**Submitted to *Manufacturing & Service Operations Management***
**manuscript MSOM-19-273.R2**

# Flexible Drug Approval Policies

Fernanda Bravo, Taylor C. Corcoran, Elisa F. Long
UCLA Anderson School of Management, Los Angeles CA 90095
fernanda.bravo@anderson.ucla.edu, taylor.corcoran.1@anderson.ucla.edu, elisa.long@anderson.ucla.edu

**Problem Definition**: To approve a novel drug therapy, the U.S. Food and Drug Administration (FDA) requires clinical trial evidence demonstrating efficacy with 2.5% statistical significance, although the agency often uses regulatory discretion when interpreting these standards. Factors including disease severity, prevalence, and availability of existing therapies are qualitatively considered, yet current guidelines fail to systematically consider such characteristics in approval decisions.

**Practical Relevance**: In making approval decisions, the FDA weighs the health benefits of introducing life-saving therapies against the potential risks of approving ineffective or harmful drugs. Tailoring approval standards to individual diseases could improve treatment options for patients with few alternatives, and further incentivize pharmaceutical companies to invest in neglected diseases.

**Methodology**: Using a novel queueing framework, we analyze the FDA's drug approval process to incorporate disease-specific factors and obsolescence—newer drugs replacing older formulas—through a set of pre-emptive $M/M/1/1$ queues. Based on public data encompassing all registered U.S. clinical trials and FDA-approved drugs, we estimate model parameters for three high-burden diseases (breast cancer, HIV, and hypertension) and solve for the optimal policy to maximize net life-years gained following FDA approval.

**Results**: The optimal policy relaxes approval standards for diseases with long trial duration, high attrition, or low R&D intensity. Results indicate that a more lenient policy is warranted for drugs targeting breast cancer or hypertension, and a more stringent policy is recommended for HIV, relative to the FDA's existing policy. If pharmaceutical firms respond to the new standards by submitting more drugs for approval—leading to an endogenous clinical trial initiation rate—the FDA's optimal policy modestly decreases for breast cancer and hypertension, with minimal change for HIV.

**Managerial Implications**: Our study offers a transparent, quantitative framework that could help the FDA develop disease-specific approval guidelines based on underlying disease-related severity, prevalence, and characteristics of the drug development process and existing market.

*Key words*: Non-Profit Management, Public Policy, Health Care Management
*History*: Submitted July 17, 2019; Revised April 4, 2020; Revised August 14, 2020; Accepted September 19, 2020.

## 1. Introduction

Since its establishment in 1906, the U.S. Food and Drug Administration (FDA) has approved over 1,500 novel therapies (Kinch et al. 2014), with many advances in recent years including 107 new drugs in 2018-19 (FDA 2020d). FDA decision-making is guided by one central tension: balancing the benefits of providing sick patients with effective therapies, against the potential danger to

consumers from taking ineffective or harmful drugs (FDA 2018). Despite undergoing rigorous evaluation, approved drugs are occasionally later found to be ineffective or detrimental to patients. One study of 108 drug withdrawals found that half were shown to be ineffective for the FDA-approved indication (Ahmed et al. 2002). In 2011, for instance, the FDA revoked Genentech's blockbuster drug Avastin® for breast cancer treatment, three years after initial approval, as evidence emerged that it was less effective at suppressing tumor growth than previously found (Pollack 2011). The drug maintains FDA-approval for treating lung, colon, and kidney cancers, permitting physicians to prescribe it off-label for breast cancer patients.

In this work, we develop a novel queueing framework to study the drug approval process, beginning with compound development through evaluation, FDA approval or rejection, and obsolescence or market expiry. Our approach can proffer insights for the agency's decision-making, by permitting flexible approval standards based on underlying disease *severity*—the impact on life expectancy; *prevalence*—the number of individuals afflicted; intensity and speed of research and development (R&D); and the number of alternative treatments available. Here, a *drug* refers to a substance intended to diagnose, cure, treat, or prevent disease; we use this synonymously with the terms medication, therapy, or compound. We do not consider FDA-regulated medical devices.

Current FDA policy requires pharmaceutical firms to demonstrate a candidate drug's *safety*, by displaying no evidence of adverse effects, and *efficacy*, by showing improvement in a health outcome related to the target condition, which are both established through a series of clinical trials. The existing approval criteria focus on minimizing a *type-I error*—approving an ineffective or harmful drug—by setting a tolerable level known as the *significance level*, $\alpha$. In their deliberations, the FDA can consider other factors including a benefit-risk assessment, but these are weighed qualitatively:

> *"For a drug to be approved for marketing, FDA must determine that the drug is effective and that its expected benefits outweigh its potential risks to patients. This assessment is informed by an extensive body of evidence about the drug's safety and efficacy... [and by] the severity of the underlying condition and how well patients' medical needs are addressed by currently available therapies..."* (FDA 2018)

Extant FDA guidelines recommend a constant threshold of $\alpha = 2.5\%$ for all diseases (FDA 2017a). Although this policy prioritizes diseases equally by holding all drugs to the same standard, no compelling rationale exists for this choice of $\alpha$ (Sterne and Smith 2001). It also ignores clinical trial nuances (e.g., rate of new compound discovery, trial duration, rate of attrition), the target population's underlying disease prevalence and severity, and the post-approval market (e.g., availability of other drugs, remaining time on patent). Notably, the FDA does not provide guidance on avoiding a *type-II error*—rejecting an effective drug—nor does this explicitly appear in their decision objective. Carpenter (2004) remarks that "*the FDA has often been excoriated for approving a*

*bad drug (or approving it too quickly) and only recently has been criticized for approving drugs too slowly.*" These criticisms—and the FDA's response in creating accelerated pathways—focus on the *speed* of approval, not the statistical burden of proof for failing to approve a drug.

Despite their aims, the FDA is often accused of fostering opaque approval policies (Downing et al. 2014). An objective model, together with existing FDA analyses, could improve transparency. The contributions of this paper are as follows:

- We develop a modeling framework to analyze FDA drug approval decisions, accounting for disease severity and prevalence, R&D intensity, trial duration and likelihood of completion, and the availability of alternative treatments. The drug development process is a series of $M/M/\infty$ queues and the post-approval market as a collection of pre-emptive $M/M/1/1$ and $M/M/\infty$ queues. To our knowledge, our study is the first to model the drug approval process as a network of queues.
- Assuming the FDA as primary decision-maker, we solve for the optimal approval threshold, by disease, to maximize expected health benefits. By adjusting the significance level by disease, our study highlights the relative importance of each disease characteristic on approval decisions. The optimal threshold increases (i.e., easier to approve) for diseases with lengthy clinical trials, high rates of attrition, and low R&D intensity. We analyze a model extension where R&D intensity is endogenously determined in response to a changing $\alpha$, as firms put forth more drug candidates for consideration, such that the net rate of *effective* drugs submitted for FDA review is unchanged.
- We construct a dataset comprising all registered clinical trials and FDA drug approvals, to illustrate our approach for three high-burden diseases: breast cancer, HIV, and hypertension. We show how the optimal significance level relates to characteristics of the development process and post-approval market. Numeric results show that a uniform approval policy is sub-optimal, highlighting the potential health gains of considering both pre- and post-approval disease characteristics.

## 2. Related Literature

Extending beyond the more traditional hospital-centered (Armony et al. 2015) or disease-specific applications of OM methodologies (Keskinocak and Savva 2020), healthcare operations management 2.0 has embraced a new ecosystem taxonomy consisting of *healthcare delivery*, *financing*, *innovation*, and *policymaking* (Dai and Tayur 2019). These authors highlight the ongoing efforts to investigate complex, interconnected decision-making by multiple stakeholders who often have competing priorities. In particular, our work presents a novel queueing-theoretic approach to optimize drug approval standards, a process at the intersection of healthcare *innovation* and *policymaking*.

**Innovation Incentives.** With so many stakeholders in the healthcare sector, regulatory incentives can help close market inefficiencies between manufacturers (e.g., pharmaceutical, medical

devices, diagnostic testing), service providers (e.g., hospitals, skilled nursing facilities, accountable care organizations (ACOs)), public health entities (e.g., state and local governments, non-governmental organizations, public-private partnerships), payers (e.g., insurers) and patients.

One mechanism to induce R&D effort is an advance market commitment (AMC), which guarantees a post-approval subsidy to drug manufacturers. Motivating investment in manufacturing capacity is critical for drugs further in development, as demonstrated by the AMC for a pneumococcal conjugate vaccine (Kremer et al. 2020). Supply capacity is a key bottleneck for manufacturers facing high yield (production) uncertainty, especially if regulators subsidize firms with more reliable production (Deo and Corbett 2009). Expected delays in vaccine delivery generate a negative feedback loop whereby purchasers reduce order quantities, dissuading manufacturers from investing in innovative production processes (Dai et al. 2016). Breaking this cycle may be achieved with contracting between manufacturers and providers, as one alternative to a government intervention. Insufficient R&D investment is compounded by a misallocation of therapies, particularly during high-demand periods like a global pandemic. Mamani et al. (2013) show that coordinating influenza vaccination between high- and low-risk countries can align incentives of the participating regions.

A wide literature has applied contracting models in healthcare settings including drug reimbursements (So and Tang 2000) and outpatient medical services (Jiang et al. 2012). More recently, performance-based contracting to incentivize high-quality, low-cost care was investigated by Aswani et al. (2019) following the launch of the Medicare Shared Savings Program, which required healthcare providers to join an ACO, a new model of integrated care delivery. Adida and Bravo (2019) also examine contracting models between tertiary care providers and referring physicians within an ACO setting. Innovative payment models, such as outcome-based pricing for drug therapies (Adida 2020), hold much promise as the alarmingly high cost of healthcare has prompted new solutions promoting "value-based care". In contrast to a direct monetary incentive or a contractual agreement, in the current paper, we focus on a varying regulatory policy to increase the availability of effective therapies.

**Clinical Trial Design.** The requisite sequence of clinical trials, although clinically necessary to establish drug safety and efficacy, is expensive and typically fails to produce a marketable therapy: only 7.5% of novel compounds that initiate clinical trials ultimately gain FDA approval (Hay et al. 2014). Several OM papers focus on optimal trial design to shorten the duration or minimize the number of volunteers exposed to a potentially unsafe drug. Deciding whether to continue, intensify enrollment, or completely abandon a Phase III trial—the most expensive segment of the clinical trial cascade—is a difficult choice faced by all pharmaceutical firms. Kouvelis et al. (2017) develop a dynamic policy, both with and without interim data on drug efficacy. Combining trial data on primary outcomes with surrogate endpoints and applying a Bayesian adaptive trial design, Anderer

et al. (2019) find a 5% efficiency gain through setting the optimal trial enrollment and stopping criteria. Chick et al. (2018) also use a Bayesian, decision-theoretic framework to design multi-arm, multi-stage trials that permit dynamic patient allocation decisions based on prior observations. Ahuja and Birge (2016) adjust randomization probabilities so that patients are treated as effectively as possible without compromising the ability to learn about efficacy. Bertsimas et al. (2015) use discrete linear optimization to construct treatment groups for small samples, allowing for more powerful statistical inference. Small-sample trial design is important for ethical reasons, but also logistically, as recruiting a large number of volunteers with a rare disease is challenging.

Other recent studies leverage existing clinical trial data to identify novel drug combinations or patient groups to target. Bertsimas et al. (2016) use machine learning to predict chemotherapy outcomes in cancer patients and suggest new drug combinations. Gupta et al. (2018) use robust optimization to identify patient subpopulations to maximize the effectiveness of an intervention. We do not explicitly model clinical trial design, but instead analyze how disease specifics drive the optimal significance level, assuming a standard balanced two-arm design.

**FDA Decision-Making.** The FDA recommends that drugs are tested against an active comparator or placebo in two randomized, double-blind trials, yet more than 60% of recently approved drugs were evaluated on the basis of a single trial, 10% of trials were not randomized, 20% were not double-blind, and 12% did not use a comparator or placebo (Downing et al. 2014). While this demonstrates flexibility in considering a range of trial evidence, it obfuscates the agency's approval criteria. Moreover, the FDA's existing expedited approval programs require manufacturers to submit post-approval outcome data demonstrating longer-term efficacy, yet only one-third of drugs are compliant, highlighting the need for other regulatory incentives (Xu et al. 2020).

Despite the importance of bringing new, effective drugs to market—and the potentially life-threatening consequences of failing to do so—few studies analyze the FDA's decision-making process. Our study relates to two papers by the same authors, who apply Bayesian Decision Analysis to jointly optimize the FDA approval policy $\alpha$, and the number of trial participants $n$. Montazerhodjat et al. (2017) minimize the expected harms associated with enrolling $n$ patients in a trial, including the costs of treatment with a toxic drug (a type-I error) or missing treatment with an effective therapy (a type-II error), as well as the post-approval impact on disability-adjusted life expectancy. In a related paper, Isakov et al. (2017) apply a similar cost-minimization framework to determine the optimal $\alpha$ and trial enrollment for 30 high-burden diseases. Our approach differs in several ways, most notably in our objective of maximizing expected net health outcomes—including the benefits of approving effective drugs—based on our assumption that the FDA is the sole decision-maker and only sets the approval policy $\alpha$. We highlight other key differences in Sections 5.1 and 6.

## 3. Drug Development Overview

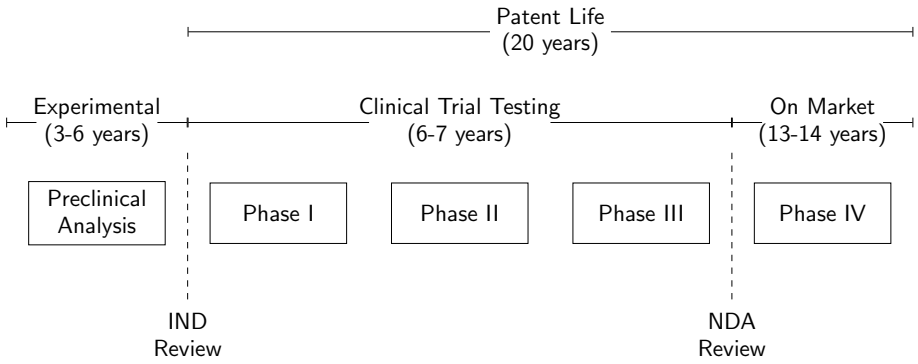
The drug approval process in the U.S. consists of a series of stages, beginning with the discovery of a new pharmaceutical compound and ending with the FDA deciding whether to grant marketing approval. See Figure 1 for a summary and average duration of each stage (PhRMA 2015).

### 3.1. Clinical Trial Testing

A novel drug compound initially undergoes preclinical analysis involving laboratory and animal testing, to screen for potential safety issues and study how the body metabolizes it (*pharmacokinetics*). If no safety concerns arise, the sponsoring firm can file an Investigational New Drug (IND) application to the FDA presenting a plan for clinical trial testing, which may begin after 30 days if the FDA raises no objections. Clinical trials typically consist of three phases designed to test safety and efficacy. Phase I trials enroll healthy volunteers to observe any side effects and pharmacokinetics. Phase II drugs are administered to volunteers diagnosed with the target illness, to establish efficacy and continue monitoring side effects. Phase III aims to establish efficacy in a large patient cohort, and to assess interactions with other drugs, responses in different subpopulations, and dosage levels. Randomized controlled trials (RCTs) are the gold standard for establishing efficacy. A common design, the two-arm balanced RCT randomly assigns participants to a *treatment* or *control* group, which are equal in size. Individuals in the treatment arm receive the experimental regimen; those in the control arm receive standard therapy or a placebo. Before the trial begins, researchers propose one or more *endpoints*—outcomes that measure direct clinical benefit—associated with the target disease that will be monitored throughout the study.

At any point in this process, the sponsoring firm may halt development for varying reasons including an inability to demonstrate efficacy, safety concerns, pharmacokinetic issues, market competition, and financial considerations (Arrowsmith and Miller 2013). After completing Phase

**Figure 1 The FDA drug development and approval process.**




*Note.* For each new compound, the FDA reviews two applications submitted by the pharmaceutical company: an IND (Investigational New Drug) and an NDA (New Drug Application).

III, a firm can submit a New Drug Application (NDA) to the FDA, consisting of trial results and a proposal for manufacturing and labeling the drug. The FDA evaluates candidate drugs using two criteria: *safety* is measured by the number and type of adverse events among trial participants, and *efficacy* is assessed by comparing the endpoint(s) in the treatment and control groups. The FDA can further request additional testing before granting marketing approval (FDA 2017b).

Drugs ultimately gaining FDA approval may be legally marketed in the U.S. and receive patenting and exclusivity rights. Patents are granted by the U.S. Patent and Trademark Office and typically expire 20 years after filing, which usually occurs before clinical trials begin. The FDA grants exclusive marketing rights to new drugs for five years upon approval. Safety and efficacy of approved drugs are continually monitored during post-marketing studies (Phase IV), with any drug-related adverse events reported to the FDA. In rare cases, drugs with harmful side effects are withdrawn from the market by the FDA or the sponsoring firm (FDA 2020a).

### 3.2. FDA Programs

Over the past four decades, the FDA has introduced programs to provide more regulatory discretion and address shortcomings of the approval process. To help offset the high costs of drug development and incentivize investment in understudied conditions, Congress passed the Orphan Drug Act of 1983, establishing tax credits and market exclusivity rights for drugs targeting rare diseases afflicting fewer than 200,000 Americans. In 2019, the FDA approved 21 new drugs with this designation (FDA 2020d), including the drug Inrebic® for treating myelofibrosis, a rare bone marrow cancer afflicting 18,000 people in the United States.

The FDA's Code of Federal Regulations addresses disease severity by stating that "*patients are generally willing to accept greater risks or side effects from products that treat life-threatening and severely-debilitating illnesses, than they would accept from products that treat less serious illnesses*" and "*the benefits of the drug need to be evaluated in light of the severity of the disease being treated*" (FDA 2019). For example, Lotronex, a drug treating irritable bowel syndrome, was voluntarily withdrawn from the market after many patients reported severe adverse reactions, but was re-approved by the FDA two years later, with restricted use, based on patient input (FDA 2020b).

Designed to address the protracted timeline required to complete clinical trial testing, lasting typically between ten and fifteen years (PhRMA 2015), the FDA has established four regulatory mechanisms: Fast Track, Accelerated Approval, Breakthrough Therapy, and Priority Review (FDA 2014). These programs benefit patients, who hopefully gain access to life-saving drugs more quickly, and pharmaceutical firms, who benefit financially from a shortened development timeline. In 2019, 60% of the 48 drugs receiving FDA-approval qualified for one or more of these programs, leaving room for improvement in the current approval process (FDA 2020d). In this paper, we explore an alternative regulatory policy: adjust the FDA's approval threshold based on disease-specific characteristics of the drug pipeline and post-approval market.

# 4. Model

We first present a standard framework for modeling drug effectiveness, drawn from the statistics literature. Next, we introduce a queueing network to model the drug development process from clinical trials to post-approval (Figure 2) followed by an analysis of the optimal approval policy and comparative statics. We extend the model to allow an endogenous clinical trial initiation rate, in response to a changing FDA approval policy. Model parameters are summarized in Table 2.

## 4.1. Drug Effectiveness

Consider a two-armed, balanced, non-adaptive clinical trial with equal enrollments in each arm. Let $X_1, \ldots, X_n$ denote independent observations of a single quantitative endpoint from $n$ patients in the treatment group, and let $Y_1, \ldots, Y_n$ denote independent observations from $n$ patients in the control group who receive standard therapy. Assume $X_i$ is drawn from a distribution with mean $\mu_X$ and variance $\sigma^2$, and $Y_i$ is drawn from a distribution with mean $\mu_Y$ and variance $\sigma^2$ (Jennison and Turnbull 2000). We assume equal variance for simplicity but this can be easily relaxed.

The quantity $\delta = \mu_X - \mu_Y$ represents the *treatment effect* of the drug under evaluation. Our analysis focuses on superiority trials, which assume that the experimental drug has no effect or a positive effect, compared to standard therapy. For each drug candidate undergoing clinical trial testing, we obtain the following one-sided hypothesis test:

$$H_0 : \delta = 0 \text{ (drug is ineffective)}$$
$$H_1 : \delta > 0 \text{ (drug is effective)}$$

The Wald statistic is computed from the observed trial data:

$$T = \frac{\left(\bar{X} - \bar{Y}\right)}{\sqrt{\frac{2\sigma^2}{n}}}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$ are the sample means. Letting $I_n = \frac{n}{2\sigma^2}$, which is known as the *information* of the sample, we rewrite this as: $T = \left(\bar{X} - \bar{Y}\right)\sqrt{I_n}$. By the Central Limit Theorem, $T$ is normally distributed with mean $\delta\sqrt{I_n}$ and variance 1. Let $z_\alpha = \Phi^{-1}(1-\alpha)$ be the critical value where $H_0$ is rejected, i.e., the drug is deemed effective. This occurs with probability

$$\mathbb{P}\left(Z > z_\alpha - \delta\sqrt{I_n}\right) = 1 - \Phi\left(\Phi^{-1}(1-\alpha) - \delta\sqrt{I_n}\right)$$

where $\Phi$ is the CDF of a standard normal random variable $Z$.

We define an *approval policy* corresponding to significance level $\alpha$ as follows: candidate drugs that complete clinical trials and undergo FDA review are approved if the $p$-value associated with the observed test statistic is less than $\alpha$, and rejected otherwise. Let $p$ be the probability that a

candidate drug is actually effective, conditional on undergoing FDA review. We obtain the following joint probability expressions:

$$\pi_{AE}(\alpha) = \left[1 - \Phi\left(\Phi^{-1}(1-\alpha) - \delta\sqrt{I_n}\right)\right] p \qquad \text{Approve effective (AE) drug} \tag{1}$$

$$\pi_{AI}(\alpha) = \alpha\,(1-p) \qquad \text{Approve ineffective (AI) drug} \tag{2}$$

$$\pi_{RE}(\alpha) = \Phi\left(\Phi^{-1}(1-\alpha) - \delta\sqrt{I_n}\right) p \qquad \text{Reject effective (RE) drug} \tag{3}$$

$$\pi_{RI}(\alpha) = (1-\alpha)\,(1-p) \qquad \text{Reject ineffective (RI) drug} \tag{4}$$

We can interpret Eqs. (1)-(4) by examining a simple numeric example (Table 1). As the theoretical treatment effect size of the candidate drug improves ($\delta$ increases) or the clinical trial enrollment $n$ increases, the term $\delta\sqrt{I_n}$ increases, resulting in a higher probability of approving an effective drug $\pi_{AE}(\alpha)$, and thus a lower probability of rejecting an effective drug $\pi_{RE}(\alpha)$. As the approval policy becomes more stringent ($\alpha$ decreases), the probability of approving an effective drug $\pi_{AE}(\alpha)$ or an ineffective drug $\pi_{AI}(\alpha)$ decreases, as expected.

We do not explicitly include drug safety given the multitude of possible adverse events, as the FDA notes: "*in typical safety assessments, there are often no prior hypotheses ... and numerous safety findings that would be of concern*" (FDA 2017a). In contrast, drug efficacy is assessed by one or more clinical endpoints, which must be objectively measured and specified before the trial commences. We assume that one quantitative *primary endpoint* needed to establish efficacy is monitored. Multiple endpoints may be used in reality, but these endpoints are often merged into a single combined endpoint (e.g., cardiovascular studies often consolidate cardiac death, heart attack, and stroke into a single compound endpoint). Finally, we assume that higher endpoint values correspond to better health outcomes, though a range of desirable values could exist.

**Table 1 Example approval policies and joint probabilities.**

| $\alpha$ | $\delta\sqrt{I_n}$ | $\pi_{AE}(\alpha)$ | $\pi_{AI}(\alpha)$ | $\pi_{RE}(\alpha)$ | $\pi_{RI}(\alpha)$ |
|---|---|---|---|---|---|
| 0.01 | 1 | 0.074 | 0.002 | 0.726 | 0.198 |
| 0.01 | 2 | 0.298 | 0.002 | 0.502 | 0.198 |
| 0.01 | 3 | 0.600 | 0.002 | 0.200 | 0.198 |
| 0.01 | 4 | 0.762 | 0.002 | 0.038 | 0.198 |
| 0.025 | 1 | 0.152 | 0.003 | 0.748 | 0.098 |
| 0.025 | 2 | 0.464 | 0.003 | 0.436 | 0.098 |
| 0.025 | 3 | 0.766 | 0.003 | 0.134 | 0.098 |
| 0.025 | 4 | 0.881 | 0.003 | 0.019 | 0.098 |
| 0.05 | 1 | 0.234 | 0.005 | 0.666 | 0.095 |
| 0.05 | 2 | 0.575 | 0.005 | 0.325 | 0.095 |
| 0.05 | 3 | 0.821 | 0.005 | 0.079 | 0.095 |
| 0.05 | 4 | 0.892 | 0.005 | 0.008 | 0.095 |

*Note.* The effectiveness probability is assumed to be $p = 0.9$.

### 4.2. Queueing Network Model

Assume that candidate drugs begin clinical trials according to a Poisson process with arrival rate $\lambda$. Combining the three clinical trial phases into a single queue simplifies the analyses and does not change key insights, as illustrated in Section 5.4. Drugs either complete clinical trial assessment, or the sponsoring firm halts the trials early. We model clinical trial duration as an exponential race between trial completion and abandonment with rates $\mu_{CT}$ and $\mu_{AB}$, supported by data from clinicaltrials.gov (see Appendix B for details). Drugs advance to FDA review with probability $\frac{\mu_{CT}}{\mu_{CT}+\mu_{AB}}$ or exit the system with probability $\frac{\mu_{AB}}{\mu_{CT}+\mu_{AB}}$. Thus, the NDA submission rate for FDA review is $\lambda_{NDA} = \lambda \frac{\mu_{CT}}{\mu_{CT}+\mu_{AB}}$. For simplicity, we assume trial completion and abandonment rates are identical across drug classes; a model extension could include parallel queues with class-specific rates. Modeling the clinical trial sequence as an $M/M/\infty$ queue captures three key elements, the initiation rate ($\lambda$), total duration ($1/\mu_{CT}$), and abandonment rate ($\mu_{AB}$), an advantage over using a single event for trial completion.

Following FDA review, a drug is approved if the $p$-value associated with the clinical trial demonstrating efficacy is less than the significance level $\alpha$. Drugs denied approval immediately exit the system. We assume the FDA's decision is instantaneous, although in reality, review lasts between six months and two years. Accounting for this delay would entail modeling the review stage as an $M/M/\infty$ queue, but would not substantially change our results. In steady state, the output of the FDA review stage constitutes a thinning of a Poisson process with the following arrival rates:

$$\lambda_{AE}(\alpha) = \lambda_{NDA}\pi_{AE}(\alpha), \quad \lambda_{AI}(\alpha) = \lambda_{NDA}\pi_{AI}(\alpha), \tag{5}$$
$$\lambda_{RE}(\alpha) = \lambda_{NDA}\pi_{RE}(\alpha), \quad \lambda_{RI}(\alpha) = \lambda_{NDA}\pi_{RI}(\alpha).$$

After gaining FDA-approval, effective and ineffective drugs are modeled separately as follows. *Ineffective* drugs join an $M/M/\infty$ queue, where "service" represents time until market withdrawal

**Figure 2** **Queueing network representing the drug development and approval process.**

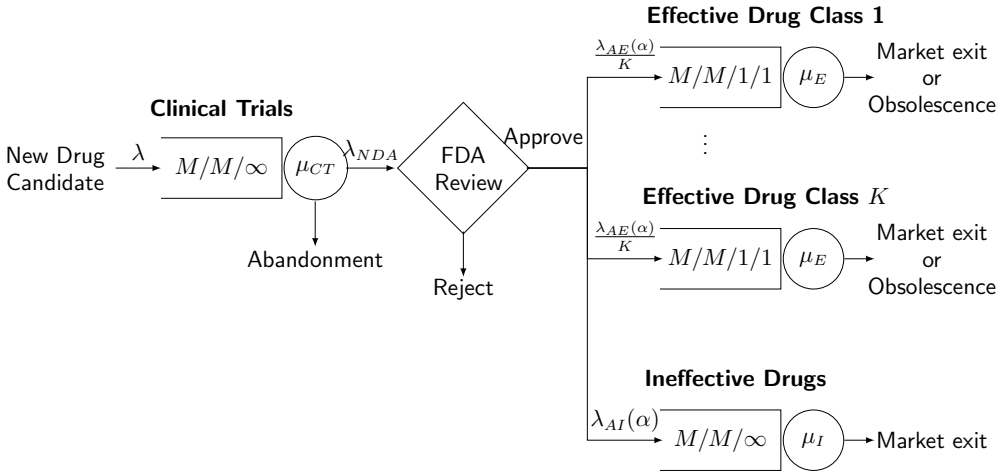


**Table 2 Summary of key model parameters.**

| Before FDA review | | After FDA review | |
|---|---|---|---|
| $\lambda$ | Phase I clinical trial initiation rate | $\alpha$ | Statistical significance level for FDA approval |
| $\mu_{CT}$ | Clinical trial completion rate | $p$ | Probability a drug is effective conditional on FDA review |
| $\mu_{AB}$ | Clinical trial abandonment rate | $K$ | Number of unique drug classes on the market |
| $\lambda_{NDA}$ | NDA submission rate for FDA review | $Q_E$ | Health benefits per effective drug on the market |
| $n$ | Clinical trial enrollment | $Q_I$ | Health harms per ineffective drug on the market |
| $\delta$ | Treatment effect size of a candidate drug | $1/\mu_E$ | Average market life of an effective drug |
| $\sigma$ | Standard deviation of treatment response | $1/\mu_I$ | Average market life of an ineffective drug |

(mean $1/\mu_I$), which is typically shorter in duration, as patients discontinue use or the drug is withdrawn by the FDA or manufacturer. Historically, few FDA-approved drugs are subsequently shown to be ineffective, and it is reasonable to expect these drugs to exit the market before patent expiry, as witnessed among each of the three diseases considered in Section 5.

*Effective* drugs comprise the vast majority of FDA-approved drugs. To capture the influx of new therapies and discontinuation of older therapies, we employ a collection of $K$ parallel pre-emptive $M/M/1/1$ queues, each representing a unique therapeutic class for a disease. A drug joins any of the $K$ queues with equal probability, implying that drug-classes capture equal market share, although we relax this assumption via simulation in Section 5.4. Here, "service" completion represents a drug exiting the market (mean $1/\mu_E$) either due to patent expiry or discontinued use, which may result in an empty queue for that particular class. This can be interpreted as a specific therapeutic class becoming outdated due to lack of recent innovation. For example, first-line treatment of HIV with a single Nucleoside Reverse Transcriptase Inhibitor is no longer standard of care. The most recent drug within the class was approved in 2003 (Table B9), yet these drugs are now typically combined with other drug classes as part of combination antiretroviral therapy.

A single server per drug class is assumed, reflecting the high market concentration within a class as a handful of drugs typically account for the majority of prescriptions. With hypertension, for example, the top five medications by market share belong to five different drug classes (ACE inhibitors, beta blockers, calcium channel blockers, diuretics, and angiotensin receptor blockers) and collectively account for more than 50% of the total market (Express Scripts Holding Company 2017). If a class contains two or more comparable drugs, market share is divided, but the net benefit to patients remains largely unchanged. Finally, service pre-emption is designed to account for older drugs becoming obsolete as newer therapies gain approval, often decades later.

The choice to model effective and ineffective drugs as separate queues is based on the historical observation that ineffective drugs rarely make it to the market and when they do, they are typically withdrawn long before patent expiry or obsolescence (Table B6). Thus, our model allows us to capture the main event triggering a market exit: an AE drug replaces another AE drug due to obsolescence. We note that granting additional ineffective drugs FDA-approval does indeed alter

the rate of effective drugs entering the market, via the $\lambda_{AE}(\alpha)$ and $\lambda_{AI}(\alpha)$ expressions (Eqs. 1-5), which effectively allows some market cannibalization.

Assuming the FDA is the sole decision-maker, we focus on their choice of significance level $\alpha$ given a fixed trial sample size $n$. The FDA's choice to approve or reject a candidate drug is based on the expected health benefits and potential risks to patients (FDA 2018). This perspective is consistent with the FDA's stated emphasis on weighing the expected health benefits of approving a drug against the potential harms (FDA 2018). As the FDA does not explicitly consider the risk of committing a type-II error, we refrain from including this in our model. The tolerance for type-II error is adjusted through experimental design, such as changing the trial sample size or decreasing measurement error, and in practice, trial enrollment is decided by the pharmaceutical company, taking into account the costs and feasibility of patient recruitment (Casella and Berger 2002).

Consistent with population-wide health benefits increasing as additional effective treatments become available—or decreasing if ineffective drugs reach the market—we assign an average health benefit $Q_E$ per effective drug class available and an average health cost $Q_I$ per ineffective drug on the market. We measure health benefits in life-years gained, which capture the additional number of years a person lives as a result of receiving treatment. The cost of an approved ineffective drug—a type-I error—is similarly measured in years of life lost, assuming a patient unknowingly receives an ineffective drug. An alternative measure that one may consider is quality adjusted life years (QALYs), which account for both length and quality of life. The unrealized value of rejecting a candidate drug is normalized to zero. For tractability, we analyze the system in steady state with time invariant parameters.

The optimal approval policy $\alpha^*$ is chosen to maximize the expected net benefit $V(\alpha)$:

$$\alpha^* = \underset{\alpha\in[0,1]}{\arg\max}\ V(\alpha) := Q_E\mathbb{E}[N_E(\alpha)] - Q_I\mathbb{E}[N_I(\alpha)], \tag{6}$$

where the expected number of effective and ineffective drugs are given by

$$\mathbb{E}[N_E(\alpha)] = \frac{K\lambda_{AE}(\alpha)}{K\mu_E + \lambda_{AE}(\alpha)}, \qquad \mathbb{E}[N_I(\alpha)] = \frac{\lambda_{AI}(\alpha)}{\mu_I}. \tag{7}$$

### 4.3. Model Analysis

We examine the structure of the optimal policy to gain insights into how the pre- and post-review characteristics affect the FDA's ultimate approval decision. All proofs are given in Appendix A.

The following result shows that the optimal significance level $\alpha^*$ is unique and is the solution to a non-linear equation.

Proposition 1. *The expected net benefit function $V(\alpha)$ is concave in $\alpha$, and the optimal policy $\alpha^*$ satisfies the following first order condition:*

$$\alpha^* = 1 - \Phi\left(\frac{1}{\delta\sqrt{I_n}}\log\left(\frac{1-p}{p}\,\frac{Q_I/\mu_I}{Q_E/\mu_E}\,\left(1+\frac{\lambda_{AE}(\alpha^*)}{K\mu_E}\right)^2\right) + \frac{\delta\sqrt{I_n}}{2}\right). \tag{8}$$

Proposition 1 demonstrates that the optimal approval policy, $\alpha^*$, weighs the steady-state health costs of approving ineffective drugs against the health benefits of approving effective drugs. Although no closed form expression for the optimal policy exists, we examine the comparative statics of $\alpha^*$ using first-order analysis.

COROLLARY 1. *The optimal approval policy $\alpha^*$ is*

*(a) increasing in $Q_E$, $\mu_I$, and $\mu_{AB}$,*

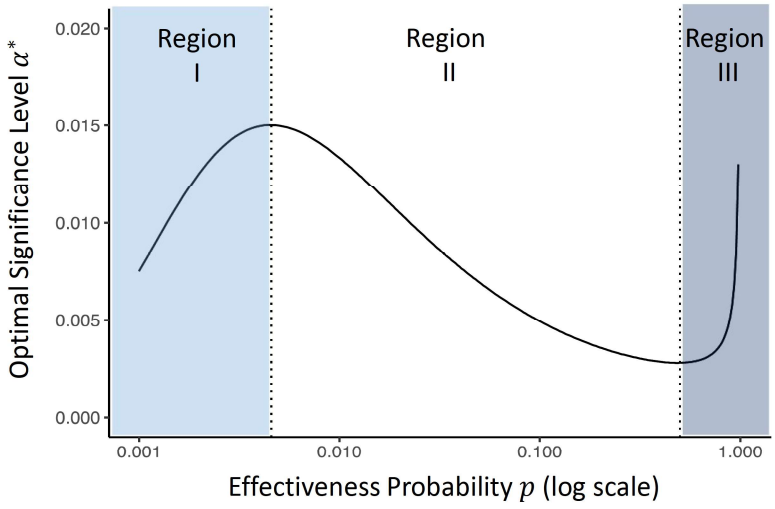
*(b) decreasing in $Q_I$, $\lambda$, and $\mu_{CT}$,*

*(c) increasing in $p$ and decreasing in $\mu_E$ if $\frac{\lambda_{AE}(\alpha^*)}{K\mu_E} < 1$.*

Corollary 1 indicates that the optimal approval policy is more stringent (lower $\alpha^*$) for diseases with many compounds in development (large $\lambda$) or short trial duration (large $\mu_{CT}$), but less stringent for those with high attrition rates (large $\mu_{AB}$). As expected, drugs with greater health benefits $Q_E$ have easier approval policies compared to those with higher health costs $Q_I$. Prolonging the time that ineffective drugs spend on the market $1/\mu_I$ increases patient harm, thus discouraging FDA approval. As the probability $p$ of effectiveness increases, or as the average time spent on the market $1/\mu_E$ increases, one might expect that approving *more* drugs is optimal. However, this intuition only holds if $\lambda_{AE}(\alpha^*)/K\mu_E < 1$; in other words, if the per-class new drug approval rate $\lambda_{AE}(\alpha^*)/K$ is less than the market exit rate $\mu_E$. Since we model the market as a collection of pre-emptive $M/M/1/1$ queues, this condition is not needed for stability; rather it serves to identify crowding in the market.

The relationship between market crowding and non-monotonicity of the optimal policy, holding all other parameters constant, is illustrated with the following simple example (Figure 3). Consider a

**Figure 3 Sensitivity of the optimal approval policy level $\alpha^*$ if condition in Corollary 1c is not satisfied.**




*Note.* $\sigma = 1$, $\delta = 0.10$, $n = 500$, $\lambda_{NDA} = 8$, $K = 1$, $Q_E = 1$, $Q_I = 0.1$, $\mu_E = 0.01$, $\mu_I = 0.10$. Region I corresponds to $0 \le p \le 0.005$, Region II to $0.005 < p \le 0.5$, and Region III to $0.5 < p \le 1$.

disease with a high rate of R&D intensity $\lambda$, and high health benefits associated with effective drugs $Q_E$ relative to the health cost of ineffective drugs $Q_I$. Drugs with a low effectiveness probability ($p < 0.5$) are defined as *long shots*, and those with high effectiveness probability ($p \geq 0.5$) are *safe bets*. A market is considered *crowded* if many available therapies are available ($\mathbb{E}[N_E(\alpha)] \approx K$) and *neglected* if few are available ($\mathbb{E}[N_E(\alpha)] << K$).

We divide Figure 3 into three regions. Region I corresponds to diseases with neglected markets and long-shot drugs. Here, as the effectiveness probability $p$ increases, the optimal policy approves *more* drugs given the limited availability of effective drugs on the market. Region II consists of long-shot drugs but a crowded market, so the potential costs of type-I error outweigh the marginal benefits of additional effective drugs as alternative therapies are available. Thus, as $p$ increases, the optimal policy approves *fewer* drugs. Finally, in Region III, the market is crowded but candidate drugs are effective with high probability, so each new approval generates a positive expected health benefit. Hence, the optimal policy in this region is to approve *more* drugs as $p$ increases.

The optimal policy $\alpha^*$ is similarly decreasing in $\mu_E$ if $\lambda_{AE}(\alpha^*)/K\mu_E < 1$, i.e., when the market is *not* crowded. In this case, as drugs spend more time on the market (higher $1/\mu_E$), the optimal policy sets more lenient approval standards to fill out the market, as these drugs are years away from obsolescence. If the market is crowded, however, this relationship is not guaranteed, and relaxing approval standards can actually trigger more drugs to exit the market to make room for newer therapies. As a result, the marginal gain in health benefits is minimal—newer therapies just displace slightly older ones—yet health costs increase due to type-I errors.

Our analysis thus far assumes a fixed number of drug classes $K$ for treating a particular disease. Increasing $K$—interpreted as approving a *first-in-class* drug with a new mechanism of action for treatment—alters the optimal policy and expected net benefits (Corollary 2). Let $\alpha_j^*$ denote the optimal policy and let $V_j^*$ denote the optimal expected net benefit when $j$ drug classes are available.

Corollary 2. *The optimal approval policies satisfy*

$$\alpha_1^* \leq \cdots \leq \alpha_K^* \leq \cdots \leq \alpha_\infty^*$$

*where* $\alpha_\infty^* = 1 - \Phi\left(\frac{1}{\delta\sqrt{I_n}} \log\left(\frac{1-p}{p} \frac{Q_I/\mu_I}{Q_E/\mu_E}\right) + \frac{\delta\sqrt{I_n}}{2}\right)$. *The optimal expected net benefit functions satisfy*

$$V_1^* \leq \cdots \leq V_K^* \leq \cdots \leq V_\infty^*,$$

*and, for any* $\alpha$ *and* $K \geq 1$,

$$V_{K+1}(\alpha) - V_K(\alpha) > V_{K+2}(\alpha) - V_{K+1}(\alpha).$$

The optimal policy is non-decreasing in the number of drug classes $K$, an intuitive result. As $K$ increases, less drug substitution and obsolescence exist (i.e., the market can sustain multiple

therapy classes) and thus the optimal policy is to ease approval standards to fill the market and obtain all the associated benefits. Further, among diseases with few available therapy classes, increasing $K$ with a first-in-class drug approval produces larger expected gains than for diseases with many existing drug classes. In other words, spurring innovation in drug development by easing approval standards is particularly beneficial for diseases with few unique drug classes.

### 4.4. Endogenous Queueing Model

We extend the base model to account for potentially endogenous effects of varying FDA approval standards on the volume and efficacy of new drugs under development. We include first-order changes, that is, we model the rate at which firms initiate Phase I clinical trials as $\tilde{\lambda}(\alpha) = \lambda \, \frac{1+\alpha}{1+\alpha_o}$ and the effectiveness probability $\tilde{p}(\alpha) = p \, \frac{1+\alpha_o}{1+\alpha}$, where $\alpha_o = 2.5\%$, the current approval policy. We assume that more relaxed approval standards will increase the volume of drugs starting clinical trials, but this increase is accompanied by a reduction in quality, so that the net NDA submission rate of *effective* drugs is unchanged, i.e., $\tilde{p}(\alpha)\tilde{\lambda}_{NDA}(\alpha) = p\lambda_{NDA}$. This assumption maintains the effective drug development output, reflecting pharmaceutical firms' limitations in identifying new effective compounds. The $(1+\alpha_o)$ normalization ensures that the expected benefits under a 2.5% approval policy in the endogenous and base model are equal. Albeit a simplified model, this captures how the optimal policy might shift given pharmaceutical firms' potentially endogenous response to such changes in approval standards. We reserve the tilde notation for parameters that depend on $\tilde{\lambda}(\alpha)$ or $\tilde{p}(\alpha)$. Thus, the endogenous objective function can be written as:

$$\tilde{V}(\alpha) = V(\alpha) + \frac{Q_I}{\mu_I} \, \frac{\lambda_{NDA}}{1+\alpha_o} \, (\alpha_o - \alpha) \tag{9}$$

and the optimal approval policy $\tilde{\alpha}^* = \arg\max_{\alpha \in [0,1]} \tilde{V}(\alpha)$.

Proposition 2. *The endogenous optimal approval policy satisfies* $\tilde{\alpha}^* \leq \alpha^*$, *and*
*(a)* $\tilde{V}^* \geq V^*$ *if* $\tilde{\alpha}^* \leq \alpha^* \leq \alpha_o$,
*(b)* $\tilde{V}^* \leq V^*$ *if* $\alpha_o \leq \tilde{\alpha}^* \leq \alpha^*$.

Accounting for potential endogenous effects in the volume of drugs starting clinical trials—and their decreased likelihood of being effective—results in a more conservative FDA's approval policy, a direct consequence of the higher rate of *ineffective* drugs $\tilde{\lambda}_{NDA}(\alpha)\,[1-\tilde{p}(\alpha)]$ being submitted for review. A more stringent approval policy $\tilde{\alpha}^*$ ensures that, as more drugs go up for review, the FDA continues to weigh the costs and benefits of approving ineffective and effective drugs. Proposition 2(a) indicates that, in some cases, the net benefits can even improve when endogenous effects are included, as we observe in the numerical study for HIV (Section 5.4).

## 5. Numerical Study

Using publicly available drug approval data, we conduct numerical analyses for three high-burden diseases (breast cancer, HIV, and hypertension), which collectively account for 10% of all drugs in development (PhRMA 2016). For each disease, we compute the optimal approval policy and compare the associated net benefits to a traditional policy of $\alpha = 2.5\%$. In sensitivity analysis, we examine how characteristics of the drug development process affect the optimal policy. We also quantify the impact of endogenous R&D intensity on the optimal policy and net benefits and conduct various robustness checks to test key model assumptions. A detailed description of our parameter estimates (Table 3) and all sources are provided in Appendix B.

### 5.1. Case Study: Breast Cancer, HIV, and Hypertension

Each year, more than 250,000 women in the U.S. are diagnosed with breast cancer and more than 40,000 die of the disease. Primary treatment consists of surgery, radiation, and/or chemotherapy and is typically completed within a year of diagnosis (Breast Cancer Society 2020). Additional hormone or targeted therapies may be prescribed for several years after primary treatment to reduce recurrence risk. Women with metastatic breast cancer may take some form of oncological therapy for the remainder of their lives. More than 1.1 million people in the U.S. are currently living with Human Immunodeficiency Virus (HIV) and more than 16,000 die each year (CDC 2020). HIV-infected individuals are prescribed antiretroviral therapy, which suppresses viral load in the body, slows disease progression, and substantially prolongs life. Chronic hypertension afflicts 106 million people in the U.S. and is a precursor for heart disease, which kills nearly 650,000 people every year, amounting to one out of every four deaths (CDC 2019). Individuals diagnosed with hypertension typically take medications to control their blood pressure throughout their life.

Significant heterogeneity exists in the R&D pipeline across diseases (Table 3). Many candidate drugs targeting breast cancer begin Phase I clinical trial testing each year ($\lambda = 30.6$), but long trial durations ($1/\mu_{CT} = 11.6$ years over Phases I-III), compounded by high attrition rates, result

**Table 3 Parameter estimates for selected diseases.**

| Parameter | Breast Cancer | HIV | Hypertension | Source |
|---|---|---|---|---|
| $\lambda$ (yearly rate) | 30.6 | 17.8 | 13.1 | National Institutes of Health (2020) |
| $\mu_{CT}$ (yearly rate) | 0.09 | 0.15 | 0.20 | National Institutes of Health (2020) |
| $\mu_{AB}$ (yearly rate) | 1.64 | 0.56 | 2.28 | Thomas et al. (2016), Wong et al. (2019) |
| $\lambda_{NDA}$ (yearly rate) | 0.96 | 1.87 | 0.55 | Calculated |
| $\delta\sqrt{I_n}$ | 3.24 | 3.24 | 3.24 | Calculated |
| $p$ | 0.90 | 0.90 | 0.90 | Assumed |
| $K$ (classes) | 12 | 8 | 9 | NCI (2020), AIDSinfo (2020), FDA (2020c) |
| $Q_E/\mu_E$ (benefit per drug) | 95,687 | 2,980,160 | 102,607,000 | Calculated |
| $Q_I/\mu_I$ (harm per drug) | 199,242 | 6,205,370 | 213,650,000 | Calculated |

*Note.* See Appendix B for calculation details.

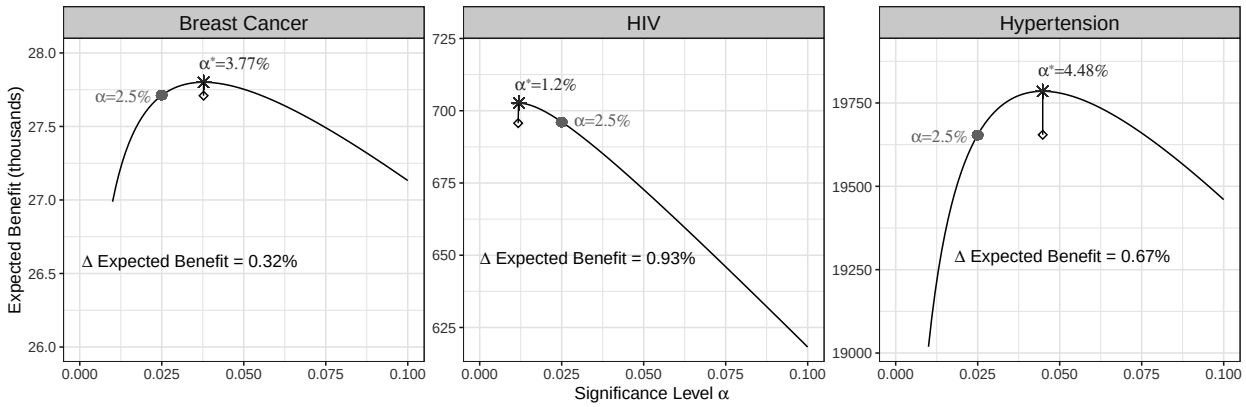
in a modest NDA submission rate of $\lambda_{NDA} = 0.96$ drugs per year. Although significantly fewer therapies targeting HIV initiate Phase I testing ($\lambda = 17.8$), the trials are shorter ($1/\mu_{CT} = 6.7$ years) with relatively low attrition, resulting in an NDA submission rate of $\lambda_{NDA} = 1.87$ drugs per year, nearly double that of breast cancer. Even fewer drugs for hypertension enter clinical trials each year ($\lambda = 13.1$), and these have the shortest average clinical trial duration ($1/\mu_{CT} = 5.0$ years), yet few drugs proceed to later stage testing, leading to the lowest NDA submission rate of the three diseases ($\lambda_{NDA} = 0.55$) drugs per year. We assume that the probability $p$ that a drug is effective, *conditional on undergoing FDA review*, equals 0.90, but we vary this in sensitivity analysis.

The post-approval market also varies by disease. Breast cancer and hypertension have had established therapies since the 1950s, with several developments in recent years (Tables B7-B9). Although the first drug for HIV, Zidovudine, was approved in 1981 (AIDSinfo 2020), considerable progress in recent years has offered patients access to entirely new classes, such as Ibalizumab (Trogarzo), which was FDA-approved in 2018. Substantial variation also exists in the incremental health benefits of approving a new therapy. Driven by differences in both market size and potential life-saving benefits, hypertension has the greatest societal per-drug health benefit $Q_E$, while breast cancer medications have the least. Hypertension drugs retain longer patent protection on the market due to shorter trial durations. Historically, ineffective hypertension drugs also spend the shortest time on the market, likely because many more patients take these drugs, leading to faster awareness of potential ineffectiveness.

We compute the optimal approval policies $\alpha^*$ for each disease, as illustrated in Figure 4. Our model suggests that a stricter policy is warranted for HIV drugs ($\alpha^* = 1.2\%$), generating nearly a 1% gain in expected health benefits as fewer ineffective drugs would make it to the market under this policy. The optimal thresholds for both breast cancer ($\alpha^* = 3.8\%$) and hypertension ($\alpha^* = 4.5\%$) are less stringent than the FDA's existing $\alpha = 2.5\%$ policy, generating improvements in population-wide health benefits as additional therapies are approved. We explore additional drivers of these results in Section 5.2.

Modifying the approval thresholds translates into a 3-4% increase in the number of effective therapies for breast cancer and hypertension entering the market each year. Given our assumption $\frac{Q_I/\mu_I}{Q_E/\mu_E} \approx 2$ across diseases, the more lenient approval standard for breast cancer and hypertension is explained by market saturation (i.e., how full the market is in steady state under the current FDA policy): $\mathbb{E}[N_E(0.025)]/K$ equals 62% (breast cancer), 84% (HIV), and 57% (hypertension). Relaxing approval standards for breast cancer and hypertension increases the net availability of effective therapies in the market, generating more health benefits than type-I error costs (e.g., for breast cancer the health benefits increase by 303 years gained while the costs only by 230 years lost). For HIV, the higher market saturation implies that approving more drugs triggers increased

**Figure 4** **Expected net benefits under the optimal policy $\alpha^*$ and current FDA policy $\alpha = 2.5\%$.**




obsolescence of existing therapies, without necessarily increasing health benefits; hence the optimal policy favors more stringent standards to mitigate type-I errors. Indeed, the optimal policy for HIV results in a cost reduction of 15,406 years while the health benefits only decrease by 8,640 years. Note, the substantial difference in life years across diseases is largely driven by the size of each patient population.

To compare our results with Isakov et al. (2017) for our three focal diseases, we set trial enrollment $n$ and information $\delta\sqrt{I_n}$ to their values, but maintain all other parameters at our base values (Table 4). We apply the queueing network model and solve for our optimal approval policy $\alpha^*$. In general, our $\alpha^*$ is more lenient for breast cancer and hypertension, yet more stringent for HIV

**Table 4** **Model results comparison.**

| | Isakov et al. (2017) | | | | Current Study | |
|---|---|---|---|---|---|---|
| Disease | $\delta$ | $n$ | $\alpha$ | $1-\beta$ | $\delta\sqrt{I_n}$ | $\alpha^*$ |
| Breast Cancer | 0.125 | 951 | 0.013 | 0.69 | 2.726 | 0.063 |
| HIV | 0.125 | 830 | 0.027 | 0.73 | 2.546 | 0.024 |
| Hypertension | 0.125 | 806 | 0.032 | 0.74 | 2.509 | 0.092 |
| Breast Cancer | 0.25 | 342 | 0.016 | 0.87 | 3.269 | 0.037 |
| HIV | 0.25 | 295 | 0.033 | 0.89 | 3.036 | 0.015 |
| Hypertension | 0.25 | 283 | 0.039 | 0.89 | 2.974 | 0.059 |
| Breast Cancer | 0.5 | 92 | 0.017 | 0.90 | 3.391 | 0.032 |
| HIV | 0.5 | 77 | 0.034 | 0.90 | 3.102 | 0.014 |
| Hypertension | 0.5 | 73 | 0.041 | 0.90 | 3.021 | 0.056 |
| Breast Cancer | 1 | 23 | 0.017 | 0.90 | 3.391 | 0.032 |
| HIV | 1 | 19 | 0.036 | 0.90 | 3.082 | 0.014 |
| Hypertension | 1 | 18 | 0.043 | 0.90 | 3.000 | 0.057 |

*Note.* Results are from Isakov et al. (2017) Tables 3-6, where $\delta$ is treatment effect size, $n$ is trial enrollment, $\alpha$ is probability of type-I error, $1-\beta$ is probability of type-II error (power), $\delta\sqrt{I_n}$ is information of the test, and $\alpha^*$ is the optimal approval policy. Isakov et al. (2017) examine Ischemic Heart Disease, for which hypertension is a precursor.

than in Isakov et al. (2017). When the treatment effect size $\delta$ is large, the optimal trial enrollment results in a statistical test with 90% power, and the differences between our studies narrow.

Several factors contribute to these different optimal policies. First, Isakov et al. (2017) take a high-level perspective to compare approval policies based solely on disease prevalence and severity. They minimize the aggregate costs of a type-I error, which includes the harm accruing to patients taking an ineffective drug during the trial and after FDA-approval, and a type-II error, measured as the potential reduction in disease severity if the drug were approved. In contrast, our approach focuses on only three diseases and we aim to maximize the expected net health benefits assuming various therapies (effective and ineffective) co-exist in the market. Our model is fundamentally a "stock and flow" approach—accounting for the number of unique drug classes available, the current rates of R&D innovation and clinical trial completion, and existence of alternative therapies—whereas Isakov et al. (2017) perform an expected cost-minimization analysis without capturing such factors. In fact, the optimal approval policy is often quite sensitive to these latter disease-specific features, as illustrated in Section 5.2 below.

Second, their results largely depend on the costs associated with type-I and type-II errors. Because they assume the per patient cost of a type-I error is identical across diseases, their policies are predominantly driven by differences in the cost of type-II error, which are nearly impossible to measure directly. This is evident in that the optimal $\alpha$ found by Isakov et al. (2017) are monotonically increasing in disease severity (i.e., $\alpha < 2.5\%$ for low-severity diseases). They estimate the type-II error cost as an arbitrary reduction in disease severity. On the other hand, we include a type-I error cost specific to the disease, but assume it is proportional to health benefits. Our objective includes aggregate expected health benefits less harms and the magnitude of $Q_E$ is one of many factors affecting $\alpha^*$; as a result, our policies are not monotonic in $Q_E$ across diseases.

Finally, to account for differential drug quality, Isakov et al. (2017) assume uninformative priors $p = 0.5$ on the effectiveness probability, and that all drugs initiating clinical trials will undergo FDA review. Our approach is more nuanced, as we directly model the entire drug development process, specifically, trial initiation, completion, and abandonment, allowing us to account for both the volume of drug candidates in the pipeline and the likelihood that a drug advances through multiple trial phases. Under the simplified assumption that approved effective drugs are modeled as a $M/M/\infty$ queue, we can essentially recover the results obtained by Isakov et al. (2017) for a given trial size, as the optimal policy would only depend on drug benefits or harms and the effectiveness probability $p$.

### 5.2. Sensitivity Analysis

We examine how the optimal approval policy changes under different assumptions, focusing on two pre-approval parameters, the trial initiation rate $\lambda$ and average trial duration $1/\mu_{CT}$, and two

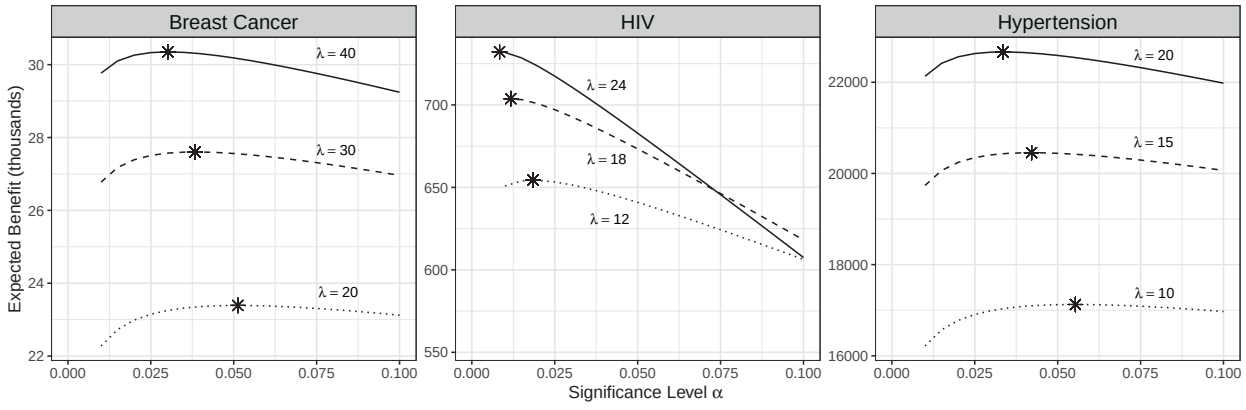
post-approval parameters, the effectiveness probability $p$ conditional on going through FDA review, and the cost-benefit ratio $\frac{Q_I/\mu_I}{Q_E/\mu_E}$ of approving an ineffective drug relative to an effective drug. For each parameter, we plot the expected net benefit as a function of $\alpha$ with the optimal significance level $\alpha^*$ highlighted, holding the axes ranges constant to facilitate comparison across parameters.

**Clinical Trial Initiation.** Modifying the clinical trial initiation rate $\lambda$ to values that are one-third higher or lower reflects the typical year to year variation in Phase I trials. As indicated by Corollary 1, the optimal approval policy $\alpha^*$ is decreasing in $\lambda$ (Figure 5). As more candidate drugs for a particular disease eventually go up for FDA review, the agency can afford to be more stringent, given the diminishing marginal returns of additional drugs treating the same underlying condition joining the market. With breast cancer, for example, if $\lambda = 40$, the optimal threshold shifts down to $\alpha^* = 3.0\%$. HIV and hypertension see similar relative reductions in $\alpha^*$ as $\lambda$ increases. Here, we assume that the $\lambda$ are exogenously determined by pharmaceutical firms. In the next sub-section, we examine the case where clinical trial investment is endogenous to the FDA's approval policy $\alpha$.

The marginal benefit of changing $\alpha$ differs by $\lambda$, most notably in the case of HIV. For high values of $\lambda$, the marginal benefits of increasing $\alpha$ drop precipitously, as the post-approval market becomes saturated with effective drugs. With HIV, if $\lambda = 24$ and $\alpha = 10\%$, then the expected number of approved effective drugs in different classes $\mathbb{E}[N_E(\alpha)] = 7.6$, nearly approaching $K = 8$. Thus, approving more candidate drugs only serves to "bump" existing effective drugs from the market. For lower values of $\lambda$, the effect is less pronounced because there is excess capacity in the market. With breast cancer or hypertension, expected benefits are less responsive to changes in $\alpha$ since $\mathbb{E}[N_E(\alpha)]$ is below $K$. One factor explaining why the optimal policy for these diseases relaxes the current 2.5% threshold is the long-run shortage of effective therapies on the market.

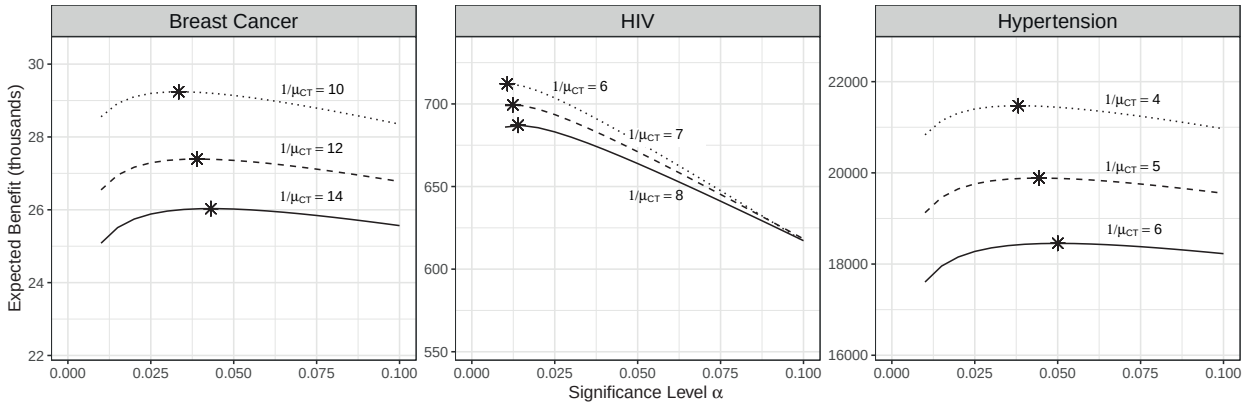
**Trial Duration and FDA Review.** In general, as the average clinical trial duration increases (i.e., $\mu_{CT}$ decreases), the optimal approval policy $\alpha^*$ increases, consistent with Corollary 1. Increasing the trial duration produces a similar effect to decreasing the Phase I initiation rate $\lambda$, given the similar effect on decreasing the net NDA submission rate $\lambda_{NDA}$.

**Figure 5** **Sensitivity of the expected net benefits and optimal policy $\alpha^*$ to the clinical trial initiation rate $\lambda$.**



Following the 1992 passage of the U.S. Prescription Drug User Fee Act (PDUFA), which established Priority Review programs to address the protracted timeline for drug development and approval (FDA 2014), the average time to complete FDA review decreased from 30 months to 15 months by the late 1990s, and further dropped to 10 months by 2011 (Darrow et al. 2014). In our model, reducing the average trial duration by one year—an effect similar to the initial change following PDUFA—generates nearly 3% more expected health benefits for breast cancer patients. This change would be equivalent to launching three additional Phase I breast cancer clinical trials every year, a 10% increase over current rates. For other diseases, shortening clinical trials by one year could increase benefits by more than 8% (hypertension) as many more patients would have access to newer therapies sooner, or only 1% (HIV) as the post-approval market is more crowded. Our model suggests that, to counter the higher NDA submission rate associated with expedited review, the FDA should simultaneously tighten the approval threshold by approximately 0.4%, to maximize expected health benefits.

**Figure 6** **Sensitivity of the expected net benefits and optimal policy $\alpha^*$ to the clinical trial duration $1/\mu_{CT}$.**

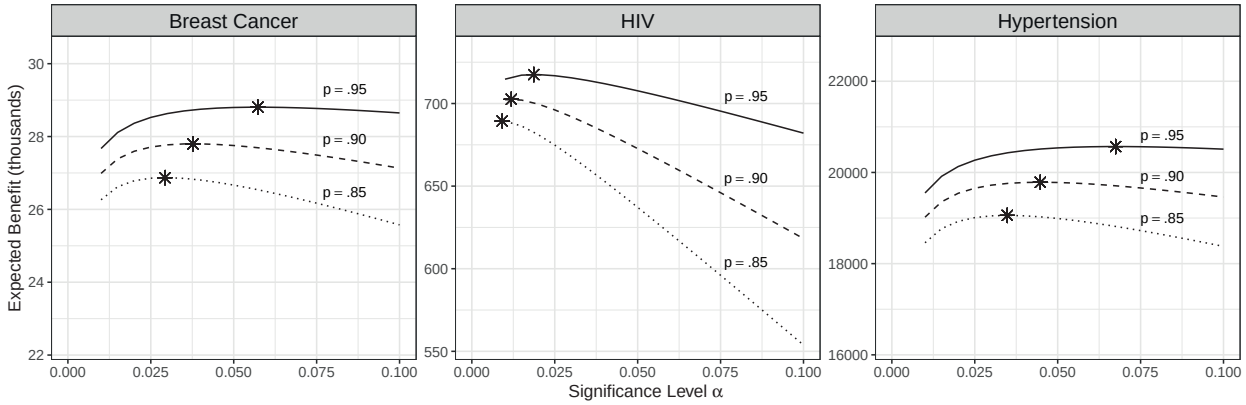



**Effectiveness Probability.** Of all the model parameters, the conditional probability that a drug is actually effective given that the sponsoring firm submits an application to the FDA, is the most challenging to estimate. We vary the probability $p \in \{0.85, 0.90, 0.95\}$, assuming that the FDA is unlikely to consider a drug candidate with a lower probability. Over this set of values, the optimal approval policy $\alpha^*$ is increasing in $p$ (Figure 7), even though the condition $\lambda_{AE}(\alpha^*)/K\mu_E < 1$ does not hold (Corollary 1).

For breast cancer, the optimal policy moves from 2.9% (if $p = 0.85$) to 5.7% (if $p = 0.95$) and for hypertension, the optimal policy moves from 3.5% (if $p = 0.85$) to 6.7% (if $p = 0.95$). For HIV, the optimal policy is less sensitive to changes in $p$, moving from 0.9% (if $p = 0.85$) to 1.9% (if $p = 0.95$). However, the expected benefits change more quickly for lower values of $p$. For $p \leq 0.85$, more ineffective HIV drugs make it to the market—and stay there for long periods of time given

the low $\mu_E$—but there is little gain to the approved effective market because $\lambda_{NDA}$ is high. In other words, given the abundance of HIV drugs in the development pipeline, the potential downside of committing a type-I error outweighs the upside of approving additional effective drugs.

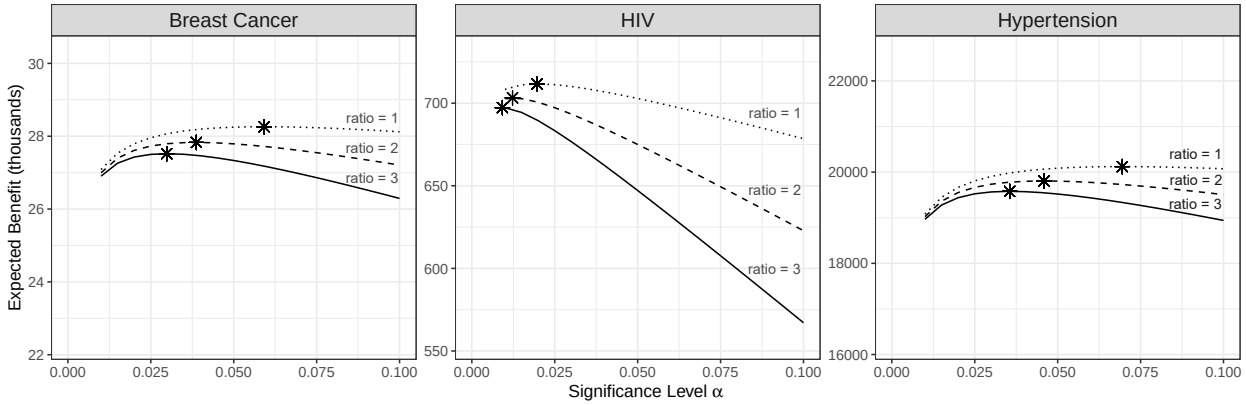
**Figure 7** **Sensitivity of the expected net benefits and optimal policy $\alpha^*$ to the effectiveness probability $p$.**




**Post-Approval Cost-Benefit Ratio.** In our base case analysis, we assume that the lifetime cost-benefit ratio of approving an ineffective drug to an effective drug, $\frac{Q_I/\mu_I}{Q_E/\mu_E}$, is approximately 2 (see Appendix B for details). This takes into account potential life years gained ($Q_E$) or lost ($Q_I$) per additional approved drug, as well as the average time the drug spends on the market before eventual exit ($1/\mu_E$) or withdrawal ($1/\mu_I$).

With a higher cost-benefit ratio, the expected benefits decrease and the optimal policy is to approve fewer drugs ($\alpha^*$ decreases), as anticipated (Figure 8). Assuming a cost-benefit ratio of 3, the optimal approval policies decrease from 3.8% to 3.0% (breast cancer), from 1.2% to 0.9% (HIV), and from 4.5% to 3.5% (hypertension). If the FDA instead assigns equal weight to approving effective and ineffective drugs (i.e., ratio = 1), then the optimal approval threshold is significantly relaxed: 5.9% (breast cancer), 2.0% (HIV), 6.9% (hypertension).

**Figure 8** **Sensitivity of the expected net benefits and optimal policy $\alpha^*$ to the cost-benefit ratio $\frac{Q_I/\mu_I}{Q_E/\mu_E}$.**

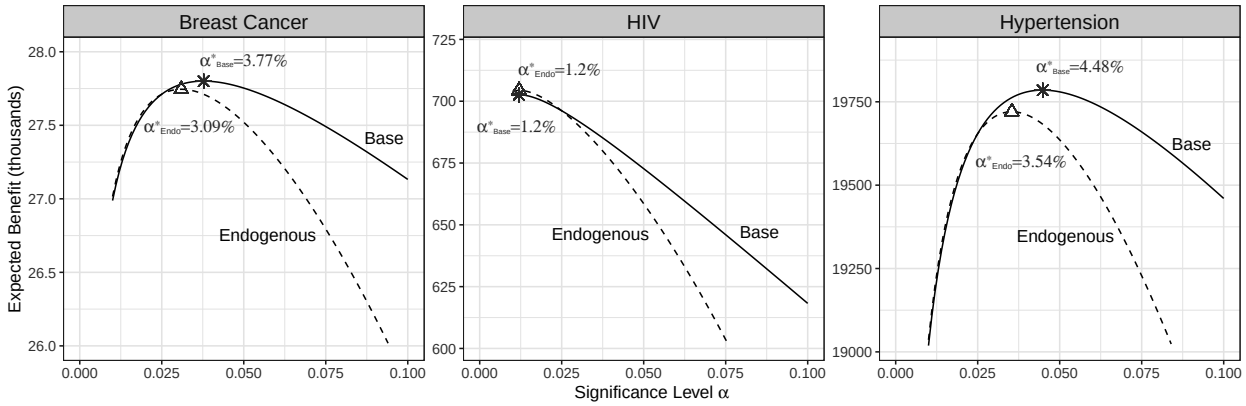


### 5.3. Endogenous Queueing Results

We next explore how an endogenous clinical trial initiation rate $\tilde{\lambda}(\alpha)$ might alter the optimal approval policy and expected benefits. Across the three diseases considered, Figure 9 illustrates how an endogenous rate modestly reduces the optimal approval threshold, to 3.1% (breast cancer), and 3.5% (hypertension), but minimally changes for HIV, results that are consistent with Proposition 2. As the approval threshold is increasingly relaxed (i.e., higher $\alpha$), the net benefits in the endogenous case and base case diverge rapidly. This occurs because the endogenous objective function $\tilde{V}(\alpha)$ includes essentially a "penalty term" $\frac{Q_I}{\mu_I}\frac{\lambda_{NDA}}{1+\alpha_o}(\alpha_o - \alpha)$, accounting for the expected health costs if additional ineffective drugs enter the market under more lenient FDA approval standards. To compensate for such costs, the endogenous model sets a lower approval policy $\tilde{\alpha}^*$.

Recent experience suggests that firms may in fact increase R&D investment in particular diseases following the introduction of new approval schemes. After the FDA launched the Breakthrough Therapy designation in 2012, the initiation rate of Phase I clinical trial for breast cancer increased by 20% (National Institutes of Health 2020). Between 2012 and 2020, five breast cancer drugs (Verzenio®, Kisqali®, Kadcyla®, Ibrance®, and Enhertu®) received FDA-approval with Breakthrough Therapy status. Although it is premature to estimate whether these successes will spur firms to further invest in other novel cancer therapies, our model highlights that firm responses to incentive programs—such as establishing Priority Review or modifying $\alpha^*$—could result in different optimal policies, warranting further investigation in firm behavior.

**Figure 9 Expected net benefits and optimal policy $\alpha^*$ under the endogenous queueing model.**




### 5.4. Robustness Checks

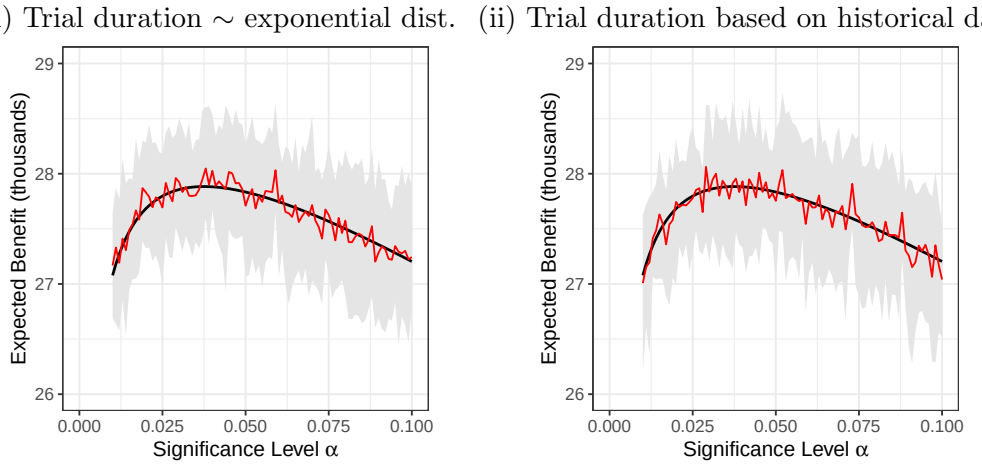
To further test our model's robustness, we simulate our model's underlying queueing network, focusing on breast cancer drugs. We relax several key assumptions and compute the expected net benefit for significance levels ranging from $\alpha = 1\%$ to $\alpha = 10\%$. We run 30 iterations and consider

a sufficiently long time horizon to achieve steady-state behavior. We burn the transient period and compute the expected net benefit using steady-state data.

**Clinical Trial Duration.** In our base model, clinical trials are modeled as a single phase with an exponential race between abandonment and service completion. We relax this assumption by considering three separate clinical trial phases and either (i) model each phase as an exponential race with specific completion and abandonment rates, or (ii) sample each phase duration using historical trial duration data from `clinicaltrials.gov`. In both scenarios, the probability of each phase successful completion is based on oncology drugs for solid tumors given in Thomas et al. (2016) and summarized in Table B2. In scenario (i), we compute the trial completion rate based on the average trial duration by phase as reported in Table B3, and the phase-specific abandonment rate to match Thomas et al. (2016); thus, the rate of drugs entering NDA review is unchanged.

Figure 10 shows that the objective function, and thus the optimal approval policy, in scenarios (i) and (ii) are virtually identical to the base model, suggesting that our earlier analysis is robust to structural and distributional assumptions in the pre-review queueing model.

**Figure 10** **Expected net benefit from simulation (red line) and base model (black line) if clinical trials (breast cancer only) are modeled as three separate phases.**



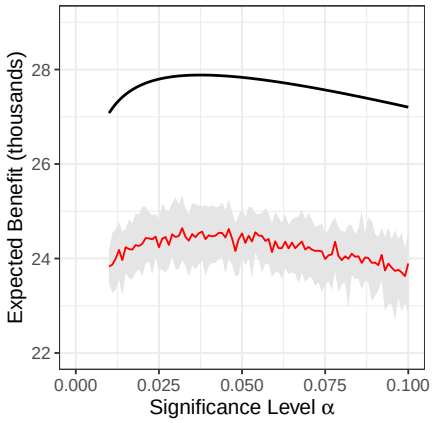

*Note.* The grey band corresponds to the 90th and 10th percentiles of the simulated objective (30 replications).

**Drug Class Distribution.** In our base model, effective drugs that gain FDA approval belong to any of $K$ drug classes, with equal probability. We relax this assumption by setting the probability distribution across drug classes using historical data on breast cancer drug approvals (Table B7). Three of the twelve approved drug classes for breast cancer account for nearly 60% of all approvals: combination chemotherapy (27%), targeted biological therapy (15.3%), and other chemotherapy (15.3%). This imbalance means that drugs in these three classes are more likely to become obsolete—as newer therapies gain FDA approval—and thus removed from the market

prematurely, before patent expiry. This observation has borne out in reality, as demonstrated by the recent surge in FDA approvals of targeted biologics for breast cancer, with six new therapies (Ibrance®, Tykerb®, Kisqali®, Nerlynx®, Verzenio®, and Lynparza®) approved since 2015.

Consequently, the remaining nine classes absorb fewer new drugs, decreasing the expected total number of approved effective drugs in the market due to eventual drug obsolescence. Given this market concentration, overall expected net benefits are lower than in our base case, yet the general shape of the objective function is largely unchanged, suggesting that the optimal policy $\alpha^*$ only modestly changes (Figure 11). Of course, in reality, different drug classes may confer different health benefits. Estimating class-specific benefits, however, is difficult because of the natural delay between the availability of first-in-class drugs to the realization of their health benefits. Moreover, disentangling the overall benefits per drug-class requires, for example, more granular clinical trial data. We predict that assuming drug class-specific health benefits will likely result in significant changes in the optimal approval policy, but this is beyond the scope of our analysis.

**Figure 11** **Expected net benefit from simulation (red line) and base model (black line) with post-market drug class distribution based on historical data (breast cancer only).**




*Note.* The grey band corresponds to the 90th and 10th percentiles of the simulated objective (30 replications).

**Time on Market.** Lastly, we relax the $M/M/1/1$ queueing assumption that the time that approved effective drugs spend on the market is exponentially distributed. Here, we assume a pre-emptive $M/G/1/1$ queue with log-normally distributed time on the market with the same mean and standard deviation as before. The longer tail of the log-normal probability distribution results in a slight upward shift of the expected health benefits as longer times on the market are more likely to realize. However, the optimal policy derived from the base case is fairly robust to the exponential assumption.

## 6. Discussion

Our queueing framework presents a novel scheme for analyzing a disease-specific FDA-approval policy, accounting for both the pre-review drug pipeline and post-approval market characteristics. Our model considers three drivers of the shortfall of therapies available to treat some diseases: (i) a lack of innovation in new drug formulation (i.e., a low *arrival rate*), (ii) lengthy clinical trials (i.e., a low *service rate*), and (iii) frequent attrition during development (i.e., a high *abandonment rate*). Together with post-approval market considerations, including the target patient population size, number of unique alternative therapies available, and expected time until drug obsolescence or patent expiry, we propose an alternative approach for the FDA: adjust the statistical significance threshold for approval by disease.

Through launching several expedited programs over the years, the FDA has existing incentives designed to spur R&D investment. The *Orphan Drug Designation*, for instance, increases research funding for rare diseases by providing monetary incentives (e.g., tax credits) to firms developing therapies for these conditions. An alternative mechanism to mitigate low R&D investment is to ease approval standards for diseases with few drugs in the early stages of development. Our framework could augment existing expedited programs, which are limited in scope, and help derive guidelines on setting disease-specific approval standards and estimating their associated health benefits and risks. A flexible approval policy offers a fundamentally different way of addressing imbalances in drug availability and serves two purposes: in the short-run, more drugs will be permitted to enter the market, and in the long-run, this may encourage pharmaceutical firms to divert resources from diseases with many competitor drugs to therapies more likely to gain approval.

Our work relates to two studies that compute the significance level $\alpha$ and trial enrollment $n$ to minimize expected patient harms via type-I and type-II errors. Montazerhodjat et al. (2017) examine drugs targeting 23 different late-stage cancers. Their optimal level for breast cancer is 17.6%—seven times higher than the traditional 2.5%—whereas our model recommends a significance level of 3.8% for this disease. These authors assume that ineffective drugs shorten a patient's life by 2 months—across *all* cancer types—despite average life expectancy ranging from 16 months (pancreatic cancer) to more than 15 years (Hodgkin lymphoma). They assume, on the other hand, that effective drugs extend life expectancy by 30% up to a maximal gain of 2.5 years, a relatively modest increase for cancers with longer survival times (Montazerhodjat et al. (2017), Table 1). The study does not include references nor background data for these assumptions, and it is perhaps unsurprising that the authors find optimal approval policies in excess of 20-30% for cancers with the worst prognoses, and a threshold of $\alpha = 48\%$ for brain cancer—essentially a coin toss. In a related paper, Isakov et al. (2017) apply a similar model to 30 high-burden diseases, and find policies more closely aligned with our results for our three focal diseases (Table 4). This study, by

the same group of authors, computes a substantially different approval threshold for breast cancer (range 1.3-1.7%), potentially due to different assumptions about the target population and costs.

Our approach differs from these two studies in several ways. First, both of these studies assign one-time costs to type-I and type-II errors, but do not consider the *health benefits* of approving drugs. In contrast, our model aggregates both the benefits of approving effective drugs and the harms of approving ineffective drugs, over the period that each type of drug spends on the market, which can differ by disease. We exclude the cost of a type-II error given the challenges in estimating this outcome, as Montazerhodjat et al. (2017) duly note: *"Type 2 harm is rarely discussed in medical and lay communities because it is difficult to quantify the number of missed opportunities..."* With trial enrollment conventionally decided by the sponsoring pharmaceutical company and $\alpha$ set by the regulatory agency, both Montazerhodjat et al. (2017) and Isakov et al. (2017), in essence, assume the role of a social planner seeking to minimize expected harms by adjusting these two levers. In contrast, we take a more limited perspective of the FDA as the sole decision-maker who only sets the approval policy $\alpha$. Finally, our queueing framework models the full R&D pipeline of the target disease (clinical trial initiation, attrition, and completion rates), as well as post-approval market dynamics (substitution between drugs within a therapeutic class and obsolescence of older therapies). Isakov et al. (2017) instead focus on a single candidate drug, without explicitly accounting for how many other related drugs are in development, approaching FDA review, or approved and on the market.

### 6.1. Limitations

Implementing a queueing model, such as the one proposed here, poses additional challenges, particularly with explaining and validating some structural assumptions (e.g., all queues are in steady state, the number of drug classes is fixed, and clinical trial attrition rates are equal across therapeutic classes within a disease). Employing more complex queueing methodology may offer more realistic assumptions, but would likely sacrifice the analytical insights gained from a more parsimonious model. A simple model is likely more interpretable to decision-makers than a complex analysis that obscures the rationale behind the optimal approval policy. Simulation results, however, suggest that our queueing model is relatively robust to several structural assumptions.

The goal of our study is to highlight how heterogeneous aspects of drug development—the target population size, alternative therapies available, speed in launching and completing clinical trial testing, attrition rates, and drug obsolescence—ought to be objectively considered by the FDA, to maximize expected net health benefits. We do not intend for the model to act as a standalone decision-support tool, nor to replace experts at the FDA who must weigh clinical trial evidence and make difficult decisions, often with life-or-death consequences. Equally important, but beyond

the scope of this paper, are questions of *fairness* (should some patient groups receive priority over others?), *equity* (should all diseases receive some minimal/maximal level of R&D investment?), and *cost-effectiveness* (should the FDA play a role in setting prices during the approval process?). This complex, high-stakes setting is ripe for future research.

We make several simplifying assumptions regarding FDA decision-making. Qualitative aspects, such as concerns over trial design or manufacturing capacity, are ignored. We do not consider that the FDA may ask a firm to revise and resubmit an NDA, which occurs in 30% of reviews (Downing et al. 2014). We assume that NDA filing and FDA review occur immediately; in reality, these reviews last an average of six to ten months. Finally, we make several assumptions when computing expected net benefits.

Drug efficacy is based on a single quantitative endpoint arising from a balanced, two-arm randomized clinical trial. Modern trial designs are often unbalanced, have more than two arms, and involve multiple endpoints. Our model could be easily adapted for unbalanced trials, but incorporating multiple arms and endpoints would require a more sophisticated hypothesis testing framework and queueing model (e.g., incorporating Bonferroni adjustment of the type-I error for multiple endpoints). Breast cancer trials, for example, often measure tumor size and time until recurrence, and establishing drug efficacy from these multiple endpoints requires multi-criteria decision-making (FDA 2017a). Such disease-specific complexity could render our model analytically intractable.

### 6.2. Future Work

Our study motivates several directions for future work. Currently, we assume that drugs are either effective or ineffective, and confer the same health benefits across patients. Possible extensions include modeling drug effectiveness as a continuous variable, modeling the post-approval health effects as random variables, or adjusting the approval threshold based on variability in treatment responses among trial participants. Tailoring drug development to better account for heterogeneity in treatment responses within a clinical trial is an burgeoning research topic, especially as personalized medicine is becoming more ubiquitous.

Another extension is to analyze the drug development process using a game theoretic model with two players: the FDA selects the significance level $\alpha$, and a pharmaceutical company selects, for instance, the clinical trial size $n$. Conditions under which a firm should not invest in clinical trials to assess drug efficacy (i.e., $n = 0$), or when they should conduct additional trials after an NDA rejection, could be explored. A multi-firm model with competitive (e.g., innovation races) or cooperative (e.g., clinical trial cost-sharing and joint marketing) players could also be developed.

We focus our analysis on FDA drug approval, but our framework could readily apply to other settings. Drugs developed in the U.S. and Europe both undergo clinical trial testing, but the review

and approval processes differ substantially. All drugs in the U.S. undergo centralized review by the FDA, whereas in Europe, there are four possible paths to approval: a centralized process overseen by the European Medicines Agency, application to the regulatory body of a single European Union (EU) state, application for approval in all EU states following approval in one state, and independent application in multiple EU states (Van Norman 2016). Our queueing framework could analyze the trade-offs of different approval pathways and to compare the European and U.S. systems.

### 6.3. Conclusions

Faced with regulating thousands of drugs in a nation where millions are newly afflicted with severe diseases each year, the FDA must find the correct balance between ensuring the safety and effectiveness of drugs while spurring development of novel therapeutics and bringing life-saving products to market in a timely manner. Our study offers a transparent, quantitative framework that can provide the FDA with insights regarding how disease severity, prevalence, and other characteristics of the drug development process and existing market could change approval standards. Such a model could augment the complex decision-making and statistical analyses conducted by the FDA, providing a more customized approach to policy-making.

## References

Adida E (2020) Outcome-Based Pricing for New Pharmaceuticals via Rebates. *Management Science* (forthcoming).

Adida E, Bravo F (2019) Contracts for Healthcare Referral Services: Coordination via Outcome-Based Penalty Contracts. *Management Science* 65(3):1322–1341.

Ahmed P, Gardella J, Nanda S (2002) Wealth Effect of Drug Withdrawals on Firms and Their Competitors. *Financial Management* 21–41.

Ahuja V, Birge JR (2016) Response-Adaptive Designs for Clinical Trials: Simultaneous Learning from Multiple Patients. *European Journal of Operational Research* 248(2):619–633.

AIDSinfo (2020) FDA-Approved HIV Medicines. URL https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/21/58/fda-approved-hiv-medicines, Accessed Aug 2020.

Anderer A, Bastani H, Silberholz J (2019) Adaptive Clinical Trial Designs with Surrogates: When Should We Bother? *Available at SSRN 3397464* .

Armony M, Israelit S, Mandelbaum A, Marmor YN, Tseytlin Y, Yom-Tov GB (2015) On Patient Flow in Hospitals: A Data-Based Queueing-Science Perspective. *Stochastic systems* 5(1):146–194.

Arrowsmith J, Miller P (2013) Trial Watch: Phase II and Phase III Attrition Rates 2011 to 2012. *Nature Reviews Drug Discovery* 12(8):569–569.

Aswani A, Shen ZJM, Siddiq A (2019) Data-Driven Incentive Design in the Medicare Shared Savings Program. *Operations Research* 67(4):1002–1026.

Bertsimas D, Johnson M, Kallus N (2015) The Power of Optimization Over Randomization in Designing Experiments Involving Small Samples. *Operations Research* 63(4):868–876.

Bertsimas D, OHair A, Relyea S, Silberholz J (2016) An Analytics Approach to Designing Combination Chemotherapy Regimens for Cancer. *Management Science* 62(5):1511–1531.

Breast Cancer Society (2020) U.S. Breast Cancer Statistics. URL https://www.breastcancer.org/symptoms/understand_bc/statistics, Accessed Aug 2020.

Carpenter DP (2004) The Political Economy of FDA Drug Review: Processing, Politics, and Lessons for Policy. *Health Affairs* 23(1):52–63.

Casella G, Berger RL (2002) *Statistical Inference* (Thomson Learning), Second edition.

Centers for Disease Control and Prevention (CDC) (2019) Heart Disease Facts. URL https://www.cdc.gov/heartdisease/facts.htm, Accessed Aug 2020.

Centers for Disease Control and Prevention (CDC) (2020) HIV in the United States: At a Glance. URL https://www.cdc.gov/hiv/statistics/overview/ataglance.html, Accessed Aug 2020.

Chick SE, Gans N, Yapar O (2018) Bayesian Sequential Learning for Clinical Trials of Multiple Correlated Medical Interventions. *Available at SSRN 3184758* .

Dai T, Cho SH, Zhang F (2016) Contracting for On-Time Delivery in the US Influenza Vaccine Supply Chain. *Manufacturing & Service Operations Management* 18(3):332–346.

Dai T, Tayur S (2019) Healthcare Operations Management: A Snapshot of Emerging Research. *Manufacturing & Service Operations Management* (forthcoming).

Darrow JJ, Avorn J, Kesselheim AS (2014) New FDA Breakthrough-Drug Category—Implications for Patients. *New England Journal of Medicine* 370(13):1252–1258.

Deo S, Corbett CJ (2009) Cournot Competition Under Yield Uncertainty: The Case of the US Influenza Vaccine Market. *Manufacturing & Service Operations Management* 11(4):563–576.

Downing NS, Aminawung JA, Shah ND, Krumholz HM, Ross JS (2014) Clinical Trial Evidence Supporting FDA Approval of Novel Therapeutic Agents, 2005-2012. *JAMA* 311(4):368.

Express Scripts Holding Company (2017) 2016 Drug Trend Report.

FDA (2014) Guidance for Industry Expedited Programs for Serious Conditions: Drugs and Biologics. URL https://www.fda.gov/media/86377/download, Accessed Aug 2020.

FDA (2017a) Multiple Endpoints in Clinical Trials Guidance for Industry. URL https://www.fda.gov/media/102657/download, Accessed Aug 2020.

FDA (2017b) The FDA's Drug Review Process: Ensuring Drugs are Safe and Effective. URL http://www.fda.gov/Drugs/ResourcesForYou/Consumers/ucm143534.htm, Accessed Aug 2020.

FDA (2018) Benefit-Risk Assessment in Drug Regulatory Decision-Making. URL https://www.fda.gov/media/112570/download, Accessed Aug 2020.

FDA (2019) Code of Federal Regulations. URL https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?, Accessed Aug 2020.

FDA (2020a) Drug Recalls. URL https://www.fda.gov/drugs/drug-safety-and-availability/drug-recalls, Accessed Aug 2020.

FDA (2020b) Drug Safety and Availability. URL https://www.fda.gov/drugs/drug-safety-and-availability, Accessed Aug 2020.

FDA (2020c) Drugs@FDA. URL https://www.accessdata.fda.gov/scripts/cder/daf/, Accessed Aug 2020.

FDA (2020d) New Drug Therapy Approvals 2019. URL https://www.fda.gov/media/134493/download, Accessed Aug 2020.

Gupta V, Han BR, Kim SH, Paek H (2018) Maximizing Intervention Effectiveness, Working Paper.

Hay M, Thomas DW, Craighead JL, Economides C, Rosenthal J (2014) Clinical Development Success Rates for Investigational Drugs. *Nature Biotechnology* 32(1):40–51.

Isakov L, Montazerhodjat V, Lo A (2017) Is the FDA Too Conservative or Too Aggressive?: A Bayesian Decision Analysis of Clinical Trial Design. *Working Paper* .

Jennison C, Turnbull W (2000) Two-Sided Tests: Introduction. *Group Sequential Methods with Applications to Clinical Trials* (Chapman and Hall/CRC).

Jiang H, Pang Z, Savin S (2012) Performance-Based Contracts for Outpatient Medical Services. *Manufacturing & Service Operations Management* 14(4):654–669.

Keskinocak P, Savva N (2020) A Review of the Healthcare-Management (Modeling) Literature Published in Manufacturing & Service Operations Management. *Manufacturing & Service Operations Management* 22(1):59–72.

Kinch MS, Haynesworth A, Kinch SL, Hoyer D (2014) An Overview of FDA-Approved New Molecular Entities: 1827-2013. *Drug Discovery Today* 19(8):1033–1039.

Kouvelis P, Milner J, Tian Z (2017) Clinical Trials for New Drug Development: Optimal Investment and Application. *Manufacturing & Service Operations Management* 19(3):437–452.

Kremer M, Levin J, Snyder CM (2020) Advance Market Commitments: Insights from Theory and Experience. *AEA Papers and Proceedings* 110:269–73.

Mamani H, Chick SE, Simchi-Levi D (2013) A Game-Theoretic Model of International Influenza Vaccination Coordination. *Management Science* 59(7):1650–1670.

Montazerhodjat V, Chaudhuri SE, Sargent DJ, Lo AW (2017) Use of Bayesian Decision Analysis to Minimize Harm in Patient-Centered Randomized Clinical Trials in Oncology. *JAMA Oncology* 3(9).

National Cancer Institute (NCI) (2020) Drugs Approved for Breast Cancer. URL https://www.cancer.gov/about-cancer/treatment/drugs/breast, Accessed Aug 2020.

National Institutes of Health (2020) ClinicalTrials.gov. URL https://clinicaltrials.gov/ct2/home, Accessed Aug 2020.

PhRMA (2015) Biopharmaceutical Research and Development: The Process Behind New Medicines.

PhRMA (2016) 2016 Profile: Biopharmaceutical Research Industry.

Pollack A (2011) FDA Revokes Approval of Avastin for Use as Breast Cancer Drug. *The New York Times* .

So KC, Tang CS (2000) Modeling the Impact of an Outcome-Oriented Reimbursement Policy on Clinic, Patients, and Pharmaceutical Firms. *Management Science* 46(7):875–892.

Sterne JAC, Smith GD (2001) Sifting the Evidence - What's Wrong with Significance Tests? *British Medical Journal* 322(8):226–231.

Thomas DW, Burns J, Audette J, Carroll A, Dow-Hygelund C, Hay M (2016) Clinical Development Success Rates 2006-2015. Technical report, Biotechnology Innovation Organization, Biomedtracker, Amplion.

Van Norman GA (2016) Drugs and Devices: Comparison of European and U.S. Approval Processes. *JACC: Basic to Translational Science* 1(5):399–412.

Wong CH, Siah KW, Lo AW (2019) Estimation of Clinical Trial Success Rates and Related Parameters. *Biostatistics* 20(2):273–286.

Xu L, Zhao H, Petruzzi NC (2020) Inducing Compliance with Postmarket Studies for Drugs Under FDAs Accelerated Approval Pathway. *Manufacturing & Service Operations Management* (forthcoming).