UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title

Feedback-Driven Methods for Evaluating and Individualizing Treatments

Permalink

https://escholarship.org/uc/item/3868164f

ISBN

9798173787606

Author

Gradu, Paula Nicoleta

Publication Date

2026-08-10

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

Feedback-Driven Methods for Evaluating and Individualizing Treatments

by

Paula Nicoleta Gradu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering and Computer Sciences

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Benjamin Recht, Co-Chair
Professor Michael I. Jordan, Co-Chair
Assistant Professor Ahmed Alaa

Summer 2026

Feedback-Driven Methods for Evaluating and Individualizing Treatments

Copyright 2026
by
Paula Nicoleta Gradu

Abstract

Feedback-Driven Methods for Evaluating and Individualizing Treatments

by

Paula Nicoleta Gradu

Doctor of Philosophy in Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Benjamin Recht, Co-Chair

Professor Michael I. Jordan, Co-Chair

Despite its many merits, the standard evidence-based approach to treatment does not naturally account for physiological and preferential differences across individuals. Currently, the main recourse clinicians have to address this limitation is to integrate patient and provider experiences into decision-making in a heuristic manner. This dissertation explores how the rich resource of *patient feedback* can be formally and explicitly harnessed to design novel methods for evaluating and individualizing treatments. The contributions aim to combine the rigour and precise theoretical guarantees of existing evidence-based approaches with the flexibility and adaptivity of more heuristic, non-statistical clinical decision making.

The first part focuses on two novel methods for evaluating treatments, one pre-deployment and the other post-deployment. On the pre-deployment front, I detail a novel sequential experiment design which iteratively uses the feedback of each patient to adjust the allocation probability of treatment vs. placebo so as to maximize statistical power. On the post-deployment end, I cover methodology that uses feedback in the form of incident reports to identify whether there are subgroups, defined by any combination of relevant features, that have disproportionately high risk of certain side effects relative to the general population.

The second part focuses on new frameworks for feedback-driven treatment protocols for two important healthcare applications which require dosage adjustments over an individual patient's trajectory. The first application considered is the problem of tapering patients off of dependency-inducing psychiatric medications or addictive substances, where individual feedback on withdrawal intensity leads to a protocol that adapts to each individual's particular circumstances and physiology. The second application is the problem of adjusting the doses of the standard, evidence-based chemotherapy regimen when confronted with individual patients exhibiting severe side effects, or, alternatively, in the setting of palliative care where the patient may have lower tolerance for quality of life decrease due to treatment.

To my mother, my father, and my cat,
to my advisors Elad, Mike and Ben,
and to my friends from all over the world.

Contents

Contents **ii**

1 Introduction **1**

I Evaluating Treatments 3

2 A Sequential Experiment Design for Adaptive Neyman Allocation **4**
2.1 Related Work 5
2.2 Preliminaries 6
2.3 Neyman Design: The Infeasible Non-Adaptive Ideal 8
2.4 Adaptive Neyman Allocation: An Online Optimization Approach 10
2.5 Inference in Large Samples 13
2.6 Notes on Alternative Designs and Ethical Considerations 15
2.7 Numerical Simulations 17
2.8 Deferred Proofs 21

3 A Reporting-Based Framework for Post-Deployment Monitoring **48**
3.1 Related Work 49
3.2 Model, Notation, and Preliminaries 50
3.3 From Preponderance To Relative Risk 52
3.4 Algorithms for Identifying Subgroups with High Reporting Overrepresentation 53
3.5 Real-World Application: Myocarditis from COVID-19 vaccines 57
3.6 Notes on Practical Considerations 60
3.7 Deferred Proofs 61

II Individualizing Treatments 70

4 Online Control for Adaptive Tapering of Medications **71**
4.1 Related Work 72
4.2 Problem Setting and Preliminaries 74
4.3 Optimal Tapering With Full System Knowledge 77

4.4 Tapering By Integral Control 80
4.5 Numerical Simulations 84
4.6 Generalization to non-linear opponent processes 88

5 A Practical Control-Theoretic Framework for Taxonomizing Equipoise in Chemotherapy Dosing 91
5.1 Proposed Framework 92
5.2 Results for Orthogonally Toxic Drugs 94
5.3 Difficulties Arising for Cross-Toxic Drugs 97
5.4 Identifying Sources of Variability 98
5.5 Validation on Retrospective Data 99
5.6 Numerical Simulations 101

Bibliography 104

Acknowledgments

First of all I would like to thank my advisors Elad, Mike, and Ben, in chronological order. As I always say, I would never be writing this dissertation without Elad noticing me while taking his class on Optimization for Machine Learning and bringing me into the world of research and academia. Online optimization is at the core of all developments covered in this dissertation, as well as the framework I continue to understand all sorts of problems through. In addition to this, Elad has always been there for me, continuing to support and advise me even after leaving Princeton. I will forever be grateful to know there is someone in the field on whom I can always rely, no matter the distance, time passed, or potential divergences in our research paths.

Next, I would also not be *here* without Mike's conviction that Berkeley was the place I should come do my PhD. The email from Mike was the first I got after sending out applications and the first professor I got to meet (via Zoom) outside of Princeton. From our very first call, Mike has been very real, empathetic, and really intuitive about very big picture considerations within academia. I feel so privileged to have had someone as legendary as Mike take such a genuine interest in me as a human, and be so down to earth in each interaction. I still cannot believe he even took the time to read a book I recommended by a Romanian author. Through SAIL and Mike I got introduced to so many brilliant and sweet people in our field. It truly made me feel like I am part of a large extended family, something I will carry with me and forever be grateful for.

Finally, but certainly not least, I want to thank Ben for being my partner in crime when it comes to the vision of this dissertation. Ben has an incredible ability to spot any trace of inauthenticity or mindlessness in research, to identify any part where you may have unknowingly cut a little corner. This has pushed me to do research that I feel is truly *honest* and especially authentic, to pursue a vision I thought I'd only have the courage to later on in my career, to disregard any incentives or desires to please and focus only on what's real. Beyond the satisfaction this has brought me in research, it has also pushed me to be even more honest with myself regarding why I am doing what I am doing, and to walk my own path in life even when it goes against what others may be expecting of me.

Moving on from advisors to an almost advisor, I would like to thank Ahmed for inviting me to give a talk in his group at the end of 2024 and for making the connection between my work on tapering and questions regarding dose reductions in chemotherapy regimens, motivating the last work covered in this dissertation. Beyond leading me to work on cancer treatment questions which has been a goal of mine ever since writing my college applications and the reason for my interest in machine learning, Ahmed has been such a present figure for the past year and a half, always shocking me with how available, reliable, helpful and inclusive he is willing to be with someone who is not even his student.

I also want to thank Travis for being the first real doctor I've had the privilege of actually working with, for his superhuman availability, and for having faith in the potential of a control-theoretic approach to real questions he and his peers were dealing with when treating

patients. In addition to the invaluable help on the research, I have also learnt so, so much about cancer, chemotherapy and medical practice themselves from him.

Beyond the help of all my advisor figures, the work covered in this dissertation is also only possible because of my wonderful collaborators. First of all I have to thank Jessica and Jenna for being my long-suffering sisters carrying so much of the load and stress of the projects covered in Chapters 3 and 5, for the never-ending multiple hour joint research or writing sessions in Berkeley Way West and Soda Hall. I also want to thank my PhD twin sister Deb, who is also the reason we joined forces to do the work in Chapter 3, and also with whom I have had so, so many chats about what meaningful and impactful research is and how to go about it. In terms of collaborators, I finally want to thank Chris and Tijana for being semi-advisors on the projects I worked with them on early during my PhD, as well as for being one of the girls (Chris) and a mama cat (Tijana).

In addition to my collaborators, my PhD experience has also been defined by my dear yacht girls Mihaela, Deb, Jess, and Holly, as well as Ian, Ayesha, Jenna and Aneesha who joined our weekly oftentimes philosophical lunches. On the SAIL side, I am so grateful for the warmth and availability of Jivat, for the inspirational love (and wedding) of Tijana and Anastasios, for Karl, for Lydia, for Ezinne, for Serena, for Antoine and Aymeric, and for everyone else who's been a part of this grand family in any shape or form. Outside of my two groups, I want to thank Katie for being so sweet and my first peer collaborator when I first joined, and to everyone else in Sergey and Pieter's groups who were all so sweet despite our different research directions. Michael and Brent too, whom I got to hang out a lot with in the Walnuts house after we ended up as roommates.

My last but deepest gratitude goes to my mother and my father for their endless support and understanding. Even as I am writing this dissertation, my mom is packing all my stuff from my room and my dad is forgiving yet another postponed phone call. I am so happy I got to celebrate this last graduation, the end of this long, long overseas educational journey, together with both of them. I can't wait for us to be only a direct flight away from now on.

Chapter 1

Introduction

While evidence-based medicine has led to impressive advancements in the safety and efficacy of treatments, it oftentimes translates to a one-size-fits-all approach that does not naturally account for physiological and preferential differences across individuals. This is partly the result of trials being limited to identifying the *average* effect of a treatment in a population, most commonly for a pre-specified specific dose or dosing rule. We thus have the dilemma:

Can we design methodology that is flexible enough to account for deviations from expected outcomes or different patient desiderata yet also benefits from rigorous safety guarantees?

This thesis explores how this dilemma can be tackled by making use of the rich resource of patient *feedback*. The resulting works formally and explicitly harness feedback to design novel methodology for *evaluating* and *individualizing* treatments. These advances aim to combine the rigour and precise theoretical guarantees of existing evidence-based approaches with the flexibility and adaptivity of more heuristic, non-statistical clinical decision making.

The first part focuses on methodology for evaluating treatments. The most classical question here is that at the root of all drug development and clinical trials: does this medication produce a statistically significant improvement over placebo or standard of treatment? This, in addition to safety profiling, is at the root of whether a new drug will be approved or not. It is also the problem which has been by far most tackled in the field of statistics and even machine learning theory, as it is most cleanly amenable to theoretical advancements. However, despite its popularity, the particular question of optimal treatment allocation (dating back to Neyman in 1934 [79]) remained unsolved. While in his work, Neyman demonstrates that sampling from treatments proportional to the within-treatment outcome variance will minimize the variance of standard estimators, such a design is unfeasible in the standard batch approach to clinical trials. In Chapter 2, I detail a novel sequential experiment design which attains this goal by iteratively using the feedback of each patient to adjust the allocation probability of treatment vs. placebo so as to maximize statistical power.

However, while much less studied, evaluation of treatments does not, or at least should not, end after approval! Due to the limitations of clinical trials, it is very possible that the

initial experiments miss more rare side effects, most especially if these side effects only affect a particular section of the population. To this end, the U.S. Food and Drug Administration maintains databases of reported adverse events for medical devices [30, 99], pharmaceuticals [84, 98] and vaccines [18, 100]. In Chapter 3, I introduce a framework for systematically using such incident report feedback to identify whether there are subgroups, defined by any combination of relevant features, that have disproportionately high risk of certain side effects relative to the general population. The hope is this provides a step toward feedback-driven automatic post-deployment evaluation/monitoring of treatments.

Beyond missing rare or subgroup-specific side effects or treatment failures, trials are also ill-suited for evaluating treatment protocols that inherently require dose adjustments over a patient's clinical trajectory. Thus, the second part of this dissertation focuses on methodology for using patient feedback to sequentially determine doses in a way which optimizes for some trajectory-level desiderata. Chapter 4 covers the problem of tapering patients off of dependency-inducing psychiatric medications or addictive substances. Prior trials on tapering protocols treat an entire dose schedule as a fixed rule: for each substance, a decay shape (e.g., linear or exponential) and fixed rate is chosen and is tested on all participants. This is inherently unsuitable for tapering as it does not account for individual physiological or preferential differences. A key issue is the lack of adaptivity to the feedback a patient would provide in the clinic regarding their day-to-day felt withdrawal intensity. By formulating tapering as an optimal control problem, we arrive at a feedback-driven protocol which provably maintains an individual's well-being during the tapering process.

Finally, in Chapter 5, I look at the application of dose reductions in chemotherapy. Here, the question of how to adjust to individuals poses further complications than in the problem of tapering. At the root of the added complexity lies the multi-drug nature of chemotherapy regimens. In this last chapter I show how an online control lens can shed light on the sources of clinical equipoise in chemotherapy dose reductions, as well as provide formal explanations for which existing practitioner heuristics are more conducive to the desired goal of staying as close as possible to the standard, evidence-based chemotherapy regimen while preventing severe, life-threatening side effects or, in the case of palliative care, preventing the treatment itself from negatively affecting quality of life beyond what is tolerable to each patient.

This dissertation is based on works co-authored with Ahmed Alaa, Jessica Dai, Jenna Fields, Chris Harshaw, Deborah Raji, Benjamin Recht, and Travis Zack [35, 26, 28].

Part I

Evaluating Treatments

Chapter 2

A Sequential Experiment Design for Adaptive Neyman Allocation

From medicine and public health to economics and public policy, randomized control trials are used in a variety of disciplines to investigate causal effects. Typically, treatment is assigned in a non-adaptive manner, where assignments are determined before any outcomes are observed. A sequential feedback-based experimental approach, which adaptively assigns treatment based on previously observed outcomes, offers the possibility of more precise or high powered estimates of relevant causal effects. While adaptive experiments have been run to develop clinical therapies for breast cancer [10], evaluate incentives to reduce partisan bias [80], and evaluate customer acquisition via online advertising [91], to name a few, their natural suitability for incorporating outcome feedback to improve power has not yet been fully explored, either by practitioners or theoreticians.

In this chapter, we tackle the above desideratum. To do so, we begin by formalizing the statistical problem at hand. First, note that an optimal non-adaptive experimental design which minimizes variance of an estimator will depend on the unknown potential outcomes, rendering it infeasible to run. However, by adaptively choosing treatment assignments in a sequential manner based on observed outcomes, we can hope to guarantee that the variance of the estimator under the adaptive design converges to the optimal non-adaptive variance. The problem of *Adaptive Neyman Allocation* then is to construct such an adaptive design which guarantees the variance converges to the (infeasible) optimal non-adaptive design.

An experimental design which sufficiently addresses the Adaptive Neyman Allocation problem offers the advantage of higher statistical power, relative to a broad class of fixed experimental designs. Practically speaking, this means that either smaller confidence intervals are obtained for a given number of experimental units, or that fewer units are required to achieve confidence intervals of a given length. In practice, this means that investigating causal effects can be cheaper—in terms of time, money, and other valuable resources—when adaptive experiments are run. Although several experimental designs have been proposed for this purpose [38, 14], none have provided formal guarantees that the optimal non-adaptive variance can be achieved and the effectiveness of such designs has recently been called into

question Cai and Rafi [17].

*This chapter is based on joint work with Jessica Dai and Christopher Harshaw, previously released as "*Clip-OGD*: An Experimental Design for Adaptive Neyman Allocation in Sequential Experiments" [26].*

2.1 Related Work

We work within the potential outcomes framework for causal inference Neyman [78], Rubin [89], and Imbens and Rubin [53]. The idea of optimal treatment allocation dates back to Neyman [79], where he demonstrates that sampling from treatments proportional to the within-treatment outcome variance will minimize the variance of standard estimators. Unfortunately, this type of design is not practically feasible when little is known about the statistics of outcomes from each treatment. Robbins [86] highlights adaptive sampling as one of the more pressing open statistical problems at the time. In Chapter 5, Solomon and Zacks [95] presents a survey of adaptive designs for survey sampling, but from a Bayesian perspective. More recently, Hahn, Hirano, and Karlan [38] proposed a two stage design in a super-population setting, where data is uniformly collected from both arms in the first stage, statistics of the treatment arm are estimated, and a fixed probability derived from estimated statistics is used in the second stage. They derive the limiting distribution of the effect estimator under the two-stage design, which has a variance that is similar to, but asymptotically bounded away from the optimal Neyman variance. In a design-based setting, Blackwell, Pashley, and Valentino [14] propose a similar two-stage approach and, through simulations, provide practical guidance on how to choose the length of the first stage. Although both of these works are motivated by achieving the Neyman variance, neither formally show that this is possible under the two-stage design.

While the goal in this work is to increase the precision of treatment effect estimates, a variety of response-adaptive designs have been developed for various objectives, including reducing mean total sample size [42] and reduction of harm reduction in null hypothesis testing [88]. Eisele [29] proposes the Doubly Adaptive Coin Based Design, which is a meta-algorithm for targeting various allocation proportions when outcomes are drawn i.i.d. from an exponential family. Hu and Rosenberger [52] critiques many response-adaptive designs as being "myopic strategies" which have "adverse effects on power", providing an asymptotic framework by which to judge adaptive design when the outcomes are i.i.d. and binary. This asymptotic evaluation framework was extended to continuous outcomes by Zhang and Rosenberger [112]. An additional line of work has developed adaptive Bayesian methods for subgroup identification [109].

Causal inference under adaptively collected data has seen a variety of recent developments which are adjacent to, but distinct from, the problem of Adaptive Neyman Allocation. One line of research has been to construct estimators via re-weighting which ensure consistency and normality when data is collected via bandit algorithms [37, 110, 111]. A second line of research has been to provide inferential methods which are valid under data-dependent

stopping times [104, 51, 39]. Finally, Offer–Westort, Coppock, and Green [80] propose an adaptive experimental design for improved selective inference, when only the effect of the best performing treatment is to be inferred.

2.2 Preliminaries

The sequential experiment takes place over T rounds, where we assume that T is fixed and known to the experimenter. At each iteration $t \in [T]$, a new experimental unit (e.g. clinical participant), enters into the experiment, so that there are T units in total. In an abuse of notation, we identify units with their respective round $t \in [T]$. The experimenter assigns a (random) treatment $Z_t \in \{0, 1\}$ (e.g. drug or placebo) to the experimental unit. The unit has two real-valued potential outcomes $y_t(1), y_t(0)$ which are unknown to the experimenter and represent the unit's measured response to the treatment assignments (e.g. measured heart rate). The term "potential" is used here because while only one treatment is assigned and thus only one outcome is observed, both outcomes have the potential to be observed. At the end of the round, the experimenter sees the observed outcome $Y_t = \mathbf{1}[Z_t = 1]y_t(1) + \mathbf{1}[Z_t = 0]y_t(0)$.

2.2.1 Potential Outcomes Framework

In this work, we adopt a *design-based framework* where the sequence of potential outcomes $\{y_t(1), y_t(0)\}_{t=1}^T$ is deterministic and the only source of randomness is treatment assignment itself. In particular, we place no assumption on the homogeneity of the outcomes: they are not necessarily related to each other in any systematic way. Although the potential outcomes are deterministic, we introduce finite population analogues of various statistics. Define the finite population second moments $S(1)$ and $S(0)$ and correlation of the treatment and control outcomes ρ to be

$$S(1)^2 = \frac{1}{T}\sum_{t=1}^{T} y_t(1)^2 \ , \quad S(0)^2 = \frac{1}{T}\sum_{t=1}^{T} y_t(0)^2 \ , \quad \text{and} \quad \rho = \frac{\frac{1}{T}\sum_{t=1}^{T} y_t(1)y_t(0)}{S(1)S(0)} \ .$$

Observe that the correlation between treatment and control outcomes is bounded $\rho \in [-1, 1]$. Although we refer to ρ as the correlation, it also known as the cosine similarity and is generally not equal to the Pearson correlation coefficient. We remark that although the potential outcomes $y_t(1)$ and $y_t(0)$ are deterministic, the observed outcome Y_t is random, as it depends on random treatment assignment. The natural filtration according to these rounds is denoted as $\mathcal{F}_1 \dots \mathcal{F}_T$, so that $\mathcal{F}_t$ captures all randomness before the sampling of Z_t, i.e. the treatments assigned and outcomes observed in previous rounds.

In this sequential setting, the mechanism for random treatment assignment can incorporate observed outcomes from previous experimental rounds. This treatment mechanism, referred to as the *experimental design*, is selected by and thus known to the exper-

imenter. Formally, the experimental design is a sequence of functions $\{\Pi_t\}_{t=1}^T$ with signature $\Pi_t : (\{0,1\} \times \mathbb{R})^{t-1} \to [0,1]$ such that treatment is assigned as $\Pr(Z_t = 1 \mid \mathcal{F}_t) = \Pi_t(Z_1, Y_1, \dots Z_{t-1}, Y_{t-1})$. We denote $P_t = \Pr(Z_t = 1 \mid \mathcal{F}_t)$ as the (random) probability of treatment assignment at iteration t, given previously observed treatment assignments and outcomes.

The causal estimand of interest is the *average treatment effect*, defined as

$$\tau = \frac{1}{T} \sum_{t=1}^{T} y_t(1) - y_t(0) \ .$$

The average treatment effect captures the average counterfactual contrast between a unit's outcomes under the two treatment assignments. For example, this could be the average contrast of a clinical participant's heart rate under the drug or placebo. Individual treatment effects are defined as $\tau_t = y_t(1) - y_t(0)$, but they cannot be estimated without strong additional assumptions, as only one outcome is observed.

A standard estimator of the average treatment effect is the Horvitz–Thompson estimator, which weights observed outcome by the probability of their observation [77, 50]. For adaptive designs, the standard Horvitz–Thompson estimator is infeasible because the marginal probability of treatment assignment $\Pr(Z_t = 1)$ depends on the unknown potential outcomes. For this reason, we investigate the *adaptive Horvitz–Thompson estimator*, which uses the random (observed) treatment probabilities used at each iteration.

$$\hat{\tau} \triangleq \frac{1}{T} \sum_{t=1}^{T} Y_t \Big(\frac{\mathbf{1}[Z_t = 1]}{P_t} - \frac{\mathbf{1}[Z_t = 0]}{1 - P_t} \Big) \ ,$$

where we recall that $P_t = \Pi_t(Z_1, Y_1, \dots Z_{t-1}, Y_{t-1})$ is the treatment probability under the experimental design given the observed data. When treatment assignments are non-adaptive and independent, then the adaptive Horvitz–Thompson estimator is the equivalent to the standard Horvitz–Thompson estimator. Such adaptively weighted estimators have been proposed previously in the literature, e.g. [15, 37]. Below, we provide positivity conditions under which the adaptive estimator is unbiased, and derive its variance.

Proposition 2.1. *If* $\min\{P_t, 1 - P_t\} > 0$ *almost surely for all* $t \in [T]$ *then the adaptive Horvitz–Thompson estimator is unbiased:* $\mathbb{E}[\hat{\tau}] = \tau$.

Proposition 2.2. *The variance of the adaptive Horvitz–Thompson estimator is*

$$T \cdot \mathrm{Var}(\hat{\tau}) = \frac{1}{T} \sum_{t=1}^{T} \Big(y_t(1)^2 \mathbb{E}\Big[\frac{1}{P_t}\Big] + y_t(0)^2 \mathbb{E}\Big[\frac{1}{1 - P_t}\Big] \Big) - \frac{1}{T} \sum_{t=1}^{T} \tau_t^2 \ .$$

2.2.2 Asymptotic Framework and Assumptions

Following the convention of design-based inference, we analyze statistical methods within an asymptotic framework [see e.g., 34, 68, 90]. This provides a formal basis for reasoning

about the performance of statistical methods as the sample size increases, giving meaning to conventional notions such as consistency and limiting distribution. Formally speaking, the asymptotic sequence of potential outcomes is a triangular array $\{\{y_{t,T}(1), y_{t,T}(0)\}_{t=1}^{T}\}_{T=1}^{\infty}$, which yields a sequence of estimands $\{\tau_T\}_{T=1}^{\infty}$ and, together with an appropriately specified sequence of experimental design, a sequence of estimators $\{\hat{\tau}_T\}_{T=1}^{\infty}$. Analysis which applies to a fixed T is said to be finite-sample (e.g. $\mathbb{E}[\hat{\tau}_T] = \tau_T$) whereas analysis which applies to the entire sequence is said to be asymptotic (e.g. $\tau_T - \hat{\tau}_T \xrightarrow{p} 0$). Although we use an asymptotic framework, we emphasize that the majority of our results are derived from finite-sample analysis and are merely interpreted through the lens of the asymptotic framework. We drop the subscript T for notational clarity.

The main regularity conditions we place on the sequence of potential outcomes is below.

Assumption 2.3. *There exist constants $0 < c \leq C$ with $c < 1$ such that for all T in the sequence:*

1. ***Bounded Moments**: $c \leq \left(\frac{1}{T}\sum_{t=1}^{T} y_t(k)^2\right)^{1/2} \leq \left(\frac{1}{T}\sum_{t=1}^{T} y_t(k)^4\right)^{1/4} \leq C \;\forall\; k \in \{0,1\}$.*
2. ***Bounded Correlation**: $\rho \geq -(1-c)$.*

The upper moment bound in Assumption 2.3 stipulates that the potential outcomes cannot grow too large with the sample size, while the lower moment bound is a type of non-degeneracy condition that prevents an increasingly large fraction of the outcomes going to zero. These assumptions are analogous to finite fourth moment and positive second moment assumptions in an i.i.d. setting. The bounded correlation assumption stipulates that the treatment and control outcomes are not exactly negatively correlated. In this work, we do not assume that these constants C and c are known to the experimenter; however, if the experimenter can correctly specify such bounds (perhaps knowing a priori the scaling of the outcomes) then some of the constant factors in our analysis can be improved. We emphasize here that Assumption 2.3 places no assumption on the order in which units arrive in the experiment. In this sense, Assumption 2.3 allows for arbitrary "non-stationarity" or "drift" in the potential outcomes over the experimental rounds. In the next section, these regularity assumptions will ensure that the Neyman variance converges to zero at the parametric rate.

2.3 Neyman Design: The Infeasible Non-Adaptive Ideal

The problem of Adaptive Neyman Allocation is to construct an adaptive experimental design that achieves nearly the same variance as an optimal non-adaptive experimental design, chosen with knowledge of all potential outcomes. The optimal non-adaptive design, referred to as the Neyman Design, is infeasible to implement because it depends on all potential outcomes, which are unknown to the experimenter at the design stage. The goal is that an adaptive experimental design—which can select treatment assignment based on observed

outcomes—can gather enough information to perform as well as the infeasible Neyman design.

In order to define the optimal non-adaptive design, we begin by defining the class of Bernoulli designs. Informally, the class of Bernoulli designs consists of non-adaptive designs where each unit receives treatment $Z_t = 1$ with probability p, independently of past treatment assignments and observations. Formally, this class is parameterized by a non-adaptive sampling probability $p \in [0, 1]$ such that for all $t \in [T]$, the treatment policy Π_t is a constant function whose value is p. Using Proposition 2.2, we can derive the variance of the Bernoulli design with parameter $p \in [0, 1]$ to be

$$T \cdot V_p = S(1)^2\Big(\frac{1}{p} - 1\Big) + S(0)^2\Big(\frac{1}{1-p} - 1\Big) + 2\rho S(1)S(0) \ .$$

From the above, we can see that in order to minimize the variance of the Horvitz–Thompson estimator under the Bernoulli design, we should set the sampling probability p so as to balance the square of the second moments of treatment and control outcomes. The Neyman Design is the Bernoulli design which minimizes the variance of the Horvitz–Thompson estimator. The corresponding optimal probability p^* and variance V_{N} are referred to as the Neyman probability and Neyman variance, respectively. The following proposition derives these quantities in terms of the potential outcomes.

Proposition 2.4. *The Neyman variance is $T \cdot V_N = 2(1+\rho)S(1)S(0)$, which is achieved by the Neyman probability $p^* = (1 + S(0)/S(1))^{-1}$.*

In order to quantify the reduction in variance achieved by the Neyman design, define the *relative Neyman efficiency with respect to* $p \in [0, 1]$ to be V_{N}/V_p. Intuitively, this ratio is a scale-free measure which captures the percent reduction in variance of the sequential Horvitz–Thompson estimator under the Neyman design. Formally, the equation for the relative Neyman efficiency is given below:

$$\frac{V_{\mathrm{N}}}{V_p} = 2(1+\rho)\left[\frac{S(1)}{S(0)} \cdot \frac{(1-p)}{p} + \frac{S(0)}{S(1)} \cdot \frac{p}{(1-p)} + \rho\right]^{-1} \ . \tag{2.1}$$

Consider the setting where outcomes are uncorrelated, and treatment outcomes are larger than control outcomes, e.g. $\rho = 0$, $S(1) = 4 \cdot S(0)$. In this case, the Neyman design is able to achieve less than half the variance of the uniform Bernoulli design (with $p = 1/2$): we can plug in to 2.1 to see that in this setting, we have $V_N/V_p = 0.47$. The improvement is larger if the experimenter makes erroneous assumptions about the relative magnitudes of the treatment and control outcomes and attempts to set p accordingly: for example, if the experimenter had set $p = 1/4$, incorrectly believing that $S(1) \leq S(0)$, then the Neyman allocation results in a sixfold improvement in variance. Blackwell, Pashley, and Valentino [14] derives qualitatively similar analysis of Neyman efficiency for stratified designs.

While the relative Neyman efficiency is helpful in determining the variance reduction afforded by the (infeasible) optimal Bernoulli design, it does not address the main question:

which adaptive experimental designs can guarantee similar variance reduction? In the next section, we propose a performance metric which better addresses this question.

2.4 Adaptive Neyman Allocation: An Online Optimization Approach

2.4.1 Neyman Ratio and Neyman Regret: New Performance Measures

Let V be the variance of the adaptive experimental design. We introduce our first performance measure of a sequential experimental design for Adaptive Neyman Allocation.

Definition 2.5. *The* Neyman ratio *of a sequential experimental design is* $\kappa_T = (V - V_N)/V_N$.

The subscript T in κ_T in included the reflect dependence of the number of rounds T. The Neyman ratio is motivated by the following relationship between the adaptive variance and the optimal Neyman variance:

$$V = \Big(\frac{V}{V_{\mathrm{N}}}\Big) \cdot V_{\mathrm{N}} = \Big(1 + \kappa_T\Big) \cdot V_{\mathrm{N}} \ . \tag{2.2}$$

Equation (2.2) shows that the adaptive design can recover the Neyman variance if and only if the Neyman ratio κ_T can be made arbitrarily small. For this reason, we propose the Neyman ratio as a performance measure of a sequential experimental design.

A natural question then becomes: how small can the Neyman ratio κ_T be made as the number of rounds T increases? To answer this question, we view the problem of minimizing the Neyman ratio through the lens of online optimization. To this end, we must re-express the variance of the sequential experimental design. For each round $t \in [T]$, define the cost function $f_t : [0, 1] \to \mathbb{R}$ as

$$f_t(p) = \frac{y_t(1)^2}{p} + \frac{y_t(0)^2}{(1-p)} \ .$$

Observe that by Proposition 2.2, the variance is given by $T \cdot \mathrm{Var}(\hat{\tau}) = \mathbb{E}[\frac{1}{T}\sum_{t=1}^{T} f_t(P_t)]$. This reformulation of variance does not allow us to minimize variance directly, for the usual reason that the outcomes, and thus the cost functions f_t, are not fully observed. On the other hand, our goal is only to show that the variance of the adaptive design is comparable to the Neyman variance.

Definition 2.6. *The* Neyman regret *of a sequential experimental design is*

$$\mathcal{R}_T = \sum_{t=1}^{T} f_t(P_t) - \min_{p \in [0,1]} \sum_{t=1}^{T} f_t(p) \ . \tag{2.3}$$

Recall that P_t is the random treatment probability at round t. The Neyman regret compares the accumulated costs $f_t(P_t)$ incurred by the adaptive design to the accumulated costs incurred by the optimal Bernoulli design which has access to all potential outcomes. The Neyman regret is random because the sequence $P_1, \dots P_T$ is random. The following theorem connects the expected Neyman regret to the Neyman ratio.

Theorem 2.7. *Under Assumption 2.3, the Neyman ratio is within a constant factor of the $1/T$-scaled expected Neyman regret: $\kappa_T = \Theta(\frac{1}{T}\mathbb{E}[\mathcal{R}_T])$.*

Theorem 2.7 demonstrates that the Neyman ratio can be made small by minimizing the expected Neyman regret in an online fashion. In particular, any sublinear bound on the expected Neyman regret ensures that the Neyman ratio goes to zero so that, in large samples, the adaptive design achieves the variance reduction of the optimal Neyman design. Any adaptive design which aims to achieve Neyman variance must, to some extent, minimize expected Neyman regret.

Fortunately, online optimization is a well-studied area with a rich source of techniques from which we may draw inspiration. However, to the best of our knowledge, existing regret minimization algorithms are not well-suited to minimizing the Neyman regret. For example, the multi-arm bandit literature typically defines regret in terms of a finite number of actions that can be taken [64] while Adaptive Neyman Allocation consists of a continuum of actions as $P_t \in [0, 1]$. This means that algorithms like UCB [6] and EXP3 [7] are not appropriate for Adaptive Neyman Allocation. Our cost objectives f_t and action space $[0, 1]$ are both convex, so the problem of Adaptive Neyman Allocation is an instance of Online Convex Optimization (OCO) [43]. Even so, the problem of minimizing Neyman regret is not immediately amenable to existing algorithms, which typically requires assumptions on the cost functions such as bounded gradients or known Lipschitz parameters. In this setting, the cost functions have gradients which blow up at the boundary and Lipschitz parameters cannot be guaranteed as they rely on the unknown heterogeneous potential outcomes. For these reasons, we must design a new algorithm specifically tailored to Adaptive Neyman Allocation.

2.4.2 Clip-OGD: A Variant of Online Stochastic Projected Gradient Descent

We present Clip-OGD, which aims to minimize the Neyman regret and thus recover the Neyman variance in large samples. The algorithm is based on the online stochastic projected gradient descent principle, but with a twist: the projection set continuously grows over the rounds. At each round t, a new treatment probability P_t is chosen by updating the previous sampling probability P_{t-1} in the negative (estimated) gradient direction of the previous cost, and then projecting to an interval $[\delta_t, 1 - \delta_t]$. Initially, this projection interval contains only the point $1/2$ and it grows as the rounds increase, allowing for larger amounts of exploitation in later rounds.

The gradient estimator G_t is obtained in the following way: the gradient of f_t at P_t is given as $f'(P_t) = -\frac{y_t(1)^2}{P_t^2} + \frac{y_t(0)^2}{(1-P_t)^2}$. Only one outcome is observed, so we used the adaptive Horvitz–Thompson principle using the conditional probability P_t to unbiasedly estimate the outcomes. CLIP-OGD is formally presented below as Algorithm 1, where the projection operator is defined as $\mathcal{P}_c(x) = \max\{c, \min\{x, 1-c\}\}$.

Algorithm 1: CLIP-OGD

Input: Step size η and decay parameter α
Initialize $P_0 \leftarrow 1/2$ and $G_0 \leftarrow 0$
for $t = 1 \dots T$ **do**
  Set projection parameter $\delta_t = (1/2) \cdot t^{-1/\alpha}$
  Compute new treatment probability $P_t \leftarrow \mathcal{P}_{\delta_t}(P_{t-1} - \eta \cdot G_{t-1})$
  Sample treatment assignment Z_t as 1 with probability P_t and 0 with probability $1 - P_t$
  Observe outcome $Y_t = \mathbf{1}[Z_t = 1]y_t(1) + \mathbf{1}[Z_t = 0]y_t(0)$
  Construct gradient estimator $G_t = Y_t^2\left(-\frac{\mathbf{1}[Z_t=1]}{P_t^3} + \frac{\mathbf{1}[Z_t=0]}{(1-P_t)^3}\right)$

Unlike the two-stage design of [38, 14], CLIP-OGD does not feature explicit explore-exploit stages, but rather performs both of these simultaneously. The trade-off is implicitly controlled through parameters η and α: smaller values of η limit the amount of that sampling probabilities can update and, likewise, larger values of α prevent extreme probabilities in earlier stages. Because the gradient of the cost functions are inversely proportional to the treatment probabilities, limiting the extremeness of the treatment probabilities in this way ensures that the gradient estimates do not increase at a fast rate. By appropriately setting input parameters, CLIP-OGD achieves $\widetilde{\mathcal{O}}(\sqrt{T})$ expected Neyman regret, where the $\widetilde{\mathcal{O}}(\cdot)$ notation hides sub-polynomial factors.

Theorem 2.8. *Under Assumption 2.3 the parameter values* $\eta = \sqrt{1/T}$ *and* $\alpha = \sqrt{5\log(T)}$ *ensure the expected Neyman regret of* CLIP-OGD *is asymptotically bounded:* $\mathbb{E}\big[\mathcal{R}_T\big] \leq \widetilde{\mathcal{O}}\big(\sqrt{T}\big)$.

Theorem 2.8 answers, in the affirmative, that it is possible to construct an adaptive experimental design whose variance recovers that of the Neyman variance, in large samples. Note that the amount of exploration (as given by the parameters η and α) should be increasing with T in order to recover these regret bounds. In Section 2.8.4, we show that CLIP-OGD is somewhat robust to different values of the decay parameter, i.e. for any value $\alpha > 5$, the expected regret will be sublinear. We also show that if the experimenter presumes to have correctly specified bounds C and c appearing in Assumption 2.3, then the step size can be modified to improve the constant factors in the Neyman regret bound, which may lead to improved performance in moderate sample sizes. We conjecture that the minimax rate for expected Neyman regret is $\mathcal{O}(\sqrt{T})$, but proving this is beyond the scope of this work—we

only remark that we do not know it to immediately follow from any existing regret lower bounds for OCO.

2.5 Inference in Large Samples

The proposed Clip-OGD was constructed to ensure that the variance of the adaptive Horvitz–Thompson estimator quickly approaches the Neyman variance. In this section, we provide confidence intervals for the average treatment effect which also enjoy reduced width compared to non-adaptive counterparts.

A necessary condition for variance estimation is that the variance itself cannot be going to zero too quickly. In design-based inference, it is common to directly posit a so-called "non-superefficient" assumption that $\mathrm{Var}(\hat{\tau}) = \Omega(1/T)$ [5, 66, 41]. The non-superefficiency assumption may be seen as an additional regularity assumption on the outcomes, e.g. preventing $y_t(1) = y_t(0) = 0$ for all $t \in [T]$. In this work, a similar lower bound on the rate of the adaptive variance is obtained through a different, perhaps more transparent, assumption on the expected Neyman regret.

Assumption 2.9. *The outcome sequence is not overly-fit to* Clip-OGD*:* $-\mathbb{E}[\mathcal{R}_T] = o(T)$.

While we have shown that $\mathbb{E}[\mathcal{R}_T] \leq \widetilde{\mathcal{O}}(\sqrt{T})$, the Neyman regret could in principle be negative if the adaptive design achieves variance which is strictly smaller than the best Bernoulli design. While this seems unlikely to happen for "typical" outcomes, it is not impossible. Assumption 2.9 rules out these edge-case settings. We suspect that Assumption 2.9 would not be necessary in an i.i.d. setting, but proving this seems beyond the scope of this work. As shown in Section 2.8.7, Assumptions 2.3 and 2.9 imply that the adaptive variance achieves the parametric rate: $\mathrm{Var}(\hat{\tau}) = \Theta(1/T)$.

2.5.1 Variance Estimation

In this section, we provide a variance estimator and show its stability in large samples. Rather than estimating the adaptive variance (which has no simple closed form), our approach is to estimate the Neyman variance directly. For an adaptive design achieving sublinear expected Neyman regret, these two quantities are asymptotically equivalent. In this way, our variance estimator may be appropriate not only for Clip-OGD, but for any adaptive design achieving sublinear expected Neyman regret.

Recall that the Neyman variance is given by $T \cdot V_{\mathrm{N}} = 2(1+\rho)S_1S_0$, where ρ is the outcome correlation, S_1 is the second moment of treatment outcomes and S_0 is the second moment of control outcomes. Unfortunately, the outcome correlation term is generally not estimable without strong assumptions in a design-based framework. Indeed, the difficulty is that terms like $y_t(1)y_t(0)$ are unobservable due to the fundamental problem of causal inference [53]. A common solution to the problem is to opt for a conservative estimate of the variance, which will ensure validity of resulting confidence intervals.

We propose estimating the following upper bound on the variance: $T \cdot \mathrm{VB} = 4S_0S_1$. This upper bound on the Neyman variance is tight (i.e. $\mathrm{VB} = V_{\mathrm{N}}$) when outcome correlation satisfies $\rho = 1$. For example, this occurs when all individual treatment effects are zero, i.e. $y_t(1) = y_t(0)$ for all $t \in [T]$. Conversely, the upper bound will become looser for smaller values of the outcome correlation. In this sense, our bound resembles both the Neyman bound and the Aronow-Samii bound [78, 4]. It may be possible to use the recent insights of Harshaw, Middleton, and Sävje [40] in order to construct variance bounds which are tight in other scenarios, but that is beyond the scope of this work. Our variance estimator is defined as

$$T \cdot \widehat{\mathrm{VB}} \triangleq 4\sqrt{\left(\frac{1}{T}\sum_{t=1}^{T} y_t^2 \frac{\mathbf{1}[z_t = 1]}{p_t}\right) \cdot \left(\frac{1}{T}\sum_{t=1}^{T} y_t^2 \frac{\mathbf{1}[z_t = 0]}{1 - p_t}\right)} \ ,$$

which is essentially a plug-in Horvitz-Thompson estimator for the second moments. Theorem 2.10 shows the error of the normalized variance estimator converges at a parametric rate.

Theorem 2.10. *Under Assumptions 2.3 and 2.9, and the parameters stated in Theorem 2.8, the error of the normalized variance estimator under* Clip-OGD *is* $T \cdot \widehat{VB} - T \cdot VB = \widetilde{\mathcal{O}}_p(T^{-1/2})$.

2.5.2 Confidence Intervals

The variance estimator may be used to construct confidence intervals for the average treatment effect. This offers experimenters standard uncertainty quantification techniques when running Clip-OGD. The following corollary shows that the resulting Chebyshev-type intervals are asymptotically valid.

Corollary 2.11. *Under Assumptions 2.3 and 2.9, and parameters stated in Theorem 2.8, Chebyshev-type intervals are asymptotically valid: for all* $\alpha \in (0, 1]$, $\liminf_{T\to\infty} \Pr(\tau \in \hat{\tau} \pm \alpha^{-1/2}\sqrt{\widehat{VB}}) \geq 1 - \alpha$.

While these confidence intervals are asymptotically valid under our regularity assumptions, they may be overly conservative in general. In particular, they will over cover when the Chebyshev tail bound is loose. We conjecture that the adaptive Horvitz–Thompson estimator under Clip-OGD satisfies a Central Limit Theorem, which would imply asymptotic validity of the narrower Wald-type intervals where $\alpha^{-1/2}$ scaling is replaced with the corresponding normal quantile, $\Phi^{-1}(1-\alpha/2)$. As discussed in Section 2.7, the adaptive estimator appears approximately normal in simulations. Until this is formally shown, we recommend experimenters exhibit caution when using Wald-type confidence intervals for the adaptive Horvitz–Thompson estimator under Clip-OGD.

2.6 Notes on Alternative Designs and Ethical Considerations

2.6.1 Explore-Then-Commit

Two-stage adaptive designs have been proposed for the purpose of variance reduction [38, 14]. Due to its similarities to algorithms in the bandits literature, we call these types of designs Explore-Then-Commit (ETC) [64]. At a high level, an Explore-then-Commit design runs the Bernoulli design with $p = 1/2$ for $T_0 \leq T$ iterations, uses the collected data to estimate p^* by $\widehat{p}^*$, and then runs the Bernoulli design with $p = \widehat{p}^*$ for the remaining $T_1 = T - T_0$ iterations. These ETC designs are conceptually simpler than Clip-OGD, and may be reasonable to apply in more restricted settings where changing the treatment probabilities is difficult or costly. However, we provide the following negative result which shows that they can suffer linear Neyman regret.

Proposition 2.12. *For all explore phase lengths T_0 satisfying $T_0 = \Omega(T^\epsilon)$ for some $\epsilon > 0$, there exist a class of potential outcomes sequences satisfying Assumption 2.3 such that the Neyman regret under Explore-then-Commit is linear: $\mathcal{R}_T = \Omega_p(T)$.*

The specific class of potential outcomes referenced in Proposition 2.12 is constructed explicitly in Section 2.8.10. ETC designs suffer larger variance when the estimated $\widehat{p}^*$ may be far from the true optimal probability p^*. In a design-based setting, this happens when the units in the explore phase are not representative of the entire sequence. Formulating conditions under which Explore-then-Commit designs achieve low Neyman regret is beyond the scope of this work, but the proof of Proposition 2.12 shows that additional regularity conditions on the order of the units will be required.

2.6.2 Multi Arm Bandit Algorithms

Multi Arm Bandit (MAB) algorithms are often used for adaptive decision making settings, from online advertising to product development. The goal of MAB algorithms is to minimize the outcome regret, which measures the contrast between the overall value obtained from the actions relative to the value of the best action. The outcome regret is conventionally defined as $\mathcal{R}_T^{\text{outcome}} = \max_{k\in\{0,1\}} \sum_{t=1}^T y_t(k) - \sum_{t=1}^T Y_t$. In certain contexts, minimizing outcome regret may be a more desirable goal than estimating a treatment effect to high precision. However, the following proposition illustrates that these two objectives are generally incompatible.

Proposition 2.13. *Let $\mathcal{A}$ be an adaptive treatment algorithm achieving sublinear outcome regret, i.e. there exists $q \in (0, 1)$ such that $\mathbb{E}[\mathcal{R}_T^{outcome}] \leq O(T^q)$ for all outcome sequences satisfying Assumption 2.3. Then, there exists a class of outcome sequences satisfying Assumption 2.3 on which $\mathcal{A}$ suffers super-linear Neyman regret, i.e. $\mathbb{E}[\mathcal{R}_T] \geq \Omega(T^{2-q})$.*

Proposition 2.13 demonstrates that the outcome regret and the Neyman regret cannot generally be simultaneously minimized. In particular, sublinear outcome regret implies that the variance of the estimator must converge slower than the $\Theta(1/T)$ parametric rate. This result contributes to a growing body of work which highlights trade-offs between various possible objectives in sequential decision making [16]. It is beyond the scope of this work to determine how such trade-offs ought to be resolved, though Section 2.6.3 discusses ethical considerations.

2.6.3 Ethical Considerations

There are—at least—two objectives when constructing an adaptive treatment allocation.

- **Minimizing Cumulative Regret**: give the "best" treatment to as many people as possible.
- **Minimizing Variance of the Effect Estimate**: estimate the effect of the treatment to as high precision as possible.

As we show in Proposition 2.13, these two objective are fundamentally incompatible: an adaptive design which aims to estimate the effect to high precision must assign treatment which has worse outcomes. Likewise, a design which seeks to maximize the utility of assigned treatments will not be able to reliably estimate causal effects to high precision. Which one of these is more ethically desirable depends on the purpose and the context of the experiment. For guidance on this ethical question, we turn to The Belmont Report.

In 1979, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research released the "Belmont Report", which has been one of the foundational texts for ethical guidance in research conducted with human subjects [85]. One of the three basic ethical principles laid out in the report is "beneficence" which is understood as the obligation the researcher has for the research to improve the general well-being of society, including those people in the experiment. The report writes:

> The Hippocratic maxim 'do no harm' has long been a fundamental principle of medical ethics. Claude Bernard extended it to the realm of research, saying that one should not injure one person regardless of the benefits that might come to others. However, even avoiding harm requires learning what is harmful; and, in the process of obtaining this information, persons may be exposed to risk of harm. Further, the Hippocratic Oath requires physicians to benefit their patients 'according to their best judgment.' Learning what will in fact benefit may require exposing persons to risk. The problem posed by these imperatives is to decide when it is justifiable to seek certain benefits despite the risks involved, and when the benefits should be foregone because of the risks.

From this perspective, it may be ethically advisable to use a variance minimizing adaptive design because it allows the researcher to learn the effect while subjecting fewer human subjects to the experimental treatments. In other words, a variance minimizing design allows researchers to learn what is harmful and what is beneficial while subjecting fewer human subjects to possible harm. An adaptive treatment plan which minimizes cumulative regret will ensure that minimal harm is done to subjects in the experiment, but will offer less certainty about the extent of the benefit or harm of the treatments. Such an approach will lead to less informative generalizable knowledge of treatment effects, possibly defeating the goal of the research study.

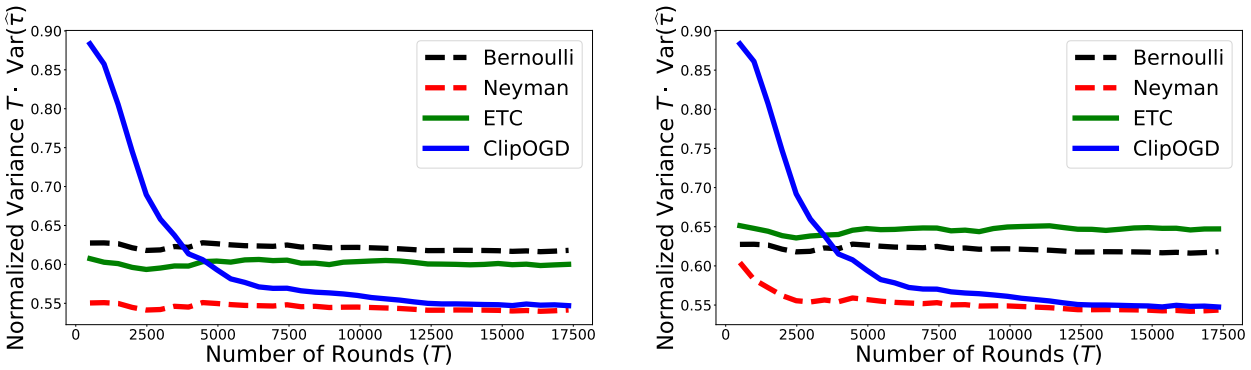
That being said, it is not our goal to suggest that one adaptive allocation plan is most ethical in all circumstances. Indeed, these ethical questions have no systematic answers which are generally applicable. It is the burden of the researchers to carefully "decide when it is justifiable to seek certain benefits despite the risks involved, and when the benefits should be foregone because of the risks." Our goal in this work is merely to provide improved statistical methodology which affords the researchers more choices when addressing these ethical questions.

2.7 Numerical Simulations

We evaluate the performance of Clip-OGD and Explore-then-Commit (ETC) for the purpose of Adaptive Neyman Allocation on the field experiment of Groh and McKenzie [36], which investigates the effect of macro-insurance on micro-enterprises in post-revolution Egypt[1]. The experimental units are 2,961 clients of Egypt's largest microfinance organization and the treatment was a novel insurance product. Several outcomes were recorded including whether the clients took on loans, introduced a new product or service, and the amount invested in machinery or equipment following treatment. To allocate treatment, Groh and McKenzie [36] use a non-adaptive matched pair experimental design. Our goal here is not to provide a new analysis of this study, but rather to construct a plausible experimental setting under which to evaluate adaptive experimental designs.

In our simulations, we focus on the numerical outcome "invested in machinery or equipment". The experimental data contains only observed outcomes, so we must impute the missing potential outcomes in order to simulate the experiment. We impute outcomes using the model $y_t(1) - y_t(0) = \tau + \gamma_t$, where $\tau = 90,000$ and $\gamma_1 \dots \gamma_T \sim \mathcal{N}(0, \sigma^2)$ are independent with $\sigma = 5,000$. This randomness is invoked only to impute potential outcomes, i.e. not re-sampled during each run of the experiment. In order to increase the sample size, we create a larger population by repeating this processes 5 times, which yields a total of $14,445$ units after those with missing entries are removed. Units are shuffled to appear in an arbitrary order and outcomes are normalized to be in the range$[0, 1]$.

[1]The following repository contains code for reproducing our simulations: `https://github.com/crharshaw/Clip-OGD-sims`.



(a) Original dataset

(b) First 100 units have flipped potential outcomes

Figure 2.1: Normalized Variance of Adaptive Estimator under Experimental Designs

Figure 2.1 presents two plots illustrating how the variance of the adaptive HT estimator varies with different designs. The x axis contains the number of rounds T and the y axis contains the normalized variance $T \cdot \mathrm{Var}(\hat{\tau})$ under the designs. For each value of T, we take the population to be the first T units in the sequence. Clip-OGD is run with the parameters recommended in Theorem 2.8 and ETC is run with $T_0 = T^{1/3}$ so that the exploration phase grows with T. The variance under Clip-OGD and ETC is estimated empirically from 50,000 runs of the experiment, while the variance under the Bernoulli and Neyman designs is computed exactly.

In Figure 2.1a, we observe that Clip-OGD requires about $T = 4,000$ samples to achieve variance equal to Bernoulli, but eventually converges to the Neyman variance. As discussed in Section 2.4.2, it may be possible to improve the convergence rate by incorporating knowledge of the outcome moments in the design parameters. On the other hand, ETC remains comparable with Bernoulli even for small values of T, but remains far away from the Neyman design for large samples. In Figure 2.1b, a similar simulation is run, except that the potential outcomes of the first 100 units are swapped, so that the first units have negative individual treatment effects. While this produces little effect on the performance of Clip-OGD, it substantially worsens the performance of ETC, which relies on the early outcomes to estimate an optimal treatment probability. In particular, ETC performs worse than Bernoulli under this minor modification—even in large samples—corroborating Proposition 2.12.

2.7.1 Confidence Intervals

We next evaluate the proposed confidence intervals, showing that Clip-OGD enjoys intervals of reduced width. We show that normal based intervals cover at the nominal level and provide further evidence that the estimator is asymptotically normal under Clip-OGD.

Table 2.1 presents the Chebyshev-based and Normal-based intervals for the Bernoulli design ($p = 1/2$) and Clip-OGD. We see that while Chebyshev over-covers, the normal-

	Chebyshev Width	Chebyshev Coverage	Normal Width	Normal Coverage
Bernoulli ($p = 1/2$)	0.0541	100%	0.0237	95.21%
Clip-OGD	0.0507	99.99%	0.0222	95.22%

Table 2.1: 95% Confidence Intervals for Bernoulli and Clip-OGD

based confidence intervals cover at the nominal level with reduced width for both designs. The relative Neyman efficiency on this dataset is somewhat close to 1, so that the reduction of the width of the confidence intervals afforded by Clip-OGD is present, though modest. The coverage of the normal-based confidence intervals provides further evidence supporting our conjecture that the adaptive Horvitz–Thompson estimator is asymptotically normal under Clip-OGD.

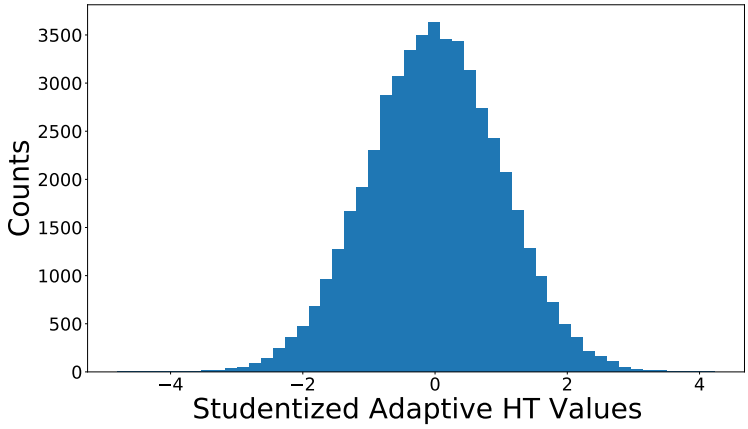


Figure 2.2: Histogram of Studentized Adaptive Horvitz–Thompson estimator under Clip-OGD ($T = 14,445$)

Figure 2.2 plots the histogram of the studentized adaptive Horvitz–Thompson estimator under Clip-OGD. By studentized, we mean that the histogram is plotting the draws of the random variable

$$Z = \frac{\tau - \hat{\tau}}{\sqrt{\mathrm{Var}(\hat{\tau})}} \ .$$

We estimate the standard deviation empirically from 50,000 runs of the experiment. The estimator is said to be asymptotically normal if $Z \xrightarrow{d} \mathcal{N}(0,1)$. Figure 2.2 provides evidence that asymptotic normality is likely to hold in this setting. Formally establishing asymptotic normality is beyond the scope of this work as it would involve very different analytic techniques than those used to establish sublinear Neyman regret.

2.7.2 Sensitivity to Step Size

We also explore through simulations how the performance of Clip-OGD depends on the step size.

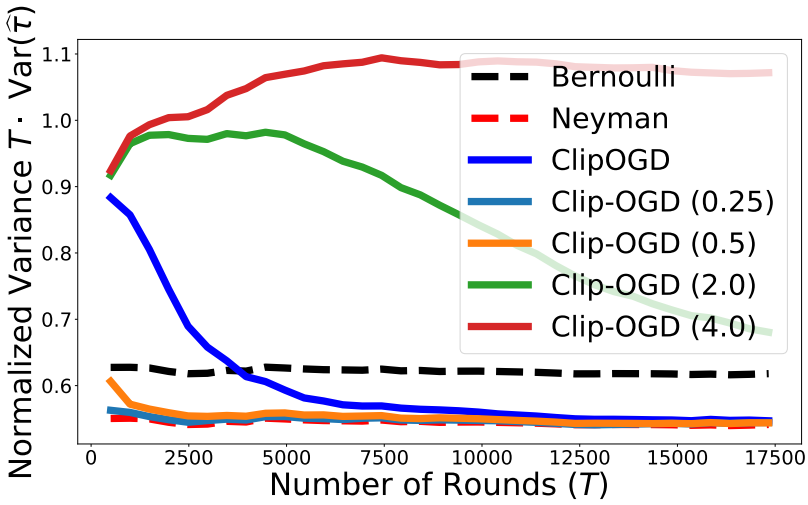
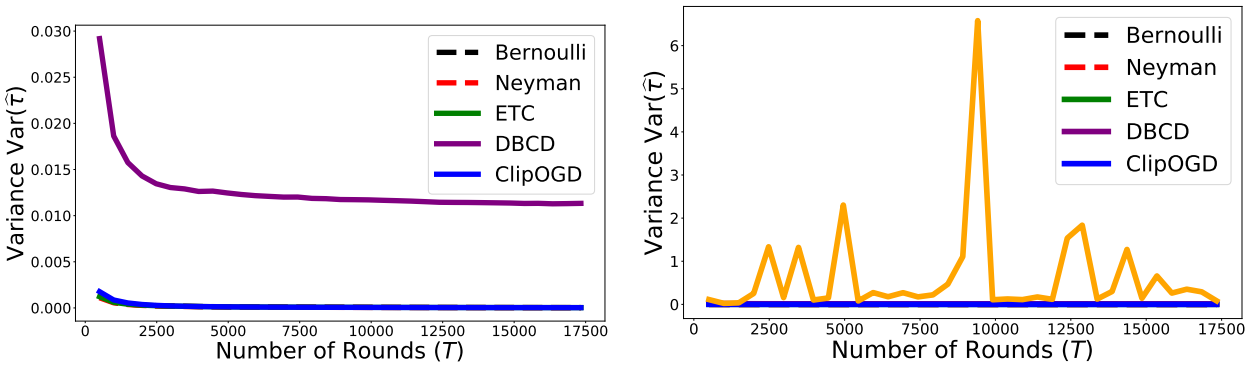


Figure 2.3: Comparing Step Sizes:

In Figure 2.3, we re-create Figure 2.1a from above but we have added instances of Clip-OGD with different step sizes of the forms $\eta = c/\sqrt{T}$ for $c \in \{0.25, 0.5, 1.0, 2.0, 4.0\}$. We find that smaller step sizes improve convergence rates, effectively removing the "overhead of adaptivity" in this example. However, because the randomized experiment can only be run once, experimenters will typically not be able to try many step sizes. While it remains an open question about how to select a step size which best mitigates the "overhead of adaptivity", our recommendation of $1/\sqrt{T}$ still maintains good convergence properties.

2.7.3 Alternative Designs

Finally, we conduct additional experiments to compare the results of Clip-OGD to alternative experimental designs which are not made for Neyman allocation. Indeed, we find that the alternative designs incur a high variance, relative to Clip-OGD and the Neyman variance.



(a) DBCD only **(b)** DBCD and EXP3

Figure 2.4: Comparison of (Unnormalized) Variances:

In Figure 2.4, we plot the variance of the adaptive designs on an unnormalized scale. We include "Doubly Biased Coin Design" proposed by Eisele [29] as `DBCD` in Fig 2.4a and both `DBCD` and `EXP3` in Fig 2.4b. We find that both `DBCD` and `EXP3` suffer from higher variance. This is because they are not designed for Adaptive Neyman Allocation as defined in this work: `DBCD` targets a different allocation rule and `EXP3` minimizes outcome regret so that essentially only one arm is pulled. Both of these algorithms let the sampling probabilities P_t get too close to the boundary of $[0, 1]$, resulting in excessively large variance.

2.8 Deferred Proofs

In this section, we provide general analysis relevant for the problem of Adaptive Neyman Allocation. In Section 2.8.1, we analyze the adaptive Horvitz–Thompson estimator. In Section 2.8.2, we derive the optimal non-adaptive Neyman design in terms of the potential outcomes. In Section 2.8.3, we show the equivalence of Neyman ratio and expected Neyman regret. For completeness, all propositions are restated below.

2.8.1 Analysis of Adaptive Horvitz–Thompson Estimator (Propositions 2.1 and 2.2)

Throughout these proofs, we break up the adaptive Horvitz–Thompson estimator into the sum of individual estimators. For each $t \in [T]$, define

$$\hat{\tau}_t = Y_t\Big(\frac{\mathbf{1}[Z_t = 1]}{P_t} - \frac{\mathbf{1}[Z_t = 0]}{1 - P_t}\Big)$$

so that the sequential Horvitz–Thompson estimator is equal to $\hat{\tau} = (1/T)\sum_{t=1}^{T} \hat{\tau}_t$. This mirrors how the average treatment effect is the average of individual treatment effects, i.e. $\tau = (1/T)\sum_{t=1}^{T} \tau_t$.

We begin by proving Proposition 2.1, which establishes the unbiasedness of the adaptive Horvitz–Thompson estimator, subject to a positivity condition.

Proposition 2.1. *If* $\min\{P_t, 1 - P_t\} > 0$ *almost surely for all* $t \in [T]$ *then the adaptive Horvitz–Thompson estimator is unbiased:* $\mathbb{E}[\hat{\tau}] = \tau$.

Proof. Observe that by linearity of expectation, we can break the expectation of the adaptive Horvitz–Thompson estimator as

$$\mathbb{E}[\hat{\tau}] = \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\hat{\tau}_t] \ .$$

Thus, it suffices to show that the individual effect estimators are unbiased: $\mathbb{E}[\hat{\tau}_t] = \tau_t$. Observe that if the positivity condition holds, then we have that the conditional expectation may be computed as

$$\begin{aligned}
\mathbb{E}[\hat{\tau}_t \mid \mathcal{F}_t] &= \mathbb{E}\Big[Y_t\Big(\frac{\mathbf{1}[Z_t = 1]}{P_t} - \frac{\mathbf{1}[Z_t = 0]}{1 - P_t}\Big) \mid \mathcal{F}_t\Big] \\
&= P_t \cdot \Big(\frac{y_t(1)}{P_t}\Big) + (1 - P_t) \cdot \Big(\frac{y_t(0)}{1 - P_t}\Big) \\
&= y_t(1) - y_t(0) \\
&= \tau_t \ .
\end{aligned}$$

The result follows by iterated expectation, $\mathbb{E}[\hat{\tau}_t] = \mathbb{E}[\mathbb{E}[\hat{\tau}_t \mid \mathcal{F}_t]] = \tau_t$. □

Next, we prove Proposition 2.2, which derives the variance of the adaptive Horvitz–Thompson estimator.

Proposition 2.2. *The variance of the adaptive Horvitz–Thompson estimator is*

$$T \cdot \mathrm{Var}(\hat{\tau}) = \frac{1}{T} \sum_{t=1}^{T} \Big(y_t(1)^2 \mathbb{E}\Big[\frac{1}{P_t}\Big] + y_t(0)^2 \mathbb{E}\Big[\frac{1}{1 - P_t}\Big]\Big) - \frac{1}{T} \sum_{t=1}^{T} \tau_t^2 \ .$$

Proof. We begin by decomposing the variance of the adaptive Horvitz–Thompson estimator as

$$\mathrm{Var}(\hat{\tau}) = \mathrm{Var}\Big(\frac{1}{T} \sum_{t=1}^{T} \hat{\tau}_t\Big) = \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \mathrm{Cov}(\hat{\tau}_t, \hat{\tau}_s) \ .$$

We now aim to compute each of these individual covariance terms. Before continuing, observe that, by construction, the individual effect estimators are conditionally unbiased $\mathbb{E}[\hat{\tau}_t \mid \mathcal{F}_t] = \tau_t$. It follows by iterated expectation that the individual effect estimators are unbiased

(unconditionally), i.e. $\mathbb{E}[\hat{\tau}_t] = \tau_t$. Suppose that $s > t$. In this case, the covariance between the individual estimators is equal to zero as,

$$\begin{aligned}\mathrm{Cov}(\hat{\tau}_t, \hat{\tau}_s) &= \mathbb{E}[\hat{\tau}_t\hat{\tau}_s] - \mathbb{E}[\hat{\tau}_t]\mathbb{E}[\hat{\tau}_s] \\ &= \mathbb{E}[\hat{\tau}_t\mathbb{E}[\hat{\tau}_s \mid \mathcal{F}_s]] - \tau_t\tau_s \\ &= \tau_s\mathbb{E}[\hat{\tau}_t] - \tau_t\tau_s \\ &= \tau_s\tau_t - \tau_t\tau_s \\ &= 0 \ .\end{aligned}$$

Now let us compute the variance of an individual effect estimator. Observe that the variance may be decomposed as

$$\mathrm{Var}(\hat{\tau}_t) = \mathbb{E}[\hat{\tau}_t^2] - \mathbb{E}[\hat{\tau}_t]^2 \ .$$

Because the individual estimator is unbiased, we have that $\mathbb{E}[\hat{\tau}_t]^2 = \tau_t^2$. Let us now analyze the first term.

$$\begin{aligned}\mathbb{E}[\hat{\tau}_t^2] &= \mathbb{E}\big[\mathbb{E}[\hat{\tau}_t^2 \mid \mathcal{F}_t]\big] && \text{(iterated expectation)} \\ &= \mathbb{E}\Big[\mathbb{E}\Big[Y_t^2\Big(\frac{\mathbf{1}[Z_t = 1]}{P_t^2} - \frac{\mathbf{1}[Z_t = 0]}{(1-P_t)^2}\Big) \mid \mathcal{F}_t\Big]\Big] \\ &= \mathbb{E}\Big[y_t(1)^2\frac{1}{P_t} + y_t(0)^2\frac{1}{1-P_t}\Big] \\ &= y_t(1)^2 \cdot \mathbb{E}\Big[\frac{1}{P_t}\Big] + y_t(0)^2 \cdot \mathbb{E}\Big[\frac{1}{1-P_t}\Big] \ .\end{aligned}$$

Thus, this establishes that the variance of an individual estimator is equal to

$$\mathrm{Var}(\hat{\tau}_t) = y_t(1)^2 \cdot \mathbb{E}\Big[\frac{1}{P_t}\Big] + y_t(0)^2 \cdot \mathbb{E}\Big[\frac{1}{1-P_t}\Big] - \tau_t^2 \ .$$

Combining terms, we have that the variance of the adaptive Horvitz–Thompson estimator is

$$\begin{aligned}T \cdot \mathrm{Var}(\hat{\tau}) &= \frac{1}{T}\sum_{t=1}^{T}\sum_{s=1}^{T}\mathrm{Cov}(\hat{\tau}_t, \hat{\tau}_s) \\ &= \frac{1}{T}\sum_{t=1}^{T}\mathrm{Var}(\hat{\tau}_t) \\ &= \frac{1}{T}\sum_{t=1}^{T}\left(y_t(1)^2 \cdot \mathbb{E}\Big[\frac{1}{P_t}\Big] + y_t(0)^2 \cdot \mathbb{E}\Big[\frac{1}{1-P_t}\Big]\right) - \frac{1}{T}\sum_{t=1}^{T}\tau_t^2 \qquad \square\end{aligned}$$

2.8.2 Derivation of the Neyman Design (Proposition 2.4)

In this section, we prove Proposition 2.4 which derives the (infeasible) non-adaptive Neyman design in terms of the Neyman probability p^* and corresponding Neyman variance V_{N}. We also show that, under Assumption 2.3, the Neyman variance achieves the parametric rate.

Proposition 2.4. *The Neyman variance is $T \cdot V_N = 2(1+\rho)S(1)S(0)$, which is achieved by the Neyman probability $p^* = (1 + S(0)/S(1))^{-1}$.*

Proof. Using Proposition 2.2, we have that the variance of the (non-adaptive) Bernoulli design with probability $p \in (0,1)$ is equal to

$$T \cdot V_p = S(1)^2\Big(\frac{1}{p} - 1\Big) + S(0)^2\Big(\frac{1}{1-p} - 1\Big) + 2\rho S(1)S(0) \ .$$

Thus, the optimal Neyman design is obtained by the p^* which minimizes the above. The first order condition stipulates that

$$\frac{\partial}{\partial p}\Big[T \cdot V_p\Big]\Big|_{p=p^*} = 0 \Leftrightarrow -S(1)^2\Big(\frac{1}{p^*}\Big)^2 + S(0)^2\Big(\frac{1}{1-p^*}\Big)^2 = 0 \ ,$$

which is solved by $p^* = \big(1 + S(0)/S(1)\big)^{-1}$. Substituting this p^* back into the variance yields the Neyman variance:

$$\begin{aligned} T \cdot V_{p^*} &= S(1)^2\Big(\frac{1}{p^*} - 1\Big) + S(0)^2\Big(\frac{1}{1-p^*} - 1\Big) + 2\rho S(1)S(0) \\ &= S(1)^2 \cdot \frac{S(0)}{S(1)} + S(0)^2 \cdot \frac{S(1)}{S(0)} + 2\rho S(1)S(0) \\ &= 2(1+\rho)S(1)S(0) \ . \end{aligned}$$

□

Next, we show that under Assumption 2.3, the Neyman variance achieves the parametric rate.

Proposition 2.14. *Under Assumption 2.3, the Neyman variance achieves the parametric rate: $V_N = \Theta(1/T)$.*

Proof. Proposition 2.4, derives the Neyman variance: $T \cdot V_{\mathrm{N}} = 2(1+\rho)S(1)S(0)$.

We begin by showing that the Neyman variance is asymptotically bounded from below. Moreover, Assumption 2.3 stipulates that there exists a constant $c > 0$ which lower bounds the second moments as $S(1) \geq c$ and $S(0) \geq c$ and the correlation as $(1+\rho) \geq c$. Thus, the normalized Neyman variance is bounded below as $T \cdot V_{\mathrm{N}} \geq 2c^3$.

Next, we show that the Neyman variance is asymptotically bounded from above at the same rate. Assumption 2.3 stipulates that there exists a constant $C > 0$ which upper bounds the second moments as $S(1) \leq C$ and $S(0) \leq C$. The correlation is bounded between $\rho \in [-1, 1]$ so that $(1+\rho) \leq 2$. These bounds together yield that the normalized Neyman variance is bounded above as $T \cdot V_{\mathrm{N}} \leq 4C^2$.

Together, these bounds establish that, under Assumption 2.3, we have that $V_{\mathrm{N}} = \Theta(1/T)$.

□

2.8.3 Equivalence of Neyman Ratio and Neyman Regret (Theorem 2.7)

In this section, we prove Theorem 2.7, which demonstrates the equivalence between the Neyman Ratio and the expected Neyman regret.

Theorem 2.7. *Under Assumption 2.3, the Neyman ratio is within a constant factor of the $1/T$-scaled expected Neyman regret: $\kappa_T = \Theta(\frac{1}{T}\mathbb{E}[\mathcal{R}_T])$.*

Proof. Recall that the Neyman ratio is defined as

$$\kappa_T = \frac{V - V_{\mathrm{N}}}{V_{\mathrm{N}}} = \frac{T \cdot V - T \cdot V_{\mathrm{N}}}{T \cdot V_{\mathrm{N}}} \ ,$$

where the second equality follows by multiplying the numerator and the denominator by T. Observe that by Proposition 2.2, the numerator is given by

$$\begin{aligned}
T \cdot V - T \cdot V_{\mathrm{N}} &= \frac{1}{T}\sum_{t=1}^{T}\left(y_t(1)^2 \cdot \mathbb{E}\Big[\frac{1}{P_t}\Big] + y_t(0)^2 \cdot \mathbb{E}\Big[\frac{1}{1-P_t}\Big]\right) \\
&\qquad - \min_{p^* \in [0,1]} \frac{1}{T}\sum_{t=1}^{T}\left(y_t(1)^2 \cdot \frac{1}{p^*} + y_t(0)^2 \cdot \frac{1}{1-p^*}\right) \\
&= \mathbb{E}\Big[\frac{1}{T}\sum_{t=1}^{T} f_t(P_t)\Big] - \min_{p^* \in [0,1]} \frac{1}{T}\sum_{t=1}^{T} f_t(p^*) \\
&= \frac{1}{T}\mathbb{E}\Big[\sum_{t=1}^{T} f_t(P_t) - \min_{p^* \in [0,1]} \sum_{t=1}^{T} f_t(p^*)\Big] \\
&= \frac{1}{T}\mathbb{E}\big[\mathcal{R}_T\big] \ .
\end{aligned}$$

Proposition 2.14 shows that under Assumption 2.3, $T \cdot V_{\mathrm{N}} = \Theta(1)$ so that the denominator is asymptotically constant. Thus, we have that $\kappa_T = \Theta(\frac{1}{T}\mathbb{E}[\mathcal{R}_T])$. □

2.8.4 Analysis of Neyman Regret for Clip-OGD

In this section, we will prove Theorem 2.8, which establish that Clip-OGD achieves $\mathcal{O}(\sqrt{T \log T})$ expected Neyman regret under our assumptions on the potential outcomes. While the main text used capital letters P_t and G_t to signify that the treatment probability and gradient estimator were random variables, we use lower case letters p_t and g_t below for the purposes of more aesthetically appealing proofs. Throughout the analysis, we define $\Delta_t = [\delta_t, 1 - \delta_t]$ and $a = 1 + C/c$ for notational convenience.

2.8.5 Proof of Theorem 2.8

The first lemma is a bound on the distance of treatment probability p_{t+1} to the optimal p^* in terms of the previous treatment probability, gradient estimate, and whether the projection interval contains the optimal p^*.

Lemma 2.15. *For each iteration $t \in [T]$,*

$$|p_{t+1} - p^*| \leq |(p_t - \eta g_t) - p^*| + \delta_t \mathbf{1}[p^* \notin \Delta_t] \ .$$

Proof. If $p^\star \in \Delta_t$, then the statement holds by Pythagorean theorem. Otherwise, note that the most that the projection operation onto Δ_t can move a point is exactly δ_t. Therefore, the distance between $\mathcal{P}_{\delta_t}(p_t - \eta g_t)$ and $p^\star$ is at most δ_t larger than that between $p_t - \eta g_t$ and $p^\star$. □

Next, we show that Assumption 2.3 implies that the optimal Neyman probability lies within an interval bounded away from zero.

Lemma 2.16. *Under Assumption 2.3, $p^* \in [\frac{1}{a}, 1 - \frac{1}{a}]$, where $a = 1 + C/c \geq 2$.*

Proof. As shown previously, the optimal Neyman probability is equal to

$$p^* = \left(1 + \sqrt{\frac{\sum_{t=1}^T y_t(0)^2}{\sum_{t=1}^T y_t(1)^2}}\right)^{-1}$$

Recall that Assumption 2.3 places the following moment conditions on the potential outcomes:

$$c \leq \Big(\frac{1}{T}\sum_{t=1}^T y_t(1)^2\Big)^{1/2} \leq C \quad \text{and} \quad c \leq \Big(\frac{1}{T}\sum_{t=1}^T y_t(1)^2\Big)^{1/2} \leq C \ .$$

This bounds p^* by

$$\Big(1 + \frac{C}{c}\Big)^{-1} \leq p^* \leq \Big(1 + \frac{c}{C}\Big)^{-1} \ .$$

The result follows by using the definition of $a = 1 + C/c$ to deduce that $(1/a) = (1 + C/c)^{-1}$ and $(1 - 1/a) = (1 + c/C)^{-1}$. □

The next lemma guarantees that after a fixed number of iterations, the projection interval will contain the Neyman optimal p^*.

Lemma 2.17. *We have that $p^* \in \Delta_t$ for all $t \geq (a/2)^\alpha$.*

Proof. Lemma 2.16 guarantees that $p^* \in [1/a, 1 - 1/a]$, where $a = 1 + C/c$. Thus, $p^* \in \Delta_t$ if $\delta_t \leq 1/a$. Using the definition of δ_t and rearranging terms, we have that

$$\delta_t \leq 1/a \Leftrightarrow (1/2)t^{-1/\alpha} \leq 1/a \Leftrightarrow t \geq (a/2)^\alpha \ .$$ □

The next lemma bounds the expected difference between the cost objective f_t evaluated at p_t and the cost objective f_t evaluated at the Neyman optimal probability p^*.

Lemma 2.18. *For each iteration $t \in [T]$, we have the bound*

$$2\mathbb{E}\Big[f_t(p_t) - f_t(p^*)\Big] \leq \frac{1}{\eta}\Big[\mathbb{E}\Big[(p_t - p^*)^2\Big] - \mathbb{E}\Big[(p_{t+1} - p^*)^2\Big]\Big] + \eta\mathbb{E}\big[g_t^2\big] + 4\mathbf{1}\Big[t < \Big(\frac{a}{2}\Big)^{\alpha}\Big]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}\mathbb{E}\Big[|g_t|\Big]\Big) \ .$$

Proof. Fix an iteration $t \in [T]$. By Lemma 2.15, and using the triangle inequality, we have that

$$\begin{aligned}
(p_{t+1} - p^*)^2 &\leq \Big(|(p_t - \eta g_t) - p^*| + \delta_t \mathbf{1}[p^* \notin \Delta_t]\Big)^2 \\
&= \big((p_t - \eta g_t) - p^*\big)^2 + \delta_t^2 \mathbf{1}[p^* \notin \Delta_t]^2 + 2|(p_t - \eta g_t) - p^*|\delta_t \mathbf{1}[p^* \notin \Delta_t] \\
&= \big((p_t - p^*) - \eta g_t\big)^2 + \mathbf{1}[p^* \notin \Delta_t]\Big(\delta_t^2 + 2\delta_t \cdot |(p_t - p^*) - \eta g_t|\Big) \\
&\leq \big((p_t - p^*) - \eta g_t\big)^2 + 2\delta_t \mathbf{1}[p^* \notin \Delta_t]\Big(\delta_t + |p_t - p^*| + \eta|g_t|\Big)
\end{aligned}$$

Because $p_t \in [\delta_t, 1 - \delta_t]$, we have that $|p_t - p^*| \leq 1 - \delta_t$ so that

$$\begin{aligned}
&\leq \big((p_t - p^*) - \eta g_t\big)^2 + 2\delta_t \mathbf{1}[p^* \notin \Delta_t]\Big(1 + \eta|g_t|\Big) \\
&=\leq (p_t - p^*)^2 + \eta^2 g_t^2 - 2\eta g_t(p_t - p^*) + 4\eta \cdot \mathbf{1}[p^* \notin \Delta_t]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}|g_t|\Big)
\end{aligned}$$

.

Rearranging terms yields that

$$2\eta g_t(p_t - p^*) \leq \Big[(p_t - p^*)^2 - (p_{t+1} - p^*)^2\Big] + \eta^2 g_t^2 + 4\eta \cdot \mathbf{1}[p^* \notin \Delta_t]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}|g_t|\Big) \ .$$

Dividing both sides by the step size 2η yields

$$g_t(p_t - p^*) \leq \frac{1}{2\eta}\Big[(p_t - p^*)^2 - (p_{t+1} - p^*)^2\Big] + \frac{\eta}{2}g_t^2 + 2\mathbf{1}[p^* \notin \Delta_t]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}|g_t|\Big) \ .$$

Using convexity of f_t, adding and subtracting terms, and using the above, we have that

$$\begin{aligned}
f_t(p_t) - f_t(p^*) &\leq \langle \nabla f_t(p_t), p_t - p^* \rangle && \text{(convexity)} \\
&= \langle g_t, p_t - p^* \rangle + \langle \nabla f_t(p_t) - g_t, p_t - p^* \rangle && \text{(adding, subtracting)} \\
&\leq \frac{1}{2\eta}\Big[(p_t - p^*)^2 - (p_{t+1} - p^*)^2\Big] + \frac{\eta}{2}g_t^2 && \text{(above)} \\
&\quad + 2\mathbf{1}[p^* \notin \Delta_t]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}|g_t|\Big) + \langle \nabla f_t(p_t) - g_t, p_t - p^* \rangle
\end{aligned}$$

By construction, we have that the gradient estimator is unbiased conditioned on D_t, i.e. $\mathbb{E}[g_t \mid D_t] = \nabla f_t(p_t)$. Thus, by iterated expectation we have that the gradient estimator is unbiased, i.e.

$$\mathbb{E}[\nabla f_t(p_t) - g_t] = \mathbb{E}[\mathbb{E}[\nabla f_t(p_t) - g_t \mid D_t]] = 0 \ .$$

Thus, taking expectations of both sides and applying Lemma 2.17 yields

$$\begin{aligned}
&2\mathbb{E}\Big[f_t(p_t) - f_t(p^*)\Big] \\
&\leq \frac{1}{\eta}\Big[\mathbb{E}\Big[(p_t - p^*)^2\Big] - \mathbb{E}\Big[(p_{t+1} - p^*)^2\Big]\Big] + \eta\mathbb{E}\big[g_t^2\big] + 4 \cdot \mathbf{1}[p^* \notin \Delta_t]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}\mathbb{E}\Big[|g_t|\Big]\Big) \\
&\leq \frac{1}{\eta}\Big[\mathbb{E}\Big[(p_t - p^*)^2\Big] - \mathbb{E}\Big[(p_{t+1} - p^*)^2\Big]\Big] + \eta\mathbb{E}\big[g_t^2\big] + 4 \cdot \mathbf{1}\Big[t < \Big(\frac{a}{2}\Big)^{\alpha}\Big]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}\mathbb{E}\Big[|g_t|\Big]\Big) \ \square
\end{aligned}$$

The following lemma derives bounds on the first and second (raw) moments of the gradient estimator at each iteration.

Lemma 2.19. *For each $t \in [T]$, the gradient estimates have bounded first and second moments:*

$$\begin{aligned}
\mathbb{E}\Big[g_t^2\Big] &\leq 2^5 t^{5/\alpha} \cdot \big(y_t(1)^4 + y_t(0)^4\big) \\
\mathbb{E}\Big[|g_t|\Big] &\leq 2^2 t^{2/\alpha} \cdot \big(y_t(1)^2 + y_t(0)^2\big)
\end{aligned}$$

Proof. We begin by handling the $\mathbb{E}\Big[g_t^2\Big]$ term. By definition of the gradient estimator, we have that the conditional expectation is at most

$$\begin{aligned}
\mathbb{E}[g_t^2 \mid D_t] &= p_t \cdot \Big(\frac{y_t(1)^2}{p_t^3}\Big)^2 + (1 - p_t) \cdot \Big(\frac{y_t(0)^2}{(1 - p_t)^3}\Big)^2 \\
&= \frac{y_t(1)^4}{p_t^5} + \frac{y_t(0)^4}{(1 - p_t)^5}
\end{aligned}$$

By definition of the algorithm, we have that $p_t \in [\delta_t, 1 - \delta_t]$ at iteration t. Thus, we may invoke the bound:

$$\begin{aligned}
&\leq \delta_t^{-5}\Big(y_t(1)^4 + y_t(1)^4\Big) \\
&= [(1/2)t^{-1/\alpha}]^{-5} \cdot \Big(y_t(1)^4 + y_t(1)^4\Big) \\
&= 2^5 t^{5/\alpha} \cdot \Big(y_t(1)^4 + y_t(1)^4\Big) \ ,
\end{aligned}$$

and the desired bound on $\mathbb{E}[g_t^2]$ follows from applying the law of iterated expectation.

The bound on the $\mathbb{E}\Big[|g_t|\Big]$ term follows in a similar way. By definition of the gradient estimator, we have that the conditional expectation is at most

$$\begin{aligned}
\mathbb{E}[|g_t| \mid D_t] &= p_t \cdot \Big|\frac{y_t(1)^2}{p_t^3}\Big| + (1-p_t)\cdot\Big|\frac{y_t(0)^2}{(1-p_t)^3}\Big| \\
&= \frac{y_t(1)^2}{p_t^2} + \frac{y_t(0)^2}{(1-p_t)^2} \\
&\leq \delta_t^{-2}\Big(y_t(1)^2 + y_t(1)^2\Big) \\
&= [(1/2)t^{-1/\alpha}]^{-2}\cdot\Big(y_t(1)^2 + y_t(1)^2\Big) \\
&= 2^2 t^{2/\alpha}\cdot\Big(y_t(1)^2 + y_t(1)^2\Big)\ ,
\end{aligned}$$

and the desired bound on $\mathbb{E}[|g_t|]$ follows from applying the law of iterated expectation. □

The following proposition bounds the expected Neyman regret for general settings of the projection parameter α.

Proposition 2.20. *Suppose Assumption 2.3 holds. Then, for any choice of projection parameter $\alpha \geq 2$ (possibly depending on T) and for the step size $\eta = \sqrt{\frac{e^\alpha}{T^{1+5/\alpha}}}$, a finite-sample bound on the expected Neyman regret incurred by* Clip-OGD *is*

$$\mathbb{E}\big[\mathcal{R}_T\big] \leq (2^2 e^{a/2} + 2^5 C^4)\sqrt{e^\alpha T^{1+5/\alpha}} + 2^2 C^2 e^{2+a/2}\sqrt{e^\alpha T}\ .$$

Proof. By Lemma 2.18, we have that the regret is at most

$$\begin{aligned}
2\mathbb{E}\Big[\mathcal{R}_T\Big] &= \sum_{t=1}^T 2\mathbb{E}\Big[f_t(p_t) - f_t(p^*)\Big] && (2.4)\\
&\leq \frac{1}{\eta}\sum_{t=1}^T\Big[\mathbb{E}\Big[(p_t-p^*)^2\Big] - \mathbb{E}\Big[(p_{t+1}-p^*)^2\Big]\Big] + \eta\sum_{t=1}^T\mathbb{E}\big[g_t^2\big] && (2.5)\\
&\quad + 4\sum_{t=1}^T \mathbf{1}\Big[t < \Big(\frac{a}{2}\Big)^\alpha\Big]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}\mathbb{E}\Big[|g_t|\Big]\Big) && (2.6)
\end{aligned}$$

Using a telescoping argument, we have that the first term is bounded by

$$\frac{1}{\eta}\sum_{t=1}^T\Big[\mathbb{E}\Big[(p_t-p^*)^2\Big] - \mathbb{E}\Big[(p_{t+1}-p^*)^2\Big]\Big] \leq \frac{1}{\eta}\mathbb{E}\Big[(p_1-p^*)^2\Big] \leq \frac{1}{\eta}\ .$$

Using Lemma 2.19 and Assumption 2.3, the sum in the second term may be bounded as

$$\begin{aligned}
\sum_{t=1}^{T} \mathbb{E}[g_t^2] &\leq \sum_{t=1}^{T} 2^5 t^{5/\alpha}(y_t(1)^4 + y_t(0)^4) && \text{(Lemma 2.19)} \\
&\leq 2^5 T^{5/\alpha}\Big(\sum_{t=1}^{T} y_t(1)^4 + \sum_{t=1}^{T} y_t(0)^4\Big) \\
&\leq 2^5 T^{5/\alpha}\Big(2C^4 T\Big) && \text{(Assumption 2.3)} \\
&= 2^6 C^4 T^{1+5/\alpha} \ .
\end{aligned}$$

Next, we deal with the third term by breaking it up into two more terms. Define $t^* = \lceil (a/2)^\alpha \rceil$. The third term can be broken into two terms:

$$4\sum_{t=1}^{T} \mathbf{1}\Big[t < \Big(\frac{a}{2}\Big)^\alpha\Big]\Big(\frac{\delta_t}{\eta} + \frac{\delta_t}{2}\mathbb{E}\Big[|g_t|\Big]\Big) = \frac{4}{\eta}\sum_{t=1}^{t^*-1}\delta_t + 2\sum_{t=1}^{t^*-1}\delta_t \mathbb{E}[|g_t|] \ .$$

The first of these two terms can be bounded in the following way. Using that $\alpha \geq 2$ we have that

$$\begin{aligned}
\sum_{t=1}^{t^*-1}\delta_t &= \sum_{t=1}^{t^*-1}(1/2)t^{-1/\alpha} \\
&= (1/2)\Big[1 + \sum_{t=2}^{t^*-1} t^{-1/\alpha}\Big] \\
&\leq (1/2)\Big[1 + \int_{x=1}^{t^*-1} x^{-1/\alpha} dx\Big] \\
&= (1/2)\Big[1 + \frac{\alpha}{\alpha-1}\cdot((t^*-1)^{(1-1/\alpha)} - 1)\Big] \\
&\leq (t^*-1)^{(1-1/\alpha)} \\
&\leq (a/2)^{\alpha(1-1/\alpha)} \\
&= (a/2)^{\alpha-1} \ .
\end{aligned}$$

The second term can be bounded using Lemma 2.19 as follows:

$$\begin{aligned}
\sum_{t=1}^{t^*-1}\delta_t \mathbb{E}[|g_t|] &\leq \sum_{t=1}^{t^*-1}(1/2)t^{-1/\alpha}2^2 t^{2/\alpha}\big(y_t(1)^2 + y_t(0)^2\big) && \text{(Lemma 2.19)} \\
&= 2\sum_{t=1}^{t^*-1} t^{1/\alpha}\big(y_t(1)^2 + y_t(0)^2\big) \\
&\leq 2\big(\sum_{t=1}^{t^*-1} t^{2/\alpha}\big)^{1/2}\Big[\Big(\sum_{t=1}^{t^*-1} y_t(1)^4\Big)^{1/2} + \Big(\sum_{t=1}^{t^*-1} y_t(0)^4\Big)^{1/2}\Big] \ , && \text{(Cauchy-Schwarz)}
\end{aligned}$$

where the last inequality follows from Cauchy–Schwarz. By extending the sum involving the outcomes to all units, we obtain an upper bound on which we can apply the bounded moment assumption:

$$\begin{aligned}
&\leq 2\Big(\sum_{t=1}^{t^*-1} t^{2/\alpha}\Big)^{1/2}\Big[\Big(\sum_{t=1}^{T} y_t(1)^4\Big)^{1/2} + \Big(\sum_{t=1}^{T} y_t(0)^4\Big)^{1/2}\Big] \\
&\leq 2\Big(\sum_{t=1}^{t^*-1} t^{2/\alpha}\Big)^{1/2}\Big[2(C^4T)^{1/2}\Big] \qquad \text{(Assumption 2.3)}
\end{aligned}$$

What remains in this step is to bound the first term above. By replacing each of the t in the sum with t^*, we obtain the following upper bound:

$$\begin{aligned}
&\leq 2(t^*-1)^{(1/2+1/\alpha)}\Big[2(C^4T)^{1/2}\Big] \\
&\leq 2((a/2)^{\alpha})^{(1/2+1/\alpha)}\Big[2(C^4T)^{1/2}\Big] \\
&= 2^2C^2\Big(\frac{a}{2}\Big)^{(1+\alpha/2)}T^{1/2} \ .
\end{aligned}$$

Using the above work, we have that the overall regret is bounded as

$$\begin{aligned}
2\mathbb{E}\Big[\mathcal{R}_T\Big] &\leq \frac{1}{\eta} + \eta 2^6 C^4 T^{1+5/\alpha} + \frac{4}{\eta}\Big(\frac{a}{2}\Big)^{(\alpha-1)} + 2^3C^2\Big(\frac{a}{2}\Big)^{(1+\alpha/2)}T^{1/2} \\
&= \frac{1}{\eta}\Big(1 + 4\Big(\frac{a}{2}\Big)^{(\alpha-1)}\Big) + \eta 2^6 C^4 T^{1+5/\alpha} + 2^3C^2\Big(\frac{a}{2}\Big)^{(1+\alpha/2)}T^{1/2} \ .
\end{aligned}$$

Next, we separate the constant a from the projection parameter α. To this end, we use the inequality $y^r \leq e^{1+y}\cdot e^r$ for all $y \in \mathbb{R}$ and $r \geq 0$, to obtain

$$\Big(\frac{a}{2}\Big)^{(\alpha-1)} \leq e^{1+a/2}e^{\alpha-1} = e^{a/2}e^{\alpha} \quad \text{and} \quad \Big(\frac{a}{2}\Big)^{(1+\alpha/2)} \leq e^{1+a/2}e^{1+\alpha/2} = e^{2+a/2}e^{\alpha/2} \ ,$$

where we have used that $a \geq 2$. Substituting these back into the above, we have that the regret bound is Substituting this quantities into the above, we have that the regret bound is

$$\begin{aligned}
2\mathbb{E}[\mathcal{R}_T] &\leq \frac{1}{\eta}\Big(1 + 4e^{a/2}e^{\alpha}\Big) + \eta 2^6 C^4 T^{1+5/\alpha} + 2^3C^2e^{2+a/2}e^{\alpha/2}T^{1/2} \\
&\leq \frac{1}{\eta}2^3e^{a/2}e^{\alpha} + \eta 2^6 C^4 T^{1+5/\alpha} + 2^3C^2e^{2+a/2}e^{\alpha/2}T^{1/2}
\end{aligned}$$

By setting $\eta = \sqrt{\frac{e^{\alpha}}{T^{1+5/\alpha}}}$ to minimize the bound above, we have that the expected Neyman regret is bounded as

$$= (2^3e^{a/2} + 2^6C^4)\sqrt{e^{\alpha}T^{1+5/\alpha}} + 2^3C^2e^{2+a/2}\sqrt{e^{\alpha}T} \ .$$

where we have used that and the result follows by dividing both sides by 2. □

Proposition 2.20 demonstrates that many different values of α will guarantee sublinear expected Neyman regret. For example, setting α to be a constant satisfying $\alpha > 5$ will ensure sublinear expected Neyman regret. However, by tuning α according to the analysis above, we can achieve $\mathcal{O}(\sqrt{T\log(T)})$ expected Neyman regret, as demonstrated by Theorem 2.8.

Theorem 2.8*. *Under Assumption 2.3 the parameter values* $\eta = \sqrt{1/T}$ *and* $\alpha = \sqrt{5\log(T)}$ *ensure the expected Neyman regret of* Clip-OGD *is bounded as*

$$\mathbb{E}\big[\mathcal{R}_T\big] \leq \Big(2^2 e^{a/2} + 2^5 C^4 + 2^2 C^2 e^{2+a/2}\Big) \cdot \sqrt{T} \cdot \exp(\sqrt{5\log(T)}) \ ,$$

which implies that $\mathbb{E}\big[\mathcal{R}_T\big] \leq \widetilde{\mathcal{O}}\big(\sqrt{T}\big)$.

Proof. Observe that for $\alpha = \sqrt{5\log(T)}$, we have that the step size posited in Proposition 2.20 (i.e. $\eta = \sqrt{\frac{e^\alpha}{T^{1+5/\alpha}}}$) is equal to $\eta = \sqrt{1/T}$. Thus, by rearranging terms and using the result of Proposition 2.20, we have that expected Neyman regret is bounded as

$$\begin{aligned}\mathbb{E}\big[\mathcal{R}_T\big] &\leq (2^2 e^{a/2} + 2^5 C^4)\sqrt{e^\alpha T^{1+5/\alpha}} + 2^2 C^2 e^{2+a/2}\sqrt{e^\alpha T} \\ &= (2^2 e^{a/2} + 2^5 C^4)\sqrt{T}\cdot\sqrt{e^\alpha T^{5/\alpha}} + 2^2 C^2 e^{2+a/2}\sqrt{T}\cdot\sqrt{e^\alpha} \ .\end{aligned}$$

The difficulty is now to find which setting of α will make this bound smallest. Observe that the real tension is in the first term and we can re-write the relevant part of this term as

$$\sqrt{e^\alpha T^{5/\alpha}} = \Big[\exp(\alpha + \log(T^{5/\alpha}))\Big]^{1/2} = \Big[\exp(\alpha + (5/\alpha)\log(T))\Big]^{1/2} \ .$$

To minimize this term, we select $\alpha = \sqrt{5\log(T)}$ which results in

$$\sqrt{e^\alpha T^{5/\alpha}} = \exp(\sqrt{5\log(T)}) \ .$$

Likewise, this choice of α results in $\sqrt{e^\alpha} \leq e^\alpha = \exp(\sqrt{5\log(T)})$. Putting this together yields the desired finite sample regret bound:

$$\mathbb{E}\big[\mathcal{R}_T\big] \leq \Big(2^2 e^{a/2} + 2^5 C^4 + 2^2 C^2 e^{2+a/2}\Big) \cdot \sqrt{T} \cdot \exp(\sqrt{5\log(T)}) \ .$$

The result follows by observing that the terms inside the parenthesis are constant by Assumption 2.3 and the function $\exp(\sqrt{5\log(T)})$ is subpolynomial. □

2.8.6 Selecting Parameters When Moment Bounds are Known

We briefly remark on how to select the step size parameter η when the experimenter can correctly specify the constants $C \geq c$ used in the Assumption 2.3. The proof of Proposition 2.20 shows that for general parameters, the Neyman regret may be bounded as

$$\mathbb{E}[\mathcal{R}_T] \leq \frac{1}{\eta} 2^2 e^{a/2} e^\alpha + \eta 2^5 C^4 T^{1+5/\alpha} + 2^2 C^2 e^{2+a/2} e^{\alpha/2} T^{1/2} \ .$$

To optimize the step size with respect to these constants, one would choose $\alpha = \sqrt{5\log(T)}$ and

$$\eta = \frac{e^{\frac{1}{4}\cdot(1+C/c)}}{2\sqrt{2}C^2} \cdot \frac{1}{\sqrt{T}} \ ,$$

where we have used that $a = 1 + C/c$. When these moment bounds are correctly specified, this choice of step size will likely yield improved convergence rates, as our bound on the Neyman regret will have a factor of C^2 rather than C^4.

2.8.7 Analysis for Inference in Large Samples

In this section, we provide the necessary statistical tools for constructing asymptotically valid confidence intervals. This can be done in two main steps. First, in Section 2.8.8 we construct a conservative variance estimator which we show is consistent in probability. Then, in Section 2.8.9, we show that the resulting Chebyshev-type intervals are asymptotically valid.

While the main text used capital letters P_t and G_t to signify that the treatment probability and gradient estimator were random variables, we use lower case letters p_t and g_t below for the purposes of more aesthetically appealing proofs. Throughout the analysis, we use the parameter settings $\eta = \sqrt{1/T}$ and $\alpha = \sqrt{5\log(T)}$, which are recommended in the main text. However, we suspect that many of our results will go through for the class of parameters $\eta = \sqrt{\frac{e^{\alpha}}{T^{1+5/\alpha}}}$ and $\alpha > 5$ which appear in Proposition 2.20.

The following lemma shows that under Assumptions 2.3 and 2.9, the variance of the adaptive Horvitz–Thompson estimator under Clip-OGD achieves the parametric rate.

Lemma 2.21. *Assumptions 2.3 and 2.9, the variance of the adaptive Horvitz–Thompson estimator under* Clip-OGD *achieves the parametric rate:* $\mathrm{Var}(\hat{\tau}) = \Theta(1/T)$.

Proof. Theorem 2.7 shows that under Assumption 2.3, they Neyman ratio is order equivalent to the $1/T$-scaled expected Neyman regret, i.e. $\kappa_T = \Theta((1/T)\mathbb{E}[\mathcal{R}_T])$. Theorem 2.8 shows that under these assumptions, Clip-OGD achieves sublinear expected Neyman regret $\mathbb{E}[\mathcal{R}_T] = o(T)$ which implies that $\limsup \kappa_T \leq 0$. Likewise, Assumption 2.9 states that the negative expected Neyman regret is sublinear, $-\mathbb{E}[\mathcal{R}_T] = o(T)$, which implies that $\liminf \kappa_T \geq 0$. Thus, we have that the Neyman ratio converges to zero, e.g. $\lim \kappa_T = 0$.

By recalling the definition of the Neyman ratio, we have that

$$0 = \lim_{T\to\infty} \kappa_T = \lim_{T\to\infty} \frac{V - V_{\mathrm{N}}}{V_{\mathrm{N}}} = \lim_{T\to\infty} \frac{T\cdot V - T\cdot V_{\mathrm{N}}}{T\cdot V_{\mathrm{N}}} \ .$$

Proposition 2.14 demonstrates that $T \cdot V_{\mathrm{N}} = \Theta(1)$. Together with the above, this implies that $T \cdot V = \Theta(1)$. □

The following lemma shows that under the recommended parameter settings, the (random) treatment probabilities are bounded away from zero and one.

Lemma 2.22. *When $\alpha = \sqrt{5\log(T)}$, we have that for all iterations $t \in [T]$, the inverse of projection parameter is bounded:*

$$\frac{1}{\delta_t} \leq 2\exp(\sqrt{\log(T^{1/5})}) = \widetilde{\mathcal{O}}(1) \ .$$

Proof. A uniform upper bound on the inverse of the projection parameters is

$$\frac{1}{\delta_t} \leq \frac{1}{\delta_T} = \frac{1}{(1/2)T^{-1/\alpha}} = 2T^{1/\alpha} = 2\exp(\frac{1}{\alpha}\log(T)) = 2\exp(\sqrt{\log(T^{1/5})}) \ .$$

To complete the proof, observe that the function $h(T) = \exp(\sqrt{1/5}\cdot\sqrt{\log(T)})$ is subpolynomial, so that we can write it as $\widetilde{\mathcal{O}}(1)$. □

2.8.8 Conservative Variance Estimator (Theorem 2.10)

In this section, we prove Theorem 2.10 which establishes that the normalized variance estimator is converges in probability to the normalized variance upper bound at a $\widetilde{\mathcal{O}}_p(T^{-1/2})$ rate. Before continuing, let us review the relevant quantities. Recall that the Neyman variance and the corresponding upper bound are given by

$$T \cdot V_{\mathrm{N}} = 2(1+\rho)S(1)S(0) \quad \text{and} \quad T \cdot \mathrm{VB} = 4S(1)S(0) \ ,$$

where the second moments $S(1)$ and $S(0)$ are defined as

$$S(1)^2 = \frac{1}{T}\sum_{t=1}^{T} y_t(1)^2 \quad \text{and} \quad S(0)^2 = \frac{1}{T}\sum_{t=1}^{T} y_t(0)^2 \ .$$

Our variance estimator is defined as

$$T \cdot \widehat{\mathrm{VB}} = \sqrt{\widehat{A(1)} \cdot \widehat{A(0)}} \ ,$$

where

$$\widehat{A(1)} = \frac{1}{T}\sum_{t=1}^{T} y_t(1)^2 \frac{\mathbf{1}[Z_t = 1]}{p_t} \quad \text{and} \quad \widehat{A(0)} = \frac{1}{T}\sum_{t=1}^{T} y_t(0)^2 \frac{\mathbf{1}[Z_t = 0]}{1-p_t} \ .$$

The random variables $\widehat{A(1)}$ and $\widehat{A(0)}$ are unbiased estimates of $S(1)^2$ and $S(0)^2$ which are based on the Horvitz–Thompson principle. However, the fact that the estimators $\widehat{A(1)}$ and $\widehat{A(0)}$ are not independent and the square root is introduced means that the variance estimator $\widehat{\mathrm{VB}}$ is not an unbiased estimator for the variance bound VB. Even so, we will show that the variance estimator is consistent for the variance bound. For aesthetic considerations, we define $A(1) = S(1)^2$ and $A(0) = S(0)^2$ so that $\widehat{A(1)}$ is an estimator for $A(1)$ and $\widehat{A(0)}$ is an estimator for $A(0)$.

Our general approach will follow in two steps. First, by bounding its bias and variance, we will show that $\widehat{A(1)} \cdot \widehat{A(0)} - A(1)A(0)$ converges as $\widetilde{\mathcal{O}}_p(T^{-1/2})$. Next, by appealing to a quantitative Continuous Mapping Theorem, we will argue that the error $\sqrt{\widehat{A(1)} \cdot \widehat{A(0)}} - \sqrt{A(1)A(0)}$ converges at the same rate. By definition, this is exactly the error of the normalized variance estimator to the normalized variance bound, i.e. $T \cdot \widehat{\mathrm{VB}} - T \cdot \mathrm{VB}$.

Before continuing, let us define new auxiliary random variables. For each $t \in [T]$, we define the variables r_t and q_t as

$$r_t = \frac{\mathbf{1}[z_t = 1]}{p_t} \quad \text{and} \quad q_t = \frac{\mathbf{1}[z_t = 0]}{1 - p_t} \ .$$

Below are basic facts about these auxiliary random variables.

Lemma 2.23. *The auxiliary random variables satisfy the following properties:*

1. $\mathbb{E}[r_t q_s] = \mathbf{1}[t \neq s]$.
2. $\mathbb{E}[r_t^2 \mid D_t] \leq \frac{1}{\delta_t}$ *and* $\mathbb{E}[q_t^2 \mid D_t] \leq \frac{1}{\delta_t}$.
3. *The covariance* $\mathrm{Cov}(r_t q_s, r_\ell q_k)$ *behave in the following ways:*

$$\begin{aligned}
&\mathrm{Cov}(r_t q_s, r_\ell q_k) = 0 && \text{if } t = s \text{ or } \ell = k \\
&\mathrm{Cov}(r_t q_s, r_\ell q_k) = 0 && \text{if } t \neq s \text{ and } \ell \neq k \text{ and } \{t, s\} \cap \{\ell, k\} = \emptyset \\
&\mathrm{Cov}(r_t q_s, r_\ell q_k) = -1 && \text{if } t \neq s \text{ and } \ell \neq k \text{ and } (\ t = k \text{ or } s = \ell\ ) \\
&\mathrm{Cov}(r_t q_s, r_\ell q_k) \leq \frac{1}{\delta_t} - 1 && \text{if } t \neq s \text{ and } \ell \neq k \text{ and } t = \ell \text{ and } s \neq k \\
&\mathrm{Cov}(r_t q_s, r_\ell q_k) \leq \frac{1}{\delta_s} - 1 && \text{if } t \neq s \text{ and } \ell \neq k \text{ and } t \neq \ell \text{ and } s = k \\
&\mathrm{Cov}(r_t q_s, r_\ell q_k) \leq \frac{1}{\delta_t \delta_s} - 1 && \text{if } t \neq s \text{ and } \ell \neq k \text{ and } t = \ell \text{ and } s = k
\end{aligned}$$

Proof. First, we show that $\mathbb{E}[r_t q_s] = \mathbf{1}[t \neq s]$. Let $t, s \in [T]$ and suppose that $t \neq s$. Without loss of generality, suppose that $t > s$. Then by using iterated expectation, we have that

$$\mathbb{E}[r_t q_s] = \mathbb{E}[q_s \mathbb{E}[r_t \mid D_t]] = \mathbb{E}[q_s] = 1 \ .$$

Otherwise, if $t = s$, then $r_t q_t = (\mathbf{1}[z_t = 1]/p_t) \cdot (\mathbf{1}[z_s = 0]/(1 - p_t)) = 0$ so that $\mathbb{E}[r_t q_t] = 0$.

Next, we show that $\mathbb{E}[r_t^2 \mid D_t] \leq \frac{1}{\delta_t}$ and $\mathbb{E}[q_t^2 \mid D_t] \leq \frac{1}{\delta_t}$. Observe that $\mathbb{E}[r_t^2 \mid D_t] = p_t(1/p_t^2) = 1/p_t \leq 1/\delta_t$, where the inequality follows by definition of Algorithm 1. A similar argument shows that $\mathbb{E}[q_t^2 \mid D_t] \leq 1/\delta_t$.

Finally, we establish the covariance terms one by one. We do this in order of the cases that they were presented in.

Case 1 $(t = s)$ **or** $(\ell = k)$: If $t = s$, then $r_t q_t$ is almost surely zero, as argued above. Likewise, if $\ell = k$ then $r_\ell q_\ell$ is almost surely zero. In either of these cases, we have that $\mathrm{Cov}(r_t q_s, r_\ell q_k) = 0$.

Case 2 $(t \neq s)$ **and** $(\ell \neq k)$ **and** $\{t, s\} \cap \{\ell, k\} = \emptyset$: Note that in this case, all the indices t, s, ℓ, and k are distinct. Without loss of generality, suppose that $t < s < \ell < k$. A repeated use of the iterated expectation yields that

$$\mathbb{E}[r_t q_s r_\ell q_s] = \mathbb{E}[r_t q_s r_\ell \mathbb{E}[q_s \mid D_s]] = \mathbb{E}[r_t q_s r_\ell] = \mathbb{E}[r_t q_s \mathbb{E}[r_\ell \mid D_r]] = \mathbb{E}[r_t q_s] = \ldots = 1 \ .$$

Because all terms of distinct, we have that $\mathbb{E}[r_t q_s] = 1$ and $\mathbb{E}[r_\ell q_k] = 1$. Thus, the covariance is equal to

$$\mathrm{Cov}(r_t q_s, r_\ell q_k) = \mathbb{E}[r_t q_s, r_\ell q_k] - \mathbb{E}[r_t q_s] \cdot \mathbb{E}[r_\ell q_k] = 1 - 1 = 0 \ .$$

Case 3 $(t \neq s)$ **and** $(\ell \neq k)$ **and** $(t = k$ **or** $s = \ell)$: Suppose that $t = k$. In this case, observe that $r_t q_k$ is zero almost surely. Thus, $\mathbb{E}[r_t q_s r_\ell q_k] = 0$. On the other hand, $t \neq s$ and $\ell \neq k$ so that $\mathbb{E}[r_t q_s] = \mathbb{E}[r_\ell q_k] = 1$. This means that

$$\mathrm{Cov}(r_t q_s, r_\ell q_k) = \mathbb{E}[r_t q_s r_\ell q_k] - \mathbb{E}[r_t q_s] \cdot \mathbb{E}[r_\ell q_k] = 0 - 1 \cdot 1 = -1 \ .$$

The same argument shows that $s = \ell$ yields the same result.

Case 4 $(t \neq s)$ **and** $(\ell \neq k)$ **and** $t = \ell$ **and** $s \neq k)$: We begin by computing the expectation of the product of these four terms. In this case, we have that $t = \ell$ so that

$$\mathbb{E}[r_t q_s r_\ell q_k] = \mathbb{E}[r_t^2 q_s q_k] \ .$$

By assumption, the indices t, s, and k are all distinct. Our approach will be to obtain an upper bound on the expectation of the project of these three terms by iterated expectation. In particular, the inequality we will use is that $\mathbb{E}[r_t^2 \mid D_t] \leq 1/\delta_t$. Suppose for now that $s < k < t$. In this case, we use iterated expectation to get

$$\mathbb{E}[r_t q_s r_\ell q_k] = \mathbb{E}[q_s q_k \mathbb{E}[r_t \mid D_t]] \leq \frac{1}{\delta_t}\mathbb{E}[q_s q_k] = \frac{1}{\delta_t}\mathbb{E}[q_s \mathbb{E}[q_k \mid D_k]] = \frac{1}{\delta_t}\mathbb{E}[q_s] = \frac{1}{\delta_t} \ .$$

In the above, we have assumes that $s < k < t$, but the same iterated expectation technique can be applied regardless of the ordering of these indices, because they are unique. Thus, in this case, $\mathbb{E}[r_t q_s r_\ell q_k] \leq 1/\delta_t$. Because $t \neq s$ and $\ell \neq k$, we have that $\mathbb{E}[r_t q_s] = \mathbb{E}[r_\ell q_k] = 1$. This means that

$$\mathrm{Cov}(r_t q_s, r_\ell q_k) = \mathbb{E}[r_t q_s r_\ell q_k] - \mathbb{E}[r_t q_s] \cdot \mathbb{E}[r_\ell q_k] \leq \frac{1}{\delta_t} - 1 \ .$$

Case 5 $(t \neq s)$ **and** $(\ell \neq k)$ **and** $t \neq \ell$ **and** $s = k)$: We begin by computing the expectation of the product of these four terms. In this case, we have that $s = k$ so that

$$\mathbb{E}[r_t q_s r_\ell q_k] = \mathbb{E}[r_t r_\ell q_s^2] \ .$$

By assumption, the indices t, ℓ, and s are all distinct. Using a similar argument as the previous case, we can use iterated expectation together with the bound $\mathbb{E}[q_s^2 \mid D_s] \leq 1/\delta_s^2$ to obtain that $\mathbb{E}[r_t q_s r_\ell q_k] \leq 1/\delta_s^2$. Because $t \neq s$ and $\ell \neq k$, we have that $\mathbb{E}[r_t q_s] = \mathbb{E}[r_\ell q_k] = 1$. This means that

$$\mathrm{Cov}(r_t q_s, r_\ell q_k) = \mathbb{E}[r_t q_s r_\ell q_k] - \mathbb{E}[r_t q_s] \cdot \mathbb{E}[r_\ell q_k] \leq \frac{1}{\delta_s} - 1 \ .$$

Case 6 $(t \neq s)$ **and** $(\ell \neq k)$ **and** $t = \ell$ **and** $s = k$**)**: Suppose without loss of generality that $t > s$. In this case, we can bound the product of the expectation of these four terms using iterated expectation and the proven inequalities:

$$\mathbb{E}[r_t q_s r_\ell q_k] = \mathbb{E}[r_t^2 q_s^2] = \mathbb{E}[q_s^2 \mathbb{E}[r_t^2 \mid D_t]] \leq \frac{1}{\delta_t} \mathbb{E}[q_s^2] \leq \frac{1}{\delta_t \delta_s} \ .$$

Because $t \neq s$, we have that $\mathbb{E}[r_t q_s] = \mathbb{E}[r_\ell, q_k] = 1$. Thus, the covariance is bounded by

$$\mathrm{Cov}(r_t q_s, r_\ell q_k) = \mathbb{E}[r_t q_s r_\ell q_k] - \mathbb{E}[r_t q_s] \cdot \mathbb{E}[r_\ell q_k] \leq \frac{1}{\delta_t \delta_s} - 1 \ . \qquad \square$$

First, we show that that the difference between the expected value of $\widehat{A(1)}\widehat{A(0)}$ and the target $A(1)A(0)$ is decreasing at a linear rate in T.

Proposition 2.24. *The absolute bias of the estimated crossing term $\widehat{A(1)}\widehat{A(0)}$ to its target value $A(1)A(0)$ is at most*

$$\left|\mathbb{E}\left[\widehat{A(1)}\widehat{A(0)}\right] - S(1)^2 S(0)^2\right| \leq \frac{C^4}{T} \ .$$

Proof. Using Lemma 2.23, we can calculate the expectation of the product $\widehat{A(1)}\widehat{A(0)}$ as

$$\begin{aligned}
\mathbb{E}\left[\widehat{A(1)}\widehat{A(0)}\right] &= \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} y_t(1)^2 r_t\right)\left(\frac{1}{T}\sum_{s=1}^{T} y_s(0)^2 q_s\right)\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} y_t(1)^2 y_t(0)^2 \mathbb{E}\left[r_t q_s\right] \\
&= \frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} y_t(1)^2 y_t(0)^2 - \frac{1}{T^2}\sum_{t=1}^{T} y_t(1)^2 y_t(0)^2 \\
&= \left(\frac{1}{T}\sum_{t=1}^{T} y_t(1)^2\right)\left(\frac{1}{T}\sum_{s=1}^{T} y_s(0)^2\right) - \frac{1}{T^2}\sum_{t=1}^{T} y_t(1)^2 y_t(0)^2 \\
&= A(1)A(0) - \frac{1}{T^2}\sum_{t=1}^{T} y_t(1)^2 y_t(0)^2 \ .
\end{aligned}$$

We complete the proof by using Cauchy-Schwarz and Assumption 2.3, to bound the absolute bias as

$$\begin{aligned}\left|\mathbb{E}\big[\widehat{A(1)}\widehat{A(0)}\big] - A(1)A(0)\right| &= \frac{1}{T^2}\sum_{t=1}^{T} y_t(1)^2 y_t(0)^2 \\ &\leq \frac{1}{T^2}\left[\Big(\sum_{t=1}^{T} y_t(1)^4\Big)\cdot\Big(\sum_{t=1}^{T} y_t(0)^4\Big)\right]^{1/2} \\ &= \frac{1}{T}\left[\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(1)^4\Big)^{1/4}\cdot\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(0)^4\Big)^{1/4}\right]^2 \\ &\leq \frac{C^4}{T}\ . \end{aligned}$$

□

Next, we show that the variance of $\widehat{A(1)}\widehat{A(0)}$ is going to zero at a sufficiently fast rate.

Proposition 2.25. *The variance of* $\widehat{A(1)}\widehat{A(0)}$ *is bounded as*

$$\mathrm{Var}(\widehat{A(1)}\widehat{A(0)}) \leq \frac{4e^{\sqrt{\log(T^{1/5})}}C^8}{T} + \frac{4C^8 e^{2\sqrt{\log(T^{1/5})}}}{T^2} = \widetilde{\mathcal{O}}\Big(\frac{1}{T}\Big)\ .$$

Proof. We begin by decomposing the variance of $\widehat{A(1)}\widehat{A(0)}$ into covariances of products of the auxiliary random variables r_t and q_s. To this end, observe that

$$\begin{aligned}\mathrm{Var}(\widehat{A(1)}\widehat{A(0)}) &= \mathrm{Var}\Big(\frac{1}{T^2}\sum_{t=1}^{T}\sum_{s=1}^{T} y_t(1)^2 y_t(0)^2 r_t q_s\Big) \\ &= \frac{1}{T^4}\sum_{t=1}^{T}\sum_{s=1}^{T}\sum_{\ell=1}^{T}\sum_{k=1}^{T} y_t(1)^2 y_s(0)^2 y_\ell(1)^2 y_k(0)^2\,\mathrm{Cov}(r_t q_s, r_\ell q_k)\end{aligned}$$

Next, we will use the result of Lemma 2.23 to handle the individual covariance terms. In particular, the first six types of terms, as described in Lemma 2.23. The first three types of terms are at most 0, so we may discard them from the sum, as they contribute no positive value. The last three terms have upper bounds, which we use here to obtain the following upper bound:

$$\leq \underbrace{\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s\in[T]\setminus\{t\}}\sum_{k\in[T]\setminus\{t,s\}} y_t(1)^4 y_s(0)^2 y_k(0)^2\Big(\frac{1}{\delta_t}-1\Big)}_{\text{Term } T_1}$$

$$+ \underbrace{\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s\in[T]\setminus\{t\}}\sum_{\ell\in[T]\setminus\{t,s\}} y_t(1)^2 y_\ell(1)^2 y_s(0)^4\big(\frac{1}{\delta_s}-1\big)}_{\text{Term } T_2}$$

$$+ \underbrace{\frac{1}{T^4}\sum_{t=1}^{T}\sum_{s\neq t} y_t(1)^4 y_s(0)^4\big(\frac{1}{\delta_t\delta_s}-1\big)}_{\text{Term } T_3}$$

Our goal will now be to bound each of these terms individually.

Terms 1 and 2: Terms 1 and 2 are similar and will be handled in the same way. Let's begin with Term 1. Observe that by Lyapunov's inequality, the moment assumptions, and Lemma 2.22, we have that

$$\begin{aligned}
T_1 &\le \frac{1}{\delta_T\cdot T^4}\Big(\sum_{t=1}^{T} y_t(1)^4\Big)\Big(\sum_{t=1}^{T} y_t(0)^2\Big)^2 \\
&\le \frac{2e^{\sqrt{\log(T^{1/5})}}}{T}\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(1)^4\Big)\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(0)^2\Big)^2 && \text{(Lemma 2.22)} \\
&= \frac{2e^{\sqrt{\log(T^{1/5})}}}{T}\left[\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(1)^4\Big)^{1/4}\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(0)^2\Big)^{1/2}\right]^4 \\
&\le \frac{2e^{\sqrt{\log(T^{1/5})}}}{T}\left[\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(1)^4\Big)^{1/4}\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(0)^4\Big)^{1/4}\right]^4 && \text{(Lyapunov's inequality)} \\
&\le \frac{2e^{\sqrt{\log(T^{1/5})}}C^8}{T} && \text{(Assumption 2.3)}
\end{aligned}$$

A similar argument shows that $T_2 \le (2e^{\sqrt{\log(T^{1/5})}}C^8)/T$.

Term 3: The third term may be upper bounded using Lemma 2.22 and the moment assumptions. Namely,

$$\begin{aligned}
T_3 &\le \frac{1}{\delta_T^2 T^4}\Big(\sum_{t=1}^{T} y_t(1)^4\Big)\Big(\sum_{t=1}^{T} y_t(0)^4\Big) \\
&= \frac{\big(2e^{\sqrt{\log(T^{1/5})}}\big)^2}{T^2}\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(1)^4\Big)\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(0)^4\Big) && \text{(Lemma 2.22)} \\
&\le \frac{4C^8 e^2\sqrt{\log(T^{1/5})}}{T^2} && \text{(Assumption 2.3)}
\end{aligned}$$

Taken together, this shows that

$$\mathrm{Var}(\widehat{A(1)A(0)}) \leq \frac{4e^{\sqrt{\log(T^{1/5})}}C^8}{T} + \frac{4C^8e^{2\sqrt{\log(T^{1/5})}}}{T^2} = \widetilde{\mathcal{O}}\Big(\frac{1}{T}\Big) \ .$$

□

This establishes the following corollary, which shows that the error $\widehat{A(1)A(0)} - A(1)A(0)$ is going to zero at a near-parametric rate, which follows from by Chebyshev's inequality from Propositions 2.24 and 2.25.

Corollary 2.26. *The following error goes to zero:* $\widehat{A(1)A(0)} - A(1)A(0) = \widetilde{\mathcal{O}}_p\big(T^{-1/2}\big)$.

Using the results derived above, we are ready to prove Theorem 2.10, which we restate here for convenience.

Theorem 2.10. *Under Assumptions 2.3 and 2.9, and the parameters stated in Theorem 2.8, the error of the normalized variance estimator under* Clip-OGD *is* $T \cdot \widehat{VB} - T \cdot VB = \widetilde{\mathcal{O}}_p(T^{-1/2})$.

Proof of Theorem 2.10. Recall that the variance estimator and the variance bound are equal to $T \cdot \mathrm{VB} = 4S(1)S(0) = 4\sqrt{A(1)A(0)}$ and $T \cdot \widehat{\mathrm{VB}} = 4\sqrt{\widehat{A(1)A(0)}}$ so that the error is given by

$$T \cdot \widehat{\mathrm{VB}} - T \cdot \mathrm{VB} = 4\Big[\sqrt{A(1)A(0)} - \sqrt{\widehat{A(1)A(0)}}\Big] \ .$$

Corollary 2.26 states that the error $\widehat{A(1)A(0)} - A(1)A(0)$ is on the order of $\widetilde{\mathcal{O}}_p\big(T^{-1/2}\big)$. By Assumption 2.3, we have that $A(1)A(0) = \sqrt{S(1)^2S(0)^2} > c^2$. Observe that the square root function $g(x) = \sqrt{x}$ is Lipschitz on the interval (c^2, ∞). Thus, by a rate-preserving Continuous Mapping Theorem, we have that the error of the normalized variance estimator is on the same order i.e. $T \cdot \mathrm{VB} - T \cdot \widehat{\mathrm{VB}} = \sqrt{A(1)A(0)} - \sqrt{\widehat{A(1)A(0)}} = \widetilde{\mathcal{O}}_p\big(T^{-1/2}\big)$. □

2.8.9 Valid Confidence Intervals (Corollary 2.11)

We now prove the asymptotic validity of the associated Chebyshev-type intervals. This proof is standard in the design-based literature, but we present it here for completeness.

Corollary 2.11. *Under Assumptions 2.3 and 2.9, and parameters stated in Theorem 2.8, Chebyshev-type intervals are asymptotically valid: for all* $\alpha \in (0, 1]$, $\liminf_{T\to\infty} \Pr(\tau \in \hat{\tau} \pm \alpha^{-1/2}\sqrt{\widehat{VB}}) \geq 1 - \alpha$.

Proof. Define the random variables

$$Z = \frac{\tau - \hat{\tau}}{\sqrt{\mathrm{Var}(\hat{\tau})}} \quad \text{and} \quad Z' = \frac{\tau - \hat{\tau}}{\sqrt{\widehat{\mathrm{VB}}}} \ .$$

Observe that they are related in the following way:

$$Z' = \frac{\tau - \hat{\tau}}{\sqrt{\widehat{\mathrm{VB}}}} = \frac{\tau - \hat{\tau}}{\sqrt{\mathrm{Var}(\hat{\tau})}} \cdot \Big(\sqrt{\frac{\mathrm{Var}(\hat{\tau})}{\mathrm{VB}}} \cdot \sqrt{\frac{\mathrm{VB}}{\widehat{\mathrm{VB}}}}\Big) = Z \cdot \Big(\sqrt{\frac{\mathrm{Var}(\hat{\tau})}{\mathrm{VB}}} \cdot \sqrt{\frac{T \cdot \mathrm{VB}}{T \cdot \widehat{\mathrm{VB}}}}\Big) \ .$$

By definition, we have that $\limsup_{T\to\infty} \mathrm{Var}(\hat{\tau})/\mathrm{VB} \leq 1$. Recall that by Proposition 2.14, $T \cdot \mathrm{VB} \geq T \cdot V_{\mathrm{N}} = \Omega(1)$ so that by Theorem 2.10 and Continuous Mapping Theorem, we have that $\sqrt{\frac{T\cdot\mathrm{VB}}{T\cdot\widehat{\mathrm{VB}}}} \xrightarrow{p} 1$. Thus, by Slutsky's theorem we have that Z' is asymptotically stochastically dominated by Z. Now we are ready to compute the coverage probability.

$$\begin{aligned}
\liminf_{T\to\infty} \Pr\Big(\tau \in \hat{\tau} \pm \alpha^{-1/2}\sqrt{\widehat{\mathrm{VB}}}\Big) &= \liminf_{T\to\infty} \Pr\Big(\Big|\frac{\tau - \hat{\tau}}{\sqrt{\widehat{\mathrm{VB}}}}\Big| \leq \alpha^{-1/2}\Big) \\
&= \liminf_{T\to\infty} \Pr\Big(\big|Z'\big| \leq \alpha^{-1/2}\Big) \\
&\geq \liminf_{T\to\infty} \Pr\Big(\big|Z\big| \leq \alpha^{-1/2}\Big) \\
&\geq 1 - \alpha \ ,
\end{aligned}$$

where the last line followed from Chebyshev's inequality and the fact that $\mathrm{Var}(Z) = 1$. □

2.8.10 Deferred Analyses of Alternative Designs

In this section, we provide analysis on the efficacy of existing adaptive experimental designs for the problem of Adaptive Neyman Allocation. To this end, we show two negative results. First, we show that the two-stage design of [38, 14] (i.e. Explore–then–Commit) can suffer linear expected Neyman Regret in the design-based framework for a large class of potential outcome sequences. Second, we show that multi-arm bandit algorithms which achieve sub-linear expected outcome regret will incur super-linear expected Neyman regret, providing further evidence that these two goals are incompatible.

Analysis of Explore-then-Commit (Proposition 2.12)

In this setting, we show that Explore-then-Commit designs can sometimes suffer linear Neyman regret, and therefore not recover the Neyman variance in large samples. We formally introduce our definition of Explore-then-Commit designs below. Let $p^*_{T_0}$ be defined as

$$p^*_{T_0} = \left(1 + \sqrt{\frac{\sum_{t=1}^{T_0} y_t(0)^2}{\sum_{t=1}^{T_0} y_t(1)^2}}\right)^{-1} \ ,$$

which is the optimal Neyman probability when considering only the sample up to T_0.

Our definition of ETC encompasses many possible ways of estimating the optimal treatment probability. The only requirement is that the estimation method will converge to $p^*_{T_0}$

Algorithm 2: EXPLORE-THEN-COMMIT

Input: Exloration phase size T_0

`// Explore Phase`

for $t = 1 \dots T_0$ **do**

  Sample treatment assignment Z_t as 1 with probability 1/2 and 0 with probability 1/2

  Observe outcome $Y_t = \mathbf{1}[Z_t = 1]y_t(1) + \mathbf{1}[Z_t = 0]y_t(0)$

Construct an estimator $\widehat{p^*_{T_0}}$ from observed data $(Z_1, Y_1, \dots Z_{T_0}, Y_{T_0})$ satisfying $\widehat{p^*_{T_0}} - p^*_{T_0} = \mathcal{O}_p(T_0^{-1/2})$.

`// Exploit Phase`

for $t = T_0 + 1 \dots T$ **do**

  Sample treatment assignment Z_t as 1 with probability $\widehat{p^*_{T_0}}$ and 0 with probability $1 - \widehat{p^*_{T_0}}$

  Observe outcome $Y_t = \mathbf{1}[Z_t = 1]y_t(1) + \mathbf{1}[Z_t = 0]y_t(0)$

at the rate $T_0^{-1/2}$. We consider $p^*_{T_0}$ rather than the true Neyman probability p^* because the observed data is informative only of the outcomes in the exploration phase T_0. Many natural estimators will fall into this class, including Horvitz–Thompson style estimators similar to those used in the construction of our variance estimator.

Before continuing, we provide a few more definitions. We define the second moments of treatment and control outcomes as well as the correlation in the exploration phase as

$$S_{T_0}(1)^2 = \frac{1}{T}\sum_{t=1}^{T_0} y_t(1)^2 \quad S_{T_0}(0)^2 = \frac{1}{T}\sum_{t=1}^{T_0} y_t(0)^2 \quad \text{and} \quad \rho_{T_0} = \frac{\frac{1}{T_0}\sum_{t=1}^{T} y_t(1)y_t(0)}{S_{T_0}(1)S_{T_0}(0)} \ .$$

We are now ready to state the formal version of Proposition 2.12.

Proposition 2.12*. *Suppose that $T_0 = \Omega(T^\epsilon)$ for some $\epsilon > 0$ and further suppose that the outcome sequence satisfies the following properties for constants $C \geq c > 0$ and $c' > 0$:*

- *The second moments $S(1)$, $S(0)$, $S_{T_0}(1)$, and $S_{T_1}(0)$ are contained in the interval $[c, C]$.*
- *The correlations are bounded away from -1, i.e. $\rho_{T_0}, \rho \geq -1 + c$.*
- *The second moments satisfy the following:*

$$S(1)^2\Big(\frac{S(0)}{S(1)} - \frac{S_{T_0}(0)}{S_{T_0}(1)}\Big) + S(0)^2\Big(\frac{S(1)}{S(0)} - \frac{S_{T_0}(1)}{S_{T_0}(0)}\Big) \geq c'$$

Then, the Neyman Regret of ETC is at least linear in probability, $\mathcal{R}_T = \Omega_p(T)$.

The first two conditions are essentially extensions of Assumption 2.3 to the exploration phase. This ensures that the probability $p^*_{T_0}$ (which is estimated in the Explore-then-Commit design) does not approach 0 or 1. The third condition is what really makes Explore-then-Commit fail to achieve sublinear Neyman regret. This condition states that the ratio of the second moments in the exploration phase is different than in the larger sequence. For example, if $S(1) = S(0) = 1$ but $S_{T_0}(0)/S_{T_0}(1) = 2$ then the condition would hold. In this case, we should not expect Explore-then-Commit to achieve the Neyman variance because the exploration phase does not contain sufficient information about the optimal Neyman probability. We now prove the proposition.

Proof. Let $p^* = \arg\min_{p\in[0,1]} \sum_{t=1}^T f_t(p)$ be the Neyman probability. We begin by re-arranging terms in the Neyman regret:

$$\begin{aligned}
\mathcal{R}_T &= \sum_{t=1}^T f_t(p_t) - \sum_{t=1}^T f_t(p^*) && \text{(def of Neyman regret)} \\
&= \sum_{t=1}^{T_0} f_t(p_t) - f_t(p^*) + \sum_{t=T_0+1}^{T} f_t(p_t) - f_t(p^*) && \text{(splitting terms by phases)} \\
&= \sum_{t=1}^{T_0} f_t(1/2) - f_t(p^*) + \sum_{t=T_0+1}^{T} f_t(\widehat{p^*_{T_0}}) - f_t(p^*) && \text{(def of ETC)} \\
&= \sum_{t=1}^{T_0} f_t(1/2) - f_t(\widehat{p^*_{T_0}}) + \sum_{t=1}^{T} f_t(\widehat{p^*_{T_0}}) - f_t(p^*) && \text{(adding + subtracting)} \\
&\geq \sum_{t=1}^{T_0} f_t(p^*_{T_0}) - f_t(\widehat{p^*_{T_0}}) + \sum_{t=1}^{T} f_t(\widehat{p^*_{T_0}}) - f_t(p^*)
\end{aligned}$$

Where the inequality follows by the optimality of $p^*_{T_0}$ on the exploration phase. By adding and subtracting the $f_t(p^*_{T_0})$ to the second sum, we obtain the following decomposition

$$= \underbrace{\sum_{t=1}^{T_0} f_t(p^*_{T_0}) - f_t(\widehat{p^*_{T_0}})}_{\text{Term 1}} + \underbrace{\sum_{t=1}^{T} f_t(\widehat{p^*_{T_0}}) - f_t(p^*_{T_0})}_{\text{Term 2}} + \underbrace{\sum_{t=1}^{T} f_t(p^*_{T_0}) - f_t(p^*)}_{\text{Term 3}}$$

We handle each of the terms separately in the remainder of the proof.

Term 1: By the assumptions on the second moments and correlation of outcomes in the exploration phase, we have that $p^*_{T_0}$ is bounded away from 0 and 1 by a constant. Furthermore, the cost functions f_t are Lipschitz on the interval $[\gamma, 1-\gamma]$ for any fixed γ. Thus, we may apply the quantitative Continuous Mapping Theorem to bound the absolute

value of Term 1 as follows:

$$\begin{aligned}
|\sum_{t=1}^{T_0} f_t(p^*_{T_0}) - f_t(\widehat{p^*_{T_0}})| &\leq \sum_{t=1}^{T_0} |f_t(p^*_{T_0}) - f_t(\widehat{p^*_{T_0}})| && \text{(triangle inequality)} \\
&\leq \sum_{t=1}^{T_0} \mathcal{O}_p(T_0^{-1/2}) && \text{(estimator property + CMT)} \\
&= \mathcal{O}_p(T_0 \cdot T_0^{-1/2}) \\
&= \mathcal{O}_p(T_0^{1/2}) \\
&= \mathcal{O}_p(T^{1/2}) \ ,
\end{aligned}$$

where the final inequality follows from the fact that $T_0 \leq T$.

Term 2: A similar argument may be applied to the second term. Again, we may apply the continuous mapping theorem as before to obtain

$$\begin{aligned}
|\sum_{t=1}^{T} f_t(p^*_{T_0}) - f_t(\widehat{p^*_{T_0}})| &\leq \sum_{t=1}^{T} |f_t(p^*_{T_0}) - f_t(\widehat{p^*_{T_0}})| && \text{(triangle inequality)} \\
&\leq \sum_{t=1}^{T} \mathcal{O}_p(T_0^{-1/2}) && \text{(estimator property + CMT)} \\
&= \mathcal{O}_p(T \cdot T_0^{-1/2}) \\
&= \mathcal{O}_p(T \cdot T^{-\epsilon/2}) \\
&= \mathcal{O}_p(T^{1-\epsilon/2}) \ ,
\end{aligned}$$

where we have used the assumption that $T_0 = \Omega(T^\epsilon)$.

Term 3: We now handle the third term. Let V_0 be the variance of adaptive Horvitz–Thompson estimator under the Bernoulli design when using the treatment probability $p^*_{T_0}$. By construction of the cost functions, we have that they are related to the variance as follows:

$$\begin{aligned}
&\sum_{t=1}^{T} f_t(p^*_{T_0}) - f_t(p^*) \\
&\quad = T \cdot \Big[T \cdot V_0 - T \cdot V_{\mathrm{N}} \Big] \\
&\quad = T \cdot \Bigg[\bigg(S(1)^2 \Big\{ \frac{1}{p^*_{T_0}} - 1 \Big\} + S(0)^2 \Big\{ \frac{1}{1 - p^*_{T_0}} - 1 \Big\} \bigg) \\
&\qquad\qquad - \bigg(S(1)^2 \Big\{ \frac{1}{p^*} - 1 \Big\} + S(0)^2 \Big\{ \frac{1}{1 - p^*} - 1 \Big\} \bigg) \Bigg] \\
&\quad = T \cdot \Bigg[S(1)^2 \Big(\frac{1}{p^*_{T_0}} - \frac{1}{p^*} \Big) + S(0)^2 \Big(\frac{1}{1 - p^*_{T_0}} - \frac{1}{1 - p^*} \Big) \Bigg]
\end{aligned}$$

$$= T \cdot \left[S(1)^2 \Big(\frac{S(0)}{S(1)} - \frac{S_{T_0}(0)}{S_{T_0}(1)} \Big) + S(0)^2 \Big(\frac{S(1)}{S(0)} - \frac{S_{T_0}(1)}{S_{T_0}(0)} \Big) \right] \ ,$$

where the last equality follows by definition of the probabilities. By Assumption, we have that this bracketed term is constant so that the third term is linear in T.

Putting these together, we have that the Neyman regret is lower bounded as

$$\mathcal{R}_T \geq \Omega(T) - \mathcal{O}_p(T^{1-\epsilon/2}) - \mathcal{O}_p(T^{1/2}) = \Omega_p(T) \ . \qquad \square$$

Analysis of Designs for Outcome Regret (Proposition 2.13)

In this section, we prove Proposition 2.13, which establishes that outcome regret and Neyman regret cannot be simultaneously minimized in general. We restate a more formal version of the proposition here. In order for a simpler proof, we make restrictions that the units have constant treatment effect and that each of the individual outcomes are more strictly bounded. We conjecture that this trade-off holds under weaker conditions.

Proposition 2.13*. *Let $\mathcal{A}$ be an adaptive treatment algorithm achieving sublinear outcome regret, i.e. there exists $q \in (0,1)$ such that $\mathbb{E}[\mathcal{R}_T^{outcome}] \leq O(T^q)$ for all outcome sequences satisfying Assumption 2.3. Consider an outcome sequence satisfying Assumption 2.3 with constants $C \geq c > 0$ and the additional conditions: (i) $\max_{1 \leq t \leq T} y_t(0)^2 \leq C^2$, and (ii) for all $t \in [T]$, $y_t(1) - y_t(0) = \tau$ and $\tau > c'$ for a constant $c' > 0$. Then, $\mathcal{A}$ suffers super-linear Neyman regret on this outcome sequence: $\mathbb{E}[\mathcal{R}_T] \geq \Omega(T^{2-q})$.*

Proof. To begin, note that the expected outcome regret is equal to

$$\begin{aligned}
\mathbb{E}[\mathcal{R}_T^{\text{outcome}}] &= \mathbb{E}\Big[\max_{k \in \{0,1\}} \sum_{t=1}^T y_t(k) - \sum_{t=1}^T Y_t \Big] && \text{(def of regret)} \\
&= \sum_{t=1}^T y_t(1) - \sum_{t=1}^T \mathbb{E}\big[Y_t\big] && (\tau > 0) \\
&= \sum_{t=1}^T y_t(1) - \sum_{t=1}^T \mathbb{E}\big[y_t(1)\mathbf{1}[Z_t = 1] + y_t(0)\mathbf{1}[Z_t = 0] \big] \\
&= \sum_{t=1}^T y_t(1) - \sum_{t=1}^T y_t(1)\mathbb{E}[p_t] + y_t(0) \cdot \big(1 - \mathbb{E}[p_t]\big) \\
&= \sum_{t=1}^T \big(y_t(1) - y_t(0)\big) \cdot \big(1 - \mathbb{E}[p_t]\big) \\
&= \tau \cdot \sum_{t=1}^T \big(1 - \mathbb{E}[p_t]\big)
\end{aligned}$$

where the last equality follow as $y_t(1) - y_t(0) = \tau$ for all $t \in [T]$ by assumption. This amounts to re-expressing the outcome regret in terms of the expected treatment probabilities played by algorithm $\mathcal{A}$. Because the outcome sequence satisfies Assumption 2.3, the expected outcome regret is at most $\mathbb{E}[\mathcal{R}_T^{\text{outcome}}] \leq \beta \cdot T^q$ for some constant β. By the above, this implies that the expectation of the sum of probabilities $1 - p_t$ must be small,

$$\sum_{t=1}^{T} \big(1 - \mathbb{E}[p_t]\big) \leq \frac{\beta}{\tau} T^q \ .$$

Next, we show that $\mathcal{A}$ must incur a large cost with respect to the functions f_t in the definition of Neyman regret. To do this, we will use a weighted version of the AM-HM inequality which states that for $x_1 \dots x_T > 0$ and $w_1 \dots w_n \geq 0$, we have that

$$\frac{\sum_{t=1}^{T} w_t x_t}{\sum_{t=1}^{T} w_t} \geq \frac{\sum_{t=1}^{T} w_t}{\sum_{t=1}^{T} \frac{w_t}{x_t}} \ .$$

The usual AM-HM inequality is recovered when $w_t = 1/T$. We now bound the expected Neyman loss, observing that

$$\begin{aligned}
\mathbb{E}\Big[\sum_{t=1}^{T} f_t(p_t)\Big] &= \mathbb{E}\Big[\sum_{t=1}^{T} \frac{y_t(1)^2}{p_t} + \frac{y_t(0)^2}{1-p_t}\Big] && \text{(def of } f_t\text{)} \\
&\geq \mathbb{E}\Big[\sum_{t=1}^{T} \frac{y_t(0)^2}{1-p_t}\Big] && \text{(non-negativity)} \\
&= \sum_{t=1}^{T} y_t(0)^2 \cdot \mathbb{E}\Big[\frac{1}{1-p_t}\Big] && \text{(linearity of } \mathbb{E}[\cdot]\text{)} \\
&\geq \sum_{t=1}^{T} y_t(0)^2 \cdot \frac{1}{\mathbb{E}\big[1-p_t\big]} && \text{(Jensen's inequality)} \\
&\geq \frac{\Big(\sum_{t=1}^{T} y_t(0)^2\Big)^2}{\sum_{t=1}^{T} y_t(0)^2 \cdot \mathbb{E}\big[1-p_t\big]} && \text{(weighted AM-HM)} \\
&\geq T^2 \frac{\Big(\frac{1}{T}\sum_{t=1}^{T} y_t(0)^2\Big)^2}{\max_{1\leq t\leq T} y_t(0)^2} \cdot \frac{1}{\sum_{t=1}^{T} \mathbb{E}\big[1-p_t\big]} \\
&= T^2 \frac{S(0)^2}{\max_{1\leq t\leq T} y_t(0)^2} \cdot \frac{1}{\sum_{t=1}^{T} \mathbb{E}\big[1-p_t\big]} \\
&\geq T^2 \frac{c^2}{C^2} \cdot \frac{\tau}{\beta} T^{-q} \\
&\geq \frac{c^2 c'}{C^2 \beta} T^{2-q} \ ,
\end{aligned}$$

where the last two inequalities follow from moment bounds in Assumption 2.3 together with the assumptions on the outcome sequence stated in the Theorem.

Next, we show that the optimal Neyman design incurs a much smaller cost. In particular,

$$\begin{aligned}
\min_{p\in[0,1]} \sum_{t=1}^{T} f_t(p) &\leq \sum_{t=1}^{T} f_t(1/2) \\
&= \sum_{t=1}^{T} \frac{y_t(1)^2}{1/2} + \frac{y_t(0)^2}{1/2} \\
&= 2T\left[\frac{1}{T}\sum_{t=1}^{T} y_t(1)^2 + \frac{1}{T}\sum_{t=1}^{T} y_t(0)^2\right] \\
&= 2T(S(1)^2 + S(0)^2) \\
&\leq 4C^2T \ .
\end{aligned}$$

Together, these facts establish that the Neyman regret for $\mathcal{A}$ is lower bounded as

$$\mathbb{E}[\mathcal{R}_T] = \mathbb{E}[\sum_{t=1}^{T} f_t(p_t) - \min_{p\in[0,1]} \sum_{t=1}^{T} f_t(p)] \geq \frac{c^2c'}{C^2\beta} T^{2-q} - 4C^2T \geq \Omega(T^{2-q}) \ . \qquad \square$$

Chapter 3

A Reporting-Based Framework for Post-Deployment Monitoring

The impact of unexpected side effects is most acutely felt by the individual patient. But if an individual experiences harm, how can they or their practitioner know whether their experience is an isolated incident or part of a larger pattern of treatment failures? Crucially, individual patients and clinicians themselves may not know whether a particular treatment reaction is generalizable and deserving of redress *until* it is placed in context with the experiences of others. This is because commonly-accepted notions of evidence in medicine are understood at an aggregate level. Therefore, since subgroup analyses are too cost prohibitive to conduct in a valid manner pre-deployment, for most treatments evidence of subgroup-specific problems can only be identified over time via *post-deployment monitoring*.

Luckily, mechanisms for individual incident reporting do already exist for medical devices [30, 99], vaccines [18, 100] and pharmaceuticals [84, 98]. A key component of these reporting databases is that information from individual reports are aggregated to build collective knowledge about specific medical devices, vaccines or pharmaceuticals—and, when applicable, this aggregated information can drive downstream decisionmaking, such as updating vaccine guidelines or drug treatment protocols (e.g., Oster et al. [82]).

However, thus far, identification of subgroup-specific risks or failures has been done in an ad-hoc manner, requiring statisticians inspecting the databases at a particular point in time based on 'hunches' derived from news reports, personal intuition, et cetera. Further, there is no established method for carefully accounting for reporting behavior, which is often necessary in order to draw statistical conclusions that count as evidence in the medical field.

In this chapter, I describe a new mechanism for post-deployment monitoring of incident reporting database in order to identify high risk subgroups. First, in Section 3.2, I describe a mathematical model of the problem described above. Then, in Section 3.3, I cover conditions on reporting behavior and show how they can be used to make inferences about rates of true harm. The proposed formalization of the problem together with the analysis of Section 3.3 allows for leveraging known approaches in sequential hypothesis testing. Section 3.4 describes how to instantiate two reasonable algorithms for the proposed test and provide theoretical

guarantees for each. Finally, in Section 3.5, the usefulness of the approach is illustrated using real-world vaccine incident report data, on which the proposed algorithm correctly identifies that young men are at a disproportionately high risk of myocarditis from COVID-19 mRNA vaccines using a smaller number of reports than the historical baseline. An overview of the framework and is given in Figure 3.1. It may be helpful to look at it quickly before continuing to the technical sections, as well as right after having concluded Sections 3.2-3.4 and before reading Sections 3.5 and 3.6.

This chapter is based on joint work with Jessica Dai, Inioluwa Deborah Raji, and Benjamin Recht, previously released as "From Individual Experience to Collective Evidence: A Reporting-Based Framework for Identifying Systemic Harms" [28].

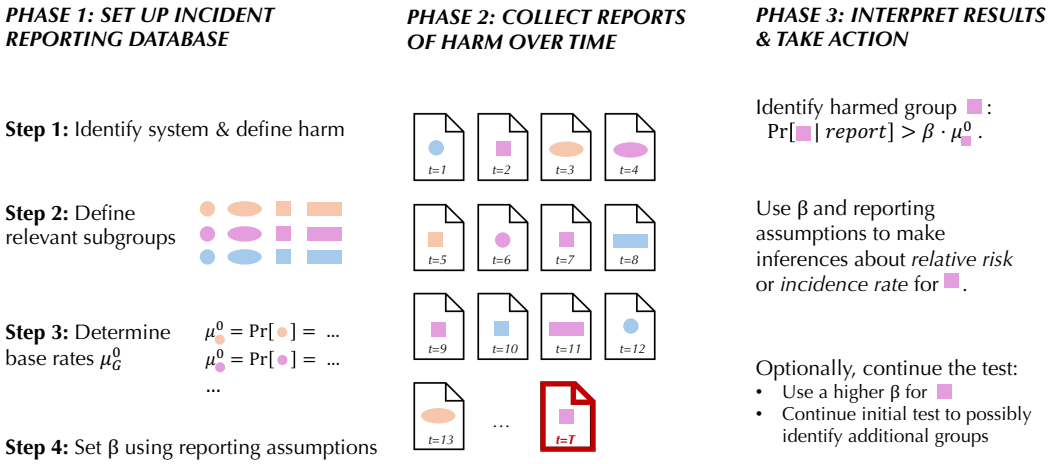


Figure 3.1: Overview of reporting database framework.

3.1 Related Work

The problem covered in this chapter lies at the intersection of various challenges addressed in fairness and statistics. We go over key related works below.

Accountability via (individual) reports. Work on identifying subgroup harm via reporting data has typically focused on learning in batch contexts, e.g. via positive-unlabeled learning for handling disparate reporting rates across subgroups (e.g., Shanmugam, Hou, and Pierson [93] and Wu and He [108]). In other works, identifying disparate reporting rates is itself is the central challenge (e.g., Liu and Garg [70] and Liu, Bhandaram, and Garg [69]). On the other hand, an emerging body of literature from the human-computer interaction community develops the concept of *contestability* (e.g., Almada [3], Vaccaro et al. [101], and Landau et al. [62]); though contestability is still typically understood in terms of individual

outcomes, we see our work as one possible path to implementing this ideal, with an eye towards empowering contestability at larger scale.

Post-deployment monitoring as hypothesis testing. Cherian and Candès [19] take a multiple testing approach for handling a large number of groups, but this test is again post-hoc (or entirely pre-deployment). Two more closely related works are that of Chugg et al. [22] and Feng et al. [31], who propose applying sequential hypothesis tests with the explicit goal of identifying problems in deployed systems in real time. However, as neither of these works study a reporting model, we propose fundamentally different tests: they test equality of means across different groups, while we compare within groups.

Identifying and defining subgroups. One approach to subgroup definition, following the line of work in multicalibration [45], is to simply enumerate over all possible combinations of covariates. For sequential problems, per-group guarantees can be provided for subgroups that are learned online [27], though these guarantees are in terms of prediction quality rather than statistical validity. For sequential experiments, Adam et al. [1] propose an approach to early stopping that does not require the experimenter to pre-specify the group experiencing harm, but instead identifies those who appear to be harmed more frequently. Though this is in spirit similar to the idea of identifying groups that report more frequently, their algorithm (and application context) is substantially different.

Sequential and multiple testing with anytime guarantees. One of our proposed tests provides anytime-validity guarantees by adapting the analysis of [54] and [8]. Our second proposed test leverages the recent literature on e-values (e.g. [105, 103]), which can be used to construct sequential tests that have validity guarantees in finite samples. While existing literature suggests methods for global null testing that can aggregate e-processes (e.g., Cho, Gan, and Kallus [21] or Chi, Ramdas, and Wang [20]), such approaches are unable to provide per-hypothesis guarantees.

Application & policy context. Sequential hypothesis tests have been used for real-world monitoring of adverse incidents in vaccines and medical devices (see, e.g., [94]). Descriptive studies have identified disparate adverse impacts in pharmaceutical [65, 106] and vaccine settings [82].

3.2 Model, Notation, and Preliminaries

The goal of constructing a reporting database is to determine whether some system that individuals interact with—for example, an (algorithmic) loan decision system, or a medical treatment—results in disproportionate harm to some meaningful subgroups. For the reporting database associated with a particular system, we will use $Y \in \{0, 1\}$ as an indicator

variable that denotes the undesirable event corresponding to that system. For example, in loan decisions, this could correspond to the event that a highly-qualified individual was denied a loan; in the medical setting, this may be an adverse physical side effect due to the treatment.

Subgroup definitions. Individuals are characterized with feature vectors $X \in \mathcal{X}$, and we index individuals as X_i ("features of individual i") or X_t ("features of the individual who reports at time t"). Every individual X_i "belongs to" at least one group G, and we will denote the event that X_i belongs to G as $\{X_i \in G\}$; we will use $\mathcal{G}$ to denote the set of all possible groups. This set of possible groups $\mathcal{G}$ can be defined arbitrarily as long as all groups can be determined as a function of covariates $\mathcal{X}$. We allow for groups to be overlapping—that is, we allow each individual X_i to be in multiple groups so that $|\{G' \in \mathcal{G} : X_i \in G'\}| \geq 1$. For example, it is possible to set $\mathcal{G} := 2^{\mathcal{X}}$ as in Hébert-Johnson et al. [45].

Reference population. The system for which the database is constructed naturally has a corresponding reference population of eligible individuals. For example, this could be everyone who has applied for a loan, or everyone who has been prescribed a certain medication. Thus, given a set of groups $\mathcal{G}$, we assume that it is possible to compute the composition of the reference population.

Assumption 3.1 (Reference population)**.** *For every $G \in \mathcal{G}$, the quantity $\mu_G^0 := \Pr[X \in G]$ is known. Throughout this work, we refer to the set $\{\mu_G^0\}_{G\in\mathcal{G}}$ as* base preponderances*.*

Probabilistic model of reporting. As the database administrator, the high-level goal is to determine whether there exists some subgroup $G \in \mathcal{G}$ where $\Pr[Y \mid X \in G]$ is abnormally high. Crucially, the database does not have access to information about every individual who interacts with the system; instead, individuals *may* report to the database if they believe that they experienced bad event Y. We thus let R_i be a random variable representing whether individual i decides to report (with $R_i = 0$ indicating no report).

Each report X_t is received sequentially, and assumed to be sampled i.i.d. from some underlying reporting distribution.[1] Given a group G, we denote its corresponding mean among reports $\Pr[X_t \in G \mid R_t = 1]$ as μ_G. We will sometimes refer to $\{\mu_G\}_{G\in\mathcal{G}}$ as (reporting) preponderances, as they represent the proportion of *reports* that each G comprises. A central claim of this work is that comparing μ_G to μ_G^0—i.e., the extent to which group G is (over)represented within the reporting database—can be a useful signal for $\Pr[Y \mid G]$ in a wide class of applications.[2] The i.i.d. model of course simplifies the analysis and exposition, but itself is not intrinsic to modeling the reporting database problem as a sequential hypothesis test. As we discuss in Section 3.6, the explicit i.i.d. assumption can be relaxed;

[1]Note that the events $\{X_t \in G\}$ and $\{X_t \in G'\}$ are correlated for any $G, G' \in \mathcal{G}$, i.e. the independence does not hold across groups. The key point in our case is independence across time.

[2]Because we allow groups to overlap, we cannot enforce $\sum_G \mu_G^0 = 1$ or $\sum_G \mu_G = 1$.

more generally, any probabilistic model for sequential testing can be adapted to reporting databases.

3.3 From Preponderance To Relative Risk

A major challenge of assessing potentially-differential rates of harm across subgroups using only reporting data is to relate the event that someone submits a report to the event that they experienced harm. That is, if someone did experience a negative outcome, how likely is it for them to have reported it, and conversely, if someone submitted a report, how likely is it to reflect "true" harm? Moreover, as is known from prior work, reporting rates themselves can vary across subgroups.

Our central proposal is to conduct a hypothesis test for each group to determine whether it is overrepresented by a factor of β among reports. That is, for each $G \in \mathcal{G}$, we test the following hypotheses:

$$\mathcal{H}_0^G : \mu_G < \beta\mu_G^0 \qquad\qquad \mathcal{H}_1^G : \mu_G > \beta\mu_G^0. \tag{3.1}$$

In Section 3.4, we will discuss concrete algorithms for conducting this test sequentially and their corresponding theoretical guarantees. Before doing so, we first argue that testing for preponderance among reports, i.e., tracking μ_G in this way, can be a meaningful way to identify subgroups with disproportionately high relative risk of side effects, even when exact reporting behavior is unknown.

Formally, we show that we can interpret our test in a way which allows us to make inferences about relative risk, the ratio between the rate of harm experienced by group G and on average over the population. The key quantity is the *report-to-incidence ratio*.

Definition 3.2 (Report-to-incidence ratio)**.** *We define the* report-to-incidence-ratio (RIR) *as* $\rho := \frac{\Pr[R=1]}{\Pr[Y=1]}$*, and the group-conditional analogue as* $\rho_G := \frac{\Pr[R=1|G]}{\Pr[Y=1|G]}$.

In Proposition 3.3, we show that if the group-conditional RIR of some group G is at most some constant multiple of the population-wide RIR, then we can convert a lower bound on report preponderance into a lower bound on true relative risk.

Proposition 3.3. *Define the relative risk of group* G *to be* $\mathrm{RR}_G := \frac{\Pr[Y=1|G]}{\Pr[Y=1]}$*. Suppose that for some group* G *we have* $\rho_G \le b \cdot \rho$*. Suppose that we determine that* $\mu_G \ge \beta\mu_G^0$ *for some* $\beta > 1$*. Then, the true relative risk experienced by* G *is at least* $\mathrm{RR}_G \ge \beta/b$.

Proof. First, note that by definition of ρ, ρ_G, and RR_G, we have

$$\rho_G \le b \cdot \rho \iff \frac{\Pr[R=1 \mid G]}{\Pr[Y=1 \mid G]} \le b \cdot \frac{\Pr[R=1]}{\Pr[Y=1]} \iff \mathrm{RR}_G \ge \frac{\Pr[R=1 \mid G]}{\Pr[R=1]} \cdot \frac{1}{b}.$$

By Bayes' rule, $\frac{\Pr[R=1|G]}{\Pr[R=1]} = \frac{\Pr[G|R=1]}{\Pr[G]} = \frac{\mu_G}{\mu_G^0}$; furthermore, by assumption, we have $\frac{\mu_G}{\mu_G^0} \ge \beta$. The result follows from combining with the previous display. □

Suppose we take $\max_G \rho_G/\rho \leq b = 1.25$, i.e., no group over-reports 25% more often than the population average. Then, if a test identifies a group G for which $\mu_G \geq 1.75 \cdot \mu_G^0$, this implies that the true relative risk for group G is at least $\text{RR}_G \geq 1.4$—that is, G experiences harm 40% more frequently relative to the population average.

3.4 Algorithms for Identifying Subgroups with High Reporting Overrepresentation

How might the test proposed in Equation (3.1) be carried out in practice, with reports arriving over time, and what properties might we want for such a test? In this section, we provide two ways to instantiate this sequential hypothesis test. For each, we provide two types of guarantees. The first is (sequential) α-validity, which, roughly speaking, guarantees correctness of groups identified in $\mathcal{G}^{\text{Flag}}$. More formally, we say that a sequential test is valid for a single group G at level α if $\Pr[\exists t : \mathcal{H}_0^G \text{ rejected}] \leq \alpha$ when $\mathcal{H}_0^G$ is true. Because we are testing for all groups in $\mathcal{G}$ simultaneously, we say that a sequential test is valid with respect to all groups $\mathcal{G}$ if $\Pr[\exists t, \exists G : \mathcal{H}_0^G \text{ erroneously rejected}] \leq \alpha$.

The second type of guarantee is power, which guarantees that the test will identify a harmed group, if one exists. In particular, we are interested in the *stopping time* T of the test, which is the number of samples required for the test to reject the first null, i.e. to raise an alarm for any group.

At a high level, our algorithms for conducting this test follow the protocol outlined in Algorithm 3. Every report X_t can be considered a binary vector indexed by the groups in $\mathcal{G}$; the G component of this vector is equal to $\mathbf{1}[X_t \in G]$. If there was only one group, we could run a sequential hypothesis test to determine whether μ_G was unacceptably large. With multiple groups, we can run $|\mathcal{G}|$ separate sequential hypothesis tests in parallel, one for each group, and correct the confidence levels for multiple hypothesis testing. For each group G, we maintain a test statistic ω_t^G that is updated as reports X_t are received over time. At each time t, each of these test statistics are compared to a threshold $\theta_t(\alpha)$, which depends on the test level α; the null hypothesis $\mathcal{H}_0^G$ for group G is rejected if $\omega_t^G > \theta_t(\alpha)$. For ease of exposition, Algorithm 3 is written so that groups corresponding to rejected nulls are collected in a set $\mathcal{G}^{\text{Flag}}$; in practice, a database administrator may choose to stop the test entirely as soon as one harmed group has been found.

Correcting for multiple hypothesis testing across groups is handled by a simple Bonferroni correction—that is, given a particular test level α, we test each individual group G at level $\alpha/|\mathcal{G}|$ rather than level α. Though Bonferroni corrections often seem onerous in non-sequential settings, we show that, for sequential problems, the Bonferroni correction incurs only a modest increase in stopping time.

In Section 3.4.1, we give a simple sequential Z-test-inspired approach which leverages a finite-time Law of the Iterated Logarithm. Section 3.4.2 presents a more complicated algorithm that leverages recent developments in anytime-valid inference. The main differences

in each algorithm lie in how they implement Lines 1 and 6 of Algorithm 3—that is, how test statistics and thresholds are computed. For each instantiation of Algorithm 3, we show validity and power guarantees. Proofs are deferred to Section 3.7.

Algorithm 3: General protocol for testing overrepresentation

Input: Set of groups $\mathcal{G}$; base preponderances $\{\mu_G^0\}_{G\in\mathcal{G}}$; test level α; relative strength β

1 Initialize test statistic ω_0^G for every $G \in \mathcal{G}$ and set threshold $\theta_0(\alpha)$;
2 Initialize set of rejected nulls (flagged groups) $\mathcal{G}^{\text{Flag}} := \emptyset$;
3 **for** $t = 1, 2, \ldots$ **do**
4 See report X_t;
5 **for** $G \in \mathcal{G}$ **do**
6 Update test statistic ω_t^G and compute threshold $\theta_t(\alpha)$;
7 **if** $\omega_t^G \geq \theta_t(\alpha)$ **then**
8 Add G to $\mathcal{G}^{\text{Flag}}$ and take requisite action for G, if applicable.

3.4.1 Sequential Z-test

One simple observation that arises from the model presented in Section 3.2 is that if each report X_t is drawn i.i.d. from some underlying distribution, then one might expect to be able to use concentration as a tool to conduct this test, since as time passes, the fraction of reports within the database from group G should converge to the true mean μ_G. We refer to this style of approach as a sequential Z-test, as it relies on measuring deviation from the mean.

Updating the test statistic ω_t^G. Given this intuition, the test statistic is a simple count of the number of times a report from each group has been seen, i.e. (with $\omega_0^G = 0$),

$$\omega_t^G \leftarrow \omega_{t-1}^G + \mathbf{1}[X_t \in G]. \tag{3.2}$$

Setting the threshold $\theta_t(\alpha)$. Given the way that ω_t^G accumulates evidence, one natural way to construct the threshold at each t is to use the mean under the alternative, plus a correction term for both sample complexity and repeated testing over time. With C set to either $\sqrt{\beta\mu_G^0(1-\beta\mu_G^0)}$ or $1/2$, the threshold (including a Bonferroni correction) is

$$\theta_t(\alpha) := t \cdot \beta\mu_G^0 + C\sqrt{2.07t \ln\left(|\mathcal{G}|\frac{(2+\log_2(t))^2}{\alpha}\right)}. \tag{3.3}$$

Theoretical guarantees. Our first guarantee is a bound on the probability that any group is incorrectly flagged.

Theorem 3.4 (Validity)**.** *Running Algorithm 3 with $\theta_t(\alpha)$ as in Equation* (3.3)*, setting $C = 1/2$, and ω_t^G updated as in Equation* (3.2)*, guarantees that the probability that $\mathcal{G}^{Flag}$ will ever contain a group G where $\mathcal{H}_0^G$ is true is at most α, i.e.*

$$\Pr\left[\exists t : \exists G \in \mathcal{G}^{Flag} \text{ s.t. } \mathcal{H}_0^G \text{ holds}\right] \leq \alpha.$$

The choice of C affects the nature of the guarantee: the true, finite-sample anytime-validity guarantee requires $C = 1/2$. If instead $C = \sqrt{\beta\mu_G^0(1-\beta\mu_G^0)}$, then, strictly speaking, the guarantee holds only asymptotically. However, a higher value of C affects stopping time unfavorably, so the asymptotic approximation can be useful practically. In this case, care must be taken to ensure that the algorithm does not erroneously reject too early due to noise; one way to implement this is to mandate a minimum stopping time.

Finally, we give a stopping time guarantee for this test.

Theorem 3.5 (Power)**.** *Let T be the stopping time of Algorithm 3 with $\theta_t(\alpha)$ as in Equation* (3.3)*, $C = 1/2$, and ω_t^G as in Equation* (3.2)*. Let $\Delta_{\max} = \max_{G\in\mathcal{G}} \mu_G - \beta\mu_G^0$. If $\Delta_{\max} > 0$, then $\Pr[T < \infty] = 1$. Furthermore, with probability $1 - \alpha/|\mathcal{G}|$, we have $T \leq \widetilde{\mathcal{O}}\left(\frac{\ln(|\mathcal{G}|)+\ln(1/\alpha)}{\Delta_{\max}^2}\right)$, and for any $\delta \in (0, \alpha/|\mathcal{G}|)$, we have with probability at least $1-\delta$ that $T \leq \widetilde{\mathcal{O}}\left(\frac{\ln(1/\delta)}{\Delta_{\max}^2}\right)$.*

3.4.2 Betting-style approach

We refer to our second algorithm as a *betting-style* approach, due to the way we construct our test statistics [92, 105, 103, 22]; one way to interpret this approach is that the test "bets against" the null hypothesis $\mathcal{H}_0^G$ being true. We direct the reader to these references for more detailed technical exposition and connection with literature on martingales, gambling, and finance. For us, these methods provide an adaptive algorithm which find a middle ground between the two approaches in the previous section: the betting-style approach achieves finite-sample validity but empirically terminates quickly when the null is false.

Updating the test statistic ω_t^G. As in the previous approach, we let ω_t^G represent some accumulated amount of evidence against the null hypothesis $\mathcal{H}_0^G$ by time t, with a higher value of ω_t^G corresponding to greater level of evidence.[3] We initialize $\omega_0^G = 0$, and use the update rule

$$\omega_t^G \leftarrow \omega_{t-1}^G + \ln\left(1 + \lambda_t^G(\mathbf{1}_{X_t\in G} - \beta\mu_G^0)\right), \tag{3.4}$$

with $\lambda_1^G, \ldots, \lambda_t^G \in [0, 1]$. Note that this expression is similar to the running sum used in Section 3.4.1. Here, the algorithm accumulates a nonlinear function, with an adaptive

[3]The quantity $\exp(\omega_t^G)$ can also be referred to as an *e-value* [103], a measure of evidence against a null hypothesis similar to a p-value.

parameter λ_t^G that weights the influence of each new sample. Our setting of λ_t is motivated by the goal of minimizing stopping time under the alternative, and thus to maximize ω_t^G. Taking $\lambda_t = 0$ means ω_t^G remains the same regardless of what new information is received at time t. On the other hand, $\lambda_t = 1/\beta\mu_G^0$ means that if we receive evidence in accordance with $\mathcal{H}_0^G$ then ω_t^G will decrease substantially; but, if we instead receive evidence *against* the null, i.e. $X_t \in G$, we maximally increase ω_t^G. Drawing from the well-studied problem of portfolio optimization in the online learning literature [24, 113, 43], we use Online Newton Step [44, 25] to ensure that ω_t^G is not too far from the best achievable in hindsight. This results in the following update for $\{\lambda_t\}_{t\geq 1}$:

$$\lambda_{t+1}^G \leftarrow \underset{[0,1]}{\text{Proj}}\left(\lambda_t^G + \frac{2}{2-\ln(3)} \cdot \frac{z_t}{1+\sum_{s\in[t]} z_s^2}\right), \tag{3.5}$$

where $z_t = \frac{\mathbf{1}[X_t \in G] - \beta\mu_G^0}{1+\lambda_t^G(\mathbf{1}[X_t\in G]-\beta\mu_G^0)}$, and $\lambda_0 = 0$.[4]

Setting the threshold $\theta_t(\alpha)$. Unlike the sequential Z-test, we use the same threshold for all timesteps. Including a Bonferroni correction, we use $\theta_t(\alpha) := \ln(|\mathcal{G}|/\alpha)$ for all t; the motivation for this setting will become clear in our discussion of Theorem 3.6.

Theoretical guarantees. We first give a validity guarantee that is essentially identical to the Sequential Z-test.

Theorem 3.6 (Validity)**.** *Running Algorithm 3 with $\theta_t(\alpha) = \ln(|\mathcal{G}|/\alpha)$ and ω_t^G updated as per Equations* (3.4) *and* (3.5) *guarantees that the probability that $\mathcal{G}^{Flag}$ will ever contains a group G where $\mathcal{H}_0^G$ is true is at most α, i.e.*

$$\Pr\left[\exists t : \exists G \in \mathcal{G}^{Flag} \text{ s.t. } \mathcal{H}_0^G \text{ holds}\right] \leq \alpha.$$

This result follows directly from the prior work referenced at the beginning of this section. At a high level, every sequence $\{\exp(\omega_t^G)\}_{t\geq 1}$ is a non-negative super-martingale under $\mathcal{H}_0^G$; informally, this means that under the null hypothesis, the sequence $\{\exp(\omega_t^G)\}_{t\geq 1}$ should be non-increasing, in expectation. This allows us to apply Ville's inequality, which guarantees that it is unlikely that $\exp(\omega_t^G)$ ever becomes too large under $\mathcal{H}_0^G$. More specifically, for any $\alpha \in (0,1)$, under the null, $\Pr[\exists t : \exp(\omega_t^G) > 1/\alpha] \leq \alpha$. Thus, maintaining a threshold of $\theta_t(\alpha) = \ln(|\mathcal{G}|/\alpha)$ is sufficient to provide a per-hypothesis $\alpha/|\mathcal{G}|$-validity guarantee, and thus α-validity overall.

We also provide the following bound on stopping time; see Section 3.7.2 for additional discussion of the $\omega_\star$ notion of gap.

[4]The constant $\frac{2}{2-\ln(3)}$ is due to Cutkosky and Orabona [25], who give a tighter version of ONS than in Hazan, Agarwal, and Kale [44].

Theorem 3.7 (Power)**.** *Let T be the stopping time of Algorithm 3 with $\theta_t(\alpha) = \ln(|\mathcal{G}|/\alpha)$ and ω_t^G updated as per Equations* (3.4) *and* (3.5)*. If $\max_{G\in\mathcal{G}} \mu_G - \beta\mu_G^0 > 0$, then, we have that $\Pr[T < \infty] = 1$. Furthermore,*

$$\mathbb{E}[T] \leq \mathcal{O}\left(\frac{1}{\omega_\star^2} + \frac{\ln(|\mathcal{G}|) + \ln(1/\alpha)}{\omega_\star}\right)$$

where $\omega_\star := \max_{G\in\mathcal{G},\lambda\in[0,1]} \mathbb{E}[\ln(1+\lambda(\mathbf{1}_{X_t\in G} - \beta\mu_G^0))]$ is the maximal expected one-step increase in ω_t^G over all groups and choices of λ.

We conclude this section with two further remarks on Theorems 3.5 and 3.7 in the context of our test. First, our modeling in Section 3.3 measures severity of harm via a *multiplicative* factor of overrepresentation. However, both notions of gap in Theorems 3.5 and 3.7 also on the absolute size of the group μ_G. Thus, for two groups G and G' with identical multiplicative gaps, i.e. $\mu_G/\mu_G^0 = \mu_{G'}/\mu_{G'}^0$, the test would stop faster in expectation for G if and only if $\mu_G^0 > \mu_{G'}^0$. That is, if two groups are "harmed" to the same extent, both algorithms will identify the larger one first.

Second, for both tests, the Bonferroni correction results in only an additive factor ($\ln(|\mathcal{G}|)/\Delta_{\max}^2$ in Theorem 3.4, and $\ln(|\mathcal{G}|)/\omega_\star$ in Theorem 3.6) in stopping time over the scenario where we had only been testing the one group with the largest gap. This means that, in terms of worst-case guarantee on stopping time, the contribution of the Bonferroni correction is small relative to the contribution of the test level α and, especially, to the gap. In fact, the impact of Bonferroni on real-world data appears to be much smaller even than this additive term.

3.5 Real-World Application: Myocarditis from COVID-19 vaccines

It is by now well-known that COVID-19 vaccines induce an elevated risk of myocarditis among young men. While initial suspicions of elevated myocarditis risk relied on case studies (e.g., Mouch et al. [76], Larson et al. [63], and Marshall et al. [75]), a more systematic understanding—including the pattern of disproportionate impact—was made possible by post-hoc analysis of reports from incident databases. Barda et al. [9] appears to be the first analysis based on a database of reports, but did not disaggregate by demographic subgroups; the confirmation of young men as the most drastically-impacted group came in later studies (e.g., Witberg et al. [107] and Oster et al. [82]).

In the U.S., these reports are collected in the Vaccine Adverse Event Reporting System (VAERS). If we had been able to run the hypothesis tests proposed in the preceding sections on the reports collected in VAERS, would we have correctly identified this problem—and if so, how quickly? Concretely, we let Y_i be the event that individual i experiences myocarditis after receiving a COVID-19 vaccine, and run the test with the end-goal of identifying elevated incidence rate $\Pr[Y_i \mid X_i \in G]$ for group(s) G corresponding to adolescent men (ages 12-17 and 18-29).

	Asymptotic Z-test		*Finite-sample Z-test*		*Betting-style test*	
	(M, 18-29)	(M, 12-17)	(M, 18-29)	(M, 12-17)	(M, 18-29)	(M, 12-17)
$\beta = 2.0$	34 (Feb. 22)	256 (May 10)	69 (Mar. 28)	530 (May 30)	61 (Mar. 23)	241 (May 8)
$\beta = 2.5$	49 (Mar. 10)	302 (May 15)	74 (Mar. 31)	546 (Jun. 1)	69 (Mar. 28)	259 (May 11)
$\beta = 3.0$	70 (Mar. 30)	324 (May 18)	111 (Apr. 20)	612 (Jun. 6)	80 (Apr. 5)	302 (May 15)

Table 3.1: On real historical sequence of myocarditis reports, time to identification of harmed groups. In each cell, we report the number of total reports to the rejection of the hypothesis corresponding to (M, 18-29) and the number of total reports corresponding to (M, 12-17). In all tests, the (M, 18-29) group is identified first—vaccines were authorized for the 12-15 age group only in May.

Data sources. The Vaccine Adverse Event Reporting System (VAERS) is a national adverse event incident reporting database for U.S.-licensed vaccines, co-managed by the Centers for Disease Control and Prevention (CDC) and the U.S. Food and Drug Administration (FDA) [18, 94]. The database is a combination of voluntary reports from patients that have received the vaccine, as well as mandatory reports from vaccine manufacturers and healthcare professionals. For this case study, we filter the set of publicly-available reports from VAERS to reports about the COVID-19 vaccine with a complaint keyword including "myocarditis." As for how a database administrator would have known to focus on myocarditis *a priori*, one might imagine, for example, that the series of case studies found in early 2021 raised the alarm that more systematic analysis was warranted for myocarditis in particular.

To determine per-demographic base rates, i.e. to compute $\{\mu^0_G\}_{G\in\mathcal{G}}$, we utilize VaxView, a government dataset tracking national vaccine coverage, managed by the CDC. VaxView does not track vaccination rates by granular subgroups, only providing coverage rates disaggregated by age, gender, and ethnicity separately. We thus impute the vaccination rates for intersections of subgroups (e.g., "12-17, M") by multiplying the known marginal rates (i.e., $\mu^0_{(12-17,M)} := \mu^0_{(12-17)} \cdot \mu^0_{(M)}$).

Defining $\mathcal{G}$. We consider (intersections of) sex and age buckets to be the subgroups of interest.[5] Age buckets are discretized into 0-4, 5-11, 12-17, 18-29, 30-39, 40-49, 50-64, 65-74, and 75+; the sex categories represented in the data are (binary) male and female. After removing groups for which no vaccines were recorded, $\mathcal{G}$ contains 29 groups.

Results. We begin by running our algorithms on the actual sequence of reports in chronological order, as received in VAERS. In particular, we consider Algorithm 3 instantiated

[5]While in principle it would have been interesting to also consider race/ethnicity, we are limited by the availability (and granularity) of the data given in VAERS, which does not include information on ethnicity/race in reports.

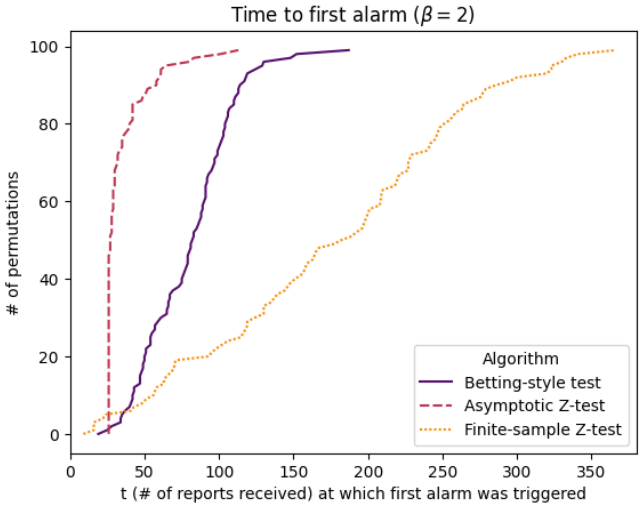


Figure 3.2: Number of reports (t) it takes for each algorithm to reject the null hypothesis for any group (i.e. first identification of harm), over 100 random permutations of COVID-19 vaccine report database. Tests are run with $\beta = 2$. Each point on the plot reflects the number of trials (out of 100) in which a rejection has occured by time t.

with ω_t^G updated according to Equation (3.2) and $\theta_t(\alpha)$ as in (3.3) and $C = 1/2$ (*Finite-sample Z-test*); with ω_t^G updated according to Equation (3.2) and $\theta_t(\alpha)$ as in (3.3) and $C = \sqrt{\beta\mu_G^0(1-\beta\mu_G^0)}$ (*Asymptotic Z-test*); and with ω_t^G updated according to Equations (3.4) and (3.5), and $\theta_t(\alpha) = \ln(|\mathcal{G}|/\alpha)$ (*Betting-style test*). For the asymptotically-valid Z-test, we require a minimum stopping time of $t = 25$, to prevent early rejections. We run all experiments for $\alpha = 0.1$.

In Table 3.1, we report the stopping time—that is, the number of reports it takes for the first null hypothesis to be rejected—of each algorithm for various values of β, as well as the corresponding date by which an alarm would have been triggered. Note that, in all tests, the (M, 18-29) group is identified first. This is consistent with the timeline of regulatory approvals: vaccines were authorized for ages 12-15 only by May 10 [72].

To explore the robustness of these results, we also run synthetic experiments, permuting the ordering of reports to get a sense of possible variance in the stopping time. We run 100 random permutations of the full set of reports. Figure 3.2 tracks the number of reports it takes for each algorithm to reject the null hypothesis for any group—that is, a scenario when the test is stopped and an alarm is raised as soon as one harmed group is identified. Each point on these plots reflects the number of trials (out of 100) in which a rejection has occurred by time t, when tests are run at $\beta = 2$.

In Figure 3.2, we compare the performance of the three algorithms. To interpret the figure, by time $t = 100$, the asymptotically-valid z-test had already identified harm in all 100 permutations; the betting-style test identified harm in around 80 permutations; and the finite-sample z-test had only identified harm in around 20 permutations. Figure 3.2 shows a clear ordering in terms of how quickly each algorithm tends to identify harm: the asymptotically-valid sequential z-test (dashed, red) is faster than the betting-style algorithm (solid, purple), which is faster than the finite-sample z-test (dotted, yellow).

We also explore the impact of Bonferroni correction for multiple hypothesis testing on stopping time. In Figure 3.3, we show the same axes—number of reports to first alarm on

the x-axis, vs. number of permutations in which an alarm was triggered on the y-axis—for the three algorithms at $\beta = 2$. As expected, the invalid version of the test, which has a lower threshold for rejecting each null, stops more quickly for all three algorithms (dashed, lighter). The difference between the invalid version and the valid version (solid, darker) is relatively minor, though the impact varies across algorithms.

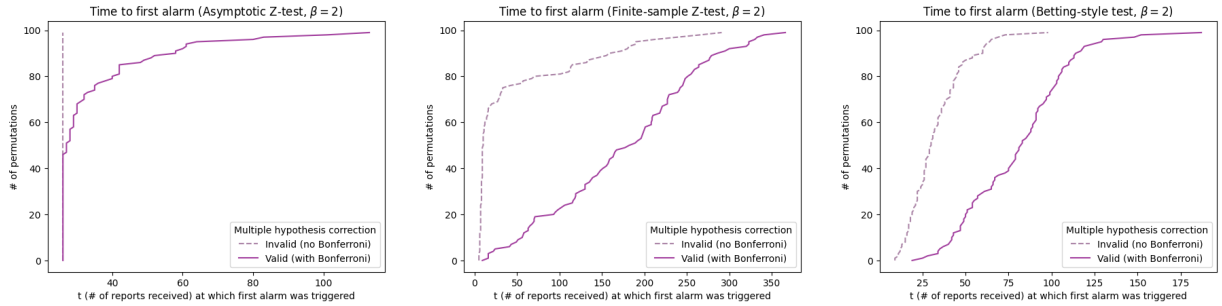


Figure 3.3: Impact of multiple hypothesis correction on stopping time across algorithms. As in Figure 3.2, each point on the plot reflects the number of trials (out of 100) in which a rejection has occurred by time t. In all plots, the lighter, dashed line reflects stopping time of the invalid test that does not correct for multiple testing; the dark, solid line reflects stopping time of the valid test including a Bonferroni correction.

Overall, our experimental results suggest that our proposed tests would in fact have been effective in determining that young men were disproportionately affected by myocarditis. Moreover, though it is difficult to determine exact timelines and the nature of clinical practice during early phases of the vaccine rollout, it is possible that such a test could have identified problems using less data—that is, more quickly—than was actually used for this finding.

Code for all experiments, including instructions for data download and pre-processing, is available online here.

3.6 Notes on Practical Considerations

Choosing $\mathcal{G}$. In our experiments in Section 3.5, we choose to define subgroups as all possible combinations of available demographic characteristics. That said, a practitioner may seek to define $\mathcal{G}$ more carefully in accordance with their application. For instance, if the goal is to illustrate discrimination in a legal sense, $\mathcal{G}$ should be defined with respect to (protected) demographic features, rather than arbitrary combinations of covariates. On the other hand, groups need not be solely demographic, which allows our approach to test for safety rather than solely fairness. For example, $\mathcal{G}$ could include which batch of a medication an individual received; our tests could then help identify whether some batches were improperly manufactured.

Baseline rates $\{\mu_G^0\}_{G\in\mathcal{G}}$. A natural question that arises from the modeling in this section is how $\{\mu_G^0\}_{G\in\mathcal{G}}$ can be determined, or if Assumption 3.1 is strictly necessary. Practically speaking, these base preponderances may be estimated, possibly with some amount of noise; however, the estimation problem can be addressed with standard techniques and is not core to our contribution. Similarly, in practice these baseline preponderances may change over time (e.g. if some subgroups increased uptake of a vaccine, or applied for loans more frequently, over time); however, such situations are relatively straightforward to handle under our algorithmic frameworks (see, e.g., the variants discussed in Chugg et al. [22]). We therefore focus on the case where we have access to the true, underlying values of $\{\mu_G^0\}_{G\in\mathcal{G}}$ for ease and clarity of exposition.

Note that testing against base preponderances of the reference population (i.e., to compare μ_G to μ_G^0) is a new test proposed by this work, and the analysis in Section 3.3 is specific to this test. Existing approaches to monitoring in reporting databases compare to different baselines, most commonly the historical overall incidence rate for the specific symptom, sometimes by subgroup [94, 61, 82]. These baselines could, in principle, be plugged into the algorithms in Section 3.4, but new analysis for (possibly group-varying) reporting rates would be necessary to draw inferences about analogous quantities of interest (e.g., RR or IR), as current approaches do not generally consider reporting behavior. In contrast, our modeling allows us to identify what quantities may affect the true incidence rate even if they may be unmeasurable.

Setting β. Finally, to run the test proposed in Equation (3.1), it is necessary to determine how to set the value of β. As we will see in Section 3.4, when β is set too high, then the test may never identify problematic groups, or identify them more slowly; on the other hand, as is clear from the previous subsections, rejecting the null hypothesis for a smaller β requires more stringent assumptions on reporting behavior. Thus, we suggest a procedure to set β as follows: (1) choose a relative risk threshold where it would be problematic for any group if RR_G surpassed that threshold; (2) make the corresponding assumptions about reporting behavior; (3) use these quantities to compute a reasonable β. We give some example computations in Section 3.5. Due to an equivalence between hypothesis testing and confidence intervals, it is statistically valid to rerun tests with different βs once data collection has begun. Thus, it may be prudent to begin by setting the lowest β that reporting assumptions would allow; then, if the tests appear to be stopping very quickly, to re-run them at higher βs, which would allow a practitioner to get a better sense of the severity of the harm.

3.7 Deferred Proofs

In this section we collect all deferred proofs.

3.7.1 Proofs for Sequential Z-test

We prove Theorem 3.4, restated above. To prove this result, we will use a foundational result known as Ville's inequality [102].

Theorem 3.8 (Ville's inequality)**.** *Let $\{M_t\}_{t\in\mathbb{N}^+}$ be a non-negative supermartingale, i.e. for all t, $M_t \geq 0$, and $\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] \leq M_t$, where $\mathcal{F}_t$ is the filtration (history) of all realizations of randomness up to and including time t. Then, for any $x \in (0,1)$, we have $\Pr[\exists t : M_t > {}^{\mathbb{E}[M_0]}/_x] \leq x$.*

The central thrust of our proof of Theorem 3.4 is due to Koolen [59] (which itself draws from Balsubramani [8], and is a refinement of Jamieson et al. [54]); we reproduce the argument in the context of our work below, though we emphasize that we do not claim the proof technique as ours.

Proof of Theorem 3.4. It is sufficient to show that for any group G where $\mathcal{H}_0^G$ holds, we have $\Pr[\exists t : G \in \mathcal{G}^{\text{Flag}}] \leq {}^{\alpha}/_{|\mathcal{G}|}$; the statement of the theorem follows from the Bonferroni correction over all $|\mathcal{G}|$ hypotheses.

Ville's inequality (Theorem 3.8) appears similar in form to the statement we hope to prove; we therefore seek to transform our test statistic $\omega_t^G = \sum_{s\in[t]} \mathbf{1}[X_s \in G]$ into a quantity that can be interpreted as a (non-negative) supermartingale. Although $\{\omega_t^G\}_{t\in\mathbb{N}^+}$ is by itself clearly not a non-negative supermartingale, each ω_t^G is the sum of t Bernoulli trials with mean μ_G, and Bernoulli random variables are sub-Gaussian with variance parameter ${}^1/_4$. Each ω_t^G therefore satisfies the property that $\mathbb{E}[\exp(\eta(\omega_t^G - \mathbb{E}[\omega_t^G])] \leq \exp(\eta^2/8)$.

This holds for any η, so we will construct a distribution ϕ on η and use it to construct a martingale M_t. In particular, note that under $\mathcal{H}_0^G$, $\mathbb{E}[\omega_t^G] < t \cdot \beta\mu_G^0$. Thus, we let $S_t \coloneqq \omega_t^G - \mathbb{E}[\omega_t^G] = \omega_t^G - t\beta\mu_G^0$. We will let $M_t = \int \phi(\eta)\exp(\eta S_t - t\eta^2/8)d\eta$. Then, for any distribution ϕ, $\{M_t\}_{t\in\mathbb{N}^+}$ is a non-negative supermartingale with respect to the randomness in realizations of reports X_t. To see this, we have

$$\begin{aligned}
\mathbb{E}[M_{t+1} \mid \mathcal{F}_t] &= \mathbb{E}\left[\int \phi(\eta)\exp\left(\eta(S_t + \mathbf{1}[X_{t+1}\in G] - \beta\mu_G^0) - \tfrac{(t+1)\eta^2}{8}\right)d\eta \;\middle|\; \mathcal{F}_t\right] \\
&= \int \phi(\eta)\exp\left(\eta S_t - \tfrac{t\eta^2}{8}\right)\mathbb{E}\left[\exp\left(\eta(\mathbf{1}[X_{t+1}\in G] - \beta\mu_G^0) - \tfrac{(t+1)\eta^2}{8}\right) \;\middle|\; \mathcal{F}_t\right]d\eta \\
&\leq \int \phi(\eta)\exp\left(\eta S_t - \tfrac{t\eta^2}{8}\right)d\eta \\
&= M_t,
\end{aligned}$$

where the inequality is due to $\frac{1}{t}\mathbb{E}[\omega_t^G] \leq \beta\mu_G^0$ and subgaussianity. It thus remains to use this martingale to compute an appropriate threshold $\theta_t(\alpha)$ on ω_t^G.

M_t will satisfy the conditions of Theorem 3.8 for any choice of ϕ, including one which puts point mass of 1 on $\eta = \eta'$ and 0 elsewhere, i.e. $\phi(\eta') = 1$ and $\phi(\eta) = 0$ for any $\eta \neq \eta'$. One path towards establishing the threshold $\theta_t(\alpha)$ is to simply pick one value of η; however, such

an η cannot depend on t and would thus result in a suboptimal threshold. Instead, we will construct ϕ such that it is a discrete distribution, indexed by $i \in \mathbb{N}^+$, so that η takes values $\eta_1, \ldots, \eta_i$ with probability $\phi_1, \ldots, \phi_i$; this allows each η_i to depend on t and therefore more finer-grained optimization of the threshold. Before committing to the exact distribution ϕ, we first illustrate how ϕ_i and η_i will be used in the threshold.

Note that $M_t = \sum_{i \in \mathbb{N}^+} \phi_i \exp(\eta_i S_t - t\eta_i^2/8) \geq \max_i \phi_i \exp(\eta_i S_t - t\eta_i^2/8)$, so for any δ, we have

$$\{M_t \geq 1/\delta\} \supseteq \{\max_i \phi_i \exp(\eta_i S_t - t\eta_i^2/8) > 1/\delta\} = \left\{S_t \geq \min_i \left(\frac{t\eta_i}{8} + \frac{1}{\eta_i} \ln \frac{1}{\phi_i \delta}\right)\right\},$$

and thus, picking $\theta_t(\alpha) = t\beta\mu_G^0 + \min_i \left(\frac{t\eta_i}{8} + \frac{1}{\eta_i} \ln \frac{1}{\phi_i \alpha/|\mathcal{G}|}\right)$ would guarantee that $\Pr[\exists t : \omega_t^G > \theta_t(\alpha)] \leq \alpha/|\mathcal{G}|$.

Finally, we must commit to ϕ_i, η_i, then optimize for i. Let $\phi_i = \frac{1}{i(i+1)}$ (note that $\sum_i \phi_i = 1$, so this is a valid distribution), $\eta_i = 2\sqrt{\frac{2\ln(1/\phi_i(\alpha/|\mathcal{G}|))}{2^i}}$, and $i = \lfloor \log_2(t) \rceil$. For $i = \log_2(t)$ (without rounding), this would have yielded $\eta_i = 2\sqrt{\frac{2\ln((\log_2(t)+1)(\log_2(t))/(\alpha/|\mathcal{G}|))}{t}}$ and $\theta_t(\alpha) = \frac{1}{2}\sqrt{2t\ln((\log_2 t)(\log_2 t + 1)/\alpha/|\mathcal{G}|)}$. The statement follows from handling the numerical impact of rounding. □

Remark 3.9. *A key constant in the proof of the version of the algorithm that is valid in finite samples is the subgaussian variance parameter, for which we used 1/4 (and which propagates to a multiplicative factor of $\sqrt{1/4} = 1/2$ on the threshold). This is because the variance any Bernoulli is at most 1/4; however, this also motivates the choice of constant for the asymptotically-valid version of the test, which instead uses the variance parameter $\beta\mu_G^0(1 - \beta\mu_G^0)$.*

We now prove the power result, Theorem 3.5, restated above.

Proof. Let $G^\star \coloneqq \arg\max_{G \in \mathcal{G}} \mu_G - \beta\mu_G^0$ and let $\Delta \coloneqq \mu_{G^\star} - \beta\mu_{G^\star}^0$. Without loss of generality, we can consider only the test corresponding to $G^\star$ (while still testing at level $\alpha/|\mathcal{G}|$). Recall that for this instantiation of Algorithm 3, the test statistic $\omega_t^{G^\star} = \sum_{s \in [t]} \mathbf{1}[X_s \in G^\star]$ is simply the number of all reports belonging to $G^\star$ by time t, and that stopping time T is the first time where $\omega_t^{G^\star}$ surpasses the threshold $\theta_t(\alpha)$, i.e., $T \coloneqq \inf_{t \in \mathbb{N}^+} \omega_t^{G^\star} > t\beta\mu_{G^\star}^0 + \frac{1}{2}\sqrt{2.06t \ln\left(|\mathcal{G}|\frac{(2+\log_2(t))^2}{\alpha}\right)}$. For ease of notation, we will denote $C_1 \coloneqq \frac{1}{2}\sqrt{2.06} = 0.718$ within this proof.

For the first claim, it is sufficient to show $\liminf_{t\to\infty} \Pr[T > t] = 0$.[6] Recall that, by our modeling, we can consider $\omega_t^{G^\star}$ to be the sum of t i.i.d. Bernoulli trials with parameter $\mu_{G^\star}$.

[6] For a simple proof of this fact, see the solution to Problem 1.13 in Bertsekas and Tsitsiklis [13].

Applying Hoeffding's inequality to this sum yields for any t that

$$\begin{aligned}\Pr[T > t] &= \Pr\left[\omega_t^{G^\star} < t\beta\mu_{G^\star}^0 + C_1\sqrt{t \cdot \ln\left(|\mathcal{G}|\frac{(2+\log_2(t))^2}{\alpha}\right)}\right] \\ &\leq \exp\left(-2\left(\Delta^2 t - 2\Delta C_1\sqrt{t \cdot \ln\left(|\mathcal{G}|\frac{(2+\log_2(t))^2}{\alpha}\right)}\right)\right).\end{aligned}$$

Note that $\frac{\sqrt{t}\ln(\log_2(t))}{t} \to 0$; it can thus be seen that $\lim_{t\to\infty}\Pr[T > t] = \lim_{t\to\infty}\exp(-t) = 1$.

For the second claim, we apply Hoeffding's inequality again to see that for all t,

$$\Pr\left[\omega_t^{G^\star} \leq \mathbb{E}[\omega_t^{G^\star}] - C_1\sqrt{t\ln\left(\frac{(2+\log_2(t))^2}{\delta}\right)}\right] \leq \Pr\left[\omega_t^{G^\star} \leq \mathbb{E}[\omega_t^{G^\star}] - \sqrt{\frac{t}{2}\ln\left(\frac{1}{\delta}\right)}\right] \leq \delta.$$

Thus, with probability at least $1-\delta$, for all t simultaneously, $\omega_t^{G^\star} > t\mu_{G^\star} - C_1\sqrt{t\ln\left(\frac{(2+\log_2(t))^2}{\delta}\right)}$. The algorithm stops at time t if and only if

$$t\mu_{G^\star} - C_1\sqrt{t\ln\left(\frac{(2+\log_2(t))^2}{\delta}\right)} > t\beta\mu_{G^\star}^0 + C_1\sqrt{t\ln\left(\frac{(2+\log_2(t))^2}{\alpha/|\mathcal{G}|}\right)}.$$

Rearranging, we have

$$\frac{t}{\left(\sqrt{\ln\left(\frac{(2+\log_2(t))^2}{\alpha/|\mathcal{G}|}\right)} + \sqrt{\ln\left(\frac{(2+\log_2(t))^2}{\delta}\right)}\right)^2} \geq \frac{C_1}{\Delta^2}.$$

Note that we can can upper bound the denominator of the left hand side as

$$\left(\sqrt{\ln\left(\frac{(2+\log_2(t))^2}{\alpha/|\mathcal{G}|}\right)} + \sqrt{\ln\left(\frac{(2+\log_2(t))^2}{\delta}\right)}\right)^2 \leq 4\ln\left(\frac{(2+\log_2(t))^2}{\min(\alpha/|\mathcal{G}|, \delta)}\right).$$

Setting $\frac{t}{4\ln\left(\frac{(2+\log_2(t))^2}{\min(\alpha/|\mathcal{G}|,\delta)}\right)} \geq \frac{C_1}{\Delta^2}$ and rearranging gives

$$\frac{t}{1+\ln((2+\log_2(t))^2)} \geq \frac{4C_1\ln(\max(\alpha/|\mathcal{G}|, 1/\delta)}{\Delta^2} \tag{3.6}$$

Thus, with probability $1-\delta$, the algorithm terminates at the smallest t satisfying Equation (3.6). The statement of the theorem follows from separating the two cases for $\delta < \alpha/|\mathcal{G}|$ and $\delta \geq \alpha/|\mathcal{G}|$, and noting that $\widetilde{O}$ notation suppresses the (negligible) log-log factor. □

3.7.2 Proofs for betting-style algorithm

We first prove Theorem 3.6, restated above.

Proof. First note that for any G for which $\mathcal{H}_0^G$ holds, the sequence $\{\exp(\omega_t^G)\}_{t\geq 0}$ is a non-negative super-martingale. The non-negative property follows directly from the quantity being an exponential of a real (albeit possibly negative) number, while the fact that it is a super-martingale follows from the computations below:

$$\begin{aligned}
\mathbb{E}[\exp(\omega_t^G)|\mathcal{F}_t] &= \mathbb{E}[\exp(\omega_{t-1}^G + \ln(1+\lambda_t^G(\mathbf{1}_{X_t\in G} - \beta\mu_G^0)))|\mathcal{F}_t] \\
&= \exp(\omega_{t-1}^G)\cdot(1+\lambda_t^G(\mathbb{E}[\mathbf{1}_{X_t\in G}|\mathcal{F}_t] - \beta\mu_G^0)) \\
&= \exp(\omega_{t-1}^G)\cdot(1+\lambda_t^G(\mu_G - \beta\mu_G^0)) \\
&\leq \exp(\omega_{t-1}^G)\cdot(1+\lambda_t^G(\beta\mu_G^0 - \beta\mu_G^0)) \\
&= \exp(\omega_{t-1}^G),
\end{aligned}$$

where the first equality follows by Eq. 3.4, the second by re-arranging and noting that all quantities except $\mathbf{1}_{X_t\in G}$ are completely determined by $\mathcal{F}_t$[7], and the third by definition (see Section 3.2). Finally, the inequality follows because $\mu_G \leq \beta\mu_G^0$ under $\mathcal{H}_0^G$ and $\lambda_t^G \geq 0$.

Next, for any group G such that $\mathcal{H}_0^G$ holds, we can apply Ville's inequality (Theorem 3.8), plugging in the super-martingale $\{\exp(\omega_t^G)\}_{t\geq 0}$ and taking x to be $\theta_t(\alpha) = \log(|\mathcal{G}|/\alpha)$. This yields the following guarantee:

$$\begin{aligned}
\Pr[\exists t : \omega_t^G > \log(|\mathcal{G}|/\alpha)] &= \Pr[\exists t : \exp(\omega_t^G) > |\mathcal{G}|/\alpha] \\
&\leq \mathbb{E}[\exp(\omega_0^G)]\cdot\alpha/|\mathcal{G}| \\
&= \alpha/|\mathcal{G}|,
\end{aligned}$$

where the final line follows because ω_0^G is initialized as 0 and hence $\exp(\omega_0^G)$ is equal to 1.

Finally, by union bound we get the desired guarantee:

$$\begin{aligned}
\Pr[\exists t : \exists G \in \mathcal{G}^{\text{Flag}} \text{ s.t. } \mathcal{H}_0^G \text{ holds}] &\leq \sum_{G \text{ s.t. } \mathcal{H}_0^G \text{ holds}} \Pr[\exists t : \omega_t^G > \log(|\mathcal{G}|/\alpha)] \\
&\leq |G \text{ s.t. } \mathcal{H}_0^G \text{ holds}|\cdot\alpha/|\mathcal{G}| \\
&\leq \alpha.
\end{aligned}$$

□

Before proving Theorem 3.7, we first state and prove some helper results.

Claim 3.10. *For any $T \geq 4$ and group G, we have that the expected value over the randomness in the realizations of each X_t of ω_T^G defined as per Equations* (3.4) *and* (3.5) *can be lower bounded as*

$$\mathbb{E}[\omega_T^G] \geq \mathbb{E}\left[\max_{\lambda\in[0,1]} \omega_T(\lambda)\right] - 2\ln T.$$

[7]In particular, it is imposed that λ_t^G be 'predictable' which precisely implies that it is fixed given $\mathcal{F}_t$.

where we define $\omega_T^G(\lambda)$ *to be the quantity obtained by applying Equation* (3.4) *with* $\lambda_t^G \coloneqq \lambda$ *for all* $t \in [T]$.

Proof. By the definition of regret we have that $\max_{\lambda\in[0,1]} \omega_T^G(\lambda) - \omega_T^G \leq R_T$. Rearranging and taking expectations, we have

$$\mathbb{E}[\omega_T^G] \geq \mathbb{E}\left[\max_{\lambda\in[0,1]} \omega_T^G(\lambda)\right] - \mathbb{E}[R_T].$$

Next, it can be verified that Equation (3.5) implements the Online Newton Step algorithm for $\ln(1+\lambda_t^G(\mathbf{1}_{X_t\in G} - \beta\mu_G^0))$ (see Appendix C of Cutkosky and Orabona [25]). We therefore have that $R_T \leq \frac{1}{2-\ln(3)}\ln(T+1)$ in general, and $R_T \leq 2\ln(T)$ for $T \geq 4$. The statement of the claim plugging this into the expression above. □

Lemma 3.11. *For each group* G, *taking* $\lambda_G^\star = \mathrm{Proj}_{[0,1]}\left[\dfrac{\mu_G - \beta\mu_G^0}{\beta\mu_G^0(1-\beta\mu_G^0)}\right]$ *maximizes expected log-wealth (at every step* t*). The resulting expected log-wealth at time* T *(had* $\lambda_G^\star$ *been used at every time* t*) is equal to*

$$\mathbb{E}\left[\omega_T^G(\lambda_G^\star)\right] = T \cdot \omega_\star^G$$

where we denote $\omega_\star^G \coloneqq \mathbb{E}[\ln(1+\lambda_G^\star(\mathbf{1}_{X_t\in G} - \beta\mu_G^0)]$ *the expected one-step wealth change under the bet* $\lambda_G^\star$.

Proof. For a fixed λ, the log-wealth at time T is given by

$$\omega_T^G(\lambda) = N_T \ln\left(1+\lambda(1-\beta\mu_G^0)\right) + (T-N_T)\ln\left(1-\lambda\beta\mu_G^0\right),$$

where $N_T = \sum_{t=1}^T \mathbf{1}_{X_t\in G}$. Taking expectations, we have that $\mathbb{E}[N_T] = T\cdot\mu_G$ and therefore

$$\mathbb{E}\left[\omega_T^G(\lambda)\right] = T\cdot\left[\mu_G\ln\left(1+\lambda(1-\beta\mu_G^0)\right) + (1-\mu_G)\ln\left(1-\lambda\beta\mu_G^0\right)\right]. \tag{3.7}$$

To maximize (3.7), we only need to find $\lambda_G^\star \in [0,1]$ that maximizes the expressions in the square brackets. Taking the derivative we see that the function is concave, and, therefore, we can solve for $\lambda_G^\star$ by setting the derivative to 0 and then projecting the resulting value to $[0,1]$. This yields $\lambda_G^\star = \mathrm{Proj}_{[0,1]}\left[\frac{\mu_G-\beta\mu_G^0}{\beta\mu_G^0(1-\beta\mu_G^0)}\right]$. Plugging this back into (3.7) we get

$$\begin{aligned}
\mathbb{E}\left[\omega_T^G(\lambda_G^\star)\right] &= T\cdot\left[\mu_G\ln\left(1+\lambda_G^\star(1-\beta\mu_G^0)\right) + (1-\mu_G)\ln\left(1-\lambda_G^\star\beta\mu_G^0\right)\right] \\
&= T\cdot\mathbb{E}[\ln(1+\lambda_G^\star(\mathbf{1}_{X_t\in G} - \beta\mu_G^0)] \\
&\coloneqq T\cdot\omega_\star^G.
\end{aligned}$$

□

Remark 3.12. *Note that we can explicitly compute*

$$\omega_\star^G = \mu_G\ln\left(1+\tfrac{\Delta_G}{\beta\mu_G^0(1-\mu_G)}\right) + \ln\left(1-\tfrac{\Delta_G}{1-\beta\mu_G^0}\right),$$

where $\Delta_G = \mu_G - \beta\mu_G^0$, *but this quantity is difficult to analyze, and it is not clear that* $\omega_\star^G$ *can be explicitly lower bounded as* $O(\Delta_G)$.

We now prove Theorem 3.7, restated above.

Proof. Let $G^\star \coloneqq \arg\max_G \omega_\star^G$ and denote the corresponding one-step wealth change $\omega_\star = \omega_\star^{G^\star}$. Note that under the alternative this will correspond to a strictly positive quantity and is equivalent to the definition in the theorem statement. We can analyze the likelihood that its null has not been rejected by time t as follows:

$$\begin{aligned}\Pr\left[\omega_t^{G^\star} < \ln(|\mathcal{G}|/\alpha)\right] &= \Pr\left[\omega_t^{G^\star} - \mathbb{E}[\omega_t^{G^\star}] < \ln(|\mathcal{G}|/\alpha) - \mathbb{E}[\omega_t^{G^\star}]\right] \\ &\le \Pr\left[\omega_t^{G^\star} - \mathbb{E}[\omega_t^{G^\star}] < \ln(|\mathcal{G}|/\alpha) - (t\cdot\omega_\star - 2\ln t)\right],\end{aligned}$$

where the inequality follows by Claim 3.10 and Lemma 3.11, and the fact that

$$\mathbb{E}[\max_{\lambda\in[0,1]} \omega_t^{G^\star}(\lambda)] \ge \mathbb{E}[\omega_t^{G^\star}(\lambda_\star^{G^\star})] = t\cdot\omega_\star.$$

Whenever t is large enough such that $\frac{\ln(t)}{t} \le \frac{\omega_\star}{4}$, we have

$$\Pr[\omega_t^{G^\star} < \ln(|\mathcal{G}|/\alpha)] \le \Pr\left[\omega_t^{G^\star} - \mathbb{E}[\omega_t^{G^\star}] < \ln(|\mathcal{G}|/\alpha) - \tfrac{3}{4}(t\cdot\omega_\star)\right]. \tag{3.8}$$

Since $\sqrt{t} \ge \ln t$ for all $t \in \mathbb{N}^\star$, this is satisfied in particular by taking $t \ge \frac{2^4}{\omega_\star^2}$. Further, note that $\ln(|\mathcal{G}|/\alpha) \le \frac{t\cdot\omega_\star}{4}$ whenever $t \ge \frac{2^2\cdot\ln(|\mathcal{G}|/\alpha)}{\omega_\star}$. So, for $t \ge \max\{\frac{2^4}{\omega_\star^2}, \frac{2^2\cdot\ln(|\mathcal{G}|/\alpha)}{\omega_\star}\}$, we have

$$\Pr[\omega_t^{G^\star} < \ln(|\mathcal{G}|/\alpha)] \le \Pr\left[\omega_t^{G^\star} - \mathbb{E}[\omega_t^{G^\star}] < -\tfrac{1}{2}(t\cdot\omega_\star)\right].$$

Now, note that since $\lambda_t^G \in [0,1]$, we have that each $\ln(1+\lambda_t^G(\mathbf{1}_{X_t\in G} - \beta\mu_G^0))$ lies in $[\ln(1-\beta\mu_G^0), \ln(2-\beta\mu_G^0)]$ and is therefore sub-Gaussian with parameter $\sigma = \frac{1}{2}\ln\left(1+\frac{1}{1-\beta\mu_G^0}\right)$; then, Hoeffding's inequality gives

$$\begin{aligned}\Pr\left[\omega_t^{G^\star} - \mathbb{E}[\omega_t^{G^\star}] < -\tfrac{1}{2}(t\cdot\omega_\star)\right] &= \Pr\left[\sum_{i\in[t]} \ln(1+\lambda_t^{G^\star}(\mathbf{1}_{X_t\in G^\star} - \beta\mu_{G^\star}^0)) - \mathbb{E}[\omega_t^{G^\star}] \le -\frac{1}{2}t\cdot\omega_\star\right] \\ &\le \exp\left(-\frac{(\frac{1}{2}t\cdot\omega_\star)^2}{\frac{1}{2}t\ln^2(1+\frac{1}{1-\beta\mu_{G^\star}^0})}\right) \\ &= \exp\left(-\frac{\omega_\star^2}{2\ln^2(1+\frac{1}{1-\beta\mu_{G^\star}^0})}\cdot t\right) \\ &\le \exp\left(-\frac{(1-\beta\mu_{G^\star}^0)^2}{2}\cdot\omega_\star^2\cdot t\right).\end{aligned}$$

where for the last inequality we used $\ln(1+x) \le x$. Now we are ready to analyze the stopping time T of Algorithm 3.

Test of power one. Let E_t be the event that we stop at time t, i.e. $E_t = \{\exists G \text{ such that } \omega_t^G \geq {}^{|\mathcal{G}|}/_{\alpha}\}$. We have that

$$\begin{aligned}
\Pr[T = \infty] &= \Pr\left[\lim_{t\to\infty} \cap_{s\leq t} \neg E_t\right] \\
&= \lim_{t\to\infty} \Pr[\cap_{s\leq t} \neg E_t] \\
&\leq \lim_{t\to\infty} \Pr[\neg E_t] \\
&= \lim_{t\to\infty} \Pr[\forall G,\, \omega_t^G < \ln({}^{|\mathcal{G}|}/_{\alpha})] \\
&\leq \lim_{t\to\infty} \Pr\left[\omega_t^{G^\star} < \ln({}^{|\mathcal{G}|}/_{\alpha})\right] \\
&\leq \lim_{t\to\infty} \exp\left(-\frac{(1-\beta\mu_G^0)^2}{2} \cdot \omega_\star^2 \cdot t\right) \\
&= 0.
\end{aligned}$$

Expected Stopping Time. Since T is a positive integer, we can express the expected stopping time as

$$\begin{aligned}
\mathbb{E}[T] &= \sum_{t=1}^{\infty} \Pr[T > t] \\
&= \sum_{t=1}^{\infty} \Pr[\neg E_1 \wedge \ldots \wedge \neg E_t] \\
&\leq \sum_{t=1}^{\infty} \Pr[\neg E_t] \\
&= \sum_{t=1}^{\infty} \Pr[\forall G,\, \omega_t^G < \ln({}^{|\mathcal{G}|}/_{\alpha})] \\
&\leq \sum_{t=1}^{\infty} \Pr\left[\omega_t^{G^\star} < \ln({}^{|\mathcal{G}|}/_{\alpha})\right] \\
&\leq \max\left\{\frac{2^4}{\omega_\star^2}, \frac{2^2 \cdot \ln({}^{|\mathcal{G}|}/_{\alpha})}{\omega_\star}\right\} + \sum_{t=1}^{\infty} \exp\left(-\frac{(1-\beta\mu_{G^\star}^0)^2}{2} \cdot \omega_\star^2 \cdot t\right) \qquad (3.9) \\
&= \max\left\{\frac{2^4}{\omega_\star^2}, \frac{2^2 \cdot \ln({}^{|\mathcal{G}|}/_{\alpha})}{\omega_\star}\right\} + \frac{1}{\exp\left((1-\beta\mu_{G^\star}^0)^2\omega_\star^2/2\right) - 1} \\
&\leq \max\left\{\frac{2^4}{\omega_\star^2}, \frac{2^2 \cdot \ln({}^{|\mathcal{G}|}/_{\alpha})}{\omega_\star}\right\} + \frac{2}{(1-\beta\mu_{G^\star}^0)^2\omega_\star^2} \qquad (3.10) \\
&\leq \mathcal{O}\left(\frac{1}{\omega_\star^2} + \frac{\ln({}^{|\mathcal{G}|}/_{\alpha})}{\omega_\star}\right)
\end{aligned}$$

where (3.9) follows from the upper bound on $\Pr\left[\omega_t^{G^\star} < \ln(|\mathcal{G}|/\alpha)\right]$ for $t \geq \max\left\{\frac{2^4}{\omega_\star^2}, \frac{2^2 \cdot \ln(|\mathcal{G}|/\alpha)}{\omega_\star}\right\}$ derived in (3.8), and (3.10) follows from $\exp(x) \geq 1 + x$. □

Part II

Individualizing Treatments

Chapter 4

Online Control for Adaptive Tapering of Medications

Tapering medications and assisting cessation of addictive substances are related challenges in health care. Relapse rates are high, and many return to drug use within weeks of entering treatment. Though no general protocols are equally valuable to all medications or all people, a personalized approach to cessation may allow caregivers flexibility to meet the diverse needs of a diverse care-seeking population. This chapter investigates adaptive tapering protocols that enable individuals to self-regulate their cessation rate. In particular, we formalize tapering as an optimization problem of minimizing the cumulative dose subject to constraints on well-being.

Such a formulation requires modeling an individual's dose-response dynamics, and we propose a mathematical formulation of the *opponent process* model of addiction due to Solomon [96]. As we discuss in Section 4.1, an opponent process consists of two competing systemic reactions to treatment. The first system governs the "positive" effect with a rapid onset and fast subsequent decay. The second system governs "negative effects" with a long latency and slow decay. In Section 4.2, we show that modeling opponent processes as LTI systems captures the qualitative behaviors specified in the psychology literature and allows us to analyze tapering as an optimal control problem.

In Section 4.3, we first derive a greedy policy that solves the optimal control problem. The dosage should be decreased by as much as possible while maintaining constraints on well-being, but no planning is needed to compute the dose. We provide a robust controller that maintains this greedy behavior and is optimal under various models of possible exogenous disturbances.

In reality, the opponent process model is an approximation, and the particular model of an individual is unknown. A possible approach informed by control theoretic practice might involve learning a dynamics model of the individual, but such learning procedures require wide varying of inputs and would not be feasible for patients attempting drug cessation. Instead of leaning on system identification, we investigate the potential of "model-free" controllers for tapering in Section 4.4. We analyze the potential of integral control, where the

error signal is the deviation from a minimally acceptable level of well-being. Using methods from the online learning literature, we show that integral control robustly reduces the dosage while minimally violating constraints on well-being. Under further mild assumptions about the time between doses, we show that the integral controller monotonically reduces the dose to zero in finite time without violating the specified constraints. In the numerical experiments of Section 4.5, we demonstrate that our adaptive protocol outperforms non-adaptive methods in terms of both maintenance of well-being and rate of dose reduction.

This chapter is based on joint work with Benjamin Recht, previously released as "Online Control For Adaptive Tapering Of Medications" [35].

4.1 Related Work

Tapering Protocols. Our work attempts to synthesize adaptive tapering protocols that can be applied both for de-prescribing medications and for assisting cessation of addictive substances. In both of these applications, most studies of tapering protocols have focused on non-personalized substance-specific procedures.

For example, for discontinuing SSRI treatment, Horowitz and Taylor [48] claim that standard recommendations for tapering plans are too rapid, and better results are achieved if tapering is performed over multiple months than over weeks. They find that recommendations for linear dose reductions by constant amounts compounds these negative effects. To address this issue, they suggest that SSRI's be tapered via exponentially decreasing regimens, reducing the dose by multiplicative rather than additive factors. They also highlight the importance of individualization for the tapering process, but note that currently there is no established methodology to this end.

Similarly, withdrawal guidelines for benzodiazepines recommend dose reductions that are proportional to the present dose (most commonly 10% reductions) per week, yielding exponentially decreasing regimens, as opposed to linear reductions. Horowitz et al. [47, 49] propose a similar method (very slow, hyperbolic) for tapering of antipsychotic medication. This work is based on case studies and notes that there is no standard guideline for tapering antipsychotic medications.

Several studies have also investigated protocols for managing withdrawal from addictive substances. In a meta-analysis of opiod tapering protocols, Berna et al. [12] found that longer tapers are typically better. The survey by Fenton et al. [32] argues that most tapering protocols which focus primarily on getting through the acute withdrawal phase may be too rapid and have negative mental health consequences. In an observational study, Agnoli et al. [2] find that tapering of opiods is significantly associated with increased risk of overdose and mental health crisis. Henry et al. [46] attempt to characterize patients' subjective experiences with opiod tapering in an effort to minimize negative tapering reactions.

Models of Drug Response and Tolerance. In the present work, we will build upon a popular model of addiction, proposed by Solomon [96], called the Opponent-Process theory of acquired motivation. In this model, the drug response is the result of an initial 'positive' effect with short lag and fast decay (A process) followed by a counter 'negative' effect with high latency and slow decay (B process). This formalizes the empirical qualitative observation that the withdrawal symptoms of a drug are characterized as opposite to its acute effects [56]. The shape of such a response is plotted in Figure 4.1.

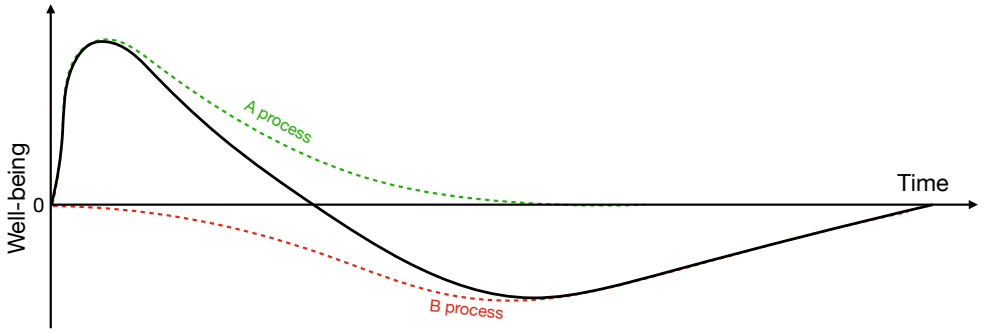


Figure 4.1: Shape of an Instantaneous Opponent Process response as described by Solomon [96].

Koob and Moal [58] and Koob [57] extend the opponent process framework to account for a chronic deviation of the regulatory system from baseline under repeated administration of the drug, which is referred to as allostasis and is illustrated in Figure 4.2. In this work, we propose a simple mathematical formulation of the opponent process that captures the salient aspects of this model.

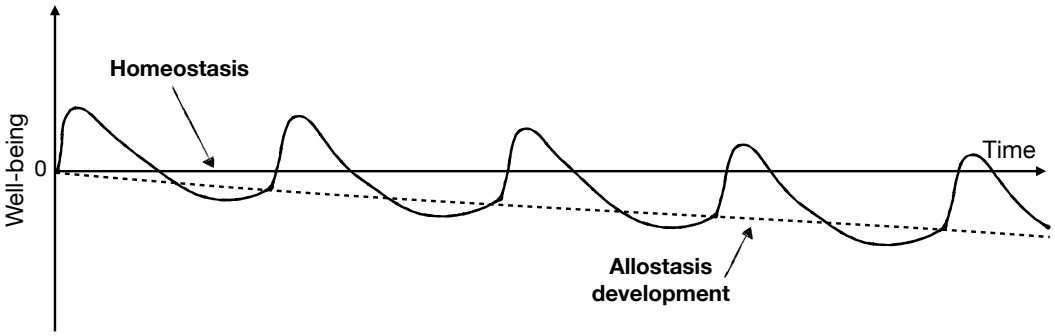


Figure 4.2: Allostasis development due to overlay of opponent processes over time as described by Koob and Moal [58].

4.2 Problem Setting and Preliminaries

To quantify a substance-agnostic drug response, we will denote an individual's *well-being* at any time t as y_t and the *dose* of a drug taken as u_t. This formulation asserts that y_t and u_t are concrete numerical quantities. For u_t this is natural as nearly all prescription/tapering schemes specify dosages in terms of raw active compound weight. For y_t, our methods can be instantiated with a wide array of imprecise quantifications of well-being. For example, for tapering anti-depressant medication, y_t could be a weekly score on the Patient Health Questionnare (PHQ-9) [60]. If we want to instead directly measure withdrawal, we could for example have y_t be the negative of the number of Discontinuation-Emergent Signs and Symptoms (DESS) [87]. [1]

We assume that the well-being y_t, given a history of doses, evolves according to a linear time-invariant (LTI) system

$$y_{t+1} = \sum_{k=0}^{t} g(k)u_{t-k} + y_{t+1}^{\text{nat}} \tag{4.1}$$

where g is the *impulse response* of the drug's effect on well-being, and y^{nat} comprises the *natural progression* of the remaining dynamics that cannot be attributed to the recorded drug intake. We focus on linear responses as they are sufficiently general to capture much of the qualitative behavior observed in the psychology literature, and discuss extensions to nonlinear systems in Section 4.6. Additionally, we allow for arbitrary deviations from the LTI dynamics through y^{nat}: any measurement noise, exogenous disturbance, and/or effects of doses taken before the time tapering protocol is adopted (set by convention to $t = 0$) are implicitly represented through this term.

Broadly, the *goal of tapering* is to minimize the consumption of a drug while keeping well-being above some specified threshold $y_{\min}$. We can directly formalize this as an online optimization problem: over some *tapering horizon* T, we aim to minimize the *cumulative dose* $\sum_{t=0}^{T-1} u_t$, while satisfying the *constraint* that $y_t \geq y_{\min}$ for all $t = \overline{1, T}$.

Modeling Opponent Processes. We now turn to giving a mathematical characterization of an opponent processes within our control-theoretic framework. As a reminder, Solomon [96] asserts that an opponent process reaction to a single dose is first an A-process where $y > 0$ and then a following B-process where $y < 0$. We summarize this property in Definition 4.1 below.

Definition 4.1 (Opponent Process)**.** *We say that g is an opponent process if there exists a time τ_0 such that $g(\tau) > 0$ when $\tau < \tau_0$ and $g(\tau) \leq 0$ if $\tau \geq \tau_0$.*

In this work, we care about opponent processes that both lead to the development of allostasis (i.e., tolerance/addiction) *and* are sufficiently well-behaved so that tapering is

[1] Since this work is framed in terms of maintaining 'well-being,' we need to take the negative of measurements of withdrawal intensity. The results can all be restated in terms of maintaining withdrawal intensity below a desired individual-specified threshold via a simple sign flip.

possible. In Definition 4.2 below, we specify a general class of systems that satisfy these desiderata.

Definition 4.2 (Linearly Progressing Opponent Process)**.** *An opponent process g is a* Linearly Progressing Opponent Process *(LPOP) if there additionally exists an $\alpha \in [0, 1)$ such that*

$$\begin{aligned} g(t+1) &\leq \alpha \cdot g(t) \quad \textit{for} \;\; t < \tau_0 - 1, \\ |g(t+1)| &\geq \alpha \cdot |g(t)| \quad \textit{for} \;\; t \geq \tau_0 \,. \end{aligned}$$

This definition quantifies opponent processes where the A-process decays exponentially, and the B-process does not decay as quickly as the A-process. Qualitatively, Definition 4.2 models a response where the positive effects decay at some rate α, then become negative, trending towards some peak of withdrawal, and finally decay back to 0, potentially at a prolonged rate.

We note that this always happens for opponent processes with $\tau_0 = 1$ since we can then take $\alpha = 0$. This observation is useful in the sequel and we flag it here specifically as a remark.

Remark 4.3. Any opponent process g with $\tau_0 = 1$ is an LPOP.

Example 4.4. Consider a linear system with impulse response

$$g(t) = \sum_{\lambda} c_\lambda \lambda^t$$

where all of the $\lambda \in [0, 1)$ and the c_λ are real valued scalars. The terms where c_λ are positive correspond to effects that increase well-being. The terms where c_λ are negative decrease well-being. With this in mind, define $\Lambda_+ = \{\lambda \;:\; c_\lambda \geq 0\}$ and $\Lambda_- = \{\lambda \;:\; c_\lambda < 0\}$. Then, the system is an LPOP if the following three conditions hold:

1. $\max_{\lambda \in \Lambda_+} \lambda \leq \min_{\lambda' \in \Lambda_-} \lambda'$
2. $\sum_\lambda c_\lambda > 0$
3. There exists some τ_0 s.t. $g(\tau_0) \leq 0$

Intuitively, this holds because the part of the system corresponding to positive effects decay more rapidly than all of the terms corresponding to negative effects. In particular, the second condition implies an initial positive response, and third condition implies that after a single dose, the well-being will eventually be negative.

The formal argument of why such systems are LPOPs proceeds by bounding $g(t+1)$ in terms of $g(t)$. If we let $\Lambda_{++} := \max_{\lambda \in \Lambda_+} \lambda$ and $\Lambda_{--} := \min_{\lambda' \in \Lambda_-} \lambda'$, then we have

$$
\begin{aligned}
g(t+1) = \sum_{\lambda} c_\lambda \lambda^{t+1} &= \sum_{\lambda_+ \in \Lambda_+} c_{\lambda_+} \lambda_+^{t+1} + \sum_{\lambda_- \in \Lambda_-} c_{\lambda_-} \lambda_-^{t+1} \\
&\leq \Lambda_{++} \cdot \sum_{\lambda_+ \in \Lambda_+} c_{\lambda_+} \lambda_+^{t} + \Lambda_{--} \cdot \sum_{\lambda_- \in \Lambda_-} c_{\lambda_-} \lambda_-^{t} \\
&\leq \Lambda_{--} \left(\sum_{\lambda_+ \in \Lambda_+} c_{\lambda_+} \lambda_+^{t} + \sum_{\lambda_- \in \Lambda_-} c_{\lambda_-} \lambda_-^{t} \right) \\
&= \Lambda_{--} g(t)
\end{aligned}
$$

Setting $\alpha \doteq \Lambda_{--}$, the above implies that for $t < \tau_0 - 1$, $g(t+1) \leq \alpha g(t)$ and for $t \geq \tau_0$, $|g(t+1)| \geq \alpha |g(t)|$. Hence g satisfies Definition 4.2.

In Figure 4.3 we plot the dose response and the effect deterioration observed when taking a fixed dose for four examples of 2-pole LTI systems, each capturing an LPOP with different time constants and relative effect magnitudes.

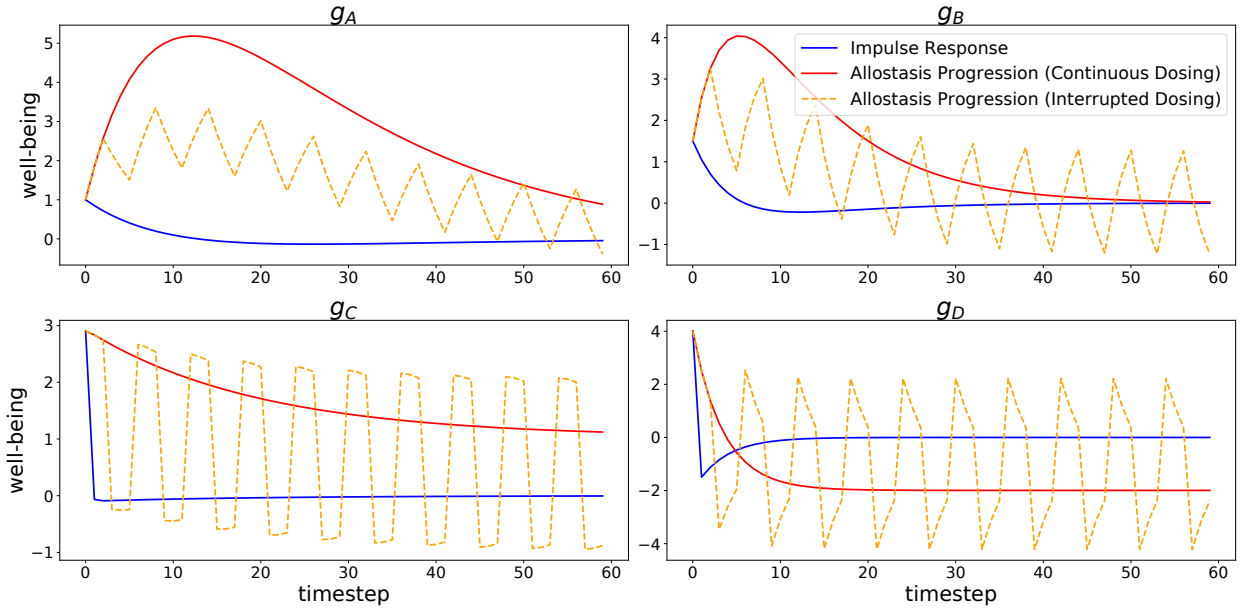


Figure 4.3: Impulse responses and allostasis progression over 60 timesteps for four opponent processes. The top panels illustrate opponent processes with slowly accumulating benefits and tolerance, characteristic of many therapeutic medications. The bottom panels capture opponent processes with immediate transient effects, and either subtle yet slowly accumulating negative effects (left) or a rapid onset of very strong negative effects (right), the latter being more characteristic of substances of abuse. The well-being is measured in arbitrary units which could be scaled to fit a particular metric of interest.

4.3 Optimal Tapering With Full System Knowledge

In this section, we derive the optimal tapering protocol when *all* the quantities described in the previous section are known. In particular, we show that the optimal solution to the tapering control problem formalized above is achieved by taking the dose that induces exact constraint matching, which is equivalent to setpoint matching with setpoint $y_{\min}$. While not directly actionable, this insight provides a path forward towards designing new adaptive tapering protocols with stronger individual guarantees, as we will see in Section 4.4.

We begin with the formal definition of the *minimum effective dose*, the smallest dose that does not lead to an immediate violation of the constraint on well-being. The main theorem of this section states that taking the minimum effective dose at every step ensures perfect constraint satisfaction with minimal cumulative dose over the entire tapering horizon .

Definition 4.5 (Minimum Effective Dose)**.** *We say that a dose u_t is the* minimum effective dose (MED) *at time t if it either leads to $y_{t+1} = y_{\min}$ or is equal to* 0*.*

Theorem 4.6. *For any LPOP g and any natural progression y^{nat}, taking the MED at every time ensures $y_t \geq y_{\min}$ for all t with minimal cumulative dose.*

Proof. Consider a dosing sequence $\{u_t^\star\}_{t=0}^{T-1}$ that maintains $y_t^\star \geq y_{\min}$ for all t with minimal cumulative dose. We will prove the theorem by contradiction: assume there exists a time t_0 when $u_{t_0}^\star$ is not the MED. Then, there must exist $\epsilon > 0$ such that replacing $u_{t_0}^\star$ with $u_{t_0}^\star - \epsilon$ would result in a slightly lower well-being measurement $y'_{t_0+1} < y^\star_{t_0+1}$ that still satisfies the constraint $y'_{t_0+1} \geq y_{\min}$. We can then consider the following modified dosing schedule $\{u'_t\}_{t=0}^{T-1}$:

$$u'_t = \begin{cases} u^\star_{t_0} - \epsilon & t = t_0 \\ u^\star_{t_0+1} + \alpha\epsilon & t = t_0 + 1 \\ u^\star_t & \text{otherwise} \end{cases}$$

where α is the LPOP's separating decay rate constant from Definition 4.2. Let $\{y'_t\}_{t=1}^{T}$ be the well-being sequence resulting from this modified dosing schedule.

Since $u'_t = u^\star_t$ for all $t < t_0$, we have that $y'_t = y^\star_t \geq y_{\min}$ for all $t \leq t_0$. Plugging in the effect of the dose modifications at t_0 and $t_0 + 1$ into the dynamics equation (4.1), we get:

$$\begin{aligned} y'_{t+1} &= y^\star_{t+1} - g(t - t_0)\epsilon + g(t - t_0 - 1)\alpha\epsilon \\ &\geq y^\star_{t+1} \\ &\geq y_{\min} \end{aligned}$$

where the first inequality follows by the definition of an LPOP (Definition 4.2). Since this holds for all $t > t_0$, we have that $y'_t \geq y_{\min}$ for all $t > t_0 + 1$ as well.

Therefore the well-being sequence $\{y'_t\}_{t=1}^{T}$ also satisfies the well-being constraints for all t. However, since $\alpha < 1$, the modified dosing schedule $\{u'_t\}_{t=0}^{T-1}$ that lead to it has strictly

smaller cumulative dose than $\{u_t^\star\}_{t=0}^{T-1}$:

$$\sum_{t=0}^{T-1} u'_t = \sum_{t=0}^{T-1} u_t^\star - \epsilon + \alpha\epsilon < \sum_{t=0}^{T-1} u_t^\star \,.$$

This is a contradiction, completing the proof. □

This theorem shows that long-term planning is not necessary to optimally taper a medication. It always suffices to take the smallest dose that ensures satisfaction of just the very next well-being constraint. Given this insight, the main question becomes how to compute the MED.

Let us first assume that we know the full specification of the impulse response g, and that we are given (or can iteratively construct) a sequence $\{y_t^{\text{nat,lb}}\}_{t=1}^T$ that lower bounds all potential underlying natural progressions y^{nat}, i.e.

$$y_t^{\text{nat}} \geq y_t^{\text{nat,lb}} \text{ for all } t. \tag{*}$$

With this information we can directly derive the MED, and therefore an explicit-form protocol given below in Proposition 4.7.

Proposition 4.7. *For any LPOP g, taking the dose*

$$u_t = \max\left\{0, \frac{y_{\min} - y_{t+1}^{\text{nat,lb}} - \sum_{k=1}^{t} g(k) u_{t-k}}{g(0)}\right\} \tag{4.2}$$

ensures $y_t \geq y_{\min}$ for all t on all natural progressions satisfying ($\star$) with minimal cumulative dose.

Proof. First we show that the doses u_t prescribed by (4.2) maintain $y_t \geq y_{\min}$ for all t under any natural progression satisfying ($\star$). To do so, we start by deriving the MED at an arbitrary time t_0 for a particular instantiation $\{\tilde{y}_t^{\text{nat}}\}_{t=1}^T$. Given any past doses $\{u_t\}_{t=0}^{t_0-1}$, the MED is equal to

$$\tilde{u}_{t_0}^{\text{MED}} = \max\left\{0, \frac{y_{\min} - \tilde{y}_{t_0+1}^{\text{nat}} - \sum_{k=1}^{t_0} g(k) u_{t_0-k}}{g(0)}\right\} \tag{4.3}$$

To see why, let $\tilde{y}_{t_0+1}$ denote the resulting well-being, and observe that the dose is either 0 or satisfies

$$g(0)\, \tilde{u}_{t_0}^{\text{MED}} + \sum_{k=1}^{t_0} g(k)\, u_{t_0-k} + \tilde{y}_{t_0+1}^{\text{nat}} = y_{\min}$$

By the dynamics equation (4.1), the left hand side is precisely equal to $\tilde{y}_{t_0+1}$. This means that if $\tilde{u}_{t_0}^{\text{MED}}$ is not equal to 0 then it results in a next iterate well-being measurement that exactly matches the constraint, which is precisely the definition of the MED.

Because, having fixed past doses, (4.3) is solely determined by $\tilde{y}^{\text{nat}}_{t_0+1}$ and is decreasing with respect to it, we have that $u_{t_0} \geq \tilde{u}^{\text{MED}}_{t_0}$ and therefore $y_{t_0+1} \geq \tilde{y}_{t_0+1} \geq y_{\min}$ for any $\tilde{y}^{\text{nat}}_{t_0} \geq y^{\text{nat,lb}}_{t_0}$. Since this holds independently for all t_0 (irrespective of past doses), we have that (4.2) maintains $y_t \geq y_{\min}$ for all t and any natural progression satisfying $(\star)$, as promised.

To see that it achieves the above with minimal cumulative dose, observe that at every time t the dose u_t prescribed is *exactly* the MED corresponding to setting $\tilde{y}^{\text{nat}}_{t+1}$ to its lowest instantiation ($y^{\text{nat,lb}}_{t+1}$) in (4.3). Therefore, Theorem 4.6 implies that any dosing sequence with a strictly lower cumulative dose will eventually lead to $y_t < y_{\min}$ on the natural progression which exactly matches the lower bound in $(\star)$. This means that the proposed protocol (4.2) ensures $y_t \geq y_{\min}$ for all t on all natural progressions satisfying $(\star)$ with minimal cumulative dose, concluding the proof. □

There are two natural approaches to lower bound y^{nat}_{t+1} at time t in order to compute u_t prescribed by (4.2). First, we may assume that $y^{\text{nat}}_{t+1} \geq y^{\text{nat}}_t$ for all t. This would correspond to assuming that the underlying natural progression is monotonically non-decreasing. In this case, the best lower bound candidate is $y^{\text{nat,lb}}_{t+1} = y^{\text{nat}}_t$ which we can compute via

$$y^{\text{nat}}_t = y_t - \sum_{k=0}^{t-1} g(k) u_{t-k-1} \,. \tag{4.4}$$

Second, we could assume that there exists a constant $L_{\text{nat}} \geq 0$ such that $y^{\text{nat}}_{t+1} \geq y^{\text{nat}}_t - L_{\text{nat}}$ for all t. This would correspond to assuming that the underlying natural progression cannot trend downwards in a single step at a rate faster than L_{nat}. In this case, the best lower bound candidate is $y^{\text{nat,lb}}_{t+1} = y^{\text{nat}}_t - L_{\text{nat}}$, where we again compute y^{nat}_t via (4.4).

Finally, one may wonder if the above principle for optimal tapering extends to more general settings. A simple yet important scenario is when the constraints are time-varying. This is relevant when a person can tolerate more intense withdrawal effects at particular times of their lives or when they can no longer tolerate an initially agreed-upon baseline. Remark 4.8 shows that optimal tapering is still possible with a time-varying lower bound and becomes equivalent to constraint tracking in this case.

Remark 4.8. The proof of Theorem 4.6 is valid even if the well-being constraints are time-varying. Therefore, optimal tapering is again achieved by taking the MED at every time t with respect to the very next well-being constraint $y^{(t+1)}_{\min}$. Furthermore, the protocol from Proposition 4.7 can be directly extended to this setting by replacing $y_{\min}$ in (4.2) with its time-varying counterpart $y^{(t)}_{\min}$.

The above remark is still true even if the dose responses g are time-varying. Beyond being able to encompass non-linear affine systems, time-varying opponent processes cover situations when there is some process other than dependency-induced allostasis altering the dose-response over time. Going even further, in Section 4.6 we consider non-linear opponent

processes and show that the main results of this section still hold under an appropriate generalization of Definition 4.2. Finally, note that by Remark 4.3 (and its extension to non-linear systems in Section 4.6), *any* opponent process discretized to have $\tau_0 = 1$ satisfies the well-behavedness assumption and therefore can be optimally tapered by taking the MED at every time. The MED is optimal under surprising generality.

4.4 Tapering By Integral Control

We saw in the previous section that optimal tapering is achieved by always taking the smallest dose that satisfies the next well-being constraint. However, computing this dose required knowledge of the model of the underlying dynamics g, as well as a way to bound the next iterate in the underlying natural progression y_{t+1}^{nat}. In practice, we would like to avoid estimating these two quantities and instead derive an update rule requiring minimal knowledge.

Since we know that optimal tapering is achieved by constraint matching (up to dose positivity) and that integral control can match setpoints asymptotically, in this section we propose and analyze a simple integral controller that takes takes the current well-being y_t as input and uses its deviation from the constraint $y_{\min}$ to iteratively compute dose corrections.

We demonstrate that this simple approach can reduce a dose while rarely violating the constraints when the gains are chosen appropriately. In fact, one needs to only know a rough range of the immediate effect of a single dose, which is equal to the value $g(0)$. Moreover, when the dynamics are discretized coarsely enough, we show that the resulting doses monotonically decrease to zero in finite time while ensuring *perfect* constraint satisfaction (i.e. $y_t \geq y_{\min}$ for all t).

To proceed, let $(x)_+ := \max(x, 0)$ and $(x)_- := \min(x, 0)$ Consider the integral control law

$$u_t = \max\left\{0, u_{t-1} - K_+(y_t - y_{\min})_+ - K_-(y_t - y_{\min})_-\right\}. \tag{4.5}$$

There are two differences between the above protocol and a canonical integral controller. First, the control action is clipped at 0 as it is impossible to take a negative dose of a medication. Second, it allows for two gains: K_+ for when y_t is above the desired "setpoint" $y_{\min}$, and K_- for when it falls below. Using two gains allows for more conservative dose corrections when $K_+ < K_-$, leading to smaller decrements when $y_t > y_{\min}$ and bigger increments when $y_t < y_{\min}$.

The main theorem of this section bounds the *long-term constraint violation* [73, 74] of the proposed integral controller. The bound we derive ensures that the running average of the well-being observations y_t up until a time T is greater than $y_{\min}$ minus a penalty decaying at rate $\frac{1}{T}$. This means that the overall well-being satisfies the constraint more strictly as time increases.

Theorem 4.9. *For any underlying y^{nat}, doses prescribed by (4.5) ensure that for any T*

$$\frac{\sum_{t=1}^{T} y_t}{T} \geq y_{\min} - \frac{y_0 - y_{\min}}{T},$$

provided $K_+ \leq g(0)^{-1} \leq K_-$.

The proof of this theorem is broadly inspired from analyses of the finite time behavior of average constraint violation in regret minimization by Mannor and Tsitsiklis [74] and Mahdavi et al. [73]. To prove the theorem, we need the following technical lemma that lower bounds the instance-wise deviation of the y_t obtained from the dosing scheme (4.5) from $y_{\min}$ if the gains K_+/K_- upper/lower bound $g(0)^{-1}$ respectively.

Lemma 4.10. *For any underlying y^{nat}, and any gains satisfying $K_+ \leq g(0)^{-1} \leq K_-$, setting u_t according to (4.5) for $t \geq 0$ yields y_{t+1} satisfying*

$$y_{t+1} \geq y_{\min} + (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) - \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k}).$$

where we adopt the convention that $u_{-1} = 0$.

Proof. We first derive an alternate expression of the (possibly negative) dose $u_t^\star$ that leads to $y_{\min}$. Concretely, using the same derivation in Propositon 4.7 of the maximally greedy protocol (4.3), we have that

$$u_t^\star = \frac{y_{\min} - y_{t+1}^{\text{nat}} - \sum_{k=1}^{t} g(k) u_{t-k}}{g(0)}$$

We aim to express this in terms of the current set-point error $y_t - y_{\min}$ and u_{t-1}, so we add and subtract y_t^{nat} and $\sum_{k=1}^{t-1} u_{t-k-1}$ (which generate $y_t - g(0)u_{t-1}$) to obtain:

$$\begin{aligned} u_t^\star = g(0)^{-1}\big(&y_{\min} - y_t^{\text{nat}} + (y_t^{\text{nat}} - y_{t+1}^{\text{nat}}) \\ &- \sum_{k=1}^{t-1} g(k) u_{t-k-1} - \sum_{k=1}^{t} g(k)(u_{t-k} - u_{t-k-1})\big) \end{aligned}$$

where we take $u_{-1} = 0$. Since $y_t^{\text{nat}} + \sum_{k=1}^{t-1} g(k) u_{t-k-1} = y_t - g(0)u_{t-1}$, we can rewrite the above as

$$\begin{aligned} u_t^\star = u_{t-1} &- g(0)^{-1}(y_t - y_{\min}) - g(0)^{-1}(y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) \\ &+ g(0)^{-1}\left(\sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k})\right) \end{aligned}$$

Since $K_+ \leq g(0)^{-1}$ and $K_- \geq g(0)^{-1}$, we have that

$$-g(0)^{-1}(y_t - y_{\min}) \leq -K_+(y_t - y_{\min})_+ - K_-(y_t - y_{\min})_-$$

And therefore, if u_t is given by (4.5), then

$$\begin{aligned} u_t^\star \leq u_t - g(0)^{-1}(y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) \\ + g(0)^{-1}\left(\sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k})\right) \end{aligned}$$

Multiplying both sides by $g(0)$ and then adding $\sum_{k=1}^{t} g(k)u_{t-k} + y_{t+1}^{\text{nat}}$, we have

$$\begin{aligned} & y_{t+1}^{\text{nat}} + g(0)u_t^\star + \sum_{k=1}^{t} g(k)u_{t-k} \\ \leq & y_{t+1}^{\text{nat}} + \sum_{k=0}^{t} g(k)u_{t-k} - (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) \\ & + \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k}) \end{aligned}$$

Since by construction the LHS is equal to exactly $y_{\min}$ and by the dynamics equation $y_{t+1}^{\text{nat}} + \sum_{k=0}^{t} g(k)u_{t-k} = y_t$, rearranging the above yields:

$$y_{\min} + (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) - \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k}) \leq y_t$$

which is precisely the stated inequality. □

With this lemma, we can now prove Theorem 4.9.

Proof of Theorem 4.9. Using Lemma 4.10, we have that

$$\begin{aligned} \sum_{t=0}^{T} y_{t+1} &\geq (T+1)y_{\min} + \sum_{t=0}^{T}\left((y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) - \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k})\right) \\ &= (T+1)y_{\min} - y_0^{\text{nat}} + y_{T+1}^{\text{nat}} + \sum_{t=0}^{T} u_t g(T-t) \\ &= (T+1)y_{\min} + y_{T+1} - y_0 \end{aligned}$$

where the first equality follows by telescoping, and the second from the implicit convention in (4.1) that $y_0^{\text{nat}} = y_0$. By cancelling y_{T+1} on both sides, we get

$$\sum_{t=0}^{T-1} y_{t+1} \geq T y_{\min} - (y_0 - y_{\min})$$

which can be immediately rearranged into the stated bound:

$$\frac{\sum_{t=1}^{T} y_t}{T} = \frac{\sum_{t=0}^{T-1} y_{t+1}}{T} \geq y_{\min} - \frac{y_0 - y_{\min}}{T}$$

□

Remark 4.11. If it is undesirable to fluctuate too much below $y_{\min}$, we can run a padded integral controller by adding δ to $y_{\min}$. Theorem 4.9 then implies

$$\frac{\sum_{t=1}^{T} y_t}{T} \geq y_{\min} + \delta - \frac{y_0 - y_{\min} - \delta}{T}\,.$$

Remark 4.12. As in the previous section, we can extend the integral controller (4.5) to time-varying constraints by plugging in $y_{\min}^{(t)}$. In this case, the guarantee from Theorem 4.9 would instead bound deviation from the average $y_{\min}$, i.e. from $T^{-1}\sum_{t=1}^{T} y_{\min}^{(t)}$.

We end the technical component of this section by deriving stronger guarantees for opponent processes with $\tau_0 = 1$, a subset of the LPOP class considered thus far. Requiring $\tau_0 = 1$ can be interpreted as an assumption on the discretization of the dose updating scheme (e.g., every day vs. every week). For example, we can turn a system g with $\tau_0 > 1$ into a system g' with $\tau_0' = 1$ by taking $g'(t) = \tau_0^{-1} \cdot \sum_{t'=t\cdot\tau_0}^{(t+1)\cdot\tau_0 - 1} g(t')$.

Our final result shows that for such suitably discretized opponent processes, when the underlying natural progression does not trend down too fast, we can guarantee that the proposed integral controller ensures *perfect* constraint satisfaction, as well as monotonically non-increasing doses that reach 0 in finite time.

Proposition 4.13. *For any g with $\tau_0 = 1$, any initial u_0 which ensures $y_1 \geq y_{\min}$, any $\delta > 0$, and any natural progression sequence satisfying $y_{t+1}^{\text{nat}} \geq y_t^{\text{nat}} - g(t)u_0 + \delta/t$, doses u_t prescribed by (4.5): (i) are non-increasing, (ii) maintain $y_t \geq y_{\min}$ for all t, and (iii) are equal to zero after some finite time T_0.*

Proof. To prove (i) and (ii) jointly, we will proceed by induction on t. The base case holds by the assumption that we have access to an initial u_0 for which $y_1 \geq y_{\min}$. Assume for the inductive hypothesis that $0 \leq u_{t-1} \leq \ldots \leq u_0$ and $y_t \geq y_{\min}$. First, observe that since $y_t - y_{\min} \geq 0$, we have

$$u_t \leq \max\{0, u_{t-1}\} = u_{t-1}$$

showing the inductive step for (i). To show the inductive step for (ii), note that by Lemma 4.10,

$$\begin{aligned} y_{t+1} &\geq y_{\min} + (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) - \sum_{k=1}^{t} g(k)(u_{t-k-1} - u_{t-k}) \\ &\geq y_{\min} + (y_{t+1}^{\text{nat}} - y_t^{\text{nat}}) + g(t)u_0 \\ &\geq y_{\min} + \delta/t \end{aligned}$$

where the second inequality follows since $g(k) < 0$ and $u_{t-k-1} \geq u_{t-k}$ for $k = \overline{1, t-1}$, while the last inequality follows from our assumption that $y_{t+1}^{\text{nat}} \geq y_t^{\text{nat}} - g(t)u_0 + \delta/t$.

Therefore, by induction, (i) and (ii) are true. Finally, by the calculation above, we know that u_t is decreasing and, further, that it can be expressed as

$$u_t = \max\left\{0, u_0 - K_+\delta \cdot \sum_{k=1}^{t} \frac{1}{t}\right\}$$

which, since $\sum_{k=1}^{t} 1/t \geq \ln t$, implies $u_t = 0$ for all $t \geq \exp\left\{\frac{u_0}{K_+\delta}\right\}$, proving (iii). □

Finally, what would actually be needed to implement this protocol? A patient well versed in subjective symptom scoring would give a minimally tolerable value for the day. They would compare this value to how well they felt yesterday. Based on this difference, their new dose would simply be given by Equation (4.5). Such a scheme could be easily implemented in a spreadsheet.

As we will show in the experiments below, the protocol is not sensitive to the choice of $K_\pm$. A rule of thumb value for setting $K_\pm$ could be devised based on the smallest dose change that yields a reasonably perceptible change in y. So, for instance, if the smallest noticeable dosage change would be 5mg and it would lead to a change in y by somewhere between 1 and 2 'units', then K_+ should be roughly 2.5 and K_- should be 5.

4.5 Numerical Simulations

In this section, we conduct numerical simulations to gain insights into the performance of the optimal and the integral-based protocols compared to that of standard (non-adaptive) tapering approaches when deployed on a population with different well-being constraints. Due to a lack of available simulators and open source data for our problem, we rely on synthetic impulse response functions, modelled to induce realistic and varied impulse responses and allostatic adaptations. We describe the experiment setup in detail below, and plot the results in Figure 4.4.

Task specification. We study the problem of tapering in the four opponent processes from Figure 4.3, starting from the allostatic state reached after taking dose $u = 1$ for $T_{\text{init}} = 60$ timesteps. For each setting, we consider tapering a population of N units with uniformly distributed constraints $y_{\min}$, over some time frame T_{taper}. Concretely, we take

$$y_{\min}^A \sim \text{Unif}(-1.5, 0.5),\ y_{\min}^B \sim \text{Unif}(-2, 0),\ y_{\min}^C \sim \text{Unif}(-1, 1),\ y_{\min}^D \sim \text{Unif}(-4.25, -2.25)$$

$$\text{and}\ \ T_{\text{taper}}^A = 180,\ T_{\text{taper}}^B = 120,\ T_{\text{taper}}^C = 90,\ T_{\text{taper}}^D = 15.$$

For a given individual and a given tapering protocol, we measure the average cumulative dose taken and the average cumulative constraint violation[2] over the tapering horizon T_{taper}. We then average these two metrics over the N units to obtain the average performance of a protocol over a population. Finally, for all setups, we also perturb y at every time with $\text{Unif}(-0.25, 0.25)$ noise in order to capture random deviations in an individual's day-to-day well-being.

Specification of tapering protocols. As baselines, we consider the linear ($u_t = u_0 - \alpha t$) and the exponential ($u_t = \alpha^t u_0$) dose decay protocols. To evaluate the protocol designs rather than a particular substance-dependent instantiation, we compute the population performance in terms of the above two metrics over a range of decay rates α. Doing so also gives us a comprehensive trade-off curve between constraint violation and dosage reduction for different rates.

To evaluate the integral controller (4.5), we need to specify the patient constraint $y_{\min}$ and the gains $K_{-/+}$. The constraint can directly be provided by the patient so we feed it directly to the protocol. Since our method can be instantiated with a padded $y^{\delta}_{\min} = y_{\min} + \delta$ to provide stronger constraint violation guarantees (Remark 4.11), we create a trade-off curve by sweeping over different settings of δ. To select the gains, we assume access to a coarse range of the instantaneous response $g(0)$. In the experiment below we specify particular settings for the sake of clarity, and provide an ablation over a multitude gain specifications a bit further down. Concretely, we take the lower and upper bounds to be 50% and 150% of the true response, which corresponds to the following conservative gain settings:

$$K_+ = (2/3) \cdot g(0)^{-1} \text{ and } K_- = 2 \cdot g(0)^{-1}.$$

Finally, we compare to the optimal tapering protocol, for which we use the full model g and the true underlying natural progression to compute the MED via (4.3).

[2]Note that this differs from the long-term constraint violation analyzed in Theorem 4.9. Here we measure the average over constraint violations $(y_{\min} - y)_+$, rather than the difference between the average well-being and $y_{\min}$.

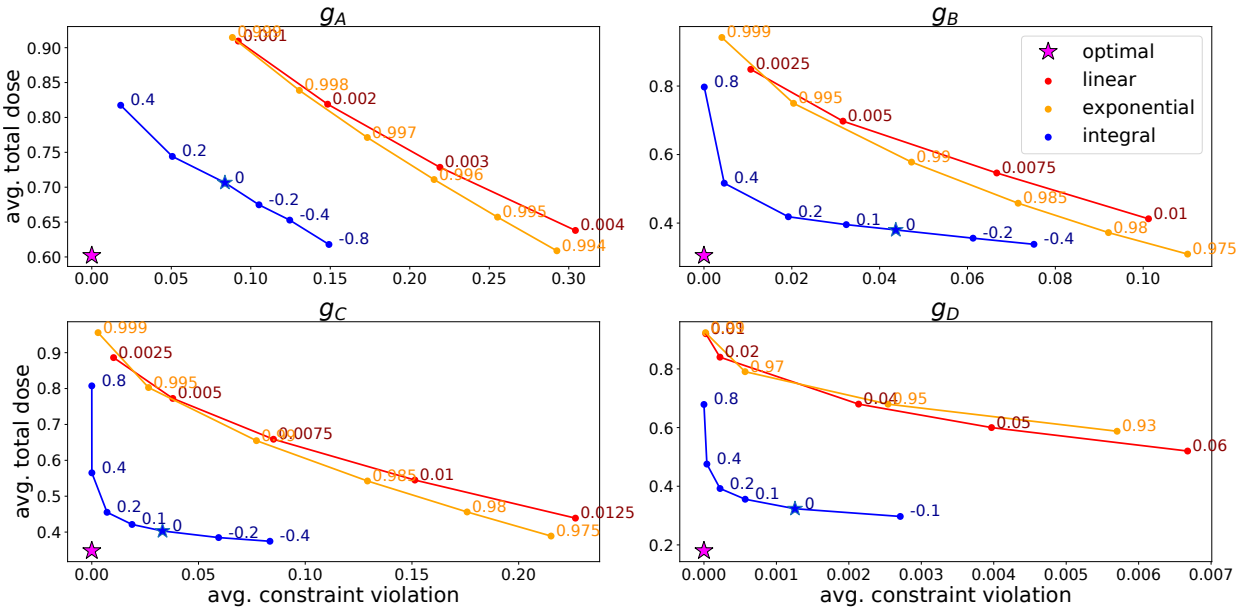


Figure 4.4: Average cumulative constraint violation (x-axis) against average cumulative dose (y-axis). Each point is averaged over $N = 100$ units.

Results. In Figure 4.4, we see that the integral control protocol dominates the baselines considered, as the curves obtained by sweeping over the various parameters do not intersect. Further, the above results corroborate the prior findings discussed in Section 4.1 that very slow rates are necessary to maintain acceptable levels of withdrawal management over a population when restricted to the existing non-adaptive approaches. Finally, there is a non-negligible performance gap between the proposed integral tapering protocol and the optimal protocol (achieved by MED dosing) which opens an avenue for future tapering research.

We conduct further experiments covering the robustness of the integral controller to even more extreme misspecification of the immediate response $g(0)$, as well as performance in terms of an additional metric of potential interest: the percentage of units that successfully taper within the specified time frame. We also invite the reader to explore different settings for the systems and/or protocols by accessing our code here.

Ablation over $g(0)$ range. In Figure 4.5, we plot the behavior of the integral protocol for the gains induced by different $g(0)$ lower/upper bounds against that of the optimal protocol and the non-adaptive baselines.[3] We see that the trends in Figure 4.4 continue to hold over wider misspecification on the range of $g(0)$. In fact, the experiments suggest an overly conservative upper bound on $g(0)$ (leading to a smaller K_-) may improve performance

[3]For the sake of readability we only plot the better baseline protocol for each opponent process (in black), i.e. exponential for $g_A/g_B/g_C$ and linear for g_D.

in certain scenarios, and that setting $K_- = K_+$ may sometimes lead to higher constraint violation in the absence of padding. Overall, we observe that the protocol can be instantiated with surprisingly rough knowledge of the immediate effect on well-being of a medication without suffering significantly in terms of performance.

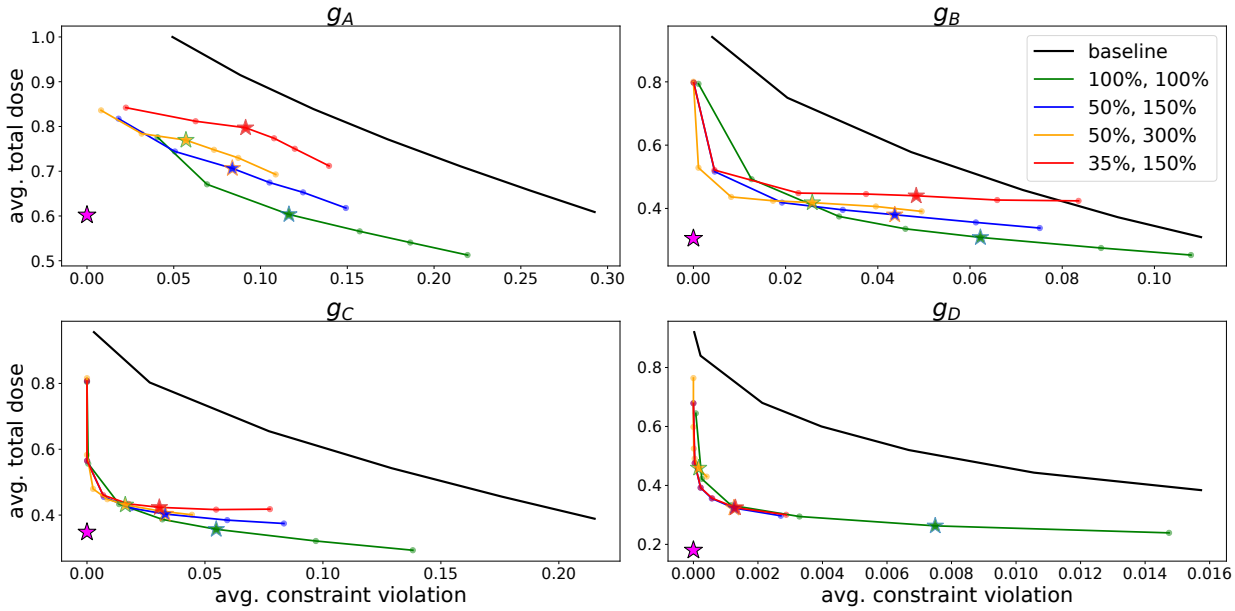


Figure 4.5: Average cumulative constraint violation (x-axis) against average cumulative dose (y-axis) for different ranges of lower and upper bounds on $g(0)$. A label of (p_1, p_2) in the legend corresponds to setting K_- to $(p_1 g(0))^{-1}$ and K_+ to $(p_2 g(0))^{-1}$.

Fraction of subjects fully tapered. In Figure 4.6, we plot the fraction of units that can be fully tapered within T_{taper} steps using the integral protocol (instantiated with the gains and padding specifications from Section 4.5). We see that our method fully tapers a significant portion of the population within some pre-specified timeline with minimal constraint violation. This is in contrast to the linear and exponential protocols which induce a 0% vs. 100% tapering success step function given a time frame T_{taper}. Across all settings considered, the baseline linear/exponential protocols would have to incur significantly larger average constraint violation (*well* beyond the range covered by our plot) to fully taper individuals within the desired time frame. Finally, we again see a gap between the tapering success of the optimal and the integral tapering schemes, and that one way to address this gap is by using negative padding to allow more constraint violation (that however is still significantly smaller than that required for full tapering via non-adaptive methods).

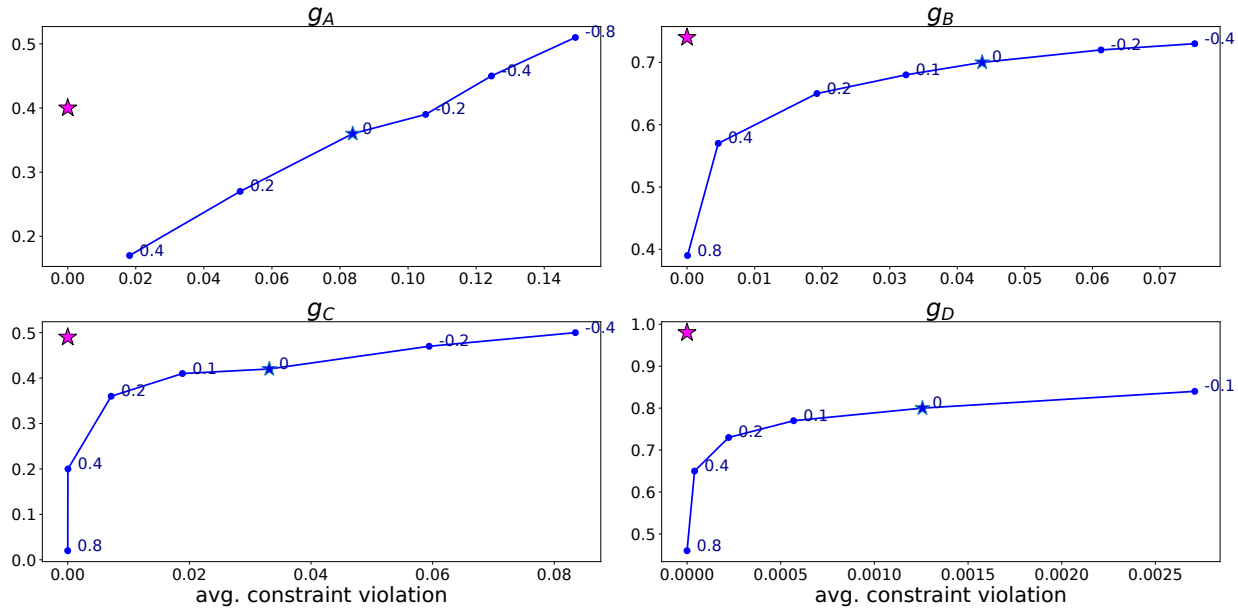


Figure 4.6: Average constraint violation (x-axis) against the fraction of units fully tapered (y-axis) for the integral (blue) and optimal (⋆) protocols.

4.6 Generalization to non-linear opponent processes

Our definition of (4.1) implies a linear contribution of the dose to the observed well-being metric. One may wonder if we can accommodate the case where the opponent process varies with u, i.e. y_t evolves according to the following generalized version of (4.1):

$$y_{t+1} = \sum_{k=0}^{t} g(k, u_{t-k}) + y_{t+1}^{\text{nat}} \tag{4.6}$$

In this subsection, we show we can, under a corresponding generalization of well-behavedness from Definition 4.2. The rate of change of each $g(t, \cdot)$ with respect to the dose u will play an important role, so we introduce the following two shorthand notations:

$$\partial_u^-(t) \doteq \inf_u \partial_u g(t, u) \qquad \partial_u^+(t) \doteq \sup_u \partial_u g(t, u)$$

Definition 4.14 (Generalized Opponent Process)**.** *We say that g is a generalized opponent process if there exists a time τ_0 such that:*

1. $\inf_u g(\tau, u) \geq 0$ *and* $\partial_u^-(t) \geq 0$ *when* $\tau < \tau_0$,
2. $\sup_u g(\tau, u) \leq 0$ *and* $\partial_u^+(t) \leq 0$ *when* $\tau \geq \tau_0$.

Definition 4.15 (Generalized LPOP)**.** *A generalized opponent process g is a generalized linearly progressing opponent process (G-LPOP) if there exists an $\alpha \in (0,1)$ such that:*

1. $\partial_u^+(t+1) \leq \alpha \cdot \partial_u^-(t)$ *when* $t < \tau_0 - 1$,
2. $\partial_u^-(t+1) \geq \alpha \cdot \partial_u^+(t)$ *when* $t \geq \tau_0$.

In this case the proof of Theorem 4.6 changes a bit, relying on Mean Value Theorem. We give the statement and proof in Theorem 4.16 below.

Theorem 4.16. *For any G-LPOP g, taking the maximally greedy dose at every time maintains $y_t \geq y_{\min}$ for all t with minimal cumulative dose.*

Proof. Consider a dosing sequence $u_t^\star$ that maintains $y_t^\star \geq y_{\min}$ for all t with minimal cumulative dose. Assume it is not maximally greedy (otherwise we would be done). This means that there exists t_0 and $\epsilon > 0$ such that $u_{t_0}^\star - \epsilon$ would produce a new $y'_{t_0+1} \geq y_{\min}$. Consider the modified dosing schedule

$$u'_t = \begin{cases} u_{t_0}^\star - \epsilon, t = t_0 \\ u_{t_0+1}^\star + \alpha\epsilon, t = t_0 + 1 \\ u_t^\star \text{ otherwise} \end{cases}$$

and let y'_t be the resulting observations. Since $u'_t = u_t^\star$ for $t < t_0$, we have that $y'_t = y_t^\star \geq y_{\min}$ up to t_0. Plugging in the effect of the dose modifications at t_0 and $t_0 + 1$ into the generalized dynamics equation (4.6), by the Mean Value Theorem, we get:

$$\begin{aligned} y'_{t+1} &= y_{t+1}^\star - g(t - t_0, u_{t_0}^\star) + g(t - t_0, u_{t_0}^\star - \epsilon) \\ &\quad - g(t - t_0 - 1, u_{t_0+1}^\star) + g(t - t_0 - 1, u_{t_0+1}^\star + \alpha\epsilon) \\ &= y_{t+1}^\star - \epsilon \cdot \partial_u g(t - t_0, u_{t_0}^\star - \delta_0\epsilon) \\ &\quad + \alpha\epsilon \cdot \partial_u g(t - t_0 - 1, u_{t_0+1}^\star + \delta_1\epsilon) \end{aligned}$$

for some $\delta_0 \in (0,1), \delta_1 \in (0,\alpha)$. Since g is a G-LPOP, this final expression is greater than or equal to $y_{t+1}^\star$ and hence is also greater than or equal to $y_{\min}$. Since this holds for all $t > t_0$, we have that y'_t satisfies the constraints for all t. To finalize, note that y'_t is achieved with smaller cumulative dose since

$$\sum_{t=1}^{T} u'_t = \sum_{t=1}^{T} u_t^\star - \epsilon + \alpha\epsilon < \sum_{t=1}^{T} u_t^\star,$$

This is a contradiction, completing the proof. □

Instantiating this in (4.2) can be done by solving for the dose which exactly matches the setpoint:

Proposition 4.17. *For any G-LPOP g, assume we know a lower bound $y_t^{\text{nat,lb}}$ such that $y_t^{\text{nat,lb}} \leq y_t^{\text{nat}}$ for all t. Then, for any instance of y_t^{nat} that is greater than the prescribed lower bound, the dosing sequence which takes u_t be the solution to*

$$g(0,u) = \max\left\{0, y_{\min} - y_{t+1}^{\text{nat,lb}} - \sum_{k=1}^{t} g(k, u_{t-k})\right\} \tag{4.7}$$

maintains $y_t \geq y_{\min}$ for all t with minimal cumulative dose.

Since $g(0,u)$ is increasing in u, we have the following guarantee:

Remark 4.18. *An ϵ-accurate solution to (4.7) can be computed in $\mathcal{O}(\log \frac{1}{\epsilon})$ steps.*

The above effectively shows that optimal tapering is equivalent to setpoint matching under non-linear g as well. While it is not as clean to derive precise guarantees for the integral controller as in the main text, the result of Theorem 4.16, combined with the robustness to $g(0)$ errors observed experimentally, strongly motivates the same approach in this more general setting.

Chapter 5

A Practical Control-Theoretic Framework for Taxonomizing Equipoise in Chemotherapy Dosing

Despite advances in targeted therapy and immunotherapy, chemotherapy remains the mainstay of oncological care [55]. These medications are increasingly being used in complex, multi-drug protocols, combining three or four overlapping antineoplastic agents [81, 23, 71], which can lead to increased rates of toxicity and thus require treatment modifications. These heuristic protocol adjustments often come with no strong guidance on the best way to maintain efficacy and minimize toxicity given the many variables of drugs, dose, and time. Furthermore, toxicity response is heterogeneous, depending largely on the individual's biology and tolerance.

All of these factors contribute to a complex landscape of clinical equipoise. Traditionally, clinical equipoise is defined as *a state of genuine uncertainty in the medical community over the optimal treatment course*, and it is used as justification for randomized control trials [33]. When combination chemotherapy regimens are administered outside of clinical trials, clinicians sequentially respond to adverse symptoms from previous treatments. Because these responses are so combinatorial in how agents interact and clinicians respond, they cannot be feasibly reflected in trials. Even in a retrospective setting, predictive models are unable to learn fine-grained impulse responses because the co-administration of several drugs makes the problem generally unidentifiable. Given these challenges, is it possible to formulate an optimization problem which will yield insights into how clinicians should be responding? Thus far the existing literature towards this end has been highly idealized and formal, in the sense that most theoretical models rely on quantities that in practice are difficult if not impossible to observe. This considerably limits the practical insights one can thus garner from such work.

We propose a control-theoretic framework for guiding dose modifications in combination therapy based entirely on *observable* quantities which are routinely collected and intuitive to clinicians. Moreover, instead of attempting to identify a single 'optimal' policy, we use this

framework to understand the landscape of optimality and identify and interpret patterns of clinical behavior. We identify three key impacts of our framework:

1. Create an increased formalism regarding what the underlying optimization problem is for clinicians and provide heuristic clinical takeaways for improving dosing regimens without prescribing specific actions to clinicians;

2. Better understand when variability can be explained by genuine clinical equipoise, and when this clinical equipoise can be resolved;

3. Identify prediction tasks which are both applicable to the clinical workflow and targeted to weaknesses in clinician decision-making.

While we present our framework as a maximization problem specific to the chemotherapy setting, our framework can generalize to any setting where the objective is to minimize or maximize dose subject to certain constraints, as in Gradu and Recht [35]. While the intuition behind these frameworks is similar, there are key differences which make maximization unique. Namely, in dose minimization problems, the initial dose is selected with no expectation that this dose is particularly efficacious over a lower (or, in some cases, higher) dose. On the other hand, in the chemotherapy objective, the initial dose administered to patients will often be both the maximum dose feasibly administered and the most efficacious dose if tolerable. All dose reductions sacrifice some level of efficacy for the ultimate goal of tumor reduction. Moreover, in the minimization case, we often are able to model as an infinite horizon and explore dose responses while gradually converging on the optimal dose for the patient. For example, in the case of psychiatric medications, the consequence for administering too high a dose is extending the time to identifying the optimal dosing protocol, and the consequence for administering too low a dose is either lowered efficacy or withdrawal symptoms. In the chemotherapy maximization case, there is often a finite horizon and administration of the treatment at the current time step may be enabling future dosing by extending patient life expectancy. Additionally, there are dose modifications from which the patient may never recover, like administering so much of a drug that a toxicity effect is saturated and a patient is unable to continue treatment.

This chapter is based on joint work with Ahmed Alaa, Jenna Fields, Benjamin Recht and Travis Zack, currently in preparation to be eventually published under the same name.

5.1 Proposed Framework

The principal aim of our framework is to fit within current practice. To address this, we propose a control-theoretic framework which:

1. Is based on routinely-collected, observable quantities;

2. Is intuitive to clinical decision makers, i.e. follows the logic of medicine;

3. Does not require deviation from standard-of-care (SOC) to such an extent that it is not implementable on a wide-scale.

This requires that our objective and constraints match that of the clinician.[1] Consulting clinicians and relevant literature, the goal of evidence-based practice is to deliver regimens which adhere as closely as possible to the protocol which has been validated by RCTs. Because there is no defined formulation of how to respond to toxicities in a way that minimally deviates from the RCT-proven regimen, clinicians have varying strategies and beliefs. We propose an optimization formulation which concretizes this desideratum.

The core idea is to define our objective to be administering a feasible dose at each cycle so as to maximize cumulative chemotherapy dosage. We define a feasible dose as a dose which (1) does not exceed the standard protocol and (2) does not, combined with all prior doses, lead to an excessive toxicity response. Notice that this *directly* translates the goal of adhering as closely as possible to the RCT-validated protocol into mathematical language. We dedicate the remainder of this section to formalizing this framework.

We encompass the fractional dose combination and toxicities at cycle t by as a dose vector $\mathbf{u}_t \in [0,1]^{d_u}$ and, respectively, a toxicity vector $\mathbf{y}_t \in \mathbb{R}^{d_y}$, where d_u is the number of drugs in the regimen and d_y is the number of toxicities tracked. Note that some entry $\mathbf{u}_t(i)$ taking some value $p \in [0,1]$ corresponds to $p\%$ of the standard dose of that particular drug being administered and, in particular, a value of 1 corresponds to administering 100% of the standard dose. This way we're not going beyond standard of care but rather offering an approach to compute the optimal dose reduction in the event that the standard dose would be too toxic for the patient at hand.

Further, we assume the toxicities $\mathbf{y}_t$ evolve according to a linear time invariant dynamical system $\mathbf{G}(t)$ where $\mathbf{G}(k)$ represents the impact of a dose combination $\mathbf{u}$ on the measured toxicities in k cycles. For example, $\mathbf{G}(k) \cdot \mathbf{u}_t$ represents the contribution of the dose combination administered at cycle t on the toxicities measured at the beginning of cycle $t+k$ (i.e. $\mathbf{y}_{t+k}$). We also assume that the toxicity impact of older doses is at most that of more recent doses, i.e. $\mathbf{G}(0) \geq \mathbf{G}(1) \geq \dots \mathbf{G}(k) \geq \dots$ for all $\mathbf{u} \in [0,1]^{d_u}$. This means that we assume that the dose administered at the beginning of cycle t, $\mathbf{u}_t$, contributes more (or, more precisely, at least as much) to the toxicity response between its administration and the beginning of cycle $t+1$ (when the next dose is determined and administered) than it does to the following toxicities responses over the following cycles $t+2, t+3$ etc.

Finally, we let T be the number of cycles we are planning to administer in sequence with no break, let $\|\cdot\|$ be any norm (can be chosen based on priors on relative importance of the drugs), $\mathbf{y}_t^{\text{nat}}$ be the natural progression of the toxicities in the absence of treatment, A be any matrix and $\mathbf{y}_{\max}$ be the tolerable threshold on the linear combinations of toxicities $A\mathbf{y}_t$.

[1] This is in contrast to existing modeling which is more idealized and looks at goals like optimizing dosing over an entire trajectory so as to minimize tumor size – while this is of course the ultimate goal, it is way too idealized to fit with current real life workflow. In addition to being difficult to impossible to translate to clinical practice, such approaches also lack explanatory power for past and ongoing clinical practice.

We therefore get the following optimization problem:

$$\begin{aligned} \max_{\mathbf{u}_{1:T}} \quad & \sum_{t=1}^{T} \|\mathbf{u}_t\| \\ \text{s.t.} \quad & \mathbf{u}_t \in [0,1]^{\mathbb{R}^{d_u}} \text{ for all } t \in \overline{1,T} \\ & \mathbf{y}_{t+1} = \sum_{k=0}^{t-1} \mathbf{G}(k) \cdot \mathbf{u}_{t-k} + \mathbf{y}_{t+1}^{\text{nat}} \text{ for all } t \in \overline{0,T-1} \\ & \mathrm{A}\mathbf{y}_t \leq \mathbf{y}_{\max} \ \ \forall\, t \in \overline{1,T} \end{aligned}$$

For easy reference, we summarize all the quantities in our framework, their notations, and the associated definitions in Table 5.1:

Notation	Definition
$\mathbf{u}_t$	fractional doses administered at cycle t
$\mathbf{y}_t$	toxicity measurements at start of cycle t
$\mathbf{y}_t^{\text{nat}}$	non-chemo-related toxicity evolution
$\mathbf{y}_{\max}$	tolerable threshold on $A\mathbf{y}_t$
$\mathbf{G}(\cdot)$	dynamical system of toxicity responses
$\|\cdot\|$	objective norm[2]
T	number of cycles in treatment

Table 5.1: Framework notations and corresponding meaning.

5.2 Results for Orthogonally Toxic Drugs

A special case of the above is when the drugs have orthogonal toxicity effects. If each drug contributes to its own unique set of toxicities, the objective becomes to maximize the cumulative dose of each drug subject to constraints on its set of toxicities. Thus, the problem turns from a multi-input optimization to multiple independent single-input optimization subproblems of the form:

[2]For example, using the ℓ_1-norm corresponds to weighing a decrease of $p\%$ in any drug as being equally undesirable, whereas using a matrix norm can encompass arbitrary weightings of the relative importance of the different drugs.

$$\begin{aligned}
\max_{u_{1:T}} \quad & \sum_{t=1}^{T} u_t \\
\text{s.t.} \quad & u_t \in [0,1] \text{ for all } t \in \overline{1,T} \\
& \mathbf{y}_{t+1} = \sum_{k=0}^{t-1} \mathbf{g}(k) \cdot u_{t-k} + \mathbf{y}_{t+1}^{\text{nat}} \text{ for all } t \in \overline{0,T-1} \\
& \mathrm{A}\mathbf{y}_t \leq \mathbf{y}_{\max} \ \ \forall\, t \in \overline{1,T}
\end{aligned}$$

where u_t is now a scalar in $[0,1]$ and $\mathbf{g}(\cdot,\cdot)$ is now a d_y-dimensional vector. In this special case, we can identify a policy which guarantees optimality. Importantly, this policy is intuitively understandable and actionable by clinicians. To proceed, we formally define a per-cycle *maximum tolerable dose*.

Definition 5.1 (Maximum Tolerable Dose)**.** *We define the maximum tolerable dose (MTD) at some time t to be the maximal dose u_t that would ensure $y_\tau \leq y^{\max}$ for all $\tau > t$ if we were to stop chemotherapy right after, i.e. take $u_{>t} \equiv 0$ afterwards.*

Intuitively, the above represents the maximum dose that would be safe *if we were to fully stop chemotherapy after*. Note that this is a local and iterative concept. It only depends on past doses and not on future doses (because by definition it is the dose that would be safe if we stopped, it does not require estimating what would happen in the future under different dosing schemes). Our main formal/theoretical contribution is the following theorem:

Theorem 5.2 (Optimality of MTD Dosing)**.** *Taking the MTD at every time t ensures $y_t \leq y^{\max}$ for every time t and achieves the maximum cumulative dose (over protocols which maintain toxicity feasibility at every step).*

Proof of Theorem 5.2. We will show the above holds by contradiction. Let $u^{(0)}_{0:T-1}$ be a dosing schedule that satisfies the toxicity constraint at every time t and results in the maximum cumulative dose achievable by feasible protocols. For the sake of contradiction, assume $u^{(0)}_{0:T-1}$ is *not* taking the MTD at every time t *and* that $u^{(0)}_{0:T-1}$ achieves a cumulative dose *strictly* larger than taking the MTD at every time t. Then, there must exist some time t_0 at which it would be feasible to increase the dose by some $\epsilon > 0$ and maintain all future toxicities below $y^{\max}$ (if we were to take no dose after). Let t_0 be the first such time (i.e. the *first* time $u^{(0)}_t$ is not the MTD) and t'_0 be the first time after t_0 where the dose is strictly positive. Let ϵ'_0 be the *highest* dose increase that maintains feasibility at $t_0 + 1$ (and therefore matches the toxicity constraint exactly). Let $\epsilon_0 = \min\{\epsilon'_0, u_{t'_0}\}$. Now consider the following dosing schedule $u^{(1)}_{0:T-1}$:

$$u^{(1)}_t = \begin{cases} u^{(0)}_t + \epsilon_0 & t = t_0 \\ u^{(0)}_t - \epsilon_0 & t = t'_0 \\ u^{(0)}_t & \text{otherwise} \end{cases}$$

and let $y'_{1:T}$ be the resulting toxicity levels. Note that $u^{(1)}_{0:T-1}$ is still feasible since

(i) For $t \leq t_0$, we have $y'_t = y_t \leq y^{\max}$ for $t \leq t_0$ by the feasibility of $u^{(0)}_{0:T-1}$ and the fact that the schedules are identical up until t_0,

(ii) For $t = t_0 + 1$, we have $y'_{t_0+1} \leq y^{\max}$ because we have precisely taken ϵ'_0 such that $u^{(1)}_{t_0} = u^{(0)}_{t_0} + \epsilon'_0$ leads to exact toxicity constraint satisfaction at $t_0 + 1$ and we have $\epsilon_0 \leq \epsilon'_0$,

(iii) For $t > t_0 + 1$ and $t < t'_0$, we have that $u^{(1)}_t = u^{(0)}_t = 0$ and by the definition of MTD, $y_t \leq y^{\max}$ in this period.

(iv) For $t > t'_0$, we have

$$\begin{aligned}
y'_{t+1} &= \sum_{k=0}^{t} g(k) \cdot u^{(1)}_{t-k} + y^{\text{nat}}_{t+1} \\
&= \sum_{k=0}^{t} g(k) \cdot u^{(0)}_{t-k} + g(t - t_0) \cdot \epsilon_0 + g(t - t'_0) \cdot (-\epsilon_0) + y^{\text{nat}}_{t+1} \\
&= \sum_{k=0}^{t} g(k) \cdot u^{(0)}_{t-k} + y^{\text{nat}}_{t+1} + (g(t - t_0) - g(t - t'_0)) \cdot \epsilon_0 \\
&= y_{t+1} + (g(t - t_0) - g(t - t'_0)) \cdot \epsilon_0 \\
&\leq y_{t+1} \\
&\leq y^{\max}
\end{aligned}$$

where we used that $g(t - t_0) \leq g(t - t'_0)$ and $\epsilon_0 > 0$, as well as the feasibility of $u^{(0)}_{0:T-1}$.

Note also that $\sum_{t=0}^{T-1} u^{(1)}_t = \sum_{t=0}^{T-1} u^{(0)}_t$. Therefore, the dosing schedules $u^{(1)}_{0:T-1}$ and $u^{(0)}_{0:T-1}$ achieve the same cumulative dose while maintaining feasibility. The key difference is that $u^{(1)}_{0:T-1}$ is prescribing the MTD up to *and including* time t_0, while the starting dosing schedule $u^{(0)}_{0:T-1}$ was only prescribing the MTD up to time $t_0 - 1$.

If $u^{(1)}_{0:T-1}$ is indeed prescribing the MTD at every step, then we are done since this implies prescribing the MTD at every step achieves the maximum cumulative dose while maintaining toxicity feasibility at every t, which leads to the desired contradiction.

Otherwise, there must again exist a time $t_1 \geq t_0$ at which it is possible to administer a higher dose while ensuring $y_{t_1+1} \leq y^{\max}$ as well as the ability to maintain future toxicities below $y^{\max}$ if we were to stop the chemo. Again we can define ϵ'_1 to be the dose increase that leads to exact next-step constraint satisfaction, let ϵ_1 be the minimum between this and the next strictly positive dose,[3] and consider an analogously modified dosing schedule

[3]If one does not exist then we are done since by definition of not being MTD, we can increase the last positive dose and strictly increase the cumulative dose.

$u^{(2)}_{0:T-1}$ which takes $u^{(2)}_{t_1} = u^{(1)}_{t_1} + \epsilon_1, u^{(2)}_{t'_1} = u^{(1)}_{t'_1} - \epsilon_1$ and keeps all other doses identical. By the exact same arguments as before we have that $u^{(2)}_{0:T-1}$ is feasible while achieving the same cumulative dose.

We can repeat this procedure, iteratively creating $u^{(k)}_{0:T-1}$ from $u^{(k-1)}_{0:T-1}$ as long as there exists some time $t_{k-1} < T$ at which $u^{(k-1)}_{0:T-1}$ is not equal to the MTD. By the above, we will iteratively have that each resulting $\sum_{t=0}^{T-1} u^{(k)}_t$ is feasible and leads to the same cumulative dose as the prior schedule, i.e. $\sum_{t=0}^{T-1} u^{(k)}_t = \sum_{t=0}^{T-1} u^{(k-1)}_t$. Note that since $t_k \geq t_{k-1}$ and we always add a strictly positive quantity to an existing dose (up to the limit of 1) this process will eventually terminate leading to a dosing schedule $u^{(K)}_{0:T-1}$ which takes the MTD at every time t, yet is feasible at all times t and achieves cumulative dose $\sum_{t=0}^{T-1} u^{(K)}_t = \sum_{t=0}^{T-1} u^{(K-1)}_t = \ldots = \sum_{t=0}^{T-1} u^{(0)}_t$. This is precisely the desired contradiction, concluding our proof. □

In this case, the results imply that if the effects were *separable* then **the problem of optimal dosing indeed can be reduced to next-step toxicity prediction**. While it is difficult in practice to know if drugs are contributing orthogonal toxicities, we note that this means for single-agent regimens, we have defined a clear optimal policy for clinicians and a prediction task which, if solved, would enable optimal treatment.

5.3 Difficulties Arising for Cross-Toxic Drugs

The story however changes when we consider the situation of drugs with toxicity interactions. At the core of the change lies the fact that in this case the maximum tolerable dose combination is not unique like in the case where we can look at the subcase of a single drug. In the case of multiple drugs, there may be multiple combinations that are maximally tolerable:

Definition 5.3 (MTD Combination)**.** *We define an MTD combination at some time t to be any dose vector $\mathbf{u}_t$ that would: (1) ensure $\mathbf{y}_\tau \leq \mathbf{y}^{\max}$ for all $\tau > t$ if we were to stop chemotherapy right after (i.e., take $\mathbf{u}_{>t} \equiv 0$ afterwards), and (2) raising any entry of $\mathbf{u}_t$ by any $\epsilon > 0$ would lead to violating (1).*

In this case we find that having a dosing trajectory where at every step the combination is MTD is *necessary but not sufficient* for optimality. This is because now there are many, many diverging trajectories which all stay within MTD. To help provide insight into why this is true, we present the dummy example in Figure 5.1. In this example, both full dose of drug 1 or full dose of drug 2 would lead to MTD. However, drug 1 toxicity decreases at a much slower rate, making drug 2 preferable assuming equal efficacy.

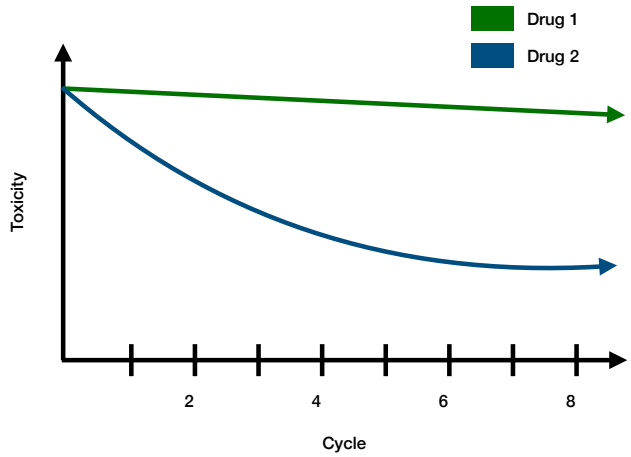


Figure 5.1: Exaggerated toxicity responses.

There are examples much more complicated than this as well, such as if the decays cross at some time. Whichever trajectory is "optimal" depends on the chosen objective norm (representing beliefs about relative efficacy) and relative toxicity decay. An intuitive illustration is depicted in Figure 5.2. While prior results still hold if we fix two drugs and optimize only one, it cannot tell us which drug it is optimal to reduce.

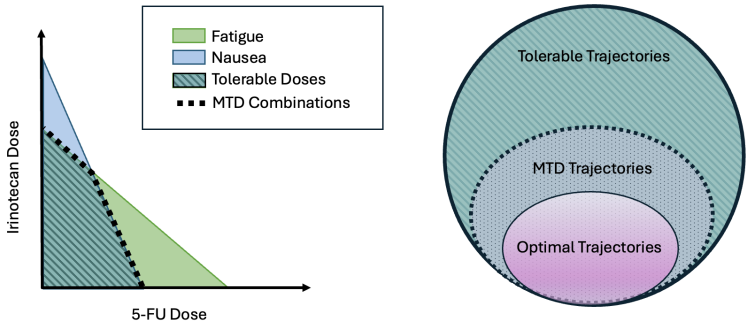


Figure 5.2: Maximally Tolerable Dose Combinations and Optimality Space

5.4 Identifying Sources of Variability

The complex nature of combination chemotherapy treatments leads to variability in clinical action. We use our framework to garner a fine-grained understanding of the sources of variability. We identify three primary sources of variability: clinician, patient, and problem. These sources, along with their mapping to the theoretic framework, are described in the subsections below.

I. Clinician-induced variability

I.a. Strategy: Clinicians believe different approaches are optimal. Our framework captures this through:

1. $\|\mathbf{u}_t\|$ allows for arbitrary weighting of drugs,
2. $\mathbf{y}_{\max}$ allows for clinician having different adverse event intensity thresholds.

I.b. Prediction: Clinicians believe in the same approach but differ in their ability to predict it. In our framework, this corresponds to clinicians differing in their ability to map $\mathbf{u}$ to $\mathbf{y}_{t+1}$

II. Patient-induced variability

II.a. Tolerance: Patients have different maximum levels of adverse effects they can tolerate. This is especially relevant in palliative settings. In our framework, this corresponds to $\mathbf{y}_{\max}$ being determined by patient.

II.b. Individual Susceptibility: Patients have different adverse responses to the same treatment. In our framework, this is captured by $\mathbf{G}$ being patient-specific.

III. Problem-induced variability

III.a. Nature of Signal: Adverse effects may be objective or subjective in how they are measured. This is allowed in our framework, since $\mathbf{y}$ may be a laboratory value or a CTCAE grade of 'nausea.'

III.b. Complex Mechanistic Uncertainty: The general landscape of treatment effect functions is unknown, including the mapping of multiple treatments and doses to adverse effects and efficacy. In our framework this means the weights for $\|\mathbf{u}_t\|$ from the efficacy of drugs and drug combinations is unknown and also that $\mathbf{G}$ responses from regimens are unknown.

5.5 Validation on Retrospective Data

While MTD is not provably strictly optimal in the multi-drug setting, because the simulations seem to indicate MTD-based decisions are closer to optimal than heuristic decisions, we assessed the correlation between MTD aligned trajectories and higher cumulative dose. To perform this analysis, we first retrospectively classified each patient cycle as MTD, below MTD, or above MTD according to patient absolute neutrophil count, platelet count, neuropathy, nausea, anorexia, and fatigue. Using clinical notes, neuropathy symptoms were

collected, as were nausea, fatigue, and anorexia. Lab values were taken from pre-cycle visit check-ins. For each toxicity, we define a set of maximally tolerable ranges according to Table 5.2.

Toxicity	**Maximally Tolerable Range**	**Drugs**
Neutrophil Count	$(1.4, 1.8)$	All
Platelet Count	$(70, 100)$	All
Neuropathy	Only fingers and/or toes	Oxaliplatin
Nausea, Anorexia, Fatigue	Grade 1	All

Table 5.2: Toxicity ranges and associated drugs

We then classify each cycle according to Table 5.3. Intuitively, if the next cycle toxicity is within the maximally tolerable range, then we classify the current cycle as a maximally tolerable dose. If the next toxicity is out of the maximally tolerable range, and the dose is greater than 0, then the current dose is more than maximally tolerable. If the next toxicity is less than the maximally tolerable range (i.e. the patient is experiencing less than maximum toxicity) and the dose is reduced, then this is classified as below maximally tolerable. We also identify for each patient the planned number of cycles from their clinical notes.

Classification	**Criteria**	**Current Doses**
Below MTD	(1) Next toxicity below minimum toxicity threshold	(1) $< 100\%$
MTD	(1) Next toxicity within toxicity range (2) Next toxicity less than or within toxicity range (3) Next toxicity above toxicity range	(1) Any (2) 100% (3) $= 0$
Above MTD	(1) Emergency visit (2) Next toxicity above toxicity range	(1) Any (2) > 0

Table 5.3: MTD classification schema

We acknowledge that this kind of classification is only retrospectively possible, and that a variety of factors can influence a provider's ability to assess whether a dose will be maximally tolerable. Rather than assess clinician performance, the goal of this analysis is to assess the

optimality of MTD. To do this, we want to test if closer adherence to MTD is correlated with higher cumulative dose.

For each patient, we calculate the fraction of cycles they received an MTD, above MTD, or below MTD dose. We then regress the administered fraction of planned dose on these fractions and baseline patient toxicities. The results of this regression can be found in Table 5.4.

Feature	**Coefficient**	**Error**
Fraction of Cycles at MTD	0.77	0.02
Fraction of Cycles Above MTD	0.80	0.04
Fraction of Cycles Sub MTD	0.68	0.03

Table 5.4: Model Coefficients

We find that a higher fraction of MTD cycles is significantly predictive of a higher cumulative dose. Importantly, we find that the MTD fraction is more predictive than the below MTD fraction. This refutes the potential argument that it is better to administer lower doses in the hopes of administering more cycles and having a higher cumulative dose. Instead, it is better to always administer as close to what the patient can tolerate as possible. Additionally, a higher fraction of above MTD cycles is not significantly more predictive of a higher cumulative dose than MTD cycles. Intuitively, this means that by always administering just as much as the patient can tolerate, you would be able to administer about as much dose as administering more than the patient could tolerate.

5.6 Numerical Simulations

The complexity of combination chemotherapy means realistic simulations are difficult. Because any simulation will then necessarily make simplifications, we focus our simulations on a simpler case where each drug has a single toxicity function g which can be considered to capture all toxicity response of that drug, with some combined maximum toxicity level. We additionally assume that toxicity response is linear in dose (i.e. $\mathbf{y}_{t+1} = \sum_{k=0}^{t-1} \mathbf{g}(k) \cdot \mathbf{u}_{t-k} + \mathbf{y}_{t+1}^{\text{nat}}$) We model the 5-FU infusion and bolus toxicity impulse responses as exponential decays and oxaliplatin as a Hill function (Figure 5.3).

Based on pharmacokinetic/pharmacodynamic literature, 5-FU bolus is set to have the shortest decay, with the 5-FU infusion decay being somewhat longer and lagging the bolus [83] [11] [97]. Oxaliplatin toxicity is often modeled with a three-compartment model and tends to have a longer tail [67]. To simulate patient variability, we sample parameters for these functions from a normal distribution at various σ values to see how the level of variability influences optimality.

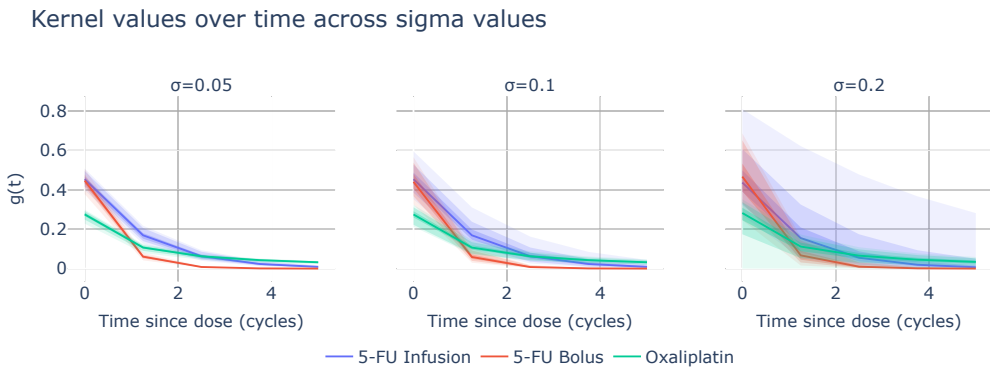


Figure 5.3: Simulated toxicity impulse responses $g(t)$ for the three FOLFOX drugs, plotted against the number of cycles since administration.

Using these toxicity functions, we are able to calculate a true optimal dosing schedule using cvxpy, which we know will be an MTD given our theoretical findings. Importantly, it will be the best possible MTD schedule, and subsequently the best possible dosing schedule of any non-violating schedule. To assess the extent to which any perfectly MTD policy deviates from this optimal value, we identify what we will call random **Oracle MTD** trajectories. **Oracle MTD** trajectories are calculated assuming full knowledge of the impulse response where $u_t^{(i)}$ indicates the dose of drug i at time t and $g^{(i)}$ is the impulse response of drug $u^{(i)}$:

$$u_t^{(i)} = \frac{\mathbf{y}_{\max} - \sum_{k=1}^{t} \boldsymbol{g}(k) \cdot \mathbf{u}_{t-k}}{\boldsymbol{g}^{(i)}(0)}$$

Once the maximum proportion of dose for each drug is calculated, if not all maximum doses can be administered, we have to select a dose from the set of possible MTDs. For the purpose of understanding the general performance of perfectly MTD trajectories, we randomly sample a viable MTD combination. We calculate the true optimal cumulative dose and 10 random Oracle MTD trajectories for 10,000 sampled impulse responses. We find that random MTD policies are still able to perform relatively well, with most above 90% of optimal (Figure 5.4).

While this result is reassuring, doctors are not randomly choosing a dose combination. As discussed previously, there is an intuition for drug efficacy and for toxicity burden. If we use any knowledge about the relative efficacy and toxicities of the drug, we are able to improve performance over random. For our specific set-up, we find that allocating drug doses according to efficacy is particularly effective in achieving near optimal performance.

This is true even if we limit the knowledge of the impulse response to previously observed data points and a population estimate. The estimation amounts to calculating the observed toxicity up until the current cycle from all previous doses (specific to a patient) and using population estimates to calculate expected toxicity decay and toxicity contribution from the

current cycle's dose. Using ANC as an example, this would mean looking at the patient's pre-cycle ANC lab and, using the general population understanding of how someone might respond to a drug, considering how much we can dose such that the patient's next ANC lab is still healthy. We call this method of dose calculation **Approximate MTD**, as these combinations are likely not true MTD combinations for the specific patient but rather estimations. Figure 5.4 compares both methods against the optimum.

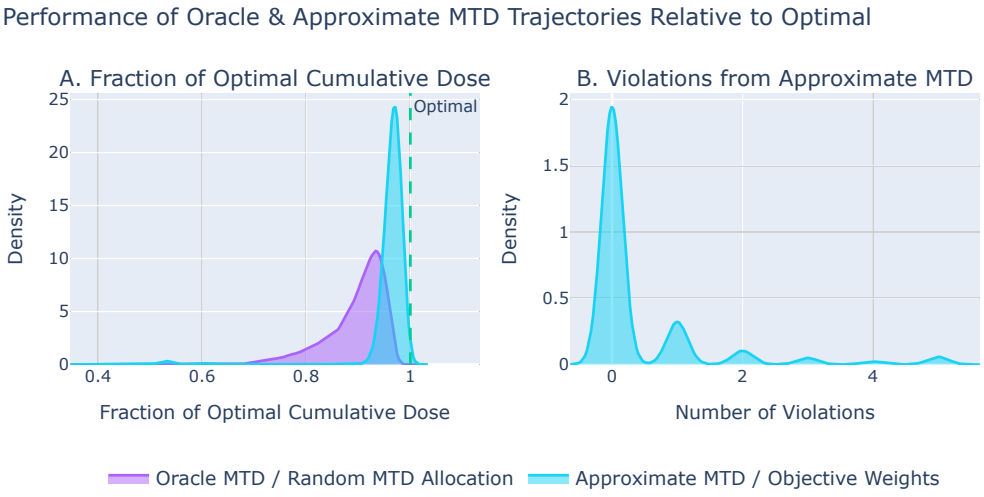


Figure 5.4: Performance of Oracle and Approximate MTD trajectories.

Taken together, these simulations suggest that most of the work is done by staying MTD: even randomly chosen MTD combinations recover the large majority of the optimal cumulative dose, and a simple efficacy-based ordering closes much of the remaining gap. Encouragingly, this holds even when the impulse responses are only estimated rather than known.

Bibliography

[1] Hammaad Adam et al. "Should I Stop or Should I Go: Early Stopping with Heterogeneous Populations". In: *Advances in Neural Information Processing Systems* 36 (2024).

[2] Alicia Agnoli et al. "Association of dose tapering with overdose or mental health crisis among patients prescribed long-term opioids". In: *JAMA* 326.5 (2021), pp. 411–419.

[3] Marco Almada. "Human intervention in automated decision-making: Toward the construction of contestable systems". In: *Proceedings of the Seventeenth International Conference on artificial intelligence and law*. 2019, pp. 2–11.

[4] P.M. Aronow and Cyrus Samii. "Conservative variance estimation for sampling designs with zero pairwise inclusion probabilities". In: *Survey Methodology* 39.1 (2013), pp. 231–241.

[5] P.M. Aronow and Cyrus Samii. "Estimating Average Causal Effects under General Interference". In: *Annals of Applied Statistics* 11.4 (2017), pp. 1912–1947. DOI: 10.1214/16-aoas1005.

[6] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. "Finite-Time Analysis of the Multiarmed Bandit Problem". In: *Machine Learning* 47.2–3 (May 2002), pp. 235–256. ISSN: 0885-6125. DOI: 10.1023/A:1013689704352.

[7] Peter Auer et al. "The Nonstochastic Multiarmed Bandit Problem". In: *SIAM Journal on Computing* 32.1 (2002), pp. 48–77. DOI: 10.1137/S0097539701398375.

[8] Akshay Balsubramani. "Sharp finite-time iterated-logarithm martingale concentration". In: *arXiv preprint arXiv:1405.2639* (2014).

[9] Noam Barda et al. "Safety of the BNT162b2 mRNA Covid-19 vaccine in a nationwide setting". In: *New England Journal of Medicine* 385.12 (2021), pp. 1078–1090.

[10] AD Barker et al. "I-SPY 2: An Adaptive Breast Cancer Trial Design in the Setting of Neoadjuvant Chemotherapy". In: *Clinical Pharmacology & Therapeutics* 86.1 (2009), pp. 97–100. DOI: 10.1038/clpt.2009.68.

[11] Jérémy Béguin et al. "Pharmacokinetics and tolerance of repeated oral administration of 5-fluorocytosine in healthy dogs". In: *BMC Veterinary Research* 17 (2021), p. 220. DOI: 10.1186/s12917-021-02927-5.

[12] Chantal Berna, Ronald J. Kulich, and James P. Rathmell. "Tapering long-term opioid therapy in chronic noncancer pain: evidence and recommendations for everyday practice". In: *Mayo Clinic Proceedings*. Vol. 90. 6. Elsevier, 2015, pp. 828–842.

[13] Dimitri Bertsekas and John N Tsitsiklis. *Introduction to probability*. Vol. 1. Athena Scientific, 2008.

[14] Matthew Blackwell, Nicole E. Pashley, and Dominic Valentino. *Batch Adaptive Designs to Improve Efficiency in Social Science Experiments*. Working Paper. Harvard University, 2022. URL: https://www.mattblackwell.org/files/papers/batch_adaptive.pdf.

[15] Jack Bowden and Lorenzo Trippa. "Unbiased estimation for response adaptive clinical trials". In: *Statistical methods in medical research* 26 (Aug. 2015).

[16] Giuseppe Burtini, Jason Loeppky, and Ramon Lawrence. "A survey of online experiment design with the stochastic multi-armed bandit". In: *arXiv preprint arXiv:1510.00757* (2015).

[17] Yong Cai and Ahnaf Rafi. "On the Performance of the Neyman Allocation with Small Pilots". arXiv:2206.04643. 2022. arXiv: 2206.04643 [econ.EM].

[18] Robert T Chen et al. "The vaccine adverse event reporting system (VAERS)". In: *Vaccine* 12.6 (1994), pp. 542–550.

[19] John J Cherian and Emmanuel J Candès. "Statistical inference for fairness auditing". In: *arXiv preprint arXiv:2305.03712* (2023).

[20] Ziyu Chi, Aaditya Ramdas, and Ruodu Wang. "Multiple testing under negative dependence". In: *arXiv preprint arXiv:2212.09706* (2022).

[21] Brian Cho, Kyra Gan, and Nathan Kallus. *Peeking with PEAK: Sequential, Nonparametric Composite Hypothesis Tests for Means of Multiple Data Streams*. June 2024. DOI: 10.48550/arXiv.2402.06122. arXiv: 2402.06122 [cs, stat]. (Visited on 11/04/2024).

[22] Ben Chugg et al. "Auditing Fairness by Betting". In: *Advances in Neural Information Processing Systems* 36 (2024).

[23] Thierry Conroy et al. "FOLFIRINOX versus Gemcitabine for Metastatic Pancreatic Cancer". In: *New England Journal of Medicine* 364.19 (2011), pp. 1817–1825. DOI: 10.1056/NEJMoa1011923.

[24] Thomas M Cover. "Universal portfolios". In: *Mathematical Finance* 1.1 (1991), pp. 1–29.

[25] Ashok Cutkosky and Francesco Orabona. "Black-box reductions for parameter-free online learning in banach spaces". In: *Conference On Learning Theory*. PMLR. 2018, pp. 1493–1529.

[26] Jessica Dai, Paula Gradu, and Christopher Harshaw. *Clip-OGD: An Experimental Design for Adaptive Neyman Allocation in Sequential Experiments*. 2023. arXiv: 2305.17187 [stat.ME]. URL: https://arxiv.org/abs/2305.17187.

[27] Jessica Dai, Nika Haghtalab, and Eric Zhao. "Learning With Multi-Group Guarantees For Clusterable Subpopulations". In: *arXiv preprint arXiv:2410.14588* (2024).

[28] Jessica Dai et al. *From Individual Experience to Collective Evidence: A Reporting-Based Framework for Identifying Systemic Harms*. 2025. arXiv: 2502.08166 [cs.LG]. URL: https://arxiv.org/abs/2502.08166.

[29] Jeffrey R. Eisele. "The doubly adaptive biased coin design for sequential clinical trials". In: *Journal of Statistical Planning and Inference* 38.2 (1994), pp. 249–261. ISSN: 0378-3758.

[30] Lisa G. Ensign and Kevin B. Cohen. "A Primer to the Structure, Content and Linkage of the FDA's Manufacturer and User Facility Device Experience (MAUDE) Files". In: *eGEMs (Generating Evidence & Methods to Improve Patient Outcomes)* 5.1 (2017), p. 12. DOI: 10.5334/egems.221.

[31] Jean Feng et al. "Monitoring machine learning-based risk prediction algorithms in the presence of performativity". In: *International Conference on Artificial Intelligence and Statistics*. PMLR. 2024, pp. 919–927.

[32] Joshua J. Fenton et al. "Trends and rapidity of dose tapering among patients prescribed long-term opioid therapy, 2008-2017". In: *JAMA Network Open* 2.11 (2019), e1916271–e1916271.

[33] Benjamin Freedman. "Equipoise and the Ethics of Clinical Research". In: *New England Journal of Medicine* 317.3 (1987), pp. 141–145. DOI: 10.1056/NEJM198707163170304.

[34] David A. Freedman. "On regression adjustments to experimental data". In: *Advances in Applied Mathematics* 40 (2008), pp. 180–193. DOI: 10.1016/j.aam.2006.12.003.

[35] Paula Gradu and Benjamin Recht. *Online Control For Adaptive Tapering Of Medications*. 2023. arXiv: 2309.11629 [eess.SY]. URL: https://arxiv.org/abs/2309.11629.

[36] Matthew Groh and David McKenzie. "Macroinsurance for Microenterprises: A Randomized Experiment in Post-revolution Egypt". In: *Journal of Development Economics* 118 (2016), pp. 13–25. DOI: 10.1016/j.jdeveco.2015.08.003.

[37] Vitor Hadad et al. "Confidence intervals for policy evaluation in adaptive experiments". In: *PNAS* 118.15 (2021). DOI: 10.1073/pnas.2014602118.

[38] Jinyong Hahn, Keisuke Hirano, and Dean Karlan. "Adaptive Experimental Design Using the Propensity Score". In: *Journal of Business & Economic Statistics* 29.1 (2011), pp. 96–108. (Visited on 12/21/2022).

[39] Dae Woong Ham et al. "Design-Based Confidence Sequences for Anytime-valid Causal Inference". arXiv:2210.08639. 2022. arXiv: 2210.08639 [stat.ME].

[40] Christopher Harshaw, Joel A. Middleton, and Fredrik Sävje. "Optimized variance estimation under interference and complex experimental designs". arXiv:2112.01709. 2021. arXiv: 2112.01709v3 [stat.ME].

[41] Christopher Harshaw, Fredrik Sävje, and Yitan Wang. "A Design-Based Riesz Representation Framework for Randomized Experiments". arXiv:2210.08698. 2022. arXiv: 2210.08698v2 [stat.ME].

[42] Lakhbir S. Hayre and Bruce W. Turnbull. "Estimation of the Odds Ratio in the Two-Armed Bandit Problem". In: *Biometrika* 68.3 (1981), pp. 661–668.

[43] Elad Hazan et al. "Introduction to online convex optimization". In: *Foundations and Trends® in Optimization* 2.3-4 (2016), pp. 157–325.

[44] Elad Hazan, Amit Agarwal, and Satyen Kale. "Logarithmic regret algorithms for online convex optimization". In: *Machine Learning* 69.2 (2007), pp. 169–192.

[45] Ursula Hébert-Johnson et al. "Multicalibration: Calibration for the (computationally-identifiable) masses". In: *International Conference on Machine Learning*. PMLR. 2018, pp. 1939–1948.

[46] Stephen G. Henry et al. "Patients' experience with opioid tapering: a conceptual model with recommendations for clinicians". In: *The Journal of Pain* 20.2 (2019), pp. 181–191.

[47] Mark Abie Horowitz, Robin M. Murray, and David Taylor. "Tapering antipsychotic treatment". In: *JAMA Psychiatry* 78.2 (2021), pp. 125–126.

[48] Mark Abie Horowitz and David Taylor. "Tapering of SSRI treatment to mitigate withdrawal symptoms". In: *The Lancet Psychiatry* 6.6 (2019), pp. 538–546.

[49] Mark Abie Horowitz et al. "A method for tapering antipsychotic treatment that may minimize the risk of relapse". In: *Schizophrenia Bulletin* 47.4 (2021), pp. 1116–1129.

[50] D. G. Horvitz and D. J. Thompson. "A generalization of sampling without replacement from a finite universe". In: *Journal of the American Statistical Association* 47.260 (1952), pp. 663–685. DOI: 10.1080/01621459.1952.10483446.

[51] Steven R. Howard et al. "Time-uniform, nonparametric, nonasymptotic confidence sequences". In: *The Annals of Statistics* 49.2 (2021), pp. 1055–1080.

[52] Feifang Hu and William F. Rosenberger. "Optimality, Variability, Power: Evaluating Response-Adaptive Randomization Procedures for Treatment Comparisons". In: *Journal of the American Statistical Association* 98.463 (2003), pp. 671–678.

[53] Guido W Imbens and Donald B Rubin. *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press, 2015.

[54] Kevin Jamieson et al. "lil'ucb: An optimal exploration algorithm for multi-armed bandits". In: *Conference on Learning Theory*. PMLR. 2014, pp. 423–439.

[55] Ronit Juthani, Sachin Punatar, and Indraneel Mittra. "New light on chemotherapy toxicity and its prevention". In: *BJC Reports* 2.1 (2024). DOI: 10.1038/s44276-024-00064-8.

[56] Harold Kalant. "Biological models of alcohol tolerance and physical dependence". In: *Alcohol Intoxication and Withdrawal I: Experimental Studies* (1973), pp. 3–14.

[57] George F. Koob. "The dark side of emotion: the addiction perspective". In: *European Journal of Pharmacology* 753 (2015), pp. 73–87.

[58] George F. Koob and Michel Le Moal. "Drug addiction, dysregulation of reward, and allostasis". In: *Neuropsychopharmacology* 24.2 (2001), pp. 97–129.

[59] Wouter M Koolen. *A quick and dirty finite time law of the iterated logarithm result*. 2017. URL: https://blog.wouterkoolen.info/QnD_LIL/post.html.

[60] Kurt Kroenke and Robert L. Spitzer. *The PHQ-9: a new depression diagnostic and severity measure*. 2002.

[61] Martin Kulldorff et al. "A maximized sequential probability ratio test for drug and vaccine safety surveillance". In: *Sequential analysis* 30.1 (2011), pp. 58–78.

[62] Susan Landau et al. "Challenging the Machine: Contestability in Government AI Systems". In: *arXiv preprint arXiv:2406.10430* (2024).

[63] Kathryn F Larson et al. "Myocarditis after BNT162b2 and mRNA-1273 vaccination". In: *Circulation* 144.6 (2021), pp. 506–508.

[64] Tor Lattimore and Csaba Szepesvári. *Bandit Algorithms*. Cambridge University Press, 2020.

[65] Katharine MN Lee et al. "A Gender Hypothesis of sex disparities in adverse drug events". In: *Social Science & Medicine* 339 (2023), p. 116385.

[66] Michael P. Leung. "Causal Inference Under Approximate Neighborhood Interference". In: *Econometrica* 90.1 (2022), pp. 267–293.

[67] F. Lévi et al. "Oxaliplatin: pharmacokinetics and chronopharmacological aspects". In: *Clinical Pharmacokinetics* 38.1 (2000), pp. 1–21. DOI: 10.2165/00003088-200038010-00001.

[68] Winston Lin. "Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique". In: *Annals of Applied Statistics* 7.1 (2013), pp. 295–318.

[69] Zhi Liu, Uma Bhandaram, and Nikhil Garg. "Quantifying spatial under-reporting disparities in resident crowdsourcing". In: *Nature Computational Science* 4.1 (2024), pp. 57–65.

[70] Zhi Liu and Nikhil Garg. "Equity in resident crowdsourcing: Measuring under-reporting without ground truth data". In: *Proceedings of the 23rd ACM Conference on Economics and Computation*. 2022, pp. 1016–1017.

[71] Fotios Loupakis et al. "Initial Therapy with FOLFOXIRI and Bevacizumab for Metastatic Colorectal Cancer". In: *New England Journal of Medicine* 371.17 (2014), pp. 1609–1618. DOI: 10.1056/NEJMoa1403108.

[72] B Lovelace. *FDA Permits Use of the Pfizer-BioNTech Covid Vaccine in Kids Ages 12 to 15*. 2021. URL: https://www.cnbc.com/2021/05/10/pfizer-covid-vaccine-fda-clears-use-in-kids-ages-12-to-15.html.

[73] Mehrdad Mahdavi, Rong Jin, and Tianbao Yang. *Trading regret for efficiency: online convex optimization with long term constraints*. 2012.

[74] Shie Mannor and John N. Tsitsiklis. "Online learning with constraints". In: *Lecture Notes in Computer Science* 4005 (2006), p. 529.

[75] Mayme Marshall et al. "Symptomatic acute myocarditis in 7 adolescents after Pfizer-BioNTech COVID-19 vaccination". In: *Pediatrics* 148.3 (2021).

[76] Saif Abu Mouch et al. "Myocarditis following COVID-19 mRNA vaccination". In: *vaccine* 39.29 (2021), pp. 3790–3793.

[77] R. Narain. "On sampling without replacement with varying probabilities". In: *Journal of the Indian Society of Agricultural Statistics* 3 (1951), pp. 169–175.

[78] Jerzy Neyman. "On the application of probability theory to agricultural experiments. Essay on principles. Section 9". In: *Statistical Science* 5.4 (1923). Reprinted in 1990., pp. 465–472.

[79] Jerzy Neyman. "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection". In: *Journal of the Royal Statistical Society* 97.4 (1934), pp. 558–625. ISSN: 09528385. (Visited on 04/25/2023).

[80] Molly Offer–Westort, Alexander Coppock, and Donald P. Green. "Adaptive Experimental Design: Prospects and Applications in Political Science". In: *American Journal of Political Science* 65.4 (2021), pp. 826–844.

[81] Do-Youn Oh et al. "Durvalumab plus Gemcitabine and Cisplatin in Advanced Biliary Tract Cancer". In: *NEJM Evidence* 1.8 (2022). DOI: 10.1056/EVIDoa2200015.

[82] Matthew E Oster et al. "Myocarditis cases reported after mRNA-based COVID-19 vaccination in the US from December 2020 to August 2021". In: *Jama* 327.4 (2022), pp. 331–340.

[83] G. J. Peters et al. "Prolonged retention of high concentrations of 5-fluorouracil in human and murine tumors as compared with plasma". In: *Cancer Chemotherapy and Pharmacology* 31.4 (1993), pp. 269–276. DOI: 10.1007/BF00685670.

[84] Emily Potter et al. "FDA Adverse Event Reporting System (FAERS) Essentials: A Guide to Understanding, Applying, and Interpreting Adverse Event Data Reported to FAERS". In: *Clinical Pharmacology and Therapeutics* 118.3 (2025), pp. 567–582. DOI: 10.1002/cpt.3701.

[85] National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: Ethical Principles and Guidelines for the Commission for the Protection of Human Subjects of Biomedical and Behavioral Research*. Tech. rep. US Department of Health, Education, and Welfare, 1978.

[86] Herbert Robbins. "Some aspects of the sequential design of experiments". In: *Bulletin of the American Mathematical Society* 58.5 (1952), pp. 527–535.

[87] Jerrold F. Rosenbaum et al. "Selective serotonin reuptake inhibitor discontinuation syndrome: a randomized clinical trial". In: *Biological Psychiatry* 44.2 (1998), pp. 77–87.

[88] W Rosenberger et al. "Optimal adaptive designs for binary response trials". In: *Biometrics* 57 (Oct. 2001), pp. 909–13.

[89] Donald B. Rubin. "Comment: Randomization analysis of experimental data". In: *Journal of the American Statistical Association* 75.371 (1980), p. 591.

[90] Fredrik Sävje, P. M. Aronow, and Michael G. Hudgens. "Average treatment effects in the presence of unknown interference". In: *Annals of Statistics* 49.2 (2021), pp. 673–701.

[91] Eric M. Schwartz, Eric T. Bradlow, and Peter S. Fader. "Customer Acquisition via Display Advertising Using Multi-Armed Bandit Experiments". In: *Marketing Science* 36.4 (2017), pp. 500–522.

[92] Glenn Shafer. "Testing by betting: A strategy for statistical and scientific communication". In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 184.2 (2021), pp. 407–431.

[93] Divya Shanmugam, Kaihua Hou, and Emma Pierson. "Quantifying disparities in intimate partner violence: a machine learning method to correct for underreporting". In: *npj Women's Health* 2.1 (2024), p. 15.

[94] Tom T Shimabukuro et al. "Safety monitoring in the vaccine adverse event reporting system (VAERS)". In: *Vaccine* 33.36 (2015), pp. 4398–4405.

[95] H. Solomon and S. Zacks. "Optimal Design of Sampling from Finite Populations: A Critical Review and Indication of New Research Areas". In: *Journal of the American Statistical Association* 65.330 (1970), pp. 653–677. (Visited on 12/21/2022).

[96] Richard L. Solomon. "The opponent-process theory of acquired motivation: the costs of pleasure and the benefits of pain". In: *American Psychologist* 35.8 (1980), p. 691.

[97] Catherine Terret et al. "Dose and time dependencies of 5-fluorouracil pharmacokinetics". In: *Clinical Pharmacology & Therapeutics* 68.3 (2000), pp. 270–279. DOI: 10.1067/mcp.2000.109352.

[98] U.S. Food and Drug Administration. *FDA Adverse Event Reporting System (FAERS) Public Dashboard*. Center for Drug Evaluation and Research. URL: `https://fis.fda.gov/sense/app/95239e26-e0be-42d9-a960-9a5f7f1c25ee/sheet/7a47a261-d58b-4203-a8aa-6d3021737452/state/analysis` (visited on 08/08/2026).

[99] U.S. Food and Drug Administration. *Manufacturer and User Facility Device Experience (MAUDE) Database*. Center for Devices and Radiological Health. URL: `https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm` (visited on 08/08/2026).

[100] U.S. Food and Drug Administration. *Vaccine Adverse Events*. Center for Biologics Evaluation and Research. URL: `https://www.fda.gov/vaccines-blood-biologics/report-problem-center-biologics-evaluation-research/vaccine-adverse-events` (visited on 08/08/2026).

[101] Kristen Vaccaro et al. "Contestability in algorithmic systems". In: *Companion Publication of the 2019 Conference on Computer Supported Cooperative Work and Social Computing*. 2019, pp. 523–527.

[102] Jean Ville. *Etude critique de la notion de collectif*. Gauthier-Villars Paris, 1939.

[103] Vladimir Vovk and Ruodu Wang. "E-values: Calibration, combination and applications". In: *The Annals of Statistics* 49.3 (2021), pp. 1736–1754.

[104] A. Wald. "Sequential Tests of Statistical Hypotheses". In: *The Annals of Mathematical Statistics* 16.2 (1945), pp. 117–186.

[105] Ian Waudby-Smith and Aaditya Ramdas. "Estimating means of bounded random variables by betting". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 86.1 (2024), pp. 1–27.

[106] Heather P Whitley and Wesley Lindsey. "Sex-based differences in drug activity". In: *American family physician* 80.11 (2009), pp. 1254–1258.

[107] Guy Witberg et al. "Myocarditis after Covid-19 vaccination in a large health care organization". In: *New England Journal of Medicine* 385.23 (2021), pp. 2132–2139.

[108] Ziwei Wu and Jingrui He. "Fairness-aware model-agnostic positive and unlabeled learning". In: *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*. 2022, pp. 1698–1708.

[109] Yanxun Xu et al. "Subgroup-Based Adaptive (SUBA) Designs for Multi-arm Biomarker Trials". In: *Statistics in Biosciences* 8 (Feb. 2014).

[110] Kelly Zhang, Lucas Janson, and Susan Murphy. "Inference for Batched Bandits". In: *Advances in Neural Information Processing Systems*. Vol. 33. Curran Associates, Inc., 2020, pp. 9818–9829.

[111] Kelly Zhang, Lucas Janson, and Susan Murphy. "Statistical Inference with M-Estimators on Adaptively Collected Data". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, 7460=–7471.

[112] Lanju Zhang and William F. Rosenberger. "Response-Adaptive Randomization for Clinical Trials with Continuous Outcomes". In: *Biometrics* 62.2 (2006), pp. 562–569. (Visited on 09/28/2023).

[113] Martin Zinkevich. "Online convex programming and generalized infinitesimal gradient ascent". In: *Proceedings of the 20th International Conference on Machine Learning*. 2003, pp. 928–936.