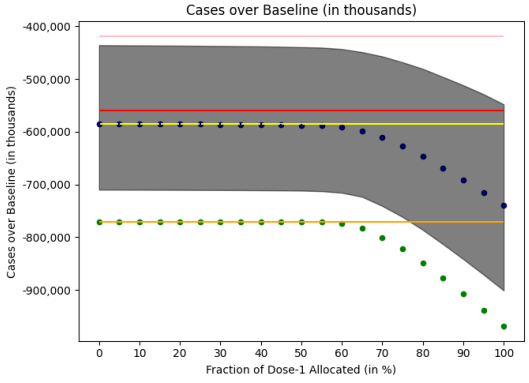
(a) Base Case: 21 days.



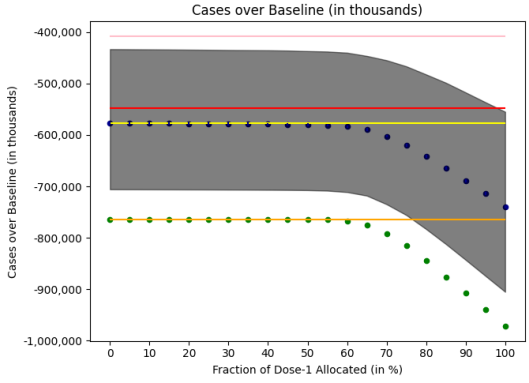
(b) Delay: 28 days.



(c) Delay: 35 days.

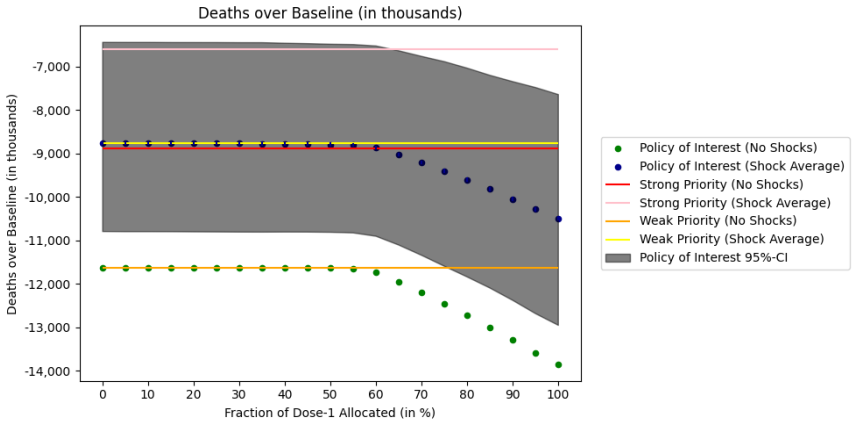


(d) Delay: 49 days.

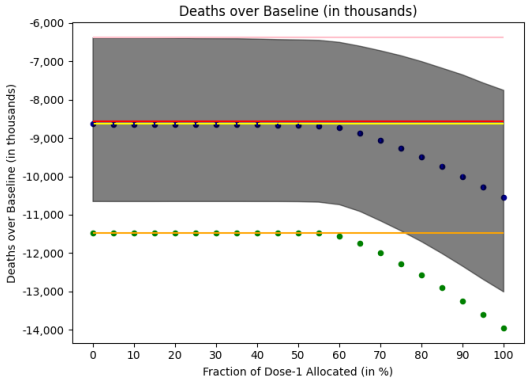


(e) Delay: 63 days.

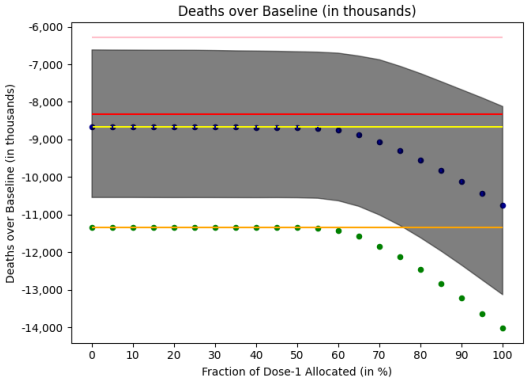
Figure D-3: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of infections** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.



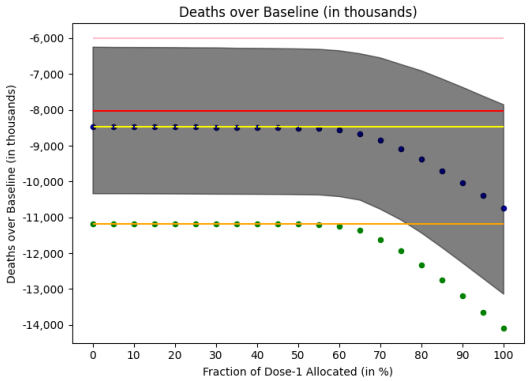
(a) Base Case: 21 days.



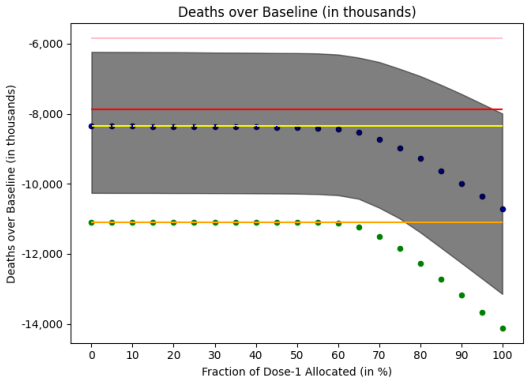
(b) Delay: 28 days.



(c) Delay: 35 days.



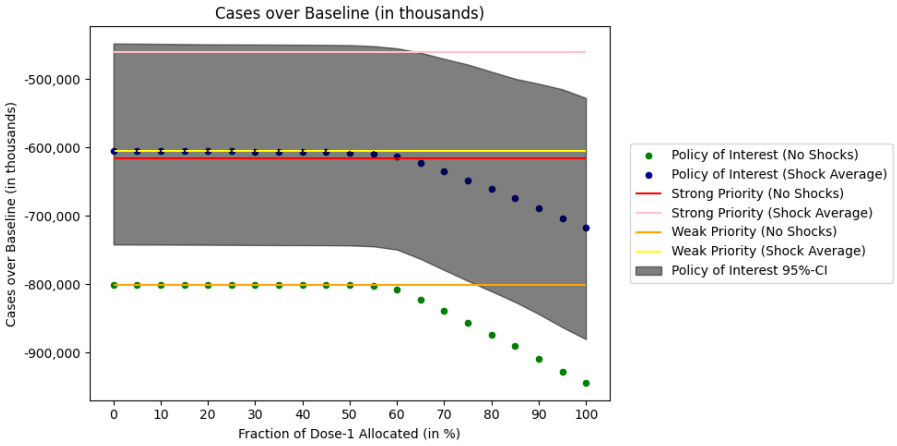
(d) Delay: 49 days.



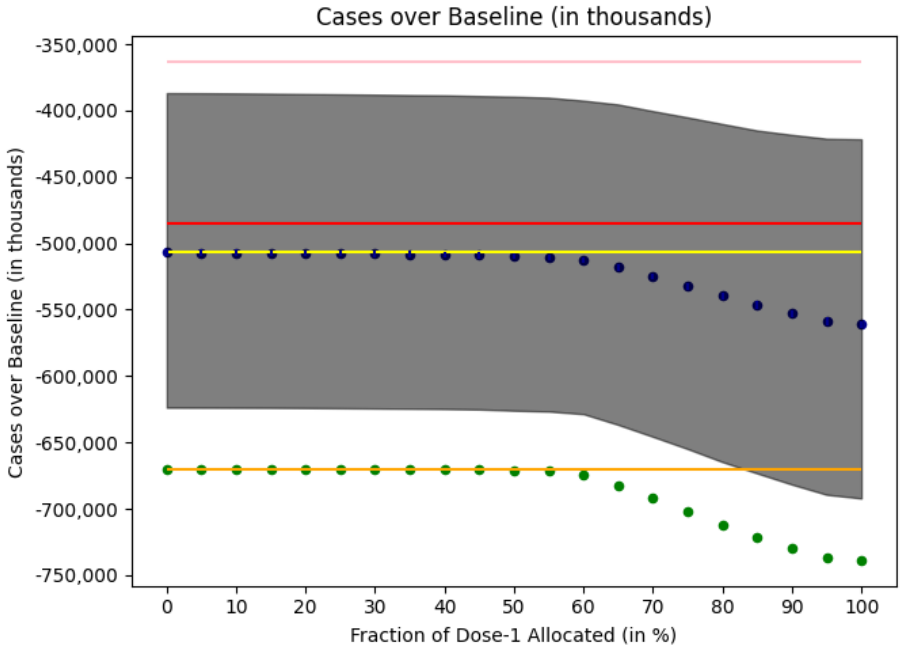
(e) Delay: 63 days.

Figure D-4: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of deaths** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.

D.2.2 Impact of Delaying the Immunity Response

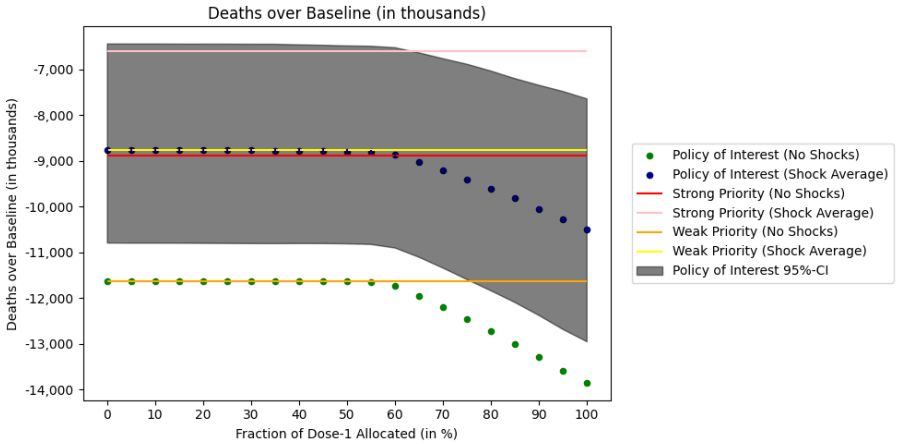


(a) Base Case: 14 days.

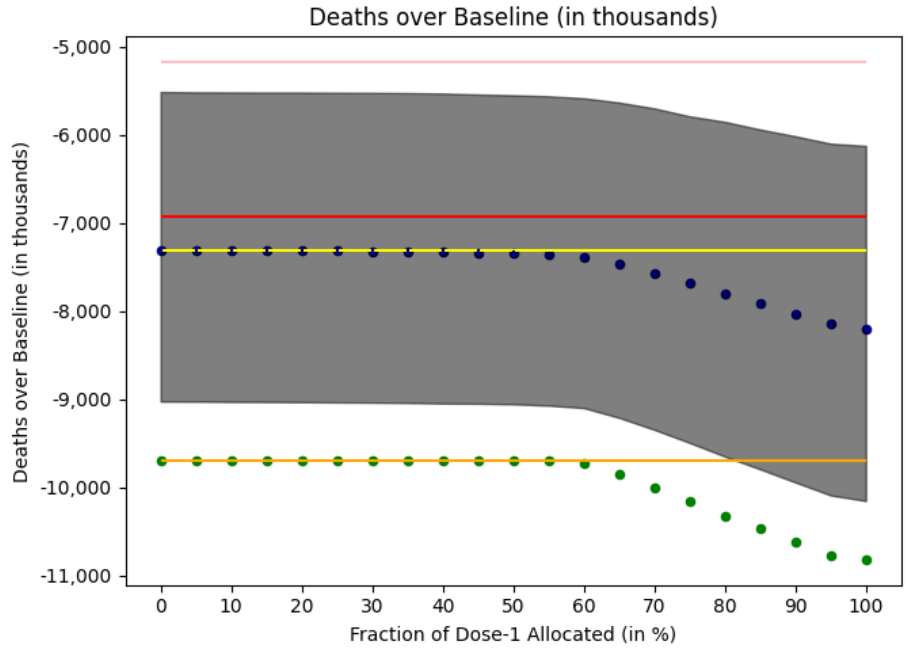


(b) Immunity: 21 days.

Figure D-5: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of infections** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.



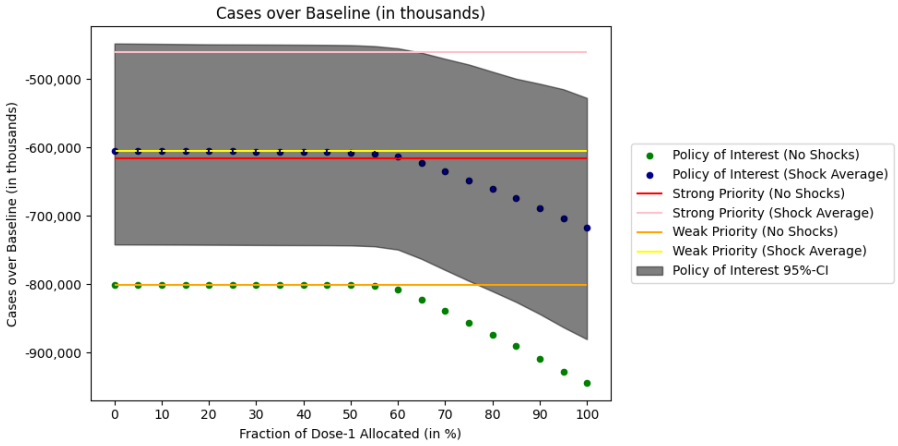
(a) Base Case: 14 days.



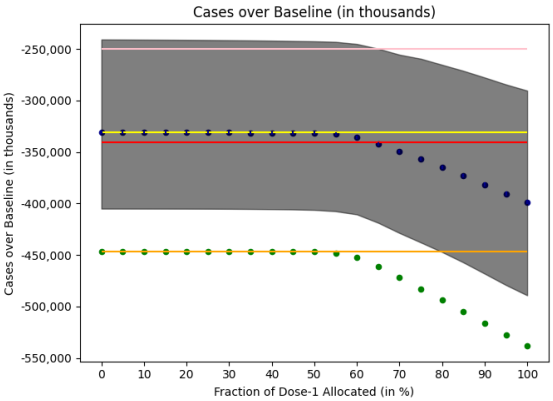
(b) Immunity: 21 days.

Figure D-6: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of deaths** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.

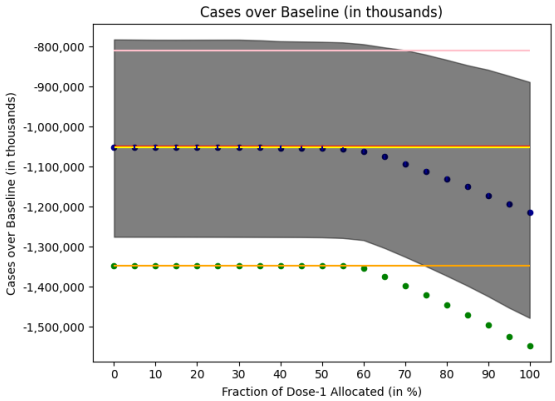
D.2.3 Impact of the Vaccine's Terminal Supply Rate



(a) Base Case: 1.5 million per day.

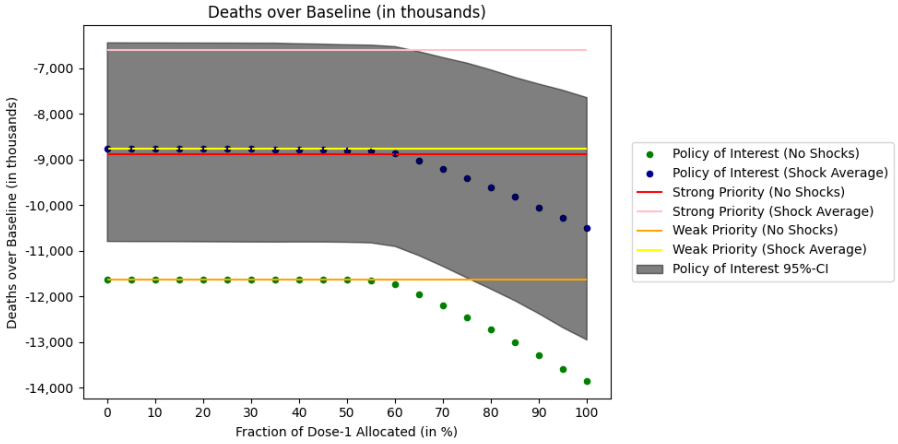


(b) Supply: 0.75 million doses per day.

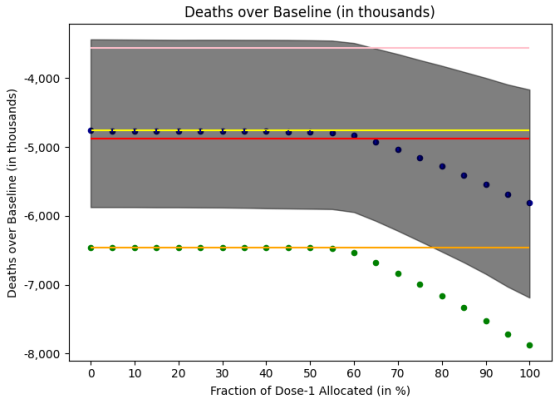


(c) Supply: 3.0 million doses per day.

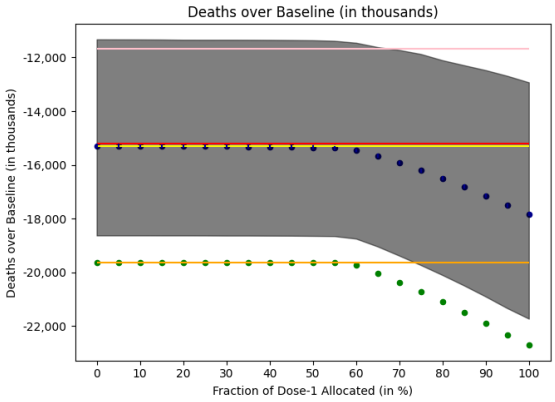
Figure D-7: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of infections** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.



(a) Base Case: 1.5 million per day.



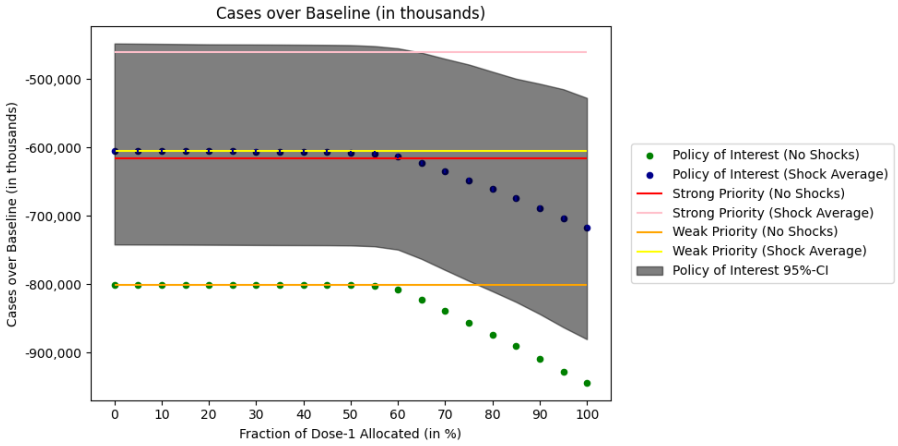
(b) Supply: 0.75 million doses per day.



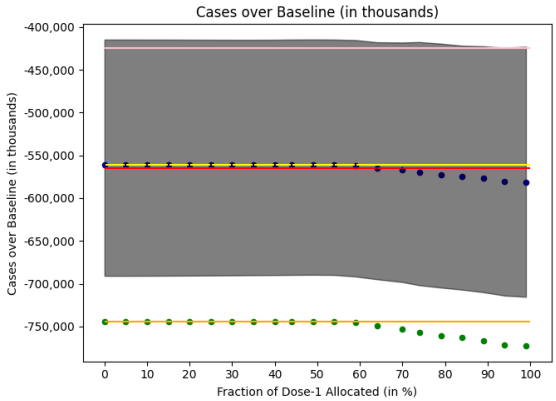
(c) Supply: 3.0 million doses per day.

Figure D-8: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of deaths** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.

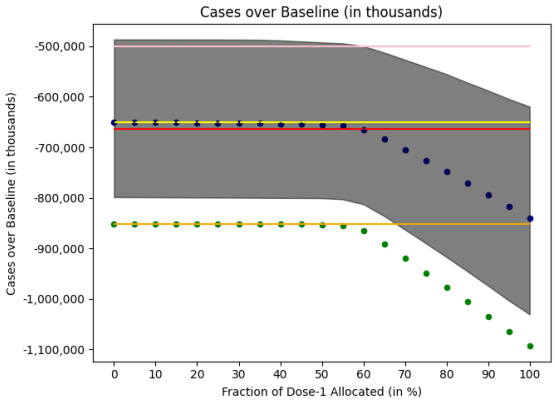
D.2.4 Impact of the First Dose Efficacy



(a) Base Case: 52% (Pfizer), 80.20% (Moderna).

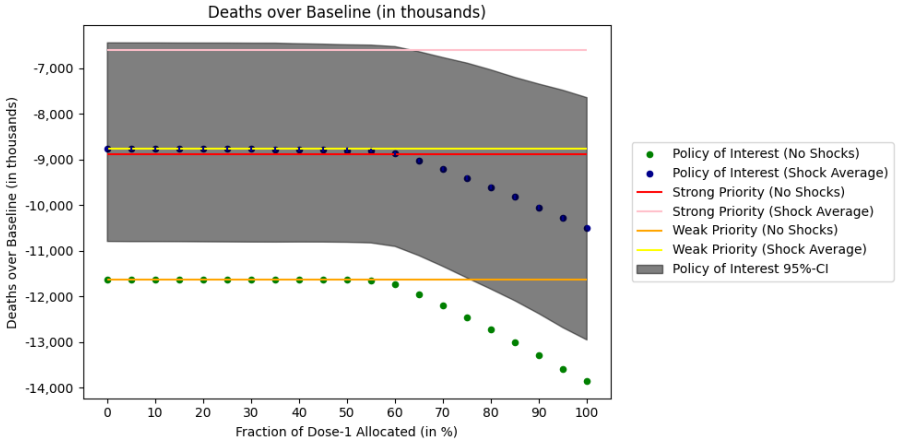


(b) Efficacy: -20%.

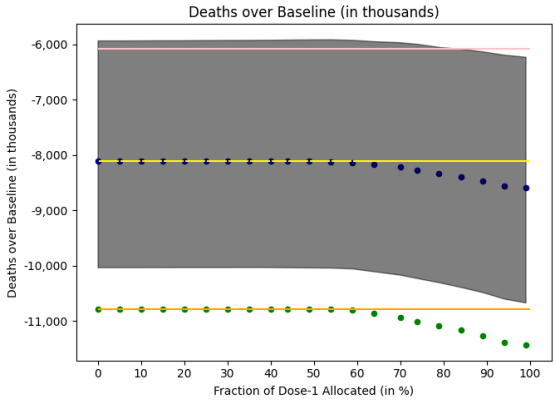


(c) Efficacy: +20%.

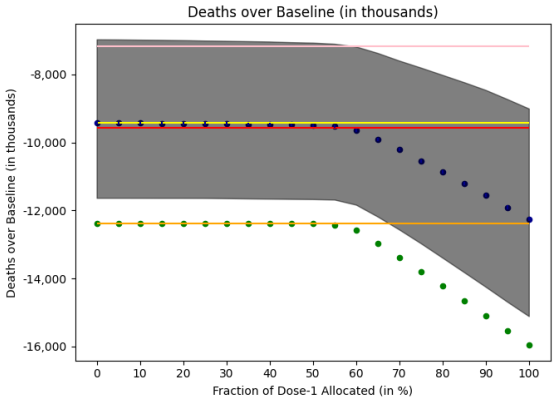
Figure D-9: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of infections** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.



(a) Base Case: 52% (Pfizer), 80.20% (Moderna).



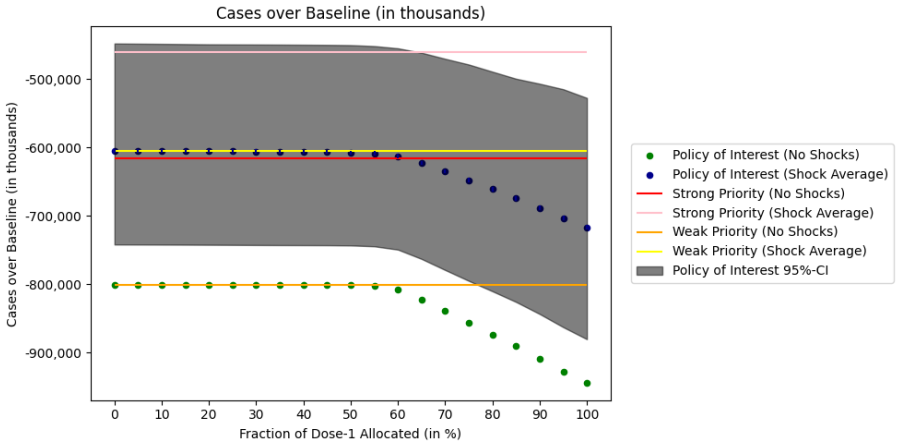
(b) Efficacy: -20%.



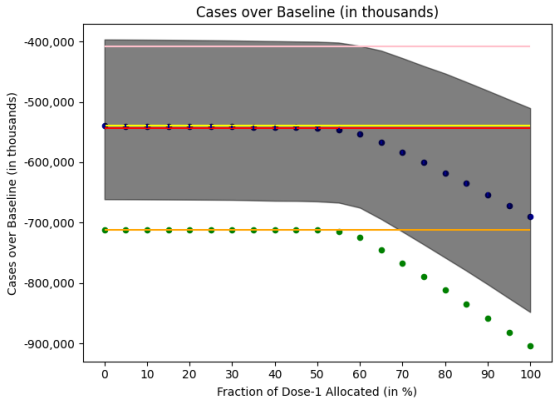
(c) Efficacy: +20%.

Figure D-10: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of deaths** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.

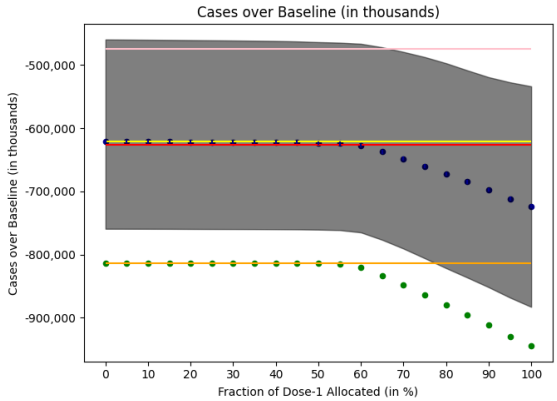
D.2.5 Impact of the Second Dose Efficacy



(a) Base Case: 92% (Pfizer), 95.60% (Moderna).

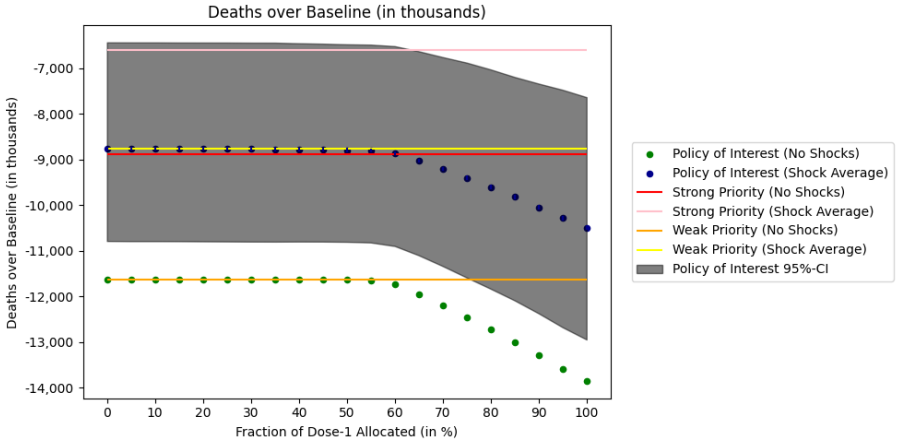


(b) Efficacy: -20%.

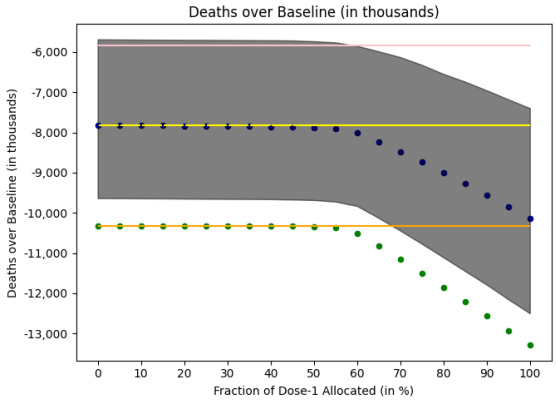


(c) Efficacy: +4%.

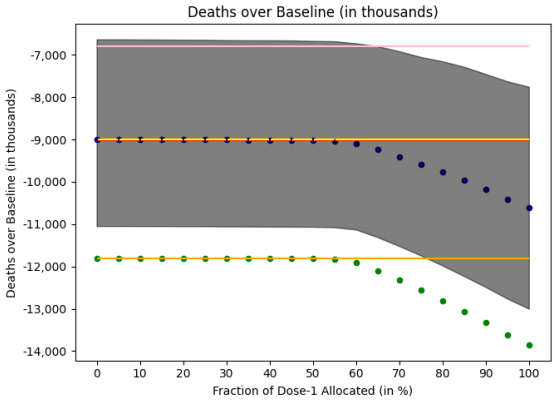
Figure D-11: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of infections** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.



(a) Base Case: 92% (Pfizer), 95.60% (Moderna).



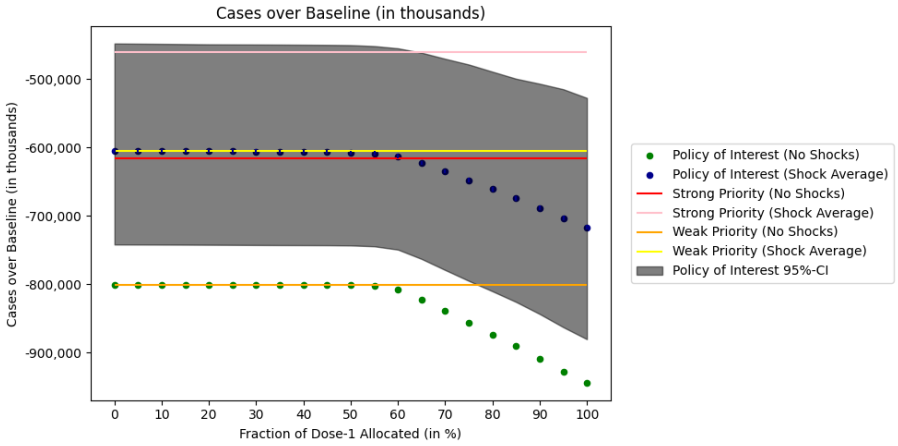
(b) Efficacy: -20%.



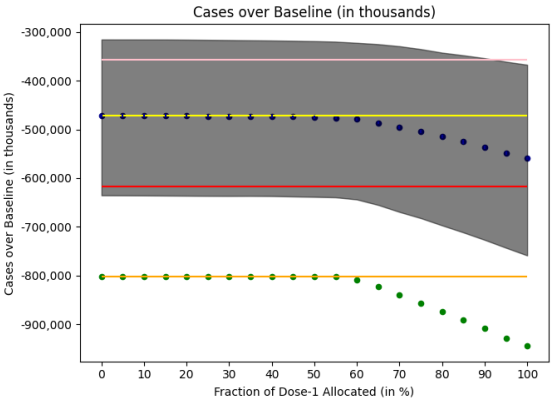
(c) Efficacy: +4%.

Figure D-12: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of deaths** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.

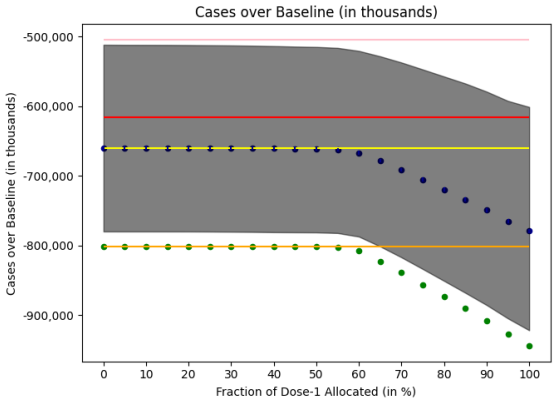
D.2.6 Impact of the Frequency of Supply Shocks



(a) Base Case: $1/30$ day^{-1}.

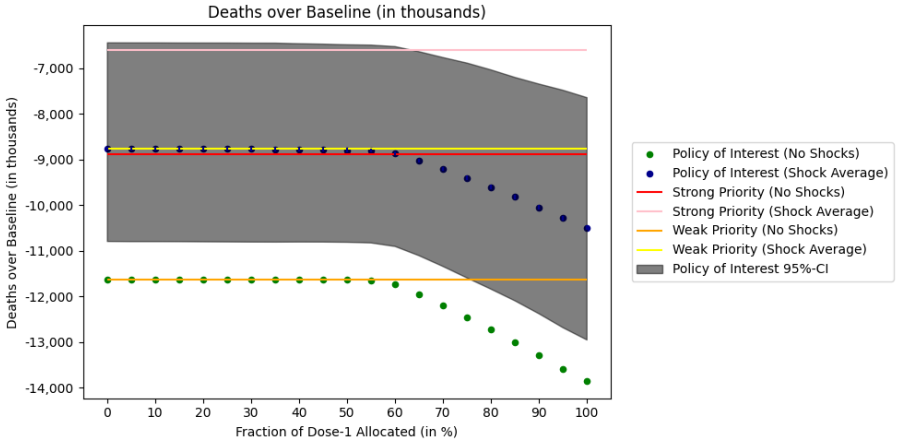


(b) Shock Frequency: $1/15$ day^{-1}.

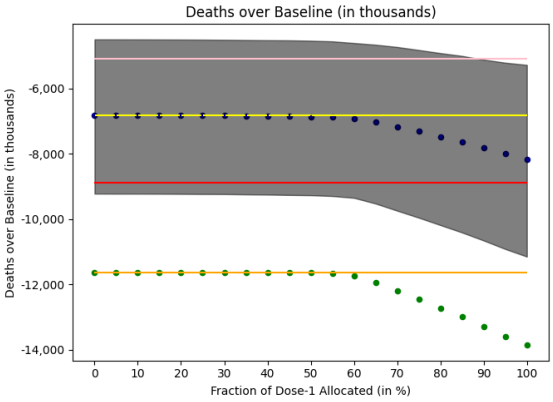


(c) Shock Frequency: $1/45$ day^{-1}.

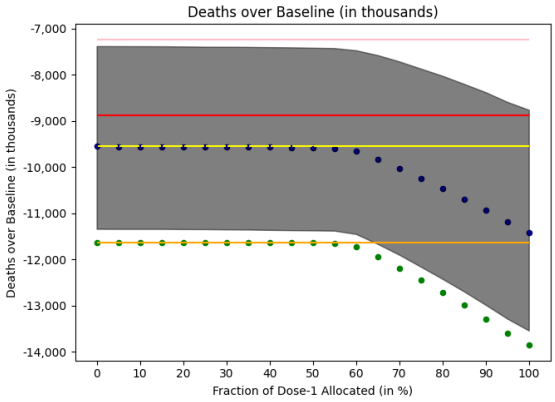
Figure D-13: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of infections** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.



(a) Base Case: $1/30$ day^{-1}.



(b) Shock Frequency: $1/15$ day^{-1}.



(c) Shock Frequency: $1/45$ day^{-1}.

Figure D-14: Simulation of the DELPHI model under supply shocks. We calculate the **cumulative number of deaths** between October 1st, 2020 and August 1st, 2021 relative to a no-vaccination baseline when a constant fraction of available doses are allocated to first-time users. Results under supply shocks are averaged over 1,000 Monte Carlo simulations. We use the February 7th, 2021 DELPHI model parameters.

Appendix E

Chapter 7 Supplementary Material

We provide mathematical proofs to propositions stated in the paper in Section E.1 of the Appendix. Section E.2 describes all the anomaly factors used in the empirical analysis. We include additional sensitivity analyses of the empirical results in Sections E.3, E.4, and E.5. Section E.3 explores the sensitivity of the latent factor models to the choice of γ in the APT loss function (Equation 7.16) using 5-fold cross-validation. We describe in Section E.4 the 60 clusters chosen for the CAE and the RCAE used in the empirical analysis. Section E.5 highlights the sensitivity of the latent factor models to the choice of activation function. Finally, Section E.6 shows the performance of the latent factor models when using a polynomial regression of order 2 (which only contains linear and interaction terms, no second order terms).

E.1 Mathematical proofs

In this section, we provide a mathematical proof of the fact that the shallow linear autoencoder with an APT loss is equivalent to the risk premia PCA introduced by [151] up to a rotation matrix.

Proof of Proposition 1. We provide a proof under the general setting when the bias parameters b^0 and b^1 are not fixed to zero. In this case, the equivalence holds when the input returns R are demeaned. Under the additional assumption that the bias terms b^0 and b^1 are set to zero, we do not need to demean the input returns R.

The APT loss (Equation 7.16) contains two terms. We consider each term separately:

$$\begin{aligned} \text{Term 1} &= \sum_{i=1}^{N}(r_{i,t} - \hat{r}_{i,t})^2, && \text{(E.1)} \\ &= \|R - \hat{R}\|_F^2, && \text{(E.2)} \\ &= \text{trace}\left((R - \hat{R})'(R - \hat{R})\right), && \text{(E.3)} \end{aligned}$$

by definition of the Frobenius norm. The second term is obtained as:

$$\begin{aligned}
\text{Term 2} &= \sum_{i=1}^{N}\left[\sum_{t=1}^{T}(r_{i,t}-\hat{r}_{i,t})\right]^2, && \text{(E.4)}\\
&= \sum_{i=1}^{N}\left[e_i'\cdot(R-\hat{R})\cdot\iota\right]^2, && \text{(E.5)}\\
&= \sum_{i=1}^{N}\iota'\cdot(R-\hat{R})'\cdot e_i\cdot e_i'\cdot(R-\hat{R})\cdot\iota, && \text{(E.6)}\\
&= \iota'\cdot(R-\hat{R})'\cdot\left[\sum_{i=1}^{N}e_i\cdot e_i'\right]\cdot(R-\hat{R})\cdot\iota, && \text{(E.7)}\\
&= \iota'\cdot(R-\hat{R})'\cdot I_T\cdot(R-\hat{R})\cdot\iota, && \text{(E.8)}\\
&= \underbrace{\iota'}_{1\times T}\cdot\underbrace{(R-\hat{R})'}_{T\times N}\cdot\underbrace{(R-\hat{R})}_{N\times T}\cdot\underbrace{\iota}_{T\times 1}, && \text{(E.9)}\\
&= \text{trace}\left(\iota'\cdot(R-\hat{R})'(R-\hat{R})\cdot\iota\right), && \text{(E.10)}\\
&= \|(R-\hat{R})\cdot\iota\|_F^2, && \text{(E.11)}
\end{aligned}$$

where e_i is the i^{th} standard basis vector in $\mathbb{R}^{T\times 1}$, $I_T\in\mathbb{R}^{T\times T}$ is the $T\times T$ identity matrix, and we used in the penultimate step the fact that the trace of a 1×1 matrix is equal to the value of the unique entry of the matrix. Hence, the objective of a shallow autoencoder is to solve the following minimization problem

$$\min_{b^0,b^1,W^0,W^1}\frac{1}{N\cdot T}\cdot\|R-\hat{R}\|_F^2+(1+\gamma)\cdot\frac{1}{N\cdot T^2}\|(R-\hat{R})\cdot\iota\|_F^2, \tag{E.12}$$

where, under a linear activation function,

$$\hat{R}=b^1\cdot\iota'+W^1\cdot b^0\cdot\iota'+W^1\cdot W^0\cdot R. \tag{E.13}$$

Recall from matrix calculus[1] that the derivative of the Frobenius norm of a matrix is given by

$$\frac{\partial}{\partial X}\|X\|_F^2=2X. \tag{E.14}$$

Taking the partial derivative of each term with respect to b^1 yields:

$$\begin{aligned}
\frac{\partial}{\partial b^1}\|R-\hat{R}\|_F^2 &= \frac{\partial}{\partial(R-\hat{R})}\|R-\hat{R}\|_F^2\cdot\frac{\partial}{\partial b^1}\left[R-b^1\cdot\iota'-W^1\cdot b^0\cdot\iota'-W^1\cdot W^0\cdot R\right], && \text{(E.15)}\\
&= -\frac{\partial}{\partial(R-\hat{R})}\|R-\hat{R}\|_F^2\cdot\frac{\partial}{\partial b^1}\left[b^1\cdot\iota'\right], && \text{(E.16)}\\
&= -2\cdot(R-\hat{R})\cdot\iota, && \text{(E.17)}
\end{aligned}$$

[1]For example, see: https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

and

$$\frac{\partial}{\partial b^1}\|(R-\hat{R})\cdot\iota\|_F^2 = \tag{E.18}$$

$$\frac{\partial}{\partial(R-\hat{R})\cdot\iota}\|(R-\hat{R})\cdot\iota\|_F^2 \cdot \frac{\partial}{\partial b^1}\left[R - b^1\cdot\iota' - W^1\cdot b^0\cdot\iota' - W^1\cdot W^0\cdot R\right] \tag{E.19}$$

$$= -\frac{\partial}{\partial(R-\hat{R})\cdot\iota}\|(R-\hat{R})\cdot\iota\|_F^2 \cdot \frac{\partial}{\partial b^1}\left[b^1\cdot\iota'\cdot\iota\right] \tag{E.20}$$

$$= -2\cdot(R-\hat{R})\cdot\iota\cdot T. \tag{E.21}$$

Hence, we need to set:

$$0 = \frac{\partial}{\partial b^1} Loss, \tag{E.22}$$

$$= \frac{-2}{N\cdot T}\cdot(R-\hat{R})\cdot\iota - 2(1+\gamma)\cdot\frac{T}{N\cdot T^2}\cdot(R-\hat{R})\cdot\iota, \tag{E.23}$$

$$= \frac{-2}{N\cdot T}\cdot(2+\gamma)\cdot(R-\hat{R})\cdot\iota, \tag{E.24}$$

implying that

$$0 = (R-\hat{R})\iota, \tag{E.25}$$

$$= \left(R - b^1\cdot\iota' - W^1\cdot b^0\cdot\iota' - W^1\cdot W^0\cdot R\right)\iota, \tag{E.26}$$

$$= R\cdot\iota - T\cdot b^1 - T\cdot W^1\cdot b^0 - W^1\cdot W^0\cdot R\cdot\iota, \tag{E.27}$$

therefore

$$b^1 = \frac{1}{T}\cdot\left(R\cdot\iota - W^1\cdot W^0\cdot R\cdot\iota\right) - W^1\cdot b^0. \tag{E.28}$$

Substituting b^1 into the $(R-\hat{R})$ expression yields:

$$R-\hat{R} = R - b^1\cdot\iota' - W^1\cdot b^0\cdot\iota' - W^1\cdot W^0\cdot R, \tag{E.29}$$

$$= R - \frac{1}{T}\left(R\iota\iota' - W^1\cdot W^0\cdot R\iota\iota' - T\cdot W^1\cdot b^0\iota'\right) - W^1\cdot b^0\cdot\iota' - W^1\cdot W^0\cdot R, \tag{E.30}$$

$$= (R - \frac{1}{T}R\iota\iota') - (W^1\cdot W^0\cdot R - \frac{1}{T}W^1\cdot W^0\cdot R\iota\iota'), \tag{E.31}$$

$$= (R - \frac{1}{T}R\iota\iota') - W^1\cdot W^0\cdot(R - \frac{1}{T}R\iota\iota'). \tag{E.32}$$

Hence, the optimal choice of b^1 demeans the returns R. However, instead of demeaning the returns, we can force the bias parameters b^0 and b^1 to zero, leading to the same expression:

$$R-\hat{R} = R - b^1\cdot\iota' - W^1\cdot b^0\cdot\iota' - W^1\cdot W^0\cdot R, \tag{E.33}$$

$$= R - W^1\cdot W^0\cdot R. \tag{E.34}$$

In the remainder of the proof, we will assume that $b^0 = 0$ and $b^1 = 0$ without loss of

generality to avoid demeaning input returns.

Taking the partial derivative of each term of the APT loss with respect to W^0 yields:

$$\begin{aligned}
\frac{\partial}{\partial W^0}\|R-\hat{R}\|_F^2 &= \frac{\partial}{\partial(R-\hat{R})}\|R-\hat{R}\|_F^2 \cdot \frac{\partial}{\partial W^0}\left[R - b^1\cdot\iota' - W^1\cdot b^0\cdot\iota' - W^1\cdot W^0\cdot R\right], && \text{(E.35)}\\
&= -\frac{\partial}{\partial(R-\hat{R})}\|R-\hat{R}\|_F^2 \cdot \frac{\partial}{\partial W^0}\left[W^1\cdot W^0\cdot R\right], && \text{(E.36)}\\
&= -2\cdot(W^1)'\cdot(R-\hat{R})\cdot R', && \text{(E.37)}
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial}{\partial W^0}\|(R-\hat{R})\cdot\iota\|_F^2 &= && \text{(E.38)}\\
\frac{\partial}{\partial(R-\hat{R})\cdot\iota}\|(R-\hat{R})\cdot\iota\|_F^2 &\cdot \frac{\partial}{\partial W^0}\left[R - b^1\cdot\iota' - W^1\cdot b^0\cdot\iota' - W^1\cdot W^0\cdot R\right]\iota, && \text{(E.39)}\\
&= -\frac{\partial}{\partial(R-\hat{R})\cdot\iota}\|(R-\hat{R})\cdot\iota\|_F^2 \cdot \frac{\partial}{\partial W^0}\left[-W^1\cdot W^0\cdot R\right]\iota, && \text{(E.40)}\\
&= -2\cdot(W^1)'\cdot(R-\hat{R})\cdot\iota\iota'\cdot R'. && \text{(E.41)}
\end{aligned}$$

Hence, we need to set:

$$\begin{aligned}
0 &= \frac{\partial}{\partial W^0}Loss, && \text{(E.42)}\\
&= \frac{-2}{N\cdot T}\cdot(W^1)'\cdot(R-\hat{R})\cdot R' - 2(1+\gamma)\cdot\frac{1}{N\cdot T^2}\cdot(W^1)'\cdot(R-\hat{R})\cdot\iota\iota'\cdot R', && \text{(E.43)}\\
&= \frac{-2}{N\cdot T}\cdot(W^1)'\cdot(R-\hat{R})\cdot(R'+\frac{1+\gamma}{T}\cdot\iota\iota'\cdot R'), && \text{(E.44)}\\
&= \frac{-2}{N\cdot T}\cdot(W^1)'\cdot(R-W^1\cdot W^0\cdot R)\cdot(R'+\frac{1+\gamma}{T}\cdot\iota\iota'\cdot R'), && \text{(E.45)}\\
&= \frac{-2}{N\cdot T}\cdot\left[(W^1)'\cdot R-(W^1)'\cdot W^1\cdot W^0\cdot R\right]\cdot(R'+\frac{1+\gamma}{T}\cdot\iota\iota'\cdot R'), && \text{(E.46)}
\end{aligned}$$

implying that

$$\begin{aligned}
0 &= (W^1)'\cdot R\cdot(R'+\frac{1+\gamma}{T}\cdot\iota\iota'\cdot R') - (W^1)'\cdot W^1\cdot W^0\cdot R\cdot(R'+\frac{1+\gamma}{T}\cdot\iota\iota'\cdot R'), && \text{(E.47)}\\
&= (W^1)'\cdot W^1\cdot\left[\left[(W^1)'\cdot W^1\right]^{-1}(W^1)'-W^0\right]\cdot(RR'+\frac{1+\gamma}{T}\cdot R\cdot\iota\iota'\cdot R'). && \text{(E.48)}
\end{aligned}$$

Assuming $[(W^1)'\cdot W^1]$ and $\left(RR'+\frac{1+\gamma}{T}\cdot R\cdot\iota\iota'\cdot R'\right)$ are full rank, we obtain

$$W^0 = \left[(W^1)'\cdot W^1\right]^{-1}(W^1)'. \qquad \text{(E.49)}$$

Substituting W^0 into the $(R - \hat{R})$ expression yields:

$$\begin{aligned} R - \hat{R} &= R - \left[(W^1)' \cdot W^1\right]^{-1} (W^1)' \cdot R, && \text{(E.50)} \\ &= M_{W^1} \cdot R, && \text{(E.51)} \end{aligned}$$

where $M_{W^1} = 1 - \left[(W^1)' \cdot W^1\right]^{-1} (W^1)'$ is the projection matrix that annihilates the subspace spanned by W.

Hence, the objective of a shallow autoencoder is to solve the following minimization problem:

$$\begin{aligned} & \min_{W^1} \frac{1}{N \cdot T} \cdot \|M_{W^1} \cdot R\|_F^2 + (1+\gamma) \cdot \frac{1}{N \cdot T^2} \|M_{W^1} \cdot R \cdot \iota\|_F^2, && \text{(E.52)} \\ = & \min_{W^1} \frac{1}{N \cdot T} \cdot \text{trace}\left(M_{W^1} \cdot RR' \cdot M'_{W^1}\right) + (1+\gamma) \cdot \frac{1}{N \cdot T^2} \text{trace}\left(M_{W^1} \cdot R \cdot \iota\iota' \cdot R' \cdot M'_{W^1}\right), && \text{(E.53)} \\ = & \min_{W^1} \frac{1}{N \cdot T} \cdot \text{trace}\left(M_{W^1} \cdot RR' \cdot M'_{W^1} + \frac{1+\gamma}{T} \cdot M_{W^1} \cdot R \cdot \iota\iota' \cdot R' \cdot M'_{W^1}\right), && \text{(E.54)} \\ = & \min_{W^1} \frac{1}{N \cdot T} \cdot \text{trace}\left(M_{W^1} \cdot R \cdot \left[I_T + \frac{1+\gamma}{T} \cdot \iota\iota'\right] \cdot R' \cdot M'_{W^1}\right). && \text{(E.55)} \end{aligned}$$

Therefore W^1 consists of the first K eigenvectors of $\frac{1}{N \cdot T} \cdot R \cdot \left(I_T + \frac{1+\gamma}{T} \iota\iota'\right) \cdot R'$, which is equivalent to the solution to the risk premia PCA proposed by [151].

□

E.2 Factor description

Table E.1: This table describes the anomaly factors used in Section 7.3. All factors are taken from [85]. The monthly returns range from July 1976 to December 2017.

Factor	Description	Factor	Description
absacc	Absolute accruals	lfe	Labor Force Efficiency
acc	Working capital accruals	lgr	Growth in long-term debt
adm	Advertising Expense-to-market	LIQ_PS	Liquidity
aeavol	Abnormal earnings announcement volume	LTR	Long-Term Reversal
age	years since first Compustat coverage	maxret	Maximum daily return
ala	Book Asset Liquidity	MktRf	Excess Market Return
ato	Asset turnover	mom36m	36-month momentum
BAB	Betting Against Beta	mom6m	6-month momentum
baspread	Bid-ask spread	moms12m	Seasonality
beta	Market Beta	ms	Financial statements score
bm_ia	Industry-adjusted book to market	mve_ia	Industry-adjusted size
cash	Cash holdings	ndf	Net debt finance
cashdebt	Cash flow to debt	ndp	Net debt-to-price
cashpr	Cash productivity	nef	Net equity finance
cdi	Composite Debt Issuance	nincr	Number of earnings increases
cei	Composite Equity Issuance	noa	Net Operating Assets
cfp	Cash flow to price ratio	nop	Net payout yield
cfp_ia	Industry-adjusted cash flow to price ratio	nxf	Net external finance
chatoia	Industry-adjusted change in asset turnover	ob_a	Order backlog
chcsho	Change in shares outstanding	ol	Operating Leverage
chempia	Industry-adjusted change in employees	op	Payout yield
chinv	Change in inventory	orgcap	Organizational Capital

Continued on next page

Table E.1 – continued from previous page

Factor	Description	Factor	Description
chmom	Change in 6-month momentum	os	Ohlson's O-score
chpmia	Industry-adjusted change in profit margin	pchcapx	Investment Growth
chtx	Change in tax expense	pchcapx_ia	Industry adjusted % change in capital expendit...
cinvest	Corporate investment	pchcapx3	Three-year Investment Growth
cinvest_a	Abnormal Corporate Investment	pchcurrat	% change in current ratio
CMA	Conservative Minus Aggressive	pchdepr	% change in depreciation
convind	Convertible debt indicator	pchgm_pchsale	% change in gross margin - % change in sales
cp	Cash flow-to-price	pchquick	% change in quick ratio
cto	Capital turnover	pchsale_pchinvt	% change in sales - % change in inventory
currat	Current ratio	pchsale_pchrect	% change in sales - % change in A/R
dcoa	Change in Current Operating Assets	pchsale_pchxsga	% change in sales - % change in SG&A
dcol	Change in Current Operating Liabilities	pchsaleinv	% change sales-to-inventory
depr	Depreciation / PP&E	pctacc	Percent accruals
dfin	Change in Net Financial Assets	pm	Profit margin
dfnl	Change in Financial Liabilities	poa	Percent Operating Accruals
divi	Dividend initiation	pps	Share price
divo	Dividend omission	pricedelay	Price delay
dnca	Change in Non-current Operating Assets	ps	Financial statements score
dncl	Change in Non-current Operating Liabilities	QMJ	Quality Minus Junk
dnco	Change in Net Non-current Operating Assets	quick	Quick ratio
dnoa	Changes in Net Operating Assets	rd	R&D increase
dolvol	Dollar trading volume	rdm	R&D Expense-to-market
dpia	Changes in PPE and Inventory-to-assets	rds	R&D-to-sales
dsti	Change in Short- term Investments	realestate_hxz	Industry-adjusted Real Estate Ratio
dwc	Changes in Net Non-cash Working Capital	retvol	Return volatility
dy	Dividend to price	RMW	Robust Minus Weak

Continued on next page

Table E.1 – continued from previous page

Factor	Description	Factor	Description
ear	Earnings announcement return	rna	Return on net operating assets
ebp	Enterprise book-to-price	roaq	Return on assets
egr	Growth in common shareholder equity	roavol	Earnings volatility
egr_hxz	Change in Book Equity	roic	Return on invested capital
em	Enterprise multiple	rs	Revenue Surprises
ep	Earnings to price	rsup	Revenue surprise
etr	Effective Tax Rate	salecash	Sales to cash
gad	Growth in advertising expense	saleinv	Sales to inventory
gma	Gross profitability	salerec	Sales to receivables
grcapx	Growth in capital expenditures	sgr	Sales growth
grltnoa	Growth in long term net operating assets	sin	Sin stocks
grltnoa_hxz	Changes in Long-term Net Operating Assets	SMB	Small Minus Big
herf	Industry Concentration	sp	Sales to price
hire	Employee growth rate	std_dolvol	Volatility of liquidity (dollar trading volume)
HML	High Minus Low	std_turn	Volatility of liquidity (share turnover)
HML_Devil	HML Devil	stdacc	Accrual volatility
HXZ_IA	HXZ Investment	stdcf	Cash flow volatility
HXZ_ROE	HXZ Profitability	STR	1-month momentum
idiovol	Idiosyncratic return volatility	sue	Unexpected quarterly earnings
ill	Illiquidity	ta	Total accruals
indmom	Industry momentum	tang	Debt capacity/firm tangibility
Intermediary	Intermediary Risk Factor	tb	Tax income to book income
invest	Capital expenditures and inventory	turn	Share turnover
IPO	New equity issue	UMD	Momentum
ivg	Inventory Growth	ww	Whited-Wu Index
kz	Kaplan-Zingales Index	zerotrade	Zero trading days

Continued on next page

Table E.1 – continued from previous page

Factor	Description	Factor	Description
lev	Leverage	zs	Altman's Z-score

E.3 Sensitivity of the APT loss

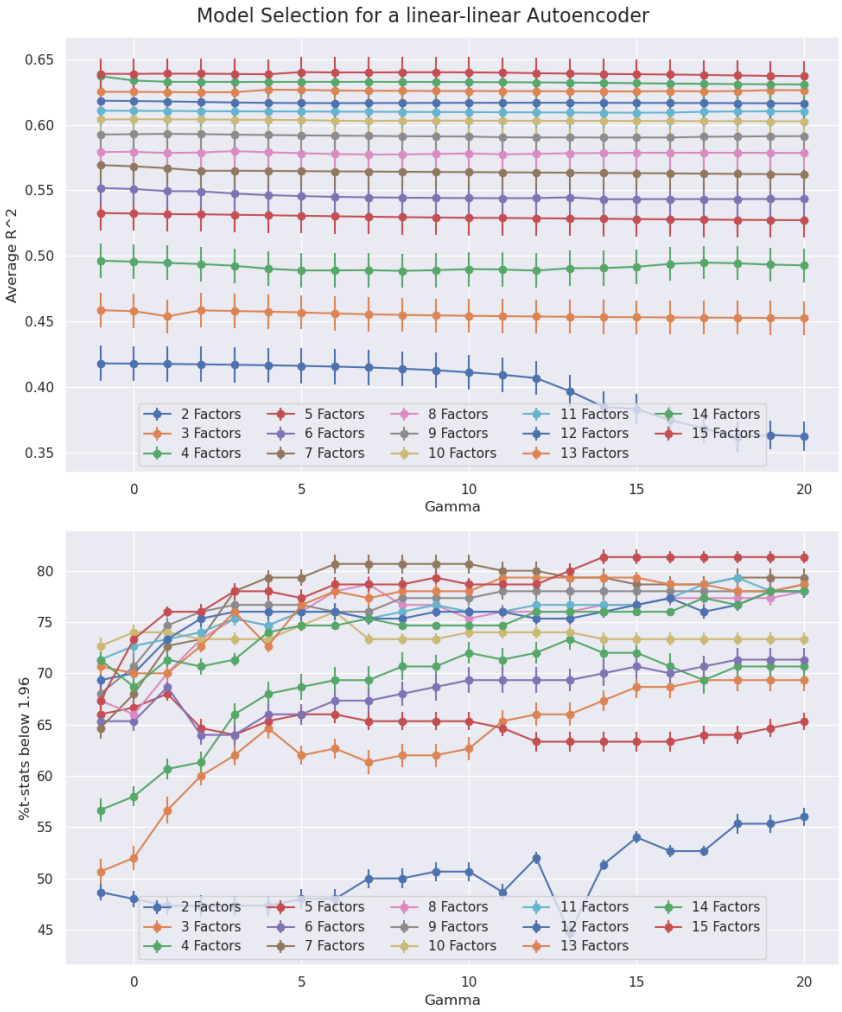


Figure E-1: Performance of latent factor models for a **shallow AE with linear activation functions** on the 150 anomaly factors as we increase the gamma parameter in the APT loss from -1 to 20. The top panel shows the average R-squared obtained across the 150 regressions and the bottom panel shows the average percentage of anomaly factors that are explained by the latent factor model, using 5-fold cross-validation. Standard errors are displayed for each model considered.

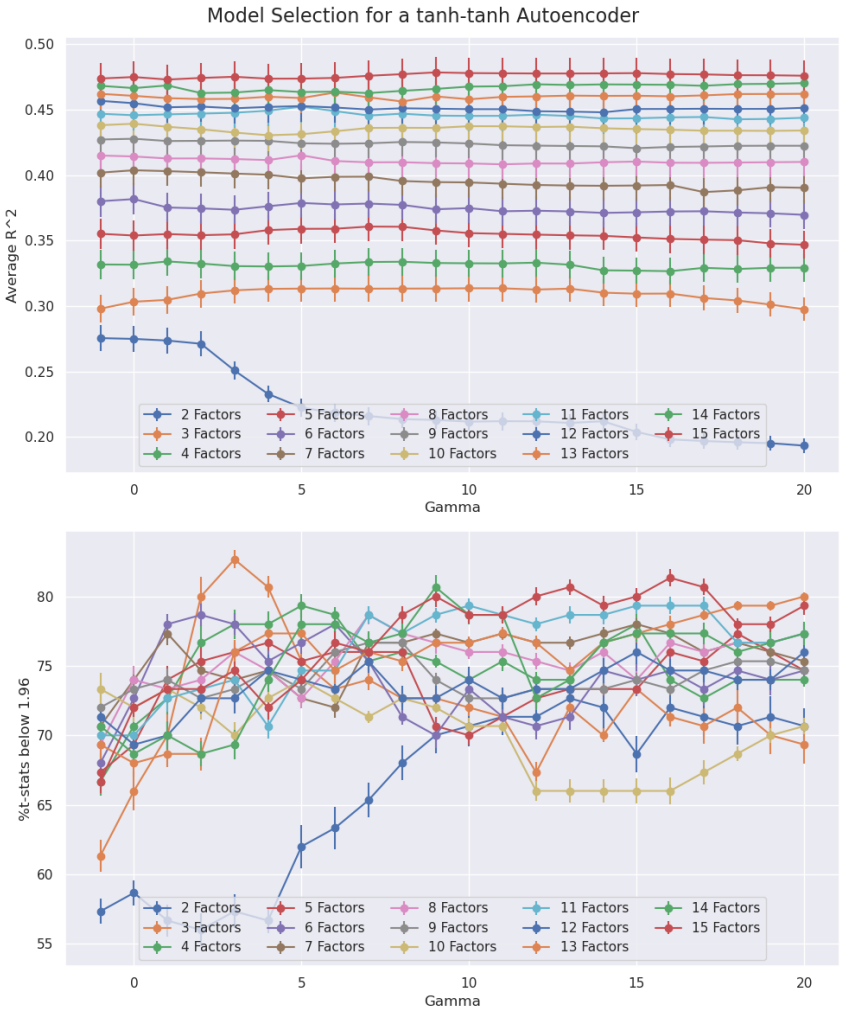


Figure E-2: Performance of latent factor models for a **shallow AE with tanh-tanh activation functions** on the 150 anomaly factors as we increase the gamma parameter in the APT loss from -1 to 20. The top panel shows the average R-squared obtained across the 150 regressions and the bottom panel shows the average percentage of anomaly factors that are explained by the latent factor model, using 5-fold cross-validation. Standard errors are displayed for each model considered.

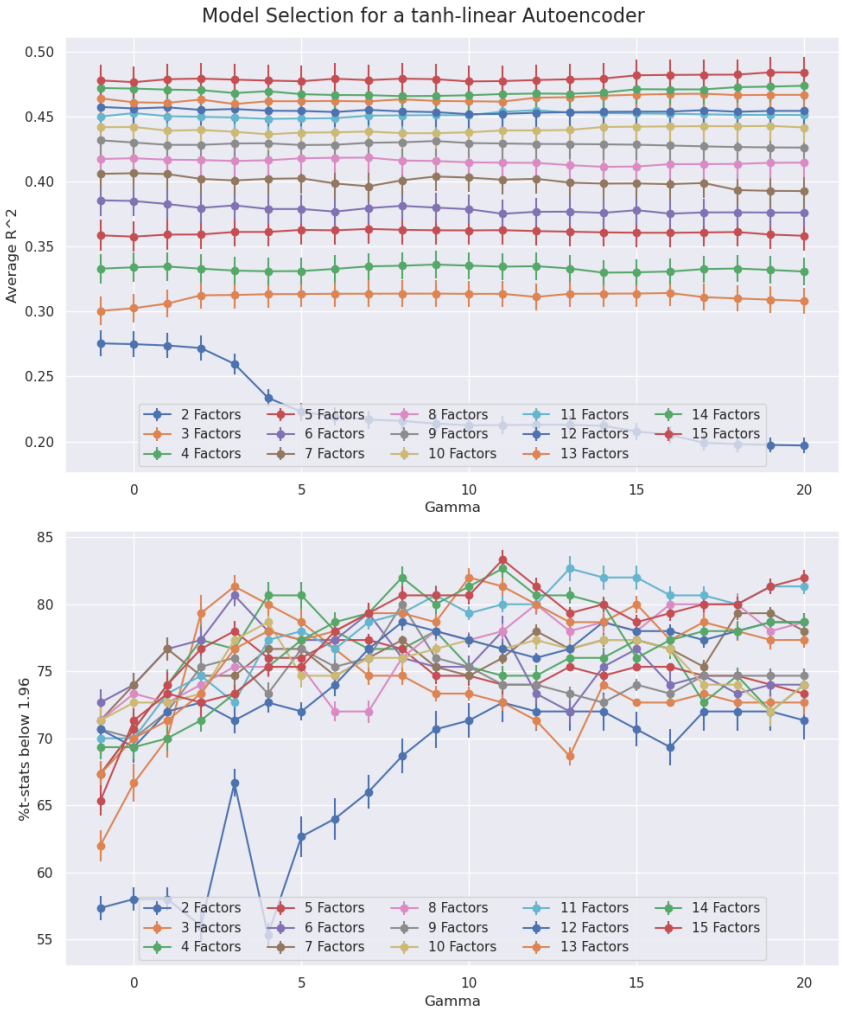


Figure E-3: Performance of latent factor models for a **shallow AE with tanh-linear activation functions** on the 150 anomaly factors as we increase the gamma parameter in the APT loss from -1 to 20. The top panel shows the average R-squared obtained across the 150 regressions and the bottom panel shows the average percentage of anomaly factors that are explained by the latent factor model, using 5-fold cross-validation. Standard errors are displayed for each model considered.

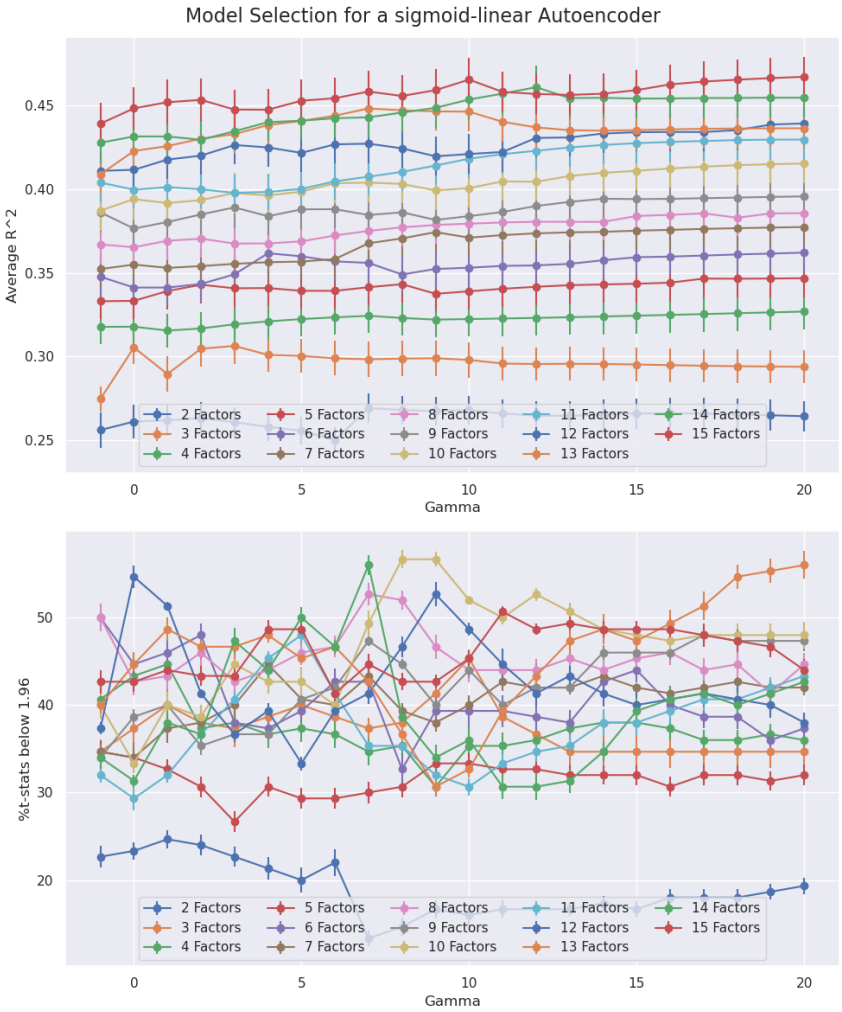


Figure E-4: Performance of latent factor models for a **shallow AE with sigmoid-linear activation functions** on the 150 anomaly factors as we increase the gamma parameter in the APT loss from -1 to 20. The top panel shows the average R-squared obtained across the 150 regressions and the bottom panel shows the average percentage of anomaly factors that are explained by the latent factor model, using 5-fold cross-validation. Standard errors are displayed for each model considered.

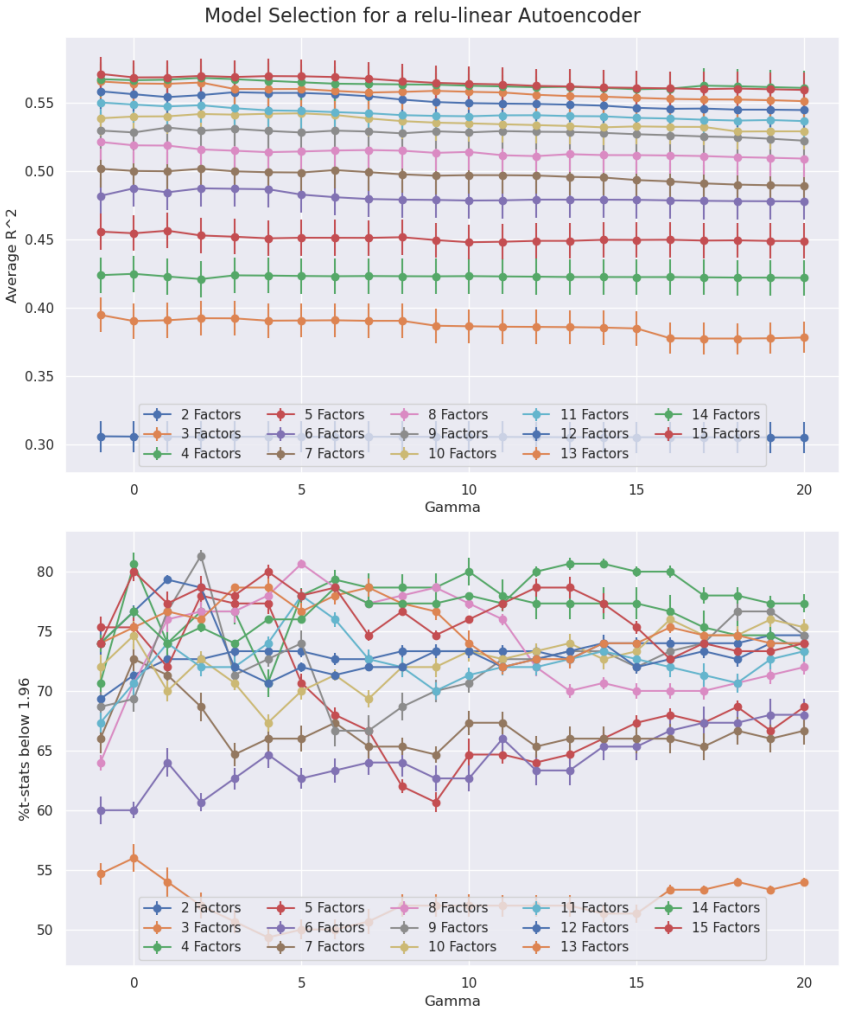


Figure E-5: Performance of latent factor models for a **shallow AE with ReLU-linear activation functions** on the 150 anomaly factors as we increase the gamma parameter in the APT loss from -1 to 20. The top panel shows the average R-squared obtained across the 150 regressions and the bottom panel shows the average percentage of anomaly factors that are explained by the latent factor model, using 5-fold cross-validation. Standard errors are displayed for each model considered.

E.4 Sensitivity to the number of clusters

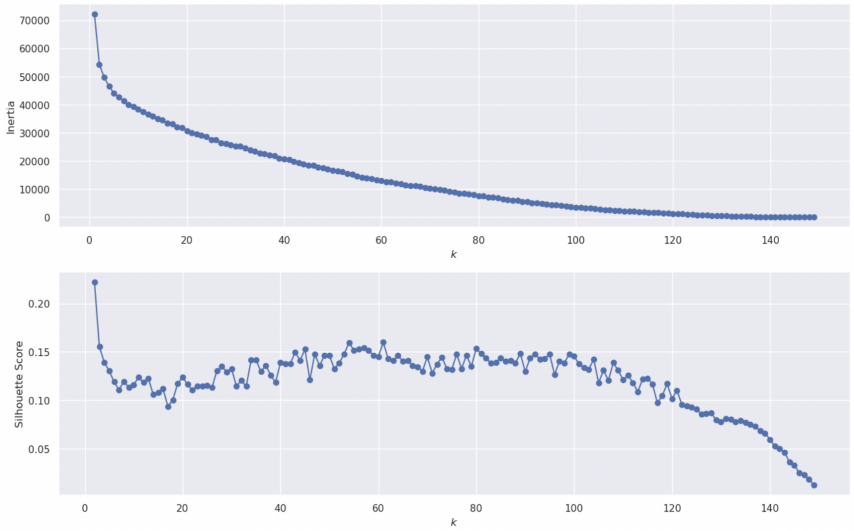


Figure E-6: Selecting the number of clusters to group anomaly factors. We plot the inertia (top panel) and the silhouette score (bottom panel) as a function of the number of clusters chosen.

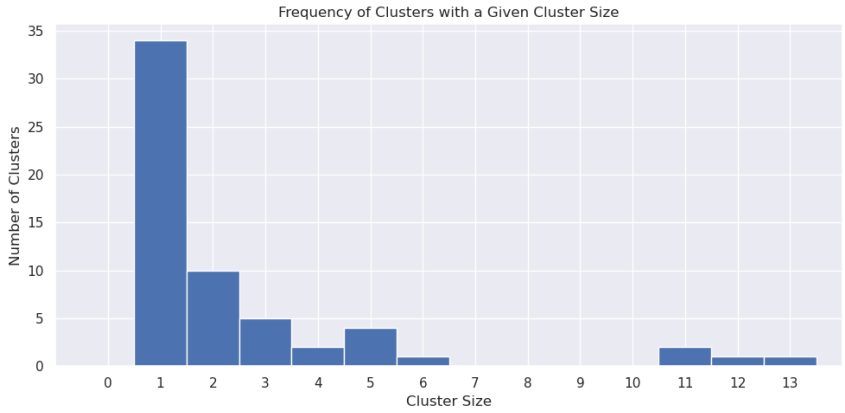


Figure E-7: Histogram of the cluster sizes obtained for $k = 60$ clusters.

Table E.2: Distribution of cluster sizes when grouping anomaly factors into $k = 60$ clusters.

Cluster Size	Cluster Assignment
1	34

Continued on next page

Table E.2 – continued from previous page

Cluster Size	Cluster Assignment
2	10
3	5
4	2
5	4
6	1
11	2
12	1
13	1

Table E.3: Clusters obtained when grouping anomaly factors into $k = 60$ clusters. The anomaly factors are described in Table E.1.

Anomaly factor	Cluster Size	Cluster	Anomaly factor	Cluster Size	Cluster
op	13	13	pps	4	10
chcsho	13	13	cto	4	8
ato	13	13	cashdebt	4	10
roavol	13	13	em	4	10
age	13	13	chinv	4	8
rna	13	13	dwc	4	8
convind	13	13	STR	3	18
nxf	13	13	pchcapx	3	38
nop	13	13	moms12m	3	11
nef	13	13	UMD	3	11
ww	13	13	mom6m	3	11
ep	13	13	rs	3	7
salerec	13	13	pchcapx3	3	38
beta	12	4	cei	3	18
IPO	12	4	grcapx	3	38
maxret	12	4	mom36m	3	18
turn	12	4	orgcap	3	14
std_turn	12	4	rsup	3	7
SMB	12	4	chtx	3	7
baspread	12	4	ala	3	14
retvol	12	4	gma	3	14
dy	12	4	currat	2	2
zerotrade	12	4	pctacc	2	29
absacc	12	4	dfnl	2	19
idiovol	12	4	dolvol	2	56

Continued on next page

Table E.3 – continued from previous page

Anomaly factor	Cluster Size	Cluster	Anomaly factor	Cluster Size	Cluster
cp	11	6	BAB	2	56
HML_Devil	11	6	ndf	2	19
cashpr	11	6	pchsaleinv	2	5
depr	11	1	aeavol	2	2
cfp	11	6	herf	2	17
quick	11	1	poa	2	29
salecash	11	1	pchsale_pchinvt	2	5
cash	11	1	dsti	2	17
rds	11	1	cinvest_a	2	27
rdm	11	1	pchdepr	2	27
bm_ia	11	6	ta	2	15
zs	11	6	Intermediary	2	25
lev	11	6	MktRf	2	25
HML	11	6	pchcurrat	2	24
adm	11	1	dfin	2	15
sp	11	6	pchquick	2	24
tang	11	1	sue	1	47
rd	11	1	cdi	1	48
ebp	11	6	chempia	1	49
nincr	11	1	ol	1	50
ndp	11	1	realestate_hxz	1	57
kz	11	6	acc	1	51
stdacc	6	58	sin	1	52
HXZ_ROE	6	58	ear	1	53
tb	6	58	divi	1	54
RMW	6	58	pricedelay	1	55
ps	6	58	cfp_ia	1	45
stdcf	6	58	saleinv	1	46
sgr	5	30	chpmia	1	28
hire	5	30	pchgm_pchsale	1	44
lgr	5	30	dnoa	1	32
dnco	5	30	ill	1	12
ms	5	0	pchsale_pchrect	1	16
dcol	5	30	ob_a	1	20
pm	5	0	etr	1	21
QMJ	5	0	indmom	1	22
HXZ_IA	5	3	lfe	1	23
egr_hxz	5	3	chmom	1	26
egr	5	3	divo	1	31
dcoa	5	3	pchcapx_ia	1	33
roic	5	0	std_dolvol	1	43

Continued on next page

Table E.3 – continued from previous page

Anomaly factor	Cluster Size	Cluster	Anomaly factor	Cluster Size	Cluster
invest	5	9	chatoia	1	34
grltnoa	5	9	dncl	1	35
dpia	5	9	cinvest	1	36
grltnoa_hxz	5	9	mve_ia	1	37
dnca	5	9	LIQ_PS	1	39
CMA	5	3	noa	1	40
os	5	0	pchsale_pchxsga	1	41
ivg	4	8	LTR	1	42
roaq	4	10	gad	1	59

E.5 Sensitivity to the activation functions

Performance of latent factor models for a various autoencoder networks with linear, tanh, sigmoid, and ReLU activation functions on the 150 anomaly factors as we increase the number of latent factors in the model from 2 to 15. In each figure, the right column shows the average adjusted R^2 obtain across the 150 regressions, and the left and middle columns show the average percentage of anomaly factors that are explained by the latent factor model (at a t-statistic threshold of 1.96 and 5 respectively). The top row corresponds to ordinary least squares (OLS) regression results, the middle row to polynomial regression results, and the bottom row to split regressions.

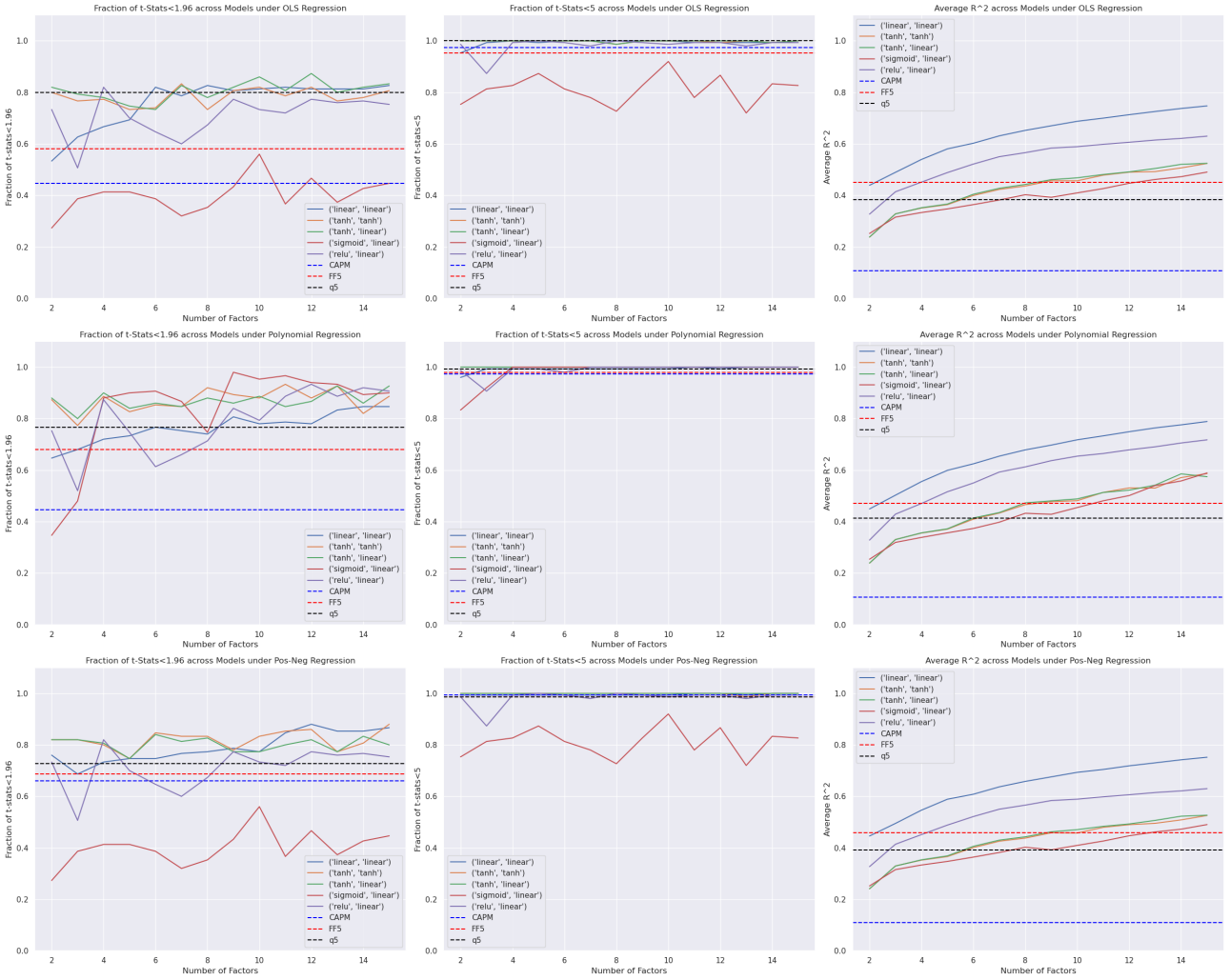


Figure E-8: Performance of **shallow AE** latent factor models with linear, tanh, sigmoid, and ReLU activation functions.

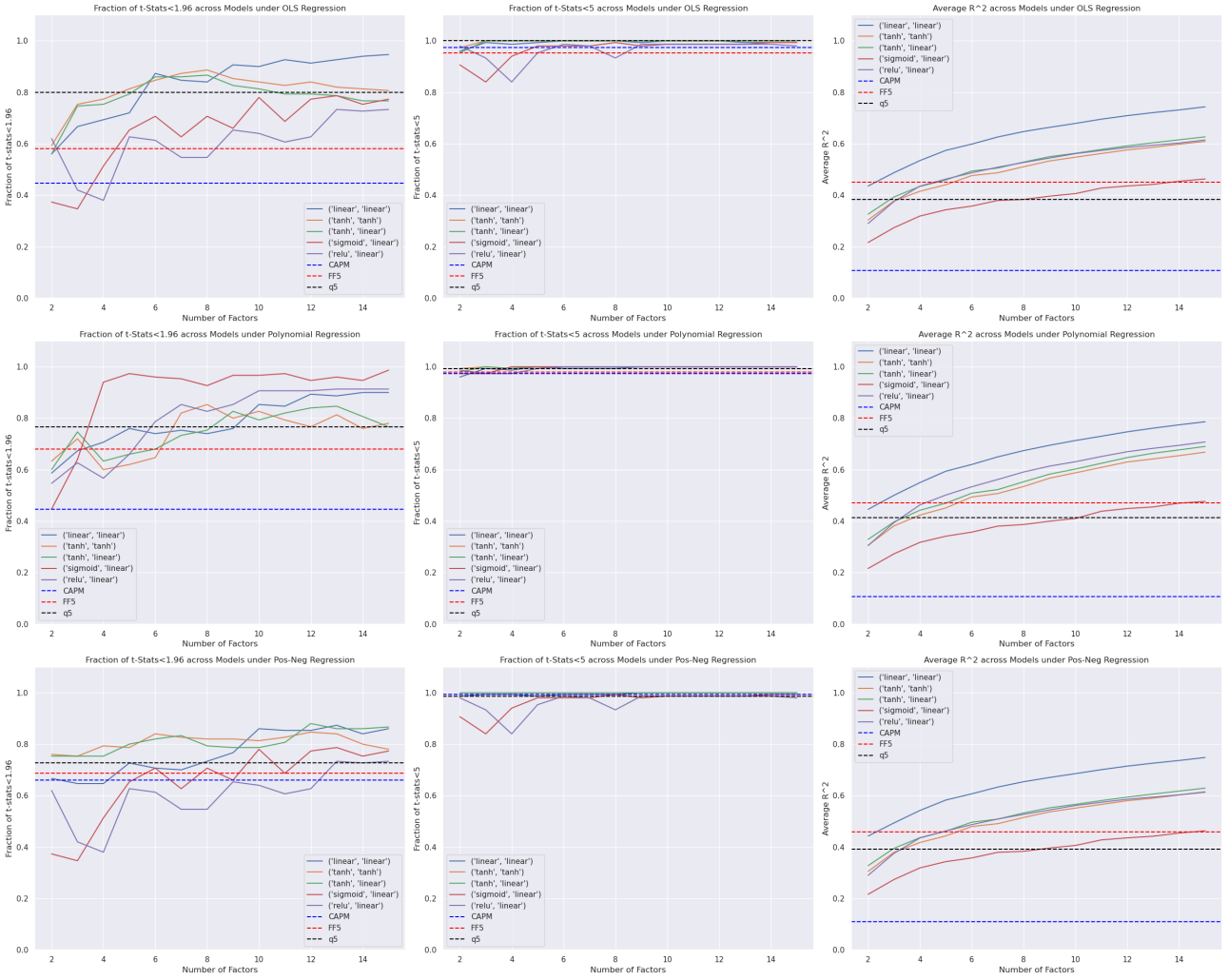


Figure E-9: Performance of **recursive shallow AE** models with linear, tanh, sigmoid, and ReLU activation functions.

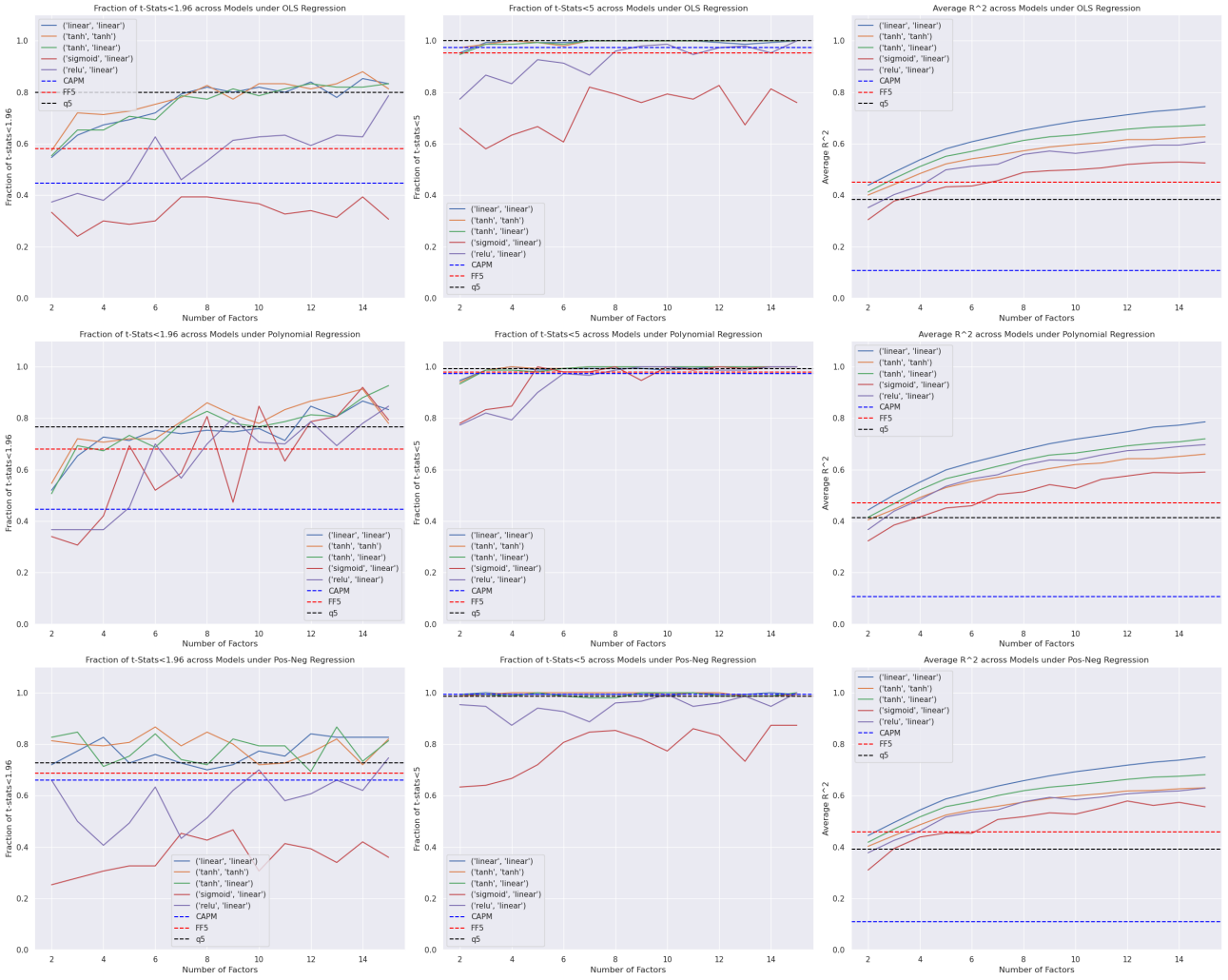


Figure E-10: Performance of **deep AE** latent factor models with linear, tanh, sigmoid, and ReLU activation functions.

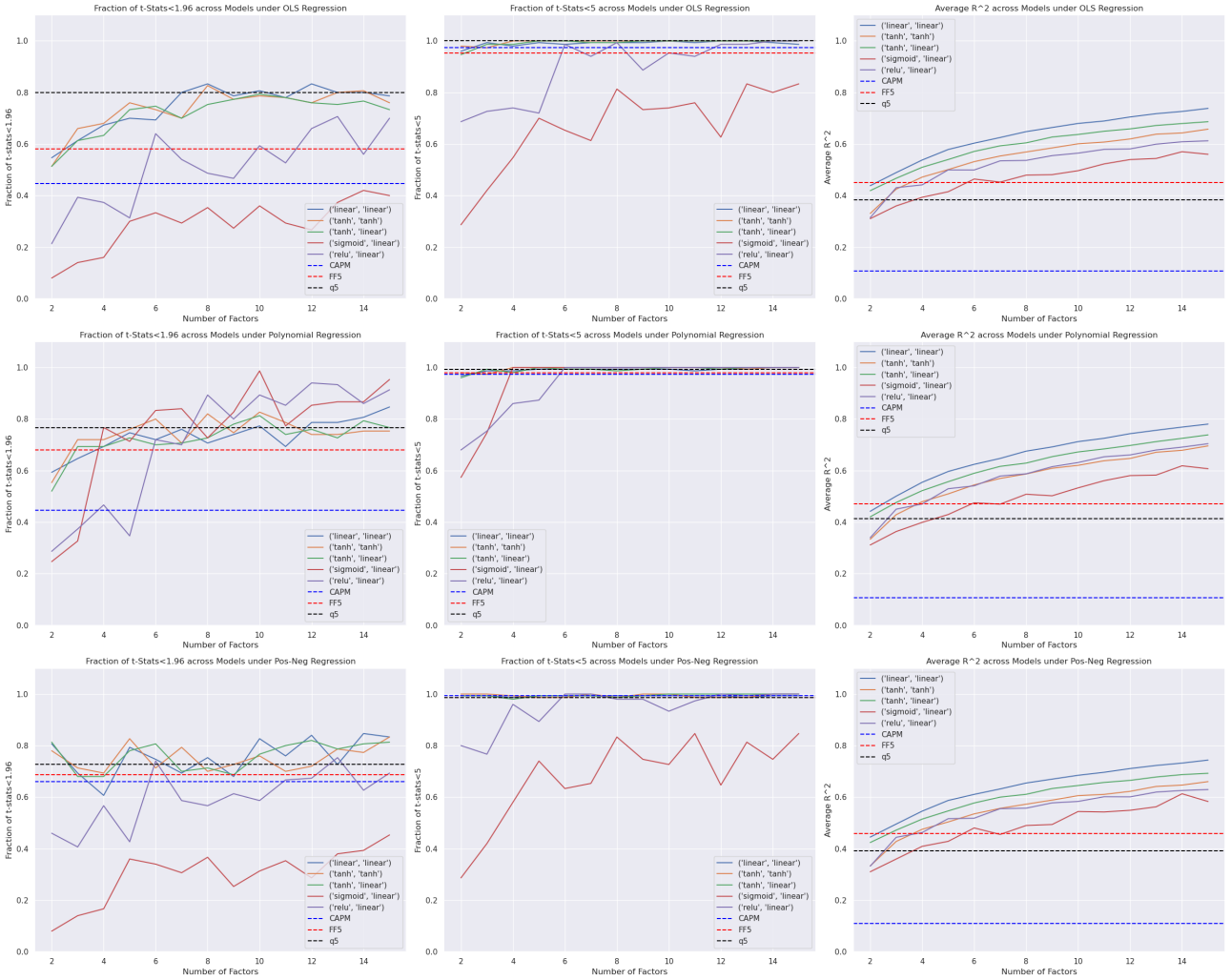


Figure E-11: Performance of **CAE** latent factor models with linear, tanh, sigmoid, and ReLU activation functions.

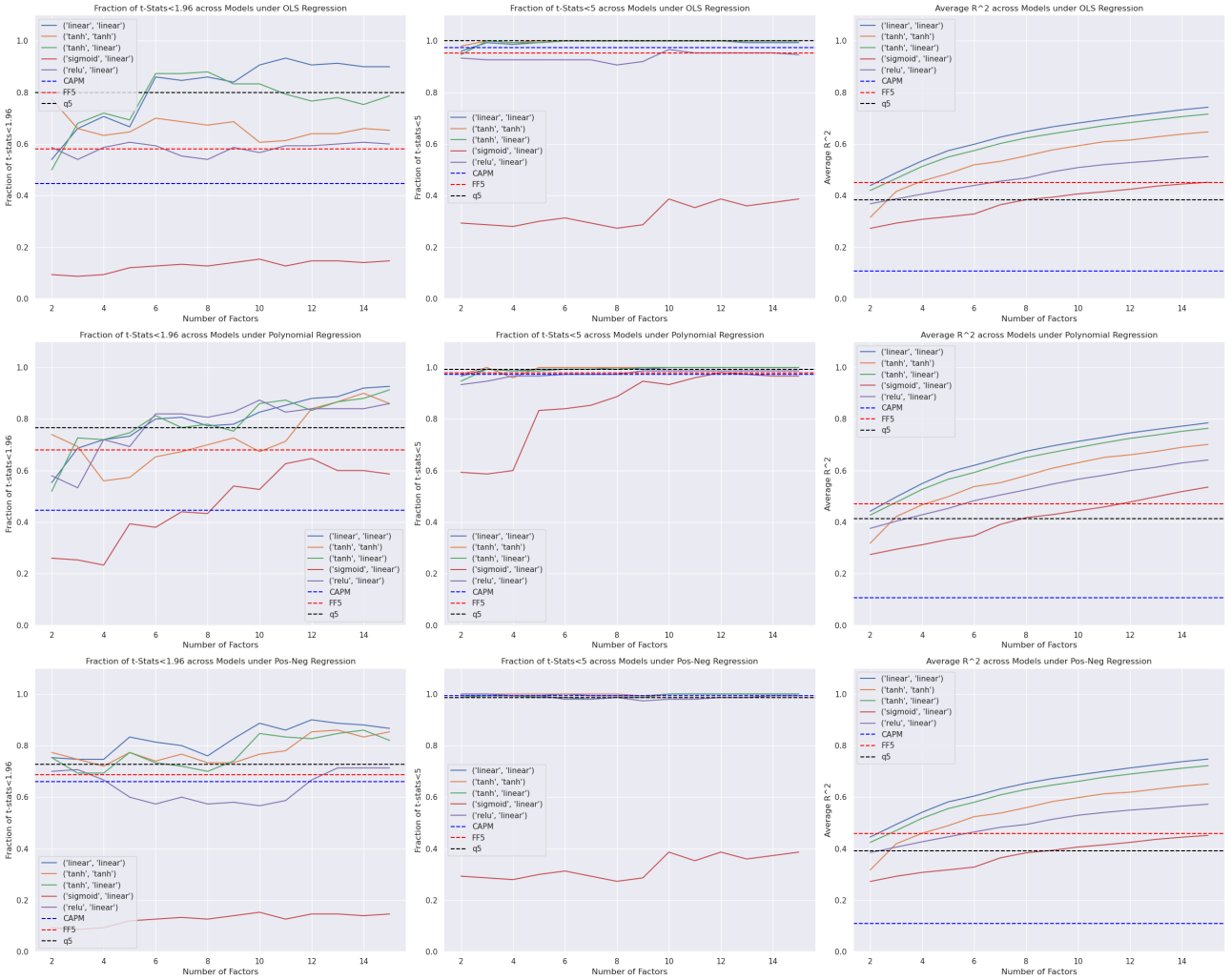


Figure E-12: Performance of **RCAE** latent factor models with linear, tanh, sigmoid, and ReLU activation functions.

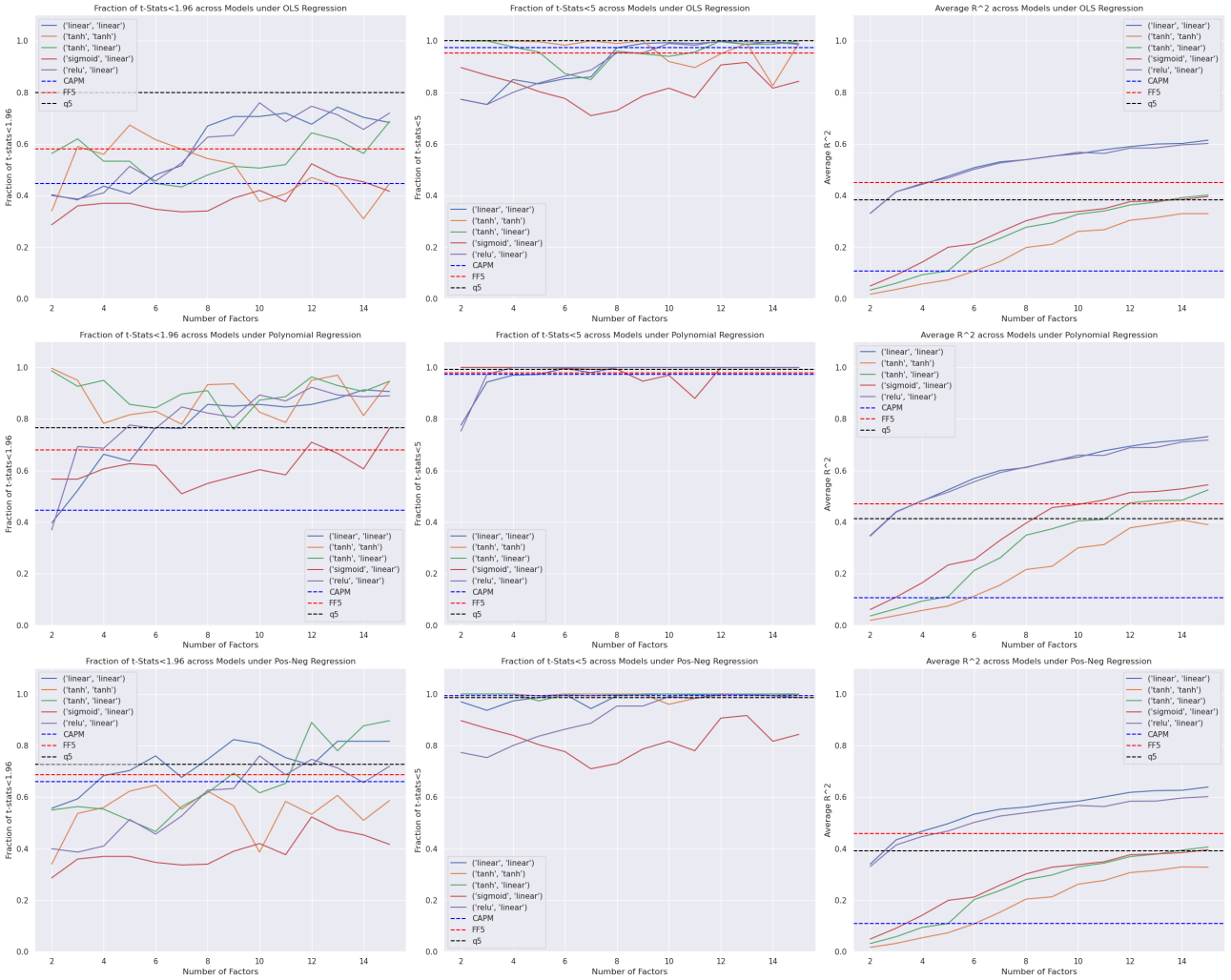


Figure E-13: Performance of **split AE** latent factor models with linear, tanh, sigmoid, and ReLU activation functions.

E.6 Polynomial regressions

We display here the performance of the latent factor models when using a polynomial regression of order 2 (which only contains linear and interaction terms, no second order terms). Tables E.4, E.5, E.6, and E.7 are related to Tables 7.1, 7.2, 7.3, and 7.4 in the main text.

Table E.4: Fraction of explained excess returns (in %). Brackets indicate the fraction obtained if we include a market factor.

Model	Factors	FF	HXZ
CAPM	50	65	48
FF5	63	80	57
q5	83	87	80
Linear 6	74 (75)	1 (65)	5 (84)
RLinear 5	75 (73)	88 (95)	69 (84)
RLinear 6	75 (71)	27 (81)	36 (80)
R-Linear 9	71 (74)	76 (96)	59 (82)
CAE 7	75 (77)	4 (88)	12 (78)
CAE 8	71 (70)	4 (87)	11 (80)
RCAE 5	73 (70)	99 (95)	87 (80)
RCAE 7	77 (79)	5 (81)	18 (81)
Deep 5	70 (65)	5 (81)	11 (76)
Deep 8	75 (75)	3 (95)	7 (73)
Tanh-Tanh 6	85 (84)	100 (99)	92 (90)
R-Tanh-Tanh 5	71 (82)	25 (79)	24 (83)
R-Tanh-Tanh 8	85 (87)	15 (95)	12 (89)
CAE 6 Tanh-Tanh	77 (77)	100 (91)	96 (82)
RCAE 4 Tanh-Tanh	70 (71)	80 (89)	63 (69)
Deep 5 Tanh-Tanh	76 (75)	21 (88)	28 (70)
Tanh 6	85 (85)	100 (99)	91 (88)
R-Tanh 4	67 (71)	97 (88)	83 (75)
CAE 6 Tanh	69 (70)	28 (49)	44 (81)
RCAE 5 Tanh	76 (71)	87 (97)	74 (84)
Deep 6 Tanh	69 (66)	5 (93)	4 (75)
ReLU 4	88 (84)	21 (44)	44 (79)
ReLU 9	82 (81)	88 (96)	85 (89)
R-ReLU 6	76 (73)	4 (59)	22 (78)
CAE 6 ReLU	75 (73)	100 (85)	97 (74)
RCAE 4 ReLU	72 (63)	16 (76)	34 (51)
Deep 5 ReLU	55 (55)	4 (41)	40 (71)
Sigmoid 4	81 (79)	100 (97)	100 (92)
R-Sigmoid 5	93 (93)	100 (77)	100 (97)

Continued on next page

Table E.4 – continued from previous page

Model	Factors	FF	HXZ
CAE 5 Sigmoid	81 (84)	71 (92)	32 (89)
RCAE 5 Sigmoid	60 (64)	32 (85)	7 (66)
Deep 5 Sigmoid	77 (75)	100 (100)	90 (82)

Table E.5: Regression adjusted R-squared (in %). Brackets indicate the fraction obtained if we include a market factor.

Model	Factors	FF	HXZ
CAPM	18	75	76
FF5	54	93	83
q5	46	90	83
Linear 6	67 (69)	63 (87)	61 (87)
RLinear 5	65 (67)	60 (86)	59 (86)
RLinear 6	67 (69)	63 (87)	61 (87)
R-Linear 9	72 (75)	66 (89)	66 (88)
CAE 7	69 (71)	63 (89)	62 (87)
CAE 8	71 (73)	65 (89)	64 (88)
RCAE 5	65 (67)	59 (86)	58 (86)
RCAE 7	69 (71)	64 (89)	63 (87)
Deep 5	65 (68)	62 (86)	60 (86)
Deep 8	71 (73)	66 (89)	65 (88)
Tanh-Tanh 6	46 (49)	50 (81)	51 (82)
R-Tanh-Tanh 5	51 (53)	58 (83)	56 (83)
R-Tanh-Tanh 8	58 (61)	61 (86)	59 (85)
CAE 6 Tanh-Tanh	59 (62)	57 (85)	56 (85)
RCAE 4 Tanh-Tanh	53 (56)	54 (83)	54 (84)
Deep 5 Tanh-Tanh	58 (61)	56 (85)	54 (85)
Tanh 6	46 (49)	50 (81)	51 (83)
R-Tanh 4	50 (53)	53 (83)	53 (83)
CAE 6 Tanh	64 (66)	59 (87)	58 (86)
RCAE 5 Tanh	62 (65)	57 (85)	57 (86)
Deep 6 Tanh	64 (67)	60 (87)	58 (86)
ReLU 4	53 (57)	44 (83)	48 (83)
ReLU 9	68 (70)	65 (88)	63 (88)
R-ReLU 6	59 (62)	52 (85)	54 (85)
CAE 6 ReLU	58 (62)	42 (84)	47 (85)
RCAE 4 ReLU	48 (54)	28 (82)	32 (82)
Deep 5 ReLU	58 (62)	48 (84)	51 (85)

Continued on next page

Table E.5 – continued from previous page

Model	Factors	FF	HXZ
Sigmoid 4	39 (43)	42 (80)	44 (81)
R-Sigmoid 5	39 (43)	40 (80)	41 (81)
CAE 5 Sigmoid	47 (51)	41 (80)	46 (82)
RCAE 5 Sigmoid	38 (43)	33 (78)	39 (80)
Deep 5 Sigmoid	50 (54)	43 (81)	49 (82)

Table E.6: Number of unexplained CAPM, FF5, and q5 factors. (*) indicates that one of the unexplained factors is the market factor.

Model	CAPM	FF5	q5
Linear 6	1*	1*	3*
RLinear 5	0	0	2
RLinear 6	1*	1*	3*
R-Linear 9	0	0	2
CAE 7	1*	1*	3*
CAE 8	1*	1*	3*
RCAE 5	0	0	2
RCAE 7	1*	1*	3*
Deep 5	1*	1*	3*
Deep 8	1*	1*	3*
Tanh-Tanh 6	0	0	1
R-Tanh-Tanh 5	1*	3*	2*
R-Tanh-Tanh 8	1*	2*	2*
CAE 6 Tanh-Tanh	0	0	1
RCAE 4 Tanh-Tanh	0	2	2
Deep 5 Tanh-Tanh	1*	2*	2*
Tanh 6	0	0	1
R-Tanh 4	0	2	1
CAE 6 Tanh	0	1	3
RCAE 5 Tanh	0	0	1
Deep 6 Tanh	1*	2*	3*
ReLU 4	1*	2*	4*
ReLU 9	0	0	0
R-ReLU 6	1*	1*	2*
CAE 6 ReLU	0	1	0
RCAE 4 ReLU	1*	2*	4*
Deep 5 ReLU	1*	3*	3*
Sigmoid 4	0	0	0

Continued on next page

Table E.6 – continued from previous page

Model	CAPM	FF5	q5
R-Sigmoid 5	0	1	1
CAE 5 Sigmoid	1*	1*	1*
RCAE 5 Sigmoid	1*	3*	3*
Deep 5 Sigmoid	0	1	2

Table E.7: Number of factors unexplained by CAPM, FF5, and q5.

Model	CAPM	FF5	q5
CAPM	—	1	0
FF5	3	—	0
q5	4	2	—
Linear 6	10	11	10
RLinear 5	5	4	4
RLinear 6	9	10	9
R-Linear 9	19	22	18
CAE 7	15	15	12
CAE 8	18	20	18
RCAE 5	8	8	9
RCAE 7	12	12	16
Deep 5	9	5	3
Deep 8	15	14	13
Tanh-Tanh 6	17	18	17
R-Tanh-Tanh 5	11	12	10
R-Tanh-Tanh 8	27	27	26
CAE 6 Tanh-Tanh	11	9	7
RCAE 4 Tanh-Tanh	8	8	7
Deep 5 Tanh-Tanh	5	4	4
Tanh 6	17	17	17
R-Tanh 4	8	9	6
CAE 6 Tanh	11	13	13
RCAE 5 Tanh	4	4	5
Deep 6 Tanh	9	8	8
ReLU 4	9	10	10
ReLU 9	43	43	41
R-ReLU 6	19	20	19
CAE 6 ReLU	10	10	10
RCAE 4 ReLU	4	5	5
Deep 5 ReLU	6	7	7

Continued on next page

Table E.7 – continued from previous page

Model	CAPM	FF5	q5
Sigmoid 4	10	10	10
R-Sigmoid 5	15	15	15
CAE 5 Sigmoid	15	15	15
RCAE 5 Sigmoid	15	15	15
Deep 5 Sigmoid	15	15	15

Appendix F

Chapter 8 Supplementary Material

In this appendix, we include the specific survey questions used for individual and institutional investors (Section F.1), and for financial advisors (Section F.2), and summarize the characteristics of the survey participants (Section F.3). In Section F.4, we provide a more detailed treatment of the risk-aversion coefficient calculation. Finally, in Section F.5, we show the year-by-year results for the asset allocation questions, and in Section F.6 we provide year-by-year results for the clustering analysis as well as a statistical significance analysis of the parameters reported.

F.1 Individual Investor and Institutional Survey Questions

The behavioral questions 1 to 3 presented in the Individual Investor and the Institutional Surveys were exactly the same. Question 4 was modified every year and was only asked to Individual Investors between 2015 and 2017.

1. Of the following six gambles, which would you prefer the most?
    a. Win $28,000 with probability 100%
    b. Win $36,000 with probability 50%; win $24,000 with probability 50%
    c. Win $44,000 with probability 50%; win $20,000 with probability 50%
    d. Win $52,000 with probability 50%; win $16,000 with probability 50%
    e. Win $60,000 with probability 50%; win $12,000 with probability 50%
    f. Win $70,000 with probability 50%; win $2,000 with probability 50%
2. How would you change your asset allocation if the S&P 500 declined between 10% and 20% during the next six months and other asset classes performed as you expected?
    a. I would do nothing
    b. I would decrease my stock or shares allocation slightly
    c. I would decrease my stock or shares allocation significantly

d. I would increase my stock or shares allocation slightly

e. I would increase my stock or shares allocation significantly

3. How would you change your asset allocation if the S&P 500 increased between 10% and 20% during the next six months and other asset classes performed as you expected?

   a. I would do nothing

   b. I would decrease my stock or shares allocation slightly

   c. I would decrease my stock or shares allocation significantly

   d. I would increase my stock or shares allocation slightly

   e. I would increase my stock or shares allocation significantly

4. (2015) Around what time during the Financial Crisis of 2007–2009 did you significantly decrease your allocation to stocks in your investment portfolio?

   a. Second half of 2007

   b. First half of 2008

   c. Second half of 2008

   d. First half of 2009

   e. Second half of 2009

   f. I did not significantly decrease my allocation to stocks during the Financial Crisis

4. (2016) In response to the recent market volatility in January of 2016, how did you change your equity allocation?

   a. I did nothing

   b. I decreased my stock or shares allocation slightly

   c. I decreased my stock or shares allocation significantly

   d. I increased my stock or shares allocation slightly

   e. I increased my stock or shares allocation significantly

4. (2017) In terms of investing, what did you do just after the U.S. presidential election on Nov. 8?

   a. I did nothing

   b. I decreased my stock or shares allocation slightly

   c. I decreased my stock or shares allocation significantly

   d. I increased my stock or shares allocation slightly

   e. I increased my stock or shares allocation significantly

F.2 Financial Advisor Survey Questions

The behavioral questions presented in the 2015 and 2016 Financial Advisor Surveys were essentially the same as the questions for individual investors. For questions pertaining to asset allocation, advisors were asked about how they would advise clients to change their allocations rather than how advisors would alter allocation in their personal investment portfolios.

1. Of the following six gambles, which would you prefer the most?
    a. Win $28,000 with probability 100%
    b. Win $36,000 with probability 50%; win $24,000 with probability 50%
    c. Win $44,000 with probability 50%; win $20,000 with probability 50%
    d. Win $52,000 with probability 50%; win $16,000 with probability 50%
    e. Win $60,000 with probability 50%; win $12,000 with probability 50%
    f. Win $70,000 with probability 50%; win $2,000 with probability 50%
2. How would you advise your clients to change their asset allocation if the S&P 500 declined between -10% and -20% during the next six months and other asset classes performed as expected?
    a. I would advise clients to do nothing
    b. I would advise clients to decrease equity allocation slightly
    c. I would advise clients to decrease equity allocation significantly
    d. I would advise clients to increase equity allocation slightly
    e. I would advise clients to increase equity allocation significantly
3. How would you advise your clients to change their asset allocation if the S&P 500 increased between -10% and -20% during the next six months and other asset classes performed as expected?
    a. I would advise clients to do nothing
    b. I would advise clients to decrease equity allocation slightly
    c. I would advise clients to decrease equity allocation significantly
    d. I would advise clients to increase equity allocation slightly
    e. I would advise clients to increase equity allocation significantly

F.3 Survey Respondents Characteristics

We provide more details on the subjects included in the surveys. Tables F.1–F.2 show the country breakdown of respondents for the Individual Investor Surveys between 2015 and 2017, while Tables F.4–F.5 do this for the Financial Advisor Surveys between 2015 and 2016. In Table F.6 we list the types of institutions included in the Institutional Investor surveys between 2015 and 2017. Finally, Table F.7 presents the definitions of the individual investor demographic categories.

Country	**# Respondents**	**Country**	**# Respondents**
Argentina	200	Japan	350
Australia	250	Mexico	350
Canada	250	Singapore	500
Chile	200	Spain	500
Colombia	200	Switzerland	350
France	500	UAE/Qatar/Kuwait	350
Germany	500	United Kingdom	750
Hong Kong	500	United States	750
Italy	500		

Table F.1: Number of respondents, by country, in the 2015 Individual Investor Survey.

Country	**# Respondents**	**Country**	**# Respondents**
Australia	300	Netherlands	250
Canada	300	Singapore	400
Chile	250	Spain	400
Colombia/Peru	300	Sweden	250
France	400	Switzerland	350
Germany	400	Taiwan	250
Hong Kong	500	UAE/Qatar/Kuwait	300
Italy	400	United Kingdom	750
Japan	350	United States	750
Mexico	300		

Table F.2: Number of respondents, by country, in the 2016 Individual Investor Survey.

Country	# Respondents	Country	# Respondents
Argentina/Uruguay	300	Korea	300
Australia	400	Mexico	300
Canada	300	Netherlands	300
Chile	300	Singapore	400
China	300	Spain	400
Colombia/Peru	300	Sweden	300
France	400	Switzerland	400
Germany	400	Taiwan	300
Hong Kong	400	UAE/Qatar/Kuwait	300
Italy	400	United Kingdom	750
Japan	300	United States	750

Table F.3: Number of respondents, by country, in the 2017 Individual Investor Survey.

Country	# Respondents	Country	# Respondents
Canada	150	Singapore	146
Chile	137	Spain	150
France	150	Switzerland	150
Germany	150	UAE/Qatar/Kuwait	144
Hong Kong	144	United Kingdom	300
Italy	150	Uruguay	140
Panama	131	United States	300

Table F.4: Number of respondents, by country, in the 2015 Financial Advisor Survey.

Country	# Respondents	Country	# Respondents
Canada	150	Singapore	150
Chile	150	Spain	150
Columbia	150	Switzerland	150
France	150	UAE/Qatar/Kuwait	150
Germany	150	United Kingdom	300
Hong Kong	150	Uruguay	150
Italy	150	United States	300
Panama	150		

Table F.5: Number of respondents, by country, in the 2016 Financial Advisor Survey.

Institution Type	2015	2016	2017
Central Bank	11		
Corporate Pension Plan	196	172	142
Insurance Company	100	112	107
Non-Profit (Endowment/Foundation)	131	146	100
Public/Government Pension Plan	140	133	116
Sovereign Wealth Fund	69	45	35
Other Institution	13	92	
Professional Buyer		200*	178
Total	660	700	700

Table F.6: Breakdown of the number of respondents by institution type, in the three Institutional Investor Surveys. * means that the number is already included in the rest of the Institution types.

Demographic	Definition
Generation Y	18–33 years old
Generation X	34–49 years old
Baby Boomers	50–68 years old
Pre-Baby Boomers	69 years old and above
Mass Market	NW: $200,000 - $300,000
Mass Affluent	NW: $300,000 - $500,000
Emerging HNW	NW: $500,000 - $1,000,000
High Net Worth	NW: $1,000,000 and above

Table F.7: Descriptions of the different demographic categories used in the Individual Investor Survey. The abbreviation *NW* means Net Worth, while *HNW* means High Net Worth.

F.4 Risk Aversion Calculation

We present here the calculation of the risk aversion coefficient in more detail. We assumed that respondents have constant relative risk aversion (CRRA) utility, and we want to estimate a group of respondent's risk aversion coefficient r and noise parameter μ. The expected utility of a subject considering a particular gamble i is given by:

$$E(U_i) = p_{i,1}U(x_{i,1}) + p_{i,2}U(x_{i,2}) \; ; \qquad U(x) = \frac{x^{1-r} - 1}{1-r} \tag{F.1}$$

where $x_{i,1}$, $x_{i,2}$ are the two possible payoffs for the gamble, and $p_{i,1}$, $p_{i,2}$ are their associated probabilities of occurring. In our case, $p_{i,1} = p_{i,2} = 1/2$, and the expected utility of a subject considering gamble i becomes:

$$E(U_i) = \frac{x_{i,1}^{1-r} + x_{i,2}^{1-r} - 2}{2(1-r)} \tag{F.2}$$

If $r = 0$, we assume $U(x) = \log(x)$ and $E(U_i) = \frac{1}{2}[\log(x_{i,1}) + \log(x_{i,2})]$. After calculating the expected utility, for each pair of gambles i, j, the subject picks gamble i with probability:

$$\mathbb{P}\left(\text{Choose gamble } i\right) = \frac{[E(U_i)]^{1/\mu}}{[E(U_i)]^{1/\mu} + [E(U_j)]^{1/\mu}} =: p_{i,j} \tag{F.3}$$

where μ is a noise parameter, since a subject may actually pick gamble j even if gamble i has higher expected utility. If we denote by η_i the percentage of respondents who preferred gamble i over gamble $i+1$, then the likelihood function for the parameters μ, r is proportional to $\prod_{i=1}^{5} p_{i,i+1}^{\eta_i} p_{i+1,1}^{1-\eta_i}$. Thus we are interested in maximizing the log-likelihood function $\sum_{i=1}^{5} \eta_i \log(p_{i,i+1}) + (1-\eta_i)\log(p_{i+1,i})$ over r and μ. For computational purposes, we also require r and μ to satisfy the following constraints:

$$\begin{cases} -20 \leq r \leq 20, \\ 0.001 \leq \mu \leq 20. \end{cases} \tag{F.4}$$

We now evaluate the inverse H^{-1} of the Hessian matrix. The standard error on r is estimated by $\sqrt{H^{-1}_{1,1}}$ and the standard error on μ is estimated by $\sqrt{H^{-1}_{2,2}}$.

F.5 S&P 500 Reactions and Risk Aversion

We expand here on the figures presented in the survey. In Figures F-1 and F-2, we present the reaction of individual investors, financial advisors and institutional investors to a 10%-20% decrease or increase in the S&P 500. We first display these results year by year, and then combine the three-years data for each group into a single plot. For reference, we provide the same plots for investors and advisors based only in the United States. In Figure F-3, we present an estimate of the risk aversion coefficient for individual investors across countries year-by-year between 2015 and

2017.

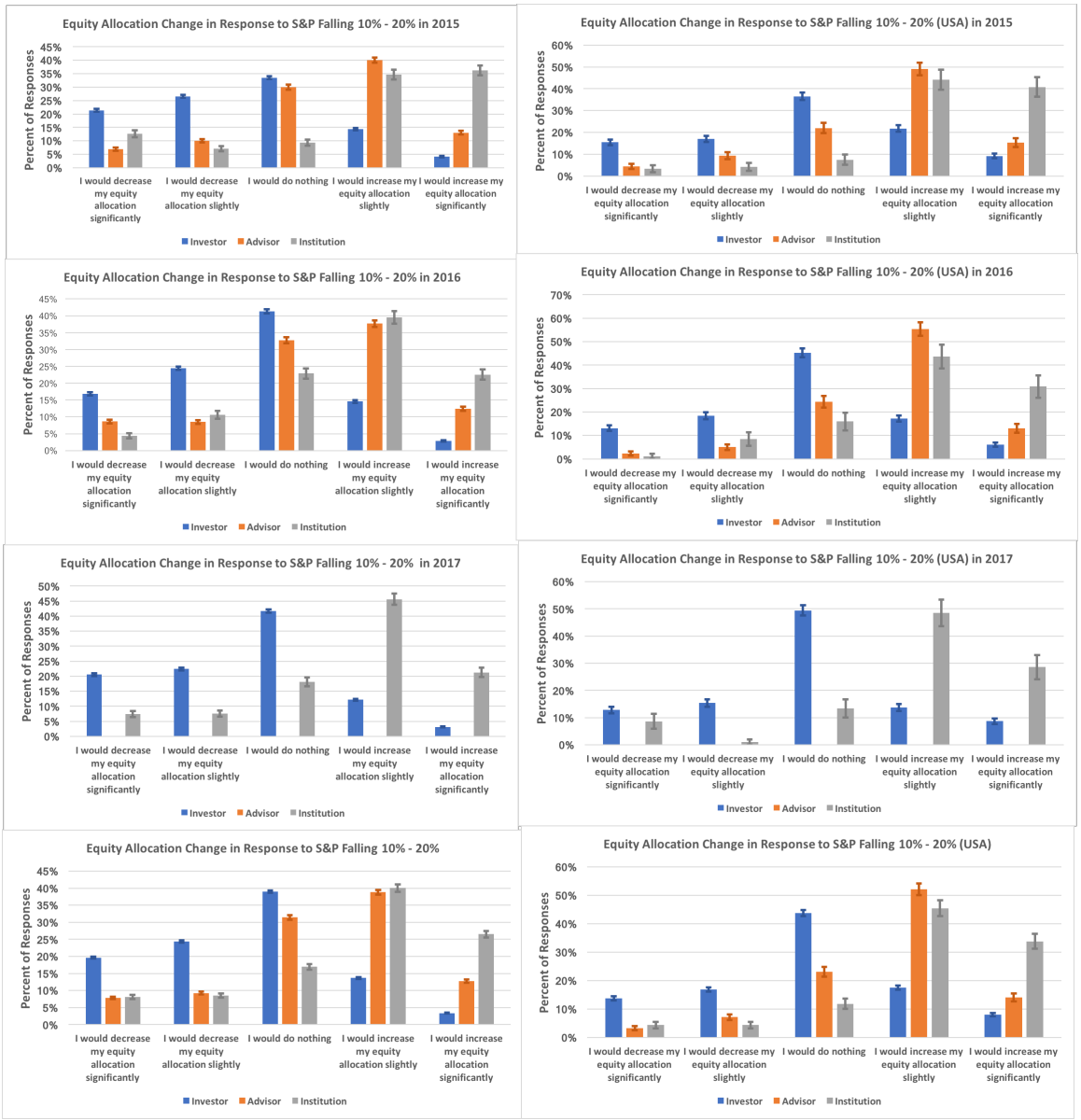


Figure F-1: Reactions to a decrease in the S&P 500 across three groups globally and in the U.S., in 2015, 2016 and 2017, and considering the three years as a single dataset. Financial Advisors were not surveyed in 2017. For each group and each possible answer, we show error bars corresponding to one standard error calculated assuming each respondent chooses either that particular answer, or any other answer.

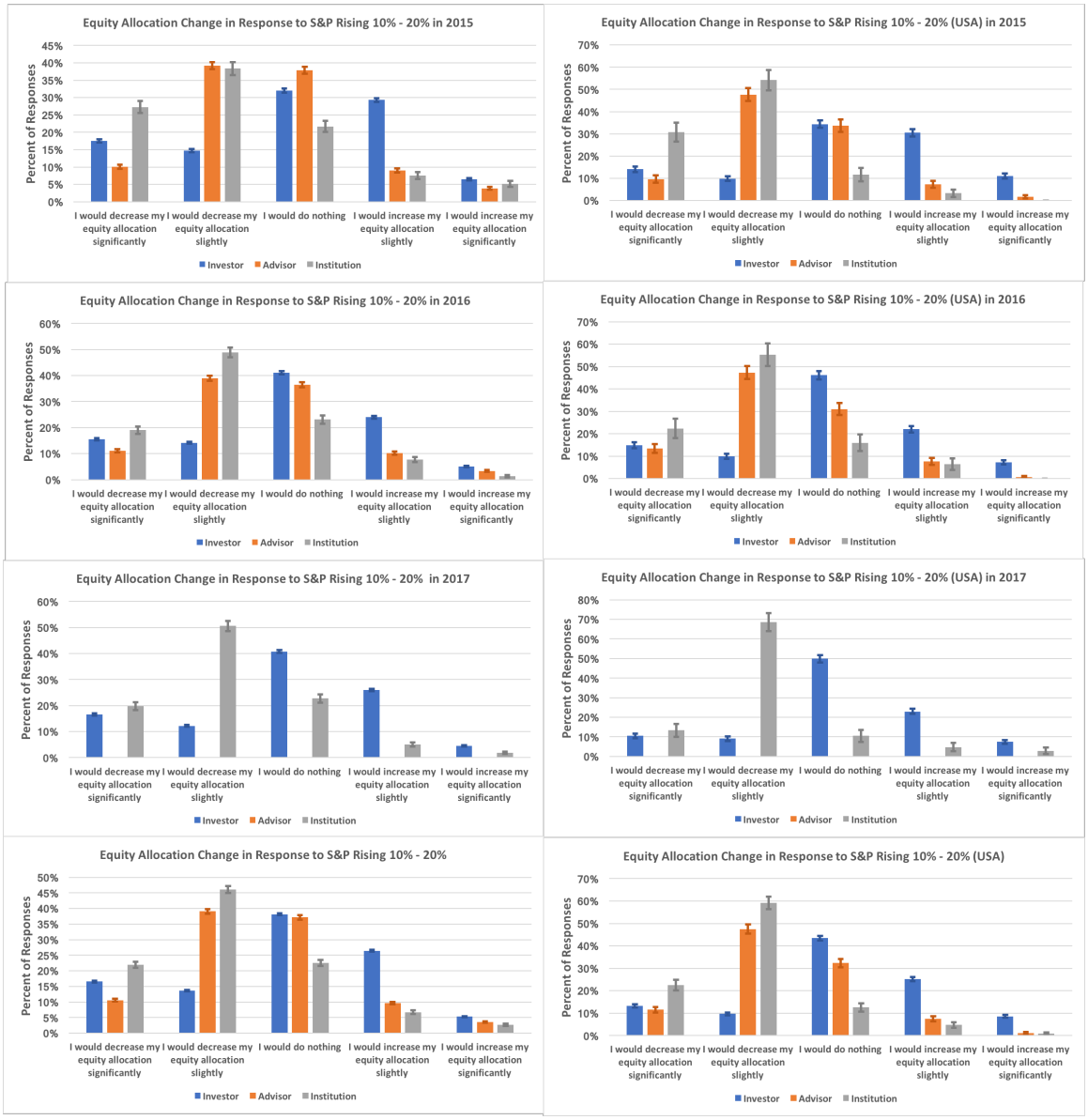


Figure F-2: Reactions to a increase in the S&P 500 across three groups globally and in the U.S., in 2015, 2016 and 2017, and considering the three years as a single dataset. Financial Advisors were not surveyed in 2017. For each group and each possible answer, we show error bars corresponding to one standard error calculated assuming each respondent chooses either that particular answer, or any other answer.

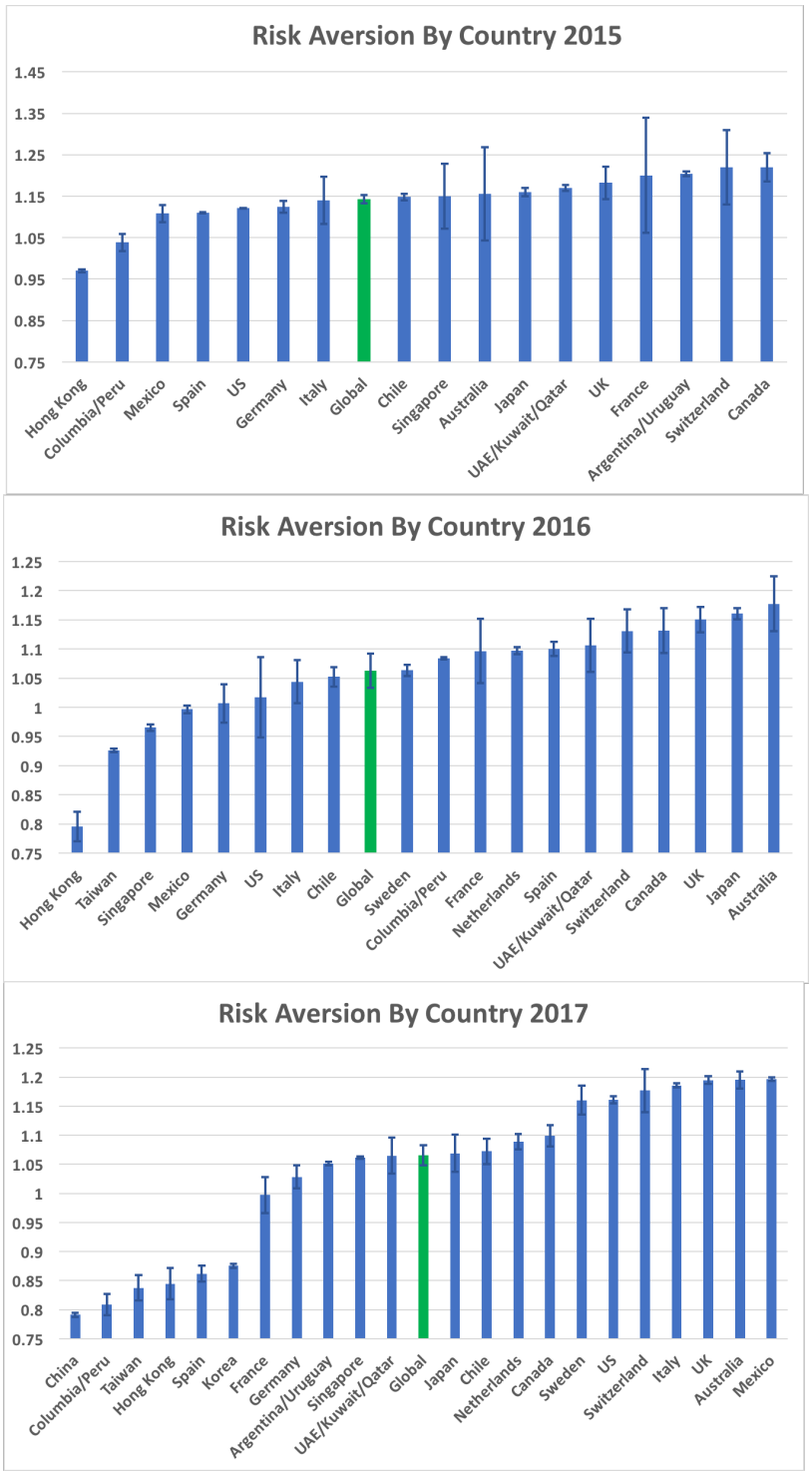


Figure F-3: Estimated risk aversion coefficients for individual investors across countries between 2015 and 2017. The risk aversion coefficients are sorted in increasing order. Error bars correspond to one standard error.

F.6 Individual Investor Clustering

We expand here on the results presented in the survey. In Tables F.8 to F.10, we summarize the results for the 2015, 2016 and 2017 Individual Investors clustering analysis. We combine the 2015–2017 results and plot the responses of each clusters in Figure F-4. We follow up on the latter plot by displaying the results of the 2015, 2016 and 2017 clusters separately in Figures F-5 to F-9. Tables F.11 to F.14 show the statistical significance when comparing the mean of demographic categories across all clusters of the combined 2015–2017 clustering analysis, as well as the 2015, 2016 and 2017 clustering analyses separately. Finally, Tables F.15 to F.16, and Tables F.17 to F.19 show the clustering results obtained in each year for Financial Advisors and Institutional advisors respectively.

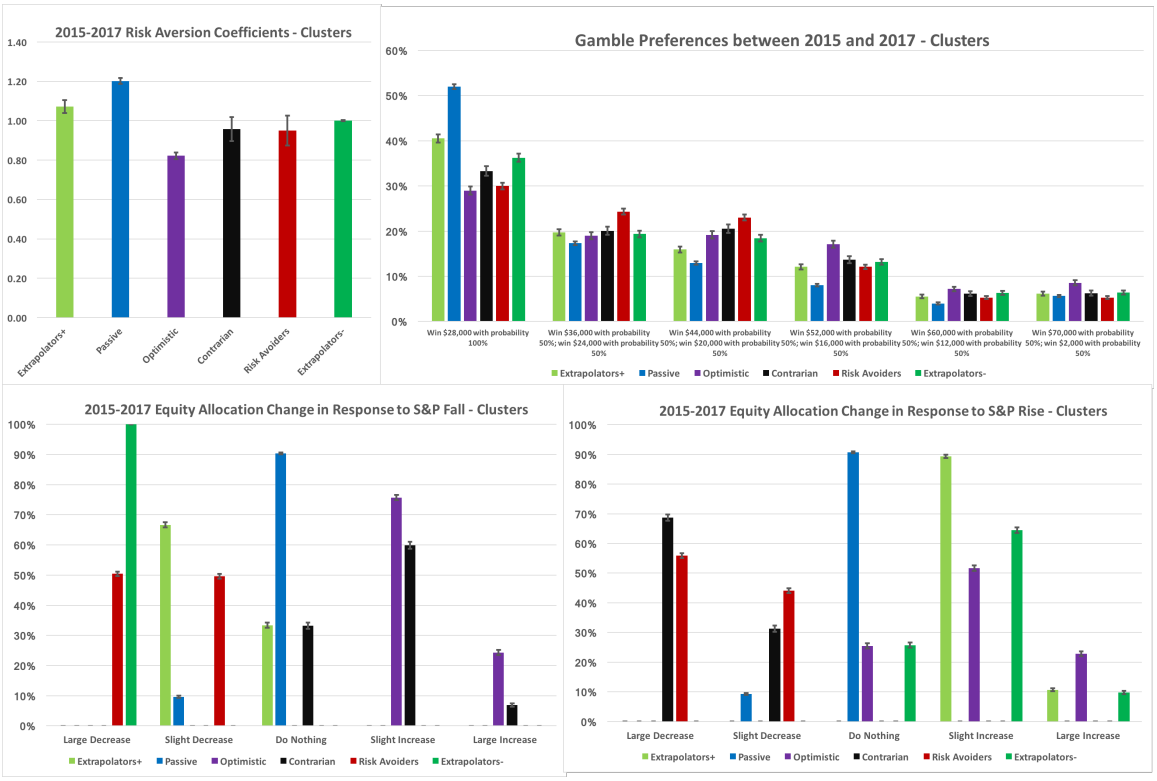


Figure F-4: Distributions of responses for the six clusters created from the 2015-2017 Individual Investor Surveys viewed as one dataset. Error bars correspond to one standard error. In the top left plot, we report the estimated risk-aversion coefficient calculated for each cluster. In the top right figure, we provide the gamble preferences of the individual investors across all groups (which was used to estimate the risk-aversion coefficients). Finally, the two lower figures show the distribution of responses for the six groups to the S&P 500 asset allocation questions.

Clustering Analysis of Individual Investors (2015)				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
% Respondents	51%	16%	14%	19%
	Allocation Decisions			
Reaction to Crisis	5.9	2.7	2.5	2.6
Reaction to S&P 500 Fall	−0.3	−1.4	−1.5	0.7
Reaction to S&P 500 Rise	0	−1.6	1	0.2
Behavior	Passive	Risk Avoiders	Extrapolators	Optimistic
	Demographics			
Gender (% of Female)	41%	42%	41%	41%
Age	1.13***	0.69***	0.86	0.8
Net Worth	1.64	1.62	1.63	1.72*
Advised	56%	57%	67%*	63%
Satisfied with 2014 Ret.	59%	44%***	55%	61%
Retired	9%***	2%	2%	3%
Gamble Preference	2.19***	2.64	2.47	2.62
Risk Aversion Coefficient	1.24***	0.98	1.08	1.01

Table F.8: Clustering of allocation decision responses from the 2015 Individual Investor Survey. For each cluster, we present the percent of respondents and the mean response based on the response coding in Table 8.3.
We also list the mean values of demographic categories across clusters. For *Age*, *Generation Y* = 0, *Generation X* = 1, *Baby Boomers* = 2, *Pre-Baby Boomers* = 3. For *Net Worth*, *Mass Market* = 0, *Mass Affluent* = 1, *Emerging HNW* = 2, *High Net Worth* = 3. The definitions of demographic categories are in Table F.7 in the Appendix. *Advised* is an indicator for if an investor uses a financial advisor. *Satisfied with 2014 Ret.* is an indicator for if an investor was satisfied with their 2014 investment returns. *Retired* is an indicator for if an investor is retired. *Gamble Preference* corresponds to the one of six gambles from Table 8.2 chosen by the investor; Responses range from 1 to 6. *Risk Aversion Coefficient* is the estimated risk aversion coefficient based on the responses in each cluster.
For each category, we color in green the cell corresponding to the cluster with the highest mean value. We test for how significant the difference is between the highest mean and second-highest mean across the clusters; the result of the test is reported in terms of number of stars in the cell. * means significance at the 5% level, *** means significance at the 0.1% level; no stars means no significance at the 5% level. We color in red the cell corresponding to the cluster with the lowest mean value, and perform the same test comparing the lowest mean and second-lowest mean across the clusters.

Clustering Analysis of Individual Investors (2016)					
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**
% Respondents	11%	22%	8%	21%	37%
	Allocation Decisions				
Reaction to January 2016	−1.6	−0.3	0.9	−0.7	0.2
Reaction to S&P 500 Fall	0.3	−1.7	1.1	−1.2	0.1
Reaction to S&P 500 Rise	0.3	0.7	1.1	−1.5	−0.2
Behavior	Semi-Passive	Extrapolators	Optimistic	Risk Avoiders	Passive
	Demographics				
Gender (% of Female)	35%	37%	37%	42%	41%
Age	43.68	44.09	43.35	42.75	46.58***
Net Worth	1.46	1.59	1.48	1.51	1.56
Advised	71%	64%	71%	69%	53%***
Satisfied with 2015 Ret.	54%	53%	60%*	56%	53%
Retired	6%	8%	6%	6%	12%***
Gamble Preference	2.48	2.49	2.81***	2.53	2.19***
Risk Aversion Coefficient	1.00	1.04	0.84***	0.96	1.18***

Table F.9: Clustering of allocation decision responses from the 2016 Individual Investor Surveys. For each cluster, we present the percent of respondents and the mean response based on the response coding in Table 8.3. We label as Semi-Passive investors, those who are passive with respect to the S&P movements, but active with respect to the January Volatility.

We also list the mean values of demographic categories across clusters. For *Net Worth*, *Mass Market* = 0, *Mass Affluent* = 1, *Emerging HNW* = 2, *High Net Worth* = 3. The definitions of demographic categories are in Table F.7 in the Appendix. *Advised* is an indicator for if an investor uses a financial advisor. *Satisfied with 2015 Ret.* is an indicator for if an investor was satisfied with their 2015 investment returns. *Retired* is an indicator for if an investor is retired. *Gamble Preference* corresponds to the one of six gambles from Table 8.2 chosen by the investor; Responses range from 1 to 6. *Risk Aversion Coefficient* is the estimated risk aversion coefficient based on the responses in each cluster.

For each category, we color in green the cell corresponding to the cluster with the highest mean value. We test for how significant the difference is between the highest mean and second-highest mean across the clusters; the result of the test is reported in terms of number of stars in the cell. * means significance at the 5% level, *** means significance at the 0.1% level; no stars means no significance at the 5% level. We color in red the cell corresponding to the cluster with the lowest mean value, and perform the same test comparing the lowest mean and second-lowest mean across the clusters.

Clustering Analysis of Individual Investors (2017)						
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**	**Cluster 5**	**Cluster 6**
% Respondents	12%	16%	36%	8%	9%	20%
	Allocation Decisions					
Reaction to U.S. Election	−0.2	−0.3	0.0	0.1	−1.5	0.2
Reaction to S&P 500 Fall	0.5	−1.5	0.1	1.3	−1.2	−1.4
Reaction to S&P 500 Rise	−1.5	−1.7	0.1	1.3	0.6	0.7
Behavior	Contrarian	Risk Avoiders	Passive	Optimistic	Extrapolators (A)	Extrapolators (P)
	Demographics					
Gender (% of Female)	39%*	44%	45%	43%	46%	44%
Age	45.57	44.58	49.58***	41.65***	44.51	46.51
Net Worth	1.48	1.41	1.37	1.51	1.60	1.45
Advised	64%	63%	55%***	75%	78%	66%
Retired	9%	6%	14%***	4%	5%	9%
From U.S.	5%***	6%	12%	18%***	7%	6%
Gamble Preference	2.52	2.50	2.18***	2.95***	2.61	2.43
Risk Aversion Coefficient	0.99	0.99	1.19*	0.73***	0.93	1.06

Table F.10: Clustering of allocation decision responses from the 2017 Individual Investor Survey. For each cluster, we present the percent of respondents and the mean response based on the response coding in Table 8.3. We label as Extrapolators (A) investors, those who are active with respect to the 2016 U.S. Presidential Election, and we label as Extrapolators (P) investors, those who are passive with respect to the 2016 U.S. Presidential Election.
We also list the mean values of demographic categories across clusters. For *Net Worth*, *Mass Market* = 0, *Mass Affluent* = 1, *Emerging HNW* = 2, *High Net Worth* = 3. The definitions of demographic categories are in Table F.7 in the Appendix. *Advised* is an indicator for if an investor uses a financial advisor. *Retired* is an indicator for if an investor is retired. *From U.S.* is an indicator for if an investor is from the United States. *Gamble Preference* corresponds to the one of six gambles from Table 8.2 chosen by the investor; Responses range from 1 to 6. *Risk Aversion Coefficient* is the estimated risk aversion coefficient based on the responses in each cluster.
For each category, we color in green the cell corresponding to the cluster with the highest mean value. We test for how significant the difference is between the highest mean and second-highest mean across the clusters; the result of the test is reported in terms of number of stars in the cell. * means significance at the 5% level, *** means significance at the 0.1% level; no stars means no significance at the 5% level. We color in red the cell corresponding to the cluster with the lowest mean value, and perform the same test comparing the lowest mean and second-lowest mean across the clusters.

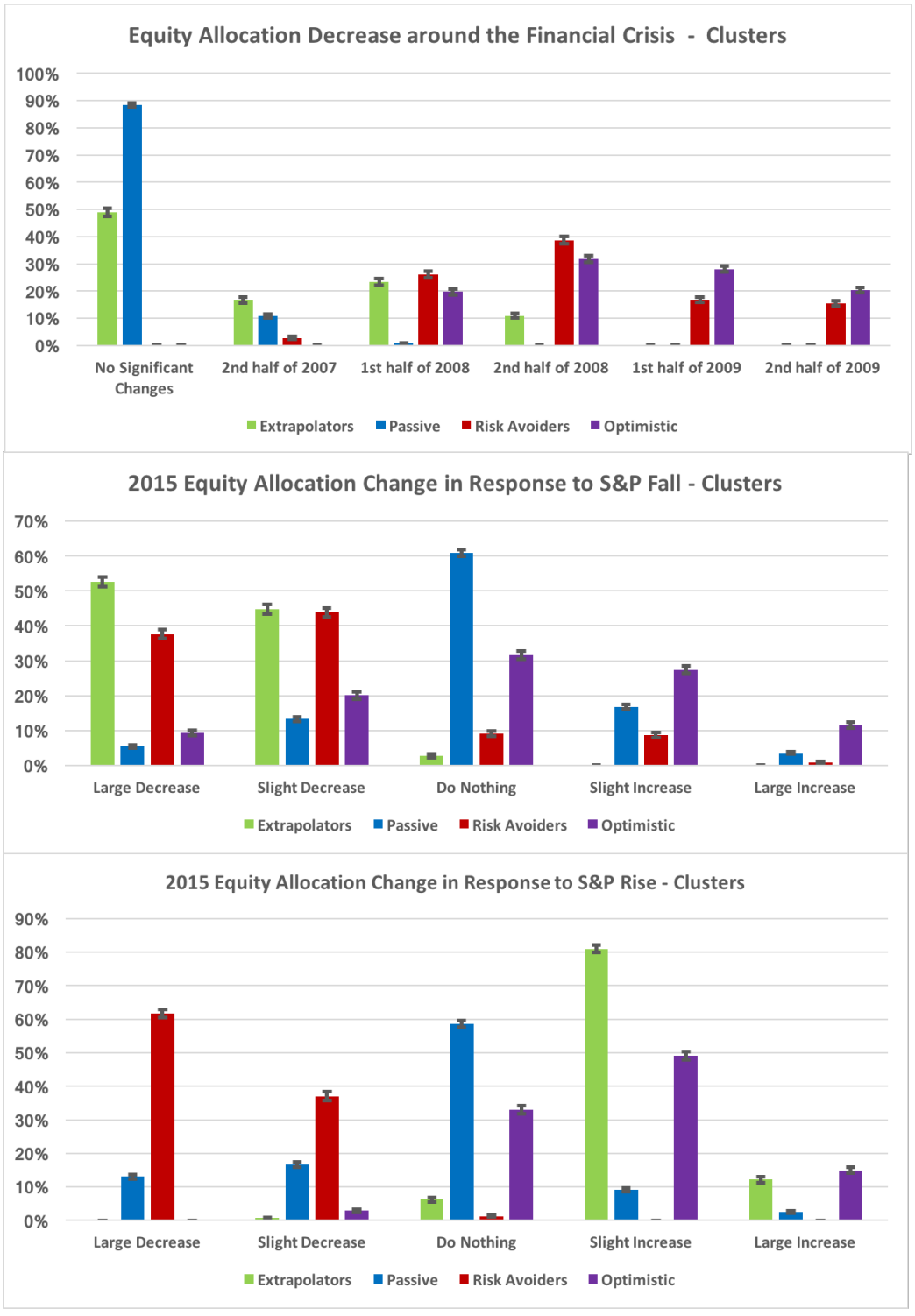


Figure F-5: Distributions of responses for the four clusters created from the 2015 Individual Investor Survey. Error bars correspond to one standard error.

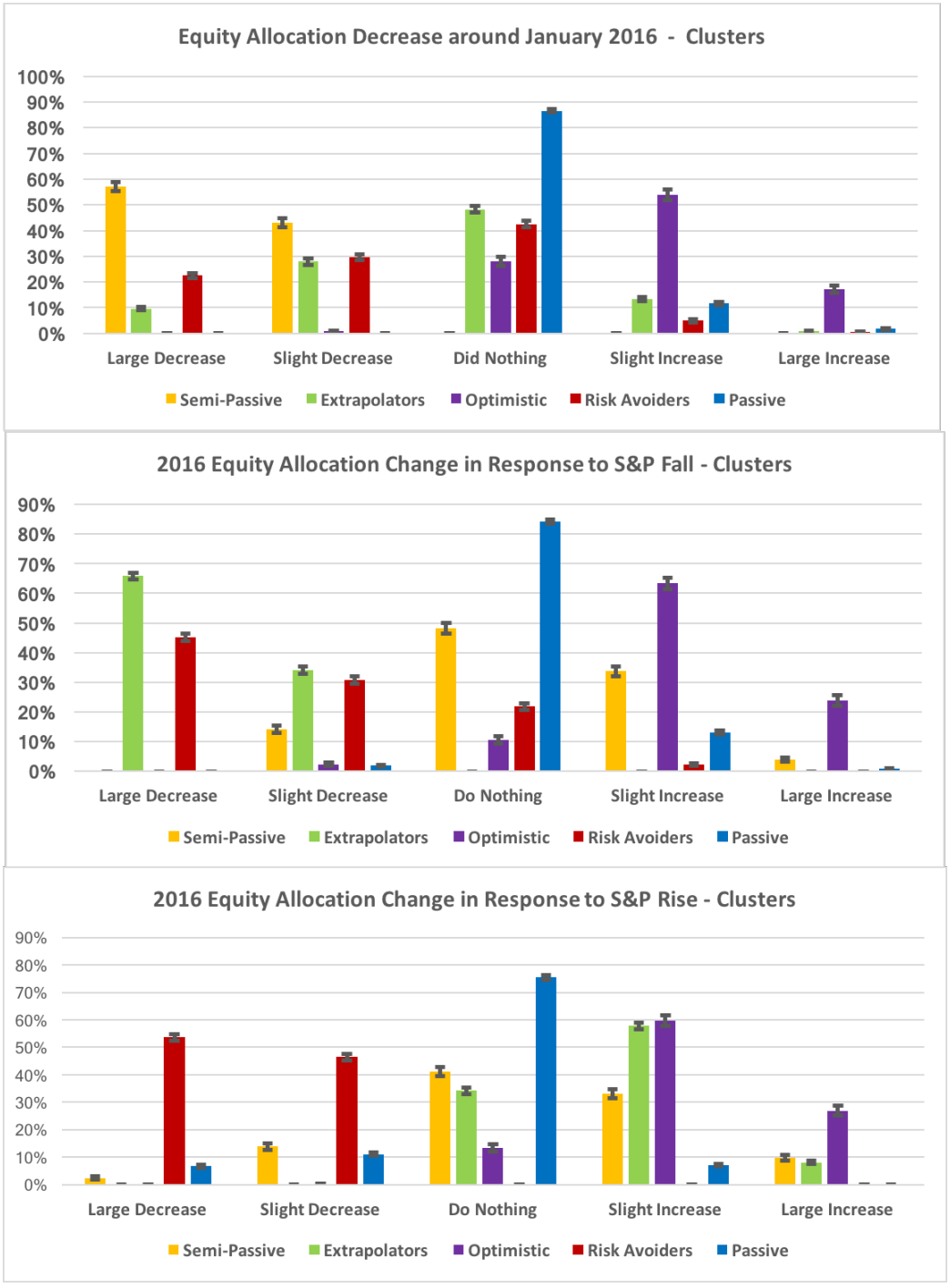


Figure F-6: Distributions of responses for the five clusters created from the 2016 Individual Investor Survey. Error bars correspond to one standard error.

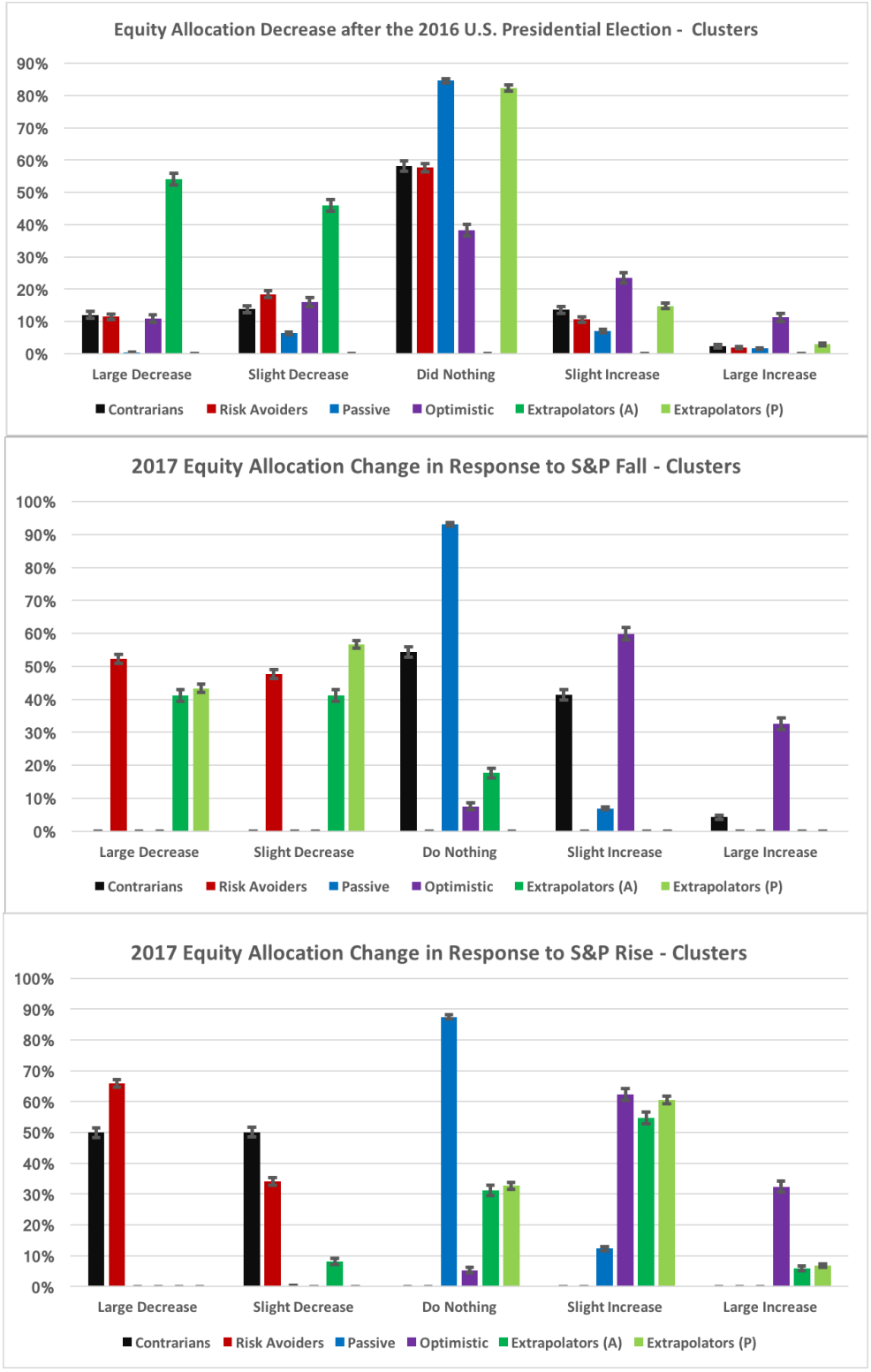


Figure F-7: Distributions of responses for the six clusters created from the 2017 Individual Investor Survey. Error bars correspond to one standard error.

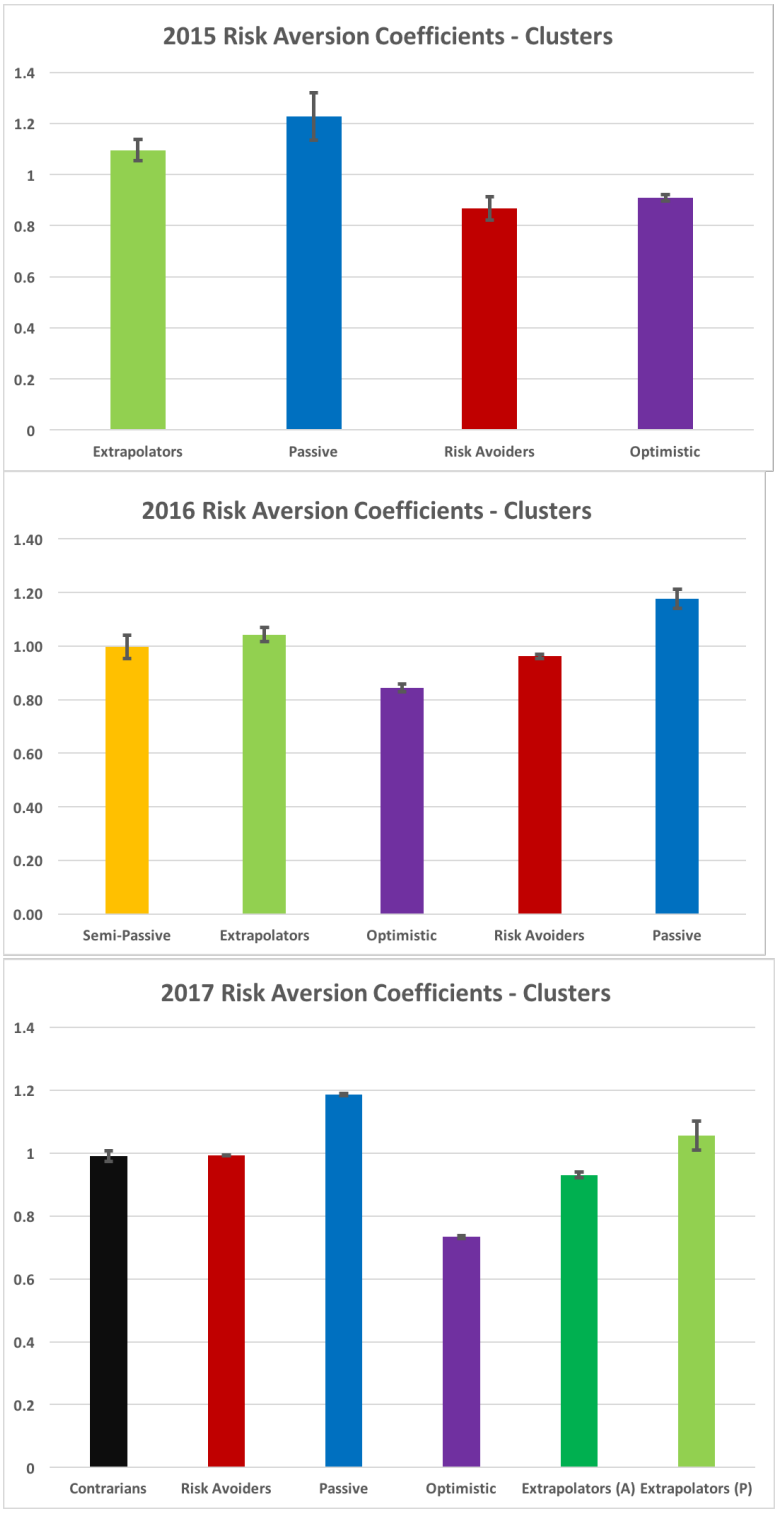


Figure F-8: Estimated risk aversion coefficients for individual investors between 2015 and 2017 across clusters. Error bars correspond to one standard error.

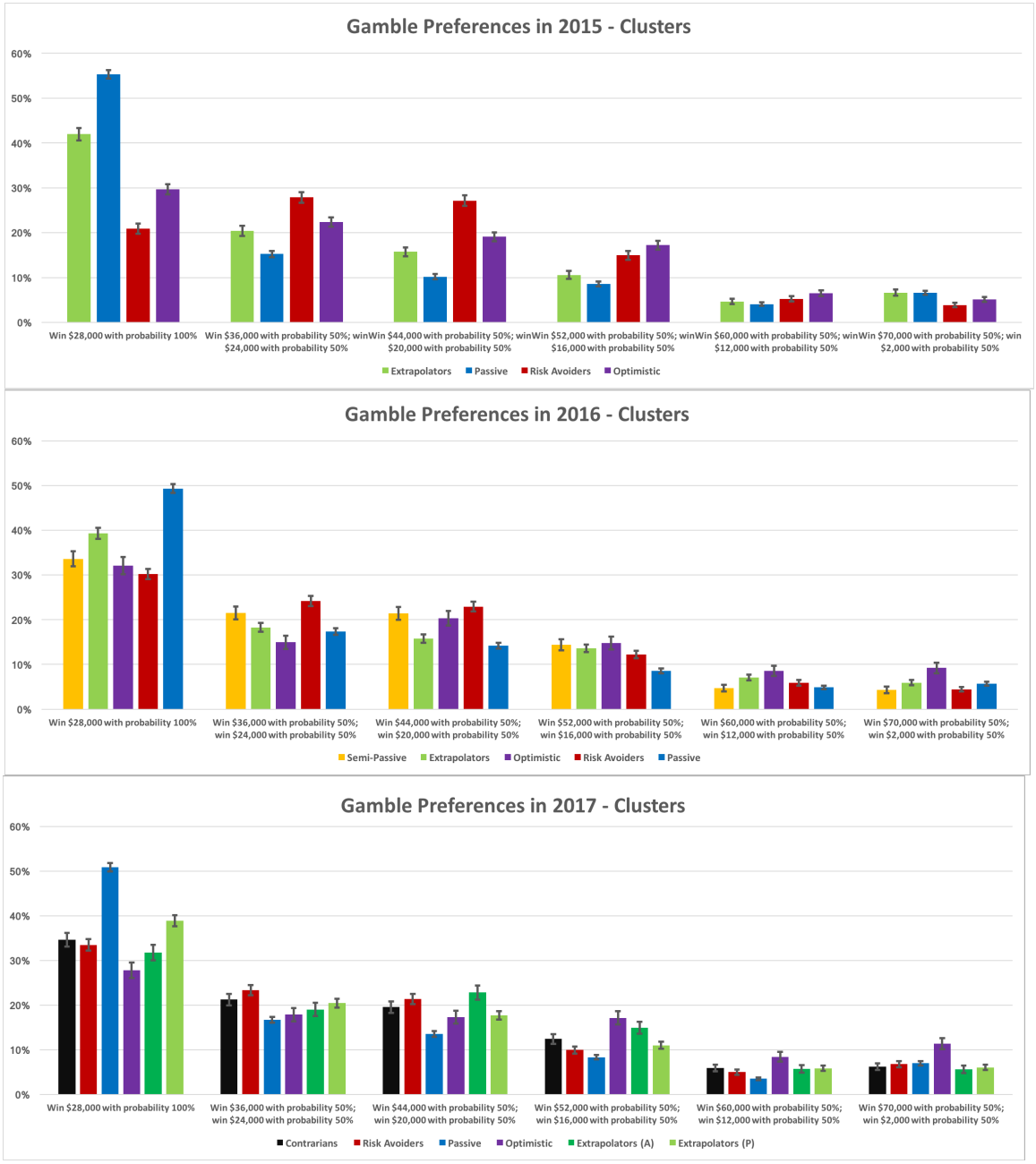


Figure F-9: Distributions of gamble preferences for individual investors between 2015 and 2017 across clusters. Error bars correspond to one standard error.

Statistical Significance for Comparisons of Mean Demographics across Clusters (2015-2017)

SP Fall	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	0.000	0.000	0.000	1.000	1.000
Passive	1.000	0.500	0.000	0.000	1.000	1.000
Opt.	1.000	1.000	0.500	1.000	1.000	1.000
Contr.	1.000	1.000	0.000	0.500	1.000	1.000
Risk Av.	0.000	0.000	0.000	0.000	0.500	1.000
Extrap.-	0.000	0.000	0.000	0.000	0.000	0.500

SP Rise	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	1.000	1.000	1.000	1.000	1.000
Passive	0.000	0.500	0.000	1.000	1.000	0.000
Opt.	0.000	1.000	0.500	1.000	1.000	1.000
Contr.	0.000	0.000	0.000	0.500	0.000	0.000
Risk Av.	0.000	0.000	0.000	1.000	0.500	0.000
Extrap.-	0.000	1.000	0.000	1.000	1.000	0.500

Gender	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	0.078	0.982	1.000	0.259	0.993
Passive	0.922	0.500	1.000	1.000	0.777	1.000
Opt.	0.018	0.000	0.500	1.000	0.002	0.621
Contr.	0.000	0.000	0.000	0.500	0.000	0.000
Risk Av.	0.741	0.223	0.998	1.000	0.500	0.999
Extrap.-	0.007	0.000	0.379	1.000	0.001	0.500

Age	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	0.000	1.000	0.998	1.000	0.997
Passive	1.000	0.500	1.000	1.000	1.000	1.000
Opt.	0.000	0.000	0.500	0.000	0.906	0.000
Contr.	0.002	0.000	1.000	0.500	1.000	0.315
Risk Av.	0.000	0.000	0.094	0.000	0.500	0.000
Extrap.-	0.003	0.000	1.000	0.685	1.000	0.500

Net Worth	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	0.808	0.000	0.368	0.785	0.170
Passive	0.192	0.500	0.000	0.134	0.505	0.023
Opt.	1.000	1.000	0.500	0.998	1.000	0.995
Contr.	0.632	0.866	0.002	0.500	0.849	0.307
Risk Av.	0.215	0.495	0.000	0.151	0.500	0.036
Extrap.-	0.830	0.977	0.005	0.693	0.964	0.500

Advised	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	1.000	0.000	0.955	0.991	0.030
Passive	0.000	0.500	0.000	0.000	0.000	0.000
Opt.	1.000	1.000	0.500	1.000	1.000	0.997
Contr.	0.045	1.000	0.000	0.500	0.597	0.000
Risk Av.	0.009	1.000	0.000	0.403	0.500	0.000
Extrap.-	0.970	1.000	0.003	1.000	1.000	0.500

Retired	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	0.000	1.000	0.973	1.000	0.991
Passive	1.000	0.500	1.000	1.000	1.000	1.000
Opt.	0.000	0.000	0.500	0.057	0.797	0.060
Contr.	0.027	0.000	0.943	0.500	0.992	0.582
Risk Av.	0.000	0.000	0.203	0.008	0.500	0.005
Extrap.-	0.009	0.000	0.940	0.418	0.995	0.500

US	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	0.000	0.000	0.252	0.398	0.915
Passive	1.000	0.500	0.000	1.000	1.000	1.000
Opt.	1.000	1.000	0.500	1.000	1.000	1.000
Contr.	0.748	0.000	0.000	0.500	0.685	0.966
Risk Av.	0.602	0.000	0.000	0.315	0.500	0.957
Extrapolators-	0.085	0.000	0.000	0.034	0.043	0.500

Gamble Pref.	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	1.000	0.000	0.000	0.000	0.001
Passive	0.000	0.500	0.000	0.000	0.000	0.000
Opt.	1.000	1.000	0.500	1.000	1.000	1.000
Contr.	1.000	1.000	0.000	0.500	0.820	0.858
Risk Av.	1.000	1.000	0.000	0.180	0.500	0.616
Extrap.-	0.999	1.000	0.000	0.142	0.384	0.500

Risk Aversion	Extrap.+	Passive	Opt.	Contr.	Risk Av.	Extrap.-
Extrap.+	0.500	1.000	0.000	0.000	0.000	0.073
Passive	0.000	0.500	0.000	0.000	0.000	0.000
Opt.	1.000	1.000	0.500	1.000	1.000	1.000
Contr.	1.000	1.000	0.000	0.500	0.240	0.825
Risk Av.	1.000	1.000	0.000	0.760	0.500	0.858
Extrap.-	0.927	1.000	0.000	0.175	0.142	0.500

Table F.11: Hypothesis tests for statistical significance when comparing means of demographic categories across the six clusters from the 2015-2017 Individual Investor Surveys. We assume that each cluster corresponds to an independent collection of investors. For every demographic, we calculate the mean response within each cluster. The resulting means are shown in Table 8.4.

For every two means, we test for statistical significance using Welch's t-test—with the exception of the risk aversion coefficient, where we employ a z-test using the risk aversion coefficients and associated standard errors from the estimation. Each cell contains the p-value associated with testing if the column cluster mean minus the row cluster mean is greater than zero. A cell is colored green if the mean in the row cluster is significantly less than the mean in the column cluster, at the 10% significance level. A cell is colored red if the mean in the row cluster is significantly greater than the mean in the column cluster, at the 10% significance level.

Extrap. is the *Extrapolators* cluster. *Extrap.+* and *Extrap.-* correspond respectively to the clusters of *Extrapolators* who react more to a rise or a decline in the S&P 500. *Risk Av.* denotes the *Risk Avoiders* cluster; *Opt.* denotes the *Optimistic* cluster; *Contr.* denotes the *Contrarians* cluster;

Risk Aversion is short for the risk aversion coefficient. *US* corresponds to the percentage of U.S. investors in the considered cluster.

Statistical Significance for Comparisons of Mean Demographics across Clusters (2015)

Fin. Crisis	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	1.000	1.000	1.000
Risk Av.	0.000	0.500	1.000	1.000
Extrap.	0.000	0.000	0.500	0.000
Optimistic	0.000	0.000	1.000	0.500

SP Fall	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	1.000	1.000	0.000
Risk Av.	0.000	0.500	1.000	0.000
Extrap.	0.000	0.000	0.500	0.000
Optimistic	1.000	1.000	1.000	0.500

SP Rise	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	1.000	0.000	0.000
Risk Av.	0.000	0.500	0.000	0.000
Extrap.	1.000	1.000	0.500	1.000
Optimistic	1.000	1.000	0.000	0.500

Gender	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	0.245	0.496	0.456
Risk Av.	0.755	0.500	0.703	0.691
Extrap.	0.504	0.297	0.500	0.470
Opt.	0.544	0.309	0.530	0.500

Satisfied 2014	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	1.000	0.974	0.064
Risk Av.	0.000	0.500	0.000	0.000
Extrap.	0.026	1.000	0.500	0.002
Opt.	0.936	1.000	0.998	0.500

Age	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	1.000	1.000	1.000
Risk Av.	0.000	0.500	0.000	0.000
Extrap.	0.000	1.000	0.500	0.977
Opt.	0.000	1.000	0.023	0.500

Retired	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	1.000	1.000	1.000
Risk Av.	0.000	0.500	0.773	0.026
Extrap.	0.000	0.227	0.500	0.004
Opt.	0.000	0.974	0.996	0.500

Net Worth	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	0.635	0.615	0.010
Risk Av.	0.365	0.500	0.488	0.017
Extrap.	0.385	0.512	0.500	0.022
Opt.	0.990	0.983	0.978	0.500

Gamble Pref.	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	0.000	0.000	0.000
Risk Av.	1.000	0.500	0.998	0.626
Extrap.	1.000	0.002	0.500	0.005
Opt.	1.000	0.374	0.995	0.500

Advised	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	0.188	0.000	0.000
Risk Av.	0.812	0.500	0.000	0.003
Extrap.	1.000	1.000	0.500	0.986
Opt.	1.000	0.997	0.014	0.500

Risk Aversion	Passive	Risk Av.	Extrap.	Opt.
Passive	0.500	0.000	0.000	0.000
Risk Av.	1.000	0.500	1.000	0.863
Extrap.	1.000	0.000	0.500	0.005
Opt.	1.000	0.137	0.995	0.500

Table F.12: Hypothesis tests for statistical significance when comparing means of demographic categories across the four clusters from the Individual Investor Survey. We assume that each cluster corresponds to an independent collection of investors. For every demographic, we calculate the mean response within each cluster. The resulting means are shown in Table F.8.

For every two means, we test for statistical significance using Welch's t-test—with the exception of the risk aversion coefficient, where we employ a z-test using the risk aversion coefficients and associated standard errors from the estimation. Each cell contains the p-value associated with testing if the column cluster mean minus the row cluster mean is greater than zero. A cell is colored green if the mean in the row cluster is significantly less than the mean in the column cluster, at the 10% significance level. A cell is colored red if the mean in the row cluster is significantly greater than the mean in the column cluster, at the 10% significance level.

Extrap. is the *Extrapolators* cluster. *Risk Av.* denotes the *Risk Avoiders* cluster; *Opt.* denotes the *Optimistic* cluster;

Risk Aversion is short for the risk aversion coefficient. *Satisfied 2014* is an indicator for if an investor was satisfied with their 2014 investment returns.

Statistical Significance for Comparisons of Mean Demographics across Clusters (2016)

Jan. 2016	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.000	0.000	0.000	0.000
Extrap.	1.000	0.500	0.000	1.000	0.000
Opt.	1.000	1.000	0.500	1.000	1.000
Risk Av.	1.000	0.000	0.000	0.500	0.000
Passive	1.000	1.000	0.000	1.000	0.500

SP Fall	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	1.000	0.000	1.000	1.000
Extrap.	0.000	0.500	0.000	0.000	0.000
Opt.	1.000	1.000	0.500	1.000	1.000
Risk Av.	0.000	1.000	0.000	0.500	0.000
Passive	0.000	1.000	0.000	1.000	0.500

SP Rise	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.000	0.000	1.000	1.000
Extrap.	1.000	0.500	0.000	1.000	1.000
Opt.	1.000	1.000	0.500	1.000	1.000
Risk Av.	0.000	0.000	0.000	0.500	0.000
Passive	0.000	0.000	0.000	1.000	0.500

Gender	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.165	0.220	0.001	0.001
Extrap.	0.835	0.500	0.503	0.004	0.004
Opt.	0.780	0.497	0.500	0.024	0.031
Risk Av.	0.999	0.996	0.976	0.500	0.634
Passive	0.999	0.996	0.969	0.366	0.500

Age	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.254	0.672	0.935	0.000
Extrap.	0.746	0.500	0.873	0.996	0.000
Opt.	0.328	0.127	0.500	0.827	0.000
Risk Av.	0.065	0.004	0.173	0.500	0.000
Passive	1.000	1.000	1.000	1.000	0.500

Net Worth	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.004	0.375	0.155	0.010
Extrap.	0.996	0.500	0.979	0.977	0.752
Opt.	0.625	0.021	0.500	0.289	0.045
Risk Av.	0.845	0.023	0.711	0.500	0.059
Passive	0.990	0.248	0.955	0.941	0.500

Advised	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	1.000	0.569	0.859	1.000
Extrap.	0.000	0.500	0.001	0.001	1.000
Opt.	0.431	0.999	0.500	0.780	1.000
Risk Av.	0.141	0.999	0.220	0.500	1.000
Passive	0.000	0.000	0.000	0.000	0.500

Satisfied 2015	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.777	0.018	0.224	0.664
Extrap.	0.223	0.500	0.001	0.034	0.310
Opt.	0.982	0.999	0.500	0.953	0.998
Risk Av.	0.776	0.966	0.047	0.500	0.940
Passive	0.336	0.690	0.002	0.060	0.500

Retired	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.026	0.344	0.410	0.000
Extrap.	0.974	0.500	0.901	0.978	0.000
Opt.	0.656	0.099	0.500	0.596	0.000
Risk Av.	0.590	0.022	0.404	0.500	0.000
Passive	1.000	1.000	1.000	1.000	0.500

Gamble Pref.	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.459	0.000	0.214	1.000
Extrap.	0.541	0.500	0.000	0.215	1.000
Opt.	1.000	1.000	0.500	1.000	1.000
Risk Av.	0.786	0.785	0.000	0.500	1.000
Passive	0.000	0.000	0.000	0.000	0.500

Risk Aversion	Semi-Pass.	Extrap.	Opt.	Risk Av.	Passive
Semi-Pass.	0.500	0.817	0.000	0.219	0.999
Extrap.	0.183	0.500	0.000	0.002	0.999
Opt.	1.000	1.000	0.500	1.000	1.000
Risk Av.	0.781	0.998	0.000	0.500	1.000
Passive	0.001	0.001	0.000	0.000	0.500

Table F.13: Hypothesis tests for statistical significance when comparing means of demographic categories across the four clusters from the Individual Investor Survey. We assume that each cluster corresponds to an independent collection of investors. For every demographic, we calculate the mean response within each cluster. The resulting means are shown in Table F.9.

For every two means, we test for statistical significance using Welch's t-test—with the exception of the risk aversion coefficient, where we employ a z-test using the risk aversion coefficients and associated standard errors from the estimation. Each cell contains the p-value associated with testing if the column cluster mean minus the row cluster mean is greater than zero. A cell is colored green if the mean in the row cluster is significantly less than the mean in the column cluster, at the 10% significance level. A cell is colored red if the mean in the row cluster is significantly greater than the mean in the column cluster, at the 10% significance level.

Extrap. is the *Extrapolators* cluster. *Semi-Pass.* corresponds to the cluster of individuals who are passive with respect to the S&P 500 movements, but are active with respect to the January Volatility. *Risk Av.* denotes the *Risk Avoiders* cluster; *Opt.* denotes the *Optimistic* cluster;

Risk Aversion is short for the risk aversion coefficient. *Satisfied 2015* is an indicator for if an investor was satisfied with their 2015 investment returns.

Statistical Significance for Comparisons of Mean Demographics across Clusters (2017)

Pres. Elec.	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.978	0.000	0.000	1.000	0.000
Risk Av.	0.022	0.500	0.000	0.000	1.000	0.000
Passive	1.000	1.000	0.500	0.138	1.000	0.000
Opt.	1.000	1.000	0.862	0.500	1.000	0.003
Extrap (A)	0.000	0.000	0.000	0.000	0.500	0.000
Extrap (P)	1.000	1.000	1.000	0.997	1.000	0.500

SP Fall	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	1.000	1.000	0.000	1.000	1.000
Risk Av.	0.000	0.500	0.000	0.000	0.000	0.000
Passive	0.000	1.000	0.500	0.000	1.000	1.000
Opt.	1.000	1.000	1.000	0.500	1.000	1.000
Extrap (A)	0.000	1.000	0.000	0.000	0.500	1.000
Extrap (P)	0.000	1.000	0.000	0.000	0.000	0.500

SP Rise	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	1.000	0.000	0.000	0.000	0.000
Risk Av.	0.000	0.500	0.000	0.000	0.000	0.000
Passive	1.000	1.000	0.500	0.000	0.000	0.000
Opt.	1.000	1.000	1.000	0.500	1.000	1.000
Extrap (A)	1.000	1.000	1.000	0.000	0.500	0.000
Extrap (P)	1.000	1.000	1.000	0.000	1.000	0.500

Gender	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.004	0.000	0.031	0.003	0.007
Risk Av.	0.996	0.500	0.308	0.643	0.307	0.610
Passive	1.000	0.692	0.500	0.786	0.435	0.807
Opt.	0.969	0.357	0.214	0.500	0.225	0.439
Extrap (A)	0.997	0.693	0.565	0.775	0.500	0.772
Extrap (P)	0.993	0.390	0.193	0.561	0.228	0.500

Age	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.960	0.000	1.000	0.949	0.044
Risk Av.	0.040	0.500	0.000	1.000	0.543	0.000
Passive	1.000	1.000	0.500	1.000	1.000	1.000
Opt.	0.000	0.000	0.000	0.500	0.000	0.000
Extrap (A)	0.051	0.457	0.000	1.000	0.500	0.000
Extrap (P)	0.956	1.000	0.000	1.000	1.000	0.500

Net Worth	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.921	0.996	0.266	0.013	0.740
Risk Av.	0.079	0.500	0.870	0.027	0.000	0.179
Passive	0.004	0.130	0.500	0.001	0.000	0.010
Opt.	0.734	0.973	0.999	0.500	0.073	0.895
Extrap (A)	0.987	1.000	1.000	0.927	0.500	0.999
Extrap (P)	0.260	0.821	0.990	0.105	0.001	0.500

Advised	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.682	1.000	0.000	0.000	0.137
Risk Av.	0.318	0.500	1.000	0.000	0.000	0.040
Passive	0.000	0.000	0.500	0.000	0.000	0.000
Opt.	1.000	1.000	1.000	0.500	0.102	1.000
Extrap (A)	1.000	1.000	1.000	0.898	0.500	1.000
Extrap (P)	0.863	0.960	1.000	0.000	0.000	0.500

Retired	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.996	0.000	1.000	0.996	0.388
Risk Av.	0.004	0.500	0.000	0.888	0.636	0.000
Passive	1.000	1.000	0.500	1.000	1.000	1.000
Opt.	0.000	0.112	0.000	0.500	0.222	0.000
Extrap (A)	0.004	0.364	0.000	0.778	0.500	0.000
Extrap (P)	0.612	1.000	0.000	1.000	1.000	0.500

US	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.000	0.000	0.000	0.000	0.000
Risk Av.	1.000	0.500	0.000	0.000	0.000	1.000
Passive	1.000	1.000	0.500	0.000	1.000	1.000
Opt.	1.000	1.000	1.000	0.500	1.000	1.000
Extrap (A)	1.000	1.000	0.000	0.000	0.500	1.000
Extrap (P)	1.000	0.000	0.000	0.000	0.000	0.500

Gamble Pref.	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.647	1.000	0.000	0.133	0.946
Risk Av.	0.353	0.500	1.000	0.000	0.062	0.911
Passive	0.000	0.000	0.500	0.000	0.000	0.000
Opt.	1.000	1.000	1.000	0.500	1.000	1.000
Extrap (A)	0.867	0.938	1.000	0.000	0.500	0.996
Extrap (P)	0.054	0.089	1.000	0.000	0.004	0.500

Risk Aversion	Contr.	Risk Av.	Passive	Opt.	Extrap (A)	Extrap (P)
Contr.	0.500	0.520	1.000	0.000	0.003	0.884
Risk Av.	0.480	0.500	0.500	0.500	0.000	0.898
Passive	0.000	0.500	0.500	0.500	0.000	0.005
Opt.	1.000	0.500	0.500	0.500	1.000	1.000
Extrap (A)	0.997	1.000	1.000	0.000	0.500	0.993
Extrap (P)	0.116	0.102	0.995	0.000	0.007	0.500

Table F.14: Hypothesis tests for statistical significance when comparing means of demographic categories across the four clusters from the Individual Investor Survey. We assume that each cluster corresponds to an independent collection of investors. For every demographic, we calculate the mean response within each cluster. The resulting means are shown in Table F.10.

For every two means, we test for statistical significance using Welch's t-test—with the exception of the risk aversion coefficient, where we employ a z-test using the risk aversion coefficients and associated standard errors from the estimation. Each cell contains the p-value associated with testing if the column cluster mean minus the row cluster mean is greater than zero. A cell is colored green if the mean in the row cluster is significantly less than the mean in the column cluster, at the 10% significance level. A cell is colored red if the mean in the row cluster is significantly greater than the mean in the column cluster, at the 10% significance level.

Contr. denotes the *Contrarians* cluster; *Risk Av.* denotes the *Risk Avoiders* cluster; *Opt.* denotes the *Optimistic* cluster; *Extrap.* is the *Extrapolators* cluster. *Extrap A* and *Extrap P* correspond respectively to the clusters of *Extrapolators* who are active or passive with respect to the 2016 U.S. Presidential Election.

Risk Aversion is short for the risk aversion coefficient. *US* corresponds to the percentage of U.S. investors in the considered cluster.

Clustering of Financial Advisors (2015)			
	Cluster 1	**Cluster 2**	**Cluster 3**
% Respondents	47%	37%	16%
	Allocation Decisions		
Behavior	Contrarian	Passive	Extrapolators
Reaction to S&P 500 Fall	1.1	0.4	-1.4
Reaction to S&P 500 Rise	-1.1	0.2	0.3

Table F.15: Clustering of allocation decision responses from the 2015 Financial Advisor Survey. For each cluster, we present the percent of respondents and mean response based on the response coding in Table 8.3.

Clustering of Financial Advisors (2016)				
	Cluster 1	**Cluster 2**	**Cluster 3**	**Cluster 4**
% Respondents	36%	34%	8%	22%
	Allocation Decisions			
Behavior	Passive	Contrarian	Optimistic	Risk Avoiders
Reaction to S&P 500 Fall	0.1	1.3	1.1	-0.8
Reaction to S&P 500 Rise	0.2	-1.1	1.3	-1.1

Table F.16: Clustering of allocation decision responses from the 2016 Financial Advisor Survey. For each cluster, we present the percent of respondents and mean response based on the response coding in Table 8.3.

Clustering of Institutional Investors (2015)			
	Cluster 1	**Cluster 2**	**Cluster 3**
% Respondents	68%	22%	10%
	Allocation Decisions		
Behavior	Contrarian	Passive/Extrapolator	Risk Avoiders
Reaction to S&P 500 Fall	1.5	-0.6	-1.4
Reaction to S&P 500 Rise	-1.1	0.7	-1.7

Table F.17: Clustering of allocation decision responses from the 2015 Institutional Investor Survey.

Clustering of Institutional Investors (2016)			
	Cluster 1	**Cluster 2**	**Cluster 3**
% Respondents	61%	34%	4%
	Allocation Decisions		
Behavior	Contrarian	Passive/Risk Avoiders	Optimistic
Reaction to S&P 500 Fall	1.3	-0.6	1.2
Reaction to S&P 500 Rise	-1.2	-0.3	1.2

Table F.18: Clustering of allocation decision responses from the 2016 Institutional Investor Survey.

Clustering of Institutional Investors (2017)			
	Cluster 1	**Cluster 2**	**Cluster 3**
% Respondents	9%	20%	71%
	Allocation Decisions		
Behavior	Risk Avoiders	Passive/Extrapolator	Contrarian
Reaction to S&P 500 Fall	-1.6	-0.3	1.2
Reaction to S&P 500 Rise	-1.1	0.4	-1.1

Table F.19: Clustering of allocation decision responses from the 2017 Institutional Investor Survey.

Bibliography

[1] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. {TensorFlow}: a system for {Large-Scale} machine learning. In *12th USENIX symposium on operating systems design and implementation (OSDI 16)*, pages 265–283, 2016.

[2] Bechara Abouarab, Christian Bazarian, Zied Ben Chaouch, Andrew W. Lo, Guillermo Mourenza, Richard Novak, and Frederic Vigneault. Financing repurposed drugs for rare diseases: A case study of unravel biosciences. Manuscript under Review, 2022.

[3] C. J. Adcock. Sample size determination: A review. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 46(2):261–283, 1997.

[4] A. E. Ades, G. Lu, and K. Claxton. Expected value of sample information calculations in medical decision modeling. *Medical decision making : an international journal of the Society for Medical Decision Making*, 2(24):207–227, 2004.

[5] Julie Agnew, Pierluigi Balduzzi, and Annika Sunden. Portfolio choice and trading in a large 401 (k) plan. *American Economic Review*, pages 193–215, 2003.

[6] Omar B. Ahmad, Cynthia Boschi-Pinto, Alan D. Lopez, Christopher J. L. Murray, Rafael Lozano, and Mie Inoue. Age standardization of rates: a new who standard. gpe discussion paper series: No 31. `http://www.who.int/healthinfo/paper31.pdf`, 2001. Accessed: 2021-04-13.

[7] Amrita Ahuja, Susan Athey, Arthur Baker, Eric Budish, Juan Camilo Castillo, Rachel Glennerster, Scott Duke Kominers, Michael Kremer, Jean Lee, Canice Prendergast, Christopher M. Snyder, Alex Tabarrok, Brandon Joel Tan, and Witold Więcek. Preparing for a pandemic: Accelerating vaccine availability. *AEA Papers and Proceedings*, 111:331–35, 05 2021.

[8] . AJMC Staff. A timeline of covid-19 vaccine developments in 2021, 04 2021.

[9] Irene Aldridge and Marco Avellaneda. *Big data science in finance*. John Wiley & Sons, 2021.

[10] Sharon Amit, Gili Regev-Yochay, Arnon Afek, Yitshak Kreiss, and Eyal Leshem. Early rate reductions of SARS-CoV-2 infection and COVID-19 in BNT162b2 vaccine recipients, 2021.

[11] B. Appleby, P. Duggan, A. Regenberg, and P. Rabins. Psychiatric and neuropsychiatric adverse events associated with deep brain stimulation: A meta-analysis of ten years' experience. *Movement disorders*, 22(12):1722–1728, 2007.

[12] A. Arora, S. Belenzon, A. Patacconi, and J. Suh. The changing structure of american innovation: Some cautionary remarks for economic growth. *Innovation Policy and the Economy*, 20:39–93, 2020.

[13] Joseph Barberio, Jacob Becraft, Zied Ben Chaouch, Dimitris Bertsimas, Tasuku Kitada, Michael L. Li, Andrew W. Lo, Kevin Shi, and Qingyang Xu. Accelerating Vaccine Innovation for Emerging Infectious Diseases via Parallel Discovery. In *Entrepreneurship and Innovation Policy and the Economy, volume 2*, volume 2 of *NBER Chapters*. National Bureau of Economic Research, Inc, July 2022.

[14] Nicholas Barberis, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. Extrapolation and bubbles. *Harvard and Yale unpublished manuscript*, 2015.

[15] Francisco Barillas and Jay Shanken. Comparing asset pricing models. *The Journal of Finance*, 73(2):715–754, 2018.

[16] Eric Bellman and Vibhuti Agarwal. India faces covid-19 vaccine shortage as cases surge., 04 2021.

[17] Zied Ben Chaouch, Andrew W. Lo, and Chi Heem Wong. Should we allocate more covid-19 vaccine doses to non-vaccinated individuals? *PLOS Global Public Health*, 2(7):1–17, 07 2022.

[18] Itzhak Ben-David and David Hirshleifer. Are investors really reluctant to realize their losses? trading responses to past returns and the disposition effect. *Review of Financial Studies*, 25(8):2485–2532, 2012.

[19] Shlomo Benartzi. Excessive extrapolation and the allocation of 401 (k) accounts to company stock. *Journal of Finance*, pages 1747–1764, 2001.

[20] Heather L. Benz, Brittany Caldwell, John P. Ruiz, Anindita Saha, Martin Ho, Stephanie Christopher, Dawn Bardot, Margaret Sheehan, Anne Donnelly, Lauren McLaughlin, Brennan Mange, A. Brett Hauber, Katrina Gwinn, William J. Heetderks, and Murray Sheldon. Patient-centered identification of meaningful regulatory endpoints for medical devices to treat parkinson's disease. *MDM Policy & Practice*, 6(1):23814683211021380, 2021. PMID: 34277950.

[21] Donald A. Berry, Scott Berry, Peter Hale, Leah Isakov, Andrew W. Lo, Kien Wei Siah, and Chi Heem Wong. A cost/benefit analysis of clinical trial designs for covid-19 vaccine candidates. *PLOS ONE*, 15(12):1–17, 12 2020.

[22] Donald A. Berry, Scott Berry, Peter Hale, Leah Isakov, Andrew W. Lo, Kien Wei Siah, and Chi Heem Wong. A cost/benefit analysis of clinical trial designs for covid-19 vaccine candidates. *PLOS ONE*, 15(12):1–17, 12 2020.

[23] Dimitris Bertsimas, Vassilis Digalakis Jr. au2, Alexander Jacquillat, Michael Lingzhi Li, and Alessandro Previero. Where to locate covid-19 mass vaccination facilities?, 2021.

[24] Dimitris Bertsimas, Leonard Boussioux, Ryan Cory-Wright, Arthur Delarue, Vasileios Digalakis, Alexandre Jacquillat, Driss Lahlou Kitane, Galit Lukin, Michael Li, Luca Mingardi, Omid Nohadani, Agni Orfanoudaki, Theodore Papalexopoulos, Ivan Paskov, Jean Pauphilet, Omar Skali Lami, Bartolomeo Stellato, Hamza Tazi Bouardi, Kimberly Villalobos Carballo, Holly Wiberg, and Cynthia Zeng. From predictions to prescriptions: A data-driven response to covid-19. *medRxiv*, 2020.

[25] Dimitris Bertsimas, Joshua Ivanhoe, Alexandre Jacquillat, Michael Li, Alessandro Previero, Omar Skali Lami, and Hamza Tazi Bouardi. Optimizing vaccine allocation to combat the covid-19 pandemic. *medRxiv*, 2020.

[26] Lindsey Bever. Why you shouldn't skip your second dose of the coronavirus vaccine, 04 2021.

[27] Karen Brooks Harper. Hundreds of thousands of covid-19 vaccine deliveries, injections delayed by winter storm, 02 2021.

[28] Svetlana Bryzgalova, Jiantao Huang, and Christian Julliard. Bayesian solutions for the factor zoo: We just ran two quadrillion models. Available at SSRN 3481736, 2019.

[29] D. Butler. Translational research: Crossing the valley of death. *Nature*, 453:840–2, 2008.

[30] E. Callaway. Dozens to be deliberately infected with coronavirus in uk 'human challenge' trials. *Nature*, 586:651–2, 2020.

[31] Mark M. Carhart. On persistence in mutual fund performance. *The Journal of Finance*, 52(1):57–82, 1997.

[32] Raf Casert. Delay in pfizer vaccine shipments frustrate europe, canada, 01 2021.

[33] . Centers for Disease Control & Prevention (CDC). Vfc vaccine management and accountability. https://www.cdc.gov/vaccines/programs/vfc/awardees/vaccine-management/, 2014. Accessed: 2022-04-14.

[34] . Centers for Disease Control & Prevention (CDC). Covid-19 vaccine inventory management best practices. https://www.cdc.gov/vaccines/covid-19/vaccine-inventory-management.html, 2021. Accessed: 2021-04-19.

[35] . Centers for Disease Control & Prevention (CDC). Covid-19 vaccines that require 2 shots. https://www.cdc.gov/coronavirus/2019-ncov/vaccines/second-shot.html, 2021. Accessed: 2021-04-19.

[36] . Centers for Disease Control & Prevention (CDC). Covid data tracker: Trends in number of covid-19 vaccinations in the us. https://covid.cdc.gov/covid-data-tracker/#vaccination-trends, 2021. Accessed: 2021-04-13.

[37] . Centers for Disease Control & Prevention (CDC). Pfizer-biontech covid-19 vaccine storage and handling summary. https://www.cdc.gov/vaccines/covid-19/info-by-product/pfizer/downloads/storage-summary.pdf, 2021. Accessed: 2021-04-19.

[38] Zied Ben Chaouch, Shomesh E. Chaudhuri, and Andrew W. Lo. Bayesian decision analysis under risk and uncertainty: A tale of two exposures. Manuscript submitted for publication, 2022.

[39] Zied Ben Chaouch, Manish Singh, Katherine Xiong, and Andrew W. Lo. Chimeras in the factor zoo. Manuscript in Progress, 2022.

[40] Zied Ben Chaouch, Qingyang Xu, Shomesh E. Chaudhuri Alice Chen, David Gebben, Raymond C. Harris, Jenny Flythe, Frank Hurst, Carol Mansfield, Anindita Saha, Murray Sheldon, Kien Wei Siah, Michelle Tarver, Katherine Treiman, Melissa West, Dallas Wood, and Andrew W. Lo. Use of bayesian decision analysis in the design of clinical trials: Application to kidney failure devices. Manuscript under preparation, 2022.

[41] N. Chaudhary, D. Weissman, and K. A. Whitehead. mrna vaccines for infectious diseases: principles, delivery and clinical translation. *Nature Reviews Drug Discovery*, 20:817–38, 2021.

[42] S. E. Chaudhuri, Cheng K., A. W. Lo, S. Pepke, S. Rinaudo, L. Roman, et al. A portfolio approach to accelerate therapeutic innovation in ovarian cancer. *Journal of Investment Management*, 17(2):5–16, 2019.

[43] Shomesh Chaudhuri, Andrew W. Lo, Danying Xiao, and Qingyang Xu. Bayesian adaptive clinical trials for anti-infective therapeutics during epidemic outbreaks. *Harvard Data Science Review*, 5 2020. https://hdsr.mitpress.mit.edu/pub/non4rfk6.

[44] Shomesh E Chaudhuri, Phillip Adamson, Dean Bruhn-Ding, Zied Ben Chaouch, David Gebben, Liliana R Gonzalez, Barry Liden, Shelby D Reed, Anindita Saha, Daniel Schaber, Kenneth Stein, Michelle E Tarver, and Andrew W. Lo. Patient-centered clinical trial design for heart failure devices via bayesian decision analysis. Manuscript submitted for publication, 2022.

[45] Shomesh E. Chaudhuri, Zied Ben Chaouch, Brennan Mange Brett Hauber, Mo Zhou, Stephanie Christopher, Dawn Bardot, Margaret Sheehan, Anne Donnelly, Lauren McLaughlin, Brittany Caldwell, Heather L. Benz, Martin Ho, Anindita Saha, Katrina Gwinn, Murray Sheldon, and Andrew W. Lo. Use of bayesian decision analysis to maximize value in patient-centered randomized clinical trials in parkinson's disease. Manuscript submitted for publication, 2021.

[46] 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.

[47] Shomesh E. Chaudhuri and Andrew W. Lo. Incorporating patient preferences via bayesian decision analysis. *Clinical Journal of the American Society of Nephrology*, 16(4):639–641, 2021.

[48] Luyang Chen, Markus Pelger, and Jason Zhu. Deep learning in asset pricing. arXiv preprint arXiv:1904.00745, 2019.

[49] James J Choi, David Laibson, Brigitte C Madrian, and Andrew Metrick. Reinforcement learning and savings behavior. *Journal of Finance*, 64(6):2515–2534, 2009.

[50] . Coalition for Epidemic Preparedness Innovations (CEPI). Our portfolio. `https://cepi.net/research_dev/our-portfolio/`, 2022. Accessed: 2022-04-14.

[51] John H Cochrane. *Asset pricing: Revised edition*. Princeton university press, 2009.

[52] John H Cochrane. Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108, 2011.

[53] Richard A Cohn, Wilbur G Lewellen, Ronald C Lease, and Gary G Schlarbaum. Individual investor risk aversion and investment portfolio composition. *The Journal of Finance*, 30(2):605–620, 1975.

[54] Gregory Connor and Robert A Korajczyk. Performance measurement with the arbitrage pricing theory: A new framework for analysis. *Journal of financial economics*, 15(3):373–394, 1986.

[55] P. T. Conway and R. Knight. Legitimization and incorporation of patient preferences. *Clinical Journal of the American Society of Nephrology*, 16(4):645–647, 2021.

[56] Stanley T. Crooke. A call to arms against ultra-rare diseases. *Nature Biotechnology*, 39(6):671–677, 2021.

[57] Noa Dagan, Noam Barda, Eldad Kepten, Oren Miron, Shay Perchik, Mark A. Katz, Miguel A. Hernán, Marc Lipsitch, Ben Reis, and Ran D. Balicer. Bnt162b2 mrna covid-19 vaccine in a nationwide mass vaccination setting. *New England Journal of Medicine*, 0(0):null, 2021.

[58] Abdi Latif Dahir. Africa's already slow vaccine drive is threatened as supplies from a stricken india are halted., 04 2021.

[59] Magnus Dahlquist, José Vicente Martinez, and Paul Söderlind. Individual investor activity and performance. Technical Report 14-08, Swedish House of Finance Research Paper, 2014.

[60] Chetan Dave, Catherine C Eckel, Cathleen A Johnson, and Christian Rojas. Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty*, 41(3):219–243, 2010.

[61] M. M. Davis, A. T. Butchart, J. R. C. Wheeler, et al. Failure-to-success ratios, transition probabilities and phase lengths for prophylactic vaccines versus other pharmaceuticals in the development pipeline. *Vaccine*, 29(51):9414–6, 2011.

[62] Werner PM De Bondt. Betting on trends: Intuitive forecasts of financial risk and return. *International Journal of Forecasting*, 9(3):355–371, 1993.

[63] Leo De Haan and Jan Kakes. Momentum or contrarian investment strategies: evidence from dutch institutional investors. *Journal of Banking & Finance*, 35(9):2245–2251, 2011.

[64] Matthew Dixon and Nick Polson. Deep fundamental factor models. *SIAM Journal on Financial Mathematics*, 11(3):SC26–SC37, 2020.

[65] Matthew F Dixon, Igor Halperin, and Paul Bilokon. *Machine learning in Finance*, volume 1406. Springer, 2020.

[66] Matthew F Dixon, Nicholas G Polson, and Kemen Goicoechea. Deep partial least squares for empirical asset pricing. arXiv preprint arXiv:2206.10014, 2022.

[67] E. Dolgin. The tangled history of mrna vaccines. *Nature*, 597:318–24, 2021.

[68] Daniel Dorn and Gur Huberman. Talk and action: What individual investors say and what they do. *Review of Finance*, 9(4):437–481, 2005.

[69] R. G. Douglas and V. B. Samant. The vaccine industry. *Plotkin's Vaccines*, 1:41–50, 2018.

[70] Daniel Ellsberg. Risk, ambiguity, and the savage axioms. *The Quarterly Journal of Economics*, 75(4):643–669, 1961.

[71] L. J. Emrich and M. R. Piedmonte. A method for generating high-dimensional multivariate binary variates. *The American Statistician*, 45(4):302–4, 1991.

[72] D. E. Fagnan, A. A. Gromatzky, R. M. Stein, J. M. Fernandez, and A. W. Lo. Financing drug discovery for orphan diseases. *Drug Discovery Today*, 19(5):533–8, 2014.

[73] Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

[74] Eugene F. Fama and Kenneth R. French. A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1–22, 2015.

[75] Eugene F. Fama and Kenneth R. French. Choosing factors. *Journal of Financial Economics*, 128(2):234–252, 2018.

[76] Eugene F Fama and James D MacBeth. Risk, return, and equilibrium: Empirical tests. *Journal of political economy*, 81(3):607–636, 1973.

[77] Jianqing Fan, Zheng Tracy Ke, Yuan Liao, and Andreas Neuhierl. Structural deep learning in conditional asset pricing. Available at SSRN 4117882, 2022.

[78] . FDA. Patient preference information — voluntary submission, review in premarket approval applications, humanitarian device exemption applications, and de novo requests, and inclusion in decision summaries and device labeling. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/benefit-risk-assessment-new-drug-and-biological-products, 2016. Accessed: 2022-05-25.

[79] . FDA. Consideration of uncertainty in making benefit-risk determinations in medical device premarket approvals, de novo classifications, and humanitarian device exemptions – guidance for industry and food and drug administration staff. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/consideration-uncertainty-making-benefit-risk-determinations-medical-device-premarket-a 2019. Accessed: 2022-05-25.

[80] . FDA. Factors to consider when making benefit-risk determinations in medical device premarket approval and de novo classifications. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/factors-consider-when-making-benefit-risk-determinations-medical-device-premarket-appr 2019. Accessed: 2022-05-25.

[81] . FDA. Requests for feedback and meetings for medical device submissions: The q-submission program—guidance for industry and food and drug administration staff. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/requests-feedback-and-meetings-medical-device-submissions-q-submission-program, 2019. Accessed: 2022-05-25.

[82] . FDA. Benefit-risk assessment for new drug and biological products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/benefit-risk-assessment-new-drug-and-biological-products, 2021. Accessed: 2022-05-25.

[83] . FDA. Benefit-risk assessment for new drug and biological products. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/benefit-risk-assessment-new-drug-and-biological-products, 2021. Accessed: 2022-01-05.

[84] . FDA. Office of neurology's summary review memorandum. https://www.accessdata.fda.gov/drugsatfda_docs/nda/2021/Aducanumab_BLA761178_Dunn_2021_06_07.pdf, 2021. Accessed: 2022-01-05.

[85] Guanhao Feng, Stefano Giglio, and Dacheng Xiu. Taming the factor zoo: A test of new factors. *The Journal of Finance*, 75(3):1327–1370, 2020.

[86] J. M. Fernandez, R. M. Stein, and A. W. Lo. Commercializing biomedical research through securitization techniques. *Nature Biotechnology*, 30(10):964–75, 2012.

[87] R. A. Fisher. *Statistical Methods for Research Workers*, pages 66–70. Springer New York, New York, NY, 1992.

[88] J. E. Flythe and M. West. Using patient preference information to inform regulatory decision making. *Clinical Journal of the American Society of Nephrology*, 16(4):642–644, 2021.

[89] Benjamin Freedman. Equipoise and the ethics of clinical research. *New England Journal of Medicine*, 317(3):141–145, 1987. PMID: 3600702.

[90] Kenneth R. French. Kenneth r. french's data library. https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. Accessed: 07-19-2022.

[91] Joachim Freyberger, Andreas Neuhierl, and Michael Weber. Dissecting characteristics nonparametrically. *The Review of Financial Studies*, 33(5):2326–2377, 2020.

[92] Michael R Gibbons, Stephen A Ross, and Jay Shanken. A test of the efficiency of a given portfolio. *Econometrica: Journal of the Econometric Society*, pages 1121–1152, 1989.

[93] Stefano Giglio, Yuan Liao, and Dacheng Xiu. Thousands of alpha tests. *The Review of Financial Studies*, 34(7):3456–3496, 2021.

[94] Stefano Giglio and Dacheng Xiu. Asset pricing with omitted factors. *Journal of Political Economy*, 129(7):1947–1990, 2021.

[95] R. Glennerster and M. R. Kremer. A better way to spur medical research and development. *Regulation*, 23(2):34–9, 2000.

[96] Zaria Gorvett. How effective is a single vaccine dose against covid-19?, 01 2021.

[97] D. Gouglas, T. T. Le, K. Henderson, A. Kaloudis, T. Danielsen, N. C. Hammersland, et al. Estimating the cost of vaccine development against epidemic infectious diseases: a cost minimisation study. *Lancet Global Health*, 6(12):e1386–96, 2018.

[98] D. Gouglas and K. Marsh. Prioritizing investments in new vaccines against epidemic infectious diseases: A multi-criteria decision analysis. *Journal of Multi-Criteria Decision Analysis*, 26(3):153–63, 2019.

[99] Dimitrios Gouglas and Kevin Marsh. Prioritizing investments in rapid response vaccine technologies for emerging infections: A portfolio decision analysis. *PLOS ONE*, 16:1–21, 02 2021.

[100] Jurgens Graham. Modelling decay of population immunity with proposed second dose deferral strategy. *medRxiv*, 2021.

[101] Robin Greenwood and Andrei Shleifer. Expectations of returns and expected returns. *Review of Financial Studies*, pages 715–746, 2014.

[102] John M Griffin, Jeffrey H Harris, and Selim Topaloglu. The dynamics of institutional and individual trading. *The Journal of Finance*, 58(6):2285–2320, 2003.

[103] Mark Grinblatt and Matti Keloharju. The investment behavior and performance of various investor types: a study of finland's unique data set. *Journal of Financial Economics*, 55(1):43–67, 2000.

[104] Mark Grinblatt and Matti Keloharju. Sensation seeking, overconfidence, and trading activity. *The Journal of Finance*, 64(2):549–578, 2009.

[105] P. M. Grundy, M. J. R. Healy, and D. H. Rees. Economic choice of the amount of experimentation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 18(1):32–55, 1956.

[106] Shihao Gu, Bryan Kelly, and Dacheng Xiu. Empirical asset pricing via machine learning. *The Review of Financial Studies*, 33(5):2223–2273, 2020.

[107] Shihao Gu, Bryan Kelly, and Dacheng Xiu. Autoencoder asset pricing models. *Journal of Econometrics*, 222(1):429–450, 2021.

[108] Jerry Halpern, Byron Wm. Brown Jr, and John Hornberger. The sample size for a clinical trial: A bayesian–decision theoretic approach. *Statistics in Medicine*, 20(6):841–858, 2001.

[109] T. Harbert. How moderna is racing to a coronavirus vaccine. https://mitsloan.mit.edu/ideas-made-to-matter/how-moderna-racing-to-a-coronavirus-vaccine, 2020. Accessed: 2022-04-14.

[110] Joop Hartog, Ada Ferrer-i Carbonell, and Nicole Jonker. Linking measured risk aversion to individual characteristics. *Kyklos*, 55(1):3–26, 2002.

[111] Campbell R Harvey and Yan Liu. Lucky factors. *Journal of Financial Economics*, 141(2):413–435, 2021.

[112] Campbell R Harvey, Yan Liu, and Heqing Zhu. ... and the cross-section of expected returns. *The Review of Financial Studies*, 29(1):5–68, 2016.

[113] Brett Hauber, Brennan Mange, Mo Zhou, Shomesh Chaudhuri, Heather L. Benz, Brittany Caldwell, John P. Ruiz, Anindita Saha, Martin Ho, Stephanie Christopher, Dawn Bardot, Margaret Sheehan, Anne Donnelly, Lauren McLaughlin, Katrina Gwinn, Andrew Lo, and Murray Sheldon. Parkinson's patients' tolerance for risk and willingness to wait for potential benefits of novel neurostimulation devices: A patient-centered threshold technique study. *MDM Policy & Practice*, 6(1):2381468320978407, 2021. PMID: 33521289.

[114] Aqil Haziq Mahmud. Delaying second covid-19 vaccine jab could allow more people in singapore to be protected and save lives, say experts, 05 2021.

[115] Anna Heath, Natalia Kunst, Christopher Jackson, Mark Strong, Fernando Alarid-Escudero, Jeremy D. Goldhaber-Fiebert, Gianluca Baio, Nicolas A. Menzies, and Hawre Jalal. Calculating the expected value of sample information in practice: Considerations from 3 case studies. *Medical Decision Making*, 40(3):314–326, 2020. PMID: 32297840.

[116] Michael Hill and Jennifer Peltz. Delay in pfizer vaccine shipments frustrate europe, canada, 01 2021.

[117] Charles A Holt and Susan K Laury. Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655, 2002.

[118] Jared Hopkins and Bojan Pancevski. Some covid-19 vaccines are effective after one dose, can be stored in normal freezers, data show, 02 2021.

[119] Kewei Hou, Haitao Mo, Chen Xue, and Lu Zhang. An Augmented q-Factor Model with Expected Growth*. *Review of Finance*, 25(1):1–41, 02 2020.

[120] Kewei Hou, Chen Xue, and Lu Zhang. Digesting anomalies: An investment approach. *The Review of Financial Studies*, 28(3):650–705, 09 2014.

[121] Kewei Hou, Chen Xue, and Lu Zhang. Replicating Anomalies. *The Review of Financial Studies*, 33(5):2019–2133, 12 2018.

[122] Paul Hunter. Delaying the second covid vaccine dose - a medical expert answers key questions, 01 2021.

[123] Paul R Hunter and Julii Brainard. Estimating the effectiveness of the pfizer covid-19 bnt162b2 vaccine after a single dose. a reanalysis of a study of 'real-world' vaccination outcomes from israel. *medRxiv*, 2021.

[124] F. P. Hurst, D. Chianchiano, L. Upchurch, B. R. Fisher, J. E. Flythe, et al. Stimulating patient engagement in medical device development in kidney disease: A report of a kidney health initiative workshop. *American Journal of Kidney Diseases*, 70(4):561–569, 2017.

[125] Securities Industry and . Financial Markets Association (SIFMA). Capital markets fact book, 2021. `https://www.sifma.org/resources/research/fact-book/`, 2021. Accessed: 2022-04-14.

[126] 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. Annals Issue in Honor of Jerry A. Hausman.

[127] Andrew Jackson. The aggregate behaviour of individual investors. *Available at SSRN 536942*, 2003.

[128] S. Jain, A. Venkataraman, M. E. Wechsler, and N. A. Peppas. Messenger rna-based vaccines: Past, present, and future directions in the context of the covid-19 pandemic. *Advanced Drug Delivery Reviews*, 179:11400, 2021.

[129] Perrine Janiaud, Telba Irony, Estelle Russek-Cohen, and Steven N. Goodman. U.s. food and drug administration reasoning in approval decisions when efficacy evidence is borderline, 2013–2018. *Annals of Internal Medicine*, 174(11):1603–1611, 2021. PMID: 34543584.

[130] Lynn A. Jansen, Daruka Mahadevan, Paul S. Appelbaum, William M. P. Klein, Neil D. Weinstein, Motomi Mori, Catherine Degnin, and Daniel P. Sulmasy. Variations in unrealistic optimism between acceptors and decliners of early phase cancer trials. *Journal of Empirical Research on Human Research Ethics*, 12(4):280–288, 2017. PMID: 28728498.

[131] S. Jarrett, S. Pagliusi, R. Park, et al. The importance of vaccine stockpiling to respond to epidemics and remediate global supply shortages affecting immunization: strategic challenges and risks identified by manufacturers. *Vaccine X*, 9:10011, 2021.

[132] S. Jarrett, L. Yang, and S. Pagliusi. Roadmap for strengthening the vaccine supply chain in emerging countries: Manufacturers' perspectives. *Vaccine X*, 5(10006):8, 2020.

[133] D. Jimenez. Covid-19: vaccine pricing varies wildly by country and company. https://www.pharmaceutical-technology.com/features/covid-19-vaccine-pricing-varies-country-company/, 2021. Accessed: 2022-04-14.

[134] Adam Jourdan and Adam Jourdan. Latin america's vaccine shortage threatens fragile revival as pandemic rages, 04 2021.

[135] Siri R. Kadire, Robert M. Wachter, and Nicole Lurie. Delayed second dose versus standard regimen for covid-19 vaccination. *New England Journal of Medicine*, 384(9):e28, 2021.

[136] Ron Kaniel, Gideon Saar, and Sheridan Titman. Individual investor trading and stock returns. *The Journal of Finance*, 63(1):273–310, 2008.

[137] K. Kelland. Gsk ends development of ebola vaccine, hands work to u.s. institute. https://www.reuters.com/article/us-health-ebola-gsk/gsk-ends-development-of-ebola-vaccine-hands-work-to-u-s-institute-idUSKCN1UW15S, 2019. Accessed: 2022-04-14.

[138] Bryan T Kelly, Seth Pruitt, and Yinan Su. Characteristics are covariances: A unified model of risk and return. *Journal of Financial Economics*, 134(3):501–524, 2019.

[139] Bryan T Kelly and Dacheng Xiu. Factor models, machine learning, and asset pricing. forthcoming in the Annual Review of Financial Economics, 2022.

[140] J. Kimmelman, K. Duckworth, T. Ramsay, T. Voss, B. Ravina, and M. Emborg. Risk of surgical delivery to deep nuclei: A meta-analysis. *Movement Disorders*, 26(8):1415–1421, 2011.

[141] Z. Kis, C. Kontoravdi, R. Shattock, and N. Shah. Resources, production scales and time required for producing rna vaccines for the global pandemic demand. *Vaccines*, 9(1), 2021.

[142] Z. Kis and Z. Rizvi. How to make enough vaccine for the world in one year? https://www.citizen.org/article/how-to-make-enough-vaccine-for-the-world-in-one-year/, 2021. Accessed: 2022-04-14.

[143] F.H. Knight. *Risk, Uncertainty and Profit*. Cosimo, Incorporated, 2006.

[144] Kishore Konda, Roland Memisevic, and David Krueger. Zero-bias autoencoders and the benefits of co-adapting features. arXiv preprint arXiv:1402.3337, 2014.

[145] Meryl Kornfield. As delays hamper some second coronavirus vaccine doses, a debate rages: Prioritize one shot or two?, 02 2021.

[146] Serhiy Kozak, Stefan Nagel, and Shrihari Santosh. Interpreting factor models. *The Journal of Finance*, 73(3):1183–1223, 2018.

[147] Naomi Kresge and Riley Griffin. Pfizer to cut vaccine shipments as belgian factory renovated, 01 2021.

[148] JL Kriss, LE Reynolds, A Wang, and et al. Covid-19 vaccine second-dose completion and interval between first and second doses among vaccinated persons - united states, december 14, 2020-february 14, 2021. *MMWR Morb Mortal Wkly Rep 2021*, 2021.

[149] Sharon LaFraniere and Noah Weiland. Covid-19: Some johnson & johnson vaccine doses on hold in u.s. after factory mix-up, 04 2021.

[150] D. Laibson. Golden eggs and hyperbolic discounting. *The Quarterly Journal of Economics*, 112(2):443–77, 1997.

[151] Martin Lettau and Markus Pelger. Estimating latent asset-pricing factors. *Journal of Econometrics*, 218(1):1–31, 2020.

[152] Michael Li, Hamza Bouardi, Omar Lami, Thomas Trikalinos, Nikolaos Trichakis, and Dimitris Bertsimas. Forecasting covid-19 and analyzing the effect of government interventions. preprint on webpage at `https://doi.org/10.1101/2020.06.23.20138693`, 06 2020.

[153] John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics*, 47(1):13–37, 1965.

[154] Edward H. Livingston. Necessity of 2 Doses of the Pfizer and Moderna COVID-19 Vaccines. *JAMA*, 325(9):898–898, 03 2021.

[155] A. W. Lo. Discussion: New directions for the fda in the 21st century. *Biostatistics, 06*, 18(3):403–407, 2017.

[156] A. W. Lo and S. E. Chaudhuri. *Healthcare Finance: Modern Financial Analysis for Accelerating Biomedical Innovation*. Princeton University Press, Princeton, NJ, 2022.

[157] A. W. Lo, C. Ho, J. Cummings, and K. S. Kosik. Parallel discovery of alzheimer's therapeutics. *Science Translational Medicine*, 6(241):241cm5–241cm5, 2014.

[158] A. W. Lo and R. Thakor. Financing medical innovation. Technical report, MIT, 2021.

[159] Andrew W Lo and Mark T Mueller. Warning: physics envy may be hazardous to your wealth! *arXiv preprint arXiv:1003.2688*, 2010.

[160] Andrew W. Lo, Alexander Remorov, and Zied Ben Chaouch. Measuring risk preferences and asset-allocation decisions: A global survey analysis. *Journal Of Investment Management*, 18(3):5–50, 2020.

[161] Jamie Lopez Bernal, Nick Andrews, Charlotte Gower, Eileen Gallagher, Ruth Simmons, Simon Thelwall, Julia Stowe, Elise Tessier, Natalie Groves, Gavin Dabrera, Richard Myers, Colin N.J. Campbell, Gayatri Amirthalingam, Matt Edmunds, Maria Zambon, Kevin E. Brown, Susan Hopkins, Meera Chand, and Mary Ramsay. Effectiveness of covid-19 vaccines against the b.1.617.2 (delta) variant. *New England Journal of Medicine*, 385(7):585–594, 2021. PMID: 34289274.

[162] Martin Makary. Op-ed: Forget second covid vaccine doses for now, 02 2021.

[163] Ulrike Malmendier and Stefan Nagel. Depression babies: Do macroeconomic experiences affect risk taking?*. *The Quarterly Journal of Economics*, 126(1):373–416, 2011.

[164] Robert Masters. Study examines investors' risk-taking propensities. *Journal of Financial Planning*, 2(3), 1989.

[165] Laura Matrajt, Julia Eaton, Tiffany Leung, and Elizabeth R. Brown. Vaccine optimization for covid-19: Who to vaccinate first? *Science Advances*, 7(6), 2020.

[166] R. Miller. New clinical trial approach for urgently needed therapies could be on horizon. Final 2020. https://medtech.pharmaintelligence.informa.com/MT103997/New-Clinical-Trial-Approach-For-Urgently-Needed-Therapies-Could-Be-On-Horizon, 2016. Accessed: 2018-12-10.

[167] Seyed M. Moghadas, Thomas N. Vilches, Kevin Zhang, Shokoofeh Nourbakhsh, Pratha Sah, Meagan C. Fitzpatrick, and Alison P. Galvani. Evaluation of covid-19 vaccination strategies with a delayed second dose. *PLOS Biology*, 19(4):1–13, 04 2021.

[168] 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, 09 2017.

[169] Vahid Montazerhodjat, John J. Frishkopf, and Andrew W. Lo. Financing drug discovery via dynamic leverage. *Drug Discovery Today*, 21(3):410–414, 2016.

[170] D. M. Morens and A. S. Fauci. Emerging pandemic diseases: How we got to covid-19. *Cell*, 182(5):1077–92, 2020.

[171] N. Nathan, M. Barry, M. Van Herp, and H. Zeller. Shortage of vaccines during a yellow fever outbreak in guinea. *Lancet*, 358(9299):2129–30, 2001.

[172] . National Advisory Committee on Immunization (NACI). Extended dose intervals for covid-19 vaccines to optimize early vaccine roll-out and population protection in canada in the context of limited vaccine supply. https://www.canada.ca/en/public-health/services/immunization/national-advisory-committee-on-immunization-naci/extended-dose-intervals-covid-19-vaccines-early-rollout-population-protection.html, 2021. Accessed: 2021-04-19.

[173] . National Brain Tumor Society (NBTS). Brain tumor investment fund. https://braintumor.org/our-research/investment-fund/, 2021. Accessed: 2022-04-14.

[174] S. Neilson, A. Dunn, and A. Bendix. Moderna's groundbreaking coronavirus vaccine was designed in just 2 days. Business Insider, https://www.businessinsider.com/moderna-designed-coronavirus-vaccine-in-2-days-2020-11, 2020. Accessed: 2022-04-14.

[175] 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.

[176] National Institute of Allergy and . Infectious Diseases (NIAID). Niaid emerging infectious diseases/pathogens. https://www.niaid.nih.gov/research/emerging-infectious-diseases-pathogens, 2018. Accessed: 2022-04-14.

[177] US Burden of Disease Collaborators. The State of US Health, 1990-2010: Burden of Diseases, Injuries, and Risk Factors. *JAMA*, 310(6):591–606, 08 2013.

[178] Anne-Marie Palsson. Does the degree of relative risk aversion vary with household characteristics? *Journal of economic psychology*, 17(6):771–787, 1996.

[179] Belluck Pam and Robbins Rebecca. F.d.a. approves alzheimer's drug despite fierce debate over whether it works, 06 2021.

[180] Belluck Pam and Robbins Rebecca. Three f.d.a. advisers resign over agency's approval of alzheimer's drug, 06 2021.

[181] N. Pardi, M. J. Hogan, F. W. Porter, and D. Weissman. mrna vaccines – a new era in vaccinology. *Nature Reviews Drug Discovery*, 17(4):261–79, 2018.

[182] H. Pezeshk. Bayesian techniques for sample size determination in clinical trials: a short review. *Statistical methods in medical research*, 6(12):489–504, 2003.

[183] S. A. Plotkin, A. A. F. Mahmoud, and J. Farrar. Establishing a global vaccine-development fund. *New England Journal of Medicine*, 373(4):297–300, 2015.

[184] Stanley A Plotkin and Neal Halsey. Accelerate Coronavirus Disease 2019 (COVID-19) Vaccine Rollout by Delaying the Second Dose of mRNA Vaccines. *Clinical Infectious Diseases*, 01 2021. ciab068.

[185] N. Pouratian, S. Thakkar, W. Kim, and J. Bronstein. Deep brain stimulation for the treatment of parkinson's disease: efficacy and safety. *Degenerative Neurological and Neuromuscular Disease*, 2:107–117, 2012.

[186] . Project ALPHA. Estimates of clinical trial probabilities of success (pos). `https://projectalpha.mit.edu/pos/`, 2021. Accessed: 2022-04-14.

[187] Esther S. Pronker, Tamar C. Weenen, Harry Commandeur, Eric H. J. H. M. Claassen, and Albertus D. M. E. Osterhaus. Risk in vaccine research and development quantified. *PLOS ONE*, 8(3):1–7, 03 2013.

[188] Kuntara Pukthuanthong, Richard Roll, and Avanidhar Subrahmanyam. A protocol for factor identification. *The Review of Financial Studies*, 32(4):1573–1607, 2019.

[189] H. Qi and D. Sun. A quadratically convergent newton method for computing the nearest correlation matrix. *SIAM Journal on Matrix Analysis and Application*, 28(2):360–85, 2006.

[190] G. Rapeport, E. Smith, A. Gilbert, A. Catchpole, H. McShane, and C. Chiu. Sars-cov-2 human challenge studies – establishing the model during an evolving pandemic. *New England Journal of Medicine*, 385(11):961–4, 2021.

[191] Evan L Ray and et al. Ensemble forecasts of coronavirus disease 2019 (covid-19) in the u.s. *medRxiv*, 2020.

[192] Antonio Regalado. The chart that shows how we'll get back to normal, 12 2020.

[193] Reuters. Researchers urge delaying pfizer vaccine's second dose as first highly effective, 02 2021.

[194] Reuters. Spain mulls delaying second dose of moderna and pfizer shots, el mundo reports, 04 2021.

[195] Rick Rojas and Giulia McDonnell Nieto del Rio. The storm is hampering coronavirus vaccination efforts., 02 2021.

[196] Stephen A Ross. The arbitrage theory of capital asset pricing. *Journal of Economic Theory*, 13(3):341–360, 1976.

[197] . RTI and . KHI. Regression paper. Manuscript submitted for publication., 2022.

[198] David Sanger and Sharon LaFraniere. As vaccine production ramps up, the white house announced the weekly supply for states will again increase., 04 2021.

[199] Robert Schlaifer and Howard Raiffa. *Applied statistical decision theory*. Royal Statistical Society, Wiley, 1961.

[200] William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk*. *The Journal of Finance*, 19(3):425–442, 1964.

[201] Billie Jean Shaw. Some south carolina patients skipping second covid vaccine dose, 03 2021.

[202] M. Sheldon. Overview of various components of the science of patient input. *Clinical Journal of the American Society of Nephrology*, 16(4):634–635, 2021.

[203] Affan Shoukat, Chad R. Wells, Joanne M. Langley, Burton H. Singer, Alison P. Galvani, and Seyed M. Moghadas. Projecting demand for critical care beds during covid-19 outbreaks in canada. *CMAJ*, 192(19):E489–E496, 2020.

[204] K. W. Siah, Q. Xu, K. Tanner, O. Futer, J. J. Frishkopf, and A. W. Lo. Accelerating glioblastoma therapeutics via venture philanthropy. *Drug Discovery Today*, 26(7):1744–9, 2021.

[205] Sheryl Stolberg. Biden raises daily vaccination target and extends travel bans, 01 2021.

[206] Jenny Strasburg. Astrazeneca vaccine shown to curb covid-19 transmission, preliminary data suggest, 02 2021.

[207] Jenny Strasburg. Can covid-19 vaccines' second dose be delayed?, 01 2021.

[208] G. T. Szabo, A. J. Mahiny, and I. Vlatkovic. Covid-19 mrna vaccines: Platforms and current developments. *Molecular Therapy*, 30:5, 2021.

[209] M. E. Tarver and C. Neuland. Integrating patient perspectives into medical device regulatory decision making to advance innovation in kidney disease. *Clinical Journal of the American Society of Nephrology*, 16(4):636–638, 2021.

[210] . The White House. Press briefing by white house covid-19 response team and public health officials. https://www.whitehouse.gov/briefing-room/press-briefings/2021/04/05/press-briefing-by-white-house-covid-19-response-team-and-public-health-officials-24/, 04 2021. Accessed: 2021-04-19.

[211] John Tozzi. Second-shot crunch leaves many without complete immunizations, 02 2021.

[212] Ashleigh Tuite, Lin Zhu, David Fisman, and Joshua Salomon. Alternative dose allocation strategies to increase benefits from constrained covid-19 vaccine supply. *Annals of Internal Medicine*, 0(0):null, 2021. PMID: 33395334.

[213] Merryn Voysey and et al. Single-dose administration and the influence of the timing of the booster dose on immunogenicity and efficacy of chadox1 ncov-19 (azd1222) vaccine: a pooled analysis of four randomised trials. *The Lancet*, 397(10277):881–891, 2021.

[214] J. Vu, B. Kaplan, S. E. Chaudhuri, M. Mansoura, and A. W. Lo. Financing vaccines for global health security. *Journal of Investment Management*, 20(2):1–17, 2022.

[215] Fei Wang and Alan E. Gelfand. A simulation-based approach to bayesian sample size determination for performance under a given model and for separating models. *Statistical Science*, 17(2):193–208, 2002.

[216] Frances M. Weaver, Kenneth Follett, Matthew Stern, Kwan Hur, Crystal Harris, William J. Marks, Johannes Rothlind, Oren Sagher, Domenic Reda, Claudia S. Moy, Rajesh Pahwa, Kim Burchiel, Penelope Hogarth, Eugene C. Lai, John E. Duda, Kathryn Holloway, Ali Samii, Stacy Horn, Jeff Bronstein, Gatana Stoner, Jill Heemskerk, Grant D. Huang, and for the CSP 468 Study Group. Bilateral Deep Brain Stimulation vs Best Medical Therapy for Patients With Advanced Parkinson Disease: A Randomized Controlled Trial. *JAMA*, 301(1):63–73, 01 2009.

[217] Noah Weiland and Sharon LaFraniere. The white house says vaccine supply is going up to 13.5 million doses a week., 02 2021.

[218] Andrew R. Willan and Eleanor M. Pinto. The value of information and optimal clinical trial design. *Statistics in Medicine*, 24(12):1791–1806, 2005.

[219] Danielle D Winchester, Sandra J Huston, and Michael S Finke. Investor prudence and the role of financial advice. *Journal of Financial Service Professionals*, 65(4), 2011.

[220] C. H. Wong, K. W. Siah, and A. W. Lo. Estimation of clinical trial success rates and related parameters. *Biostatistics*, 20(2):273–86, 2019.

[221] J. Woodcock and L. M. LaVange. Master protocols to study multiple therapies, multiple diseases, or both. *New England Journal of Medicine*, 377(1):62–70, 2017.

[222] . World Health Organization (WHO). 2019 yellow fever icg annual meeting report. `https://www.who.int/publications/i/item/9789240008236`, 2020. Accessed: 2022-04-14.

[223] . World Health Organization (WHO). A brief guide to emerging infectious diseases and zoonoses. `https://apps.who.int/iris/handle/10665/204722`, 2021. Accessed: 2022-04-14.

[224] Qingyang Xu, Joonhyuk Cho, Zied Ben Chaouch, and Andrew W. Lo. Incorporating patient preferences and burden-of-disease in evaluating als drug candidate amx0035: A bayesian decision analysis perspective. Manuscript under Review, 2022.

[225] Lu Zhang. Lu zhang's data library. http://global-q.org/testingportfolios.html. Accessed: 07-19-2022.